Assignment 01

Use case: Is there a relationship between humidity and temperature? What about between humidity and apparent temperature? Can you predict the apparent temperature given the humidity? https://www.kaggle.com/budincsevity/szeged-weather?select=weatherHistory.csv

Produce Python Notebook, which performs the following task

(a) Preprocess the dataset as specified in the data mining process.

I. Handle Missing Values and Outliers if any
II. Produce Q-Q Plots and Histograms of the features, and apply the transformations if required.
III. If it is required, apply suitable feature coding techniques.
IV. Scale and/or standardized the features, produce relevant graphs to show the scaling/ standardizing effect.
V. If necessary, apply feature discretization, and produce a relevant graph to show the discretization

(b) Perform Feature Engineering by executing the following task:

I. Appropriately use PCA (Principal Component Analysis) or SVD (Singular Value Decomposition) for feature reduction.
II. Identify significant and independent features using appropriate techniques. Show how you selected the features using suitable graphs.

(c) Apply the following techniques to predict the value of Y for the test dataset (Training Dataset to Test Dataset is 0.8 to 0.2 ratio)

Linear Regression Model

The Assignment Report (Submit as a Report / Submit as a Blog)

You need to formulate solutions for each of parts (a) through (e) above, clearly explaining your Python code and specifying the outputs produced by the code for the dataset given in a *Jupyter Notebook* named *Solution_StudentNumber.ipynb*. For each such part, a descriptive summary with an interpretation should be given for the output obtained after each executable cell.