Modeling Crime in Colorado

Flatiron School

Capstone Project

Authors: Elena Kazakova

Cohort: DS02222021

Instructor: James Irving

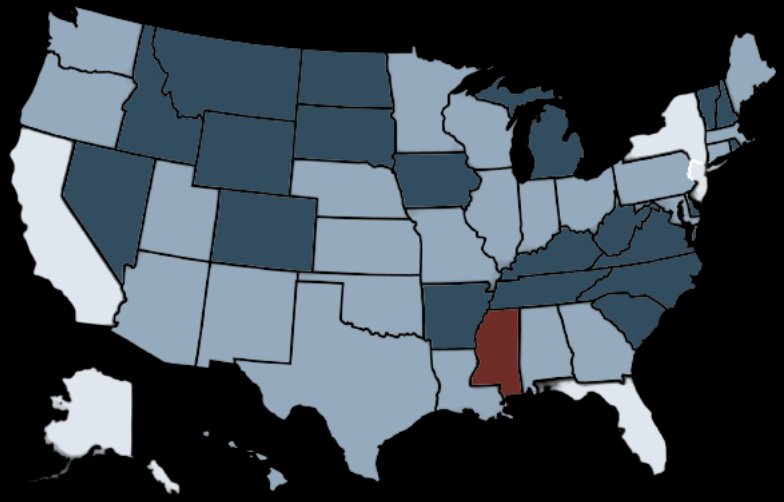
Project Goals

- Modeling crime rates-foundation for shaping:
  - **resources allocation for law enforcement**
  - **public policy**
  - **preventive measures**
- General and categorical crime rates:
  - **provide transparency**
  - **create easier access**
  - **expand awareness**
- Web application:
  - **data available through visualizations and statistics**

