# Machine Learning

**SIES (NERUL) COLLEGE OF ARTS SCIENCE AND**

**COMMERCE**

**NAAC ACCREDITED ‘A’ GRADE COLLEGE** **(ISO**

**9001:2015 CERTIFIED INSTITUTION)**

# NERUL, NAVI MUMBAI - 400706



**DEPARTMENT OF INFORMATION**

**TECHNOLOGY**

**PRACTICAL BOOK ON**

**Machine Learning**

**SUBMITTED TO,**

**UNIVERSITY OF MUMBAI**

**BY**

# VISHAL VIJAY VARMA

**MASTER OF INFORMATION**

**TECHNOLOGY**

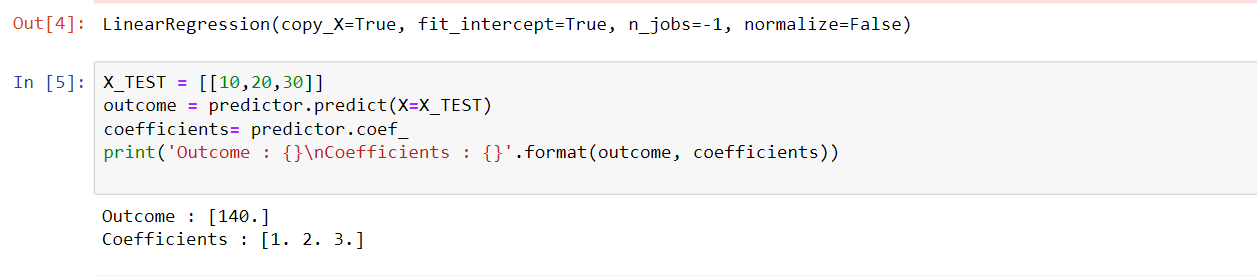
**PART - II (SEMESTER - III)**

**(2022-2023)**

**INDEX**

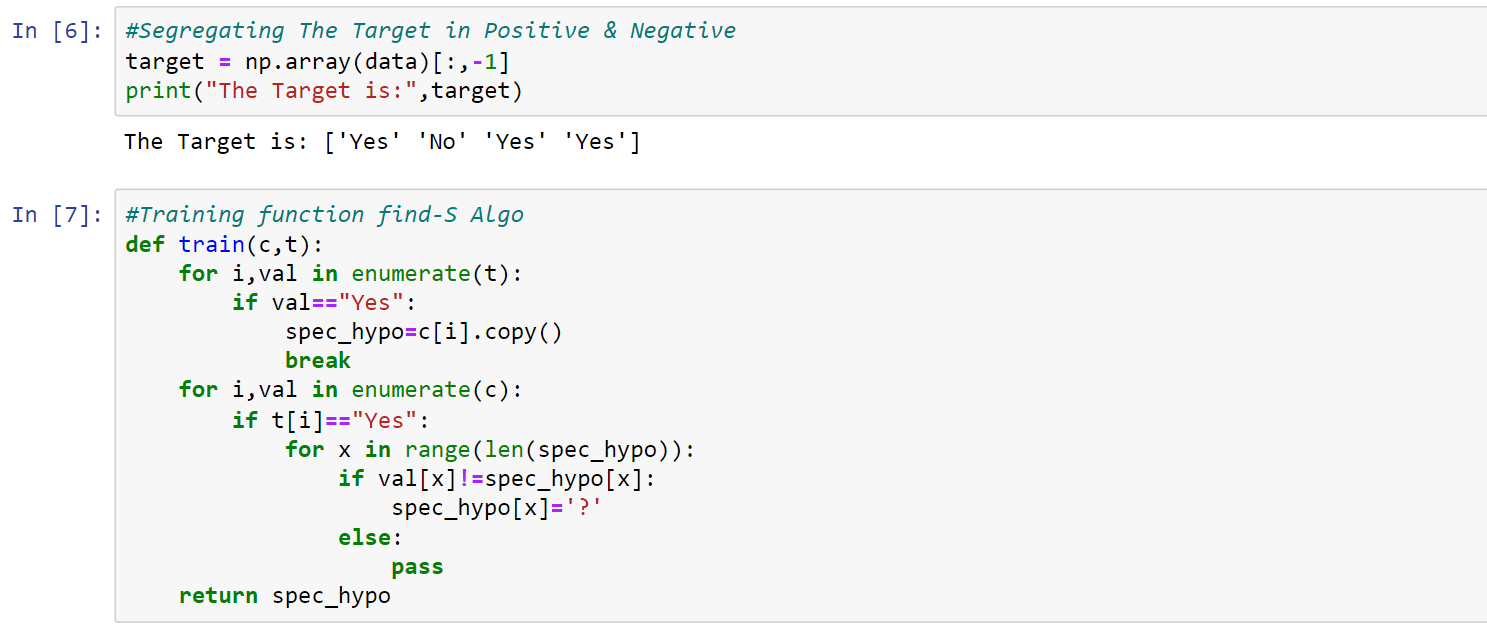
|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Practical** | | **Page No** |
|  | **1A** | Design a simple machine learning model to train the training instances and test the same. |  |
|  | **1B** | Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file |  |
|  | **2A** | Perform Data Loading, Feature selection (Principal Component analysis) and Feature Scoring and Ranking. |  |
|  | **2B** | For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples. |  |
|  | **3A** | Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets. |  |
|  | **3B** | Write a program to implement Decision Tree and Random forest with Prediction, Test Score and Confusion Matrix. |  |
|  | **4A** | For a given set of training data examples stored in a .CSV file implement Least Square Regression algorithm. |  |
|  | **4B** | For a given set of training data examples stored in a .CSV file implement Logistic Regression algorithm. |  |
|  | **5A** | Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample. |  |
|  | **5B** | Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. |  |
|  | **6A** | Implement the different Distance methods (Euclidean) with Prediction, Test Score and Confusion Matrix. |  |
|  | **6B** | Implement the classification model using clustering for the following techniques with K means clustering with Prediction, Test Score and Confusion Matrix. |  |
|  | **7A** | Implement the classification model using clustering for the following techniques with hierarchical clustering with Prediction, Test Score and Confusion Matrix |  |
|  | **7B** | Implement the Rule based method and test the same. |  |
|  | **8A** | Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. |  |
|  | **8B** | Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs. |  |
|  | **9A** | Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets. |  |
|  | **9B** | Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task. |  |
|  | **10A** | Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample. |  |
|  | **10B** | Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets. |  |

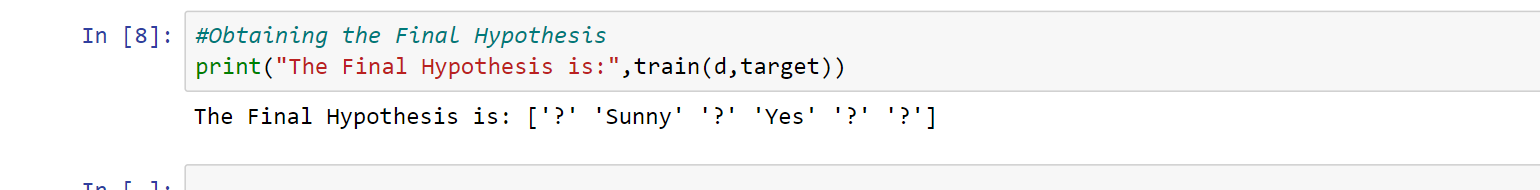




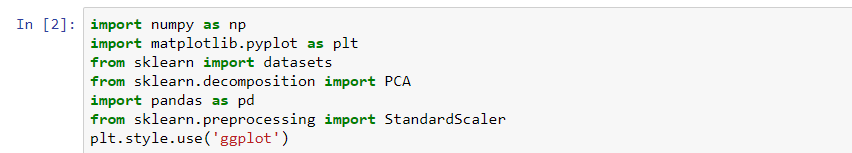
**1 B :- Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file**

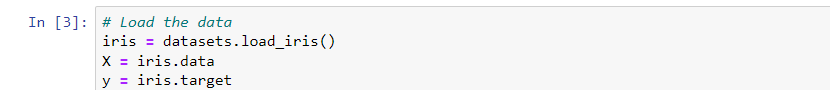


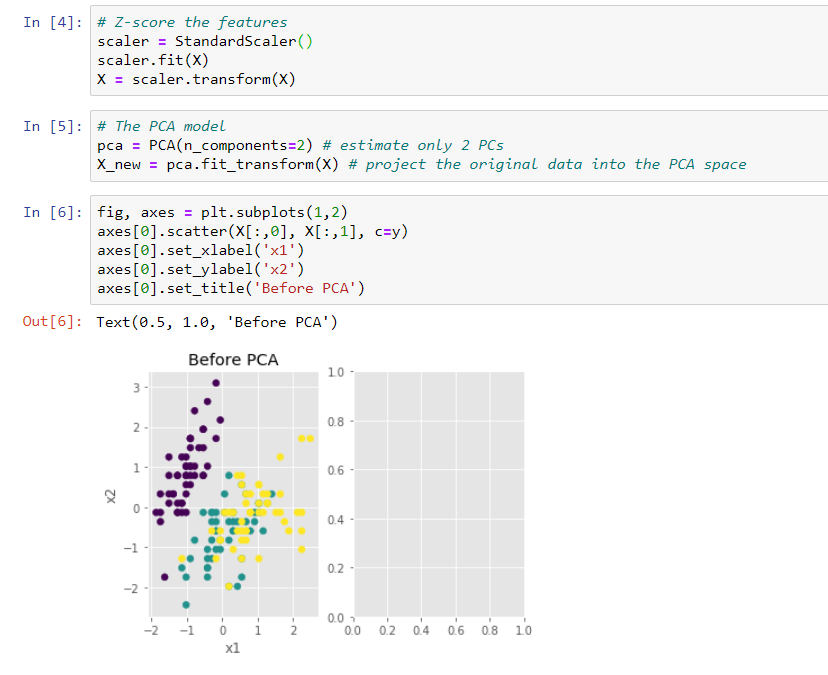


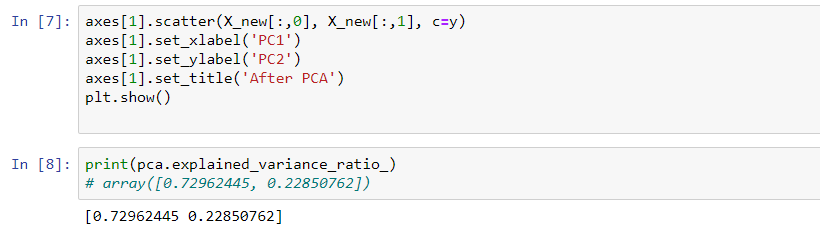


**2 A**







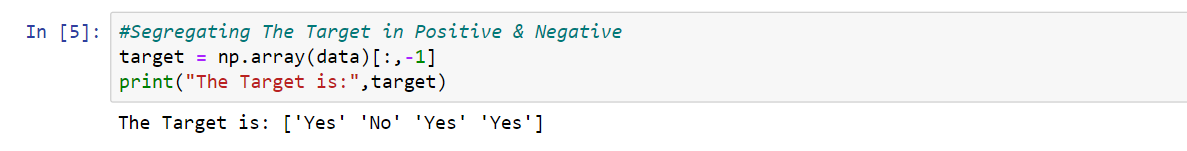




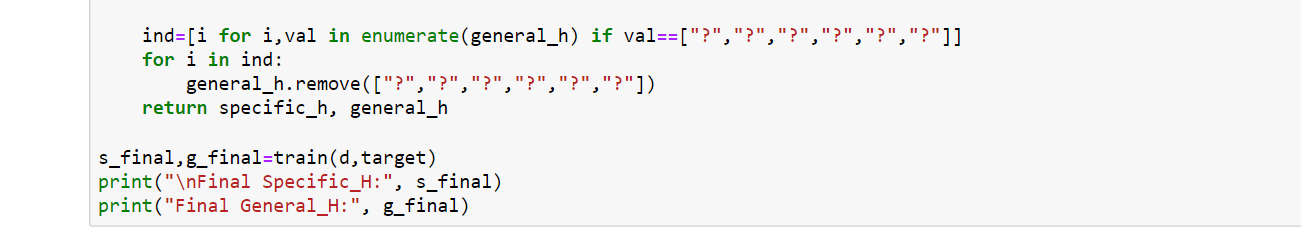


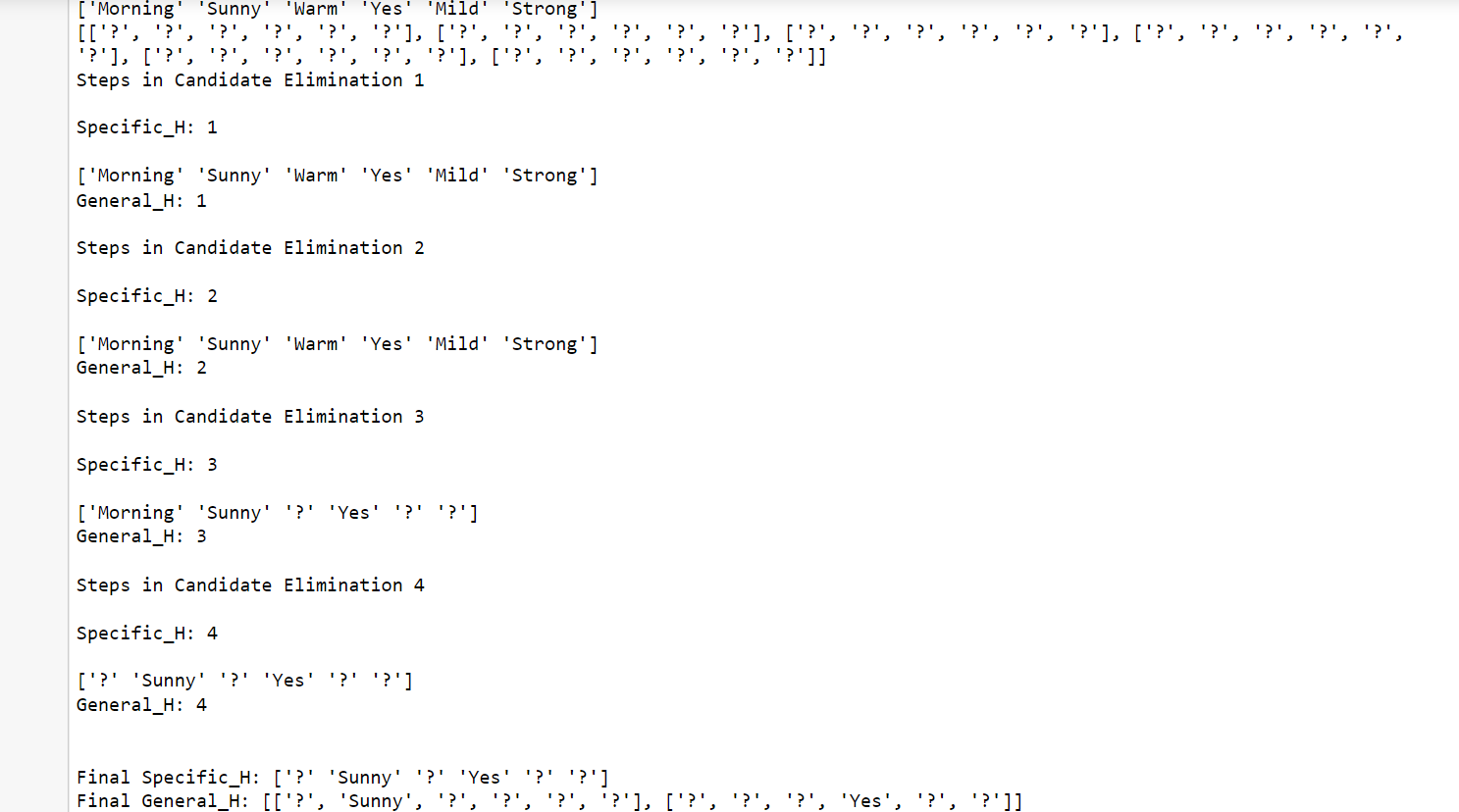
**2 B :- For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples**



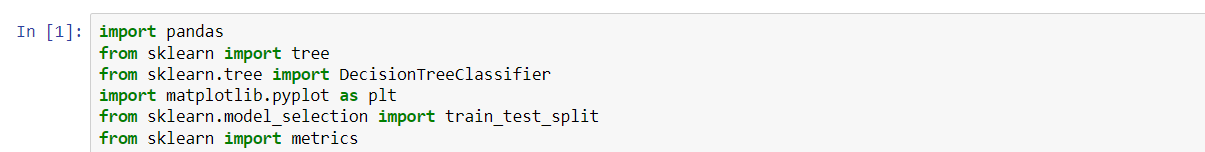
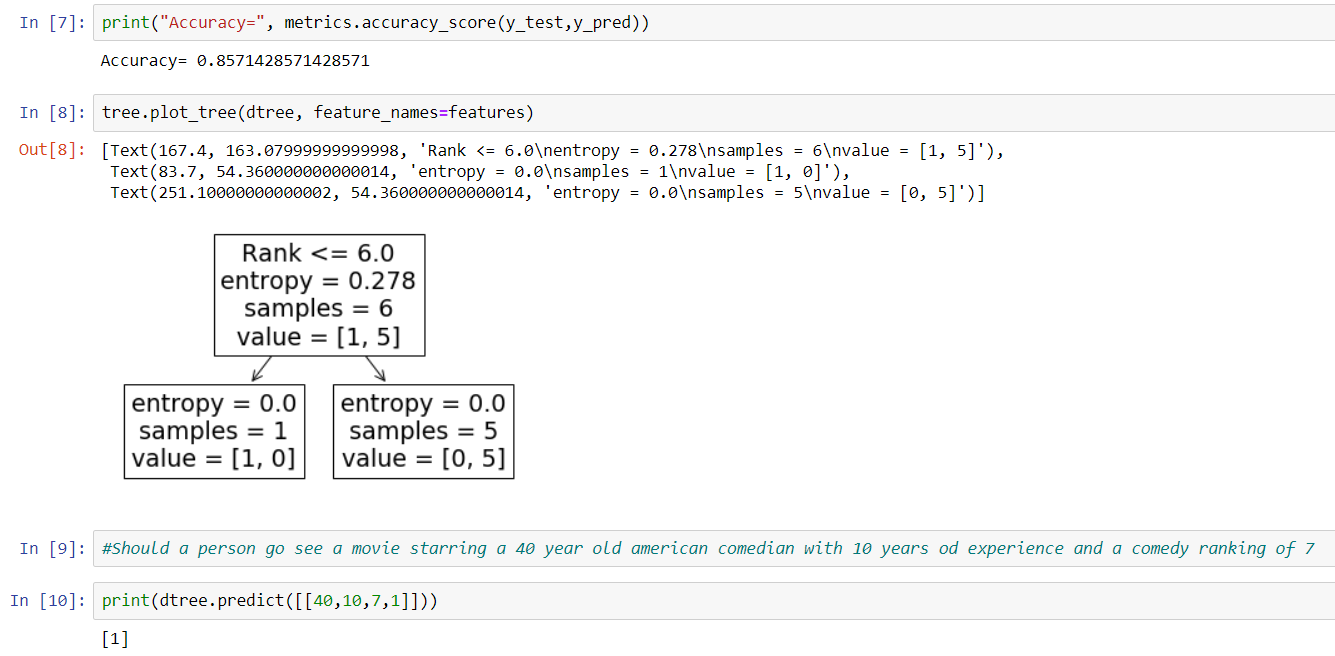


