



RELATÓRIO DE ANÁLISE PARA  
INVESTIMENTO EM KING COUNTY

# ROOF IMÓVEIS

---

MAR 2023// PREPARADO POR MIKAELY PEDROSA



# ANÁLISE PARA INVESTIMENTO EM **KING COUNTY**, WASHINGTON.

---

## A QUESTÃO DO NEGÓCIO:

Com o propósito de expansão do negócio no Condado de County, em Washington, a Roof Imóveis deseja saber **quais os cinco imóveis recomendados e quais os cinco que não estão recomendados** para realizar novos investimentos.

Será feita um análise exploratória dos imóveis da região utilizando linguagem de programação Python e suas bibliotecas.

Definiremos critérios para seleção dos imóveis com base nas características da população local de acordo com o censo e também de acordo com a avaliação dos imóveis disponíveis.

## O ENTENDIMENTO DO NEGÓCIO:

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. (Wikipédia)



## DADOS DO ÚLTIMO CENSO:

Households, 2017-2021	902,308
PEOPLE	
Families & Living Arrangements	
Households, 2017-2021	902,308
Persons per household, 2017-2021	2.44
Living in same house 1 year ago, percent of persons age 1 year+, 2017-2021	81.9%
Language other than English spoken at home, percent of persons age 5 years+, 2017-2021	28.9%

<https://www.census.gov/quickfacts/fact/table/kingcountywashington/HSD410221>

Segundo o último censo, **a média de pessoas por casa (lar) é de 2.44 pessoas**. Isto nos dá um norte sobre o tamanho do imóvel que poderemos recomendar.

Housing	
Housing units, July 1, 2021, (V2021)	985,351
Owner-occupied housing unit rate, 2017-2021	56.6%
Median value of owner-occupied housing units, 2017-2021	\$651,900
Median selected monthly owner costs -with a mortgage, 2017-2021	\$2,753
Median selected monthly owner costs -without a mortgage, 2017-2021	\$919
Median gross rent, 2017-2021	\$1,801
Building permits, 2021	19,549

A média do valor das casas ocupadas pelo proprietário é de **\$ 651.900**  
E a média gasta com aluguel bruto é de **\$ 1801**

## DADOS DISPONÍVEIS:

O Dataset disponível contém 21 colunas com características dos imóveis e possui 21613 entradas.

df.head(6)

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15
0	7129300520	20141013T000000	221900.0	3	1.00	1180	5650	1.0	0	0	...	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5650
1	6414100192	20141209T000000	538000.0	3	2.25	2570	7242	2.0	0	0	...	7	2170	400	1951	1991	98125	47.7210	-122.319	1690	7639
2	5631500400	20150225T000000	180000.0	2	1.00	770	10000	1.0	0	0	...	6	770	0	1933	0	98028	47.7379	-122.233	2720	8062
3	2487200875	20141209T000000	604000.0	4	3.00	1960	5000	1.0	0	0	...	7	1050	910	1965	0	98136	47.5208	-122.393	1360	5000
4	1954400510	20150218T000000	510000.0	3	2.00	1680	8080	1.0	0	0	...	8	1680	0	1987	0	98074	47.6168	-122.045	1800	7503
5	7237550310	20140512T000000	1225000.0	4	4.50	5420	101930	1.0	0	0	...	11	3890	1530	2001	0	98053	47.6561	-122.005	4760	101930

6 rows x 21 columns

(Dataset que foi disponibilizado para a consulta dos imóveis)

### Legenda das variáveis:

id = Identificador único do imóvel

date = Data da Venda

price = Preço de venda

bedrooms = N° de Quartos

bathrooms = N° 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

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

## TRATAMENTO DOS DADOS

### Informações sobre o dataset:

```
df.info()

<class 'pandas.core.frame.DataFrame'>
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
```

Após conhecer as colunas e linhas presentes, adicionamos uma coluna extra para identificar a cidade a que pertence cada imóvel por meio do código postal utilizando a função `Us Zipcode`.

zipcode	lat	long	sqft_living15	sqft_lot15	city
98178	47.5112	-122.257	1340	5650	Seattle
98125	47.7210	-122.319	1690	7639	Seattle
98028	47.7379	-122.233	2720	8062	Kenmore
98136	47.5208	-122.393	1360	5000	Seattle

O código completo está disponível ao fim do relatório via link do Colab.

