

# Análise Bagy

May 8, 2022

**Teste Técnico Bagy** Neste documento está contida a análise do dataset fornecido pela Bagy. O objetivo desta análise é responder as seguintes perguntas: 1. Quais as 10 lojas com maior faturamento em pedidos? Sumarize os dados dessa loja apresentando os seguintes campos: - StoreId - Volume total de itens vendidos - Valor total vendido 2. Qual ticket médio mensal dos pedidos apresentados? Qual o volume médio mensal de vendas (todas as lojas)? 3. Existe alguma correlação entre o país e o volume de itens vendidos? Qual o item mais vendido e o menos vendido no país com maior volume de vendas? 4. Extra - A partir dos números apresentados e o seu conhecimento prévio, apresente ideais/estratégias que podem ser implementadas para auxiliar nossos lojistas a aumentarem o volume de vendas?

## Carregando o dataset e criando as colunas necessárias

```
[ ]: import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline

# para ler o arquivo corretamente é preciso levar em conta que ele está
→ separado por ';'
# e que o encoding adequado não é utf-8 como de costume
df = pd.read_csv('teste_dados_ecommerce.csv', sep=';', encoding='latin-1')

# criando uma coluna com o valor total de cada pedido ao multiplicar a
→ quantidade de unidades
# pelo valor de cada uma
df['Total'] = df['Quantity'] * df['UnitPrice']

# convertendo a coluna de data de string para TIMESTAMP
df['InvoiceDate'] = pd.to_datetime(df['InvoiceDate'])
```

## Pergunta 1

```
[ ]: # separando o dataset por id das lojas e calculando a soma das colunas
→ 'Quantity' e 'Total'
lojas_total = df.groupby(by=['StoreId'])[['Quantity', 'Total']].sum()

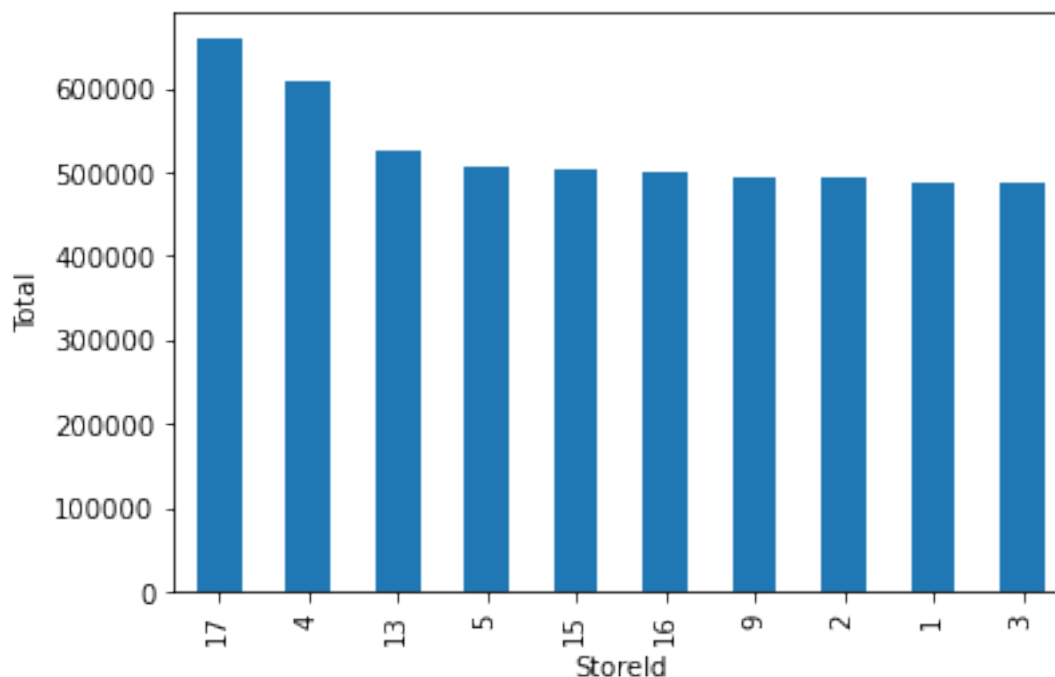
# ordenando os valores e selecionando as top 10 lojas
lojas_total.sort_values(by=['Total'], ascending=False).head(10)
```

```
[ ]:      Quantity      Total
StoreId
17      337514  659675.740
4       350583  607875.170
13      266709  525823.970
5       251249  505066.590
15      268690  504586.530
16      259878  501436.810
9       271067  494288.550
2       241355  493832.540
1       236687  488837.071
3       245406  488437.800
```