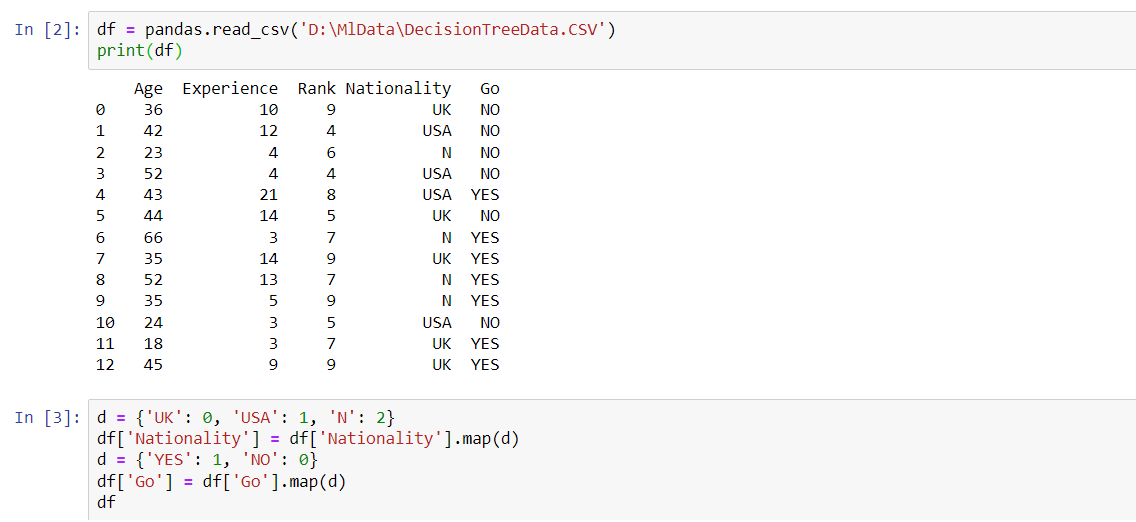


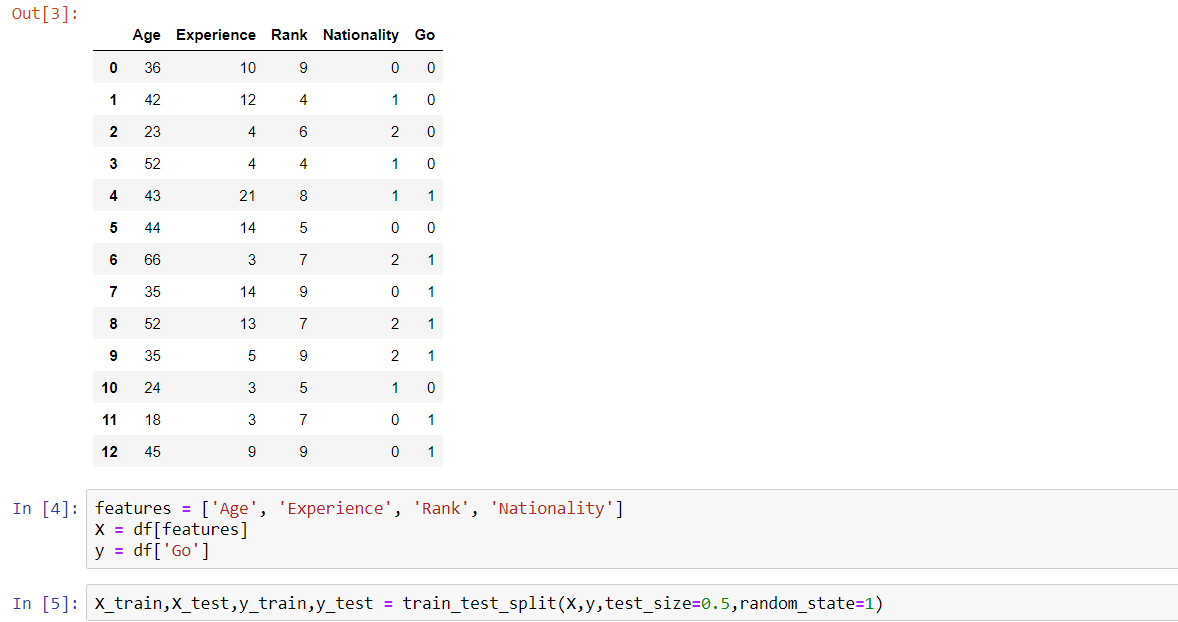


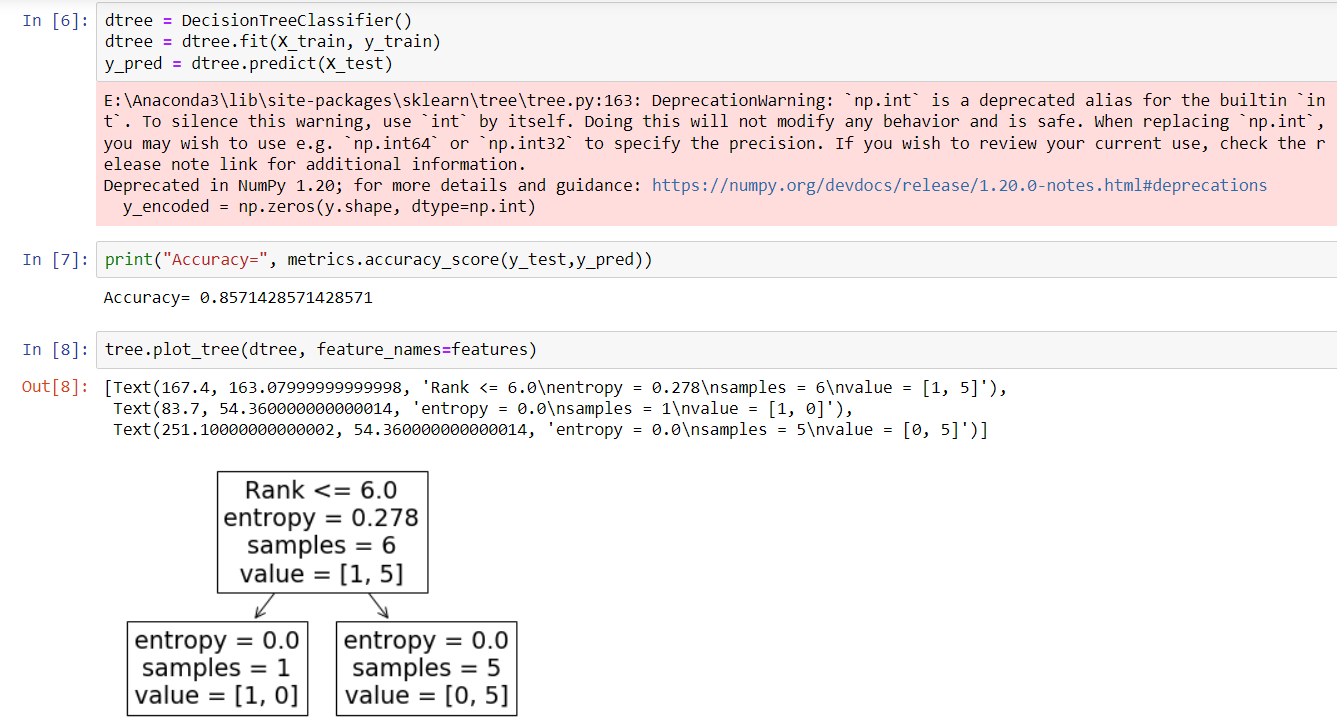


**3A**

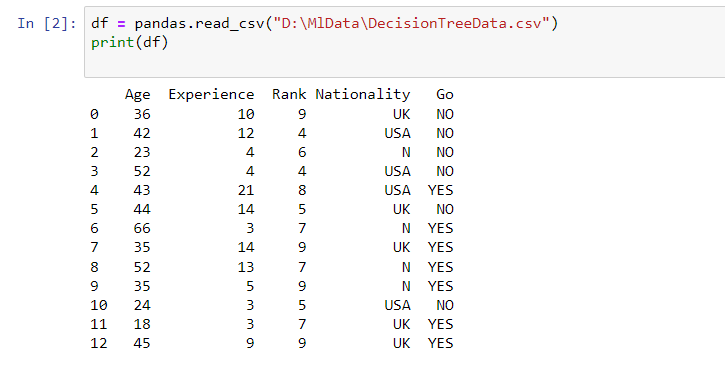


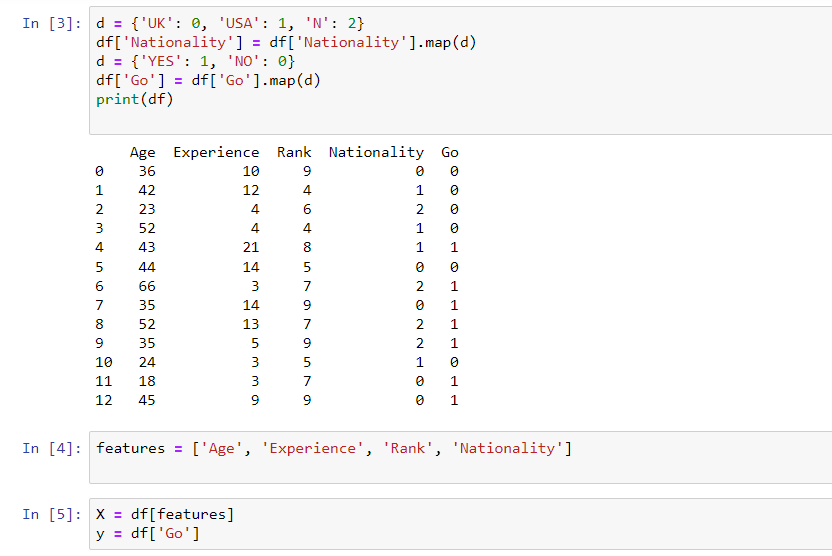


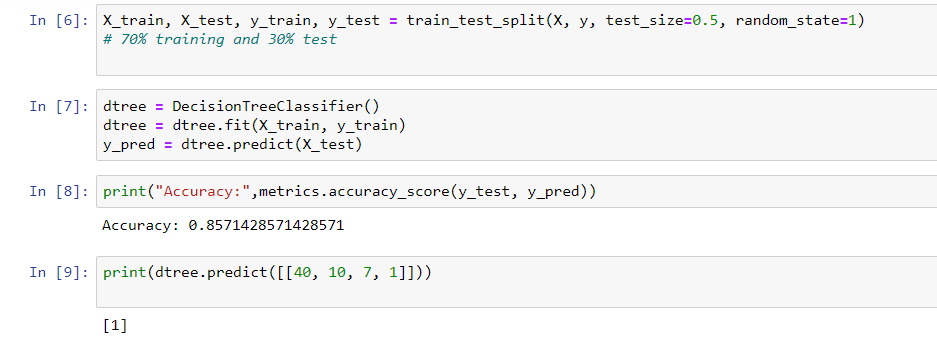


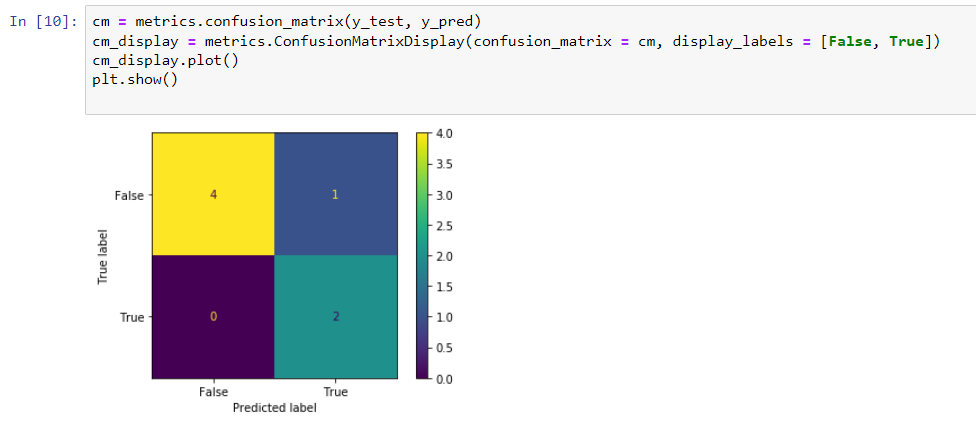


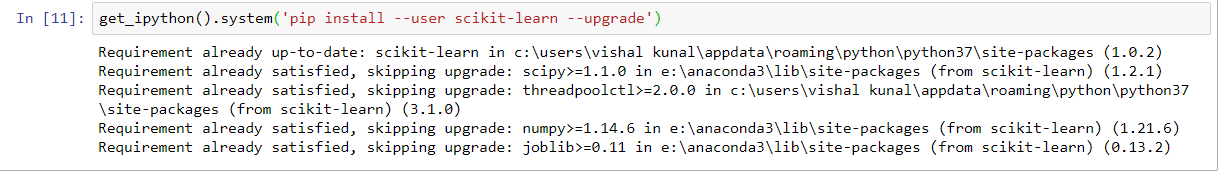
**3 B :- Write a program to implement Decision Tree and Random forest with Prediction, Test Score and Confusion Matrix.**



