**Capstone Project**

**IBM Data Science Professional Certificate**

**By: Ajeo Mathew Cherukathara Mathew**

**INTRODUCTION**

Crimes happen ubiquitously no matter how large or educated the people are. Crimes committed across the world are increasing every day and it is difficult to tackle this problem. Traditional approaches like being cautious and improving the safety can’t guarantee us in reducing the crime rates. It is very difficult to understand the patterns of crimes but with the help of predictive analysis in machine learning, we can take necessary measures in order to prevent the crimes.

In a recent report by the Los Angeles Times, we could see that the L.A. homicides are down again and police credit thousands of extra patrol hours. Assuming that police can more often patrol a district assigned to them or the area around their police stations, the business question that I’m trying to answer is “Does the presence of constant police patrolling impact the number of homicides within the police district?”. In this analytics project,

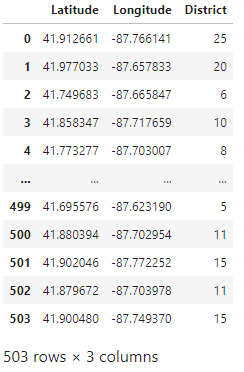
**DATA SCOURCE**

The data for this analysis project is taken from multiple sources.

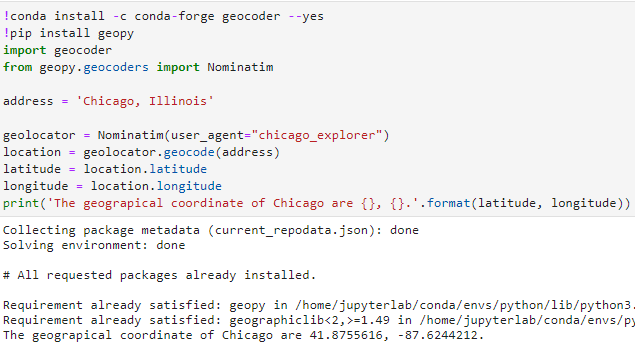
* Chicago Crimes\_2019 Dataset accessed from the Chicago Data Portal. This dataset is an open dataset which has records of crimes which happened in the year 2019 within Chicago. The dataset is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. We will be more interested in the Latitude, Longitude, and District fields of the dataset where the Primary Type field is Homicide.
* For studying the different police districts along with their geographic boundaries, we use the Boundaries - Police Districts (current) dataset from the Chicago Data Portal. The data will be a ‘.geojson’ file type.
* Finally to access information about locations of the different police stations in the city of Chicago, we use the FourSquare API.

**METHODOLOGY**

We started working with the Chicago Crimes\_2019 dataset importing the .csv file and filtering out rows where the primary category value is Homicide. I had previously worked on this dataset to perform predictive analysis for crimes that will happen in the city of Chicago. We did some initial data preparation and removed unnecessary columns and rows having null values that aren’t required. We ended up a dataframe with Latitude, Longitude and District fields.

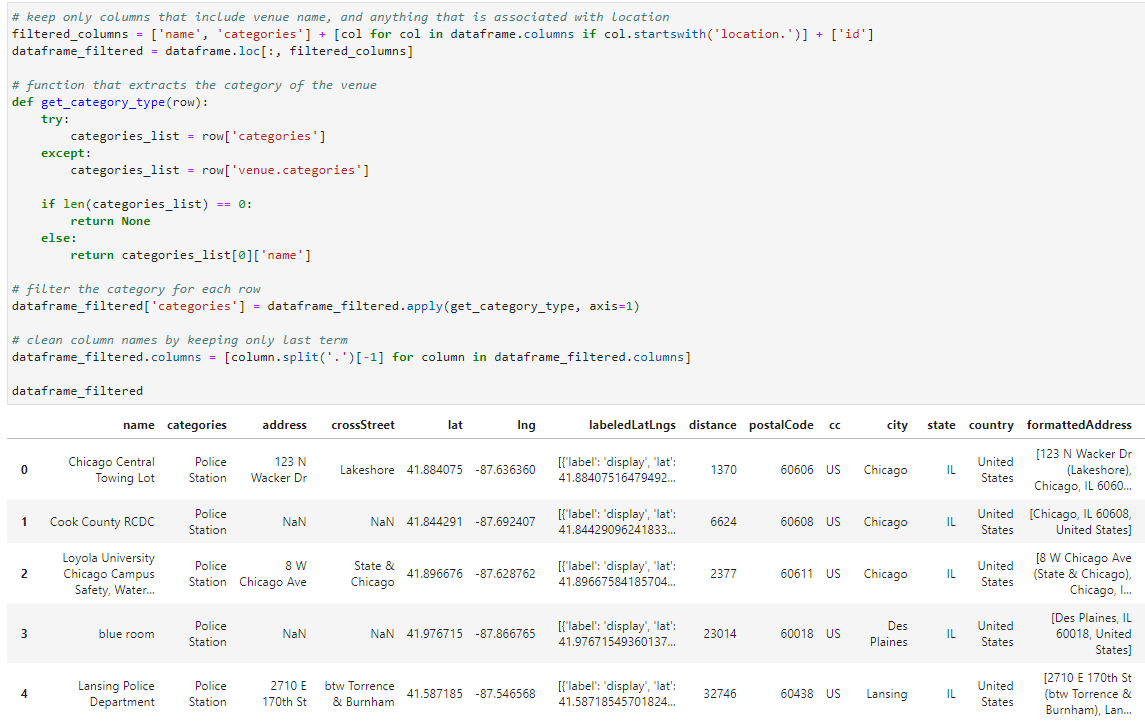


We then extracted the geographical coordinates for Chicago, Illinois with the help of geopy package.