Como pode ser observado, na tabela acima temos a ID da cada loja seguido pelo número total de pedidos e o faturamento total. Nessa tabela estão presentes apenas as **10 lojas com maior faturamento**, ranqueadas por faturamento total. **A loja com maior faturamento foi a 17**, que embora não tenha sido a que mais vendeu foi a que mais faturou.

```
[ ]: # novamente ordenando os valores e selecionando as top 10 lojas, mas gerando um
      ↳ grafico de barras
lojas_total.sort_values(by=['Total'], ascending=False)['Total'].head(10).
      ↳ plot(kind='bar', ylabel='Total')
```

```
[ ]: <AxesSubplot:xlabel='StoreId', ylabel='Total'>
```



## Pergunta 2

```
[ ]: # agrupando os dados por ano e mês
dates = df.set_index('InvoiceDate')
# gerando o total de pedidos e faturamento por mês
ticket_medio = dates.groupby(by=[dates.index.year, dates.index.
    ↳month])[['Quantity', 'Total']].sum()

# criando uma coluna com o valor do ticket médio mensal
ticket_medio['Average Ticket'] = ticket_medio['Total'] /
    ↳ticket_medio['Quantity']
ticket_medio
```

```
[ ]:
```

		Quantity	Total	Average Ticket
InvoiceDate	InvoiceDate			
2010	12	342228	748957.020	2.188474
2011	1	308966	560000.260	1.812498
	2	277989	498062.650	1.791663
	3	351872	683267.080	1.941806
	4	289098	493207.121	1.706021
	5	380391	723333.510	1.901553
	6	341623	691123.120	2.023058
	7	391116	681300.111	1.741939
	8	406199	682680.510	1.680655
	9	549817	1019687.622	1.854595
	10	570532	1070704.670	1.876678
	11	740286	1461756.250	1.974583
	12	226333	433668.010	1.916062

Na tabela acima, temos a quantidade total de pedidos para cada mês do ano e o faturamento total mensal. Os dados disponíveis compreendem o período entre dezembro de 2010 e dezembro de 2011. Com esses valores foi possível calcular o *ticket médio de cada mês*, com os valores variando entre 1,7 e 2,1 dependendo do mês.

```
[ ]: # agrupando os dados por ano e mês
qtd_media_vendas_mensais = dates.groupby(by=[dates.index.year, dates.index.
    ↳month])

# separando a coluna de quantidade, somando o total para cada mês e gerando a
    ↳média
qtd_media_vendas_mensais['Quantity'].sum().mean()
```

```
[ ]: 398188.46153846156
```

```
[ ]: # agrupando os dados por ano e mês
total_media_vendas_mensais = dates.groupby(by=[dates.index.year, dates.index.
    ↳month])
```

```
# separando a coluna do total, somando o total para cada mês e gerando a média
total_media_vendas_mensais['Total'].sum().mean()
```

```
[ ]: 749826.7641538461
```

