



CURSO: SISTEMA DE INFORMAÇÃO

COMPETÊNCIA: **ELABORAR PROJETO DE BUSINESS INTELLIGENCE**

DOCENTE: MARCELLA SIQUEIRA

# **Projeto de Business Intelligence: Análise de Varejo para Megaloja Global**

DISCENTES:

Henrique Albuquerque Araújo

Adeval Neto Cordeiro de Albuquerque

Gabriel Barbosa Claudino

Caio Henrique

CAMPINA GRANDE-PB

2025

## SUMÁRIO

<b>1 - INTRODUÇÃO E OBJETIVOS ESTRATÉGICOS</b>	<b>3</b>
<b>2 - ANÁLISE E ESPECIFICAÇÃO DA BASE DE DADOS</b>	<b>3</b>
2.1 - FONTE E DESCRIÇÃO DOS DADOS	4-6
2.2 - VOLUME E ESTRUTURA INICIAL	6
2.3 - DIAGNÓSTICO DE QUALIDADE E OBSERVAÇÕES INICIAIS	6-7
<b>3 - PERGUNTAS ESTRATÉGICAS DE NEGÓCIO</b>	<b>7</b>
3.1 - ANÁLISE DE CLIENTES	7
3.2 - ANÁLISE DE EFICIÊNCIA OPERACIONAL	7
3.3 - ANÁLISE DE DESEMPENHO DE VENDAS E PEDIDOS	7
<b>4 - USUÁRIOS E STAKEHOLDERS</b>	<b>8</b>
4.1 - USUÁRIOS DIRETOS	8
4.2 - STAKEHOLDERS ESTRATÉGICOS	8
<b>5 - KEY PERFORMANCE INDICATORS (KPI'S)</b>	<b>8-10</b>
<b>6 - PLANO DE TRABALHO TÉCNICO</b>	<b>10</b>
6.1 - ESTRATÉGIA DE COLETA E PREPARAÇÃO DOS DADOS	10
6.2 - MODELAGEM E TRANSFORMAÇÃO DE DADOS	10-11
6.3 - FERRAMENTAS UTILIZADAS	12

# 1. Introdução e Objetivos Estratégicos

Este documento detalha o plano técnico e estratégico para o desenvolvimento de uma solução de Business Intelligence (BI) a partir do conjunto de dados de varejo de uma megaloja global, cobrindo um período de quatro anos. O objetivo principal deste projeto é transformar dados brutos em insights acionáveis que permitam otimizar as operações, aprofundar a compreensão do comportamento do cliente e impulsionar o desempenho geral de vendas. A iniciativa visa capacitar a tomada de decisão em múltiplos níveis da organização por meio de análises precisas e visualizações de dados intuitivas.

Os objetivos centrais do projeto foram definidos para responder a perguntas críticas de negócio, organizadas nas seguintes frentes de análise:

- **Análise de Clientes:** Desenvolver uma compreensão profunda do valor e dos segmentos de clientes, identificando os perfis mais valiosos e seus padrões de compra.
- **Eficiência Operacional:** Avaliar e medir a performance dos processos de envio e entrega, focando na agilidade e consistência dos prazos logísticos.
- **Desempenho de Vendas:** Identificar padrões de vendas, tendências regionais e a performance de pedidos para otimizar a alocação de recursos e estratégias de mercado.

Para alcançar estes objetivos, uma análise rigorosa da base de dados é o ponto de partida indispensável, definindo o potencial e os limites de nossa iniciativa analítica.

## 2. Análise e Especificação da Base de Dados

A compreensão aprofundada da fonte de dados é um pilar crítico para o sucesso deste projeto. A análise da estrutura, volume e qualidade dos dados existentes serve como alicerce para a modelagem dimensional e garante a confiabilidade de todas as análises futuras. Um diagnóstico preciso nesta fase inicial mitiga riscos e direciona as etapas de preparação e transformação de dados de forma eficiente.

### 2.1. Fonte e Descrição dos Dados

A origem dos dados é um dataset de varejo de uma megaloja global, que consolida informações de transações ao longo de um período de 4 anos. A seguir, detalhamos a estrutura original das colunas e sua relevância analítica.

