

## Análise Integrada: Dashboard Power BI + Predição de Lucro

- Introdução e Definição do Problema:

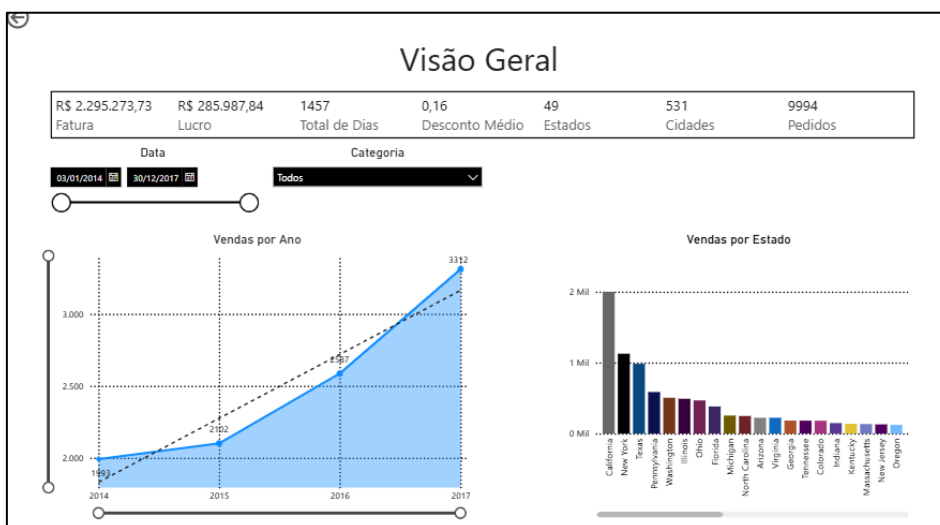
Este projeto tem como objetivo analisar dados de vendas de uma loja fictícia, utilizando tanto técnicas de visualização com Power BI quanto modelos de machine learning com Python (Jupyter Notebook) para prever o lucro (Profit). A base utilizada foi extraída do Superstore Dataset no Kaggle (<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final>), contendo informações sobre: pedidos, categorias, clientes, estados, datas, descontos, entre outros fatores. O foco principal da predição foi compreender como diferentes variáveis influenciam o lucro e como essas influências se expressam visualmente. A seguir irei mostrar Análise descritiva com Power BI, Predição de Lucro com Machine Learning e Conclusão.

- Análise Descritiva com Power BI:

Para esta análise eu dividi os dados em 6 páginas que são Visão Geral, Categoria, Sub-Categoria, Desconto, Localização e Transporte.

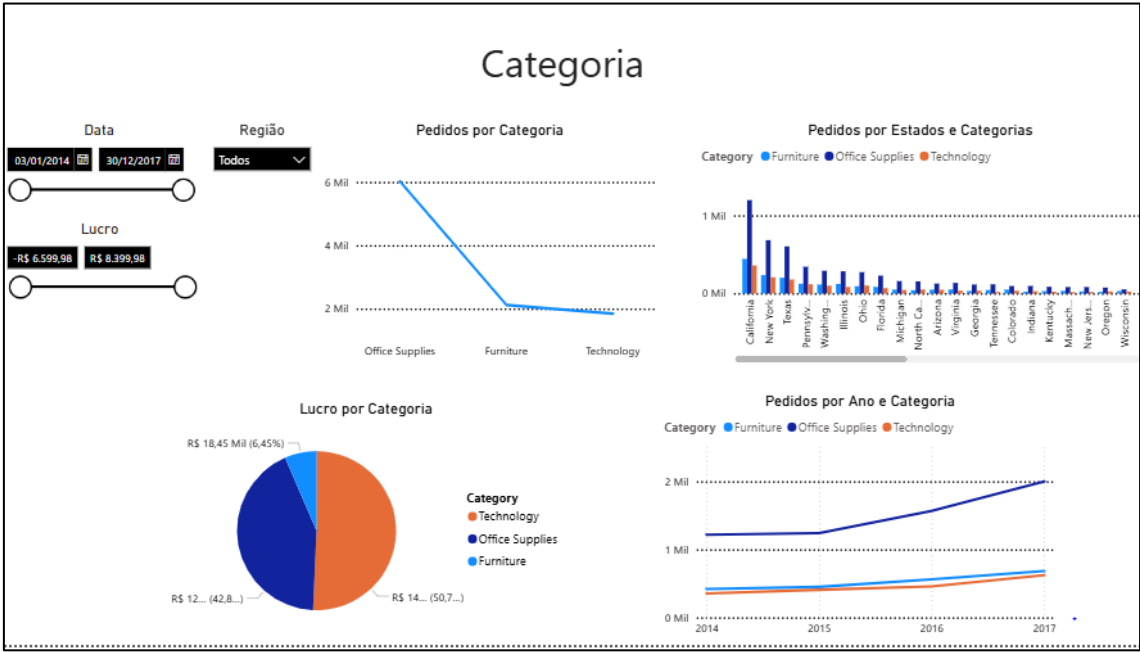
Visão Geral:

Como pode-se ver na imagem abaixo, a progressão de vendas por ano, como também fatura, lucro e outras informações importantes sobre o panorama geral que podemos analisar dos dados. Aqui destaquei junto com vendas por ano, vendas por estado e o estado da California é consideravelmente onde mais se vende dentro desses quatros anos. Nesta página aparece filtros de data e categoria para uma busca mais específica.



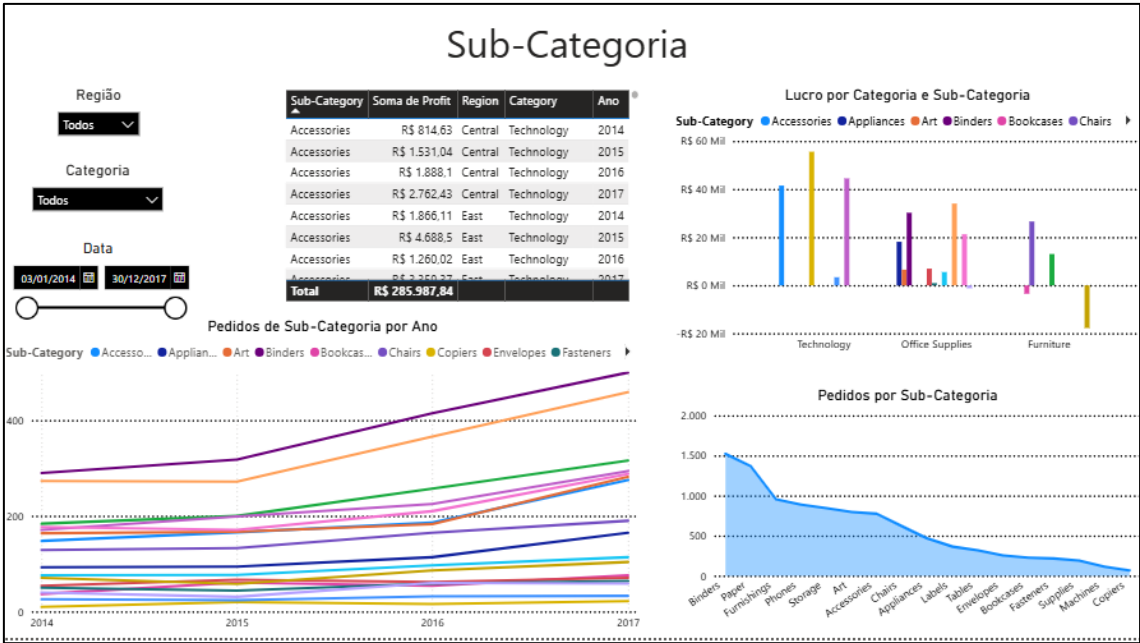
Categoria:

Nesta página de Categorias, rapidamente vemos que Office Supplies é a mais vendida em todos os anos e principalmente para o estado que mais compra desta loja. Porém no gráfico de lucros por categoria, technology é a mais lucrativa. Aqui nesta página também se pode filtrar a pesquisa com data, região e lucro.