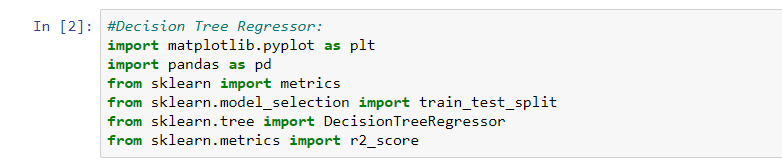


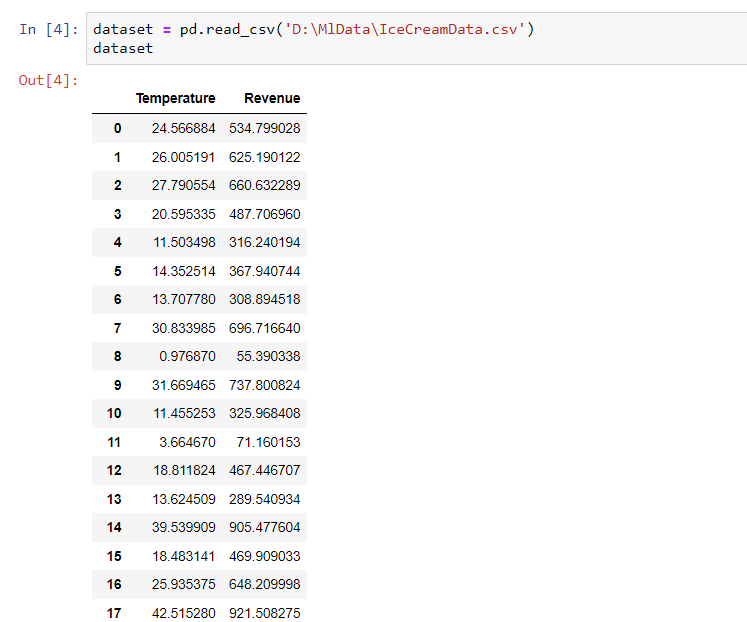


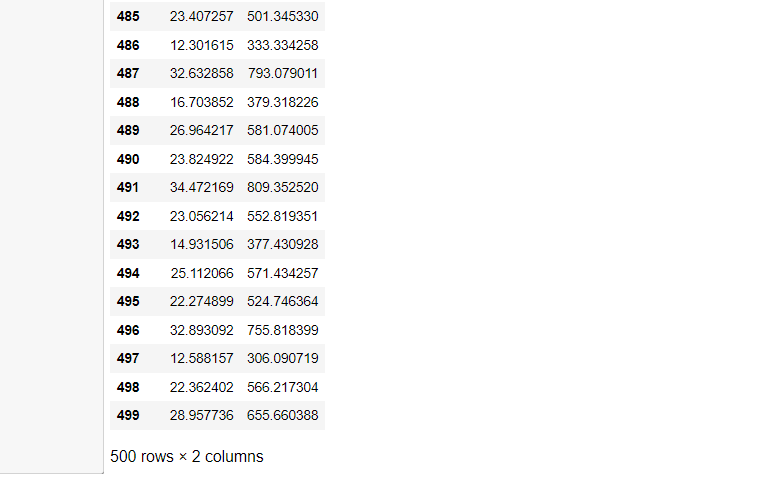


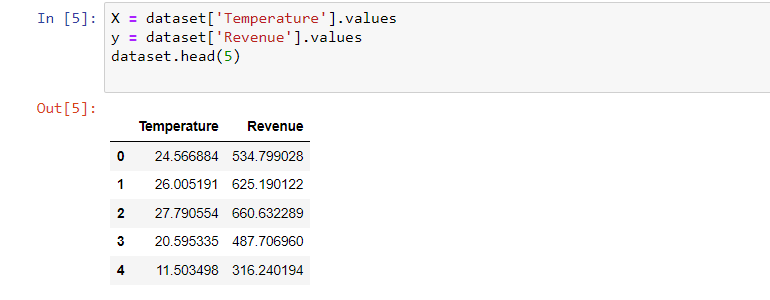


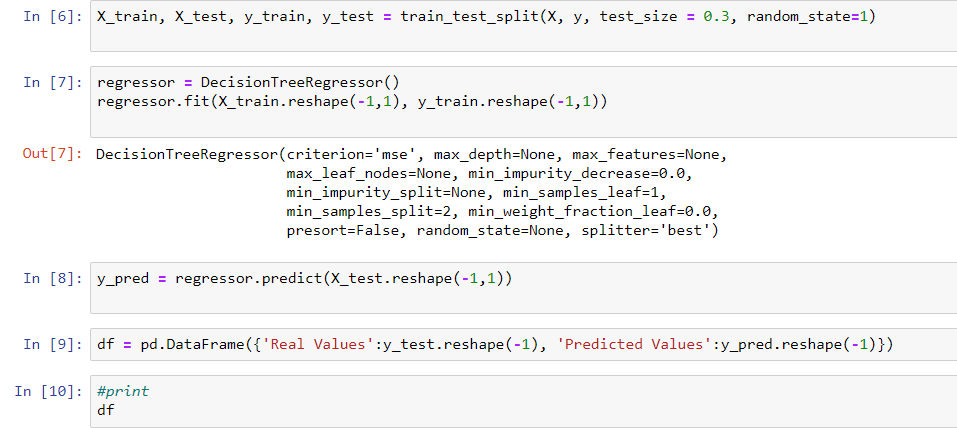
**DECISION TREE**

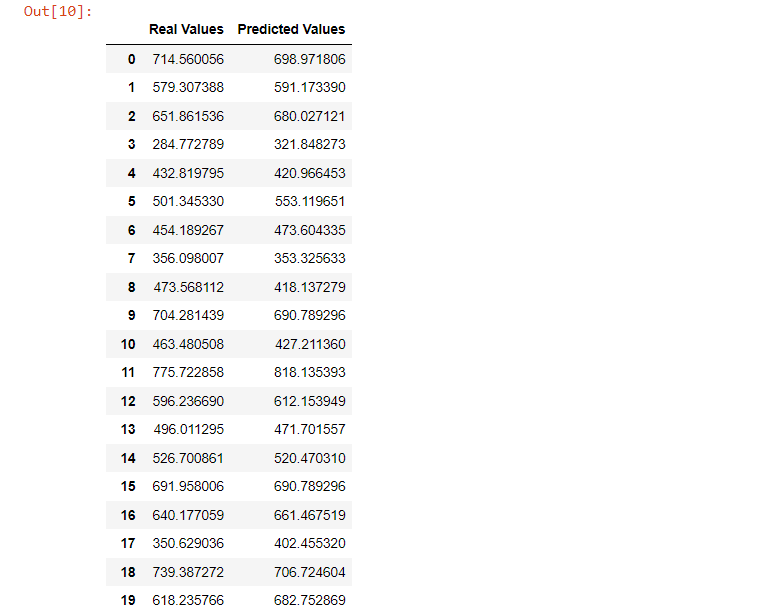


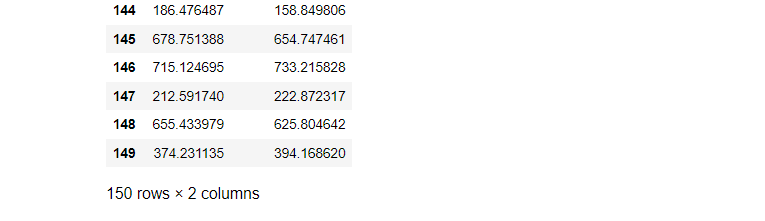


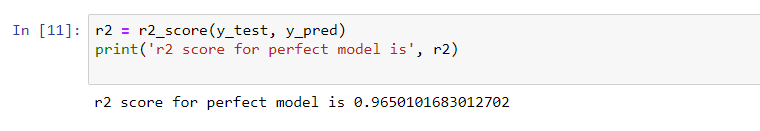




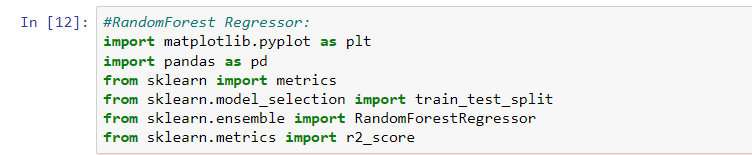


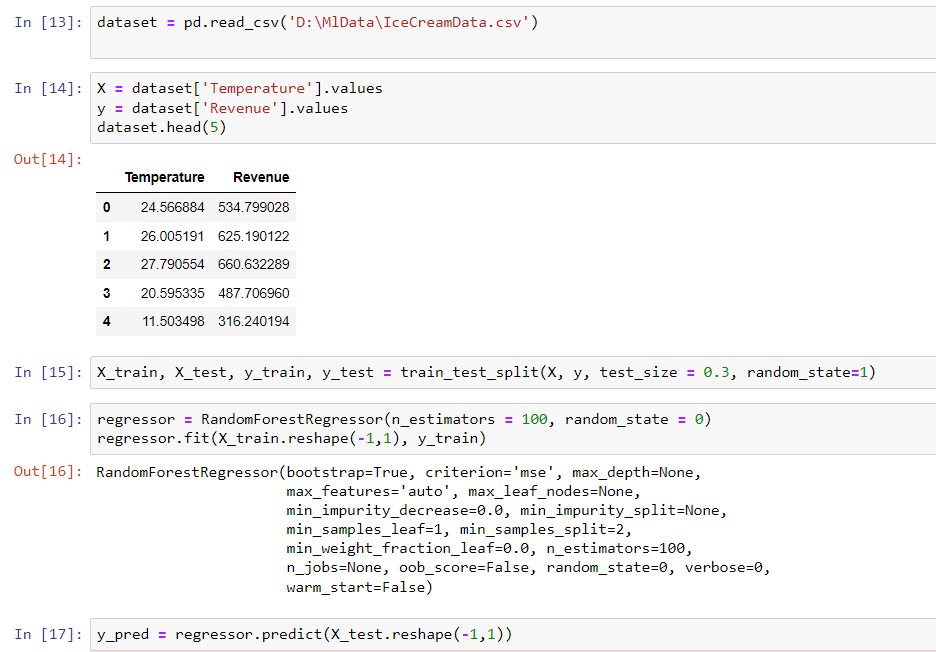




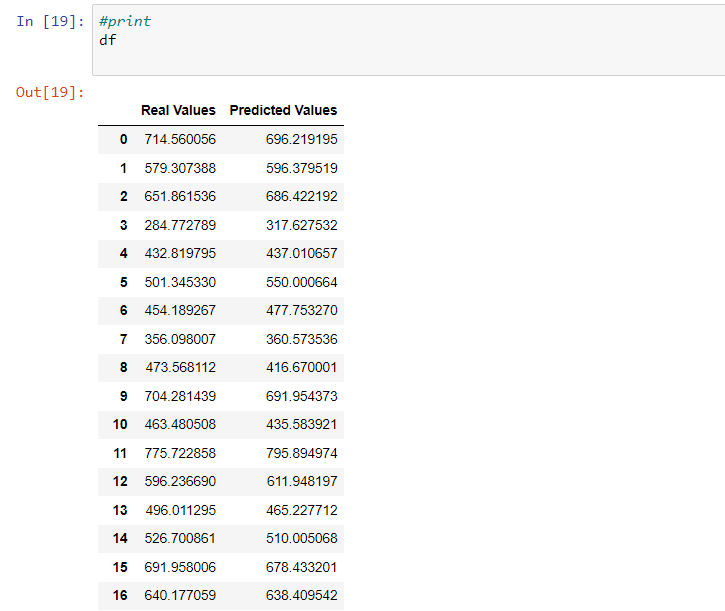


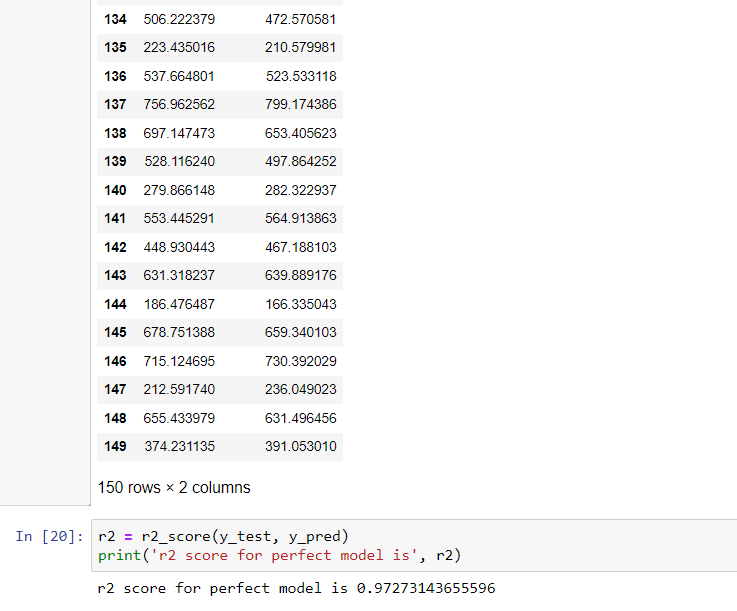
**RANDOM FOREST**



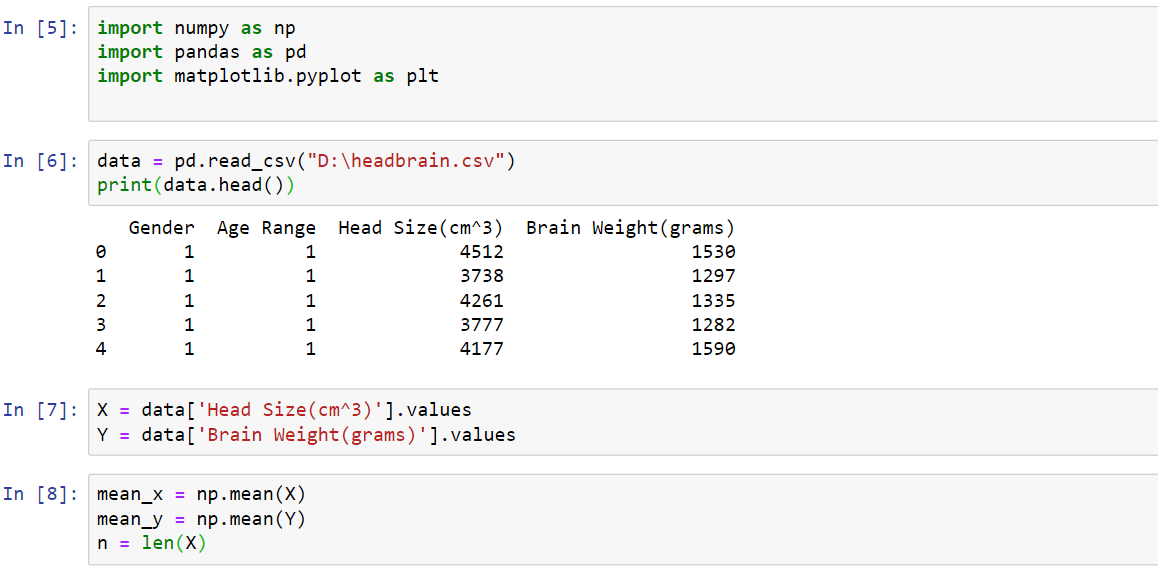


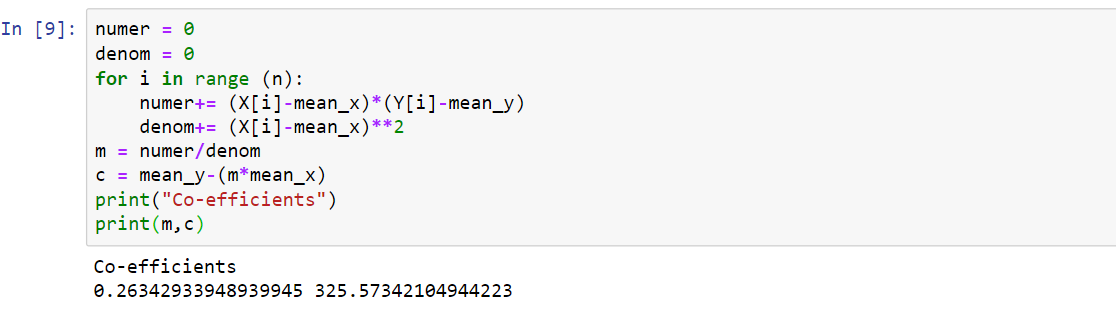


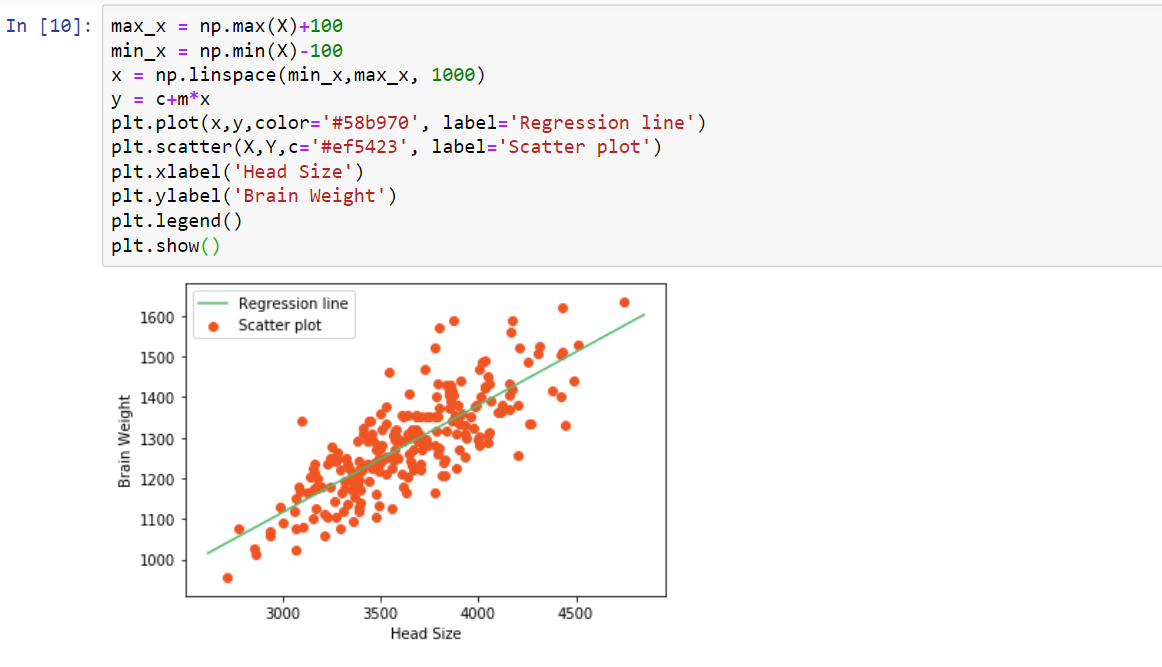




**4A**

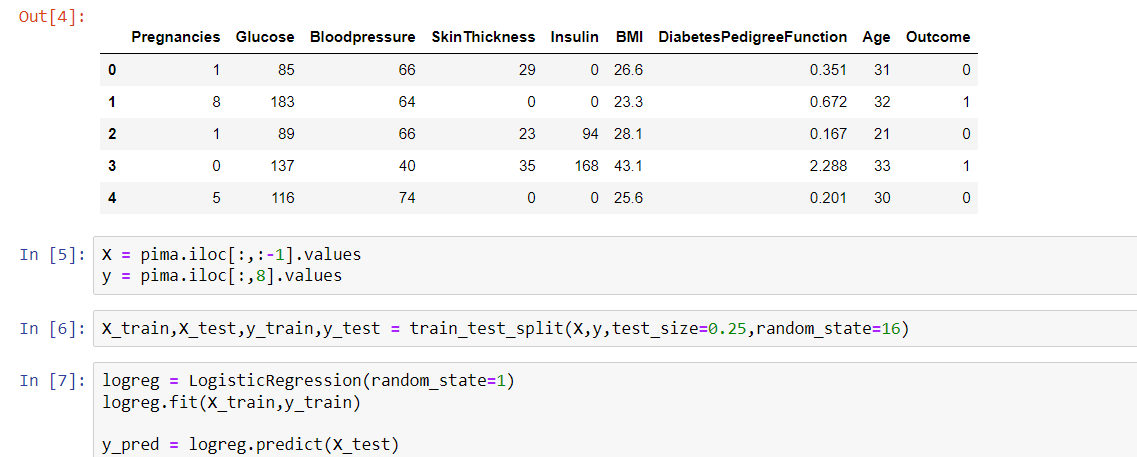


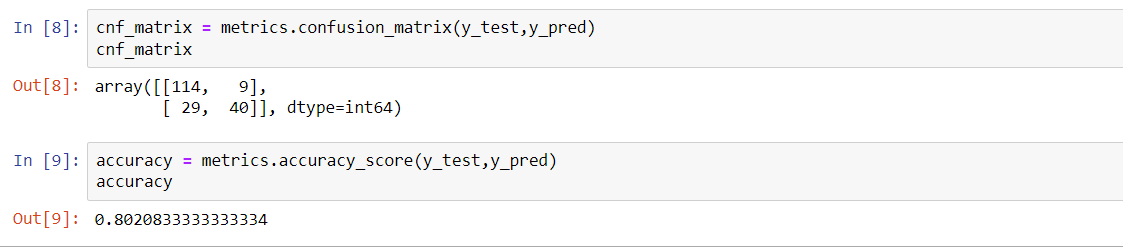




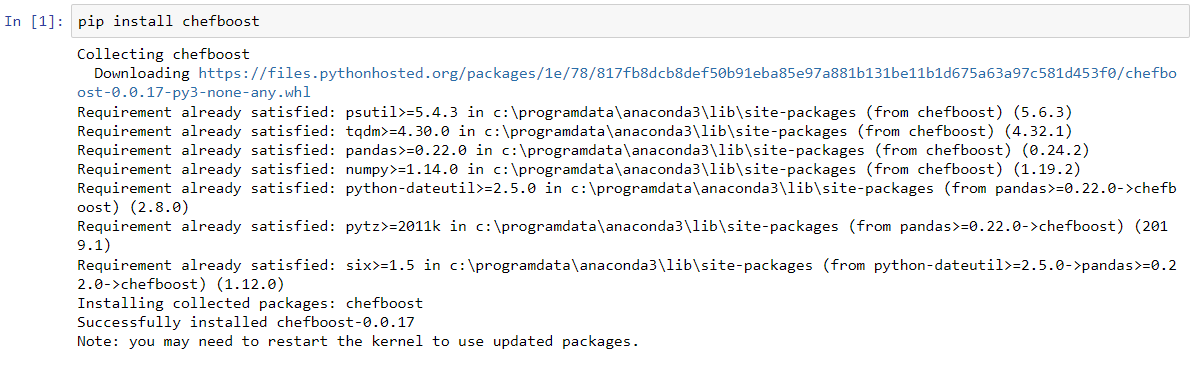
**4 B :- For a given set of training data examples stored in a .CSV file implement Logistic Regression algorithm.**

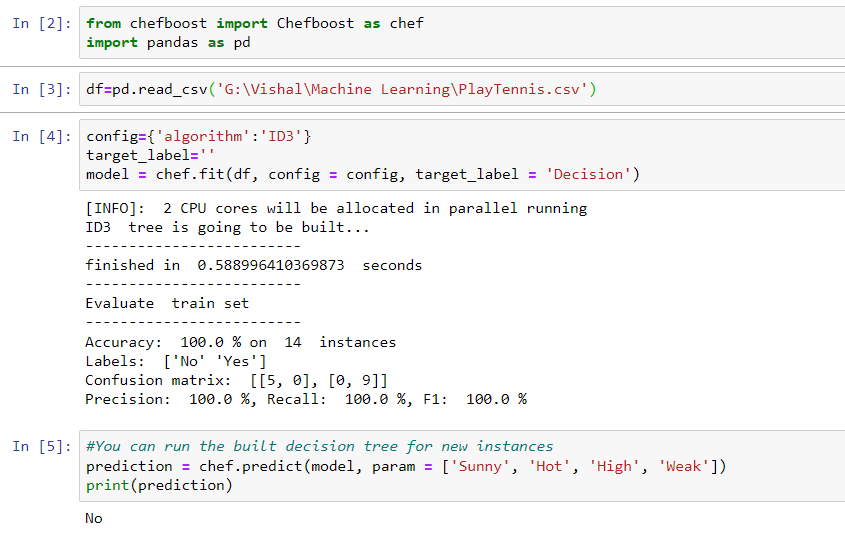


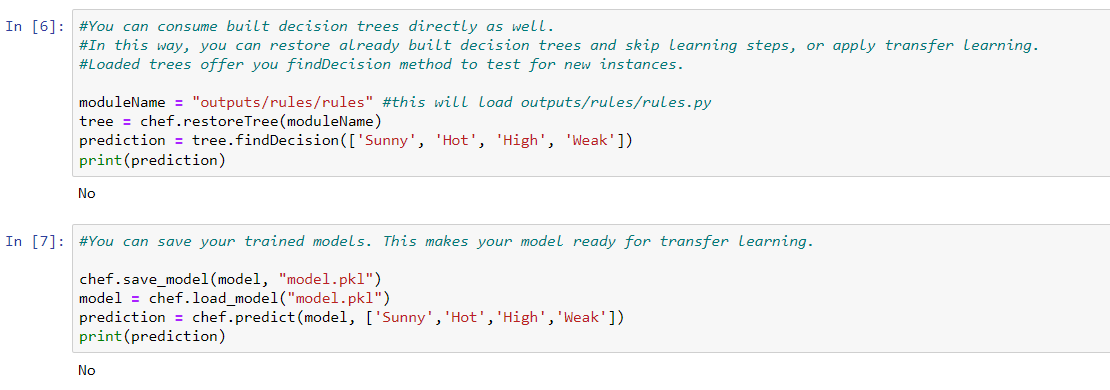




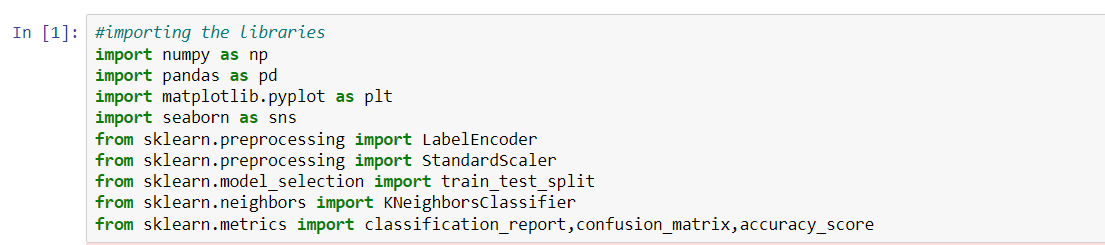
**5A**







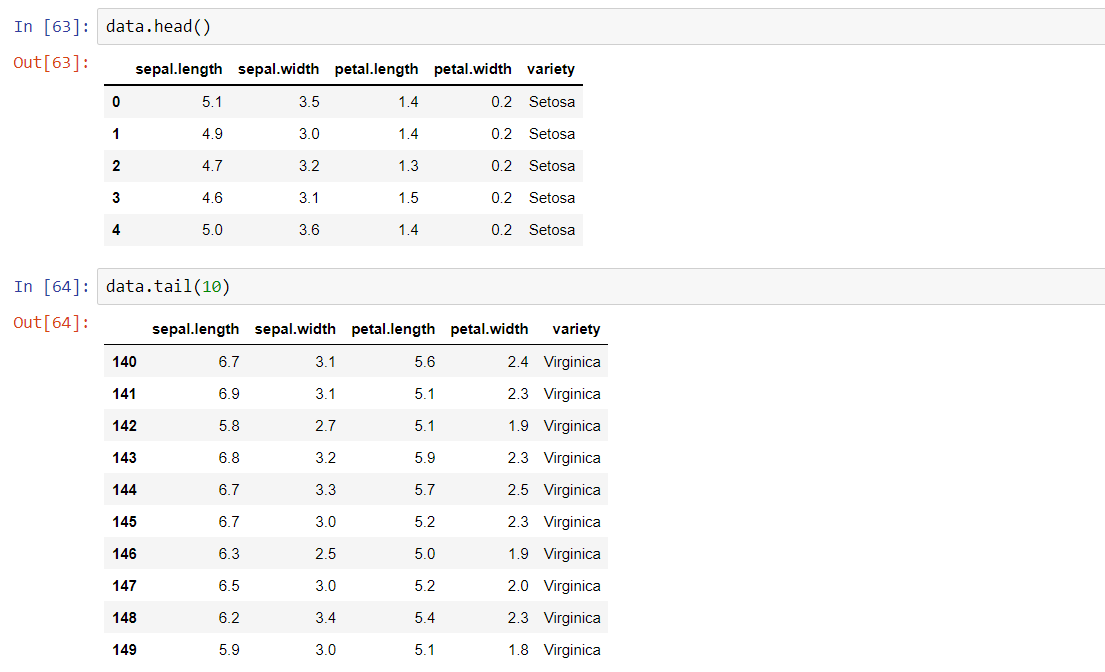
**5 B :- Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set.**