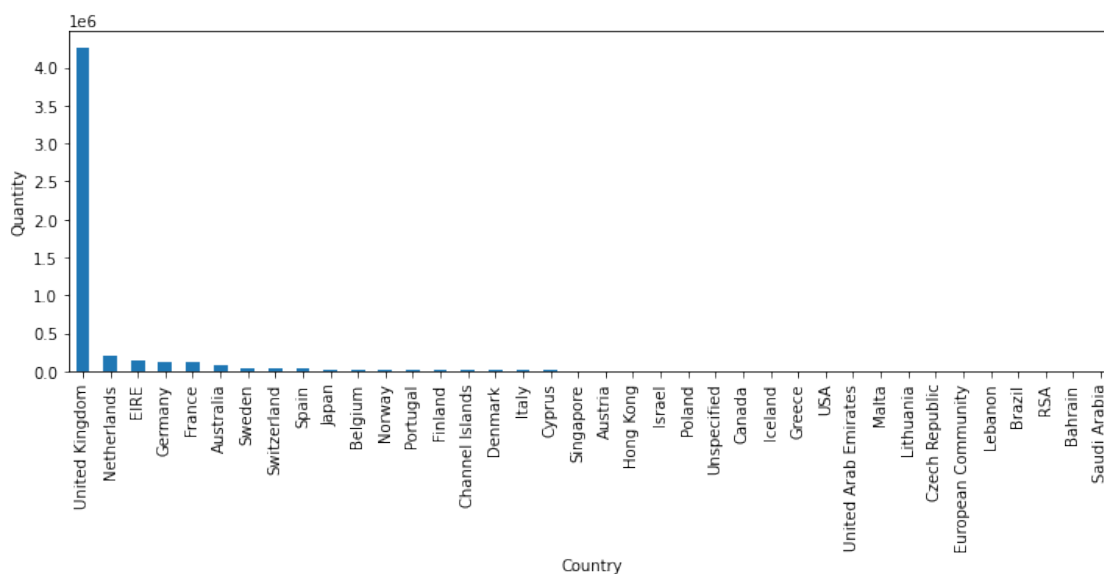
Com base nisso, podemos ver que o ***volume médio de vendas mensal*** (considerando todas as lojas) foi de 398.188 unidades em quantidade e \$749.826,76 em faturamento.

### Pergunta 3

```
[ ]: # agrupando dados por países e somando o total de unidades vendidas em cada um
vendas_por_pais = df.groupby(by=['Country'])['Quantity'].sum()

# ordenando o resultado pelo total de unidades e gerando um grafico de barras
vendas_por_pais.sort_values(ascending=False).plot(kind='bar', figsize=(12, 4),
↪ylabel='Quantity')
```

```
[ ]: <AxesSubplot:xlabel='Country', ylabel='Quantity'>
```



Fica evidente que a maioria das vendas se concentra no Reino Unido, além de alguns poucos países europeus. Como a concentração no Reino Unido é muito alta, é possível inferir que esta plataforma de vendas é muito mais conhecida e acessada lá do que em qualquer outro lugar no mundo. Além disso, fica evidente que o público em geral é europeu e que se existe interesse em expandir e diversificar esses clientes será preciso focar bastante no marketing em outros países.

```
[ ]: # agrupando os dados por país e item vendido, depois somando o total de vendas
↪para cada item
mais_vendido_uk = df.groupby(by=['Country', 'Description'])[['Quantity']].sum()
```

```
# selecionando o resultado do reino unido
mais_vendido_uk.loc['United Kingdom'].sort_values('Quantity')
```

```
[ ]:                                     Quantity
Description
printing smudges/thrown away          -19200
Unsaleable, destroyed.                 -15644
check                                  -12030
Printing smudges/thrown away          -9058
Damaged                                -7540
...                                     ...
WHITE HANGING HEART T-LIGHT HOLDER    33193
ASSORTED COLOUR BIRD ORNAMENT         33679
POPCORN HOLDER                        34365
JUMBO BAG RED RETROSPOT               43167
WORLD WAR 2 GLIDERS ASSTD DESIGNS     48326
```

```
[4202 rows x 1 columns]
```

A tabela acima apresenta o item ***WORLD WAR 2 GLIDERS ASSTD DESIGNS*** como o mais vendido no Reino Unido com um total de ***48.326 unidades***. Já o item menos ‘vendido’ foram itens defeituosos com erros que foram descartados, representado um total negativo de ***-19.200 unidades***. Se descon- sideramos itens negativos, o resultado é que existiram vários itens sem nenhuma venda registrada.

#### Pergunta 4 - Extra

- Tendo em vista a análise realizada, os lojistas que buscam aumentar suas vendas deveriam focar primariamente no mercado do Reino Unido. Isso significa investir mais em anúncios neste país e ter sempre em mente os indicadores de mercado para este país, já que é nele onde se concentram a esmagadora maioria dos consumidores e consequentemente das vendas.
- Direcionar uma parte do investimento para outros mercados pode ser uma opção inteligente para conquistar novos espaços de menor concorrência, mas é necessário ter cautela, uma vez que primeiro se faz útil entender o motivo da concentração num só país para a partir disso pensar em como diversificar. Um cenário de aparente pouca concorrência pode ser enganoso, já que podem existir fatores ocultos não compreendidos nos dados em questão, como uma pla- taforma concorrente nos outros países que concentra as vendas neste.