

Olá, caro participante!

## Contexto Driva Tecnologia LTDA

Aqui na Driva trabalhamos ativamente com centenas de bases de dados. Neste contexto utilizamos base de dados públicas para enriquecer informações das empresas através de alguma chave em comum, por exemplo, o CNPJ. Posteriormente utilizamos estes datapoints para auxiliar nossos clientes nos desafios de marketing e vendas.

## Instruções iniciais

Partindo deste princípio, você deverá ter um pé em análise e visualização de dados! Seja por meio de um software de modelagem e análise de dados (no caso o **Power BI**) ou uma linguagem de modelagem/análise de dados (**Python, R, entre outras**). Podendo até mesmo utilizar algum software de planilha (**Excel, Google planilhas, entre outros**)! O objetivo aqui é transformar os dados em informação apresentável para o cliente.

## Contexto teste prático

Ter um negócio na feira é algo extremamente desafiador do ponto de vista de dados. No entanto, um comerciante encontrou recentemente um meio de coletar dados a partir de suas vendas para ter um controle maior da sua própria operação.

Após desvendar como implementar um processo de coleta que não impactasse negativamente sua operação, ele se viu novamente em apuros porque descobriu que coletar os dados não era suficiente e que agora precisa de ajuda para transformá-los em informações que agreguem valor à sua estratégia de gestão.

Depois de algum tempo procurando por alguém que fosse capaz de ajudá-lo, chegou até você com os seguintes **desafios**:

1. Como está a evolução do meu faturamento em geral (todos os produtos)

```
1) Como está a evolução do meu faturamento em geral (todos os produtos)
```

```
[9] soma_valores = vendasdf['VALOR_VENDA'].sum()
soma_arredondada = round(soma_valores, 2)
print("A soma dos valores é:", soma_arredondada)
```

A soma dos valores é: 160383.88

Valor identico ao do excel: ☒ Soma: 160383,876

2. Qual foi o meu melhor dia em vendas?

```
2) Qual foi o meu melhor dia em vendas?
```

```
grupo_por_dia = vendasdf.groupby(vendasdf['DATA'].dt.date).agg({'VALOR_VENDA': 'sum'})

melhor_dia = grupo_por_dia.idxmax()[0]
valor_vendas_melhor_dia = grupo_por_dia.max()[0]

print('O melhor dia de venda foi:', melhor_dia)
print('Com um total de vendas de R$', round(valor_vendas_melhor_dia, 2))
```

O melhor dia de venda foi: 2021-10-26  
Com um total de vendas de R\$ 2365.34

3. Qual foi o dia em que eu mais vendi Bananas?

3)Qual foi o dia em que eu mais vendi Bananas?

```
[11] # Filtra o DataFrame para incluir somente as linhas com ID_PRODUTO = 1
df_filtrado = vendasdf[vendasdf['ID_PRODUTO'] == 1]

# Agrupa o DataFrame filtrado por dia e soma a coluna 'VALOR_VENDA' de cada grupo
grupo_por_dia = df_filtrado.groupby(df_filtrado['DATA'].dt.date).agg({'VALOR_VENDA': 'sum'})

# Encontra o dia em que a soma foi maior
melhor_dia = grupo_por_dia.idxmax()[0]
valor_vendas_melhor_dia = grupo_por_dia.max()[0]

print('O melhor dia de venda para o produto 1 foi:', melhor_dia)
print('Com um total de vendas de R$', round(valor_vendas_melhor_dia, 2))
```

O melhor dia de venda para o produto 1 foi: 2021-10-28  
Com um total de vendas de R\$ 291.72

4. Qual produto eu vendo em maior quantidade (kg)?

4)Qual produto eu vendo em maior quantidade (kg)?

```
vendasdf['VALOR_VENDA'] = vendasdf['VALOR_VENDA'].astype(str)
vendasdf['VALOR_VENDA'] = vendasdf['VALOR_VENDA'].str.replace('.', '', 1).astype(float)
produtosdf[['PREÇO_KG', 'PESO_MEDIO_UNITARIO_KG']] = produtosdf[['PREÇO_KG', 'PESO_MEDIO_UNITARIO_KG']].apply(lambda x: x.str.replace('.', '', 1).astype(float))

[14] df = vendasdf.merge(produtosdf, on='ID_PRODUTO')

df['QUANTIDADE_KG'] = df['VALOR_VENDA'] * df['PESO_MEDIO_UNITARIO_KG']

quantidade_por_produto = df.groupby('NOME_PRODUTO')['QUANTIDADE_KG'].sum()

produto_mais_vendido = quantidade_por_produto.idxmax()
produto_mais_vendido

'Uva'
```

