Project Report

Programming and Database Fundamentals for Data Scientists (EAS 503)

Jin Chen

Abstract

My purpose for this project is to help people to better understand the crimes in New York City. In order to do that, I am going to analyze the demographic pattern of suspects, time frame of the crime, and the types of crimes under each region.

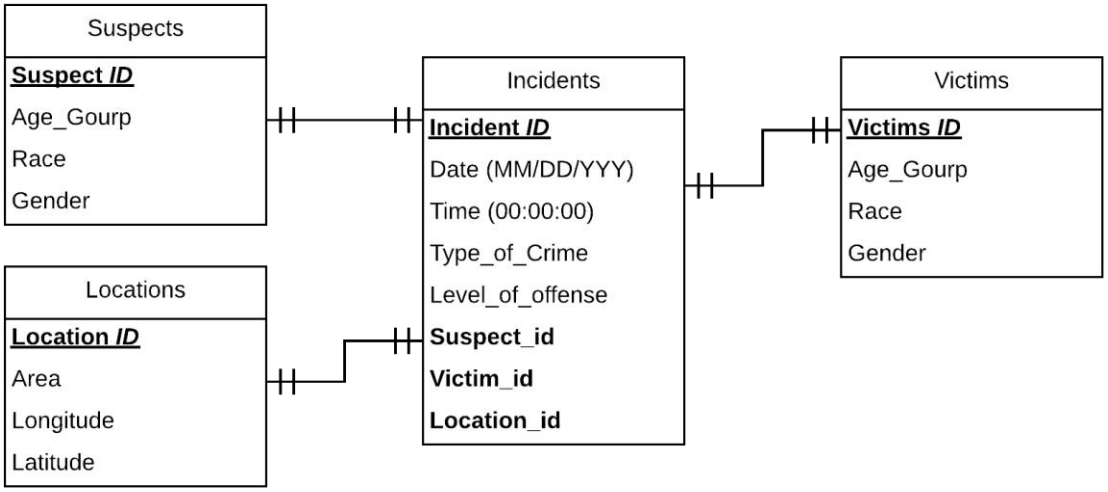
Start Up

First of all, please install Plotly and Matplotlib, before running Visualization.ipynb:

plotly.py may be installed:

- using pip: $ pip install plotly==4.3.0
- or conda: $ conda install -c plotly plotly=4.3.0

After we pick the columns that we need, we created a relational database based on the ER Diagram here (see file Create_database.ipynb):



Before answering any the question below, we wrote SQL-Query (see Visualization.ipynb) to load the data to python memory, and then put it into Panda-DataFrame. Then we manipulate the data and make some changes to transform the the data into the way we want in order to visualize the data. We use two libraries to visualize the data: Plotly, Matplotlib.

Questions with our Findings

All the charts can be found in *Visualization.ipynb* *in order

1. *The demographic pattern of the crime such as gender, race, and age group*

Half of the suspects (54.5%) are 25~44 years old. Age group of 45~64 and 18~24 are around 20%. The half of the suspects (54.5%) are Black. The White and White Hispanic occupies 36.6% of the suspects, and the others are around 13%. The 60.7% suspects are male, and female only have 18.7%. The rest are unknow.

2. *The region has the highest crime rate*

***Formula**:*

Crime rate = number of crime in each region / population of the each region

Staten Island has the lowest crime rate which is lower than 0.17%, and Brooklyn (1.16%) and Manhattan (1.01%) take top two places. Then, the crime rate of Bronx and Queens are around 0.8%, respectively.

***The results after normalized is very close to unnormalized results.**

3. *What time during the day happens the most frequently*

The lowest crime rate is happened at 5:00 AM during the day. Criminals are active between 12PM and 18PM, and the crime rate reaches its peak at 15PM. The crimes mostly occurred between 15:00PM and 19:30PM on Thursday and Friday. Moreover, there is less crimes happen around Tuesday. Furthermore, there are three levels of the crimes which are Felony, Misdemeanor and Violation. The huge amount of Felony occurs at 12PM, 15PM to 19PM on Thursday and Friday, at 12PM on Monday, and at 12am every day. There is less Felony happening on Tuesday. The Misdemeanor and Violation have the similar pattern as the Felony besides that the crime rates of these two levels of crime are not very high at 12am.

4. *Which type of crime happens most frequently*

The top three are “Harassment, SUBD 3,4,5” (like stalking), “Assault 3” and “Larceny, Petit From Store-Shopl” (stealing from shops).

The detail of the crime can be found by using the function ‘get_detail()’ at the end of *Visualization.ipynb

5. *Which type of crime happens most frequently in a certain borough*

In Bronx, Brooklyn, Queens and Staten Island, “Harassment, SUBD 3,4,5” happens mostly comparing with other types of crime, respectively. In Brooklyn, the crime rate of “Harassment, SUBD 3,4,5” is higher than the rates in other regions. In Manhattan, “Larceny, Petit From Store-Shopl” and the crime rate nearly twice the second highest crime rate of “Larceny, Petit From Store-Shopl”.

Conclusion

1. **Try not go out at 12am because it is the time the Felony happens quite often comparing with the other levels of the crime.**
2. **Staten Island is a relatively safe place, so feel free to travel or live there.**
3. **Stay away from the random guys who come and talk to you, and be careful with stalking.**
4. **In Manhattan, the shops need to pay attention to the stealing.**
5. **Avoid disputes with other people all the time.**

Data Source:

1. NYPD Complaint Data Current (Year To Date): https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243

2. Population of each borough in NYC: https://en.wikipedia.org/wiki/Boroughs_of_New_York_City