

CICLO DE VIDA DOS DADOS

ALUNOS

Alexandre Yudi
SP3046923

Alkindar José
SP3029956

André Correia
SP305201X

Cecília Braz
SP3049876

Marcelo Carlos
SP3046583



OBJETIVO

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 tendência 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.



BASES UTILIZADAS

BASE DE POLUIÇÃO NO AR EM 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.

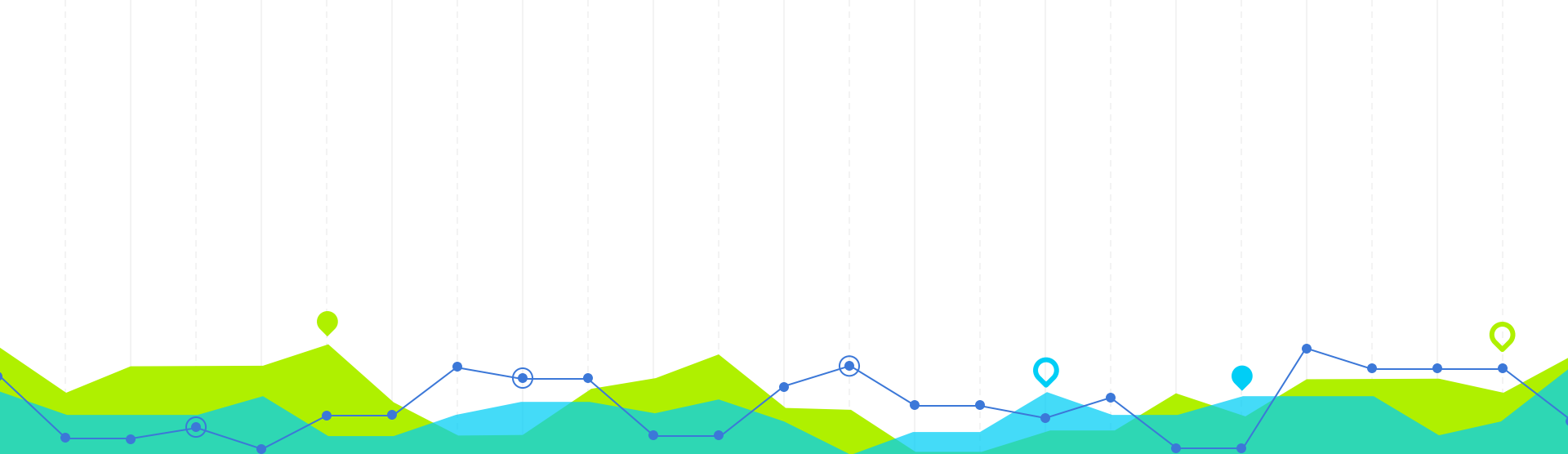


BASES UTILIZADAS

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.





Análise Explícita

1

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 todos os registros do mesmo dia em um único e, conseqüentemente, somar a emissão de um poluente alvo. Essa operação foi feita para cada poluente.

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

```
> 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
```

EXPLÍCITA

Já na base de aluguéis, primeiro foi feito um filtro para apenas consultar as moradias que estavam disponíveis para aluguel e não o que estavam disponíveis para venda. Após isso ainda foi realizado um **summary** que pode ser consultado no código.

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




Análise Exploratória 2

EXPLORATÓRIA

BASE DE POLUIÇÃO NO AR EM SÃO PAULO

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.



EXPLORATÓRIA

```
> 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

EXPLORATÓRIA

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

EXPLORATÓRIA

BASE DE VALORES IMOBILIÁRIOS

