**Predicting Crime Rate using Machine Learning Model**

**Problem statement**: Crime rate prediction using Machine Learning

Use machine learning to predict the percentage of crime taking place in the metropolitan areas of US by using the dataset provided.

Description of dataset is as follows -

1- land area : size in square miles

2-percent\_city: percent of population in central city/cities

3-percent\_senior: percent of population ≤ 65 years

4-physicians: number of professionally active physicians

5-hospital\_beds: total number of hospital beds

6-graduates: percent of adults that finished high school

7-work\_force: number of persons in work force in thousands

8-income: total income in 1976 in millions of dollars

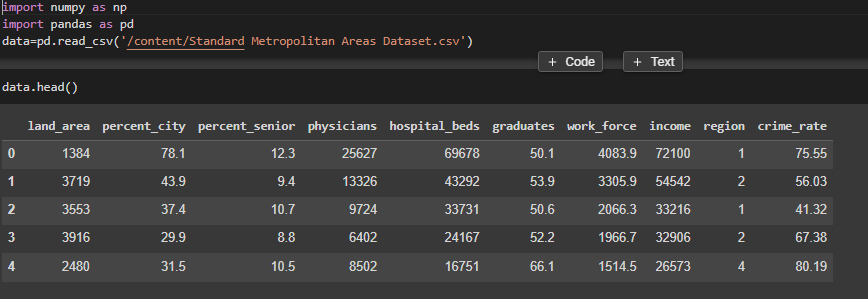
9-crime\_rate: Ratio of number of serious crimes by total population

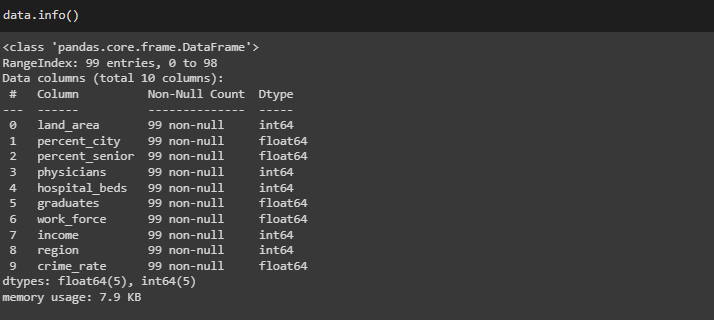
10-region: geographic region according to US Census

The file contains data of 99 standard metropolitan areas in the US. The data set provides information on 10 variables for each area for the period 1976-1977. The areas have been divided into 4 geographic regions: 1=North-East, 2=North-Central, 3=South, 4=West.

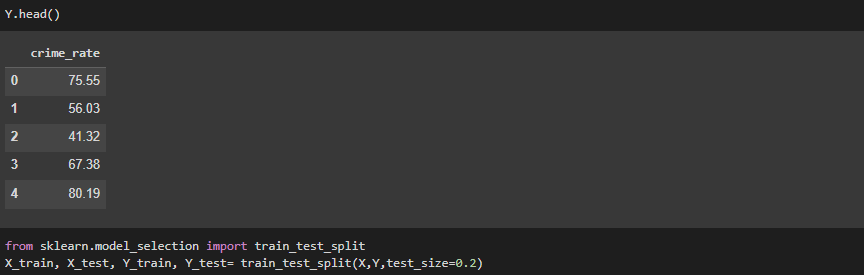
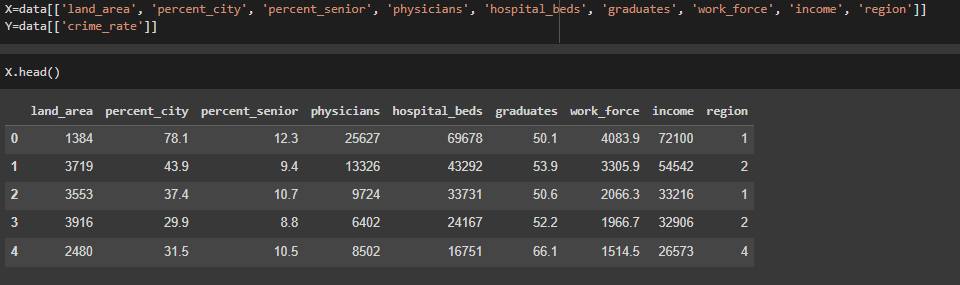
**Machine Learning Model using Linear Regression:**

Step1: Gathering the data set, Loading and Reading the data set.