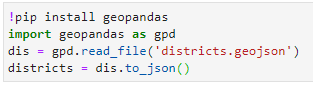


We used FourSquare credentials along with the Category ID = 4bf58dd8d48988d12e941735 to identify all the police stations within the radius of 25000 city center which we identified earlier. With the help of requests package, we connected with the FourSquared API to fetch the results of 30 police stations in the city of Chicago.





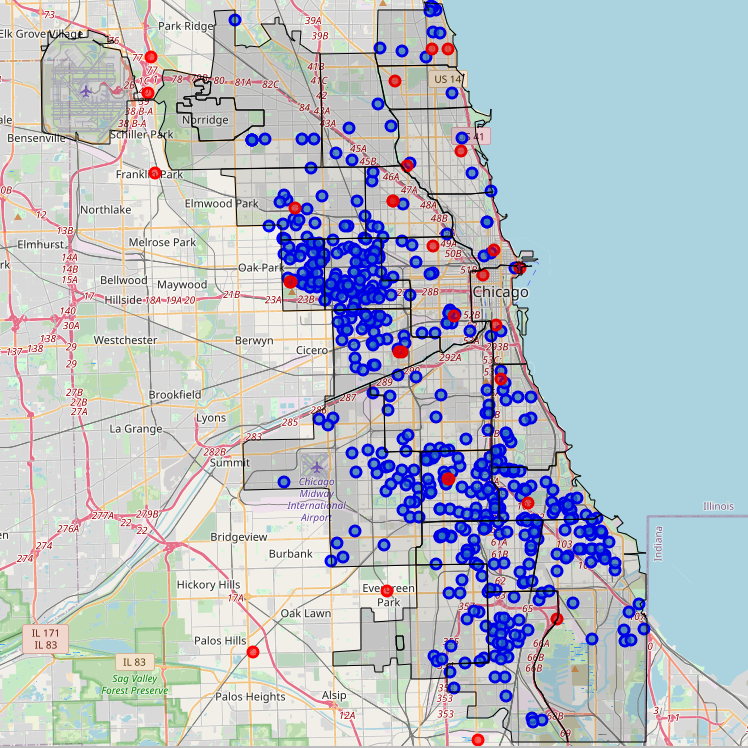
We imported the .geojson file which contains the boundary details of each of the police districts in Chicago. Once imported, we converted the data to .json file for plotting the data.



**RESULTS**

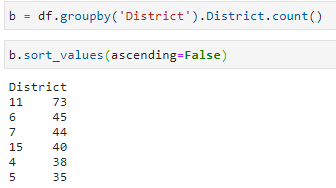
Using all the above data collected, we used Folium package to create the below map of Chicago. We added the locations of where the homicides happed in the year 2019 within the city identified as blue circles, added the locations of the police stations within the city identified as red circles, and finally added the police district boundaries to separate the cities in order to identify patterns.





**DISCUSSION**

* From the above graph, we were able to see that four out of the 25 districts had no police stations monitoring the region.
* There were clusters of homicide incidents especially Oak Park region and in East side of Chicago region.
* There were considerably lesser number of homicides in the districts closer to downtown Chicago attributing to the fact that there are larger number of police stations and high patrolling in the region.
* However same cant be said about districts such as 11, 6 and 7. These were the top three districts where the highest number of homicides were reported in the city.



* Surprisingly there are no police stations centered around all these three regions. This could have directly resulted in more crimes committed in these neighborhoods. Other studies have shown that most of the crimes committed are in the Oak View region where 11 and 15 districts are located with only one police station monitoring the region and West side of the city where 6, 7, 4 and 5 are located again with only two stations monitoring the three regions.
* We can confidently say and suggest to the police department in Chicago to increase patrolling in the Oak View and East side of Chicago. If the reports are true for the city of Los Angeles, we can definitely bring down the crime rates within the city.

**CONCLUSION**

Crimes are ever increasing around us but with the help of predictive analytics, we can be ever more cautious. United States have always been in the center of homicide repots whether its through gun violence or domestic violence.

Recent reports by CNN has shown a huge decline in the number of crimes happening inside the city. The report cites Chicago Police credit the drop-in violence partly to the investments in data-driven policing and the creation of strategic decision support centers in 20 of the city's 22 police districts. The report quotes

“Data-driven policing uses information from multiple sources to help police learn where crimes are happening and where they are likely to happen. Police said the decision support centers harness technology such as crime cameras and gunshot detection systems to help officers prevent crime and increase response times.”

The scope of crime data analytics is endless and there is always room for improvement.

**REFERENCES**

* <https://www.latimes.com/california/story/2020-01-16/l-a-homicides-are-down-again-police-credit-thousands-of-extra-patrol-hours> , report on the homicide rate in Los Angeles
* <https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Districts-current-/fthy-xz3r> , for the police district location data
* <https://data.cityofchicago.org/Public-Safety/Crimes-2019/w98m-zvie> , for the crimes 2019 dataset
* <https://www.cnn.com/2018/12/31/us/chicago-murders-drop-2018/index.html>, News article regarding the decrease in crimes in the City of Chicago