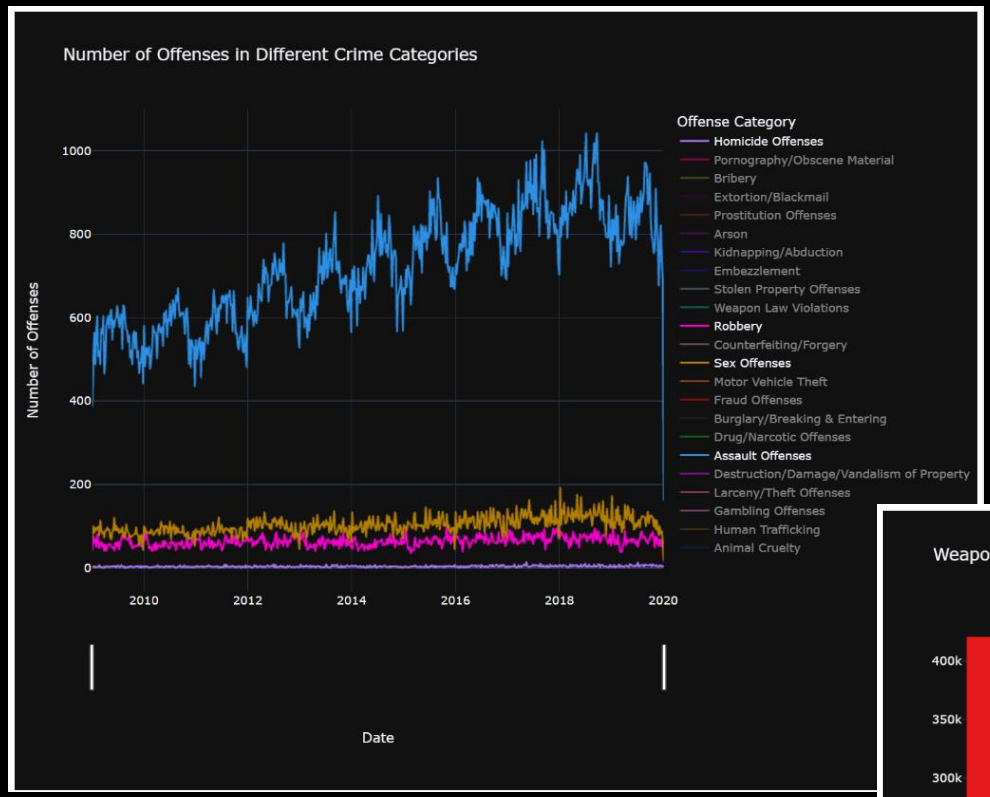
Data

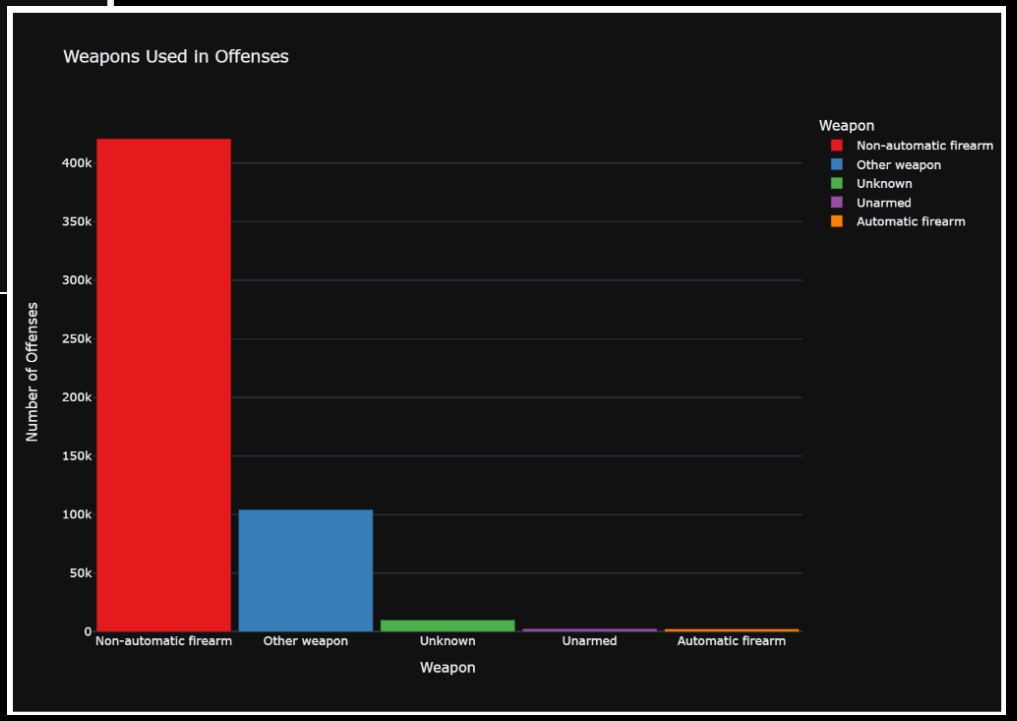
- 44 tables available for download
- API available (but no current data)
- Data:
  - Offense information
    - ✓ **Date/Time**
    - ✓ **Category**
    - ✓ **Weapon**
    - ✓ **Bias**
  - Reporting agencies information
    - ✓ **Geographic data**
  - Victim demographic data
  - Offender demographic data

FBI
Crime Data Explorer
(2009-2019)

