Project Documentation

Store Sales Analysis

Submitted by: SAAS Team

Abstract

This project focuses on the foundational phase of a data analysis workflow using Power BI, emphasizing the definition of analytical objectives and the exploration of a comprehensive sales dataset. The primary goals are to identify key business questions, determine appropriate analytical methodologies, and assess the structure and quality of the data to ensure accurate insights. Through Power BI, we explore patterns, trends, and potential data issues across various dimensions such as customer behavior, shipping performance, product categories, and regional sales performance. The dataset includes transactional and customer-level information, capturing critical attributes like order and shipment timelines, customer demographics, and sales figures, which together provide a rich context for business intelligence reporting and strategic decision-making.

Introduction

In today's data-driven business environment, understanding customer behavior, product performance, and operational efficiency is critical for maintaining a competitive edge. This project initiates a comprehensive data analysis process using Power BI, a powerful business intelligence tool, to transform raw sales data into meaningful insights. The analysis begins by clearly defining the project’s objectives, identifying key business questions, and determining the appropriate analytical approach.

The dataset at the center of this analysis includes detailed information on orders, customers, products, and shipping logistics. By exploring this data, we aim to uncover trends, patterns, and anomalies that can inform strategic decisions. This introductory phase also involves evaluating the data’s structure, completeness, and accuracy to ensure the reliability of any conclusions drawn.

Ultimately, this project serves as a foundation for building dynamic dashboards and reports that support data-driven decision-making across various business functions.

Problems to be Addressed

1️⃣ Data Cleaning

- Converted date fields ("Order Date", "Ship Date") to proper Date format.
- Rounded Sales values to 2 decimal places for clarity.
- Removed unnecessary columns like "Row ID" and "Country" to streamline the dataset.

2️⃣ Sales Prediction Using Machine Learning

- Built a machine learning model to forecast future sales using features like product category, region, segment, and time-based variables.
- Steps: feature engineering, data splitting, model training (e.g., Linear Regression, Random forest), and evaluation (RMSE/MAE).
- Predictions can be integrated into Power BI for actionable insights.

3️⃣ Dashboard Design

- Created an interactive Power BI dashboard highlighting key metrics (sales, profit, top products, delivery).
- Used visual elements like charts, slicers, and maps for clarity.
- Dashboard designed for different user roles with filters and performance optimization.

Goals/Objectives

Our goal is to analyze store sales data using Power BI to uncover actionable insights, such as: Identifying sales trends and peak performance periods. Evaluating product and store performance. Understanding customer behavior and profitability.

Milestone and Deadline

Week 1: Requirement Gathering & Planning

- Define the project scope, objectives, KPIs, and identify data sources.

Week 2: Data Collection & Cleaning

- Gather raw data and perform necessary cleaning and preprocessing tasks.

Week 3: Data Modeling & DAX Development

- Build data models and create calculated columns/measures using DAX in Power BI.

Weeks 4–5: Dashboard Design & Development

- Design user-friendly, interactive dashboards tailored to business needs.

Week 6: Testing & Validation

- Validate dashboard accuracy and ensure alignment with stakeholder expectations.

Week 7: Deployment & Training

- Final review, documentation, and user training for dashboard adoption.

Week 8: Project Completion

- Official handover and closure of the project.

Tools & Technologies

1. Power BI
   - Used for building interactive dashboards and reports to visualize insights effectively.
2. Power Query Editor
   - Utilized for data cleaning, transformation, and preparation within Power BI.
3. DAX (Data Analysis Expressions)
   - Enables advanced calculations, custom measures, and dynamic reporting in Power BI.
4. SQL (SQL Server DBMS)
   - Serves as the primary relational database for structured data storage and querying.
5. Python
   - Key libraries used:
     - Pandas for data manipulation and analysis
     - NumPy for numerical operations
     - Matplotlib/Seaborn for data visualization
     - Scikit-learn for building and evaluating machine learning models
     - https://www.kaggle.com/code/tarekomar55/store-sales-analysis-prediction
6. Notion
   - Supports project planning, documentation, and team collaboration throughout the project lifecycle.
   - https://www.notion.so/Store-Sales-Analysis-1955947f86aa80ffb562da77a4befbfc?pvs=4

SQL/DAX Queries

SQL commands:

Total Sales (Sum of Sales):

- SELECT SUM(Sales) AS TotalSales
- FROM [Superstore Sales Dataset];

Top 5 Sub-Categories by Sales:

- SELECT TOP 5 SubCategory, SUM(Sales) AS SumOfSales
- FROM [Superstore Sales Dataset]
- GROUP BY SubCategory
- ORDER BY SumOfSales DESC;

Top 10 Customers by Sales:

- SELECT TOP 10 CustomerName, SUM(Sales) AS TotalSales
- FROM [Superstore Sales Dataset]
- GROUP BY CustomerName
- ORDER BY TotalSales DESC;

Total Sales by Segment:

- SELECT Segment, SUM(Sales) AS TotalSales
- FROM [Superstore Sales Dataset]
- GROUP BY Segment;

Total Sales by Ship Mode:

- SELECT ShipMode, SUM(Sales) AS TotalSales
- FROM [Superstore Sales Dataset]
- GROUP BY ShipMode;

Total Sales by Category:

- SELECT Category, SUM(Sales) AS TotalSales
- FROM [Superstore Sales Dataset]
- GROUP BY Category;

DAX commands:

- Total Sales = SUM('Superstore Sales Dataset'[Sales])
- Count Of Customers = DISTINCTCOUNT('Superstore Sales Dataset'[Customer Name])
- Count Of Orders = COUNT('Superstore Sales Dataset'[Order ID])

Key Personnel

Who is on our team?

Tarek Omar
Team Leader

Retaj Yasser
Dashboard Developer

Yassin Ahmed
Dashboard Developer

Eman Abdelrahim
Data Analyst