

ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
MESTRADO EM CIÊNCIA DA COMPUTAÇÃO

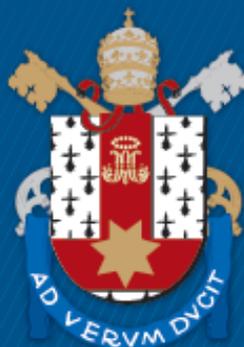
ÂNDERSON PINTO ALVES

**UM MODELO DE ANÁLISE VISUAL DE DADOS DE ENERGIA PARA EDIFÍCIOS E
CIDADES INTELIGENTES**

Porto Alegre

2020

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica
do Rio Grande do Sul

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**UM MODELO DE ANÁLISE
VISUAL DE DADOS DE
ENERGIA PARA EDIFÍCIOS E
CIDADES INTELIGENTES**

ÂNDERSON PINTO ALVES

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Isabel H. Manssour

**Porto Alegre
2020**

DEDICATÓRIA

Dedico este trabalho aos meus pais e aos meus amigos.

*“So understand
Don’t waste your time always searching for
those wasted years
Face up, make your stand
And realize you’re living in the golden years”
(Adrian Smith - Iron Maiden)*

AGRADECIMENTOS

Agradeço primeiramente a Deus por proporcionar um momento ímpar na minha vida. Somente ele sabe o tamanho do sonho que estou realizando.

Agradeço ao meu pai Hildo Severo Alves e minha mãe Melânia Terezinha Pinto Alves pelo bem maior da vida. Fico extremamente feliz por orgulhá-los por ser a primeira pessoa da família a atingir este título.

Aos meus filhos caninos Gigante e Mel que os amo tanto. Sair e chegar em casa e ver os rabinhos deles balançando de felicidade ao me encontrar é algo indescritível.

À Hewlett-Packard Brasil por fornecer a bolsa de estudo para este curso de mestrado e a Pontifícia Universidade Católica do Rio Grande do Sul por todos os recursos físicos, humanos e tecnológicos que disponibiliza a seus pesquisadores.

Ao Programa de Pós-Graduação em Ciência da Computação e a todos os professores dos quais pude conviver durante esse período, que me ensinaram muito mais do que apenas aquilo que o ofício requer, me instruíram a pensar e refletir, pois o mundo é muito maior que imaginamos.

Aos professores e colegas de mestrado por sempre estarem disponíveis para ajudar. Uma menção especial aos colegas do grupo de pesquisa DaVInt pelas incansáveis ajudas e apoios psicológicos.

A todos aqueles que de alguma forma contribuíram ou torceram pela concretização desta pesquisa e desta trajetória. Agradeço também aos participantes desse estudo que dedicaram seu tempo para validar minha pesquisa. Seus feedbacks foram importantes para o desenvolvimento deste estudo.

Por fim, mas não menos importante, a minha ilustre orientadora Prof^a. Isabel H. Manssour, que me recebeu na PUCRS, que me orientou com muita dedicação, confiança, conhecimento e atenção na elaboração desta pesquisa. Agradeço por acreditar em mim e por permitir a oportunidade de participar de deste grupo fantástico de pesquisadores. Ver meu nome na lista de aprovados no processo seletivo deste curso foi um dos dias mais felizes da minha vida. Muito obrigado.

UM MODELO DE ANÁLISE VISUAL DE DADOS DE ENERGIA PARA EDIFÍCIOS E CIDADES INTELIGENTES

RESUMO

Devido à recente evolução tecnológica, novos sensores e dispositivos estão sendo incorporados em edifícios e cidades inteligentes, para facilitar o entendimento da sua dinâmica e melhorar seu gerenciamento, bem como sua relação custo-benefício. Neste contexto, a combinação de tecnologia da informação e dispositivos capazes de capturar e compartilhar informações com outros dispositivos pode ajudar a coletar e compreender dados de energia. Esta é uma tarefa importante para avaliar a eficiência energética, ajudando a resolver problemas relacionados à energia. Porém, pode-se tornar um desafio analisar grandes volumes de dados que são coletados e armazenados ininterruptamente para confirmar tendências, identificar padrões ocultos e valores discrepantes que ajudem na tomada de decisões. O uso de representações gráficas pode auxiliar neste processo, mas a análise visual de grandes volumes de dados de energia pode não ser uma tarefa simples de ser executada, pois muitas das ferramentas de visualização existentes não foram planejadas com esta finalidade, dificultando tanto uma análise interativa com diferentes níveis de granularidade ao longo do tempo, como a comparação entre dados meteorológicos com diferentes dados de energia. Assim, o objetivo deste trabalho é apresentar um modelo para análise visual de dados de consumo ou geração de energia para cidades e edifícios inteligentes. Este modelo permite carregar, analisar e comparar dados de energia e dados meteorológicos ao longo do tempo para, por exemplo, identificar padrões de consumo com diferentes condições climáticas e valores discrepantes. Além disso, oferece diversas formas de explorar e compreender padrões entre diferentes conjuntos de dados, incorpora quatro algoritmos para realizar análises preditivas e permite avaliar os dados com diferentes níveis de granularidades de tempo, através de uma abordagem interativa baseada na técnica de detalhes sob demanda integrada com *coordinated multiple views*. Uma avaliação com especialistas de domínio demonstra a viabilidade, além das vantagens de usar esse modelo para explorar, monitorar e comparar dados de energia.

Palavras-Chave: Análise Visual, Consumo Energético, Geração de energia, Cidades Inteligentes, Edifícios Inteligentes.

ABSTRACT

Due to the recent technological evolution, new sensors and devices are being incorporated in buildings and smart cities, to facilitate the understanding of its dynamics and improve its management, as well as its cost-benefit ratio. In this context, the combination of information technology and devices capable of capturing and sharing information with other devices can help to collect and understand energy data. This is an important task to assess energy efficiency, helping to solve energy-related problems. However, it can become a challenge to analyze large volumes of data that are collected and stored continuously to confirm trends, identify hidden patterns and outliers that help in decision making. The use of graphical representations can assist in this process, but the visual analysis of large volumes of energy data may not be a simple task to be performed, as many of the existing visualization tools were not designed for this purpose, making it so difficult for an interactive analysis with different levels of granularity over time, such as comparing meteorological data with different energy data. Thus, the objective of this work is to present a model for visual analysis of consumption data or energy generation for cities and smart buildings. This model allows to load, analyze and compare energy data and meteorological data over time to, for example, identify consumption patterns with different climatic conditions and outliers. In addition, it offers several ways to explore and understand patterns between different sets of data, incorporates four algorithms to perform predictive analysis and allows to evaluate data with different levels of time granularities, through an interactive approach based on the integrated on-demand detail technique com coordinated multiple views. An evaluation with domain experts demonstrates the feasibility, in addition to the advantages of using this model to explore, monitor and compare energy data.

Keywords: Visual Analysis, Energy Consumption, Energy Generation, Smart Cities, Smart Buildings.

LISTA DE FIGURAS

Figura 2.1 – Visualização de gráfico de barras ao longo do tempo. Barras ascendentes representam condições mais quentes, enquanto barras descendente representam verões mais frios de um conjunto de dados de dados climáticos. [3].	30
Figura 2.2 – Exibição de dados de consumo de energia em gráficos de visualização de dados de séries temporais, que são: A) Linhas; B) Polar; C) Plot; D) Barras; E) Área. No eixo horizontal são mostradas os dias da semana e no eixo vertical as faixas de consumo.	31
Figura 2.3 – O processo do <i>Visual Analytics</i> é caracterizado pela interação entre dados, visualizações, modelos sobre os dados e os usuários, a fim de fomentar a descoberta do conhecimento [64].	32
Figura 2.4 – O escopo do <i>Visual Analytics</i> [60].	33
Figura 2.5 – <i>Calendar View</i> representando a média da demanda de energia agrupada por dias de um ano [44].	33
Figura 2.6 – Visualização de padrões recursivos dos preços diários de ações de 100 ações em 20 anos [57].	34
Figura 2.7 – <i>Heatmap</i> exibindo dados de genes, evidenciando a diferença do grupo de genes de pacientes sem câncer no lado esquerdo e pacientes com câncer no lado direito [41].	35
Figura 2.8 – Os principais componentes de um <i>boxplot</i> : mediana, quartis, limite inferior e limite superior [40].	36
Figura 2.9 – Exemplos de <i>Dashboards</i> : Os <i>Dashboards</i> 1 e 5 ilustram painéis direcionados à tomada de decisões, enquanto os <i>Dashboards</i> 3 e 4 dividem em blocos as informações para facilitar a análise dos dados. O <i>Dashboard</i> 2 tem como alvo o cenário autônomo quantificado, embora <i>Dashboard</i> 6 represente painéis projetados para comunicação de propósito geral. O <i>Dashboard</i> 7 captura novas extensões para painéis tradicionais [98].	38
Figura 2.10 – Exibição do uso do modelo de interação <i>Brushing and Linking</i> . Ao fazer uma seleção em qualquer um dos gráficos, os demais gráficos exibirão dados referente a seleção aplicada [115].	39
Figura 2.11 – <i>Coordinated Multiple Views</i> exibindo 8 visualizações distintas e integradas [73].	40
Figura 2.12 – Exemplo de uso do <i>Details on demand</i> , onde foi refinado o campo visual para obter maiores detalhes dos dados das subredes selecionadas [36].	41

Figura 3.1 – Comparando o consumo semanal de energia de vários edifícios de acordo com suas modalidades de uso [1].	50
Figura 3.2 – Visualização do mapa de calor da demanda de energia de um edifício do campus. As colunas representam horas do dia, da meia-noite até 23h. As linhas representam os dias da semana, com os finais de semana na parte inferior. Cada quadrante em uma série de mapas de calor representa o ano em segmentos de três meses. Cada ponto corresponde à média de todos os dias desse período, em média por hora do dia e dia da semana [127].	51
Figura 3.3 – Painel online exibindo o consumo dos edifícios do campus universitário [107].	51
Figura 3.4 – Histogramas exibindo dados de consumo de energia de edifícios públicos [16].	52
Figura 3.5 – Tela do protótipo, fornecendo uma visão geral do consumo de energia de uma sala (a), o nível de conforto (b) e os tempos de operação do HVAC (c) [87].	53
Figura 3.6 – Técnicas de análise visual mostrando dados de consumo de energia de três formas diferentes: (a) Consumo total de energia (cinza claro), (b) Consumo de energia no edifício e aparelhos elétricos (vermelho) e (c) conforto de ocupação (magenta) [51].	53
Figura 3.7 – Apresentação visual do tempo de duração de uma anomalia [53].	54
Figura 3.8 – Comparação dos escores de anomalia resultantes com base nos métodos propostos. A terceira segunda-feira mostra um comportamento incomum sendo refletido nas pontuações de anomalia [31].	55
Figura 3.9 – Exibição do consumo energético de uma vizinhança com valores de anomalia calculados por diferentes métodos de detecção de anomalias. O gráfico mostra várias instâncias de anomalias e como o valor de anomalia computado difere usando as informações temporais e de vizinhança [9].	55
Figura 3.10 – Exibição da demanda de consumo semanal em um <i>Heatmap</i> [9].	56
Figura 3.11 – Operação diária em um edifício: amostragem para vários dias da semana, usando dois anos de dados [84].	57
Figura 3.12 – Painel de consumo de energia detalhando o uso de energia em casa, bem como recomendações para reduzir o consumo [13].	57
Figura 3.13 – O número de conexões <i>Wi-Fi</i> (como uma estimativa da ocupação humana) estratificadas por andar [72].	58
Figura 3.14 – Aplicativo web para monitoramento de energia em um complexo hospitalar, exibindo o consumo do hospital em diferentes gráficos [44].	58

Figura 3.15 – Plataforma visual mostrando a visualização de radar de geração de energia comparativa de vários tipos de recursos energéticos. Cada fatia é dividida em vários recursos de energia, indicando quanto desse recurso contribuiu para a energia total produzida no tempo especificado. Todas as fatias juntas mostram a geração de energia nas últimas 24 horas [30].	59
Figura 3.16 – Visão geral do sistema: (A) Exibição de marcadores por "Subclasse" e "Séries temporais"; (b) Apresentação dos números de verificações e séries temporais; (c) Amostra da porcentagem de dados afetados; (d) Exibição do consumo ao longo do tempo; (e) Amostra da severidade das indicações de verificação. (f) Exibição do filtro de verificações. (g) Apresentação das anomalias em uma usina de energia; (h, i) Visualizações das anomalias de dados para validação [8].	60
Figura 3.17 – Exibição da abordagem utilizada, onde é mostrado a avaliação das técnicas de redução de dimensionalidade. (a) Confiança. (b) Continuidade [76].	61
Figura 3.18 – A combinação dos diferentes mapeamentos visuais e a integração das técnicas de interação permitem analisar periodicamente diferentes intervalos de tempo por meio de consultas visuais. À esquerda, uma seleção visual é mostrada com os pontos destacados em vermelho para a exploração do consumo nesses pontos. Além disso, manter este destaque em uma mudança de codificação permite a ligação entre as codificações para os pontos selecionados à direita, com os pontos selecionados [89].	62
Figura 3.19 – Exibição do protótipo: (a) Mostra o consumo energético em um layout geográfico. (b) Mostra um layout do protótipo com pesos adicionais aplicados aos nós [123].	62
Figura 3.20 – A tela inicial apresenta dados de consumo de eletricidade (canto superior esquerdo), água da torneira quente (meio superior) e aquecimento (superior direito), acompanhados, respectivamente, de comparações históricas (pontos verdes) com consumo de hora em hora/diário/semanal/mensal clicando nos medidores de energia, e comparações com outras casas (pontos amarelos). Um feedback sobre a distribuição de eletricidade, água quente e consumo de aquecimento mensais, é fornecido na parte inferior esquerda, e comparações históricas diárias no canto inferior direito. Além disso, também há menus para acessar diferentes recursos da casa na parte central inferior [80].	63

Figura 3.21 – Exibição do <i>Dense Pixel Display</i> mostrando o consumo de energia permitindo que dados modelados sejam carregados e reordenados para que os padrões de consumo semanal dos dispositivos possam ser comparados. Várias opções de coloração seqüenciais são fornecidas, tais como linha, coluna ou célula selecionada, permitindo que esquemas divergentes enfatizem a diferença dos itens selecionados. Os padrões nos dados diários (Iluminação), sazonal (Aquecimento) e modelados (Lavagem e Limpeza) são claros, assim como as diferenças nos finais de semana (duas linhas inferiores de cada célula), como o atraso no pico do aquecimento da manhã e mais cozimento durante o dia nos fins de semana [45].	64
Figura 3.22 – Apresentação do <i>Dashboard</i> exibindo dados de umidade, temperatura, dióxido de carbono, eletricidade, gás e água no sensor instalado na cozinha de uma residência [125].	65
Figura 3.23 – Exploração de dados com diferentes níveis de zoom. Em todas as figuras (a) - (c), o gráfico na parte inferior é o histórico completo dos dados, enquanto o gráfico diretamente acima mostra os dados correspondentes à janela de tempo selecionada pela janela transparente na visão geral gráfico. Em cada parcela, as regiões azuis são devidas a horários de pico, enquanto as regiões vermelhas correspondem a horários de pico [2].	66
Figura 3.24 – Amostragem utilizada: a) Padrões de atividade entre homens e mulheres em domicílios com dois adultos entre 20 e 35 anos e crianças (= fig. 3). (b) Curvas de carga geradas por atividade para eletricidade entre os homens no grupo, e (c) entre as mulheres no grupo [38].	67
Figura 3.25 – Ferramenta mostrando um gráfico de séries temporais de consumo de energia que pode ser anotado pelos usuários para ajudá-los a entender seus próprios dados [33].	67
Figura 3.26 – Frequência de erros e erros para o consumo de energia estimado pelo <i>Energy plus</i> [128].	68
Figura 3.27 – Exibição das diferenças entre os perfis de carga e o consumo real de uma família <i>Load profile generator</i> [83].	69
Figura 3.28 – Visualização gerada pelo <i>tableau</i> mostrando dados de vendas e lucro e quatro dimensões: Tempo, Estado, Produto e Fornecedor [116].	70
Figura 4.1 – Etapas do desenvolvimento do modelo.	75
Figura 4.2 – . Estrutura do JSON para carregar dados: A) Exibe a formatação dos dados de energia. B) Apresenta a formatação dos dados meteorológicos.	79
Figura 4.3 – Fluxo de atividades da análise visual interativa.	81
Figura 4.4 – Componentes do modelo proposto.	82
Figura 4.5 – Interface completa do modelo desenvolvido.	84

Figura 4.6 – Exibição do <i>Calendar View</i> . Abaixo de cada mês é exibido o valor de consumo ou da geração semanal e o número da semana do ano.	85
Figura 4.7 – Exibição dos meses de janeiro e fevereiro no <i>Calendar View</i> , onde são apresentados no rodapé de cada mês a semana do ano e a média de consumo ou de geração de energia.	86
Figura 4.8 – Representação dos dados de consumo de energia em um ano: as horas de um dia são exibidas no eixo vertical, e os meses ao longo do ano são mostrados no eixo horizontal.	87
Figura 4.9 – Resultado de um <i>zoom in</i> , após a seleção de uma área no componente.	87
Figura 4.10 – Exibição dos dados filtrados iguais ou superiores ao valor contido no <i>range</i>	88
Figura 4.11 – <i>Boxplot</i> exibindo o consumo energético mensal.	89
Figura 4.12 – Mapa de calor mostrando dados de energia de uma semana.	90
Figura 4.13 – Mapa de calor exibindo os marcadores que identificam em quais horas de uma semana houve o maior (triângulo vermelho) e o menor valor de energia (triângulo azul).	91
Figura 4.14 – Mapa de calor com os marcadores que mostram o horário de maior (triângulo vermelho) e menor valor de energia (triângulo azul) em cada dia de uma semana.	91
Figura 4.15 – Exibição dos gráficos para visualização de séries temporais, que são: A) Gráfico de barras, B) Gráfico de linhas, C) Gráfico polar (ou radar), D) Gráfico de plot, E) Gráfico de área.	93
Figura 4.16 – Exibição dos gráficos de dias e horas mostrando valores de consumo de acordo com a seleção feita, e variância, desvio padrão, erro padrão e coeficiente de variação do conjunto de dados.	94
Figura 4.17 – Exibição dos gráficos de dia e horas mostrando valores de consumo, mínimos, médios, medianas e máximos.	95
Figura 4.18 – Fluxo de exibição para visualização com diferentes granularidades de tempo: A) Seleção de um dado a partir do <i>Calendar View</i> ; B) Seleção de um dado a partir do <i>Dense Pixel Display</i> ; Qualquer seleção de uma das etapas anteriores carregam os dados semanais. A partir da seleção de uma célula no <i>Heatmap</i> , as etapas C e D ocorrem, exibindo os dados em dias e horas.	96
Figura 4.19 – Exibição dos gráficos para análise de consumo anuais.	97
Figura 4.20 – Exibição dos energias do conjuntos de dados, neste caso há apenas uma fonte de energia.	97

Figura 4.21 – Janela <i>pop-up</i> mostrando as opções disponíveis (mensal, semanal e diário) para visualizar os dados a partir do <i>Calendar View</i>	98
Figura 4.22 – Janela <i>pop-up</i> mostrando as opções disponíveis (mensal, semanal, diário e horas) para visualizar os dados a partir do <i>Dense Pixel Display</i>	98
Figura 4.23 – Apresentação das janelas <i>pop-up</i> em diferentes exemplos e mostrando dados de diferentes conjuntos de dados.	99
Figura 4.24 – Exibição do gráfico para análise de consumo anual dos anos de 2011 à 2016.	101
Figura 4.25 – Exibição do gráfico para análise de consumo anual do ano de 2017.	101
Figura 4.26 – Exibição do gráfico para análise de consumo anual do ano de 2017 juntamente com os valores de predição gerada pelos quatro algoritmos.	102
Figura 4.27 – Exibição do gráfico para análise das predições geradas para o ano de 2018.	102
Figura 4.28 – Exibição dos valores dos conjuntos gerados com as estimativas de erros em porcentagem para cada algoritmo preditivo.	103
Figura 5.1 – <i>Calendar view</i> exibindo o consumo de energia ao longo de um ano.	109
Figura 5.2 – <i>Dense Pixel Display</i> para análise do conjunto de dados B.	110
Figura 5.3 – Exibição dos dados de todas as fontes de consumo de energia do conjunto de dados B.	111

LISTA DE TABELAS

Tabela 3.1 – Conjunto inicial de artigos selecionados.	47
Tabela 3.2 – Artigos do <i>Snowballing Forward</i>	48
Tabela 3.3 – Artigos do <i>Snowballing Backward</i>	49
Tabela 3.4 – Artigos removidos desta pesquisa.	49
Tabela 3.5 – Tabela dos artigos agrupados por diferentes categorias, onde cada artigo exibe os nomes dos autores, ano da publicação e objetivo do artigo. .	72
Tabela 3.6 – Tabela comparativa de técnicas de visualização de dados, modelos de interação e outros recursos disponíveis (Fornece <i>Dashboard</i> , Permite Comparações e Algoritmos de detecção de anomalias) utilizados nos 29 artigos selecionados.	73
Tabela 4.1 – Exibição dos valores gerados pelos algoritmos de predição e as taxas de erros de cada um deles em cada mês do ano.	104
Tabela 5.1 – Perfil dos participantes, incluindo ferramentas usadas e formatos dos dados utilizados.	115

LISTA DE SIGLAS

AR - *AutoRegressive*

CI - Cidade Inteligente

CMV - *Coordinated Multiple Views*

EI - Edifício Inteligente

HVAC - *Heating, Ventilation, Air Conditioning*

MAE - *Mean Absolute Error*

SAR - *Sarima*

TIC - Tecnologia da Informação e Comunicação

SUMÁRIO

1	INTRODUÇÃO	25
2	FUNDAMENTAÇÃO TEÓRICA	29
2.1	VISUALIZAÇÃO DE DADOS	29
2.1.1	VISUALIZAÇÃO DE DADOS DE SÉRIES TEMPORAIS	29
2.2	ANÁLISE VISUAL (<i>VISUAL ANALYTICS</i>)	30
2.2.1	<i>CALENDAR VIEW</i>	33
2.2.2	<i>DENSE PIXEL DISPLAYS</i>	34
2.2.3	<i>HEATMAP</i>	34
2.2.4	<i>BOXPLOT</i>	35
2.3	CIDADES E EDIFÍCIOS INTELIGENTES	36
2.4	<i>DASHBOARD</i>	37
2.5	MODELOS DE INTERAÇÃO	37
2.5.1	<i>BRUSHING AND LINKING</i>	38
2.5.2	<i>COORDINATE MULTIPLE VIEWS</i>	39
2.5.3	<i>DETAILS ON DEMAND</i>	39
2.6	ALGORITMOS DE ANÁLISE PREDITIVA	40
2.6.1	<i>AUTOREGRESSIVE</i>	41
2.6.2	<i>HOLT-WINTERS</i>	42
2.6.3	<i>SARIMA</i>	42
2.6.4	<i>PROPHET</i>	43
3	TRABALHOS RELACIONADOS	45
3.1	METODOLOGIA DE PESQUISA	45
3.2	EDIFÍCIOS INTELIGENTES	48
3.3	DETECÇÃO DE ANOMALIAS	54
3.4	EFICIÊNCIA ENERGÉTICA	56
3.5	REDES INTELIGENTES	60
3.6	RESIDÊNCIAS INTELIGENTES	61
3.7	FERRAMENTAS	67
3.8	ANÁLISE DOS ARTIGOS RELACIONADOS	69

4	DESCRIÇÃO DO MODELO	75
4.1	METODOLOGIA	75
4.2	LEVANTAMENTO DE REQUISITOS	76
4.3	ENTRADA DE DADOS	78
4.4	COMPONENTES DO MODELO E TÉCNICAS DE INTERAÇÃO	80
4.5	IMPLEMENTAÇÃO E TÉCNICAS DE VISUALIZAÇÃO	83
4.5.1	GRÁFICOS PARA DADOS ANUAIS	83
4.5.2	GRÁFICOS PARA DADOS MENSAIS	88
4.5.3	GRÁFICOS PARA DADOS SEMANAIS	89
4.5.4	GRÁFICOS PARA DADOS DIÁRIOS E EM HORAS	90
4.5.5	ANÁLISES COMPARATIVAS	95
4.5.6	PREDIÇÃO DE CONSUMO	99
5	RESULTADOS E AVALIAÇÃO DO MODELO	107
5.1	CONJUNTOS DE DADOS	107
5.2	RESULTADOS	108
5.3	ENTREVISTAS COM ESPECIALISTAS	109
5.4	EXPERIMENTOS E LIMITAÇÕES	114
6	CONCLUSÕES E TRABALHOS FUTUROS	117
	REFERÊNCIAS	119
	APÊNDICE A – ENTREVISTA PARA COLETA DE DADOS JUNTO A ESPECIALISTAS	131
	APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)	135

1. INTRODUÇÃO

O conceito de Cidade Inteligente (CI) tem sido bastante difundido nos últimos anos [32, 77] e está relacionado com infraestrutura de tecnologia da informação e comunicação (TIC) [48]. Ações de várias áreas do conhecimento são combinadas em uma cidade inteligente que é caracterizada por avanços tecnológicos e desenvolvimento urbano [6], visando sustentabilidade e melhoria na qualidade de vida dos cidadãos [4].

O aperfeiçoamento da tecnologia proporcionou muitas vantagens para as cidades com a geração de sistemas integrados a sensores que também começaram a ser utilizados nas construções e no gerenciamento de edifícios. Os chamados Edifícios Inteligentes (EI) incorporam dispositivos de controle automático aos seus sistemas técnicos e administrativos. O conceito de EI é antigo, surgiu pela primeira vez nos EUA na década de 80, quando sistemas de automação de segurança e iluminação começaram a ser conectados [75], mas recentemente receberam maior visibilidade. Um edifício é considerado inteligente quando é capaz de oferecer um ambiente produtivo e com uma boa relação custo-benefício pela otimização de sistemas, serviços, estrutura e gerenciamento [56, 81].

Um dos desafios, tanto para uma CI quanto para um EI, é analisar os dados coletados, pois estes dados são gerados por diversas fontes, disponibilizados em diferentes formatos e podem formar um grande conjunto de dados. A maioria destes dados correspondem a dados de séries temporais, que são conjuntos de dados coletados sequencialmente ao longo do tempo, consistindo em observações discretas de um fenômeno, armazenados em intervalos de tempo consistentes [37]. A análise de dados de séries temporais desempenha um papel importante para a investigação de padrões, modelagem e previsão do comportamento. Este tipo de análise disponibiliza uma visão da evolução ao longo do tempo ou de mudanças incomuns [112].

Dados de consumo de energia que têm como origem diferentes aparelhos eletrônicos, eletrodomésticos e de ar condicionado, podem ser coletados em um EI. Da mesma forma, dados de geração de energia, provenientes de diversas fontes, tais como energia solar ou eólica, também podem ser coletados por empresas ou até mesmo em um EI. Compreender estes dados é uma tarefa importante para avaliar a eficiência energética, que visa diminuir o consumo de energia de modo eficiente, usando energias renováveis ou reduzindo custos mantendo o funcionamento dos serviços [5]. Sendo assim, observamos que um dos principais problemas no desenvolvimento de cidades e edifícios inteligentes é transformar uma grande quantidade de dados coletados em informações e, finalmente, em decisões estratégicas e táticas.

Com o aumento do volume de dados, torna-se importante buscar formas de analisá-los. O uso de técnicas de visualização, associando dados às representações visuais, facilita a sua análise e a obtenção de informações sobre os mesmos [91]. Segundo Card et al. [22]

o principal objetivo da visualização de dados é comunicar a informação de maneira clara e eficaz através de representações visuais interativas e suportadas por um computador, visando ampliar a cognição. Estas representações visuais não necessitam ter um visual muito sofisticado, porém devem fornecer o entendimento do conjunto de dados, comunicando seus aspectos de forma objetiva para o usuário. Assim, possibilita-se a visualização de grandes quantidades de dados de maneira rápida e eficiente, auxiliando tanto na compreensão do problema, como na descoberta de um novo entendimento sobre os dados, e, conseqüentemente, na tomada de decisões.

A utilização de técnicas de visualização de dados pode trazer benefícios tanto para cidades quanto para EI. Foram encontrados trabalhos na literatura que possibilitam fazer uma análise visual de dados de energia [110, 88]. Entretanto, observamos que muitos destes trabalhos não possibilitam algumas análises interativas que podem facilitar a tomada de decisão, tais como: comparações dinâmicas entre diferentes fontes de consumo de energia ou entre o consumo de energia associada a dados meteorológicos; visualizações com diferentes granularidades, isto é, em diferentes escalas de tempo; análises estatísticas e algoritmos de predição.

Neste contexto, surge a motivação deste trabalho, que visa apresentar um modelo para análise visual de dados de energia para cidades e edifícios inteligentes. Assim, definimos a seguinte questão de pesquisa para estudo: **Como é possível auxiliar o estudo e a tomada de decisão relacionada a dados de consumo ou de geração de energia através de um modelo de análise visual?** Como objetivos específicos, temos: a criação de um modelo para visualização, análise e comparação de dados de consumo ou de geração de energia e dados meteorológicos; teste do modelo com dados de consumo ou de geração de energia gerados por Cidades e/ou Edifícios Inteligentes; Utilização de algoritmos de análise preditiva; Validação do modelo com especialistas.

Para atingir os objetivos propostos, desenvolvemos um modelo para análise e comparação visual de dados de energia que podem ser obtidos ao longo do tempo. Este modelo apresenta uma abordagem interativa baseada na técnica de detalhes sob demanda integrada com *Coordinated Multiple Views* (CMV), e disponibiliza quatro diferentes algoritmos que permitem fazer uma análise preditiva, além de várias técnicas de visualização. Realizamos uma avaliação qualitativa com usuários especialistas de domínio, que demonstrou a viabilidade e as vantagens de usá-lo para monitorar dados de energia e dados meteorológicos. Assim, as principais contribuições deste trabalho são: a possibilidade de analisar e comparar dados meteorológicos e dados de energia de diferentes fontes ao longo do tempo e com diferentes granularidades; a disponibilidade de técnicas de interação que suportam a análise visual; o fornecimento de algoritmos de predição; e avaliação com especialistas que demonstra as vantagens do uso do modelo. Para melhor compreensão deste estudo, consideramos como fontes de consumo de energia diferentes aparelhos eletrônicos, tais como aparelhos de ar condicionado, lâmpadas e eletrodomésticos.

O restante do documento está organizado da seguinte maneira: o Capítulo 2 apresenta uma fundamentação teórica com a descrição de alguns conceitos, algoritmos e técnicas usadas no desenvolvimento do trabalho; O Capítulo 3 aborda alguns trabalhos relacionados a este estudo; A descrição do modelo proposto é apresentada no Capítulo 4 e o Capítulo 5 apresenta os resultados obtidos e uma avaliação preliminar do modelo; Por fim, no Capítulo 6, são apresentadas as conclusões e os trabalhos futuros para esta dissertação de mestrado.

2. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo é apresentada a base teórica dos temas relacionados a este estudo. São abordadas as seguintes áreas: Visualização de dados, Cidades e Edifícios Inteligentes, (*Visual Analytics*), *Dashboard*, Modelos de Interação e Algoritmos de análise preditiva.

2.1 Visualização de dados

Segundo Sadiku et al. [96] a visualização de dados envolve a apresentação de dados em forma gráfica, auxiliando o entendimento e a análise da informação. Ajuda a explicar fatos e beneficia qualquer campo de estudo que exija formas inovadoras de apresentar informações grandes e complexas [96]. As representações visuais surgiram antes mesmo da escrita, milhares de anos atrás, quando os acontecimentos do cotidiano eram desenhados visando repassar conhecimento para as próximas gerações [66].

Existem diferentes técnicas de visualização que são aplicadas para análise de dados. O uso de cada técnica depende de alguns fatores, como o tipo de dado, quais informações se deseja obter com o uso das técnicas e qual é o público alvo. Ao longo dessa seção, serão destacadas algumas categorias de visualização de dados que se aplicam a este estudo, que são: Visualização de dados de séries temporais, *Heatmap* e *Dense Pixel Displays*.

2.1.1 Visualização de dados de séries temporais

Dados de séries temporais são conjuntos de valores que mudam ao longo do tempo. Através da visualização destes dados podem ser descobertas informações, como padrões frequentes [43] e detecção de *outliers* [71].

Na última década, houve muito interesse no desenvolvimento de técnicas de visualização de dados de séries temporais [49]. Há exemplos do uso deste tipo de visualização na área de: finanças, como análise dos preços diários de ações e taxas de câmbio; ciências, para verificar temperaturas e níveis de poluição que variam de hora em hora; e políticas públicas, como análise de taxas de criminalidade e segurança. Desse modo, as visualizações auxiliam na interpretação das informações com maior precisão. Analistas de dados podem comparar informações ao longo tempo, conforme ilustra a Figura 2.1, recurso de grande benefício prático para obter um resultado significativo. Uma ferramenta de visualização de dados de séries temporais deve ser fácil e simples de ser operada através da interface com o usuário, e os resultados exibidos devem ser de rápida compreensão e entendimento.

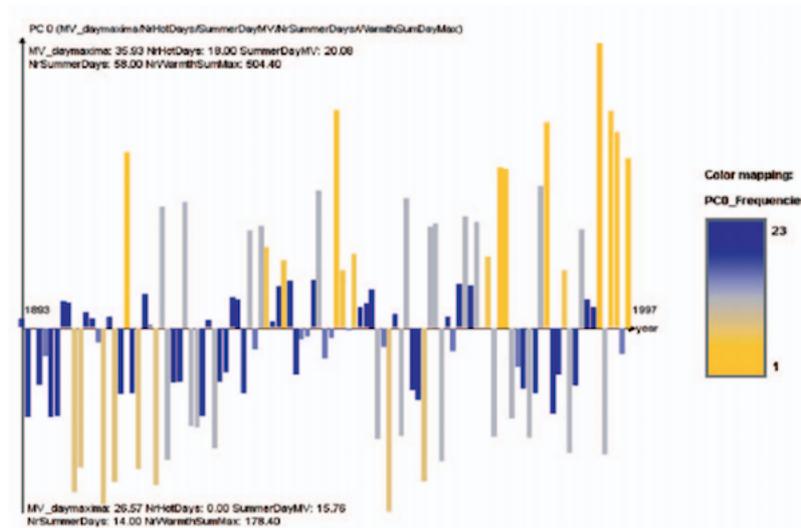


Figura 2.1 – Visualização de gráfico de barras ao longo do tempo. Barras ascendentes representam condições mais quentes, enquanto barras descendente representam verões mais frios de um conjunto de dados de dados climáticos. [3].

A exibição efetiva de séries temporais múltiplas é um problema de pesquisa na área de visualização: Como aumentar a quantidade de dados que os analistas humanos podem efetivamente interpretar e compreender? Tufte [109] aconselha *designers* a maximizar a densidade de dados (representações visuais de dados por área de gráfico). Essas abordagens se destacam no aumento da quantidade de informações que podem ser mapeadas dentro de um espaço de exibição. Este aumento da densidade de dados permite a exibição de mais gráficos em uma área fixa, tornando mais fácil para os usuários compararem informações de vários gráficos [99, 109].

As técnicas de visualização de dados de séries temporais mais conhecidas são gráficos de linhas, polar, plot, barras e área. A Figura 2.2 ilustra estes gráficos.

2.2 Análise Visual (*Visual Analytics*)

Segundo Thomas e Cook [105], *Visual Analytics*, ou Análise Visual, "é a ciência do raciocínio analítico facilitado por uma interface visual interativa, fornecendo um meio eficaz para interagir dinamicamente e visualmente, para explorar e analisar dados grandes e complexos". Permite a compreensão dos dados e facilita o desenvolvimento de estratégias relevantes de mitigação de riscos. É eficiente na detecção de comportamentos esperados e, mais importante, na descoberta do inesperado. É uma área multidisciplinar que engloba [105]:

- Técnicas de raciocínio analítico que permitem aos usuários obter *insights* que apoiam diretamente a avaliação, o planejamento e a tomada de decisões;



Figura 2.2 – Exibição de dados de consumo de energia em gráficos de visualização de dados de séries temporais, que são: A) Linhas; B) Polar; C) Plot; D) Barras; E) Área. No eixo horizontal são mostrados os dias da semana e no eixo vertical as faixas de consumo.

- Representações gráficas e técnicas de interação que permitam aos usuários diferentes visualizações, novas formas de exploração dos dados e entendimento de grandes quantidades de informações imediatamente;
- Técnicas de apoio à produção, apresentação e disseminação dos resultados obtidos através de uma análise, visando comunicar a informação identificada sobre um determinado contexto.

Em outras palavras, *Visual Analytics* é um processo interativo que envolve coleta de informações, pré-processamento de dados, visualização, interação, representação do conhecimento e tomada de decisão, conforme ilustra a Figura 2.3. O objetivo final é obter uma visão sobre um determinado problema, que pode ser descrito, por exemplo, por uma grande quantidade de dados científicos, forenses ou comerciais de fontes heterogêneas. Os métodos de descoberta de conhecimento em bancos de dados, estatística e matemática são competências importantes dentro deste promissor campo de pesquisa [60].

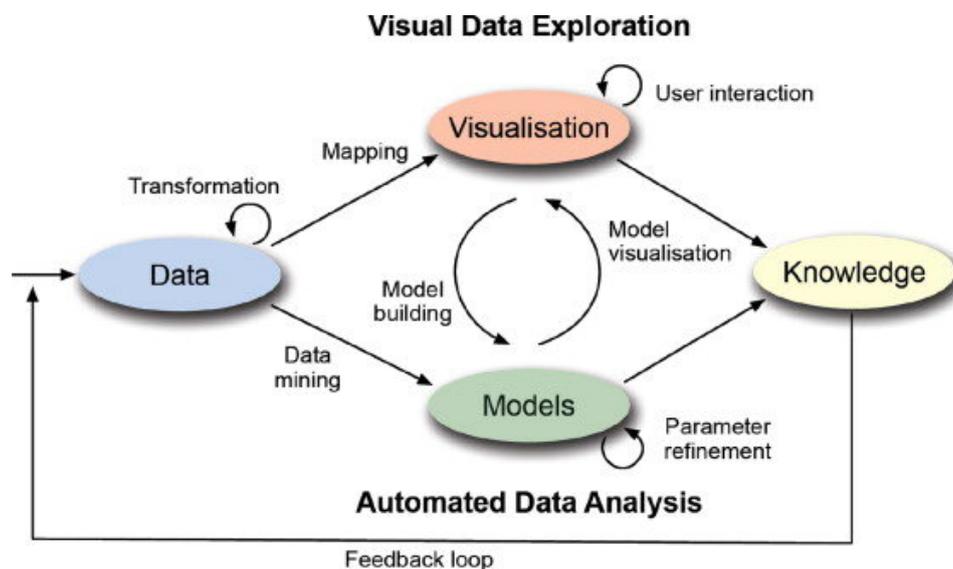


Figura 2.3 – O processo do *Visual Analytics* é caracterizado pela interação entre dados, visualizações, modelos sobre os dados e os usuários, a fim de fomentar a descoberta do conhecimento [64].

Visual Analytics é mais que apenas visualização, pois combina visualização, fatores humanos e análise de dados. A Figura 2.4 ilustra o escopo do *Visual Analytics*, que integra diferentes análises e fatores humanos (interação, cognição, percepção, etc.), desempenhando um papel fundamental na comunicação entre humanos e computadores, bem como no processo de tomada de decisão. Este campo multidisciplinar visa auxiliar na compreensão dos resultados analíticos de um contexto, usando termos que sejam significativos para eles, para disseminar as informações com o público-alvo [124]. Sobre a análise de dados, o *Visual Analytics* reforça a importância das metodologias desenvolvidas nos campos de gerenciamento de informações, descoberta do conhecimento e análise estatística. Nota-se que o *Visual Analytics*, não é susceptível de se tornar um campo de estudo separado [106], mas sua influência irá se espalhar sobre as áreas de pesquisa que compreende.

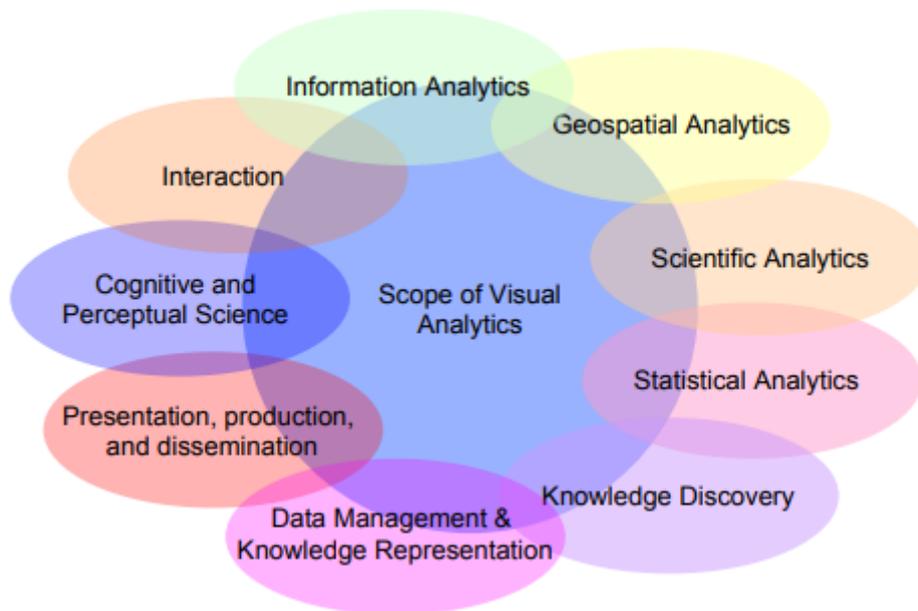


Figura 2.4 – O escopo do *Visual Analytics* [60].

2.2.1 *Calendar View*

Calendar View é uma técnica de visualização que permite exibir dados em diferentes períodos de tempo e divididos em unidades, como dias, semanas, meses e anos [119], como mostra a Figura 2.5. É baseado em tabelas que organizam eventos ao longo de eixos de tempo representativos, como as horas do dia ao longo do eixo vertical e os dias da semana ao longo do eixo horizontal [97].

Essa técnica é interessante para renderizar grandes conjuntos de dados porque permite exibir milhares de valores em uma única tela. Diferentemente do *Dense Pixel Displays* (descrito na Seção 2.2.2), ele agrupa dados diários. A identificação de padrões em várias granularidades de tempo é possível quando combinada com técnicas de interação como *Details on demand* e *Brushing and Linking*, apresentadas, respectivamente, nas Seções 2.5.3 e 2.5.1.

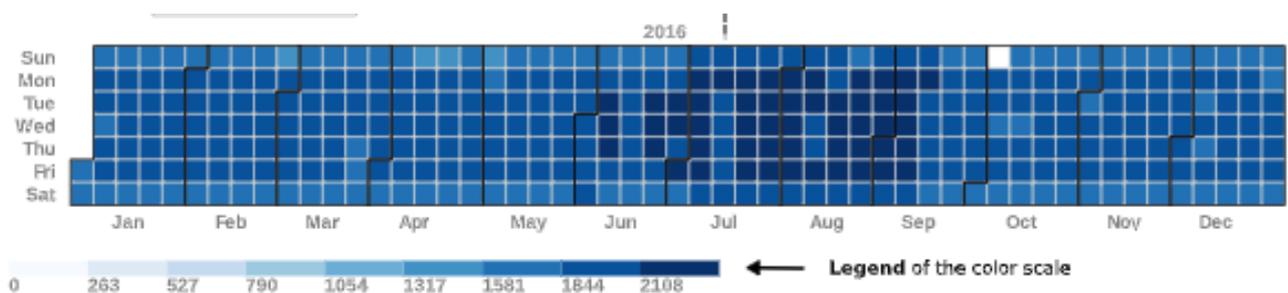


Figura 2.5 – *Calendar View* representando a média da demanda de energia agrupada por dias de um ano [44].

2.2.2 *Dense Pixel Displays*

Dense Pixel Displays, também conhecida como técnica orientada a pixel, é uma junção entre métodos baseados em pontos e regiões. Essa técnica mapeia cada valor em um pixel individual, criando um polígono preenchido para representar cada dimensão de dados [57, 59]. Este tipo de gráfico usa o máximo do espaço em tela, permitindo que conjuntos de dados com milhões de valores sejam exibidos em uma única tela. Cada valor de dado define a cor de um único pixel. Este gráfico pode revelar novos recursos e padrões dos dados.

Ward et al. [114] explicam que cada dimensão de um conjunto de dados gera uma sub-imagem separada dentro da visualização, como ilustra a Figura 2.6, que mostra que cada dimensão é independente e determina a cor dos pixels correspondentes [114]. Esta técnica fornece a capacidade de ampliar áreas de visualização. Se houver uma sobreposição de informações nessas áreas, o sistema ou modelo que aplica esta técnica deverá oferecer filtros para que o usuário possa obter uma visão mais detalhada destes pixels [39].

