# Python Lab Assignment -1

***Team-10 Members****:*

**Sai Rohith Guntupally** – 8

**Chaitanya Mallepudi** – 14

**Madhu Varma Rudraraju** – 20

**Github**:

<https://github.com/MRChaitanya/PythonDL_programming_ICP/tree/master/Module1_Lab1>

**Video Link**:

https://drive.google.com/open?id=1eVUOlDu2lZx5DrpsNbZBS3cHLVGtYmKo

***Introduction***:

In this assignment we use the dictionaries, functions, Inheritance in python and built the models from the dataset using scikit-learn library. We have taken data set from Kaggle and analyzed the data and use various techniques like Regression, Clustering and Classification and visualized the data.

**Objective**:

To implement the below provided tasks:

1. Operations on Tuples, Sets, Lists and Dictionaries.
2. Inheritance, Method overloading and overriding.
3. Web scraping using Beautiful Soup.
4. Classification techniques using Naïve-Bayes, SVM and K-Nearest Neighbor.
5. K- Means clustering and visualizing the clusters and evaluating the silhouette score.
6. Natural Language Processing.
7. Implement Multiple regression and calculate RMSE and Variance.

**Requirements**:

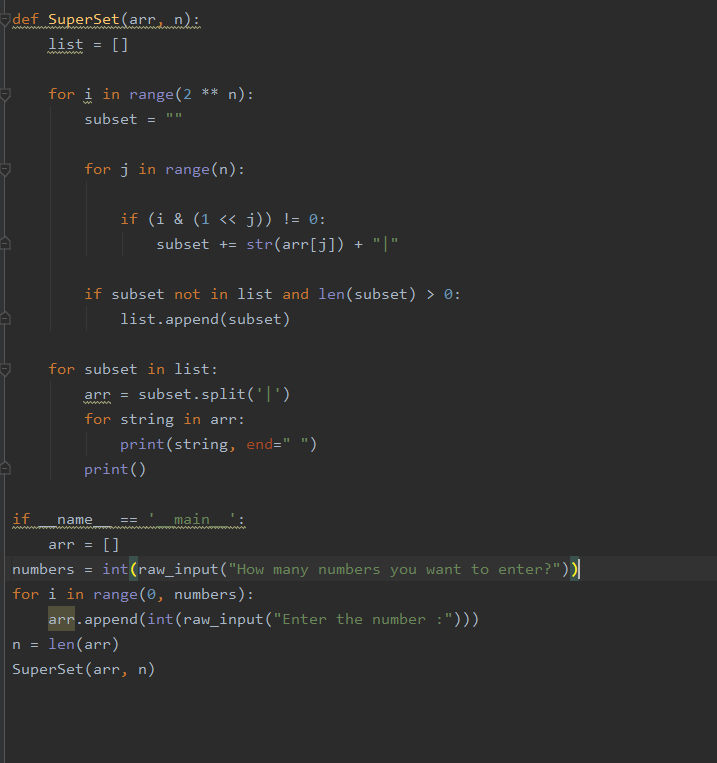
1. PyCharm IDE or Jupyter Notebook
2. Python 3.5 or above compiler
3. Anaconda Interpreter.

**Approaches and Methods:**

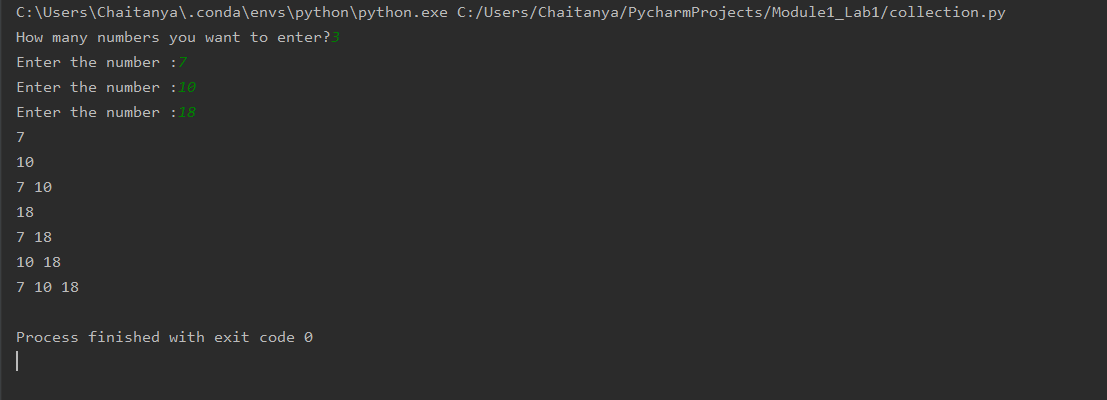
1. For operations of tuples, lists and dictionaries we used the python inbuilt data types and functions.
2. For Inheritance we used concept of class, method overriding.
3. For the EDA we have either filled the null values with mean values or dropped the null values.
4. For the ML techniques like regression, classification and clustering we use the scikit-learn library.
5. For Calculating the RMSE, variance and Silhoutte scores we used the predefined methods in the scikit-learn library.

