

# Estruturação de Dados e Análise de Vendas da Super Store com ETL e Comparação com Concorrentes

## Equipe:

Ester Linderos Saravia

## Data:

28/11/2024

## Ficha Técnica:





ETL

# Introdução



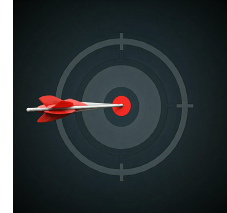
ETL

Este projeto tem como objetivo otimizar a gestão de dados da Super Store, uma empresa do setor varejista. O objetivo foi transformar esses dados em um formato estruturado, facilitando o processo de tomada de decisão, análises de desempenho e comparações com concorrentes.

A solução proposta é um sistema de ETL (Extração, Transformação e Carga) que organiza os dados em tabelas de fatos e dimensões, resultando em um Data Warehouse bem estruturado. Este sistema foi desenvolvido utilizando **Python**, **SQLite3**, **Power BI** e **BeautifulSoup**, com migração a **PostgreSQL**.



# Objetivo do Projeto



- Estruturar dados da Super Store em um Data Warehouse funcional.
- Criar um pipeline de ETL para integrar dados de vendas da Super Store com informações de concorrentes.
- Permitir a análise e visualização de dados em tempo real utilizando Power BI.
- Comparar o desempenho da Super Store com grandes concorrentes no setor varejista.



# Ferramentas Utilizadas

- **Python com SQLite3:**

Utilizado para implementação do banco de dados, desenvolvimento do sistema ETL, e manipulação de dados. SQLite3 foi escolhido como solução local para gerenciar as tabelas de fatos e dimensões.

- **Power BI:**

Utilizado para a criação de dashboards interativos, permitindo a visualização dinâmica dos dados de vendas, tendências por região, desempenho por produto, entre outros.

- **BeautifulSoup (Web Scraping):**

Para coletar dados de concorrentes (como empresas multinacionais) de uma página da Wikipédia, visando comparações de desempenho.

- **PostgreSQL:**

Adotado como banco de dados relacional para armazenar os dados migrados do SQLite3, garantindo maior robustez, escalabilidade e suporte a consultas complexas. Facilitou a integração com o Power BI e permitiu a centralização de dados estruturados em um ambiente mais adequado para operações corporativas.

# Procedimentos Realizados



## 1. Criação das Tabelas de Fatos e Dimensões

- **Tabelas de Fatos:**
  - **fact\_vendas:** Contém os dados principais de vendas, como quantidade vendida, receita, lucro e IDs de referência.
- **Tabelas de Dimensões:**
  - **dim\_cliente:** Informações sobre os clientes.
  - **dim\_produto:** Dados detalhados sobre os produtos.
  - **dim\_localidade:** Dados sobre as localizações geográficas das vendas.

## 2. Importação e Tratamento de Dados

- Importação dos dados para o banco de dados.
- Tratamento de valores nulos e duplicados.
- Padronização de variáveis categóricas e numéricas para garantir a integridade dos dados.

### **3. Web Scraping com BeautifulSoup**

- Extração de informações sobre concorrentes do setor varejista de uma tabela na Wikipédia.

Os dados foram integrados ao banco da Super Store para comparações de desempenho.

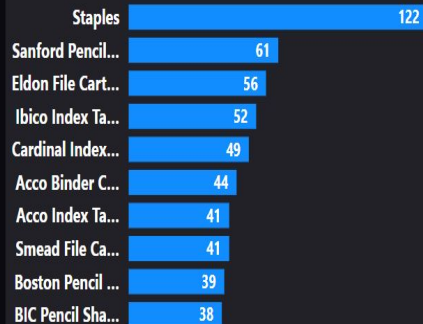
### **4. Análise Exploratória de Dados**

- Realização de análise de vendas, lucro e volume de pedidos no Power BI.
- Segmentação por categoria, região e data.
- Comparação de desempenho com concorrentes usando as informações extraídas via web scraping.

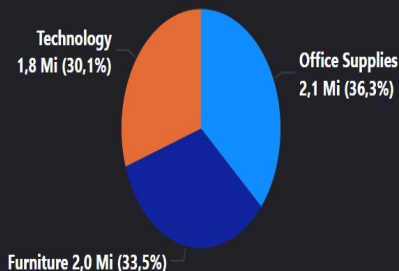
### **5. Agendamento de Atualizações de Tabelas**

- Planejamento para atualização periódica dos dados no banco, garantindo que os dados sejam atualizados conforme novas transações.

