**Machine Learning II - Project Proposal - Predicting Crime in LA & NY**

Team 20 - Bukrinsky, Jaju, Kalha, Mei, Singh

**Business Problem**

The crime rate in the United States significantly impacts people's lives, the stability of society, and the amount of money spent on law enforcement and criminal justice (DOJ, n.d.). High crime rates can lead to increased fear and anxiety among individuals and insecurity and mistrust in their communities (Wike et al., 2008). The cost of policing, prosecuting, and incarcerating criminals can also significantly burden taxpayers, taking resources away from other important areas such as education and healthcare (Collins, 2020). Additionally, high crime rates can lead to a lack of economic investment and opportunities, as businesses may be hesitant to invest in areas with high crime rates (Pimental, 2022). According to the University of Chicago, the estimated annual cost of crime is up to $5.76 trillion. Thus, addressing the root causes of crime and predicting the type and number of crimes in certain cities allow relevant parties to issue the proper task forces and prevent crime rates from increasing.

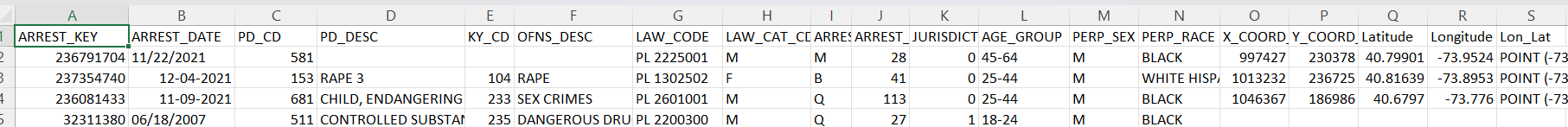
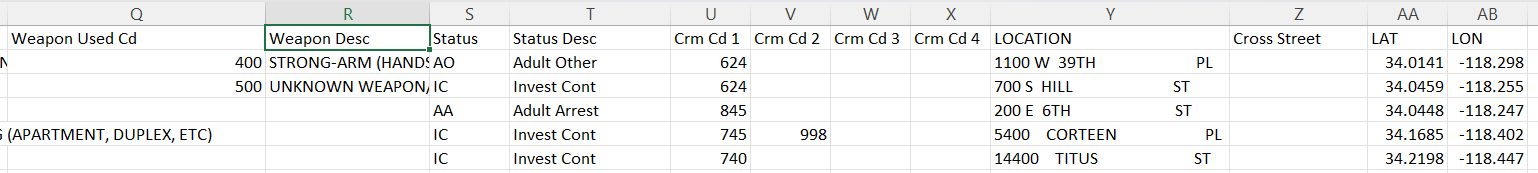
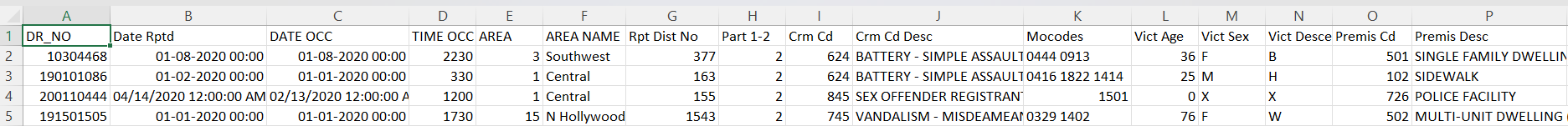
**Managerial Application and Decision**

In this project, by analyzing the crime data in New York City and Los Angeles, along with data on the city’s education, parks, and demographics, we seek to examine and draw insights pertaining to (1) predicting the number of crimes, (2) the factors that have the most impact on the number of crimes, and if possible (3) predicting the types of crimes more prevalent in each city. We hope to establish models that predict the number and types of crimes that may happen based on certain attributes and factors. The possible implications of this analysis could significantly impact local governments and police departments to make more informed decisions on behalf of their communities. For example:

*Police Departments:* As commissioners and departments work to understand how to serve their communities best (Norton, 2022), police departments and communities can utilize the results of our analysis to implement relevant task forces and personnel requirements and assign/request a budget for equipment and number of services.

*Local Governments:* Local governments are responsible for budgeting and advocating for their community and can use the analysis results to understand what factors cause crime and invest in prevention.

**Data Source and Sample Data**

In this project, the data is obtained from the United States government's open data websites for New York City and Los Angeles. We have focused on the “NYPD Arrests Data (Historic)” dataset and the “LA-Crime data from 2020 to present” dataset to make our predictions.  
The first dataset, the NYPD Arrests Data (Historic), provides comprehensive records (6 million rows) of arrests made by the New York City Police Department (NYPD) over a period of 15 years. This also includes information on the specific crimes and crime categories with demographic information about perpetrators.  
  
  
The second dataset, the Crime Data from the Los Angeles Police Department, contains 2 million rows of similar information on crime incidents reported to the LAPD.

Together, these two datasets provide a wealth of information that can be used to explore similarities and differences in crime patterns and trends across different regions and investigate how social and demographic factors influence criminal behavior and the criminal justice system. Once the data is grouped areawise, we will combine the data with the US census and health information from the city Open Data website to gain additional insights.

We will use various modeling techniques to carry out our predictions, and performance evaluation will be conducted on these models to ensure the best model is selected.

**References**

<https://cops.usdoj.gov/RIC/Publications/cops-w0669-pub.pdf>

<https://www.pewresearch.org/global/2008/04/15/where-trust-is-high-crime-and-corruption-are-low/>

<https://www.vox.com/the-highlight/21430892/defund-the-police-funding-abolish-george-floyd-breonna-taylor-daniel-prude>

<https://scholars.unh.edu/cgi/viewcontent.cgi?article=1676&context=honors>

<https://www.journals.uchicago.edu/doi/full/10.1086/715713#:~:text=The%20estimated%20annual%20cost%20of,%243.92%20trillion%20net%20of%20transfers>.

<https://norton.house.gov/media-center/press-releases/norton-introduces-bill-to-create-local-task-forces-on-policing-0>

<https://catalog.data.gov/dataset/nypd-arrests-data-historic>

<https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

<https://data.lacity.org/>

<https://opendata.cityofnewyork.us/>