**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Business Intelligence e Analytics**

**Isabel Marcolino da Silva**

**VENDAS EM *MARKETPLACE*: UMA ANÁLISE DO *E-COMMERCE* OLIST STORE**

Belo Horizonte

2021

**Isabel Marcolino da Silva**

**VENDAS EM *MARKETPLACE*: UMA ANÁLISE DO *E-COMMERCE* OLIST STORE**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Business Intelligence e Analytics como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2021 **RESUMO**

Com a expansão e a popularização do acesso à internet, surge em paralelo o comércio eletrônico. *Marketplace* é definido como um shopping virtual, uma plataforma online em que o proprietário permite a ofertas de produtos e serviços de diferentes vendedores. Este trabalho analisa dados da *startup* Olist, empresa brasileira que atua no segmento de *e-commerce*, por meio de *marketplace*. Foram elaborados painéis estratégico, tático e operacional e também realizadas análises de Machine Learning (regra de associação e K-Means) utilizando a linguagem Python. A análise dos dados permitiu conclusões em níveis financeiro, marketing e logístico. Há oportunidade para incentivo de formas de pagamento pouco utilizadas e *marketing* direcionado tanto para aumento de faturamento em meses com menor demanda quanto para o incentivo de compras associadas (mais de um produto por ordem). Sob o ponto de vista logístico, há oportunidades para reduzir o tempo e o custo de entrega dos produtos.

Palavras-chave: Análise de dados. *Dashboard*. *E-commerce*. Vendas.

**SUMÁRIO**

[1. Introdução 5](#_Toc67504144)

[1.1. Contexto 5](#_Toc67504145)

[1.2. Público alvo 6](#_Toc67504146)

[2. Modelo de Dados 6](#_Toc67504147)

[2.1. Modelo Dimensional 6](#_Toc67504148)

[2.2. Fatos e Dimensões 7](#_Toc67504149)

[3. Processo ETL 8](#_Toc67504150)

[4. Camada de Apresentação 8](#_Toc67504151)

[4.1. Painel Estratégico 8](#_Toc67504152)

[4.2. Painel Tático 10](#_Toc67504153)

[4.3. Painel Operacional 11](#_Toc67504154)

[4.4. Análises de Machine Learning 12](#_Toc67504155)

[5. Registros de Homologação 14](#_Toc67504156)

[6. Conclusões 15](#_Toc67504157)

[7. Links 17](#_Toc67504158)

[REFERÊNCIAS 18](#_Toc67504159)

# 1. Introdução

## 1.1. Contexto

Com a expansão e a popularização do acesso à internet, surge em paralelo o comércio eletrônico (*e-commerce*). Comércio eletrônico é toda cadeia de valor e processos transacionados entre dois parceiros de negócio ou entre um negócio e seus consumidores, em ambiente eletrônico através de tecnologias e sistemas de informação (ROSA, 2019).

*Marketplace*, por sua vez, pode ser definido como um shopping virtual, uma plataforma online em que o proprietário permite a ofertas de produtos e serviços de diferentes vendedores. Assim, produtos e serviços de diferentes vendedores serão exibidos no mesmo local e todo o processo de compra (desde o pedido até o pagamento) é realizado no mesmo ambiente – do proprietário (BALTAZAR, 2020; ROSA, 2019). A comercialização através do *marketplace* permite expandir o mercado de consumidores disponíveis para um vendedor. Em contrapartida, isso acarreta despesas para este que terá de arcar com algumas despesas como, por exemplo uma taxa fixa de participação na plataforma, ou porcentagem de faturamento com a plataforma (TAKAHATA, 2017).

Neste trabalho, serão analisados dados da *startup* Olist Store, empresa brasileira que atua no segmento de *e-commerce*, por meio de *marketplace*. Resumidamente, o Olist Store concentra vendedores que desejam anunciar em *marketplaces* como Mercado Livre, B2W, Via Varejo e Amazon (figura 1).



