**Project – Exam1**

**CSEE 5590/CS490: Python and Deep Learning**

# **Group 8:**

Medavarapu Sumanth: 29

Roshna Toke: 50

Salkuti Sandeep Reddy: 41

Patchala Sai Chand: 34

Ramineni Pardha Saradhi: 38

### **Introduction:**

This project is about five different questions which we must solved using given methods and importing respective libraries and methods.

### **Objectives:**

The objective of this project is it will cover all the concepts which we have learned in the class. To learn different algorithms and its advantages, disadvantages, and its accuracy. We can compare best algorithm which suits to our dataset.

### **Approaches/Methods & Workflow:**

**Q.2 Apply K-means on the dataset in this** [**link**](https://umkc.box.com/s/lwr3s70prbe3tdifzx6fblq7wqln54xq) **and visualize the clusters using matplotlib or seaborn.**

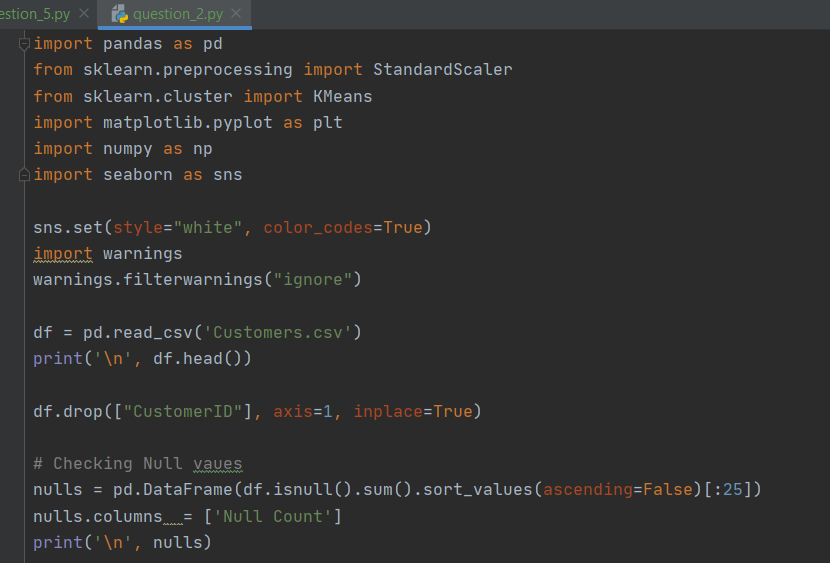
## **Our goal is to cluster our customers into buying groups based off of their Annual Income and Spending Scores**

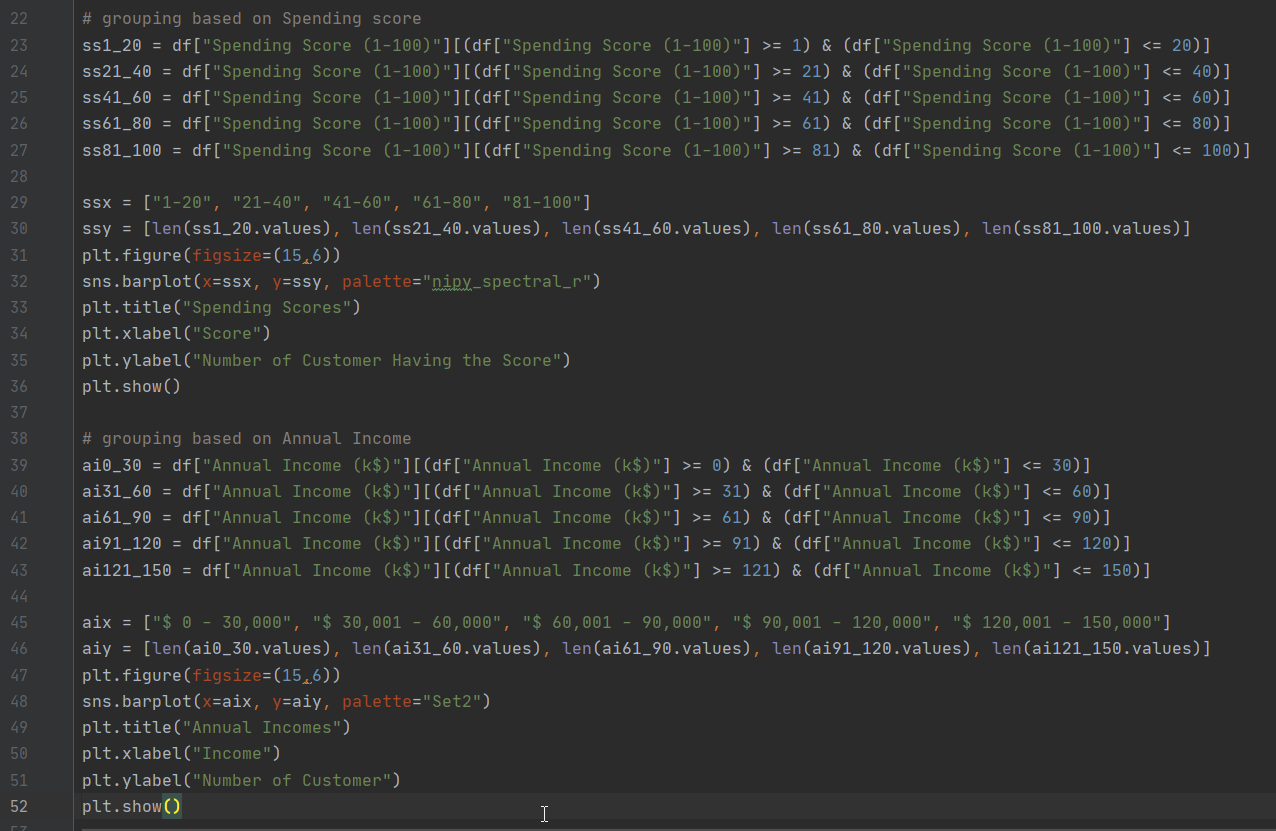
1. Report which K is the best using the elbow method.
2. Evaluate with silhouette score or other scores relevant for unsupervised approaches (before applying clustering clean the data set if needed)
3. Can you interpret the clustering result that you have visualized?

