Data Cleaning and Integration

File Topic	Name of File	Source
Number of total Crime	NYPD_Complaint_Data_Historic.csv (file too big to upload)	https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i
Number of hate crimes	NYPD_Hate_Crimes.csv	https://data.cityofnewyork.us/Public-Safety/NYPD-Hate-Crimes/bqiq-cu78
Coronavirus Daily Cases	daily_cases.scv	https://github.com/nychealth/coronavirus-data/blob/master/trends/tests.csv
Total population change		https://en.wikipedia.org/wiki/Demographic_history_of_New_York_City https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork/PST045219
2017, 2018 hate crime numbers	hate-crime-complaints-by-motivation-annual-2017.xlsx hate-crime-complaints-by-motivation-annual-2018.xlsx	https://www1.nyc.gov/site/nypd/stats/reports-analysis/hate-crimes-archive-2017.page https://www1.nyc.gov/site/nypd/stats/reports-analysis/hate-crimes-archive-2018.page
NYC COVID-19 Timeline		https://www.investopedia.com/historical-timeline-of-covid-19-in-new-york-city-5071986

- All the files with file names are in "datasets/crimes".

The file NYPD_Complaint_Data_Historic.csv is too big to upload (2GB). We used PySpark to analyze the data to deal with the large size.

We had mainly used NYPD_Hate_Crimes.csv to analyze hate crime in NYC. It had rows with duplicate "Full Complaint ID" and "Arrest ID". We presumed as below:

- no "Arrest ID" → no arrest
- same "Full Complaint ID", different "Arrest ID" → different arrests for different people in same crime incident
- different "Full Complaint ID", same "Arrest ID" → multiple complaint calls for the same incident

However, this information was insignificant for us, so we decided to delete the duplicate rows. Initially, we tried using OpenRefine to cleanse the data, but for some reason, dealing with duplicates did not work. That is why we implemented data cleansing code in Juypter instead. However, we managed to cleanse the data in OpenRefine, so the data is cleansed but has extra code in Juypter to keep data clean.

Also, there were two rows with cases with "Complaint Year Number" being 2021 but "Month Number" being 12, which means December. I had changed the "Month Number" to 1, according to "Record Create Date". The increase in number could have been due to an actual increase in the crime, or an increase in awareness of Asian hate crime due to this year's "Stop Asian Hate" movement.