Figura 1. Estrutura de negócio do Olist.   
Fonte: <https://blog.olist.com/olist-store-como-funciona/>, acessado em 01/03/2021.

Os dados referem-se a vendas realizadas entre 2016 e 2018, e foram extraídos do site Kaggle, plataforma que hospeda datasets públicos úteis para análises de dados, Machine Learning, Big Data e outros. Os *dashboards* foram elaborados utilizando o software Power BI Desktop e apresenta análises em níveis Estratégico, Tático e Operacional. Por fim, foram realizadas análises de Machine Learning utilizando a linguagem Python, na ferramenta Jupyter Notebook.

## 1.2. Público alvo

O público alvo abrange todos os *stakeholders* da cadeia de valor: dos vendedores à alta gestão. A solução foi desenvolvida para ser utilizada sem a necessidade de conhecimentos prévios de programação, porém o ideal é que o usuário tenha conhecimentos mínimos de sistema operacional para navegar nos paineis. Entretanto, para os paineis voltados ao nível gerencial, é interessante que se tenha um conhecimento mínimo de análise de dados para obtenção de *insights* para o negócio e também no auxílio na tomada de decisão. De forma semelhante, os *notebooks* criados no Jupyter Notebook foram pensados para o público gerencial, como forma de obtenção de *insights* e resposta a questões de negócio.

# 2. Modelo de Dados

## 2.1. Modelo Dimensional

Para este trabalho, o modelo dimensional utilizado foi o *star schema* (uma tabela fato ligada a várias tabelas dimensão).

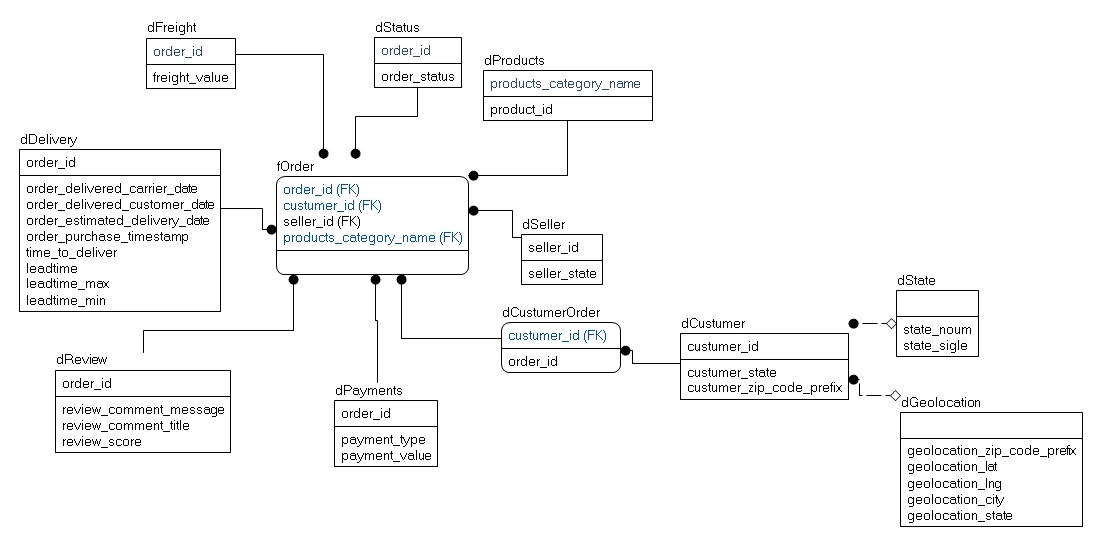


Figura 2. Diagrama dimensional, elaborado com a ferramenta Erwin Data Modeler.

