

Formação Em

DADOS DEX

Roof Imóveis



— A Questão de Negócio:

A Roof Imóveis busca expandir sua atuação no mercado imobiliário internacional e contratou uma consultoria estratégica para investir em imóveis no Condado de King, nos Estados Unidos. Com base no dataset disponibilizado, foram aplicados os conceitos de CRISP-DM para identificar os 5 imóveis ideais para investimento e os 5 que não seriam recomendados.



_ O Entendimento do Negócio:



— Observações:

Para obtermos uma análise mais precisa, é necessário definir qual é o objetivo da compra do imóvel: revenda, aluguel ou uso próprio. Além disso, é importante identificar o público-alvo, como casais sem filhos, casais com filhos, solteiros ou estudantes. Também é necessário definir o tipo de imóvel desejado, se é de alto padrão ou popular, e qual o valor máximo do investimento.

Em relação aos dados, seria mais simples se tivéssemos acesso ao valor do metro quadrado por cidade. É importante observar se os imóveis recomendados estão bem localizados, por exemplo, se há escolas próximas e se estão próximos ao centro da cidade, entre outros fatores.

— METRICAS

- **Localização:** o imóvel deve estar em uma das 10 cidades mais populosas de Washington.
- **Valor do metro quadrado:** deve ser menor que a média da cidade.
- **Condição do imóvel:** a condição do imóvel deve ser igual a 5.
- **Tamanho do terreno:** deve ser maior que a média da cidade.

— METRICAS

- **Quantidade de quartos:** deve ser maior que a média da cidade.
- **Quantidade de banheiros:** deve ser maior que a média da cidade.
- **Ano de construção:** deve ser maior ou igual a 1980.

A Coleta de Dados:

Coleta e avaliação dos dados



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 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 cima 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

— Variáveis escolhidas:

	id	city	price	sqft_lot	condition	yr_built	bedrooms	bathrooms
0	2144800146	Seattle	257500.0	9334	5	1981	3	2
1	3348401382	Seattle	318000.0	12662	5	1982	3	2
2	2826049260	Seattle	482500.0	7626	5	1990	4	3
3	4022902715	Seattle	525000.0	10277	5	1993	5	3
4	5067400032	Seattle	550000.0	14400	5	1985	3	2

A Limpeza de Dados:

Limpeza e padronização dos dados.



— Alterando para int a coluna bathrooms:

1- Ao iniciar a análise dos dados, deparei-me com a coluna "bathrooms" e percebi que havia informações referentes a casas com, por exemplo, 2.25, 4.50, 1.75 banheiros. Esses valores não faziam sentido para a minha análise. Então, optei por convertê-los em números inteiros. Lembrando que pesquisei e descobri que esses dados significam que o banheiro em questão não é completo, ou seja, possui todos os itens que um banheiro normalmente tem, exceto pelo chuveiro, a pia ou o vaso sanitário. Apesar disso, concluí que esses dados não eram relevantes para a minha análise.

bathrooms	
0	1.00
1	2.25
2	1.00
3	3.00
4	2.00
...	...
21608	2.50
21609	2.50
21610	0.75
21611	2.50
21612	0.75
21613 rows × 1 columns	

— Alterando para int a coluna floors:

2 - Posteriormente, encontrei uma situação semelhante na coluna "floors", onde havia informações referentes a casas com 1.5, 2.5, 3.5 andares. Como essa coluna indicava o número de andares da casa, optei por convertê-los em números inteiros também.

floors	
18477	3.5
875	3.5
20308	3.5
11594	3.5
20772	3.5
...	...
4600	1.0
12432	1.0
12434	1.0
12435	1.0
0	1.0
21613 rows × 1 columns	

IDs repetidos:

3 - Ao analisar o data frame, identifiquei 353 imóveis que haviam sido vendidos mais de uma vez. Uma das métricas sugeridas foi a valorização desses imóveis. No entanto, optei por ignorar essa métrica por dois motivos. Primeiramente, uma valorização passada não necessariamente indica uma valorização futura. Além disso, dos 21613 imóveis analisados, apenas 353 haviam sido vendidos mais de uma vez. Portanto, optei por remover os IDs duplicados.

id	
93	6021501535
94	6021501535
313	4139480200
314	4139480200
324	7520000520
...	...
20670	8564860270
20779	6300000226
20780	6300000226
21580	7853420110
21581	7853420110
353 rows × 1 columns	

A Exploração de Dados:

Avaliando as informações.

