

Kimball Lifecycle Proposal

CIS 9440-UWA

Final Project Milestone 1

Group Number - 19

Students: Hanqing Chen (hanqing.chen@baruchmail.cuny.edu)

Xinyue Chen (xinyue.chen@baruchmail.cuny.edu)

Candice Lee (candice.lee@baruchmail.cuny.edu)

Data Warehouse Project Title:

NYC Crimes and COVID-19 Cases Data Warehouse

Motivation for Project idea:

The U.S is seeing a massive spike in hate crime since the COVID-19 pandemic. There appears to be more violent attacks against certain minority groups recently, making these types of groups easier targets and more vulnerable to hate crimes. As students and residents in New York City, we are eager to help change the situation and stop the violence.

Description of the issues or opportunities the project will address:

The COVID-19 has been spreading in the United States for almost a year, and many assumed Asian communities (and many other groups, including women and elderly people) have ever since been facing a surge of hate crimes targeting them. With categories of different crimes committed, our project seeks to analyze datasets extracted from public sources to verify whether these types of assumptions are true or not and how crime rate is impacted by COVID-19. If true, by using our analysis we hope to push local authorities and volunteer groups to better utilize their resources to protect our residents.

Business Justification:

High-level Business Initiative:

Figure out the correlation between crimes and COVID-19 cases, and try to reduce crime in the future.

BI Sponsors and Stakeholders (who will own this project?)

Government departments such as the New York Police Department.

What's the Business Value?

Reduce the crime rates and provide a more secure living environment for residents. It will be beneficial in improving local business and property prices.

How long will this take? How much will this cost?

It will take 6 data engineers 3 months, and cost around \$800,000.

Technical Justification:

Which data sources do we already have for this project?

Dataset 1: New York State Statewide COVID-19 Testing_NY state

<https://health.data.ny.gov/Health/New-York-State-Statewide-COVID-19-Testing/xdss-u53e>

Dataset 2: NYC Complaint Data__NYC Open Data (Recent)

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

Dataset 3: NYC Complaint Data__NYC Open Data (Historic)

<https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>

What new data sources do we need (if any)?

Dataset 1: NYC Counties_NYC 311

<https://portal.311.nyc.gov/article/?kanumber=KA-02877>

Is the data we have conformed, consistent, and current? (data quality)

Yes, the data is conformed, consistent, and current.

What technical skills will we need to complete this project?

ETL creation

Data modeling

Dimensional modeling

Business Intelligence practices

Data engineering

Data warehousing reporting

Will we need any new types of technologies?

Big Data technologies

Cloud data warehousing

Key Performance Indicators (KPI's) your Data Warehouse will display:

1. Pre-pandemic crime rate vs Crime rate in pandemic era
2. Crime rate by area
3. COVID-19 cases rate by area
4. Victims grouped by ethnicity
5. Victims grouped by gender
6. Victims grouped by age
7. Proportion changes of the crimes among different categories
8. Percentage change in the crime rate across different categories
9. Crime rate grouped by different time of the day