INSTITUTO FEDERAL DE SÃO PAULO
CAMPUS SÃO PAULO
CURSO DE ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

RELATÓRIO ICDA6

Alexandre Yudi I. de Oliveira - SP3046923
Alkindar José Ferraz Rodrigues - SP3029956
André Correia Zarzur - SP305201X
Cecília Braz da Silva - SP3049876
Marcelo Carlos Olímpio Junior - SP3046583

São Paulo-SP
2022

Sumário

1 Introdução

Este relatório procura estabelecer uma relação entre a poluição na cidade de São Paulo e tendências do mercado imobiliário da cidade. Para tanto, utilizamos uma base com o histórico de poluição desde 2013 e uma outra com um snapshot do mercado imobiliário em abril de 2019. Esperamos demonstrar que locais com tendencia de alta poluição apresentam um mercado imobiliário menos oportuno e com menor valorização e estabelecer um modelo que permita localizar o valor de um imóvel com base na tendência de poluição da região onde ele se encontra. As duas bases pertencentes ao estudo foram selecionadas no *website* Kaggle (KAGGLE, 2022).

Portanto, procuramos entender a distribuição dos poluentes para o mês de referência, as caracteristícas do mercado imobiliário de São Paulo para o mesmo mês. Para isso, olha-se os valores de alguel e condomínio praticados, a área dos imóveis, número de quartos, vagas de garagem e banheiros; e na presença de comodidades, com piscina e elevadores. Por fim, estabelece-se um modelo de regra de associação, que relaciona caracteristicas dos imóveis, separadas em faixas, com faixas de medidas dos poluentes registradas na estação de medição mais próxima, calculado com base nas coordenadas do imóvel e da estação.

Os resultados das análise supracitadas estão disponíveis na sessão 2.4.

2 Ciclo de Vida dos Dados

O objetivo desse capítulo é realizar a descrição do ciclo de vida dos dados utilizados para o desenvolvimento do relatório de acordo com a Figura 1, passando desde a produção até o descarte dos mesmos.

Figura 1 – Ciclo de Vida dos Dados

Fonte: Sirqueira (2018).

2.1 Produção

A etapa de Produção tem como objetivo indicar como foi feita a coleta das informações independente do meio utilizado. Como estão sendo utilizadas duas bases, a etapa de Produção ocorreu de duas formas diferentes:

- Base de poluição no ar de São Paulo
  - Os dados de poluição no ar foram coletados através de estações de medição de poluição presentes no estado de São Paulo. Neste caso específico, os dados foram produzidos e provêm da Companhia Ambiental do Estado de São Paulo (CETESB), agência do governo do estado responsável pelo controle, fiscalização, monitoramento e licenciamento de atividades geradoras de poluição em São Paulo.
- Base de valores imobiliários
  - Os dados de valores imobiliários foram coletados de diversas fontes diferentes, as principais sendo *websites* imobiliários. Apesar de não termos a proveniência exata dos dados utilizados, eles foram produzidos em diversos meios de aplicações do mercado imobiliário em São Paulo.

2.2 Armazenamento

A etapa de Armazenamento se refere a como os dados foram retidos independente do meio utilizado. No caso das duas bases, temos o mesmo processo de armazenamento, os dados produtivos tanto da poluição do ar de São Paulo, quanto do mercado imobiliário foram agrupados em uma base de dados relacional, e a sua utilização para a realização de análises através eles se encontram organizados em um documento eletrônico em formato CSV.

2.3 Transformação

A etapa de Transformação é essencialmente a alteração da estrutura dos dados para a adequação a processos específicos. De acordo com os autores dos conjuntos de dados obtidos no Kaggle, no caso das duas bases selecionadas, temos a obtenção dos dados que estão armazenados e estruturados em bancos de dados relacionais (RDBMS) através de técnicas de raspagem de dados, para que seja feita a transformação em uma estrutura mais simples e concisa de ser analisada, arquivos CSV, como anteriormente citado.