Dataset: SuperStore Sales Dataset

Origem: Kaggle (<https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>)

Coluna Original	Descrição e Relevância Analítica
<b>Row ID</b>	Identificador sequencial para cada linha. Considerada uma coluna redundante, sem valor analítico direto, sendo candidata à exclusão.
<b>Order ID</b>	Identificador único do pedido (nota fiscal). É um campo essencial para a contagem distinta de pedidos e para a estruturação da tabela fato.
<b>Order Date</b>	Data de processamento do pedido. Fundamental para análises temporais e para o cálculo de métricas de eficiência operacional.
<b>Ship Date</b>	Data de envio da mercadoria. Em conjunto com <b>Order Date</b> , permite calcular o tempo de processamento. Adicionalmente, possibilita mapear a influência de feriados e fins de semana nos prazos de envio.
<b>Ship Mode</b>	Modalidade do envio ( <b>Standard Class, Second Class, First Class, Same Day</b> ). Permite analisar a eficiência por tipo de serviço, cujos prazos exploratórios variam de 4-7 dias ( <b>Standard Class</b> ) a envios no mesmo dia ( <b>Same Day</b> ).
<b>Custome</b>	Identificador único do cliente. Chave primária ideal para a dimensão de clientes ( <b>dClientes</b> ), garantindo a identificação

<b>r ID</b>	precisa em vez do nome.
<b>Customer Name</b>	Nome do cliente. Utilizado para identificação e visualização em relatórios.
<b>Segment</b>	Segmento de mercado do cliente ( <b>Consumer, Corporate, Home Office</b> ). Permite análises de comportamento e é fundamental para responder à pergunta de negócio sobre o ticket médio por perfil de cliente.
<b>Country</b>	País da transação. Coluna com valor único ("Estados Unidos"), tornando-a redundante e passível de exclusão.
<b>City</b>	Cidade do cliente. Nível de granularidade geográfica que pode ser removido em favor de <b>State</b> e <b>Region</b> para análises mais agregadas.
<b>State</b>	Estado do cliente. Campo geográfico relevante para análises de desempenho em um nível intermediário de granularidade.
<b>Postal Code</b>	Código postal (CEP). Apresenta valores ausentes e possui menor utilidade analítica que <b>State</b> ou <b>Region</b> para este escopo.
<b>Region</b>	Região geográfica ( <b>East, Central, South, West</b> ). Nível de granularidade mais alto, ideal para análises estratégicas de desempenho de vendas.
<b>Product ID</b>	Identificador único do produto. Essencial como chave primária para a dimensão de produtos ( <b>dProdutos</b> ).

<b>Category</b>	Categoria principal do produto ( <b>Office Supplies, Furniture, Technology</b> ). Permite análises de vendas por linha de produto.
<b>Sub-Category</b>	Subcategoria do produto. Oferece um nível de detalhe adicional para a análise de performance de produtos.
<b>Product Name</b>	Nome específico e extenso do produto. Considerado de baixa utilidade para visualização devido à sua extensão, sendo candidato à remoção.
<b>Sales</b>	Valor da venda. Métrica quantitativa central para a maioria das análises financeiras e de desempenho.

## 2.2. Volume e Estrutura Inicial

O dataset inicial é composto por **9.800 registros e 18 atributos originais**. Após o diagnóstico inicial, o plano de modelagem se concentrará em **13 atributos principais** após a exclusão de colunas redundantes ou de baixa relevância, conforme detalhado na próxima seção.

## 2.3. Diagnóstico de Qualidade e Observações Iniciais

A análise exploratória inicial revelou os seguintes pontos que requerem atenção durante a fase de preparação dos dados:

- **Colunas Redundantes:** As colunas **Row ID** e **Country** serão descartadas. **Row ID** é um simples contador de linhas sem valor de negócio, e **Country** contém um único valor, não oferecendo variabilidade para análise.
- **Dados Ausentes:** A coluna **Postal Code** apresenta valores vazios. A solução proposta é substituir os nulos pelo CEP **5401**, que corresponde à cidade e estado informados nas mesmas linhas, embora a coluna inteira seja candidata à remoção por baixa relevância analítica.
- **Granularidade e Utilidade:** Colunas como **Product Name** (muito extenso e específico para visualizações agregadas) e **City** (menos útil

para análises estratégicas que **State** ou **Region**) serão removidas para simplificar o modelo e focar em dimensões mais relevantes.

