**Crimes in India - Capstone Project**

Indian Crime project is one of best project to understand machine learning and data analysis. India crime project is one of hard project for me. In this project we understand about how to collect data and, how to process data how to analyze data and how to prepare data for machine learning task.

We have given four phase to create this project and I we explain it in this blog.

In phase 1 we have given that collect Indian population data and Indian literacy rate data and Indian state wise area data.

In this we learn how to collect data and understand data marge data and more.

Also, institute tell us to collect every year data but data is given only 2001 and 2011 because Indian government organize population counting in every 10 years. Institute also give us so many crime csv files.

In second phase we have to do analysis of given statement. This task not easy for me because I have only two years data and given file have every year data. To analysis this to given statement, I edit data in excel. First statement is Analysis of Literacy Rate vs Total Crimes which is takes to so many times to complete it. I plot a scatter plot to analysis how literacy rate and crime are related. And concussion that literacy rate is high their crime rate is low. There has some exception also like Maharashtra has more than 90% literacy rate but crime rate is high. Second statement is Analysis of the type of crime vs each state vs Literacy rate, this task is also hard and time taking for me. Third Analysis of year-on-year total crime rate this task is easy for me. 2011 crime rate is lower than 2001 crime rate it because of increasing literacy rate and awareness of take action against crime. forth task is Analysis of area vs overall crime. Area is also important in analysis of crime rate because where area is high means population is high. And high area means population is high and high population means high crime rate. Fifth task is Analysis of Population vs overall Crime. This is also similar that high population means high crime. And last task is Each state crime report. And I got that Madhya parades has high crime rate in India.

We have to perform third phase in MySQL. We know that SQL is most important skill for data analysis. In this phase we learn how to write query to get necessary data and how to import data and how to export data. Fist task is to import data from csv file this is very easy task. Second and third task is that to write queries to extract data from table. In this phase we also have to do export table after creating queries.

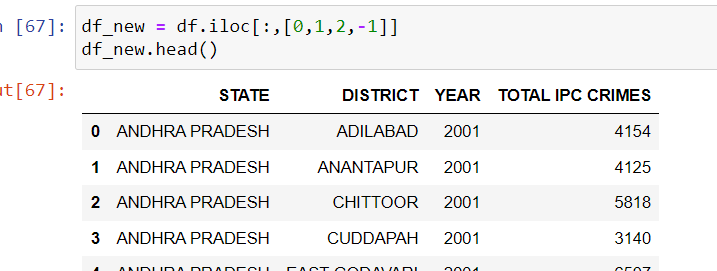
Fort phase is to create three cluster of like

