
Análise de Dados: Vendas Internacionais

Débora Ferreira, Israel Argôlo e Edilberto

Silva FACSSENAC-DF, Brasília-DF

deboranunesferreira@hotmail.com; israelargolo0809@gmail.com;

prof.edilberto.silva@gmail.com

Abstract: This article analyzes global data from the fictional company Inter Tech to assist in strategic decision-making. The company sells products in the categories of 'Furniture', 'Technology', and 'Supplies' across various countries. The research aims to identify the country with the highest sales volume and provide detailed insights through a Business Intelligence (BI) model using ETL, Data Warehouse, and interactive dashboards in Power BI. Utilizing the CRISP-DM model, the study covers all phases of the data mining lifecycle. Results highlight the US as the sales leader, with technology being the top-selling category, indicating areas for future investment.

Resumo. Este artigo analisa dados globais da fictícia empresa Inter Tech para auxiliar na tomada de decisões estratégicas. A empresa vende produtos nas categorias 'Móveis', 'Tecnologia' e 'Suprimentos' em diversos países. A pesquisa visa identificar o país com maior volume de vendas e fornecer insights detalhados através de um modelo de Business Intelligence (BI) usando ETL, Data Warehouse e dashboards interativos no Power BI. Utilizando o modelo CRISP-DM, o estudo abrange todas as fases do ciclo de vida da mineração de dados. Resultados destacam os EUA como líder em vendas, com a tecnologia sendo a categoria mais vendida, indicando áreas para investimentos futuros.

Palavras-chave: Análise de Dados, Inter Tech, ETL, Business Intelligence, Data Warehouse, Power BI, Vendas Globais, Estratégia de Decisão, Modelagem Multidimensional, Dashboards Interativos, Mineração de Dados, CRISP-DM.

1. Introdução

Este artigo aborda uma análise de dados global destinada a auxiliar na tomada de decisões estratégicas. Na Inter Tech, que uma empresa fictícia, os produtos são vendidos categorizados por diferentes tipos e subtipos. Por exemplo, na categoria 'Tecnologia', há diversas subcategorias, como 'Celulares'. A empresa opera com três categorias principais: 'Móveis', 'Tecnologia' e 'Suprimentos', distribuindo suas vendas em vários países ao redor do mundo.

A empresa em questão possui vastos conjuntos de dados oriundos de suas operações internacionais. No entanto, enfrenta dificuldades na análise eficaz desses dados para tomada de decisões estratégicas. Para enfrentar os desafios mencionados, este estudo se propõe a responder a seguinte pergunta de negócio:

- Qual país registrou o maior volume de vendas globalmente?

O objetivo principal deste estudo é fornecer insights valiosos para a empresa por meio de uma análise detalhada de seus dados globais de vendas. Para alcançar esse objetivo, foram estabelecidos os seguintes objetivos específicos:

1. Obtenção e tratamento da massa de dados necessária para responder ao objetivo geral.
2. Realização do entendimento dos dados.
3. Criação de uma área de preparação (Staging Area) para importação de todos os dados necessários.
4. Importação dos dados da área de preparação para o Modelo Multidimensional (MMD).

5. Desenvolvimento do Modelo Multidimensional (MMD).
6. Criação de Dashboards para visualização dos dados.
7. Validação das informações obtidas e ajustes em conjunto com a área demandante/cliente.

No contexto da análise de dados para tomada de decisões estratégicas, o processo de Business Intelligence (BI) desempenha um papel crucial. Esse processo é iniciado com a coleta, limpeza e transformação dos dados, conhecido como ETL (Extract, Transform, Load). O ETL é composto por um conjunto de ferramentas especializadas que automatizam essas etapas, permitindo a integração de dados de diversas fontes.