- **Duplicidade Funcional:** A repetição de valores em **Order ID**, **Order Date** e **Ship Date** é um comportamento esperado e funcional. Isso ocorre porque um único pedido (nota fiscal) pode conter múltiplos produtos, cada um representado por uma linha no dataset, mantendo os mesmos dados de cabeçalho do pedido.

Este diagnóstico não apenas dita as etapas de preparação de dados, mas também valida a viabilidade de responder às perguntas de negócio a seguir, garantindo que as análises futuras sejam construídas sobre uma base de dados íntegra e relevante.

### 3. Perguntas Estratégicas de Negócio

Esta seção representa o núcleo do projeto, definindo as perguntas estratégicas que orientarão todo o desenvolvimento técnico. As respostas a estas questões fornecerão os insights necessários para a tomada de decisão, direcionando desde a modelagem dos dados até a construção dos dashboards finais.

#### 3.1. Análise de Clientes

- Qual o gasto médio por cliente (**Ticket Médio Cliente**)?
- Quais são os clientes mais valiosos para a empresa?
- Qual o ticket médio por segmento de cliente (**Consumer, Corporate, Home Office**)?

#### 3.2. Análise de Eficiência Operacional

- Qual o tempo médio entre o processamento do pedido e o envio?
- Qual a consistência e previsibilidade dos prazos de envio para cada modalidade (**Ship Mode**)?

#### 3.3. Análise de Desempenho de Vendas e Pedidos

- Qual o número total de pedidos realizados por período?
- Qual o ticket médio por pedido em um determinado período?
- Quais regiões (**East, Central, South, West**) geram o maior volume de vendas?

As respostas a essas perguntas serão consumidas por diferentes públicos dentro da organização, conforme detalhado a seguir.



## 4. Usuários e Stakeholders

A definição clara dos públicos-alvo da solução de BI é fundamental para garantir que os dashboards e relatórios sejam relevantes, úteis e alinhados com as necessidades estratégicas e operacionais da empresa. A identificação dos usuários e stakeholders direciona o design da interface, o nível de detalhe das informações e a forma como os insights são comunicados.

### 4.1. Usuários Diretos

Este grupo é composto por profissionais que utilizarão a solução para investigações aprofundadas e análises ad-hoc, explorando os dados para descobrir insights não previstos nos dashboards principais. Perfis incluem **Analistas de Negócios** e **Analistas de Dados**.

### 4.2. Stakeholders Estratégicos

Este grupo é formado por gestores e diretores que consumirão os insights consolidados para a tomada de decisão estratégica. Eles necessitam de visões de alto nível e indicadores-chave para monitorar a performance e direcionar suas equipes.

- **Gerência de Vendas e Marketing:** Utilizarão as análises de **Ticket Médio por Cliente**, **Total de Vendas por Região** e desempenho por **Segmento** para refinar estratégias de mercado, personalizar campanhas e alocar recursos de vendas de forma mais eficaz.
- **Gerência de Operações e Logística:** Consumirão diretamente os KPIs de **Tempo Médio de Envio** e **Desvio Padrão do Tempo de Envio** para otimizar processos, renegociar contratos com transportadoras e garantir a previsibilidade das entregas, impactando diretamente a satisfação do cliente.
- **Diretoria Executiva:** Necessitam de visões consolidadas do desempenho geral do negócio, incluindo tendências de vendas, rentabilidade por região e eficiência operacional, para direcionamento estratégico de longo prazo.

Os Key Performance Indicators (KPIs) a seguir serão as métricas formais para atender às necessidades informacionais desses usuários e stakeholders.

## 5. Key Performance Indicators (KPIs)

Os KPIs (Indicadores-Chave de Performance) são as métricas quantificáveis que permitirão monitorar o desempenho em relação aos objetivos de negócio



estabelecidos. Eles traduzem as perguntas estratégicas em valores mensuráveis, possibilitando um acompanhamento contínuo e objetivo. A tabela abaixo formaliza o cálculo e o propósito de cada indicador que será implementado na solução de BI.

