MINI PROJECT

Phase-1

1) Define Problem statement (Online retail sales dataset)

The online retail sales dataset under consideration is a comprehensive collection of transactions from an e-commerce platform. It encompasses a wide range of products, each identified by a StockCode, described briefly, and quantified in terms of Quantity. The dataset also includes essential information such as Invoice details, Customer IDs, and geographical location (Country). This project's primary objective is to conduct a thorough exploratory data analysis (EDA) and statistical analysis of the online retail sales data to extract actionable insights and draw meaningful conclusions.

2)Project Plan and Product Backlog: Online Retail Shop Analysis

Phase 1: Data Exploration and Preparation

Task 1: Define Project Objectives
>> Clearly state the goals and objectives of the analysis for the online retail dataset.

Task 2: Data Acquisition
>> Gather and load the provided dataset into the analysis environment.

Task 3: Initial Data Inspection
>>Get an overview of the dataset's structure, including columns, data types, and any immediate data quality issues.

Task 4: Data Cleaning and Preprocessing
>> Address missing values, duplicates, and any data inconsistencies.
Format data types for analysis.

Phase 2: Exploratory Data Analysis (EDA)

Task 5: Univariate Analysis
>> Analyze individual variables (e.g., Quantity, Price) to understand their distributions and characteristics.

Task 6: Bivariate Analysis
>> Explore relationships between pairs of variables to uncover potential correlations or patterns.

Task 7: Visualizations
>>Create relevant visualizations (e.g., histograms, scatter plots) for better data understanding.

Phase 3: Statistical Analysis and Hypothesis Testing

Task 8: Frame Analytical Questions
>> *Formulate specific questions related to the online retail data (e.g., Are there significant differences in purchase quantities across different products?).*

Task 9: Calculate Dispersion
>> *Measure variability in key parameters (e.g., Quantity, Price).*

Task 10: Data Distribution Analysis
>> *Assess how data is distributed for different variables.*

Task 11: Hypothesis Testing
>> *Conduct relevant tests (e.g., t-tests, ANOVA) to answer formulated questions.*

Phase 4: Customer Segmentation (if applicable)

Task 12: Define Segmentation Criteria
>> *Determine factors for customer segmentation (e.g., purchasing behavior, demographics).*

Task 13: Segment Identification
>> *Use clustering techniques to identify customer segments.*

Phase 5: Documentation and Reporting

Task 14: Document Insights
>> *Summarize key findings, insights, and methodologies used during the analysis.*

Task 15: Prepare Summary Report
>> *Create a report with visualizations and key takeaways.*

Phase 6: Project Review and Next Steps

Task 16: Review Project Progress
>> Evaluate the project's achievements and assess if objectives were met.
Task 17: Identify Next Steps
>> Determine potential follow-up analyses or actions based on the findings.

Product Backlog:

1.Define project objectives and scope.

2.Acquire and load the online retail dataset.

3.Inspect initial dataset structure and quality.

4.Clean and preprocess the data.

5.Perform univariate and bivariate analysis.

6.Create relevant visualizations for data exploration.

7.Formulate specific analytical questions.

8.Calculate dispersion for key parameters.

9.Analyze data distribution for different variables.

10.Conduct hypothesis testing.

11.Define customer segmentation criteria (if applicable).

12.Identify customer segments using clustering techniques (if applicable).

13.Document insights and methodologies.

14.Prepare a summary report with visualizations.

15.Review project progress and adjust plan if necessary.

16. Identify potential follow-up analyses or actions.

3.   Creation of Git Repository