```
> 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		

EXPLORATÓRIA

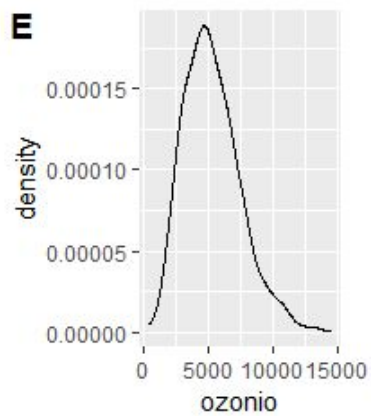
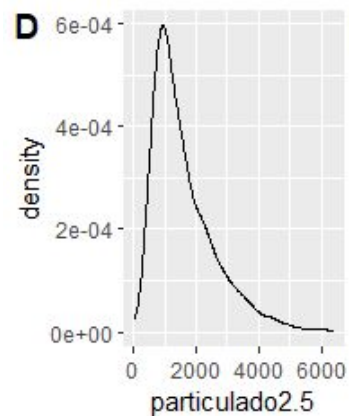
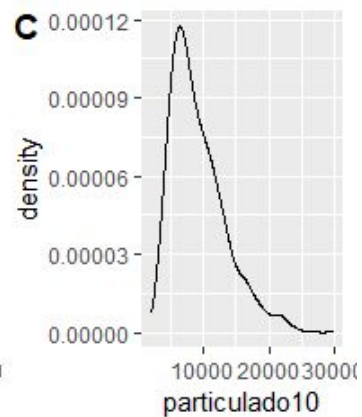
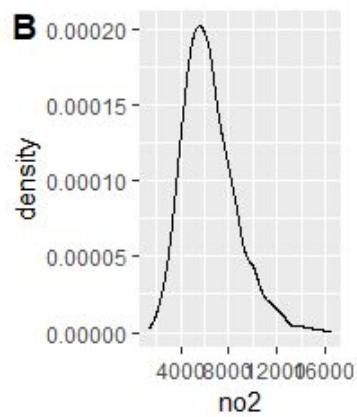
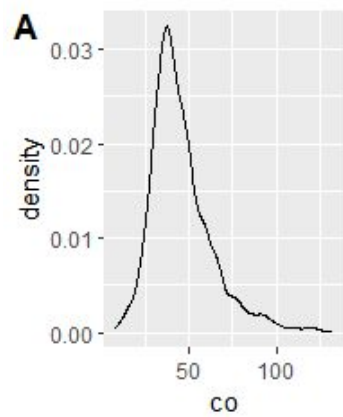
Suites	Parking	Elevator	Furnished
Min. :0.000	Min. :0.000	Mode :logical	Mode :logical
1st Qu.:1.000	1st Qu.:1.000	FALSE:5061	FALSE:5978
Median :1.000	Median :1.000	TRUE :2167	TRUE :1250
Mean :1.024	Mean :1.452		
3rd Qu.:1.000	3rd Qu.:2.000		
Max. :5.000	Max. :9.000		
Latitude	Longitude		
Min. :-46.75	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		

EXPLORATÓRIA

BASE DE POLUIÇÃO NO AR EM SÃO PAULO

Foi feita a soma dos dados de um poluente por dia ao longo dos anos de 2013 a 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.





EXPLORATÓRIA

BASE DE POLUIÇÃO NO AR EM SÃO PAULO

Para a base de aluguéis, o ggplot foi utilizado para construir um gráfico de pontos, o qual relaciona o preço, e se a moradia é mobiliada, com outro fator.