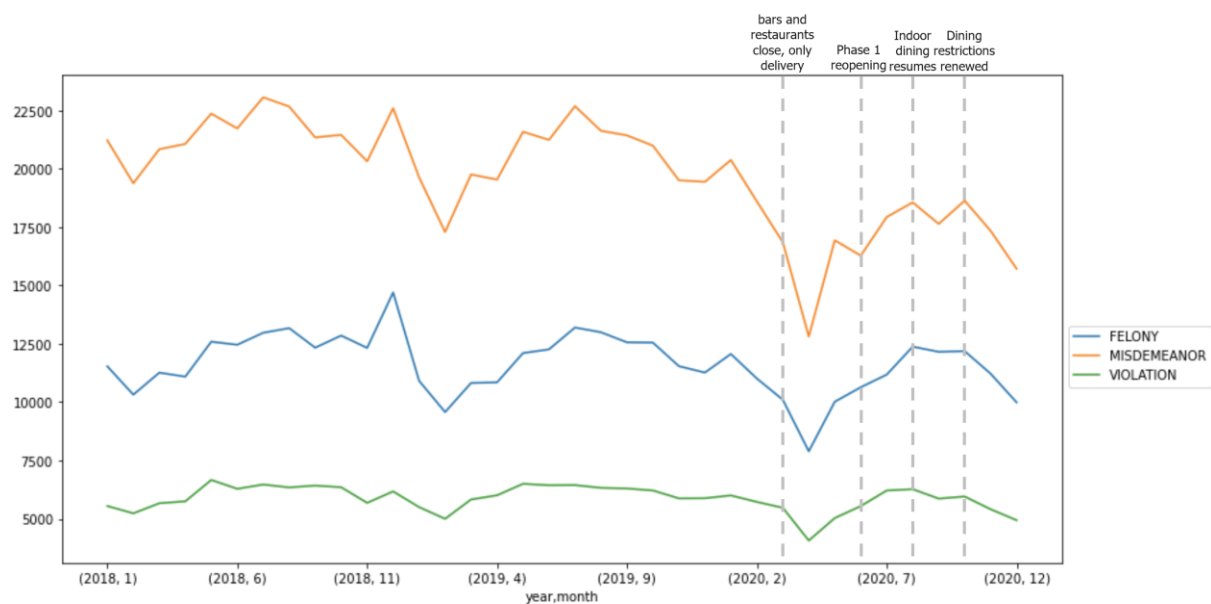**Data Analysis and Findings**

Crime Overall

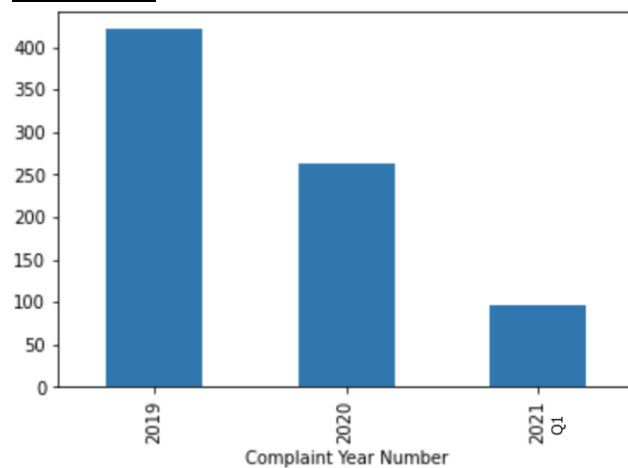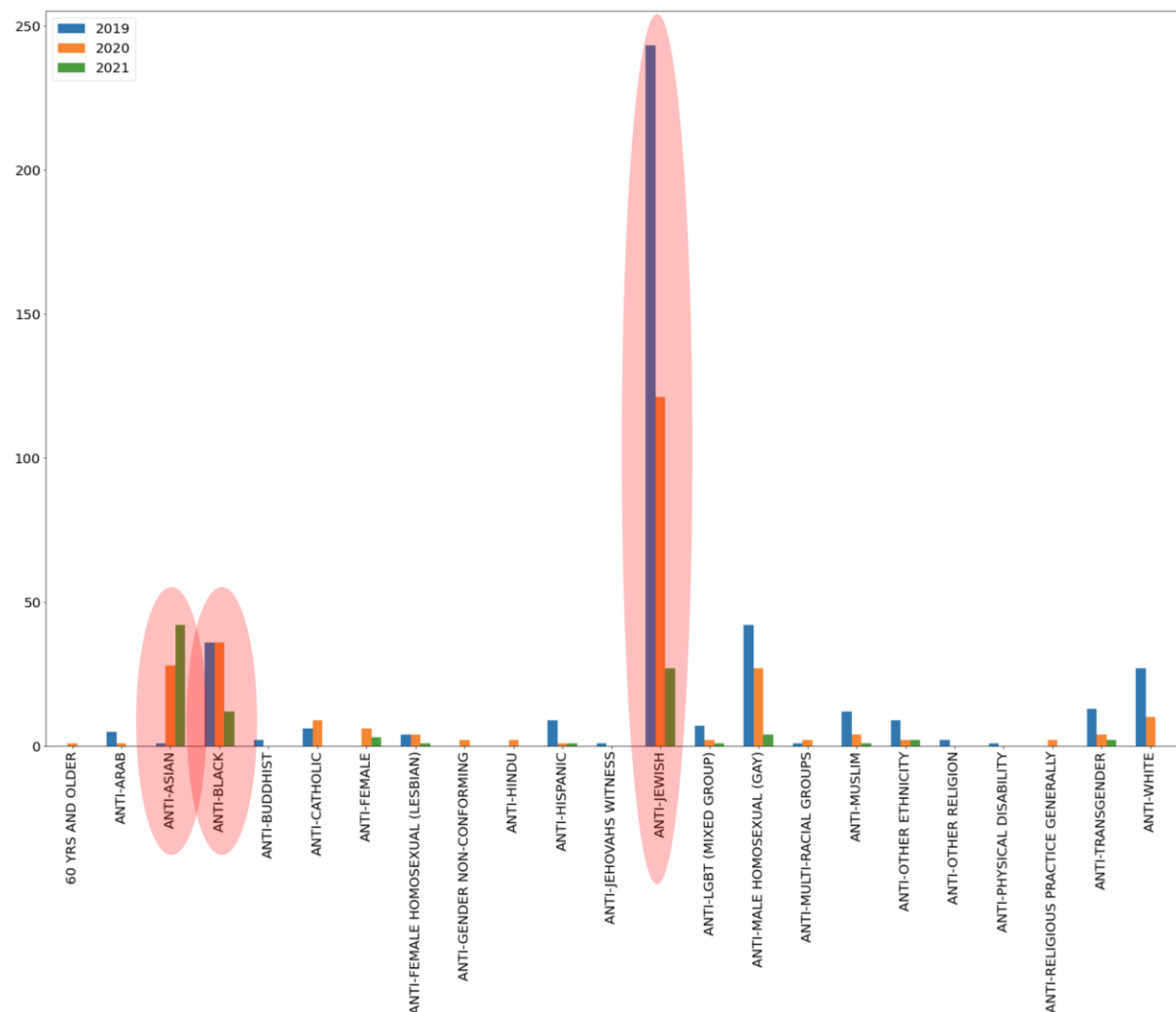|	cnt
year	
2018	478119
2019	457543
2020	404892

