Introduction:

The given lab assignment is associated with programming with Python. This explores the various machine and deep learning concepts of handling data in terms of classification and/or regression using supervised and unsupervised approaches.

Objectives:

The lab work involves a collection of tasks to be accomplished listed as below.

* Creation of dictionary using keys and their corresponding values using a list of tuples containing information of students
* Evaluation of a string to identify the longest sub-string existing within without the repetition of characters resulting in the derivation of the sub-string with new revised length
* Development of an airline booking system that enables the user to reserve an airline ticket per user needs
* Performing multiple regression of a data set and evaluating its performance (accuracy) using Root Mean Squared Error (RMSE) and R Squared (R2) methods of performance evaluation
* Using a data set containing both qualitative and quantitative data, perform an exploratory analysis to identify the most correlated features associated with the target, to remove null values associated with the features, if any, and to convert any categorical features into numerical features
* Perform classification on data set comprising of both qualitative and quantitative data using the algorithms: Naive Bayes, Support Vector Machines and K-Nearest Neighbors
* Cluster analysis on a data set of our choice, computing their performance using the silhouette score and visualizing their results, along with application of elbow method to identify the ideal number of clusters

Concepts Incorporated:

The following concepts have been explored and utilized to work on data of different forms to accomplish different objectives.

* Dictionary, list and tuples
* String and sub-string concepts
* Class and their corresponding functions and approaches
* Multiple Regression along with RSME and R2 methods of performance evaluation
* Data Analysis
* Classification - Naive Bayes, Support Vector Machine (Linear SVM) and K-Nearest Neighbors
* K-Means Clustering along with silhouette score approach of performance evaluation and elbow method to determine optimal clusters

Data Sets Used:

Multiple data sets have been to accomplish different objectives.

* Diabetes (Clustering)
* Boston Housing (Multiple Regression)
* Bank (Classification)