## 2.2. Fatos e Dimensões

Para este trabalho foram utilizadas as seguintes tabelas:

Tabela 1 - Descrição das tabelas fato e dimensão.

|  |  |  |
| --- | --- | --- |
| **Tabela** | **Descrição** | **Tipo** |
| dCustomerOrder | Contém o id da ordem (PK) e o id do cliente. | Dimensão |
| dCustumers | Contém o id do cliente (PK) e a sigla do seu estado. | Dimensão |
| dDelivery | Contém o id da ordem (PK), datas de compra, de envio e de entrega (real e estimada). Também contém as colunas calculadas, criadas diretamente no Power BI: tempo demorado para despachar a ordem e lead time (real, mínimo e máximo). | Dimensão |
| dFreight | Contém o id da ordem (PK) e o valor do frete. | Dimensão |
| dPayments | Contém o id da ordem (PK), o tipo de pagamento (crédito, débito, voucher e boleto) e o valor do pagamento. | Dimensão |
| dProducts | Contém o id do produto (PK) e sua categoria. | Dimensão |
| dReview | Contém o id da ordem (PK), o título e o texto da avaliação do cliente e a nota atribuída pelo cliente. Tabela em formato .csv. | Dimensão |
| dSellers | Contém o id do vendedor (PK) e a sigla do seu estado. | Dimensão |
| dStateSigle | Contém a sigla do estado e o respectivo nome por extenso. | Dimensão |
| dStatus | Contém o id da ordem (PK) e o seu status. | Dimensão |
| dGeolocation | Contém o CEP (PK), latitude, longitude, cidade e estado. Tabela em formato .csv, utilizada apenas na análise de K-Means no Jupyter Notebook. | Dimensão |
| fOrder | Contém o id da ordem (FK), do produto e do vendedor. | Fato |

# 3. Processo ETL

Os *datasets* foram obtidos da plataforma Kaggle e a definição de tabelas fato e dimensão foram realizadas em Excel. O restante do processo ETL foi realizado no software Power BI Desktop. As planilhas foram carregadas no Power BI em formato .xlsx e, em seguida, foi realizada a transformação de dados no Power Query: ajuste de títulos de colunas, adição de metadados (em *Applied steps*), criação de colunas calculadas.

# 4. Camada de Apresentação

A página inicial apresenta os links para cada um dos painéis: estratégico, tático e operacional. Basta clicar sobre um desses termos que o usuário é direcionado para o *dashboard* correspondente. Em cada um dos outros painéis, no canto superior direito, há um ícone de casa que redireciona o usuário à página inicial.

## 4.1. Painel Estratégico

O painel estratégico apresenta uma visão gerencial dos dados. Na porção superior há *cards* sumarizando alguns aspectos do negócio.

* Faturamento: trata-se da soma do valor de todas as ordens;
* Ticket médio: valor médio das ordens – no painel tático esse item é visto sob perspectiva mensal;
* Clientes: quantidade de clientes registrados na base;
* Avaliação média: média das avaliações dos clientes.

Também há na porção superior uma visualização de KPI: o faturamento anual (*year to date*), isto é, faturamento do ano e também a meta do faturamento para o ano e a porcentagem restante para atingir a meta. Ressalta-se que a medida criada como meta para foi estipulada (10M BRL).

Na porção do meio do painel apresenta-se um *multi-row card*, sumarizando as seguintes informações:

* Categorias: número total de categorias de produtos registradas na base – no painel tático essa informação é vista com perspectiva anual;
* Vendedores: quantidade de vendedores registrados na base;
* Ordens por vendedor: quantidade média de ordens por vendedor.

Um gráfico de barras mostra as três categorias com maior faturamento e um gráfico de pizza apresenta o quanto do faturamento representa frete e quanto representa produto.

O gráfico de Gauge (velocímetro) apresenta o *lead time* médio em dias para entrega do pedido. Considera-se *lead time* o tempo decorrido entre a data de compra e data de entrega. Nos painéis tático e operacional outro aspecto desse tema é apresentado sobre o ponto de vista tempo médio para envio.