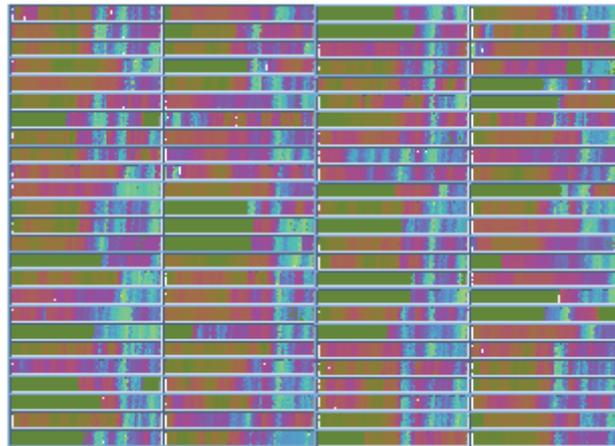


Figura 2.6 – Visualização de padrões recursivos dos preços diários de ações de 100 ações em 20 anos [57].

2.2.3 *Heatmap*

Outra técnica de visualização bastante utilizada é o *Heatmap*, ou mapa de calor. *Heatmap* é uma das representações mais efetivas para a visualização de dados de séries temporais. Esta visualização pode exigir um maior espaço para sua exibição quando um conjunto de dados de entrada possuir uma grande quantidade de informações. Muitas vezes, pode ser necessário disponibilizar mecanismos para filtrar de forma interativa da-

dos não importantes, ou períodos de tempo, para que possam formar mapas de calor de tamanhos apropriados e focar em pontos de interesse [67].

Esta técnica de visualização possui vantagens sobre a utilização do espaço de exibição em relação a outras representações, conforme ilustra a Figura 2.7, que evidenciou uma variância de genes entre si e só foi possível esta verificação com o uso do *heatmap* [41]. A visualização de dados de séries temporais com *Heatmap* é útil para o desenvolvimento de ferramentas de análise visual [27].

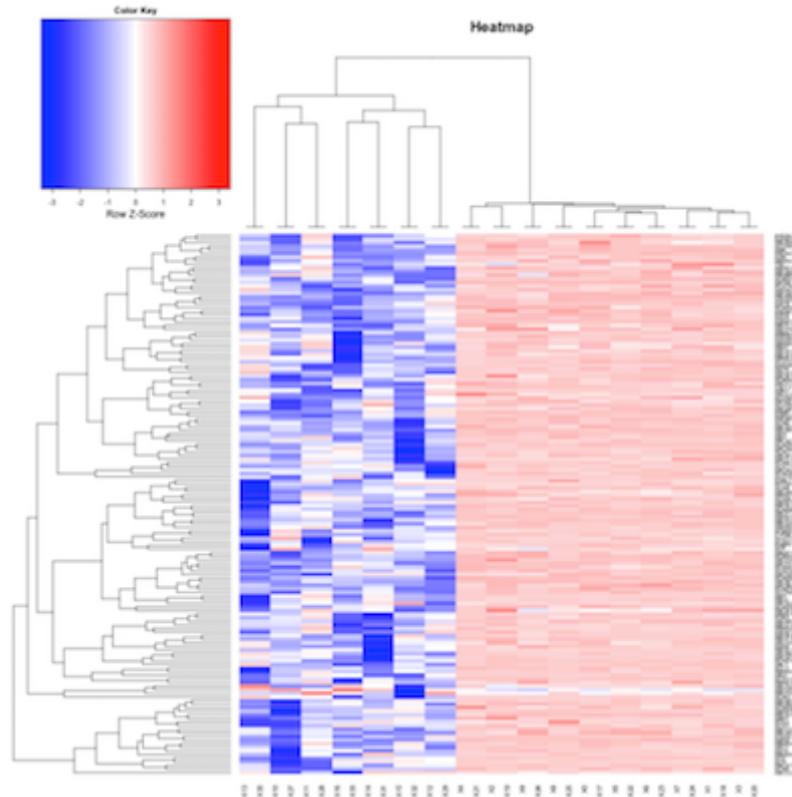


Figura 2.7 – *Heatmap* exibindo dados de genes, evidenciando a diferença do grupo de genes de pacientes sem câncer no lado esquerdo e pacientes com câncer no lado direito [41].

2.2.4 *Boxplot*

Boxplots são representações gráficas baseadas em estatísticas que fornecem uma visão geral e um resumo numérico de um conjunto de dados [40]. Normalmente mostram a mediana, a média, os intervalos de confiança e os valores discrepantes de um conjunto de dados, sendo útil para comparar distribuições entre grupos [118]. A Figura 2.8 detalha a sua definição. A altura da caixa corresponde à distribuição dos valores, ou seja, um *boxplot* com uma caixa pequena significa que há uma concentração de dados em um intervalo menor de valores, e um *boxplot* com uma caixa maior representa uma distribuição mais uniforme [62]. Os quartis, que são uma forma de variação de dados observados, são representados pelas

variáveis Q1, Q2 e Q3, sendo que o Q1 apresenta 25% dos menores valores na amostra, o Q2 ou mediana, que representa a divisão da metade superior (ou 50%) da metade inferior da amostragem e o Q3 exibe dados igual ou superior a 75% da amostragem.

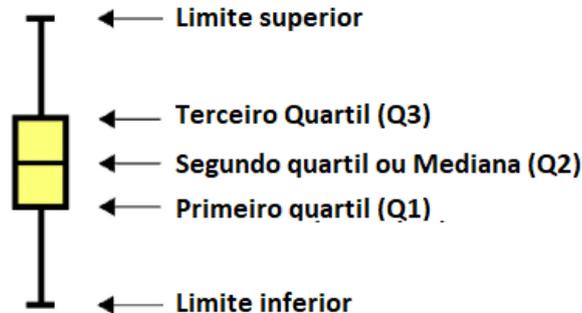


Figura 2.8 – Os principais componentes de um *boxplot*: mediana, quartis, limite inferior e limite superior [40].

2.3 Cidades e Edifícios Inteligentes

Há um consenso de que uma Cidade Inteligente (CI) é um conceito muito amplo, representando inovação não apenas em infraestrutura física, mas também em fatores humanos e sociais, visando melhorar a qualidade de vida dos cidadãos [78]. Um dos seus principais objetivos é fornecer informações em tempo real, a fim de lidar com desafios futuros e gerenciar os recursos disponíveis de maneira mais eficiente [24]. Para que esse processo seja funcional, é necessário trabalhar com sensores capazes de coletar e compartilhar informações [47]. Esses sensores enviam dados, por exemplo, para um centro de controle que tem a capacidade de monitorar, gerenciar e controlar informações de edifícios (como temperatura externa e interna) ou condições de tráfego. Após a coleta e análise das informações, as tecnologias existentes em tal centro de controle permitiriam, então, gerenciar a operação dos condicionadores de ar e ajustar o intervalo dos semáforos [61].

Como o avanço tecnológico oferece muitas vantagens para as cidades com o desenvolvimento de sistemas de sensores integrados, não demorou muito para que esses recursos começassem a ser usados na construção e operação de edifícios. Desta forma, Edifícios Inteligentes, que incorporam dispositivos de controle automático aos seus sistemas técnicos e administrativos, passaram a receber maior visibilidade. O conceito de Edifício Inteligente (EI) não é novo, surgiu nos EUA nos anos 80, quando os sistemas de automação de segurança e iluminação começaram a ser usados [75].

No final da década de 1980, surgiram os primeiros prédios com automação de energia, combinando sistemas existentes de automação predial com outros recursos tecnológicos, como medidores inteligentes, para permitir o gerenciamento do consumo de energia

[56, 81]. Um edifício é considerado inteligente quando é capaz de oferecer um ambiente produtivo e econômico otimizando sistemas, serviços, estrutura e gerenciamento. Assim, o “inteligente” de um edifício não pode ser avaliado apenas pelo número de sistemas automatizados, pois um edifício pode ter um alto nível de automação e não ser inteligente. Portanto, um EI deve apresentar uma integração de sistemas, tais como: sistemas de coleta de dados; automação e gerenciamento predial; sistema de telecomunicações; e um sistema de computador que gerencia todos os anteriores.

2.4 *Dashboard*

Não há uma definição formal do que é um *dashboard*, mas se pode afirmar que *Dashboards* e visualização são ferramentas cognitivas que visam melhorar o “alcance de controle” sobre muitos dados corporativos [63]. Eles podem fornecer um meio único e poderoso para apresentar informações, mas eles raramente correspondem ao seu potencial. A maioria dos *dashboards* falham na comunicação, não por causa de tecnologia inadequada, mas por causa de implementações mal projetadas [108]. Wexler et al. [117] relata que um *Dashboard* requer visualizações de dados atualizadas em uma única página, além da exibição de dados usados para monitorar condições e/ou facilitar a compreensão, podendo incluir elementos de infográfico ou visualizações narrativas, conforme demonstrado na Figura 2.9.

Dashboards são um dos casos de uso mais comuns para visualização de dados, e seu *design* e contextos de utilização são consideravelmente diferentes das ferramentas de visualização exploratória. Eles são construídos e empregados para apoiar a tomada de decisão orientada por dados em difentes áreas do conhecimento. Por exemplo, eles podem ser usados pelos alunos para acompanhar e monitorar o consumo de energia.

2.5 Modelos de interação

Além de projetar as formas de visualização de dados, é necessário identificar maneiras para que o usuário possa explorar os dados de forma adequada, permitindo a realização de seleção, filtragem, agrupamento e outras diversas operações nos conjuntos de dados, para obter diferentes perspectivas visuais. Os modelos de interação auxiliam neste processo, permitindo uma maior flexibilidade para analisar os dados de diferentes maneiras e fornecendo uma seleção interativa dos dados a serem visualizados mais detalhadamente. Dada a importância deste processo, foram selecionados três modelos de interação para serem descritos: *Brushing and Linking*, *Coordinated Multiple Views (CMV)* e *Details on demand*.



Figura 2.9 – Exemplos de *Dashboards*: Os *Dashboards* 1 e 5 ilustram painéis direcionados à tomada de decisões, enquanto os *Dashboards* 3 e 4 dividem em blocos as informações para facilitar a análise dos dados. O *Dashboard* 2 tem como alvo o cenário autônomo quantificado, embora *Dashboard* 6 represente painéis projetados para comunicação de propósito geral. O *Dashboard* 7 captura novas extensões para painéis tradicionais [98].

2.5.1 *Brushing and Linking*

Brushing and Linking é uma técnica de interação para explorar diferentes visualizações de um conjunto de dados [58], que normalmente está integrada com *Coordinated Multiple Views* [93] e é amplamente adotada por ferramentas de visualização, como *Tableau*¹, *SpotFire*² e *D3.js*³ [65].

Brushing and Linking permite ao usuário selecionar alguns dados de um gráfico e visualizar como esta seleção é refletida em outros gráficos vinculados, como demonstrado na Figura 2.10. Nesta figura, quatro gráficos diferentes estão vinculados. Assim, quando houver a seleção em um gráfico, automaticamente os outros gráficos mostrarão dados da seleção realizada, otimizando a análise dos dados [115].

É uma técnica de interação comum em ferramentas de visualização modernas para dados tabulares e uma estratégia eficaz para entender as relações entre subconjuntos de dados e dimensões.

¹<https://www.tableau.com/>

²<https://www.tibco.com/products/tibco-spotfire>

³<https://d3js.org/>

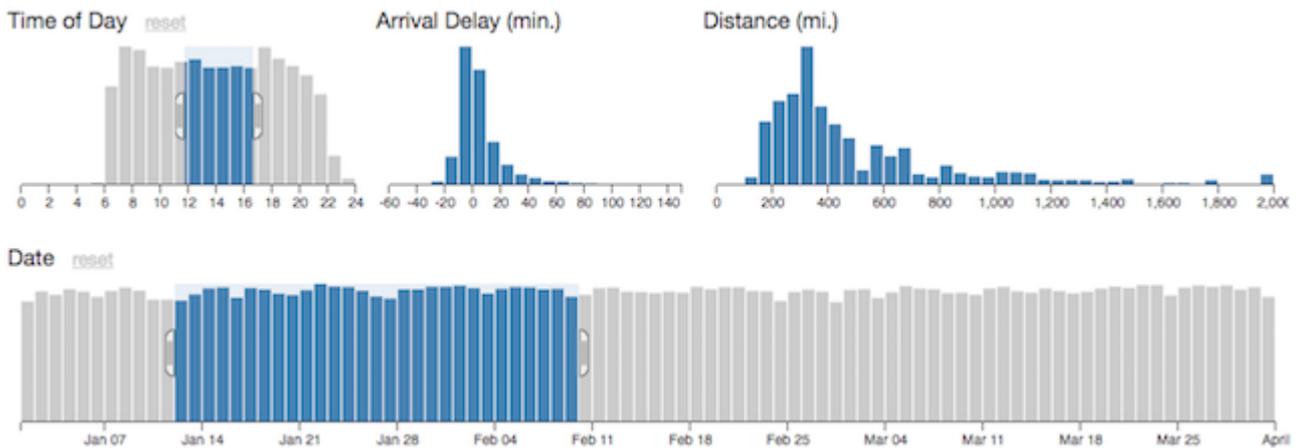


Figura 2.10 – Exibição do uso do modelo de interação *Brushing and Linking*. Ao fazer uma seleção em qualquer um dos gráficos, os demais gráficos exibirão dados referente a seleção aplicada [115].

2.5.2 *Coordinate multiple views*

Coordinated Multiple Views (CMV) é uma técnica de visualização exploratória que permite aos usuários visualizarem seus dados de diferentes formas. A premissa desta técnica é que os usuários entendam seus dados, interajam com as informações apresentadas e visualizem através de diferentes representações [93], conforme mostra a Figura 2.11. Esta figura exhibe a combinação de diferentes visualizações de um mesmo conjunto de dados. Ela é gerada através de uma seleção interativa de dados, permitindo que os usuários construam visualizações integradas em diferentes blocos dinâmicos. Quando um bloco recebe uma nova seleção ou é atualizado, todas as outras visualizações refletem aos dados selecionados [73].

O principal objetivo do CMV é encontrar informações e entender o grande volume de conjuntos de dados potencialmente diversos de vários componentes e tipos. O usuário deseja entender tendências, localizar anomalias, isolar e reorganizar informações, comparar e deixar claras quaisquer diferenças ou similaridades entre conjuntos de dados [105]. Além disso, o usuário pode examinar cenários e desenvolver hipóteses através de uma pesquisa sistemática. Para conseguir isso, a visualização precisa estar em um ambiente altamente interativo que permita ao usuário encontrar e descobrir informações.

2.5.3 *Details on demand*

Details on demand é uma técnica que permite a seleção interativa de partes de dados a serem visualizadas mais detalhadamente, ao mesmo tempo em que fornece uma

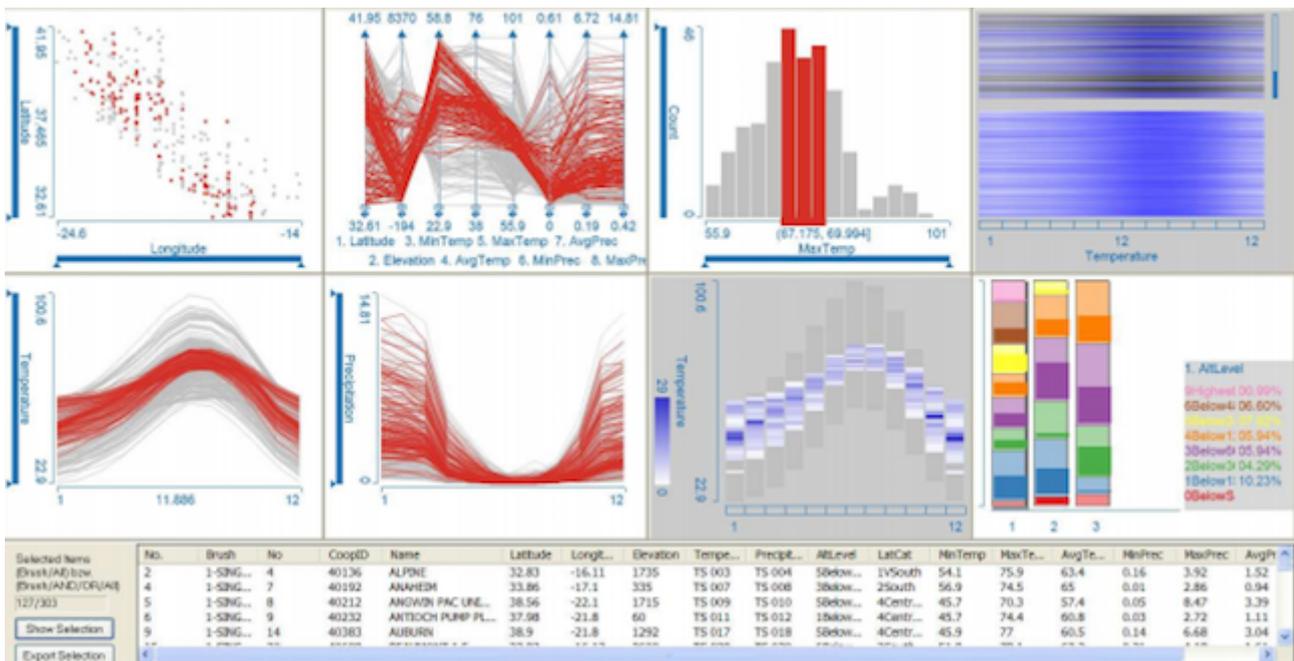


Figura 2.11 – *Coordinated Multiple Views* exibindo 8 visualizações distintas e integradas [73].

visão geral de todo o contexto. Possui métodos de refinamento sob demanda, nos quais o usuário solicita maior detalhe e refinamento progressivo [90], como mostra a Figura 2.12, na qual é selecionada uma região específica e automaticamente são exibidas mais informações da região selecionada.

É um modelo de interação que fornece mais detalhes sobre o conjunto de dados somente depois que o usuário as solicitou. Normalmente, isso permite que o usuário tenha uma visão geral melhor do conjunto de dados, devido à quantidade reduzida de informações, ao mesmo tempo que compreenda os “detalhes sob demanda”. Essa técnica fornece informações adicionais, sem exigir uma mudança de visão. Isso pode ser útil para relacionar as informações detalhadas ao resto do conjunto de dados ou para resolver rapidamente determinadas tarefas, como identificar um elemento de dados específico entre muitos, ou relacionar atributos de dois ou mais pontos de dados.

2.6 Algoritmos de análise preditiva

Análise preditiva é a combinação do uso de dados, algoritmos de estatísticas e técnicas de aprendizado de máquina para identificar a probabilidade de resultados futuros baseados em dados históricos. Os modelos preditivos usam os resultados conhecidos para prever valores para diferentes ou novos dados. A modelagem resulta em previsões que representam uma probabilidade de possíveis valores de dados. A análise preditiva é usada para prever tendências, melhorar o desempenho, orientar a tomada de decisões e prever o

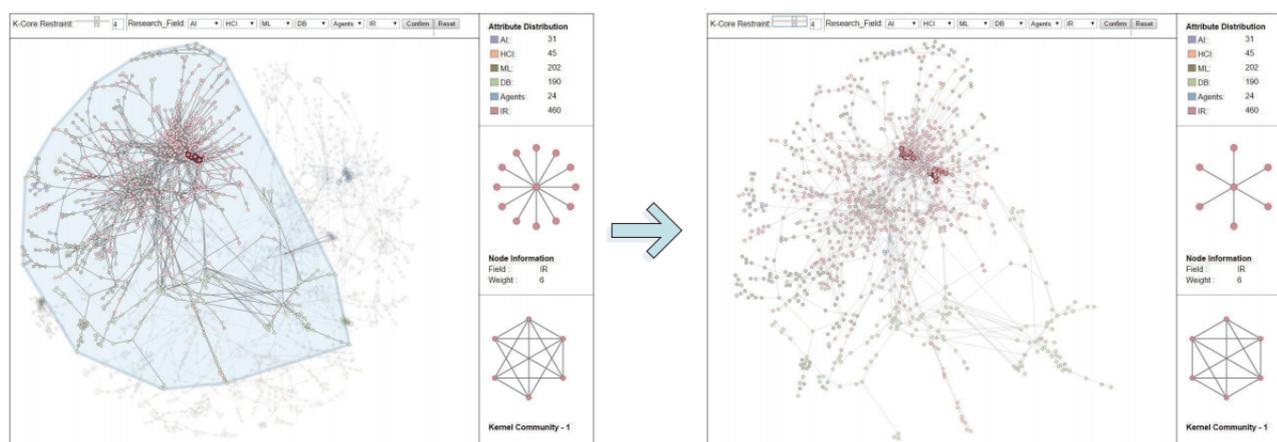


Figura 2.12 – Exemplo de uso do *Details on demand*, onde foi refinado o campo visual para obter maiores detalhes dos dados das subredes selecionadas [36].

comportamento [12]. A principal vantagem é a possibilidade de deixar de tomar decisões baseadas na intuição, para tomar decisões com base em algoritmos estatísticos e dados históricos.

Para utilizar um algoritmo de análise preditiva alguns critérios devem ser analisados, tais como: disponibilidade dos dados, tipo de informação, estrutura dos dados, contexto dos dados, tamanho do conjunto de dados, definição de métricas de desempenho, entre outros. Dado este contexto, o primeiro passo foi identificar qual tipo de informação é fornecido pelos conjuntos de dados utilizados neste estudo. Mais informações sobre estes conjuntos podem ser vistos na Seção 5.1.

O tipo dos conjuntos de dados utilizados é dados de séries temporais, que é caracterizado por uma sequência de observações feitas em um período de tempo, como apresentado na Seção 2.1.1. A análise de séries temporais permite descrever a dependência dos dados em relação ao tempo e prever valores futuros [14].

A previsão de uma série temporal é o estabelecimento dos valores futuros da série, sendo uma previsão da estimativa acerca da verossimilhança de eventos futuros, baseados na informação atual e histórica. Pressupõe a modelagem matemática do fenômeno, obtenção de conclusões e avaliação do modelo em termos de precisão. Neste estudo, foram abordados quatro diferentes modelos de algoritmo de análise preditiva, que são: *Autoregressive*, *Holt-Winters* e *Sarima*. Estes modelos são descritos a seguir.

2.6.1 *Autoregressive*

Um modelo *Autoregressive*, ou autoregressivo (AR), prevê um comportamento futuro baseado no comportamento passado [28]. É usado para previsão quando existe alguma correlação entre os valores em uma série temporal e os valores que os precedem e

sucedem. Ele tem sido utilizado em muitas aplicações de diversas áreas de pesquisa. Sua configuração é baseada no uso de dados observados no passado para prever o próximo valor [28].

A modelagem AR é geralmente definida por três etapas. A primeira é a identificação dos valores que o modelo utilizará, visando determinar a ordem do modelo AR. O segundo estágio é a estimativa do modelo, estimando os coeficientes de AR. E o último estágio é geralmente examinar a adequação do modelo estimado [113].

2.6.2 Holt-Winters

O método de *Holt-Winters* é um algoritmo de análise preditiva para uso exclusivo com dados de séries temporais [55]. Ele já foi aplicado em redes inteligentes para prever a demanda de eletricidade [100]. Também tem sido extensivamente usado para vários outros problemas de pesquisa para fazer previsões, como, por exemplo, para vendas da empresa e tráfego de rede [100, 103].

O método de *Holt-Winters* é a extensão dos métodos de *Holt* [50] e *Winters* [121] para identificar sazonalidade, que é responsável pelas variações em uma série temporal e representa flutuações intra-anuais mais ou menos estáveis ano após ano em relação ao tempo, direção e magnitude. Neste método estendido, as previsões são feitas usando outras variáveis como tendência e a própria sazonalidade. O método tem duas variantes, que são métodos aditivos e multiplicativos. Eles diferem com base na variação de componentes sazonais. Quando as variações sazonais são aproximadamente constantes, o método aditivo é usado. O método multiplicativo é preferencialmente usado quando as variações sazonais estão mudando proporcionalmente ao nível da série.

2.6.3 Sarima

ARIMA é um método de previsão para dados de série temporal univariados. Como o próprio nome sugere, ele suporta elementos autorregressivos. Um problema com o ARIMA é que ele não suporta dados sazonais. Essa é uma série temporal com um ciclo de repetição. O ARIMA espera que os dados não sejam sazonais ou que tenham o componente sazonal removido, por exemplo, ajustados sazonalmente por meio de métodos como a diferenciação sazonal [29].

A média móvel integrada autoregressiva sazonal, SARIMA ou ARIMA Sazonal, é uma extensão do ARIMA (média móvel integrada autoregressiva) que suporta explicitamente dados de séries temporais univariadas com um componente sazonal. É o modelo

linear mais popular para previsão de séries temporais sazonais que alcançou grande sucesso tanto em pesquisas acadêmicas, quanto em aplicações industriais durante as últimas três décadas [29].

2.6.4 *Prophet*

O *Prophet* é um algoritmo de análise preditiva disponível nas linguagens *Python* e *R* para previsão de dados de séries temporais. Ele foi publicado pela equipe principal de ciência de dados do *Facebook* sendo usado para analisar a falta de dados, capturar as mudanças nas tendências e identificar grandes *outliers*. Além disso, oferece uma estimativa razoável dos dados sem precisar de esforço manual [104].

O *Prophet* possui sua própria estrutura de dados especialmente para lidar com séries temporais e com sazonalidades. Ela possui duas colunas básicas: Uma colunas é denominada de 'ds' que armazena séries temporais de datas. A outra coluna é 'y' e armazena os valores correspondentes das séries temporais. Assim, a estrutura fornecida funciona bem em séries temporais e fornece algumas opções para lidar com a sazonalidade anual, semanal e diária. Devido ao fornecimento dessas opções, um analista de dados pode escolher a granularidade do tempo disponível para o modelo de previsão no conjunto de dados [104].

3. TRABALHOS RELACIONADOS

Neste capítulo são apresentados os trabalhos relacionados com este estudo. Inicialmente, é explicada a estratégia de pesquisa, quais foram as bases de dados utilizadas, as definições das palavras-chaves, e os critérios de inclusão e exclusão que levaram à seleção de 29 artigos. Estes artigos foram selecionados através de palavras chaves e leitura do resumo, e posteriormente classificados e agrupados considerando suas características da seguinte maneira: Edifícios Inteligentes, Detecção de anomalias, Eficiência Energética, Redes Inteligentes e Residências Inteligentes. Além disso, também foram exploradas algumas ferramentas relacionadas com a temática deste estudo e que podem auxiliar na visualização de dados de energia. As seções a seguir apresentam a metodologia de pesquisa, os artigos e ferramentas selecionados, identificando suas funcionalidades, vantagens e desvantagens.

3.1 Metodologia de pesquisa

O primeiro passo para realização deste estudo foi a definição de um escopo adequado que inicialmente foi configurado para abranger trabalhos da área de *Visual Analytics* relacionados com o tema de consumo de energia. Isso levou à definição de questões preliminares de pesquisa e à identificação das palavras-chave que seriam usadas. Dessa forma, ficou estabelecido que seriam usadas as seguintes palavras-chaves para a busca de trabalhos relacionados: ***Visual Analytics, Energy Consumption***.

Estas palavras-chaves foram usadas para busca de artigos nas seguintes bases de dados e bibliotecas digitais: IEEE ¹, ACM ², Elsevier ³, Springer ⁴, Research Gate ⁵ e o portal Scopus ⁶. Com base nos resultados obtidos em uma busca inicial, o escopo e as palavras-chave foram aprimoradas e as buscas foram reconduzidas. Além disso, foram definidas as regras de inclusão e exclusão apresentadas a seguir, que foram aplicadas considerando o título e o resumo dos artigos.

Critérios de inclusão:

- Ter o texto completo disponível na Internet;
- Apresentar as palavras-chaves de busca no título ou no resumo;
- Ter sido publicado após o ano de 2000;

¹<https://ieeexplore.ieee.org/>

²<https://dl.acm.org/>

³<https://www.sciencedirect.com/>

⁴<https://link.springer.com/>

⁵<https://www.researchgate.net/>

⁶<https://www.scopus.com/>

- Estar escrito em português ou inglês;
- Possuir mais de 4 páginas.

Critérios de exclusão:

- Não estar disponível na Internet, ou seja, nas bibliotecas digitais e portais que se tem acesso;
- Não abordar o tema da pesquisa;
- Não apresentar uma técnica de visualização de dados.
- Não abordar ao menos um dos seguintes temas de pesquisa: *Dashboard*, análise visual interativa e análises preditivas.

O resultado deste processo levou a seleção de um conjunto de artigos, a partir dos quais foi aplicado o método de pesquisa de *Snowballing*. Esta metodologia é bastante utilizada para encontrar artigos que não são retornados por um processo de busca automático, mas que podem ser interessantes para a realização da pesquisa. Greenhalgh e Peacock [46] afirmam que utilizaram três diferentes métodos de busca em suas pesquisas e concluíram que as abordagens de busca por protocolo não são necessariamente mais eficientes, pois algumas fontes são encontradas pelo próprio conhecimento do pesquisador. Por isso, *Snowballing* é considerada uma abordagem muito eficaz.

A primeira etapa do *Snowballing* consiste na seleção de um conjunto inicial de artigos. Estes artigos podem ser selecionados através do conhecimento do pesquisador, ou identificados a partir de buscas em algumas bases de dados. A partir dessa seleção inicial, conforme descrito por Wholin [122], existem dois subprocessos para serem realizados, que são denominados como *Forward* e *Backward*. A etapa de *Forward* começa a partir do conjunto inicial e consiste em localizar novos artigos que os citem, realizando várias iterações, até que não sejam localizados novos artigos. Desta forma, são identificados artigos que citaram os artigos aceitos na iteração anterior, devendo sempre ser realizado de acordo com critérios de inclusão e exclusão. Da mesma forma que o *Forward*, o subprocesso *Backward* também consiste em iterações, porém, nesta etapa são realizadas iterações a partir das referências mencionadas nos artigos, também respeitando critérios de inclusão e exclusão. O processo deve ser repetido inúmeras vezes, até que não sejam localizados artigos que possam ser incluídos na pesquisa. As buscas realizadas inicialmente, considerando as palavras-chaves e os critérios de inclusão e exclusão apresentados anteriormente, resultaram na seleção de 13 artigos, que estão listados na Tabela 3.1.

A partir destes artigos foram aplicados os subprocessos de *Forward* e *Backward*. A etapa de *Forward* foi realizada em dois ciclos: no primeiro, foram analisados 348 artigos e foram aceitos 13; no segundo, foram analisados 41 artigos e nenhum foi aceito. Os

ID	Artigo	Biblioteca	Ano
P1	Chin et al. [30]; A visual Analytics platform and advanced visualization tool for interpreting and analyzing wind energy time-series data.	Elsevier	2018
P2	Janetzko et al. [53]; Anomaly detection for visual analytics of power consumption data.	Elsevier	2014
P3	L. Bartram [13]; Design Challenges and for Eco-Feedback in the Home.	IEEE	2015
P4	Jui-Sheng et al. [31]; Early-warning application for real-time detection of energy consumption anomalies in buildings.	Elsevier	2017
P5	S. Timm e B. Deal [107]; Effective or ephemeral? The role of energy information dashboards in changing occupant energy behaviors.	Elsevier	2016
P6	Blanco et al. [16]; Energy analytics in public buildings using interactive histograms.	Elsevier	2017
P7	Xiong et al. [126]; Model-based shading and lighting controls considering visual comfort and energy use.	Elsevier	2016
P8	Agarwal et al. [1]; The Energy Dashboard: Improving the Visibility of Energy Consumption at a Campus-Wide Scale.	ACM	2009
P9	Pérez et al. [89]; Power-Consumption Analysis through Web-Based Visual Data Exploration.	Elsevier	2014
P10	Nilsson et al. [80]; Smart homes, home energy management systems and real-time feedback.	Elsevier	2018
P11	Prouzeau et al. [87]; Visual Analytics for Energy Monitoring in the context of Building Management.	IEEE	2018
P12	Goodwin et al. [45]; Creative User-Centered Visualization Design for Energy Analysts and Modelers.	IEEE	2013
P13	Brehmer et al. [19]; Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis.	IEEE	2016

Tabela 3.1 – Conjunto inicial de artigos selecionados.

13 artigos aceitos nesta etapa estão apresentados na Tabela 3.2. A etapa de *Backward* também foi realizada em dois ciclos: no primeiro, foram analisados 454 artigos e foram aceitos 7; no segundo, foram analisados 169 artigos e nenhum foi aceito. Dessa forma, o *Snowballing Backward* resultou nos 7 artigos apresentados na Tabela 3.3.

A soma dos 13 artigos aceitos pelo subprocesso *Snowballing Forward*, com os 7 artigos identificados pelo subprocesso *Snowballing Backward*, mais os 13 artigos iniciais, resultou em 33 artigos relacionados ao tema de pesquisa que foram lidos na íntegra. Durante a leitura do texto de cada artigo, foram aplicados novamente os critérios de exclusão apresentados anteriormente. Assim, foram removidos 3 artigos do ciclo de *Forward* e 1 artigo do ciclo de *Backward* pelo fato deles se encaixarem em um dos critério de exclusão que foram apresentados na Seção 3.1, resultando em 29 artigos selecionados. Os artigos apresentados na Tabela 3.4 foram removidos desta pesquisa porque não abordaram ao menos um dos seguintes itens: *Dashboard*, análise visual interativa e análises preditivas.

ID	Título do artigo	Biblioteca	Ano
F1	Arjunan et al. [9]; Multi-User Energy Consumption Monitoring and Anomaly Detection with Partial Context Information.	ACM	2015
F2	Arbesse et al. [8]; Visplause: Visual Data Quality Assessment of Many Time Series Using Plausibility Checks.	IEEE	2016
F3	Pickering et al. [84]; Building electricity consumption: Data analytics of building operations with classical time series decomposition and case based subsetting.	Elsevier	2018
F4	Cui et al. [34]; Anomaly Detection System for School Electricity Consumption Data.	IEEE	2017
F5	Iyengar et al. [52]; Watt Home: A Data-driven Approach for Energy Efficiency Analytics at City-scale.	ACM	2018
F6	Berry et al. [15]; Finding faults and influencing consumption: the role of in-home energy feedback displays in managing high-tech homes.	Springer	2016
F7	Garcia et al. [44]; Interactive visualization for NILM in large buildings using non-negative matrix factorization.	Elsevier	2018
F8	Ioannidis et al. [51]; Occupancy driven building performance assessment.	Elsevier	2016
F9	Wood et al. [125]; Sensors, sense-making and sensitivities: UK household experiences with a feedback display on energy consumption and indoor environmental conditions.	Elsevier	2019
F10	N. Ahmed e K. Mueller [2]; EnergyScout: A Consumer Oriented Dashboard for Smart Meter Data Analytic.	Research Gate	2019
F11	Çakmak et al. [129]; A New Framework for the Analysis of Large Scale Multi-Rate Power Data.	Research Gate	2014
F12	Li Sun et al. [102]; Regional difference of household electricity consumption: An empirical study of Jiangsu, China.	Elsevier	2018
F13	N. Sisworahardjo and A. Saad; [101]; Spatio-Temporal Context Anomaly Detection for Residential Power Consumption.	Research Gate	2017

Tabela 3.2 – Artigos do *Snowballing Forward*

3.2 Edifícios inteligentes

Um edifício inteligente é aquele que usa a tecnologia para compartilhar informações entre sistemas sobre o que está acontecendo no prédio, de modo a otimizar o seu desempenho. Estas informações são usadas para automatizar vários processos, desde aquecimento e ventilação até ar condicionado e segurança. Eles são compostos por vários dispositivos controlados pelo sistema de automação. Esses dispositivos podem variar

ID	Título do artigo	Biblioteca	Ano
B1	K. Ellegård e J. Palm [38]; Visualizing energy consumption activities as a tool for making everyday life more sustainable.	Elsevier	2011
B2	Wong et al. [123]; A Novel Visualization Technique for Electric Power Grid Analytics.	IEEE	2009
B3	Yarbrough et al. [127]; Visualizing building energy demand for building peak energy analysis.	Elsevier	2015
B4	A. Morán et al; [76]; Analysis of electricity consumption profiles in public buildings with dimensionality reduction.	Elsevier	2013
B5	C. Martani et al. [72]; ENERNET: Studying the dynamic relationship between building occupancy and energy consumption.	Elsevier	2012
B6	E. Costanza et al. [33]; Understanding domestic energy consumption through interactive visualisation: A field study.	ACM	2012
B7	M. Nielsen e K. Grønbæk. [79]; Towards Highly Affine Visualizations of Consumption Data from Buildings.	Research Gate	2015

Tabela 3.3 – Artigos do *Snowballing Backward*.

ID	Título do artigo	Biblioteca	Ano
F11	Çakmak et al. [129]; A New Framework for the Analysis of Large Scale Multi-Rate Power Data.	Research Gate	2014
F12	Li Sun et al. [102]; Regional difference of household electricity consumption: An empirical study of Jiangsu, China.	Elsevier	2018
F13	N. Sisworahardjo and A. Saad. [101]; Spatio-Temporal Context Anomaly Detection for Residential Power Consumption.	Research Gate	2017
B7	M. Nielsen e K. Grønbæk. [79]; Towards Highly Affine Visualizations of Consumption Data from Buildings.	Research Gate	2015

Tabela 3.4 – Artigos removidos desta pesquisa.

de dispositivos simples, como *gadgets*, sensores, luzes e outros, a dispositivos elétricos inteligentes [82].

As despesas decorrentes do uso de uma casa ou edifício podem gerar custos bem significativos para seus proprietários, principalmente se houver desperdício. Isto ocorre, por exemplo, no caso das luzes estarem acesas em salas não utilizadas, ou no caso de haverem espaços aquecidos quando não há pessoas ao redor para desfrutar do calor. Logo, se houverem sensores instalados que detectam a presença de pessoas, automaticamente elas identificariam o momento exato de acender as luzes ou ligar aquecedores. Com esse gerenciamento, geraria uma economia de energia quando não haver presença de pessoas e diminuição de custos financeiros. A principal motivação por trás do edifício inteligente é evitar esse desperdício de energia e recursos, tanto para cortar custos, quanto para melho-

rar a eficiência energética. Foram encontrados vários artigos na literatura que abordam o uso de visualização de dados para edifícios inteligentes, conforme apresentado a seguir.

O artigo P10 de Nilsson et al. [80] teve como objetivo otimizar o uso de energia em edifícios inteligentes. Foram apresentados dados coletados de quatro edifícios selecionados, desde residências a centros de dados, conforme mostrado na Figura 3.1, que apresenta dados de energia destes edifícios para uma semana em agosto de 2009. Foi utilizada a unidade de *Watts/square feet* como a métrica de comparação, pois os edifícios são de tamanhos diferentes.

Em particular, o artigo concentrou-se em edifícios de uso misto, nos quais o consumo de energia de equipamentos de TI são responsáveis por mais de um quarto do uso total de energia e gerou observações detalhadas que identificaram os componentes primários do uso de energia e picos no consumo de energia. Surpreendentemente, o artigo forneceu *insights* sobre como reduzir significativamente o consumo de energia. [1].

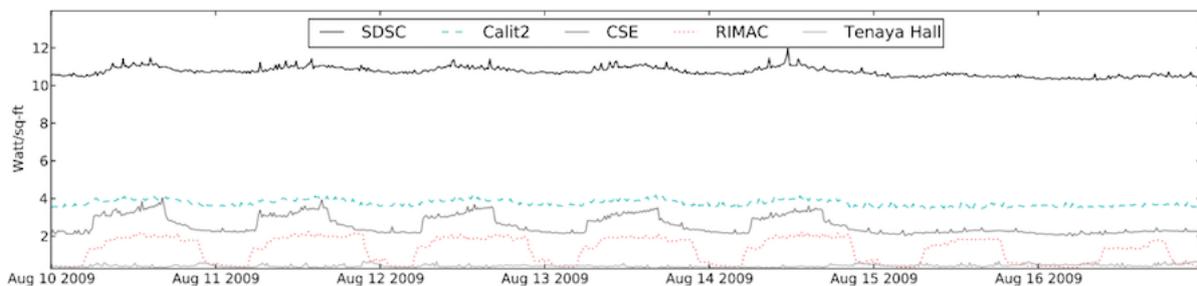


Figura 3.1 – Comparando o consumo semanal de energia de vários edifícios de acordo com suas modalidades de uso [1].

A pesquisa P13 de Brehmer et al. [19] afirma que o consumo energético de grande edifícios é desafiador para analisar e monitorar, pois as ferramentas de análise atuais não são escalonáveis. Visando solucionar este problema, foi realizado um estudo de *design* de visualização, começando com uma análise completa do domínio do estudo e uma caracterização de abstrações de dados e tarefas. A principal contribuição do artigo foi oferecer um conjunto de opções e diretrizes de *design* generalizáveis. Elas foram enquadradas em termos de correspondência, incompatibilidades entre abstrações e idiomas de codificação visual para dados orientados ao tempo, diretrizes que são transferidas além do domínio da energia.

O artigo B3 de Yarbrough et al. [127] teve como objetivo identificar a redução dos custos de demanda de picos de energias em um campus universitário, no qual foi investigado o uso de energia ao longo do dia. Para facilitar essa avaliação, foi desenvolvida uma ferramenta de análise visual, permitindo comparações de edifícios dentro campus.

Uma técnica de visualização de dados utilizada neste trabalho foi o *Heatmap*, como ilustra a Figura 3.2, na qual é mostrada a demanda diária de energia do edifício em um mês. No entanto, esta ferramenta de visualização é útil apenas para investigar a

demanda de energia em um prédio e não para comparar padrões de energia de vários edifícios, especialmente ao longo de um ano inteiro.

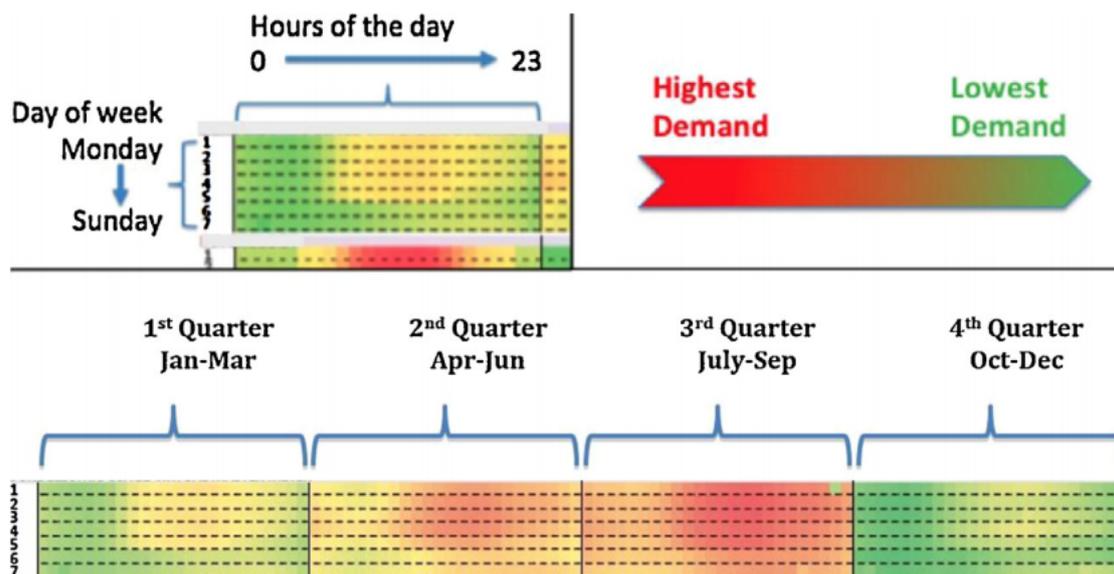


Figura 3.2 – Visualização do mapa de calor da demanda de energia de um edifício do campus. As colunas representam horas do dia, da meia-noite até 23h. As linhas representam os dias da semana, com os finais de semana na parte inferior. Cada quadrante em uma série de mapas de calor representa o ano em segmentos de três meses. Cada ponto corresponde à média de todos os dias desse período, em média por hora do dia e dia da semana [127].

A pesquisa P5 de Timm e Deal [107] abordou o uso de energia em tempo real para analisar os comportamentos de estudantes e docentes em relação ao uso de energia através de um painel de energia online, como ilustra a Figura 3.3. As descobertas realizadas mostram que, embora o estudo resultou em economia significativa de energia, as diferenças nas atitudes ou comportamentos de estudantes e docentes se mostraram insignificantes.



Figura 3.3 – Painel online exibindo o consumo dos edifícios do campus universitário [107].

O artigo P6 de Blanco et al. [16] propôs uma abordagem de análise visual baseada em cubos de dados para fornecer uma análise perspicaz de como a energia está sendo

usada em um grupo de edifícios públicos. A análise foi feita por meio de uma interface visual baseada na web. Nesta interface, havia visualizações coordenadas, de acordo com diferentes atributos, sob diferentes cenários. Eles eram definidos por filtros direcionados pelo usuário sobre esses atributos, como ilustrado na Figura 3.4.

