



Data Science & Machine Learning

**dnc**

# Mini Projeto - Roof Imóveis

---

Deliverable 01 - Slot 03

# Contextualização

Roof Imóveis é uma das maiores empresas do ramo imobiliário Brasileiro e quer expandir sua área de atuação fazendo um investimento internacional, com isso, ela contratou você, para uma consultoria estratégica.

A empresa irá investir em imóveis no Condado de County, nos Estados Unidos.

Utilizando o dataset disponibilizado, você irá aplicar os conceitos de CRISP-DM aprendidos até o momento em um projeto prático, que busca desenvolver seu pensamento analítico, insights e tomada de decisões.

# King County - Condado de King - Washington

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](#)

O Dataset apresenta a venda de imóveis e suas características na região de Maio de 2014 à Maio de 2015. Tem um total de 21613 registros com um total de 21 atributos.

**Dataset:** <https://www.kaggle.com/harlfoxem/housesalesprediction>

**Dataset detalhado:** <https://geodacenter.github.io/data-and-lab/KingCounty-HouseSales2015/>

# O Que Devo Fazer?

A empresa te contratou pois quer saber qual seria os **5 imóveis que ela deveria investir e o porquê e quais 5 você não recomendaria o investimento de jeito nenhum.**

Você deve montar um colab bem redigido mostrando o passo a passo das etapas de análise do case, fazendo uma introdução e dados atuais sobre o problema.

Também é importante ressaltar/explicar passagens complexas ou insights que você tiver durante a resolução desse projeto.

# Sobre os Dados

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

# Descrição do Projeto

O Projeto de Insight cobre 5 questionamentos basicamente (Que você estabelecer):

## **\* A Questão de Negócio:**

- Qual a minha meta?
- Como posso chegar?

## **\* O Entendimento do Negócio:**

- Quais dados eu tenho disponível?
- Eles são relevantes para o problema?
- Eles me trazem uma solução direta?
- Qual o meu setor ou range de negócio?

### **\* A Coleta de Dados:**

- Os dados que eu tenho fazem sentido?
- Estão no formato que eu gostaria?
- O que mais eu consigo obter de informação destes dados?

### **\* A Limpeza de Dados:**

- Remoção de dados outliers e valores ausentes

### **\* A Exploração de Dados:**

- Visualização dos dados
- Quais as minhas métricas essenciais
- Baseado nos dados, qual imóvel ou tipo de imóveis eu deveria comprar.

Baseado nestes pensamentos que você utilizou para responder as perguntas acima, qual seria os **5 imóveis que a empresa deveria investir e o porquê e quais 5 você não recomendaria o investimento de jeito nenhum.**

# Dicas de Insights

- Encontre a cidade de cada imóvel;
- Calcule o preço por área de imóvel;
- Calcule o lucro/ perda de investimento caso o imóvel seja vendido mais de uma vez naquele período.





# ENTREGA

O Mini Projeto deve ser submetido no portal do deliverable conforme indicado no modulo.

A submissao deve conter, o colab com todos os detalhes do desenvolvimento do mini projeto + breve apresentação do que foi feito em forma de relatório.

Um integrante do grupo somente precisa fazer a submissão.

# A Questão de Negócio:

A meta desse projeto é identificar quais imóveis tem um potencial maior para investimento para que haja um retorno plausível e sem perdas a longo prazo, localizar também locais e imóveis que não terão um retorno viável de investimento. A coleta dessas informações foi realizada através de um conjunto de dados disponibilizados no kagle.

## O Entendimento do Negócio:

Através dos dados disponibilizados no kaggle foi possível fazer um levantamento, mapeamento e filtragem, para solucionar a questão em si em cima do negócio pautado, pois contém o preço de venda, entre outras variáveis, das casas nos meados de 2014 e 2015, bem relevantes para avaliação um modelo de regressão na área do setor imobiliário.

