

Mini Projeto – Roof Imóveis

A Questão do Negócio

Com a intenção de entrar no mercado internacional de imóveis, mais especificamente no mercado americano, a Roof pretende investir em 5 imóveis e descartar outros 5.

O Entendimento do Negócio

King County é um dos 39 condados do estado de Washington, nos Estados Unidos, possui entre as principais cidades Seattle Bellevue, Mercer Island, Kirkland, Vashon, Sammamish, Federal Way, Auburn, Redmond, Kent, Kenmore e Enumclaw. Sua população atual é de aproximadamente 2,3 milhões de habitantes com uma renda per capita de US\$ 50.000 (*U.S. Census Bureau*).



(Fig 1)

Coleta dos Dados

Para chegar na escolha dos imóveis, foi disponibilizado um Dataset com 21613 imóveis contendo as seguintes variáveis:

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 ft ²
sqft_lot	Tamanho do terreno em ft ²
floors	Número de andares
waterfront	Indicativo se o imóvel é a beira-mar
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 cima do solo em ft ²
sqft_basmt	Área em abaixo do solo em ft ²
yr_built	Ano de construção
yr_renov	Ano de restauração, caso o contrário, 0.
zipcode	Zip Code 5
lat	Latitude
long	Longitude
sqft_liv15	Média da área habitável dos 15 imóveis mais próximos em ft ²
sqft_lot15	Média da área do lote dos 15 imóveis mais próximos em ft ²

Através do sistema Python e suas bibliotecas foi feito o tratamento dos dados e a seleção das variáveis para tomada de decisão sobre quais imóveis a Roof deveria investir : ID, Price, Bedrooms, Bathrooms, Sqft_ltr, View, Condition e ZipCode.

Limpeza de Dados

Iniciando com a procura de valores ausentes e duplicados, com o resultado negativo para ambos os casos, foi usado a informação Zipcode para definir as cidades em que os imóveis estão localizados, incluindo uma nova coluna chamada, city(cidade).

```
cities = []
for zipcode in df['zipcode']:
    try:
        zc = zcbd[zipcode]
        cities.append(zc.city)
    except KeyError:
        cities.append(float('nan'))
df['city'] = cities
df
```

(Fig 2)

Com a inclusão da coluna city(cidade) foi feita o tratamento do Dataframe, excluindo colunas que não seriam utilizadas na escolha dos imóveis. Como a empresa Roof não nos passou nenhuma característica específica do imóvel ou sua localização, o grau de importância foi definido por mim.

```
[52] #Escolher quais itens serão escolhidos para definir piores e melhores imóveis
#Id - Identificação
#Price
#Bedrooms
#Bathrooms
#sqft_lot
#condition
#view
#city

df = df.drop(['date', 'sqft_living', 'floors', 'waterfront', 'grade', 'sqft_above', 'sqft_basement', 'yr_built', 'zipcode', 'lat', 'long', 'sqft_living15', 'sqft_lot15', 'yr_renovated'], axis=1)
```

```
[53] df
```

	id	price	bedrooms	bathrooms	sqft_lot	view	condition	city
0	7129300520	221900.0	3	1.00	5650	0	3	Seattle
1	6414100192	538000.0	3	2.25	7242	0	3	Seattle
2	5631500400	180000.0	2	1.00	10000	0	3	Kenmore
3	2487200875	604000.0	4	3.00	5000	0	5	Seattle
4	1954400510	510000.0	3	2.00	8080	0	3	Sammamish
...
21608	263000018	360000.0	3	2.50	1131	0	3	Seattle
21609	6600060120	400000.0	4	2.50	5813	0	3	Seattle
21610	1523300141	402101.0	2	0.75	1350	0	3	Seattle
21611	291310100	400000.0	3	2.50	2388	0	3	Issaquah
21612	1523300157	325000.0	2	0.75	1076	0	3	Seattle

21613 rows x 8 columns

(Fig 3)

Conforme a figura acima, a coluna bathroom possui informações do tipo *float* que foi entendido de modo que o número cheio representa o banheiro completo enquanto o valor fracionado representa lavabo.

