

# **Relatório de Score de atratividade de investimentos em imóveis de King County (WA)**

Matheus Lamas Machado

---

**Link Colab Notebook:**

[https://colab.research.google.com/drive/1PcBT\\_RmGeji9xOnkuejgH27wD-5DoaJV?usp=sharing](https://colab.research.google.com/drive/1PcBT_RmGeji9xOnkuejgH27wD-5DoaJV?usp=sharing)

---

## **1. BUSINESS UNDERSTANDING**

### **1.1. Objetivos do Negócio**

A Roof Imóveis, uma das maiores empresas do ramo imobiliário Brasileiro visa expandir sua área de atuação iniciando um projeto de investimento no Condado de King (WA, EUA). O objetivo da empresa é farejar oportunidades de investimentos do ramo imobiliário residencial, estabelecendo critérios que atendam aos desejos do consumidor, e.g., preço, qualidade do imóvel, vista, vizinhança etc.

O relatório tem o objetivo de indicar 5 imóveis que ela poderá investir e quais 5 não são recomendados como investimento e sua justificativa, com uma análise das características particulares dos grupos de imóveis.

### **1.2. Plano do Projeto**

O plano vai seguir a estratégia de entender a natureza das variáveis que afetam a precificação do m<sup>2</sup> dos imóveis e de características subjetivas, como: vista, qualidade estrutural do imóvel etc, seguindo um modelo de estatística descritiva básica.

## **2. ENTENDIMENTO DO NEGÓCIO**

O Condado de King é um dos 39 condados do estado americano de Washington. Hospedando a cidade mais populosa do condado, Seattle. Com mais de 2,2 milhões de habitantes. O Dataset a ser utilizado para a análise apresenta a venda de imóveis e suas

características desde Maio de 2014 à Maio de 2015. Possuindo um total de 21613 registros com um total de 21 atributos.

Atributos da base de dados:

```
'id', 'date', 'price', 'bedrooms', 'bathrooms', 'sqft_living',  
'sqft_lot', 'floors', 'waterfront', 'view', 'condition', 'grade',  
'sqft_above', 'sqft_basement', 'yr_built', 'yr_renovated',  
'zipcode', 'lat', 'long', 'sqft_living15', 'sqft_lot15'
```

### **2.1. Características relevantes para a exploração**

Partindo do pressuposto que a empresa tenha interesse em investir em oportunidades de negócio sem levar em considerações aspectos subjetivos e pessoais dos compradores, foram escolhidas características que podem ser consideradas como universais, ou seja, que de maneira geral o comprador, independente do gosto dele, vai valorizar.

Dentre estas características foram definidas: preço, idade do imóvel, vista, condição do imóvel, qualidade estrutural da casa, ano de reforma, e se tem vista para corpos d'água.

Apesar de “vista” e “estar de frente para água” serem características de cunho pessoal, são critérios que possuem certa concordância entre o público alvo da empresa e foram levados em conta.

Alguns atributos serão trabalhados mais a fundo a fim de discriminar eventuais vieses que o dado possa conter.

## **3. A COLETA DE DADOS**

O dado apresenta-se coeso, sem valores nulos, sem colunas faltantes, pronto para ser processado, com exceção da data, os dados estão no formato *Integer ou Float*, e foram importados no formato DataFrame através da biblioteca Pandas.

Uma breve análise foi realizada para entender a amplitude dos atributos e como é a sua distribuição.

## 4. A LIMPEZA DOS DADOS

### 4.1. Remoção de dados inconsistentes

Após a análise de cada variável, observou-se que os atributos “*bathrooms*” e “*floors*” apresentam valores decimais, o que a princípio não faz sentido, pois não é possível ter meio quarto. Como essa codificação não foi descrita na documentação, optou-se pela retificação desses valores para o número inteiro abaixo do decimal, ou seja, e.g., 4.5 → 4.

