

## 1 PERGUNTAS

### 1.1 QUAL FOI O FATURAMENTO TOTAL NO PERÍODO?

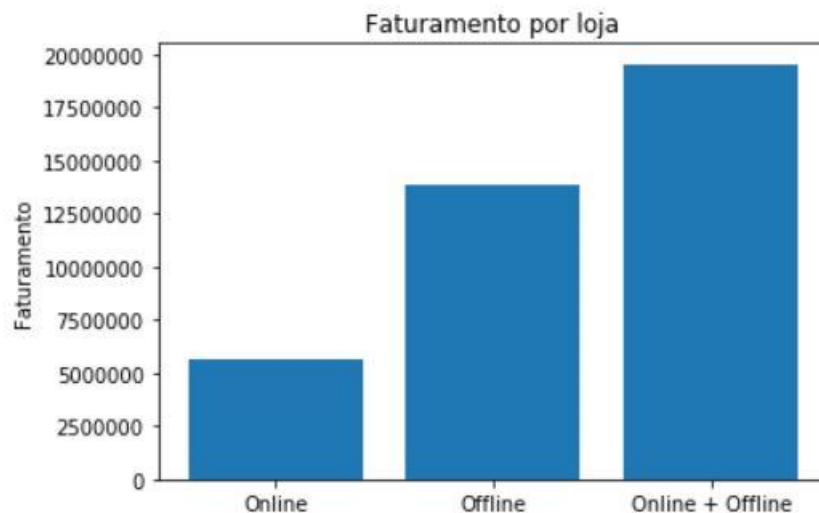
É preciso pegar o valor total das vendas do `offline_data` e `onlineorders_data`. Foi verificado que existem “quantity” nos dois dataframes que são maiores que 1 e também que a coluna “price” só contava o preço unitário, ou seja, não estava sendo multiplicado “quantity” por “price”. Então foi feita uma nova coluna nos dois dataframes para multiplicar as duas colunas (“total”). Feito isso, foi realizado o somatório da coluna “total” nos dois dataframes e posteriormente somando os dois resultados para obter o faturamento total.

O resultado da análise está abaixo.

LOJA	FATURAMENTO (R\$)
ONLINE	5649323
FÍSICA	13903005.32
ONLINE + FÍSICA	19552328.32

A loja online faz parte de 28.9% do faturamento total da rede de lojas. Em contrapartida, 71.1% do faturamento da rede vem das lojas físicas.

Foi plotado um gráfico também para visualizarmos melhor o resultado.

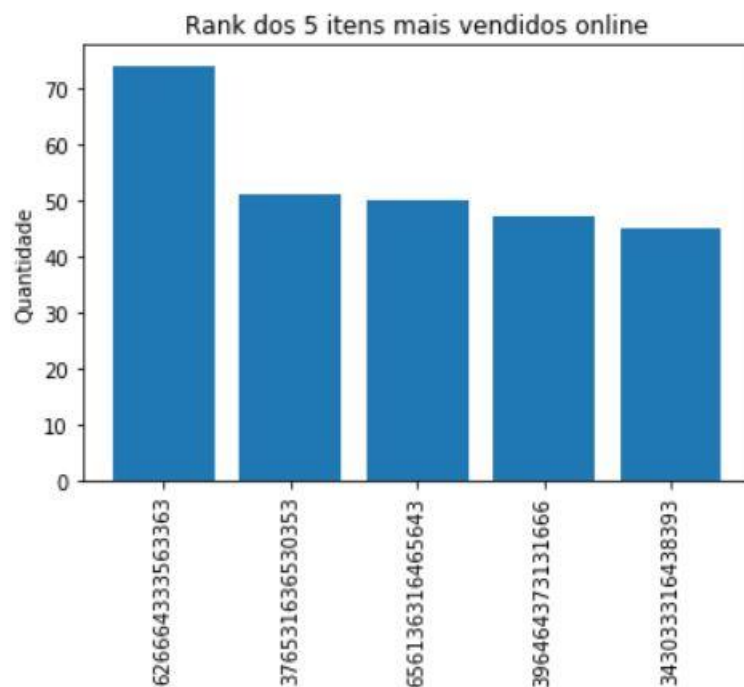


### 1.2 QUAL O PRODUTO MAIS COMPRADO ONLINE?

Para fazer essa análise, primeiramente foi feita uma limpeza do dataframe `onlineorders`, descartando as colunas que não seriam necessárias e deixando apenas “on\_product\_id” e “quantity”. Após isso, foi agrupado as linhas com o mesmo “on\_product\_id” e somando “quantity”.

Feito isso, o item com “quantity” maior (ou seja, o que mais vendeu) foi o de id 626664333563363. Foi também criado um ranking com os 5 produtos que mais venderam, sendo eles apresentados no quadro e gráfico abaixo.