As variáveis, bedrooms, bathrooms, view e condition foram utilizados para fazer uma filtragem mais refinada, diminuindo consideravelmente a quantidade de imóveis a serem pesquisados.

```
[54] # Fazer um filtro por quarto, banheiro, vista e condição do imóveis
```

```
df = df.loc[((df['bedrooms']>=3))]
df = df.loc[((df['bathrooms']>=2.5))]
df = df.loc[((df['condition']>= 4))]
df = df.loc[((df['view']>= 3))]
df
```

	id	price	bedrooms	bathrooms	sqft_lot	view	condition	city
58	7922800400	951000.0	5	3.25	14342	4	4	Bellevue
312	1860600135	2384000.0	5	2.50	9050	4	5	Seattle
409	1560930070	840000.0	4	3.50	40139	4	4	Maple Valley
457	2122039094	705000.0	3	3.00	20978	3	4	Vashon
540	622049114	2125000.0	3	2.50	24069	4	4	Seattle
...
18828	1455100355	1675000.0	3	2.50	8343	4	4	Seattle
18912	2425049066	1920000.0	4	2.50	34412	3	4	Medina
19068	6613001241	1415000.0	4	3.00	4408	3	4	Seattle
19215	4131900042	2000000.0	5	4.25	10862	3	4	Mercer Island
19562	3271800910	1356920.0	4	3.50	5800	3	5	Seattle

202 rows x 8 columns

(Fig 4)

Após realizado todos os filtros e eliminação de informações não utilizadas, chegamos a um número de 202, e começamos a tratar os dados achando o preço médio dos imóveis por cidade e a quantidade de imóveis por cidade.

```
[55]
```

```
df2 = df.groupby('city')['price'].mean().round()
df2
```

city	
Auburn	413667.0
Bellevue	1405570.0
Enumclaw	450000.0
Federal Way	546000.0
Kenmore	805000.0
Kent	697500.0
Kirkland	1694050.0
Maple Valley	840000.0
Medina	2780450.0
Mercer Island	2357357.0
Redmond	1950000.0
Renton	589950.0
Sammamish	1690578.0
Seattle	1428630.0
Vashon	714000.0

Name: price, dtype: float64

(Fig 5)

```
[56] # Quantidade de imóveis em cada cidade
```

```
n_city = df['city'].value_counts()
n_city
```

Seattle	103
Bellevue	45
Mercer Island	15
Kirkland	10
Vashon	5
Sammamish	5
Federal Way	4
Auburn	3
Redmond	2
Kent	2
Kenmore	2
Enumclaw	2
Medina	2
Maple Valley	1
Renton	1

Name: city, dtype: int64

(Fig 6)

Com base nas informações retiradas de www.census.gov, a escolha de Seattle como a cidade para investimento deu-se pelo fato dela ser a maior do condado de King County, ser a mais populosa com 769.000 pessoas e ter uma renda per capita de U\$ 65.000, tendo portanto uma renda maior que a do condado com U\$ 50.000.

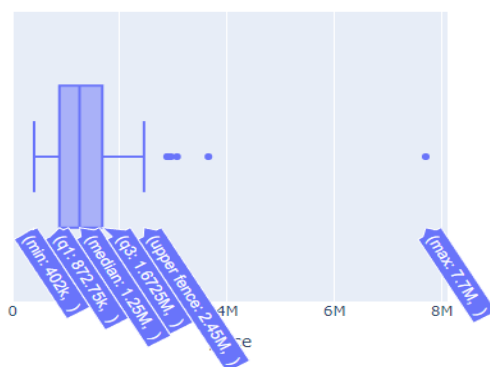
Definimos os 20 melhores imóveis a partir do preço, pelo critério de localizar o imóvel com preço abaixo do valor médio (conforme fig 5) na cidade de Seattle.

```
df_sorted0a = df_Seattle.sort_values('price')
df_top= df_sorted0a.head(20)
print(df_top)
```

	id	price	bedrooms	bathrooms	sqft_lot	view	condition	\
4429	2011400583	402000.0	3	2.50	9994	3	4	
4576	7203600750	421000.0	3	2.50	4505	3	4	
1271	7893804340	470000.0	4	2.50	8062	3	4	
1755	7784400060	545000.0	3	2.50	9000	3	4	
11110	8859000045	545000.0	4	2.75	8480	3	4	
8791	5067400032	550000.0	3	2.50	14400	3	5	
7412	7130300170	552000.0	4	2.75	8429	3	4	
18079	2769600035	612000.0	4	2.50	3626	3	4	
9101	7129301001	675000.0	4	2.75	6780	3	5	
18097	1455600062	689000.0	3	2.50	9612	3	4	
6023	6641800020	700000.0	4	2.50	9650	3	4	
10772	7774200070	725000.0	4	2.50	13950	3	4	
13348	9285800330	732000.0	3	3.75	6517	4	4	
11844	2623039082	770000.0	3	3.50	21744	4	4	
3791	4077800017	775000.0	4	2.75	6200	3	4	
1062	3524039204	790000.0	4	2.75	11900	3	4	
15727	6116500290	799950.0	6	2.75	36721	3	4	
5747	2771104830	800000.0	4	3.75	4000	3	4	
3464	4154305290	805000.0	5	3.00	6480	3	4	
10382	579002220	808000.0	3	2.50	6240	3	5	

(Fig 7)

Os gráficos abaixo representam a faixa de preço dos imóveis na cidade de Seattle, e a faixa de preço entre os 20 imóveis selecionados (fig7) pela ordem de preço.