Os valores discretizados para os atributos citados:

```
bathrooms → {0.75, 1.0, 2.0, 2.25, 3.0, 4.5, 1.5,
2.5, 1.75, 2.75, 3.25, 4.0, 3.5, 4.75, 5.0, 0.0,
3.75, 8.0, 4.25, 5.25, 5.5, 5.75, 1.25, 6.0, 0.5,
6.75, 6.25, 7.5, 7.75, 6.5}
floors → {1.0, 2.5, 2.0, 3.0, 1.5, 3.5}
```

## 5. EXPLORAÇÃO DE DADOS

### 5.1. Compreendendo as variáveis

A fim de compreender a natureza do atributo “*zipcode*” foi plotado suas coordenadas num gráfico *scatter* (Figura 1). Que permitiu definir que os zipcodes representam compartimentações do Condado, ou melhor, Bairros. Desta forma faz sentido procurar oportunidades de investimento comparando atributos em relação a sua vizinhança, levando em consideração o *zipcode*.

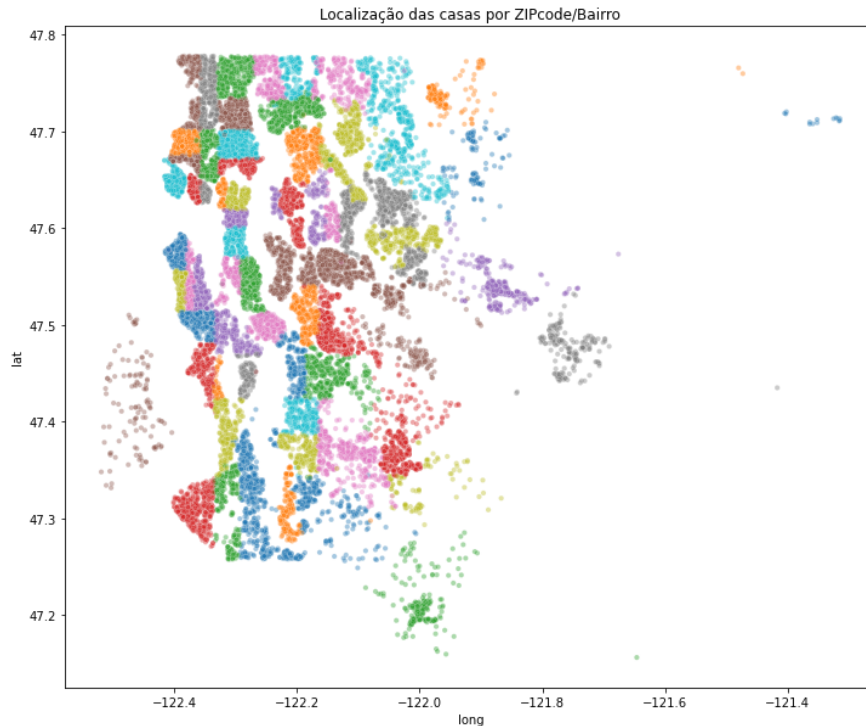


Figura 1. Mapa de localização dos imóveis discriminados por ZIPcode/Bairro.

## 5.2. Variáveis chave

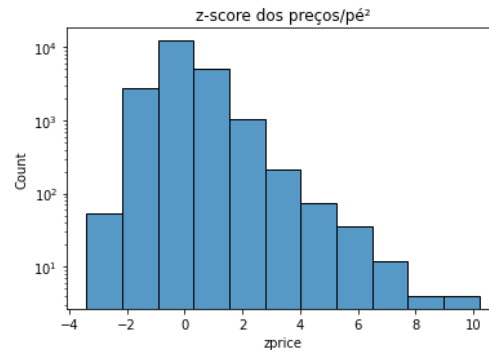
Partindo desta ideia, primeiro calculou-se o valor do pé<sup>2</sup> (*price\_sqft*), para atribuir valor proporcionalmente a área do imóvel (neste caso optou-se por considerar apenas a área útil, e não o “quintal”).