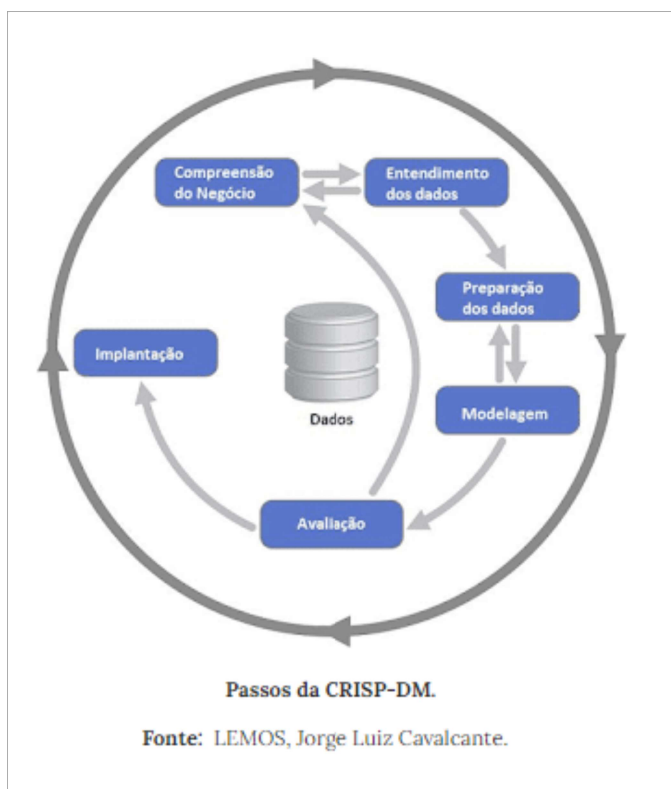
Uma vez que os dados foram coletados e preparados, eles são carregados em um Data Warehouse (Armazém de Dados). O DW funciona como um repositório central ou depósito de dados, onde os dados são associados e armazenados de forma organizada. Isso permite que os usuários do BI tenham acesso a uma fonte única de informações confiáveis e consistentes, facilitando a análise e geração de insights.

Dashboards de tomada de decisão são interfaces visuais que apresentam dados de forma simples e intuitiva para facilitar a interpretação e decisões rápidas. Eles destacam métricas-chave e tendências, utilizando ferramentas como PowerBI para criar visualizações interativas.

No decorrer do projeto, será utilizado um banco de dados relacional chamado MySQL para estabelecer os relacionamentos entre as tabelas do conjunto de dados e realizar a modelagem multidimensional. Posteriormente, o Power BI será empregado para apresentar visualmente os dados de maneira compreensível e interativa.

A compreensão do negócio e dos dados é crucial para a análise de dados e modelagem. É necessário se familiarizar com o contexto e objetivos do negócio, coletar, limpar e explorar os dados para extrair informações significativas. A preparação dos dados inclui processamento, limpeza e transformação para serem usados nos modelos analíticos. [1]

A modelagem utiliza os dados preparados para construir modelos analíticos que geram conhecimento e previsões relevantes. A etapa de avaliação é crítica, avaliando a eficácia e precisão dos modelos. A implementação é a fase final, onde o conhecimento e previsões são colocados em prática para gerar valor para o negócio, integrando modelos, criando dashboards, automatizando processos, entre outras ações.



2. CRISP-DM

O CRISP-DM (Cross-Industry Standard Process for Data Mining) é um modelo de processo amplamente utilizado para guiar projetos de mineração de dados. Ele foi desenvolvido para fornecer uma estrutura robusta e flexível para as diferentes etapas do ciclo de vida de um projeto de data mining. O modelo CRISP-DM é composto por seis fases principais: compreensão do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação.

3. Resultados

3.1 Compreensão do Negócio

A compreensão abrangente do negócio começa pela exploração dos dados, analisando-se a estrutura, distribuição e padrões iniciais. No contexto das vendas globais da empresa, essa análise envolve a identificação dos países e regiões com maior volume de vendas, além da investigação das categorias e subcategorias mais populares. Esses insights iniciais são fundamentais para entender as tendências do mercado e identificar oportunidades de crescimento.

O objetivo principal deste trabalho é, em primeiro lugar, compreender o panorama das vendas, identificando os segmentos, categorias e subcategorias mais comercializados em cada país e região. Ao entender esses aspectos, podemos também analisar quais países e produtos estão registrando vendas mais baixas e investigar as razões por trás dessas discrepâncias, porque o país X vende mais do que o país Y. Com base nessa análise, pretendemos desenvolver estratégias que promovam equilíbrio nas vendas e, ao mesmo tempo, ampliem a prospecção de clientes.

Com base na análise dos dados, é possível desenvolver estratégias sólidas para otimizar as vendas globais da empresa. Isso inclui a comparação de desempenho entre países, a identificação de áreas de oportunidade e desafios, e o desenvolvimento de planos de ação específicos. Ao implementar essas estratégias e monitorar continuamente o desempenho, a empresa pode se posicionar de forma mais competitiva no mercado global e alcançar seus objetivos de crescimento.

A empresa é fictícia e os dados são hipotéticos, no ramo de varejo com foco em produtos de tecnologia, móveis e suprimentos.