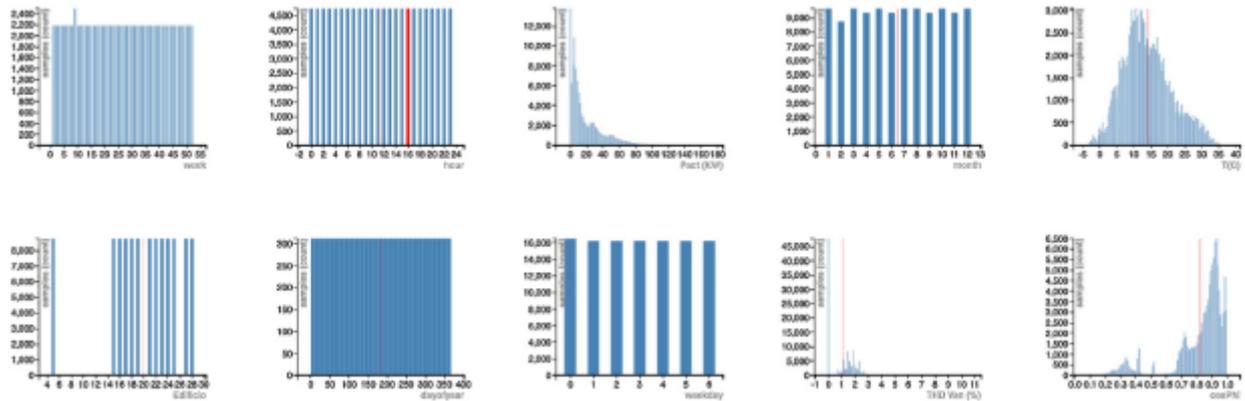


Figura 3.4 – Histogramas exibindo dados de consumo de energia de edifícios públicos [16].

O artigo P11 de Prouzeau et al. [87] descreve um protótipo de interface para um sistema de gerenciamento de aquecimento, ventilação e ar condicionado (HVAC), que são os tipos mais comuns de sistemas de conforto ambiental em edifícios [69]. Este sistema fornece visualizações interativas de dados como temperatura e consumo de energia, além de dados sobre feedback de conforto dos ocupantes e ocupação, conforme ilustra a Figura 3.5. Este protótipo foi avaliado para mostrar como ele poderia ser usado para planejar uma estratégia de economia de energia a longo prazo, e também com possíveis usuários, gerentes de instalações em uma universidade. Como resultado, observou-se que ele poderia ser usado para planejar uma estratégia de economia de energia a longo prazo.

O foco do artigo F8 de Ioannidis et al. [51] foi a avaliação de desempenho de um edifício usando técnicas de *Big Data* e análise visual impulsionadas pela ocupação do edifício. Neste artigo, várias técnicas de análise visual são exploradas, considerando as cognição humana, percepção e tomada de decisão. A ferramenta, apresentada na Figura 3.6, foi testada em informações de dados reais adquiridas de um edifício localizado no sul da Europa, demonstrando a sua eficácia e a sua usabilidade para os gestores de edifícios.

A pesquisa F5 de Iyengar et al. [52] apresentou uma ferramenta para identificar os edifícios menos eficientes de um grande número de edifícios em uma cidade ou região. Porém, este artigo não apresentou muitas técnicas de visualização de dados. Os artigos P2 [53] e P4 [31] abordam o tema de Edifícios inteligentes, mas também foram categorizados como Detecção de anomalias. Por isso, mais informações sobre estes trabalhos são apresentadas na Seção 3.3.

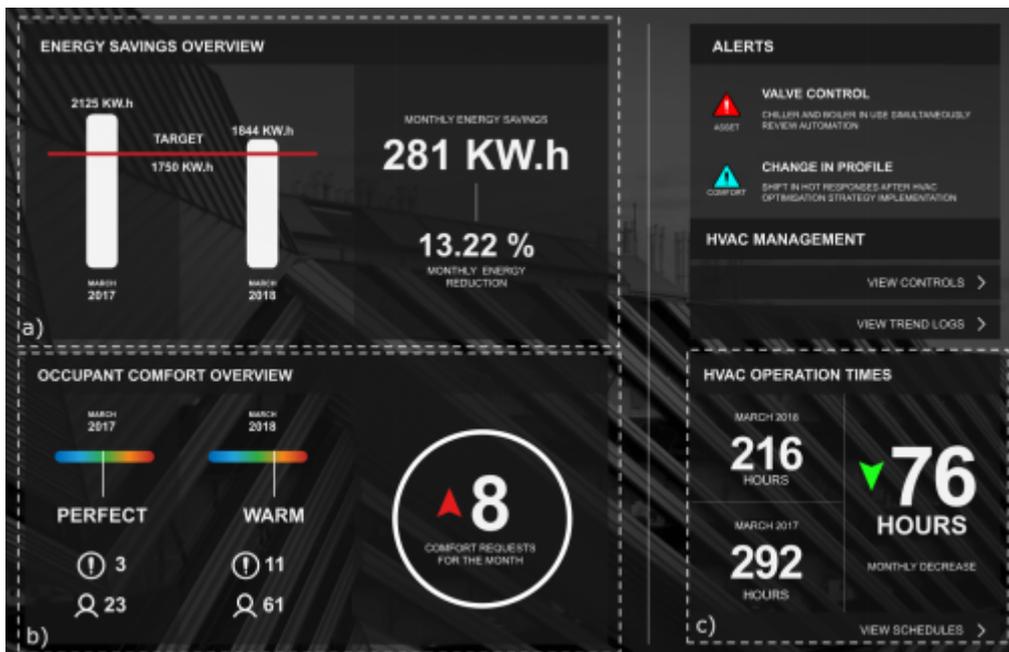


Figura 3.5 – Tela do protótipo, fornecendo uma visão geral do consumo de energia de uma sala (a), o nível de conforto (b) e os tempos de operação do HVAC (c) [87].

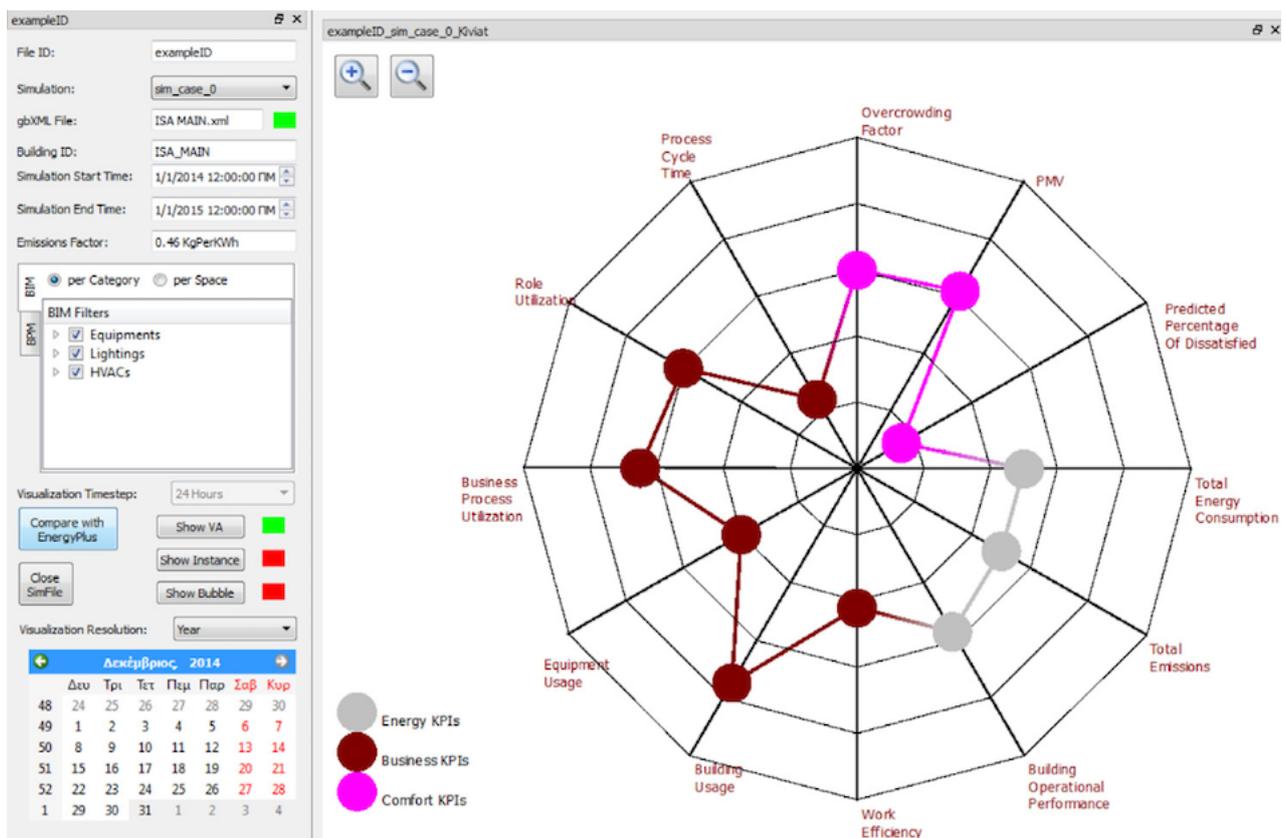


Figura 3.6 – Técnicas de análise visual mostrando dados de consumo de energia de três formas diferentes: (a) Consumo total de energia (cinza claro), (b) Consumo de energia no edifício e aparelhos elétricos (vermelho) e (c) conforto de ocupação (magenta) [51].

3.3 Detecção de anomalias

Anomalias são padrões ou eventos em um conjunto de dados que não correspondem a um conceito bem definido de normalidade [26]. São instâncias que não seguem um comportamento padrão esperado e podem estar contidas nos dados por motivos distintos, dependendo do uso que está sendo aplicado.

Neste contexto, o objetivo do trabalho P2 de Janetzko et al. [53] foi permitir que um analista compreenda o comportamento do consumo de energia em um edifício e que consiga identificar valores inesperados de consumo de energia [53]. O artigo introduziu um novo algoritmo de detecção de anomalia não supervisionada e foram apresentadas diferentes possibilidades de visualização dos dados das anomalias, visando auxiliar o analista, como, por exemplo, a exibição do tempo de duração de uma anomalia ilustrado na Figura 3.7.

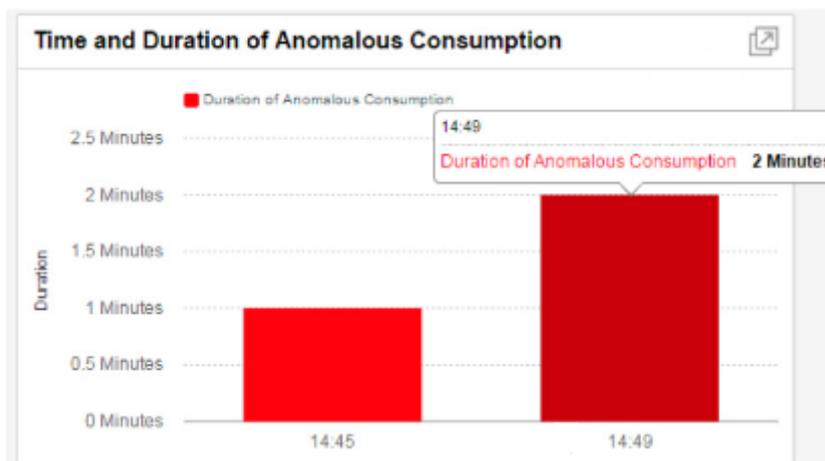


Figura 3.7 – Apresentação visual do tempo de duração de uma anomalia [53].

A principal contribuição do artigo P4 de Jui-Sheng et al. [31] foi a criação de uma arquitetura de aplicação de alerta antecipado que apresenta visualmente o consumo anômalo de energia em tempo real dentro de um edifício. Neste caso, foram usados dados obtidos de vários instrumentos (medidores e sensores inteligentes) para os ocupantes de um escritório [31]. O consumo anômalo do painel foi projetado para garantir que ocupantes do escritório com habilidades técnicas limitadas entendam os dados de consumo de energia apresentados, como ilustrado na Figura 3.8.

Detecção de anomalias também foi abordado nos artigos F1 [9] e F4 [34]. O artigo F1 de Arjunan et al. [9] apresentou um sistema para monitorar o consumo de energia de vários usuários em uma vizinhança e um novo algoritmo para detectar anomalias combinando dados de múltiplos usuários. Uma das técnicas de visualização utilizada neste sistema foi o *Dense Pixel Display*, como exibido na Figura 3.9, que mostra várias instâncias de anomalias no bairro analisado.

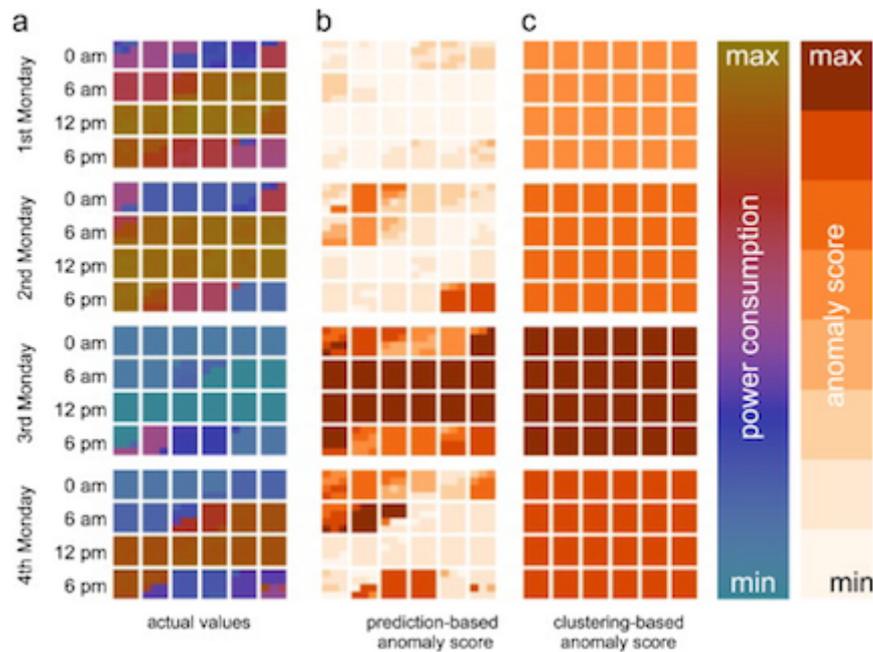


Figura 3.8 – Comparação dos escores de anomalia resultantes com base nos métodos propostos. A terceira segunda-feira mostra um comportamento incomum sendo refletido nas pontuações de anomalia [31].

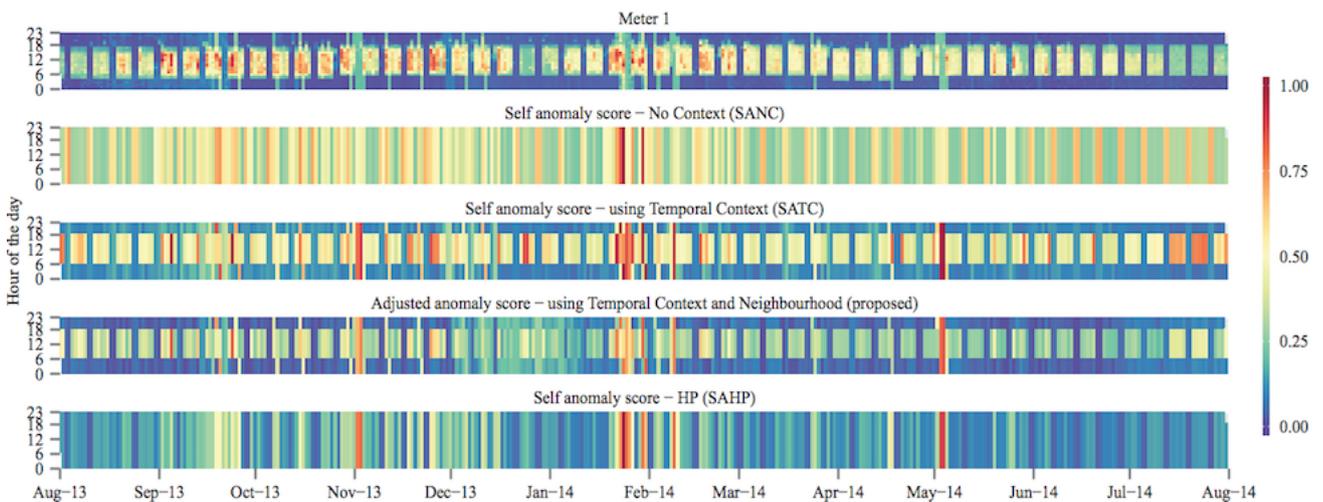


Figura 3.9 – Exibição do consumo energético de uma vizinhança com valores de anomalia calculados por diferentes métodos de detecção de anomalias. O gráfico mostra várias instâncias de anomalias e como o valor de anomalia computado difere usando as informações temporais e de vizinhança [9].

O artigo F4 de Cui et al. [34] concentrou-se no gerenciamento de instalações remotas que identificam eventos anômalos em edifícios, visando detectar anomalias de dados de energia. Assim como os artigos anteriores, utilizou diversas técnicas de visualização de dados para exibir o consumo energético e dados de anomalias, entre elas está o *Heatmap*, mostrado na Figura 3.10, que exibe dados de consumo de eletricidade semanal de uma escola.

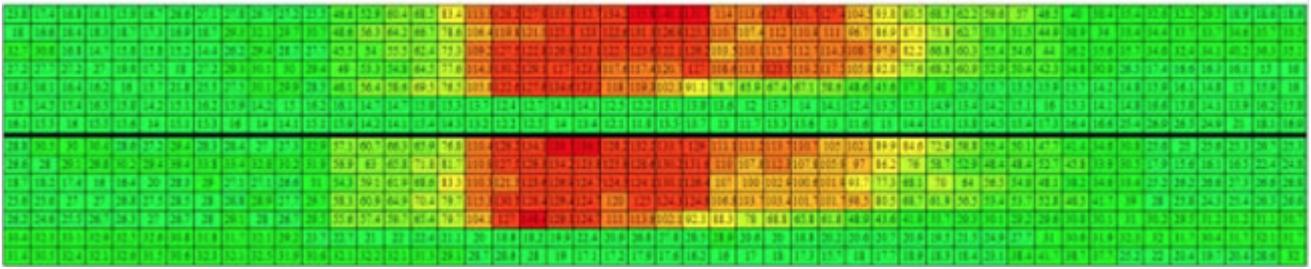


Figura 3.10 – Exibição da demanda de consumo semanal em um *Heatmap* [9].

3.4 Eficiência energética

Eficiência energética consiste na forma com que um serviço ou atividade é feito, utilizando a menor quantidade de energia possível e preocupando-se com a relação entre a quantidade de energia empregada e a disponibilizada para sua realização. Pode ser implementada em edifícios residenciais e comerciais, buscando obter mais aproveitamento da luz e ventilação natural, para evitar gastos com iluminação. Além da obtenção de um serviço com baixo gasto de energia, tem por objetivo propiciar um uso otimizado oferecendo conforto térmico, visual e acústico aos usuários [68].

Um artigo que abordou o tema de eficiência energética foi o F3 de Pickering et al. [84]. Eles analisaram seis edifícios comerciais e identificaram várias de suas características, incluindo o potencial de economia de mais de 700MWh por ano, avaliado em US \$92.000, a partir de um reescalonamento. Com acesso apenas a dados de medidores inteligentes de todo o edifício, esses resultados são obtidos de forma online e instantânea, como ilustrado na Figura 3.11. Isso justifica uma abordagem rigorosa de análise de dados para mostrar o potencial de eficiência energética de edifícios [84].

O artigo P7 de Xiong et al. [126] apresentou o desenvolvimento e implementação de algoritmos de controle visando minimizar o uso de energia de iluminação. Neste trabalho foram usados apenas gráficos de barras, linhas e plot para visualizar as informações de energia solar.

O trabalho P3 de L. Bartram [13] apresenta uma série de questões relacionadas à exibição efetiva de dados de uso doméstico de energia. Também analisa as abordagens de feedback para o usuário doméstico e detalha alguns dos problemas que impedem a eficácia dessas abordagens, enquadrando a discussão em torno das questões maiores sobre como a tecnologia e os dados se encaixam na prática doméstica. O artigo apresenta um painel de consumo de energia que utiliza algumas técnicas de visualização como gráficos de barras, linhas e pizza, em uma interface visual integrada com *Dashboard*, como mostra a Figura 3.12.

O artigo B5 de Martani et al. [72] visou criar um novo método para medir a atividade de ocupação humana, através do uso de conexões *Wi-Fi*. Os dados coletados de consumo

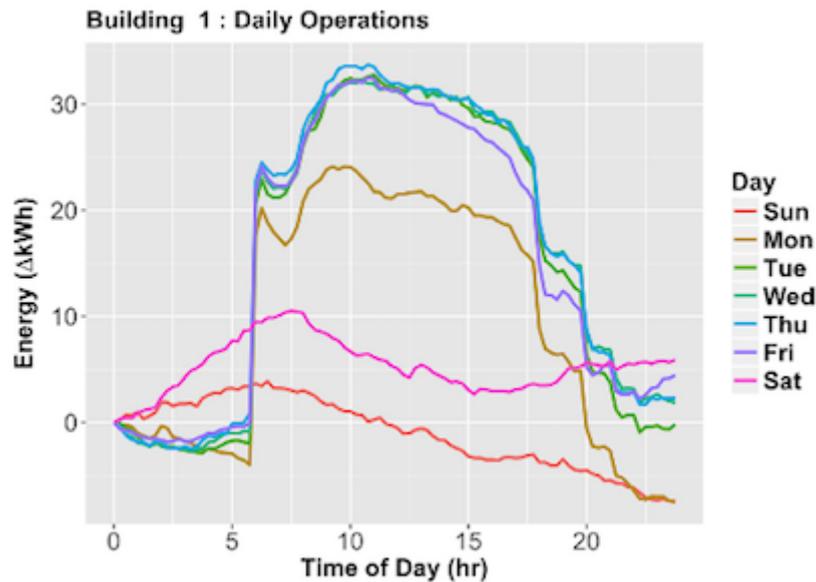


Figura 3.11 – Operação diária em um edifício: amostragem para vários dias da semana, usando dois anos de dados [84].

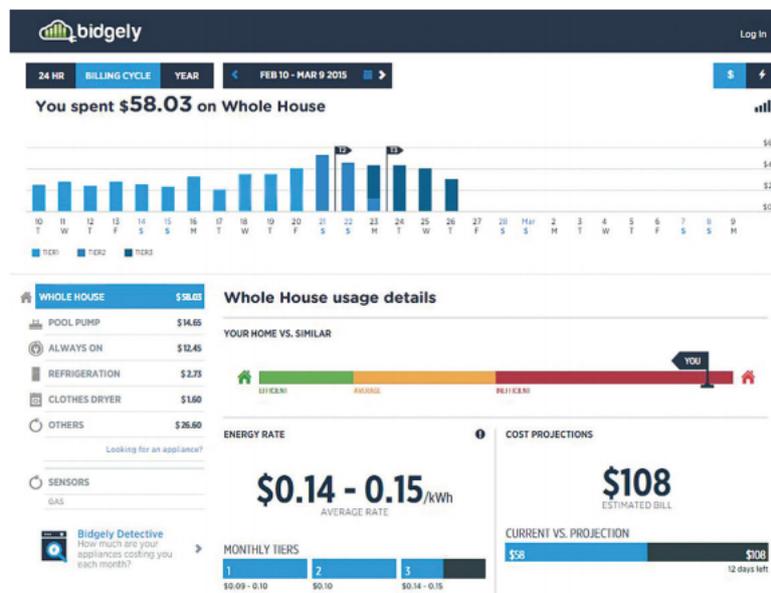


Figura 3.12 – Painel de consumo de energia detalhando o uso de energia em casa, bem como recomendações para reduzir o consumo [13].

de energia foram comparados em dois edifícios dentro do campus do *Massachusetts Institute of Technology*. Os autores mostraram os dados de consumo de energia coletados com diferentes técnicas de séries temporais. Um destas técnicas está ilustrada na Figura 3.13, que exibe o número de conexões *WiFi* em um andar de um edifício.

A partir da identificação da taxa de ocupação das salas, combinada com outros fatores como temperatura externa, foi possível melhorar o funcionamento dos sistemas de aquecimento, ventilação e ar condicionado (HVAC). Outro destaque foi que uma pequena parte dos níveis de consumo de eletricidade se correlacionou com a ocupação.

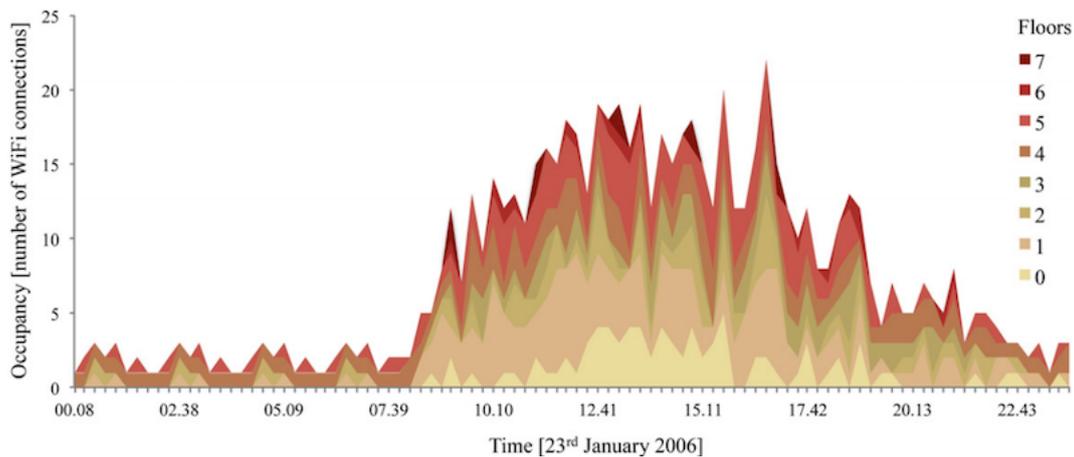


Figura 3.13 – O número de conexões *Wi-Fi* (como uma estimativa da ocupação humana) estratificadas por andar [72].

Integrar o conhecimento dos usuários por meio de um aplicativo web com gráficos de barras, heatmaps e calendários, ilustrado na Figura 3.14, foi o objetivo do artigo F7 de Garcia et al. [44]. A abordagem proposta foi testada com dados reais de demanda de energia elétrica de um complexo hospitalar, mostrando como a interpretação destes dados foi melhorada através de técnicas de visualizações interativas.

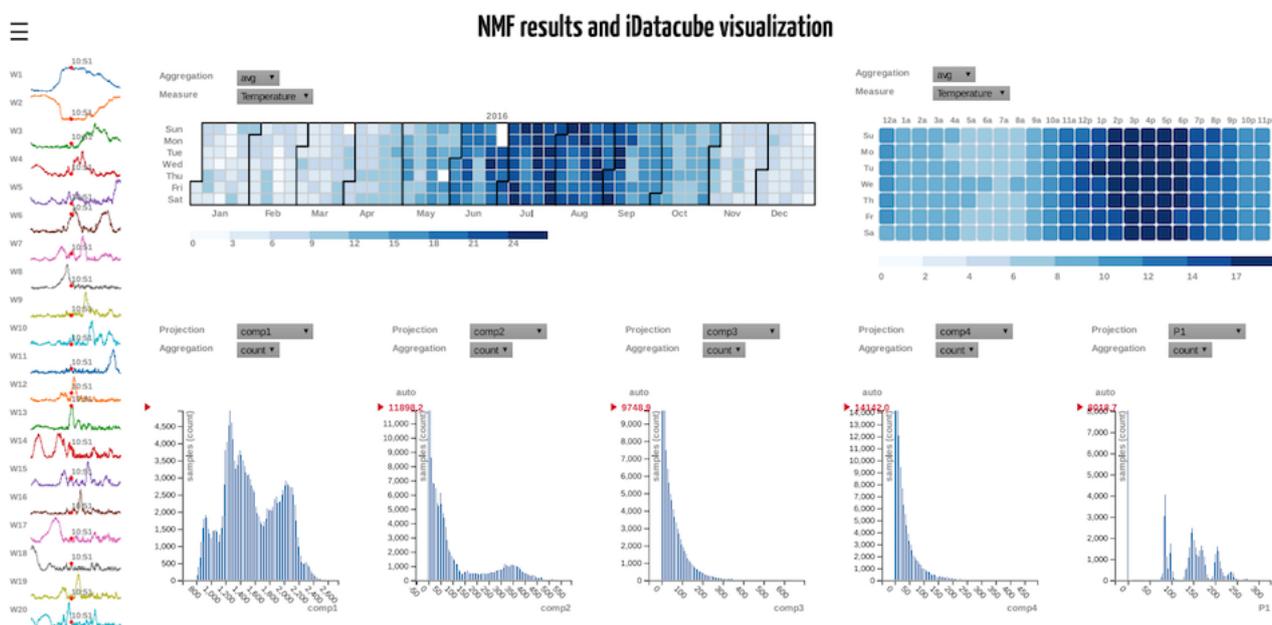


Figura 3.14 – Aplicativo web para monitoramento de energia em um complexo hospitalar, exibindo o consumo do hospital em diferentes gráficos [44].

Chin et al. [30] afirmam no artigo P1 que a energia eólica está produzindo grandes quantidades de dados em tempo real que precisam ser eficientemente monitorados, a fim de combinar e equilibrar essa fonte variável de energia com outros recursos de energia para atender às demandas. Para apoiar o monitoramento de energia eólica, desenvolveram uma plataforma de análise visual capaz de conectar-se a fontes de dados de operações eólicas

existentes para integra-los e atender eficientemente à demanda de energia. Além disso, também desenvolveram um conjunto de ferramentas avançadas de visualização para esta plataforma para suportar diretamente as necessidades de análise e operações de energia eólica, conforme mostra a Figura 3.15.

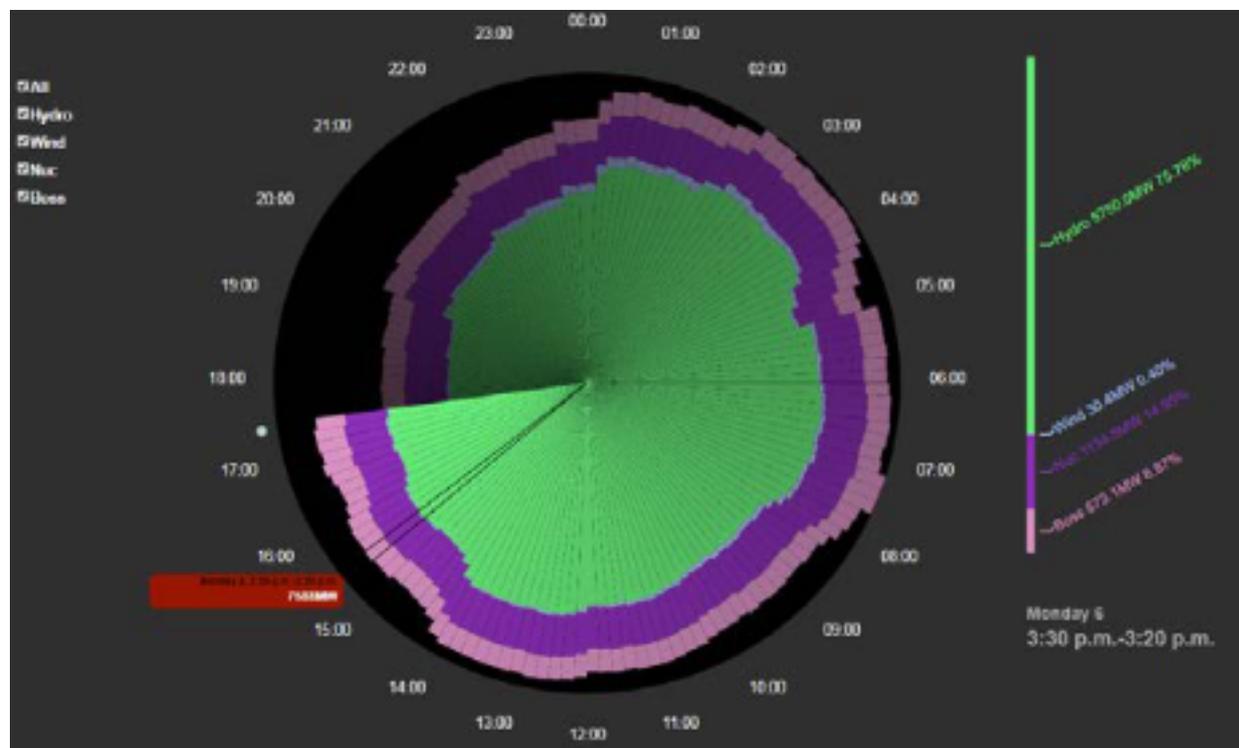


Figura 3.15 – Plataforma visual mostrando a visualização de radar de geração de energia comparativa de vários tipos de recursos energéticos. Cada fatia é dividida em vários recursos de energia, indicando quanto desse recurso contribuiu para a energia total produzida no tempo especificado. Todas as fatias juntas mostram a geração de energia nas últimas 24 horas [30].

A principal contribuição da pesquisa F2 realizada por Arbesse et al. [8] foi a criação de um sistema para suportar uma inspeção eficiente de problemas de qualidade de dados para muitas séries temporais. O foco foi utilizar meta-informações relativas à semântica de dados de séries temporais para apresentar resultados de maneira flexível. As visualizações permitem que os usuários inspecionem anomalias em detalhes e gerem hipóteses sobre possíveis causas. O sistema apresenta diferentes técnicas de visualizações de dados de séries temporais, conforme mostrado na Figura 3.16, e foi guiado por metas estabelecidas a partir de uma análise abrangente de tarefas com especialistas de domínio no setor de energia.

Os dois últimos artigos que foram identificados nesta categoria são os trabalhos F5 de Iyengar et al. [52] e B4 de Morán et al. [76]. Como o primeiro também é voltado para edifícios inteligentes, ele foi descrito na Seção 3.2. A pesquisa de Morán analisou o perfil diário de consumo elétrico de um edifício e sua correlação com fatores ambientais que permitem examinar e estimar sua demanda de eletricidade e criou uma nova abordagem



Figura 3.16 – Visão geral do sistema: (A) Exibição de marcadores por "Subclasse" e "Séries temporais"; (b) Apresentação dos números de verificações e séries temporais; (c) Amostra da porcentagem de dados afetados; (d) Exibição do consumo ao longo do tempo; (e) Amostra da severidade das indicações de verificação. (f) Exibição do filtro de verificações. (g) Apresentação das anomalias em uma usina de energia; (h, i) Visualizações das anomalias de dados para validação [8].

para fornecer uma análise detalhada e visual das correlações entre consumo e variáveis ambientais. Como os perfis de consumo podem ser caracterizados por muitos componentes, a entrada é de alta dimensionalidade. Por essa razão, foi necessário aplicar técnicas de redução de dimensionalidade que permitam a projeção desses dados em um espaço 2D facilmente interpretável, como mostra a Figura 3.17. Posteriormente, a abordagem proposta utiliza o algoritmo escolhido para analisar a influência das variáveis ambientais no consumo de eletricidade em prédios de uma unidade.

3.5 Redes Inteligentes

Segundo Bamberger et al. [11], uma rede inteligente, ou *Smart Grid*, tem como objetivo principal promover o uso de energia renovável, de modo que a indústria de energia atenda aos requisitos eficientes, flexíveis, ecológicos e condizentes com o mercado, melhorando níveis de serviço e a interação com o usuário.

Jin [54] define *Smart Grid* como uma rede elétrica moderna totalmente automatizada, qual cada usuário pode ser monitorado em tempo real, e certificar-se do fluxo bidirecional da corrente e informações em cada ponto de usinas para o cliente. Através de alguns fatores, como o uso de inteligência distribuída, comunicações de banda larga e integração do sistema de controle automático, uma rede inteligente pode garantir que as transações de mercado online em tempo real sejam efetuadas.

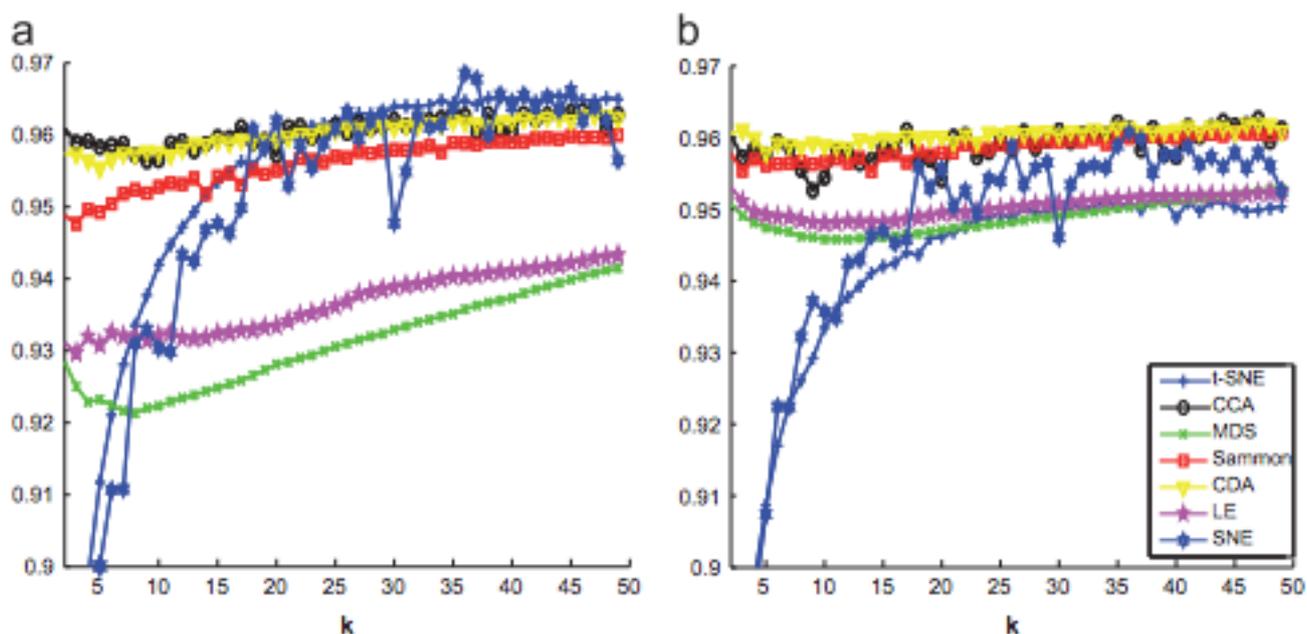


Figura 3.17 – Exibição da abordagem utilizada, onde é mostrado a avaliação das técnicas de redução de dimensionalidade. (a) Confiança. (b) Continuidade [76].

Um artigo que abordou o tema de *Smart Grid* foi o P8 de Agarwal et al. [1], cujo objetivo consistiu na análise da demanda de energia elétrica em edifícios universitários. Os resultados foram apresentados através de uma aplicação web cuja interface pode ser vista na Figura 3.18. A visualização dos dados e os mapas de similaridade são representados com técnicas interativas, permitindo uma exploração dinâmica dos dados [89].

Wong et al. B2 [123] afirma em sua pesquisa que a visualização da informação é uma grande promessa para a indústria de energia elétrica, mas seu potencial até agora não foi suficientemente explorado pela comunidade de visualização. O artigo apresenta o desenvolvimento de um novo protótipo de sistema de visualização para explorar o planejamento e o monitoramento da infra-estrutura de eletricidade da América do Norte. Além disso, descreve detalhes de implementação da ferramenta, que é exibida na Figura 3.19, e desempenho e avalia seus pontos fortes e fracos em relação à atual visualização da rede elétrica.

3.6 Residências Inteligentes

Uma residência inteligente é uma aplicação de computação abrangente que visa facilitar tarefas domésticas. Vários sinônimos são usados para a residência inteligente, como, por exemplo, automação residencial, casa adaptativa e casa consciente. O conceito de residência inteligente é a integração de diferentes serviços dentro de uma casa usando um sistema de comunicação comum. Assim, é possível assegurar uma operação econô-

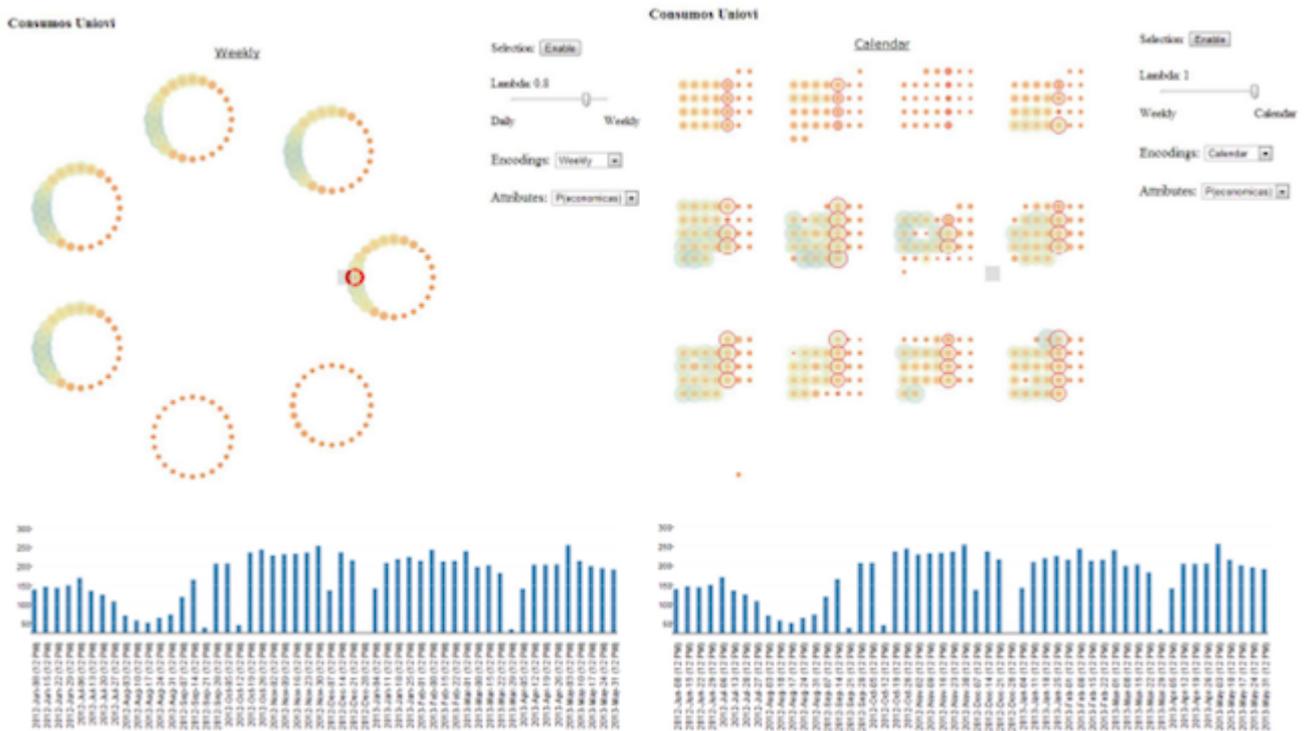


Figura 3.18 – A combinação dos diferentes mapeamentos visuais e a integração das técnicas de interação permitem analisar periodicamente diferentes intervalos de tempo por meio de consultas visuais. À esquerda, uma seleção visual é mostrada com os pontos destacados em vermelho para a exploração do consumo nesses pontos. Além disso, manter este destaque em uma mudança de codificação permite a ligação entre as codificações para os pontos selecionados à direita, com os pontos selecionados [89].



Figura 3.19 – Exibição do protótipo: (a) Mostra o consumo energético em um layout geográfico. (b) Mostra um layout do protótipo com pesos adicionais aplicados aos nós [123].

mica, segura e confortável da casa, incluindo um alto grau de funcionalidade e flexibilidade inteligentes [85].

Neste contexto, Pérez et al. [89] investigou no artigo P9 o potencial das residências inteligentes para promover reduções no uso de energia, focando no segmento populacional de famílias de alta renda. Foi feita uma avaliação durante um ano com 154 famílias de um distrito urbano sustentável em Estocolmo. A avaliação baseou-se em análises de dados de consumo de eletricidade e água quente, e em entrevistas para fornecer uma compreensão de como os *feedbacks* e as características de consumo são percebidas e influenciadas, além dos efeitos no comportamento e na consciência sobre o consumo. Os resultados mostram que o impacto no consumo de energia varia entre domicílios, com respostas individuais sobre *feedback* energético [80]. Uma aplicação denominada *Tingco Home* foi desenvolvida, com opções de monitoramento de recursos por meio de telas e um aplicativo móvel. A exibição dos dados coletados de uma residência é apresentada na Figura 3.20, incluindo funcionalidades de *feedback* de energia, tomadas inteligentes, entre outras opções.