Em seguida, foi utilizado um tratamento utilizando a técnica de z-score. O z-score, ou pontuação padrão, é o número de desvios padrão pelos quais o valor de uma pontuação crua (ou seja, um valor observado ou ponto de dados) está acima ou abaixo do valor médio do que está sendo observado ou medido. Valores acima da média têm scores positivos, enquanto aqueles abaixo da média têm scores negativos. Pode ser exemplificado pela equação (1):

$$(1) \ z = \frac{x - \mu}{\sigma}$$

Onde  $x$  é o valor observado,  $\mu$  é a média da população,  $\sigma$  é o desvio padrão da população e  $z$  é o score daquele indivíduo.

Os z-scores dos valores por pé<sup>2</sup> foram calculados considerando a média e o desvio padrão por bairro e anexados como um novo atributo (**zprice**) a base de dados e sua distribuição está exemplificada na Figura 2.



*Figura 2. Z-Score dos preços/pé<sup>2</sup>.*

Um segundo score foi produzido em cima dos valores médios do m<sup>2</sup> dos imóveis da vizinhança. Partindo do pressuposto que um morador queira se sentir integrado a vizinhança, ele evitaria adquirir uma casa fora dos padrões socioeconômicos da região, portanto criou-se um score que compara o quão o imóvel em questão é diferente em tamanho das 15 casas vizinhas ('*sqft\_living15*').

Os valores foram elevados ao quadrado para retirar os valores negativos, ou seja, independente se o imóvel é maior ou menor que a redondeza, ele é diferente de qualquer forma; e para tirar o efeito linear da diferença dos imóveis, ou seja, se a diferença for pouca, ela se torna imperceptível, a diferença dos imóveis vai impactar no enquadramento dele na vizinhança caso a metragem seja muito discrepante. O resultado aparece na Figura 3 (Atributo **sqft15\_score**).

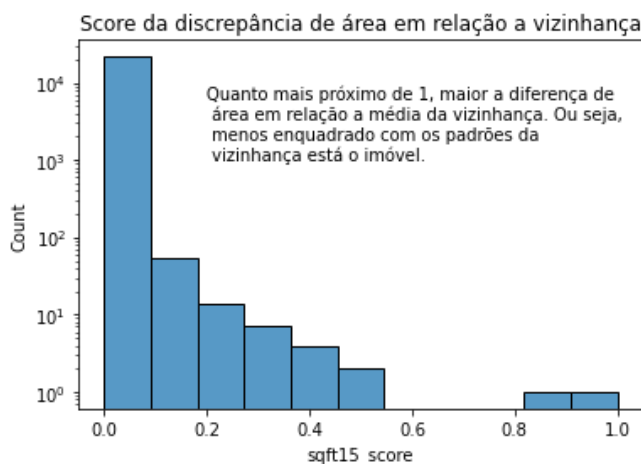


Figura 3. Score da discrepância da área útil do imóvel em relação aos imóveis das redondezas.

### 5.3. Score de atratividade

O score final de atratividade foi construído através de uma simples equação, basicamente somando todos os atributos de atratividade de interesse do cliente. Para que todos os atributos exercessem a mesma influência sobre o score, cada um deles foi normalizado a 100%, ou seja, com valores entre 0-1 (exceto **zprice**).

A fim do cliente ajustar através de sua experiência quais atributos tem maior peso aos olhos do cliente, foi adicionado a opção de atribuir pesos a cada atributo, e.g., neste caso em específico foi atribuído peso 1 para o ano de construção, e peso 0.5 para ano de reforma, porque a reforma, em geral, é incapaz de apagar resquícios de um imóvel muito antigo, então a pontuação da reforma contribui com o score final, mas em menor escala.

Como o objetivo é garimpar oportunidades de investimento, o fator preço foi preponderante, e neste caso optou-se por dividir o valor do score total pelo z-score do preço/pé<sup>2</sup> (**zprice**), desta forma reduzindo drasticamente o score de imóveis superfaturados do radar da Imobiliária. O score final pode ser resumido pela equação (2).

$$(2) \text{ SCORE} = \frac{(\text{waterfront} + \text{view} + \text{condition} + \text{grade} + \text{yr\_built} + \text{yr\_renovated}) - \text{sqft15\_score}}{\text{zprice} + (1 + \min(\text{zprice}))}$$

Como os valores de **zprice** atingem valores negativos, todos eles foram somados a 1 + o menor valor do **zprice** afim do denominador sempre ser maior que 1, desta forma, garantindo que o denominador sempre divida a soma do numerador por um valor maior que 1.

