**Universidade Federal Rural de Pernambuco - UFRPE**

**Departamento de Estatística e Informática - DEINFO**

**Bacharelado em Sistemas de Informação - BSI**

**3ª Avaliação**



**Modelagem de dados**

**Equipe:**

Andressa Santana

Jonathas Silva

Lhaíslla Cavalcanti

Marianna Alves

Vitor Olimpio

**Professora:**

Maria da Conceição de Moraes Batista

Recife/PE

2022

**Conjunto de dados públicos de comércio eletrônico brasileiro por Olist**

O conjunto de dados escolhido trata de dados públicos de comércio eletrônico brasileiro por Olist, contendo informações de 100 mil pedidos de 2016 a 2018 feitos em vários marketplaces no Brasil.

A base dados aberta obtida por meio do kaggle e fornecida pela Olist,a maior loja de departamentos dos marketplaces brasileiros, responsável por conectar pequenas empresas de todo o brasil a canais de comunicação em um único contrato, onde os comerciantes podem vender seus produtos através da Olist Store e enviá-los diretamente aos clientes usando os parceiros de logística da Olist. Depois que um cliente compra o produto da Olist Store, um vendedor é notificado para atender esse pedido. Assim que o cliente recebe o produto, ou vence a data prevista de entrega, o cliente recebe uma pesquisa de satisfação por e-mail onde pode dar uma nota da experiência de compra e anotar alguns comentários.

Desta forma a base de dados apresenta informações de pedidos a partir de múltiplas dimensões: desde o status do pedido, preço, pagamento e desempenho do frete até a localização do cliente, atributos do produto e avaliações escritas pelos clientes. Também um conjunto de informações de geolocalização que relaciona os CEPs brasileiros às coordenadas de latitude e longitude.

### **Características sobre a base de dados:**

### Um pedido pode ter vários itens de um mesmo produto ou diferentes produtos.

1. Tratam-se de dados comerciais reais, anonimizados.

**Perguntas das Consultas 1°VA:**

1. Qual a porcentagem das avaliações dos pedidos ?
2. Qual o número de clientes e vendedores por estado?
3. Qual a quantidade de pedidos, seus tipos de pagamento e o valor total referente a cada dia em relação ao dia que ocorreu mais vendas? Que insights tiramos?
4. Qual o item mais caro por estado e categoria? Qual a quantidade de items vendidos? Qual o total de vendas?
5. Qual a média do frete de cada cidade / estado do vendedor pelas categorias do produto?

**Perguntas das Consultas ao DW 3°VA:**

**Obs: O resultado das consultas variam conforme o filtro escolhido**

1. Qual a quantidade de pedidos realizados por consulta de acordo com os dias da semana ou final de semana?

A partir da observação da quantidade de pedidos efetuados por semana, tanto por meio do histograma quanto pelo scatter é possível identificar que os dias onde são realizados a maior quantidade de pedidos são Tuesday com 16.448k de pedidos e wednesday com 15.926k, ambos os dias são considerados meio da semana.

1. Os períodos que ocorrem maior quantidade vendas estão relacionados a datas comemorativas?

Datas comemorativas são períodos em que há um grande fluxo de movimentação financeira, desta forma identificar as datas que possuem maior quantidade de pedidos possibilita uma série de abordagens em tais datas que podem potencializar as vendas, a partir disto as datas comemorativas com maior quantidade de pedidos realizada foram:

24/11 : 1231 pedidos (Período do feriado de ação de graças/black friday)

07/08: 548 pedidos (Período do feriado do dia dos pais)

16/05 : 530 pedidos (Período do feriado do dia das mães)

15/05: 528 pedidos (Período do feriado do dia das mães)

25/11: 519 pedidos (Período do feriado de ação de graças/black friday)

1. Qual a média de avaliação dos pedidos por tipo selecionado?

Os resultados obtidos demonstram a média de avaliação entre a consulta selecionada.

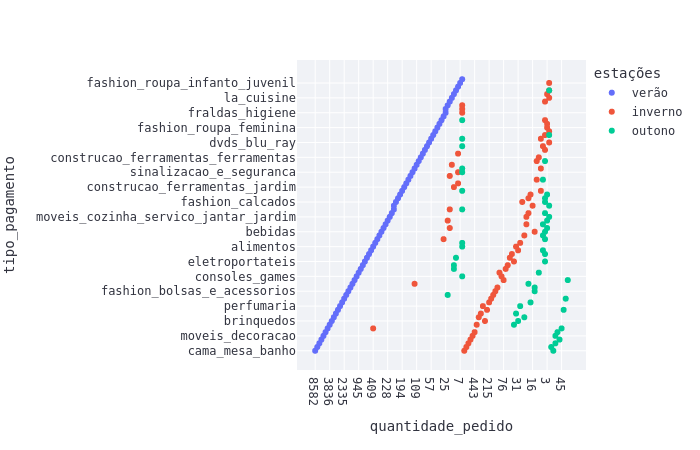
Nessa questão são exibidos 02 gráficos onde podemos visualizar de formas diferentes a relação entre o eixo x e y.