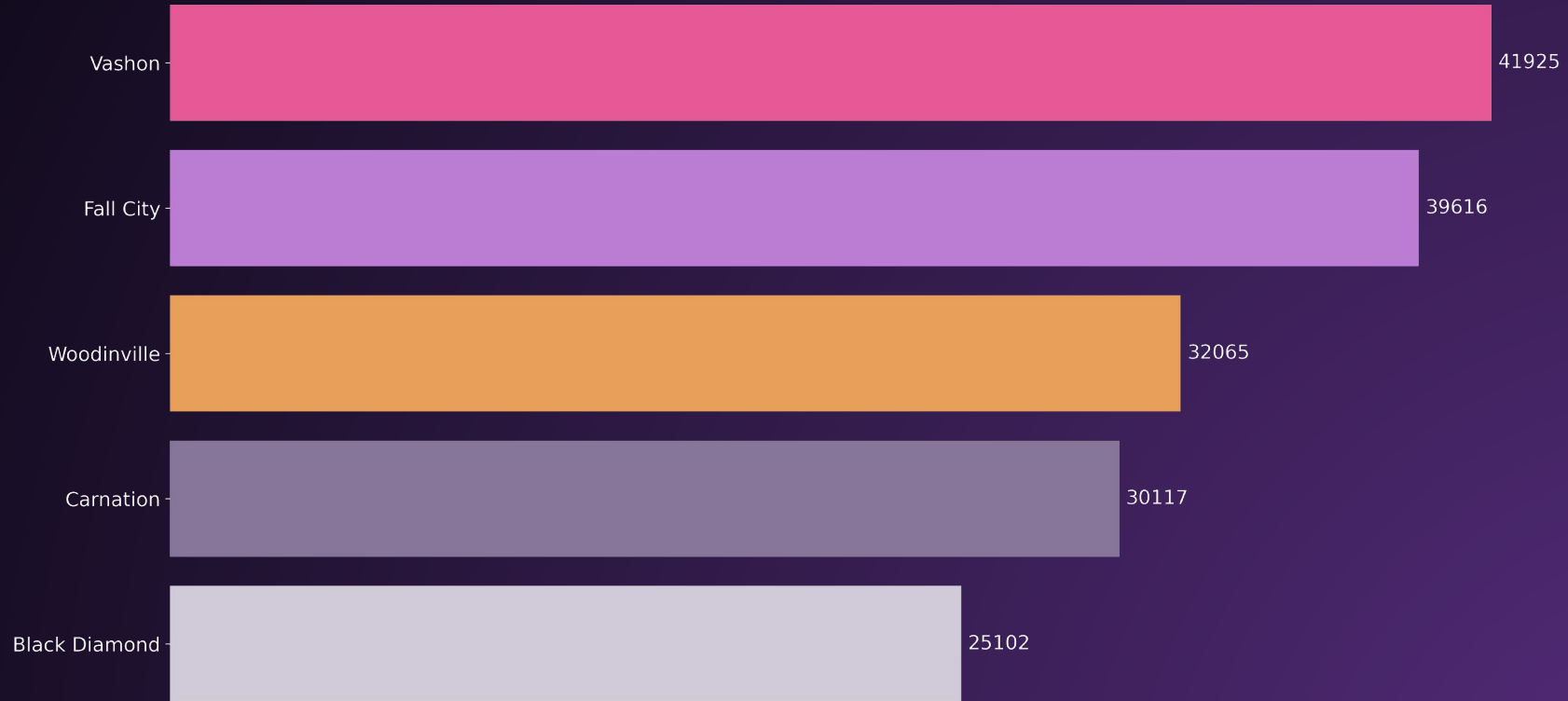




— As 5 cidades com a maior concentração de imóveis.



— As 5 cidades com o ft2 mais caro:



— As 5 cidades com os maiores terrenos.

— **Ranqueamento das lojas:**

Por fim, mas não menos importante, criei um algoritmo de ranqueamento de imóveis. Esse algoritmo consiste em atribuir notas aos imóveis filtrados previamente. Cada critério estipulado tem um peso maior ou menor dependendo do caso, sendo que a nota máxima que um imóvel pode receber é 100.

