Visualizing Crime distribution in Chicago

Familusi Oluwatosin Adekunle

March 1, 2020

1. Introduction

1.1 Background

Chicago has many things that make it unique, including Lake Michigan, a booming population, a thriving manufacturing industry, a prime location in the heart of the country and much more. Running a small business in the city is unlike anywhere else. Here is a sample of what makes Chicago the perfect place in which to own and operate your small business. Each year, there are numbers of crimes committed against business which can sometimes lead to foreclosure, therefore, it is beneficial to assess the current security threat level before starting a business or before choosing a location for a business.

1.2 Problem

Data to be visualized are the reported crimes within Chicago including specific locations. This project is focused on revealing areas of high crime rate to give investors or entrepreneurs an idea of what type of crime and how frequently such crime would be expected to happen in such neighborhood.

1.3 Interest

Investors and entrepreneurs looking to locate favorable areas for business would be very much interested in the prospect of understanding the security history of where a business is to be located. Other possible interest might arise for new home owners and tourist who would love to have a brief review of the locality.

2. Data Acquisition and cleaning

2.1 Data Sources

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent

seven days. **Crimes reported are with coordinates are found on here and the extract can be found here.**

2.2 Data Cleaning

Data downloaded from multiple sources were cleaned by selecting the needed rows for visualization. Some rows under coordinate columns ('LONGITUDE', 'LATITUDE') contained 'NaN' values and were dropped so as to make visualization of such attributes possible.

I made a connection request using generated CLIENT_ID and CLIENT_SECRET credentials by passing 'CHICAGO, IL' as the address of the location to be explored using foursquare geolocator to get the coordinates for the city of Chicago which the crime data will be superimposed on.

An instance of the Chicago metro was generated using 'Stamen Toner' mode. Maps generated are related to the cleaned data to visualize the specific location of each crime.

2.3 Feature Selection

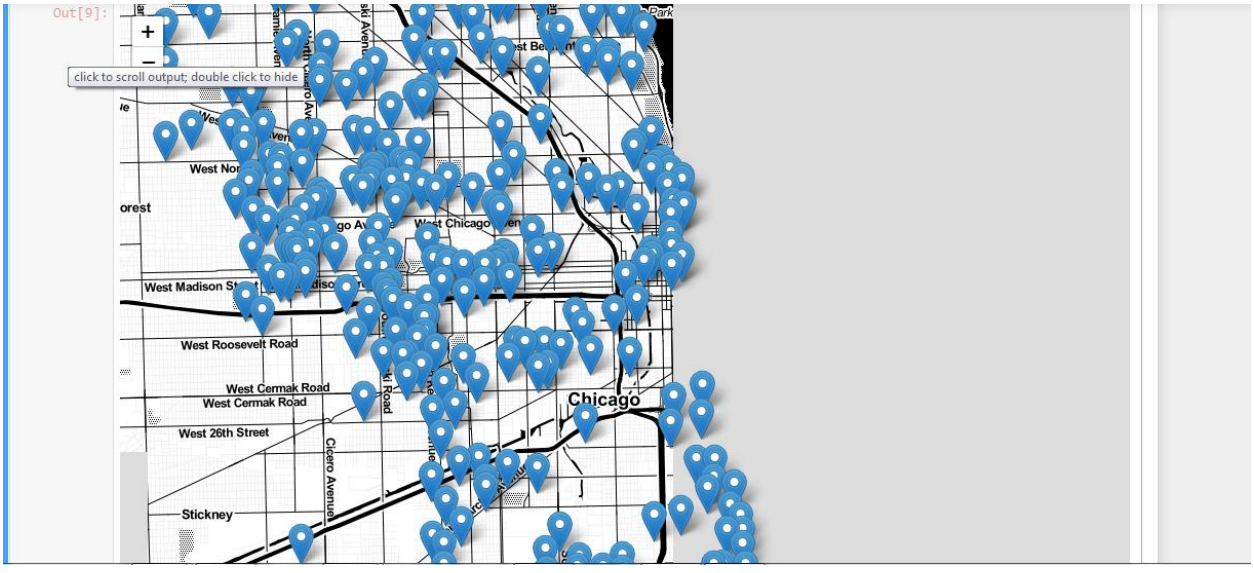
Some features were selected for analysis and are briefly summarized in the table below