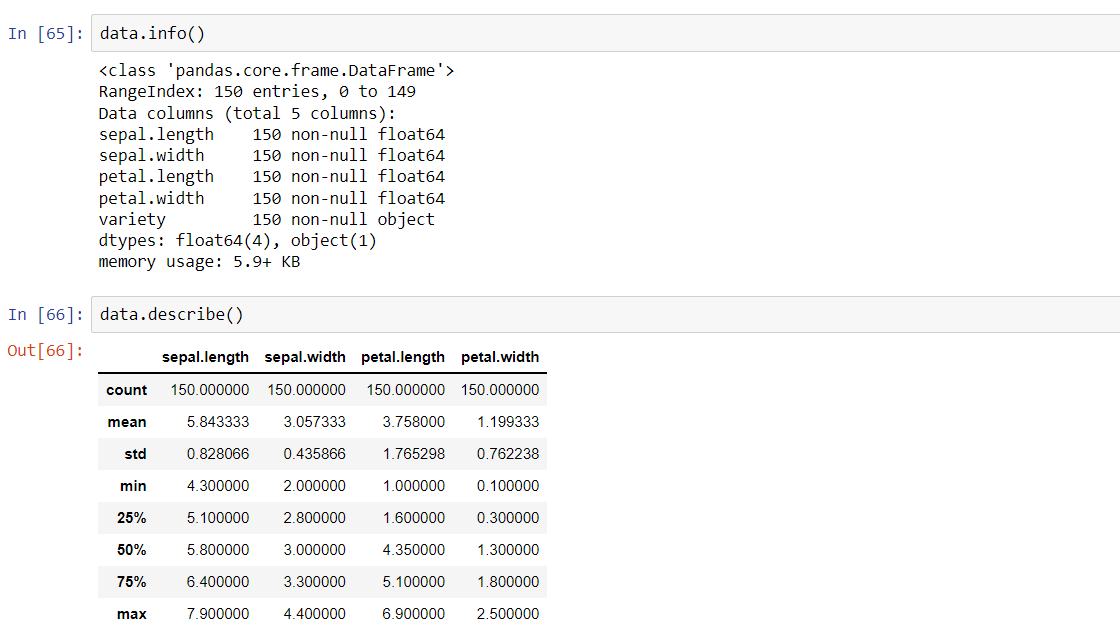


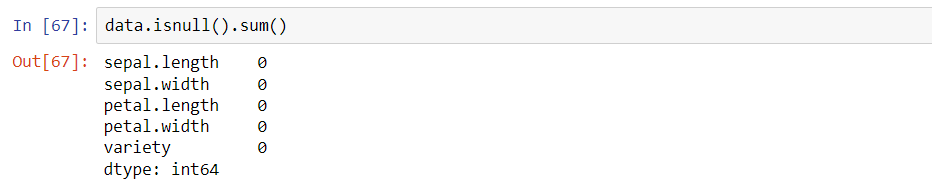






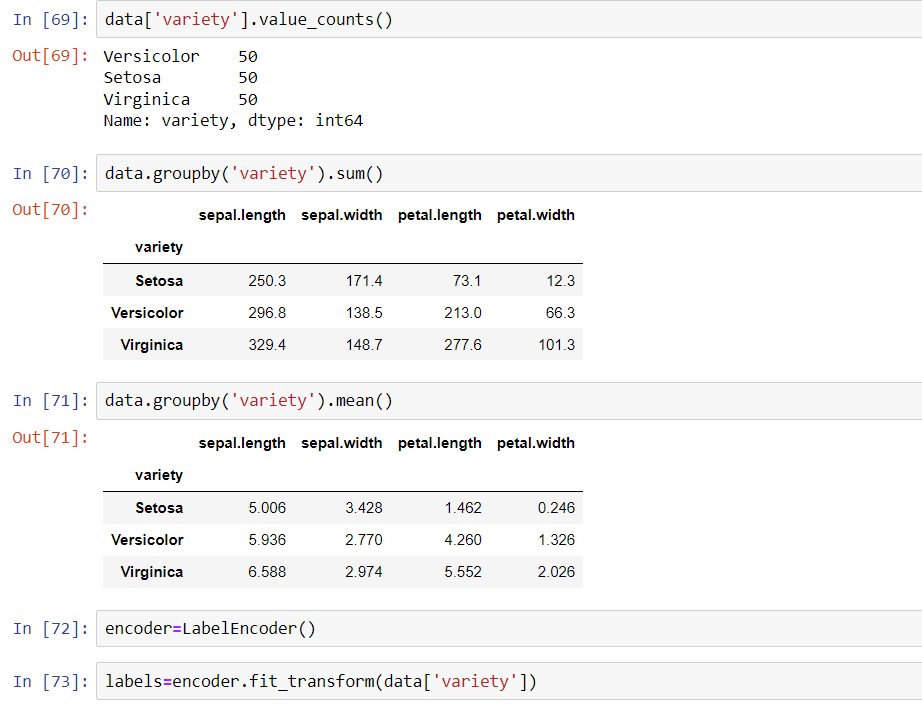


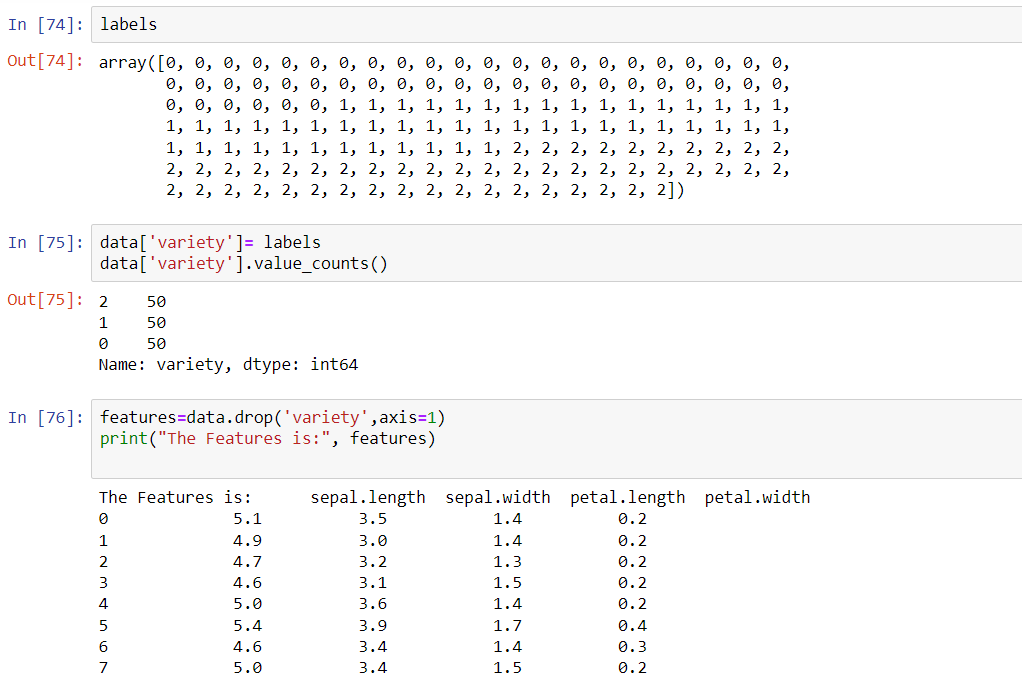




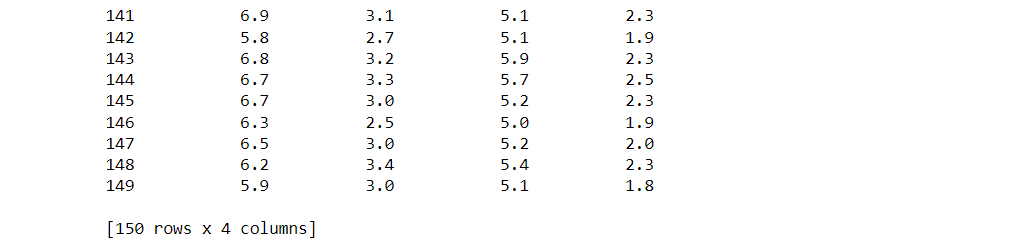




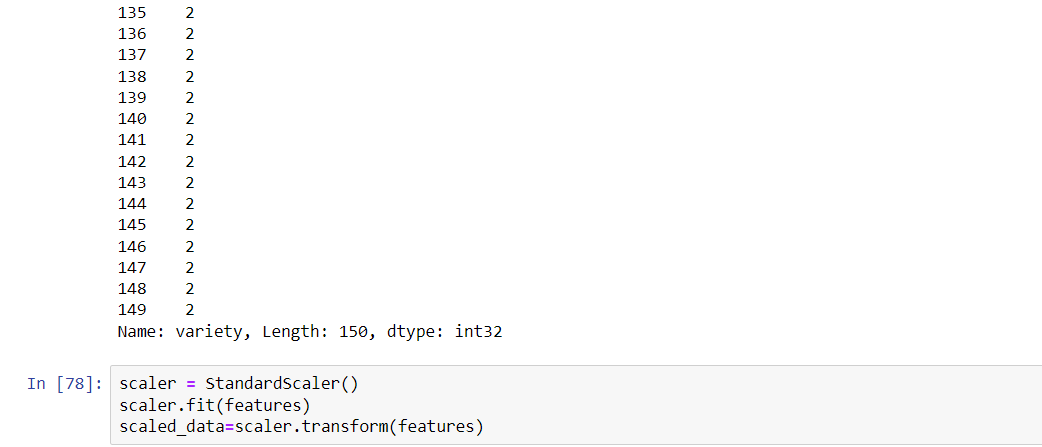


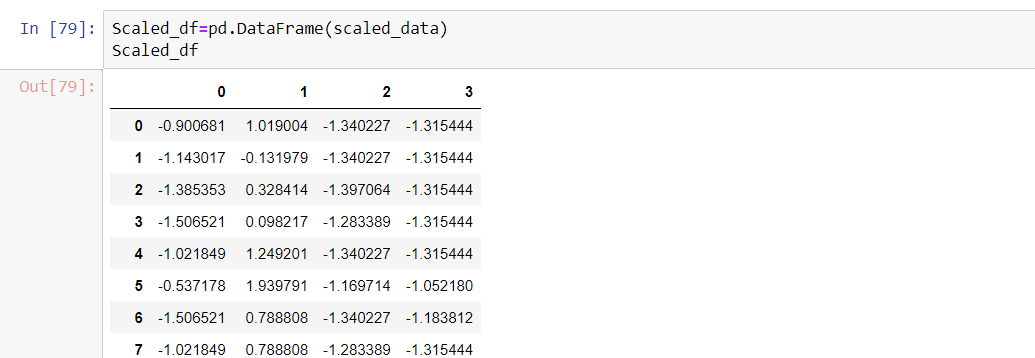




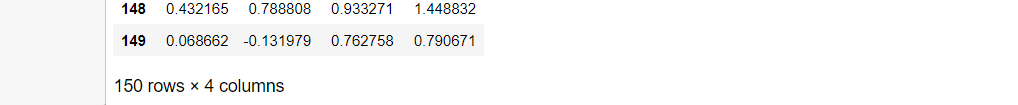


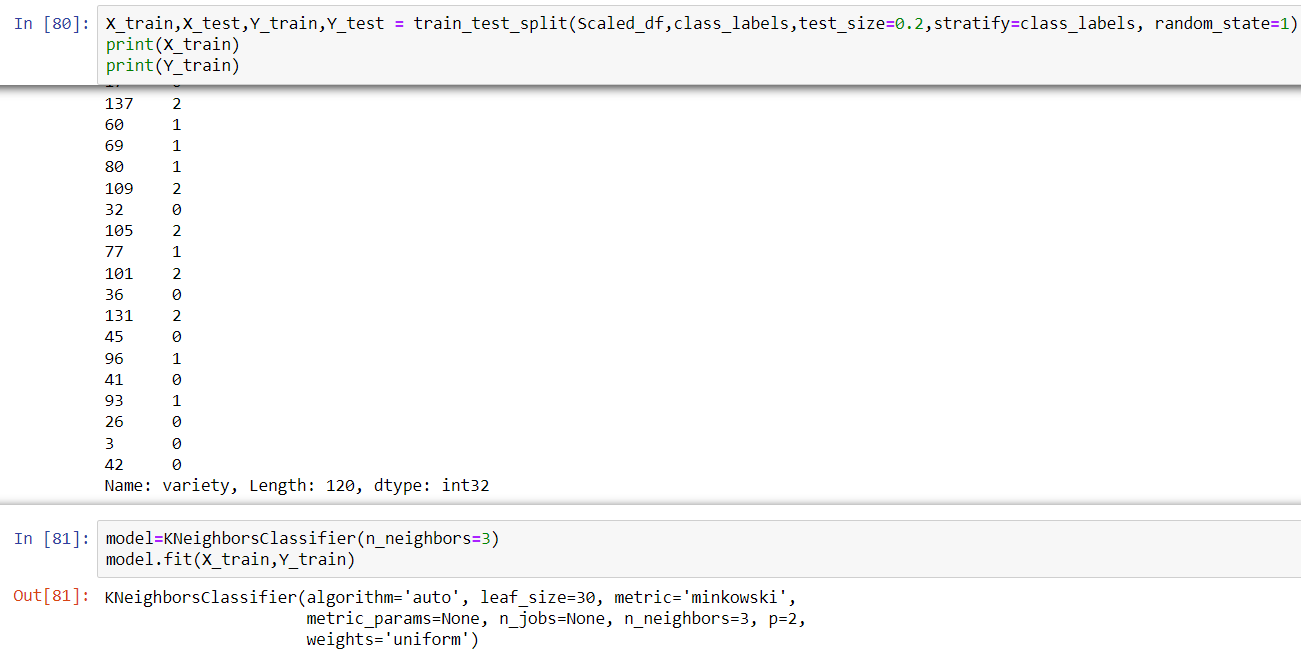


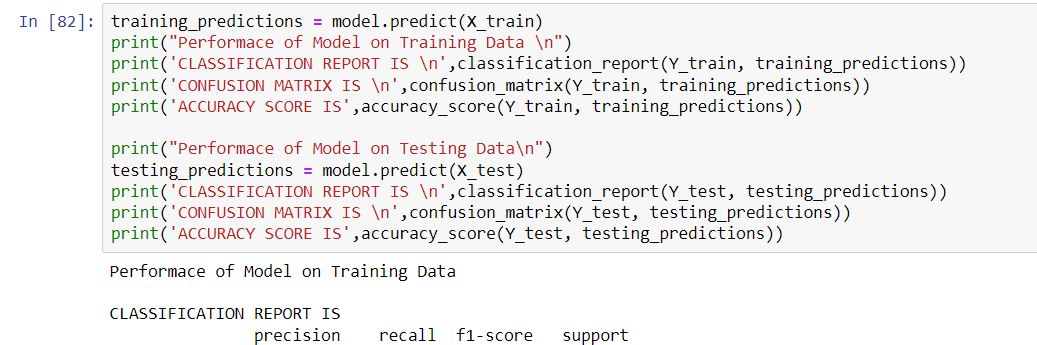


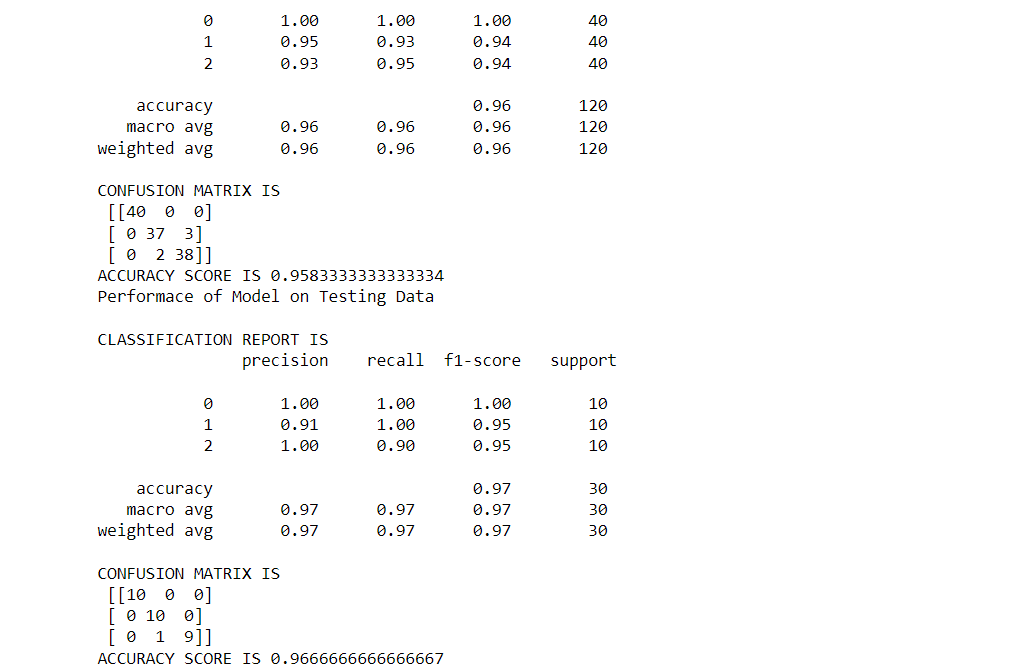


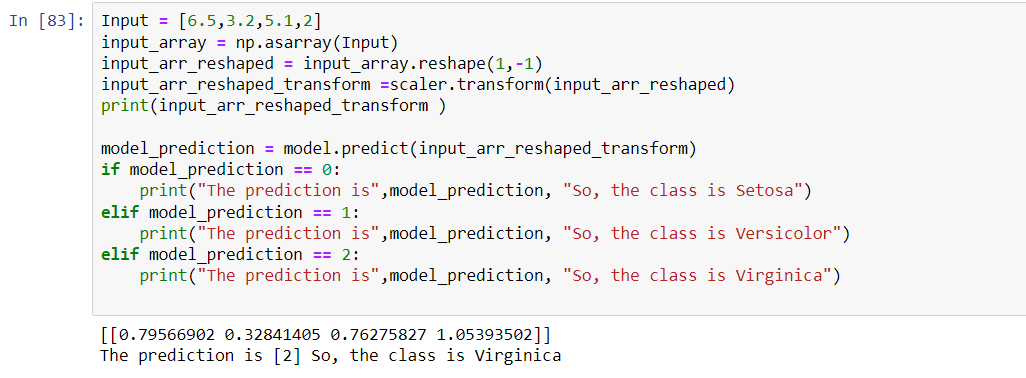






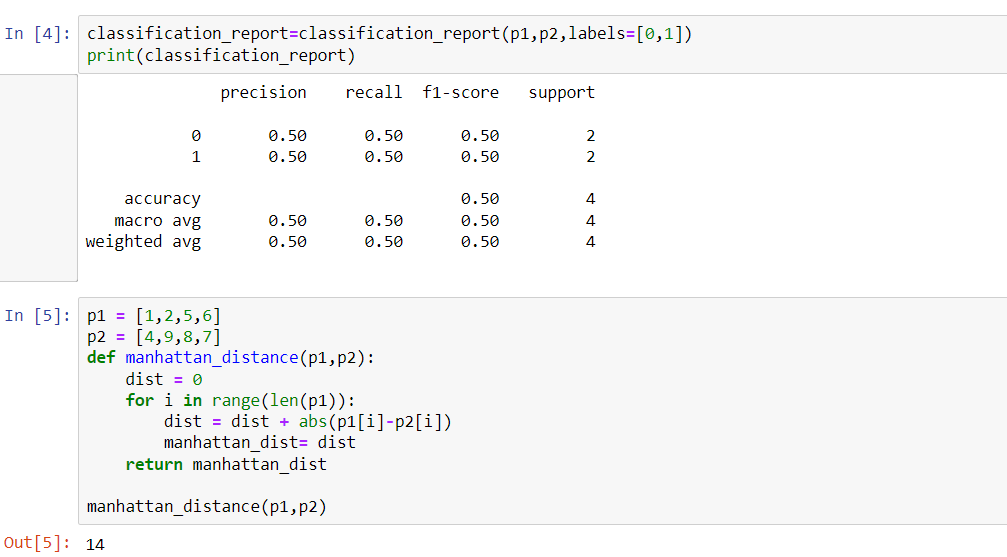


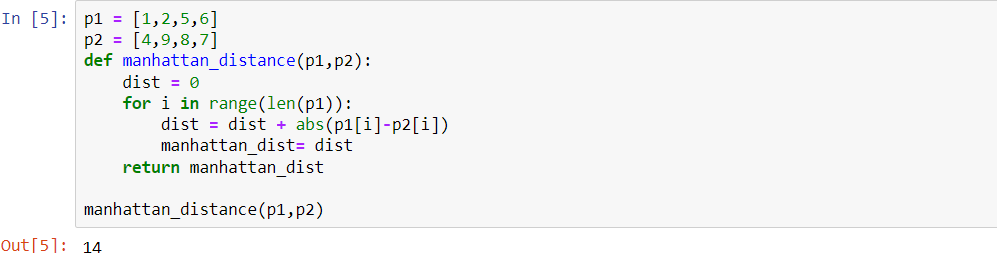


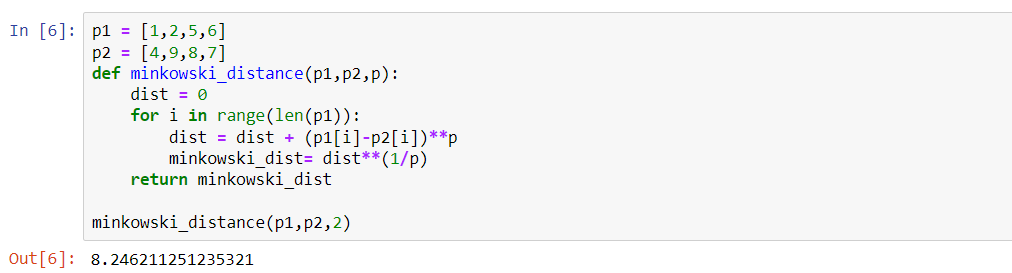


**6A**

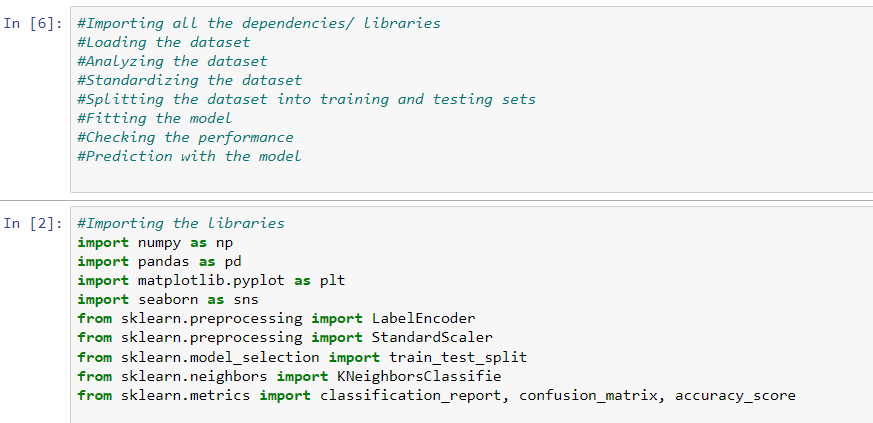


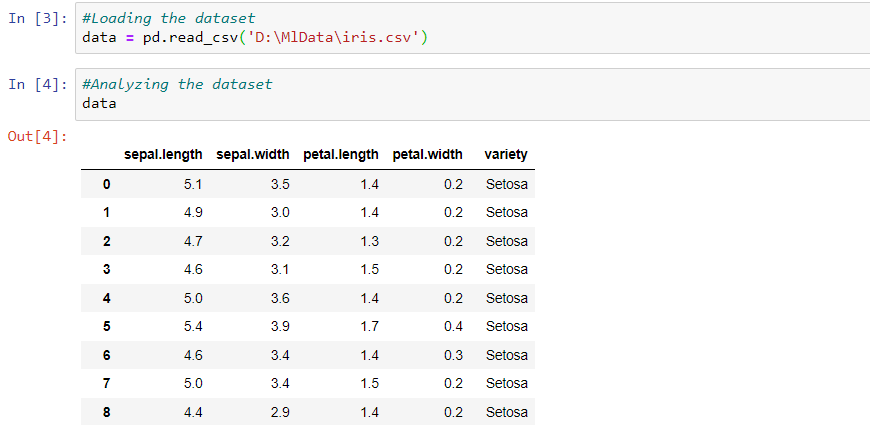




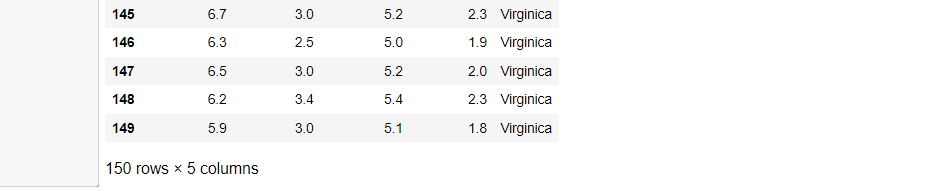


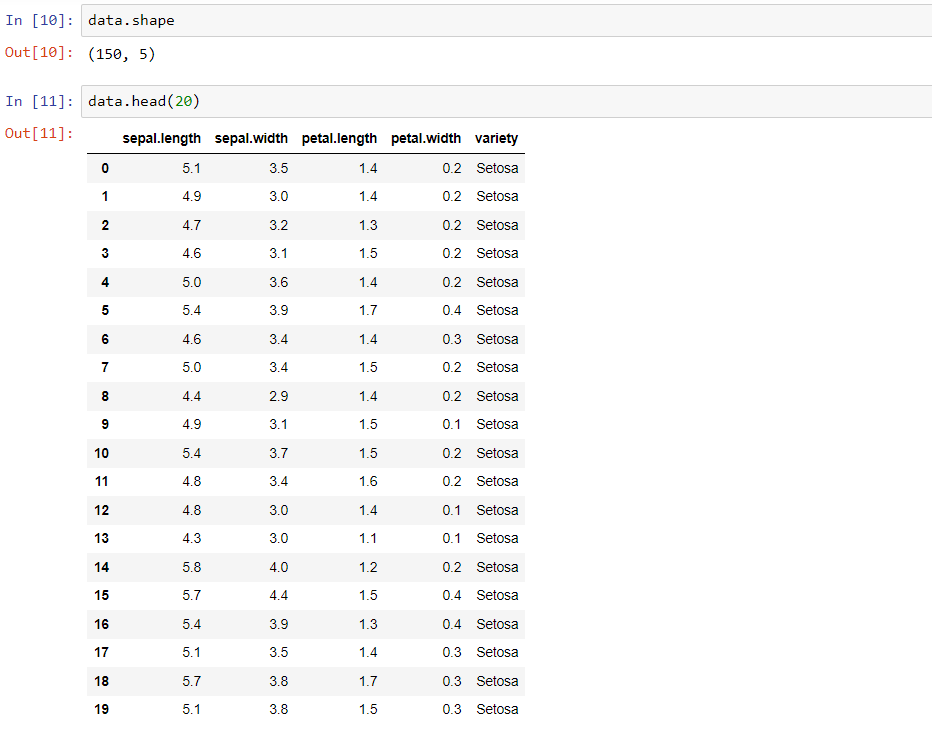
**6 B :- Implement the classification model using clustering for the following techniques with K means clustering with Prediction, Test Score and Confusion Matrix.**