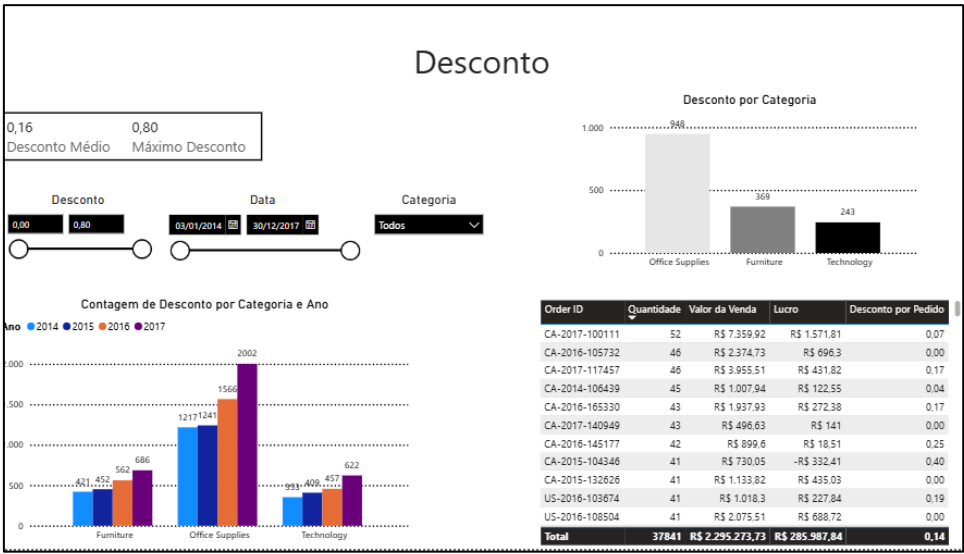
Sub-Categoria:

Aqui entramos numa ramificação da categoria, e podemos ver novamente a techonology sendo mais lucrativa. Tendo gráficos informando sobre a quantidade de pedidos, e pedidos por ano, junto com uma tabela para uma análise mais ampla acompanhado com os filtros de região, categoria e data.



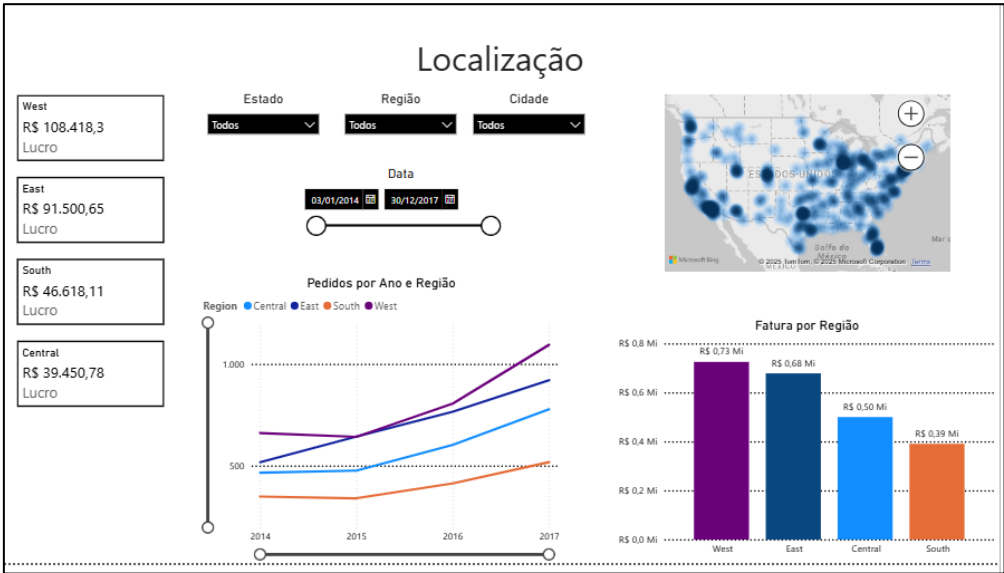
Desconto:

Aqui nesta página denominada desconto. Tem gráficos informando desconto por categorias por ano e desconto por categoria durante os quatro anos. Junto com uma tabela que correlaciona 5 colunas para uma análise mais ampla que será bem utilizada usando os filtros de desconto, data e categoria.



