BATTLE OF THE NEIGHBORHOODS

THE RISE OF GULLMAN

Coursera Capstone Project

IBM Data Scientist Professional Certificate

by Anthony Chacon

Introduction

Setting the Stage

Target Audience

Ronnie Gulliver is a multi-millionaire, international playboy, philanthropist, bodybuilder, two-time kite-surfing champion, and CEO of **Gulliver Investments INC.** He's looking to break up the monotony of the day-to-day and has decided to start fighting crime. He's decided that there's no better place to start than… **NEW YORK CITY**!

Ronnie Gulliver

GULLMAN Costume

Background

I was recently picked up for a data science internship at **Gulliver Investments INC** and was selected to serve as **Mr. Gulliver's** personal data scientist. He told me that I have "a trustful face" *(whatever that means)* and let me know about his plans to save **NEW YORK CITY** as... **GULLMAN**! He showed me the costume and ran his plan down.

According to the **FBI**, crime rates in major cities within the United States have been increasing exponentially. While the violent crime rate average for the country was around **200** per 100,000 inhabitants in 1997, today the number sits in the high **700s**. However, there are still some cities who face a higher than average crime rate - notably **NEW YORK CITY**!

Problem

Most sociologists view crime is a social problem, with many of the motivations between crimes committed including factors such as lack of police funding, issues with armed robberies, gang-related violence and more. But not in **NEW YORK CITY**! Crooked cops, corrupt politicians, gangs of themed thugs roaming the streets -- these are a constant problem in **NEW YORK CITY**! Well... not anymore! **GULLMAN** is ready to crack some skulls and stop crime through vigilante justice! He just needs to know where and if it's worth the trouble…

Mr. Gulliver explained to me that it would look bad for the brand if **GULLMAN** started from the bottom and gradually worked his way up. Therefore, **Mr. Gulliver** wanted me to find the most crime-ridden part of **NEW YORK CITY** and present the information back to him, so that he can make the call to suit up or stick to being a millionaire.

Data & Methodology

Finding the Right Fight

In order to determine the best location for **GULLMAN** to start fighting crime, I need to examine criminal statistics for **NEW YORK CITY** and explore neighborhoods to find the locations where crime is most prevalent. In order to do so, the following steps will be taken:

1.) Download, Explore, and Clean the Criminal Dataset.
2.) Analyze the data by comparing crime statistics based on **Borough**, **Precinct**, **Offense**, and **Locations** using **Matplotlib** to graphically represent the data for easy consumption.
3.) Map out the area using **Folium,** so that **GULLMAN** can get started.