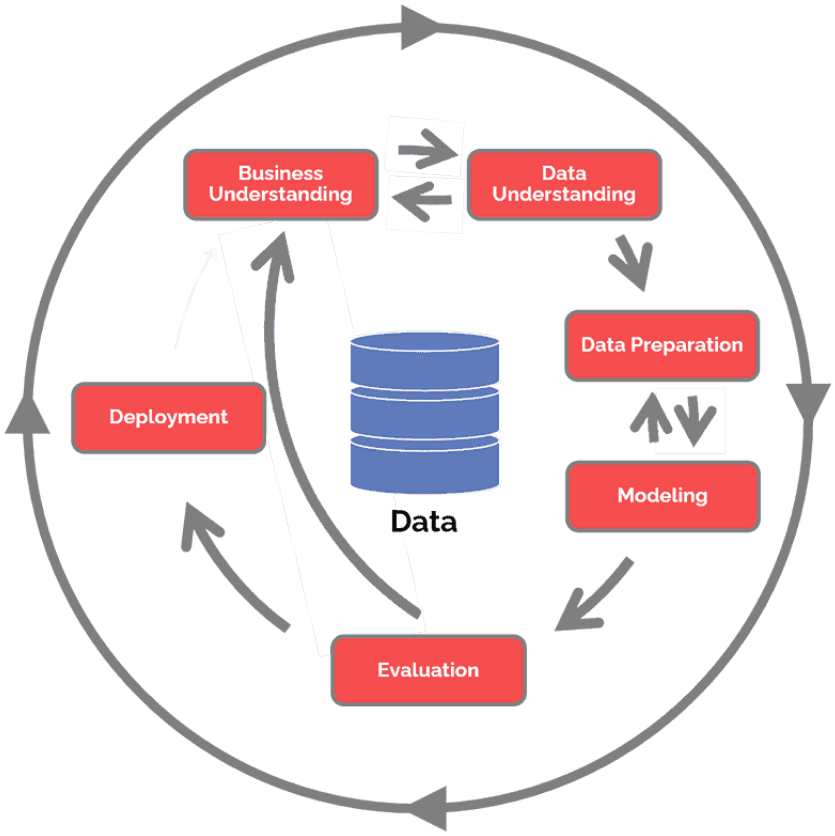
2.4 Análise

A análise de dados é uma etapa crucial para a obtenção de modelos através do uso de técnicas estatísticas, já que é a etapa do ciclo de vida com o maior esforço para o presente relatório. Aqui utilizamos de ferramentas e processos para a extração de informação e conhecimento a partir das duas bases de dados selecionadas por meio da análise exploratória, explícita e implícita. Com elas, podemos encontrar padrões nos dados e avaliar os modelos de acordo com a necessidade do estudo. No entanto, será realizado o trabalho de analisar os dados para observar a relação entre a poluição do ar com o mercado imobiliário da cidade de São Paulo.

2.4.1 CRISP-DM

O CRISP-DM é um dos padrões de modelos de *datamining* tradicionais, sendo o mais conhecido e adotado. Sua estrutura é composta por 6 fases conforme a Figura 2

Figura 2 – Modelo CRISP-DM



Fonte: Hotz (2022).

2.4.1.1 *Bussiness Understanding*

O *Bussiness Understanding* (Entendimento do Negócio) é a fase em que deverá ser compreendido as características e as necessidades do negócio. Para o projeto, o negócio sendo trabalhado faz parte do mercado imobiliário, sendo assim, é importante entender as tendências que influenciam no sucesso das vendas. Como esse negócio tem um perfil naturalmente relacionado a localização, foi decidido trabalhar com bases que utilizam esse tipo de informação. O objetivo da análise é de estabelecer uma relação entre a poluição na cidade de São Paulo e tendências do mercado imobiliário da cidade, tendo assim como objetivo demonstrar que locais com tendencia de alta poluição apresentam um mercado imobiliário menos oportuno e com menor valorização.

2.4.1.2 *Data Understanding*

É na fase de *Data Understanding* (Entendimento dos Dados) que ocorrerá a avaliação dos dados selecionados para a criação do modelo. Os dados de poluição estão divididos em duas bases, uma referente à bairros no estado de São Paulo junto com o seu índice de poluição e a sua localização em latitude e longitude (estacoes-cetesb.csv), e outra contendo as informações detalhas desse índice citado (cetesb.csv). Essas duas bases estão relacionadas através de um id.

Já os dados imobiliários estão contidos em uma base única, essa base contendo informações que caracterizam o imóvel assim como a sua localização em forma tanto de bairro como de latitude e longitude (sao-paulo-properties-april-2019.csv).

Levando todas as bases em consideração, os principais dados que serão utilizados serão as colunas referentes à localização dos imóveis, as informações relacionadas ao índice de poluição dos bairros, e os dados que caracterizam o imóvel, como por exemplo a coluna *Furnished* que indica se um imóvel estará mobiliado ou não.

Deverá ser realizado um *Join* entre as as bases de poluição utilizando os ids em comum e, após isso, outro *Join* entre essa nova base criada e a base imobiliária através das informações de latitude e longitude.

2.4.1.3 *Data Preparation*

Na *Data Preparation* (Preparação dos Dados) os dados serão selecionados e organizados. Como citado na sessão de *Data Understanding* 2.4.1.2, será realizado o *Join* entre as bases de poluição através das colunas de id e, posteriormente, mais um *Join*, dessa vez entre a nova base de poluição e a de dados imobiliários atráves das informações de latitute e longitude.

Porém, antes desses *Joins*, foram realizados alguns filtros e transformações nas bases. Primeiramente, na base de poluição referente aos bairros (estacoes-cetesb.csv), foi realizado um filtro para selecionar apenas bairros na cidade de São Paulo. Na base de poluição referente aos dados detalhados dos índices (cetesb.csv), foram selecionadas apenas as colunas dos poluentes, a do tempo e a do id, sendo realizado juntamente com esse *select* uma transformação da coluna de tempo em data.

Na base de dados imobiliários (sao-paulo-properties-april-2019.csv) foram realizadas diversas transformações. Primeiro, foi feito um filtro para selecionar apenas os imóveis de aluguel. Após isso, os campos *Swimming Pool*, *Elevator*, *Furnished*, *Property* Type e *New* foram transformados para *boolean*. Por útltimo, foi realizado um *factor* na coluna *Negotiation Type*.

Em relação aos dados *missing*, eles estavam presentes somente nas bases de poluição. Para realizar a tratativa deles, foi utilizado o método *pairwise deletion*, em que é excluído apenas as observações que possuem dados *missing* na variável utilizada para análise. (RODRIGUES, 2020)

2.4.1.4 *Modeling*

O tópico de *Modeling* (Modelagem) é a fase na qual são aplicadas as técnicas de mineração de dados para assim realizar a análise deles.

- **Análise Explícita**

  Para fazer a análise explícita foram feitas operações de baixa complexidade. Assim sendo, na base dos registros de poluição, foi utilizado um *aggregate* para juntar todas todos os registros do mesmo dia em um único e, consequentemente, somar a emissão de um

poluente alvo. Essa operação foi feita para cada poluente, mas, para fim de exemplo, veja as Figuras 3 e 4.