KEPT FEATURES	DROPPED FEATURES	REASON
PRIMARY_TYPE, BEAT, DISTRICT, WARD,	ID, CASE_NUMBER	ID and CASE_NUMBER are unique identifier codes given to each arrested suspect and their respective case files. This attribute does not offer much information with regards to the focus of the project
	DATE, UPDATEDON	DATE AND UPDATEON are datetimeobjects from when the crime was documented and the day month and year on which official changes were made to those records
LOCATION_DESCRIPTION, COMMUNITY_AREA_NUMBER	BLOCK, ARREST, FBI_CODE	BLOCK are house addresses or categorical descriptions of crime

		venues, ARREST is represented by a either TRUE or FALSE to show if the offender was apprehended or not.
	IUCR,	
LATITUDE, LONGITUDE,	X_COORDINATE, YCOORDINAT, LOCATION.	COORDINATES and LOCATION DATA are replicas of LONGITUDE AND LATITUDE used for visualization.

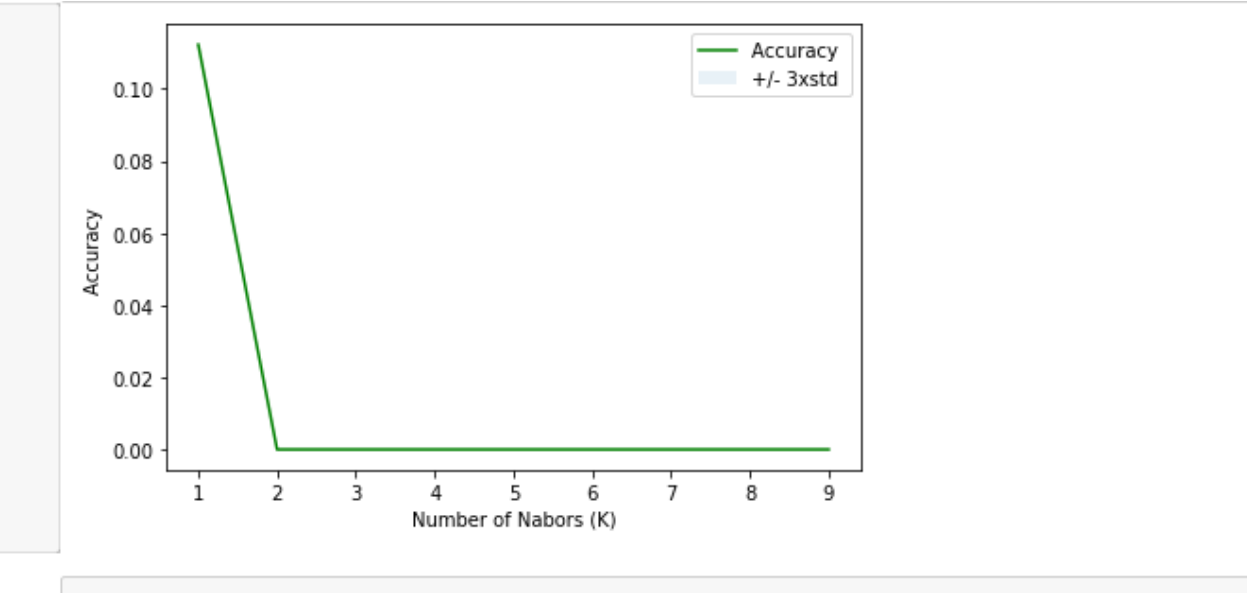
2. EXPLORATORY DATA ANALYSIS

Map visualization indicated areas with high incidence of crime in the past years. Also, the map revealed areas of less crime activity, this areas are more abundant towards the western side of the metropolis.



KNN

Aim of using K Nearest Neighbors is to determine of Data (form 'DISTRICT', 'WARD', 'COMMUNITY_AREA_NUMBER', 'X_COORDINATE', 'Y_COORDINATE', 'LONGITUDE', 'LATITUDE') can be used to predict the type of crime in areas of the city at the optimum K of 2, the model returned a prediction accuracy of .3 show it as very weak.



Conclusion

In conclusion, the visualized data shows areas located farther from the west are more crime free in contrast to the areas such as Lincolnwood and west Chicago. The data also shows some disparity in between 2 neighboring communities for example where there appears to be less crime activity. This can be as a result of the population or other factors.

Future Directions.

The data utilized gets updated dynamically, hence conducting such visualizations or even creating software that automatically indicate recent crime activity via the foursquare API might be a very good way forward.

The analysis can be taken further through the use of DBSCAN for grouping neighborhoods with similar crime statistics. This can help the authorities as well to

know the most expected crime for each neighborhood and areas where they can be expected using the BLOCK feature.