**INTRODUÇÃO**

Este é um conjunto de dados públicos de comércio eletrônico brasileiro de pedidos feitos na Olist Store. O conjunto de dados contém informações de 100 mil pedidos realizados entre 2016 e 2018 em diversos marketplaces brasileiros. Seus recursos permitem visualizar um pedido em diversas dimensões, desde o status do pedido, preço, desempenho de pagamento e frete até a localização do cliente, atributos do produto e avaliações escritas pelos clientes.

Também lançamos um conjunto de dados de geolocalização que relaciona os CEPs brasileiros às suas coordenadas de latitude e longitude (lat/lng).

Estes são dados comerciais reais, foram anonimizados e as referências às empresas e parceiros no texto da revisão foram substituídas pelos nomes das grandes casas de Game of Thrones.

Depois que um cliente compra o produto na Olist Store, um vendedor é notificado para atender ao pedido. Assim que o cliente recebe o produto, ou vence a data estimada de entrega, o cliente recebe por e-mail uma pesquisa de satisfação onde pode dar uma nota sobre a experiência de compra e anotar alguns comentários.

O arquivo [dataset.png] mostra a modelagem do dataset.

**DESENVOLVIMENTO**

ETL

O processo de tratamento dos dados se iniciou optando pela não utilização do arquivo CSV que traduzia as categorias de produtos para o inglês. Seguindo nesta linha, as colunas e valores de todas as tabelas que continham nomes em inglês foram traduzidos para o português. Para padronizar os dados de maneira rápida e eficiente, foi criada uma função ‘SemAcentos’ para remover os acentos das cidades na tabela de Geolocalização. As colunas ‘Cidade’ e ‘UF’ também foram removidas das tabelas Clientes e Vendedores para evitar redundâncias nos dados. Optou-se por manter apenas o CEP, que seria utilizado para fazer a relação com a tabela de Geolocalização, que já fornecia essas informações, assim como latitude e longitude.

Ainda visando manter dados úteis e bem estruturados na análise, optou-se por excluir colunas de especificação do produto (Tamanho, peso, etc.) da tabela Produtos, a coluna de sequência de pagamentos da tabela Pagamentos e a coluna de identificação da ordem do pedido da tabela Pedidos.

MODELAGEM

Criação de medidas e colunas calculadas utilizando fórmulas DAX, para auxiliar nas futuras visualizações do dashboard.

Ocultação de dados sensíveis como as ID’s que não serão utilizadas e não precisam estar disponíveis aos usuários.

Criação de uma tabela de calendário utilizando a função CALENDARAUTO, que facilita a criação de relatórios e análises baseadas em tempo. Isso permite a identificação de tendências ao longo do tempo, sazonalidade e outras métricas temporais, porque garante que todas as datas possíveis estejam presentes.

VISUALIZAÇÃO

Apesar deste ser um dataset aberto ao público, o nome da Olist não foi utilizado no dashboard e optou-se pela criação de uma logo fictícia.

As artes utilizadas como tela de fundo do dashboard foram feitas no Figma.

Foram desenvolvidas 3 páginas de relatório:

Na página Visão Geral, foram disponibilizados os principais indicadores e informações mais relevantes para a empresa: Faturamento, número de vendas e produtos mais vendidos e que mais faturam. A visualização possibilita a interação entre as categorias e os valores aos quais estão associadas.

O Perfil Regional fornece uma maior compreensão entre as regiões e seus impactos nas vendas e faturamento. Ela serve como um complemento para as informações disponíveis na primeira página do relatório.

Na Página de Indicadores, o foco já não é mais apenas informar sobre o andamento do negócio, mas sim compreendê-lo, levando a um olhar mais crítico na busca por possíveis melhorias.

**CONCLUSÕES**

Claramente um negócio em ascensão, o ecommerce vem em uma crescente no faturamento sendo liderado pelas categorias de Beleza Saúde, Cama Mesa Banho e Esporte Lazer que figuram entre as categorias de maior faturamento como também nas mais vendidas, possibilitando uma maior compreensão do perfil dos clientes. Vale ressaltar que uma análise rápida nas linhas de tendências observando a análise temporal por categoria, mostra que a Relógios Presentes possui o maior crescimento no período disposto.

As 2 regiões de maior faturamento são o Sudeste e o Sul, respectivamente. No Sudeste lidera a categoria de Relógios Presentes citada no parágrafo anterior, região onde São Paulo tem a maior representatividade no faturamento. No Sul é possível observar o Paraná como sendo líder de faturamento onde a categoria Informatica Acessorios tem a maior representatividade.

A avaliação média dos clientes supera a meta estabelecida de 4, o que indica uma boa satisfação dos clientes em relação aos produtos e serviços ofertados. Prova disso é que o ecommerce possui uma baixa taxa de cancelamento: inferior a 1% em relação ao total pedidos realizados. Porém a taxa de retenção que indica quantos clientes repetiram sua compra é de apenas 3%, o que levanta questionamentos como:

- Se o produto/serviço ofertado está sendo bem avaliado, porque os clientes não estão retornando?

