

1 Business Understanding

1.1 Identificando o Negócio

A Roof Imóveis é uma das maiores empresas do ramo imobiliário do Brasil, com mais de 50 anos de atuação.

A empresa é especializada na compra e venda de imóveis, focando na compra, reforma e posterior vendas destes imóveis. Mas também possui atuação nos segmentos de compra e venda de terrenos, locações de imóveis residenciais, comerciais e industriais, intermediações e avaliações.

Possui cerca de 300 colaboradores divididos nos setores de vendas, administração, finanças e recursos humanos. Também emprega de forma indireta mais de 50 colaboradores com os serviços de reforma.

1.2 Objetivos do Negócio

Com planos de expansão, a Roof quer investir em mercados internacionais. Após algumas avaliações e com base na expertise dos gestores da empresa, optaram por escolher o Condado de King County, no estado de Washington, nos Estados Unidos.

O Condado de King é um dos 39 condados do estado americano de Washington. A sede e cidade mais populosa do condado é Seattle. Foi fundado em 1852. Com mais de 2,2 milhões de habitantes, de acordo com o censo nacional de 2020, é o condado mais populoso do estado e o 12º mais populoso do país.

A empresa deseja obter um relatório detalhado mostrando 5 imóveis para investir e também 5 imóveis nos quais a empresa não deve investir. Também foi solicitado pela empresa que qualquer informação relevante ou insight obtido durante a elaboração desse estudo fosse adicionado ao relatório.

2 Data Understanding

Para a análise, será utilizada uma base de dados de vendas de imóveis obtidas entre maio/2014 e maio/2015 no Condado de King. A base de dados possui 21 variáveis que estão listadas a seguir:

Variável	Descrição
id	Identificador único do imóvel
date	Data da Venda
price	Preço de venda
bedrooms	Nº de Quartos
bathrooms	Nº de Banheiros
sqft_liv	Tamanho de área habitável em ft²
sqft_lot	Tamanho do terreno em ft²
floors	Número de andares
waterfront	Indicativo se o imóvel é a beira-mar

Variável	Descrição
view	Grau de quão belo é a vista do imóvel (0 a 4)
condition	Condição da casa (1 a 5)
grade	Classificação por qualidade de material utilizado na construção
sqft_above	Área em acima do solo em ft²
sqft_basmt	Área em abaixo do solo em ft²
yr_built	Ano de construção
yr_renov	Ano de restauração, caso o contrário, 0.
zipcode	Zip Code 5 - Similar ao CEP
lat	Latitude
long	Longitude
sqft_liv15	Média da área habitável dos 15 imóveis mais próximos, em ft²
sqft_lot15	Média da área do lote dos 15 imóveis mais próximos, em ft²

Após aberto o DataFrame, podemos ver uma breve descrição dos dados conforme a figura a seguir:

```

RangeIndex: 21613 entries, 0 to 21612
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     21613 non-null  int64
1   date                   21613 non-null  object
2   price                  21613 non-null  float64
3   bedrooms               21613 non-null  int64
4   bathrooms              21613 non-null  float64
5   sqft_living            21613 non-null  int64
6   sqft_lot               21613 non-null  int64
7   floors                 21613 non-null  float64
8   waterfront             21613 non-null  int64
9   view                   21613 non-null  int64
10  condition               21613 non-null  int64
11  grade                  21613 non-null  int64
12  sqft_above             21613 non-null  int64
13  sqft_basement          21613 non-null  int64
14  yr_built               21613 non-null  int64
15  yr_renovated           21613 non-null  int64
16  zipcode                21613 non-null  int64
17  lat                    21613 non-null  float64
18  long                   21613 non-null  float64
19  sqft_living15          21613 non-null  int64
20  sqft_lot15             21613 non-null  int64
dtypes: float64(5), int64(15), object(1)
memory usage: 3.5+ MB

```

Podemos observar que a base de dados possui 21613 imóveis registrados, divididos em 3 tipos de dados: 5 do tipo *float* (número decimal), 15 do tipo *int* (número inteiro) e 1 do tipo *object* (string).

O próximo passo foi observar se havia dados nulos ou faltantes no *dataframe*. Não havia dados nulos ou faltantes.