De forma geral pode-se resumir que imóveis com preços por pé<sup>2</sup> muito baixo (em relação ao bairro) terão os seus scores praticamente intocados e aparecerão como boas

oportunidades de investimento. A ideia é que um imóvel em ótimas condições, bem fabricado, jovem e com boa vista, mas que esteja supervalorizado não seja apresentado como oportunidade de investimento ao cliente.

## 6. ANÁLISE FINAL + SUGESTÃO DE IMÓVEIS

### 6.1. Score de atratividade

A definição do score dos imóveis por bairro foi crucial para determinar oportunidades de negócio balanceadas pela condição socioeconômica de cada bairro de King County. Primeiramente antes da atual análise, havia se calculado o z-price para todo o King County, o que acabou gerando uma pontuação enviesada que condenava imóveis de bairros nobres a 'parecerem' super precificados, e imóveis de bairros humildes como boas oportunidades (Figura 4), o resultado final da primeira análise refletiu na realidade, não oportunidades de negócios, mas sim num mapa de calor do preço/pé<sup>2</sup>.

O score ponderado pelas médias e desvios dos bairros apresentaram valores de atratividade que são indiferentes à classe econômica, ou seja, permite que existam boas oportunidades de negócios tanto em bairros mais ricos ou mais pobres. A Figura 5 exemplifica como diferentes bairros apresentam distribuição de score parecidas, independente da classe social do bairro.

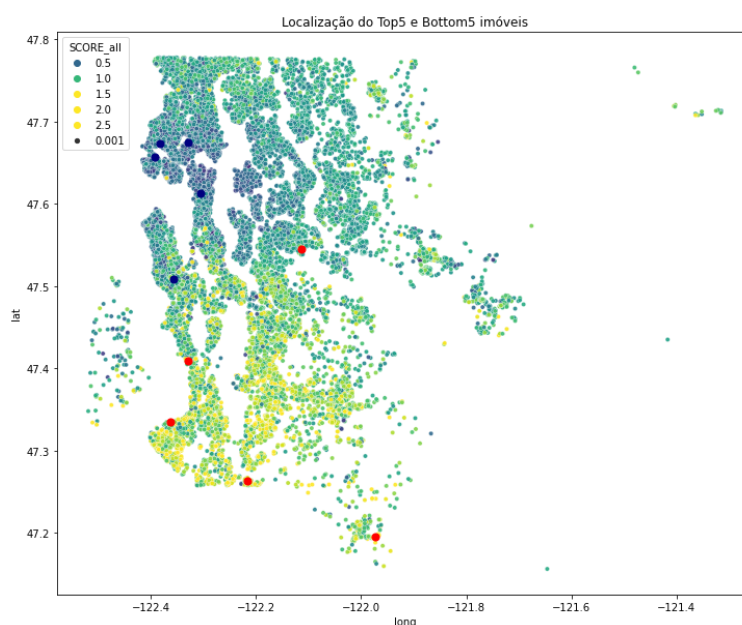
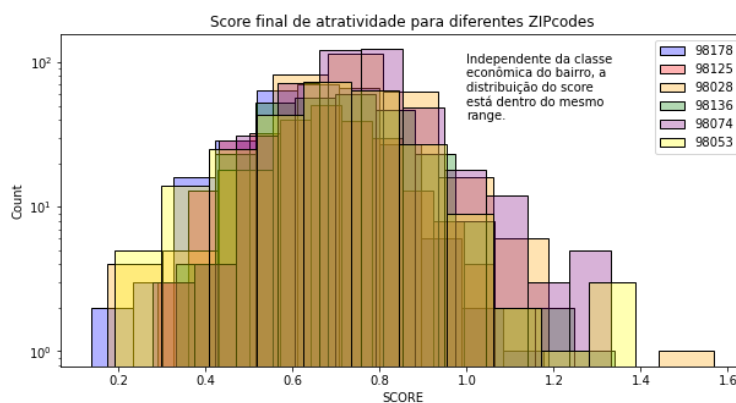
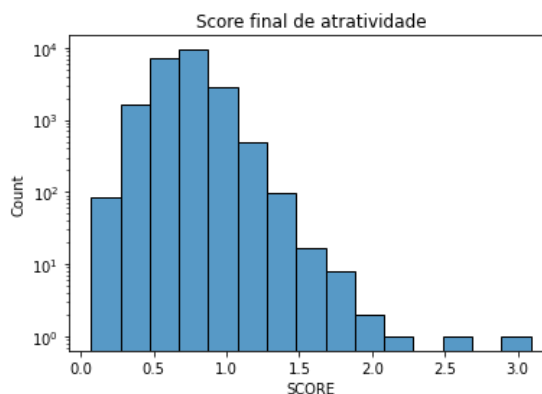