Figura 3 – Análise Explícita - Agregar por Data

```
# somando o total de co registrado por dia
agr <- aggregate(tratado["co"], by=tratado["data"], sum)
head(agr)
```

Fonte: Autores do documento

Figura 4 – Análise Explícita - Agregar por Data: Resultado

```
> head(agr)
        data   co
1 2013-05-09 54.3
2 2013-05-10 55.4
3 2013-05-11 52.3
4 2013-05-12 67.2
5 2013-05-13 88.7
6 2013-05-14 96.8
```

Fonte: Autores do documento

Já na base de alugueis, primeiro foi feito um filtro para apenas consultar as moradias que estavam disponíveis para aluguel e não o que estavam disponíveis para venda, por exemplo. Após isso ainda foi realizado um *summary* que pode ser consultado no código. Veja a operação do filtro e o resultado, respectivamente, nas Figuras 5 e 6.

Figura 5 – Análise Explícita - Exemplo de *Filter*

```
# selecionar apenas os registros de aluguel
apenas_alugueis <- aluguel %>% filter(`Negotiation Type` == "rent")
apenas_alugueis
```

Fonte: Autores do documento

Figura 6 – Análise Explícita - Resultado do *Filter*

```
`Negotiation Type`
<fct>
rent
rent
rent
rent
rent
rent
```

Fonte: Autores do documento

- **Análise Exploratória**

Na Análise Exploratória serão utilizados métodos visuais e quantitativos para resumir os conjuntos de dados com o objetivo de simplificar a leitura da base sem fazer suposições sobre o seu conteúdo.

Primeiramente, foi realizado um *Summary* das estatísticas de cada coluna da base de poluição (após *Join* das duas). Nesse resumo pode ser visto todas as estatística básicas de cada um dos poluentes na base. O resultado pode ser visto nas Figuras 7 e 8.

Figura 7 – Análise Exploratória - Estatíticas Poluição 1

```
> summary(pollution_records)
      idt              data                nome                co
 Min.   : 8.00   Min.   :2013-05-09   Length:660101      Min.   :0.0
 1st Qu.:22.00   1st Qu.:2014-12-02   Class :character   1st Qu.:0.4
 Median :30.00   Median :2016-06-27   Mode  :character   Median :0.6
 Mean   :30.17   Mean   :2016-06-27                      Mean   :0.8
 3rd Qu.:36.00   3rd Qu.:2018-01-21                      3rd Qu.:1.0
 Max.   :53.00   Max.   :2019-08-17                      Max.   :8.0
                 NA's   :5                               NA's   :525893
```

Fonte: Autores do documento

Figura 8 – Análise Exploratória - Estatíticas Poluição 2

```
      no2          particulado10    particulado2.5       ozonio
 Min.   :  0.00   Min.   :  0.00   Min.   :  0.0    Min.   :  0.00
 1st Qu.: 20.00   1st Qu.: 17.00   1st Qu.:  9.0    1st Qu.: 11.00
 Median : 35.00   Median : 29.00   Median : 15.0    Median : 30.00
 Mean   : 39.49   Mean   : 39.94   Mean   : 19.4    Mean   : 35.32
 3rd Qu.: 53.00   3rd Qu.: 49.00   3rd Qu.: 25.0    3rd Qu.: 49.00
 Max.   :278.00   Max.   :978.00   Max.   :920.0    Max.   :347.00
 NA's   :288541   NA's   :133950   NA's   :469030   NA's   :315185
```

Fonte: Autores do documento

Depois, foi feito a mesma coisa, dessa vez com a base dos dados imobiliários. O resultado pode ser visto nas Figuras 9 e 10.

Figura 9 – Análise Exploratória - Estatíticas Imobiliária 1

```
> summary(apenas_alugueis)
     Price            Condo             Size            Rooms            Toilets
 Min.   :  480   Min.   :   0.0   Min.   : 30.00   Min.   : 1.000   Min.   :1.000
 1st Qu.: 1350   1st Qu.: 395.8   1st Qu.: 52.00   1st Qu.: 2.000   1st Qu.:2.000
 Median : 2000   Median : 595.0   Median : 67.00   Median : 2.000   Median :2.000
 Mean   : 3078   Mean   : 825.2   Mean   : 89.49   Mean   : 2.304   Mean   :2.106
 3rd Qu.: 3300   3rd Qu.: 990.0   3rd Qu.:100.00   3rd Qu.: 3.000   3rd Qu.:2.000
 Max.   :50000   Max.   :9500.0   Max.   :880.00   Max.   :10.000   Max.   :8.000
 Swimming Pool       New             District         Negotiation Type Property Type
 Mode :logical   Mode :logical   Length:7228        rent:7228        Mode :logical
 FALSE:3701      FALSE:7222      Class :character   sale:   0        FALSE:7228
 TRUE :3527      TRUE :6         Mode  :character
```

