Project On

“Sales Performance Analysis and Forecasting”

Prepared By:

Babita Acharya

Submitted To:

Mindrisers Institute of Technology

Table of Content

Introduction

This project focuses on the analysis of sales performance using the Superstore Sales Dataset. Sales performance analysis is essential for understanding business growth, identifying profitable products and regions, and improving strategic decision-making. The dataset provides detailed information on orders, customers, products, regions, shipping modes, sales, profit, discounts, and time-related attributes. By analyzing this data, meaningful insights can be extracted to support data-driven business decisions.

Project Overview

The objective of this project is to evaluate sales performance and uncover trends, patterns, and key business insights using data analysis techniques. The analysis includes **data cleaning, data transformation, exploratory data analysis (EDA), and visualization** to better understand sales behavior and profitability.

Objective

The main objectives of this project are:

- To analyze sales performance over time
- To identify top and low-performing products and regions
- To explore sales trends using visualizations

Dataset Description

Source: Kaggle (Datasets)
Type: Retail Sales Data
Rows: 9994
Columns: 21

Features

i. Sales Performance Metrics

- **Sales:** Total revenue generated from orders.
- **Profit:** Net profit earned after costs.
- **Quantity:** Number of items sold.
- **Discount:** Discount applied to orders.

ii. Product Analysis

- **Category:** Furniture, Office Supplies, Technology.
- **Sub-Category:** Detailed product classification to identify top and low-performing items.

iii. Regional Analysis

- **Region:** Central, East, South, West.
- **State & City:** Used to evaluate geographical performance and market strength.

iv. Customer Segmentation

- **Segment:** Consumer, Corporate, Home Office.
- Helps analyze customer behavior and purchasing trends.

v. Shipping & Delivery Insights

- **Ship Mode:** Standard Class, Second Class, First Class, Same Day.
- Used to understand delivery efficiency and its impact on sales and profit.

vi. Time-Based Features

- **Order Date:** Used for trend and seasonality analysis.

- **Year, Month, Quarter:** Derived features for time series analysis.

Methodology

The project followed the steps outlined below:

I. **Data Collection and Loading:**
The Superstore dataset was loaded into a Pandas DataFrame for analysis.

II. **Data Cleaning and Preprocessing:**
The dataset was checked for missing values and duplicates. Date columns were converted into proper datetime format, and new time-based features such as year, month, and quarter were created.

III. **Outlier Detection and Treatment:**
Box plots were used to identify outliers in Sales and Profit. Outliers were retained as they represent legitimate high-value transactions and loss-making orders.

IV. **Exploratory Data Analysis (EDA):**
Statistical summaries and visualizations such as line charts, bar charts, box plots, and heatmaps were used to explore trends, product performance, regional performance, shipping modes, and correlations.

V. **Visualization and Insights:**
Visual analysis was used to identify peak sales periods, high-performing categories, loss-making products, and the negative impact of discounts on profit.

Conclusion

The Sales Performance Analysis provides valuable insights into business performance and customer behavior. The analysis shows that certain product categories and regions consistently generate higher sales and profit, while excessive discounting negatively affects profitability. Standard Class shipping contributes the highest sales volume, indicating customer preference for cost-effective delivery. Seasonal trends reveal peak sales periods that can be leveraged for marketing and inventory planning.

Overall, this project demonstrates how data-driven analysis can support strategic decision-making, optimize pricing and logistics, and improve overall sales performance.

GitHub Repository

The complete project including datasets, notebooks, and documentation is maintained in a GitHub repository.