Name: Pavan Deore

Course: DANLC

Batch: ANP-C8220

Trainer: Rajshri Thete Ma'am

Synopsis:

1. **Project Title: Data Analysis of SuperStore Sales using Python and MySQL**

2. **Objective:**

The primary objective of this project is to conduct an in-depth analysis of the "Sample Superstore" dataset to uncover key insights related to sales, profit, discounts, and other factors that influence the performance of a retail business. The goal is to identify trends, correlations, and patterns that can guide strategic business decisions aimed at increasing profitability and operational efficiency.

3. Dataset Overview:

The dataset, "Sample Superstore," (.sql) file contains detailed records of transactions from a retail store. It includes 13 columns and 9994 entries, capturing various attributes such as:

Ship Mode: The mode of shipping used for the order.

Segment: The customer segment (Consumer, Corporate, Home Office).

Country, City, State, Postal Code: Geographic information.

Region: The region where the order was placed.

Category: The product category (Furniture, Office Supplies, Technology).

Sub-Category: More specific product categories (e.g., Chairs, Bookcases).

Sales: The sales amount for each transaction.

Quantity: The quantity of items sold.

Discount: The discount applied to the order.

Profit: The profit generated from the order.

4. Scope of Work:

The project will involve the following tasks:

- **Data Exploration:** Understanding the dataset, including the features and target variable.
- **Data Preprocessing:** Cleaning the dataset by handling missing values, removing outliers, and normalizing/standardizing the data.
- **Feature Selection:** Identifying the most significant features influencing wine quality.
- **Data Visualization:** Using plots and graphs to visualize the relationship between features and wine quality.
- **Model Building:** Building and evaluating machine learning models to predict wine quality.
- **Interpretation of Results:** Analyzing the output of the models and drawing conclusions.
- **Reporting:** Documenting the findings and preparing a final report.

5. Methodology:

The project is divided into the following steps:

1. **Data Cleaning and Preprocessing:**
   - Handling missing values (if any) and ensuring data types are correctly assigned.
   - Performing initial exploratory data analysis (EDA) to understand the basic structure and distribution of the data.

2. **Exploratory Data Analysis (EDA):**
   - Conducting visual and statistical analysis to identify trends and patterns. Analysing sales and profit distribution across different categories, regions, and customer segments.
   - Investigating the impact of discounts on sales and profitability.
   - Exploring correlations between numeric variables such as sales, profit, quantity, and discount.

3. **Data Visualization:**
   - Creating a series of visualizations to present the findings effectively, including bar plots, scatter plots, heatmaps, and distribution plots.
   - Key visualizations include sales and profit by category, region, state, and sub-category, as well as the impact of discounts on sales and profitability.

4. **Conclusion and Insights:**
   - Summarizing the key findings from the analysis.
   - Providing actionable insights that can inform strategic decisions in pricing, marketing, and inventory management.

5. **Evalution and Interpretation:**

➢ Compare model performance.
➢ Interpret the results to understand the impact of different features on wine quality.

6. **Visualization:**

➢ Generate charts and graphs to visualize the findings.

7. **Reporting:**

➢ Compile the analysis, results, and insights into a comprehensive report.

6. Tools and Technologies:

Programming Language	Python, MySQL
Libraries	Pandas, NumPy, Matplotlib, Seaborn,
IDE:	Jupyter Notebook , MySQL Workbench

7. Expected Outcomes:
➢ Identification of key factors influencing sales and profitability.
➢ Understanding the impact of discounts on business performance.
➢ Recommendations for improving sales strategies and operational efficiency.

8. Timeline:

Week 1: Data collection and cleaning.
Week 2: Exploratory data analysis and feature selection.
Week 3: Modeling and visualization.
Week 4: Final analysis, reporting, and presentation of findings.