Fonte: Autores do documento

Figura 10 – Análise Exploratória - Estatíticas Imobiliária 2

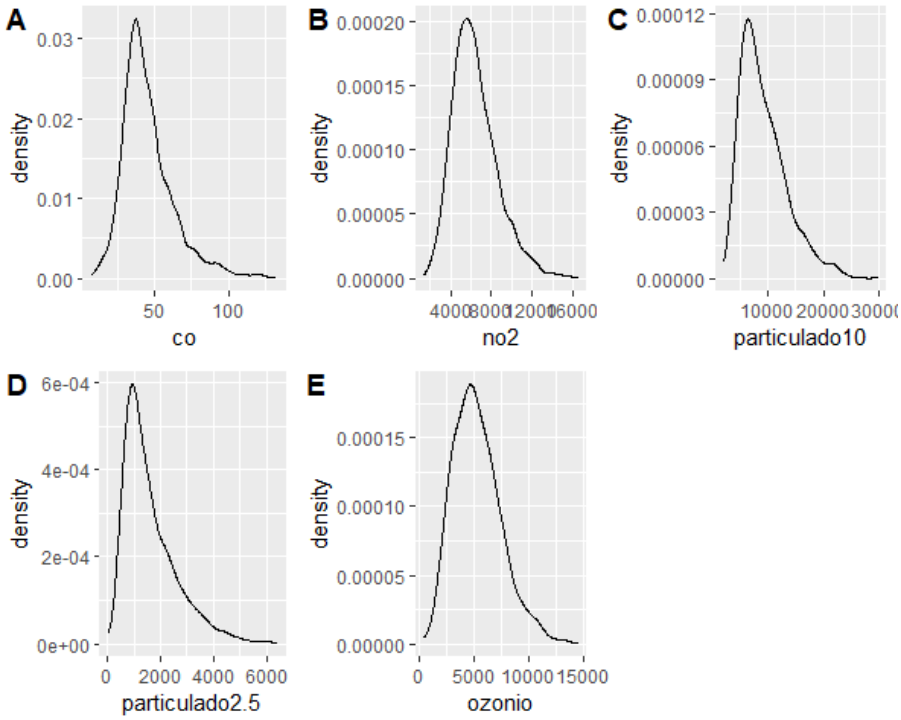
```
     Suites          Parking         Elevator        Furnished
 Min.   :0.000   Min.   :0.000   Mode :logical   Mode :logical
 1st Qu.:1.000   1st Qu.:1.000   FALSE:5061      FALSE:5978
 Median :1.000   Median :1.000   TRUE :2167      TRUE :1250
 Mean   :1.024   Mean   :1.452
 3rd Qu.:1.000   3rd Qu.:2.000
 Max.   :5.000   Max.   :9.000
     Latitude        Longitude
 Min.   :-46.75   Min.   :-58.36
 1st Qu.:-23.60   1st Qu.:-46.69
 Median :-23.56   Median :-46.64
 Mean   :-22.03   Mean   :-43.50
 3rd Qu.:-23.52   3rd Qu.:-46.59
 Max.   :  0.00   Max.   :  0.00
```

Fonte: Autores do documento

Voltando para a base de poluição, foi feita a soma dos dados de um poluente por dia ao longo dos anos de 2013 à 2019, a fim de visualizar a densidade do valor dele com a biblioteca ggplot2. A operação foi feita para todos os poluentes gerando 5 gráficos de densidade que podem ser vistos na Figura 11.

Dessa forma ficou mais simples de visualizar as médias e os *outliers* de cada poluente mesmo que seus valores sejam muito diferentes em números absolutos. Lembrando que para cada poluente foi feito um tratamento dos dados *missing* antes da operação.

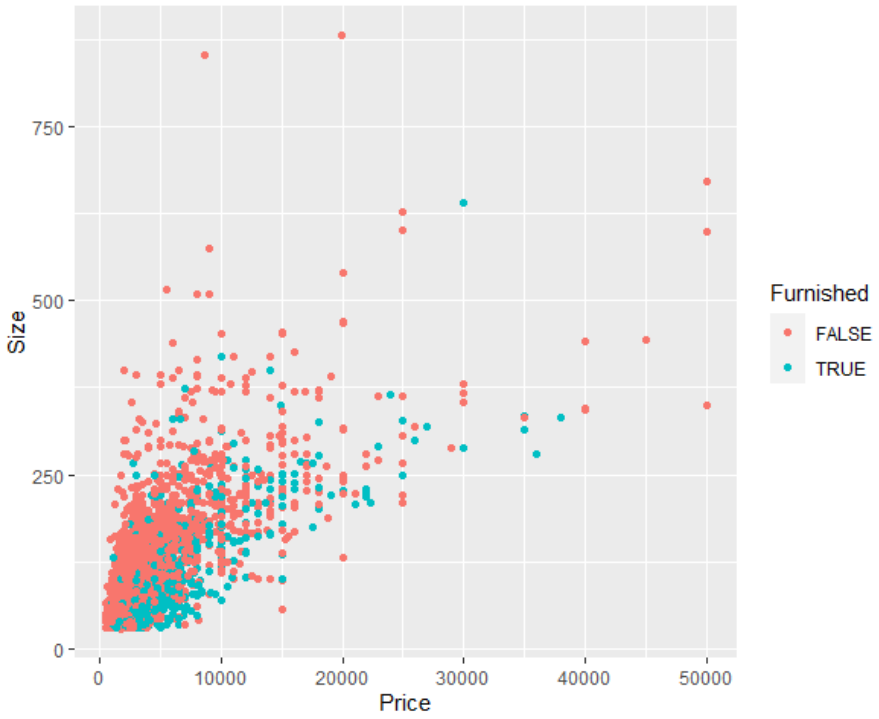
Figura 11 – Análise Exploratória - Densidade de Poluentes



Fonte: Autores do documento

Para a base de alugueis, o ggplot foi utilizado para construir um gráfico de pontos o qual relaciona o Preço, e se a moradia é mobiliada, com outro fator. Começando pelo tamanho, veja a relação entre eles na Figura 12.

