Data Preprocessing and Transformation

Objective

The objective of this assignment is to practice data preprocessing techniques on a real-world dataset. You will handle missing values, detect and treat outliers, normalize and discretize data, apply encoding techniques, and save the processed data.

Dataset

- Use a dataset "**Employees.xlsx**" that includes various types of columns (numeric, categorical) and contains missing values and potential outliers.

Tasks

Write a Python script that reads the `employees.xlsx` file and applies the following tasks:

1. **Identify Missing Values**
   - Calculate and report the ratio of missing values for each column in the dataset.
2. **Handle Missing Values**
   - For each column with missing values, use the following approach based on the data type:
     - **Numeric Columns**: Replace missing values with the mean of the column.
     - **Categorical Columns**: Replace missing values with the mode (most frequent value) of the column.
3. **Handle Negative Values**
   - For all numeric features in the dataset, convert any negative values to their absolute values.
4. **Outlier Detection and Management**
   - **Detect Outliers**:
     - Use **boxplots** and the **Z-score** method to detect outliers in numeric columns.
   - **Handle Outliers**:
     - For each identified outlier, replace it with the median of the values in the column (excluding the outliers).
5. **Normalization**
   - Normalize the values in the `Salary` column to be within the range [0, 1000].
6. **Discretization**
   - Discretize the `Performance Score` column into **4 bins** (e.g., Low, Medium, High, Excellent) based on the score distribution. Each bin should represent a range of performance scores.
7. **Encoding Categorical Variables**
   - Apply **Label Encoding** to the following categorical columns:
     - `Department`
     - `Education Level`
     - `Seniority Level`
8. **Save the Processed Data**
   - Save the fully processed dataset to a new file named **Processed_Employee_Dataset.xlsx**.

After completing the tasks, upload your Python script to the e-learning portal.