Como parâmetros de mínimo e máximo considerou-se:

* *Lead time* mínimo: tempo decorrido entre a data de compra e a data de entrega à transportadora;
* *Lead time* máximo: tempo decorrido entre a data de compra e data estimada de entrega.

Ao final, o faturamento é apresentado por ano e também a previsão para os próximos cinco anos. No painel tático esse tópico é visto sob perspectiva mensal.

Nesse painel é possível filtrar a visualização por anos (considerando a data do pedido). O filtro localiza-se no canto superior esquerdo.

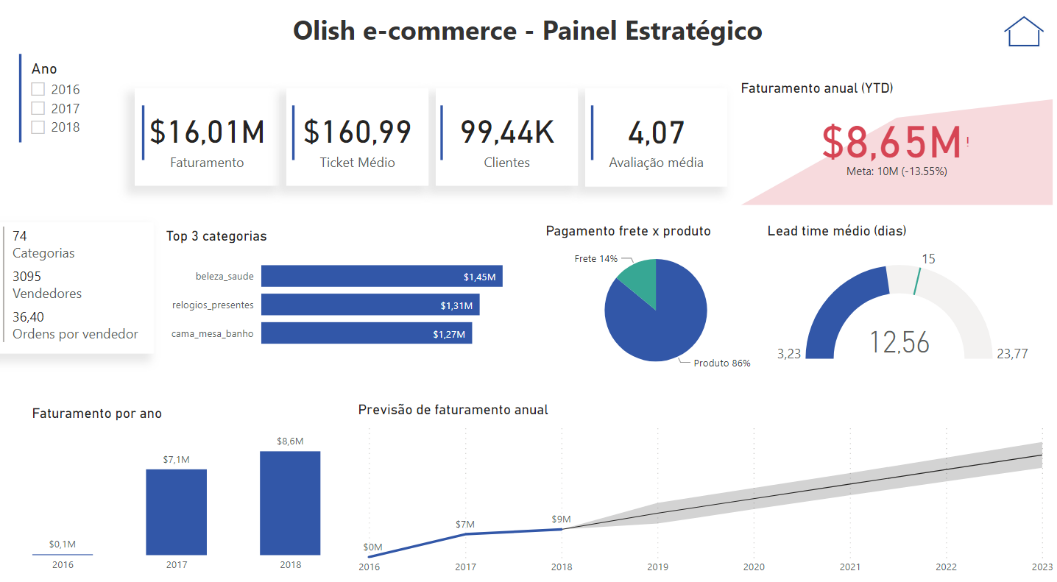


Figura 3. *Dashboard* estratégico

## 4.2. Painel Tático

O painel tático apresenta na porção superior o faturamento por mês em colunas e o ticket médio mensal no eixo secundário, oferecendo assim uma visão de se o aumento ou a diminuição do faturamento acompanha o aumento ou a diminuição do ticket médio.

Abaixo, é apresentada outra visão do faturamento, desta vez por categoria. Visualiza-se assim que, ao longo dos anos, algumas categorias aumentam seu faturamento enquanto outras diminuíram. Nota-se que as categorias de beleza/saúde e relógio/presentes aumentaram seu faturamento, em contraste com cama/mesa/banho e esporte/lazer que decresceram o faturamento ao longo dos anos. Similarmente, é apresentada uma tabela sumarizando o valor das 3 categorias de maior faturamento.

A avaliação média apresentada no painel estratégico retorna ao painel tático na forma de visualização em barras das três categorias pior avaliadas. De forma semelhante, a questão lead time aparece no painel tático sob aspecto de análise do tempo médio para envio, que diminui nos meses de menor faturamento.

Um *treemap* apresenta os três estados do Brasil com maior contributo para o faturamento: São Paulo, Rio de Janeiro e Minas Gerais, nessa ordem. Também é apresentado ao lado um mapa do Brasil com o valor médio do frete por estado. Nota-se que os estados da região Norte e a porção norte da região Nordeste possui a média de frete mais alta em comparação aos estados do Sul e Sudeste (principalmente São Paulo).

