Assignment: Data Preparation Basics

Tasks:

➔ You will find credit_customers dataset on the drive use it to solve the assignment.

Data Loading:

- Load the dataset into a Pandas DataFrame.

Exploratory Data Analysis (EDA):

- Perform EDA to understand the structure and characteristics of the dataset.
- Use descriptive statistics and visualizations to analyze data distribution, and understand relationships between variables, with a description for the relationships and how each feature effect in the target.

Handling Missing Values:

- Identify and handle missing values in the dataset. Choose appropriate techniques such as imputation or deletion based on the nature of missing data and describe why did you choose this technique.

Outlier Detection and Handling:

- Detect outliers using statistical methods and visualize them, with explain why and where.
- Decide on a strategy to handle outliers, whether it's removing them, or imputing them, with explain the result (comment, visualization).

Imbalance problem and Handling:

- Check balance between classes and visualize them, with explain why and where.
- Decide on a strategy to handle imbalance, with explain the result (comment, visualization) and describe why did you choose this technique.

Data Transformation:

- Apply feature scaling (Min-Max scaling or Z-score normalization) to numerical features.
- Convert categorical variables into numerical format using techniques like one-hot encoding or label encoding.

Classification Model:

- Build a model using a suitable library (e.g., scikit-learn).
- Train the model on the training dataset.

Model Evaluation:

- Evaluate the model's performance on the testing set.
- Use appropriate classification metrics.

Conclusion:

- Summarize the findings from the classification task.
- Reflect on the model's performance and potential areas for improvement.