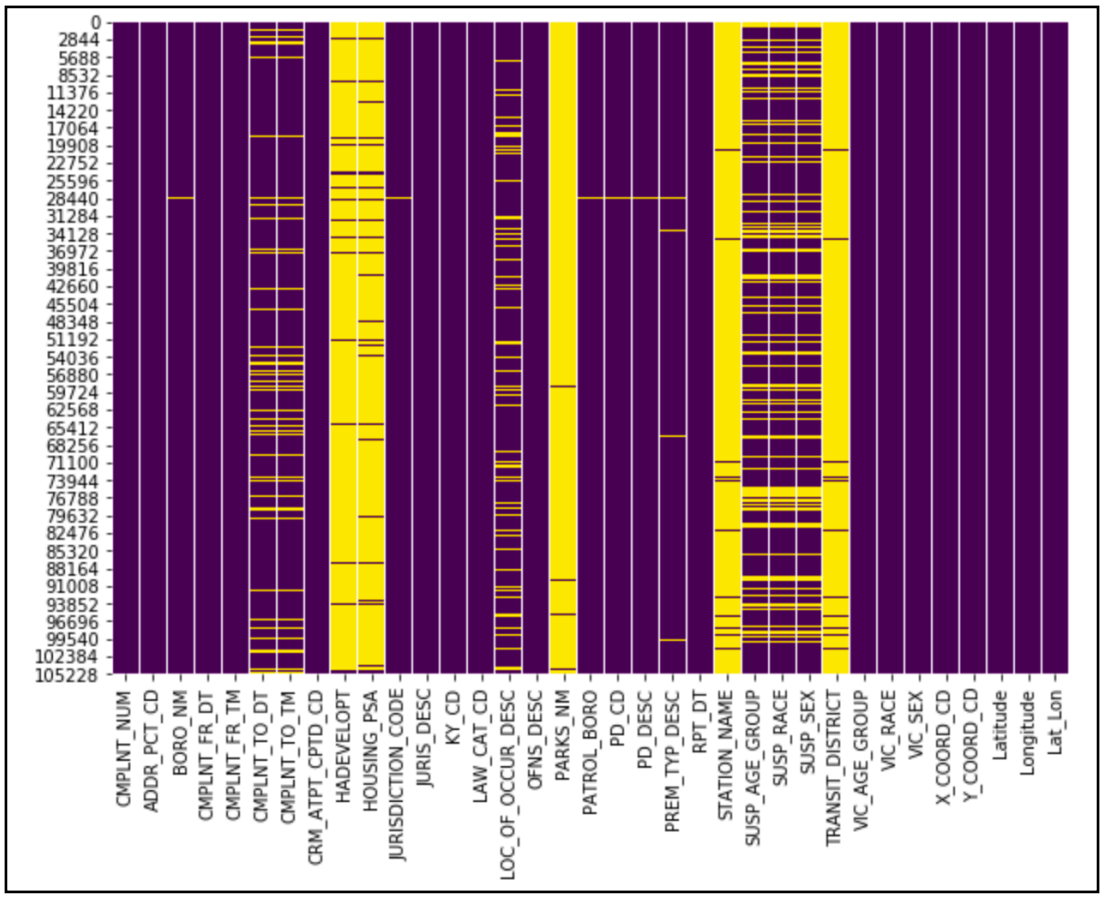
Download & Explore Criminal Dataset

In order to gather the information I needed, I used the **GULLMAN *"CRIME-PUTER"*** to hack the NYPD Database*. From there, I explored the dataset and began clean up of the DF.

CMPLNT_NUM	ADDR_PCT_CD	BORO_NM	CMPLNT_FR_DT	CMPLNT_FR_TM	CMPLNT_TO_DT	CMPLNT_TO_TM	CRM_ATPT_CPTD_CD	HADEVELOPT
466502077	75	BROOKLYN	03/30/2020	17:30:00	03/31/2020	06:53:00	COMPLETED	NaN
303191835	77	BROOKLYN	03/28/2020	19:30:00	03/28/2020	20:30:00	COMPLETED	NaN
735488557	43	BRONX	03/29/2020	14:10:00	NaN	NaN	COMPLETED	NaN
315962428	40	BRONX	03/29/2020	07:10:00	03/29/2020	07:16:00	COMPLETED	NaN
165437868	114	QUEENS	03/27/2020	13:15:00	03/27/2020	14:00:00	COMPLETED	NaN

(*Source: https://data.cityofnewyork.us/resource/5uac-w243.json)

There was a lot of **NaN**, so I used **SNS** to build a heatmap to ID the relevant data and start cleaning up.



(**Compiled SNS showing NaN Data**)