**Approaches:**

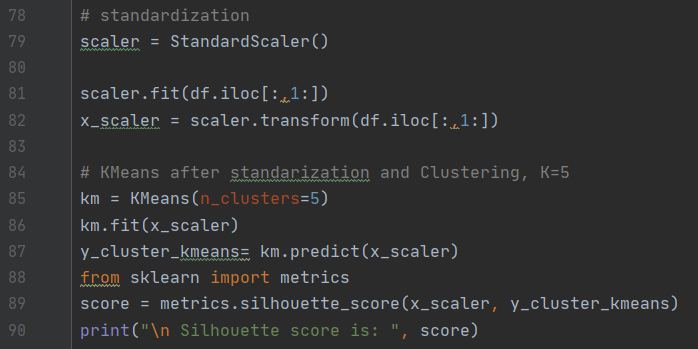
1. Imported required libraries and loaded given dataset.
2. Dropped ‘CustomerId’ column.
3. Checked for Null Values for data cleaning.
4. Showing grouping of customers based on Spending score and Annual income by using seaborn barplot.
5. Written one function for mapping gender value. (M=0, F=1)
6. Performed K-Means Clustering using Elbow method. K = 5.
7. Scaled the data and calculated silhouette score.
8. From box plot visualization, we can interpret that range of spending score is more than annual income.
9. Shown interpretation based on three parameters. (Age, Annual Income, Spending score)

