



Graduação em Ciência da Computação

José Rodolfo de Lima Farias

**SWEET HOME: UMA FERRAMENTA DE VISUALIZAÇÃO DE
DADOS FOCADA NO MERCADO IMOBILIÁRIO**

Trabalho de Graduação



Universidade Federal de Pernambuco
secgrad@cin.ufpe.br
www.cin.ufpe.br/~secgrad

RECIFE
2017



Universidade Federal de Pernambuco
Centro de Informática
Graduação em Ciência da Computação

José Rodolfo de Lima Farias

**SWEET HOME: UMA FERRAMENTA DE VISUALIZAÇÃO DE
DADOS FOCADA NO MERCADO IMOBILIÁRIO**

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: *Nivan Roberto Ferreira Júnior*

RECIFE
2017

Agradecimentos

Primeiramente agradeço a toda minha família. Em especial para minha mãe e meu pai, onde mesmo com dificuldades sempre julgaram minha educação de fundamental importância. Dando-me suporte em todos esses anos, e me mostrando o caminho certo a ser seguido. Amo vocês.

Quero agradecer também a todos os meus amigos que sempre me fizeram mais feliz durante nossos momentos juntos. Em especial a dois grupos que conheci durante a faculdade. Um abraço ao BdR e "Polemicas e Dragões", espero estar sempre com vocês.

Por ultimo quero agradecer a todos os professores que passaram pela minha vida. Cada um foi de fundamental importância na definição do meu caráter e das minhas escolhas. Em especial ao meu professor e orientador Nivan.

Resumo

O processo de escolha de um imóvel para moradia é um processo complexo no qual vários critérios são considerados. Além de características do imóvel, como número de quartos, tamanho e preço, uma grande importância é dada para fatores relacionados ao contexto urbano no qual o imóvel está inserido. Por exemplo, fatores como disponibilidade de lazer, atividades culturais, segurança, saúde, transporte e acessibilidade são de grande importância nesta decisão. Apesar da importância, as ferramentas atuais como OLX, ZapImoveis e VivaReal ignoram ou, nos melhores casos, só consideram parcialmente tais fatores.

Para resolver este problema, neste trabalho é proposto o Sweet Home, uma ferramenta interativa para a exploração de dados de ofertas imobiliárias complementadas com dados sobre o contexto urbano. Para este fim, nossa ferramenta integra diversas coleções de dados urbanos e ofertas imobiliárias a técnicas de visualização para explorar esta complexa coleção de dados. Através do uso das técnicas de visualização é possível analisar valores socio-econômicos das áreas de interesse na cidade e bem como sua relação com os imóveis a venda.

Nós demonstramos a eficiência do Sweet Home por meio de uma série de estudos de caso que mostram o poder do Sweet Home no processo de procura de imóveis.

Palavras-chave: Visualização de dados imobiliários, Dados imobiliários, Visualização de dados espaciais

Abstract

The process of choosing a place to live is a complex one, whose many criteria must be taken into consideration before a final decision. Besides the property's characteristics, like the number of bedrooms, its size, and its price, the urban context the property is located in is also very relevant. For instance, recreation and entertainment availability, cultural activities, safety, healthy environment, transportation and accessibility are very important aspects to be contemplated. Despite its importance, modern search tools, as OLX, ZapImoveis and Viva Real, ignore or only partly consider such relevant aspects. In order to solve this problem, we propose Sweet Home an interactive visualization tool whose goal is to support data exploration of real estate offers along with its urban context. In order to do so, our tool integrates several urban datasets together with real estate offers data. It is possible to analyse socio-economical values of any areas in the city, as well as their relation to the properties for sale, through the usage of visualization techniques. We demonstrate the efficiency of the Sweet Home through a series of case studies that prove its capability during the process of choosing a property to live.

Keywords: Real Estate Data Visualization, Real Estate Data, Spatial Data Visualization

Lista de Figuras

2.1	Site da construtora Moura Dubeux mostrando mostrando as características e a localização de um imóvel.	12
2.2	Sites de venda de imóveis tradicionais	12
3.1	Computação dos escores de características urbanas. Os elementos de cada categoria dentro de um raio dado (1 km) são utilizados para computar o escore correspondente à categoria.	19
4.1	Interface do Sweet Home	21
4.2	Exemplo de mudança de opacidade	23
5.1	Valores inconsistentes encontrados através das visualizações do Sweet Home.	25
5.2	Anomalias encontradas na base de dados	26
5.3	Heatmaps de Recife	27
5.4	Busca utilizando filtros relacionados a estrutura física do imóvel	28
5.5	Busca utilizando filtros relacionados as características externas	28
5.6	Busca encontrando um apartamento maior com as mesmas características externas	28

Lista de Tabelas

2.1	Comparação de ferramentas de exploração entre <i>websites</i> de venda de imóveis.	11
3.1	Exemplo do <i>dataset</i> antes da filtragem dos dados	14

Lista de Acrônimos

JSON	<i>JavaScript Object Notation</i>
HTML	<i>HyperText Markup Language</i>
API	<i>Application program interface</i>
ODbL	<i>Open Database License</i>
OSM	<i>Open Street Maps</i>

Sumário

1	Introdução	9
1.1	Contribuições	10
1.2	Estrutura do trabalho	10
2	Trabalhos Relacionados	11
3	Análise e tratamento dos dados	13
3.1	<i>Dataset</i> de Anúncios de Imóveis	13
3.2	<i>Dataset</i> de Características Urbanas	14
3.3	Integração dos Dados	18
4	Sweet Home	20
4.1	Tarefas Analíticas	20
4.2	Sweet Home	20
4.2.1	Vista Geográfica	20
4.2.2	Vista de Atributos dos Imóveis	22
4.2.3	Painel de Controles	23
5	Casos de uso	24
5.1	Detectando Erros nos Dados	24
5.2	Entendendo a Cidade	25
5.3	Buscando Imóveis	26
6	Conclusão	29
6.1	Melhorias e Trabalhos futuros	29
	Referências	30
7	Apêndice	31
7.1	Consultas ao <i>Open Street Maps</i> (OSM)	31

1

Introdução

O mercado imobiliário movimentou 325 bilhões de reais por ano, o que equivale a 6,4% do PIB brasileiro. Mesmo com o mercado em crise o número de imóveis para venda ou locação se mantém em crescimento. De fato, nos 7 primeiros meses de 2016, foram lançados 33.975 novas unidades imobiliárias, número 13,8% superior ao observado nos primeiros 7 meses de 2015¹.

Dada a grande popularidade deste mercado, foi criado um grande número de aplicações de venda de imóveis. Exemplos de tais aplicações são OLX², ZapImoveis³ e VivaReal⁴. Uma característica comum a esse tipo de ferramenta é o fato de que elas se concentram em explorar informações do imóvel em si (quantidade de quartos, banheiros, área). Entretanto, existem muitos outros fatores envolvidos no processo de escolha de uma moradia ou compra de imóvel. Em particular, são de extrema importância fatores ligados ao contexto urbano no qual o imóvel está inserido. Fatores como segurança, lazer, cultura, transporte, diversidade populacional, presença de áreas verdes, dentre outros SILVA (2009); FERREIRA et al. (2015); DESIMONE (2017) são cruciais para a escolha de onde morar. Apesar da importância, as ferramentas atuais como ZapImoveis ou VivaReal ignoram ou, nos melhores casos, só consideram parcialmente tais fatores.

