# Case OLIST

Foram criadas visualizações que auxiliam no monitoramento das categorias dos produtos vendidos pelo e-commerce, com valores mínimos e máximo e totalidade de vendas por estado. Uma visualização onde se pode acompanhar o total de vendas feitas por estado assim como as compras feita por localidade e notas por categoria dos produtos. Uma sessão criada para compara as avaliações com a quantidade de comentários dos produtos. Na última tela pode-se acompanhar a média de tempo gasto para se realizar uma entrega e a estimativa para ela, assim como o total de itens já entregues, total de itens por status da entrega e as notas dos itens entregues.

Na página 1 do dashboard temos as seguintes KPIs Total de Categorias de Produtos onde vemos o total de categorias existente na base de dados, Menor valor Produto com o menor valor de um produto e Maior valor Produto com o produto de maior valor. Toda essa *view* pode ser manipulado pelo filtro Categoria de Produto. Podemos visualizar as categorias e quantidades de venda em uma tabela e visualizar o total de compras feitas por estado no mapa.

Na página 2 temos o total de vendas por estado em um gráfico de funil e 4 mapas onde neles acompanhamos total de clientes por estado, vendedores por estado e compras e vendas por UF.

Na página 3 é visível uma tabela com as informações referente a quantidade de notas dadas por avaliações e quantidade de comentários feitos de acordo com cada nota de 1 a 5. Temos um gráfico de área com a proporção avaliação x comentários. Também temos um gráfico de barras com as notas por categoria dos produtos.

Na página 4 a visualização é focada nos prazos de entrega onde nelas podemos acompanhar média de tempo que se leva para ser feita uma entrega desse a data de aprovação do pedido e a média real do tempo levado para essa entrega ser realizada levando em conta a data da aprovação do pedido para a data da entrega efetuada. Temos uma tabela com os status de entrega e o total de cada. E um gráfico de barras com as notas por itens entregues.

A fonte de dados é baseada em arquivos do tipo *csv* foram todos carregados através do Pentaho Data Integration para que fossem transformados e carregados no bando de dados PostegreSQL. Foram identificadas as seguintes tabelas como fato: order\_items, order\_payments, order\_reviews e orders. As tabelas a seguir foram identificadas como dimensão: customers, geolocation, products, sellers e category\_name\_translation. Foi criada uma dimensão tempo através da stored procedure UP\_CARGA\_DIMENSAO\_TEMPO desenvolvida previamente. Tudo armazenado em 2 data Warehouse: dw\_olist e dw\_entrega.

Na ETL foi optado por criar 1 *transformation* para cada arquivo csv dando a carga em uma tabela no banco de dados, dividindo as *transformation* através dos Jobs para dimensão e para fato. A dimensão tempo foi criada através de uma *proccedure* e populada com o período entre o ano 2000 e o ano 2100, visto que não existem dados anteriores a 2017 e superiores a 2018 nos dados entregues. O dw\_entrega foi criado a parte para ser tratado apenas as informações referentes as entregas.

Foi criado um modelo de predição em Python carregando os dados do dw\_entrega onde nele tenta prever a chance de um pedido ser entregue com sucesso ou cancelado, entretanto as informações contidas na base de dados não foi o suficiente para gerar um modelo de aprendizado confiável. Sendo assim não foi visto a utilidade de usá-lo como fonte no dashboard.

# Arquitetura

Diagrama

Descrição gerada automaticamente