ON_PRODUCT_ID	QUANTIDADE
626664333563363	74
376531636530353	51
656136316465643	50
396464373131666	47
343033316438393	45

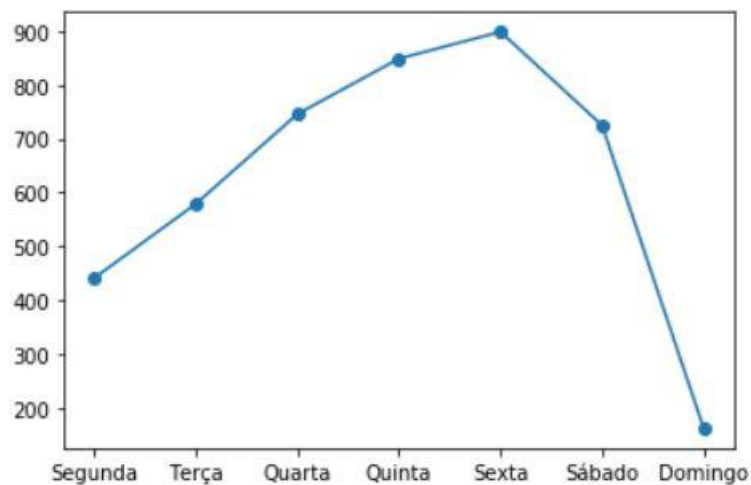


### 1.3 CARIOCAS GOSTAM DE COMPRAR NO FIM DE SEMANA?

Para responder essa pergunta, primeiramente é necessário transformar o “date” no dataframe de vendas offline e inserir o resultado em outra coluna. Optei por fazer uma filtragem no dataframe, iniciando por tirar os estados diferentes de “RJ” e também as colunas que não seriam necessárias para a análise. Feito isso, como a sale\_id pode aparecer em mais de uma linha (no caso de ter mais de um produto na mesma venda), deixei apenas valores únicos para contabilizar as vendas por dia.

Os resultados obtidos pela contagem por dia da semana são mostrados abaixo.

DIA DA SEMANA	VENDAS
SEGUNDA	441
TERÇA	579
QUARTA	746
QUINTA	849
SEXTA	899
SÁBADO	725
DOMINGO	162



Os cariocas fizeram 3514 compras durante a semana, com uma média de 702.8 compras por dia da semana. Já no final de semana, o total de compras foi de 887, com uma média diária de 443.5 compras. A média de compras durante o final de semana é de 63.10% em relação à média durante a semana. Com esses resultados, verificou-se que os cariocas preferem comprar durante a semana. Além disso, percebe-se no gráfico acima que o número de compras cresce ao longo da semana, começando a decair no sábado e atingindo o menor valor no domingo.

#### 1.4 É COMUM ESCOLHER ONLINE E TERMINAR A COMPRA NA LOJA FÍSICA?

Para a pergunta 4 "É comum escolher online e terminar na loja física?" verificarei os `customer_id` que são iguais entre o dataframe de `online_pageviews` e `offline`. Antes disso, é preciso verificar se existem `customer_id` em comum entre o `pageviews` e `onlineorders` (uma vez que podem ter terminado a compra do `pageviews`), deletando os iguais da análise posterior entre `pageviews` e `offline`. Essa é uma das simplificações realizadas para esta análise. A etapa que eu gostaria de fazer mas não consegui seria a de fazer uma

verificação das datas, por exemplo: apenas contabilizar uma "visualização online e compra na loja física" caso a pageview tenha data menor que a do offline.

O primeiro passo será filtrar alguns dados do dataframe, uma vez que nem todos serão necessários e isso também diminuirá o processamento do mesmo.

Ainda é preciso filtrar os `customer_id` que estão como NaN, pois esses não são possíveis de identificar se visitaram a loja física.

serão utilizados apenas os dados referentes aos "pageType" que são "product", "checkout" ou "cart", pois essas são as páginas que o cliente poderia ter visto algum item e analisado se gostaria de comprar ou não antes de ir na loja física. O próximo passo será verificar quais `customers_id` estão presentes neste dataframe.

Utilizando apenas as pageviews entre "product", "checkout" e "cart", 5913 `customers_id` diferentes acessaram essas páginas. Esses serão os ids que verificarei no `onlineorders` e posteriormente a diferença destes no offline. Antes de fazer a diferença entre dataframes, irei deixar apenas valores únicos na coluna `customer_id`.

Irei deixar apenas as colunas de `customer_id` nos dataframes e esses serão os passos dados a seguir. Outro passo será retirar os NaN do "customer\_id" do `online_orders`.