- Como é solicitada a avaliação ao cliente no site/aplicativo do ecommerce?

Nota-se que as categorias que mais faturam não figuram nem entre as melhores e nem entre as piores avaliadas e isso seria um ponto de atenção.

Os gráficos de rankeamento das categorias podem ser utilizados em uma decisão sobre quais produtos devem ser mais ofertados ou participarem das publicidades da empresa e quais produtos devem passar por avaliações aprofundadas em busca de melhorias.

Em relação as entregas, elas levam em média 12 dias para serem efetuadas e 6% delas são realizadas com atraso.

Já os horários onde ocorrem a maior quantidade de pedidos são perfeitos para a empresa realizar ofertas e promoções para os clientes. Já os horários que possuem menor quantidade de pedidos servem para a realização de manutenção no site e aplicativo da empresa a fim de trazer uma melhor experiência aos usuários.

**INTRODUCTION**

This is a public dataset of Brazilian e-commerce orders placed on the Olist Store. The dataset contains information on 100,000 orders from 2016 to 2018 made on various marketplaces in Brazil. Its features allow you to visualize an order in multiple dimensions: from order status, price, payment and shipping performance to customer location, product attributes and finally customer reviews. We also released a geolocation dataset that links Brazilian ZIP codes to lat/lng coordinates.

This is real commercial data, it has been anonymized and references to companies and partners in the review text have been replaced with the names of the great houses from Game of Thrones.

After a customer purchases a product from the Olist Store, a seller is notified to fulfill the order. As soon as the customer receives the product, or the estimated delivery date passes, the customer receives an email satisfaction survey where they can rate the shopping experience and write down some comments.

The image [dataset.png] shows the modeling of this dataset.

**DEVELOPMENT**

ETL

The data processing process began by opting not to use the CSV file that translated product categories into English. Following this line, the columns and values of all tables that contained English names were translated and a 'SemAcentos' function was created to remove the accents from the cities in the Geolocation table in order to standardize the data quickly and efficiently. The 'City' and 'UF' columns were also removed from the Customers and Sellers tables to avoid data redundancy, opting to keep only the ZIP code that would be used to make the relationship with the Geolocation table that already provided this information as well as latitude and longitude.

MODELING

Creation of measures and calculated columns using DAX formulas to assist in future dashboard visualizations.

Hiding sensitive data such as IDs that will not be used and do not need to be available to users.

Creation of a calendar table using the CALENDARAUTO function which facilitates the creation of reports and time-based analyses, such as trends over time, seasonality, and other temporal metrics, because it ensures that all possible dates are present.

VISUALIZATION

Despite this being a public dataset, the Olist name was not used in the dashboard and a fictitious logo was created instead.

The artwork used as the dashboard background was made in Figma.

Three report pages were developed:

The Overview page provides the main indicators and most relevant information for the company: Revenue, number of sales and best-selling and highest-grossing products. The visualization allows interaction between the categories and the values associated with them.

The Regional Profile provides a greater understanding of the regions and their impacts on sales and revenue. It serves as a complement to the information available on the first page of the report.

On the Indicators Page, the focus is no longer just on informing about the progress of the business, but on understanding it, leading to a more critical look in the search for possible improvements.

**CONCLUSIONS**

Clearly a rising business, e-commerce has been growing in revenue, being led by the categories of Beauty and Health, Bed, Bath and Table and Sports and Leisure, which are among the highest revenue and best-selling categories, allowing a better understanding of the customer profile. It is worth noting that a quick analysis of the trend lines observing the time analysis by category, shows that Watches and Gifts has the highest growth in the period disposed.

The two regions with the highest revenue are the Southeast and South, respectively. In the Southeast, the Watches and Gifts category mentioned in the previous paragraph leads, a region where São Paulo has the highest revenue representation. In the South, Paraná can be observed as the leader in revenue, where the Computers and Accessories category has the highest representation.

The average customer rating exceeds the established target of 4, which indicates good customer satisfaction with the products and services offered. Proof of this is that the e-commerce has a low cancellation rate: less than 1% in relation to the total orders placed. However, the retention rate, which indicates how many customers repeated their purchase, is only 3%, which raises questions such as:

If the product/service offered is being well evaluated, why are customers not returning?

How is the customer evaluation requested on the e-commerce website/application?

It is noted that the categories that generate the most revenue are neither among the best nor the worst evaluated, and this would be a point of attention.

The category ranking graphs can be used to decide which products should be offered more or participate in the company's advertising and which products should undergo in-depth evaluations in search of improvements.

In relation to deliveries, they take an average of 12 days to be made and 6% of them are made with delay.

\*\*The times when the most orders occur are perfect for the company to make offers and promotions for customers. The times with the least number of orders are used to perform maintenance