Dada a iniciativa global de abertura de dados urbanos, existe uma grande oportunidade para melhor dar suporte a decisão na escolha de imóveis para moradia. Este suporte se dará por um melhor entendimento da cidade, através do uso de dados que descrevem múltiplos aspectos da cidade em estudo. Existem diversas aplicações focadas em visualizar conjuntos de dados urbanos isoladamente. Por exemplo, TaxiVis FERREIRA et al. (2013) foca na visualização de dados de transporte, enquanto que o trabalho de Janetzko et al. JANETZKO et al. (2014) foca na exploração de dados de consumo de energia. Entretanto, poucos sistemas focam em visualizar os vários componentes de uma cidade e suas interações. Exemplos destes sistemas são Urbane FERREIRA et al. (2015) e Legible Cities CHANG et al. (2007). Entretanto, nenhuma

¹Notícia no portal G1. Link: <https://goo.gl/f14B7b>

²<http://www.olx.com.br/imoveis>

³<https://www.zapimoveis.com.br/>

⁴<https://www.vivareal.com.br/>

delas foca no domínio de imóveis.

1.1 Contribuições

Neste trabalho apresentamos Sweet Home, uma ferramenta de visualização interativa de dados para dar suporte a escolha de um imóvel por meio de um melhor entendimento de dados urbanos. Técnicas de visualização interativa de dados são reconhecidas como ferramentas eficientes que auxiliam na exploração e identificação de padrões em coleções de dados complexos, o que as fazem ideal para o problema com o qual estamos lidando MUNZNER (2014). Para alcançar este objetivo, o Sweet Home integra coleções diversas de dados urbanos (obtidos de fontes abertas) e de ofertas imobiliárias e, assim, possibilita analisar valores socio-econômicos das áreas de interesse na cidade e bem como sua relação com os imóveis a venda. O sistema Sweet Home foi implementado usando tecnologia web. Isto facilita o acesso ao sistema. Como ferramenta principal no desenvolvimento foram utilizadas as bibliotecas D3⁵ e Leaflet⁶.

1.2 Estrutura do trabalho

A organização deste documento está descrita a seguir. O capítulo 2 apresenta trabalhos relacionados na literatura, bem como ferramentas comerciais disponíveis para a compra de imóveis. O capítulo 3 discute o processo de coleta da base de dados utilizado no projeto, bem como o tratamento necessário para a base ser utilizada no Sweet Home. O design do sistema Sweet Home é apresentado no capítulo 4, discutindo as tarefas analíticas a serem executadas, assim como as técnicas de visualização utilizadas. O capítulo 5 apresenta alguns casos de uso que mostram o Sweet Home em ação. Finalmente, no capítulo 6, são feitas considerações sobre o que o projeto alcançou e quais serão os possíveis trabalhos futuros.

⁵<https://d3js.org/>

⁶<http://leafletjs.com/>

2

Trabalhos Relacionados

Este capítulo discute trabalhos relacionados ao tema de análise visual de dados focados no mercado imobiliário.

Como falado na introdução, o mercado apresenta um crescimento constante. Como consequência desse crescimento é natural o aparecimento de ferramentas para auxiliar na tomada de decisão na compra de imóveis. De forma geral, cada empreendimento imobiliário é veiculado pelos sites das construtoras, como forma de propaganda do mesmo. Por exemplo, veja a Figura 2.1. Neste características do imóvel e do seu entorno são apresentadas. Entretanto, estes sites não dão suporte à comparação de imóveis e, portanto, não suprem a necessidade que buscamos.

Sites como OLX Imóveis(Figura 2.2a), VivaReal(Figura 2.2b), ZapImoveis(Figura 2.2c) são focados na venda de residências. Todos eles apresentam uma interface de certa forma semelhante. Esta interface apresenta uma lista de imóveis com uma foto e informações ligadas diretamente a infraestrutura da residência, informações como número de quartos, suítes e vagas. Porém há uma pobreza de detalhes quando o assunto tratado é localização. Por exemplo o OLX Imóveis não possui nenhuma ferramenta de inspeção da localização dos imóveis. VivaReal e ZapImoveis possuem um mapa, porém o mesmo mostra exclusivamente só o um imóvel selecionado, logo deixando muito complicado para comparar localizações entre as diversas escolhas disponíveis para o usuário. Outro problema encontrado nos sites citados é a falta de informação em relação as características externas. Como por exemplo se o apartamento está próximo a escolas ou pontos de ônibus. Este fato é a motivação principal do presente trabalho.

Tabela 2.1: Comparação de ferramentas de exploração entre *websites* de venda de imóveis.

Website	Mapa	Comparação	Filtros internos	Filtros externos
VivaReal	Sim	Não	Sim	Não
ZapImoveis	Sim	Não	Sim	Não
OLX Imóveis	Não	Não	Sim	Não

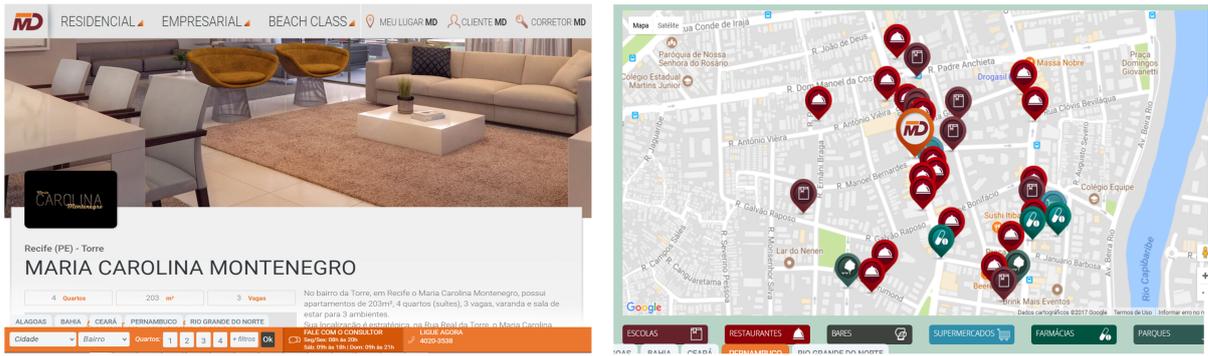
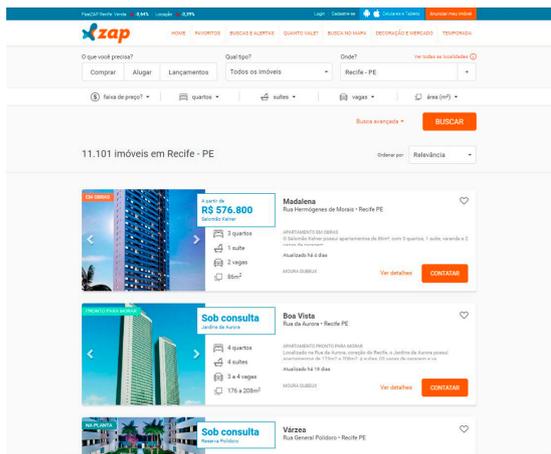
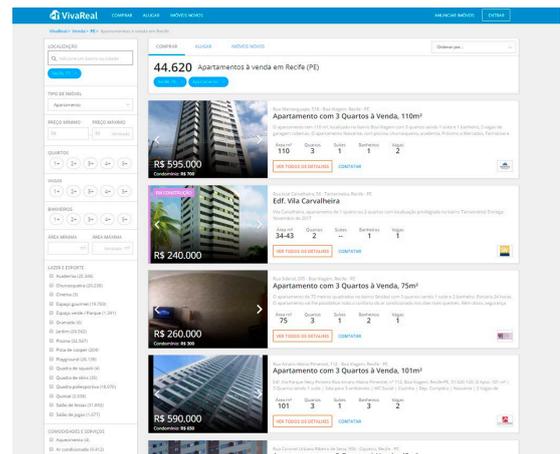


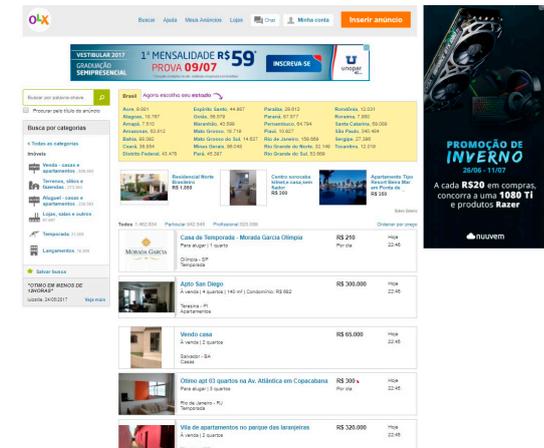
Figura 2.1: Site da construtora Moura Dubeux mostrando mostrando as características e a localização de um imóvel.