## Tratando a coluna 'Date'

```
df['date'] = pd.to_datetime(df['date'])
print(df.dtypes)
```

```
id                int64
date             datetime64[ns]
price            float64
bedrooms         int64
bathrooms        float64
sqft_living      int64
sqft_lot         int64
floors           float64
waterfront       int64
view             int64
condition        int64
grade            int64
sqft_above       int64
sqft_basement    int64
yr_built         int64
yr_renovated     int64
zipcode          int64
lat             float64
long            float64
sqft_living15    int64
sqft_lot15       int64
dtype: object
```

## Verificando os valores nulos:

```
[ ] ## Verificação de Valores nulos:
    df.isna().sum()
```

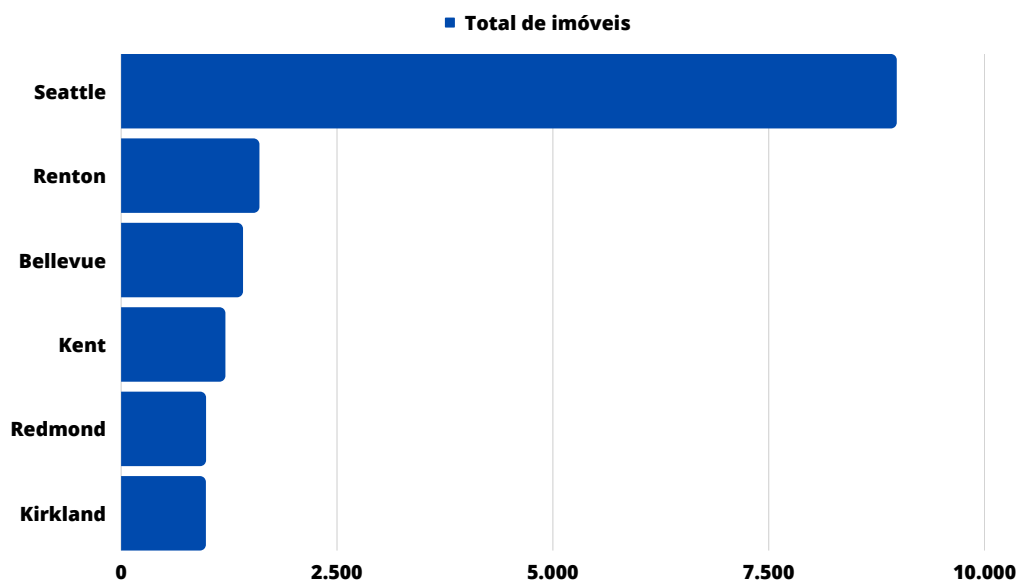
```
date            0
price           0
bedrooms        0
bathrooms       0
sqft_living     0
sqft_lot        0
floors          0
waterfront      0
view            0
condition       0
grade           0
sqft_above      0
sqft_basement   0
yr_built        0
yr_renovated    0
zipcode         0
lat             0
long            0
sqft_living15   0
sqft_lot15      0
city            0
dtype: int64
```

## EXPLORAÇÃO DOS DADOS:

### Distribuição dos imóveis por cidade:

```
idades = df['city'].value_counts().to_frame()
idades = idades.sort_values(by='city', ascending=False)
idades
```

	city
Seattle	8977
Renton	1597
Bellevue	1407
Kent	1203
Redmond	979
Kirkland	977
Auburn	912
Sammamish	800
Federal Way	779
Issaquah	733
Maple Valley	590
Woodinville	471
Snoqualmie	310
Kenmore	283
Mercer Island	282
Enumclaw	234
North Bend	221
Bothell	195
Duvall	190
Carnation	124
Vashon	118
Black Diamond	100
Fall City	81
Medina	50



## Número de andares prevalentes:

```
df['floors'].value_counts().to_frame()
```

**floors**

1.0	10680
2.0	8241
1.5	1910
3.0	613
2.5	161
3.5	8

### Mediana nº de andares:

```
mediana_floors = np.median(df['floors'])  
print(mediana_floors)
```

1.5

## Número de quartos prevalentes:

```
[ ] df['bedrooms'].value_counts().to_frame()
```

**bedrooms**

3	9824
4	6882
2	2760
5	1601
6	272
1	199
7	38
0	13
8	13
9	6
10	3
11	1
33	1

### Mediana nº de quartos:

```
mediana_quartos = np.median(df['bedrooms'])  
print(mediana_quartos)
```

3.0



## Número de banheiros prevalentes:

↳	bathrooms		
	2.50	5380	4.25 79
	1.00	3852	0.75 72
	1.75	3048	4.75 23
	2.25	2047	5.00 21
	2.00	1930	5.25 13
	1.50	1446	0.00 10
	2.75	1185	5.50 10
	3.00	753	1.25 9
	3.50	731	6.00 6
	3.25	589	0.50 4
	3.75	155	5.75 4
	4.00	136	6.75 2
	4.50	100	8.00 2
	4.25	79	6.25 2
	0.75	72	6.50 2
	4.75	23	7.50 1
			7.75 1