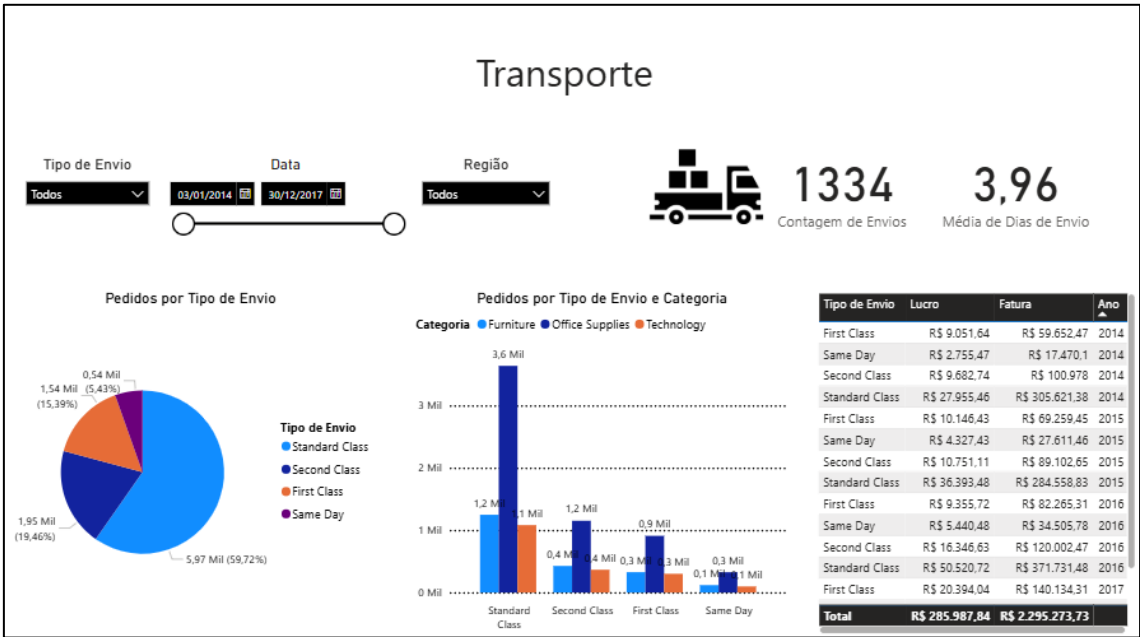
Localização:

Na penúltima página, temos a localização informando as vendas nas regiões. Tendo um gráfico de mapa, informando os lugares com as vendas, e podemos ver a parte central com menos vendas comparada as partes mais próximas da água, informando um aspecto geográfico já conhecido da região onde menos civis vivem nesta parte por conta das montanhas impedirem a passagem dos ventos, logo refletindo a falta de negociações nessas áreas.



Transporte:

Nesta última página, se informa sobre os meios de transporte com os gráficos e tabelas relacionando com outros temas já vistos anteriormente, porém sem um enfoque na distribuição.



- Predição de Lucro com Machine Learning:

A parte de modelagem foi realizada no Jupyter Notebook, com foco em prever a variável Profit a partir de diversas variáveis disponíveis, se divide em Pré-processamento dos Dados, Análise Exploratória Inicial, Modelagem Preditiva de Lucro e Interpretação dos gráficos do modelo XGBoost.

Pré-processamento dos Dados:

Leitura do Dataset Superstore.csv.

Correção do nome da coluna Profit.

Conversão correta das datas em Order Date e Ship Date.

Limpeza de dados em Profit (remoção de caracteres e conversão para float).

Criação de variáveis derivadas: Order\_Year, Order\_Month, Quarter, Delivery\_Days (tempo de entrega) e Discounted\_Value (valor total de desconto por venda).

Análise Exploratória Inicial:

Impressão de estrutura e valores ausentes.

Matriz de correlação entre variáveis numéricas: Sales, Quantity, Discount, Profit, Quarter, Discounted\_Value.

Identificou correlações relevantes com Profit, para orientar a modelagem.

A correlação negativa de Discount com Profit reforça a observação feita no Power BI.

As seguintes colunas 'City', 'State', 'Region', 'Category', 'Sub-Category', 'Segment' foram convertidas para predição usando a técnica One-Hot Encoding.

## Modelagem Preditiva de Lucro:

Foram aplicados diferentes algoritmos para prever Profit, comparando desempenho:

### a) Regressão Linear Simples

Modelo básico com avaliação por MSE: 80500.09 e  $R^2$ : 0.69.

Usando as colunas 'City', 'State', 'Region', 'Category', 'Sub-Category', 'Segment', 'Order Date', 'Sales', 'Quantity', 'Discount', 'Profit', 'Quarter', 'Discounted\_Value'.

### b) LassoCV (Regularização L1)

Para seleção de variáveis relevantes automaticamente.

Sendo as mais importantes Sales, Sub-Category\_Copiers da Categoria mais lucrativa technology e Sub-Category\_Binders da Categoria Office Supplies que é a mais vendida, para a predição de Lucro (Profit).

### c) Regressão Polinomial

RMSE: 99.43 e  $R^2$ : 0.5705.

Usando as colunas 'Sales', 'Discount', 'Quantity' para a predição de Profit (Lucro). Onde teve gráfico informando Previsão vs. Valor real e Distribuição dos Resíduos.

