**Project 4 Proposal**

**Research Question(s)**

1. What is the structure of crime in Los Angeles?
   1. Where are crimes most likely to occur?
   2. What kinds of crimes are most common?
   3. How frequent are arrests made for crimes?
   4. What demographics are most frequently victimized?
   5. In what kinds of locations (e.g., homes, public spaces, stores, etc.) are crimes most likely to occur?
2. Are there any relationships between victim demographics and crime type?
3. Are there any relationships between victim demographics and arrest likelihood?
4. Can we accurately predict whether arrests are likely to occur or not from the available data?

**Data Set**

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

Data comes from a data set a Los Angeles data file from data.gov.

**Data Cleaning and Transformation**

Some preliminary cleaning will be done within the excel file. The remainder of the data will be cleaned in Jupyter. The data will be exported as a cleaned excel file for use in subsequent analyses.

**Tools**

Python/Jupyter Notebooks

Pandas

Scikit-learn

Tableau

**Analytics Plan**

We plan to develop the following:

* Descriptive visualizations in Tableau covering the following:
  + Bar graph of crime types
  + Pie chart of arrest rates
  + Bars graph showing demographic breakdown of victims by age, race, and sex
* A map in Tableau showing the locations of crimes, with color and size coded for arrest made and type of crime and filterable by demographics of victim.
* Inferential analysis examining factors that correlate to victim demographics
* A supervised learning model to predict the likelihood of an arrest.

**Timetable**

Finish Data Cleaning (Dan)- By end of class on February 5

Create All Tableau Visualizations (Fesseha)- By end of class on February 7

Complete Inferential Analyses (Dan)- By end of class on February 7

Make Tableau Map and Finalize Tableau Dashboards (Fesseha)- By end of day on February 11

Complete Supervised Learning Model (Dan)- By end of day on February 11

Draw Conclusions from Analysis (Both)- By beginning of class on February 12

Plan Presentation- A Tableau Story (Both)- By end of class on February 12

Write Up Report (Dan)- By February 13