Na coluna merge, as linhas com `left_only` são as que só aparecem na pageviews e não no `onlineorders`, sendo que estes `customer_id` são os necessários para a análise. Desta forma, irei deletar os `customer_id` com merge diferente de `left_only` e depois deletar a coluna merge. 2383 `customers` visitaram "product", "checkout" e "cart" e não concluíram a compra (não aparecem em `onlineorders`). Esse último resultado será analisado juntamente com os valores de `customer_id` do dataframe offline.

Serão realizados os mesmos procedimentos no dataframe do offline, sendo deletados os `customer_id` como NaN, deletando as colunas que são diferentes de `customer_id` e também deixando apenas valores únicos na mesma coluna.

Agora irei realizar o mesmo passo de unir os dataframes e verificar a diferença.

A contagem de both é 135, ou seja, os consumidores que visitaram a loja online e que também visitaram a loja física. Os que só visitaram a loja física (`right_only`) são 9766. Os que só viram online e não visitaram a loja física (`left_only`) são 2248.

1.11% dos consumidores visitaram online e compraram na loja física. Ou seja, não é comum ver online e comprar na loja física.

## **1.5 O TIME DE MARKETING DESTA REDE QUER FAZER UMA CAMPANHA OFERECENDO UM CUPOM DE 20% NAS COMPRAS DE LOJA FÍSICA PARA QUEM VISITOU O SITE E ABANDONOU UM CARRINHO COM OS PRODUTOS. ESTIME O RESULTADO DESSA CAMPANHA.**

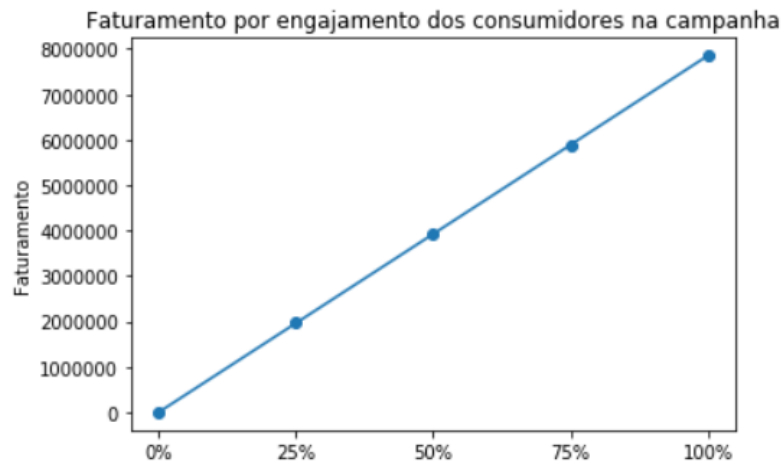
Para a análise da pergunta 5, será pegar quem visitou o "pageType" cart no dataframe pageviews mas não aparece no `onlineorders` e também deletar algumas colunas nesse primeiro passo para diminuir o processamento. Para a análise, deixarei apenas os `customer_id` tanto no pageviews quanto no `online_orders`.

Foi feito um merge entre os dataframes para verificar se o valor do `customer_id` estava contido no outro. Desta forma, só foi mantido o que não estava no `onlineorders`.

Após isso, foram selecionados quais on\_product\_id esses customer\_id apareciam, posteriormente pegando o “price” dos on\_product\_id selecionados no pageviews do da coluna correspondente do onlineorders.

O somatório de preço dos produtos vistos no pageviews pelo customer\_id que parou na página “cart” e não finalizou a compra foi de 9813668 reais. Aplicando o desconto da campanha de 20%, o faturamento total seria de 7850934.4 reais.

Foi analisado também o engajamento dos clientes na campanha. Foram feitos cálculos caso 75%, 50% e 25% do valor dos produtos fossem comprados.



## 2 OUTRAS ANÁLISES

### 2.1 QUAL PLATAFORMA FOI MAIS UTILIZADA PARA COMPRAS ONLINE?

Será analisado qual foi a plataforma mais utilizada para compras online. Será utilizado valores únicos de order\_id. Feito isso, foi realizado o value\_counts() por deviceType.

Desta forma, foi obtido como resultado que mobile foi responsável por 4300 vendas (58.96%) e desktop 2993 (41.04%).

### 2.2 EM QUAIS ESTADO A LOJA FÍSICA VENDE MAIS? EM QUAIS VENDE MENOS?

Outra análise que será feita é qual estado vende mais e qual vende menos. Novamente, será contabilizado apenas sale\_id único.

O estado que mais vendeu é o RJ (4401), já o que menos vendeu foi MS (149).

Segue o gráfico com o resultado total por estado.



### 2.3 QUAIS DIAS DA SEMANA A LOJA OFFLINE VENDE MAIS?

Outra análise será quais dias da semana a loja offline mais vendeu, analisando `sale_id` único novamente.

O dia com mais vendas é sexta (2999) e a com menos venda é domingo (1143).

Segue o gráfico com o resultado total por dia da semana.