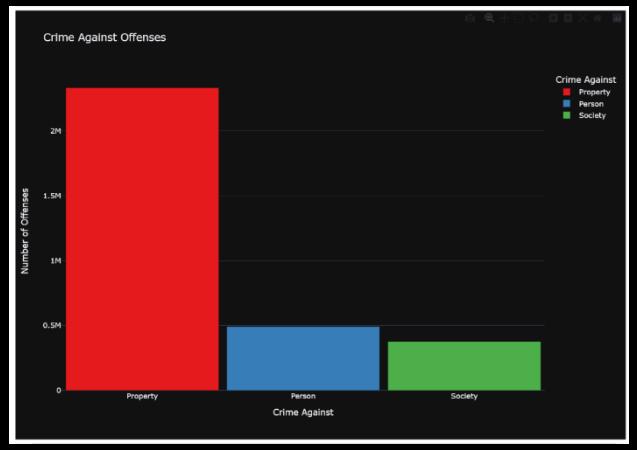


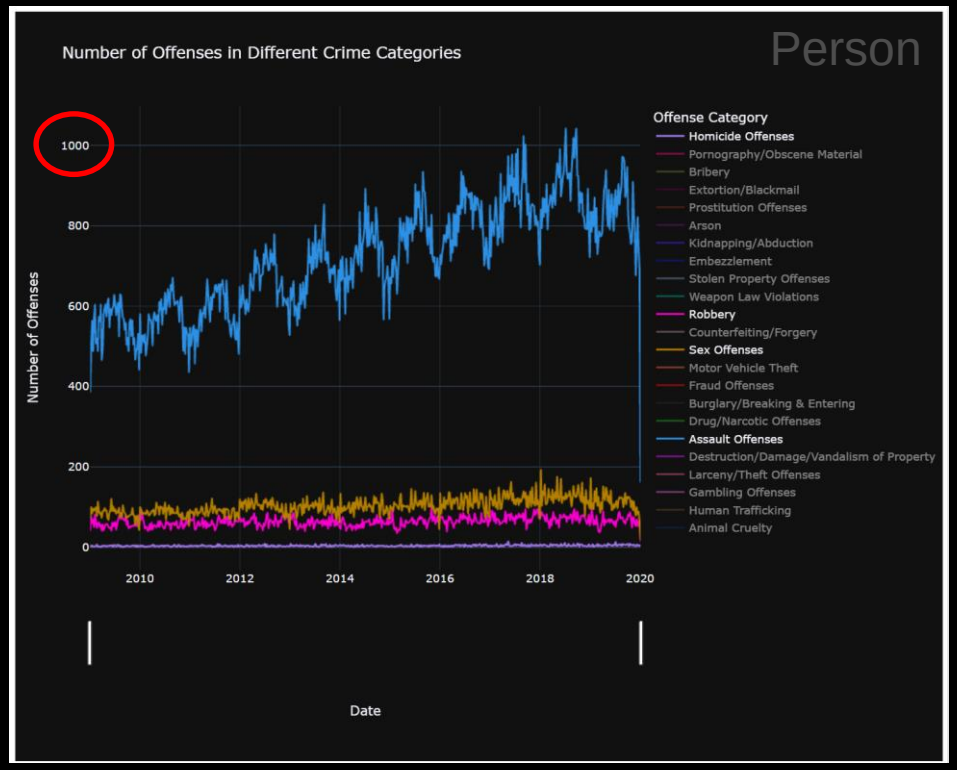
Violent Crimes

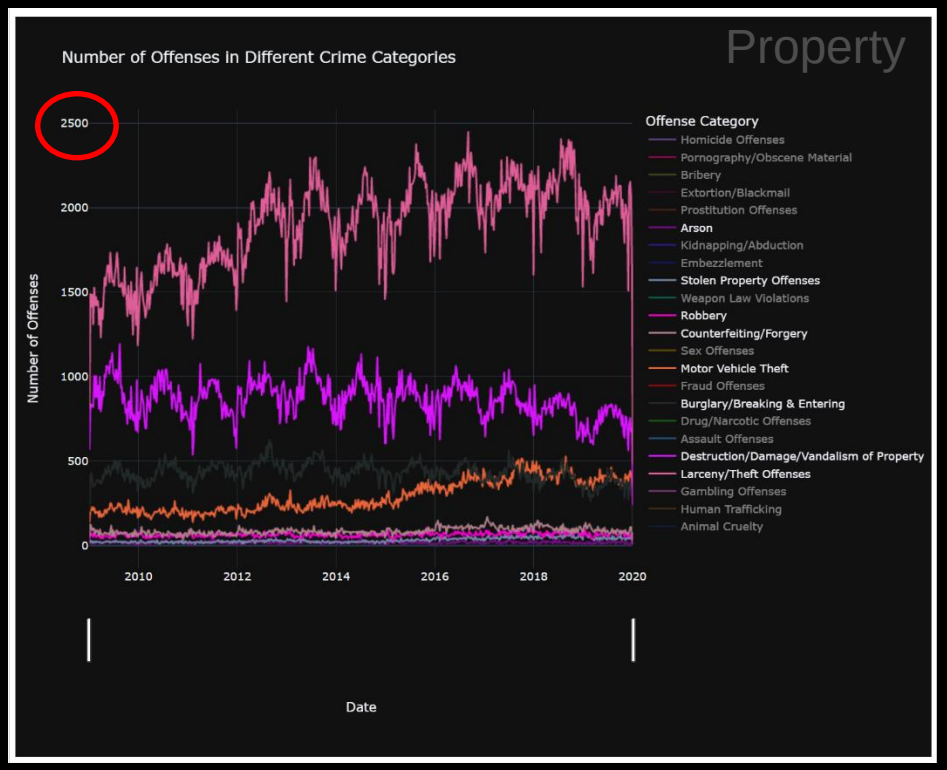




Crime Against Categories





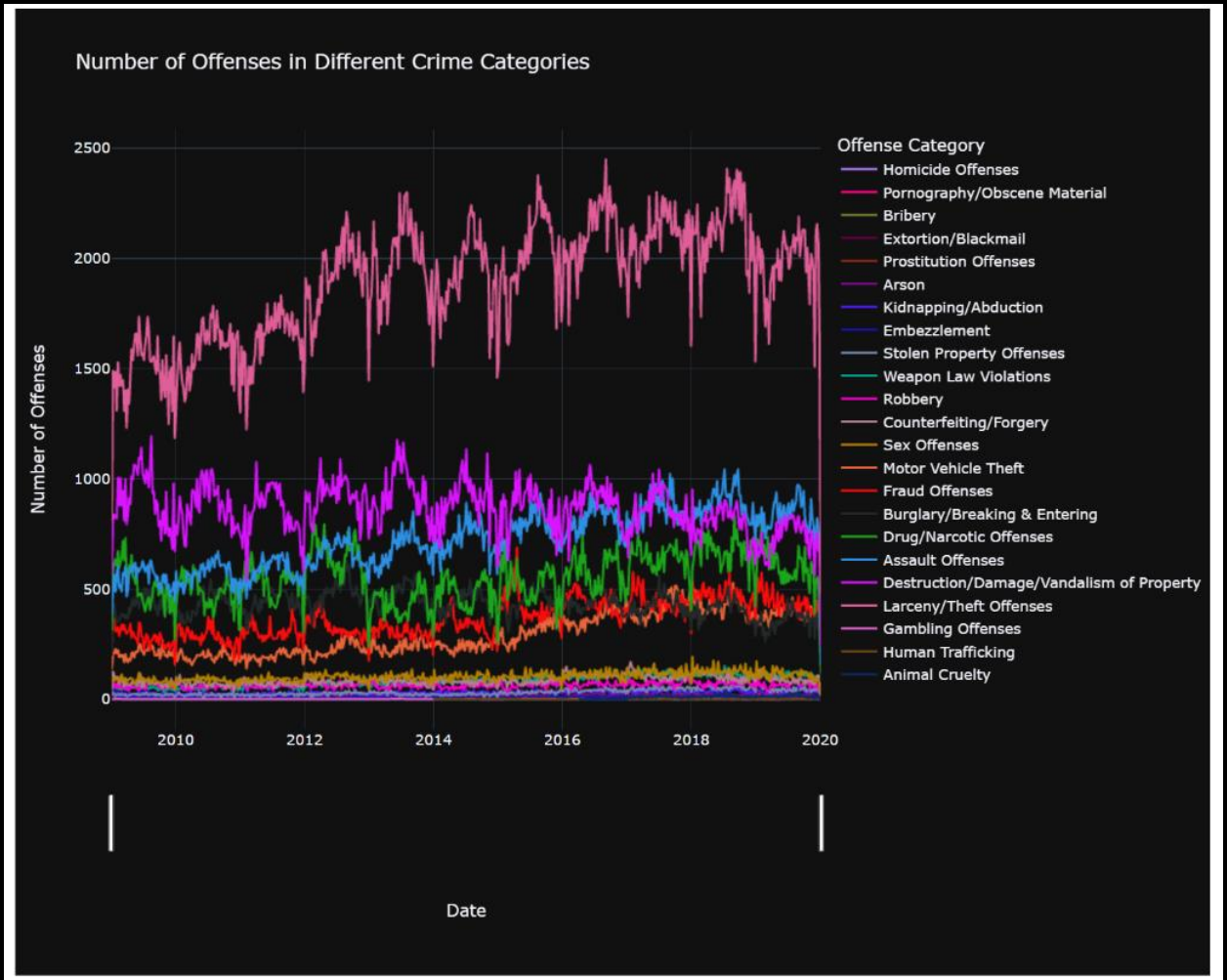


Categories of Offenses

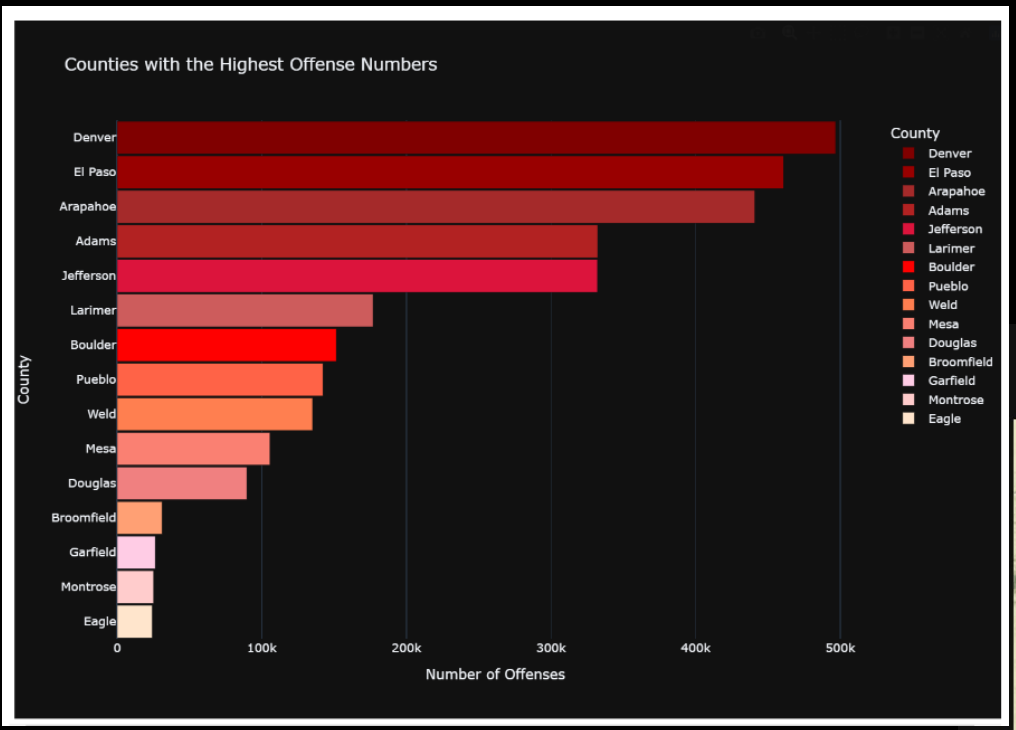