(a) OLX Imóveis



(b) Viva Real



(c) Zap Imóveis

Figura 2.2: Sites de venda de imóveis tradicionais

3

Análise e tratamento dos dados

Este capítulo descreve os dados utilizados neste trabalho, bem como o processamento feito nos mesmos para que pudessem ser visualizados no Sweet Home. A fim de analisar e dar um melhor suporte à tomada de decisão na compra de imóveis, Sweet Home integra dois tipos de dados: dados referentes a anúncio de imóveis, descritos na seção 3.1, e dados de características externas, ou seja dados do ambiente urbano no qual os imóveis estão inseridos, descritos na seção 3.2.

3.1 *Dataset de Anúncios de Imóveis*

Como fonte de dados para informações de características e preços de imóveis, foi usado um conjunto de dados obtido a partir de um *webscraper* no site Zap imóveis. *Webscraper* é uma ferramenta com finalidade de extrair data a partir de um site *HyperText Markup Language* (HTML). O *webscraper* foi desenvolvido pelo usuário do github pauloromeira ¹. O mesmo disponibilizou de forma aberta e gratuita os dados em uma plataforma de compartilhamento chamada data.world. ² Tal base de dados é relevante, haja vista que o ZapImóveis é um site consolidado no mercado. Fundado em 2000 e adquirido pelo grupo Globo em 2001 ³, o Zap Imóveis veio a se tornar líder no mercado imobiliário em 2005 e desde então se mantém na liderança.

O conjunto de dados utilizado se baseia em anúncios na cidade do Recife contendo 7203 entidades. Cada entidade contém título (nome do imóvel), preço, tipo (residencial ou comercial, rua, bairro, CEP, latitude, longitude, tipo vendedor (pessoa física ou imobiliária), número de quartos, número de suítes, área útil, área total, número de vagas, taxa condomínio, IPTU, e data da coleta dos dados (Tabela 3.1).

Como todo processo automática para coleta de dados, o webscraper descrito acima gerou dados com problemas de qualidade. Para resolver este problema, aplicamos alguns processos de limpeza de dados. A primeira filtragem foi remover instâncias que tinham valores nulos

¹<https://github.com/pauloromeira/realestate-scraper>

²<https://data.world/pauloromeira/anuncios-de-imoveis-recife-pe>

³https://pt.wikipedia.org/wiki/Zap_Imóveis

Tabela 3.1: Exemplo do *dataset* antes da filtragem dos dados

Preço	Lat	Long	Quartos	Área	Condomínio
380000	-26.9225709	-49.0606445	2	62	420
380000	-8.027619	-34.9055794	3	62	500
340000000	-8.0465796	-34.9116897	2	72	550
400000	-8.0334804	-34.9104243	2	69	590
550000	-8.1185721	-34.8959562	4	188	400
295000	-8.0492426	-34.9315404	3	70	220
9000000	-8.0295649	-34.9545166	8	910	
215000	-8.0395071	-34.9090258	2	56	260

(faltando) para atributos que julgamos fundamentais para nossa análise. Os atributos fundamentais escolhidos foram: Título, número de quartos, preço, área útil, número de vagas, taxa condomínio, latitude e longitude. A segunda filtragem se deu em remover todos os estabelecimentos comerciais, pois o foco de nosso trabalho é dar suporte na escolha de imóveis residenciais. Em um terceiro passo, realizamos um processo de limpeza de acordo com a do imóvel, de acordo com as coordenadas geográficas. Imóveis cuja localização se encontrava fora de um retângulo em volta da cidade do Recife foram removidos da base de dados. Finalmente, uma filtragem em imóveis que seus valores para os atributos fundamentais eram maiores que 95 % do atributos. Essa filtragem foi necessária por que alguns imóveis estavam com valores claramente errados nos atributos fundamentais, como pode ser visto na figura 5.1b. Ao fim da aplicação de todos os filtros citados anteriormente, o *dataset* de imóveis apresentou 2642 entidades.

3.2 Dataset de Características Urbanas

A fim de dar melhor suporte ao processo de decisão na compra de imóveis, é necessário ser capaz de comparar imóveis de acordo com características externas na qual o imóvel está localizado. Para tanto precisamos de uma fonte de dados que contenha a localização de várias *commodities urbanas* para diversas áreas em uma cidade de interesse. Para este fim foi selecionado o *OpenStreetMaps* (OSM) HAKLAY; WEBER (2008), que é um projeto de geolocalização baseado em *crowdsourcing*. Este projeto tem como principal objetivo manter e disponibilizar de forma gratuita dados espaciais e uma infraestrutura para mapeamento destes dados. Assim como outros projetos *crowdsourcing*, os dados contidos no *Open Street Maps* (OSM) são oriundos de contribuições voluntária e devem respeitar *Open Database License* (ODbL). O ODbL define o uso, modificação e o compartilhamento livre desses dados desde que essa mesma liberdade seja repassada à quem usar sua plataforma. Logo toda informação contida no OSM é livre e gratuita. Outro ponto que chamou atenção no OSM, além do fato de ser totalmente gratuito, foi o número elevado de contribuições. O número total de contribuintes ultrapassa quatro milhões, já o número total de objetos geográficos se aproxima de quatro bilhões de ⁴.

⁴http://www.openstreetmap.org/stats/data_stats.html

A aquisição dos dados do OSM foi via uma *Application program interface* (API) chamada Overpass⁵. A Overpass API funciona a base de *queries* de pesquisa, ou seja uma *query* é enviado ao servidor e se obtém uma resposta. Essa resposta pode ser de vários formatos, mas por questão de facilidade optou-se por manter o padrão que é GeoJson. GeoJson (Listing 3.2) é uma formato de dados semelhante a *JavaScript Object Notation* (JSON), porém com a peculiaridade de uma tag especificadamente para geoposicionamento contendo latitude e longitude. Os dados do OSM são categorizados pela união de elementos e tags. Os elementos são organizados em:

- *Nodes* são localizações pontuais no mapa, esses pontos podem ser isolados ou conectados. Cada *node* apresenta latitude, longitude, identificador e um conjunto de tags. A função de cada tag será explicada mais abaixo;
- *Ways* são criados a partir de ligações entre os *nodes*, podendo dar origem a rotas ou rios por exemplo;
- *Closed ways* são criados a partir de um *loop* de *ways*, por exemplo uma *area* ou barreiras;
- *Areas* são também formados a partir de *ways*, mas obrigatoriamente possuem a tag *area* com o valor positivo;
- *Relation* são usados para criar formas geométricas mais complexas, ou expressar relacionamento entre os elementos citados anteriormente sem a necessidade de uma ligação física entre eles;
- *Tags* são propriedades referentes ao elemento vinculado, por exemplo um *node* restaurante contém a *tag* com identificador *amenity* igual a *restaurant*.