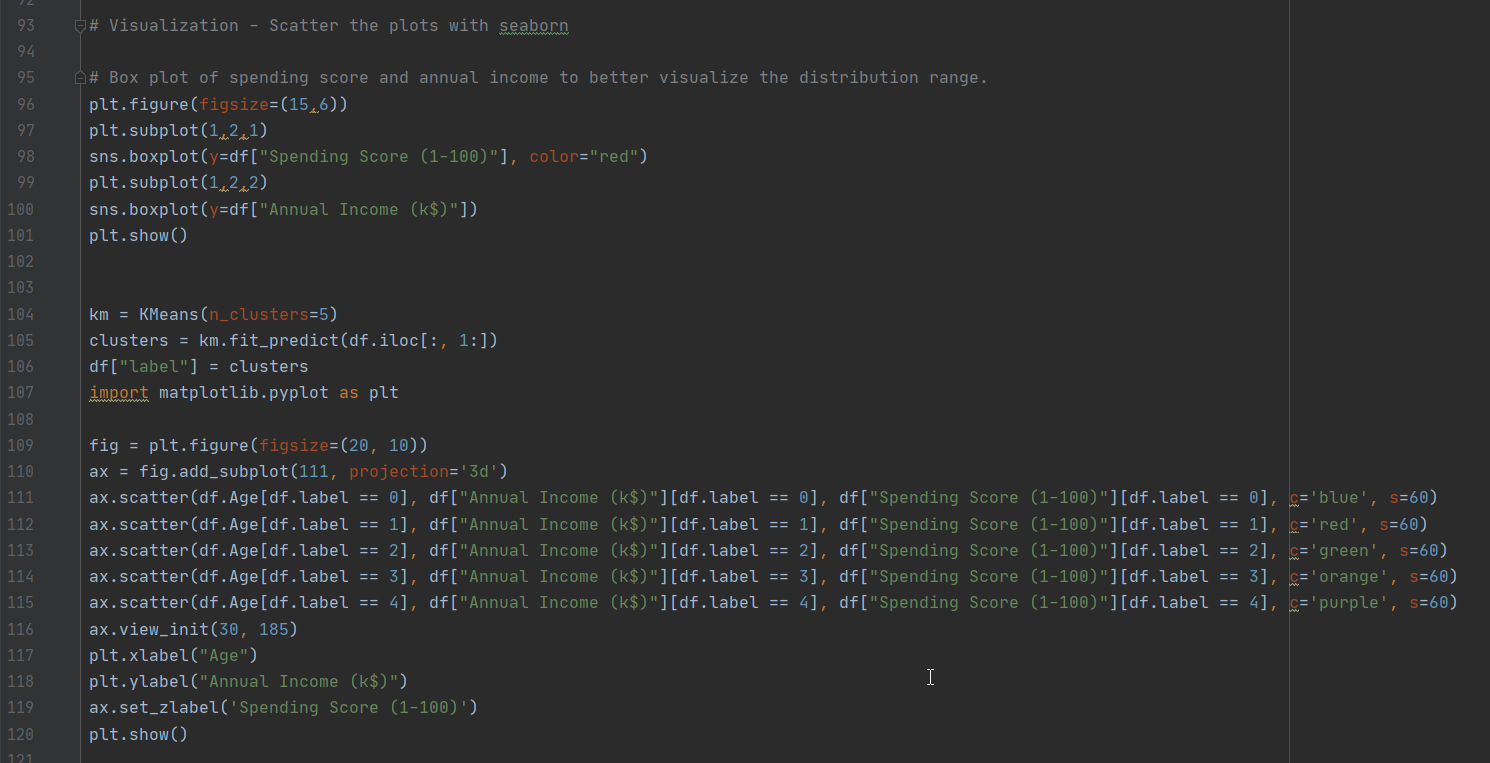
**Workflow:**

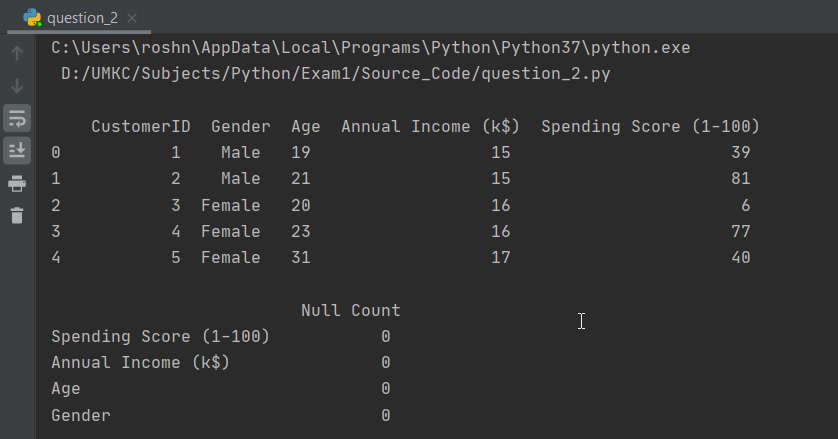




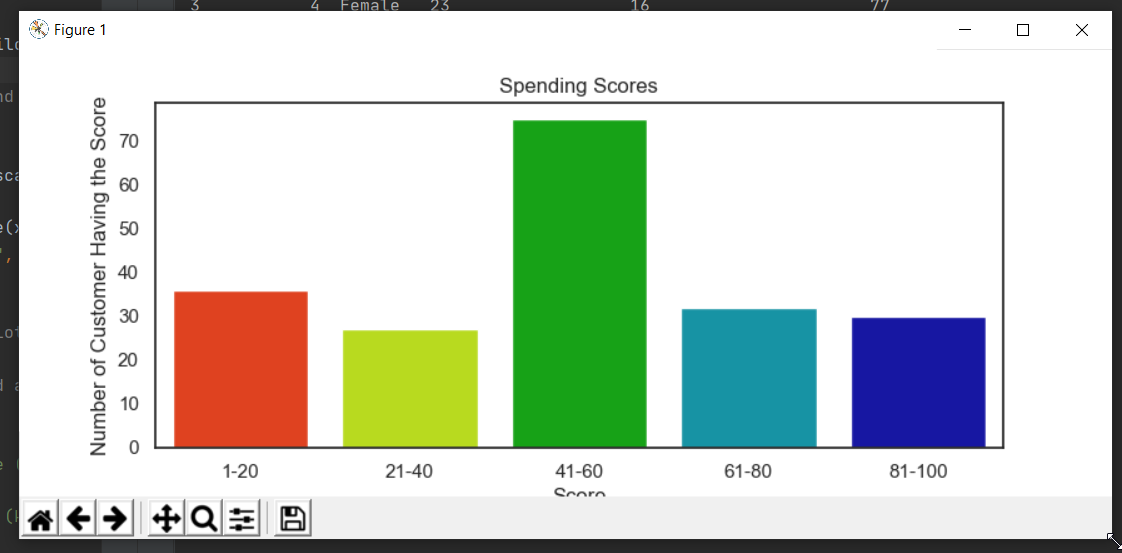