### d) Random Forest Regressor

RMSE: 81.17 e  $R^2$ : 0.7138.

Desta vez usando 'Sub-Category\_Copiers' da Categoria mais lucrativa technology e 'Sub-Category\_Binders' da Categoria Office Supplies que é a mais vendida, mostrado pelo LassoCV (Regularização L1).

Usando as colunas 'Sales', 'Discount', 'Quantity', 'Sub-Category\_Copiers', 'Sub-Category\_Binders', 'Discounted\_Value', para prever Profit. Tendo os gráficos Importância das Variáveis - Random Forest, Distribuição dos Resíduos e PDPs para analisar o modelo. Tendo uma melhora de performance bem perceptível.

#### d) XGBoost Regressor

81.17 e  $R^2$ : 0.7138 (mesmo que o anterior).

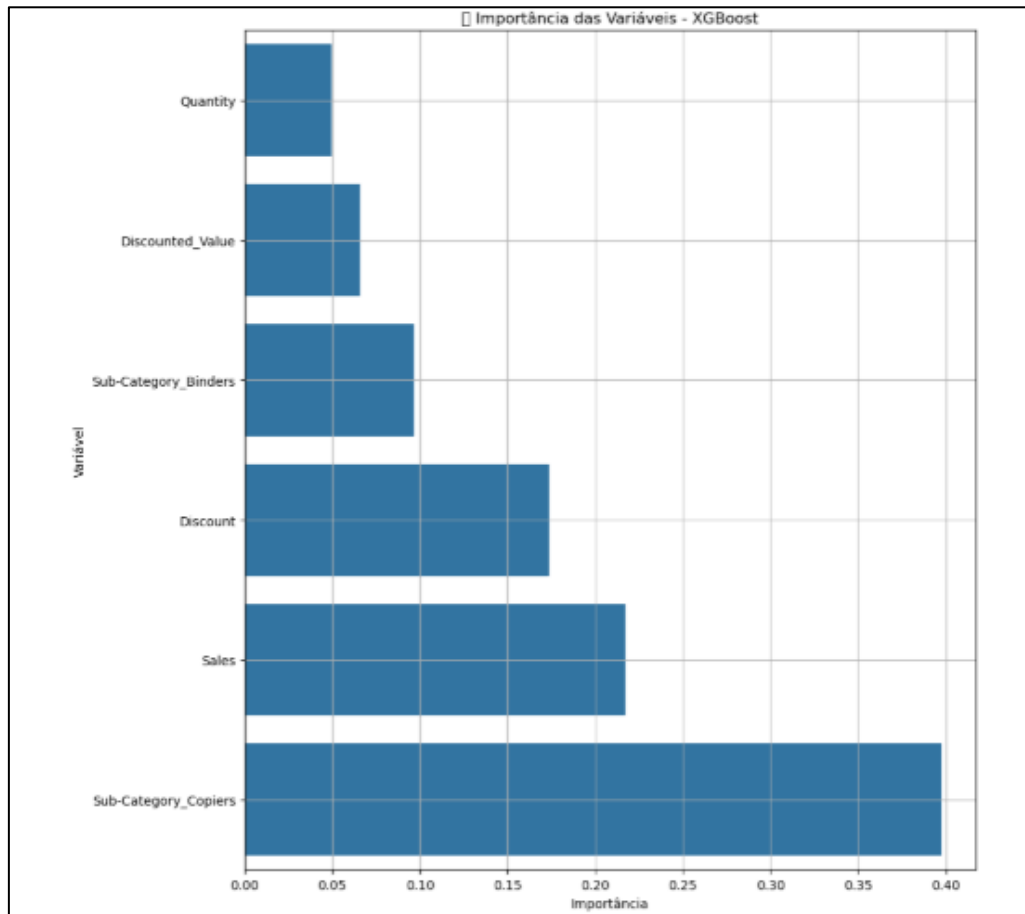
Usando as colunas 'Sales', 'Discount', 'Quantity', 'Sub-Category\_Copiers', 'Sub-Category\_Binders', 'Discounted\_Value' para a predição. Tendo gráficos PDPs, Importância das Variáveis – XGBoost, Gráfico SHAP e Distribuição dos Resíduos para sua análise.

Aqui nesta parte foram feitos dois hiperparâmetros, um menos amplo que deu RMSE: 142.8304522295174, e outro mais amplo que seu resultado foi 81.17 e  $R^2$ : 0.7138 o mesmo sem o ajuste de hiperparâmetros.