3 Data Preparation

O Primeiro ajuste a ser feito foi com a coluna 'date' que foi ajustada para que ficasse em um formato mais adequado, extraíndo o horário (pois todos estavam zerados) e separando a data em 3 novas colunas: dia da semana, mês e ano. Após isso, através da coluna 'zipcode', foi encontrada a cidade a qual cada imóvel pertence.

Logo após, foram calculados os preços médios x área para servir de parâmetro para futuras análises.

4 Modeling

Após as alterações feitas no *dataframe*, ele ficou com 28 colunas. O primeiro passo feito for gerar uma tabela com os valores das correlações das variáveis com o preço de cada imóvel, juntamente com um gráfico para verificar quais itens tem relação.

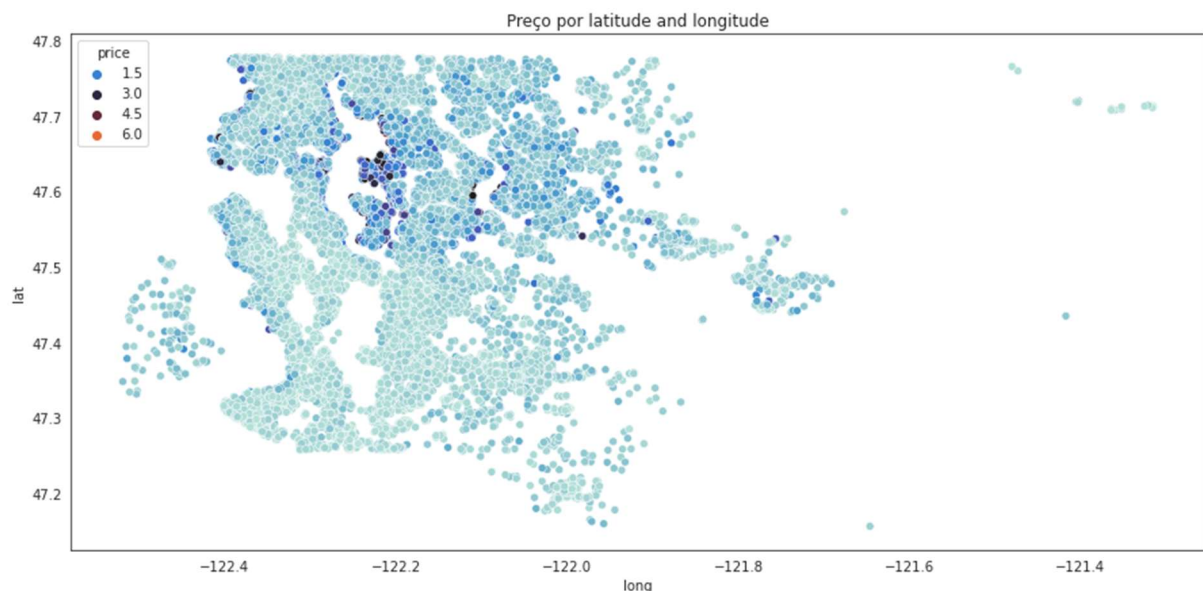
A partir disso, começamos o processo de procurar por *outliers*, que são dados que se diferenciam drasticamente de todos os outros. Em outras palavras, um *outlier* é um valor que foge da normalidade e que pode (e provavelmente irá) causar anomalias nos resultados obtidos por meio de algoritmos e sistemas de análise.

O primeiro encontrado foi referente ao número de quartos, que estavam variando de 0 a 33 quartos. Como casas com menos de 1 quarto não fazem sentido, todas as casas com menos de 1 quarto foram retiradas da análise. Da mesma forma, as casas com mais de 6 quartos somavam apenas 62 unidades, uma parte muito pequena do total, e levando em conta que o mercado para esse tipo de residência ser muito restrito, foram retiradas também todas com mais de 6 quartos. Após isso, foi feita uma análise parecida com a quantidade de banheiros, porém aqui cabe uma observação: no *dataframe*, os banheiros vieram com valores flutuantes, ou seja, com valores depois da vírgula (divididos em .25, .50 e .75). Isso se deve ao fato de uma métrica usada para classificar os banheiros nos Estados Unidos, onde cada fração de .25 corresponde a um aparelho: vaso sanitário, pia, chuveiro e banheira. E um banheiro completo, ou master bathroom, só é considerado como 1 unidade inteira se tiver esses 4 itens. Logo, uma residência que possui 1.5 banheiros possui 1 banheiro com os 4 itens (master) e outro com apenas 2 itens (nesse caso provavelmente um vaso e uma pia, ou lavabo).

