**NYC 2019 Crime Trends Analysis**

Our project will be focused on analyzing crime data in the five boroughs of NYC from 2019. Our hypothesis is that there is a positive correlation between crime rate and the distance from streetlights and wifi hotspots.  We believe the reason being that streetlights and wifi hotspots draw more foot traffic. In order to test our hypothesis, we have identified three datasets that provide us the details we need. The three datasets we used are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Dataset** | **Cleansing / Filtering** | **Datapoints Used** |
| 1 | **NYPD Crime Data**  <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i/data>  This data set includes NYPD crime reports for the 5 boroughs of NYC. | 1. Crimes in 2019 2. Drop all NA values for latitude and longitude 3. Filter by Borough 4. Adding Day/Night column 5. Renamed columns | 1. Complaint time 2. Complaint date 3. Crime category 4. Borough 5. Latitude 6. Longitude |
| 2 | **NYC Wifi Hotspots**  <https://data.cityofnewyork.us/City-Government/NYC-Wi-Fi-Hotspot-Locations/yjub-udmw/data>  Includes locations of publicly available Wifi across NYC | 1. Renamed columns 2. Filter by Borough | 1. Borough 2. Latitude 3. Longitude |
| 3 | **NYC Mobile Telecommunications**  [https://data.cityofnewyork.us/City-Government/Mobile-Telecommunications-Franchise Pole-Reservati/tbgj-tdd6/data](https://data.cityofnewyork.us/City-Government/Mobile-Telecommunications-Franchise%20Pole-Reservati/tbgj-tdd6/data)  Locations of street light poles, traffic light poles and utility poles reserved by companies authorized by the NYC Department of Information Technology and Telecommunications  Note: This dataset does not contain all streetlights | 1. Renamed columns 2. Filter by Borough 3. Filter Pole Class for “City” (identifies streetlights) | 1. Pole Class 2. Borough 3. Latitude 4. Longitude |

We broke down the datasets to the following:

1. By boroughs
2. Focused only on felonies
3. Day crimes in relation to wifi hotspots
4. Night crimes in relation to streetlights

We expect to see less crime in an area that experiences more wifi hotspots and has more streetlights than in areas without.  We will be comparing data and creating charts / graphs to illustrate our findings.

**Instructions to run .py file:**

1. Download all 3 final datasets
2. Install geoplotlib and pyglet (pip install geoplotlib, pip install pyglet)
3. Restart kernel
4. Run code

**Mapping / Plotting -** Our goal was to create a plot / heat map of the 5 boroughs, pinpointing locations of crime as well as the locations of each streetlight and wifi hotspot from our datasets.

1. Colors were assigned for each variable (Crime, Streetlight, Wifi)
2. 3 plots were combined to form one large map plot.
3. The **darkmatter** tool altered our map to black, creating a more welcoming backdrop for our multi-colored plot, which was drowned out on the original white.
4. The **alphaCrime** function adjusted the transparency of our map, wherein we were able to create more of a heat-map, which was especially important given the frequency of our red plot points.
5. The **geoplotlib** function allowed us to zoom in and out of our plots as needed, enabling us to focus on certain boroughs or more specific areas in a given borough.

**Calculations / Analysis -** In order to perform our analysis, we calculated the number of crimes that occurred in relation to the distance of streetlights and wifi hotspots. In order to make these calculations, we did the following:

1. Took each crime and calculated the distance of each streetlight and wifi
2. Kept the shortest distance of both for each crime
3. Created a dataframe that contains all crimes plus the distance of the closest streetlight and wifi for each
4. Filtered for felonies during the day in relation to distance from nearest wifi hotspot and filtered for felonies during the night in relation to distance from nearest streetlight
5. Narrowed the focus to a 1-block perimeter of streetlight / wifi (this is due to the need to narrow down the focused area because we did not have information for all the streetlights in NYC)
6. Used **matplotlib** to graph all results for each borough
7. Plotted the trendline to better determine if there is a positive, negative, or no correlation
8. Calculated the slope for each trend line to understand the rate of increase / decrease

**Findings / Conclusions:**

1. More crimes occur at night vs. during the day
2. Brooklyn and The Bronx have very similar crime rates
3. Manhattan has the highest rate of crime
4. Mostly very clear positive correlation between streetlights and crime rate
5. Mostly positive correlation between wifi and crime rate, although a lot weaker correlation than streetlights
6. Manhattan is the only borough where there is a negative correlation between streetlights / wifi and crime rate
7. Within a perimeter of approx 0.2 miles (~106 ft), there is a very strong correlation between streetlight / wifi and crime rate (for all boroughs)

In conclusion, these correlations support our hypothesis, which is, more crime would occur when further away from streetlights or wifi, possibly due to the presence of more people in those areas.