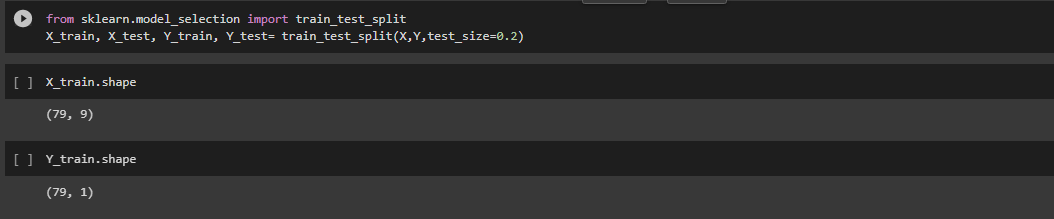




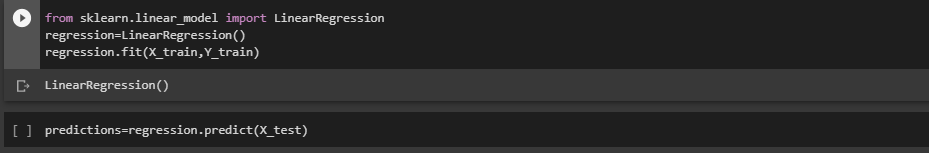
Step2: Divide the data into Dependent(Y) and Independent(X) variables.



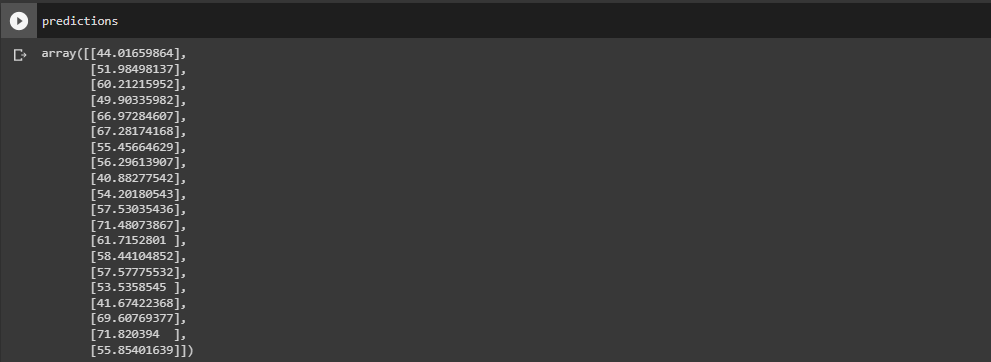
Step 3: Splitting the data into Training and Testing Set.



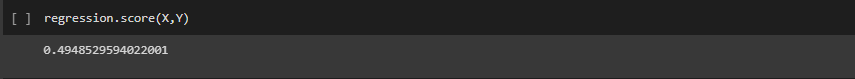
Step 4: Creating the Machine Learning model by using Linear Regression Algorithm.



Predicting the Crime rate in all the Region.

