

Name: Srikar Bajjuri

MAIN TASK - Data Analytics | Python

- **Data Exploration:** Thoroughness in understanding the dataset and initial insights.
- **Data Visualization:** Clarity, accuracy, and effectiveness of visualizations.

Explanation:

Data Analytics | Python Task

Name: Srikar Bajjuri

Details: VIT - 20MIC0106

Dataset - "Global-Superstore" Data source

: https://docs.google.com/spreadsheets/d/1KagwoQLy1quKvT_82amuS-x3UnsoIX4J6p02ewbjQNA/edit?usp=sharing

- Data Exploration: Thoroughness in understanding the dataset and initial insights.
 - Data Visualization: Clarity, accuracy, and effectiveness of visualizations.
-

To perform this given task we will be following the below main steps:

1. libraries importing and initialization
2. Dataset loading
3. Pre-processing and cleaning of dataset
4. Main Exploration Analysis;
 - EDA - Exploration and data analysis
 - DV - Data Visualization
5. Feature Engineering and visualization

KEYWORDS IN THE DOCUMENT

EDA - Exploratory Data Analysis

VIZ - Visualization

Feature Engineering and visualization

Feature engineering is the process of using domain knowledge to create new features (input variables) from raw data to improve the performance of machine learning models. In the context of Exploratory Data Analysis (EDA), feature engineering involves transforming and enhancing the dataset to uncover insights and patterns that may not be immediately apparent.

Benefits of Feature Engineering in EDA:

Improved Model Performance: Well-engineered features can enhance the predictive power of machine learning models. **Better Insights:** New features can reveal hidden patterns and relationships within the data. **Enhanced Understanding:** Helps in understanding the underlying structure of the data and the factors influencing the target variable. In summary, feature engineering is a crucial step in EDA and the overall data preprocessing pipeline, enabling more effective analysis and model building by transforming raw data into meaningful features.

some additional features after feature engineering:

Sales per Day: Calculate the sales per day for each order.

Profit per Day: Calculate the profit per day for each order.

Discounted Sales: Calculate the effective sales after considering the discount.

High Discount Flag: Create a flag indicating whether a discount is higher than a certain threshold (e.g., 20%).

Insight Analysis of feature engineering above

Sales per Day and Profit per Day: Analyzing these distributions can help identify the efficiency of processing orders. Higher values might indicate faster processing times with substantial sales or profit.

Discounted Sales: This can show how discounts impact actual sales. Comparing this with the original sales can help measure the effectiveness of discount strategies.

High Discount Flag Impact on Profit: Understanding how high discounts affect profit can help in making strategic decisions about discount offerings.

POWER BI TASK – Data Visualization

Power BI Exercise

To perform the task given I have followed the below steps and implemented to achieve the requirements specified;

Key Words to know: [

Step 1: From the link of dataset given, I have downloaded the file as .xlsx file to support Power BI platform

Data source link given;

https://docs.google.com/spreadsheets/d/1KagwoQLy1quKvT_82amuS-x3UnsoIX4J6p02ewbjQNA/edit?usp=sharing

Step 2: We have loaded the dataset, now pre-dominantly if we are accurately familiar with the data in our file and if we have an idea of what Key Performance Indicators (KPIs) we going to visualize we can drop the columns which are not necessary, for example; the Postal Code column and other few ID columns are not required so we can drop them, but as I am not the curator of the dataset, I would keep all the columns as it is.

Note: If we drop columns that are not required, we can decrease the load on the Power BI processing.

Step 3: Next, we will have to identify all the KPIs which we are going to visualize and create the dashboard with, in this exercise we have already been provided with the KPIs by the company in the task instructions.

KPIs: Key Metrics:

Display the following key metrics:

- Total Sales Revenue
- Sales by Product Category
- Top Selling Products
- Sales Trend Over Time
- Monthly Sales Comparison

Step 4: In this exercise I will be creating 5 pages for each KPI, and a cumulative dashboard page including a few main visualizations from each KPI dashboard to make it concise in one page.

To implement the whole KPIs I have used multiple charts and visualizations;

In the main cumulative dashboard I have used one or two visualization from each KPI dashboard