Analysis of NYC Crime Incidents in 2020 with Covid-19

Wes Wang, Kyle Ma

2021/10/13

1 Project Description

Our project aims to provide insights on the emerging trend of crime committed in New York, 2020, with COVID-19 as background. We seek to detect if there is correlation between crime frequency in New York and COVID-19. Our pre-analysis hypothesis is that there would be a positive correlation, because this unprecedented pandemic has caused massive public insecurity and deaths. We believe these unstable factors in this sensitive year may lead to more crimes and less order. In previous analysis, researchers ended up with mixed results. In Spatial Patterns of COVID-19 Incidence in Relation to Crime Rate Across London [1], it is claimed that most types of crimes including violence were not correlated with COVID at all, while in COVID and Crime: An Early Empirical Look [2], it is indicated that violence had a negative correlation with COVID-19. We plan to evaluate these results with our data in New York area.

Our analysis will be based on two datasets provided by NYC Open Data and Google, which will be detailed in later sections. We focus on both overall trend, i.e., how the count of all types of crimes react to the spread of COVID-19, as well as how the counts of crimes in each sub-category like Hate Crime react to it. Moreover, we will do correlation analysis in different time settings: days, weeks and months. In addition, we will also conduct analysis with the lagging mode we designed. In this mode, we match crime incidents with COVID-19 incidents/counts from several units of time ago. This is because a sharp increase of COVID-19 patients might not affect the current month. Instead, it might affect the crime rates in the next week/month. We hope these configurations generate a more thorough results and macro view of the subject. Lastly, the project will produce a measure of the correlation and a visualization of the trend.

Note that this project does not aim to do causal inferences on COVID-19 and crime rates. This is because there are too many confounders about this matter in various fields: politics, social media, medical, etc. We do not intend to draw conclusions on whether COVID-19 leads to an increase or decrease in crime rates. We plan to analyze the correlation between them.

2 Target Audience of This Project

The project would provide insight on the correlation between the crime rate in NYC and Covid-19 cases.
It is especially beneficial for:

- Good citizens to know when to raise awareness.
- Police department to be prepared for potential crime outburst.
- Scholars interested in the relationship between Covid cases and crime rate.

3 Potential Insights

- We would be able to answer our own hypothesis of whether there is a correlation.
- With the comprehensive data we have, we are able to analyze more closely – how each type of crimes interacts with the pandemic.
- Since our data on crime incidents also contain pre-COVID era, we will be able to see if there is difference in the overall trend before and after.
- We can also see that within COVID era, whether the trend is consistent from beginning to end, or if there is variation and anomalies along.

4 Assessment of Results & Analytics

We will compare the results to the previous papers, and see if the trend we produced is generally consistent with theirs. In addition, since we have multiple configurations in factors like types of crimes, time settings, etc., it is important to report all results respectively and cross validate with one another to see if they are consistent. Lastly, it is necessary to validate our results by inspecting the visualizations generated based on them.

5 Data Source

5.1 Data Source One: NYPD Complaint Data Historic (Wes Wang)

This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to 2020.

- Source: Police Department (NYPD)
- Link: https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i
- Schema:
  1. NYPD_Complaint_Data_Historic.csv
     - CMPLNT_NUM: Number Randomly generated persistent ID for each complaint
     - CMPLNT_FR_DT: Date Exact date of occurrence for the reported event
     - CMPLNT_FR_TM: Text Exact time of occurrence for the reported event
     - CMPLNT_TO_DT: Date Ending date of occurrence for the reported event, if exact time of occurrence is unknown
     - CMPLNT_TO_TM: Text Ending time of occurrence for the reported event, if exact time of occurrence is unknown
     - ADDR_PCT_CD: Number The precinct in which the incident occurred
     - RPT_DT: Date Date event was reported to police
     - KY_CD: Number Three digit offense classification code
     - OFNS_DESC: Text Description of offense corresponding with key code
     - PD_CD: Number Three digit internal classification code
     - PD_DESC: Text Description of internal classification corresponding with PD code
     - CRM_ATPT_CPTD_CD: Text Indicator of whether crime was successfully completed or attempted
     - LAW_CAT_CD: Text Level of offense
     - BORO_NM: Text The name of the borough in which the incident occurred
     - LOC_OF_OCCUR_DESC: Text Specific location of occurrence
     - PREM_TYP_DESC: Text Specific description of premises