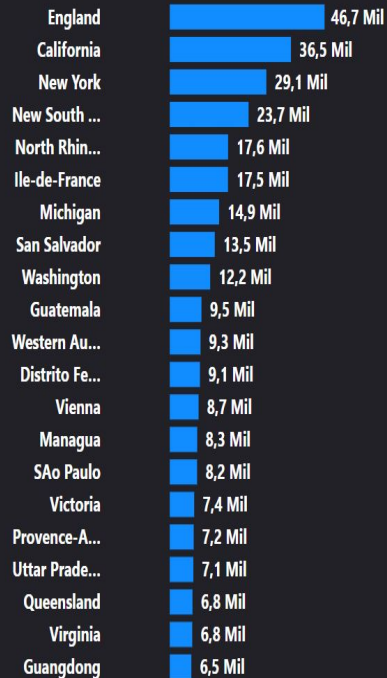
## Volume de Pedidos por Produto



## Vendas por Categoria



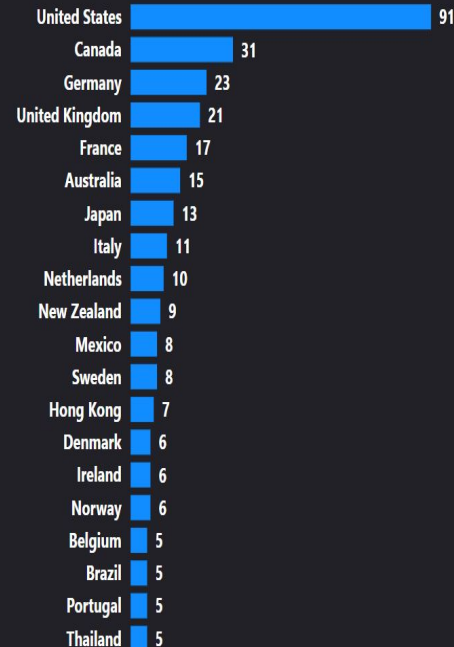
## Lucro por Estado



## Quantidade de Empresas Presentes na Super Store



## Empresas Presentes na Super Store por Região/Pais



# Conclusão

O projeto resultou em um sistema de gestão de dados robusto e bem estruturado para a Super Store, centralizando informações anteriormente dispersas. O Data Warehouse criado possibilita análises de vendas em tempo real, reduzindo significativamente a necessidade de intervenções manuais no processamento de dados. Além disso, o uso de ferramentas como Python e Power BI assegurou um pipeline de ETL eficiente, proporcionando visualizações dinâmicas e insights acessíveis.

Embora a comparação detalhada com concorrentes tenha sido limitada pela falta de dados externos padronizados, o projeto estabeleceu uma base sólida para futuras integrações e análises. O Power BI revelou-se uma ferramenta poderosa para explorar tendências, segmentações e desempenho por produto, oferecendo suporte crucial à tomada de decisão estratégica.







# Recomendações e Próximos Passos

## 1. Ampliar a Comparação com Concorrentes

- Investir em fontes de dados externas mais completas e padronizadas para melhorar a análise comparativa. A integração com APIs ou dados de mercado pode oferecer uma visão mais detalhada do setor.

## 2. Expansão de Dashboards e Métricas

- Incorporar métricas avançadas nos dashboards, como análise de satisfação do cliente, ticket médio e margens de lucro por categoria.
- Explorar dados de feedback de clientes para identificar oportunidades de melhoria.

## 3. Automação Completa do Pipeline

- Implementar scripts e agendamentos automatizados para atualizar as tabelas e dashboards regularmente, garantindo análises sempre baseadas em dados atualizados.

## 1. **Implementação de Modelos Preditivos**

- Desenvolver modelos para previsão de vendas, segmentação de clientes e identificação de churn, utilizando aprendizado de máquina. Isso permitirá ações mais proativas e baseadas em dados.

## 2. **Treinamento e Disseminação Interna**

- Capacitar equipes internas para usar os dashboards de forma autônoma e explorar os dados para diferentes áreas estratégicas, como marketing, logística e vendas.

## 3. **Integração com Novos Sistemas**

- Integrar o Data Warehouse com outros sistemas de gestão, como CRM ou ERP, para uma visão ainda mais abrangente da operação.