A nuvem de palavras foi elaborada com base no texto de avaliação dos clientes. Destacam-se algumas palavras: entrega; produto; não; chegou; adorei; recomendo; prazo; rápida; aguardando. Uma análise superficial desse conjunto de palavras parece demonstrar que a logística de entrega é uma área ambígua: recebe críticas negativas de não entrega, mas também elogios de rápida entrega.

Finalmente, são exibidas visualizações das formas de pagamento: o ticket médio do cartão de crédito o maior, seguido de boleto e cartão de débito e, por fim, voucher, que apresenta o menor valor. Considerando o volume de pagamentos, os pagamentos com cartão de crédito configuram o maior faturamento (12,5M), seguido do boleto.

Nesse painel estão disponíveis filtros de data, ano, trimestre (*quarter*) e mês. Os filtros localizam-se na lateral esquerda do painel.

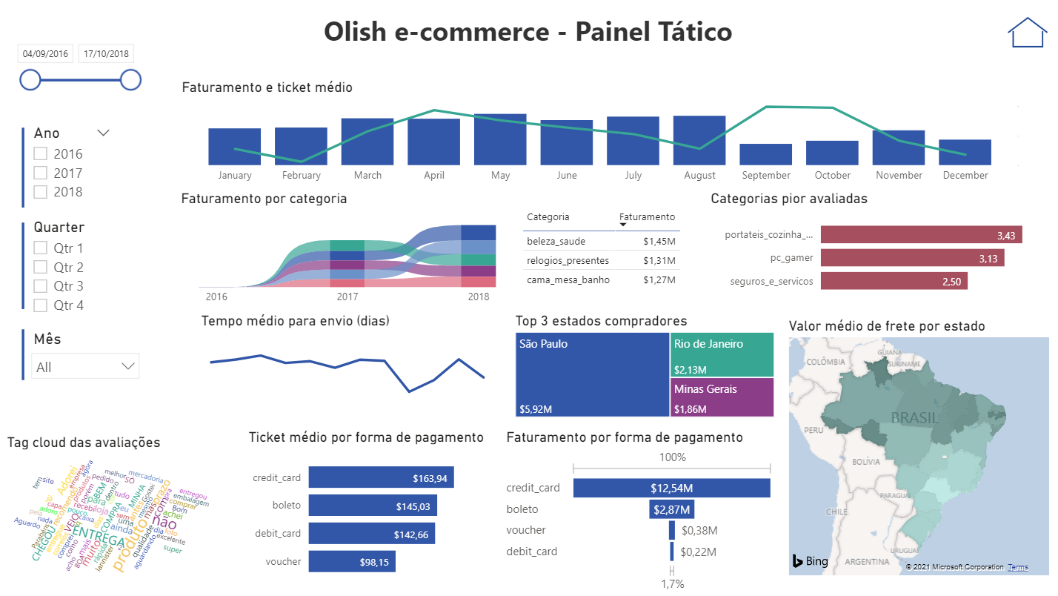


Figura 4. *Dashboard* tático

## 4.3. Painel Operacional

O painel operacional apresenta os dados focado no curto prazo. A ideia é que o responsável por enviar ou gerenciar quem despacha possa ter essas informações com maior facilidade.

Na porção superior há um *card* com o total de ordens que ainda não foram despachadas, isto é, ordens com status aprovado; cancelado; criado; faturado; processando e indisponível. Ao lado, há um filtro com todos os status.

O gráfico de Gauge (velocímetro) apresenta o tempo médio para envio, em dias. Isto é, o tempo decorrido entre a compra e a entrega do pedido para a transportadora.

Em seguida, são apresentadas três tabelas: um sumário de da quantidade de ordens ainda não despachadas por status; relação de datas para entrega e a quantidade de ordens correspondente; e por fim, a quantidade de ordens por estado e categoria.