3.2 Entendimento de Dados

Os dados utilizados neste estudo foram extraídos do blog da Data Science Academy[2] na página sobre datasets públicos em abril de 2024. Especificamente, foram obtidos um conjunto de dados sobre vendas globais. O arquivo "vendas_internacionais.xlsx" contém 51.291 linhas e 14 colunas, fornecendo informações detalhadas sobre as vendas em vários países, categorizadas em três tipos principais de produtos: tecnologia, móveis e suprimentos.

Neste artigo, focaremos na compreensão dos dados disponíveis para análise. Os dados, de natureza internacional, abrangem os seguintes campos:

Data_Pedido, ID_Cliente, Segmento, Região, País, Product ID, Categoria, Subcategoria, Total_Vendas, Quantidade, Desconto, Lucro e Prioridade. Eles foram coletados e preparados meticulosamente para análise, visando responder a pergunta de negócio estabelecida anteriormente e garantir a relevância dos resultados. Para este trabalho serão utilizados somente os campos descritos abaixo:

Campo	Tipo	Descrição
ID_PEDIDO	Numérico	Número identificador único do pedido
DATA_PEDIDO	Data	Data em que o pedido foi realizado
ID_CLIENTE	Numérico	Número identificador único do cliente
SEGMENTO	Caracter	Segmento do produto 1 – Consumidor 2 – Corporativo 3 – Home Office
REGIÃO	Caracter	Região de venda, representada pelo estado do país
PAÍS	Caracter	País onde a venda ocorreu
PRODUCT_ID	Numérico	Número identificador único do produto
CATEGORIA	Caracter	Categoria do produto 1 – Tecnologia 2 – Móveis 3 – Suprimentos
SUBCATEGORIA	Caracter	Subcategoria do produto Ex: 1 – Accessories 2 – Phones 3 – Tables
TOTAL_VENDAS	Numérico	Valor total das vendas em moeda local
QUANTIDADE	Numérico	Quantidade de produtos vendidos
DESCONTO	Numérico	Valor do desconto aplicado na venda
LUCRO	Numérico	Lucro obtido com a venda
PRIORIDADE	Caracter	Prioridade atribuída à venda 1 – Alto 2 – Baixo 3 – Médio 4 – Crítico

3.3 Preparação de Dados

A preparação de dados é uma etapa essencial no processo de Business Intelligence, onde os dados são tratados antes de serem submetidos à análise. No contexto da solução estudada, a preparação de dados contempla as seguintes fases.

3.3.1 Elaboração do Staging Area

Primeiro, uma staging area foi criada no MySQL para servir como área temporária de armazenamento e processamento dos dados. Em seguida, os dados foram importados de arquivos CSV para o MySQL, garantindo que todas as informações necessárias fossem carregadas corretamente. Após a importação, os dados foram transformados para garantir a compatibilidade e facilitar a análise posterior. Um exemplo desse tratamento foi a conversão de valores textuais em valores numéricos, como a transformação da categoria 'Tecnologia' em '1'.

3.3.2 ETL

O processo de ETL (Extract, Transform, Load) é fundamental para a preparação dos dados. Este processo envolve:

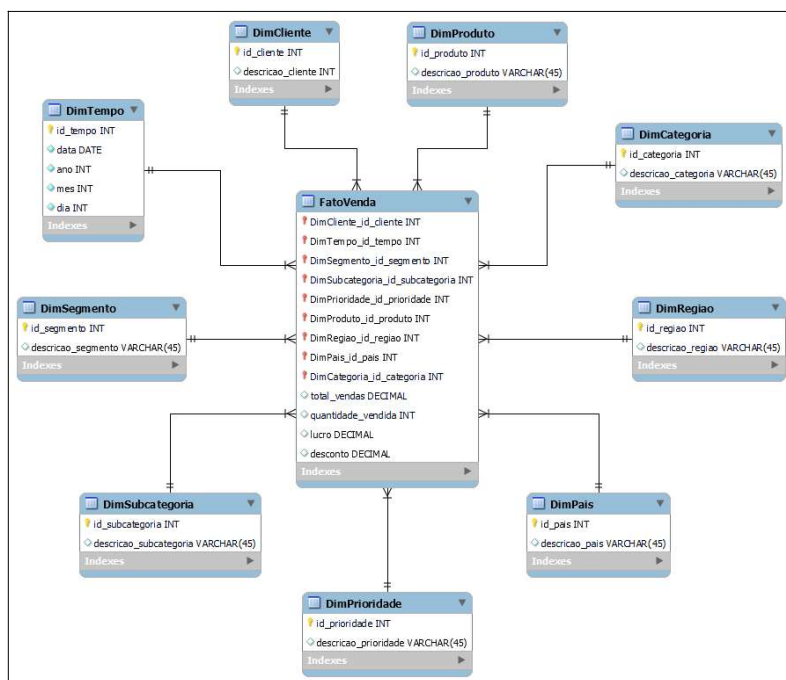
1. **Extração:** Os dados foram extraídos dos arquivos CSV importados para a staging area.
2. **Transformação:** A transformação incluiu a limpeza dos dados, a normalização dos valores e a conversão de strings em inteiros conforme necessário.
3. **Carga:** Os dados transformados foram carregados nas tabelas do Data Warehouse para posterior análise.