Figura 4. Mapa de localização dos imóveis mais atrativos (Vermelho) e menos atrativos (Azul) pelo método 1.



*Figura 5. Histograma da atratividade de 6 bairros aleatórios de King County.*

A Figura 6 resume a distribuição da pontuação de todos os imóveis considerando seus respectivos bairros, e os atributos citados anteriormente. E a Figura 7, a localização dos 5 imóveis mais atrativos e os 5 menos. Neste mapa é possível observar que a atratividade não se relaciona mais com a região onde encontra-se o imóvel, é possível ver imóveis com scores altos (em amarelo) por todo o condado. Muito pelo contrário, na região onde anteriormente os imóveis estavam com scores baixos, devido ao seu alto preço/pé<sup>2</sup>, agora concentra os 5 imóveis de mais atratividade.



*Figura 6. Histograma do score final de atratividade dos imóveis de King County.*

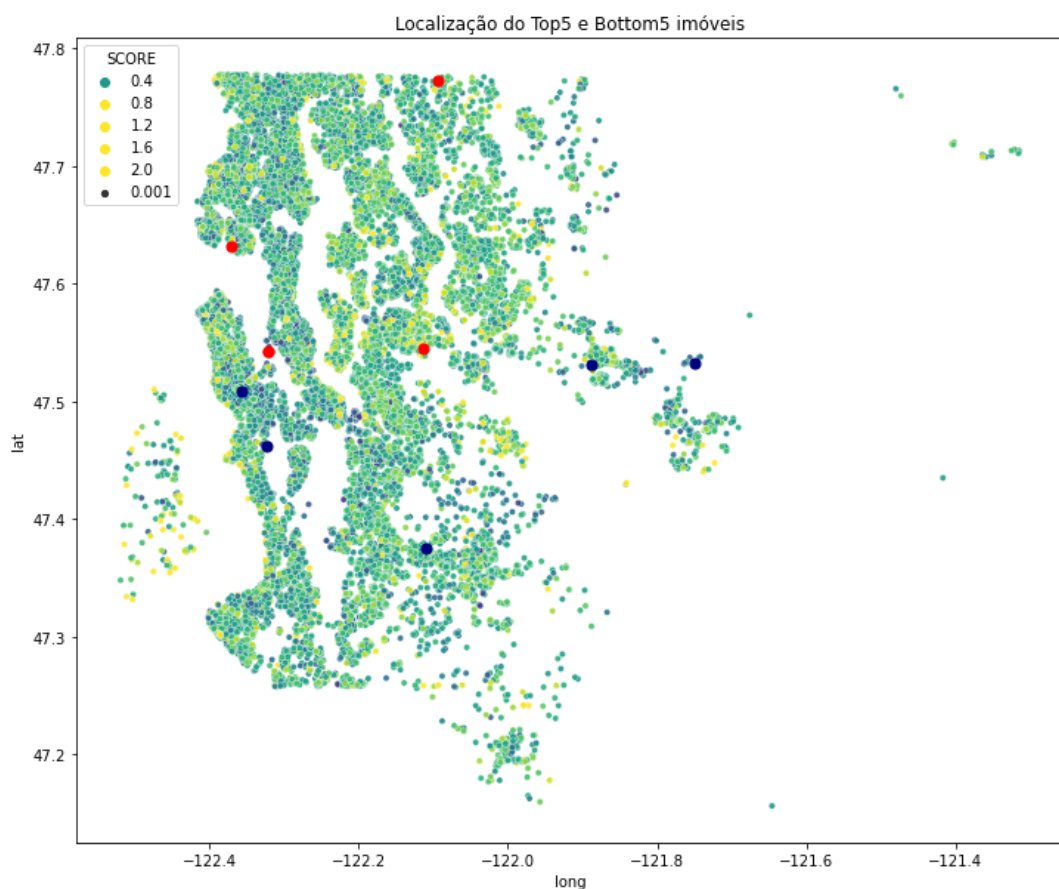
A Tabela 1 e Tabela 2 resumem os 10 imóveis a serem destacados para o relatório. Podemos observar algumas características em comum entre eles, que estão intimamente ligadas a composição do score de atratividade final observado.