**WorkFlow:**

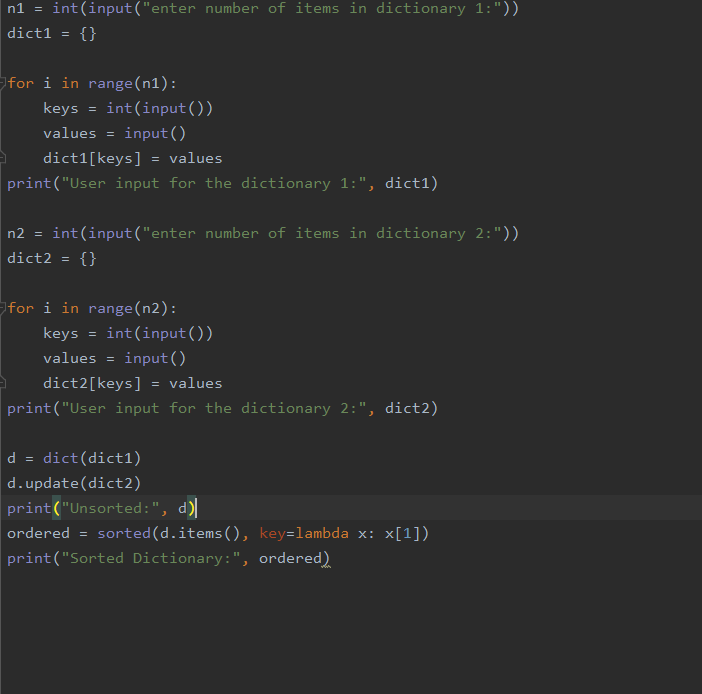
**Task 1:**



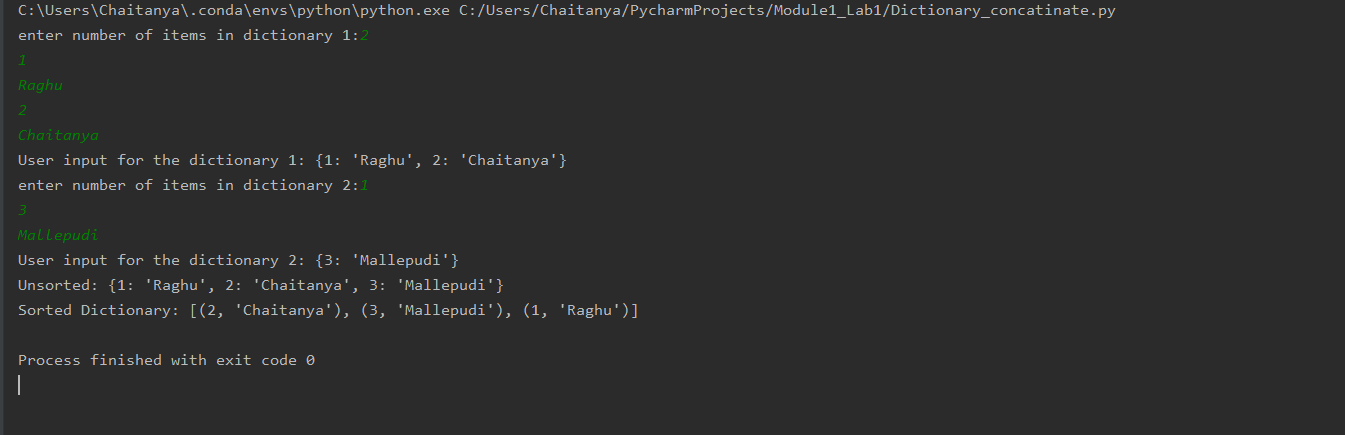
Output:



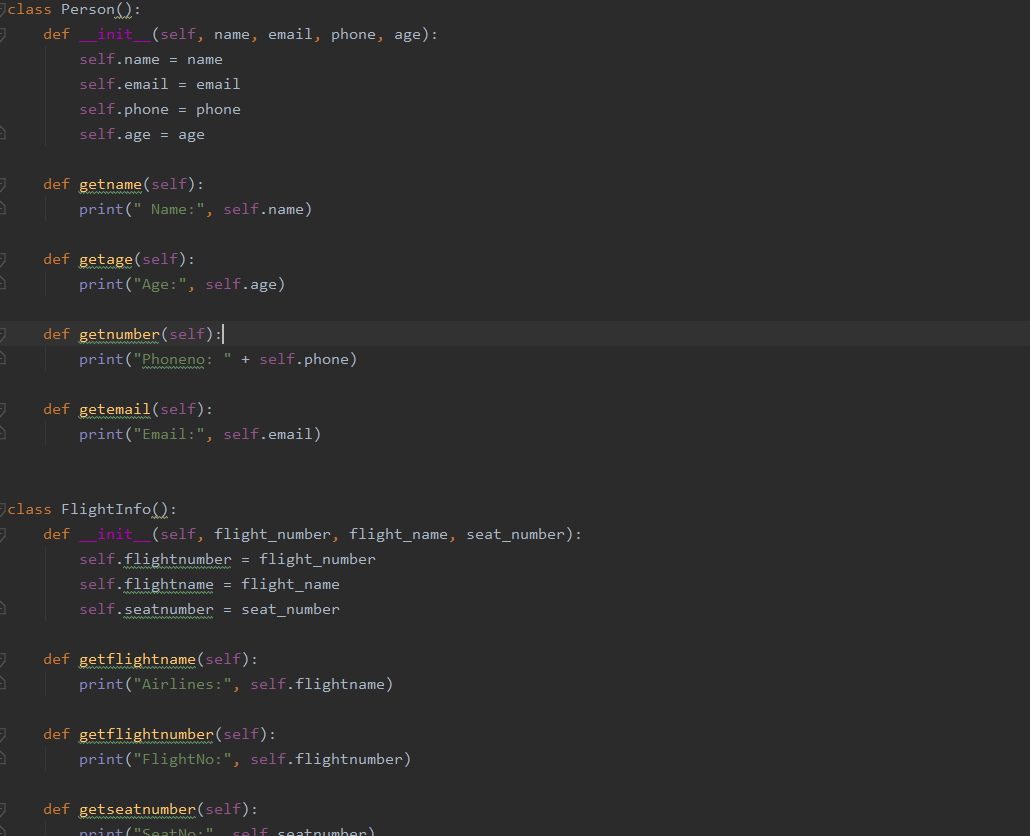
**Task 2**:

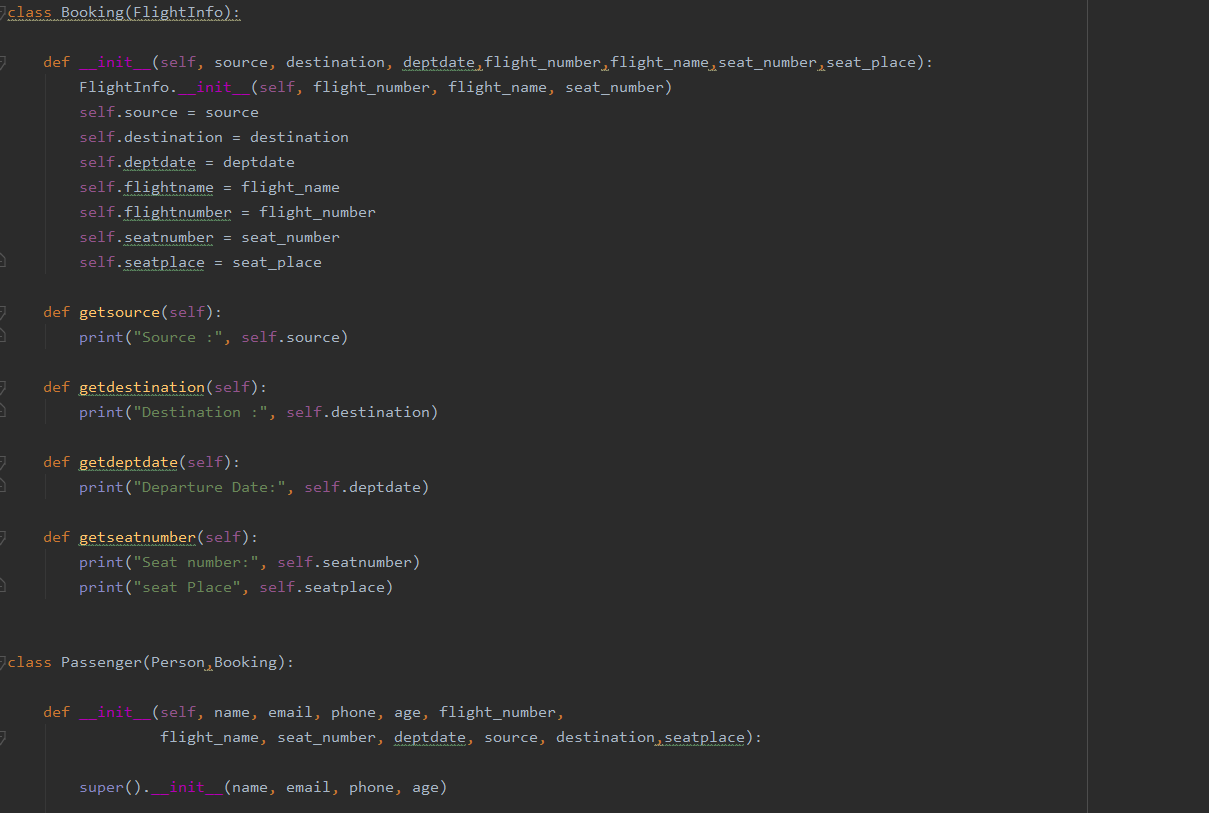


Output:



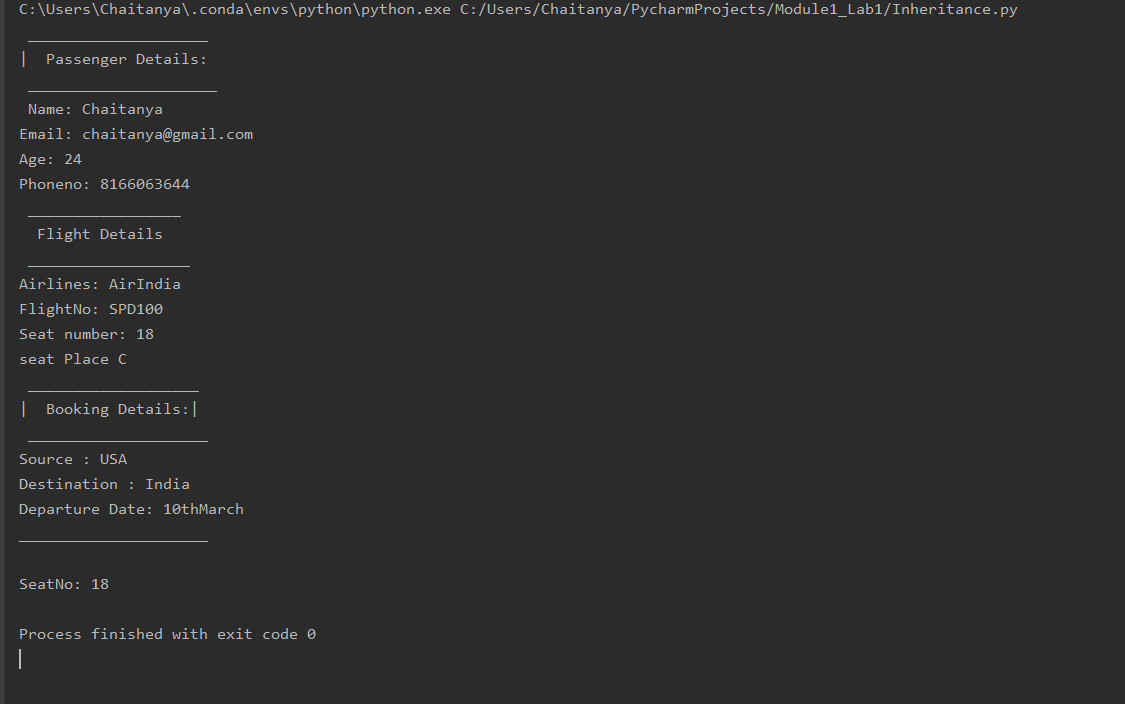
**Task3**:







Output:

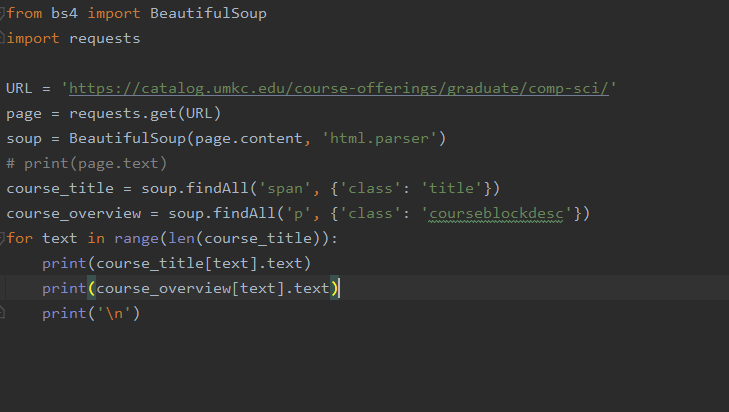


**Observation**

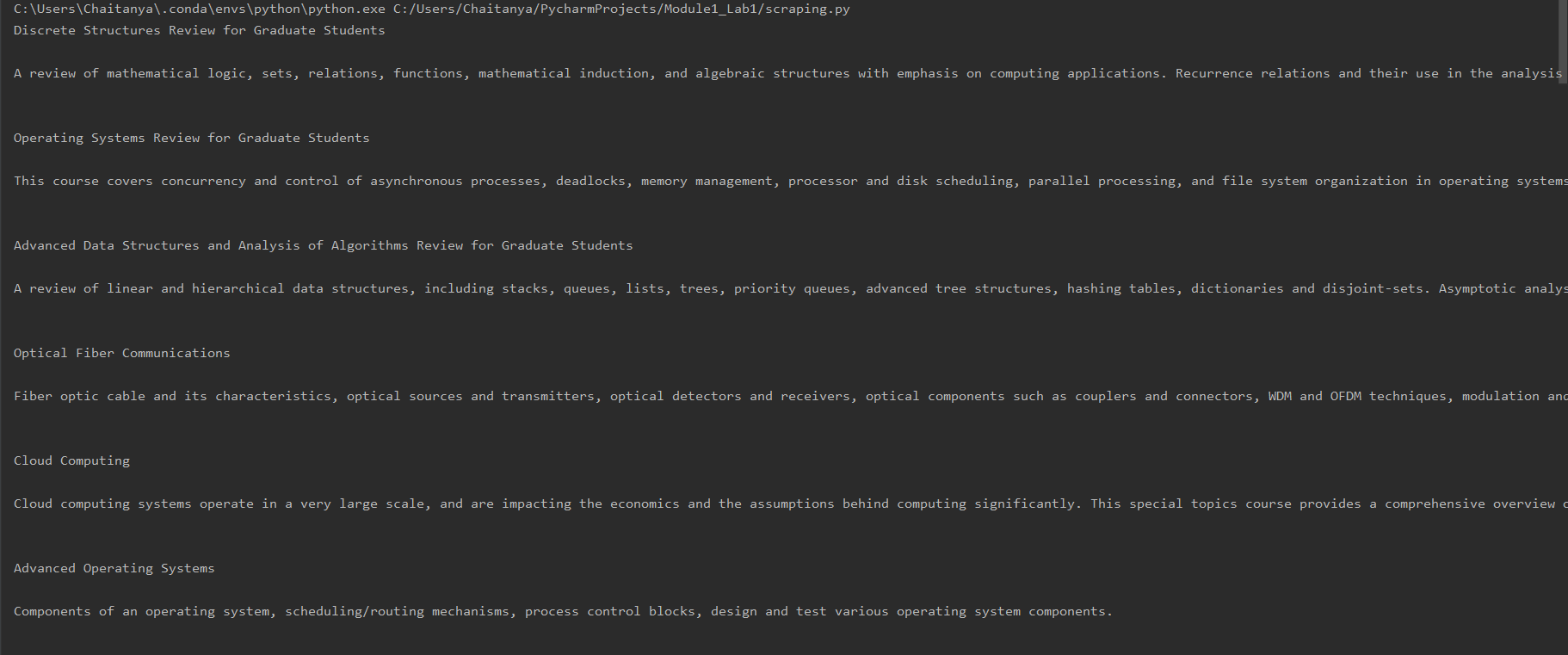
By applying the method overriding we can see the method getseatnumber() got different output when we call it using passenger object and flightinfo object.

We can see the multiple inheritance in the passenger class which inherits fightinfo, person, Booking, Booking inherits the flightinfo class.

Task 4:



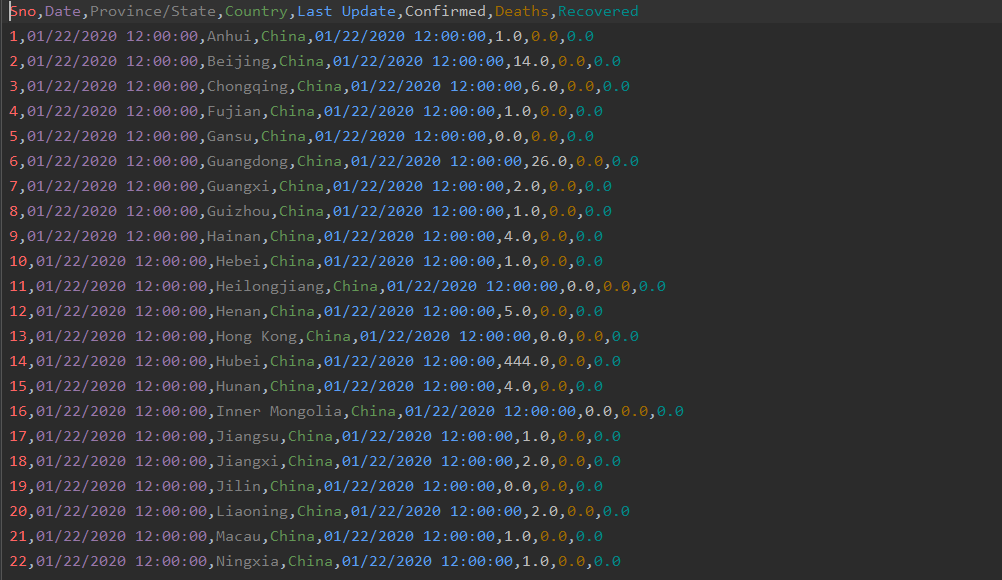
Output:



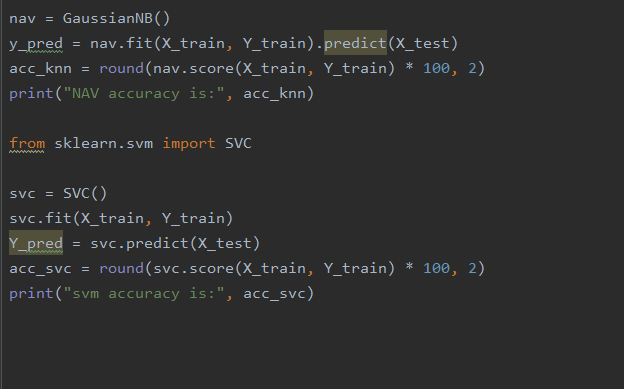
**Task 5**:

**Dataset**: corona\_data

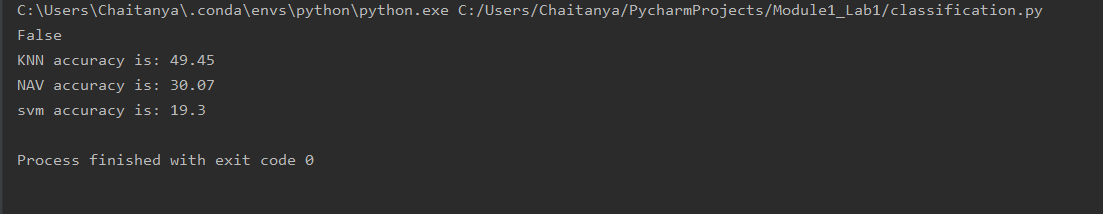
You can get this data in <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#2019_nCoV_data.csv>.



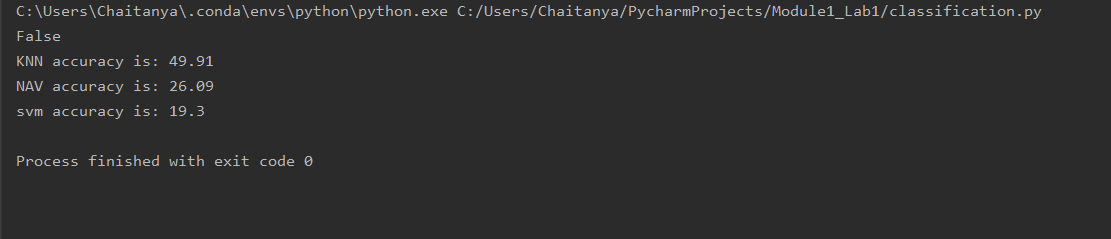




Output Before EDA:



Output After EDA:



**Observation**:

In this we can see that there are no null values in the data set. But there is some noise in the data.

So, we can clearly see that there is no major difference in the accuracy of SVM; But the Naïve Bayes classification has a minor difference.

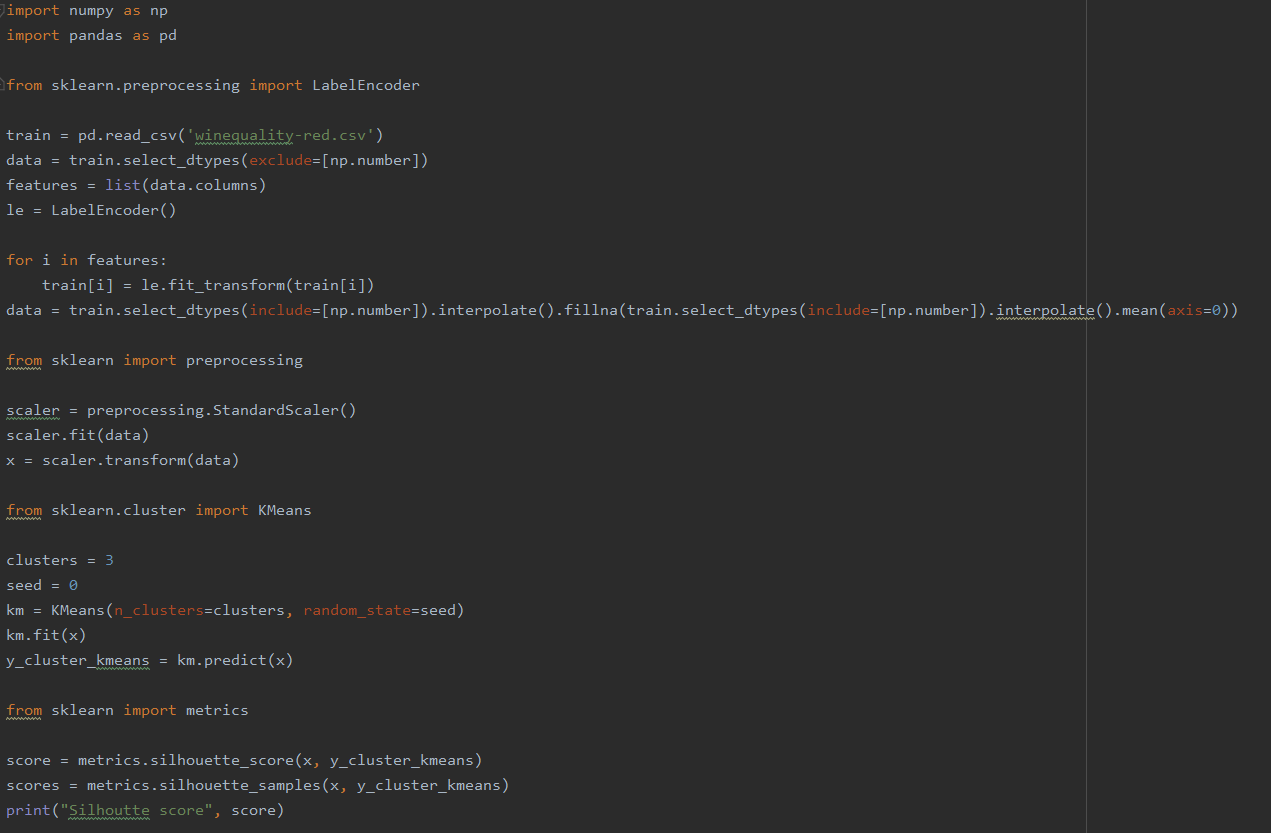
KNN also sees a slight difference before and after EDA.