Classificamos os dados de características urbanas em seis classes: Educação, Entretenimento, Financeiro, Saúde, Alimentação e Transporte. Para cada uma das classes, usamos a API Overpass para acessar elementos geográficos do OSM associados a cada uma delas. Em mais detalhes, para cada classe temos uma *query* :

- Educação (Listing 3.1) foi buscado *nodes* com a *tag amenity* igual a colégios, universidades, escola de idiomas e outros locais ligados diretamente a estudo;
- Entretenimento foi buscado *nodes* com as *tag amenity* ou *tag leisure* igual a centro de artes, cinema, teatros e outros locais ligados diretamente a diversão ou entretenimento;
- Financeiro foi buscado *nodes* com a *tag amenity* igual a caixa eletrônico, bancos ou casa de câmbio;

⁵<http://overpass-api.de/>

- Saúde foi buscado *nodes* com a *tag amenity* igual a clínicas, hospitais, farmácias e outros locais ligados diretamente a saúde;
- Alimentação foi buscado *nodes* com a *tag amenity* igual a restaurantes, cafés, bares e outros locais ligados diretamente a alimentação;
- Transporte foi buscado *nodes* com a *tag amenity* ou *tag leisure* igual parada de ônibus, táxi, metrô e outros locais ligados diretamente a transporte;

Um exemplo de código para a query de educação pode ser visto no Quadro 3.1. O resultado desta query pode ser visto no Quadro 3.2. O exemplo das outras *queries* pode ser encontrado no apêndice (Seção 7.1).

Listing 3.1: Query para coleta de dados em educação

```
responseEducation = api.Get("""
    area[name = "%s"]->.a;
    (
        node(area.a)[amenity~"college|kindergarten|library|
            public_bookcase|school|music_school|driving_school
            |language_school|university"];
    )
    "" % city)
```

Já em relação as *tags* foi utilizado as seguintes *keys*: *Area* que define a área dos *nodes*, que no caso específico desse trabalho são *nodes* com valor de *area* igual a Recife; *Amenity*, *nodes* que contém essa *tag* são referentes a locais de uso público ou de interesse comum. *Leisure*, *nodes* que contém *leisure* como *tag* são referentes a locais de entretenimento. *Highway* foi utilizado pois contém as paradas de ônibus.

Listing 3.2: Parte do GeoJson retornando pela *query* de educação

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "id": 560398646,
      "geometry": {
        "type": "Point",
        "coordinates": [
          -34.8987418,
          -8.0620287
        ]
      }
    }
  ]
}
```

```
    ]
  },
  "properties":{
    "amenity":"school",
    "name":"col\u00e9gio GGE"
  }
},
{
  "type":"Feature",
  "id":560404588,
  "geometry":{
    "type":"Point",
    "coordinates":[
      -34.9034684,
      -8.0608496
    ]
  },
  "properties":{
    "amenity":"university",
    "name":"FCAP - UPE"
  }
},
{
  "type":"Feature",
  "id":560404598,
  "geometry":{
    "type":"Point",
    "coordinates":[
      -34.9034427,
      -8.0594984
    ]
  },
  "properties":{
    "amenity":"university",
    "name":"Escola Polit\u00e9cnica - UPE"
  }
},
{
  "type":"Feature",
```