Figura 3.20 – A tela inicial apresenta dados de consumo de eletricidade (canto superior esquerdo), água da torneira quente (meio superior) e aquecimento (superior direito), acompanhados, respectivamente, de comparações históricas (pontos verdes) com consumo de hora em hora/diário/semanal/mensal clicando nos medidores de energia, e comparações com outras casas (pontos amarelos). Um feedback sobre a distribuição de eletricidade, água quente e consumo de aquecimento mensais, é fornecido na parte inferior esquerda, e comparações históricas diárias no canto inferior direito. Além disso, também há menus para acessar diferentes recursos da casa na parte central inferior [80].

Desenvolver um processo de design focado no usuário com técnicas de visualização que provem a criatividade para identificar oportunidades de visualizar dados gerados por um grande fornecedor de energia foi o objetivo principal do artigo de Goodwin et al P12 [45]. Os protótipos de visualização desenvolvidos foram eficazes em uma situação em que os conjuntos de dados são amplamente desconhecidos e os requisitos abertos - permitindo uma exploração bem-sucedida das possibilidades de visualização na análise de dados do *Smart Home*. O artigo, que usou técnicas de visualização como o *Dense Pixel Display* apresentada na Figura 3.21, constituiu o primeiro projeto de visualização sistemática de uma fonte rica em dados importante para fornecedores e consumidores de energia, pois a tecnologia *Smart Meter* é amplamente implementada. Ao empregar técnicas de criatividade no estágio de requisitos do design e desenvolvimento da visualização, surgiu o caminho para o uso e estudo de métodos de criatividade no *design* de visualização.

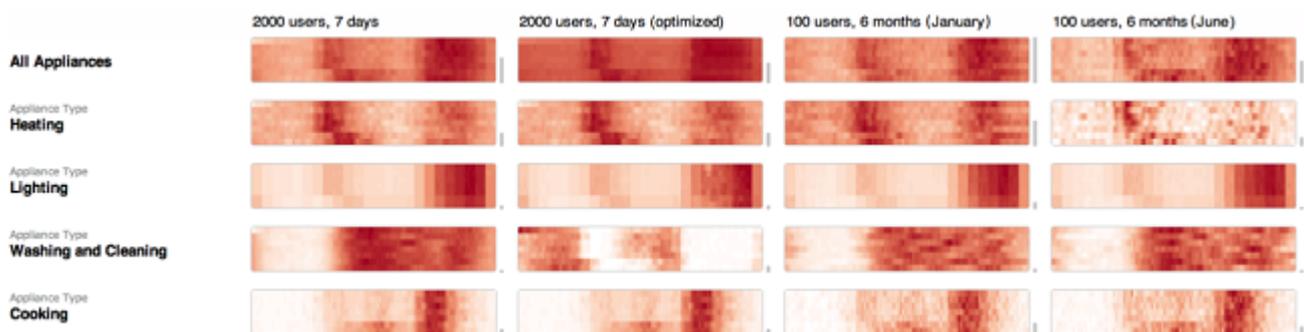


Figura 3.21 – Exibição do *Dense Pixel Display* mostrando o consumo de energia permitindo que dados modelados sejam carregados e reordenados para que os padrões de consumo semanal dos dispositivos possam ser comparados. Várias opções de coloração seqüenciais são fornecidas, tais como linha, coluna ou célula selecionada, permitindo que esquemas divergentes enfatizem a diferença dos itens selecionados. Os padrões nos dados diários (Iluminação), sazonal (Aquecimento) e modelados (Lavagem e Limpeza) são claros, assim como as diferenças nos finais de semana (duas linhas inferiores de cada célula), como o atraso no pico do aquecimento da manhã e mais cozimento durante o dia nos fins de semana [45].

Outro artigo relacionado a residências inteligentes é o F6 de Berry et al. [15], que descreve um sistema de monitoramento instalado em uma propriedade residencial e fornece um resumo dos dados de uso de energia gerados. Neste trabalho foram realizadas 25 entrevistas para discutir as atitudes dos residentes em relação a experiências de interação com o sistema. Os residentes descrevem como o *feedback* é exibido para ajudá-los a entender seu comportamento energético de uso final, reduzir o uso de energia e avaliar se os eletrodomésticos e os sistemas de energia renovável estão operando corretamente. Além disso, o papel da identificação de falhas no sistema de energia é destacado por muitos entrevistados, sendo que os *feedbacks* recebidos fornecem os meios para monitorar o desempenho do sistema, identificar falhas do sistema e manter resultados de baixa energia.

A utilização de *Smart Meter* em casas para coletar dados de energia a fim de identificar o comportamento das famílias foi o objetivo do artigo F9 de Wood et al. [125]. A pesquisa investigou como os membros das famílias respondem quando os dados de consumo de energia são integrados a informações sobre condições ambientais internas e acompanhadas de conselhos sobre ações de economia de energia. O artigo usou um sistema com *Dashboard* integrado de medidores de energia e sensores de presença que permitiu realizar a coleta de *feedback*. O aplicativo de visualização foi desenvolvido para exibir dados de consumo de eletricidade e gás, bem como dados de três dados meteorológicos: umidade relativa, dióxido de carbono e temperatura, como exibido na Figura 3.22.

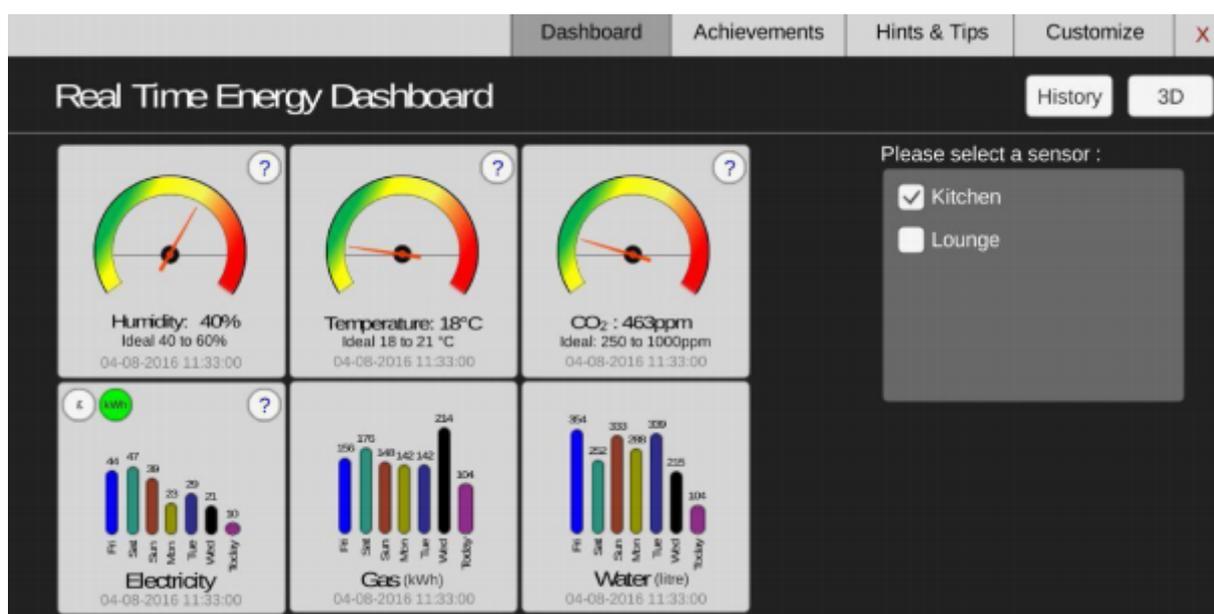


Figura 3.22 – Apresentação do *Dashboard* exibindo dados de umidade, temperatura, dióxido de carbono, eletricidade, gás e água no sensor instalado na cozinha de uma residência [125].

Explorar uma interface que busca o equilíbrio entre complexidade e usabilidade foi o tema do artigo F10 de Ahmed e Mueller [2]. Eles projetaram um sistema de painel versátil interativo com o qual os consumidores podem entender o consumo de energia de seus dispositivos domésticos para descobrir o impacto de seus padrões de uso, conforme mostra a Figura 3.23. Além de compará-los com os padrões de uso do passado, e por análise hipotética, identificam quais os efeitos que uma modificação desses padrões pode haver.

A necessidade de analisar e compreender o consumo de energia em relação aos padrões de atividade das famílias foi o objetivo do artigo das pesquisadoras Ellegård e Palm B1 [38], que afirmam que este entendimento é vital para o desenvolvimento de políticas que contribuam para uma vida com eficiência energética e para o que as pessoas considerem como uma vida cotidiana “boa”. Para atingir este objetivo, elas usaram a abordagem do diário geográfico do tempo juntamente com entrevistas para analisar a vida cotidiana como

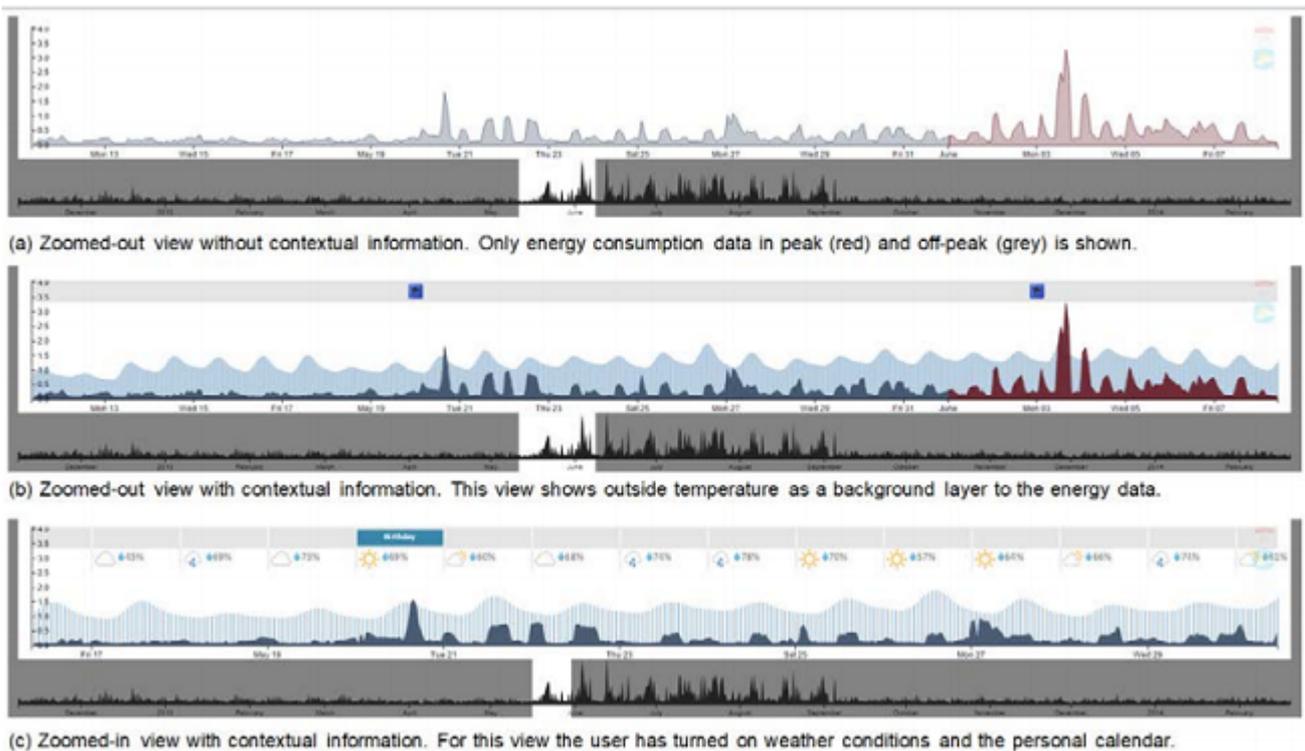


Figura 3.23 – Exploração de dados com diferentes níveis de zoom. Em todas as figuras (a) - (c), o gráfico na parte inferior é o histórico completo dos dados, enquanto o gráfico diretamente acima mostra os dados correspondentes à janela de tempo selecionada pela janela transparente na visão geral gráfico. Em cada parcela, as regiões azuis são devidas a horários de pico, enquanto as regiões vermelhas correspondem a horários de pico [2].

uma totalidade. A partir dos diários dos membros da família, aprenderam sobre quando, onde e quais atividades relacionadas à energia ocorrem em um contexto familiar, conforme mostrado na Figura 3.24 e por quem (e em que contexto social) elas são realizadas.

Motivado pela necessidade de melhor gerenciar a demanda de energia em casa, o objetivo de pesquisa de Costanza et al. B6 [33] foi integrar sistemas de visualizações interativas de consumo de energia, que permitiram aos usuários interagir e entender seus dados de consumo, relacionando-os as atividades concretas em sua vida. Para isso, apresentaram o projeto, a implementação e a avaliação de uma ferramenta, onde utilizaram visualizações interativas, como exibido na Figura 3.25, que permitiram aos usuários anotar e manipular uma representação gráfica de seus próprios dados de consumo de eletricidade e, portanto, compreender seu uso de energia no passado e entender quando como, e para que fim, alguma quantidade de energia foi usada.

Os resultados sugerem que em geral a abordagem de anotação é bem-sucedida: ao se envolver com os dados, os usuários começaram a relacionar o consumo de energia às atividades, e não apenas aos aparelhos. Além disso, eles foram capazes de descobrir que alguns aparelhos consomem mais do que o esperado [33].

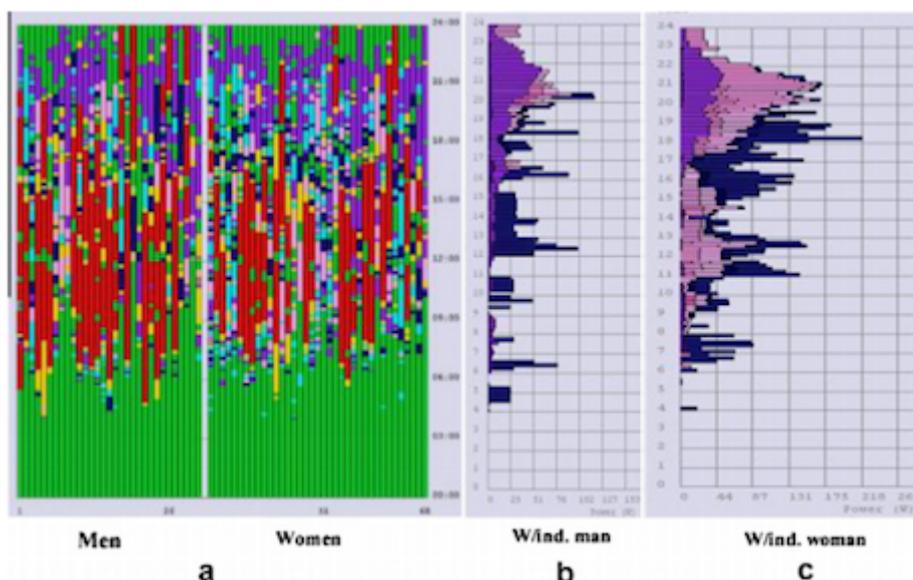


Figura 3.24 – Amostragem utilizada: a) Padrões de atividade entre homens e mulheres em domicílios com dois adultos entre 20 e 35 anos e crianças (= fig. 3). (b) Curvas de carga geradas por atividade para eletricidade entre os homens no grupo, e (c) entre as mulheres no grupo [38].



Figura 3.25 – Ferramenta mostrando um gráfico de séries temporais de consumo de energia que pode ser anotado pelos usuários para ajudá-los a entender seus próprios dados [33].

3.7 Ferramentas

Existem algumas ferramentas que permitem a análise e o monitoramento de consumo de energia. Entre as mais conhecidas estão *Energy plus*⁷, *Load profile generator*⁸ e o *Tableau*⁹. Visando identificar as funcionalidades que cada uma possui, foi realizada uma análise de cada uma delas e os resultados obtidos são apresentados a seguir.

⁷<https://energyplus.net/>

⁸<https://www.loadprofilegenerator.de/>

⁹<https://www.tableau.com/>

Energy plus. É um programa de simulação de energia para edifícios que engenheiros, arquitetos e pesquisadores usam para modelar tanto o consumo de energia (para aquecimento, resfriamento, ventilação e iluminação), quanto o consumo de água. Alguns dos recursos e capacidades incluem uma solução integrada e simulações instantâneas das condições da zona térmica do sistema HVAC - gerando cálculos de iluminação, brilho para relatar conforto visual e controles de luz, entre outras funcionalidades. O programa permite a criação de ambientes internos para gerar simulações de consumo de energia e fornece algumas técnicas de visualização de séries temporais, conforme exibido na Figura 3.26. Essa ferramenta não apresenta modelos de interação para visualizar dados ao longo do tempo e não suporta a inclusão de algoritmos de predição de consumo ou geração de energia.

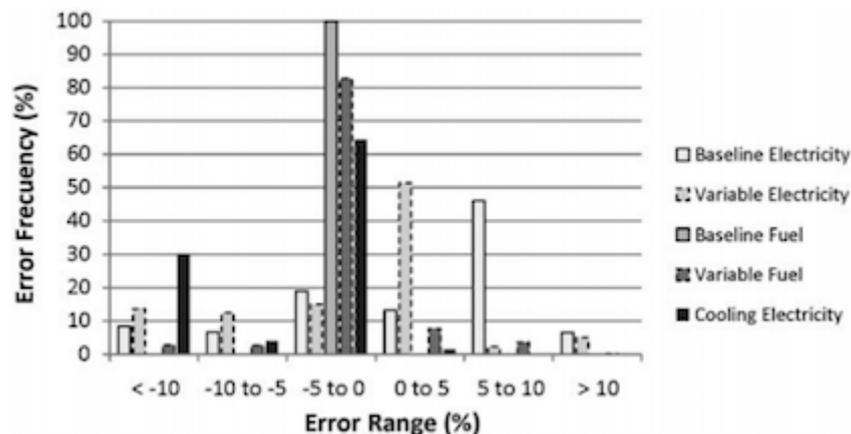


Figura 3.26 – Frequência de erros e erros para o consumo de energia estimado pelo *Energy plus* [128].

Load profile generator. É uma ferramenta para modelagem de consumo de energia residencial que realiza simulações de comportamento de pessoas em uma residência. Ele contém alguns lares predefinidos para gerar simulações e permite gerar arquivos CSV para utilizar os dados simulados em outros softwares de simulação. Além disso, gera diferentes relatórios para ajudar na modelagem e exibição dos dados simulados pela ferramenta, como ilustrado na Figura 3.27. Essa ferramenta não fornece *Dashboard* com gráficos interativos e, assim como *Energia Plus*, não suporta a inclusão de algoritmos de predição.

Tableau. É uma plataforma de análise visual avançada e flexível que suporta diferentes conjuntos de dados. É um sistema comercial que oferece conexão com diversas fontes de dados, propiciando a análise de diferentes dados, como ilustra a Figura 3.28. Ele fornece uma interface para criação de *dashboard* possibilitando arrastar e soltar componentes e visualizações para exploração interativa de dados. O *Tableau* não é uma ferramenta específica para analistas de consumo de energia (com o foco específico em monitoramento ou demanda energética), mas contém funcionalidades que servem a este propósito. O *Tableau* não oferece muitos modelos de interação, como por exemplo *Details on Demand*. Os gráficos gerados não permite gerar interatividade entre si, impedindo de apresentar,

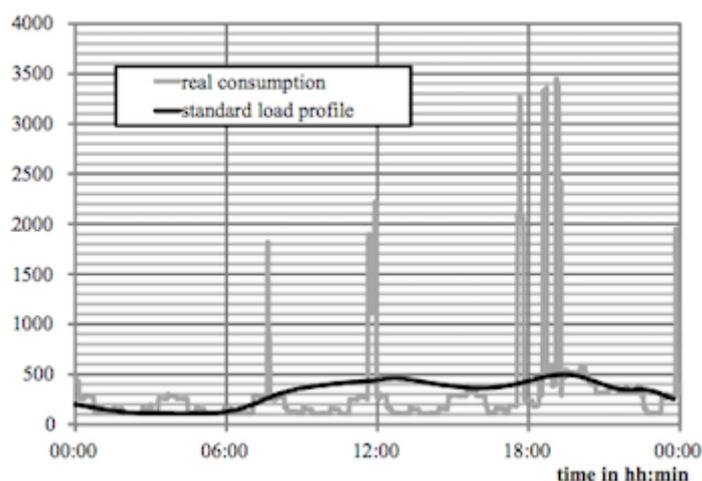


Figura 3.27 – Exibição das diferenças entre os perfis de carga e o consumo real de uma família *Load profile generator* [83].

por exemplo, um refinamento de diferentes níveis de granularidade dos dados ao longo do tempo.

3.8 Análise dos artigos relacionados

Após a apresentação de diferentes trabalhos é importante fazer um estudo comparativo entre eles para destacar a relevância de cada um para análise de dados de energia. De alguma forma, os artigos abordados neste capítulo proporcionam métricas que podem contribuir para a redução ou melhorar o consumo de energia elétrica.

Os artigos analisados se destacaram não somente por estarem dentro dos critérios de pesquisa. Eles correspondem a resultados obtidos por pesquisas sobre o tema de consumo energético, abrangendo a análise do perfil do público alvo, o entendimento das variações climáticas, o conhecimento dos equipamentos utilizados, entre outras atividades. Este estudo observa fatos e fenômenos da maneira como ocorrem na realidade e também coletam dados referentes aos elementos observados para, posteriormente, analisá-los e interpretá-los. O objetivo é compreender e explicar o problema que é objeto de estudo desta pesquisa, e os seus resultados podem gerar valor para a sociedade.

Este estudo também permitiu identificar quais técnicas de visualização de dados são mais utilizadas. As técnicas que mais apareceram nos artigos foram gráfico de linhas e de barras. A aplicação destas técnicas nos trabalhos analisados visavam exibir dados em horas, dias e, em alguns casos, em semanas. Para exibir dados com uma granularidade maior que dias e horas, como dados anuais e mensais, as técnicas de visualização que foram mais usadas, conforme mostra a Tabela 3.6, foram *Heatmap*, *Calendar View* e *Dense Pixel Display*. Por fim, a tabela mostra que 1/3 dos artigos selecionados utilizaram algum modelo de interação e/ou usaram *Dashboard*.



Figura 3.28 – Visualização gerada pelo *tableau* mostrando dados de vendas e lucro e quatro dimensões: Tempo, Estado, Produto e Fornecedor [116].

Entre as vantagens de analisar estes artigos, estão a identificação dos seus objetivos e das várias formas de visualização de dados de dados utilizadas. A tabela 3.5 exibe os artigos agrupados pelas categorias identificadas, incluindo também o objetivo de cada um. As técnicas de visualização que foram utilizadas nestes artigos, bem como as formas de interação, podem ser visualizadas na tabela 3.6. Estas técnicas são usadas para um melhor entendimento do consumo de energia e auxiliam os gestores na tomada de decisão. Outro fator importante é referente aos benefícios que podem gerar para a sociedade de modo geral.

Através da análise destes artigos, observamos a ausência de ferramentas ou modelos visuais que forneçam dados estatísticos para auxiliar no monitoramento do consumo. A utilização destes recursos é comum em ferramentas de *Visual Analytics*, como mencio-

nado na Seção 2.2, e no contexto aqui abordado permite ampliar a análise visual do gasto energético, fornecendo *insights* importantes para tomada de decisão.

Outra limitação dos trabalhos analisados é a não disponibilidade de algoritmos de análise preditivas. Estes algoritmos combinam dados, estatística e técnicas de aprendizado de máquina para identificar a probabilidade de resultados futuros baseados em informações históricas. Consideramos isso importante para prever dados e cenários futuros, principalmente para auxiliar os gestores na tomada de decisão. Além disso, não identificamos um trabalho que permita carregar diferentes conjuntos dados de energia e dados meteorológicos, e que possibilite compará-los em diferentes períodos ao longo do tempo.

Analisando os artigos abordados, percebe-se que poucos deles permitem realizar comparativos de consumo de energia com dados meteorológicos. No artigo P1, Chin et al. [30] explicam como é realizada a coleta de dados de fontes eólicas, mas não possibilitam comparar e relacionar outras variáveis, tais como temperatura e umidade do ar, e como isso implica diretamente na coleta de energia. Da mesma forma, Jui-Sheng et al. [31] no artigo P4 disponibilizam um *Dashboard* que identifica consumos anômalos, mas não demonstram a relação de nenhum dado meteorológico, como, por exemplo, temperatura, e como isso influencia a geração de um evento anômalo. Já Pickering et al. [84] exibem gráficos que relacionam dados de temperatura com dados de aquecimento e refrigeração, porém não mostram o impacto desses valores para o consumo de energia. O modelo aqui proposto permite gerar visualizações e análises comparativas de diferentes dados em diferentes níveis de granularidade ao longo do tempo. Não há suporte para detecção de anomalias, mas é possível realizar a identificação visual de picos de consumo de energia associado com qualquer dado meteorológico que tenha sido passado para o modelo. Essas análises são importantes, por exemplo, para permitir a identificação de que forma esses dados meteorológicos influenciam o consumo ou a geração de energia e, conseqüentemente, podem se transformar em decisões e estratégias para diminuir o consumo ou redistribuir a energia gerada.

ID	Referência	Ano	Objetivo
Edifícios inteligentes			
P2	Janetzko et al. [53]	2014	Apresenta um novo algoritmo de detecção de anomalia
P4	Jui-Sheng et al. [31]	2017	Desenvolve uma aplicação para analisar o consumo de energia
P5	Timm e Deal. [107]	2016	Apresenta formas de uso de energia para ocupantes em edifícios
P6	Blanco et al. [16]	2017	Avalia o consumo de energia em edifícios
P10	Nilsson et al. [80]	2018	Apresenta um sistema com consumo energético de quatro edifícios
P11	Prouzeau et al. [87]	2018	Descreve um protótipo visual para diferentes energias
P13	Brehmer et al. [19]	2016	Fornecer um conjunto de opções e diretrizes de design visual para dados orientados ao tempo no domínio de energia.
F5	Iyengar et al. [52]	2018	Apresenta uma ferramenta que visa identificar os edifícios menos eficientes
F8	Ioannidis et al. [51]	2018	Avalia o desempenho de um edifício usando análise visual impulsionadas pela ocupação do edifício.
B3	Yarbrough et al. [127]	2015	Desenvolve uma ferramenta visual para energias
Deteção de anomalias			
P2	Janetzko et al. [53]	2014	Apresenta um novo algoritmo de detecção de anomalia
P4	Jui-Sheng et al. [31]	2017	Desenvolve uma aplicação para analisar o consumo de eletricidade
F1	Arjunan et al. [9]	2015	Apresenta um sistema para monitorar o consumo de energia de vários usuários em uma vizinhança
F4	Cui et al. [34]	2017	Identifica eventos anômalos em edifícios
Eficiência energética			
P1	Chin et al. [30]	2018	Desenvolve uma plataforma de análise visual para exibir dados de energias eólicas
P3	L. Bartram [13]	2015	Aborda um estudo para conservação e redução de demanda de energia doméstica
P7	Xiong et al. [126]	2016	Desenvolve e implementa algoritmos de controle para sombreamento e operação de iluminação
F2	Arbesser et al. [8]	2017	Exibe um sistema para suportar uma inspeção de consumo eficiente
F3	Pickering et al. [84]	2018	Analisa dados de consumo de seis edifícios comerciais
F5	Iyengar et al. [52]	2018	Apresenta uma ferramenta que visa identificar os edifícios menos eficiente
F7	Garcia et al. [44]	2018	Apresenta um aplicativo de análise visual para usuários de domínio
B4	Morán et al. [76]	2013	Fornecer uma análise visual das correlações entre consumo e variáveis ambientais
B5	Martani et al. [72]	2012	Propõe um novo método para analisar o consumo de energia
Redes inteligentes			
P8	Agarwal et al. [1]	2009	Desenvolve um sistema para apoio no consumo de energias
B2	Wong et al. [123]	2009	Descreve a implementação e avalia a visualização de rede elétricas
Residências inteligentes			
P9	Pérez et al. [89]	2014	Investiga o potencial de um sistema de energia para reduzir o consumo
P12	Goodwin et al. [45]	2013	Desenvolve um processo de design focado no usuário para identificar oportunidades e visualização de dados de energia
F6	Berry et al. [15]	2016	Descreve um sistema de monitoramento de energia instalados em uma casa
F9	Wood et al. [125]	2019	Investiga como as pessoas respondem aos dados de energia em ambientes internos
F10	Ahmed e Mueller. [2]	2019	Exploram uma interface buscando equilibrar a troca entre complexidade e usabilidade para análise visual de energia
B1	Ellegård and Palm. [38]	2011	Aborda a necessidade de analisar e compreender o consumo energético de famílias
B6	Costanza et al. [33]	2012	Apresenta uma ferramenta de análise visual para monitorar o consumo de eletricidade

Tabela 3.5 – Tabela dos artigos agrupados por diferentes categorias, onde cada artigo exibe os nomes dos autores, ano da publicação e objetivo do artigo.

Artigos	Técnicas de visualização de dados													Modelos de Interação			Outros Recursos			
	Dense Pixel Display	Heatmap	Line chart	Radar/Polar chart	Coordenadas Paralelas	Bar chart	Plot chart	Calendar view	Spiral	Treemap	Solid gauge	Buble chart	Pie chart	Combinções integradas	Coordinate Multiple Views	Brushing and Linking	Details on Demand	Fornece Dashboard	Permite Comparações	Algoritmos de detecção
P1				✓	✓									✓			✓	✓		
P2	✓		✓				✓	✓	✓					✓	✓			✓	✓	
P3						✓								✓				✓		
P4			✓		✓										✓	✓	✓	✓		✓
P5						✓				✓				✓				✓		
P6			✓			✓													✓	
P7						✓									✓	✓			✓	
P8			✓			✓									✓	✓			✓	
P9			✓			✓		✓						✓	✓	✓				
P10						✓	✓		✓	✓								✓		
P11			✓			✓	✓							✓						
P12	✓					✓													✓	
P13	✓	✓	✓		✓	✓		✓							✓	✓	✓	✓		
F1	✓					✓													✓	
F2			✓			✓								✓	✓	✓	✓		✓	
F3			✓			✓													✓	
F4		✓	✓			✓													✓	✓
F5						✓	✓												✓	✓
F6			✓			✓								✓					✓	✓
F7		✓	✓			✓	✓							✓	✓	✓		✓		
F8		✓	✓	✓		✓					✓	✓				✓	✓		✓	✓
F9	✓	✓	✓		✓									✓				✓	✓	
F10			✓			✓		✓										✓	✓	
B1		✓	✓																✓	
B2		✓												✓		✓			✓	
B3		✓	✓																✓	
B4		✓	✓			✓													✓	
B5			✓			✓	✓							✓				✓	✓	
B6			✓			✓								✓				✓	✓	

Tabela 3.6 – Tabela comparativa de técnicas de visualização de dados, modelos de interação e outros recursos disponíveis (Fornece *Dashboard*, Permite Comparações e Algoritmos de detecção de anomalias) utilizados nos 29 artigos seleccionados.

4. DESCRIÇÃO DO MODELO

O modelo proposto disponibiliza diferentes técnicas de visualização e interação para permitir a análise de dados de energia ao longo do tempo. Na Seção 4.1, detalhamos a metodologia utilizada e as fases de execução da pesquisa. Na seção 4.2 apresentamos os requisitos levantados para o projeto e o desenvolvimento do modelo proposto. Na seção 4.3 descrevemos a estrutura de dados que o modelo suporta. Por fim, as Seções 4.4 e 4.5 descrevem, respectivamente, as técnicas de interação e visualização escolhidas e implementadas no protótipo do modelo.

4.1 Metodologia

De acordo com Carpendale [23] há uma variedade de metodologias de pesquisa voltadas para visualização de informações, cada uma com seus benefícios e desvantagens. Para aplicar uma metodologia adequada, é importante considerar o objetivo do trabalho, bem como a possibilidade dele auxiliar na compreensão e entendimento da visualização em diferentes contextos.

Seguindo os conceitos apresentados por Denzin et al. [35], este trabalho se caracteriza como um estudo experimental e qualitativo. Portanto, foi definido um conjunto de atividades para alcançar os objetivos propostos. Estas atividades foram organizadas nas 10 etapas apresentadas na Figura 4.1, numeradas de acordo com a sua ordem de execução.

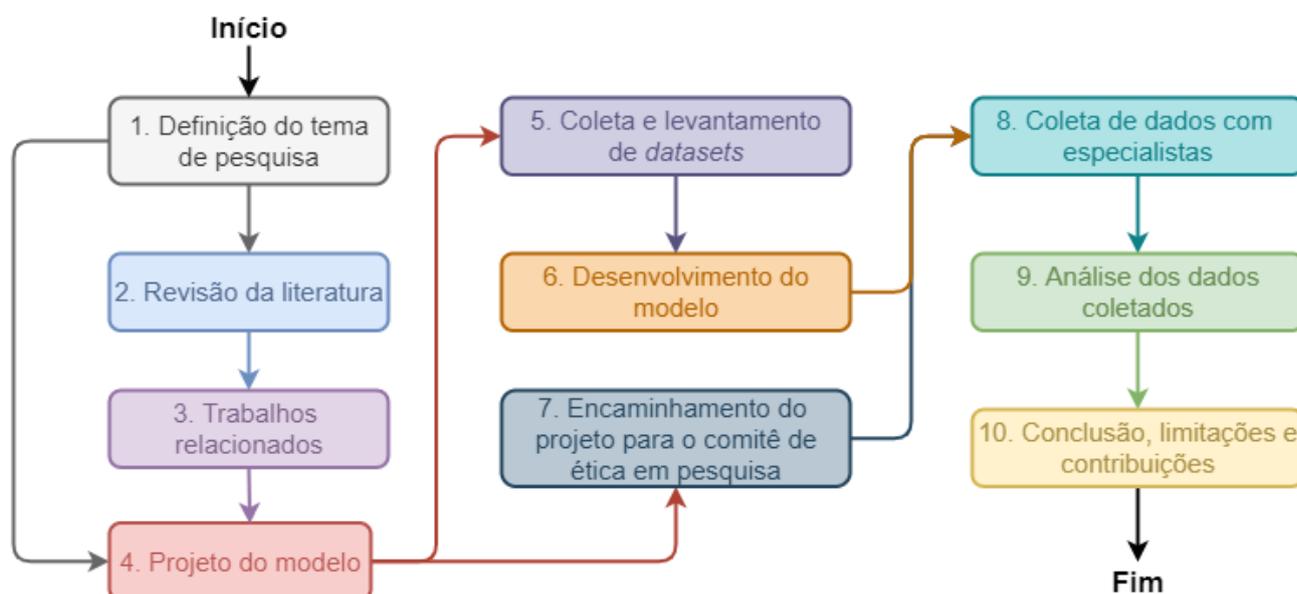


Figura 4.1 – Etapas do desenvolvimento do modelo.

Conforme ilustra a Figura 4.1, após a definição do tema de pesquisa foi feita uma revisão da literatura e de trabalhos relacionados, que foram contempladas pelas etapas 2 e 3, conforme descrito no Capítulo 3. As etapas 4, 5 e 6 compreendem, respectivamente, o projeto do modelo, a definição e busca de conjuntos de dados a serem utilizados, e a implementação da estrutura necessária para visualizar, analisar e interagir com os dados. A etapa 7 corresponde ao encaminhamento do protocolo de pesquisa ao Comitê de Ética em Pesquisa (CEP), para que fosse possível validar o estudo realizado com especialistas. Após a aprovação no CEP, foram executadas as etapas 8 e 9, que englobam a coleta e análise de dados através de entrevistas com especialistas usando um questionário semiestruturado [74], e são descritas no Capítulo 5. Por fim, a etapa 10, apresentada no Capítulo 6, consiste na análise final do modelo desenvolvido, suas limitações, contribuições e trabalhos futuros da pesquisa.

4.2 Levantamento de Requisitos

Inicialmente, consideramos dois cenários de uso hipotético para identificar os requisitos do modelo. Neste contexto, dois usuários especialistas precisam analisar os dados de energia de um EI para encontrar padrões de consumo e anomalias, a fim de adotar medidas para reduzir o gasto de energia. O principal objetivo destes usuários é entender como os padrões de consumo variam ao longo do tempo e em quais situações houve um gasto discrepante, para auxiliar na definição de possíveis ações a serem tomadas. Assim, foram criados os seguintes cenários (C) hipotéticos:

- C1 - José é um engenheiro civil que foi contratado para melhorar a eficiência energética de um EI. Para isto, ele está considerando realizar algumas ações imediatas, como a instalação de painéis solares. Porém, ele não tem conhecimento sobre sistemas de análise visual de energia que poderiam ser adotados para facilitar o entendimento do consumo energético atual deste edifício e também monitorar futuramente a captação e energia solar. Ele demonstra preocupação em compreender como se comportam os dados meteorológicos na região, como a temperatura, umidade do ar e períodos de maior e menor consumo de energia. Além disso, ele considera que prever comportamentos futuros pode ser essencial para tomada de decisões, como, por exemplo, colocação de vidros duplos que impeçam a entrada de frio ou calor nos ambientes, visando diminuir o gasto de energia oriunda de aparelhos de ar condicionado.
- C2 - Helena é uma engenheira elétrica que atua como gestora em uma concessionária municipal de energia. Esta companhia possui diferentes dados de energia, porém, devido ao grande volume, ela tem dificuldade para analisá-los. Ela lamenta a falta de estudo destes dados, pois acredita que se houvesse um sistema para auxiliar nesta

tarefa ela conseguiria identificar possíveis anomalias e diminuir o desperdício de energia, bem como de recursos públicos. Ela reconhece que cada região da cidade tem suas peculiaridades e há diferentes padrões de gasto ao longo do ano, mas ressalta que poderia ter mais ações públicas se conseguisse identificar tanto os dias e horários dos picos de consumo em cada região, como as zonas que mais apresentam valores discrepantes de consumo. Além disso, também seria interessante comparar os padrões de consumo entre diferentes regiões, fazendo uma associação com as variações de temperatura ao longo do ano.

Diante destes casos, foram levantadas as seguintes questões (Q) para serem respondidas:

- Q1 - O padrão de consumo varia em diferentes períodos de tempo e para diferentes fontes? É necessário fornecer maneiras de analisar e compreender o gasto de energia em diferentes períodos de tempo (por exemplo, diário, semanal, mensal ou anual) e para diferentes fontes de energia (por exemplo, aquecedor elétrico, ventilador e ar condicionado) visando propor ações que reduzam o gasto de energia.
- Q2 - Quando ocorrem *outliers*, isto é, consumos fora do padrão, e em que situações? Neste caso, é preciso fornecer representações visuais que tanto facilitem a rápida identificação de valores discrepantes, como permitam fazer uma análise estatística que ajude a identificar o motivo que levou à ocorrência do *outlier*.
- Q3 - Como é possível visualizar os dados de energia de forma interativa? Essa questão inclui a identificação das técnicas de interação mais apropriadas para ajudar na análise, considerando diferentes níveis de granularidade de tempo dos dados.
- Q4 - Como realizar comparações entre diferentes fontes de energia e dados meteorológicos em diferentes períodos? Para isto, é necessário apresentar maneiras para permitir a visualização de dados de diferentes conjuntos e para diferentes períodos ao longo do tempo.
- Q5 - Qual é a tendência de gasto de energia nas próximas semanas? Para responder a esta questão é preciso fornecer algoritmos para prever o consumo futuro de energia, analisando os dados do passado.

Com base nessas questões, identificamos a importância da análise exploratória de dados para entender os dados de energia. Assim, definimos os sete requisitos (R) listados a seguir para o modelo proposto:

- R1 - Oferecer um formato de dados de entrada simples que permite carregar vários conjuntos de dados;

- R2 - Fornecer representações visuais de dados para diferentes níveis de granularidade de tempo (anual, mensal, semanal, diária ou por hora);
- R3 - Prover uma abordagem interativa;
- R4 - Fornecer diferentes técnicas de visualização para exibir os mesmos dados em diversas representações visuais;
- R5 - Permitir visualizar e comparar dados de consumo de energia obtidos de diferentes fontes (por exemplo, aquecedor elétrico, ar condicionado ou eletrodomésticos), dados de geração de energia (obtidos, por exemplo, através de painéis solares ou sensores fotovoltaicos) e dados meteorológicos (por exemplo, dados de temperatura, umidade e pressão do ar);
- R6 - Fornecer informações adicionais, como o uso de funções estatísticas, para facilitar a compreensão dos dados;
- R7 - Permitir a aplicação de diferentes algoritmos de predição para o gasto de energia.

Após o planejamento destes cenários hipotéticos, com as questões geradas e os requisitos criados, iniciamos a etapa de definição e implementação da entrada de dados, incluindo a especificação da estrutura de dados que o modelo suporta para gerar as representações visuais e as técnicas de interações. Essas informações são apresentadas na seção 4.3. Para validar estes cenários hipotéticos, elaboramos algumas perguntas no questionário de avaliação com usuários, conforme demonstrado na seção 2 do Apêndice B. Esses questionamentos foram criados com o objetivo de realizar um levantamento de práticas e necessidades que os especialistas enfrentam para analisar e explorar dados de energia.

4.3 Entrada de dados

O presente modelo foi projetado para representar especificamente dois tipos de dados: Dados de energia e dados meteorológicos. Dados de energia podem ser obtidos por dados de consumo de energia, que são gerados através de casas residenciais, condomínios ou edifícios, ou por dados de geração de energia, fornecido por fontes de energia renováveis, como energia eólica ou solar. Dados meteorológicos são dados relacionados a partir de diversas variáveis climáticas, como temperatura, vento, umidade, precipitação, radiação atmosférica e pressão da superfície, entre outros.

Independentemente do tipo de dado de energia que for carregado no modelo, é necessário que ele seja fornecido com um formato padrão e que os valores escalares sejam normalizados, para que as representações visuais sejam apresentadas. Portanto, o formato

de entrada de dados do modelo foi definido para permitir a entrada de diversos valores de uma maneira simples e eficaz.

Considerando que os dados de energia correspondem a dados de séries temporais, é preciso identificar o nome e o valor consumido ou gerado de cada fonte de energia. Assim, essas informações devem estar em um arquivo no formato JSON [70] com a estrutura apresentada na Figura 4.2-A, que é o tipo de arquivo que pode ser carregado pelo modelo (atendendo ao requisito R1). O formato JSON foi escolhido por ser um formato intercâmbio de dados leve e fácil de ler e escrever [70]. Caso os dados forem disponibilizados em outro formato, será necessário realizar um pré-processamento para deixá-los no formato de leitura que o modelo suporta. Para carregar dados meteorológicos deve ser informada a mesma estrutura de dados, sendo necessário identificar o tipo de dado (temperatura, pressão, etc.) e o valor associado a ele, como mostra a Figura 4.2-B. A primeira linha de cada conjunto de dados, tanto de energia como de dados meteorológicos, deve conter a identificação dos campos, sendo data e hora sempre o dois primeiros. As demais linhas, portanto, contêm os valores associados a cada campo.

A)			
DATA	HORA	AR CONDICIONADO	VENTILADOR
2019-11-08	0	20w	30w
2019-11-08	1	21w	28w
2019-11-08	2	20w	29w
2019-11-08	3	22w	30w
2019-11-08	4	21w	28w
B)			
DATA	HORA	TEMPERATURA	VELOCIDADE DO VENTO
2019-11-08	0	20°C	5m/s
2019-11-08	1	21°C	8m/s
2019-11-08	2	19°C	4m/s
2019-11-08	3	18°C	7m/s
2019-11-08	4	20°C	6m/s

Figura 4.2 – . Estrutura do JSON para carregar dados: A) Exibe a formatação dos dados de energia. B) Apresenta a formatação dos dados meteorológicos.

Para melhor visualização através do modelo, recomenda-se que o conjunto de dados fornecido contenha informações de mais de um ano. O modelo suporta a visualização de dados de diferentes fontes de energia, em diferentes períodos ao longo do tempo,

como dados anuais, mensais, semanais e diários. No entanto, o modelo também permite a visualização dos dados se forem fornecidos apenas dados de alguns meses do ano, por exemplo.

Os valores de cada fonte de energia e de qualquer dado meteorológico devem ser fornecidos em valores escalares padrão. Por exemplo, se dados de consumo de energia variarem de 0 a 100 watts, o valor 0 estará representando o menor consumo e o valor 100 indica o maior gasto energético. Da mesma forma, se forem carregados dados meteorológicos de temperatura de -5° a 40° graus, o modelo irá gerar as representações visuais de acordo com essa faixa de valores. No entanto, para serem carregados no modelo, os valores escalares no arquivo JSON não precisam necessariamente estar entre 0 e 100 watts para dados de energia e nem na faixa de -5° a 40° graus para dados meteorológicos de temperatura. Porém, precisam estar em um intervalo escalar. Isso ocorre porque o modelo requer um intervalo escalar numérico para gerar a escala de cores de acordo com o valor numérico.