Tamanho X Preço X Mobiliado

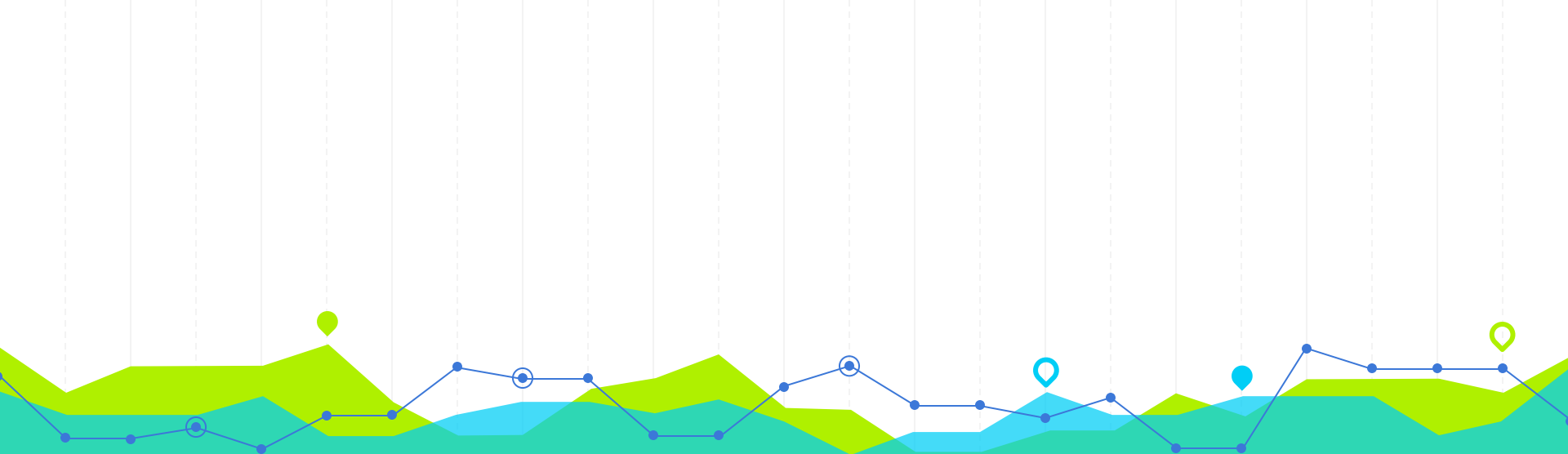


Quartos X Preço X Mobiliado



Suítes X Preço X Mobiliado





Análise Implícita

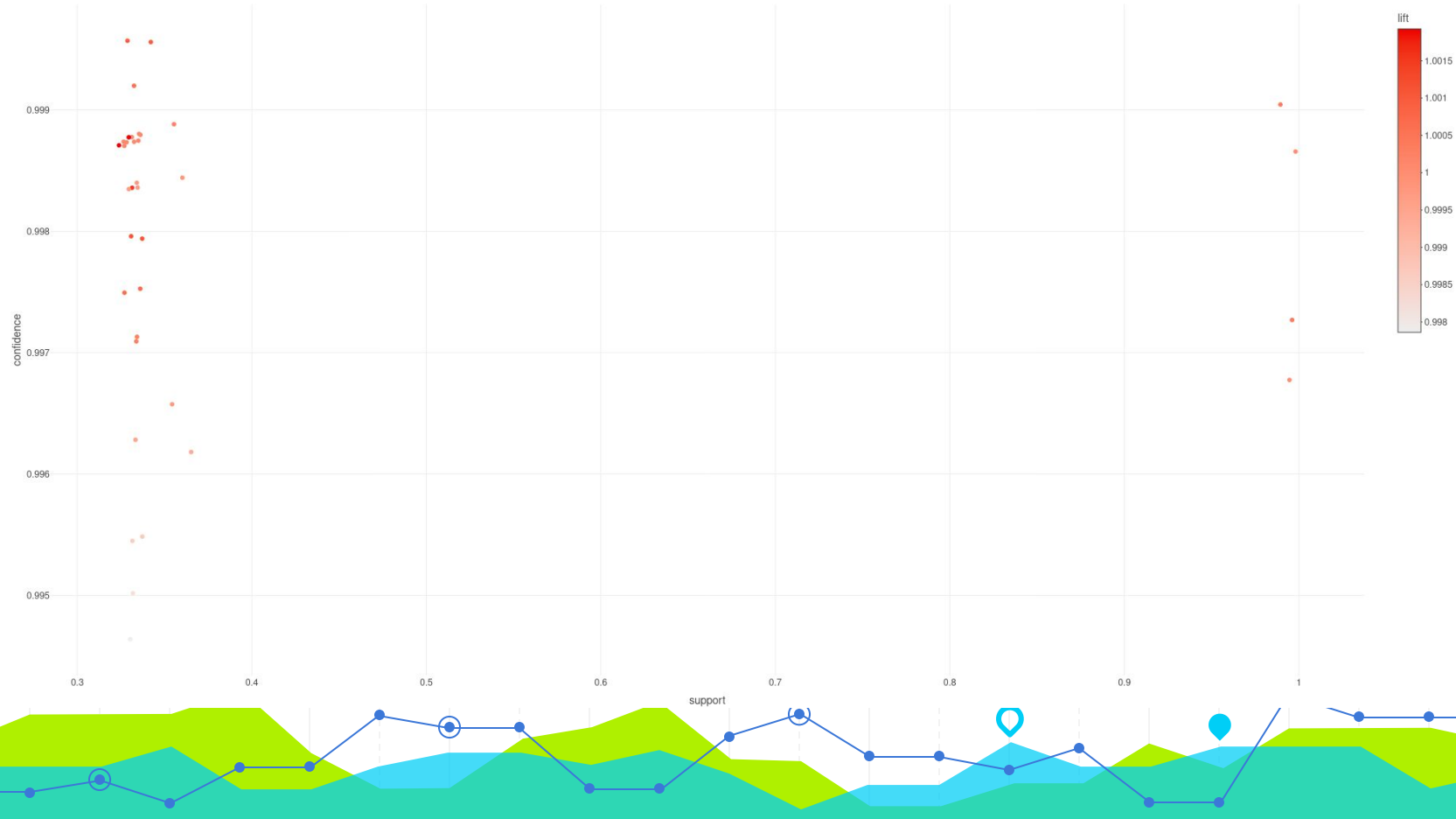
3

IMPLÍCITA

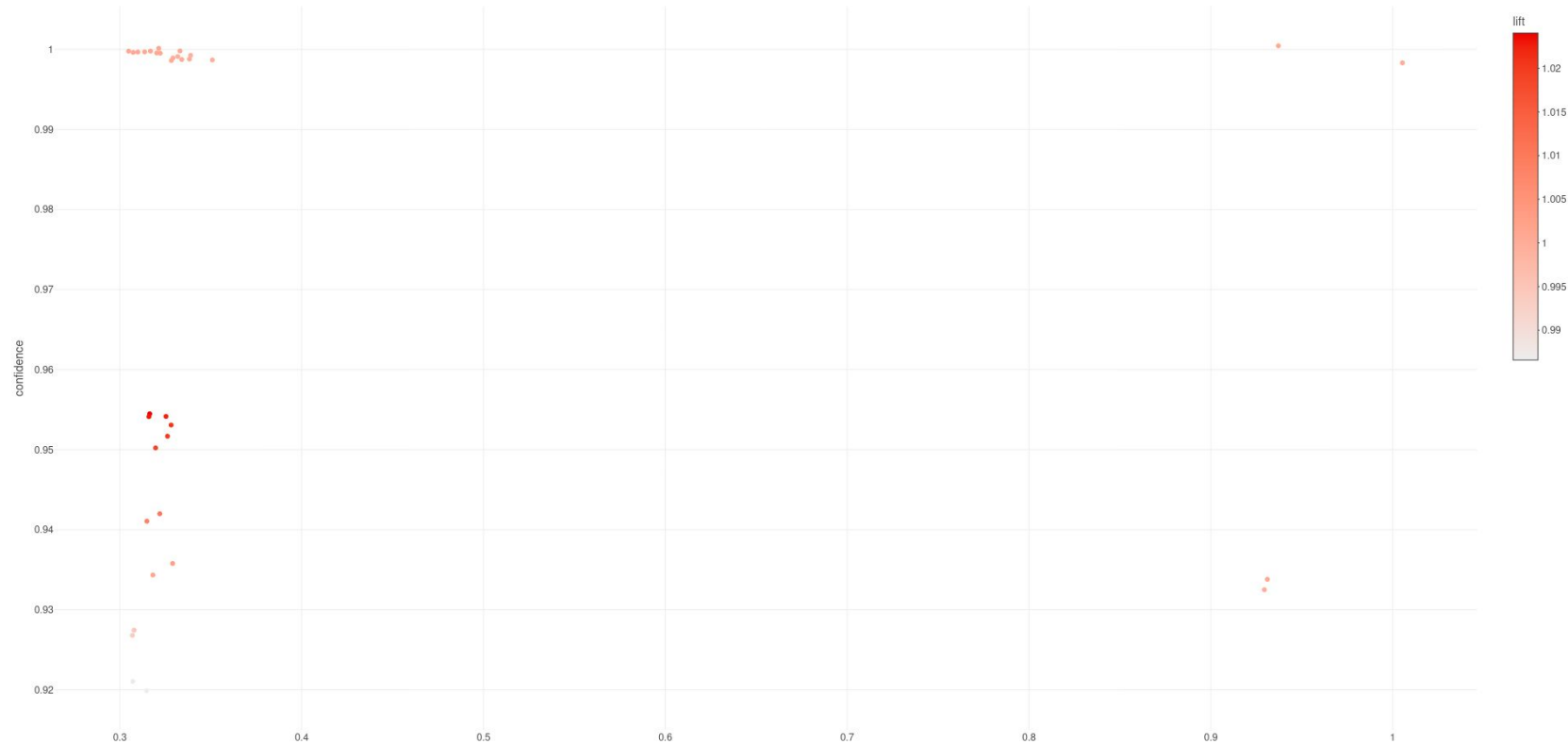
Como forma de descobrir alguma correlação entre os valores de poluição e os valores de venda e aluguel praticados na cidade de São Paulo, os dados foram unidos conforme a estação de medição mais próxima e foi estabelecido um modelo de regras de decisão para identificar a faixa de preço mais provável correlacionada a fatores como tamanho do imóvel e níveis de poluentes nas proximidades.



Valores de confiança e suporte para regras envolvendo alugueis



Valores de confiança e suporte para regras envolvendo vendas



IMPLÍCITA

As próximas figuras 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 níveis dos poluentes NO2, CO e particulados. Apesar da notação 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 associadas 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



Inspeção das regras encontradas para vendas

```
@implicit.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
```

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

THANKS!

