

Name: Pranav Nair
Roll No.: 2019130042
Batch and A.Year.: Batch-A 2021-22
Course: Data Analytics

Experiment No.: 1

Aim: Perform EDA such as number of data samples, number of features, number of classes, number of data samples per class, removing missing values, conversion to numbers, using seaborn library to plot different graphs.

Code:

<https://colab.research.google.com/drive/1hitq9Trxv2c9MV2GVuNRxJR8Ek48eiHp?usp=sharing>

Conclusion:

In this experiment, the objective was to perform exploratory data analysis on the given dataset. The dataset for this experiment is about criminal data of the city of Denver between Jan,2012 and Oct,2016. Initially, I have performed basic operations like finding the data frame shape, finding the missing values and then removing them, number of classes and samples available per class. The classes were then converted into numeric representation. Using seaborn library, graphs are plotted of types like histogram, bar plot, distribution plot, scatterplot and line plot. From these graphs some of the inference made were:

- from the scatter plot, which shows distribution of crime across Denver, most of the crime took place in the central region and the number decreases as it moves to the corners
- from the line plots, one can see how the crimes increased/decreased from Jan,2012 to Oct,2016 and also how it changed on yearly basis
- the total number of offenses when viewed on the monthly basis has an uneven distribution but when viewed yearly, it shows increase in cases for 3 years with peak at 2014 and decline in the next 2 years.
- from the line plot between year of incidents and delay in reporting a crime, it is observed that the average time(number of days) between an incident occurring and a report being filed decreased.
- from the distribution plot for districts in Denver, we can see that most of the crimes that took place in each district were of the category 'Burglary'

- from the distribution plots, we can see that for incident address, district, precinct and offense code, how the total criminal acts that took place is distributed.
- from the histogram plot between average delay in reporting a crime and offense category, the category 'Larceny' has the highest average delay
- the address '745_S_BRYANT_ST' had reported an offense of Burglary, of type 'Burglary_No_Force' located in the district with id '4' and at precinct id '412' on '2015-02-02'. The crime took place on '2014-01-28'
- the precinct with id 323 has the highest average delay having a crime reported
- the district with id 2 and precinct with id 212 had the highest number of crimes reported
- neighbourhood - 'Elyria Swansea' had reported highest number of criminal offenses
- most of the incidents reported had connections to marijuana industry (~800/1253).
- 'burglary' is most repeated offense, the district with id '3' has seen a lot of cases of burglary while district with id '2' least.

Dataset: https://data.world/denver/marijuana-related-crime/workspace/file?filename=crime_marijuana.csv

GitHub: <https://github.com/pranav567/Data-analytics-lab-2021-22>