```
"id":560628925,
"geometry":{
  "type":"Point",
  "coordinates":[
    -34.9282542,
    -8.0477011
  ]
},
"properties":{
  "amenity":"school",
  "name":"Col\u00e9gio Sancarlos"
}
},
```

Cada *query* dará origem a um sub conjunto de dados que será usado para criar escores de proximidade para a classe dado um imóvel (próxima seção) ou uma localidade como veremos na próxima seção.

3.3 Integração dos Dados

A fim de poder comparar imóveis de acordo com as características urbanas de seu entorno, usamos as queries definidas na seção anterior para definir escores para cada uma das classes Educação, Entreterimento, Financeiro, Saúde, Alimentação e Transporte. Isto foi feito adicionando seis atributos ao *dataset* de imóveis. Cada escore foi calculado em relação a o quão próximo um imóvel (ou localidade) está de elementos geográficos associados a cada classe. Para tanto, para cada classe e cada imóvel, foram filtrados todos os elementos do OSM dentro de um radio de 1 Km. Para cada um dos elementos resultantes foi calculada a distância para o imóvel e estas foram agregadas de acordo com a fórmula abaixo:

$$E = \sum_x \frac{r - d_x}{r},$$

onde $r = 1km$ e d_x representa a distância do imóvel ao elemento x A Figura 3.1 ilustra este processo. Para este exemplo existem quatro elementos (pontos verdes) próximos ao ponto de interesse (ponto preto). Neste caso, o escore é dado por:

$$E = \frac{r - d_1}{r} + \frac{r - d_2}{r} + \frac{r - d_3}{r} + \frac{r - d_4}{r}$$

Como resultado desse processo foi obtido um *dataset* com 2642 entidades e com valores concretos para título, quartos, preço, área útil, vagas, taxa condomínio, educação, entretenimento, financeiro, saúde, alimentação e transporte.

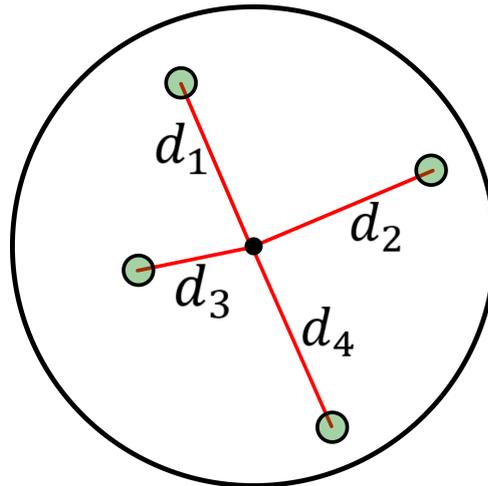


Figura 3.1: Computação dos escores de características urbanas. Os elementos de cada categoria dentro de um raio dado (1 km) são utilizados para computar o escore correspondente à categoria.

4

Sweet Home

Este capítulo apresenta o design do Sweet Home. Este processo de design se deu de forma iterativa e incremental, no qual tarefas analíticas foram elicitadas (Seção 4.1) e estas foram usadas na escolha das visualizações e interações implementadas no sistema.

4.1 Tarefas Analíticas

Partindo da premissa que a principal característica do Sweet Home é o entendimento do imóvel em relação a suas características externas, o sistema Sweet Home foi desenvolvido para suportar as seguintes tarefas analíticas:

- (T1) Conhecer a cidade a partir da identificação das características externas do imóvel.
- (T2) Possibilitar a identificação de imóveis com preço abaixo do valor em relação as valores das demais variáveis;
- (T3) Entender como as variáveis do ambiente influenciam o preço da residência;
- (T4) Possibilitar a manipulação de filtros, afim de possibilitar tanto filtros ligados as características internas quando características externas.

4.2 Sweet Home

Nesta seção é descrito o design do Sweet Home para suportar as tarefas analíticas descritas na seção anterior. A Figura 4.1 apresenta a interface do sistema. Esta interface é composta por três componentes principais: a vista geográfica, a vista de atributos de imóveis e um painel de controles. Estes componentes são descritos em detalhes no restante deste capítulo.

4.2.1 Vista Geográfica

O componente de visão geográfica é constituído de um mapa que mostra os aspectos espaciais dos dados. Este componente possibilita ver os imóveis presentes no conjunto de dados

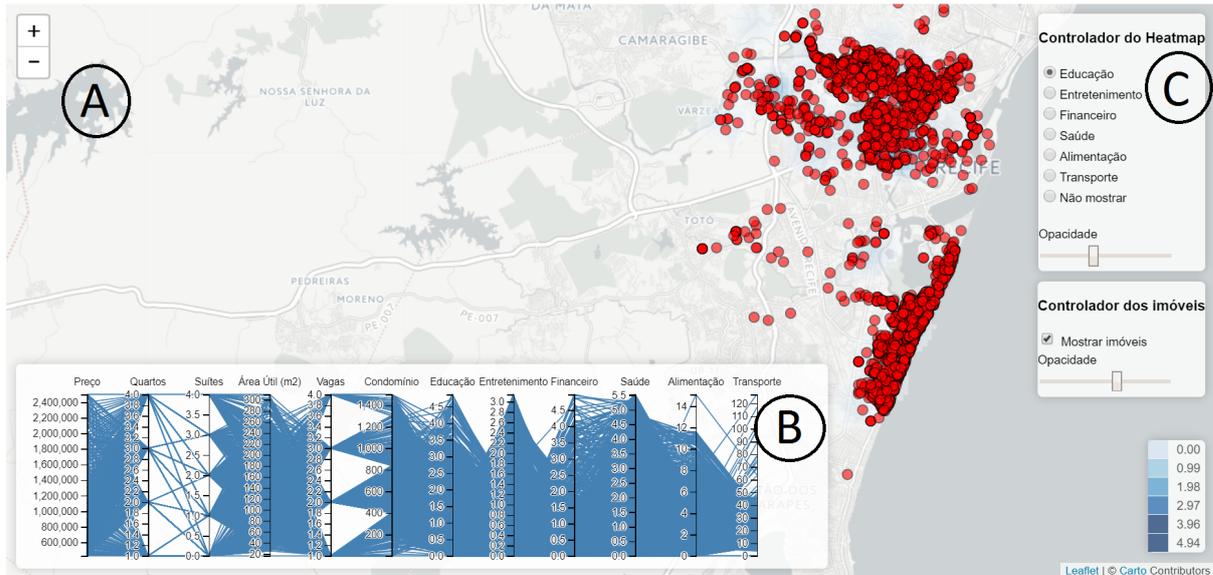


Figura 4.1: Interface do Sweet Home é composta por 3 componentes: a vista geográfica (A), a vista de atributos dos imóveis (B) e o painel controles do sistema (C).

de anúncios e as características externas dos imóveis por meio de *heatmaps*. Os imóveis são apresentados como círculos sobre o mapa. A fim de ver detalhes sobre cada imóvel, ao passar o mouse por cima do círculo correspondente, um *tooltip* é apresentado com os seus atributos. Para suportar a tarefa analítica T1 foi necessário a criação de *heatmaps*. Os *heatmaps* são apresentados em uma escala de cores sequencial obtida do ColorBrewer BREWER; HARROWER (2017) (canto direito inferior da Figura 4.1). Com as informações da base de dados recolhida do OSM, foi criado então um *grid*, de área retangular, sobre a cidade do Recife. O *grid* apresenta 10000 células, onde cada célula tem altura de aproximadamente 250 metros e largura de 180 metros. O tamanho da célula foi determinado empiricamente para satisfazer critérios de desempenho computacional e precisão geográfica. Cada célula é colorida com uma intensidade de cor referente a ao valor da célula. A função utilizada corresponde a mesma função citada na seção 3.1. O valor se refere à quantidade de itens que estão na área de um quilômetro de raio do centro da célula, com uma função de decaimento linear baseado na proximidade com o centro da célula. O valor da célula é pre-calculado e colocado em uma estrutura de dados, para então quando a página for carregada, o *web browser* irá colorir a célula com a cor referente ao valor anteriormente calculado na estrutura de dados. Se a célula não tem valor, a mesma não será colorida, ou seja não aparecerá no mapa. No controle do *heatmap*, existem seis opções de *heatmap* ou a opção de não mostrar o mesmo. Então para cada *heatmap* selecionado, o javascript irá remover e colorir novamente a célula com o novo valor referente ao seu centro. Afim de adaptar a necessidades específicas, é possível mudar a opacidade do *heatmap*. No canto inferior direito, existe uma legenda para cada intensidade de cor no *heatmap*, a mesma se modifica de acordo com a seção escolhida. A implementação do mapa foi feita utilizando a biblioteca *Leaflet.js*, direcionada à construção de aplicações *web* com mapas interativos. Além de proporcionar as interações de *pan* e *zoom* YOU et al. (2007), a biblioteca proporciona funções para a composição