Os imóveis mais atrativos possuem condições medianas, mas em contrapartida foram construídos com materiais de mais alta qualidade (7-12), também são imóveis mais jovens, mas não necessariamente são imóveis à beira d'água (provavelmente o que acaba supervalorizando e consequentemente reduzindo o score). A vista pode ser um fator



importante, mas não necessariamente, já que 3 imóveis possuem vista 0. Destaca-se, como esperado, o z-price, dos 5 imóveis todos com valores muito abaixo da média, entre -2.2 e -3.4 desvios abaixo da média do preço/pé<sup>2</sup> do bairro.

Dentre os imóveis indesejados para investimento, pode-se destacar, que são imóveis antigos (<1970), sem reformas, em condições de conservação baixas a médias, e qualidade dos materiais de construção baixas (1-5). Todos não se encontram com vista para água, nem possuem pontuação de vista. Além disso, são imóveis que apresentam preço por pé<sup>2</sup> muito alto, entre 3.2 a 10 vezes o desvio padrão acima da média em relação ao bairro, em resumo com valores extremamente inflacionados para os padrões dos bairros.



*Figura 7. Mapa de localização dos imóveis mais atrativos (Vermelho, dois imóveis são vizinhos, portanto os pontos estão sobrepostos) e menos atrativos (Azul).*

*Tabela 1. Tabela de imóveis mais atrativos.*

	<b>waterfront</b>	<b>view</b>	<b>condition</b>	<b>grade</b>	<b>yr_built</b>	<b>yr_renovated</b>	<b>sqft15_score</b>	<b>zprice</b>	<b>SCORE</b>
326	0	0	3	9	1987	0	0.000033	-3.406637	3.089822
21372	0	2	3	10	2015	0	0.051392	-2.798331	2.517313
21050	0	4	3	12	2007	0	0.179331	-2.196254	2.164529
3844	0	0	3	7	2007	0	0.001356	-2.968844	2.037205
11436	0	0	3	7	2007	0	0.001531	-2.910999	1.958298

*Tabela 2. Tabela de imóveis menos atrativos.*

	<b>waterfront</b>	<b>view</b>	<b>condition</b>	<b>grade</b>	<b>yr_built</b>	<b>yr_renovated</b>	<b>sqft15_score</b>	<b>zprice</b>	<b>SCORE</b>
19452	0	0	1	1	1963	0	0.023424	3.231770	0.068653
13629	0	0	1	5	1932	0	0.006305	5.423579	0.095483
8623	0	0	2	4	1953	0	0.008057	6.132391	0.137851
5832	0	0	2	3	1950	0	0.002014	4.770025	0.137970
8606	0	0	3	5	1953	0	0.017210	10.080076	0.145673