## Mediana nº de banheiros:

```

✓ [37] # Mediana do nº de banheiros:
1s   mediana_banheiro = np.median(df['bathrooms'])
      print(mediana_banheiro)

2.25

```

## Mediana do tamanho da área construída:

```

▶ # Mediana da área construída:
  mediana_area_constr = df_final['sqft_living'].median()
  print(mediana_area_constr)

1910.0

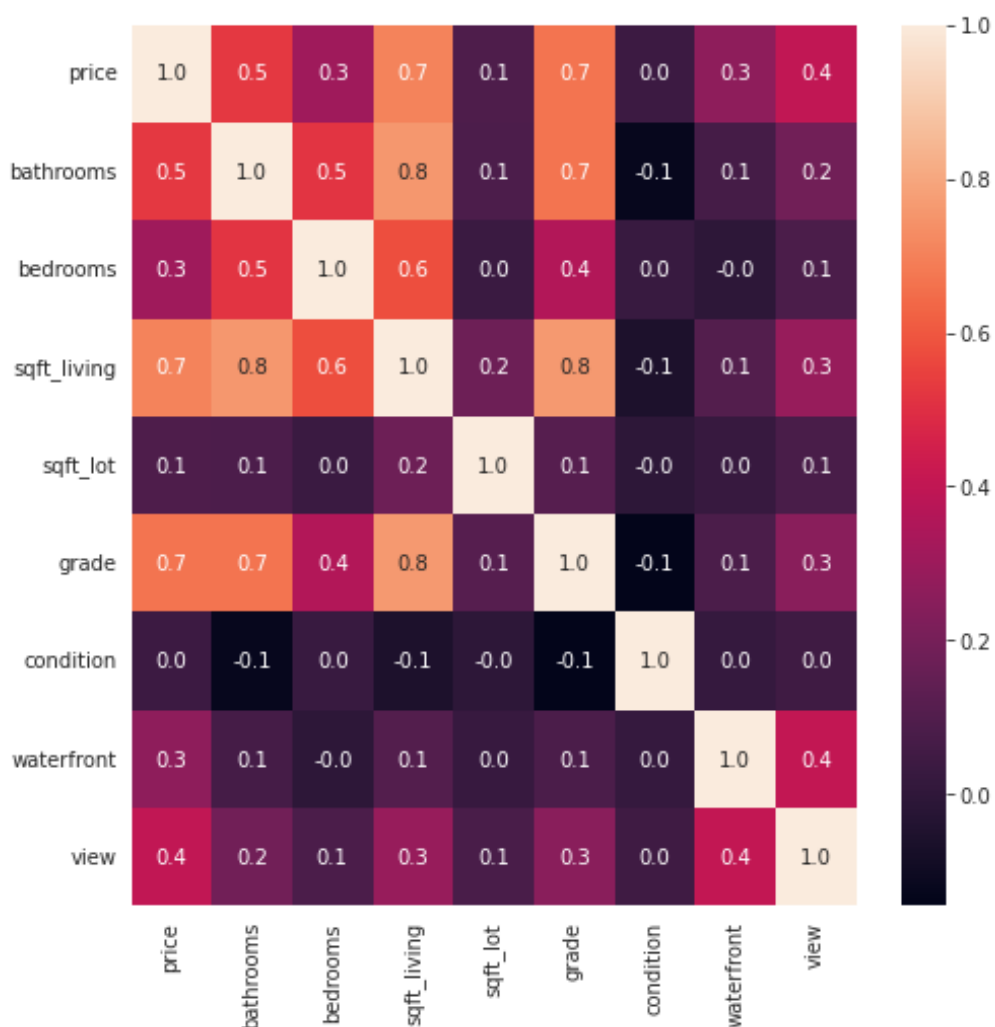
```

## Mediana do preço do imóvel:

```
# Mediana de preços:
mediana_preco = df_final['price'].median()
print(mediana_preco)
```

450000.0

## Correlações entre as variáveis:



O Heatmap avalia a intensidade que uma variável se correlaciona com outra. Quanto mais próximo do valor 1.0, maior a correlação. Aqui iremos focar nas variáveis que estão mais se correlacionando com o preço: sqft\_living (área construída), grade (qualidade do material) e bathrooms (banheiros).

## DEFININDO OS CRITÉRIOS DE SELEÇÃO DOS IMÓVEIS

Selecionaremos os critérios com base em três principais fatores: características da população do Condado, fatores que influenciam no preço do imóvel e com base em algumas medianas.