Gráficos do modelo XGBoost:

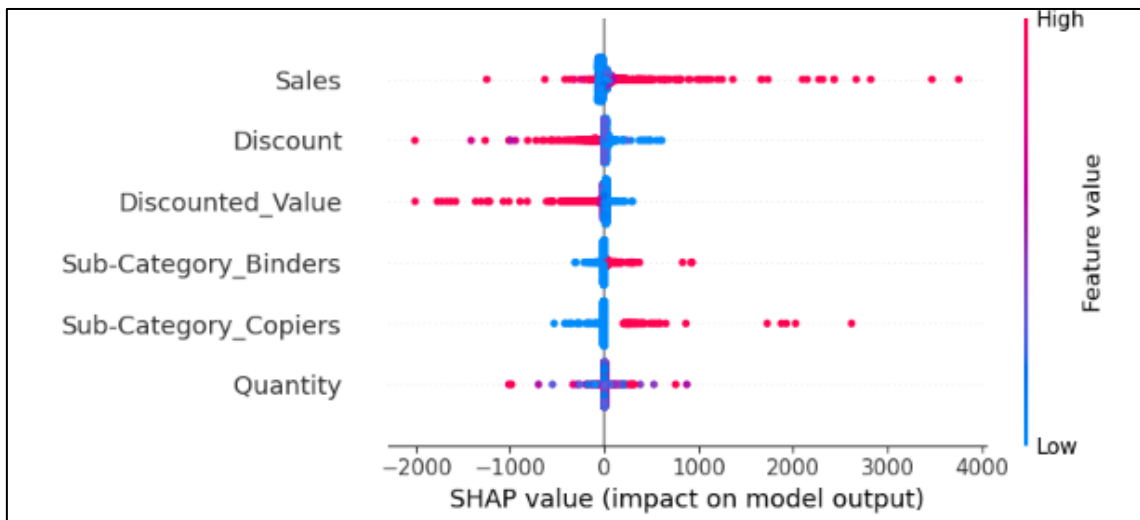
Escolhi este modelo pois fiz por último e seria mais simples visualizar seus gráficos. E não teria mudança significativa dos gráficos para o modelo anterior Random Forest.

#### Importância das Variáveis



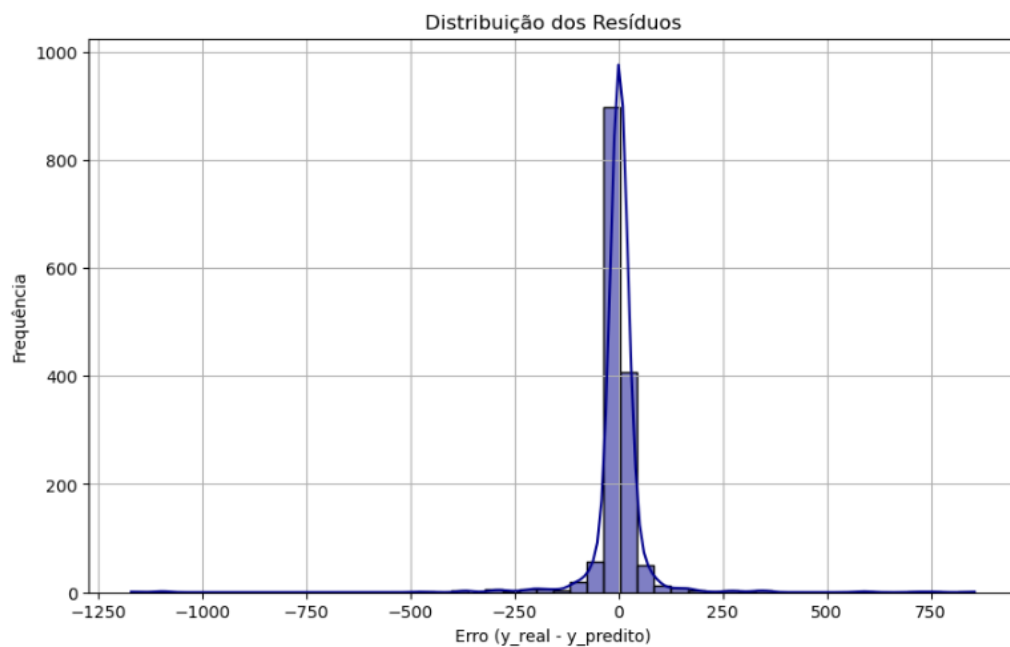
## SHAP

Mostra como cada variável afetou individualmente a previsão.