These classification methods, gives different accuracies for different datasets.

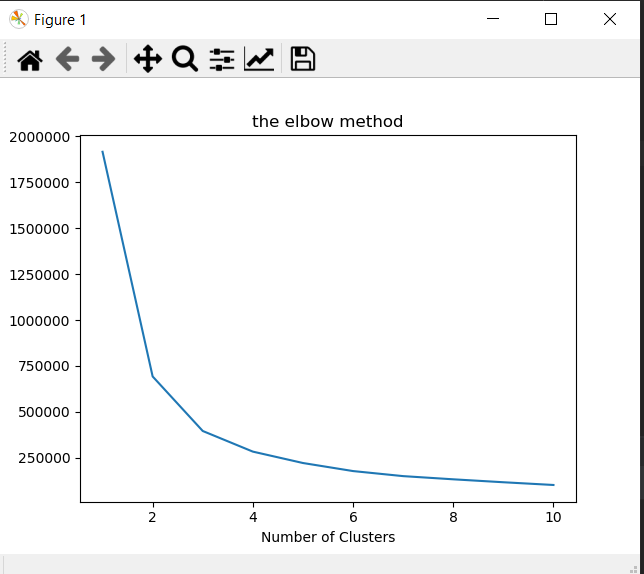
**Task 6**:

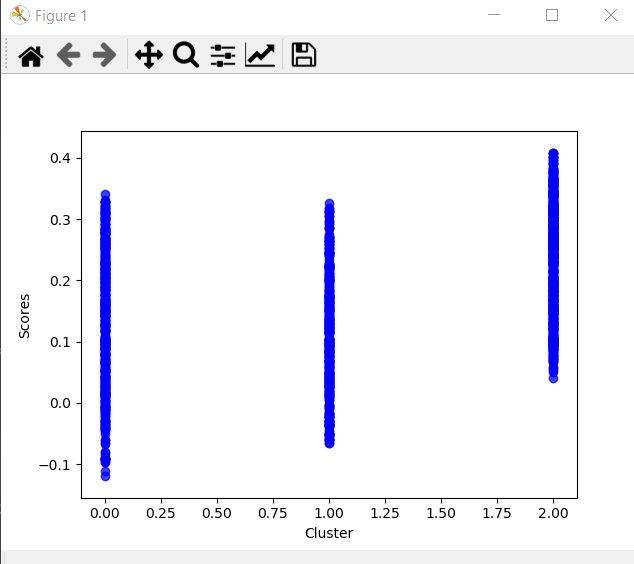
**DataSet**

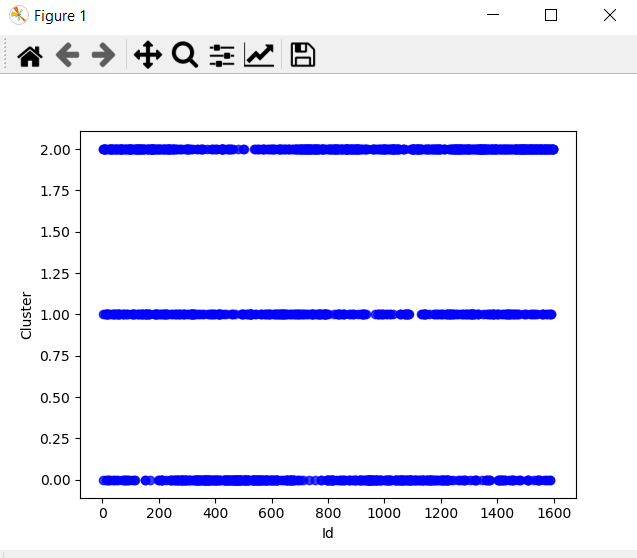


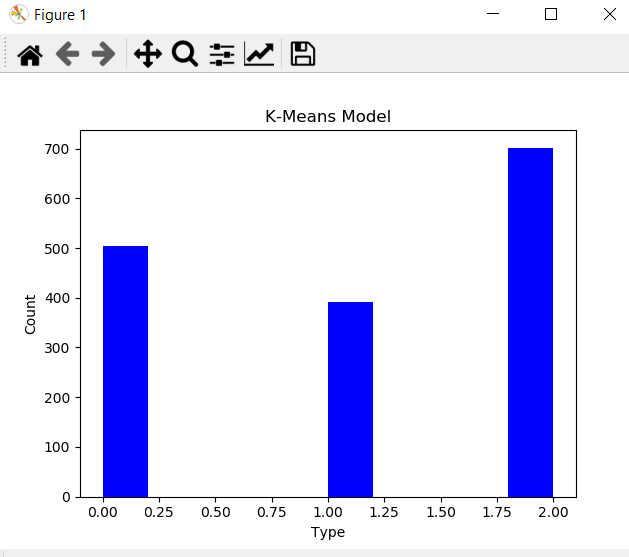


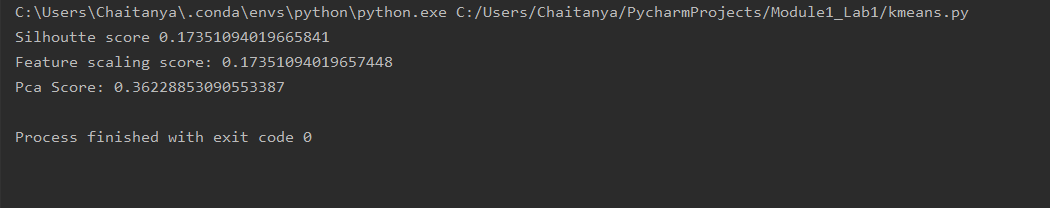












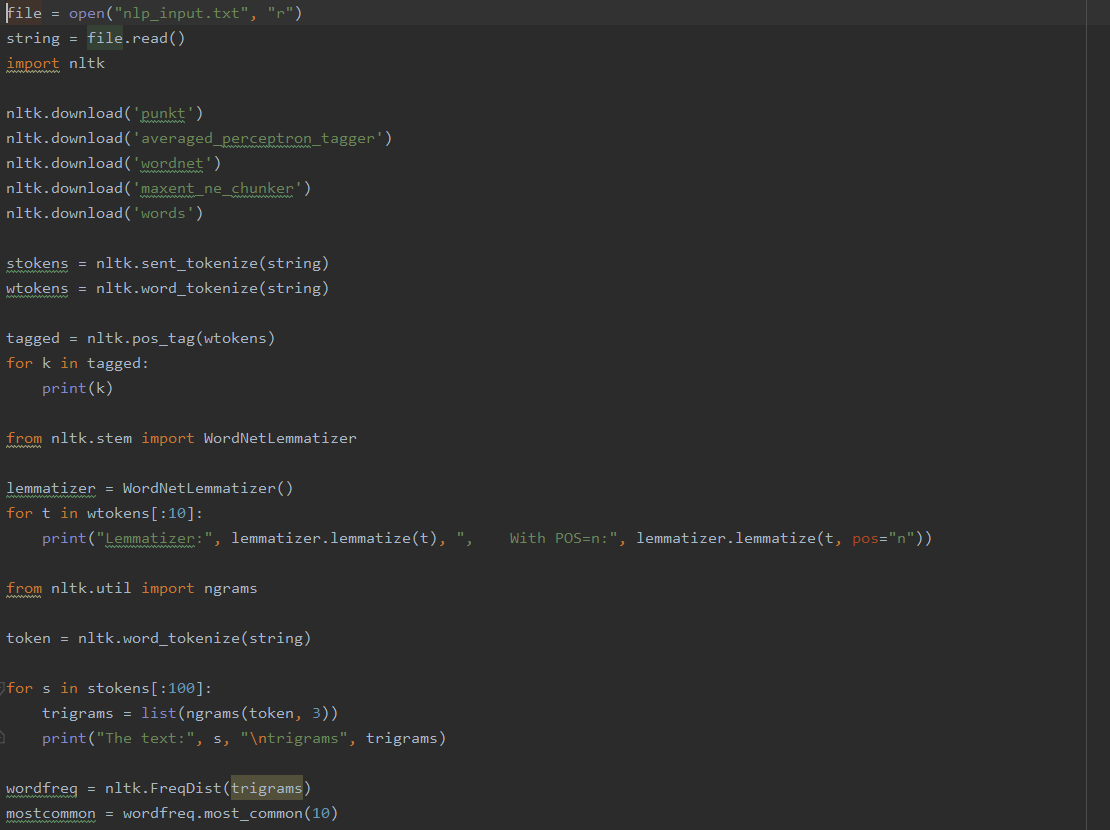
**Observation:**

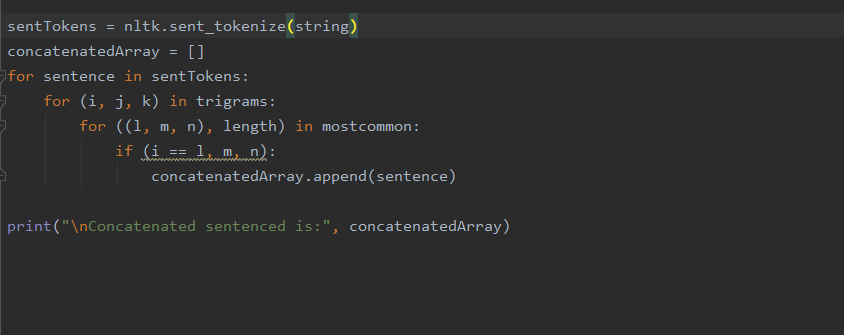
In this we can say that elbow method is used to find the optimal number of clusters. By looking at the above graph we can find that the number of optimal clusters is 3.

By looking at the silhoutte score we can say that the scores are all approximately nearer to 0.

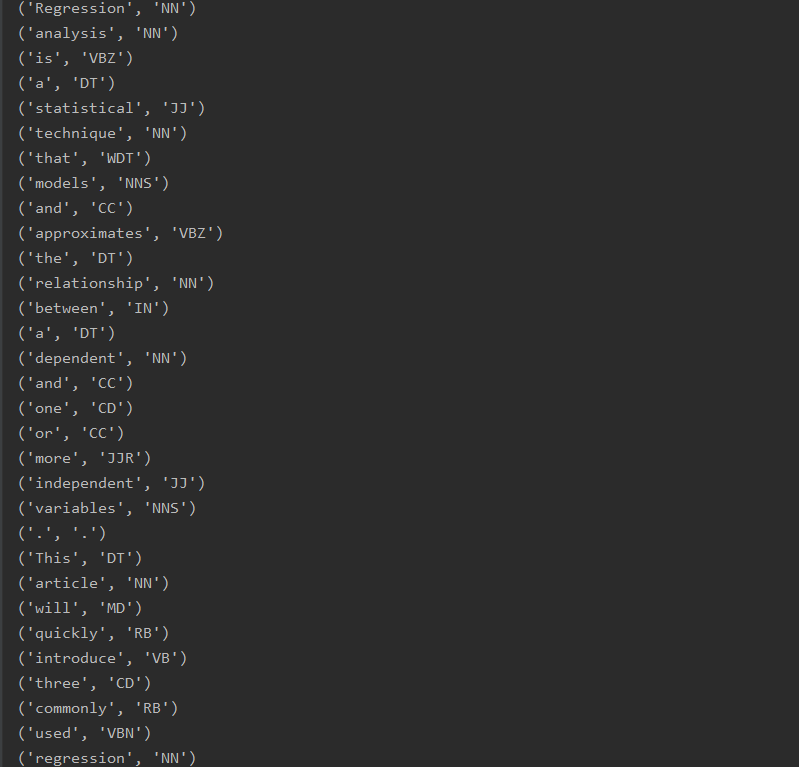
So the data points in the clusters are nearer to the decision boundary.