3.3.3 Modelo Multidimensional

Para suportar análises complexas, foi desenvolvido um Modelo Multidimensional (MMD), que incluiu as seguintes etapas:

Primeiramente, foram criadas as tabelas fato e dimensões no MySQL, facilitando a organização e a consulta dos dados. Em seguida, os relacionamentos entre as tabelas foram estabelecidos para garantir a integridade e a coesão dos dados. Uma atenção especial foi dada às tabelas **DIM_TEMPO** e **FATO_VENDA**, que são cruciais para as análises temporais e de vendas. Finalmente, os dados da staging area foram importados para as tabelas do modelo multidimensional, permitindo a realização de consultas e análises avançadas.

O Script SQL para carga de dados e criação da(s) tabela(s) do Modelo Multidimensional encontram-se no compartilhamento [GitHub Débora](#)



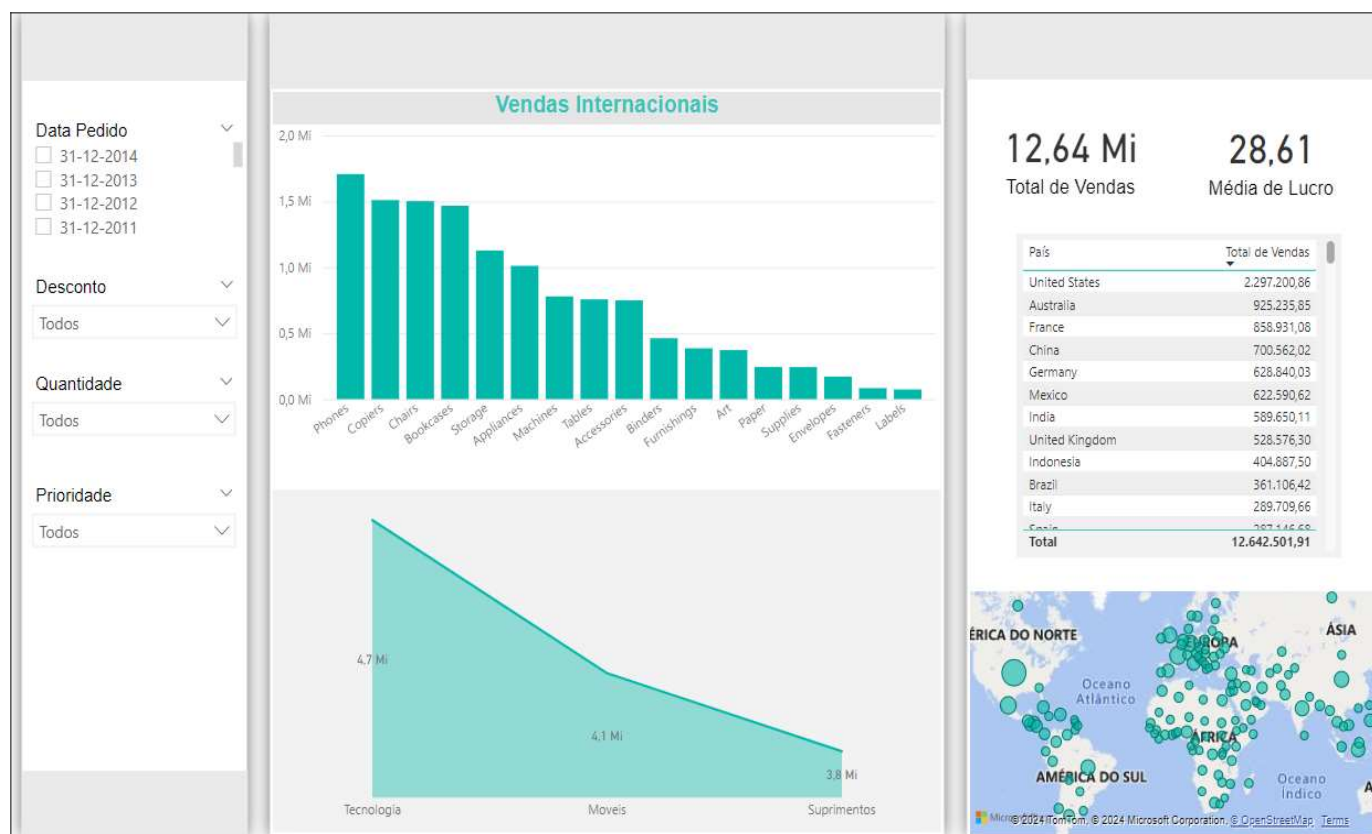
3.4 Modelagem

Na seção de modelagem, os dashboards foram desenvolvidos para fornecer uma visualização interativa e intuitiva dos dados.

O arquivo **PowerBI** encontra-se no compartilhamento **GitHub Débora**

Nesta estrutura de BI incluem métricas como o total de vendas, média de lucro e tabela detalhando o total de vendas por país. Além disso, há gráficos que destacam os produtos mais vendidos em subcategorias específicas e as categorias de produtos com maior volume de vendas. Essas visualizações oferecem uma compreensão abrangente do desempenho das vendas e das áreas mais lucrativas da empresa.

3.4.1 Dashboard



3.5 Avaliação

Ao analisarmos os países com maior volume de vendas, observamos que o primeiro lugar foi ocupado pelos Estados Unidos, seguido pela Austrália em segundo lugar, França em terceiro, e China e Alemanha em quarto e quinto lugar, respectivamente.

Em relação às categorias de produtos, constatamos que a Tecnologia foi a categoria com o maior índice de vendas, seguida por Móveis e Suprimentos. Quanto às subcategorias, os produtos mais vendidos incluem Telefones, Copiadoras, Cadeiras, Estantes, Armazenamento, entre outros. Para uma visualização mais detalhada da distribuição geográfica das vendas, fornecemos um mapa que destaca os países onde ocorreram as vendas.