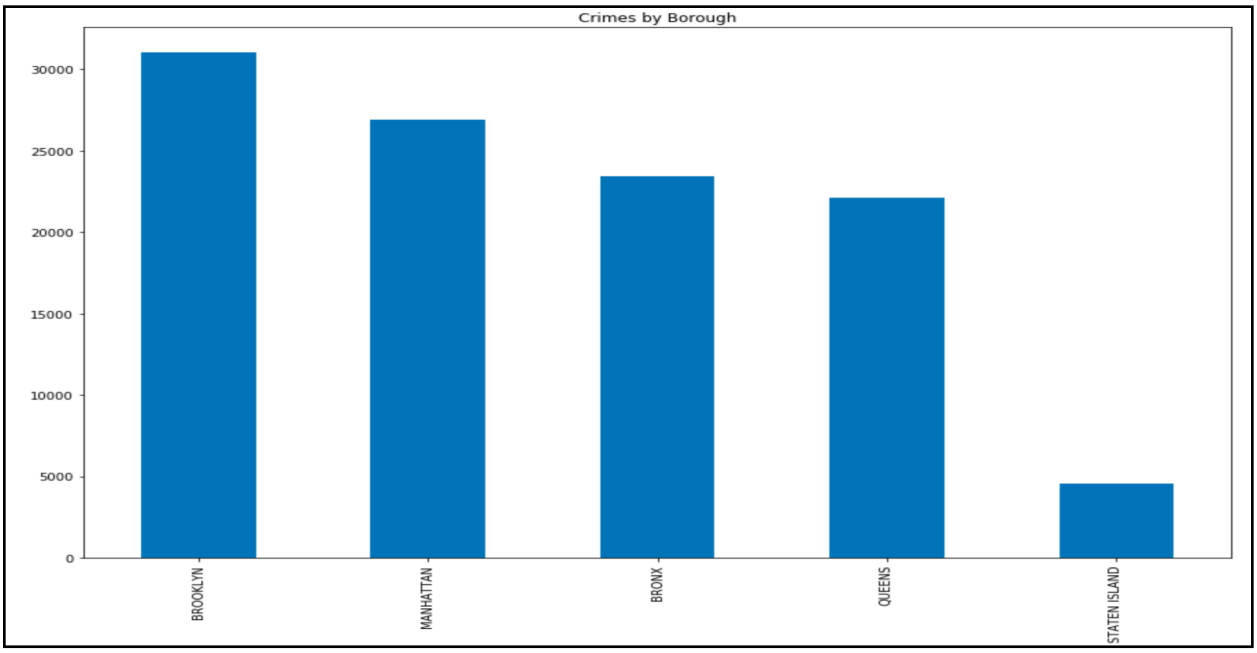
From there, I built a DF list, dropped most of the columns and left the relevant data. I continued organizing, renaming, and pairing down the data until I was able to build a DF that was easy to read and understand.