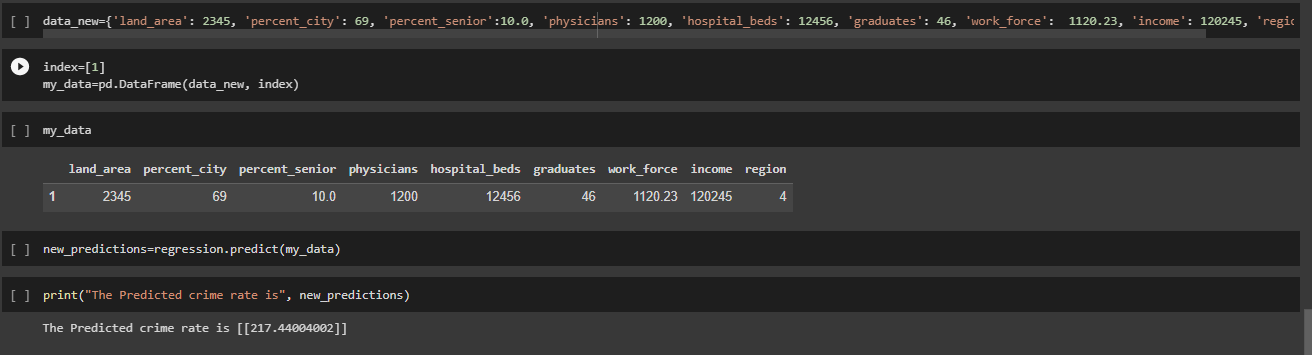


Checking the accuracy of the Predictions:



So the accuracy of the Prediction using Linear Regression is 49%. This Accuracy rate is very less to trust the predictions of ML model.

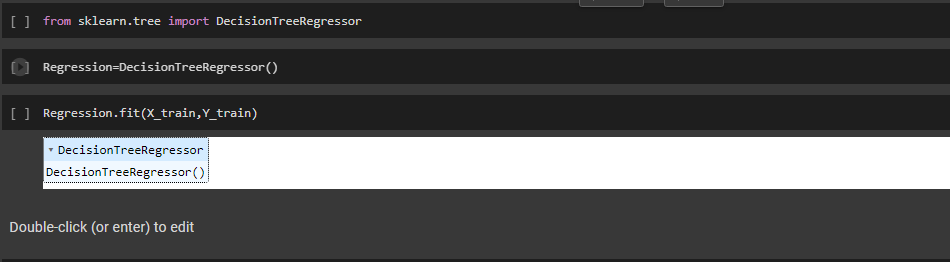
Lets try providing the New Data set and check what is the Predicted Value.

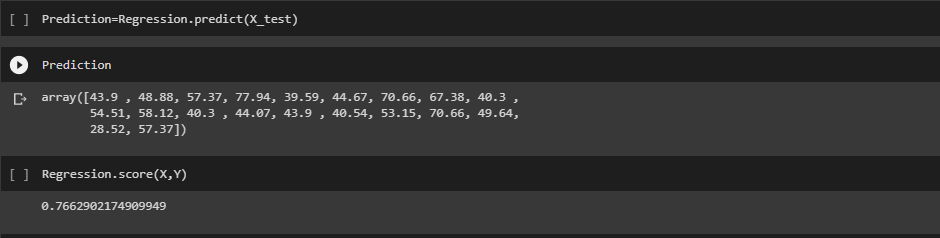


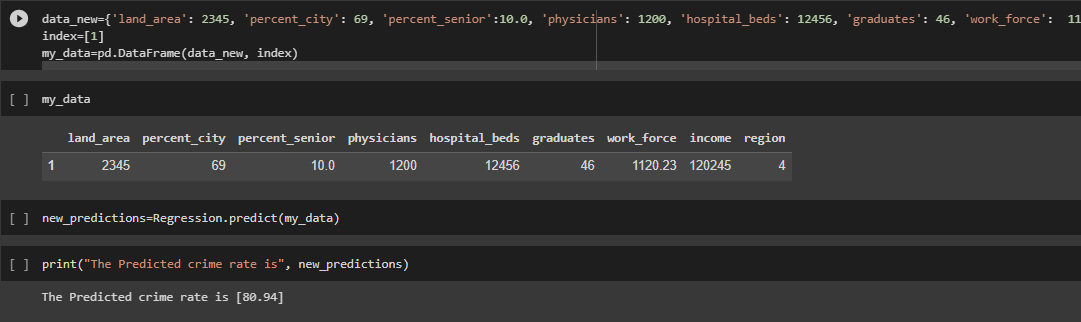
So the predicted Crime Rate is 217.44

Since the Accuracy of the Model is very less so lets try out using other ML Model if we get better Results