Offenses
3201143

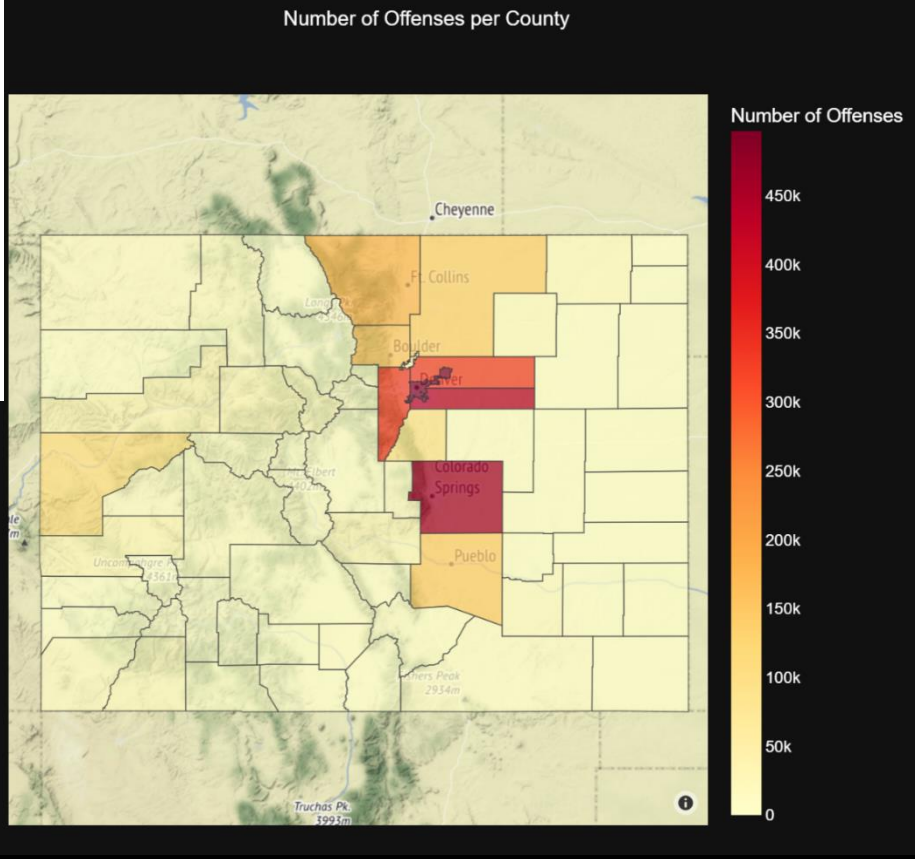
Victims
3229640

Offenders
3197991



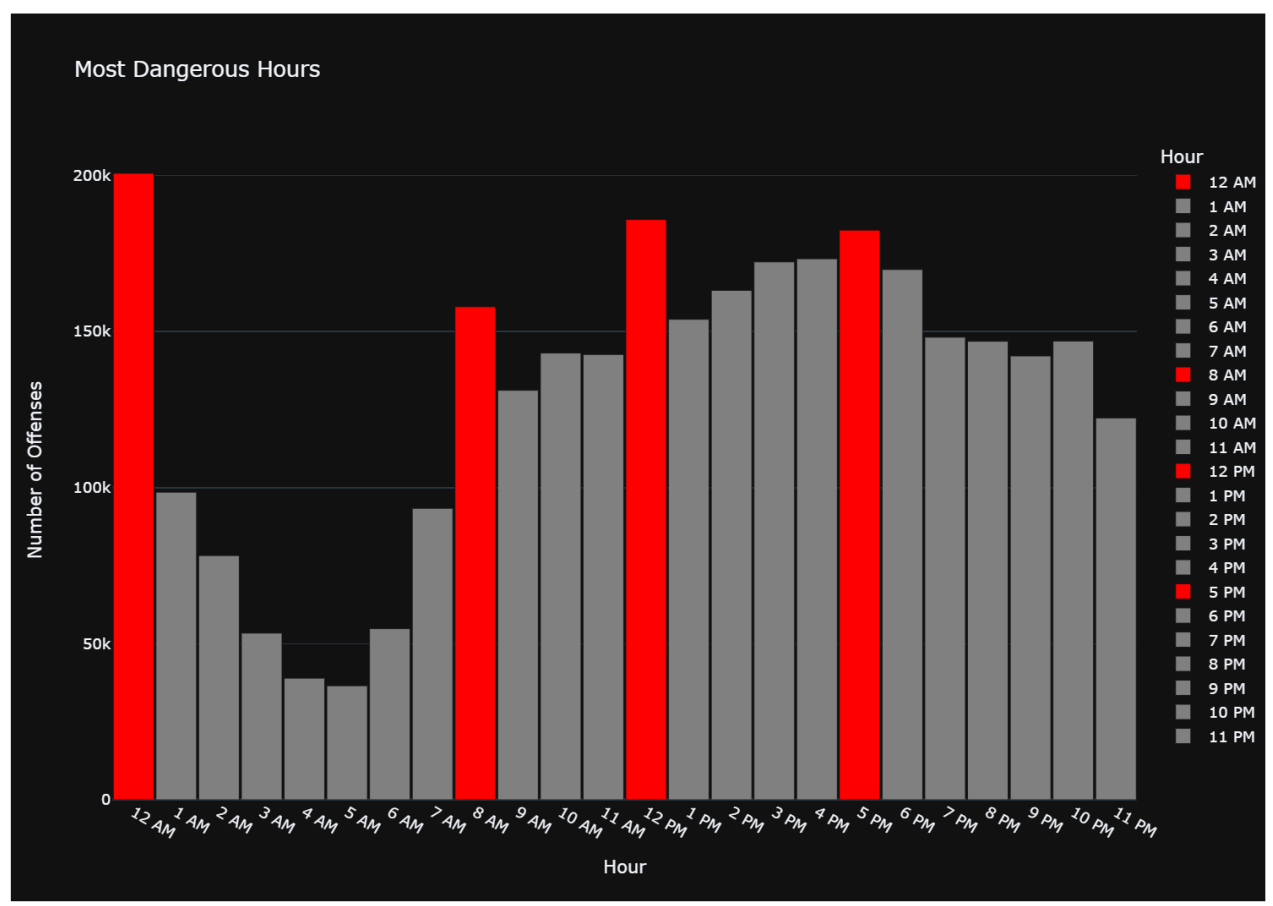
Geography of Crimes





Crime Timing

- ✓ Midnight
- ✓ 8 AM
- ✓ 12 PM
- ✓ 5 PM

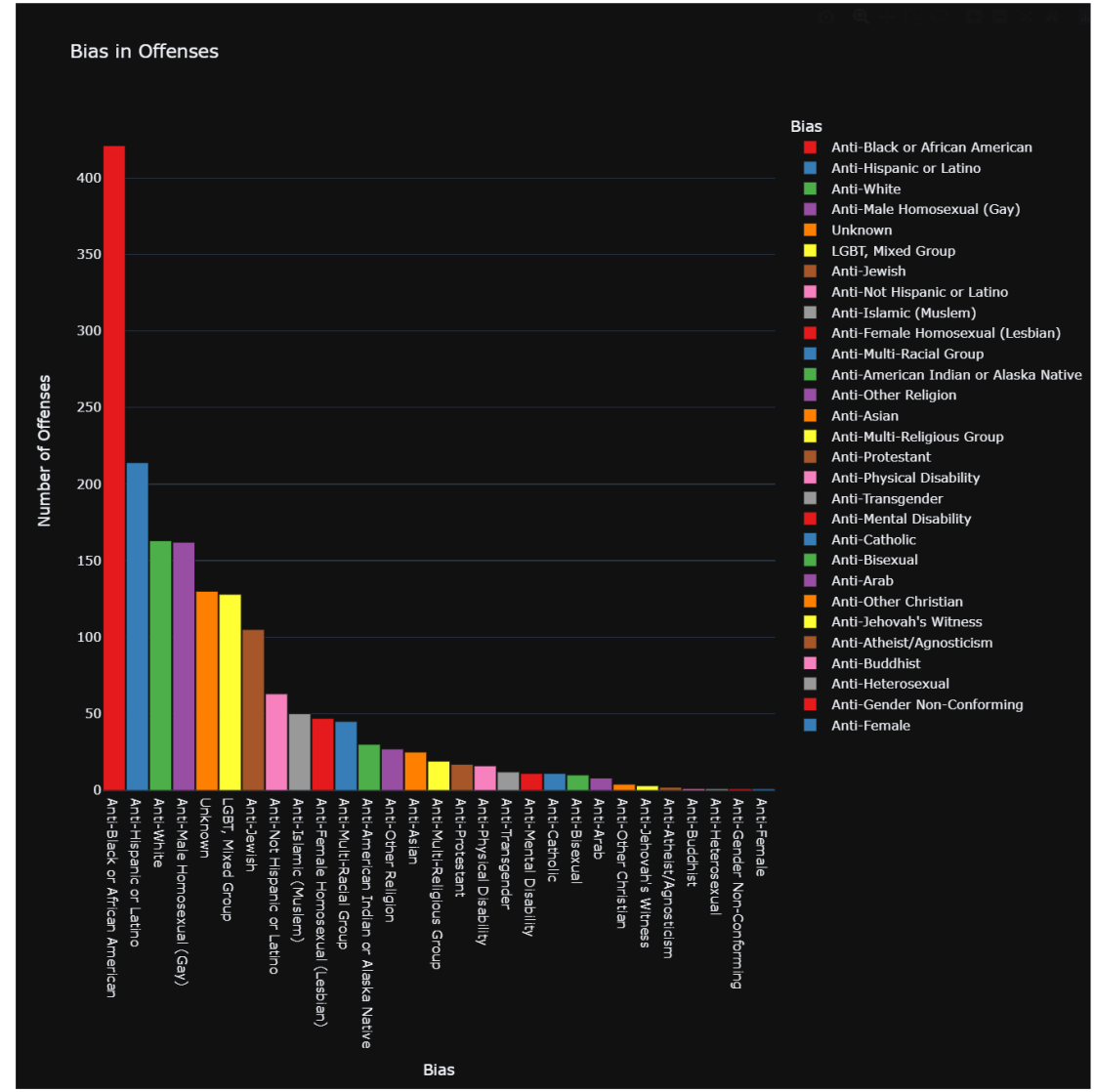


Conclusions (general)

1. The project demonstrated that the actual crime data could be analyzed and modeled with significant accuracy
2. Generated models forecast general crime rates and categories crime rates
3. The Exploratory Data Analysis identified:
   - **Geographic areas with higher crime rates can help plan for law enforcement resources and preventive programs**
   - **Crimes' targets and weapons used**
   - **Offenders' and victims' demographics**
4. The dataset, results of EDA, and the models are the base for a web-based dashboard built with the dash python package