The overall crime rate shows that COVID-19 has indeed had a benign impact on the crime rate in New York City. The total number of crimes in NYC decreased during 2020 dropping from 46K cases to 40K (11.5% decrease). From the start of the pandemic in March 2020 to the reopening in June 2020, the crime rate was below what NYC has experienced for the last two years. Especially in April 2020, it was a record drop in the number of crimes. From June to November, the number steadily grew, and as NYC again restricted restaurants and public schools, the numbers dwindled. This trend did not differ among felony, misdemeanor, and violation. This may indicate that NYC's restriction in the movement of its citizens due to COVID-19, has had a considerable impact on the number of crimes that occur.

Hate Crime

The most interesting impact of COVID-19 was on hate crime numbers. Overall, The Number of hate crimes dropped from 421 to 264 cases (38% decrease).

However, the numbers differed among what kind of hate crime it was. There was a noticeable, almost a half decrease in the number of "Anti-Jewish" crimes. Among the decreased number of total hate crime cases from 2019 to 2020 which is 160 cases, 121 cases were "Anti-Jewish", so it can be inferred that the decrease in hate crime was mainly due to decrease in "Anti-Jewish" crime.

The number of "Anti-Black" crimes remained the same. Though the city had experienced a wide movement of "Black-Lives-Matter", it does not seem to have contributed to a decrease in crimes against the Black people. Since it is expected that the Black population decreased 3% since 2010, the no-change in the number could actually mean an increase in the number per Black person. It could indicate the BLM movement has raised the racial conflict to the surface, or it could also be a rationale as to why we need movements like BLM.

The most apparent change in the number would be the "Anti-Asian" hate crime number. While in 2019, there was only 1 case, in 2020, it grew to 28 cases, and just in the first quarter of 2021, it had reached 42, the highest number of hate crimes in the city. The number in 2017 and 2018 is 6 and 5 respectively, so it is a clear indication that the year 2020 had a major impact on "Anti-Asian" crime.

The Asian population is expected to have grown 12.6% since 2010, so it is not sufficient to say that "Anti-Asian" crime increased due to an increase in the Asian population in New York City. Also, since the overall crime number has decreased in 2020, making the city a "safer" place in terms of crime rate, it was not the case for Asian hate crimes. A surge in the number in March 2020 reflects that COVID-19 could be the most significant factor in the increase in Asian hate crime. It is alarming to see a surge in the criminal cases as NYC starts to reopen with vaccinations, a possible indication that "Anti-Asian" crime could become the new norm. The increase in number could have been due to an actual increase in the crime, or an increase in awareness of Asian hate crime due to this year's "Stop Asian Hate" movement.