**Machine Learning Model using Decision Tree Regression:**

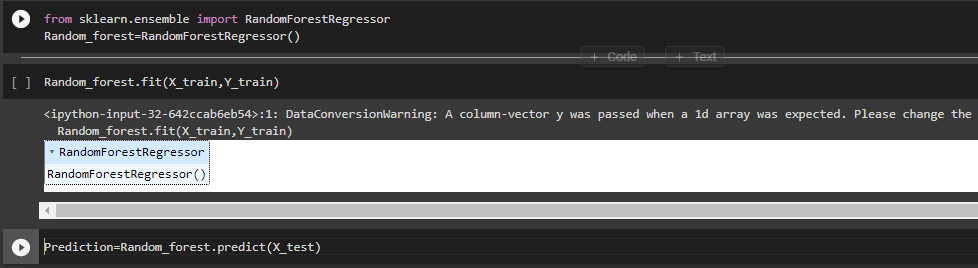
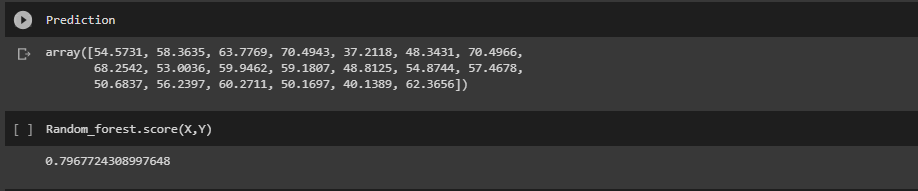
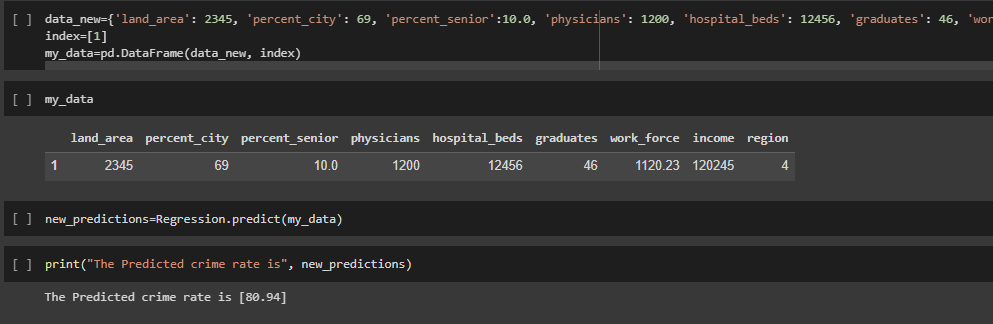






So in case of Decision Tree Regressor, we can see the Accuracy Rate is 76.6% and is better than Linear Regressor. With the same new data set, the predicted Crime Rate is 80.94.

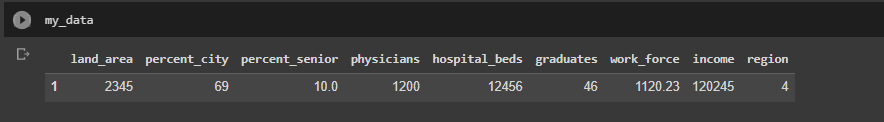
**Machine Learning Model using Random Forest Regression:**

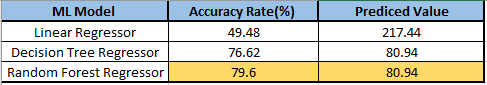
  

So in case of Random Forest Regressor, we can see that the Accuracy Rate is 79.6 and is better than both Linear Regressor and Decision Tree Regressor. With the same new data set, the predicted Crime Rate is 80.94.