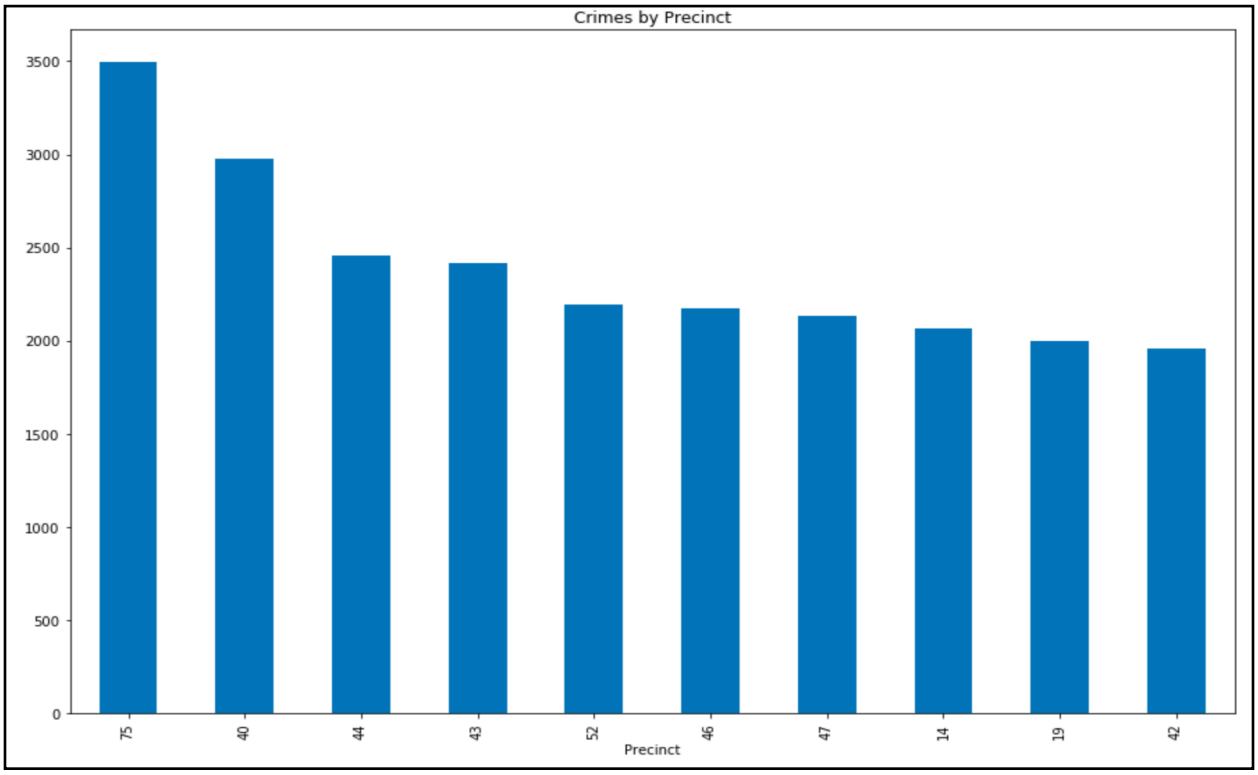
	Borough	Precinct	Location	Offense	Latitude	Longitude	Lat_Lon
0	BROOKLYN	75	STREET	PETIT LARCENY	40.656991	-73.876574	(40.65699087900003, -73.87657444799999)
1	BROOKLYN	77	STREET	RAPE	40.674583	-73.930222	(40.67458330800008, -73.93022154099998)
2	BRONX	43	DRUG STORE	PETIT LARCENY	40.830443	-73.871349	(40.83044253800006, -73.871349147)
3	BRONX	40	GROCERY/BODEGA	PETIT LARCENY	40.817878	-73.916957	(40.817877907000025, -73.91695668199996)
4	QUEENS	114	OTHER	ASSAULT 3 & RELATED OFFENSES	40.752011	-73.935872	(40.75201086000004, -73.93587196099996)