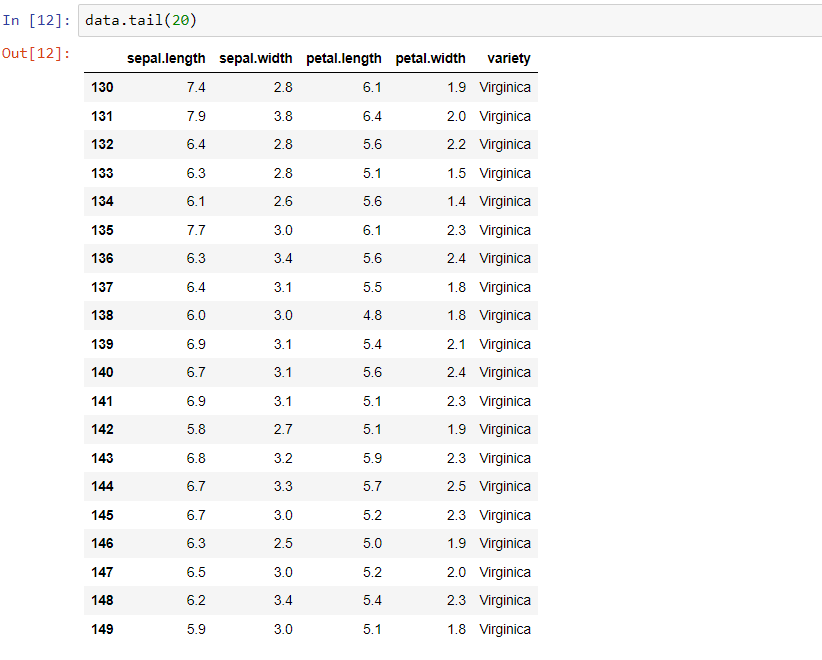


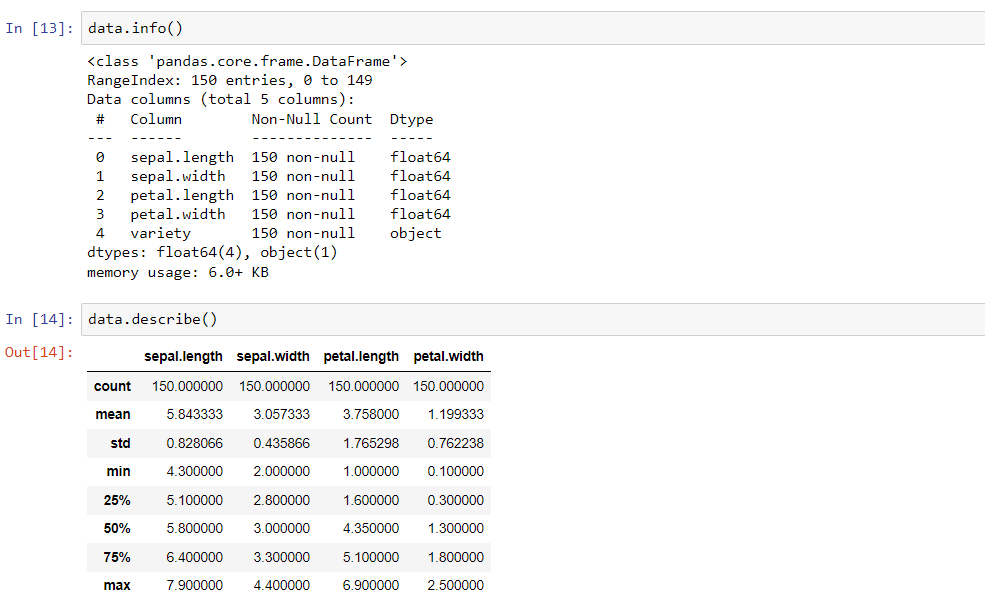


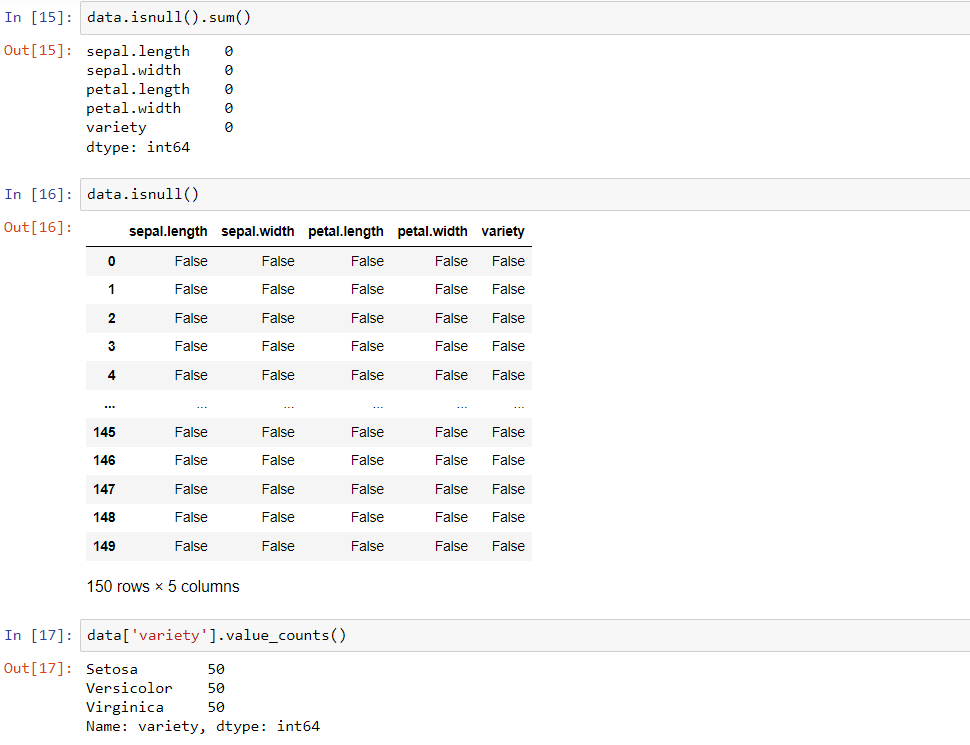


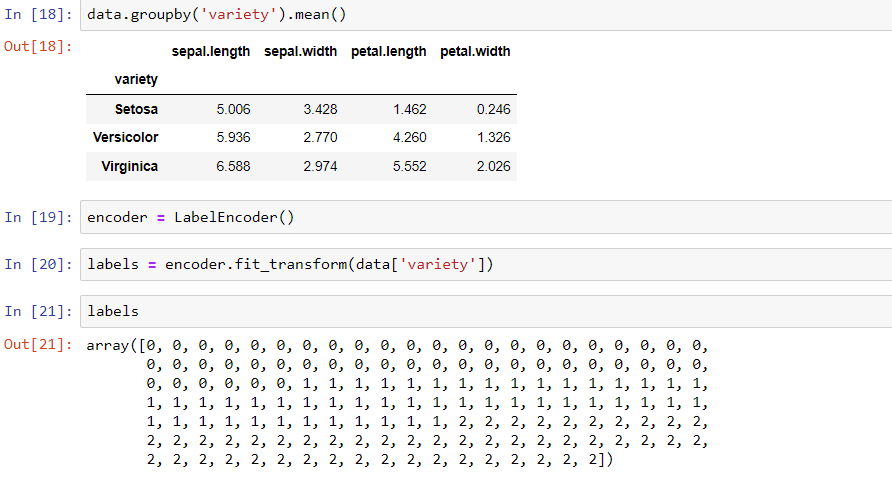


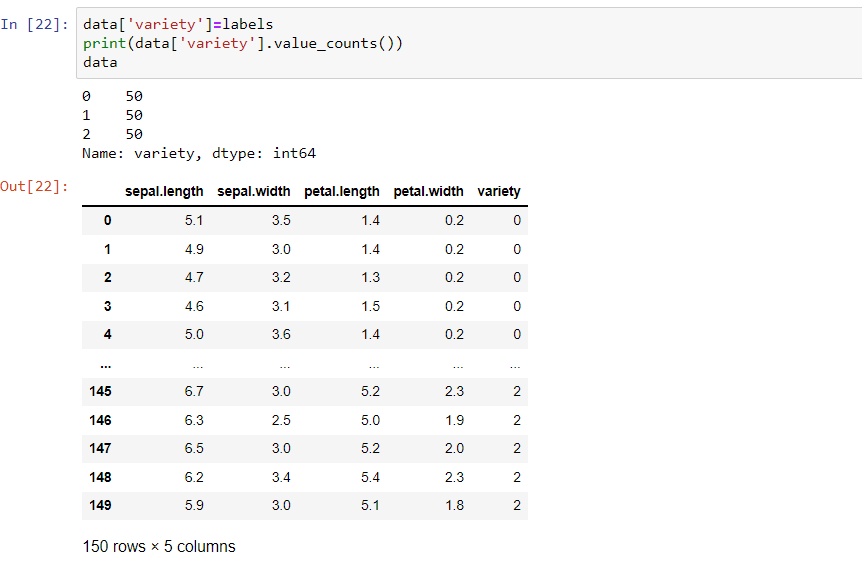


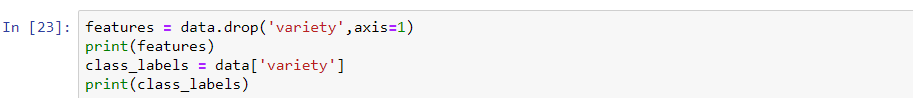


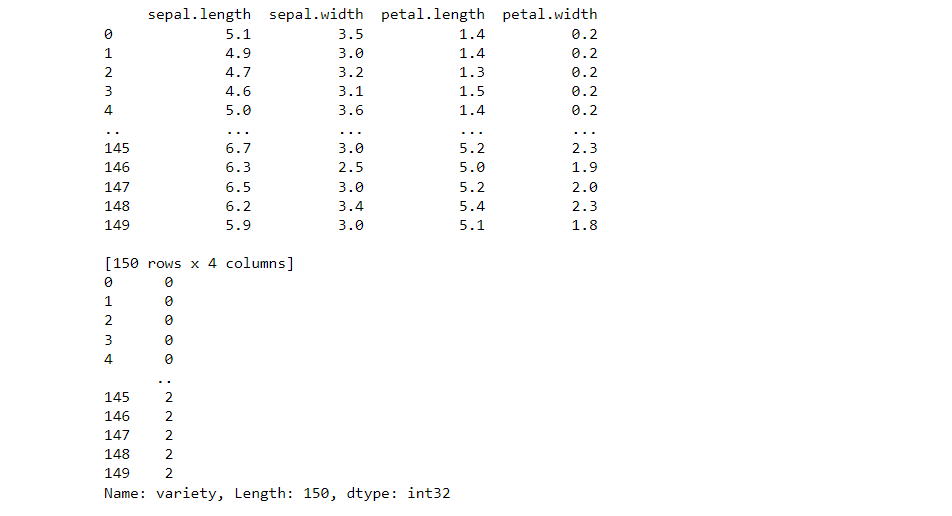


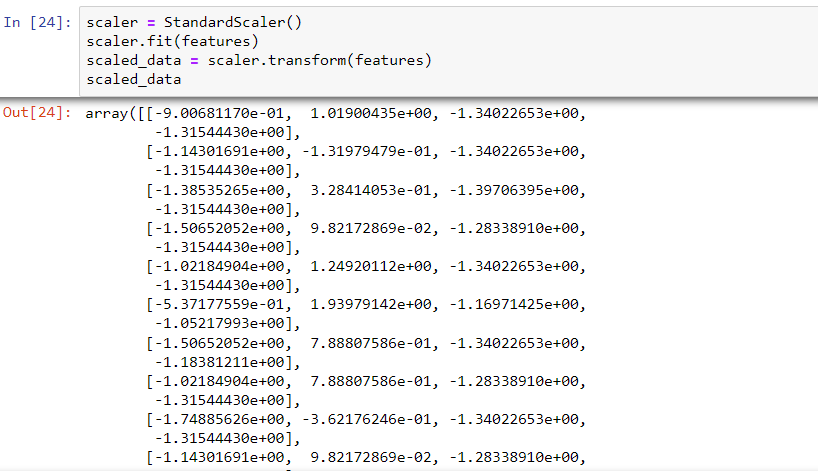


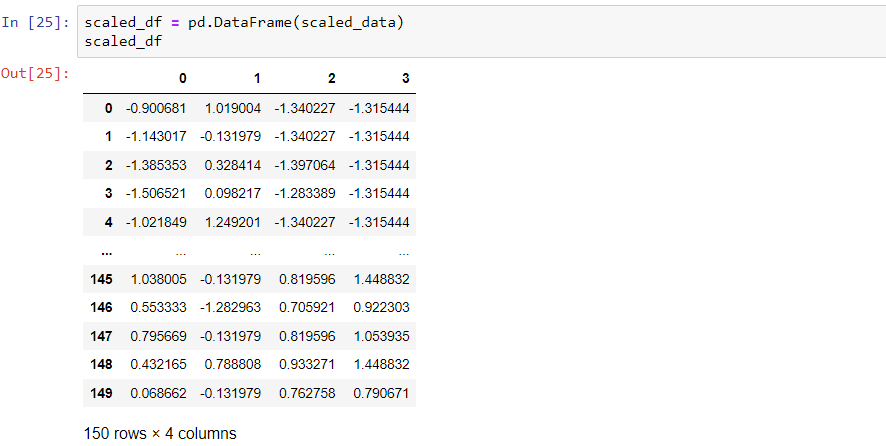










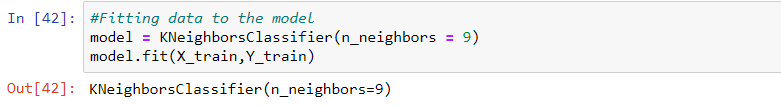


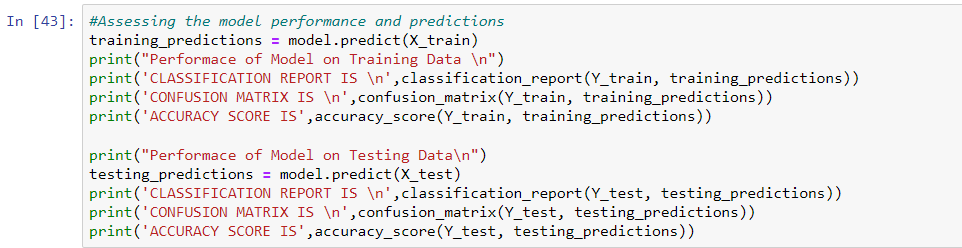
X\_train,X\_test,Y\_train,Y\_test = train\_test\_split(scaled\_df, class\_labels, random\_state=2, test\_size = 0.2, stratify =class\_labels)

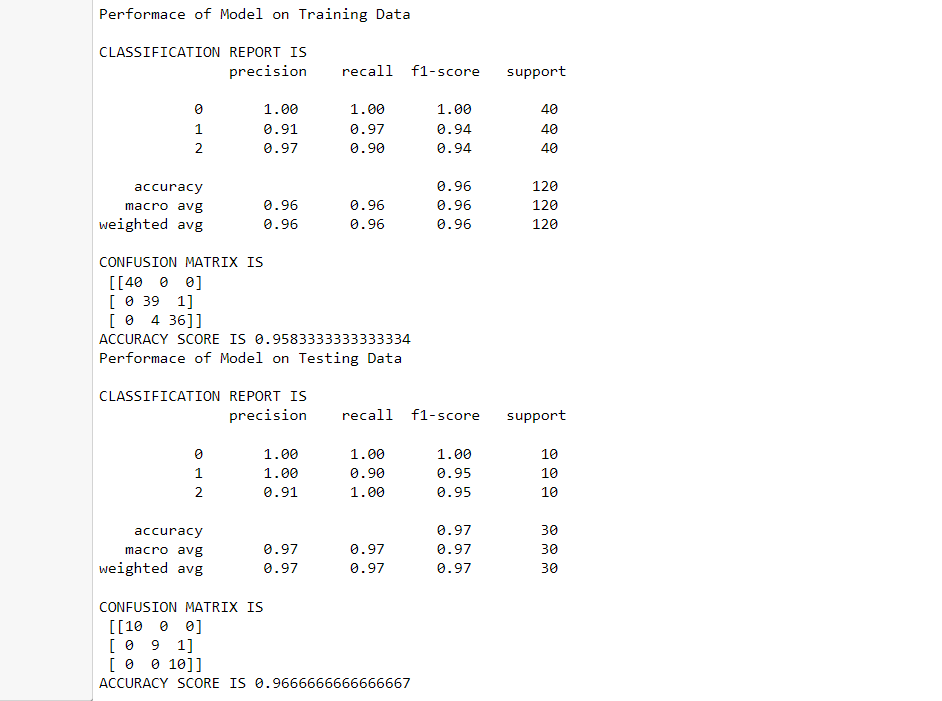
print(X\_train)

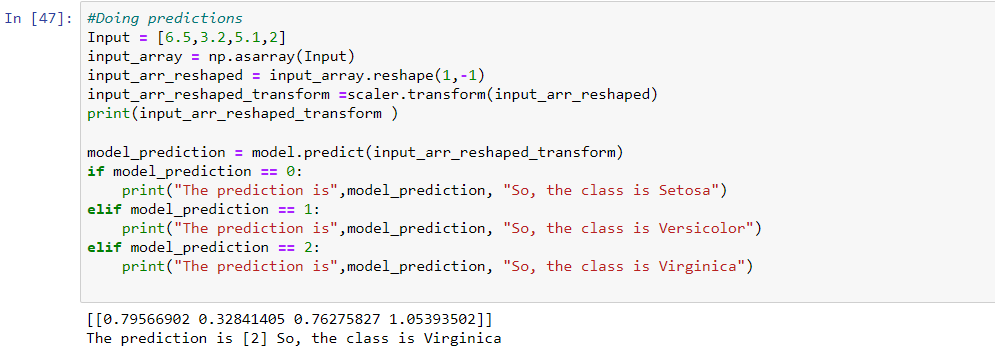
print(Y\_train)