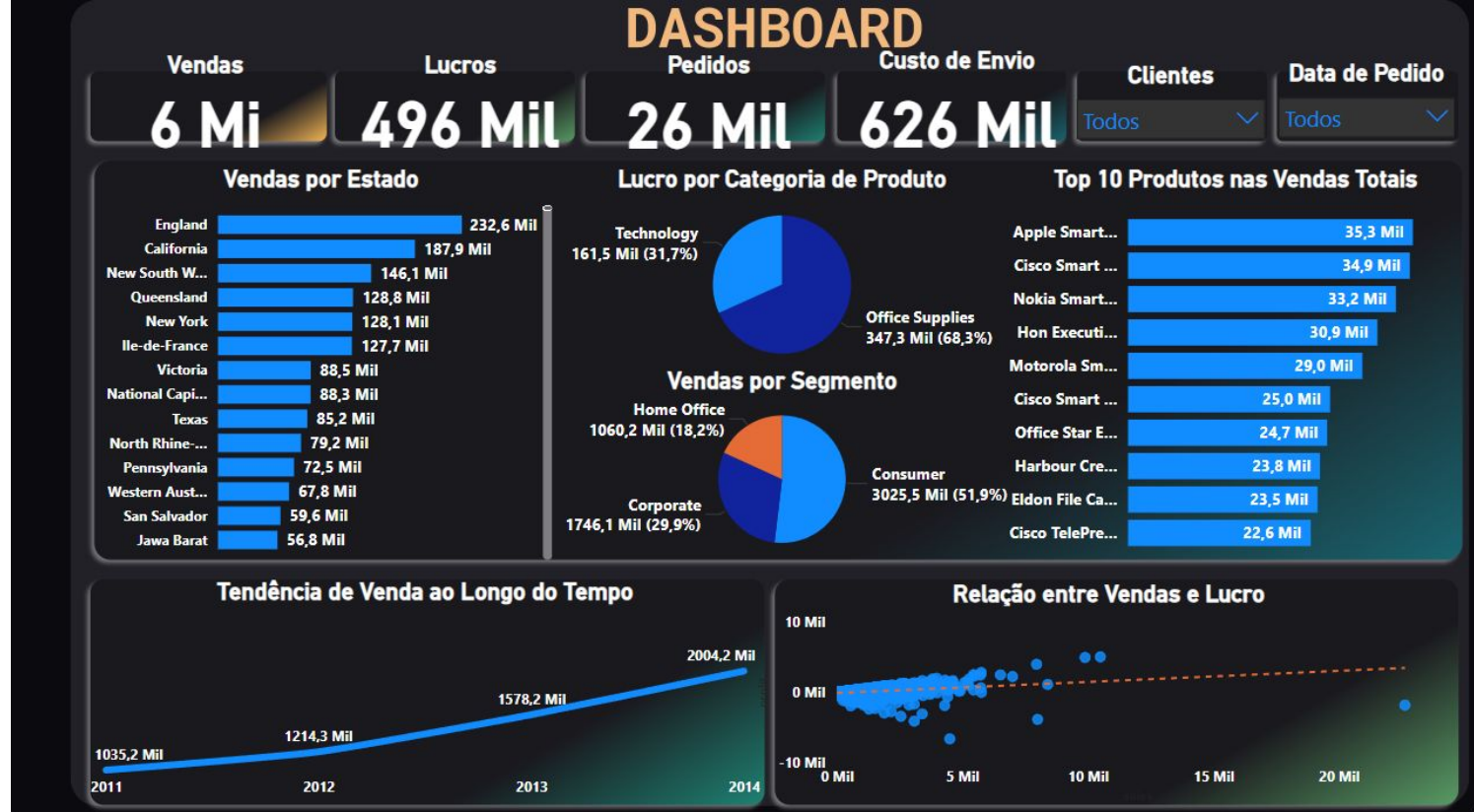


## Considerações Finais

Este projeto é um marco no processo de transformação digital da Super Store. Ele demonstra o impacto positivo da centralização e estruturação dos dados para a análise de desempenho e planejamento estratégico. A base estabelecida permite avanços significativos em termos de eficiência, integração e previsibilidade, criando oportunidades de expansão e competitividade no mercado varejista. O próximo passo será explorar ainda mais o potencial dos dados para fortalecer a posição da Super Store no setor.

# Gráficos Recomendados para o Dashboard

1. **Gráfico de Barras ou Map (Vendas por Estado):** Para analisar o desempenho de vendas por localização.
2. **Gráfico de Linha (Tendência de Vendas ao Longo do Tempo):** Para acompanhar as variações de vendas ao longo dos meses.
3. **Gráfico de barra (Participação de Produto nas Vendas):** Para visualizar quais produtos têm maior participação nas vendas totais.
4. **Gráfico de Dispersão (Relação entre Vendas e Lucro):** Para analisar o impacto das vendas no lucro.
5. **Gráfico de Comparação (Vendas da Super Store vs Concorrentes):** Para comparar o desempenho da Super Store com os concorrentes.



Com essa estrutura, a Super Store agora possui um sistema organizado e eficiente, facilitando consultas e gerando insights acionáveis sobre o mercado e o comportamento dos consumidores.

# DASHBOARD