de mapas interativos através da definição de camadas de informações geográficas indicando as coordenadas (latitue e longitude) de objetos a serem visualizados.

4.2.2 Vista de Atributos dos Imóveis

Para solucionar as tarefas analíticas T2, T3 e T4 foi necessário a criação de uma um gráfico de coordenadas paralelas INSELBERG; DIMSDALE (1990). Este gráfico é uma técnica de visualização de dados multi-dimensionais tradicional e bastante poderosa. Nela dimensões, atributos dos imóveis, são representadas por linhas verticais paralelas e cada ponto multi-dimensional (imóvel) é representado por uma linha poligonal ligando as coordenadas correspondentes. Interações de *brushing* podem ser utilizadas para selecionar pontos nesta visualização. Além disso, correlações podem ser investigadas por inspeção do padrão visual das linhas entre eixos consecutivos.

No Sweet Home, o gráfico apresenta doze dimensões (Figura 4.1 (B)) representando tanto atributos do imóvel quanto atributos do entorno urbano:

- Preço, valor anunciado do imóvel;
- Quartos, número de quartos do imóvel;
- Suítes, número de suítes do imóvel;
- Área útil, valor de área do imóvel em metros quadrados;
- Vagas, número de vagas na garagem;
- Condomínio, valor da taxa de condomínio em reais;
- Educação, valor referente ao retorno da função de proximidade com as escolas, universidades, jardim de infância e etc;
- Entretenimento, valor referente ao retorno da função de proximidade com os cinemas, teatros, centro de artes e etc;
- Financeiro, valor referente ao retorno da função de proximidade com os caixas eletrônicos, bancos e casas de câmbio;
- Saúde, valor referente ao retorno da função de proximidade com os hospitais, clínicas, farmácias e etc;
- Alimentação, valor referente ao retorno da função de proximidade com os bares, restaurantes, cafés e etc;
- Transporte, valor referente ao retorno da função de proximidade com as paradas de ônibus, paradas de metro, posto de combustível e etc.

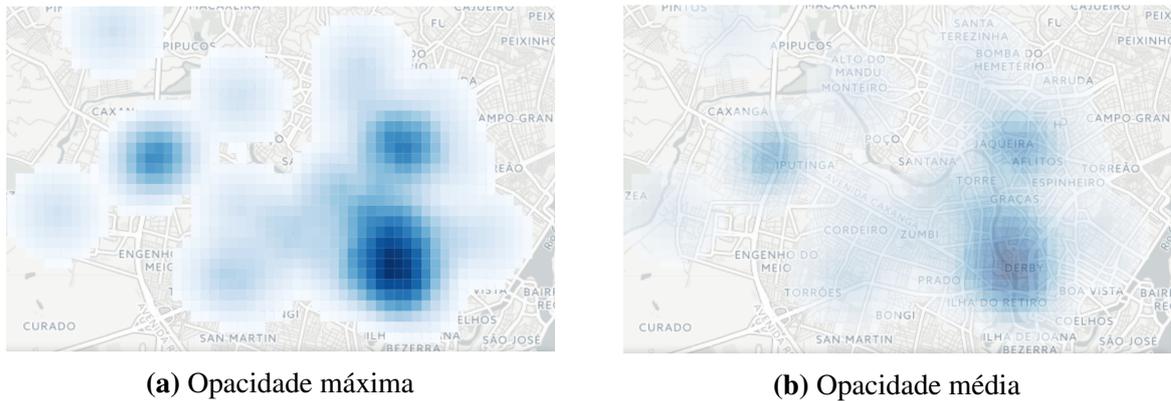


Figura 4.2: Exemplo de mudança de opacidade

4.2.3 Painel de Controles

O painel de controle é composto por dois controladores, um para os imóveis e o outro para os *heatmaps*. No controlador dos imóveis é possível remover os círculos do mapa, ou alterar a opacidade do mesmo, afim de atender a necessidades específicas. Já no controlador do *heatmap*, existem seis opções de *heatmap* ou a opção de não mostrar o mesmo. Cada opção de visualização do *heatmap* está ligado a uma categoria citada na Seção 3.2. Então para cada *heatmap* selecionado, o sistema irá atualizar a visualização de acordo com a escolha do usuário. Além disso, é possível ajustar a opacidade do heatmap para ser possível ver o mapa renderizado abaixo da camada do *heatmap*.

5

Casos de uso

Este capítulo demonstra a utilidade do Sweet Home através de uma série de casos de uso. Estes demonstram o poder e a versatilidade do sistema em objetivos indo da limpeza de dados até a identificação de padrões.

5.1 Detectando Erros nos Dados

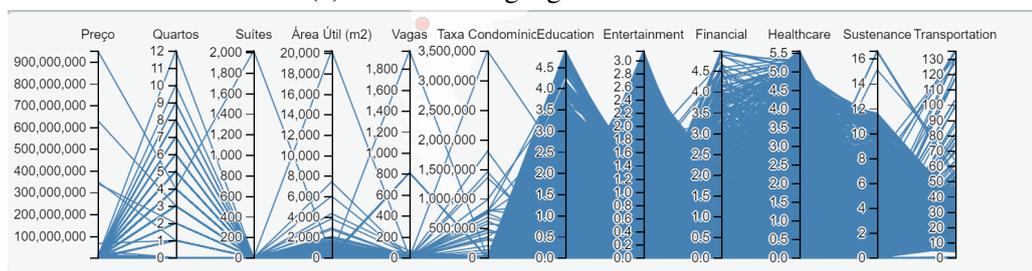
Através das ferramentas de visualização do Sweet Home é possível encontrar erros e inconsistências nos dados. Como um primeiro exemplo disso, somente observando a distribuição geográfica dos imóveis é possível ver que o processo de mineração usados para extrair os dados de imóveis em Recife. Na Figura 5.1a, é possível ver que foram minerados imóveis em outras localidades ou até mesmo com latitude e longitude incoerentes. Analogamente, inspecionando os atributos dos imóveis através do gráfico de coordenadas paralelas também é possível encontrar problemas nos dados. Vemos que em todas as coordenadas do conjunto de dados de anúncio (Preço, Quartos, Suítes, Área Útil, Vagas e Condomínio) existem problemas. Por exemplo, imóveis com área útil de 20000 metros quadrados ou com taxa de condomínio de mais de 3 milhões de reais. Estes erros foram a motivação para o processo de limpeza de dados descrito na Seção 3.1.

Em qualquer processo automático de curagem de dados (como o descrito acima), ainda é possível que instâncias de dados inconsistentes

Mesmo depois do processo de curagem dos dados descrito acima, Com o uso do gráfico de coordenadas paralelas é possível filtrar os dados de forma rápida. Ganhando uma facilidade na hora de identificar possíveis erros dos dados. Como exemplo as figuras 5.2 foi possível identificar incoerências entre a base de dados. Selecionando um vértice do gráfico é de esperar um número condizente com esse vértice, como por exemplo área útil está diretamente ligada ao número de quartos. No exemplo mostrado, foi selecionado apartamentos com áreas úteis entre 16 e 40 metros quadrados, porém uma linha evidencia que existe um apartamento com 3 quartos e área útil de 16 metros quadrados, logo a aplicação mostra de maneira fácil as anomalias na base de dados.



(a) Incoerências geográficas.

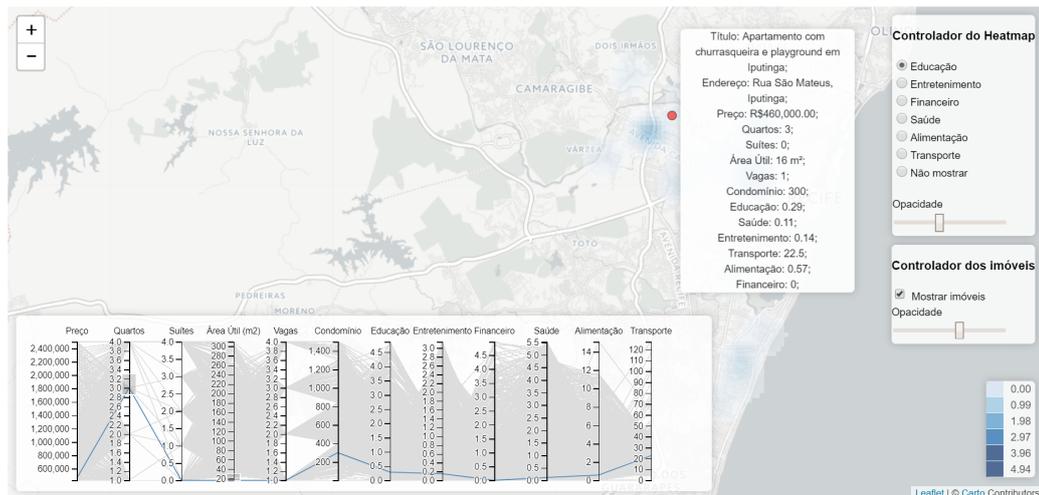


(b) Incoerências nos valores dos atributos.

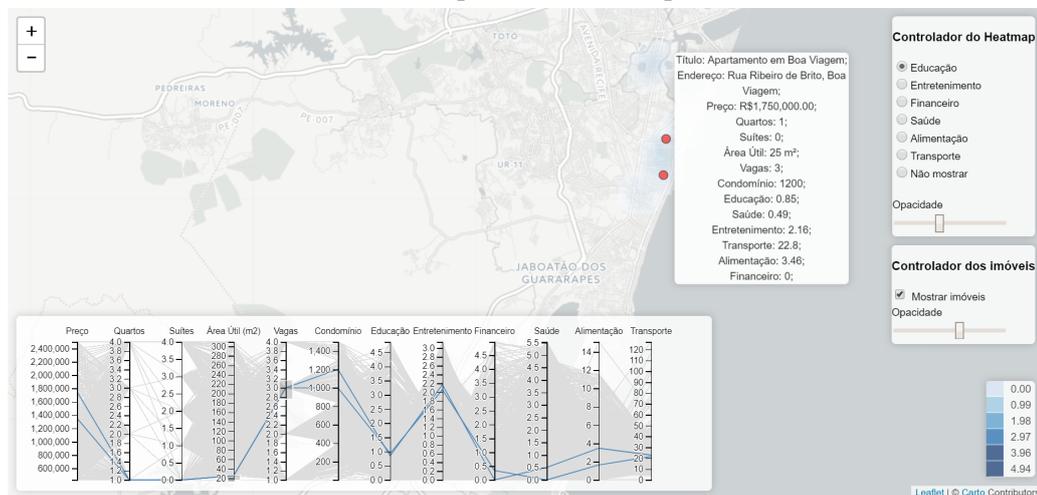
Figura 5.1: Valores inconsistentes encontrados através das visualizações do Sweet Home.

5.2 Entendendo a Cidade

O Sweet Home possibilita um maior entendimento da cidade. Como por exemplo, por meio dos *heatmaps* é possível conhecer a cidade e suas regiões. Na figura 5.3f é possível identificar o centro da cidade por meio da maior intensidade de transporte do mesmo. Outra inferência que é possível fazer se dá na forma do *heatmap* financeiro 5.3d. Analisando o *heatmap* financeiro 5.3d é possível identificar áreas com maior movimentação financeira ou de maior comércio. É natural o surgimento de bancos ou caixas eletrônicos em área de maior movimentação monetária. Com a união de *heatmaps* é possível extrair informações mais complexas. Um exemplo é observação de que transporte, financeiro e alimentação aparecem de forma mais intensa no centro da cidade. Pois o centro é onde se acumula o maior número de trabalhadores diários, logo é natural o aparecimento de restaurantes para suprir essa demanda alimentícia. Já em outras áreas com muito entretenimento e alimentação mediana dar-se a entender que são locais mais focados em finais de semanas, e não para o dia-a-dia. Outra forma de conhecer a cidade é por meio dos imóveis. Tipicamente cada cidade tem suas áreas mais "nobres", áreas que não necessariamente estão ligadas a proximidade com características externas, mas sim ligadas a uma convenção social que agrega aquela área um valor econômico.