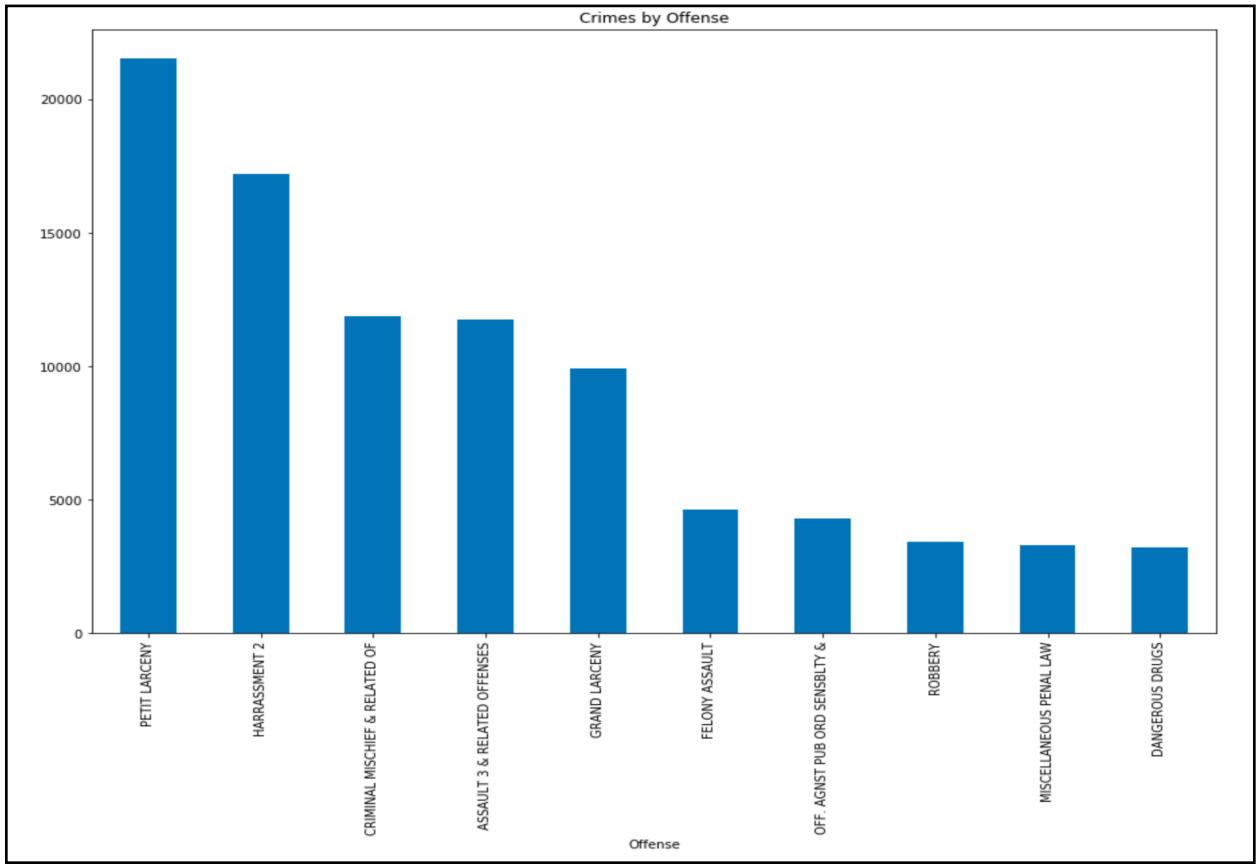
(**Final DF with new COL names and order. All others were discarded.**)