- JURIS_DESC: Text Description of the jurisdiction code
- JURISDICTION_CODE: Number Jurisdiction responsible for incident.
- PARKS_NM: Text Name of NYC park, playground or greenspace of occurrence
- HADEVELOPT: Text Name of NYCHA housing development of occurrence,
- HOUSING_PSA: Text Development Level Code
- X_COORD_CD: Number X-coordinate for New York State Plane Coordinate System
- Y_COORD_CD: Number Y-coordinate for New York State Plane Coordinate System
- SUSP_AGE_GROUP: Number Suspect's Age Group
- SUSP_RACE: Text Suspect's Race Description
- SUSP_SEX: Text Suspect's Sex Description
- TRANSIT_DISTRICT: Number Transit district in which the offense occurred.
- Latitude: Number Midblock Latitude coordinate for Global Coordinate System
- Longitude: Number Midblock Longitude coordinate for Global Coordinate System
- Lat_Lon: Object Geospatial Location Point (latitude and Longitude combined)
- PATROL_BORO: Text The name of the patrol borough in which the incident occurred
- STATION_NAME: Text Transit station name
- VIC_AGE_GROUP: Text Victim's Age Group
- VIC_RACE: Text Victim's Race Description
- VIC_SEX: Text Victim's Sex Description

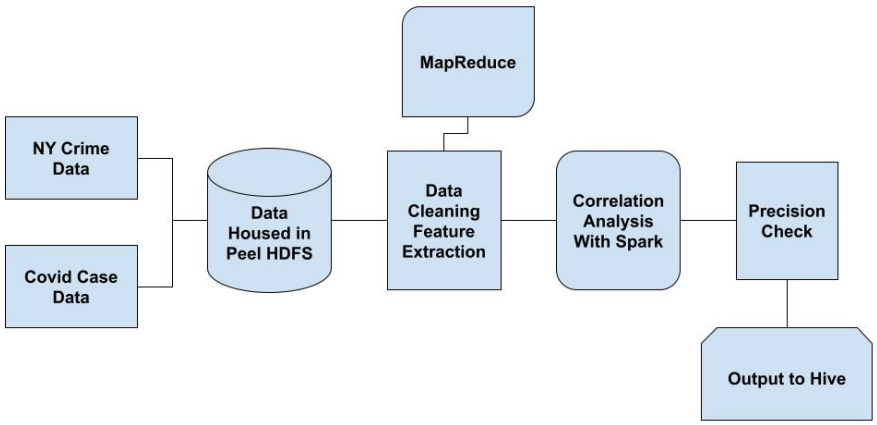
5.2 Data Source Two: Covid-19 Open Data Repository (Kyle Ma)

The COVID-19 Open Data Repository provides one of the largest aggregations of COVID-19 data available for technical users, with information uploaded daily from hundreds of sources.

- Source: Google
- Link: https://health.google.com/covid-19/open-data/raw-data
- Schema:

1. epidemiology.csv
   - date: string ISO 8601 date (YYYY-MM-DD) of the datapoint
   - key: string Unique string identifying the region
   - new_confirmed: integer Count of new cases confirmed after positive test on this date
   - new_deceased: integer Count of new deaths from a positive COVID-19 case on this date
   - new_recovered: integer Count of new recoveries from a positive COVID-19 case on this date
   - new_tested: integer Count of new COVID-19 tests performed on this date
   - cumulative_confirmed: integer Cumulative sum of cases confirmed after positive test to date
   - cumulative_deceased: integer Cumulative sum of deaths from a positive COVID-19 case to date
   - cumulative_recovered: integer Cumulative sum of recoveries from a positive COVID-19 case to date
   - cumulative_tested: integer Cumulative sum of COVID-19 tests performed to date

6 Initial Design Diagram



References

[1] Y. Sun, Y. Huang, K. Yuan, T. O. Chan, and Y. Wang, "Spatial patterns of covid-19 incidence in relation to crime rate across london," *ISPRS International Journal of Geo-Information*, vol. 10, no. 2, 2021. [Online]. Available: https://www.mdpi.com/2220-9964/10/2/53

[2] D. S. Abrams, "Covid and crime: An early empirical look," *Journal of Public Economics*, vol. 194, p. 104344, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0047272720302085