(a) 60 metros quadrados com 7 quartos



(b) Apartamento de 25 metros quadrados com 1 quarto e 3 vagas na garagem

Figura 5.2: Anomalias encontradas na base de dados

5.3 Buscando Imóveis

Essa seção será dedicada para mostrar exemplos de buscas relacionadas na escolha do imóvel.

A figura 5.4 exemplifica uma busca por valores referentes a estrutura do imóvel, como por exemplo número de vagas, quartos, suítes e quartos. Nesse tipo de buscar, o usuário deseja buscar um imóvel de acordo com suas características internas, desprezando toda as características externas. Possibilitando ao usuário buscar da mesma forma que os outros sites fazem.

A figura 5.5 exemplifica uma busca por valores referentes às características externas do imóvel. Com a opção de filtrar por características externas, o usuário pode selecionar características do seu interesse pessoal.

A figura 5.6 exemplifica uma busca com filtros de características internas e externas. Esse tipo de busca apresenta um ganho em relação a busca tradicional pois acrescenta valores

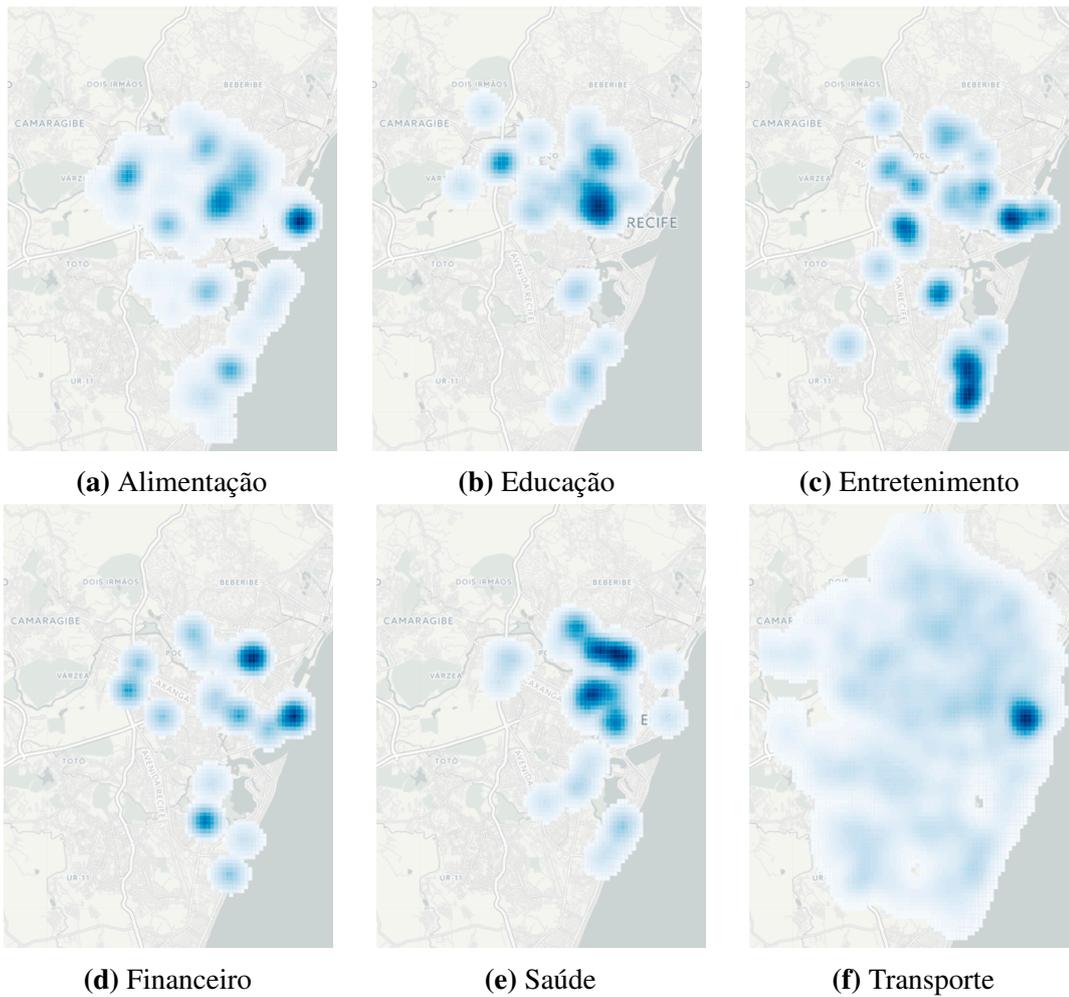


Figura 5.3: Heatmaps de Recife

externos a já conhecida busca tradicional com características internas. No exemplo da figura, o usuário seleciona características internas e externas, e é possível ver que área útil apresenta dois imóveis com 280 e 200 metros quadrados, enquanto os outros imóveis ficam 80 e 60 metros quadrados. Como consequência disso o usuário pode comprar um imóvel com as mesmas características, mas com um ganho em outras características. Sendo assim o usuário encontra um imóvel melhor.

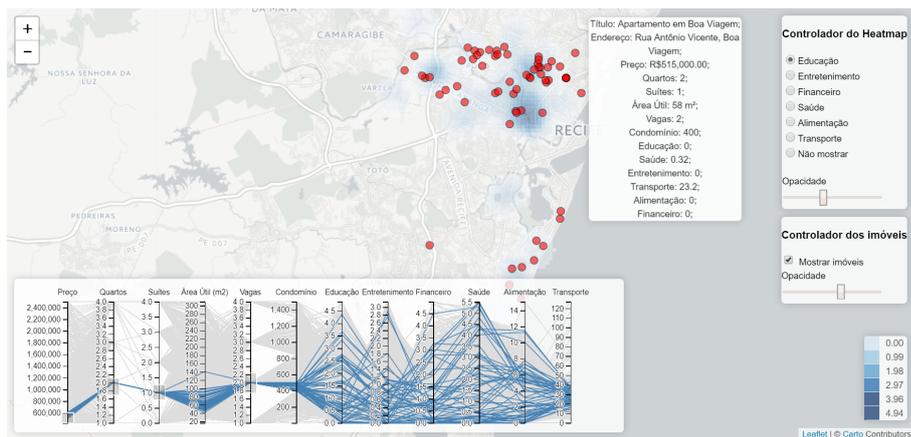


Figura 5.4: Busca utilizando filtros relacionados a estrutura física do imóvel

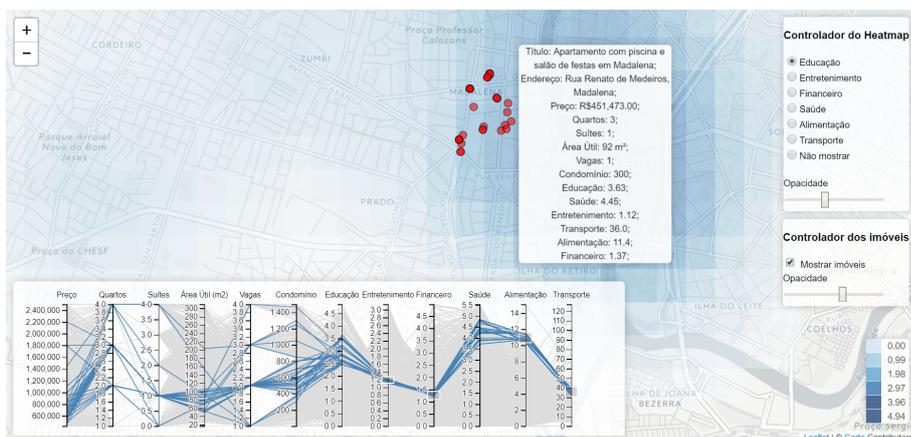


Figura 5.5: Busca utilizando filtros relacionados as características externas

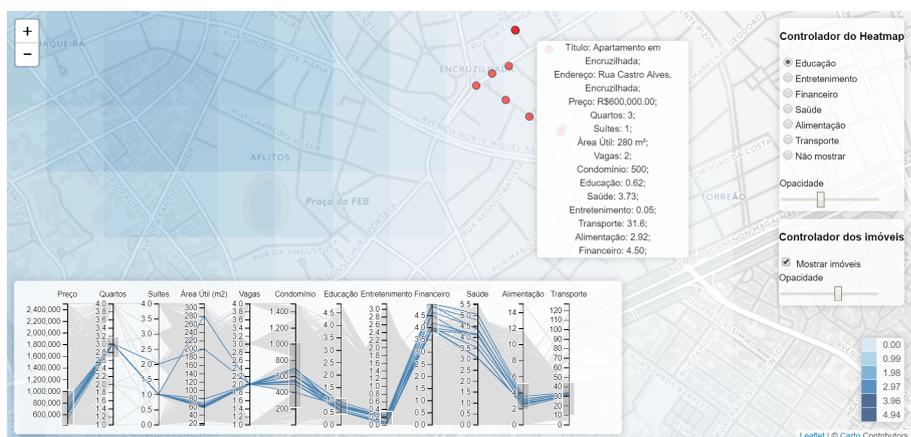


Figura 5.6: Busca encontrando um apartamento maior com as mesmas características externas

6

Conclusão

Esse trabalho propôs a criação de um ferramenta web chamada Sweet Home. na qual se concentra na visualização de dados de imóveis para dar suporte a decisão na hora de compra. Como estudo de caso foi usado a cidade do Recife pois tanto o orientando quanto o orientador possuem conhecimento da cidade. O Sweet Home cumpriu o objetivo proposto por que o mesmo além de cumprir todas tarefas analíticas (Seção 4.1) como explanado durante toda a seção 4.2 ainda mostra de maneira fácil e rápida um grande volume de dados. Como consequência dessa facilitação é possível uma comparação rápida e eficiente entre imóveis. Com o poder de melhor comparação o usuário consegue um melhor suporte no momento de busca para compra de imóvel.

6.1 Melhorias e Trabalhos futuros

Mesmo cumprindo o seu objetivo, o Sweet Home deixa espaço para melhorias. Tais melhorias se dariam em uma interação entre o servidor e a pagina web em si. No qual o servidor contaria novamente os rates em relação a localidade a partir de um tempo proposto. Outra possível melhoria seria na interface gráfica. Um ciclo de testes com usuário afim de entender pontos que não estão tão claros para eles. Outra possível melhoria seria uma forma de extrair a base de imóveis periodicamente afim de expandir para mais cidades e contextos. Uma melhoria ligada diretamente a base de dados seria utilizar outra fonte de dados em relação as características externas. Mesmo o OSM tendo uma boa quantidade de informação, ele não é tão popular aqui no Brasil como o Google Maps, então uma melhoria futura seria a utilização de outro mapa com mais informação, como por exemplo o Google Maps.

Referências

- BREWER, C. A.; HARROWER, M. ColorBrewer 2.0. **Web site: [http://www. ColorBrewer.org/](http://www.ColorBrewer.org/). Acessado Julho**, [S.l.], 2017.
- CHANG, R. et al. Legible cities: focus-dependent multi-resolution visualization of urban relationships. **IEEE transactions on visualization and computer graphics**, [S.l.], v.13, n.6, p.1169–1175, 2007.
- DESIMONE, B. Why Location Matters in Real Estate. **Web site: <http://www.foxbusiness.com/features/2013/10/22/why-location-matters-in-real-estate.html>. Acessado Julho**, [S.l.], 2017.
- FERREIRA, N. et al. Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v.19, n.12, p.2149–2158, 2013.
- FERREIRA, N. et al. Urbane: a 3d framework to support data driven decision making in urban development. In: VISUAL ANALYTICS SCIENCE AND TECHNOLOGY (VAST), 2015 IEEE CONFERENCE ON. **Anais...** [S.l.: s.n.], 2015. p.97–104.
- HAKLAY, M.; WEBER, P. Openstreetmap: user-generated street maps. **IEEE Pervasive Computing**, [S.l.], v.7, n.4, p.12–18, 2008.
- INSELBERG, A.; DIMSDALE, B. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: VISUALIZATION'90, 1. **Proceedings...** [S.l.: s.n.], 1990. p.361–378.
- JANETZKO, H. et al. Anomaly detection for visual analytics of power consumption data. **Computers & Graphics**, [S.l.], v.38, p.27–37, 2014.
- MUNZNER, T. **Visualization Analysis and Design**. [S.l.]: CRC Press, 2014.
- SILVA, D. S. d. Comportamento do consumidor: critérios de escolha adotados para compra de imóveis residenciais no plano piloto em Brasília df. , [S.l.], 2009.
- YOU, M. et al. A usability evaluation of web map zoom and pan functions. **International Journal of Design**, [S.l.], v.1, n.1, 2007.

7

Apêndice

7.1 Consultas ao OSM

Listing 7.1: Query para coleta de dados em educação