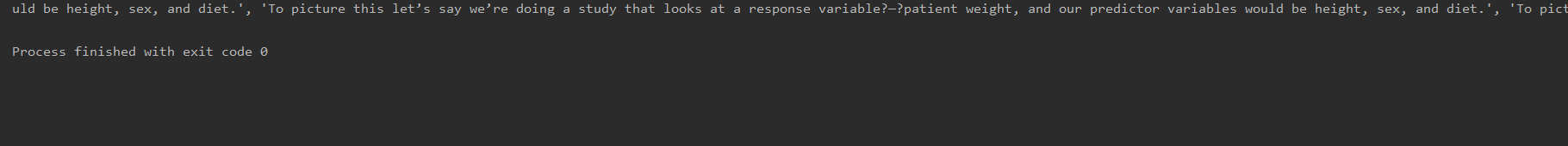
**Task 7**:





Output:

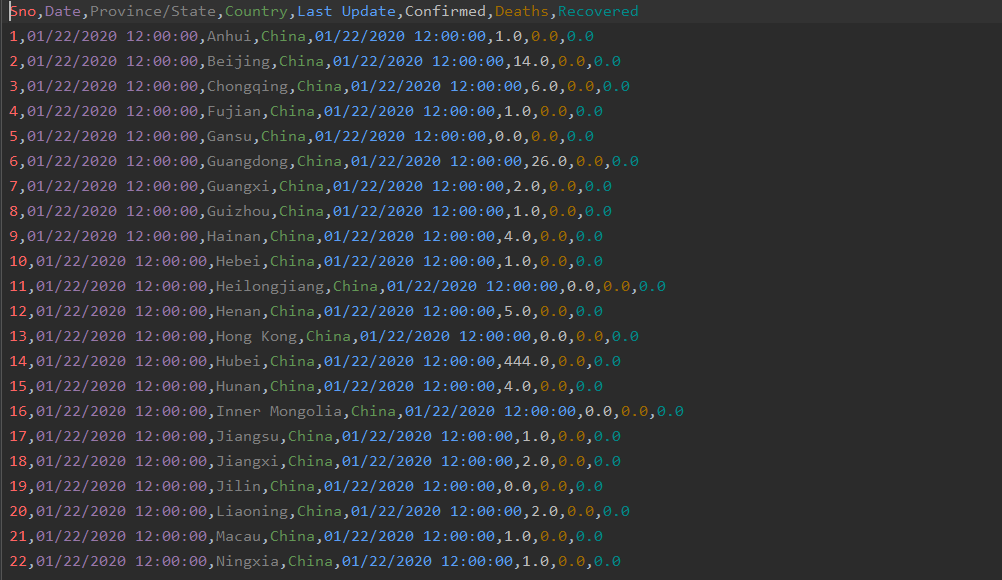


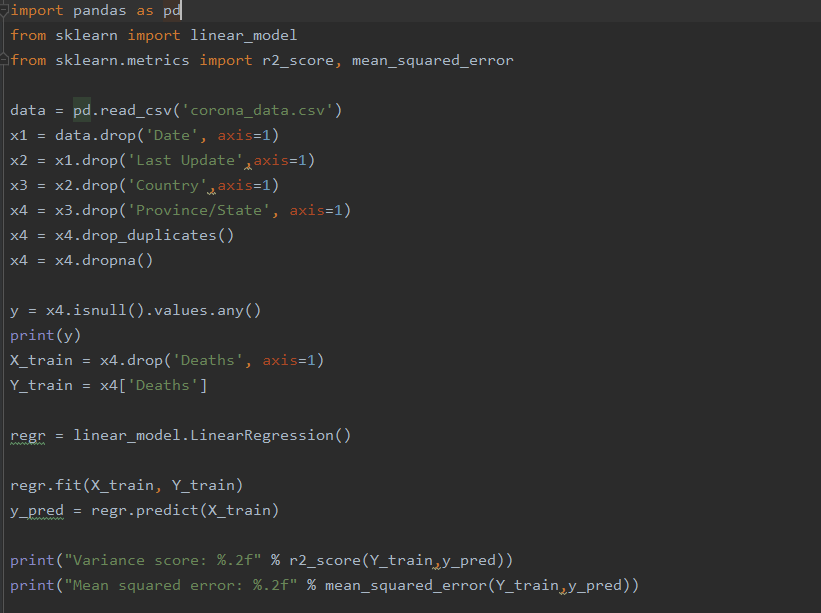


**Task 8**:

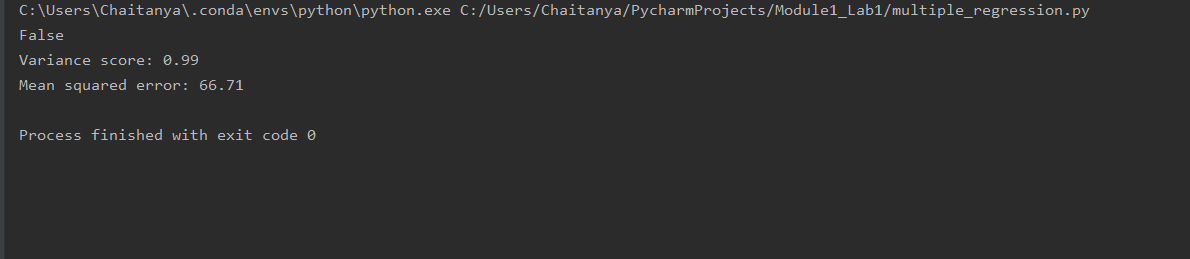
**Dataset**: corona\_data

You can get this data in <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#2019_nCoV_data.csv>.





Output:



**Observation:**

We can see there is a high variance, so the complexity of the model is high. So we can say the data provided is overfitted.

**Conclusion:**

In this Module we know to use the python datatypes, functions, classes and inheritance and came to know how to use the ML techniques such as Regression, Classification, Clustering using scikit-learn library and various methods in it.