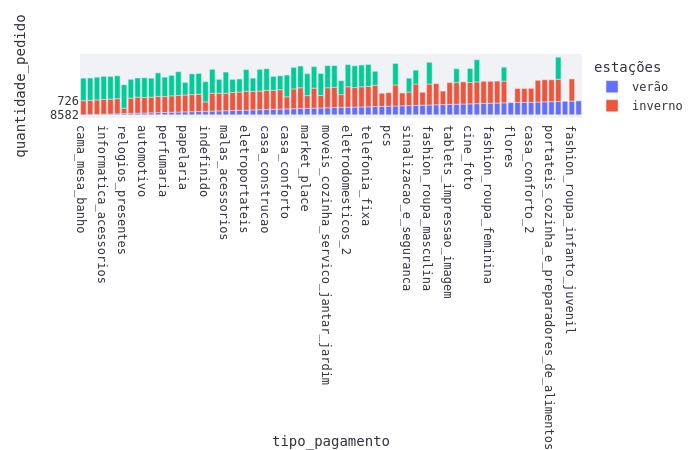
Em ambos os gráficos podemos relacionar a quantidade de pedidos com a média da avaliação de acordo com a opção escolhida. Em geral, quanto maior a quantidade de pedidos, a média da avaliação fica entre a média das notas, quando a quantidade de pedidos é muito pouca, às vezes ocorre do valor ser mais próximo às extremidades.

1. Qual a quantidade de pedidos em relação ao seu ao campo de consulta por estado ?

**Tomando como base o filtro “tipo\_pagamento”**A forma de pagamento é essencial no meio dos setores e-commerce pois é possível identificar a classificação de poder aquisitivo dos consumidores do pedido, uma segmentação da análise dos pedidos por estado possibilita uma atuação de propaganda mais específica para os consumidores. Desta forma é possível identificar um maior número de pedidos que foram pagos com credit\_card, seguido por pagamentos através do boleto. Também é possível identificar que o estado que apresenta a maior quantidade de pedidos é o estado de São Paulo, tanto com as formas de pagamento em credit\_card quanto em boleto ( também na quantidade geral de pedidos).

1. Qual a quantidade de categorias de pedido vendidas por estação do ano?





Por meio desta consulta é possível identificar qual a estação do ano que apresenta a maior quantidade de pedidos e se também a categoria do pedido realizado por estação do ano. Desta forma a estação do ano que apresenta a maior quantidade de pedidos realizada é o outono com 3104 pedidos, o que corresponde a 30% dos pedidos realizados, também é possível identificar que a categoria mais vendida nesta estação é a de cama\_mesa\_banho, que também lidera a quantidade pedidos realizados no inverno com a segunda maior quantidade de pedidos feitos no total. A categoria com maior quantidade de pedidos é beleza\_saude, também na estação do outono.

**Problemas encontrados:**

1. As médias dos pedidos não são proporcionais à quantidade de pedidos efetuados, por exemplo os pedidos que apresentam as maiores médias não apresentam a maior quantidade total.
2. As datas comemorativas que mais movimentam o comércio não são datas fixas, são tadas que alternam conforme o ano, exemplo o dia dos pais e o dia das mães que acontecem no segundo domingo dos respectivos meses de agosto e maio.
3. Os dias da semana em que são realizados o maior número de pedidos no e-commerce são no meio da semana e não nos fins de semana, como é de costume geral da população supor. É necessário uma investigação mais aprofundada para que haja a compreensão do motivo, mas uma das suposições é que por se tratar de um e-commerce torna a compra muito prática, sem necessidade de deslocamento para efetuar a transação.
4. Na construção do Dw pedido pelo pentaho a carga com os id\_ da base de origem causavam repetições das informações.
5. As chaves originais da base de dados de origem eram do tipo varChar desta forma foi necessário alterar o id para auto\_increment para realização da carga do dw.

**Dataset:** [Conjunto de dados públicos de comércio eletrônico brasileiro por Olist](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=olist_customers_dataset.csv)

**Dicionário de Dados :**[Dicionário de dados](https://docs.google.com/spreadsheets/d/1bKNCcatDhR2HYB42XBZFxb1vpTE5EDF2u0sYKsv7yy4/edit#gid=0)

**Modelo ER:** [Modelo ER](https://miro.com/welcomeonboard/UnY5ZFhuWGJHVndPbTZpb252RGZnd2xPY1VRdjZZYzdMektnd05lRjVXdk1kcFVwTXAyTHpIcmp3NnBWOHllcHwzMDc0NDU3MzY2Mjk0NzE0MDQ4?share_link_id=364424513153) **Github:** [Repositório do github](https://github.com/jonathasfsilva/md-grupo1) **(branch: master)**