# A Coleta de Dados:

Nessa etapa foi realizado a importação do arquivo csv para realizar o tratamento dos dados da melhor maneira possível

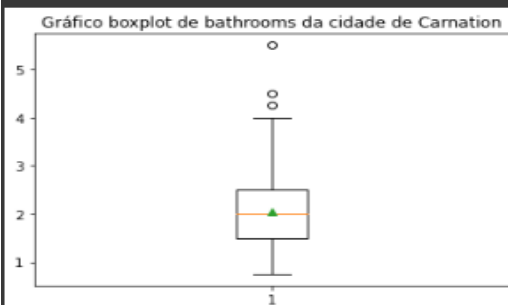
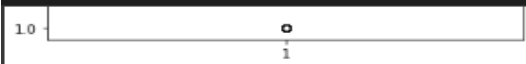
```
df = pd.read_csv('kc_house_data.csv', parse_dates=['date', 'yr_built'], decimal=',.')  
df.head()
```

bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	...	grade	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat	long	sqft_li
3	1.00	1180	5650	1.0	0	0	...	7	1180	0	1955-01-01	0	98178	47.5112	-122.257	
3	2.25	2570	7242	2.0	0	0	...	7	2170	400	1951-01-01	1991	98125	47.7210	-122.319	
2	1.00	770	10000	1.0	0	0	...	6	770	0	1933-01-01	0	98028	47.7379	-122.233	
4	3.00	1960	5000	1.0	0	0	...	7	1050	910	1965-01-01	0	98136	47.5208	-122.393	
3	2.00	1680	8080	1.0	0	0	...	8	1680	0	1987-01-01	0	98074	47.6168	-122.045	

# A Limpeza de Dados:

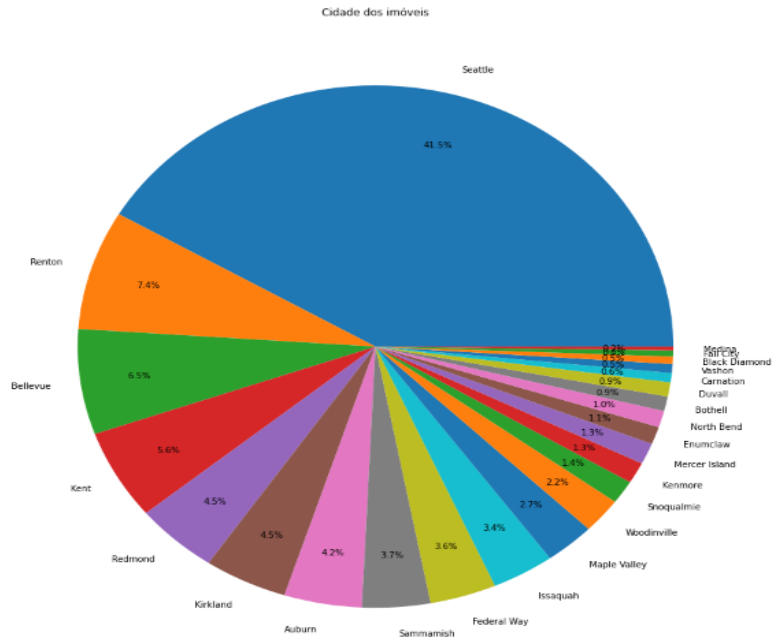
Foi realizada a remoção e visualização dos outliers que são dados que se diferenciam drasticamente de todos os outros.

```
for a in range(len(df['City'].unique())):  
    for b in ['price', 'bedrooms', 'bathrooms', 'sqft_living']:  
        plt.boxplot(df[df['City'] == df['City'].unique()[a]][b].values, showmeans=True)  
        plt.title(f"Gráfico boxplot de {b} da cidade de {df['City'].unique()[a]}")  
        print()  
    plt.show()
```



# A Exploração de Dados:

Verificando o percentual de imóveis por cidade. Grande maioria dos imóveis está localizada em Seattle, Renton, Bellevue, Kent, Redmond, Kirkland e Auburn (>70%).



# Análise final + Sugestão de imóveis

Imóveis que mais deram lucro

	id	venda_1	venda_2	lucro
113	6021500970	345000.0	874950.0	529950.0
147	7977201085	350000.0	740000.0	390000.0
131	7409700215	550000.0	921500.0	371500.0
100	4443800385	410000.0	778100.0	368100.0
57	2568300045	305000.0	649950.0	344950.0

Imóveis que mais deram prejuízo

	id	venda_1	venda_2	lucro
127	7167000040	740000.0	700000.0	-40000.0
61	2726049071	510000.0	489950.0	-20050.0
88	4139420590	1212500.0	1200000.0	-12500.0
63	2767603612	500000.0	489000.0	-11000.0
159	8682262400	430000.0	419950.0	-10050.0



# OBIGADO!



©dnc group. Todos os direitos reservados.