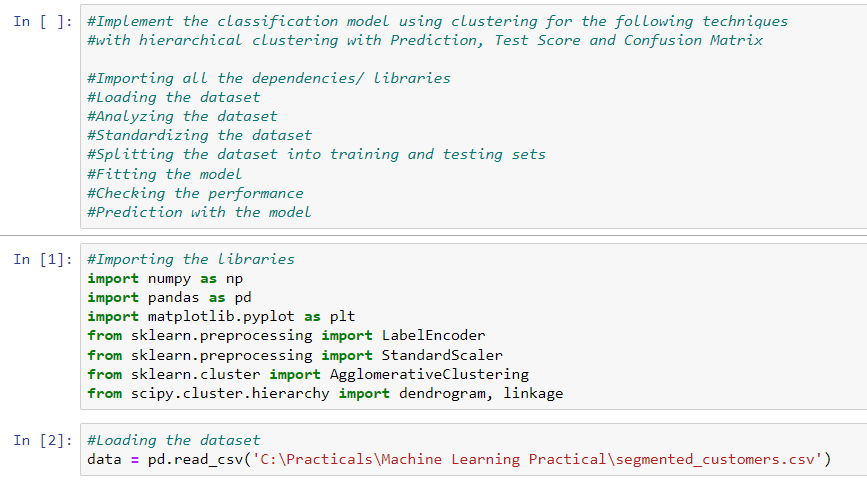


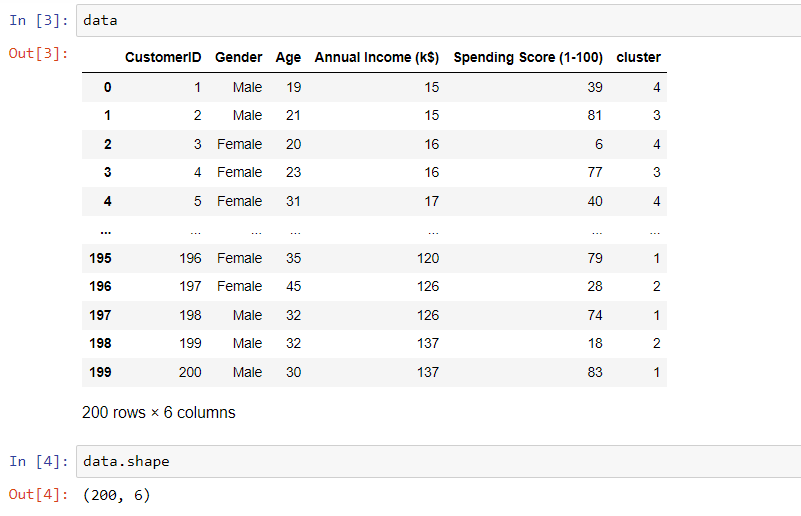


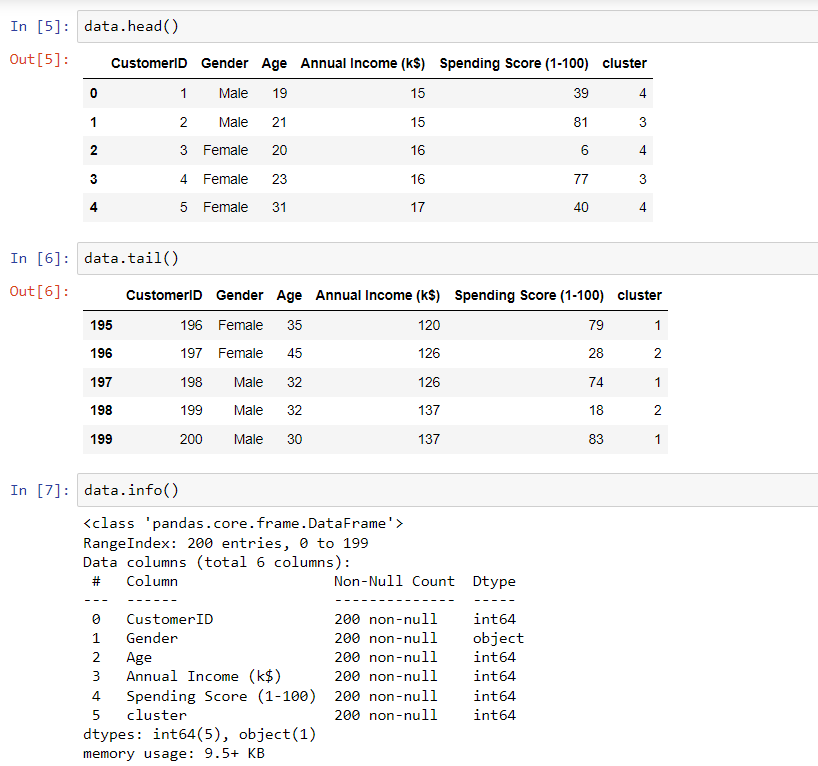


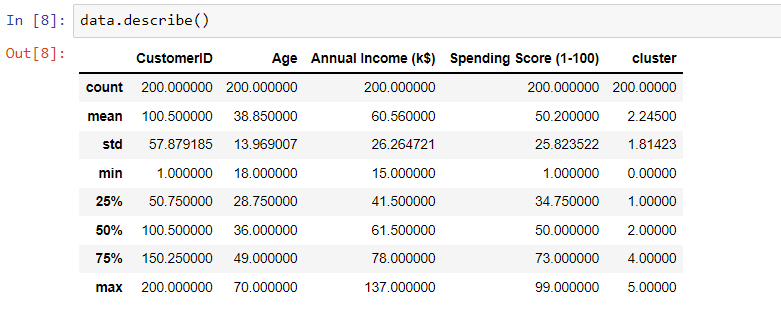


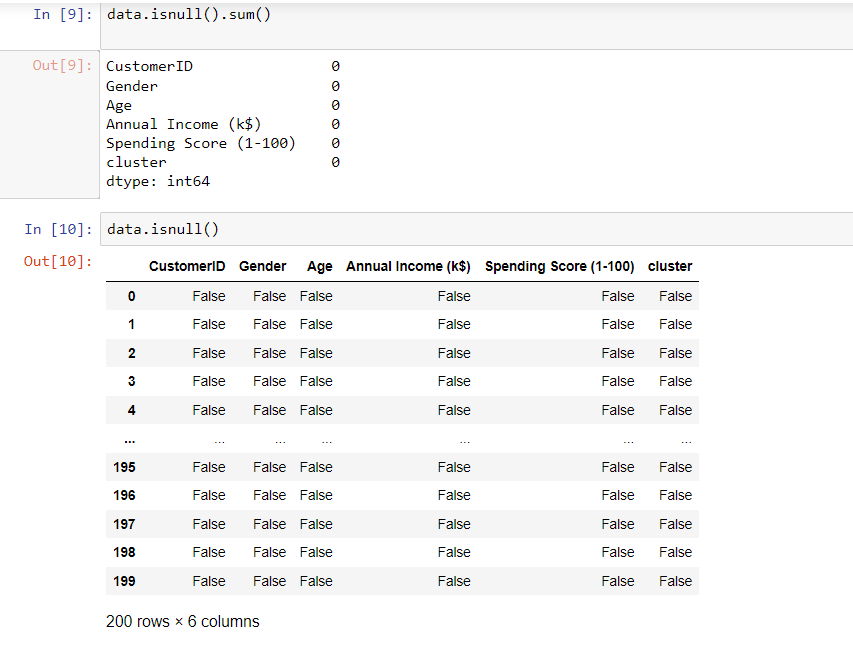
**7A**

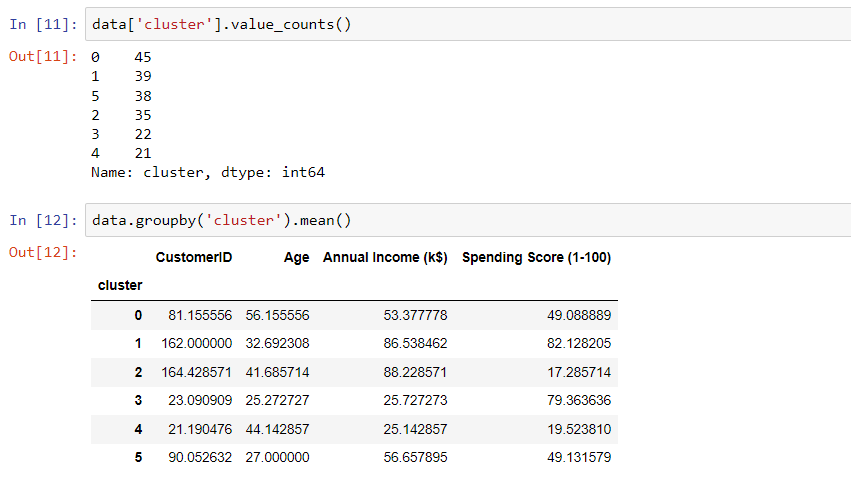


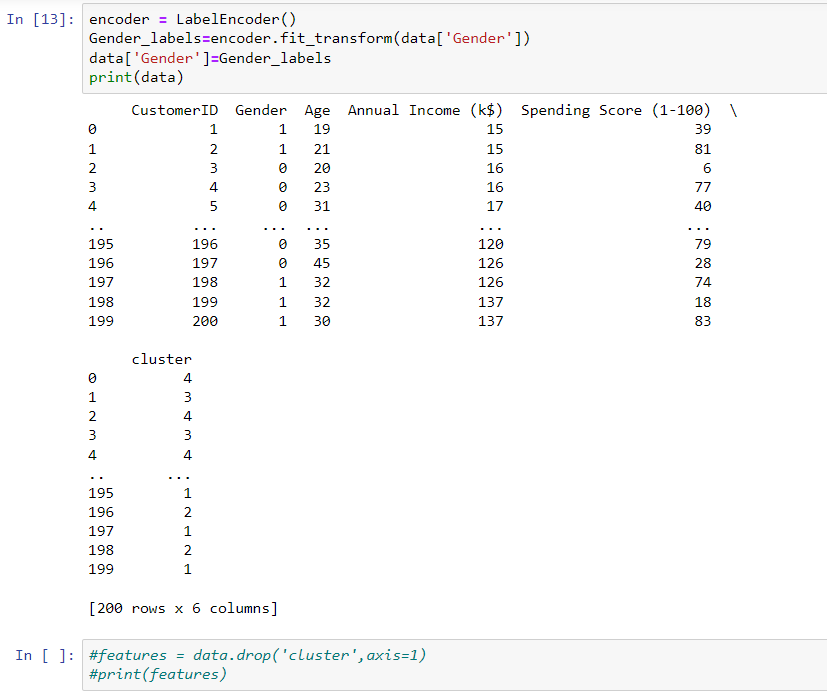


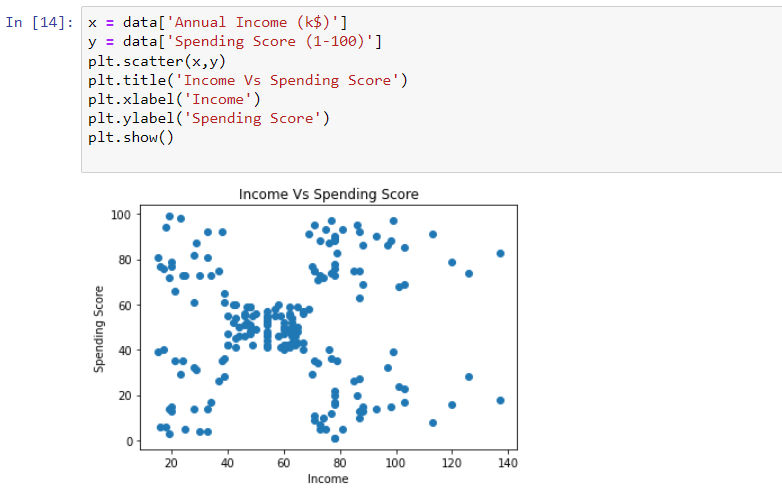


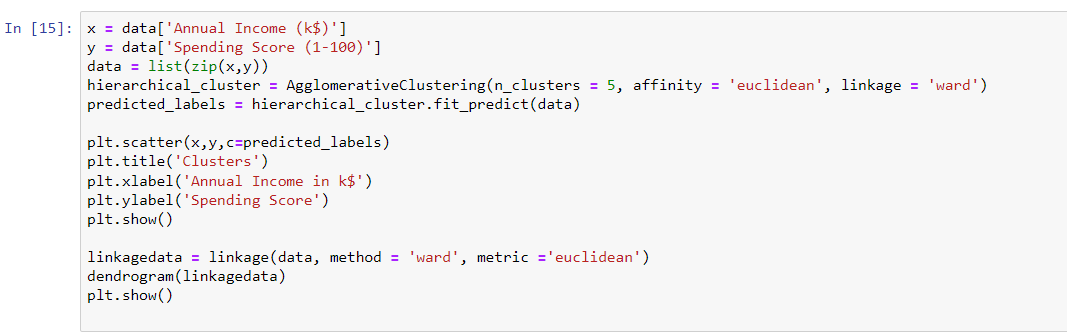


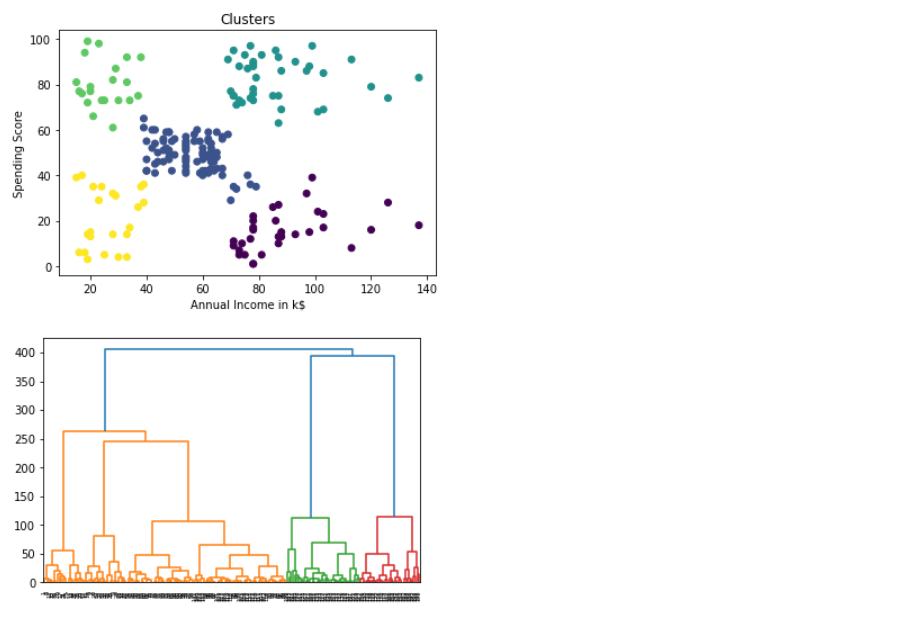




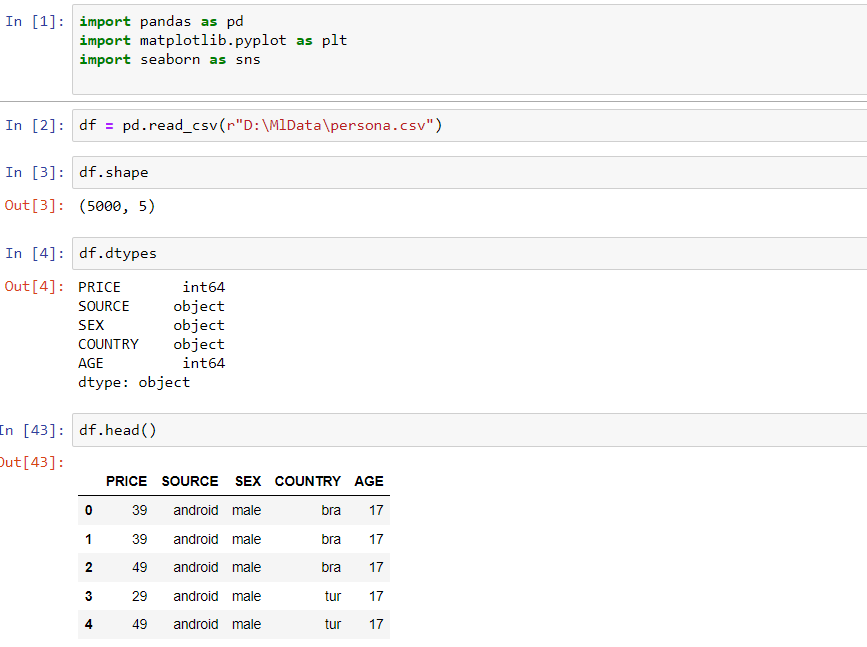


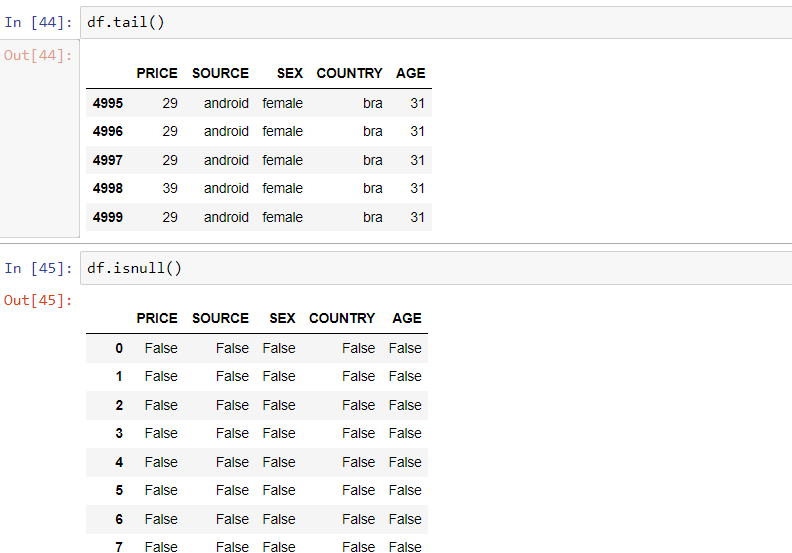




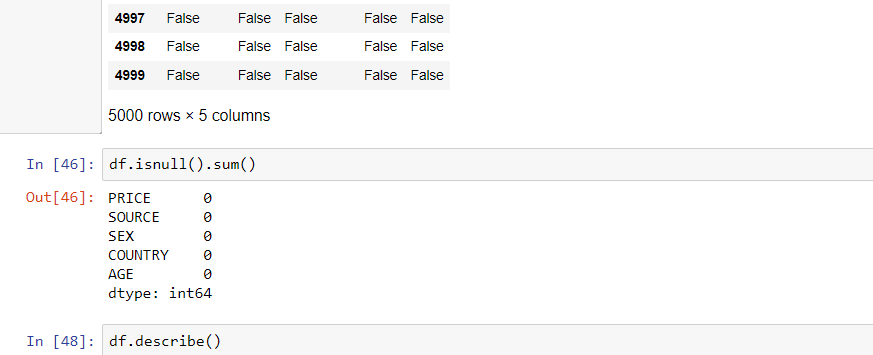


**7 B :- Implement the Rule based method and test the same.**

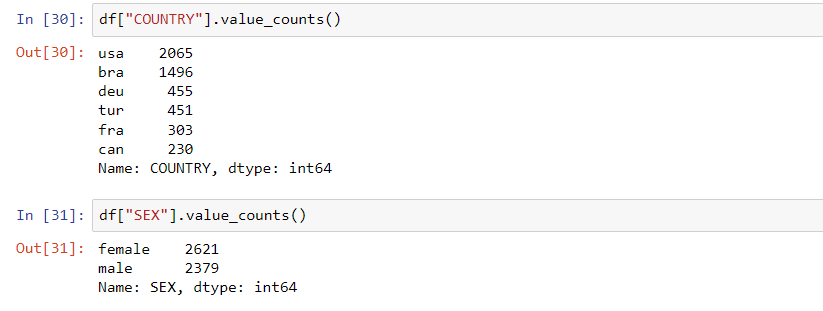


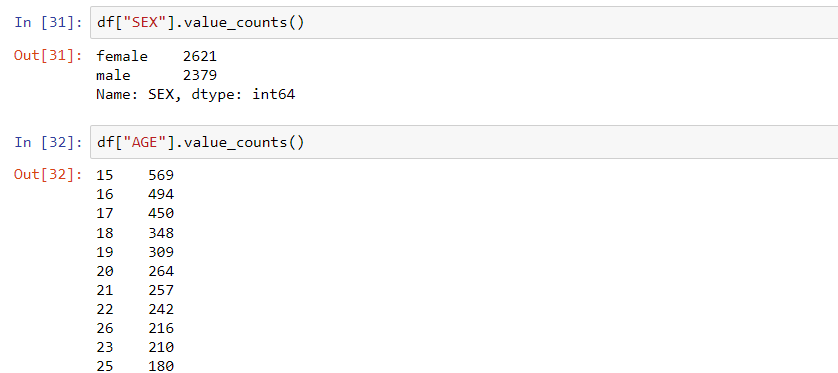








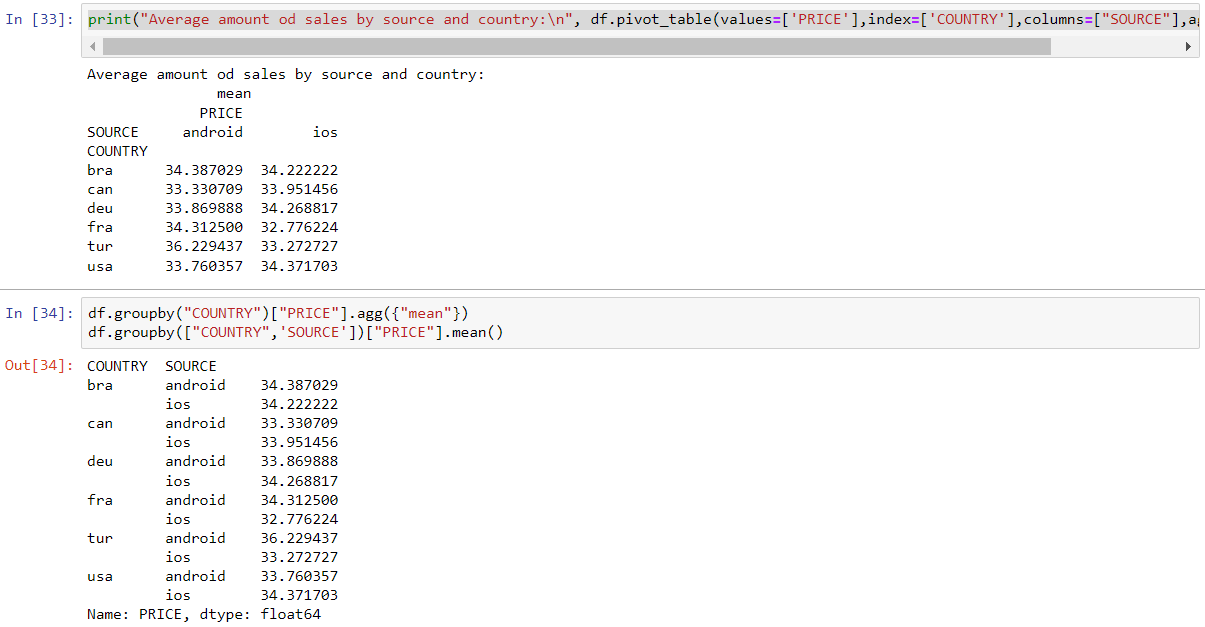


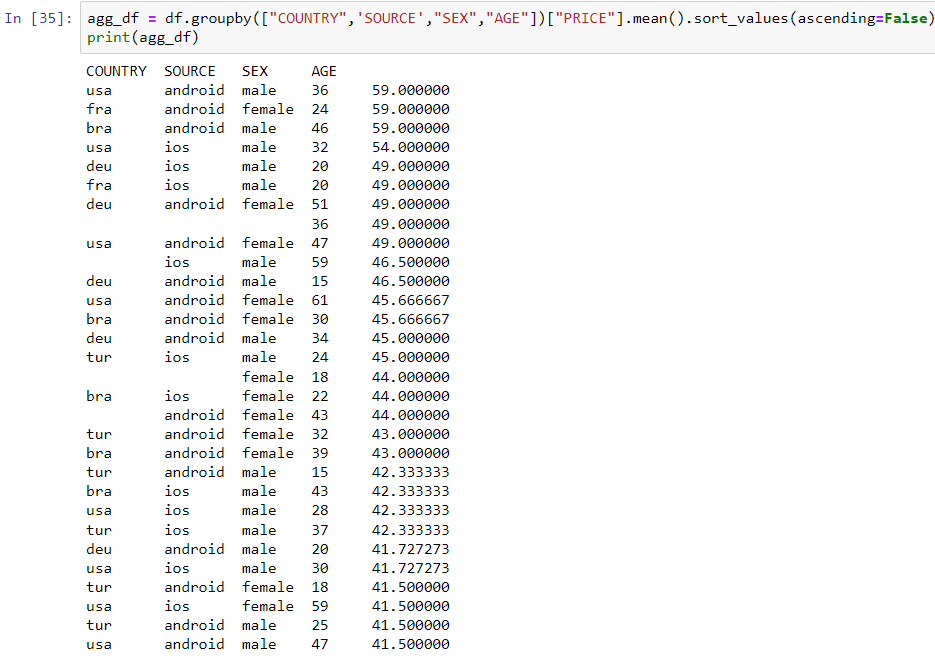


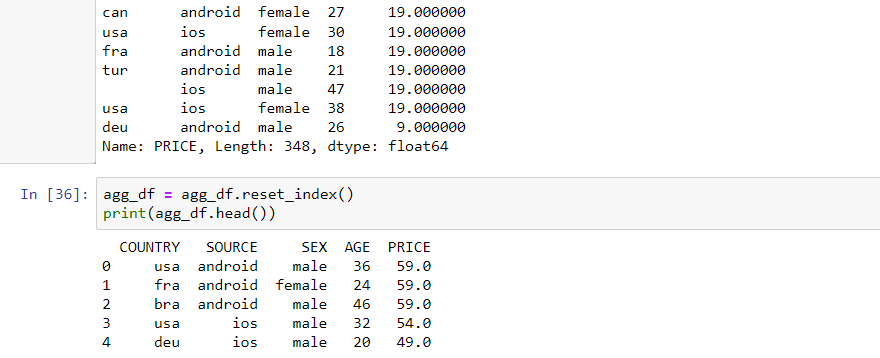


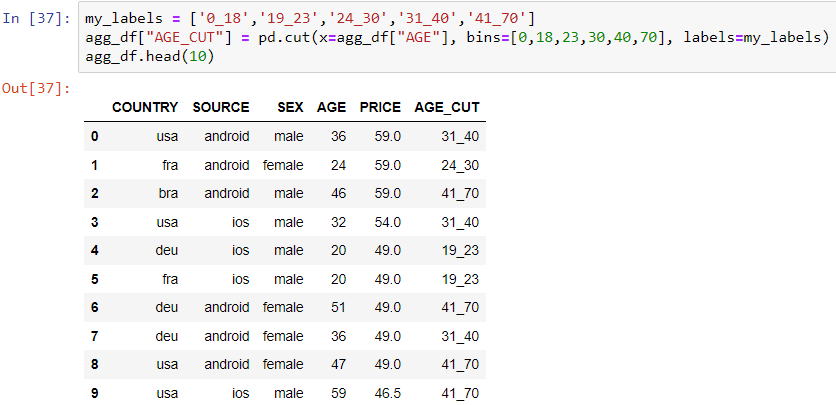


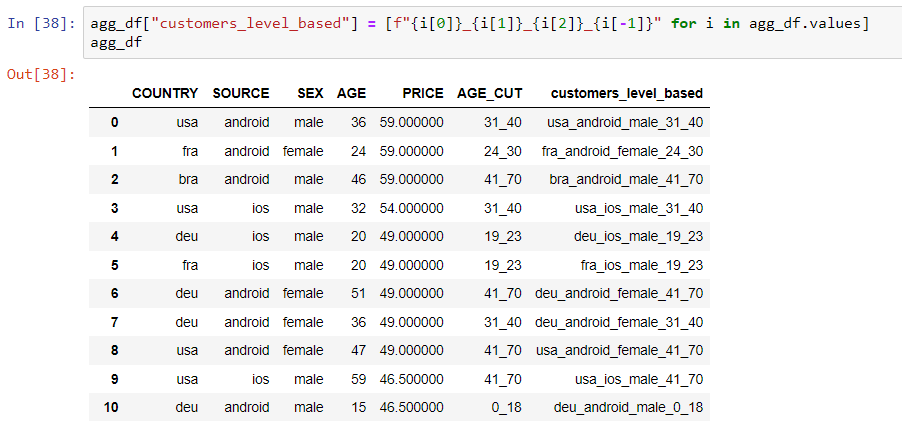
print("Average amount od sales by source and country:\n", df.pivot\_table(values=['PRICE'],index=['COUNTRY'],columns=["SOURCE"],aggfunc=["mean"]))



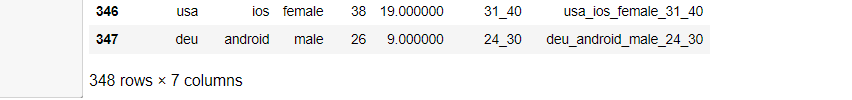










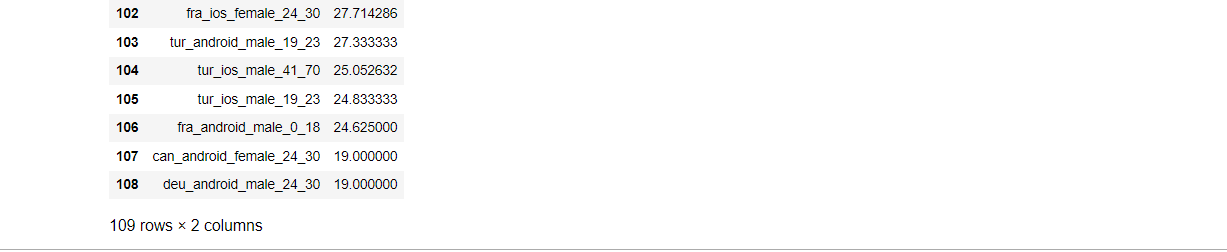


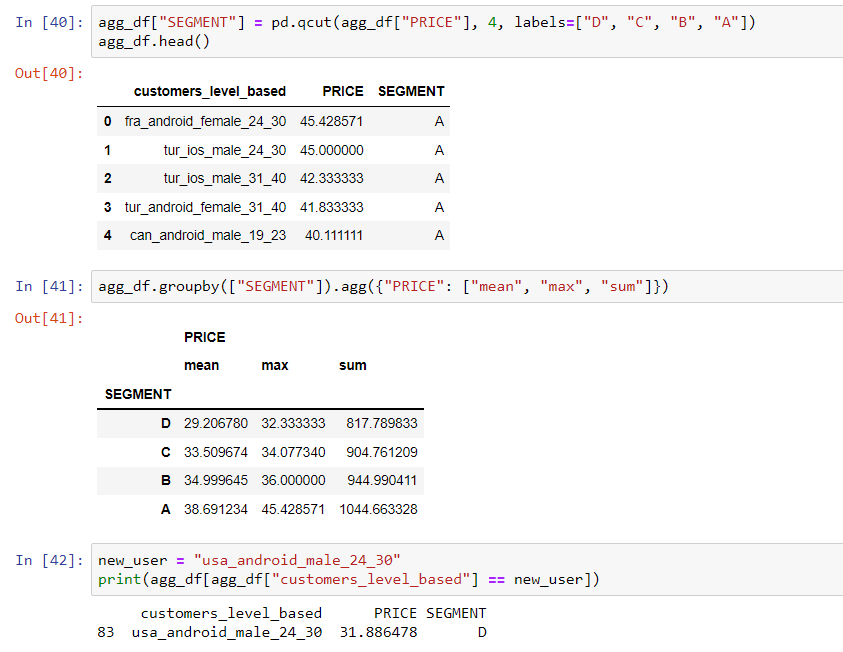
agg\_df = agg\_df.loc[:, ["customers\_level\_based", "PRICE"]].groupby("customers\_level\_based").agg({"PRICE":"mean"}).sort\_values(by="PRICE", ascending=False).reset\_index()

agg\_df["customers\_level\_based"].head(10)