4.4 Componentes do modelo e Técnicas de interação

Para possibilitar uma análise visual interativa, foi definido um fluxo de atividades a serem realizadas, conforme ilustra a Figura 4.3. A primeira atividade (Figura 4.3a) consiste em selecionar obrigatoriamente um conjunto de dados de energia e o ano a ser analisado. Após isso, o modelo gera as visualizações correspondente às informações do conjunto de dados de energia e executa a média geral do consumo ou geração de energia, ou seja, realiza a soma de todos valores de energia e divide pela quantidade de energias carregadas. Dados meteorológicos são opcionais e somente serão apresentados caso o ano do conjunto de energia seja igual ao ano dos dados meteorológicos. O modelo proposto aceita qualquer conjunto que contenha dados de séries temporais de energia no formato apresentado na Seção 4.3. Portanto, é necessário ter campos de data e hora, associados a pelo menos um dado de energia para cada registro. Esses dados correspondem a valores escalares normalizados. Após a seleção dos dados de energia, o modelo automaticamente identifica se há dados meteorológicos para o ano selecionado e disponibiliza os dados meteorológicos que podem ser visualizados.

Após carregar os dados, a segunda atividade (Figura 4.3b) corresponde à seleção das fontes de energia para explorar, analisar, visualizar e comparar, com a possibilidade de, posteriormente, associar uma representação visual diferente para cada uma das fontes de energias selecionadas. A última atividade (Figura 4.3c) corresponde às visualizações interativas, considerando as seleções realizadas anteriormente. O modelo é todo interativo, ou seja, se o conjunto de dados (Figura 4.3a) ou a seleção das fontes de energia (Figura

4.3b) forem alteradas, ou se houver alguma alguma interação nas visualizações, estas são automaticamente atualizadas (Figura 4.3c).

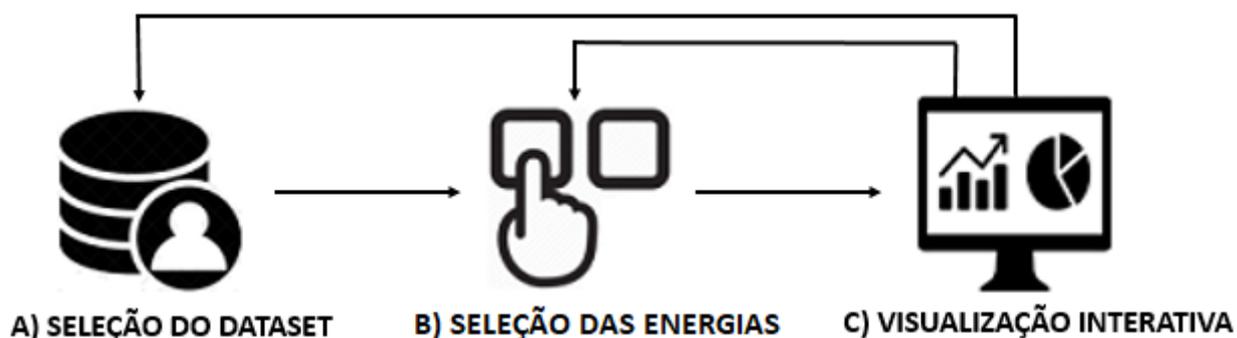


Figura 4.3 – Fluxo de atividades da análise visual interativa.

A Figura 4.4 apresenta os componentes presentes no modelo, considerando também os seus posicionamentos. Dada a premissa que o modelo deve exibir dados em diferentes granularidades de tempo, a seleção e a exibição dos mesmos deve ser eficaz e interativa. Por isso, as representações gráficas são exibidas simultaneamente, divididas em vários blocos, cada um com uma representação diferente do mesmo conjunto de dados.

O três primeiros blocos de cima para baixo correspondem às visualizações anuais dos dados de energia e dados meteorológicos, e por isso ocupam uma área maior. Os blocos abaixo apresentam, da esquerda para a direita, os dados semanais, diário e de hora em hora. A interação ocorre da seguinte maneira: sempre que é feita uma seleção nos gráficos anuais de energia, os respectivos dados semanais da seleção feita são exibidos, e sempre que é feita uma seleção no gráfico semanal, os respectivos dados do dia selecionado são exibidos, e de forma análoga, podem ser exibidos dados de hora em hora ao longo de um dia selecionado. Na sequência tem dois blocos lado a lado, que representam o consumo energético mensal, mas com focos distintos: o gráfico do lado esquerdo apresenta os dados totais de energia para cada mês ao longo de um ano com a possibilidade de incluir informações de previsão. O gráfico do lado direito apresenta o consumo total mensal da energia selecionada. No último bloco, são exibidas informações com os totais de energia por mês do ano selecionado. Assim, é possível começar a análise com uma visualização de dados anual e diminuir gradualmente a granularidade das informações, passando por meses, semanas, dias e chegando em horas, que é a menor granularidade fornecida pelo modelo (suportando o requisito R2).

As técnicas de interação devem facilitar o uso do modelo, para que através de poucos cliques seja possível explorar os dados de diferentes maneiras. Por isso, foram projetadas diferentes formas de interação, permitindo que um grande conjunto de dados seja analisado e explorado em diferentes períodos ao longo do ano, possibilitando a utilização de filtros com efeito de “*zoom in/zoom out*” [7] e fornecendo diferentes formas de comparar dados.

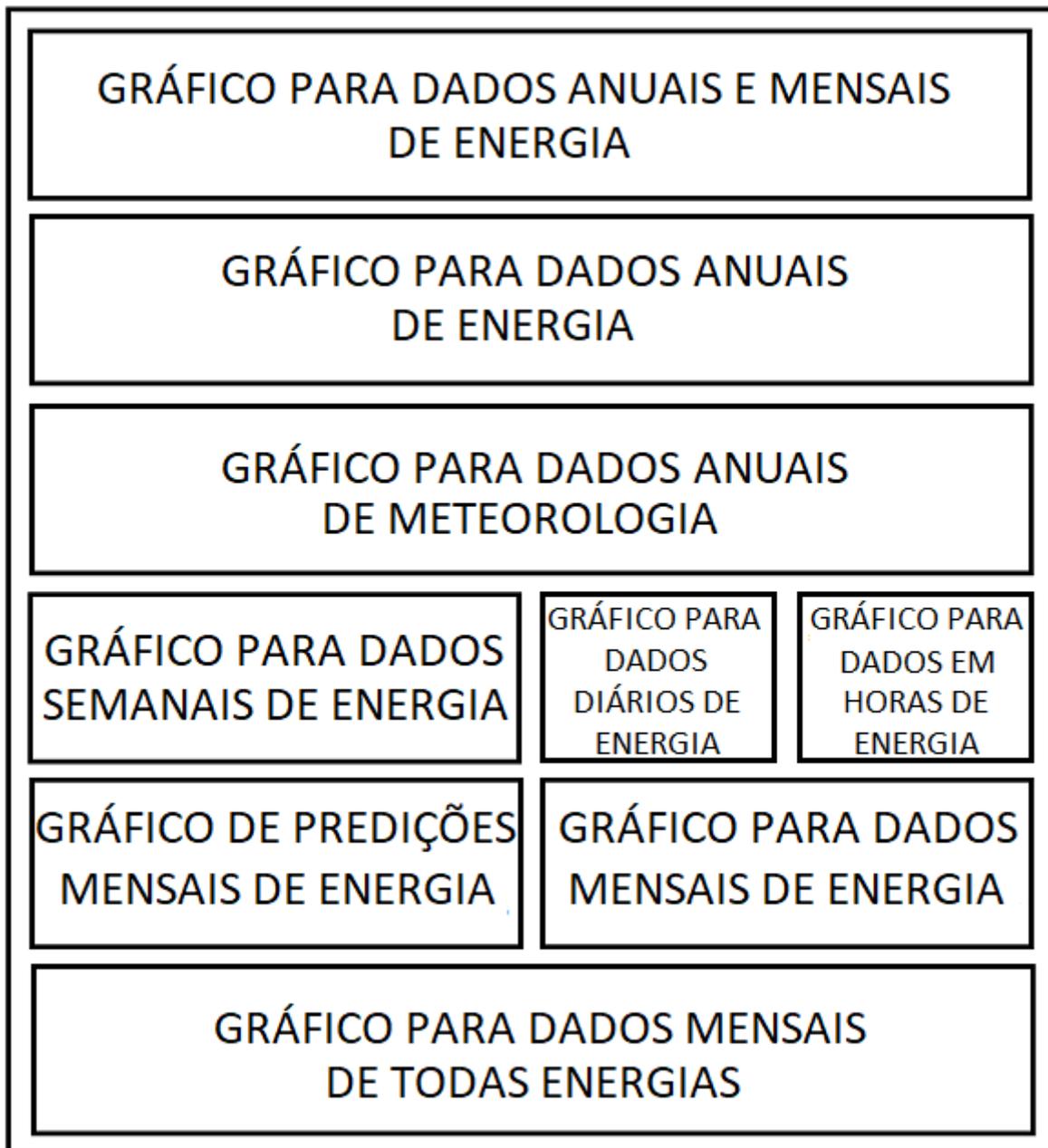


Figura 4.4 – Componentes do modelo proposto.

Portanto, uma técnica de interação suportada pelo modelo é *Details on demand*, pois ela permite a seleção interativa de dados a serem visualizados mais detalhadamente. Algumas informações sobre essa técnica são apresentadas na Seção 2.5.3. Esta técnica foi integrada com *Coordinated Multiple Views* (CMV) [92, 94, 10], ou múltiplas visões coordenadas, que é usada em diversas aplicações para dar suporte a uma exploração interativa [17, 65, 111] (atendendo o requisito R3). Por fim, o modelo também suporta abrir ilimitadas janelas com diferentes visualizações, uma abordagem interativa que facilita a comparação diferentes dados de energia e dados meteorológicos. Este sistema de janelas foi baseado na proposta apresentada por Cava [25].

Estas técnicas foram selecionadas para serem integradas ao modelo porque se mostraram adequadas para a análise visual de diferentes dados de energia. Como o modelo foi projetado para ser bastante interativo, em qualquer bloco apresentado na figura 4.4 pode ser feita uma seleção. Dessa forma, sempre que é feita uma seleção em qualquer visualização fornecida, todos os gráficos associados são atualizados para mostrar apenas os dados relacionados à seleção.

4.5 Implementação e Técnicas de Visualização

Nesta seção é descrita a implementação do modelo e as razões que levaram à utilização de cada técnica de visualização e interação. A Figura 4.5 mostra a interface completa do modelo. Cada técnica de visualização que compõe o modelo está associada a uma granularidade de tempo: anual, mensal, semanal, diária e em horas. Para as visualizações anuais, foram utilizadas as técnicas *Calendar View* e *Dense Pixel Display*. Para as granularidades mensais, o próprio *Calendar View* auxilia nesta análise, mas também foram disponibilizados gráficos de linhas e *boxplot*. A granularidade semanal é representada pelo *Heatmap*. Por fim, as granularidades diárias e em horas são apresentadas em gráficos de linhas, barras, plot e *boxplot*. Todas as figuras exibidas ao longo desta seção foram geradas usando o conjunto de dados A, descrito na Seção 5.2. A implementação foi feita utilizando as seguintes tecnologias: Javascript¹, jQuery², Python³, PHP⁴ e a biblioteca para geração de gráficos *Highcharts*⁵.

4.5.1 Gráficos para Dados Anuais

Calendar View é uma técnica de visualização que agrupa dados diários ao longo do ano dividido por meses. Cada dia é colorido de acordo com uma escala de cores predefinida, visando representar diferentes faixas de valores. Isso facilita a localização de dias com características semelhantes e o reconhecimento de padrões visuais [21]. É uma técnica que permite aos usuários explorar informações em diferentes escalas de tempo, como ano, mês, semana e dia [42] (suportando o requisito R2). Mais detalhes dessa técnica foram apresentados na Seção 2.2.1.

O uso desta técnica no modelo permite ao usuário compreender comportamentos, padrões, tendências e identificar *outliers* e anomalias em uma perspectiva anual dos dados.

¹<https://www.javascript.com/>

²<https://jquery.com/>

³<https://www.python.org/>

⁴<https://www.php.net/>

⁵<https://www.javascript.com/>

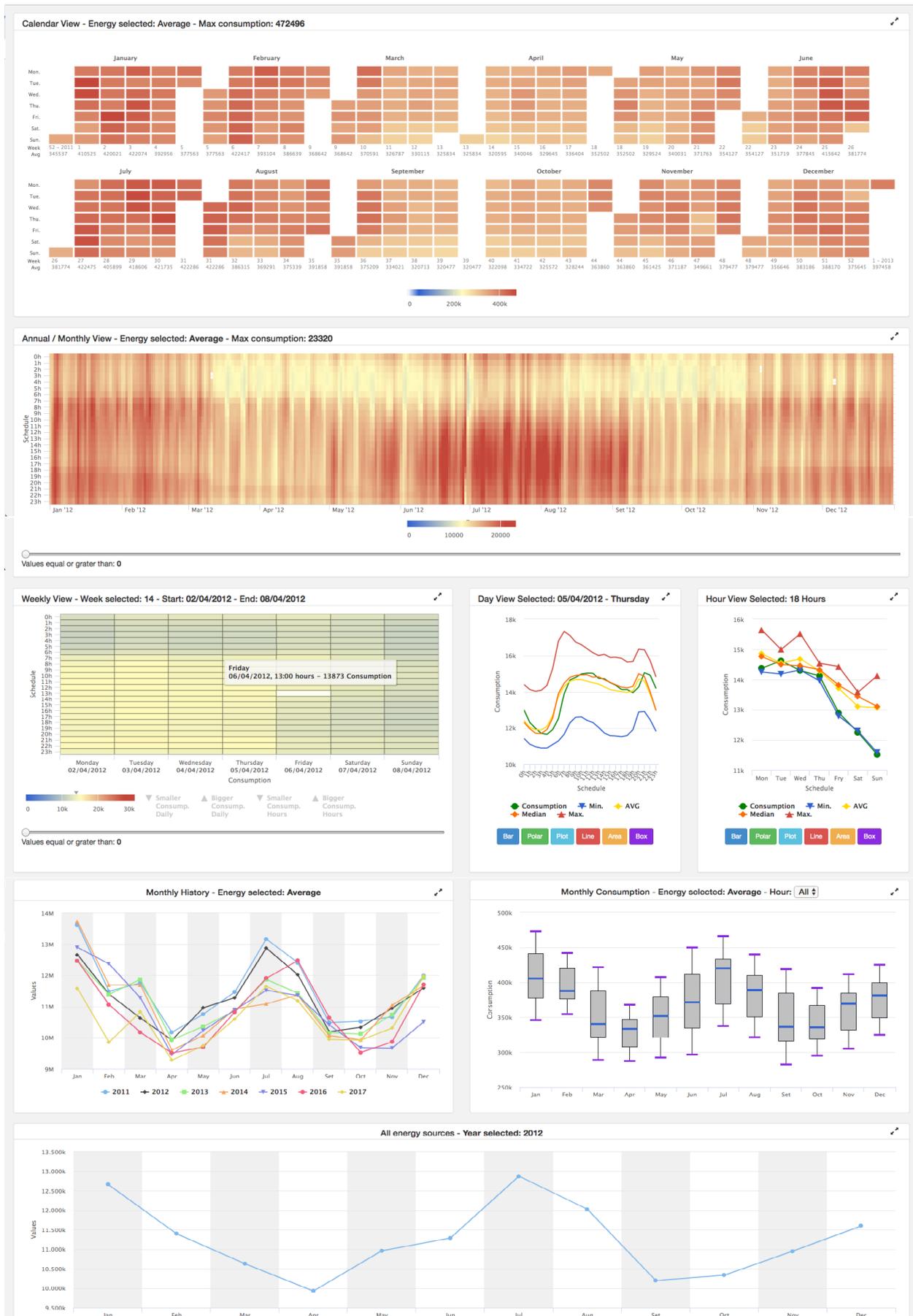


Figura 4.5 – Interface completa do modelo desenvolvido.

A partir desta visualização do conjunto de dados, é possível refinar a granularidade da informação de interesse em um período específico, partindo de uma visão anual e avançando para outras granularidades de tempo, como períodos semanais, diários e em horas (Um exemplo de como ocorre este refinamento é demonstrado na Figura 4.18).

Na implementação do modelo, *Calendar View* mostra o consumo ou geração de energia diariamente, ou seja, a soma da energia em horas do dia. Para melhor exibição dos dados, ele este dividido em dois semestres e é exibido o nome do dia da semana ao lado esquerdo. Além disso, abaixo de cada mês, é exibido o valor médio de consumo energético semanal e o número da semana do ano, conforme mostrado na Figura 4.6.

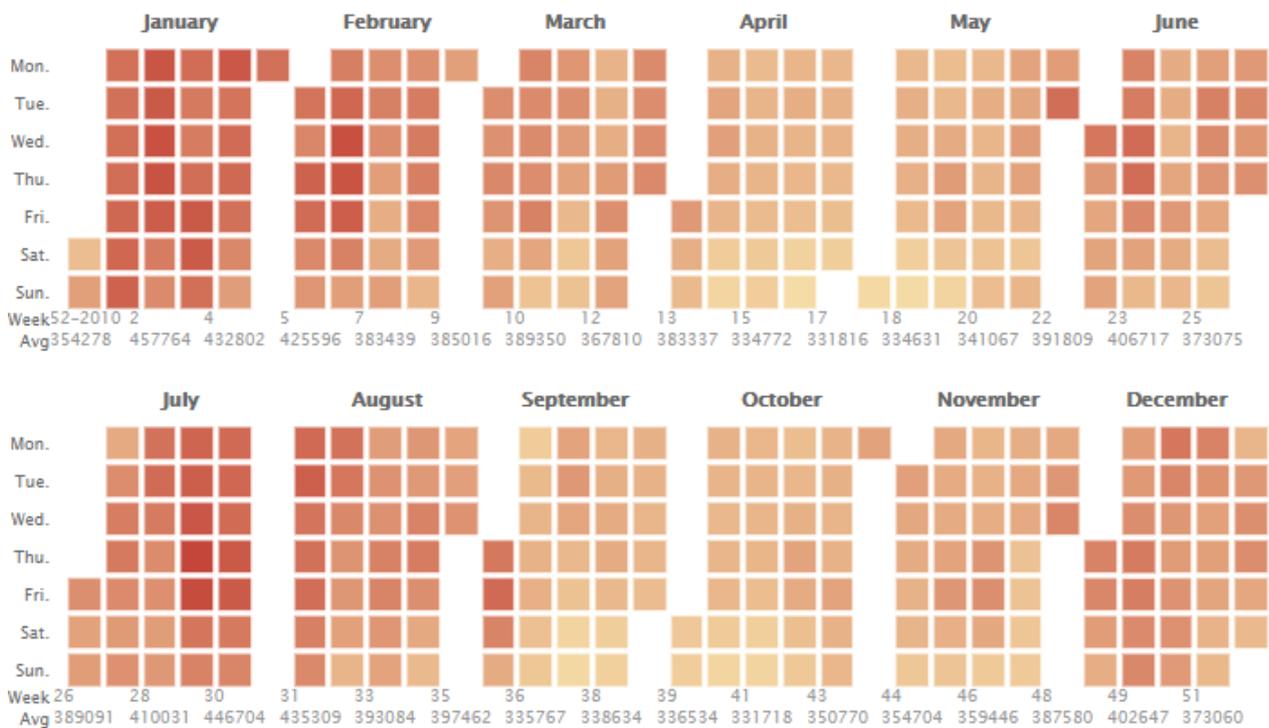


Figura 4.6 – Exibição do *Calendar View*. Abaixo de cada mês é exibido o valor de consumo ou da geração semanal e o número da semana do ano.

Diferentemente de um calendário impresso que apresenta a passagem dos dias ao longo do ano da esquerda para a direita, este calendário foi desenvolvido para apresentar os dias de cima para baixo. A justificativa para este padrão é que isso facilita a identificação da semana do ano no componente e também a exibição da média de consumo semanal no rodapé de cada mês, como mostra a Figura 4.7.

Dense Pixel Display, ou técnicas orientadas a pixels descritas na Seção 2.2.2, é interessante para representar grandes conjuntos de dados porque permite que milhares de valores sejam exibidos em uma única tela. Neste caso, cada valor controlará a cor de um único pixel. Assim, modificar a seleção de dados nesse gráfico pode revelar novos recursos e padrões de dados.

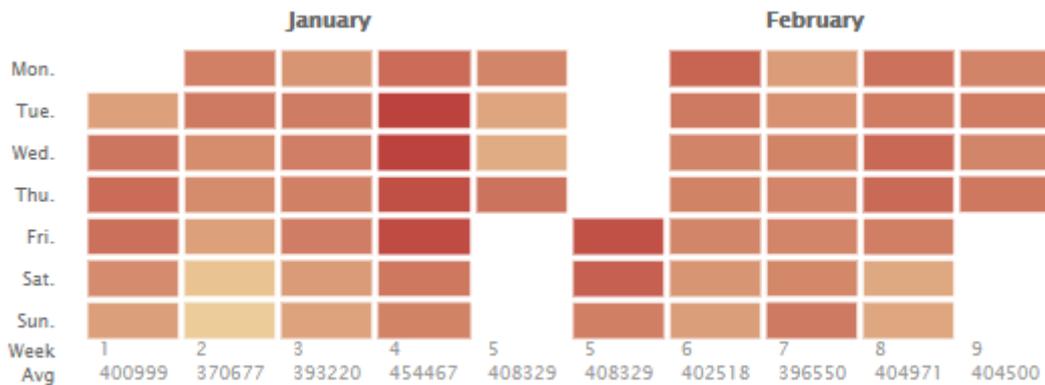


Figura 4.7 – Exibição dos meses de janeiro e fevereiro no *Calendar View*, onde são apresentados no rodapé de cada mês a semana do ano e a média de consumo ou de geração de energia.

Assim como *Calendar View*, o *Dense Pixel Display* também permite compreender comportamentos, padrões, tendências e identificar *outliers* e anomalias em uma perspectiva anual dos dados. Na implementação do modelo, essa técnica exibe os dados a partir da seleção de uma fonte de energia e é responsável por exibir informações de energia de qualquer conjunto de dados coletado em um intervalo de tempo específico não superior a um ano. Por exemplo, 8.760 registros de um ano (8.784 registros para anos bissextos) podem ser exibidos, representando o valor da energia de cada hora por dia durante um ano. Devido ao número de registros a serem visualizados e a quantidade máxima usada em espaço na tela, tornou-se necessário o uso dessa técnica para mostrar dados de energia e dados meteorológicos. A Figura 4.8 como mostra um exemplo de dados de consumo de energia.

Para interpretar esta visualização, a seguinte análise deve ser feita: cada hora é representada por uma célula, que exibe o valor dos dados de consumo de energia. Este valor corresponde a uma escala de cores, que começa com uma cor azul forte representando baixo consumo de energia e termina com uma cor vermelha escura representando os dados de consumo de energia mais altos. Outras cores, como amarelo e laranja, representam valores intermediários entre baixo e alto consumo.

Outro recurso disponível para este gráfico é a seleção de um determinado período e a possibilidade de filtrar dados por um valor específico no conjunto de dados. Ao selecionar um período, o efeito *zoom in* será realizado, permitindo uma visão mais precisa dos valores selecionados. As células no gráfico aumentarão proporcionalmente ao tamanho da área selecionada, ou seja, quanto maior a área selecionada, menor o tamanho da célula, como mostra a figura 4.9. O efeito de *zoom out* ocorre somente quando houver a seleção de todo conjunto.

Outra opção fornecida pelo modelo é a exibição das informações a partir de um valor específico. Isto permite mostrar dados iguais ou superiores ao valor especificado, como ilustrado na figura 4.10.

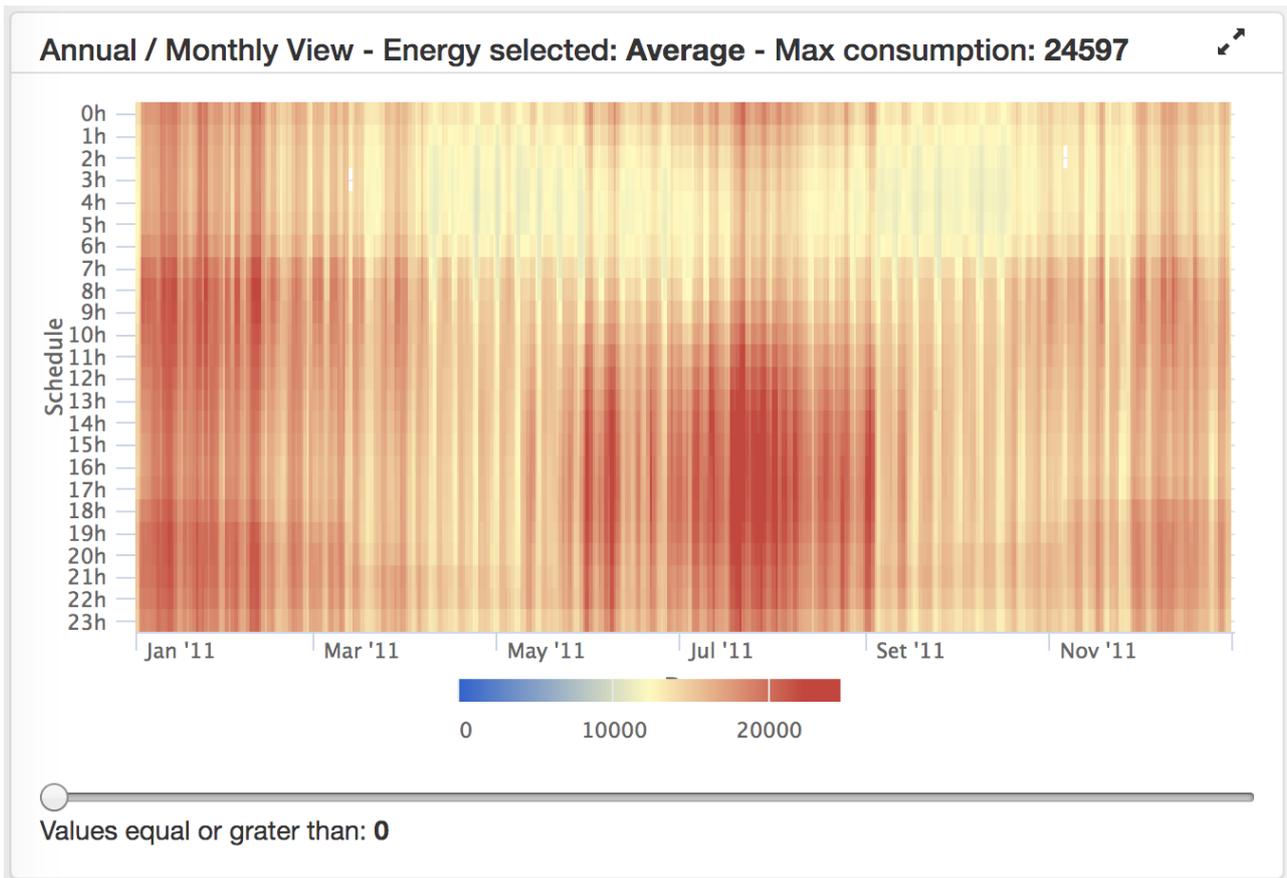


Figura 4.8 – Representação dos dados de consumo de energia em um ano: as horas de um dia são exibidas no eixo vertical, e os meses ao longo do ano são mostrados no eixo horizontal.

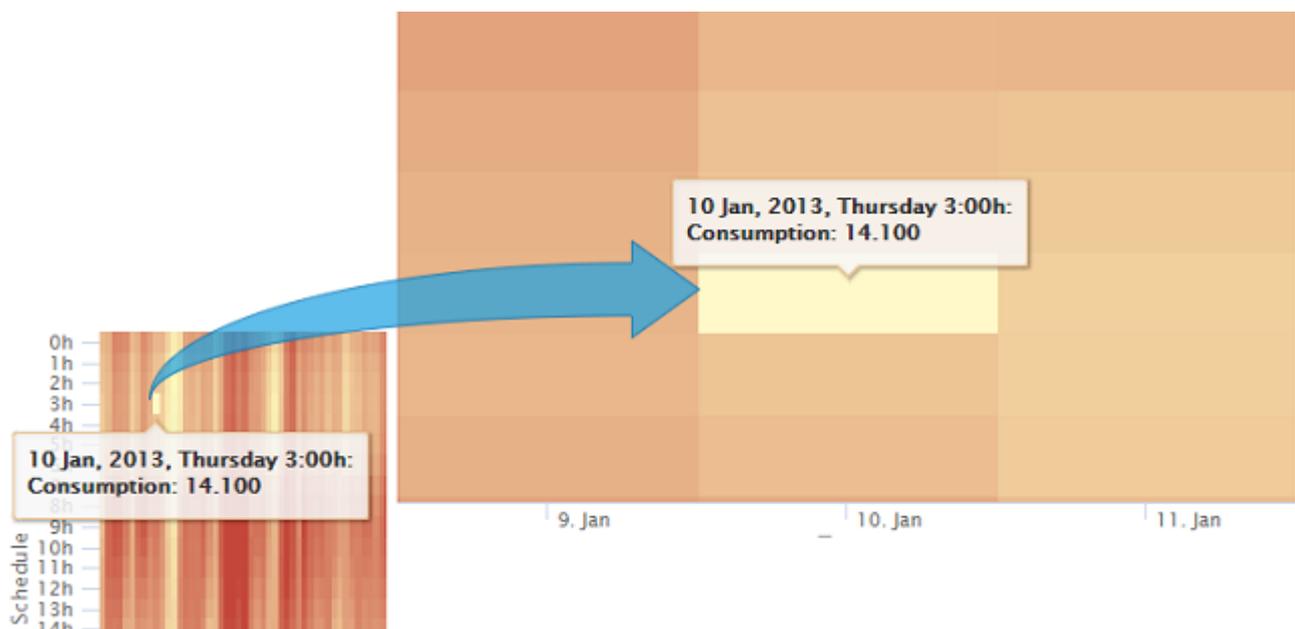


Figura 4.9 – Resultado de um *zoom in*, após a seleção de uma área no componente.

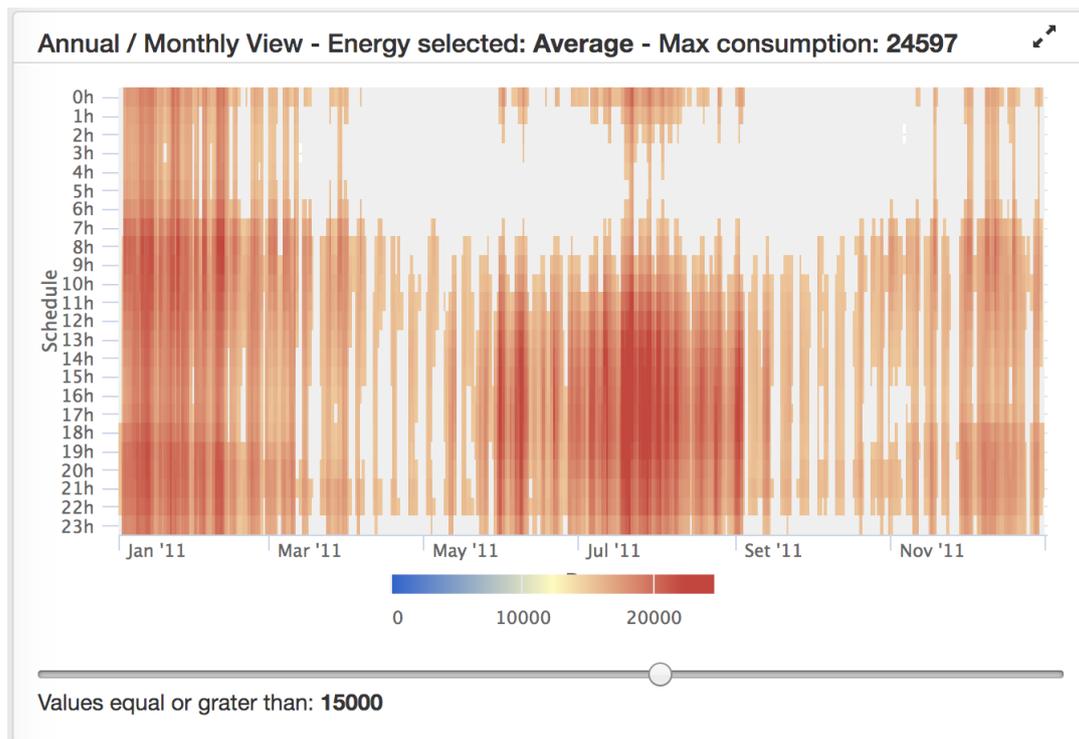


Figura 4.10 – Exibição dos dados filtrados iguais ou superiores ao valor contido no *range*.

4.5.2 Gráficos para Dados Mensais

Boxplot é uma técnica de visualização que permite a análise exploratória de dados envolvendo o uso de técnicas estatísticas para identificar padrões que podem estar ocultos em um conjunto de dados numéricos. É usado para resumir visualmente e comparar grupos de dados. O *Boxplot* usa a mediana, os quartis aproximados e os dados mais baixos e mais altos para transmitir o nível, a dispersão e a simetria de uma distribuição de valores de dados [120]. Ele também pode ser usado para facilmente identificar *outliers*. Além disso, tornou-se uma técnica padrão para apresentar cinco informações: valores mínimo e máximo do intervalo, os quartis superior e inferior e a mediana [86]. Mais detalhes dessa técnica foram apresentados na Seção 2.2.4.

Considerando as características oferecidas pelo *Boxplot*, ele foi aplicado no modelo para visualizar dados em horas e em dias a partir de uma semana selecionada e para visualizar o valor energético total mensal, como mostra a Figura 4.11. Inicialmente, são apresentados os consumos mensais ao longo do ano. O modelo oferece um filtro que permite gerar a mesma visualização para uma hora selecionada.

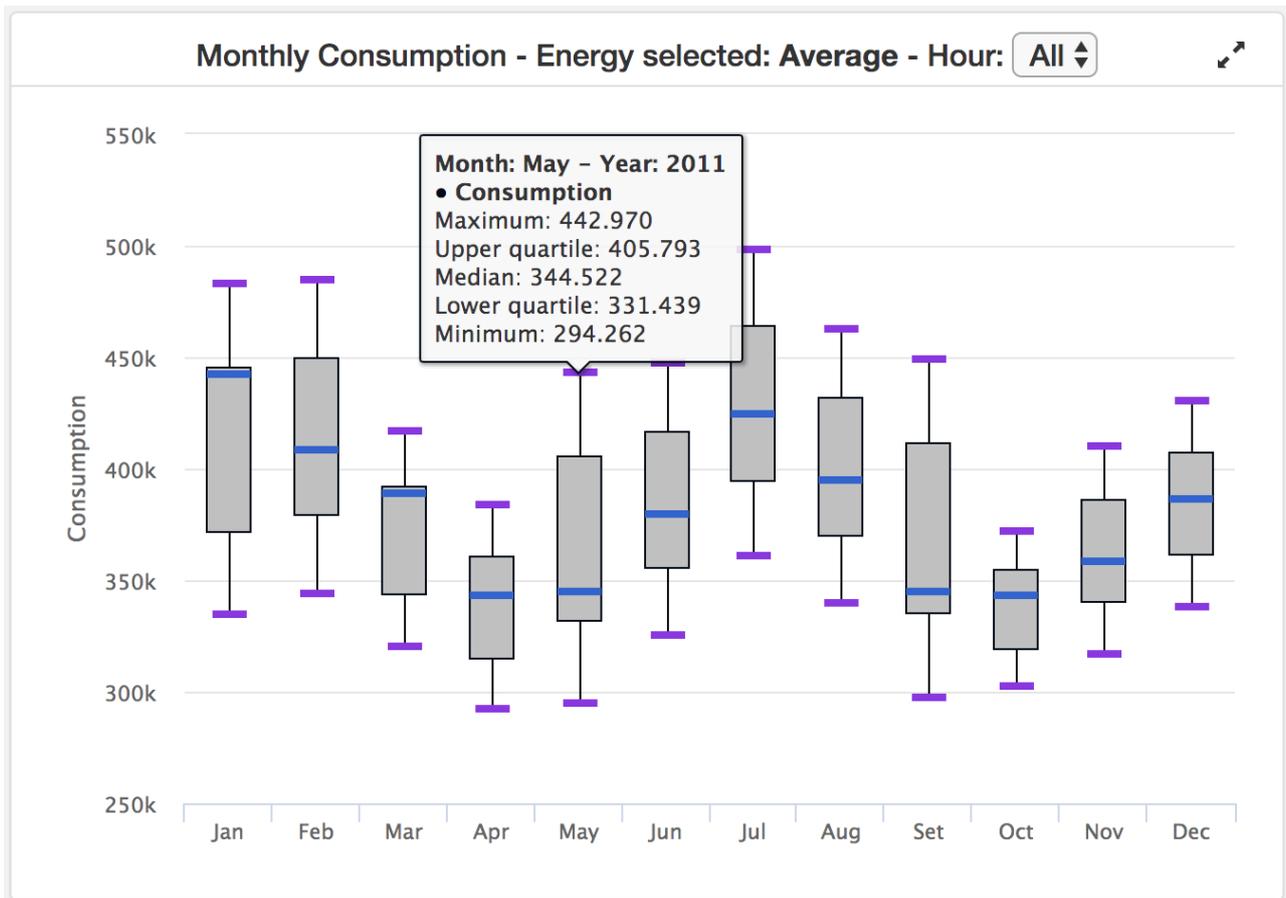


Figura 4.11 – *Boxplot* exibindo o consumo energético mensal.

4.5.3 Gráficos para Dados Semanais

Para mostrar os dados como uma granularidade semanal, foi escolhida a técnica de visualização *Heatmap*, ou mapa de calor, que é útil para o desenvolvimento de ferramentas de análise visual [27] (informações sobre esta técnica são apresentados na Seção 2.2.3). A aplicação desta técnica no modelo possibilita a exibição dos dados em um período semanal, colaborando para atender o requisito R4. Ou seja, enquanto o *Dense Pixel Display* mostra os dados de um ano, o *Heatmap* exibe os dados de uma semana, possibilitando compreender os comportamentos mencionados anteriormente (padrões, tendências, *outliers* e anomalias) em um período semanal.

Neste componente, são exibidos 168 registros que representam o consumo hora a hora, em uma semana, como mostra a figura 4.12. Neste caso, o eixo vertical contém as horas ao longo do dia e os dias da semana são exibidos no eixo horizontal. Dado o número de registros a serem exibidos, o uso do mapa de calor tornou-se importante, pois oferece vantagens sobre outras representações considerando o uso do espaço de exibição. Além disso, também é possível inserir marcadores no *Heatmap*, como mostram as Figuras 4.13

e 4.14. Estes marcadores identificam os horários de maior e menor de energia durante a semana selecionada.

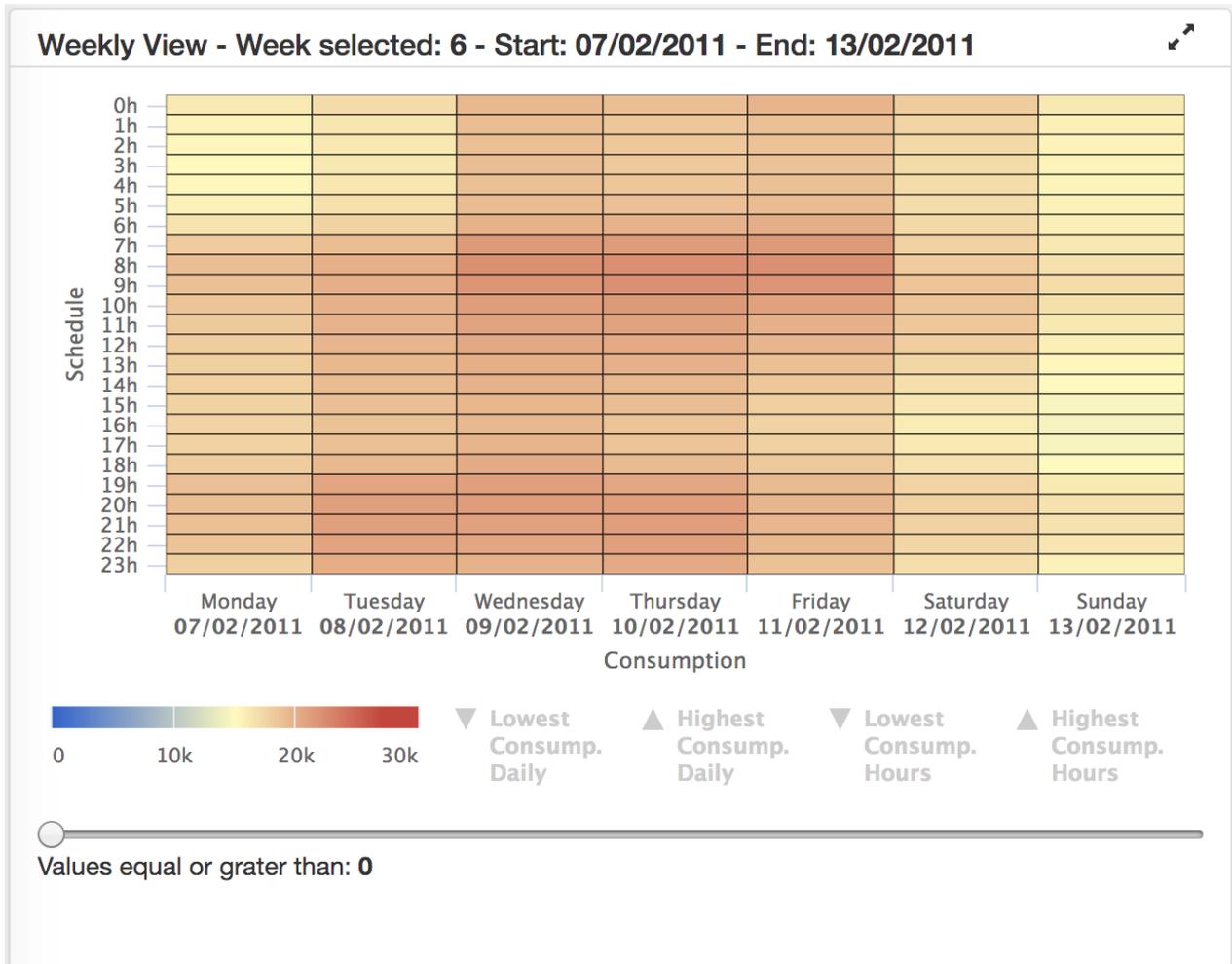


Figura 4.12 – Mapa de calor mostrando dados de energia de uma semana.

Os marcadores permitem realizar análises comparativas e mostrar ao usuário quais são os horários de maior consumo ou geração de energia. A Figura 4.13 mostra os marcadores posicionados nos horários de maior e menor consumo durante uma semana. São exibidos 48 marcadores no gráfico, sendo 24 para os horários de menor consumo (independente do dia da semana) e 24 para os horários de maior consumo. De maneira análoga, na Figura 4.14 também são apresentados os horários de maior e menor consumo para cada dia de uma semana. Neste cenário, são apresentados 14 marcadores no gráfico, sendo 7 para as horas de menor valor e 7 para os picos de energia em cada dia da semana.