Com isso, como uma casa com apenas 1 aparelho (.25) ou 1 lavabo (.50) também não faz muito sentido, todas que possuem esse valor foram retiradas. Foram retiradas todas com

mais de 6 banheiros, pelo mesmo fator citado quanto ao dos quartos, ou seja, a dificuldade de comercialização destes imóveis.

Após isso, foi criado um gráfico para verificar os preços dos imóveis com base na sua localização (latitude e longitude), o que demonstrou que os imóveis mais valiosos se concentram a beira mar, como podemos ver abaixo:



Como ainda não encontramos qualquer sugestão com base nas análises acima, prosseguimos com os estudos. O passo seguinte foi encontrar imóveis que tivessem sido vendidos mais de uma vez no período, para verificar se eles tiveram lucro ou prejuízo nesse período, o que dará uma boa base de valorização ou desvalorização dos imóveis. Para isso, foram filtrados os *id's*, ou seja, os números únicos de registros dos imóveis, que aparecem mais de uma vez no *dataframe*. Com isso, obtivemos 175 imóveis com essas características.

Após isso, deu-se início ao processo de cálculo desses lucros ou prejuízos, que consiste em subtrair a segunda venda (a mais recente) da primeira venda (mais antiga) para obtermos a diferença entre os valores e suas respectivas porcentagens. O resultado foram variações de -5,41% (negativa) a 321,80% (positiva). Com isso, conseguimos ter parâmetros das características dos imóveis que mais valorizaram e desvalorizaram nesse período de um ano, como também as suas localizações.

5 Deployment

Com bases nesses dados daremos inícios as nossas recomendações de imóveis para investir e não investir seguindo esse critério: procurando residências com características muito próximas as que tiveram maior valorização, como número de banheiros, quartos, andares,

tamanho internos dos imóveis e de seus lotes. Tendo como base também a localização próxima dos imóveis de referência.

5.1 Imóveis para investir

5.1.1. Primeiro imóvel

O imóvel que teve a maior valorização foi o de *id* 2023049218, com valor inicial de \$ 105.500,00 e uma valorização de 321,80%, com lucro de \$339.500,00 dólares em apenas 4 meses. Ele possui 2 quartos, 1 banheiro, 1 andar, não possui porão, 930 pés quadrados de área interna, 7.740 pés quadrados de terreno, não possui vista para o mar ou lago, 1 na escala de condição e 5 na escala de qualidade do material de construção, foi construída no ano de 1932 e não foi reformada.

Com base nessas informações, foi possível identificar um imóvel muito parecido com esse, com 2 quartos, 1 banheiro, 1 andar, não possui porão, 930 pés quadrados de área interna, 7.623 pés quadrados de terreno, não possui vista para o mar ou lago, 2 na escala de condição e 6 na escala de qualidade do material de construção, foi construída no ano de 1942 e também não foi reformada. O preço dela é de \$ 100.000,00 e fica localizada a apenas 1,9km do imóvel de referência. Seu *id* é o 1900000195.

5.1.2. Segundo imóvel

Seguindo a mesma orientação do imóvel mais lucrativo, o de *id* 2023049218, temos outro imóvel nas imediações com características parecidas.

O imóvel possui 2 quartos, 1 banheiro, 1 andar, 950 pés quadrados de área interna, 8.100 pés quadrados de terreno, não possui vista para o mar ou lago, 3 na escala de condição e 6 na escala de qualidade do material de construção, foi construída no ano de 1955 e também não foi reformada. O preço dela é de \$ 200.000,00, quase o dobro do imóvel de referência, mas isso se deve ao fato de ter uma condição melhor, uma posição acima na escala de qualidade e uma área interna um pouco maior e um lote maior. Seu *id* é o 2968801105 e se localiza a apenas 2km do imóvel de referência.

5.1.3. Terceiro imóvel

A terceira indicação parte do 4º imóvel mais rentável no período, de *id* 2422049104, com valor inicial de \$ 85.000,00 valor de venda de \$ 235000, ou seja, uma valorização de 176,47% no período de 3,5 meses. Ele possui características parecidas com o imóvel de *id* 2023049218, o que mostra uma possível valorização de imóveis com essas características.

