Data Mining - EX3

Deadline: Wednesday, Aban 2, 1402 - October 23, 2024

Consider the data in the file **"Churn(preprocessing)"** and answer the following questions. Upload your answer to the E-learn system in the form of a notebook (.ipynb format) including the executed code, outputs, and explanations for each question.

1. **Handling Missing Data:**
   - **a.** Remove rows where the **"Int'l Plan"** column contains missing values.
   - **b.** Replace the missing values in the **"VMail Plan"** column with the mode (most frequent value).
   - **c.** For the columns **"VMail Message"**, **"Day Mins"**, and **"Day Calls"**, determine which method for handling missing values is most appropriate. Implement that method.
   - **d.** Impute the missing data in the **"Night Mins"** column randomly, using a normal distribution.

2. **Identifying Outliers:**
   - **a.** Remove the missing values in the **"Day Charge"** column, and then identify its outliers using a graph.
   - **b.** Identify the outliers in the **"Day Charge"** column using the **Z-score** method.
   - **c.** Identify the outliers once again using the **IQR** method.

3. **Data Normalization:**
   - **a.** Normalize the data in the **"Night Mins"** column using the **min-max** method. (Use the completed data from question 1.)
   - **b.** Standardize the data in the **"Night Mins"** column using the **Z-score** method.
   - **c.** Compare the results of both normalization methods using a graph.

- **d.** Calculate and compare the skewness of the data before and after applying both transformations.

4. Normality Check:

Check the normality of the data in the **"Eve Mins"** column (only available values) using visualizations such as histograms or Q-Q plots.