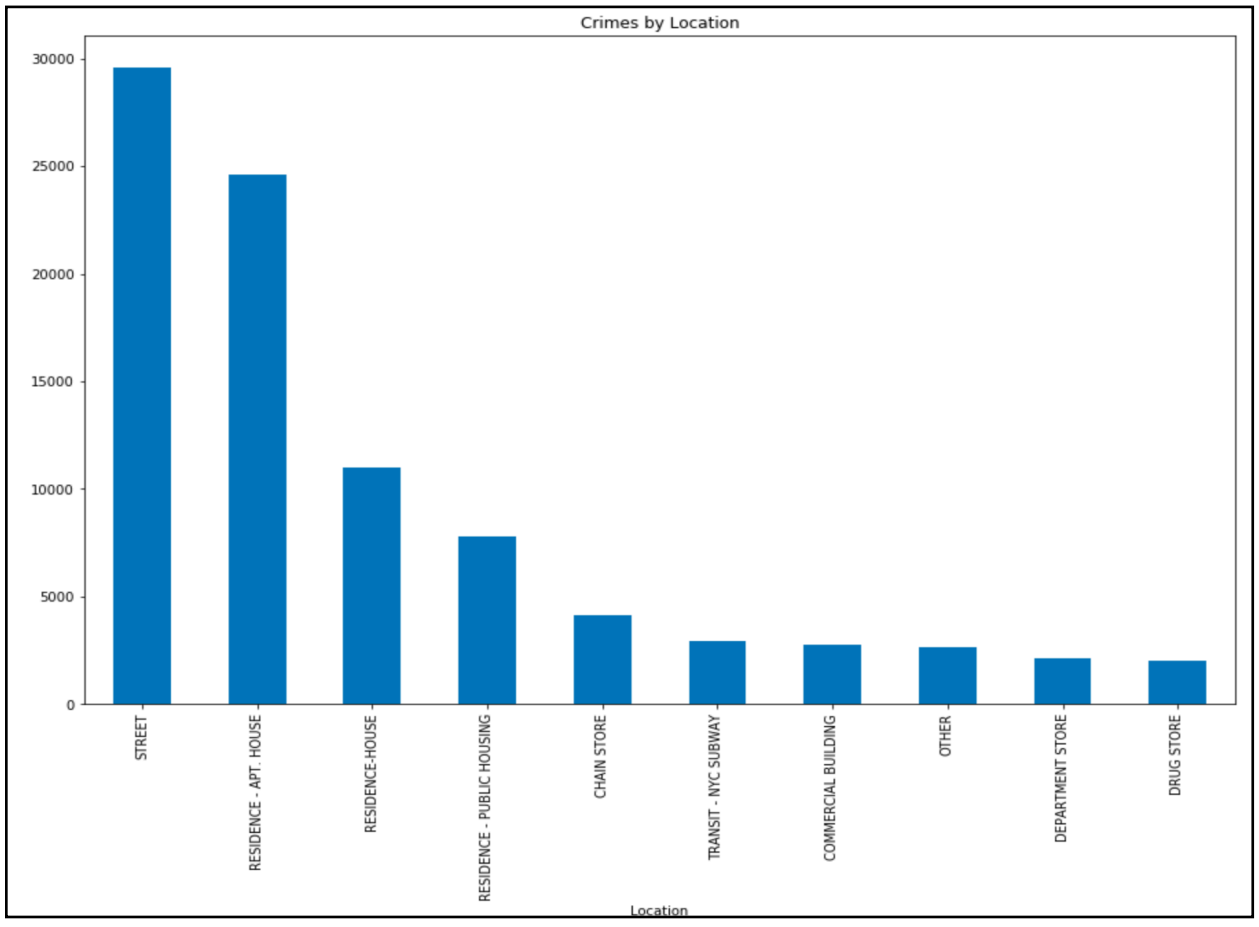
Analysis of NEW YORK CITY Crime Statistics

After cleaning up the data and gathering the right information, I used **MATPLOTLIB** to create a series of charts to pair down the relevant information in order to determine the highest prevalence of crime by Borough, Precinct, Offense, and Location.






Crimes by Offense
20000
15000
10000
5000
0
PETIT LARCENY
HARRASSMENT 2
CRIMINAL MISCHIEF & RELATED OF
ASSAULT 3 & RELATED OFFENSES
GRAND LARCENY
FELONY ASSAULT
OFF. AGNST PUB ORD SENSBLTY &
ROBBERY
MISCELLANEOUS PENAL LAW
DANGEROUS DRUGS
Offense

