

ANALYSIS OF SUPERSTORE SALES DATASET

Submitted by,
Smruthi S Kumar



OBJECTIVE:

Analyze retail sales data to derive insights into customer behavior, popular products, and sales trends.

TECH STACK USED:

- Jupyter Notebook
- Data Visualization Libraries of python like Numpy, Pandas, Matplotlib and Seaborn



CONTENTS

- Import the data
- Data Exploration
- Data Cleaning
- Exploratory Data Analysis (EDA)
- Recommendations

IMPORT THE DATASET

- The name of the dataset is Superstore Sales Data
- The dataset is imported into the Jupyter notebook using Pandas

DATA EXPLORATION

- Dataset has 9800 rows and 18 columns.
- The data types of the column include int, object and float.
- Total number of unique values in each column was found out.
- Statistical description of the data was analyzed including mean, median, mode, standard deviation etc.



DATA CLEANING

- The column “Postal Code” has null values present, it has been dropped.
- Checked if duplicate values are present.
- Converted data types.

EXPLORATORY DATA ANALYSIS (EDA)

Some of the insights derived are as follows:

- The total sales is 2252607.4127
- It is observed that maximum shipment mode was through standard class.
- Consumers have done the highest sales.
- Only United States was the country present in the sales column.
- Highest sales have been done in the western region of United States.
- Office supplies was sold in maximum number.
- Out of the office supplies category binders was more sold and copiers was the least sold.



- Customer segmentation was done by calculating recency, frequency and monetary value for each customer and was clustered as 0, 1 and 2.
- Cluster 0 may represent regular or moderate-value customers who make purchases less frequently compared to other segments.
- Cluster 1: This segment represents high-value customers who make frequent purchases.
- Cluster 2: This segment represents customers who are moderately engaged and have a moderate spending level.
- Product analysis was done and the top 10 selling products was identified.
- Time series analysis was done on the basis of day, month and year. Maximum number of sales was happened in 2017 and later.



RECOMMENDATIONS

- **Target High-Value Customer Segments:** Identify and prioritize high-value customer segments, such as those with high frequency and monetary value.
- **Improve Customer Experience:** Focus on enhancing the overall customer experience across all touchpoints, including online platforms, physical stores, and customer service interactions.
- **Monitor and Adjust Strategies:** Continuously monitor key performance indicators (KPIs) such as sales volume, customer retention rate, and average order value.

THANK YOU