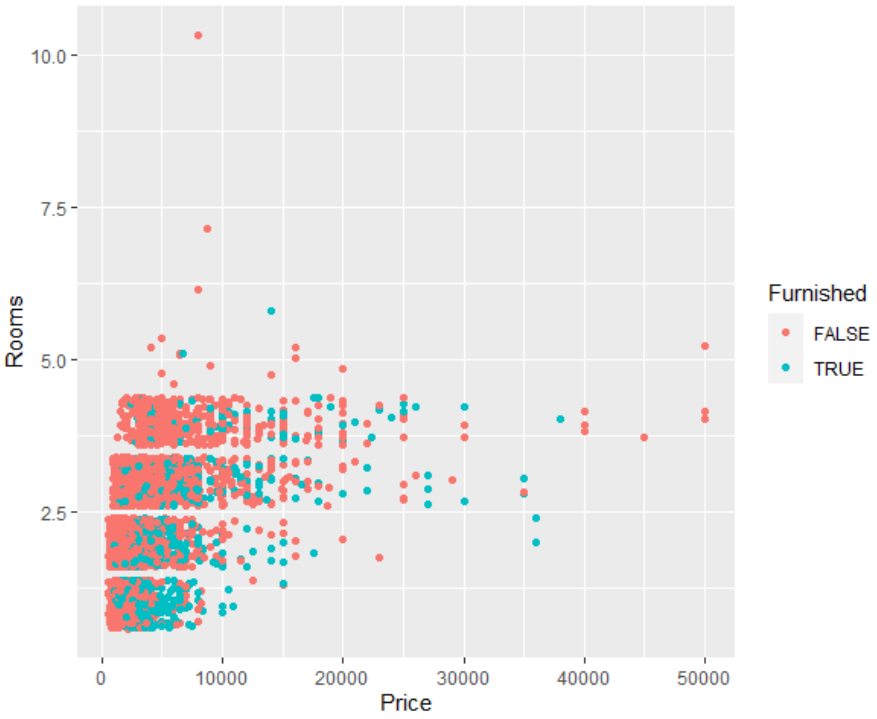
Figura 12 – Análise Exploratória - Tamanho X Preço X Mobiliado



Fonte: Autores do documento

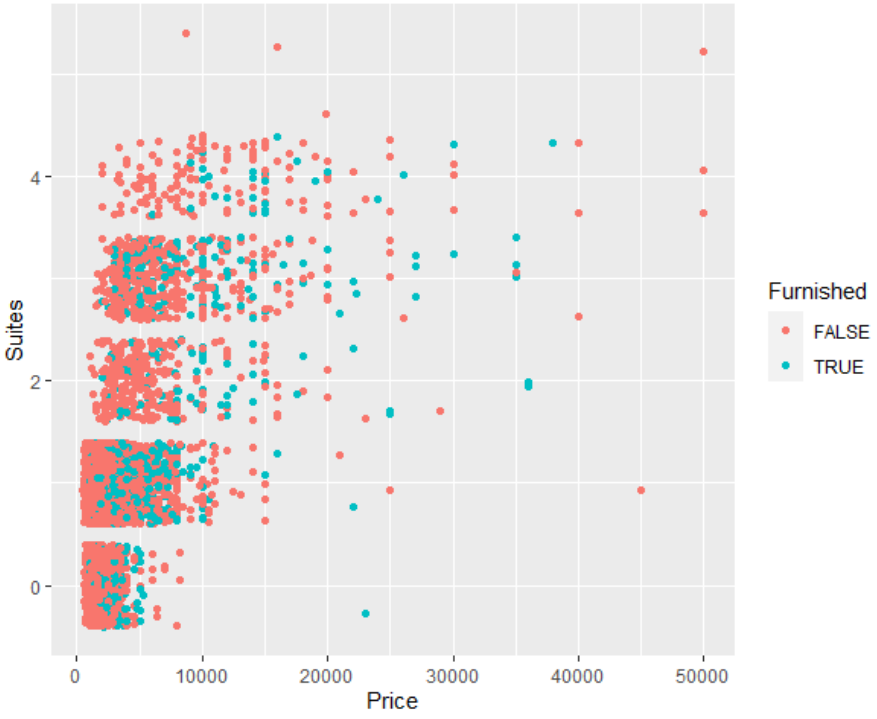
A mesma operação foi feita para os outros fatores como quantidade de quartos, quantidade de suítes e quantidade de vagas de estacionamento. Todos podem ser vistos respectivamente nas Figuras 13, 14 e 15.