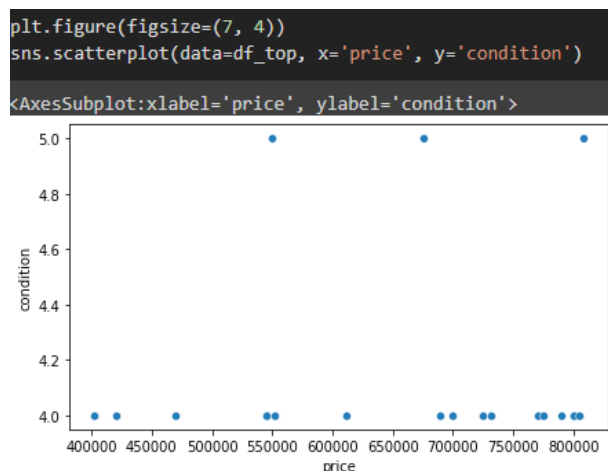


(Fig 8)

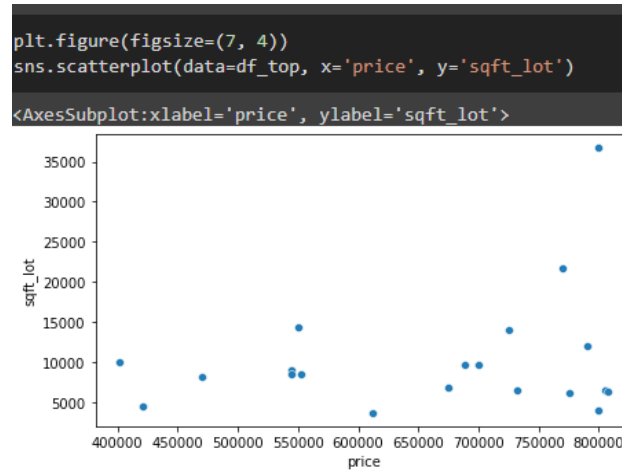


(Fig 9)

Nos próximos dois gráficos foram selecionados, condição X preço e a seguir foi selecionado a condição x preço, localizando 3 dos 20 imóveis com a pontuação máxima.



(Fig 10)



(Fig 11)

Análise Final mais Sugestão de Imóveis

Os parâmetros de escolha dos imóveis em relação aos gráficos postados foi analisado da seguinte forma:

1. Faixa de preço da média dos 20 imóveis mais baratos de Seattle (máximo 808K);
2. Condição do Imóvel, preferência 5;
3. O Tamanho do Lote;
4. Número de Quartos;
5. Número de Banheiro;
6. Vista.

As sugestões de Imóveis para compra pelo potencial de valorização e retorno financeiro.

ID	Preço(U\$)	Condição	Lote	Quarto/Banheiro	Vista
5067400032	550.000,00	5	14400	3 / 2.50	3
2623039082	770.000,00	4	21744	3 / 3.50	4
7774200070	725.000,00	4	13950	4 / 2.50	3
3524039204	790.000,00	4	11900	4 / 2.75	3
6116500290	799950,00	4	36721	6 / 2.75	3

A escolha dos piores imóveis para investimento está definido pelos piores resultados encontrados nas variáveis do Dataset.

ID	Preço(U\$)	Condição	Lote	Quarto/Banheiro	Vista	Cidade
8056000075	1.065.000,00	3	9466	2 / 1.75	0	Bellevue
1224049005	1.087.500,00	3	11340	2 / 2.00	0	Mercer Island
2954400190	1.295.650,00	3	28008	0 / 0	0	Redmond
1526059051	995.000,00	2	64468	2 / 2.00	0	Woodinville
8835770330	1.057.000,00	3	36721	2 / 1.50	0	North Bend

Rodolfo Antonio Raffaele – RID #83528

COLAB:

https://colab.research.google.com/drive/1oKitoI0KLmWj4TPX_ly9-Ur1-lvpWO48?usp=sharing