```
responseEducation = api.Get("""
    area[name = "%s"]->.a;
    (
        node(area.a) [amenity~"college|kindergarten|library|
            public_bookcase|school|music_school|driving_school
            |language_school|university"];
    )
    "" " % city)
```

Listing 7.2: Query para coleta de dados em entretenimento

```
responseEntertainment = api.Get("""
    area[name = "%s"]->.a;
    (
        node(area.a) [amenity~"arts_centre|brothel|casino|
            cinema|community_centre|fountain|gambling|
            nightclub|planetarium|social_centre|stripclub|
            studio|swingerclub|theatre"];
        node(area.a) [leisure];
    )
    "" " % city)
```

Listing 7.3: Query para coleta de dados em financeiro

```
responseFinancial = api.Get("""
    area[name = "%s"]->.a;
```

```
(
  node(area.a) [amenity~"atm|bank|bureau_de_change"];
)
"" % city)
```

Listing 7.4: Query para coleta de dados em saúde

```
responseHealthcare = api.Get("""
  area[name = "%s"]->.a;
(
  node(area.a) [amenity~"baby_hatch|clinic|dentist|
  doctors|hospital|nursing_home|pharmacy|
  social_facility|veterinary|blood_donation"];
)
"" % city)
```

Listing 7.5: Query para coleta de dados em alimentação

```
responseSustenance = api.Get("""
  area[name = "%s"]->.a;
(
  node(area.a) [amenity~"bar|bbq|biergarten|cafe|
  drinking_water|fast_food|food_court|ice_cream|pub|
  restaurant"];
)
"" % city)
```

Listing 7.6: Query para coleta de dados em transporte

```
responseTransportation = api.Get("""
  area[name = "%s"]->.a;
(
  node(area.a) [amenity~"bicycle_parking|
  bicycle_repair_station|bicycle_rental|boat_sharing
  |bus_station|car_rental|car_sharing|car_wash|
  charging_station|ferry_terminal|fuel|grit_bin|
  motorcycle_parking|parking|parking_entrance|
  parking_space|taxi"];
  node(area.a) [highway ~ bus_stop];
)
"" % city)
```