Figura 13 – Análise Exploratória - Quartos X Preço X Mobiliado



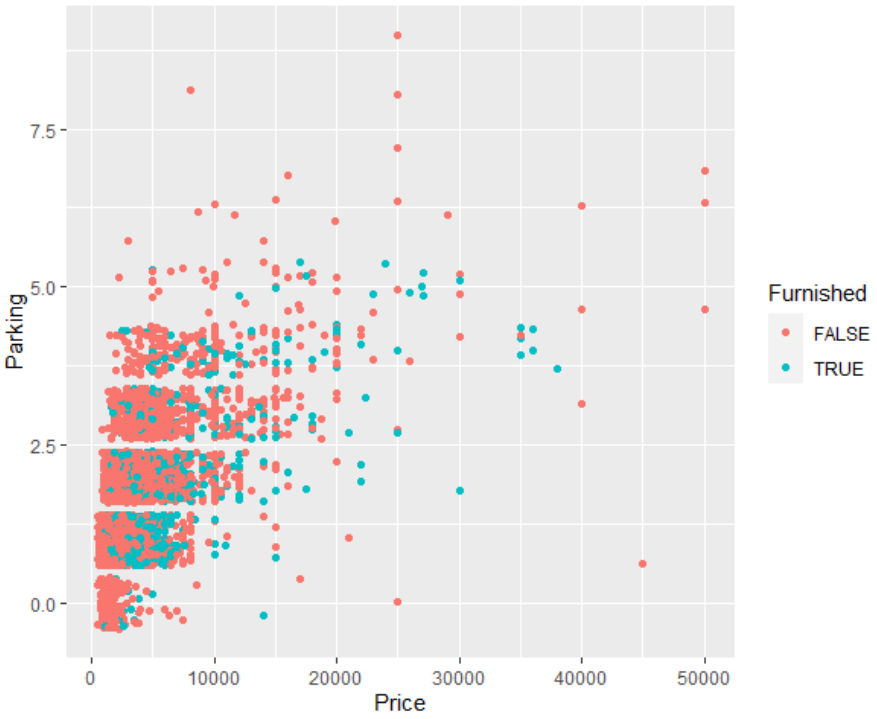
Fonte: Autores do documento

Figura 14 – Análise Exploratória - Suítes X Preço X Mobiliado



Fonte: Autores do documento

Figura 15 – Análise Exploratória - Vagas X Preço X Mobiliado

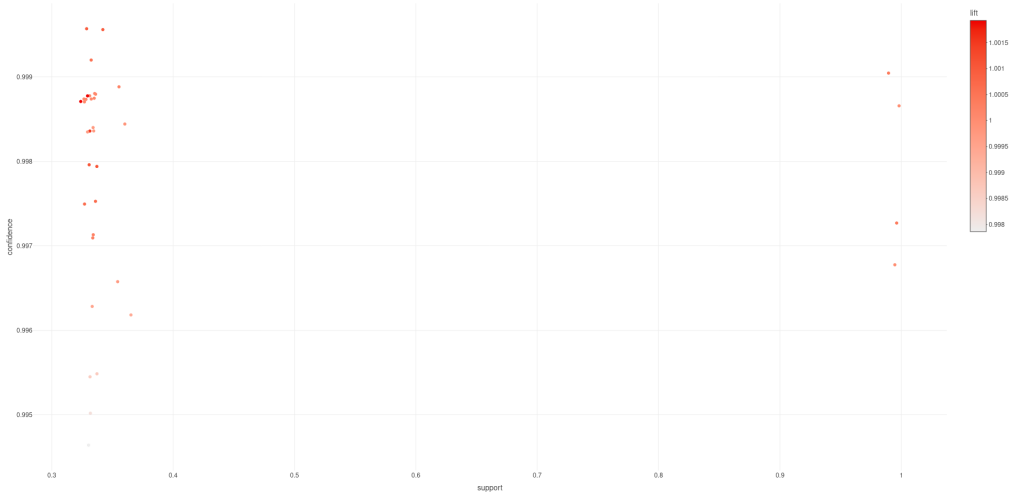


Fonte: Autores do documento

- **Análise Implícita**

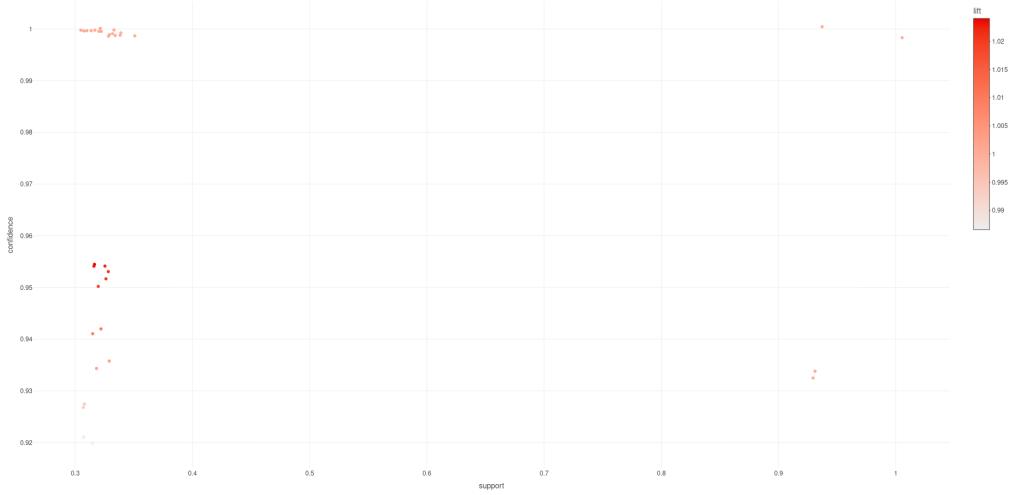
  Como forma de descobrir alguma correlação entre os valores de poluição e os valores de venda e aluguel praticados na cidade de São Paulo, os dados foram unidos conforme a estação de medição mais próxima e foi estabelecido um modelo de regras de decisão para identificar a faixa de preço mais provavel correlacionada a fatores como tamanho do imóvel e níveis de poluentes nas proximidades. A demonstração dos valores de confiança e suporte pode ser encontrada nas Figuras 16 e 17.

Figura 16 – Análise implícita - valores de confiança e suporte para regras envolvendo aluguéis



Fonte: Autores do documento

Figura 17 – Análise implícita - valores de confiança e suporte para regras envolvendo vendas



Fonte: Autores do documento

Outra forma de visualizar o modelo está nos arquivos `output/salesRulesGraph.html` e `output/rentRulesGraph.html`, que apresentam uma visão interativa das regras descobertas. Uma análise mais aprofundada delas é apresentada na sessão *Evaluation* (2.4.1.5).

2.4.1.5 *Evaluation*

As figuras 18 e 19 demonstram as regras encontradas para cada tipo de negociação. Podemos notar que em ambos se estabelece uma correlação entre a faixa de preço do imóvel e o niveis dos poluentes NO_2, CO e particulados. Apesar da notaçõa científica usada para imprimir os valores monetários, pode-se distinguir duas faixas de preço entre R$42,000.00 e R$290,000.00,

e entre R$550,000.00 e R$1,000,000.00 mais propensas a estarem asssociadas a zonas poluídas com dióxido de nitrogênio e material particulado inalável. Outras associações encontradas a estes poluentes envolvem condomínios baixos de R$0.00 a R$560.00 e tamanhos relativamente grandes, entre 74 e 620 metros quadrados.