Ele possui 2 quartos, 1 banheiro, 1 andar, 830 pés quadrados de área interna e 9.000 pés quadrados de terreno, não possui vista para o mar ou lago, 3 na escala de condição e 6 na escala de qualidade do material de construção e também não foi reformado.

O imóvel indicado é o de *id* 1863900190. Ele possui 2 quartos, 1 banheiro, 1 andar, 840 pés quadrados de área interna e 7.200 pés quadrados de terreno, não possui vista para o mar ou lago, 4 na escala de condição e 5 na escala de qualidade do material de construção e não foi reformado. Ele fica a apenas 660 metros de distância do imóvel de *id* 2422049104 e seu preço é de \$ 202.000,00

5.1.4. Quarto imóvel

A quarta indicação parte do imóvel de *id* 3883800011 como referência. Ele possui as seguintes características: 3 quartos, 1 banheiro, 1 andar, 860 pés quadrados de área interna, 10.426 pés quadrados de lote, não possui vista para o mar ou lago, 3 na escala de condição e 6 na escala de qualidade do material de construção, não possui porão, não foi reformada e foi construída no ano de 1954. Ela obteve uma valorização de 168.17%, passando de \$ 82.000,00 para \$ 219.900,00 em 5 meses.

O imóvel encontrado foi o de *id* 723049219. Ele possui 3 quartos, 1 banheiro, 1 andar, 880 pés quadrados de área interna, 10.800 pés quadrados de lote, não possui vista para o mar ou lago, 3 na escala de condição e 6 na escala de qualidade do material de construção, não possui porão, não foi reformada e foi construída no ano de 1942. Fica a apenas 480 metros de distância do imóvel de *id* 3883800011 por um preço de \$ 210.000,00.

5.1.5. Quinto imóvel

A quinta indicação tem como referência o imóvel de *id* 6021500970. Ele tem 2 quartos, 1 banheiro, 1.080 pés quadrados de área interna, 4.000 pés quadrados de área externa, 1 andar, não possui vista para o mar ou lago, 3 na escala de condição e 7 na escala de qualidade do material de construção, não possui porão, não foi reformada e foi construída no ano de 1940. Foi comprada pelo valor de \$ 3450.000,00 e vendida por \$ 874.950,00 cerca de um ano depois, com um lucro de aproximadamente 153,61%.

O imóvel encontrado foi o de *id* 6021501320. Ele tem 2 quartos, 1 banheiro, 1.030 pés quadrados de área interna, 4.365 pés quadrados de área externa, 1 andar, não possui vista para o mar ou lago, 3 na escala de condição e 7 na escala de qualidade do material de construção, não possui porão, não foi reformada e foi construída no ano de 1942. Possui o valor de \$ 450.000,00 e fica a apenas 300 metros do imóvel referência.

5.2. Imóveis para não investir

Todos os imóveis listados a seguir possuem o mesmo motivo para a indicação de não investimento: tiveram uma desvalorização em um pequeno período de tempo. Tiveram variação de -1,03% a -5,41%. O que indica uma tendência de desvalorização dos imóveis. São os de *id*: 4139420590, 2767603612 ,8682262400, 2726049071 e 7167000040.

6 Conclusões finais

Com isso, concluímos a análise com as 5 indicações de investimento e as 5 indicações para não investir. Inicialmente pensei que haveria forte relação de imóveis reformados em relação ao seu preço, assim como o de imóveis mais novos, mas isso não se mostrou válido ao analisar os dados. Uma conclusão adicional que podemos ter é quanto ao dia da semana e ao mês mais indicado para se comprar e vender um imóvel. O mês de abril é o melhor para se vender um imóvel, enquanto o mês de fevereiro é o melhor para se comprar, onde as informações se invertem no quesito venda. Quanto ao dia da semana, o melhor para se vender é o sábado e o pior a quinta-feira, onde as informações também se invertem para o quesito venda.

Ismael Paulo Refosco Pauchner

RID 32967

Link para o Colab com as análises:

<https://colab.research.google.com/drive/1JdxD81fohYT4n6lLI9m7Z3RZLJmU850M?usp=sharing>.