## Distribuição de Resíduos

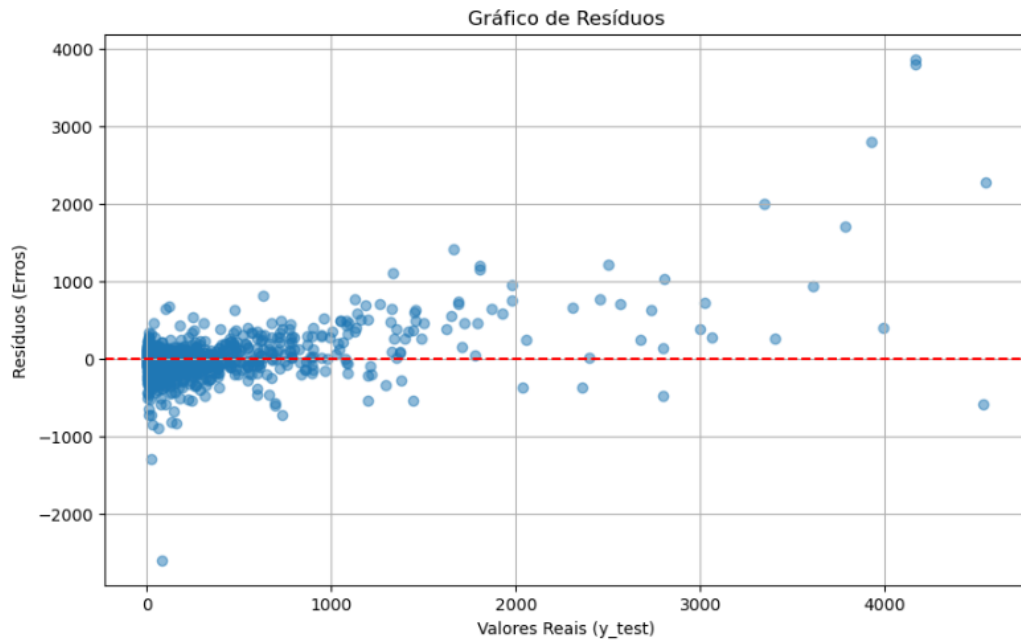
Frequência do resultado real menos o previsto.





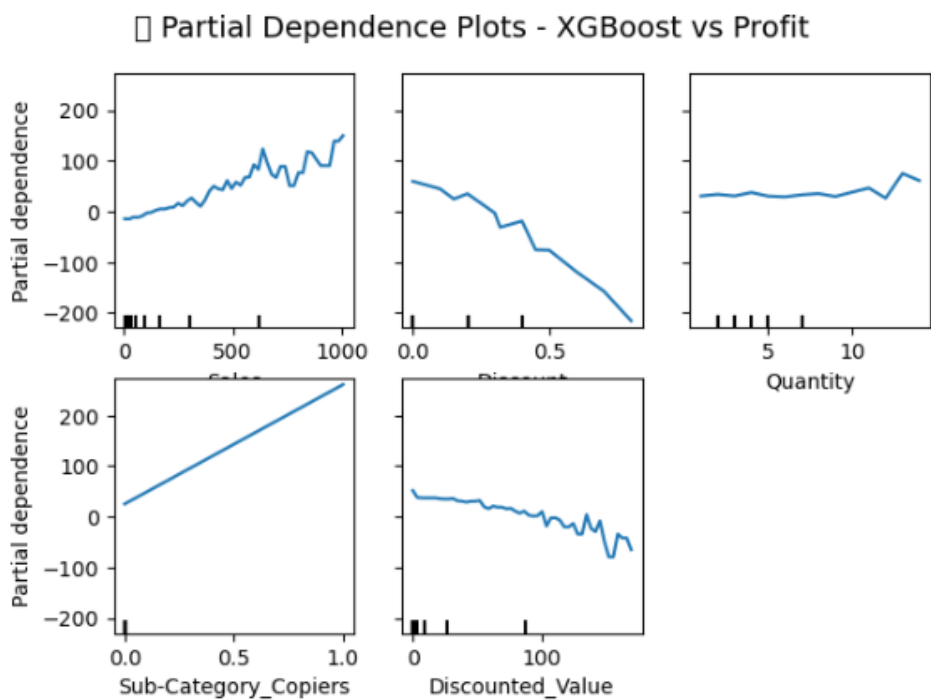
## Gráfico de Resíduos

Mostra a dispersão dos resíduos e os valores reais.



