Project Proposal

Kadin Wilkins, Matilda Orona, Vikram Magal, Anushka Vazirani
kwilkins@berkeley.edu, matilda_orona@berkeley.edu, vikram_magal@berkeley.edu, avazirani@berkeley.edu

Motivation:

Through this project, we want to answer whether we can predict levels of violent and non-violent crime based on location and time. This is interesting because it could help differently inform law enforcement and allow them to more appropriately respond to crime. Understanding patterns of crime could provide valuable insights for resource allocation as well, promoting the use of community policing for non-violent crimes.

Data:

Our dataset is published by LAPD and contains crime data for the city of Los Angeles from 2020 to present and is updated on a biweekly basis.
Link to dataset: https://catalog.data.gov/dataset/crime-data-from-2020-to-present
Each crime resides within a row containing data on the time of the crime including the date and hour-minute of the crime, the location of the crime by cross-street, lat/lon, and area name. Each crime also includes data pertaining to whether a weapon is used in the crime, and if so, what weapon was used in each case. In addition, this dataset gives us access to the crime code, and their descriptions. Our dataset contains 1,004,991 rows and 28 columns.

We also plan to incorporate population statistics and socioeconomic information from US Census data. Census data is available at various levels of granularity - we will likely use it at the zip code level to join to our crime dataset and aggregate afterwards to the geographic level we choose to predict crime statistics at.
Link to US Census data: https://data.census.gov/all

Related work:

- Predicting the Probability of Crime Related Danger in Los Angeles
  This paper attempts to utilize multiple machine learning algorithms to accurately estimate the danger of crime by area within Los Angeles.
- Prevention Is better than cure: Predicting violent crime in US counties using machine-learning methods
  This project evaluated multiple ml models including Linear Regression, KNN, and Gradient Boosting(XGB) on data from over 2100 US counties.
- Using Machine Learning Algorithms to Analyze Crime Data
  This research focuses on predicting violent crime specifically in the state of Mississippi.

Methodology:

We plan to apply several algorithms ranging in complexity. We'll employ linear regression, random forest, and neural networks to predict the number of crimes occurring in each area in a specific time period. With our data, it may be possible to specify our predictions from as broadly as months, to as granular as hours.

Project GitHub repo: https://github.com/vkrmgl/DS207_Final_Project