Nesse painel estão disponíveis filtros de status (já mencionados anteriormente), e também de data. Além disso, pode-se filtrar o painel por ano, trimestre (*quarter*), mês, dia e estado.

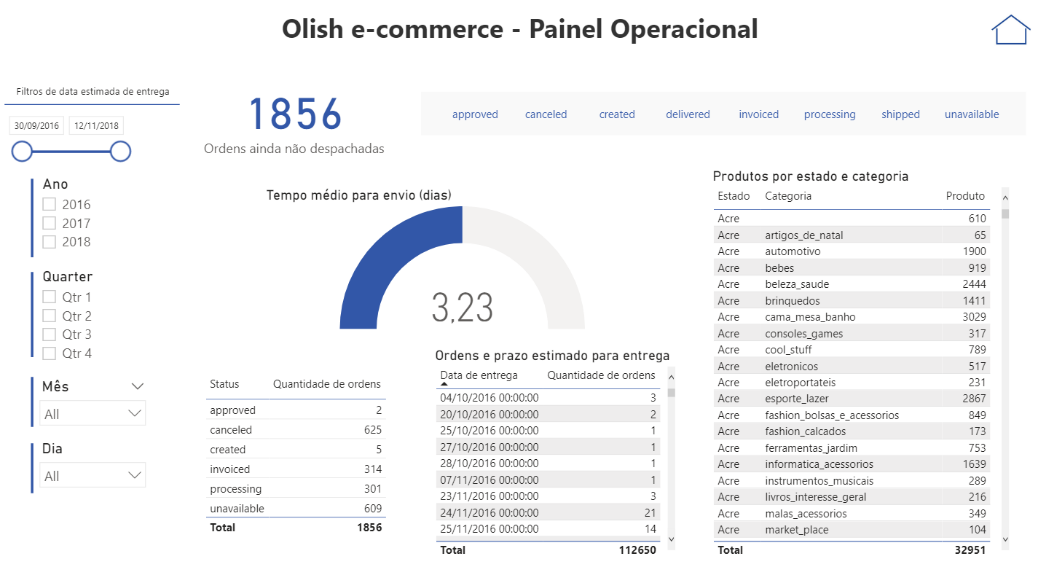


Figura 5. *Dashboard* operacional

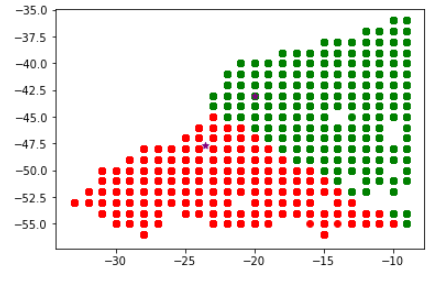
## 4.4. Análises de Machine Learning

A fim de complementar a apresentação de dados em *dashboard* no Power BI, foram realizadas duas análises utilizando a linguagem Python, no Jupiter Notebook.

Primeiramente, foi realizada uma análise de regra de associação para tentar identificar relacionamentos ou padrões frequentes na base de dados. Verificou-se que o ambos os índices de suporte e confiança são baixos, porque a maior parte das ordens registradas no *dataset* foram com realizadas com apenas um item ou mais de uma unidade do mesmo item. Assim, para essa análise, foram excluídos esses casos e chegou-se a uma regra de associação entre as categorias casa/conforto (antecessor) e cama/mesa/banho (consequente), tendo um lift de 1,8, o que indica associação (lift > 1,0). Dentro do universo de ordens contendo itens de mais de uma categoria, quem compra um item do setor de casa/conforto tipicamente compra um item de casa/mesa/banho.

Outra análise realizada foi o agrupamento K-Means, método de clusterização não-supervisionada. Para tal, estabeleceu-se uma problemática que seria a criação de dois centros de distribuição para melhor atender a logística de entrega e qual seria a melhor localização desses. Os parâmetros utilizados para tal, foram os dados de geolocalização dos clientes cadastrados na base de dados (latitude e longitude). A análise dos dados mostrou que não há a formação bem definida de dois *clusters*. Isso se deve à ampla distribuição dos clientes em todo o território.

