

**NOME: ADJAIR NASCIMENTO BEZERRA**

**RID: 37827**

**LINK COLAB:** [https://colab.research.google.com/drive/1XKI3VLjS\\_zc3KaHOy0B\\_C0HTp16uFHEM?usp=sharing](https://colab.research.google.com/drive/1XKI3VLjS_zc3KaHOy0B_C0HTp16uFHEM?usp=sharing)

## **RESPONDENDO AOS QUESTIONAMENTOS**

- **A Questão de Negócio:**

- Qual a minha meta?

Levantar um faturamento das lojas nos USA e mostrar qual loja é a melhor para expandir o seu tamanho.

- Como posso chegar?

Verificando quais as médias semanais de vendas por loja, verificar o mínimo, máximo e a média ao longo do tempo e com isso escolher a melhor loja para expansão.

- **O Entendimento do Negócio:**

- Quais dados eu tenho disponível?

Na base de dados, temos as informações de vendas semanais, métricas econômicas e meteorológicas de 45 lojas da rede Walmart dos USA, onde contém número da loja, semana de venda, venda naquela semana, se a semana é com feriado ou não, temperatura do dia em °F, preço do combustível por região da loja, também há dados de índice de preços ao consumidor e taxa de desemprego.

- Eles são relevantes para o problema?

Quase todos os atributos neste caso são relevantes, principalmente aqueles que podemos associar com preço de vendas semanais e vendas de todo o período da loja e índice de preços ao consumidor.

- Eles me trazem uma solução direta?

Eles me trazem diversas opções e noções de que ainda não é possível escolher uma loja sem saber a média de vendas e o mínimo, máximo também.

- Qual o meu setor ou range de negócio?

No geral, o meu setor de varejo da Walmart são 45 lojas espalhadas pelo USA. Para a tomada de decisão, filtrei a média, mínima e máxima de vendas por loja ao longo do tempo, onde a loja 4,10,14 e 20 foram umas das que mais se destacaram.

- A Coleta de Dados:

- Os dados que eu tenho fazem sentido?

Sim, fazem sentido, não contém valores ausentes, detalhe no dado de temperatura por região, dependendo da temperatura o valor de vendas poderá diminuir ou aumentar.

- Estão no formato que eu gostaria?

Alguns dados precisavam ser formatados, como no caso do Data.

### Transformando o Date em tipo data

```
[20] 1 df['Date'] = pd.to_datetime( df['Date'].dt.strftime('%m/%d/%Y'))
```

```
[21] 1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6435 entries, 0 to 6434
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Store            6435 non-null   int64
1   Date             6435 non-null   datetime64[ns]
2   Weekly_Sales     6435 non-null   float64
3   Holiday_Flag     6435 non-null   int64
4   Temperature      6435 non-null   float64
5   Fuel_Price       6435 non-null   float64
6   CPI              6435 non-null   object
7   Unemployment     6435 non-null   float64
dtypes: datetime64[ns](1), float64(4), int64(2), object(1)
memory usage: 402.3+ KB
```

- O que mais eu consigo obter de informação destes dados?

Alguns insights aplicados no problema foram os valores de vendas por semana, a correlação dos dados, preço do combustível por região, taxa de desemprego e temperatura por região.

- A Limpeza de Dados:
- Data

A data veio em um formato object, portanto foi feita uma tratativa para limpar os dados.

- Remoção de dados outliers e valores ausentes:  
Não contém valores ausentes.

- A Exploração de Dados:

- Visualização dos dados

É possível montar diversas visualizações de dados e análises, porém limitou-se a ver a melhor loja para expandir, através das análises por média de vendas, temperatura da região e min e max de vendas por loja.

- Quais as minhas métricas essenciais?

Para a escolha da loja a expandir, foi considerado o valor que a loja mostrou perante as outras de acordo com a sua capacidade de vendas.

- Baseado nos dados, qual loja se deve investir?

Com base nas análises as lojas 4,10,14 e 20 foram destaques, mas a loja 20 mostrou o maior poder de vendas e melhor rendimento em valores de vendas min e max dentre as lojas, fortemente recomendo a expansão da loja 20 para uma melhoria jamais vista dentre todas as lojas do varejo.

Vendas máximas, mínimas e médias de todas as lojas ao longo do tempo.

```
1 vendas = df.groupby('Store').agg({'Weekly_Sales':['max', 'min', 'mean']})
2 vendas.head(20)
```

	Weekly_Sales		
	max	min	mean
Store			
1	2387950.20	1316899.31	1555264.40
2	3436007.68	1650394.44	1925751.34
3	605990.41	339597.38	402704.44
4	3676388.98	1762539.30	2094712.96
5	507900.07	260636.71	318011.81
6	2727575.18	1261253.18	1564728.19
7	1059715.27	372673.61	570617.31
8	1511641.09	772539.12	908749.52
9	905324.68	452905.22	543980.55
10	3749057.69	1627707.31	1899424.57
11	2306265.36	1100418.69	1356383.12
12	1768249.89	802105.50	1009001.61
13	3595903.20	1633663.12	2003620.31
14	3818686.45	1479514.66	2020978.40
15	1368318.17	454183.42	623312.47
16	1004730.69	368600.00	519247.73
17	1309226.79	635862.55	893581.39
18	2027507.15	540922.94	1084718.42
19	2678206.42	1181204.53	1444999.04
20	3766687.43	1761016.51	2107676.87

#### ▼ 4.1 Verificação da média de vendas semanais por cada loja

```
[42] 1 media_semanal = df.groupby(['Store'])['Weekly_Sales'].mean().round(2).sort_values(ascending = False)  
      2 media_semanal
```

```
Store  
20  2107676.87  
4   2094712.96  
14  2020978.40  
13  2003620.31  
2   1925751.34  
10  1899424.57  
27  1775216.20  
6   1564728.19
```