4.5.4 Gráficos para Dados Diários e em Horas

Outros gráficos, como gráficos de linha, pizza e barra, podem ser usados para identificar padrões frequentes e anormais [43, 71]. Esses gráficos foram incorporados ao

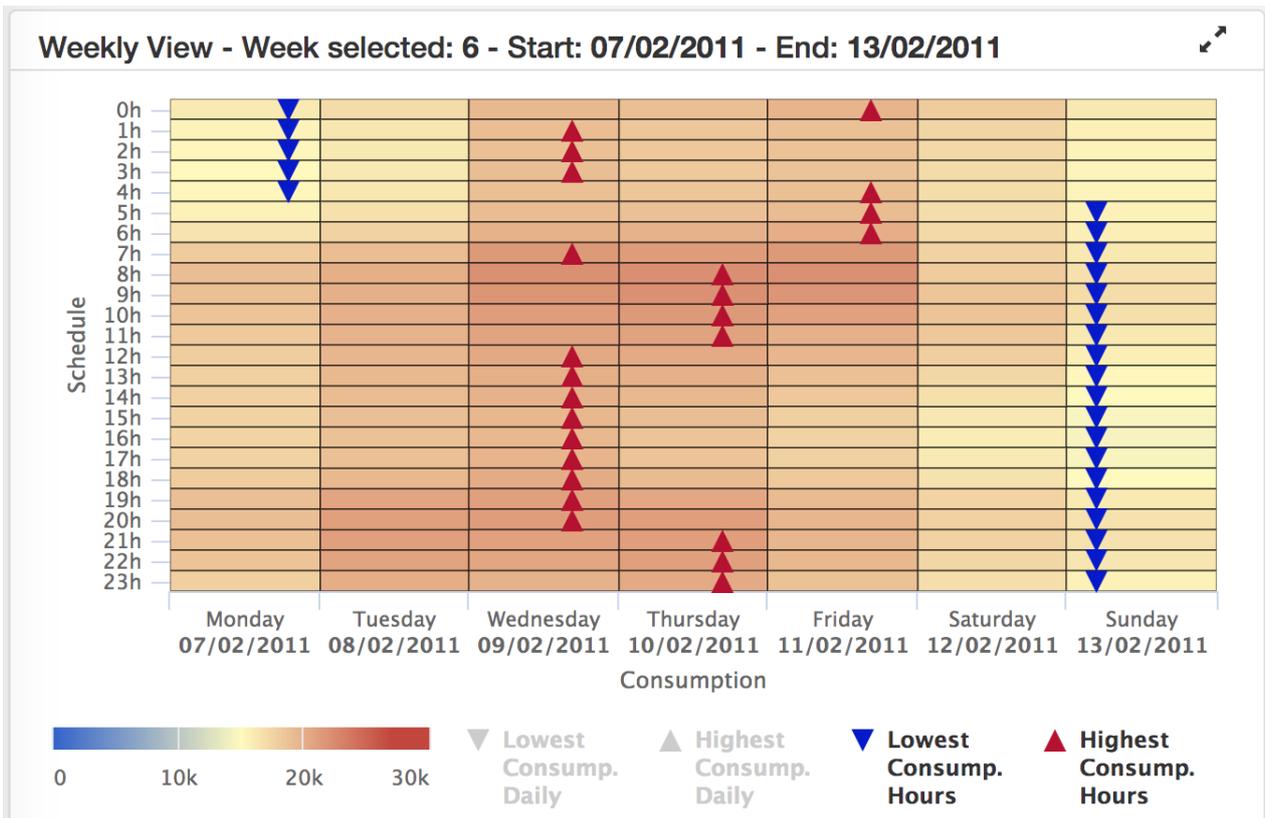


Figura 4.13 – Mapa de calor exibindo os marcadores que identificam em quais horas de uma semana houve o maior (triângulo vermelho) e o menor valor de energia (triângulo azul).

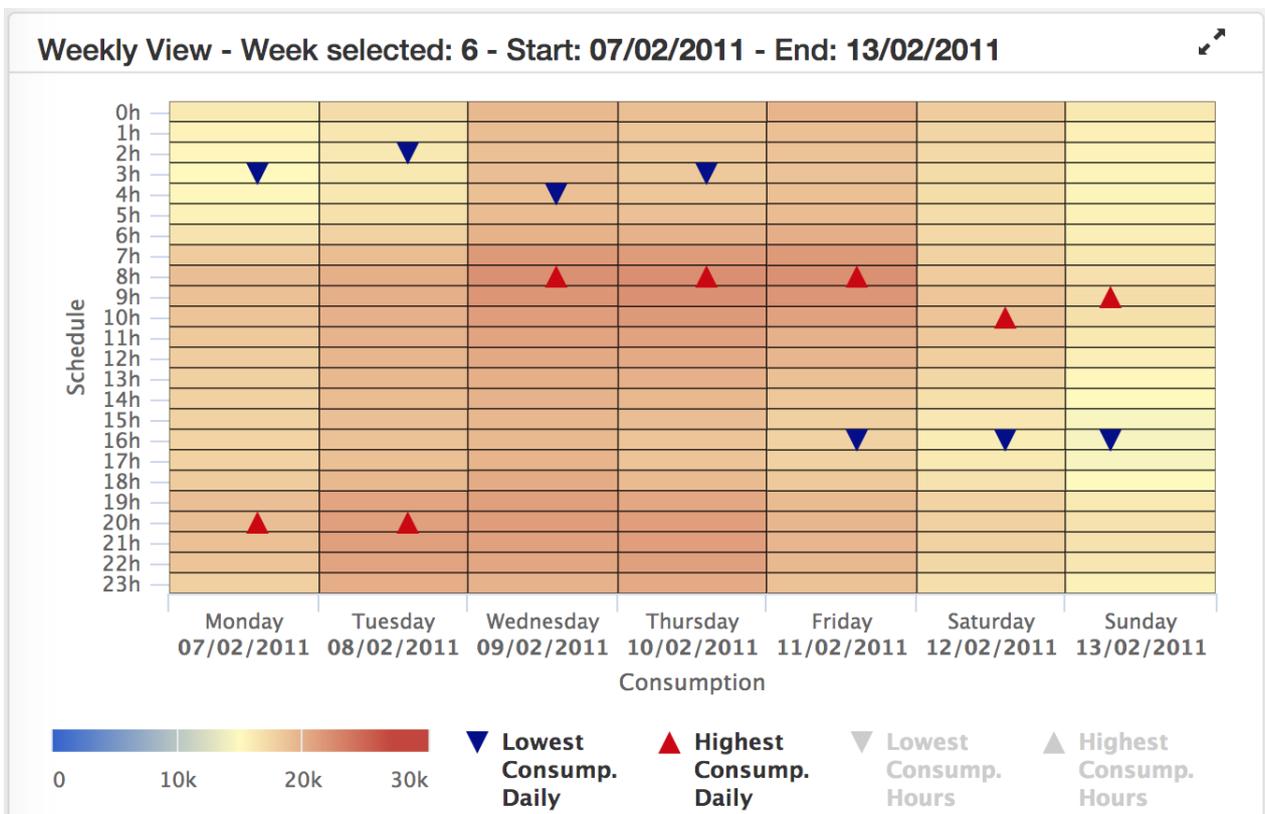


Figura 4.14 – Mapa de calor com os marcadores que mostram o horário de maior (triângulo vermelho) e menor valor de energia (triângulo azul) em cada dia de uma semana.

modelo para permitir realizar uma análise na última granularidade que o modelo proposto oferece, que é a visualização dos dados por dia e por hora. Seguindo o propósito de oferecer diferentes estágios de granularidade dos dados ao longo do tempo, estas visualizações são exibidas a partir da seleção de um dia da semana no *Heatmap*, ou seja, parte de um período semanal para uma visualização diária e por hora. A Figura 4.15 ilustra os mesmos valores de consumo em diferentes gráficos, sendo no eixo horizontal o consumo ou geração de energia e no eixo vertical as escalas de valores.

Nestes gráficos estão incluídas algumas funções estatísticas para auxiliar na análise e na compreensão dos padrões de consumo da semana selecionada. A Figura 4.16 apresenta a visualização do consumo tanto para dias, como para horas, e os seguintes valores estatísticos são mostrados:

- Mediana: Conforme relatam Bussab e Morettin [77, p 35], “A mediana é a realização que ocupa a posição central da série de observações, quando estão ordenadas em ordem crescente”.
- Variância: Segundo Morettin [76, p 49], a variância é “a medida de dispersão (ou de concentração) de probabilidade em torno da média”.
- Desvio padrão: Conforme Triola [111, p 38], “O desvio-padrão de um conjunto de valores amostrais é uma medida da variação dos valores em relação à média”.
- Erro padrão: Segundo descreve Montgomery [75, p 225], “O erro padrão é uma estimativa do desvio padrão. Se o erro padrão envolve parâmetros desconhecidos que podem ser estimados, a substituição desses valores produz um erro padrão estimado”.
- Coeficiente de variação: Conforme afirma Brown [20, p 155] “O coeficiente de variação é atraente como ferramenta estatística, porque aparentemente permite a comparação de variáveis livres de efeitos de escala”.

O uso destas medidas no modelo auxilia a identificar se os valores apresentados em um conjunto de dados estão dispersos ou não, e o quão distantes eles podem estar um do outro. A variância, o desvio padrão, o erro padrão e o coeficiente de variação foram aplicados para análise do consumo (suportando o requisito R6). Estas informações são exibidas quando é selecionado qualquer dado de consumo no gráfico, conforme mostra a Figura 4.16. Como estas informações foram aplicadas considerando o consumo de energia no período do conjunto de dados, independente do dado selecionado, elas serão as mesmas, não fazendo nenhuma referência a hora do consumo.

A utilização destas medidas no modelo foram aplicada da seguinte maneira:

- Para dias: Para exibir os valores da mediana, mínimos, médios e máximos, o modelo automaticamente busca todos os valores do mês selecionado e faz os cálculos, como

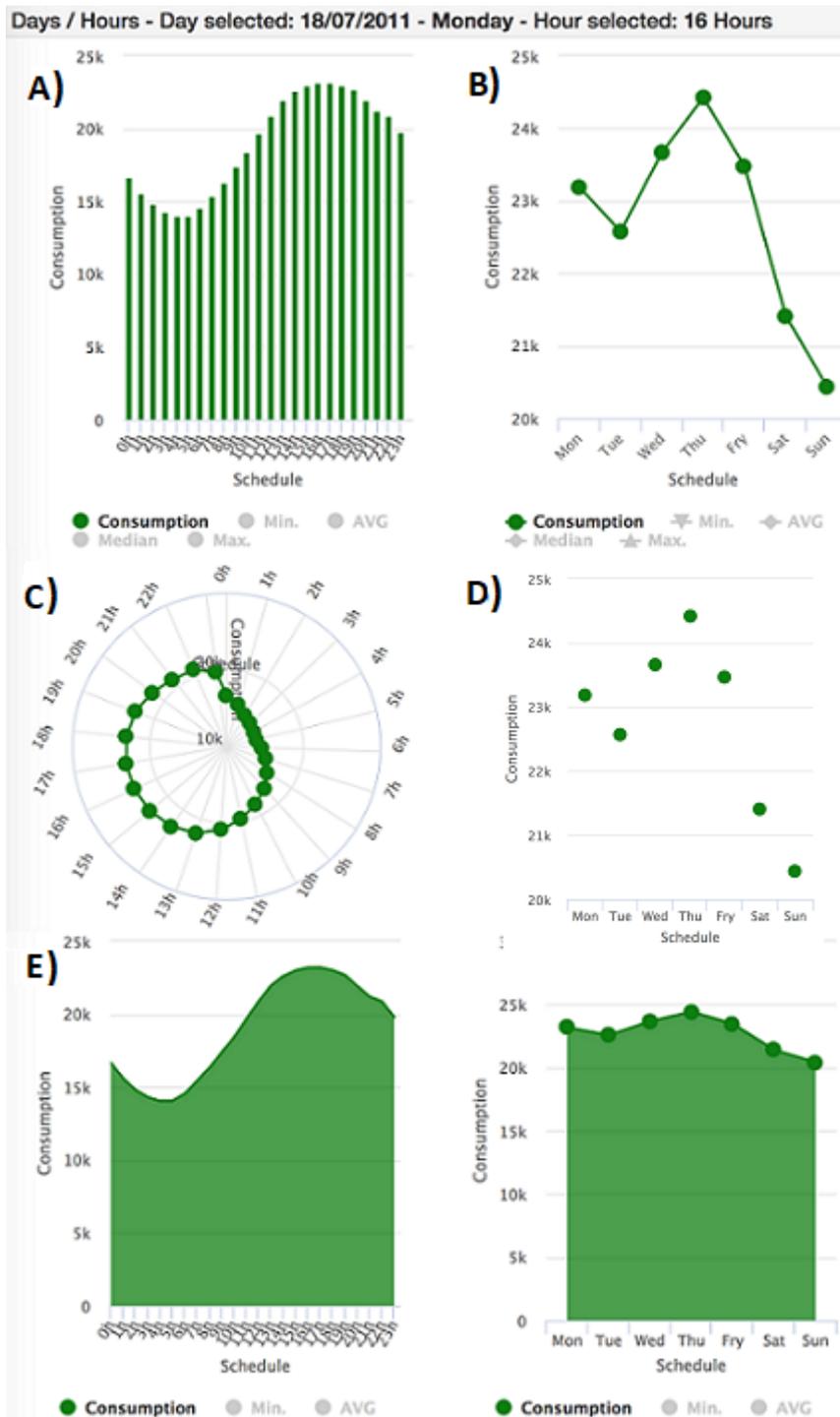


Figura 4.15 – Exibição dos gráficos para visualização de séries temporais, que são: A) Gráfico de barras, B) Gráfico de linhas, C) Gráfico polar (ou radar), D) Gráfico de plot, E) Gráfico de área.

é mostrado na Figura 4.17-A. Em outras palavras, os dados são obtidos através do mês selecionado para depois calcular e exibir os valores de mínimo, média, mediana e máximo deste mês;

- Para horas: Neste caso, o modelo automaticamente identifica os dias da semana do mês selecionado e realiza os cálculos, conforme mostra a Figura 4.17-B. Em outras

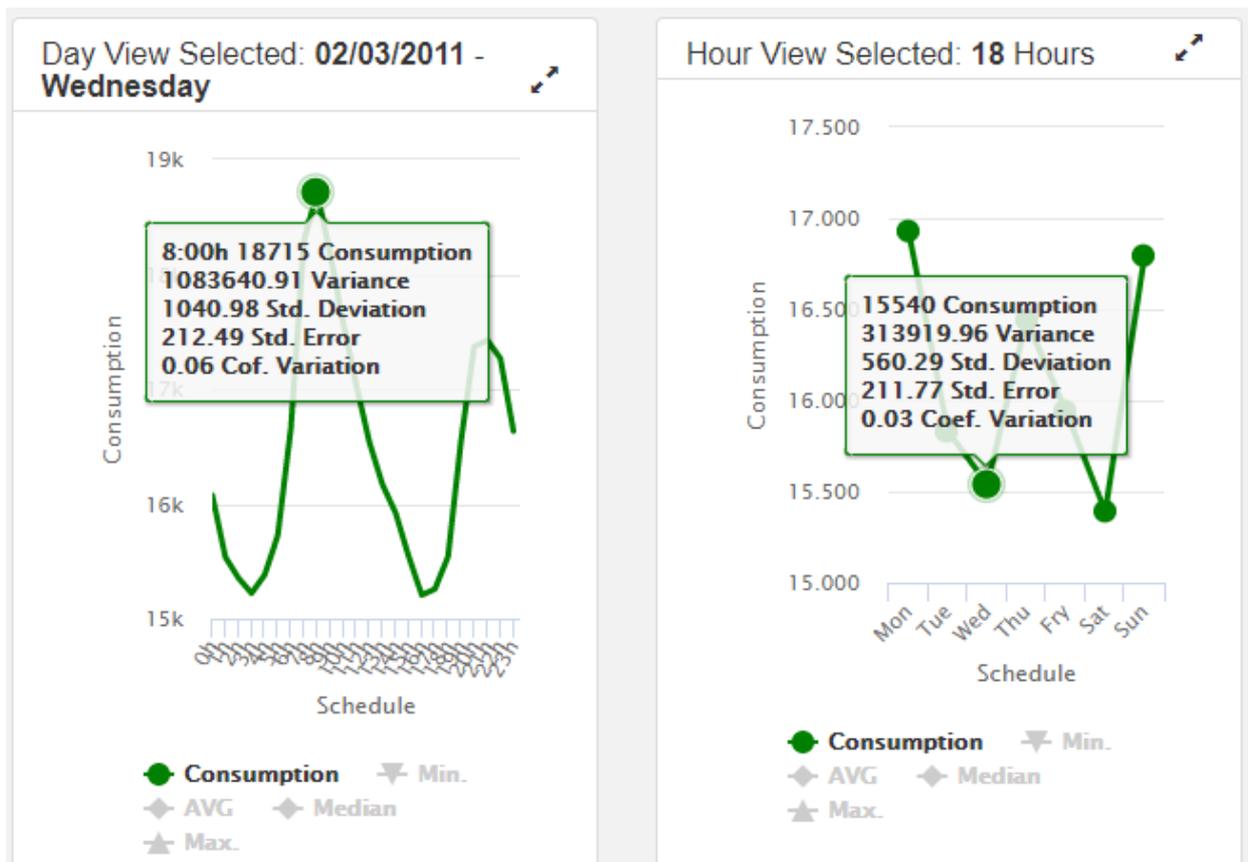


Figura 4.16 – Exibição dos gráficos de dias e horas mostrando valores de consumo de acordo com a seleção feita, e variância, desvio padrão, erro padrão e coeficiente de variação do conjunto de dados.

palavras, o modelo busca os dados por dias da semana do mês selecionado, calcula todos valores separadamente por dia (Por exemplo: calcula o mínimo, o máximo, a média e a mediana para cada dia da semana) e exhibe no gráficos.

Para ser gerada esta visualização no modelo é necessário selecionar diferentes estágios de granularidades. A Figura 4.18 ilustra este avanço. A estrutura do modelo começa em uma granularidade anual de dados apresentados pelos componentes *Calendar View* e *Dense Pixel Display*. Após a seleção de uma célula de qualquer uma destas técnicas, como mostra a Figura 4.18-A e 4.18-B, o *Heatmap* mostra os dados semanais correspondentes à semana daquela célula. Dessa forma, o modelo avançou da perspectiva anual e mensal atingindo a granularidade semanal. As etapas na Figura 4.18-C e 4.18-D representam o avanço do estágio de diferentes granularidades. Neste caso, saindo de uma perspectiva semanal e indo para uma visualização em dias e horas.

A figura 4.19 permite a análise do consumo mensal de todos os conjuntos de dados carregados. O modelo automaticamente resume todo o consumo mensal em diferentes anos através da fonte de energia selecionada. Neste exemplo, seis anos do mesmo conjuntos de dados foram carregados e o modelo apresenta a soma mensal dos dados de energia para cada ano. O último bloco desenvolvido no modelo apresenta separadamente o con-

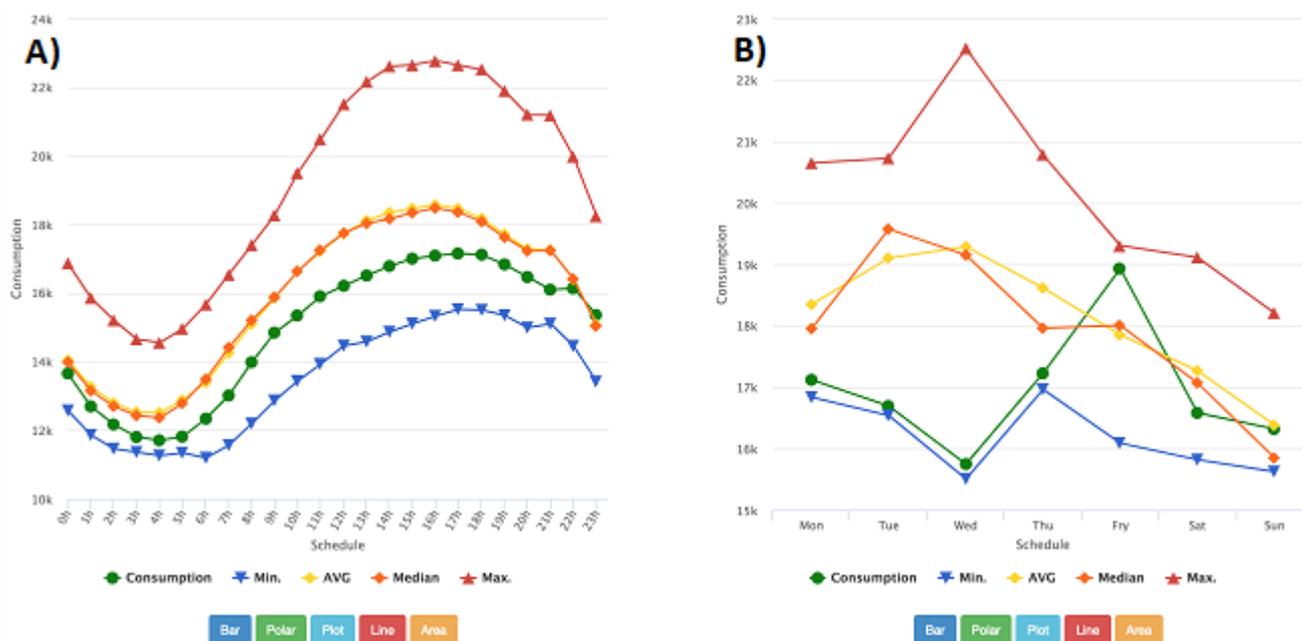


Figura 4.17 – Exibição dos gráficos de dia e horas mostrando valores de consumo, mínimos, médios, medianas e máximos.

sumo anual de todas as energias contidas no conjunto de dados para o ano selecionado, conforme exibe a figura 4.20.

4.5.5 Análises comparativas

Realizar comparações entre diferentes fontes de consumo ou geração de energia possibilita realizar diversas análises. Para analisar estes dados ao longo do tempo, é interessante possibilitar uma fácil interação entre diferentes visualizações. No presente modelo, esse recurso foi desenvolvido visando explorar conjuntos de dados distintos em diferentes anos e em diversos períodos ao longo do ano (atendendo o requisito R5).

O recurso de comparação de diferentes fontes de energia foi implementado no modelo através de janelas *pop-up* móveis, com a possibilidade de customizar o seu tamanho de acordo com a preferência do usuário e interagir com os dados de várias formas. Como não há um limite para a criação de janelas no modelo, inúmeros *pop-up* de diferentes fontes de energia podem ser criados. Porém, neste caso, para uma análise visual eficaz, é necessário ter um monitor grande. Somente assim será possível criar e colocar lado a lado várias janelas *pop-up*.

Para fazer uma comparação no modelo é necessário selecionar um dia ou horário específico no *Calendar view* ou no *Dense Pixel Display* e será mostrado quais os períodos que o usuário pode visualizar os dados. No *Calendar view*, as opções de períodos disponíveis são mensal, semanal e diário, conforme ilustra a Figura 4.21. No *Dense Pixel Display*

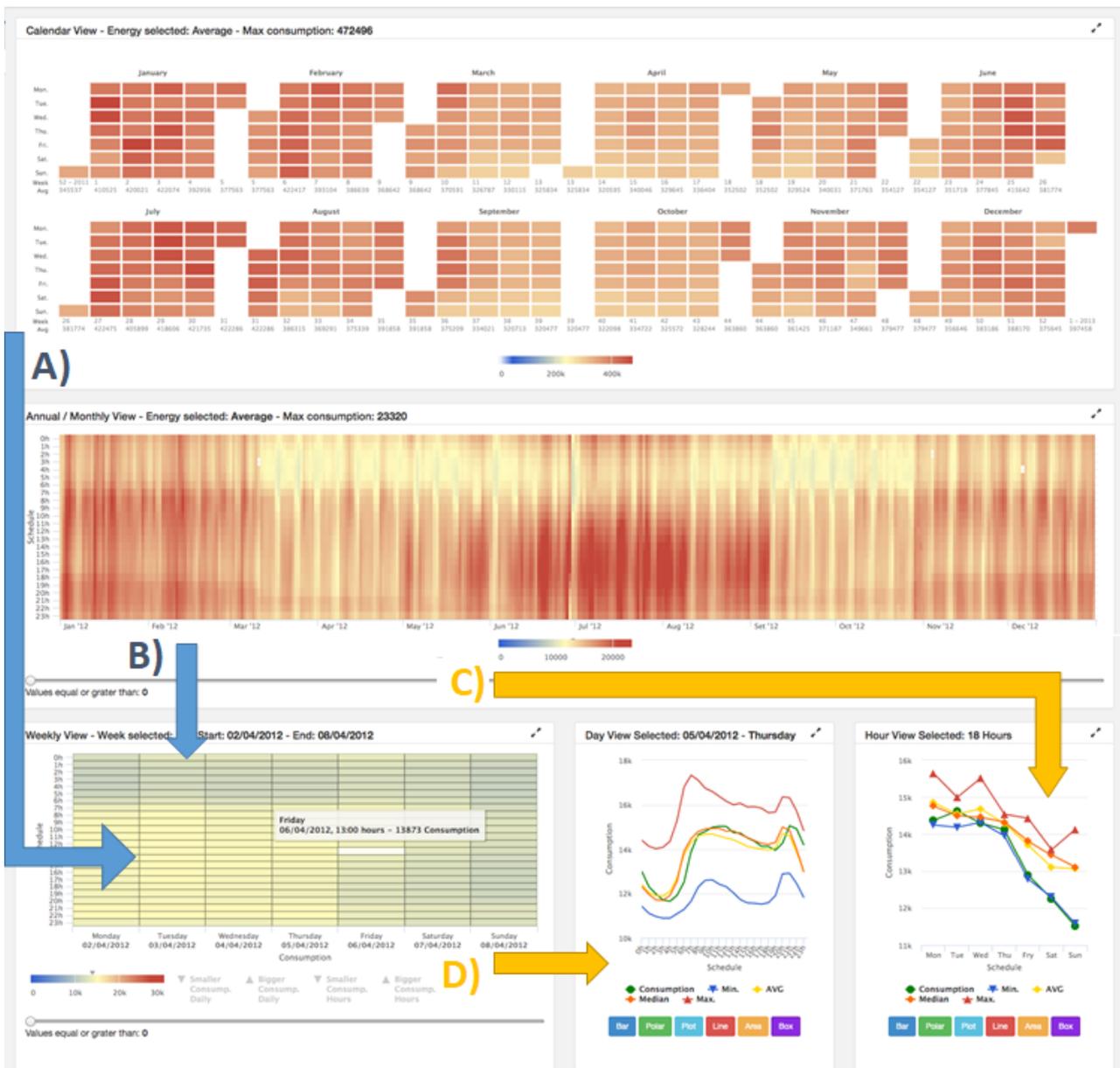


Figura 4.18 – Fluxo de exibição para visualização com diferentes granularidades de tempo: A) Seleção de um dado a partir do *Calendar View*; B) Seleção de um dado a partir do *Dense Pixel Display*; Qualquer seleção de uma das etapas anteriores carregam os dados semanais. A partir da seleção de uma célula no *Heatmap*, as etapas C e D ocorrem, exibindo os dados em dias e horas.

são exibidas as mesmas opções com a adição da opção em horas, conforme ilustra a Figura 4.22.

Após a escolha do período desejado para visualizar os dados, uma nova janela *pop-up* é exibida com os dados carregados. Não há limites para criar novas janelas *pop-up*. Com isso, é possível gerar visualizações de diferentes períodos ao longo do ano, de diversas fontes de energia e/ou dados meteorológicos e de conjuntos de dados distintos. Para exemplificar a análise comparativa do modelo, foram criadas duas tarefas através dos conjuntos de dados deste estudo (ver subseção 5.1).

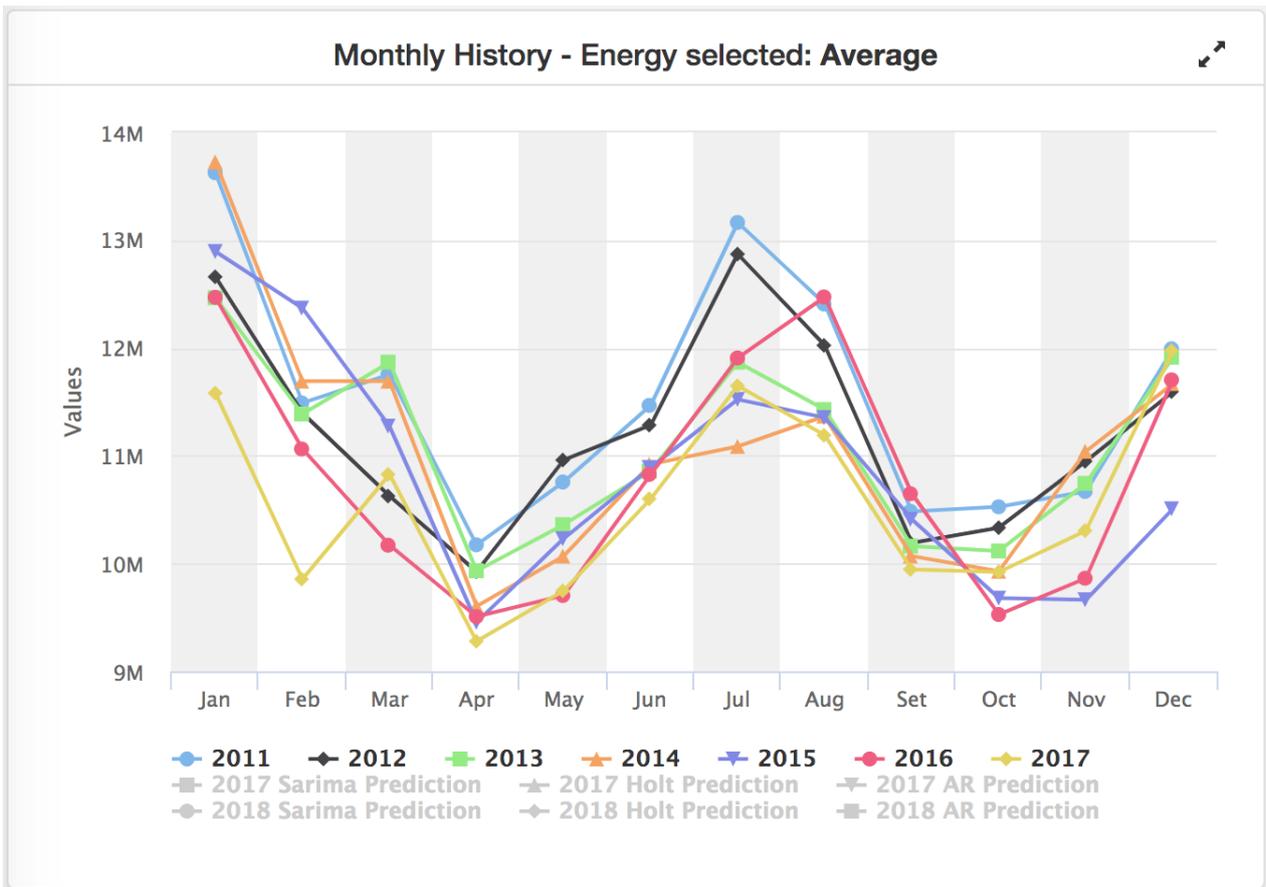


Figura 4.19 – Exibição dos gráficos para análise de consumo anuais.

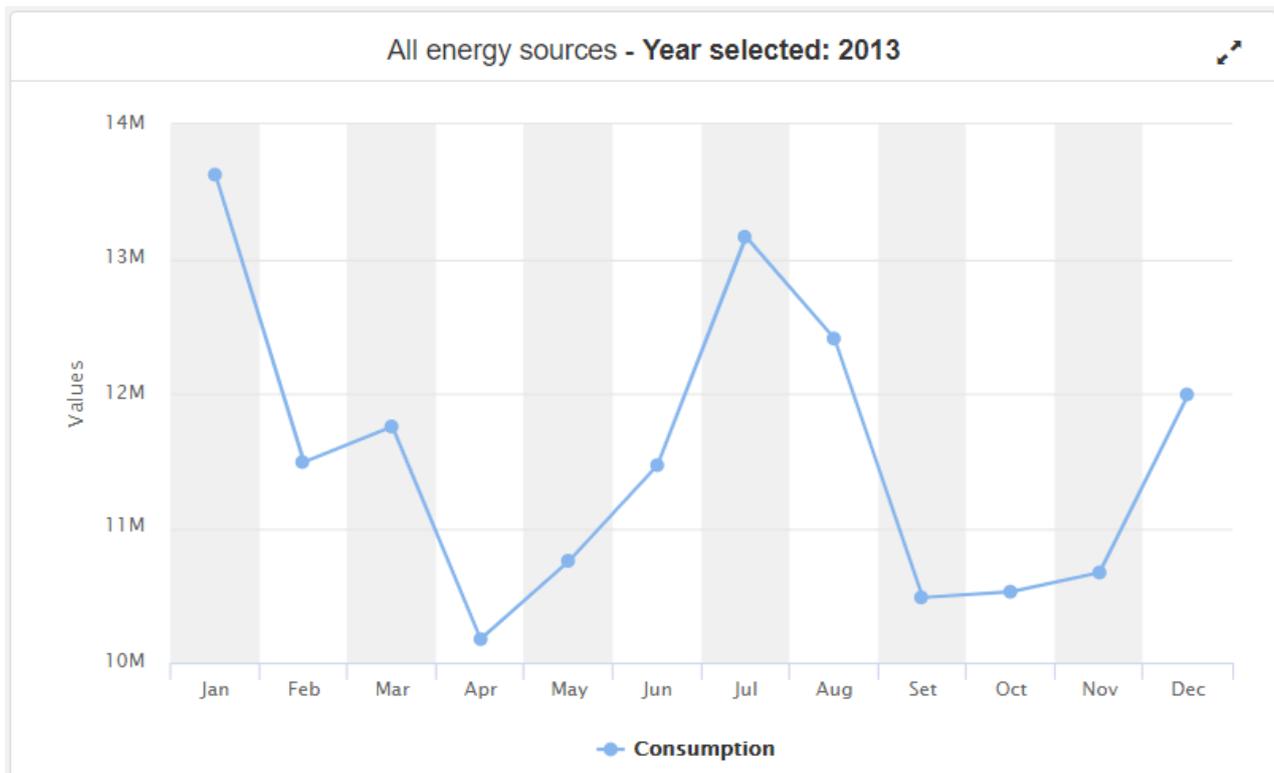


Figura 4.20 – Exibição dos energias do conjuntos de dados, neste caso há apenas uma fonte de energia.

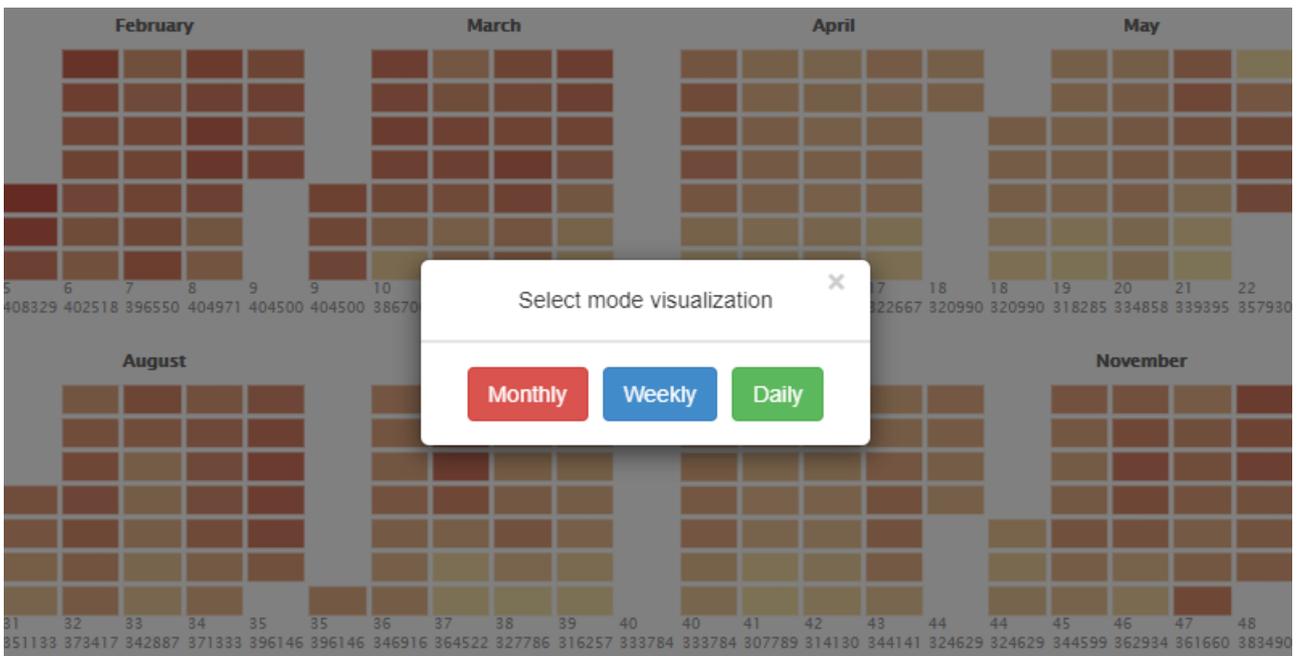


Figura 4.21 – Janela *pop-up* mostrando as opções disponíveis (mensal, semanal e diário) para visualizar os dados a partir do *Calendar View*.

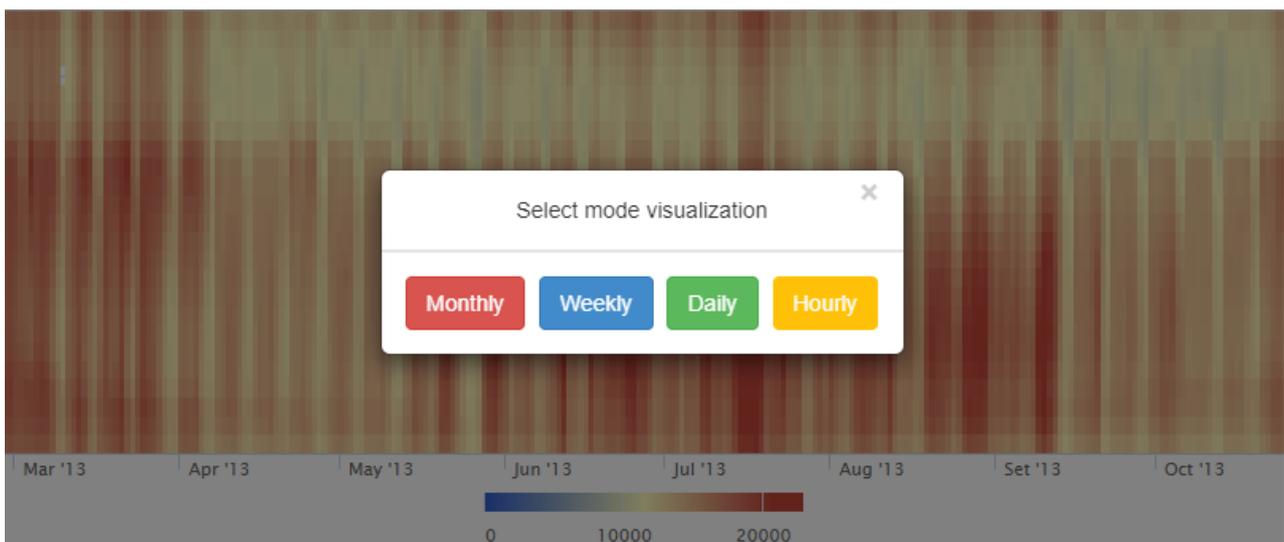


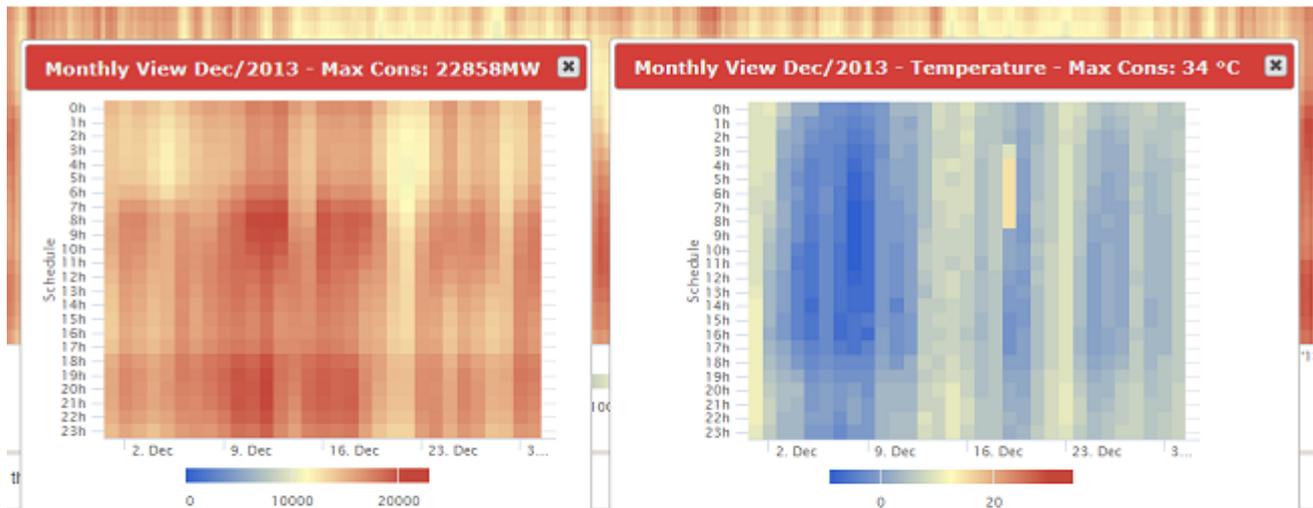
Figura 4.22 – Janela *pop-up* mostrando as opções disponíveis (mensal, semanal, diário e horas) para visualizar os dados a partir do *Dense Pixel Display*

Para ambas tarefas foram geradas duas janelas *pop-up*. No lado esquerdo da Figura 4.23-A são mostrados os dados de consumo de energia do mês de dezembro do ano de 2013 do conjunto de dados A. Do lado direito, exibe dados meteorológicos de temperatura do mesmo mês e ano do conjunto de dados C. Ao analisar os dados, podemos afirmar que a primeira quinzena deste mês foi mais fria e conseqüentemente houve maior consumo de energia no inverno americano, que é mostrado na janela à esquerda.

Na Figura 4.23-B foram utilizados dados do conjunto de dados B. Do lado esquerdo são mostrados os dados da fonte de energia eletricidade do mês de janeiro do ano de 2004.

Do lado direito, exibe dados da fonte de energia aquecimento à gás. Ao analisar essas janelas *pop-up*, podemos verificar que não há consumo de eletricidade e aquecimento à gás no prédio comercial entre 1 e 4 horas da manhã. Além disso, verificamos que houve um consumo constante de eletricidade em diferentes faixas de horários, diferentemente do aquecimento à gás, que na primeira quinzena apresentou um consumo maior.

A)



B)

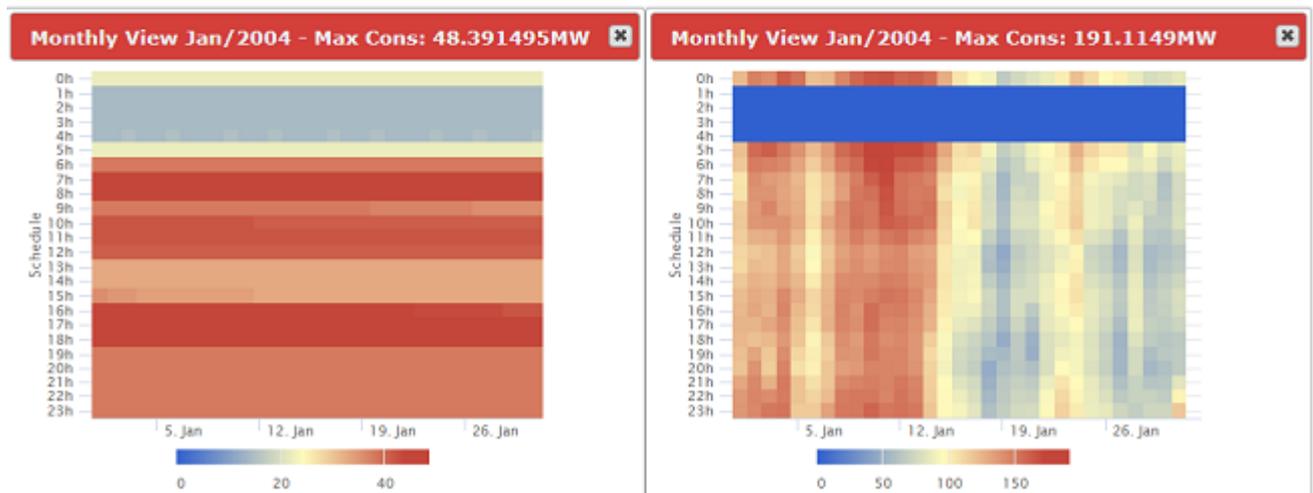


Figura 4.23 – Apresentação das janelas *pop-up* em diferentes exemplos e mostrando dados de diferentes conjuntos de dados.

4.5.6 Predição de consumo

Para incluir os algoritmos de predições no modelo, foi realizada uma análise para identificar em qual gráfico seria mais adequado incorporar a funcionalidade de predição.

Após esta análise, foi identificado que a melhor abordagem seria incluí-los no gráfico de dados mensal, o qual apresenta a soma mensal de cada dado de energia, como mostra a Figura 4.19-A. Foi adotada esta estratégia porque este gráfico não exibe uma grande quantidade de dados, possibilitando um maior foco nos dados de predição (suportando o requisito R7).

Posteriormente, foi analisada qual seria a melhor forma de gerar as predições destes valores. Por exemplo, no caso do conjunto de dados A utilizado neste trabalho (ver subseção 5.1) que possui dados coletados de 2011 à 2017, deve ser feita a predição dos valores de consumo para o ano de 2018, além de descobrir a taxa de assertividade, o que é um desafio. Um importante item a ser destacado é que o modelo gera predições apenas para o ano seguinte do conjunto de dados informado. Por exemplo, se fosse informado ao modelo um conjunto com dados de 2010 à 2015, o modelo automaticamente identifica que deve ser gerado os valores de predição para o ano de 2016.