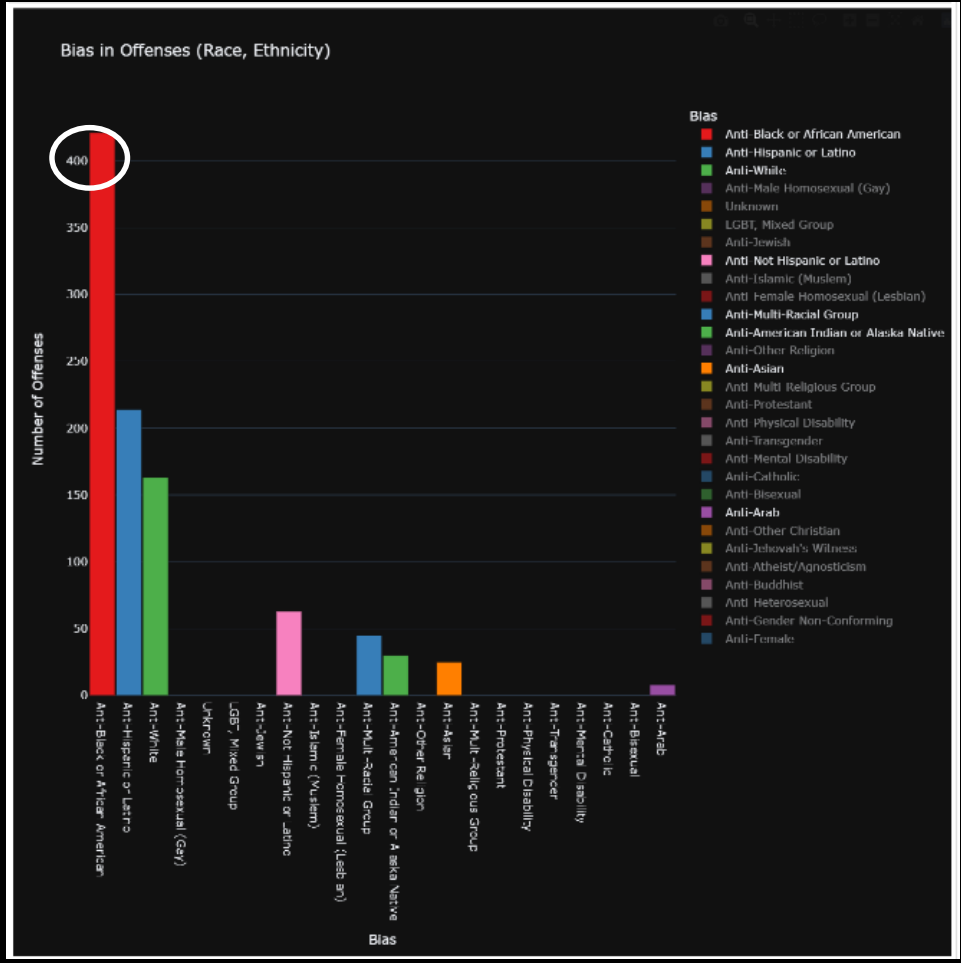
Bias Motivation

Offenses with Bias Recorded
1727

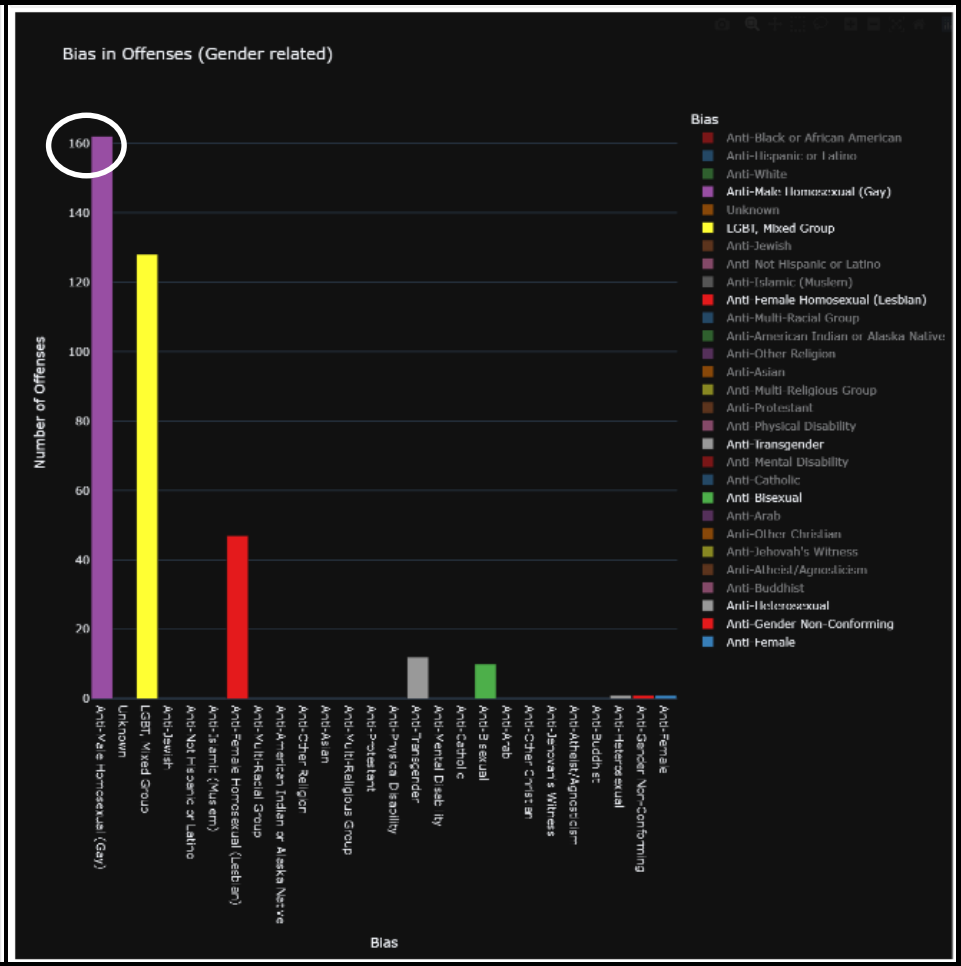


Bias Categories

Race/Ethnicity



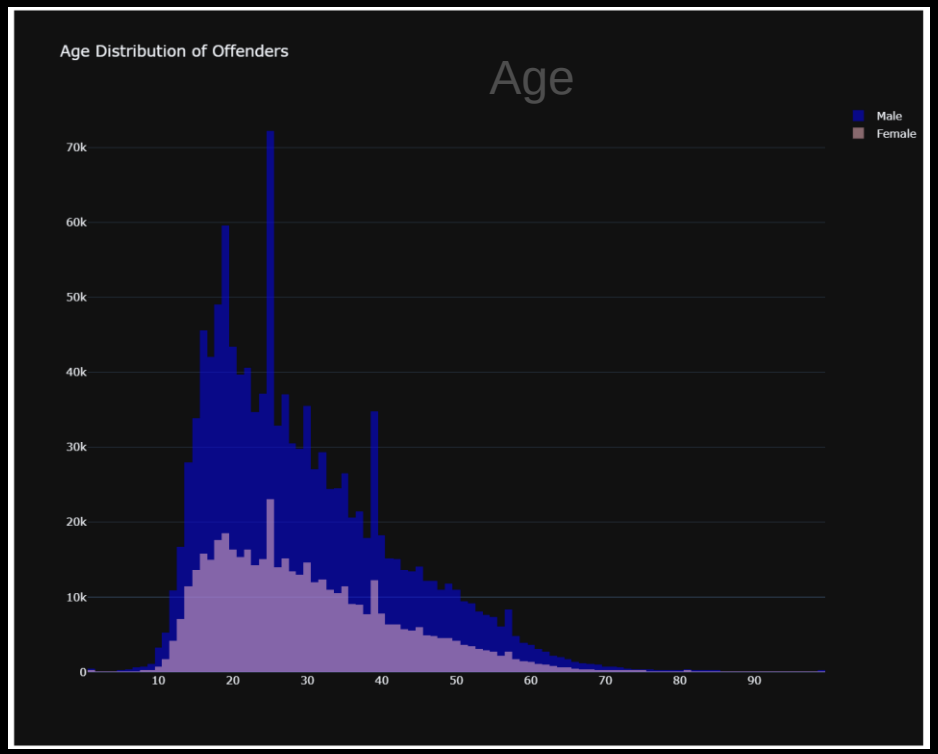
Sexual Orientation

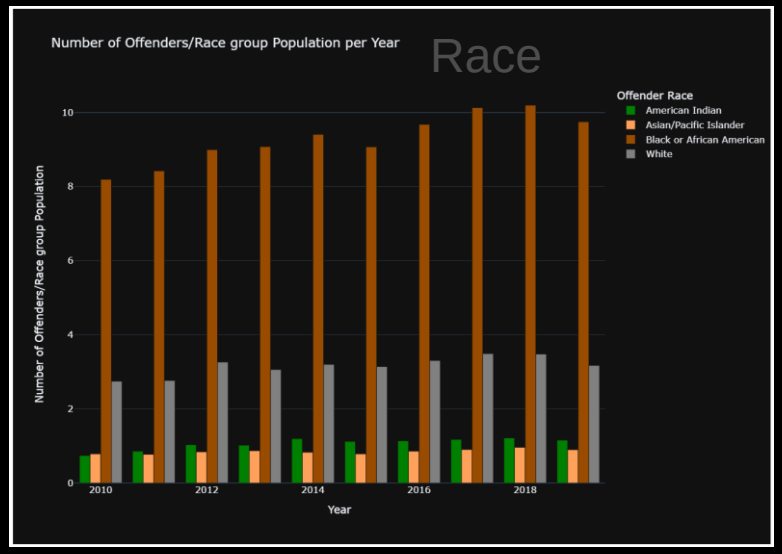


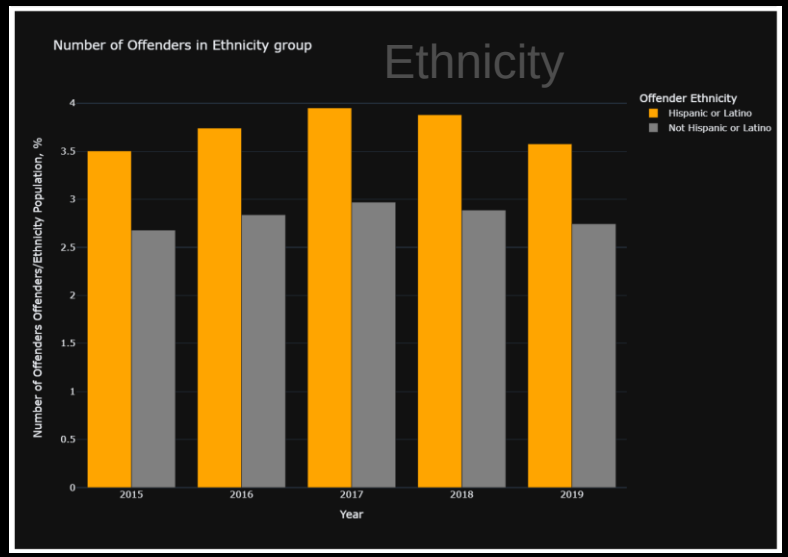
Conclusions (crime details)

1. The most prevalent crime categories in Colorado are crimes against property:
   - **Larceny/Theft, Damage/Vandalism of Property, Burglary/Breaking & Entering**