KPI	Fórmula/Cálculo	Justificativa Estratégica	Área/Usuário Impactado
<b>Ticket Médio por Cliente</b>	<b>DIVIDE(SUM(fVendas[Sales]), DISTINCTCOUNT(fVendas[Customer ID]))</b>	Medir o valor médio de compra por cliente para otimizar estratégias de retenção e marketing.	Vendas, Marketing
<b>Tempo Médio de Envio</b>	<b>AVERAGEX(fVendas, DATEDIFF(fVendas[Order Date], fVendas[Ship Date], DAY))</b>	Avaliar a eficiência do processo logístico desde o pedido até o despacho.	Operações, Logística
<b>Ticket Médio por Pedido</b>	<b>DIVIDE(SUM(fVendas[Sales]), DISTINCTCOUNT(fVendas[Order ID]))</b>	Entender o valor médio de cada transação para análises de rentabilidade.	Vendas, Finanças
<b>Desvio Padrão do Tempo de Envio</b>	<b>STDEVX.P(fVendas, DATEDIFF(fVendas[Order Date], fVendas[Ship Date], DAY))</b>	Medir a consistência e previsibilidade das entregas, identificando instabilidades no processo.	Operações, Logística

<b>Total de Vendas por Região</b>	<b>SUM( fVendas[ Sales ] ) agrupado por dClientes[ Region ]</b>	Identificar as regiões com melhor desempenho para alocação de recursos e estratégias de expansão.	Vendas, Estratégia
<b>Quantidade e de Pedidos</b>	<b>DISTINCTCOUNT( fVendas[ Order ID ] )</b>	Acompanhar o volume de transações para entender a demanda e a capacidade operacional.	Vendas, Operações

O plano técnico a seguir detalha como esses KPIs serão implementados a partir da transformação dos dados brutos.

## 6. Plano de Trabalho Técnico

Esta seção apresenta o roteiro para a implementação técnica do projeto de Business Intelligence. Serão detalhados os passos para transformar o conjunto de dados brutos em um modelo de dados otimizado, bem como as ferramentas que serão utilizadas para desenvolver e disponibilizar as análises.

### 6.1. Estratégia de Coleta e Preparação dos Dados

Nesta fase inicial do projeto, o foco não será na coleta de dados adicionais, mas sim na preparação e limpeza do dataset existente para garantir a qualidade e a relevância das informações. As seguintes ações de tratamento de dados serão executadas:

- **Exclusão de Colunas:** As colunas **RowID**, **Country**, **Postal code**, **Product Name**, e **City** serão removidas do modelo final. Essas exclusões são justificadas pela redundância (**RowID**, **Country**), baixa relevância analítica (**Postal Code**, **City**) ou dificuldade de visualização (**Product Name**).

## 6.2. Modelagem e Transformação de Dados (Esquema Estrela)

Para otimizar o desempenho das consultas e proporcionar clareza analítica, será implementado um modelo dimensional do tipo Esquema Estrela (Star Schema). Este modelo organiza os dados em uma tabela de fatos central, que contém as métricas de negócio, e múltiplas tabelas de dimensão, que descrevem o contexto dessas métricas. A estrutura planejada é a seguinte:

*Nota: (PK) denota Primary Key, o identificador único da tabela dimensão. (SK) denota Surrogate Key, a representação da chave primária na tabela fato.*

- **Tabela Fato (fVendas):** Esta tabela centralizará as métricas quantitativas (**Sales**) e as chaves estrangeiras que a conectam ao contexto descritivo das dimensões.
  - **Order ID**
  - **Order Date**
  - **Ship Date**
  - **Ship Mode**
  - **Sales**
  - **Customer ID (SK)**
  - **Product ID (SK)**
- **Dimensão de Clientes (dClientes):** Agrupará todos os atributos relacionados aos clientes.
  - **Customer ID (PK)**
  - **Customer Name**
  - **Segment**
  - **State**
  - **Region**
- **Dimensão de Produtos (dProdutos):** Consolidará as informações descritivas dos produtos.
  - **Product ID (PK)**
  - **Category**
  - **Sub-Category**
- **Dimensão de Calendário (dCalendario):** Uma tabela de data dedicada será criada e relacionada a **Order Date** e **Ship Date**. Esta é uma prática recomendada de modelagem que permite análises temporais robustas (ano-contra-ano, mês-a-mês, dia da semana) de forma performática e sem a necessidade de cálculos complexos nos relatórios.

### 6.3. Ferramentas Previstas

- **Power BI:** Será a principal ferramenta do projeto, utilizada para a modelagem de dados, criação de cálculos e medidas em DAX (Data Analysis Expressions), e para o desenvolvimento dos dashboards e relatórios visuais interativos.