Figura 6. Resultado da clusterização K-Means com base em dados de geolocalização



# 5. Registros de Homologação

Os testes de homologação foram realizados com consultas às tabelas SQL Server e os resultados confrontados com a visualização apresentada em Power BI. Os primeiros testes são com relação ao valor do faturamento e o ticket médio. A consulta em SQL apresenta os valores de 16008872,1199989 e 160,990266693473 enquanto no Power BI é esses valores são apresentados arredondados: 16M e 160,99.

Figura 7. *Query* em SQL Server

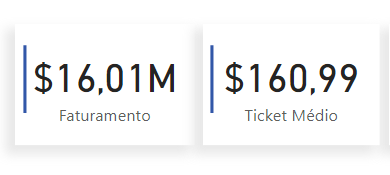
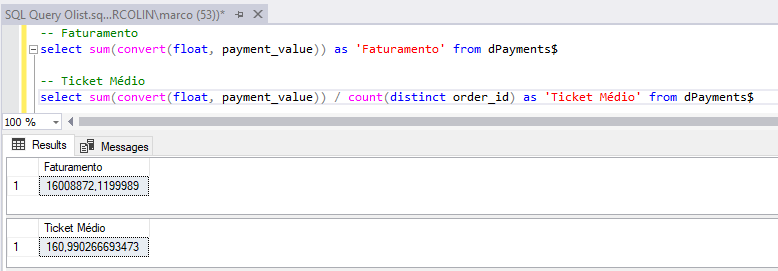


Figura 8. Valores de faturamento e ticket médio apresentado no painel estratégico em Power BI

Ainda, foi realizada uma consulta apresentando o ticket médio por forma de pagamento, consonante com o gráfico de barras apresentado no painel tático em Power BI.

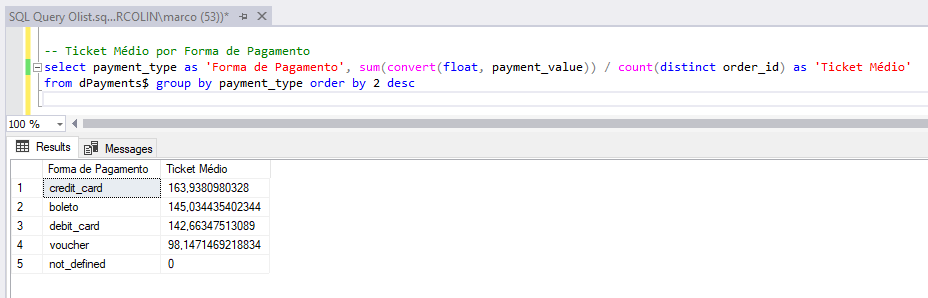


Figura 8. *Query* em SQL Server

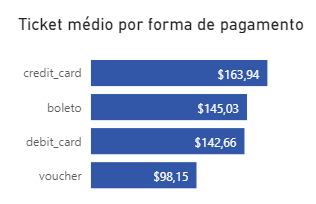


Figura 9. Valores de ticket médio por forma de pagamento

# 6. Conclusões

A análise dos dados de venda do *e-commerce* Olist permitiu análise em diferentes níveis. Em nível financeiro, nota-se que há uma crescente do faturamento, tendendo ao faturamento de $31M em 2023. Com relação às formas de pagamento utilizadas pelos clientes, o cartão de crédito assume o primeiro lugar respondendo por quatro vezes mais do que a forma de pagamento boleto (em segundo lugar). Considerando as taxas que esse tipo de pagamento pode incorrer para o vendedor e a lentidão para que o saldo chegue em conta, os pagamentos em boleto ou mesmo débito podem ser mais interessantes. Assim, tem-se uma oportunidade para incentivo ao cliente de uso dessas outras formas (por exemplo, 5% de desconto em compras à vista).