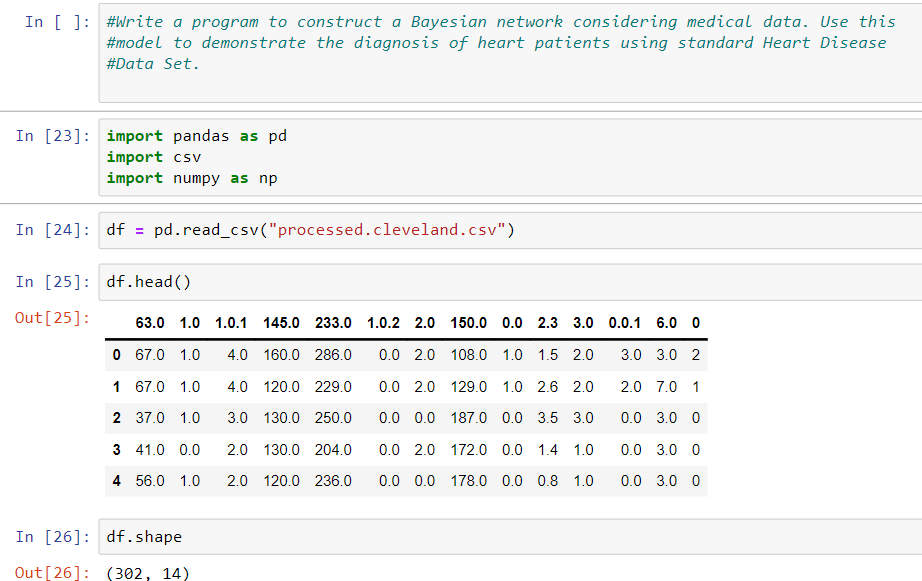
agg\_df

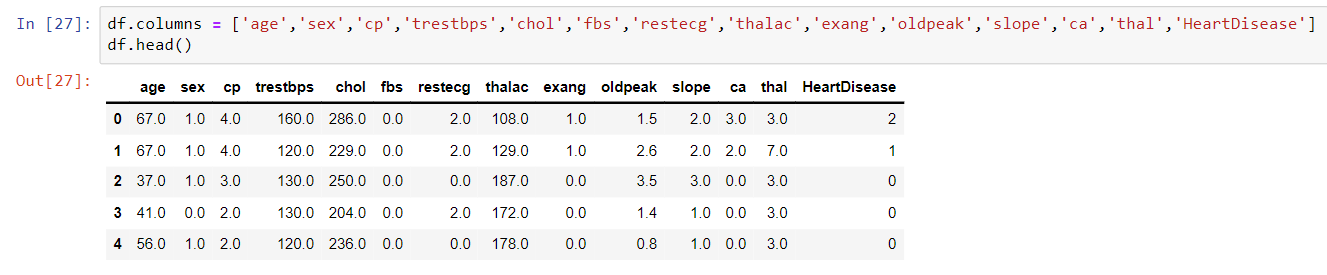


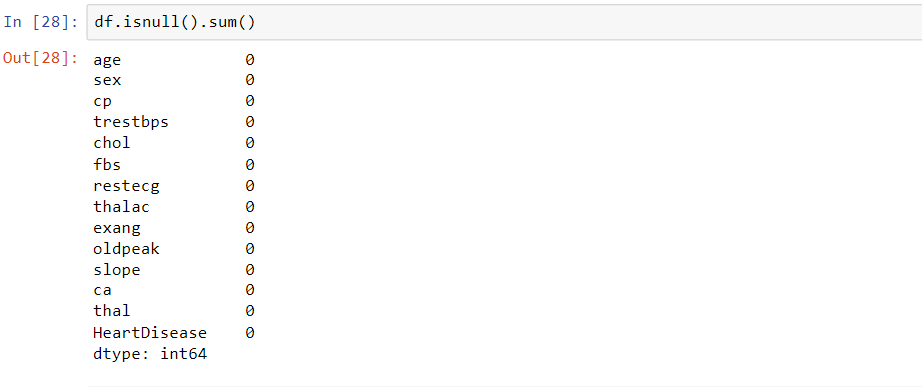


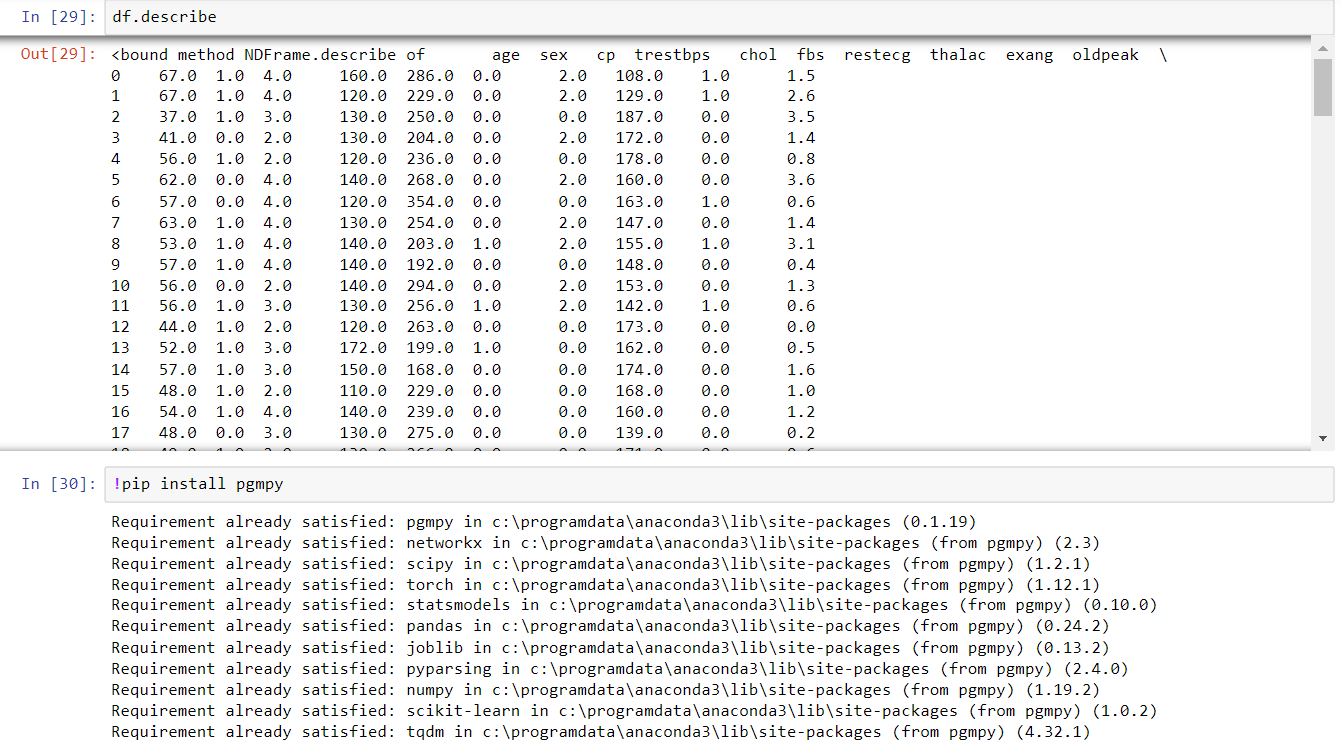


**8A**

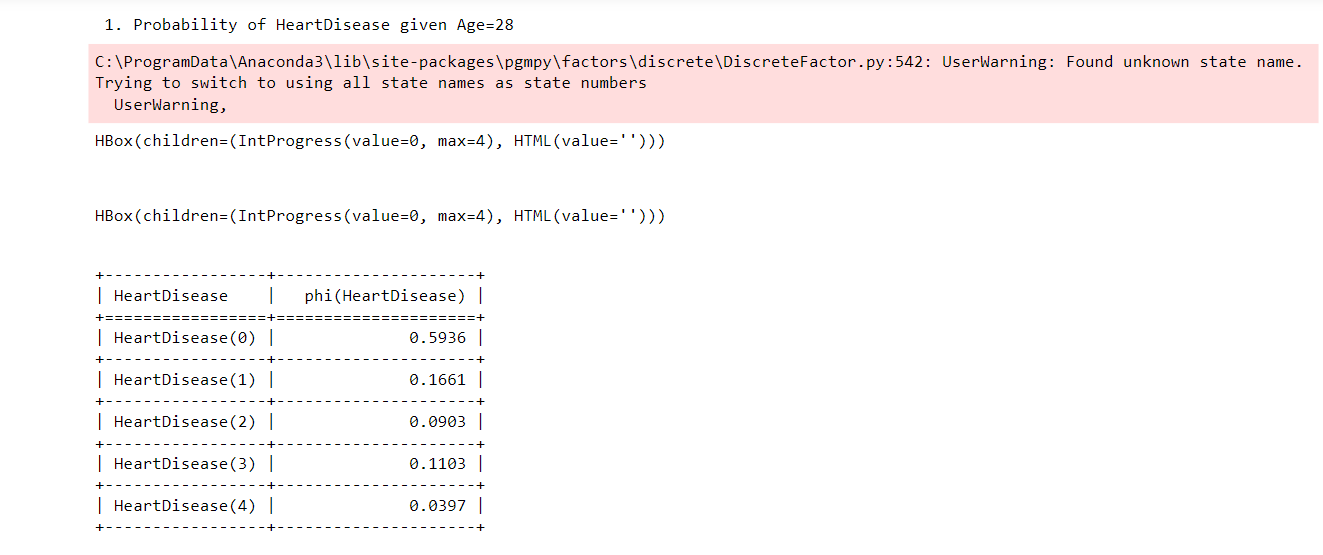


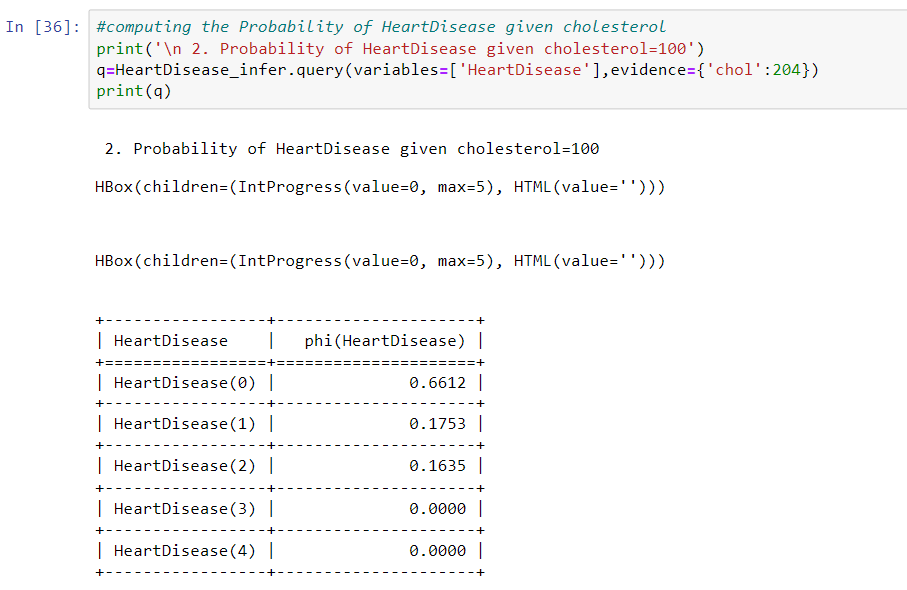


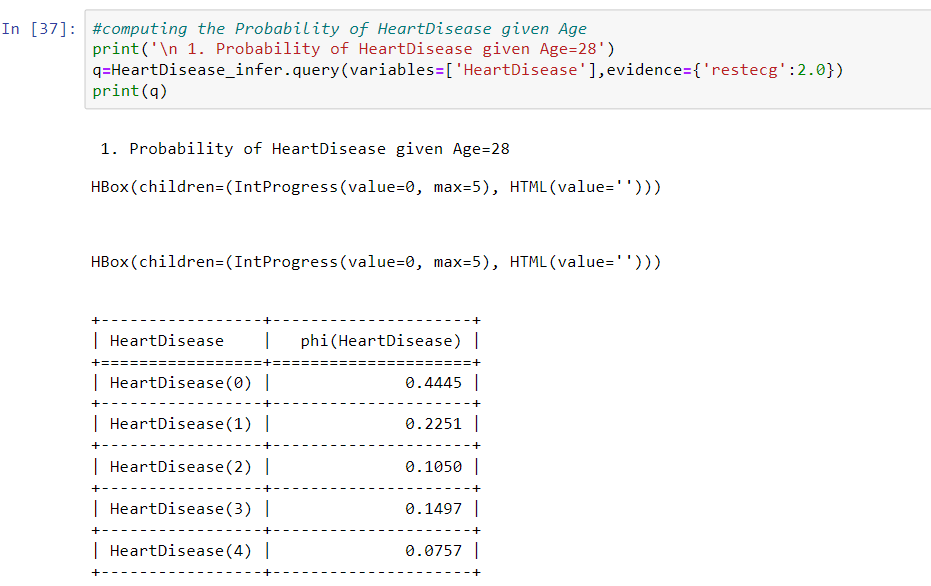








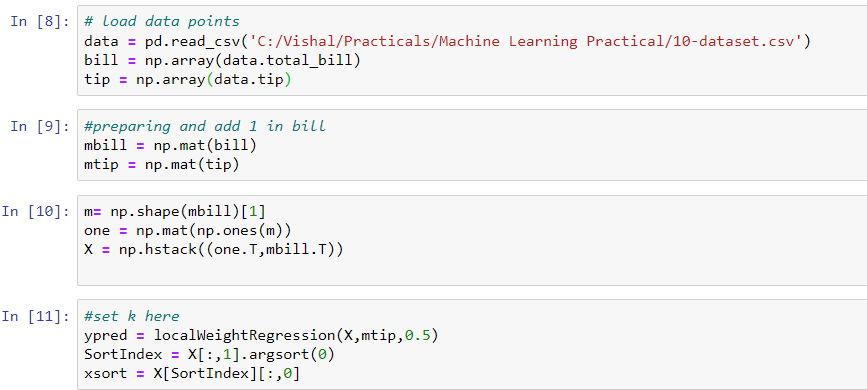


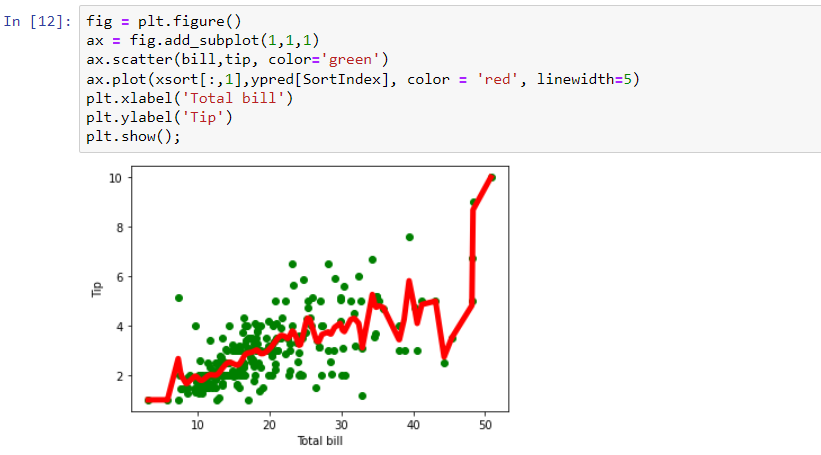




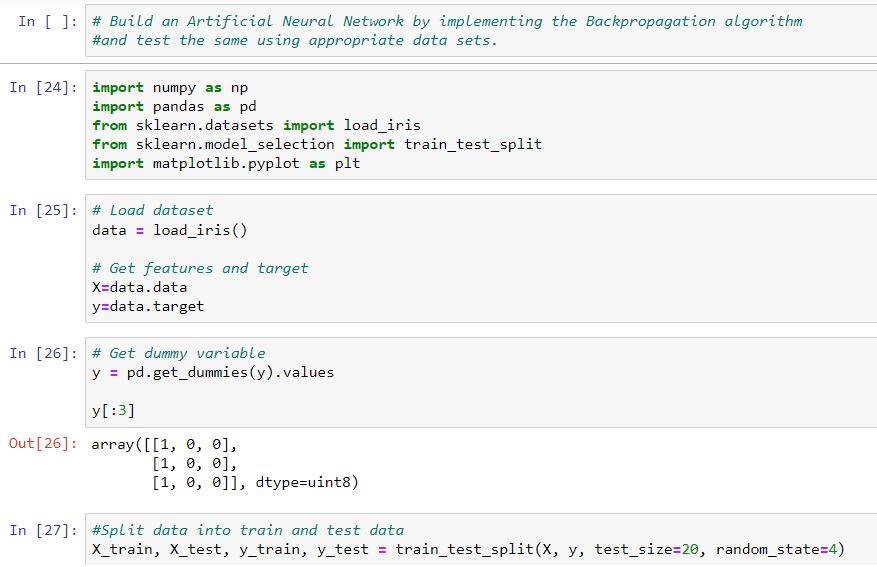
**8 B :- Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points. Select appropriate data set for your experiment and draw graphs.**

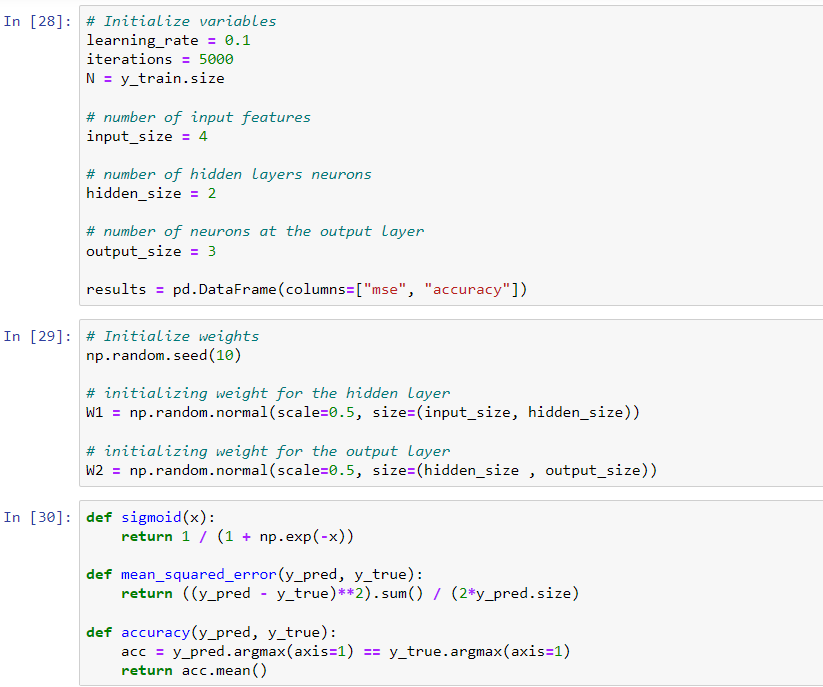