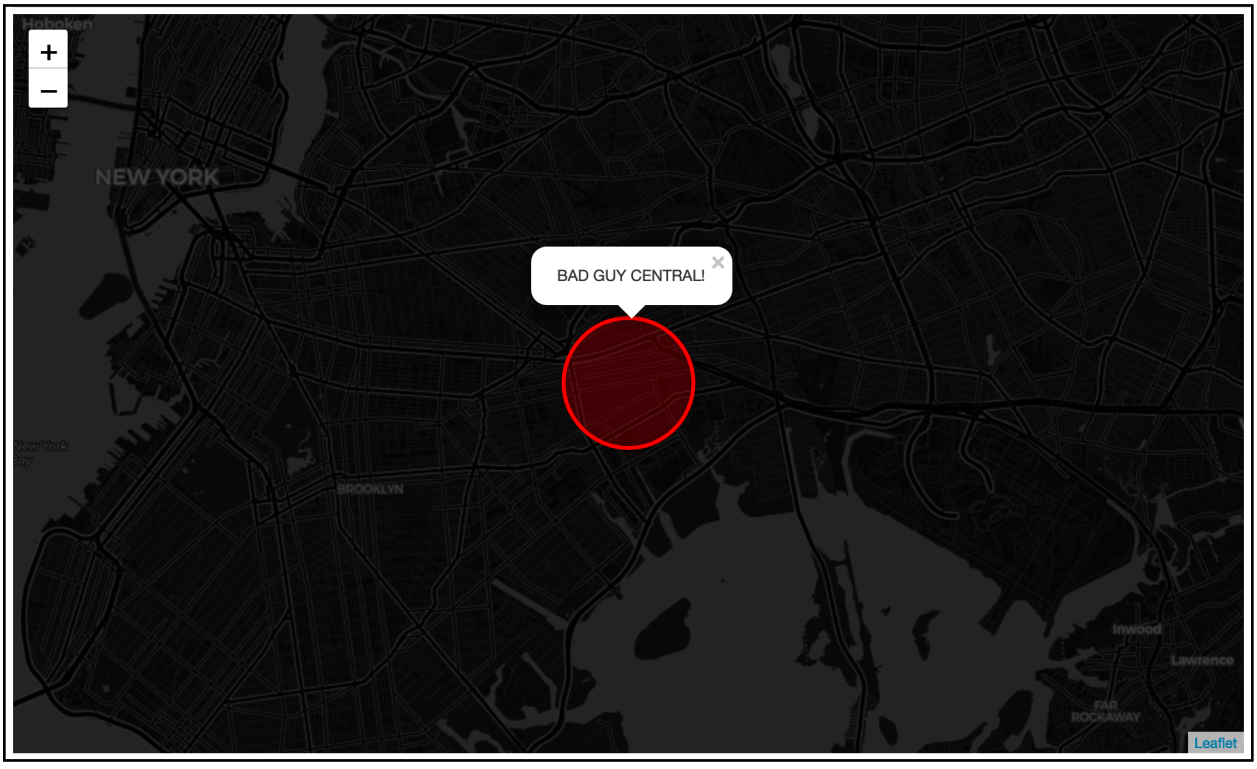

Crimes by Location
30000
25000
20000
15000
10000
5000
0
STREET
RESIDENCE - APT. HOUSE
RESIDENCE-HOUSE
RESIDENCE - PUBLIC HOUSING
CHAIN STORE
TRANSIT - NYC SUBWAY
COMMERCIAL BUILDING
OTHER
DEPARTMENT STORE
DRUG STORE
Location

Results: Highest Prevalence

Borough: Brooklyn
Precinct: 75th Precinct
Location: Streets
Offense: Petty Larceny

Building a Map

After determining the best course of action, I had to figure out where this place is on a map! I searched Google Maps for the exact address to the 75th Precinct and used geolocator to convert it to LAT/LONG. I set the geocoords based on the location of the 75th Precinct and used **Folium** to generate a map with a working radius in which to operate.



(Folium Map with Radius centered on 75th Precinct location in Brooklyn)

Discussion & Conclusions

Hard Choices

Discussion Presented to Ronnie Gulliver

Following **Mr. Gulliver's** instructions, I compiled and presented the information to him and tabled two options for his consideration:

1.) **Take Flight and Become the GULLMAN**: fighting **Petty Larceny on the Streets of Brooklyn's 75th Precinct** is the best place for **GULLMAN** to start his vigilante crime patrols and make a name for himself.

2.) **Remain Ronnie Gulliver**: life would remain the same and Brooklyn would continue to exist as a wretched hive of petty larceny. But **Mr. Gulliver** would be largely unaffected - he can continue running a multi-million dollar portfolio and driving his expensive car.

Conclusions and Final Thoughts

Only **Ronnie Gulliver** knows what **Ronnie Gulliver** wants. While the information was presented in a clear and concise manner, there are several variables that are left to examine. Given the initial request, I think this product satisfies all requirements, while leaving room to explore other options. If anything, the information shows me where I can improve as a data scientist. And I am ready for the challenge. **FLY! GULL! FLY!**