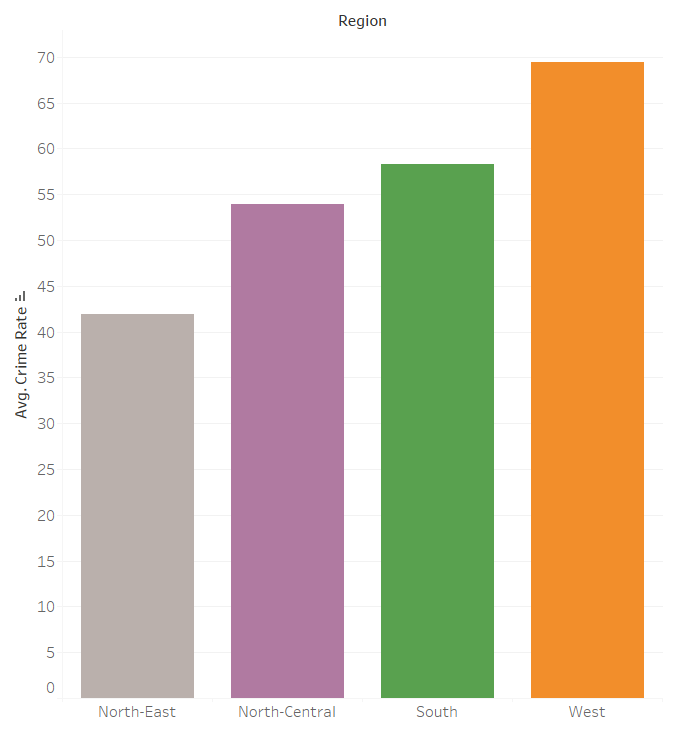
**Inference:**

By providing the same new data set to all the three Models, we can observe that Random Forest Regressor is able to Predict the Value with more accurately so my choice would be **Random Forest Regressor.**