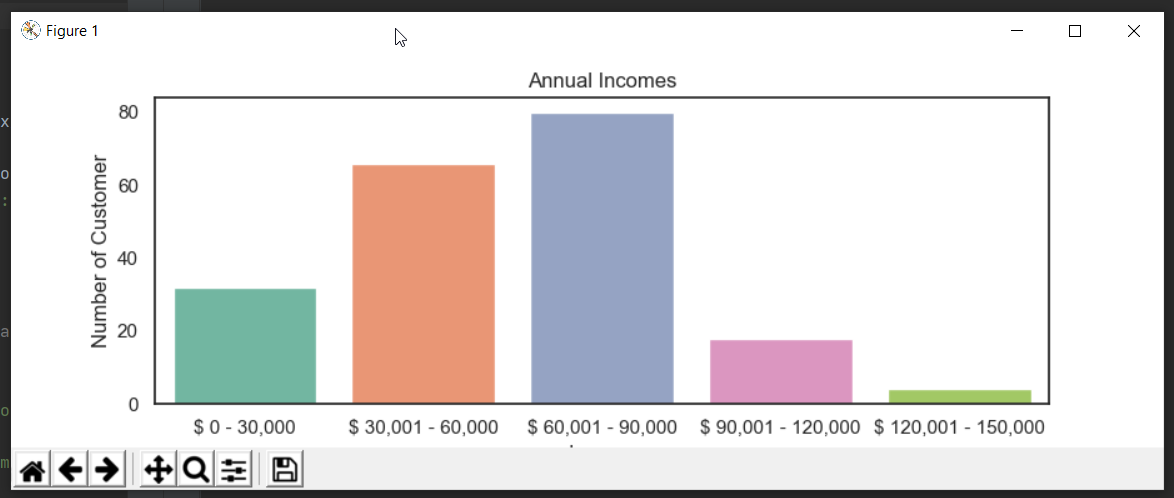


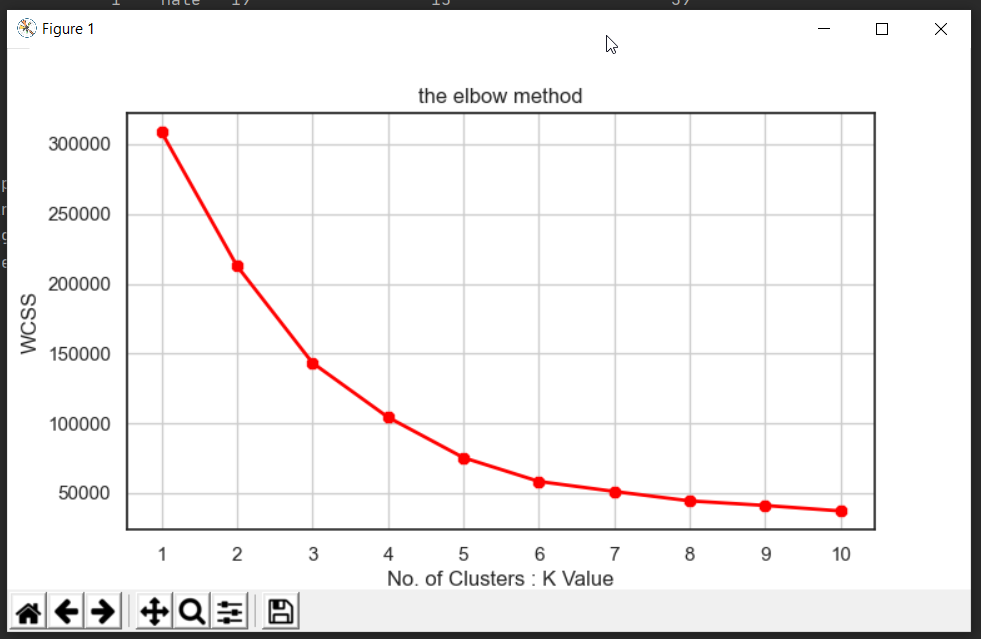


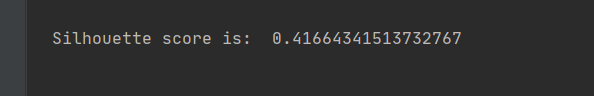


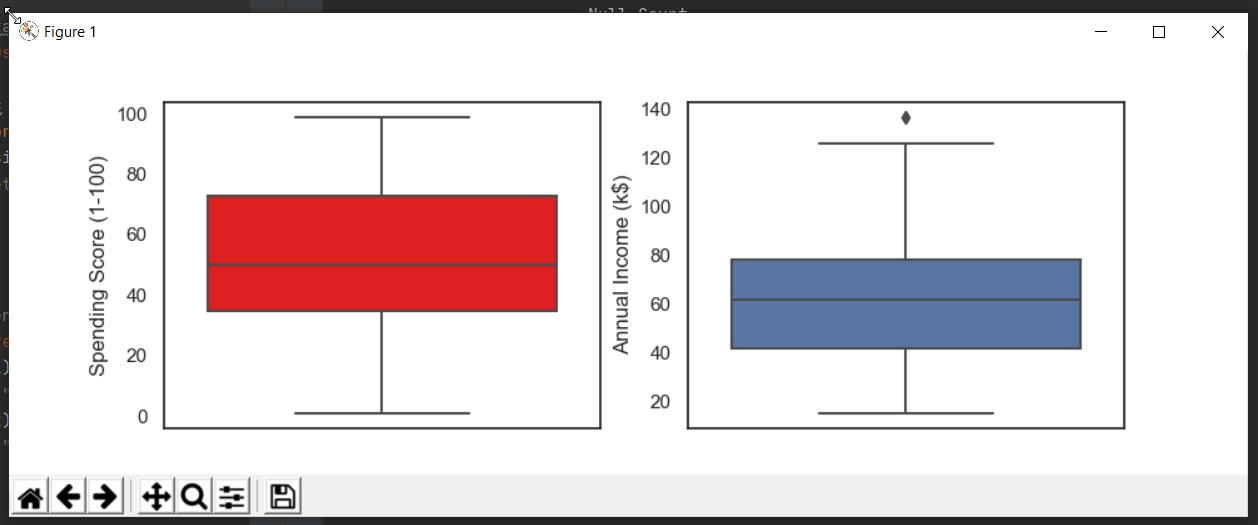
Output:

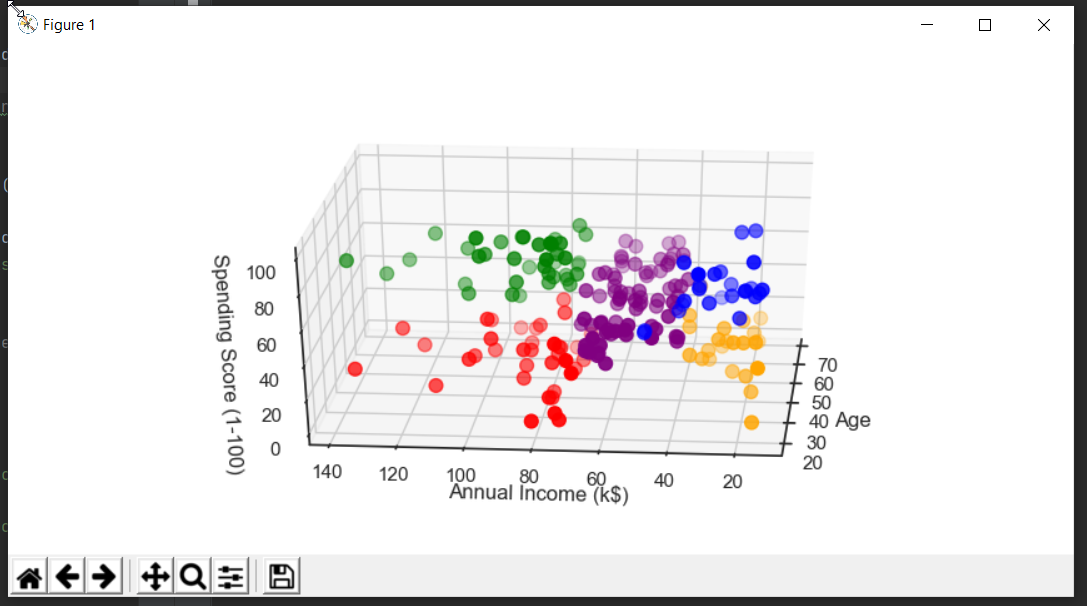












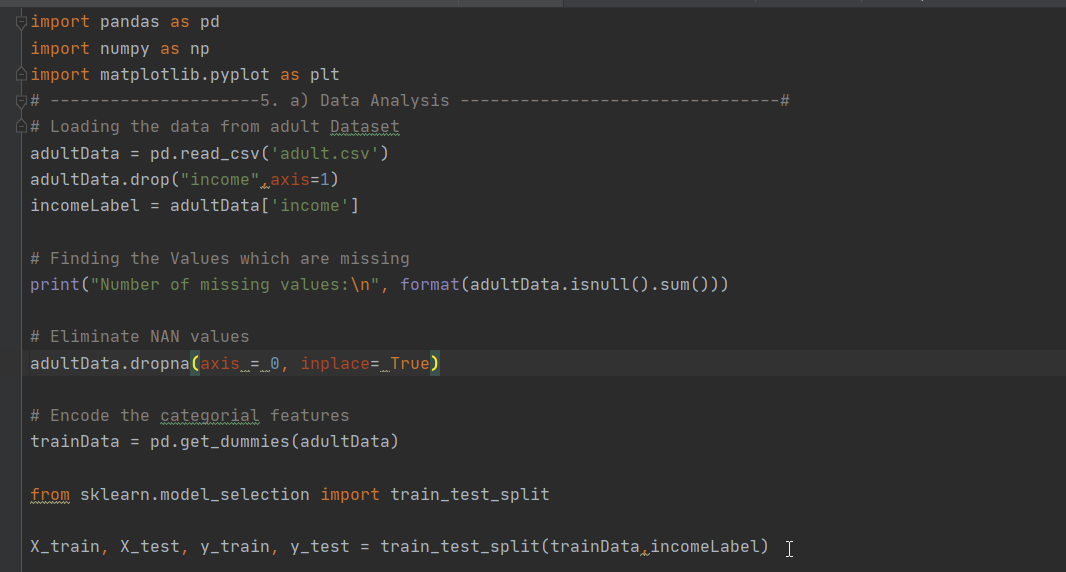
**Q.5 Pick any dataset online for the classification problem which includes both numeric and non-numeric features**

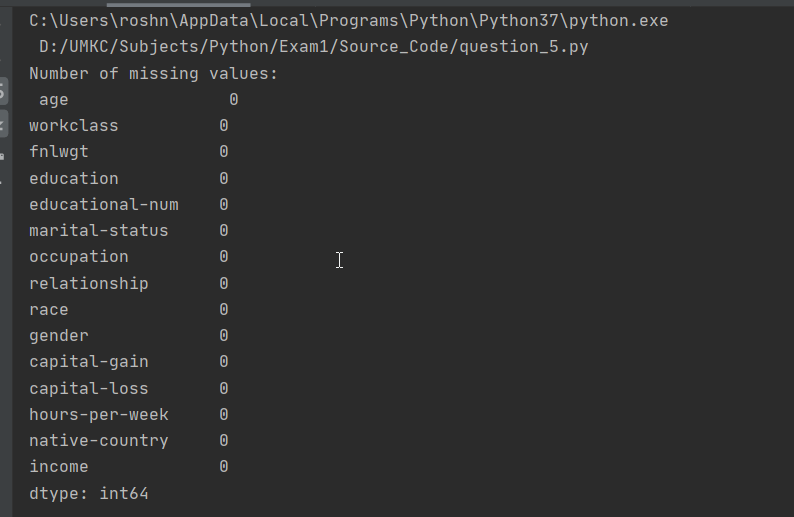
1. Perform exploratory data analysis on the data set (it can be anything on your choice that gives insight about the dataset)