Antes de atribuirmos notas aos imóveis, precisamos filtrar os dados. Como temos mais de 21 mil linhas, utilizei os seguintes filtros: condição mínima igual a 5, ano de construção maior ou igual a 1980, número de quartos maior ou igual a 1, número de banheiros maior ou igual a 1 e preço máximo menor ou igual a 1 milhão. Após a filtragem, chegou o momento de ranquear os imóveis.

— Ranqueamento das lojas:

Os critérios utilizados para ranqueá-los foram:

- **A Localização**, se o imóvel está dentro das 10 cidades mais populosas, ele ganha 30 pontos;
- **O valor do ft²**, se o imóvel estiver com um valor do ft² abaixo da média da cidade, concluo que ele está subvalorizado e ganhará mais 30 pontos;
- **O tamanho do terreno**, se o imóvel for maior que a média da cidade, ganhará mais 20 pontos;
- **O número de quartos**, se o imóvel tiver mais quartos que a média da cidade, ganhará mais 10 pontos;
- **O número de banheiros**, se o imóvel tiver mais banheiros que a média da cidade, ganhará mais 10 pontos.

— Ranqueamento das lojas:

Critérios	Notas
Cidade	30
Valor do ft2	30
Tamanho do terreno	20
Quartos	10
Banheiros	10



— Sugestão de imóveis:

Os 5 recomendados e os 5 que NÃO recomendados.



— Os 5 recomendados:

1. **O imóvel de ID: 2144800146**, está localizado em Seattle, sendo vendido por \$257500, com valor do pé quadrado de \$27, tamanho do terreno de 9334, 3 quartos e 2 banheiros, com área habitável de 1300.
2. **O imóvel de ID: 3348401382**, também está localizado em Seattle, com preço de venda de \$318000, valor do pé quadrado de \$25, tamanho do terreno de 12662, 3 quartos e 2 banheiros, com área habitável de 1690.
3. **O imóvel de ID: 2826049260**, também em Seattle, com preço de venda de \$482500, valor do pé quadrado de \$63, tamanho do terreno de 7626, 4 quartos e 3 banheiros, com área habitável de 1630.

— Os 5 recomendados:

4. **O imóvel de ID: 4022902715**, também em Seattle, com preço de venda de \$525000, valor do pé quadrado de \$51, tamanho do terreno de 10277, 5 quartos e 3 banheiros, com área habitável de 2480.
5. **O imóvel de ID: 5067400032**, ainda em Seattle, com preço de venda de \$550000, valor do pé quadrado de \$38, tamanho do terreno de 14400, 3 quartos e 2 banheiros, com área habitável de 3070.

— Os 5 que NÃO recomendados:

1. **O imóvel de ID: 3523029059**, está localizado em Vashon, com preço de venda de 181000, valor do pé quadrado de \$16, tamanho do terreno de 10807, 2 quartos e 1 banheiro, com área habitável de 1560.
2. **O imóvel de ID: 913000340**, está localizado em Seattle, com preço de venda de \$252000, valor do pé quadrado de \$153, tamanho do terreno de 1638, 1 quarto e 1 banheiro, com área habitável de 680.
3. **O imóvel de ID: 9407110710**, está localizado em North Bend, com preço de venda de \$322000, valor do pé quadrado de \$38, tamanho do terreno de 8400, 3 quartos e 1 banheiro, com área habitável de 1510.

— Os 5 que NÃO recomendamos:

4. **O imóvel de ID: 2028700265**, está localizado em Seattle, com preço de venda de \$505000, valor do pé quadrado de \$132, tamanho do terreno de 3816, 2 quartos e 1 banheiro, com área habitável de 1310.
5. **O imóvel de ID: 4083802195**, está localizado em Seattle, com preço de venda de \$578888, valor do pé quadrado de \$144, tamanho do terreno de 4000, 2 quartos e 2 banheiros, com área habitável de 1060.

— Informações do Autor:

- **Nome:** Samuel Barbosa Evangelista
- **RID:** 86386
- **GitHub:** [Link](#)
- **LinkedIn:** [Link](#)



— TEAM PRESENTATION



Samuel Barbosa

Data Scientist

Não confie. Verifique!