5. Quais estratégias você sugeriria para um faturamento maior?

5)Quais estratégias você sugeriria para um faturamento maior?

```
#quantidade que cada produto lucrou em cada mês ordenando o mais vendido

df = vendasdf.merge(produtosdf, on='ID_PRODUTO')

df['LUCRO'] = df['VALOR_VENDA'] * df['PESO_MEDIO_UNITARIO_KG'] - df['PREÇO_KG']

df['DATA'] = pd.to_datetime(df['DATA'], format='%d/%m/%Y')

df_grouped = df.groupby([df['DATA'].dt.month, 'NOME_PRODUTO'])['LUCRO'].sum().reset_index()

df_sorted = df_grouped.sort_values(by='LUCRO', ascending=False)

print(df_sorted)
```

DATA	NOME_PRODUTO	LUCRO
10	Abacaxi	3153.773520
9	Abacaxi	2943.122400
8	Banana	2534.149000
8	Abacaxi	2530.826400
10	Banana	2283.249080
9	Banana	2089.396800
10	Uva	612.047220
8	Limão	496.648480
10	Limão	422.121440
9	Limão	372.131400
10	Laranja	143.429746
8	Laranja	107.122720
9	Laranja	97.595440
10	Mamão	90.163200
8	Mamão	77.774520
9	Mamão	66.961800
10	Caju	-69.230898
9	Caju	-97.639230
8	Caju	-97.747590
9	Uva	-99.269100
8	Uva	-192.310500
10	Tangerina	-245.580840
9	Maçã	-279.343100
10	Maçã	-301.177660
8	Maçã	-377.769200
8	Tangerina	-382.214440
9	Tangerina	-425.556240
8	Abacate	-628.094000
10	Abacate	-681.369200
9	Abacate	-822.890000
9	Pêssego	-1115.015250
10	Pêssego	-1188.782994
8	Pêssego	-1273.540710
10	Ameixa	-2456.658470
8	Ameixa	-2797.625950
9	Ameixa	-2920.810200
10	Pêra	-3418.626368
9	Pêra	-3972.400550
8	Pêra	-4157.789210
10	Carambola	-9423.287800
8	Carambola	-11131.157000
9	Carambola	-11968.613000

No dataset coletado em Agosto, Setembro e Outubro, as frutas Caju, Uva, Tangerina, Maçã, Abacate, Pêssego, Ameixa, Pêra e Carambola tiveram um saldo negativo. Para contornar esse problema, é possível fazer promoções com marketing para obter resultados mais positivos, desde que se saiba o preço adquirido do fornecedor para evitar prejuízos. No entanto, a uva foi o único produto que teve resultado positivo no último mês, enquanto nos meses anteriores foi negativo. Laranja e Mamão mantiveram-se acima do negativo, mas ainda há espaço para melhorar os resultados.

Minha sugestão é coletar dados adicionais, como o prazo de validade das frutas, o preço adquirido do fornecedor e outros gastos, e criar um algoritmo que sugira maneiras de estabilizar o preço antes que os produtos sejam completamente perdidos. Isso envolveria tentar vender as frutas o mais rápido e lucrativamente possível, sem perdas desnecessárias.

### Observações e instruções:

Embase sua estratégia sempre com visuais (por meio de gráficos), dados (tabelas) e descrições qualitativas.

1. O conteúdo deste arquivo é totalmente fictício

3. Considere que todas as vendas são feitas dentro de uma mesma região.

**faça uma cópia base de dados e trabalhe a partir dele, no final envie o arquivo respondido para o e-mail [elimar@driva.com.br](mailto:elimar@driva.com.br) e [patrick@driva.com.br](mailto:patrick@driva.com.br)**

**Dicionário de dados e Base de dados:**  Teste modelagem de dados