Após a realização de alguns testes, decidimos utilizar os dados de 2011 à 2016 e deixar o ano de 2017 para validação. Assim, ao gerar a predição para o ano de 2017, foi possível comparar os valores gerados com os dados reais existentes para o ano de 2017 e calcular o valor de erro.

Antes de definir quais algoritmos de predição poderiam ser utilizados no modelo, foi realizada uma pesquisa para compreender os comportamentos dos mesmos com o tipo de dado suportado (dados de séries temporais). Alguns algoritmos não tiveram um resultado satisfatório, como o algoritmo de predição *Arima*, que ao ser testado gerou predições com baixas taxas de assertividade. Devido a este critério, tanto o *Arima*, como outros modelos de predição, foram descartados para serem usados no modelo.

Os algoritmos que apresentaram maior taxa de assertividade e possibilitavam trabalhar com dados de séries temporais foram aplicados no modelo. Assim, foram selecionados os seguintes algoritmos de predição: *Autoregressive*, *Holt-Winter*, *Sarima* e *Prophet*. Algumas informações sobre cada um destes algoritmos foram apresentadas nas subseções 2.6.1 à 2.6.4. Para explicar a adoção deste processo, foram incluídas diferentes figuras, começando pela Figura 4.24 e terminando na Figura 4.28. No rodapé de cada gráfico são exibidos os anos dos conjuntos dos dados, bastando clicar em cada um para que o modelo possa esconder ou mostrar um conjunto específico, oferecendo diferentes formas de visualização. A Figura 4.24 exibe os dados do conjunto de dados A do ano de 2011 a 2016.

A Figura 4.25 exibe isoladamente os dados do conjunto de dados A do ano de 2017. A Figura 4.26 exibe os dados do conjunto de dados A do ano de 2017 junto com dados dos três algoritmos de predições. Nesta imagem, é possível analisar os dados que foram predicionados. Apesar da distância dos valores de predição no mês de fevereiro, nos outros meses há uma boa proximidade dos valores gerados em comparação com os valores

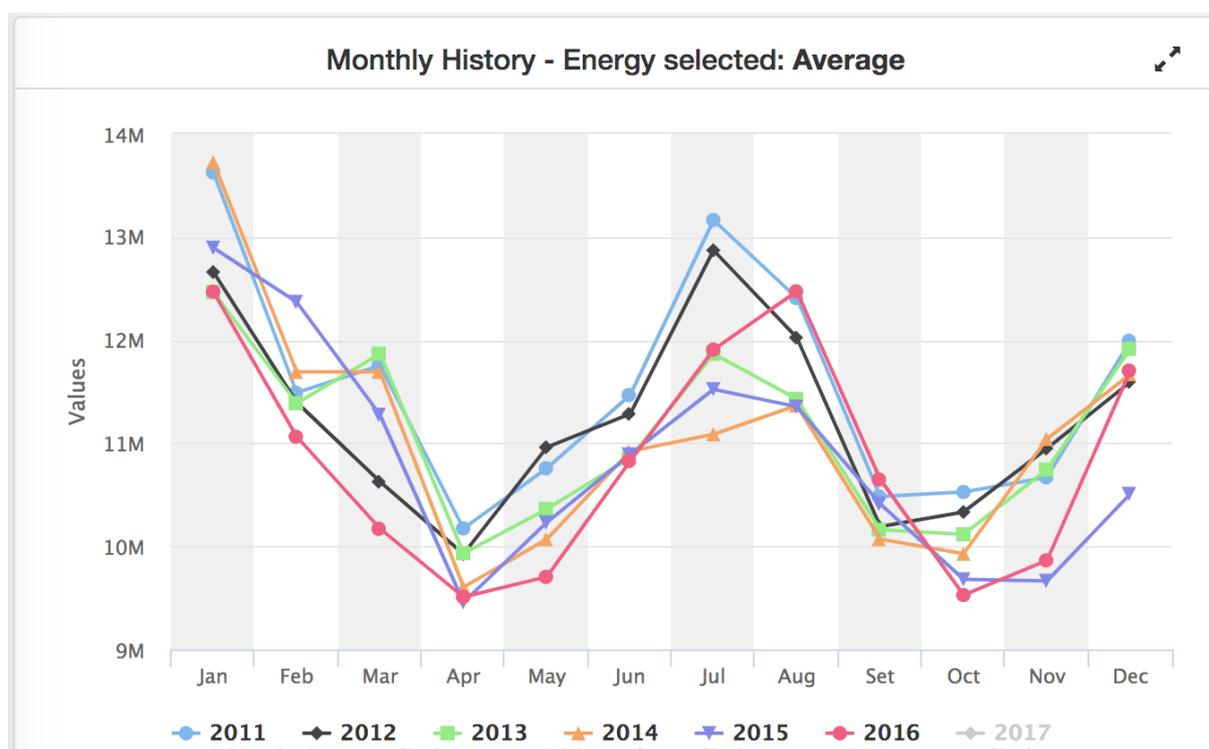


Figura 4.24 – Exibição do gráfico para análise de consumo anual dos anos de 2011 à 2016.

reais, como os meses de maio, junho, agosto, setembro, outubro e novembro. A Figura 4.27 apresenta as previsões geradas para o ano de 2018.

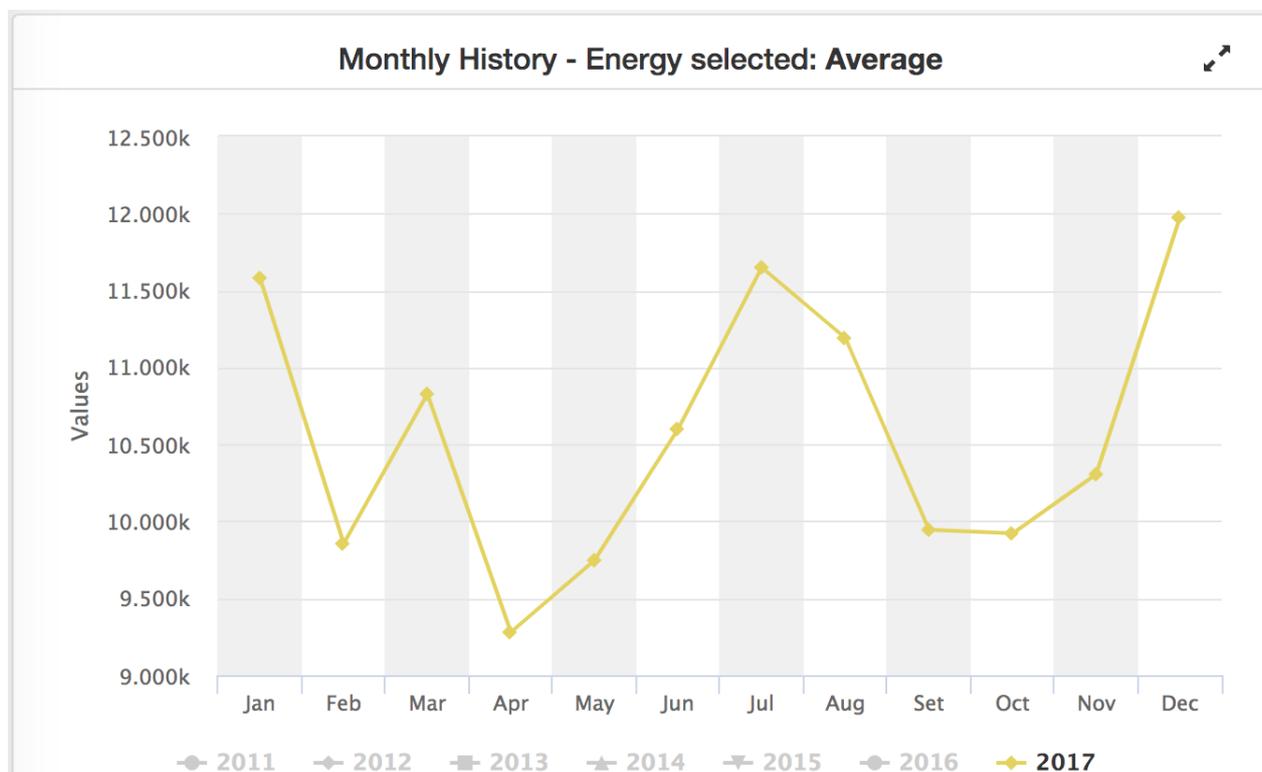


Figura 4.25 – Exibição do gráfico para análise de consumo anual do ano de 2017.

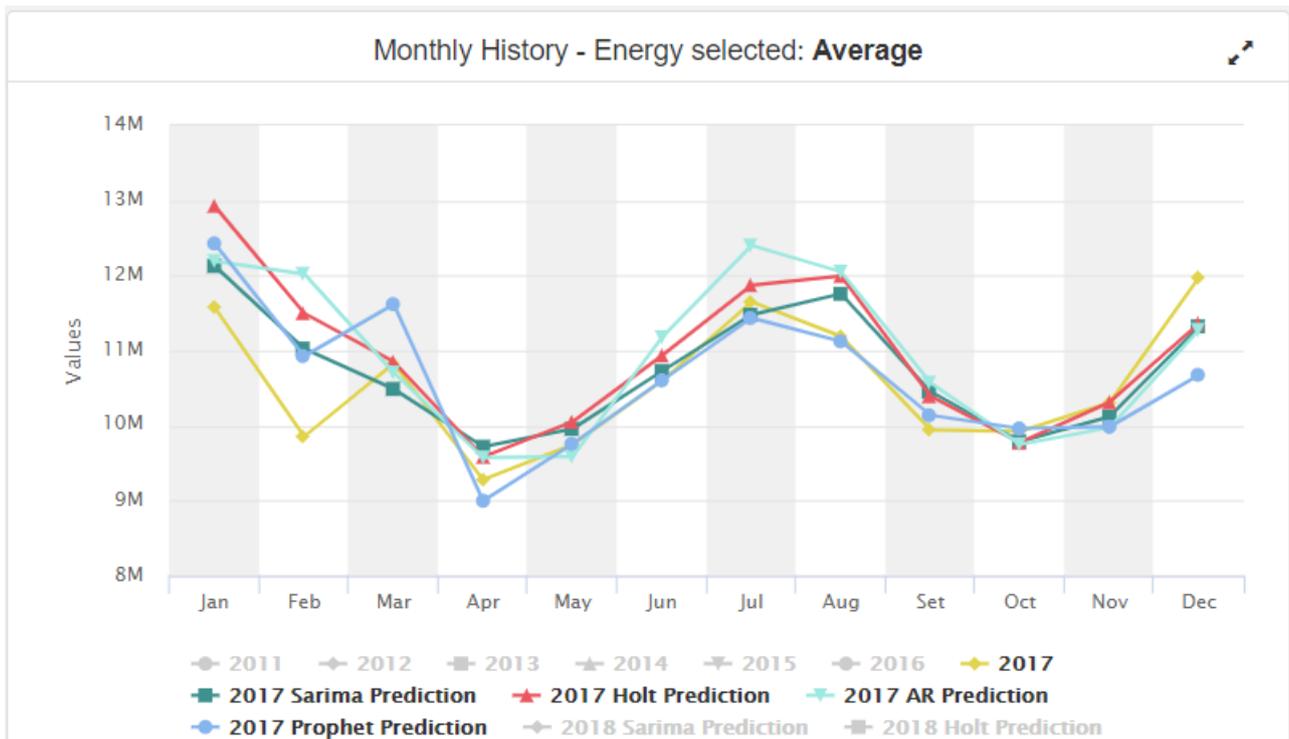


Figura 4.26 – Exibição do gráfico para análise de consumo anual do ano de 2017 juntamente com os valores de predição gerada pelos quatro algoritmos.

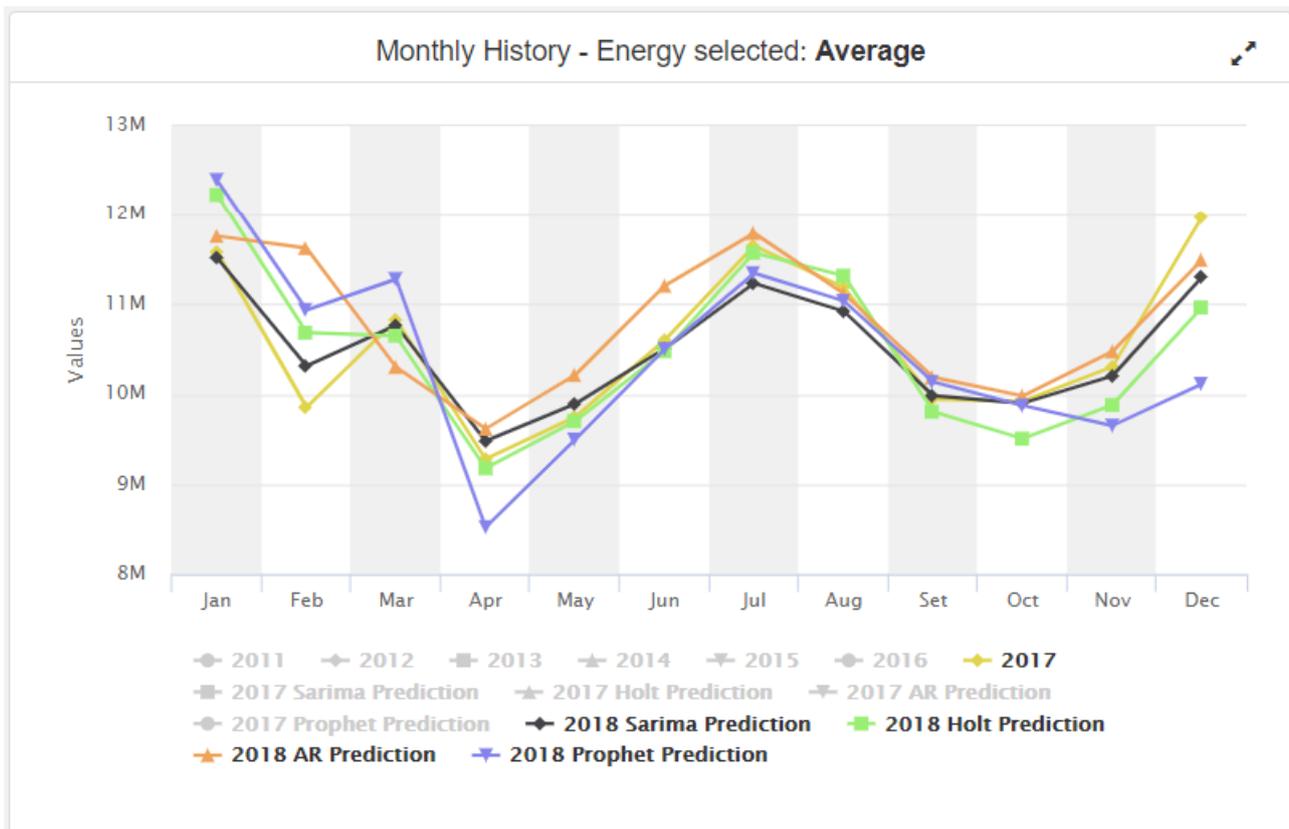


Figura 4.27 – Exibição do gráfico para análise das predições geradas para o ano de 2018.

A Figura 4.28 exibe todos valores do conjunto, as métricas mencionadas e os dados gerados pelas predições para um mês selecionado (no caso o mês de maio). Estas

métricas são informadas para cada mês neste gráfico, bastando apenas selecionar o mês que se deseja analisar. Neste exemplo, é possível notar que o algoritmo que teve melhor taxa de assertividade para o mês selecionado foi o algoritmo *Prophet*.

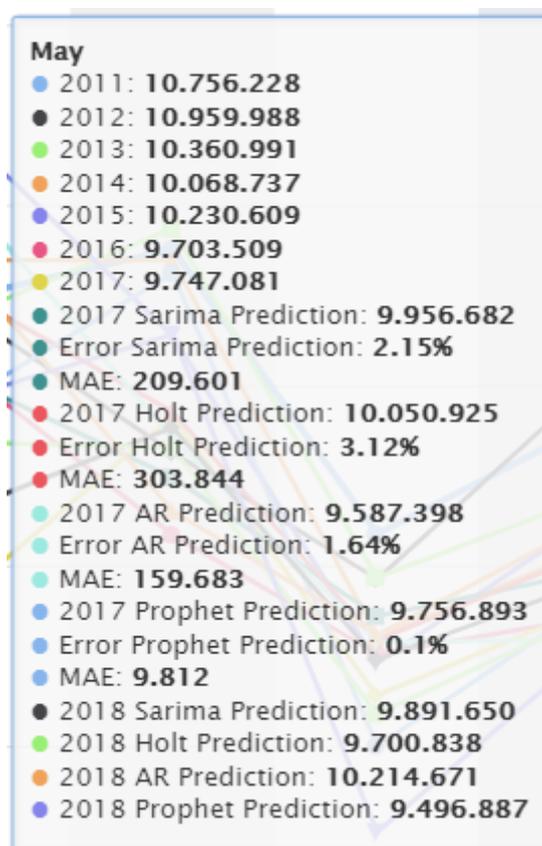


Figura 4.28 – Exibição dos valores dos conjuntos gerados com as estimativas de erros em porcentagem para cada algoritmo preditivo.

Para obter uma melhor compreensão dos dados gerados, foi criada uma tabela para mostrar a taxa de aproximação dos valores de predição em comparação dos valores reais. Na tabela 4.1 foram inseridos os valores do ano de 2017 e os valores de predição dos quatro algoritmos. Para exibir uma porcentagem de erros desses valores, foi utilizada a seguinte fórmula:

$$\text{ValorAbsoluto (ValorPrevisao - ValorReal) / ValorReal,}$$

na qual *ValorPrevisao* significa o valor de uma predição e o *ValorReal* consiste no valor mensal de 2017.

Ao analisar os dados contidos na Tabela 4.1, notamos que os valores de predição que ficaram mais distantes dos valores reais foram para o mês de fevereiro, sendo que a pior predição foi gerada pelo algoritmo *Autoregressive*, cujo valor da predição foi de 21.99%, uma taxa de erro considerada elevada. Em contraste a este valor, algumas predições ficaram em torno da faixa de 1% de erro, que foram nos meses de junho, julho, outubro e novembro para o algoritmo *Sarima*. Para o algoritmo *Holt-Winters* os meses de março, julho, outubro e novembro também ficaram nesta faixa. O modelo de predição *Prophet*

		JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Dados 2017		11581251	9855340	10827644	9279596	9747081	10600150	11649628	11190134	9945277	9921554	10305415	11974478
Predição SAR 2017		12126697	11026722	10489018	9718628	9956777	10727078	10727078	11756884	10454446	9788410	10119525	11322657
Predição HW 2017		12930052	11499776	10851868	9584311	10050906	10935913	11873585	11995158	10392502	9773306	10309716	11362265
Predição AR 2017		12192809	12022755	10706595	9575184	9587398	11178104	12401992	12051466	10572099	9748854	9972855	11276042
Predição Prophet (PRO) 2017		12428834	10927523	11620031	8998798	9756893	10604431	11434551	11124374	10137715	9962428	9985176	10676775
	Média final												
Percentual de erro SAR	4.00%	4.71%	11.89%	3.13%	4.73%	2.15%	1.20%	1.48%	5.06%	5.12%	1.34%	1.80%	5.44%
Percentual de erro HW	4.87%	11.65%	16.69%	0.22%	3.28%	3.12%	3.17%	1.92%	7.19%	4.50%	1.49%	0.04%	5.11%
Percentual de erro AR	5.83%	5.28%	21.99%	1.12%	3.19%	1.64%	5.45%	6.46%	7.70%	6.30%	1.74%	3.23%	5.83%
Percentual de erro PRO	3.95%	7.32%	10.88%	7.32%	3.03%	0.1%	0.04%	1.85%	0.59%	1.93%	0.41%	3.11%	10.84%
Média de erro mensal		7.24%	15.36%	2.94%	3.55%	1.75%	2.46%	2.92%	5.13%	4.46%	1.24%	2.04%	6.80%

Tabela 4.1 – Exibição dos valores gerados pelos algoritmos de predição e as taxas de erros de cada um deles em cada mês do ano.

apresentou ótimos resultados para os meses de maio, junho, agosto e outubro, mostrando taxas de erro abaixo de 1%, sendo respectivamente 0,1%, 0,04%, 0,59% e 0,41%. Por fim, para o algoritmo *Autoregressive* os meses que ficaram com erro nesta faixa em torno de 1% foram os meses de março, maio e outubro. Estes valores em questão são considerados muito bons, pois a assertividade dos valores gerados pelas predições ficaram próximos de 99%.

Para calcular a média de erro de cada algoritmo e verificar qual foi o algoritmo que melhor gerou as predições, foram somados os valores de todos os meses e foram divididos pelos 12 meses do ano. Dessa forma, os resultados foram de 5.83% para o algoritmo *AutoRegressive* - algoritmo que, em média, teve a menor taxa de assertividade -, 4.87% para o algoritmo *Holt-Winters*, 4% para o algoritmo *Sarima* e 3,95% para o algoritmo *Prophet*, algoritmo que, em média, teve a maior taxa de assertividade e dessa forma foi considerado o melhor algoritmo aplicado neste estudo.

Na última linha da Tabela 4.1, foi inserida a média mensal de erro. A pior média foi no mês de fevereiro com valor de 15.36%. Porém, nos meses de março, maio, junho, julho,

outubro e novembro a média de erro ficou abaixo da faixa de 3% (valores que estão apresentados em negrito). Acreditamos que as taxas de assertividade dos quatro algoritmos foram melhorando durante o ano porque quanto maior a quantidade de dados que cada algoritmo recebe, mais próximo o valor da predição estará do valor real. Além disso, diferentes conjuntos de dados podem ser carregados, é possível que ocorram variações nas taxas de assertividade dos algoritmos.

5. RESULTADOS E AVALIAÇÃO DO MODELO

Nesta seção, descrevemos em detalhes os conjuntos de dados usados em nossos experimentos, mostrando os resultados obtidos com a análise feita através do nosso modelo. Além disso, é apresentado o resultado de um estudo realizado com usuários especialistas.

5.1 Conjuntos de dados

Como mencionado anteriormente na seção 1, houve dificuldade em encontrar trabalhos relacionados que abordassem um modelo visual projetado para análise de dados de consumo de energia de edifícios e cidades inteligentes. Da mesma forma, foi difícil encontrar conjuntos de dados públicos que pudessem ser usados no nosso modelo. Após uma extensa pesquisa, encontramos dois conjuntos de dados de energia que puderam ser usados no trabalho. Estes conjuntos de dados são referenciados como “conjunto de dados A” e “conjunto de dados B”. Além disso, também conseguimos dados sobre o clima, como temperatura e pressão, referenciados aqui como “conjunto de dados C”.

O conjunto de dados A contém dados da *PJM Interconnection LLC*¹, que é uma organização de transmissão regional nos Estados Unidos. Ela faz parte da rede de Interconexão Oriental que opera um sistema de transmissão elétrica que atende os estados americanos de Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, Nova Jersey, Carolina do Norte, Ohio, Pensilvânia, Tennessee, Virgínia, Virgínia Ocidental e Distrito de Columbia. Os dados de consumo de energia por hora vêm do site da PJM, estão em megawatts, foram coletados de 2011 à 2017 e possuem apenas dados de uma única fonte, no caso, consumo elétrico.

O conjunto de dados B contém arquivos diferentes com dados de hora em hora para 16 tipos de prédios comerciais e prédios residenciais. Os dados estão disponíveis para um ano meteorológico típico em locais nos Estados Unidos². O tamanho total desse conjunto de dados é de aproximadamente 4,8 GB compactado ou 19 GB não compactado. Foram coletados dados do ano de 2004 de diferentes fontes de consumo de energia, tais como aquecedor de água, eletricidade, aquecimento de gás, entre outros.

O conjunto de dados C possui dados de alta resolução temporal (medições por hora) de vários atributos climáticos, como temperatura, umidade, pressão do ar e velocidade do vento. Esses dados estão disponíveis para 30 cidades dos EUA e do Canadá, além de

¹<https://www.kaggle.com/robikscube/hourly-energy-consumption>

²<https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states>

6 cidades israelenses³. Estes dados foram adquiridos usando a API do tempo no site do OpenWeatherMap e está disponível sob a licença ODbL⁴.

5.2 Resultados

Para demonstrar que o modelo suporta o uso de diferentes dados, os conjuntos de dados A e B descritos na Seção 5.1 foram carregados, um de cada vez. Os resultados da análise visual feita através do modelo usando o conjunto de dados A podem ser vistos nas Figuras 4.6 a 4.28.

Os resultados da análise visual feita tendo como entrada o conjunto de dados B são apresentados a seguir. A Figura 5.1 mostra o gráfico *Calendar view*, no qual é possível verificar um consumo médio elevado de energia nos meses de novembro a fevereiro. Este fato possivelmente está relacionado com o rigoroso inverno americano, que leva ao intensivo uso de aquecedores e aparelhos de ar condicionado. Em contrapartida, os meses de maio à setembro apresentaram um baixo consumo de energia. Isso está possivelmente relacionado temperaturas amenas da primavera e verão.

Já no gráfico *Dense Pixel Display* apresentado na Figura 5.2 é possível notar uma faixa azul contínua, representando o consumo médio (*Average*) de todas as fontes de energia do conjunto de dados B. Isso mostra que não há, ou há um consumo muito baixo de energia entre 1 e 4 horas da manhã. Como são dados de um edifício comercial, faz sentido que quase não haja consumo nesta faixa de horário. Ainda, um deslocamento nesta faixa azul nos meses de março e novembro refletem, respectivamente, o início e o fim do horário de verão. Outra questão que pode ser destacada é o alto consumo de energia entre às 5 horas da manhã e a meia-noite de outubro a abril. Esse alto consumo de dados possivelmente está relacionado ao inverno americano que ocorre durante este período, pois as baixas temperaturas fazem com que aumente o consumo de energia. Nesse período, também identificamos que há alguns picos de consumo de energia mais altos por volta das 8, 12 e 17 horas. Isto pode estar relacionado ao alto uso de elevadores, respectivamente, com o início da jornada de trabalho do prédio e a hora do almoço. De maio a outubro o consumo de energia é menor, possivelmente devido ao clima, pois abrange as estações de primavera e verão que dispensam o uso de aquecedores, além de pegar época de férias, o que pode fazer com que diminua o número de pessoas no prédio.

O gráfico da Figura 5.3 exhibe o consumo total por mês de todas fontes de energia do conjunto de dados B, sendo que as linhas na cor amarela e verde representam, respectivamente, os dados de consumo de gás e de aquecimento a gás. Um *insight* interessante que surgiu ao analisar este gráfico, foi verificar que no período de inverno americano (entre

³<https://www.kaggle.com/selfishgene/historical-hourly-weather-data>

⁴<https://opendatacommons.org/licenses/odbl/index.html>

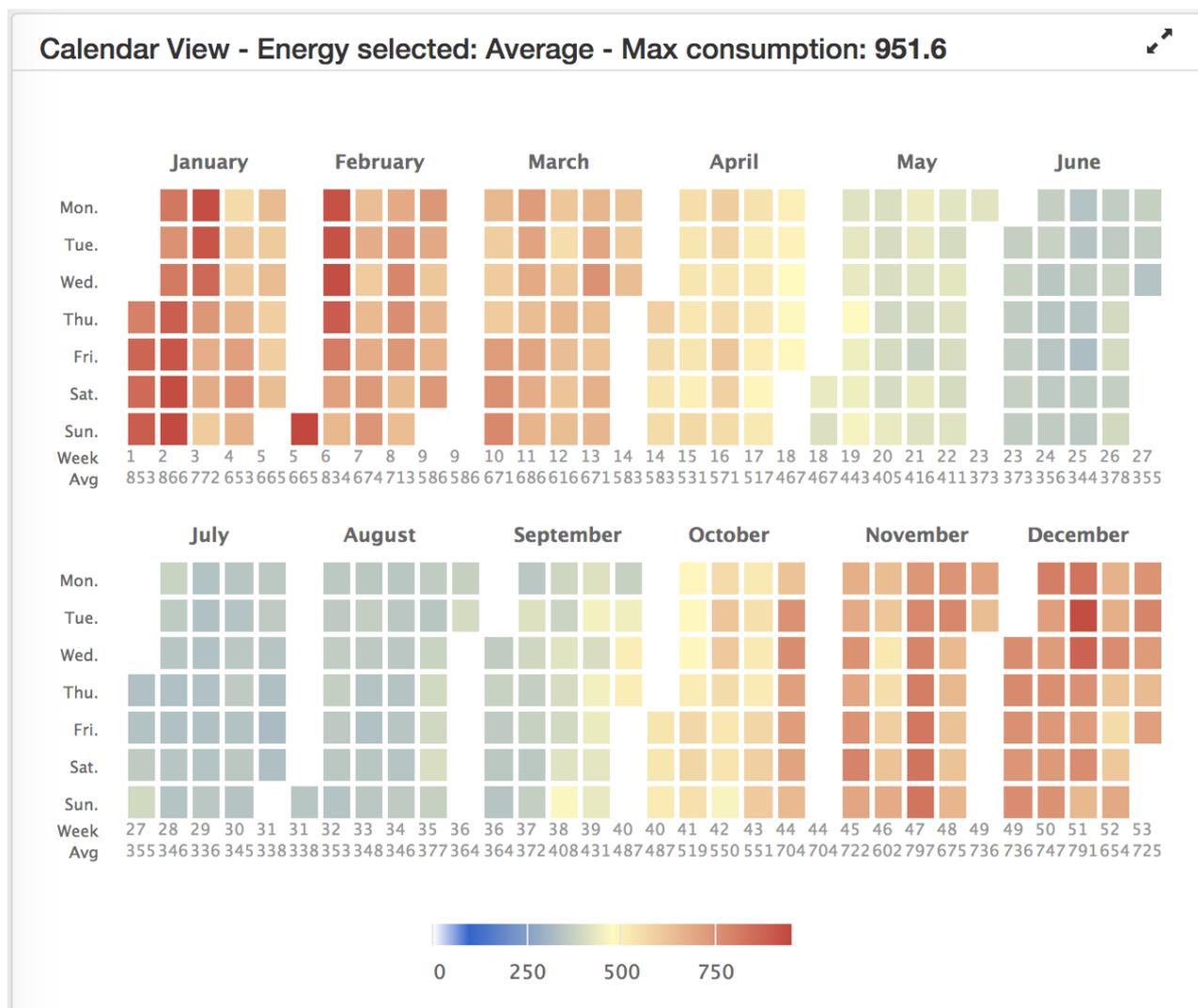


Figura 5.1 – *Calendar view* exibindo o consumo de energia ao longo de um ano.

dezembro e março) ocorre um elevado consumo destas fontes de energia em comparação às outras, como refrigeração, iluminação e eletricidade. Nos meses do outono, a partir de outubro, também se verifica o crescimento do consumo. O menor consumo pode ser observado entre os meses de abril à setembro. Por fim, também é possível verificar neste gráfico que quase não há alterações de consumo das outras fontes de energia ao longo do ano.

5.3 Entrevistas com especialistas

De acordo com Boyce e Neale [18], entrevistas com especialistas é uma técnica de pesquisa qualitativa que envolve a realização de entrevistas individuais intensivas com um pequeno número de entrevistados para explorar suas perspectivas sobre uma determinada ideia, programa ou situação. Assim, para avaliar a eficácia do nosso modelo, realizamos entrevistas com quatro especialistas de domínio da área de eficiência energética. Eles foram

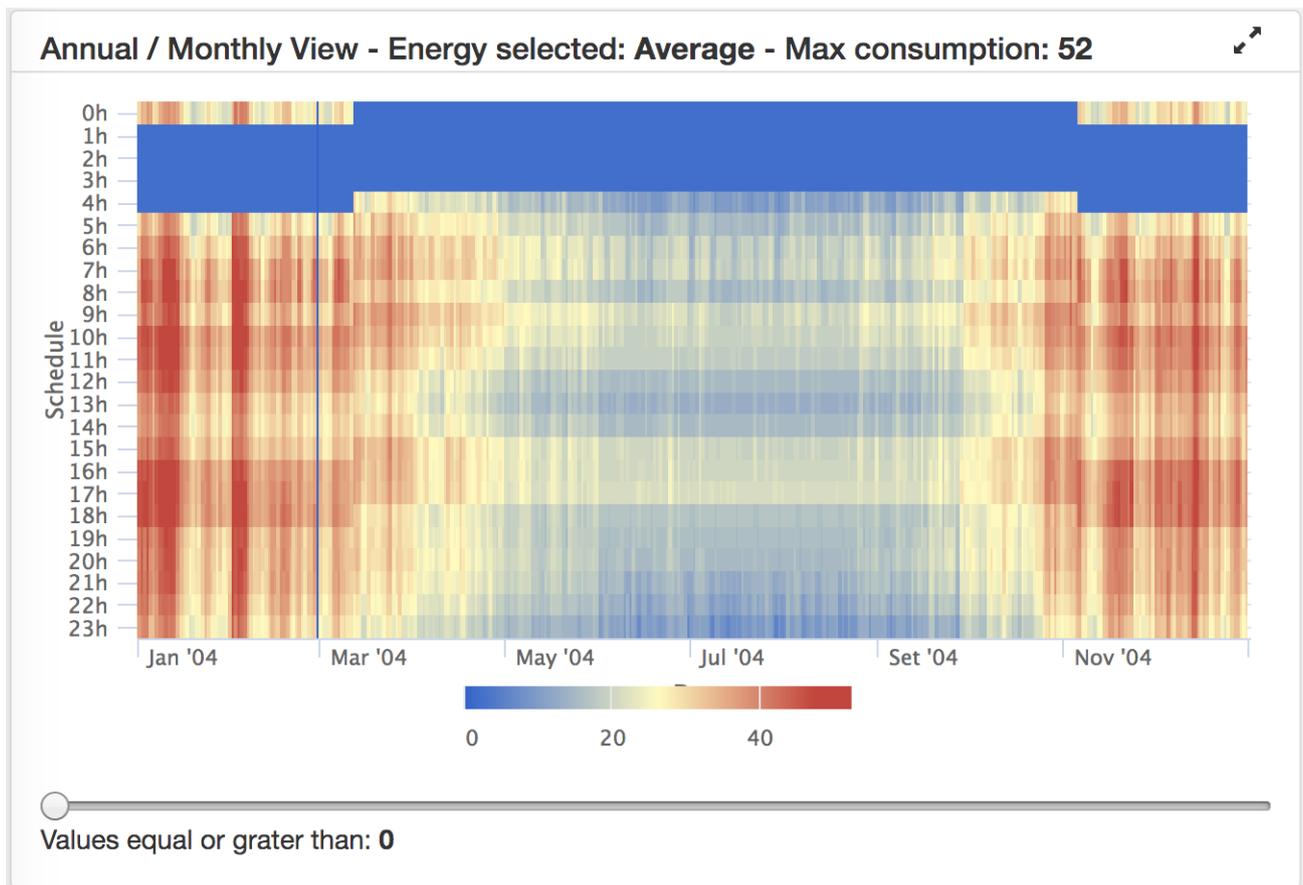


Figura 5.2 – *Dense Pixel Display* para análise do conjunto de dados B.

selecionados por causa de sua experiência em eficiência energética. Utilizamos entrevistas semiestruturadas com um conjunto de questões objetivas, permitindo espaço suficiente para a abertura de depoimentos detalhados do entrevistado [20]. Três participantes participaram das entrevistas de forma online e remota, e um participante realizou a entrevista presencialmente. Porém, nenhum participante manipulou diretamente a implementação do modelo.

O perfil dos participantes, referenciados aqui como P1, P2, P3 e P4, é apresentado na Tabela 5.1. Esta tabela contém informações detalhadas sobre os entrevistados, como idade, formação acadêmica, cargo que ocupa e ferramentas que utilizam em suas atividades. Os quatro participantes atuam em instituições privadas, possuem diferentes cargos e têm como principal ferramenta de análise visual de dados o Excel.

A entrevista foi realizada em três etapas que ocorreu da seguinte maneira. Na primeira etapa, apresentamos o formulário de consentimento que está Apêndice A e descrevemos como seria conduzida a entrevista. Em seguida, iniciamos a entrevista que foi guiada por um questionário semiestruturado em 3 partes, com 27 questões sobre suas experiências profissionais e conhecimento em eficiência energética. Este questionário, definido com base no trabalho de Milani [74], pode ser visto no Apêndice B. Depois disso, eles tiveram entre 1 e 2 minutos para realizar as duas tarefas contidas na seção 3 questi-

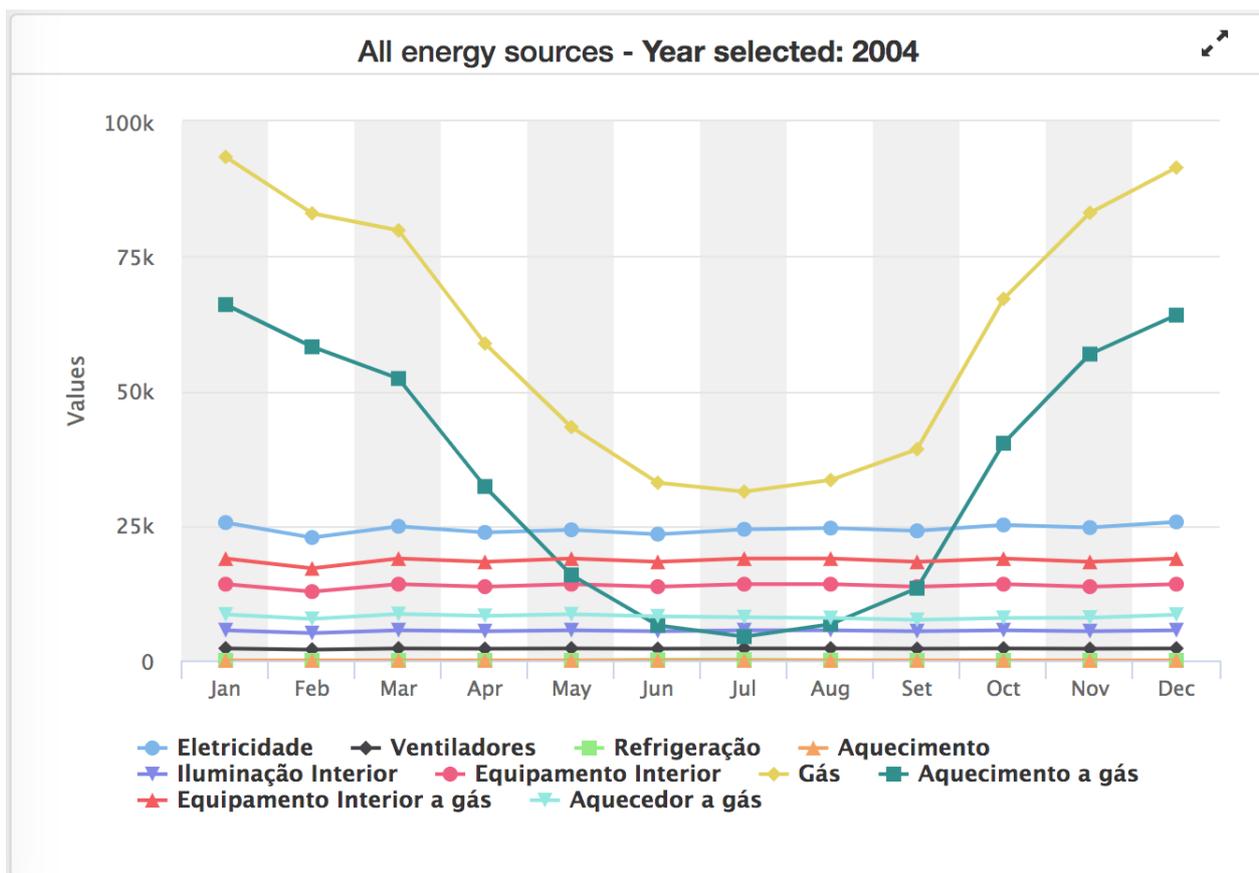


Figura 5.3 – Exibição dos dados de todas as fontes de consumo de energia do conjunto de dados B.

onário, visando manipular o modelo proposto e conhecer suas funcionalidades. Por fim, o especialista respondeu perguntas diretamente relacionadas ao modelo.

A participante P1 é uma analista de eficiência energética com mestrado em engenharia mecânica e que possui cerca de 3 anos e meio de experiência em eficiência energética. Ela relatou que entre suas atividades profissionais está a análise de dados de energia gerados por fontes limpas e dados de consumo de energia de prédios. Esses dados são extraídos através de medidores de potência, equipamento de sensores, horímetro, alicate de medição, batímetro e medidor de vazão de chuveiro. Ela não soube estimar a quantidade de dados que utiliza, mas disse que são fornecidos nos formatos XLS, CSV e TXT.

P1 relata que ao receber os dados sua primeira atividade é verificar se a medição obtida está correta e se há anormalidade, dados faltantes ou *outliers*, além de analisar a média de consumo. Ela afirma que o *insight* mais importante em um primeiro momento é verificar se há algum comportamento anormal, fora da curva de consumo ou algum desvio. Para verificar isso, são usados vários sistemas fornecidos pelos próprios fabricantes dos medidores, que auxiliam na análise visual. Além disso, por trabalhar em vários projetos simultaneamente, ela relata algumas dificuldades, como usar uma ferramenta que consiga visualizar um projeto como um todo e explorar dados de energia de diferentes projetos ao mesmo tempo em *Dashboards* interativos, os quais ela prefere utilizar.

Após a execução das duas tarefas contidas no Apêndice A, P1 destacou a facilidade e eficiência que o modelo proporciona em visualizar o menor e o maior consumo de energia, e o refinamento sequencial ao longo do tempo. Adicionalmente elogiou a interatividade do modelo e o modo de seleção e comparação dos dados de energia. Ela conseguiu identificar, sem dificuldade, as tendências de consumo ao longo do ano e os padrões de consumo nas quatro diferentes estações. Reconheceu a importância do uso de algoritmos de predição, o que seria interessante para suas atividades diárias, principalmente pela taxa de erros, pois ajuda a prever tanto consumos futuros, como a compra e venda de energia. Destacou positivamente a aplicação de funções estatísticas no modelo, pois é algo que ela costuma utilizar diariamente. P1 ressaltou dificuldade no entendimento da visualização *Boxplot*, pois não tem muita familiaridade do uso dessa técnica, e sugeriu a exibição de qual algoritmo gerou as melhores predições.

O participante P2 é um especialista em sistema de distribuição que possui mestrado em engenharia elétrica e atua por cerca de 25 anos na área de eficiência energética. Ele informou que entre suas tarefas diárias profissionais está a análise de dados elétricos, dados de temperatura e dados pluviométricos. Esses dados são coletados de subestações públicas e geradores privados, são recebidos em arquivos nos formatos XLS, CSV e TXT e podem chegar a um total de 25 Gigabytes.

P2 relatou que enfrenta diariamente problema com dados faltantes por diversos motivos, tais como problemas no medidor, demora na coleta de informações e falhas de comunicação. Devido a estes fatores, afirma que ter dados para trabalhar é de fundamental importância. Entre as ferramentas que utiliza para realizar estas análises, está o Excel e sistemas próprios fornecidos pelos fabricantes de medidores. Relata que estas ferramentas são ricas em informações, eficazes e de fácil usabilidade, contemplando suas necessidades diárias.

Depois de executar as tarefas propostas, P2 elogiou a estrutura do modelo por ser de fácil entendimento e uso, além de afirmar que não houve dificuldade para visualizar os maiores e menores consumo de energia ao longo do ano. Sentiu facilidade para visualizar o consumo de energia e os dados meteorológicos ao longo do ano, o que considerou importante para uma possível tomada de decisão. Ao analisar o gráfico com as predições, informou que ele teria aplicação na sua empresa, pois possibilitaria estimar os custos operacionais de compra e venda de energia. P2 acredita que o modelo está bem estruturado, que todas as visualizações são eficazes e estão bem empregadas. Ele sugeriu que seria de grande valor se o modelo pudesse suportar algoritmos de detecção de algum consumo fora do padrão e que fosse indicado qual é o melhor algoritmo de predição, algo que foi destacado pela P1.

O participante P3 é um coordenador de serviços de recuperação de energia com mestrado em gestão educacional e que possui especializações em gestão empresarial e gestão de pessoas. Trabalha há 25 anos na empresa, mas atua somente há 2 anos dire-

tamente com eficiência energética. Entre as suas demandas profissionais está a análise de consumo de eletricidade. Os dados são recebidos através de sistemas proprietários dos fabricantes de sensores em arquivos Excel com o volume de megabytes. P3 considera importante visualizar a curva de carga e a curva de consumo de energia, bem como a variância e a tendência de ambos. Inicialmente, ele mostrou interesse em trabalhar com um *Dashboard* interativo, pois auxiliaria na compreensão dos dados de consumo de energia e seus comportamentos diários, tanto para ele, quanto para os clientes e consumidores.