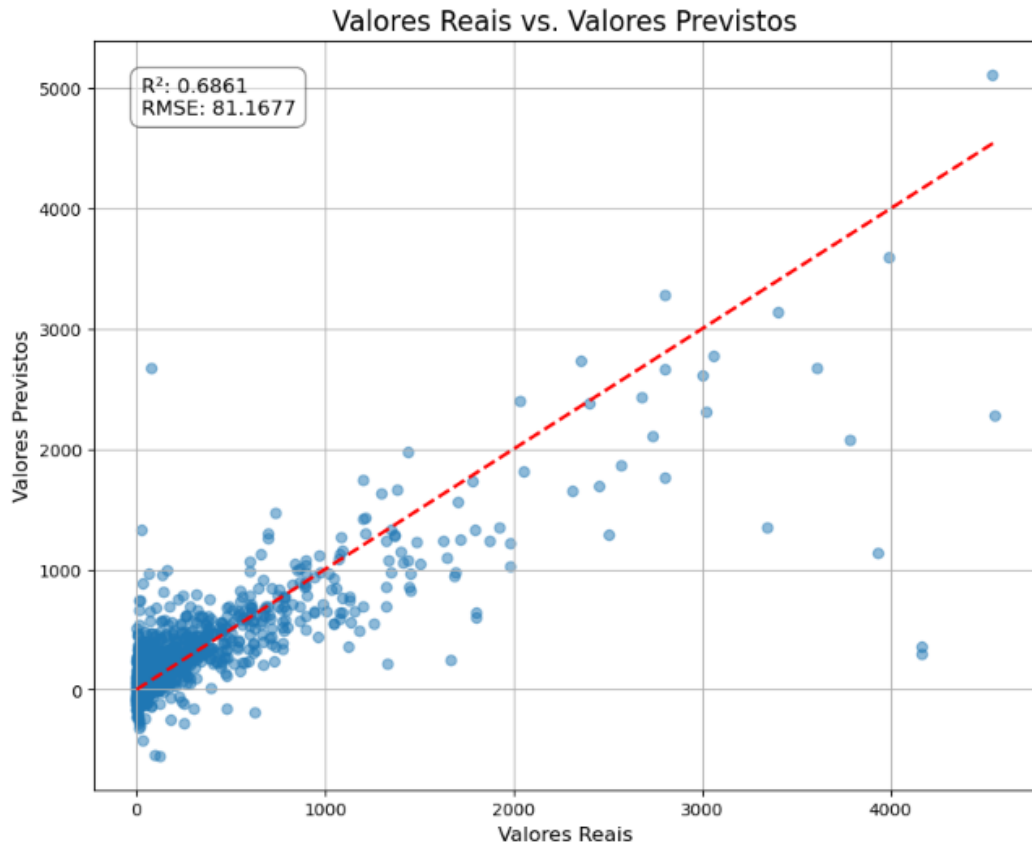
## PDP's

Ajudará a entender relações não lineares. Estão sendo divididos por Sales, Discount, Quantity, Sub-Category\_Copiers e Discount\_Value, respectivamente.



## Visão Geral do Modelo

Mostra como o modelo consegue generalizar os valores.



- Conclusão:

Este projeto mostrou como a combinação de análise visual interativa e modelagem preditiva pode gerar insights poderosos. Power BI revelou padrões importantes com clareza e interatividade. Modelos preditivos identificaram variáveis-chave e permitiram simular cenários futuros. Integração entre métodos garantiu consistência e maior valor analítico.

### Insights:

A empresa vende mais com o passar do tempo.

A categoria mais vendida é Office Supplies, porém o maior lucro vai para a categoria technology.

As sub-categorias Copiers e binders estão correlacionadas de forma positiva com o lucro.

A empresa deve avaliar os descontos acima de 35% pois afeta diretamente a margem de lucro.

A Sub-Categoria bookcases e Tables da categoria Furniture, revelam saldo negativo, causando prejuízo.

A zona oeste é a região que tem mais lucros e é onde fica o estado da Califórnia com mais lucros também.

O envio Standard Class nos quatro anos foram as mais lucrativas e as mais requisitadas.

Standard Class foi muito acionado com a categoria Office Supplies.