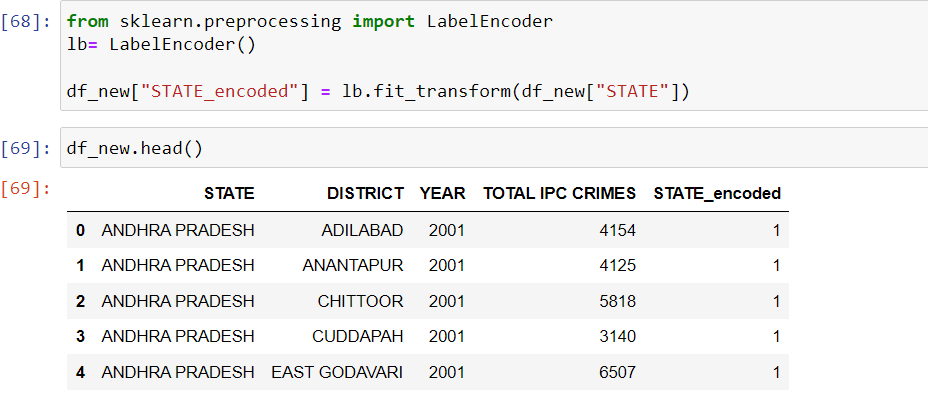
1. Sensitive Area's

2. Moderate Area's

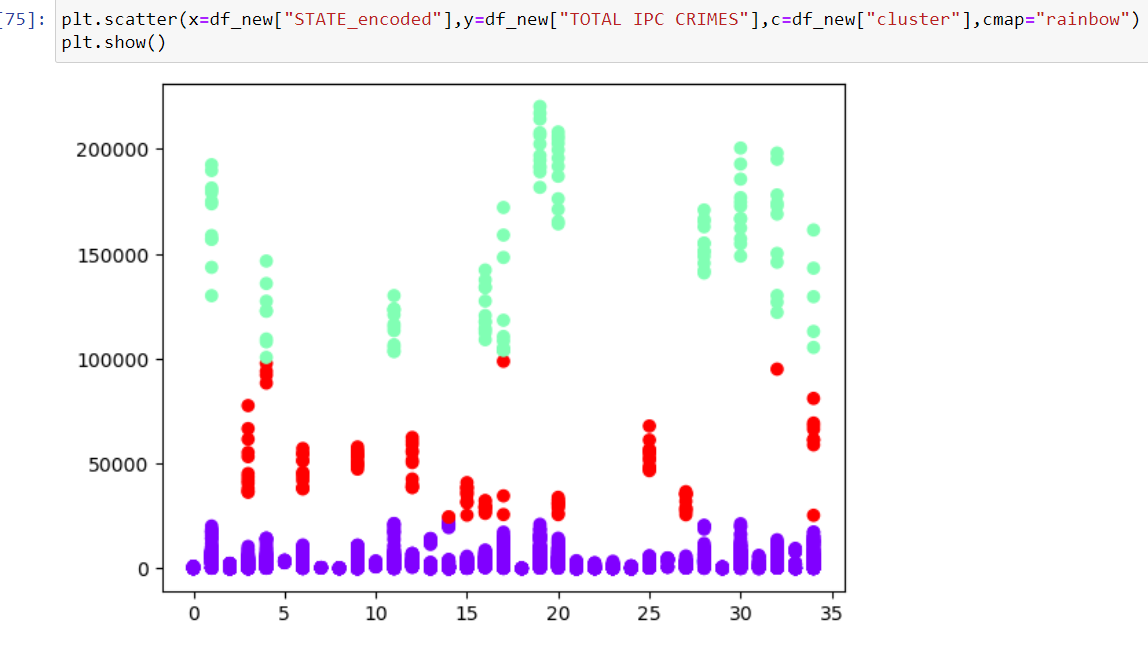
3. Peaceful Area's

In this task we have to apply general knowledge of how to select dataset and how to select feature and etc.

I select this 01 District wise crimes committed IPC 2001 2012 file for cluster task. I will create sub dataset for clustering.

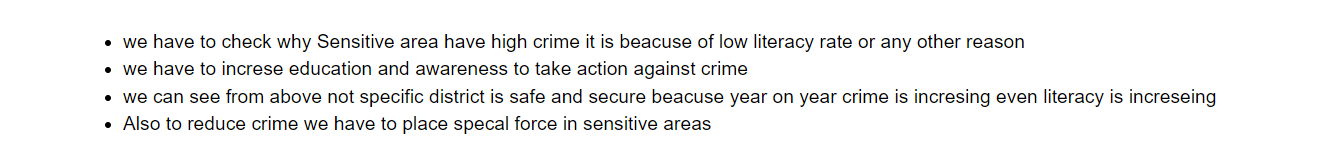
I convert state column in encoding task and add an new column .

I select TOTAL IPC CRIME column for cluster task and make cluster from area. And graph is given below .



We can see that in purple colour this area is a Peaceful Area's, in red colour this area is Moderate Area's and in light green colour area is Sensitive Area's.

Next task is to separate data and store in another data frame to analysis on it why area is different from each other. And lastly give concussion is given below.



Finaly this project is change of view of analysis. And processing of data to create machine learning model002E