Nesta seção apresentamos as descobertas relacionadas ao desempenho de vendas da empresa ao longo de um período de 4 anos. Destacamos que o total de vendas alcançou a cifra de 12,64 milhões, com uma média de lucro de 28,61.

Este projeto demonstra que a integração de ferramentas avançadas de BI, como Power BI, aliada a um processo estruturado de ETL e modelagem multidimensional, transforma vastos conjuntos de dados em insights acionáveis. A capacidade de visualizar e interpretar essas informações em tempo real através de dashboards interativos é um diferencial competitivo que permite à Inter Tech adaptar-se rapidamente às mudanças do mercado e sustentar seu crescimento a longo prazo.

4. Conclusão

Neste projeto, destacamos a importância da análise de dados para decisões estratégicas baseadas em fatos concretos. Observamos que os Estados Unidos lideram em volume de vendas, o que nos permite desenvolver estratégias replicáveis para outros mercados. É crucial compreender as estratégias de sucesso nos Estados Unidos e adaptá-las para diferentes contextos globais.

Identificamos que a categoria de tecnologia apresenta o maior percentual de vendas, indicando a necessidade de aumentar os investimentos nessa área para maximizar os lucros e atender à demanda do mercado.

Compreender o que, onde e porque certos padrões de vendas ocorrem é essencial para ajustar estratégias e impulsionar as vendas. A implementação de estratégias baseadas em dados sólidos pode alavancar ainda mais as vendas globais da Inter Tech, tornando-a mais competitiva e eficiente internacionalmente.

Além dos benefícios diretos nas vendas, a análise de dados proporciona uma visão abrangente das operações da empresa, identificando ineficiências e oportunidades de melhoria em processos internos. Por exemplo, revela tendências sazonais, preferências regionais de produtos e a eficácia das campanhas de marketing, facilitando a criação de estratégias personalizadas para diferentes mercados e segmentos de clientes.

5. Referências

- [1]. Professor Germano Crispim Vasconcelos (2018). Centro de Informática; Graduação em Ciência da Computação, Universidade Federal de Pernambuco. Disponível: https://www.cin.ufpe.br/~tg/2018-2/TG_CC/tg_mvgn.pdf
- [2]. <https://blog.dsacademy.com.br/blog/>