### CRITÉRIO 1:

**Imóveis que comportem uma família pequena (entre 2 e 4 pessoas)  
Serão selecionados os imóveis que possuam entre 2 e 4 quartos.**

```
criterio1 = (df['bedrooms'] <=2) & (df['bedrooms'] <=4)
```

### CRITÉRIO 2:

**Imóveis com 1 ou 2 andares.**

```
criterio2 = (df['floors'] >= 1) & (df['floors'] <= 2)
```

### CRITÉRIO 3:

**Imóveis a partir de 2 banheiros .**

```
criterio3 = (df['bathrooms'] >= 2)
```

### CRITÉRIO 4:

**Como a média do valor dos imóveis habitados pelo proprietário gira em torno de \$651900 (Censo), selecionaremos os imóveis com valor menor que média para que haja margem para lucro.**

```
criterio4 = (df['price'] <= 651900)
```

**CRITÉRIO 5:**

O valor da área construída maior que a média. Em conjunto com um preço menor que a média, selecionaremos imóveis que possam ser mais valorizados.

```
criterio5 = (df['sqft_living'] >= 1910)
```

**CRITÉRIO 6:**

Qualidade do material de construção. Varia de 1 a 13. Selecionaremos os que tenha nota de 6 para cima.

```
criterio6 = (df['grade'] >= 6)
```

**CRITÉRIO 7:**

Vista do imóvel. Varia de 0 a 4. Selecionaremos os de nota 3 para cima.

```
criterio7 = (df['view'] >= 3)
```

**APLICANDO OS CRITÉRIOS, CHEGAMOS A 5 IMÓVEIS QUE SERÃO RECOMENDADOS PARA INVESTIMENTO:**

```
df_criterios = df_criterios.sort_values('price', ascending=True)
df_criterios.head()
```

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_living15	sqft_lot15	city
3829	8550001515	2014-10-01	429592.0	2	2.75	1992	10946	1.5	1	4	...	1288	704	1903	0	98070	47.3551	-122.475	1110	8328	Vashon
8325	567000025	2015-04-02	577500.0	2	2.50	2330	3000	2.0	0	3	...	2330	0	1915	1994	98144	47.5953	-122.294	1760	4000	Seattle
13312	1822039138	2015-02-27	600000.0	2	2.25	2320	18919	2.0	1	4	...	2320	0	1976	0	98070	47.3905	-122.462	1610	18919	Vashon
16674	623039026	2014-11-25	645000.0	2	2.25	2770	11884	1.0	0	3	...	1570	1200	1969	0	98070	47.5098	-122.474	2310	17097	Vashon
7044	8886000005	2015-03-09	649000.0	2	2.75	2090	23962	2.0	0	3	...	2090	0	1988	0	98070	47.4145	-122.440	1820	32340	Vashon

5 rows x 22 columns



	id	date	price	bedrooms	bathrooms	sqft_living
3829	8550001515	2014-10-01	429592.0	2	2.75	1992
8325	567000025	2015-04-02	577500.0	2	2.50	2330
13312	1822039138	2015-02-27	600000.0	2	2.25	2320
16674	623039026	2014-11-25	645000.0	2	2.25	2770
7044	8886000005	2015-03-09	649000.0	2	2.75	2090

rows x 22 columns

## IMÓVEIS NÃO RECOMENDADOS

Excluiremos aqueles imóveis cujas características fujam muito do padrão anteriormente traçado:

1. Preço muito elevado, pois pode dificultar a negociação visto que foge do padrão comercial do Condado; (price > 900000);
2. Com uma vista que não é considerada agradável (view < 2), fator que pode dificultar a venda posteriormente;
3. Quantidade pequena de banheiros, pelo preço cobrado (bathrooms < 2)
4. Material utilizado com qualidade inferior, pelo preço que é cobrado (grade < 7)

```

[71] excluir1 = (df['price'] > 900000)
[72] excluir2 = (df['view'] < 2)
[73] excluir3 = (df['bathrooms'] < 2)
[74] excluir4 = (df['grade'] < 7)
[75] df_excluidos = df.loc[excluir1 & excluir2 & excluir3 & excluir4]
    df_excluidos.head(10)

```

Chegamos então a 5 imóveis que **NÃO** recomendamos para investimento:



	id	date	price	bedrooms	bathrooms
6243	5536100010	2015-02-04	1050000.0	4	1.0
7319	5101403915	2015-04-03	970000.0	2	1.0
8387	4389201021	2014-07-01	1014250.0	3	1.0
9314	4389200610	2014-12-01	903000.0	2	1.5
18685	3325069064	2015-03-26	1052000.0	3	1.0

5 rows x 6 columns

RELATÓRIO DE ANÁLISE PARA  
INVESTIMENTO EM KING COUNTY

# ROOF IMÓVEIS

---

MAR 2023//  
PREPARADO POR MIKAELY PEDROSA



[LINK DO CÓDIGO](#)

**AGRADECEMOS A CONFIANÇA!**