Após a execução das atividades propostas usando o modelo, P3 relatou que ele atendeu todas as suas expectativas, principalmente pelo *Dashboard* interativo, pelos recursos de comparação que o modelo oferece e pela granularidade de tempo em que os dados podem ser analisados. Relata ainda que o modelo desenvolvido supre todas suas necessidades de análises visuais. P3 afirmou não ter dificuldade para visualizar os picos de energia e identificou quais seriam os melhores algoritmos de predição por mês. Ele disse ter dificuldade para compreender a visualização *Dense Pixel Display*, pois não o conhecia e a quantidade de informações que ele contém é muito grande. Por fim, sugeriu que o modelo poderia fornecer uma associação de consumo de energia com dados sobre custos financeiros.

O participante P4 possui mestrado em engenharia elétrica e meio ambiente e coordena um laboratório de eficiência energética. Está há 22 no cargo atual e possui mais de 25 anos de experiência em eficiência energética. Ele trabalha com dados de diversas empresas privadas e órgãos públicos e governamentais. Estes dados são provenientes de sensores e geradores de energia, e correspondem a potência ativa, reativa, tensão elétrica, eletricidade, entre outros. Normalmente são recebidos no formato XLS e CSV, e não possuem um tamanho muito grande, abaixo de 1 megabyte. Esses dados são carregados e explorados em planilhas eletrônicas e sistemas proprietários dos fabricantes de sensores e geradores de energia.

Sua principal atividade ao realizar uma análise de dados de energia é verificar a curva de carga, juntamente com a tendência de consumo. Ele afirma que gosta de trabalhar exclusivamente com planilhas eletrônicas, pois elas fornecem funções estatísticas importantes e de fácil manipulação, recurso que, em sua opinião, é faltante nas ferramentas que já trabalhou.

Após a execução das duas tarefas propostas, ele afirmou que o modelo é completo. A visualização ao longo do tempo é didática, pois as interações e possibilidades de visualização em diferentes granularidades de tempo facilitam o refinamento e a identificação do consumo de energia. P4 não teve dificuldade para visualizar o maior e o menor consumo de energia, os picos semanais e quais seriam os melhores valores gerados pelos algoritmos de predição. Considerou positivo o fato do modelo disponibilizar a comparação de várias informações em diferentes conjuntos de dados e, principalmente, oferecer algumas funções estatísticas que as planilhas eletrônicas fornecem. P2 sugeriu que o modelo

poderia utilizar os dados meteorológicos para gerar previsões futuras, visto que diversas empresas necessitam destes dados para gerar lucros financeiros.

Após a realização das quatro entrevistas, o modelo mostrou-se aplicável em diferentes contextos de uso. Todos participantes ficaram interessados em usar o modelo em suas atividades diárias e, de certa forma, surpresos com os recursos disponibilizados, principalmente pela refinamento sequencial do longo do tempo, interatividade, possibilidade de comparar diferentes tipos de dados e algoritmos de previsão utilizados. Dessa forma, o modelo desenvolvido mostrou-se adequado e eficiente para visualizar dados de energia e meteorológicos.

X

5.4 Experimentos e limitações

Conforme descrito na seção 4.4, o modelo possui um recurso para obter a média geral de todas fontes de energia contidas no conjunto de dados. A média é gerada pela soma de todas as fontes de energia dividida pela quantidade de fontes informadas. Através deste recurso, foi feito um experimento para determinar a quantidade de dados de entrada que o modelo suporta gerenciar.

Para isso, foi criado um algoritmo para a simulação de dados, que gerou um conjunto de dados. Este arquivo continha 1140 fontes de consumo de energia, informando aleatoriamente os valores de consumo para cada hora ao longo do ano. Somando os valores de data, hora para as 1140 fontes de consumo de energia, o total de dados gerado foi 10.003,920 (dez milhões e três mil e novecentos e vinte) para cada conjunto criado, que neste caso foram 5 conjuntos, totalizando 50.019.600 (cinquenta milhões e dezenove mil e seiscentos) dados.

O experimento foi realizado em um *notebook Macbook pro* de 2.2GHz, com processador I7 e memória de 16GB, e o tempo de resposta do modelo foi de aproximadamente 30 segundos. Foram feitas algumas tentativas para aumentar esta quantidade de dados, buscando conhecer a capacidade total do modelo. Mas, ao aumentar este volume de dados o modelo apresentou uma demora excessiva e algumas funcionalidades não responderam de maneira adequada. Sendo assim, é possível afirmar que a capacidade de dados que o modelo suporta é de aproximadamente 50 milhões, o que equivale a 40 megabytes.

Participantes	P1	P2	P3	P4
Perfil				
Gênero	♀	♂	♂	♂
Idade	26	50	41	50
Formação acadêmica - Escolaridade				
Graduação				
Engenharia de Energia	✓			
Engenharia Elétrica		✓		✓
Contabilidade			✓	
Mestrado	✓	✓	✓	✓
Cargo - Experiência em eficiência energética				
Analista de eficiência energética	✓			
Especialista e sistema de distribuição		✓		
Coordenador dos serviços de recuperação de energia			✓	
Professor universitário				✓
Anos de experiência eficiência energética	3.5	25	2	25
Ferramentas utilizadas				
Excel	✓	✓	✓	✓
Energy Plus	✓			
Formato dos dados				
XLS/XLSX/CSV	✓	✓	✓	✓
TXT	✓	✓		

Tabela 5.1 – Perfil dos participantes, incluindo ferramentas usadas e formatos dos dados utilizados.

6. CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho teve como objetivo a criação de um modelo de análise visual de dados de consumo energético e dados meteorológicos para edifícios e cidades inteligentes. O desenvolvimento deste trabalho visou elaborar um estudo para auxiliar a tomada de decisão de dados de energia.

Através de uma revisão da literatura foi possível obter uma maior compreensão sobre o tema e identificar quais técnicas de visualização seriam mais adequadas para incorporar no modelo. A partir da análise destes trabalhos, foi possível obter uma maior compreensão sobre técnicas de análise visual usadas com dados de energia.

Para facilitar a análise e a tomada de decisão, o modelo desenvolvido apresenta os dados em diferentes níveis de granularidades ao longo do tempo utilizando diferentes representações visuais e formas de interação, além de disponibilizar quatro diferentes algoritmos de análise preditiva. O modelo foi testado com três conjuntos de dados: um gerado por uma distribuidora de energia, outro que contém o consumo de energia de um edifício comercial e um terceiro que contém dados meteorológicos de diversas cidades. Além disso, foi feita uma avaliação preliminar por especialistas da área de eficiência energética.

As principais contribuições deste trabalho são: Análise e comparação de dados meteorológicos e dados de energia ao longo do tempo e com diferentes granularidades; Técnicas de interação que suportam a análise visual; Disponibilidade de algoritmos de predição; Avaliação com especialistas comprova as vantagens do uso do modelo. Outras contribuições incluem a abordagem interativa baseada em detalhes sob demanda integrada com CMV, a possibilidade de fazer uma análise preditiva e os resultados positivos de uma avaliação preliminar com usuários. Através do modelo proposto é possível identificar anomalias, tendências e padrões de comportamentos. Além disso, auxilia na predição de consumo, na identificação de consumos médios mensais por diferentes fontes de consumo de energia, e na visualização do consumo anual de todas fontes de consumo de energia.

Como trabalhos futuros, pretendemos implementar no modelo as sugestões dadas pelos participantes desta pesquisa e incluir no modelo outros algoritmos de predição para este tipo específico de dado, como o modelo de previsão de séries temporais de temperatura interna usando redes neurais profundas pré-treinadas [95]. Além disso, podem ser realizados testes de usabilidade e entrevistas com outros usuários especialistas para verificar a possibilidade de adicionar novos recursos e funções estatísticas ao modelo. Outra melhoria a ser realizada é a inserção de *frameworks*, tal como o *Elastic Search*¹ para possibilitar o recebimento de dados em tempo real.

¹<https://www.elastic.co>

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Agarwal, Y.; Weng, T.; Gupta, R. “The energy dashboard: Improving the visibility of energy consumption at a campus-wide scale”. In: Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, 2009, pp. 55–60.
- [2] Ahmed, N.; Mueller, K. “Energyscout: A consumer oriented dashboard for smart meter data analytics”, *ArXiv Preprint*, vol. 1, Nov 2019, pp. 1–23.
- [3] Aigner, W.; Miksch, S.; Müller, W.; Schumann, H.; Tominski, C. “Visual methods for analyzing time-oriented data”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, Jan 2008, pp. 47–60.
- [4] Albino, V.; Berardi, U.; Dangelico, R. “Smart cities: Definitions, dimensions, performance, and initiatives”, *Journal of Urban Technology*, vol. 22, Feb 2015, pp. 2015.
- [5] Ang, B.; Choong, W.; Ng, T. “Energy security: Definitions, dimensions and indexes”, *Renewable and Sustainable Energy Reviews*, vol. 42, Feb 2015, pp. 1077–1093.
- [6] Angelidou, M. “Smart cities: A conjuncture of four forces”, *Elsevier Cities*, vol. 47, May 2015, pp. 95–106.
- [7] Anthes, G. “Zoom in, zoom out”, *Communications of the ACM*, vol. 55, Dec 2012, pp. 18–19.
- [8] Arbesser, C.; Spechtenhauser, F.; Mühlbacher, T.; Piringer, H. “Visplause: Visual data quality assessment of many time series using plausibility checks”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, Jan 2017, pp. 641–650.
- [9] Arjunan, P.; Khadilkar, H.; Ganu, T.; Z. Charbiwala, M. Z.; Singh, A.; Singh, P. “Multi-user energy consumption monitoring and anomaly detection with partial context information”. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, 2015, pp. 35–44.
- [10] Baldonado, M.; Woodruff, A.; Kuchinsky, A. “Guidelines for using multiple views in information visualization”. In: Proceedings of the working conference on Advanced visual interfaces, 2000, pp. 110–119.
- [11] Bamberger, Y.; Baptista, J.; Belmans, R.; Buchholz, B.; Chebbo, M.; Doblado, J.; Efthymiou, V.; Gallo, L.; Handschin, E.; Hatziaargyriou, N. “Vision and strategy for europe’s electricity networks of the future: European technology PlatformSmartGrids”. Office for Official Publications of the European Communities, 2006, 584-591p.

- [12] Banumathi, S.; Aloysius, A. “Predictive analytics concepts in big data: A survey”, *International Journal of Advanced Research in Computer Science*, vol. 8, Sep 2015, pp. 61–68.
- [13] Bartram, L. “Design challenges and opportunities for eco-feedback in the home”, *IEEE Computer Graphics and Applications*, vol. 35, Jul 2015, pp. 52–62.
- [14] Becker, M. “Modelos para previsão em séries temporais: Uma aplicação para a taxa de desemprego na região metropolitana de porto alegre”, Monografia, Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Departamento de Estatística. Curso de Estatística: Bacharelado, Dec 2010, 46p.
- [15] Berry, S.; Whaley, D.; Saman, W.; Davidson, K.; Berry, S. “Finding faults and influencing consumption: the role of in-home energy feedback displays in managing high-tech homes”, *Energy Efficiency*, vol. 10, Aug 2017, pp. 787–807.
- [16] Blanco, I.; Vega, A.; López, D.; González, M.; Castro, S.; Medrano, M. “Energy analytics in public buildings using interactive histograms”, *Energy and Buildings*, vol. 134, Jan 2017, pp. 94–104.
- [17] Bostock, M.; Ogievetsky, V.; Heer, J. “D-3: Data-driven documents”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, Dec 2011, pp. 2301–2309.
- [18] Boyce, C.; Neale, P. “Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input”. Pathfinder International, 2006, 16p.
- [19] Brehmer, M.; Ng, J.; Tate, K.; Munzner, T. “Matches, mismatches, and methods: Multiple-view workflows for energy portfolio analysis”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, Jan 2016, pp. 449–458.
- [20] Buber, R.; Holzmüller, H. “Qualitative marktforschun”. Gabler, 2008, 1152p.
- [21] Buono, P.; Balducci, F. “A web app for visualizing electronic nose data”. In: Proceedings of the 22nd International Conference Information Visualisation, 2018, pp. 198–203.
- [22] Card, S.; Mackinlay, J.; Shneiderman, B. “Readings in information visualization: Using vision to think”. Morgan Kaufmann Publishers Inc., 1999, 712p.
- [23] Carpendale, S.; Kerren, A.; Stasko, J.; Fekete, J.; North, C. “Evaluating Information Visualizations”. Springer Berlin Heidelberg, 2008, cap. 1, pp. 19–45.
- [24] Carrato, T.; Nesbitt, P.; Kehoe, M. “Ibm intelligent operations center providing insight, oversight, and smooth collaboration”. Capturado em: <https://www.ibm.com/developerworks/industry/library/ind-intelligent-operations-center/>, Jul 2019.

- [25] Cava, R. “Abordagens heterogêneas para a exploração interativa de grafos multivariados”, Tese de doutorado, Universidade Federal do Rio Grande do Sul. Instituto de Informática. Programa de Pós-Graduação em Computação., Jan 2017, 145p.
- [26] Chandola, V.; Banerjee, A.; Kumar, V. “Anomaly detection: A survey”, *ACM Computer Survey*, vol. 41, Jul 2009, pp. 1–58.
- [27] Chang, R.; Ghoniem, M.; Kosara, R.; Ribarsky, W.; Yang, J.; Suma, E.; Ziemkiewicz, C.; Kern, D.; Sudjianto, A. “Wirevis: Visualization of categorical, time-varying data from financial transactions”. In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, 2007, pp. 155–162.
- [28] Chen, J.; Wu, C.; Wu, K.; Litva, J. “Combining an autoregressive (ar) model with the fd-td algorithm for improved computational efficiency”. In: *Proceedings of the Microwave Symposium Digest, IEEE MTT-S International*, 1993, pp. 749–752.
- [29] Chen, K.; Wang, C. “A hybrid sarima and support vector machines in forecasting the production values of the machinery industry in taiwan”, *Expert Systems with Applications*, vol. 32, Jan 2007, pp. 254–264.
- [30] Chin, G.; Chen, Y.; Fitzhenry, E.; McGary, B.; Pirrung, M.; Bruce, J.; Winner, S. “A visual analytics platform and advanced visualization tools for interpreting and analyzing wind energy time-series data”, *IFAC-PapersOnLine*, vol. 51, Jan 2018, pp. 480–485.
- [31] Chou, J.; Telaga, A.; Chong, W.; Gibson, G. “Early-warning application for real-time detection of energy consumption anomalies in buildings”, *Journal of Cleaner Production*, vol. 149, Apr 2017, pp. 711–722.
- [32] Chourabi, H.; Nam, T.; Walker, S.; Gil-Garcia, J. R.; Mellouli, S.; Nahon, K.; Pardo, T. A.; Scholl, H. J. “Understanding smart cities: An integrative framework”. In: *Proceedings of the 45th Hawaii International Conference on System Sciences*, 2012, pp. 2289–2297.
- [33] Costanza, E.; Ramchurn, S.; Jennings, N. “Understanding domestic energy consumption through interactive visualisation: A field study”. In: *Proceedings of the ACM Conference on Ubiquitous Computing*, Sep 2012, pp. 216–225.
- [34] Cui, W.; Wang, H. “Anomaly detection and visualization of school electricity consumption data”. In: *Proceedings of the IEEE 2nd International Conference on Big Data Analysis*, 2017, pp. 606–611.
- [35] Denzin, N. K.; Lincoln, Y. S. “The Sage handbook of qualitative research”. SAGE Publications Inc, 2011, 766p.

- [36] Du, X.; Wei, Y.; Wu, L. “Interactive details on demand visual analysis on large attributed networks”, *Chinese Journal of Electronics*, vol. 27, Sep 2018, pp. 900–909.
- [37] Ehlers, R. “Análise de séries temporais”, *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, vol. 1, Jan 2007, pp. 1–118.
- [38] Ellegård, K.; Palm, J. “Visualizing energy consumption activities as a tool for making everyday life more sustainable”, *Applied Energy*, vol. 88, Jan 2011, pp. 1920–1926.
- [39] Fernandez, R.; Fetais, N. “Short survey of techniques for visualizing events recorded in logs”. In: *Proceedings of the 9th IEEE-GCC Conference and Exhibition, 2017*, pp. 1–9.
- [40] Ferreira, J.; Pinheiro, M.; Santos, W.; Maia, R. “Graphical representation of chemical periodicity of main elements through boxplot”, *Educación Química*, vol. 27, Jul 2016, pp. 209–216.
- [41] Ferretti, Y. “Ferramenta computacional para análise integrada de dados clínicos e biomoleculares”, *Dissertação de mestrado, Universidade de São Paulo*, Nov 2015, 1-79p.
- [42] Fu, S.; Zhao, J.; Cheng, H.; Zhu, H.; Marlow, J. “T-cal: Understanding team conversational data with calendar-based visualization”. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems, 2018*, pp. 1–13.
- [43] Fu, T.; Chung, F.; Kwok, K.; Ng, C. “Stock time series visualization based on data point importance”, *Engineering Applications of Artificial Intelligence*, vol. 21, Dec 2008, pp. 1217–1232.
- [44] García, D.; Díaz, I.; D, P.; Cuadrado, A.; Domínguez, M.; Morán, A. “Interactive visualization for nilm in large buildings using non-negative matrix factorization”, *Energy and Buildings*, vol. 176, Oct 2018, pp. 95–108.
- [45] Goodwin, S.; Dykes, J.; Jones, S.; Dillingham, I.; Dove, G.; Duffy, A.; Kachkaev, A.; Slingsby, A.; Wood, J. “Creative user-centered visualization design for energy analysts and modelers”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, Dec 2013, pp. 2516–2525.
- [46] Greenhalgh, T.; Peacock, R. “Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources”, *BMJ (Clinical research ed.)*, vol. 331, Dec 2005, pp. 1064–5.
- [47] Gupta, R. “Abc of internet of things: Advancements, benefits, challenges, enablers and facilities of iot”. In: *Proceedings of the Symposium on Colossal Data Analysis and Networking, 2016*, pp. 1–5.

- [48] Harrison, C.; Eckman, B.; Hamilton, R.; Hartswick, P.; Kalagnanam, J.; Paraszczak, J.; Williams, P. "Foundations for smarter cities", *IBM Journal of Research and Development*, vol. 54, Jul 2010, pp. 1–16.
- [49] Heer, J.; Kong, N.; Agrawala, M. "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations". In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 1303–1312.
- [50] Holt, C. "Forecasting seasonals and trends by exponentially weighted moving averages", *International Journal of Forecasting*, vol. 20, Jan 2004, pp. 5–10.
- [51] Ioannidis, D.; Tropios, P.; Krinidis, S.; Stavropoulos, G.; Tzovaras, D.; Likothanasis, S. "Occupancy driven building performance assessment", *Journal of Innovation in Digital Ecosystems*, vol. 3, Dec 2016, pp. 57–69.
- [52] Iyengar, S.; Lee, S.; Irwin, D.; Shenoy, P.; Weil, B. "Wathome: A data-driven approach for energy efficiency analytics at city-scale". In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery 38; Data Mining, 2018, pp. 396–405.
- [53] Janetzko, H.; Stoffel, F.; Mittelstädt, S.; Keim, D. "Anomaly detection for visual analytics of power consumption data", *Computers Graphics*, vol. 38, Feb 2014, pp. 27–37.
- [54] Jin, X.; Zhang, Y.; Wang, X. "Strategy and coordinated development of strong and smart grid". In: Proceedings of the IEEE PES Innovative Smart Grid Technologies, 2012, pp. 1–4.
- [55] Kalekar, P. S. "Time series forecasting using holt-winters exponential smoothing", *Kanwal Rekhi school of information Technology*, vol. 4329008, Dec 2004, pp. 1–13.
- [56] Katipamula, S. "Smart buildings can help smart grid: Transactive controls". In: Proceedings of the IEEE PES Innovative Smart Grid Technologies, 2012, pp. 1–1.
- [57] Keim, D. "Designing pixel-oriented visualization techniques: Theory and applications", *IEEE Transactions on Visualization and Computer Graphics*, vol. 6, Jan 2000, pp. 59–78.
- [58] Keim, D. "Information visualization and visual data mining", *IEEE Transactions on Visualization and Computer Graphics*, vol. 8, Jan 2002, pp. 1–8.
- [59] Keim, D.; Kriegel, H. "Visdb: Database exploration using multidimensional visualization", *IEEE Computer Graphics and Applications*, vol. 14, Sep 1994, pp. 40–49.

- [60] Keim, D.; Mansmann, F.; Schneidewind, J.; Thomas, J. "Visual analytics: Scope and challenges", *Lecture notes in computer science*, vol. 1, abr 2008, pp. 76–90.
- [61] Kim, T.; Ramos, C.; Mohammed, S. "Smart city and iot", *Future Generation Computer Systems*, vol. 76, Nov 2017, pp. 159–162.
- [62] Kitchenham, B.; Pickard, L.; MacDonell, S.; Shepperd, M. "What accuracy statistics really measure", *IEEE Proceedings - Software*, vol. 148, Jun 2001, pp. 81–85.
- [63] Kitchin, R.; Maalsen, S.; McArdle, G. "The praxis and politics of building urban dashboards", *Geoforum*, vol. 77, Dec 2016, pp. 93–101.
- [64] Kohlhammer, J.; Keim, D.; Pohl, M.; Santucci, G.; Andrienko, G. "Solving problems with visual analytics", *Procedia Computer Science*, vol. 7, Dec 2011, pp. 117–120.
- [65] Koytek, P.; Perin, C.; Vermeulen, J.; André, E.; Carpendale, S. "Mybrush: Brushing and linking with personal agency", *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, Jan 2018, pp. 605–615.
- [66] Kumar, D.; Rajasegarar, S.; Leckie, C.; Gubbi, J.; Bezdek, J.; Palaniswami, M.; Chan, J. "Adaptive cluster tendency visualization and anomaly detection for streaming data", *ACM Transactions on Knowledge Discovery from Data*, vol. 11, Dec 2016, pp. 24–40.
- [67] Kumatani, S.; Itoh, T.; Motohashi, Y.; Umezu, K.; Takatsuka, M. "Time-varying data visualization using clustered heatmap and dual scatterplots". In: *Proceedings of the 20th International Conference Information Visualisation*, 2016, pp. 63–68.
- [68] Lamberts, R.; Dutra, L.; Oliveira, F. "Eficiência energética na arquitetura". Eletrobras/Procel, 1999, 382p.
- [69] Li, T.; Alavy, M.; Siegel, J. "Measurement of residential hvac system runtime", *Building and Environment*, vol. 150, Mar 2019, pp. 99–107.
- [70] Lin, B.; Chen, Y.; Chen, X.; Yu, Y. "Comparison between json and xml in applications based on ajax". In: *Proceedings of the International Conference on Computer Science and Service System*, 2012, pp. 1174–1177.
- [71] Lin, J.; Keogh, E.; Lonardi, S.; Lankford, J.; Nystrom, D. "Visually mining and monitoring massive time series". In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004, pp. 460–469.
- [72] Martani, C.; Lee, D.; Robinson, P.; Britter, R.; Ratti, C. "Enernet: Studying the dynamic relationship between building occupancy and energy consumption", *Energy and Buildings*, vol. 47, Apr 2012, pp. 584–591.

- [73] Matkovic, K.; Freiler, W.; Gracanin, D.; Hauser, H. "Comvis: A coordinated multiple views system for prototyping new visualization technology". In: Proceedings of the 12th International Conference Information Visualisation, 2008, pp. 215–220.
- [74] Milani, A. "Preprocessing profiling model for visual analytics", Dissertação de mestrado, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), 2019, 119p.
- [75] Morvaj, B.; Lugaric, L.; Krajcar, S. "Demonstrating smart buildings and smart grid features in a smart energy city". In: Proceedings of the 3rd International Youth Conference on Energetics, 2011, pp. 1–8.
- [76] Morán, A.; Fuertes, J.; Prada, M.; Alonso, S.; Barrientos, P.; Díaz, I.; Domínguez, M. "Analysis of electricity consumption profiles in public buildings with dimensionality reduction techniques", *Engineering Applications of Artificial Intelligence*, vol. 26, Sep 2013, pp. 1872–1880.
- [77] Naphade, M.; Banavar, G.; Harrison, C.; Paraszczak, J.; Morris, R. "Smarter cities and their innovation challenges", *Computer*, vol. 44, Jul 2011, pp. 32–39.
- [78] Neirotti, P.; Marco, A.; Cagliano, A.; Mangano, G.; Scorrano, F. "Current trends in smart city initiatives: Some stylised facts", *Cities*, vol. 38, Jun 2014, pp. 25–36.
- [79] Nielsen, M.; Grønbæk, K. "Towards highly affine visualizations of consumption data from buildings", *Information Visualization Theory and Applications*, Jan 2015, pp. 247–255.
- [80] Nilsson, A.; Wester, M.; Lazarevic, D.; Brandt, N. "Smart homes, home energy management systems and real-time feedback: Lessons for influencing household energy consumption from a swedish field study", *Energy and Buildings*, vol. 179, Nov 2018, pp. 15–25.
- [81] Pedrasa, M.; Spooner, T.; MacGill, I. "Coordinated scheduling of residential distributed energy resources to optimize smart home energy services", *IEEE Transactions on Smart Grid*, vol. 1, Sep 2010, pp. 134–143.
- [82] Peña, K.; Yoseba, B.; Borges, C.; Cruz, H.; Haase, J.; Bruckne, D. "Smart buildings and the smart grid". In: Proceedings of the IECON - 39th Annual Conference of the IEEE Industrial Electronics Society, 2013, pp. 113–117.
- [83] Pflugradt, N.; Teuscher, J.; Platzer, B.; W.Schufft. "Analysing low-voltage grids using a behaviour based load profile generator". In: Proceedings of the International Conference on Renewable Energies and Power Quality, 2013, pp. 1–5.

- [84] Pickering, E.; Hossain, M.; French, R.; Abramson, A. "Building electricity consumption: Data analytics of building operations with classical time series decomposition and case based subsetting", *Energy and Buildings*, vol. 177, Oct 2018, pp. 184–196.
- [85] Pishva, D.; Takeda, K. "A product based security model for smart home appliances". In: Proceedings of the 40th Annual International Carnahan Conference on Security Technology, 2006, pp. 234–242.
- [86] Potter, K.; Hagen, H.; Kerren, A.; Dannenmann, P. "Methods for presenting statistical information: The box plot", *Visualization of large and unstructured data sets*, vol. 4, Jan 2006, pp. 97–106.
- [87] Prouzeau, A.; Dharshini, M.; Balasubramaniam, M.; Henry, J.; Hoang, N.; Dwyer, T. "Visual analytics for energy monitoring in the context of building management". In: Proceedings of the International Symposium on Big Data Visual and Immersive Analytics, 2018, pp. 1–9.
- [88] Pujani, V.; Akbar, F.; Nazir, R. "Management review of energy consumption: The energy saving opportunity in university buildings". In: Proceedings of the 5th International Conference on Industrial and Business Engineering, 2019, pp. 110–116.
- [89] Pérez, D.; Díaz, I.; Cuadrado, A.; García-Fernández, F.; Diez, A.; Domínguez, M. "Power-consumption analysis through web-based visual data exploration", *IFAC Proceedings Volumes*, vol. 47, Aug 2014, pp. 11257–11262.
- [90] Rauschenbach, U. "Progressive image transmission using levels of detail and regions of interest". In: Proceedings of the IASTED Conference on Computer Graphics and Imaging, 1998, pp. 1–4.
- [91] Rhyne, T. M.; Chen, M. "Cutting-edge research in visualization", *IEEE Journals Magazines*, vol. 46, May 2013, pp. 22–24.
- [92] Roberts, J. "On encouraging multiple views for visualization". In: Proceedings of the IEEE Conference Information Visualization, 1998, pp. 8–14.
- [93] Roberts, J. "State of the art: Coordinated multiple views in exploratory visualization". In: Proceedings of the Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization, 2007, pp. 61–71.
- [94] Roberts, J.; Wright, M. "Towards ubiquitous brushing for information visualization". In: Proceedings of the Information Visualization, Tenth International Conference on, 2006, pp. 151–156.
- [95] Romeu, P.; Zamora-Martinez, F.; Botella-Rocamora, P.; Pardo, J. "Time-series forecasting of indoor temperature using pre-trained deep neural networks". In:

Proceedings of the Artificial Neural Networks and Machine Learning, 2013, pp. 451–458.

- [96] Sadiku, M.; Shadare, A.; Musa, S.; Akujuobi, C.; Perry, R. “Data visualization”, *International Journal of Engineering Research and Advanced Technology*, vol. 2, Dec 2016, pp. 2454–6135.
- [97] Sandnes, F.; Lundh, M. “Calendars for individuals with cognitive disabilities: A comparison of table view and list view”. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers Accessibility, 2015, pp. 329–330.
- [98] Sarikaya, A.; Correll, M.; Bartram, L.; Tory, M.; Fisher, D. “What do we talk about when we talk about dashboards?”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, Jan 2019, pp. 682–692.
- [99] Shneiderman, B. “Tree visualization with tree-maps: 2-d space-filling approach”, *ACM Transactions on Graphics*, vol. 11, Jan 1992, pp. 92–99.
- [100] Silva, P.; Ilić, D.; Karnouskos, S. “The impact of smart grid prosumer grouping on forecasting accuracy and its benefits for local electricity market trading”, *IEEE Transactions on Smart Grid*, vol. 5, Jan 2014, pp. 402–410.
- [101] Sisworahardjo, N.; Saad, A. “Spatio - temporal context anomaly detection for residential power consumption”, *International Journal on Electrical Engineering and Informatics*, vol. 9, Dec 2017, pp. 776–785.
- [102] Sun, L.; Zhou, K.; Yang, S. “Regional difference of household electricity consumption: An empirical study of jiangsu, china”, *Journal of Cleaner Production*, vol. 171, Jan 2018, pp. 1415–1428.
- [103] Szmit, M.; Szmit, A. “Usage of modified holt-winters method in the anomaly detection of network traffic: Case studies”, *Journal of Computer Systems Networks and Communications*, vol. 5, May 2012, pp. 1–5.
- [104] Taylor, S. J.; Letham, B. “Forecasting at scale”, *The American Statistician*, vol. 72, abr 2018, pp. 37–45.
- [105] Thomas, J. “Illuminating the path: The research and development agenda for visual analytics”. Department of Homeland Security, 2005, 186p.
- [106] Thomas, J.; Cook, K. “A visual analytics agenda”, *IEEE Computer Graphics and Applications*, vol. 26, Jan 2006, pp. 10–13.
- [107] Timm, S.; Deal, B. “Effective or ephemeral? the role of energy information dashboards in changing occupant energy behaviors”, *Energy Research Social Science*, vol. 19, Sep 2016, pp. 11–20.

- [108] Toasa, R.; Maximiano, M.; Reis, C.; Guevara, D. “Data visualization techniques for real-time information — a custom and dynamic dashboard for analyzing surveys’ results”. In: Proceedings of the 13th Iberian Conference on Information Systems and Technologies, 2018, pp. 1–7.
- [109] Tufte, E. “The visual display of quantitative information”. Graphics Press, 1986, 197p.
- [110] Uno, F.; Dahlan, N. “Development of energy management system gui for office building”, *IET Seminar Digest*, vol. 2014, Jan 2014, pp. 1–6.
- [111] Viau, C.; McGuffin, M. “Connectedcharts: Explicit visualization of relationships between data graphics”, *Computer Graphics Forum*, vol. 31, Jun 2012, pp. 1285–1294.
- [112] von Landesberger, T.; Bremm, S.; Andrienko, N.; Andrienko, G.; Tekušová, M. “Visual analytics methods for categoric spatio-temporal data”. In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, 2012, pp. 183–192.
- [113] Wang, J.; Zhang, T. “Degradation prediction method by use of autoregressive algorithm”. In: Proceedings of the IEEE International Conference on Industrial Technology, 2008, pp. 1–6.
- [114] Ward, M.; Grinstein, G.; Keim, D. “Interactive data visualization: Foundations, techniques, and applications”. A. K. Peters, Ltd., 2015, 578p.
- [115] Weaver, C. “Cross-filtered views for multidimensional visual analysis”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, Mar 2010, pp. 192–204.
- [116] Wesley, R.; Eldridge, M.; Terlecki, P. “An analytic data engine for visualization in tableau”. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, 2011, pp. 1185–1194.
- [117] Wexler, S.; Shaffer, J.; Cotgreave, A. “The big book of dashboards: Visualizing your data using real-world business scenarios”. Wiley Publishing, 2017, 448p.
- [118] Wickham, H.; Stryjewski, L. “40 years of boxplots”, *The American Statistician*, vol. 1, Dec 2011, pp. 1–9.
- [119] Wijk, V.; Jarke, J.; Selow, V.; Edward, R. “Cluster and calendar based visualization of time series data”. In: Proceedings of the IEEE Symposium on Information Visualization, 1999, pp. 1–4.
- [120] Williamson, D.; Parker, R.; Kendrick, J. “The box plot: A simple visual method to interpret data”, *Annals of Internal Medicine*, vol. 110, Jun 1989, pp. 916–921.

- [121] Winters, P. "Forecasting sales by exponentially weighted moving averages", *Management Science*, vol. 6, Apr 1960, pp. 324–342.
- [122] Wohlin, C. "Guidelines for snowballing in systematic literature studies and a replication in software engineering". In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, 2014, pp. 1–10.
- [123] Wong, P.; Schneider, K.; Mackey, P.; Foote, H.; Chin, G.; Guttromson, R.; Thomas, J. "A novel visualization technique for electric power grid analytics", *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, May 2009, pp. 410–423.
- [124] Wong, P.; Thomas, J. "Visual analytics", *IEEE Computer Graphics and Applications*, vol. 24, Sep 2004, pp. 20–21.
- [125] Wood, G.; Day, R.; Creamer, E.; Horst, D.; Hussain, A.; Liu, S.; Shukla, A.; Iweka, O.; Gaterell, M.; Petridis, P.; Adams, N.; Brown, V. "Sensors, sense-making and sensitivities: Uk household experiences with a feedback display on energy consumption and indoor environmental conditions", *IEEE Transactions on Visualization and Computer Graphics*, vol. 55, Sep 2019, pp. 93–105.
- [126] Xiong, J.; Tzempelikos, A. "Model-based shading and lighting controls considering visual comfort and energy use", *Solar Energy*, vol. 134, Sep 2016, pp. 416–428.
- [127] Yarbrough, I.; Sun, Q.; Reeves, D.; Hackman, K.; Bennett, R.; Henshel, D. "Visualizing building energy demand for building peak energy analysis", *Energy and Buildings*, vol. 91, Mar 2015, pp. 10–15.
- [128] Zhou, Y.; Wu, J.; Wang, R.; Shiochi, S. "Energy simulation in the variable refrigerant flow air-conditioning system under cooling conditions", *Energy and Buildings*, vol. 39, Feb 2007, pp. 212–220.
- [129] Çakmak, H.; Maass, H.; Bach, F.; Kühnapfel, U. "A New Framework for the Analysis of Large Scale Multi-Rate Power Data". KIT Karlsruhe, 2014, 0-42p.

APÊNDICE A – ENTREVISTA PARA COLETA DE DADOS JUNTO A ESPECIALISTAS

Anexo 1 – Instrumentos para coleta de dados: Entrevistas

Instrumento	Roteiro semi-estruturado – Questionário
Formato	<ul style="list-style-type: none"> ● Roteiro semiestruturado para entrevista presencial ou via videoconferência. ● Este questionário está dividido em 3 etapas: <ul style="list-style-type: none"> ○ Etapa 1: Levantamento de perfil; ○ Etapa 2: Levantamento de Práticas e Necessidades dos especialistas; ○ Etapa 3: Avaliação do Modelo Proposto ● Com duração prevista entre 1h e 1h e 30 minutos, incluindo a leitura do TCLE e introdução do estudo que está sendo realizado.
Objetivo	<ul style="list-style-type: none"> ● Identificar as práticas e necessidade dos especialistas e as técnicas de visualização de dados adequadas para análise de consumo ou produção de energia. ● Avaliar o modelo proposto, verificando a adequação das visualizações e interações propostas.
Participantes	<ul style="list-style-type: none"> ● Serão recrutados de 2 a 5 voluntários para participarem das entrevistas. ● Critério de inclusão: profissionais de tecnologia da informação (TI) com experiência em sistemas computacionais que envolvam análise de dados de eficiência energética. Mínimo de 1 ano de experiência dedicado na função de analista de dados ou que possa ser relacionado como cientista de dados, analista de BI, etc. ● Critério de exclusão: menores de idade ou que necessitem de algum tipo de amparo (físico ou cognitivo), e profissionais sem experiência prévia de manipulação e transformação de dados.
Questões	<p><u>Questionário - Parte 1 - Levantamento de Perfil</u></p> <ol style="list-style-type: none"> 1. Gênero; 2. Idade; 3. Escolaridade (nível mais alto e área); 4. Local de Trabalho (por ex. se Academia/Pesquisa ou Industrial/Empresa) ; 5. Qual o cargo que ocupa; 6. Quanto tempo de experiência com análise de dados de eficiência energética. <p><u>Questionário - Parte 2 - Levantamento de Práticas e Necessidades dos especialistas</u> Considerando a análise de dados de energia que realiza, por favor responda:</p> <ol style="list-style-type: none"> 7. Como os dados são adquiridos? (ex. dados abertos do governo, sistema proprietário da empresa onde trabalha, etc.) 8. Quais são os tipos de dados usados para análise? (ex. dados de energia gerados por fontes de energia limpa, dados de consumo de energia de um prédio, etc.) 9. Qual é o formato dos dados? (ex. csv, json, tabelas, etc.) 10. Qual é o volume dos dados usados para análise? (ex. MB, GB, TB, etc.)

11. Quais são as principais atividades/tarefas desempenhadas no processo de análise destes dados?
12. O que é importante entender/visualizar para identificar um problema ou obter *insights* dos dados?
13. Qual(is) sistema(s) é(são) usado(s) para análise destes dados?
14. Este sistema disponibiliza alguma técnica de visualização para auxiliar na análise e interpretação dos dados?
[SIM] Quais (por exemplo, um gráfico de barras)? Entre as técnicas disponibilizadas, quais você mais utiliza? Por quê?
15. Na sua opinião, quais tipos de análises deveriam ser suportadas por representações visuais?
16. Há alguma forma de visualização que você acha que poderia facilitar o processo de análise destes dados?

Questionário - Parte 3 - Avaliação do Modelo Proposto

Execute as tarefas a seguir para analisar dois conjunto de dados no modelo, explorando suas possibilidades de visualização e interação, e responda as respectivas questões.

Tarefa 1: Analise o conjunto de dados carregado no modelo referente ao consumo de energia de uma concessionária de energia elétrica americana. O mesmo possui dados de consumo de energia de um período de 6 anos (De 2011 a 2016). Explore as funcionalidades do modelo para:

- Visualizar os dados de forma anual, através das técnicas de visualização anuais aplicadas no modelo;
- Identificar o mês com o maior e o menor consumo mensal;
- Comparar o consumo entre dois meses;
- Comparar o consumo entre duas semanas;
- Analisar as previsões geradas.

Responda as perguntas a seguir.

17. A visualização dos dados facilitou a identificação do maior e o menor consumo mensal deste conjunto de dados?
18. Alterando os anos do dataset, é possível identificar quais foram os períodos com o maior e menor consumo?
19. Houve dificuldade para identificar o maior e o menor consumo de energia em uma semana? Qual a importância deste recurso?
20. Você identificou alguma tendência, padrão ou *outlier* de consumo de energia?
21. Sobre as previsões geradas, qual algoritmo apresentou melhor resultado? Quais as vantagens dessa funcionalidade?

Tarefa 2: Analise o conjunto de dados carregado no modelo referente ao consumo de energia de um edifício nos Estados Unidos. O mesmo possui dados do ano de 2004 com 10 fontes de energias diferentes. Explore as funcionalidades do modelo para:

- Visualizar os dados de forma anual, através das técnicas de visualização anuais aplicadas no modelo;
- Comparar o consumo entre duas fontes de energia no mesmo período;
- Identificar os maiores e menores consumos em uma determinada semana do ano para duas fontes de energias.

Responda as perguntas a seguir.

22. Alterando as fontes de energia do dataset, é possível identificar quais foram os períodos com o maior e menor consumo de uma determinada energia?
23. Você acha importante comparar diferentes fontes de energia? Você considera que esta funcionalidade é um diferencial do modelo?

Responda às questões finais (Após executar e responder as perguntas relativas às tarefas 1 e 2):

24. A forma como são apresentadas as técnicas de visualização no modelo auxilia a análise e consumo de energia ao longo do tempo?
25. Quais técnicas de visualização você considera útil e quais você considera que não está agregando conhecimento?
26. Você considera que as funções estatísticas disponíveis no modelo auxiliam na análise de dados?
27. Existe alguma informação adicional não mencionada que você gostaria de comentar relacionada a este assunto? Teria alguma sugestão de melhoria ou funcionalidade que poderia ser acrescentada no modelo proposto?

**APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(TCLE)**

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Nós, Anderson Pinto Alves (aluno de mestrado) e Isabel Harb Manssour (professora orientadora), responsáveis pela pesquisa denominada de um modelo de análise visual de dados de energia para edifícios e cidades inteligentes, estamos fazendo um convite para você participar como voluntário neste estudo visando entender e analisar conjuntos de dados.

Esta pesquisa visa identificar as necessidades de usuários que trabalham com a exploração e análise de dados relacionados à eficiência energética, buscando desenvolver um modelo de visualização de dados para auxiliar na sua exploração e análise visual, a fim de encontrar padrões. Para isso, precisamos mapear atividades e recursos de visualização de dados utilizados durante esse processo e avaliar o modelo proposto. Não há benefícios a curto prazo para os participantes dessa pesquisa. Porém, ao término desse estudo são esperadas duas contribuições principais: proposta de um modelo que propicie a visualização e a análise de dados em diferentes períodos ao longo do tempo e de diferentes fontes de energia, além de validar as técnicas de análise visual propostas.

Para a coleta dos dados poderão ser utilizadas diferentes técnicas, tais como: entrevista seguindo roteiro semi estruturado e observação do desenvolvimento das atividades pré-estabelecidas no sistema sob análise. Entendemos que há riscos mínimos durante essas atividades como: divulgação de dados confidenciais (quebra de sigilo) e desconforto ou constrangimento durante gravações de áudio e/ou vídeo. Lembrando que o objetivo deste estudo não é avaliar o participante, mas, sim, avaliar os processos de trabalho e o sistema computacional que o participante estará usando durante o teste. O uso que se faz dos registros efetuados durante o teste é estritamente limitado a atividades acadêmicas e buscaremos garantir o anonimato e confidencialidade dos participantes.

Outras informações importantes:

- Durante todo o período da pesquisa você tem o direito de esclarecer qualquer dúvida ou pedir qualquer outro esclarecimento, bastando para isso entrar em contato com o pesquisador Anderson através do telefone (51) 99299-8095.
- Você tem garantido o seu direito de não aceitar participar ou de retirar sua permissão, a qualquer momento, sem nenhum tipo de prejuízo ou retaliação, pela sua decisão.
- As informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos participantes, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre sua participação.
- Ao assinar este termo de consentimento, você não abre mão de nenhum direito legal que teria de outra forma.
- Não assine este termo de consentimento a menos que tenha tido a oportunidade de fazer perguntas e tenha recebido respostas satisfatórias para suas dúvidas.
- Se você concordar em participar deste estudo, você rubricará todas as páginas e assinará e datará duas vias originais deste termo de consentimento. Você receberá uma das vias para seus registros e a outra será arquivada pelo responsável pelo estudo.

Rubrica do participante

Rubrica dos pesquisadores responsáveis

Eu, _____ , após a leitura deste documento e de ter tido a oportunidade de conversar com o pesquisador responsável, para esclarecer todas as minhas dúvidas, acredito estar suficientemente informado, ficando claro para mim que minha participação é voluntária e que posso retirar este consentimento a qualquer momento sem penalidades ou perda de qualquer benefício. Estou ciente também dos objetivos da pesquisa, dos procedimentos aos quais serei submetido, dos possíveis danos ou riscos deles provenientes e da garantia de confidencialidade e esclarecimentos sempre que desejar.

Diante do exposto expressei minha concordância de espontânea vontade em participar deste estudo.

Assinatura do participante da pesquisa

Assinatura de uma testemunha

DECLARAÇÃO DO PROFISSIONAL QUE OBTIVE O CONSENTIMENTO

Expliquei integralmente este estudo ao participante. Na minha opinião e na opinião do participante, houve acesso suficiente às informações, incluindo riscos e benefícios, para que uma decisão consciente seja tomada.

Data: _____

Anderson Pinto Alves
Aluno de Mestrado em Ciência da Computação
PPGCC – Escola Politécnica

Isabel Harb Manssour
Professora do PPGCC – Escola Politécnica



Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br