2. Apply the three classification algorithms Naïve Bayes, SVM and KNN on the chosen data set and report which classifier gives better result.
3. Try SVM with linear and non-linear kernel and report which one gives better performance

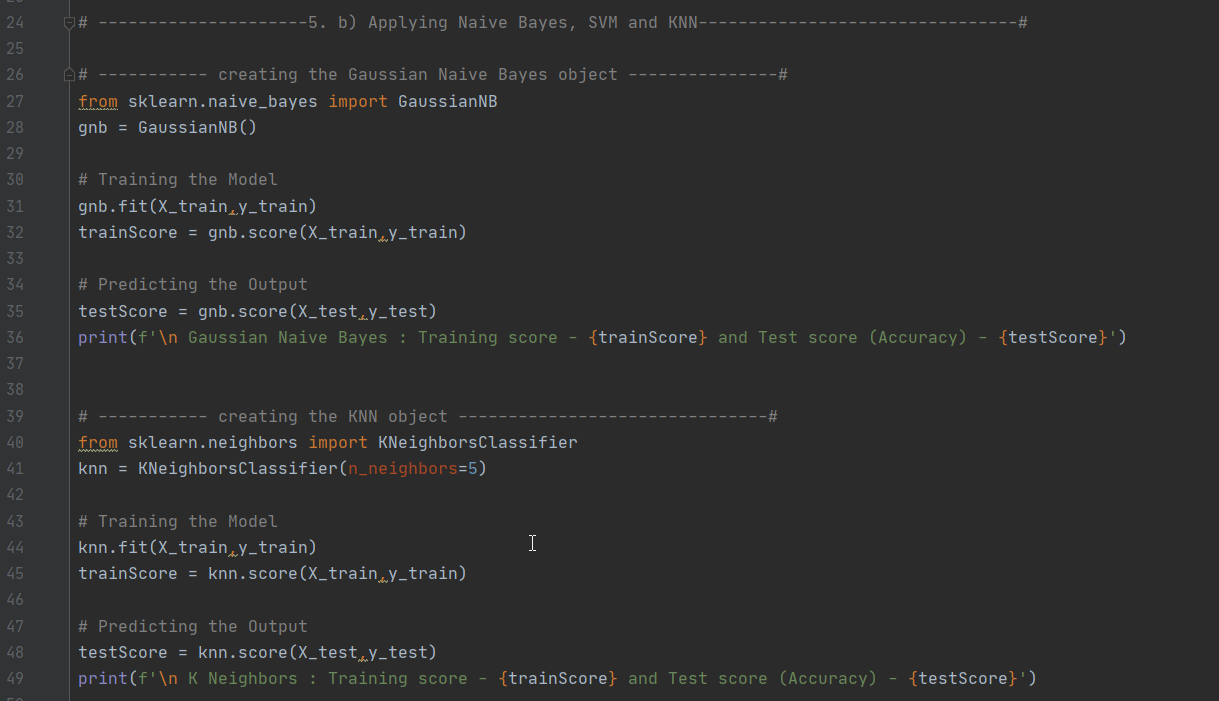
Approaches:

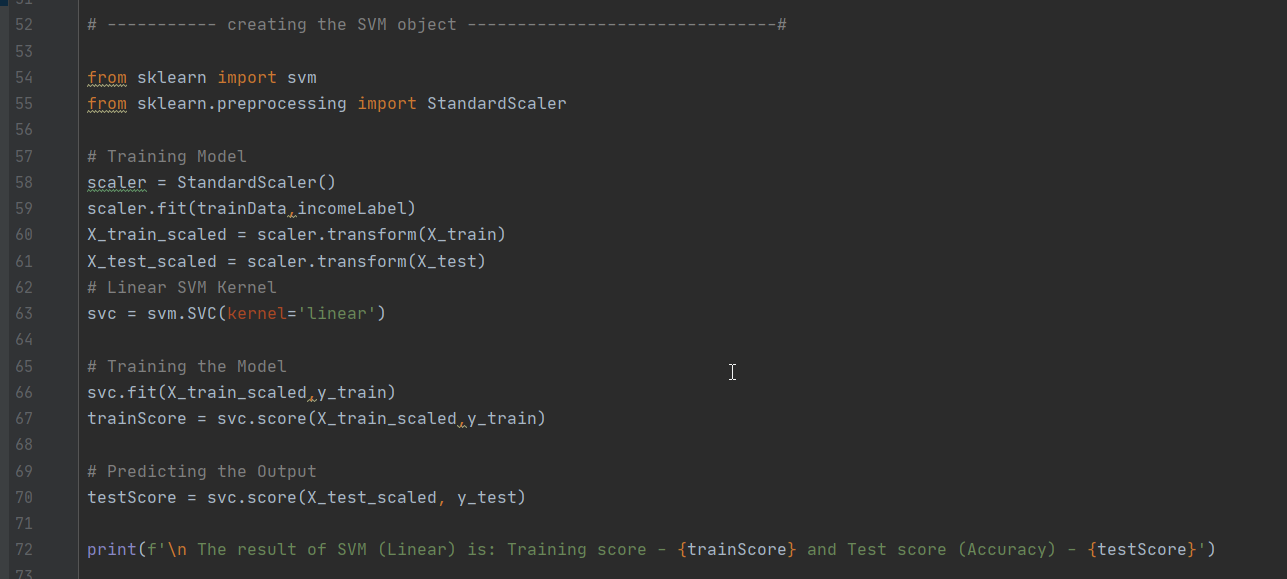
1. We used data from link: <https://www.kaggle.com/wenruliu/adult-income-dataset> to predict whether income exceeds $50K per year.
2. Importing required libraries.
3. Extracting required training data and target column (income).
4. We checked whether null values are present or not, but we didn’t find any null value.
5. Also, we are removing NAN values by using ‘dropna’ method by using ‘axis=0’(row).
6. We are using Encoding categorical features.
7. Used train\_test\_split for splitting data into respective test, train variables.
8. Trained the model using given algorithms (GNB, SVM and, KNN) and calculated its accuracy.
9. Predicted best accuracy algorithm. (SVM=1)
10. Checked the performance of Linear and Non-Linear SVM Kernel and plotted on bar graph.

