

TASK 6.1 : SUPERSTORE SALES DATA ANALYSIS

ABOUT THE DATA:

Superstore is an office supplies store spread across the US. With changing times they want to focus on increasing profits and maintaining customer loyalty. They want a comprehensive analysis of their past data, to get a forecast for the future sales. Where to focus more.

They want to know who are their top customers, what cities and states are having maximum sales, sales growth over the years and so on.

DATA SOURCE:

The data for this project is an open-source data downloaded from <https://www.kaggle.com/datasets/rohitsahoo/sales-forecasting>

DATA CLEANING:

1. Removed the unnecessary columns (not relevant for our analysis) – 'Row ID', 'Order ID', 'Customer ID', 'Customer Name' (because PII)
2. Found 1 duplicate row, which was removed.
3. Found 11 missing values in the column 'Postal Code'. These were segregated and since all missing codes were from the same city 'Burlington' in Vermont State, the postal code was replaced after searching google.
4. Date formats (in order date and shipping date columns) we changed and updated in correct format.
5. For ease of analysis, month and year columns were created.

DATA PROFILE:

1. The original data set had 9800 rows and 18 columns, which after cleaning are now 9799 rows and 16 columns.

DATA LIMITATIONS & ETHICS:

1. The Shipment date and Order date were listed in different format. This makes it difficult to analyze.
2. The Postal code of one of the cities was missing so I had to replace it with the correct code.
3. In order to understand monthly/yearly sales, I had to create new columns month & Year from the Order date column.
4. There is no data on profit margin for each product so It is difficult to understand profitability.
5. Shipment mode description is not clear enough to understand the delivery time.

TASK 6.1 : SUPERSTORE SALES DATA ANALYSIS

COLUMN DETAILS

COLUMN	COLUMN DESCRIPTION	DATA TYPE	TIME VARIANT
Order ID	The number assigned to the order	Quantitative, Discrete	no
Order Date	Date of the order placed	Quantitative, Discrete	yes
Ship Date	Date of the shipping	Quantitative, Discrete	yes
Ship Mode	Type of shipment method	Qualitative, Ordinal	no
Customer ID	The number assigned to the customer	Quantitative, Discrete	no
Customer Name	Name of the customer	Qualitative, Nominal	no
Segment	Segment of the product	Qualitative, Nominal	no
Country	Name of the country	Qualitative, Nominal	no
City	Name of the city	Qualitative, Nominal	no
State	Name of the state	Qualitative, Nominal	no
Postal Code	The code assigned to each area	Quantitative, Discrete	no
Region	The region of the country	Qualitative, Nominal	no
Product ID	The number assigned to the product	Quantitative, Discrete	no
Category	Category of the product	Qualitative, Ordinal	no
Sub-Category	Sub-category of the product	Qualitative, Ordinal	no
Product Name	Item description	Qualitative, Nominal	no
Sales	Sales of the order Id	Quantitative, Continuous	yes

QUESTIONS TO BE ANSWERED:

1. What day of the week are our sales high on average?
2. What is monthly average sales?
3. What are the year over growth year/ total sales over the years?
4. Who are the top customer of all times?
5. What are the highest sales by category?
6. Which city does have the highest sales?