Figura 18 – Inspeção das regras encontradas para vendas

```
implicita.R  51:41  83%
>(inspect(rules.sales))
     lhs                                rhs                                support confidence  coverage      lift count
[1]  {}                              => {no2=[51,82]}                     0.9327823  0.9327823 1.0000000 1.0000000  5981
[2]  {}                              => {particulado10=[39,63]}           0.9990643  0.9990643 1.0000000 1.0000000  6406
[3]  {Size=[30,52]}                  => {particulado10=[39,63]}           0.3137867  1.0000000 0.3137867 1.0009366  2012
[4]  {Condo=[0,270)}                 => {no2=[51,82]}                     0.3048971  0.9269796 0.3289145 0.9937792  1955
[5]  {Condo=[0,270)}                 => {particulado10=[39,63]}           0.3286026  0.9990517 0.3289145 0.9999874  2107
[6]  {Price=[4.2e+04,2.9e+05)}       => {particulado10=[39,63]}           0.3301622  0.9995279 0.3303182 1.0004640  2117
[7]  {Price=[5.5e+05,1e+07]}         => {no2=[51,82]}                     0.3192452  0.9543124 0.3345290 1.0230816  2047
[8]  {Price=[5.5e+05,1e+07]}         => {particulado10=[39,63]}           0.3342171  0.9990676 0.3345290 1.0000033  2143
[9]  {Condo=[270,560)}               => {no2=[51,82]}                     0.3078603  0.9202797 0.3345290 0.9865965  1974
[10] {Condo=[270,560)}               => {particulado10=[39,63]}           0.3342171  0.9990676 0.3345290 1.0000033  2143
[11] {Price=[2.9e+05,5.5e+05)}       => {no2=[51,82]}                     0.3155022  0.9413681 0.3351528 1.0092045  2023
[12] {Price=[2.9e+05,5.5e+05)}       => {particulado10=[39,63]}           0.3346850  0.9986040 0.3351528 0.9995393  2146
[13] {Condo=[560,8.92e+03]}          => {no2=[51,82]}                     0.3200250  0.9508804 0.3365565 1.0194023  2052
[14] {Condo=[560,8.92e+03]}          => {particulado10=[39,63]}           0.3362445  0.9990732 0.3365565 1.0000090  2156
[15] {Size=[74,620]}                 => {no2=[51,82]}                     0.3239239  0.9523155 0.3401435 1.0209408  2077
[16] {Size=[74,620]}                 => {particulado10=[39,63]}           0.3396756  0.9986245 0.3401435 0.9995598  2178
[17] {Size=[52,74)}                  => {no2=[51,82]}                     0.3233001  0.9342046 0.3460699 1.0015248  2073
[18] {Size=[52,74)}                  => {particulado10=[39,63]}           0.3456020  0.9986480 0.3460699 0.9995834  2216
[19] {no2=[51,82]}                   => {particulado10=[39,63]}           0.9327823  1.0000000 0.9327823 1.0009366  5981
[20] {particulado10=[39,63]}         => {no2=[51,82]}                     0.9327823  0.9336559 0.9990643 1.0009366  5981
[21] {Condo=[0,270),
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3048971  1.0000000 0.3048971 1.0009366  1955
[22] {Condo=[0,270),
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3048971  0.9278595 0.3286026 0.9947225  1955
[23] {Price=[5.5e+05,1e+07],
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3192452  1.0000000 0.3192452 1.0009366  2047
[24] {Price=[5.5e+05,1e+07],
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3192452  0.9552030 0.3342171 1.0240364  2047
[25] {Condo=[270,560),
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3078603  1.0000000 0.3078603 1.0009366  1974
[26] {Condo=[270,560),
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3078603  0.9211386 0.3342171 0.9875172  1974
[27] {Price=[2.9e+05,5.5e+05),
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3155022  1.0000000 0.3155022 1.0009366  2023
[28] {Price=[2.9e+05,5.5e+05),
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3155022  0.9426841 0.3346850 1.0106153  2023
[29] {Condo=[560,8.92e+03],
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3200250  1.0000000 0.3200250 1.0009366  2052
[30] {Condo=[560,8.92e+03],
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3200250  0.9517625 0.3362445 1.0203480  2052
[31] {Size=[74,620],
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3239239  1.0000000 0.3239239 1.0009366  2077
[32] {Size=[74,620],
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3239239  0.9536272 0.3396756 1.0223470  2077
[33] {Size=[52,74),
      no2=[51,82]}                   => {particulado10=[39,63]}           0.3233001  1.0000000 0.3233001 1.0009366  2073
[34] {Size=[52,74),
      particulado10=[39,63]}         => {no2=[51,82]}                     0.3233001  0.9354693 0.3456020 1.0028807  2073
NULL
>
*R:trabalho*  31:0  All
```

Fonte: Autores do documento

Já para alguéis, as faixas de preço encontradas são de R$1,500.00 a R$2.699.00 e R$2,700.00 a R$5,000.00. Os tamanhos são menores que os apartamentos a venda, variando entre 30 e 82 metros quadrados, com um subgrupo menos relevante de excedentes, que variam de 82 a 800

metros quadrados. Sendo assim, obteve-se uma regra segundo a qual pode-se associar um perfil de imóveis a áreas poluídas.