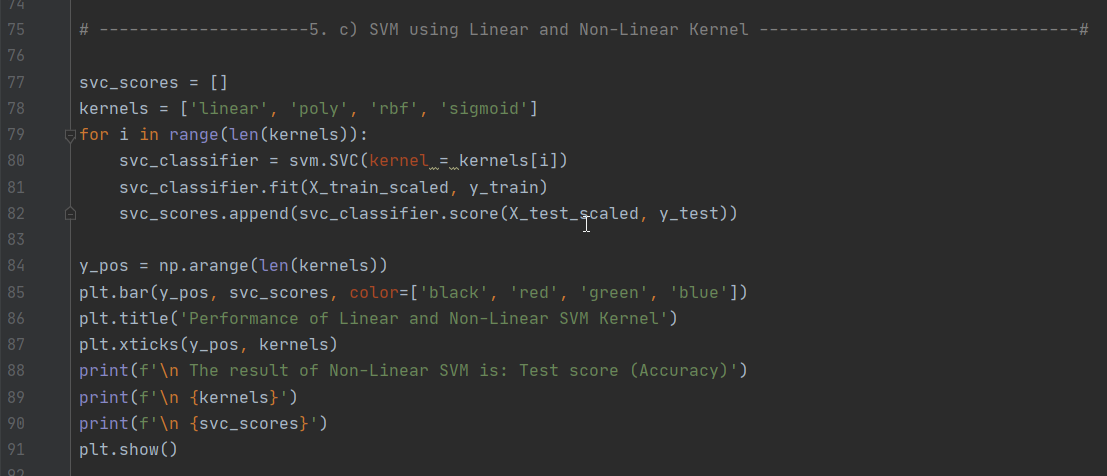
Workflow:



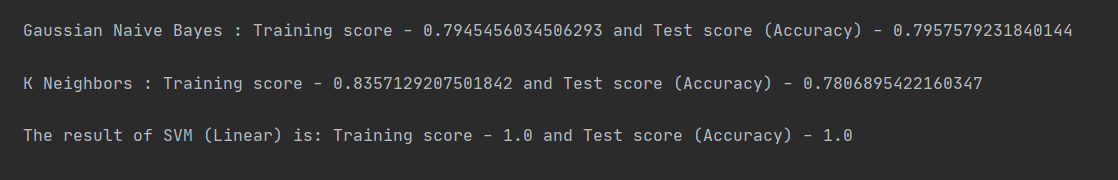


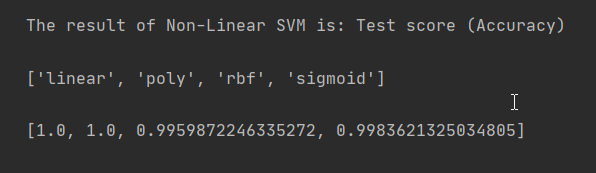


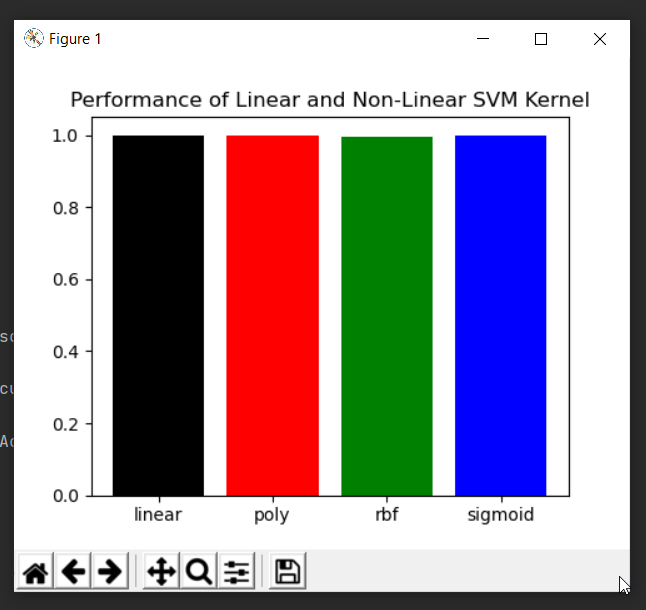




Output:







### **Datasets:**

The Dataset we used in this project was:

2. <https://umkc.app.box.com/s/lwr3s70prbe3tdifzx6fblq7wqln54xq>

5. <https://www.kaggle.com/wenruliu/adult-income-dataset>

### **Parameters:**

In this project we have used multiple parameters in all questions like,

n\_clusters, init, df.iloc, X\_train, X\_test, y\_train, y\_test, n\_neighbors, kernel, color, axis, inplace, style, color\_codes, ascending, linewidth, marker, figsize etc.

### **Evaluation & Discussion:**

Evaluation of all the questions are shown in Workflow feature.

All the team members have discussed on all the dataset that we used, solving questions and its evaluation.

### **Conclusion:**

We learned how to apply machine learning concepts and algorithms to capture required data from dataset and predict the respective output. By solving all the questions, we got enough knowledge of Python and Deep Learning.