Sob perspectiva de *marketing*, há muitas oportunidades para aumento do faturamento ou atingimento de novos públicos. Observando-se o gráfico de colunas e linha intitulado “Faturamento e ticket médio”, nota-se que os meses de menor faturamento são setembro e outubro, em oposição, são esses também os meses de maior ticket médio. Esta é então uma oportunidade para uma campanha de *marketing* direcionado a um público similar ao desses compradores já que são os dispostos a pagar mais em uma compra (ticket médio) em relação aos dos outros meses. Outro ponto relevante que pode ser destacado nesse gráfico é que o mês de dezembro é um dos de menor faturamento. Notadamente dezembro é um mês de grande aquecimento de vendas devido às festas de final de ano em que ocorre compra de presentes para troca entre pessoas queridas, no entanto, os registros do Olist apresentam um faturamento menor do que a maioria dos demais meses.

A partir da aplicação de Machine Learning na base de dados, verificou-se que a maior parte dos usuários adquire apenas um item em seus pedidos. Levando-se em conta que os itens são ofertados em um *marketplace*, há a oportunidade de campanhas direcionadas, a fim de incentivar a compra de mais de um produto (descontos progressivos, por exemplo).

Finalmente, é possível extrair análises sob o ponto de vista de logística. Nota-se que outubro, novembro e dezembro são os meses maior com maior tempo médio para envio (em dias). Dadas as peculiaridades de uma atuação em *marketplace* como é o Olist, não seria possível sugerir a contratação de trabalhadores temporários para aliviar a pressão para envio nessa época. Esse constitui-se então um desafio sobre o qual os times de logística podem se debruçar em busca de alternativas.

Outro aspecto logístico levantado por esse trabalho foi a oportunidade o estabelecimento de dois centros de distribuição. Isto é, similar ao que o Mercado Livre faz na opção full – em que esse se responsabiliza pelo armazenamento, embalagem e envio dos produtos. No entanto, a análise por K-Means utilizando a geolocalização dos clientes não respondeu bem a essa proposição. Uma oportunidade para futuras análises seria utilizar apenas os pares de latitude e longitude duplicados (ou seja, locais com recorrência de compra). Ou ainda, tentar identificar *clusters* de clientes por outros critérios, como, por exemplo, valor do frete.

O trabalho aqui realizado pode aprofundar-se em análises relacionados a todos os pontos abordados aqui (financeiro, marketing e logística), seja refinando as métricas e os algoritmos utilizados quanto aplicando outros. Podem ser aplicadas técnicas de processamento de linguagem natural (NPL) a partir das avaliações dos clientes. Também é possível aprofundamento do entendimento do público do Olist a partir da utilização de outros algoritmos de clusterização, a fim de implementar sistemas de recomendação mais efetivos para os clientes. Ainda, é possível expandir a utilização dos painéis estratégico, tático e operacional ao utilizar em dados de negócio em tempo real.

# 7. Links

* Vídeo: <https://youtu.be/1EatVPZFZG4>
* Dashboard em Power BI e código fonte: <https://drive.google.com/drive/folders/11sSVzZXtO-tWmd89gNc4O6A-Y9QIHnt->
* Jupyter Notebook (Python 3): <https://github.com/isabelmarcolino/Olist_TCC_BIA_Notebooks>
* Base de dados Kaggle: <https://www.kaggle.com/olistbr/brazilian-ecommerce>

# REFERÊNCIAS

BALTAZAR, Gonçalo Alexandre Amaral. **Estratégia de crescimento da Worten: evolução para o marketplace.** Lisboa: 2020.

ROSA, João Roberto Conceição. ***Marketplace* no Brasil: desafios, vantagens e tendências deste modelo de negócio para empresas varejistas.** São Paulo: 2019.

TAKAHATA, Leonardo Eidy. **E-commerce no Marketplace.** Santa Catarina: 2017.