Figura 19 – Inspeção das regras encontradas para aluguel

```
> (inspect(rules.aluguel))
     lhs                                rhs                                support confidence  coverage      lift count
[1]  {}                              => {no2=[12,51]}                      0.9968179  0.9968179 1.0000000 1.0000000  7205
[2]  {}                              => {particulado10=[39,69]}            0.9986165  0.9986165 1.0000000 1.0000000  7218
[3]  {Size=[57,82]}                  => {no2=[12,51]}                      0.3285833  0.9987384 0.3289983 1.0019266  2375
[4]  {Size=[57,82]}                  => {particulado10=[39,69]}            0.3285833  0.9987384 0.3289983 1.0001221  2375
[5]  {Condo=[0,453]}                 => {no2=[12,51]}                      0.3316270  0.9979184 0.3323188 1.0011040  2397
[6]  {Condo=[0,453]}                 => {particulado10=[39,69]}            0.3319037  0.9987510 0.3323188 1.0001347  2399
[7]  {Size=[30,57]}                  => {no2=[12,51]}                      0.3314887  0.9970870 0.3324571 1.0002699  2396
[8]  {Size=[30,57]}                  => {particulado10=[39,69]}            0.3319037  0.9983354 0.3324571 0.9997185  2399
[9]  {Condo=[796,9.5e+03]}           => {no2=[12,51]}                      0.3317654  0.9950207 0.3334256 0.9981971  2398
[10] {Condo=[796,9.5e+03]}           => {particulado10=[39,69]}            0.3328722  0.9983402 0.3334256 0.9997234  2406
[11] {Condo=[453,796]}               => {no2=[12,51]}                      0.3334256  0.9975166 0.3342557 1.0007009  2410
[12] {Condo=[453,796]}               => {particulado10=[39,69]}            0.3338406  0.9987583 0.3342557 1.0001420  2413
[13] {Size=[82,880]}                 => {no2=[12,51]}                      0.3367460  0.9946874 0.3385445 0.9978626  2434
[14] {Size=[82,880]}                 => {particulado10=[39,69]}            0.3381295  0.9987740 0.3385445 1.0001577  2444
[15] {Price=[2.7e+03,5e+04]}         => {no2=[12,51]}                      0.3378528  0.9963280 0.3390980 0.9995085  2442
[16] {Price=[2.7e+03,5e+04]}         => {particulado10=[39,69]}            0.3385445  0.9983680 0.3390980 0.9997512  2447
[17] {Price=[1.5e+03,2.7e+03]}       => {no2=[12,51]}                      0.3613724  0.9961861 0.3627559 0.9993662  2612
[18] {Price=[1.5e+03,2.7e+03]}       => {particulado10=[39,69]}            0.3622025  0.9984744 0.3627559 0.9998578  2618
[19] {no2=[12,51]}                   => {particulado10=[39,69]}            0.9958495  0.9990285 0.9968179 1.0004125  7198
[20] {particulado10=[39,69]}         => {no2=[12,51]}                      0.9958495  0.9972291 0.9986165 1.0004125  7198
[21] {Size=[57,82],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3281682  0.9987368 0.3285833 1.0001205  2372
[22] {Size=[57,82],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3281682  0.9987368 0.3285833 1.0019250  2372
[23] {Condo=[0,453],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3312120  0.9987484 0.3316270 1.0001321  2394
[24] {Condo=[0,453],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3312120  0.9979158 0.3319037 1.0011014  2394
[25] {Size=[30,57],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3310736  0.9987479 0.3314887 1.0001316  2393
[26] {Size=[30,57],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3310736  0.9974990 0.3319037 1.0006832  2393
[27] {Condo=[796,9.5e+03],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3313503  0.9987490 0.3317654 1.0001326  2395
[28] {Condo=[796,9.5e+03],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3313503  0.9954281 0.3328722 0.9986057  2395
[29] {Condo=[453,796],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3332872  0.9995851 0.3334256 1.0009699  2409
[30] {Condo=[453,796],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3332872  0.9983423 0.3338406 1.0015292  2409
[31] {Size=[82,880],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3366076  0.9995892 0.3367460 1.0009740  2433
[32] {Size=[82,880],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3366076  0.9954992 0.3381295 0.9986770  2433
[33] {Price=[2.7e+03,5e+04],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3375761  0.9991810 0.3378528 1.0005653  2440
[34] {Price=[2.7e+03,5e+04],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3375761  0.9971394 0.3385445 1.0003225  2440
[35] {Price=[1.5e+03,2.7e+03],
      no2=[12,51]}                   => {particulado10=[39,69]}            0.3609574  0.9988515 0.3613724 1.0002353  2609
[36] {Price=[1.5e+03,2.7e+03],
      particulado10=[39,69]}         => {no2=[12,51]}                      0.3609574  0.9965623 0.3622025 0.9997435  2609
NULL
>

 *R:trabalho*  12:105 All
```

Fonte: Autores do documento

2.4.1.6 *Deployment*

Com o estudo realizado, pudemos observar a correlação entre preços de imóveis com o nível de poluição em São Paulo. Uma vez obtidos os resultados dos modelos criados, a ideia seria fazer o deploy dele em um ambiente produtivo para a tomada de decisões. Portanto, por

se tratar de um estudo acadêmico, foi desenvolvido *shinydashboard* utilizando a linguagem de programação R, de modo a obter uma melhor visualização das informações obtidas.

2.5 Descarte

Por fim, na etapa de descarte dos dados, é necessária a avaliação em relação à vida útil do dado. A partir de determinado momento os dados obtidos nas bases utilizadas para o estudo não serão mais viáveis para um modelo de análise com o propósito de tomada de decisções. Quando isso ocorrer, o descarte, seguindo as normativas da Lei Geral de Proteção de Dados (LGPD) terá de ser realizado, de modo que os dados sejam excluídos de forma segura.

3 Conclusões

Quais são as conclusões do trabalho? Que análise você faz dos resultados obtidos? O que o seu trabalho representa para área de estudo? Sim, você fará spoiler de tudo isso no resumo.

Referências

HOTZ, N. **What is CRISP DM?** 2022. Disponível em: <https://www.datascience-pm.com/crisp-dm-2/>.

KAGGLE. **Datasets**. 2022. Disponível em: <https://www.kaggle.com/datasets>.

RODRIGUES, G. **Valores missing - Parte 2**. 2020. Disponível em: <https://medium.com/psicodata/valores-missing-parte-2-d2e0b832ce14>.

SIRQUEIRA, T. **NoSQL e a Importância da Engenharia de Software e da Engenharia de Dados para o Big Data**. 2018. Disponível em: <https://www.researchgate.net/figure/Figura-22-Ciclo-de-vida-do-dado_fig1_327035187>.