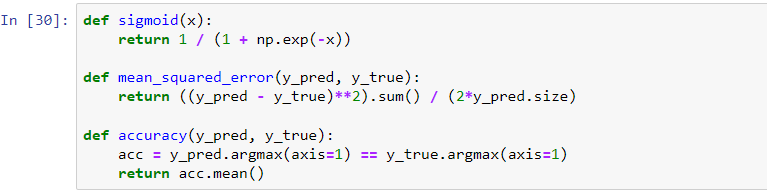


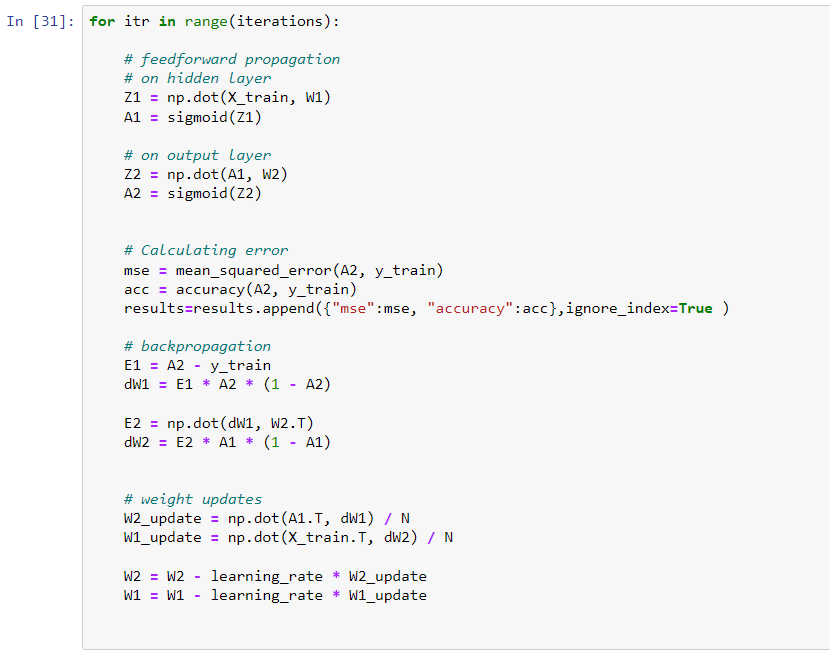


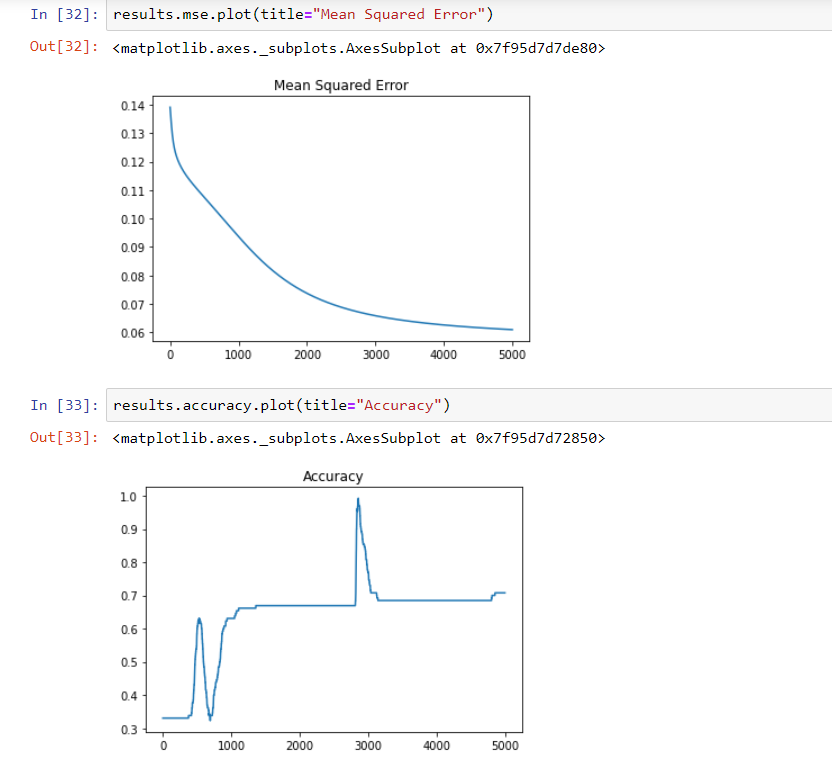
**9A**

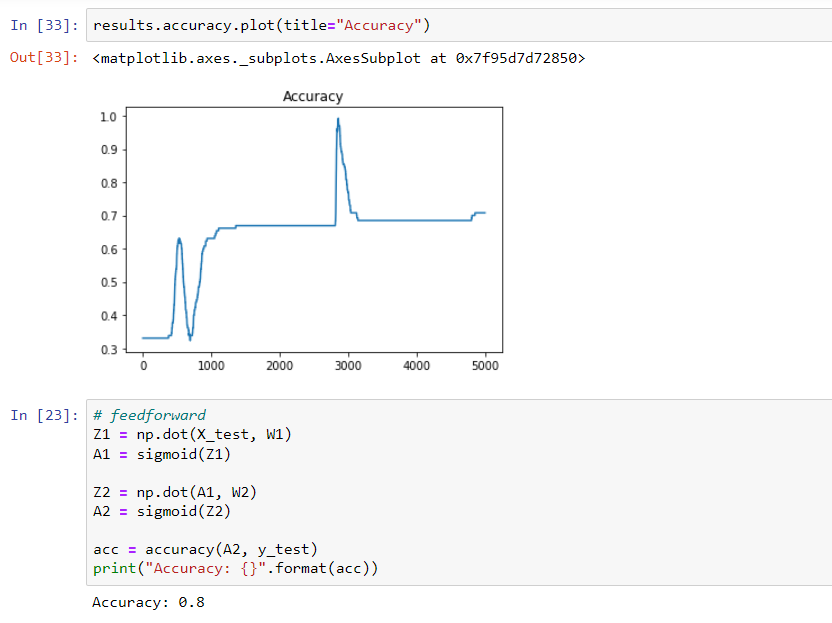








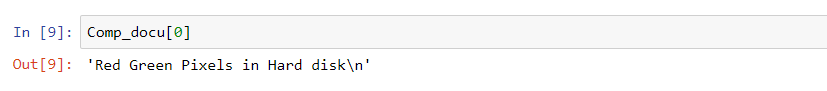




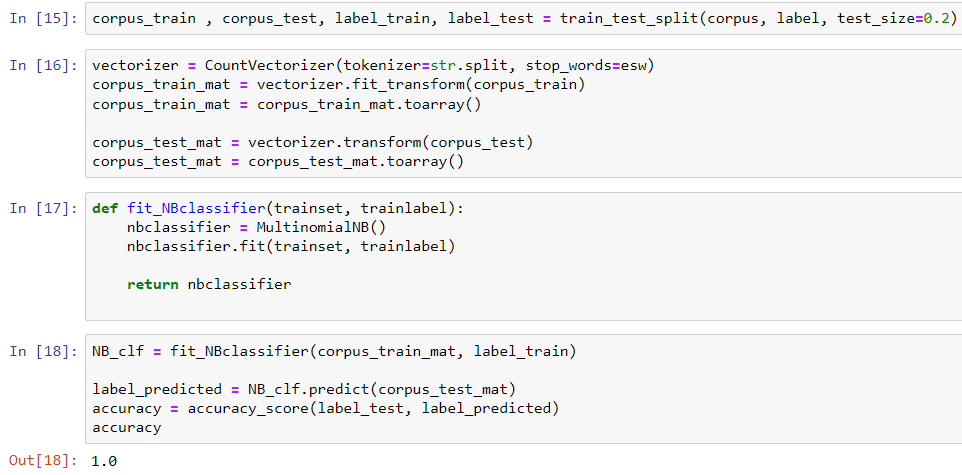
**9B:- Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to perform this task.**

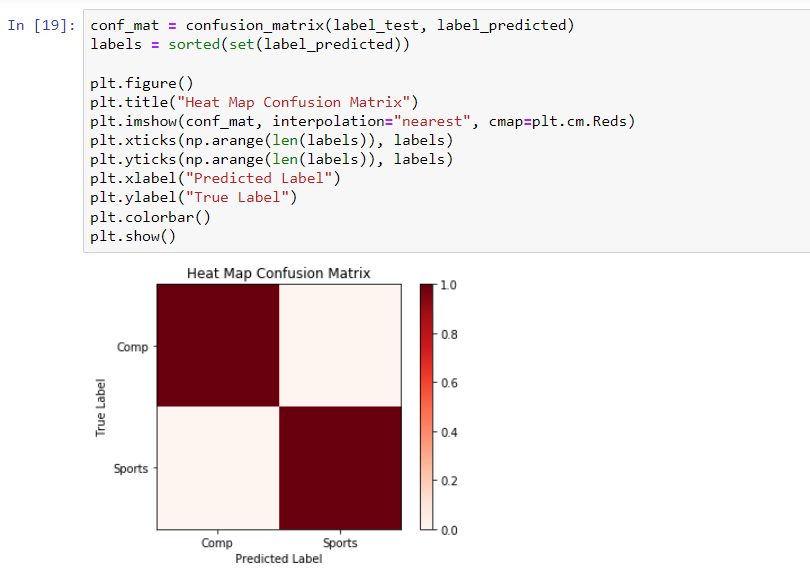




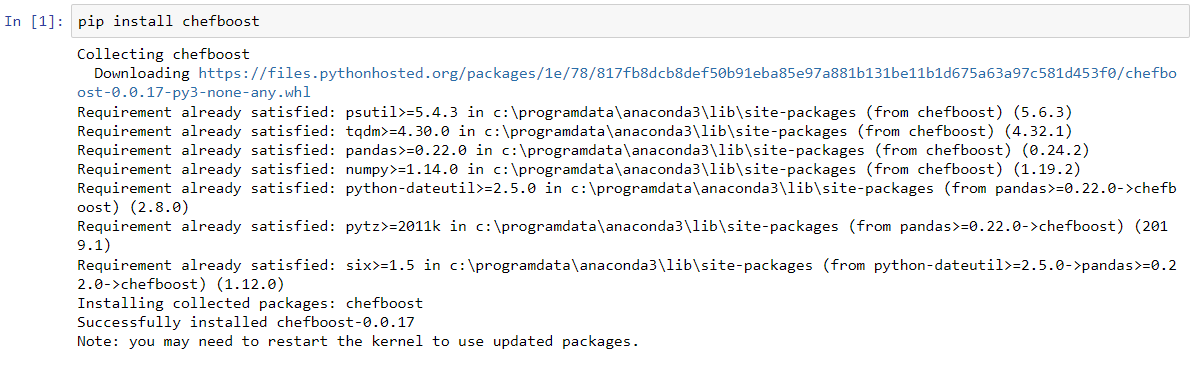


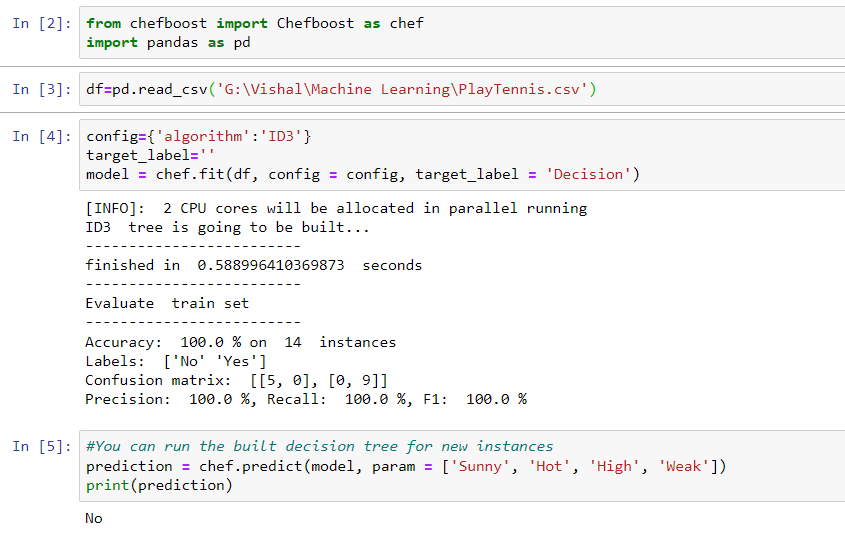


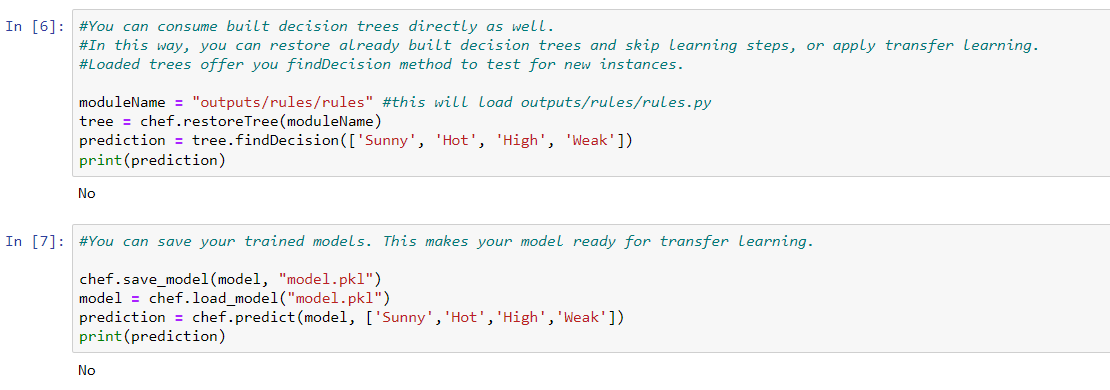




**10A**







**10B :- Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.**