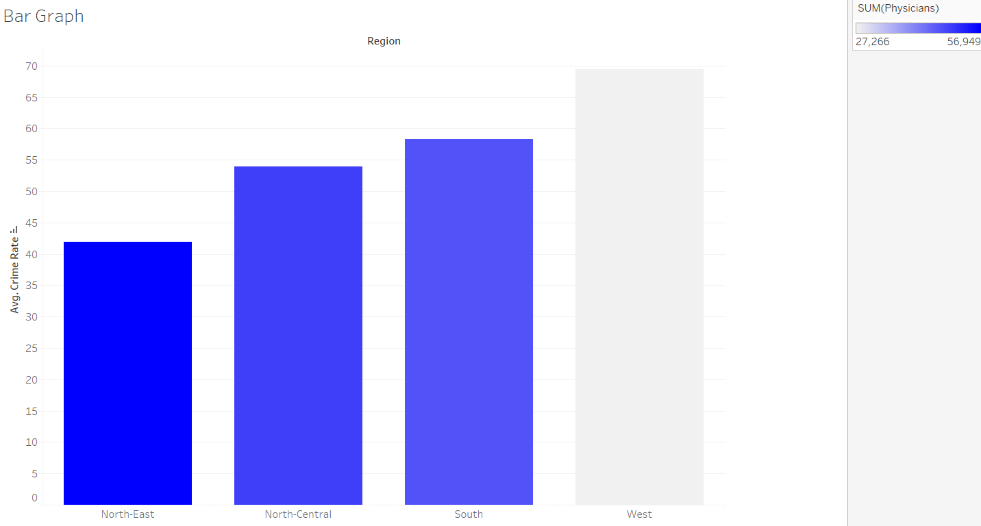




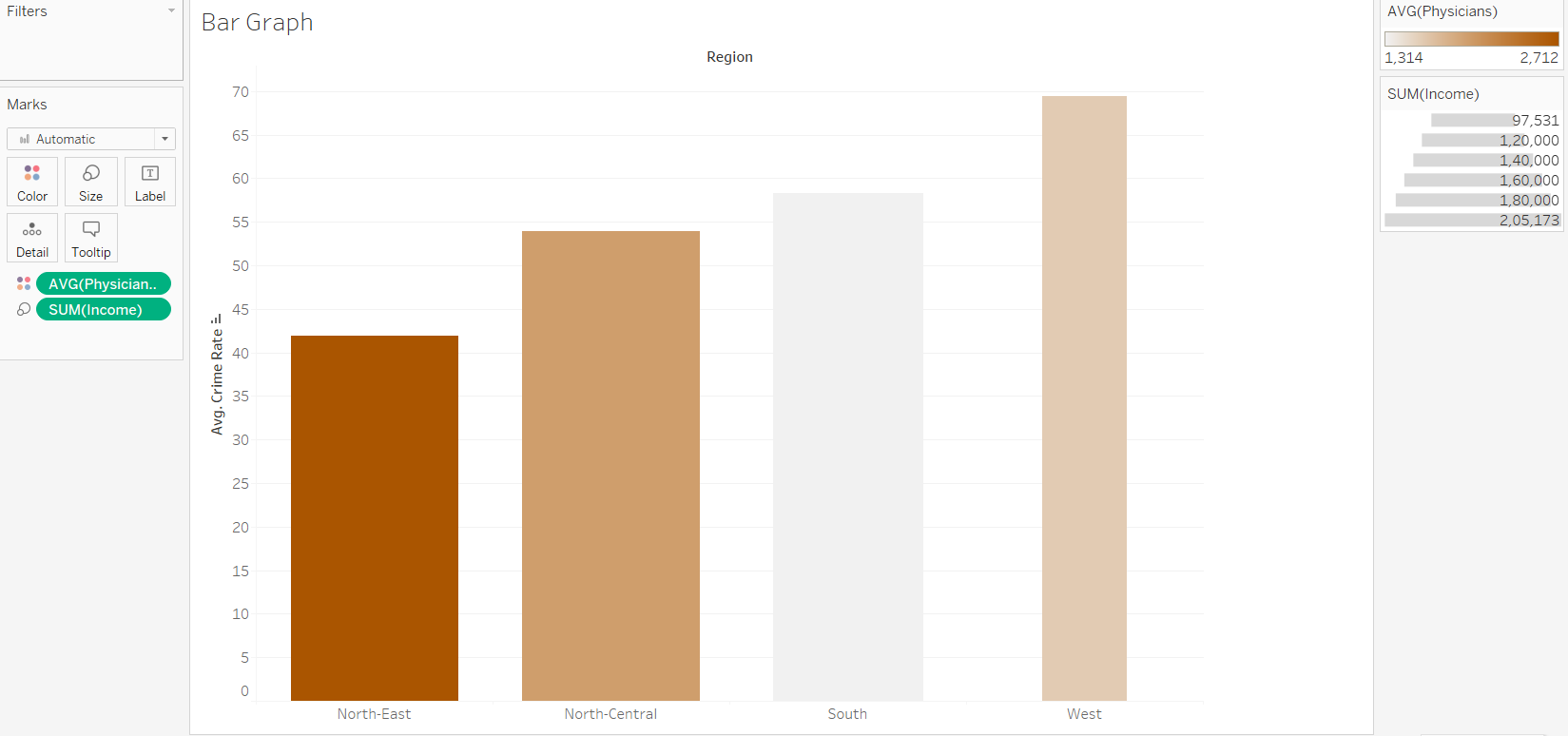
Visualizations of Data:



From this Graph we can understand that the Average crime rate is Highest in **West region** compared to all the other regions.



From the above graph we can understand that More Physicians reside in Northeast region. Since well educated people are less prone to involve in crime, the Northeast Region has the very less Average crime rate compared to all the Other Region.



From the above graph we can understand that since the Income in West region is very less people are involved in more crimes for survival. Because of this reason the Average Crime rate is Highest in West Region.