2. Firearms are used in violent crimes four times more frequently than all other weapons
3. Counties with most offences are:
   - **Denver, El Paso and Arapahoe**
4. Hate crime statistics shows that race and sex orientation motivated offences are the most prevalent.
   - **Race motivated offenses committed against Black victims twice as frequently as against any other race**
   - **Most frequent sexual orientation motivated offenses are committed against gay men**

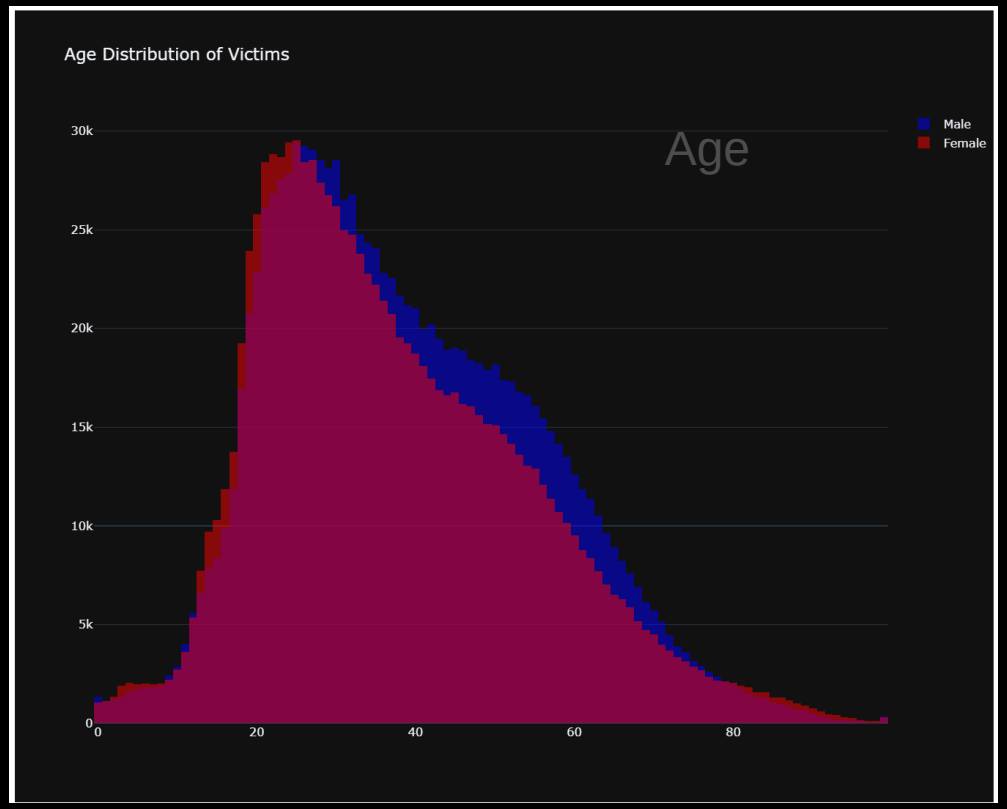
Demographics of Offenders

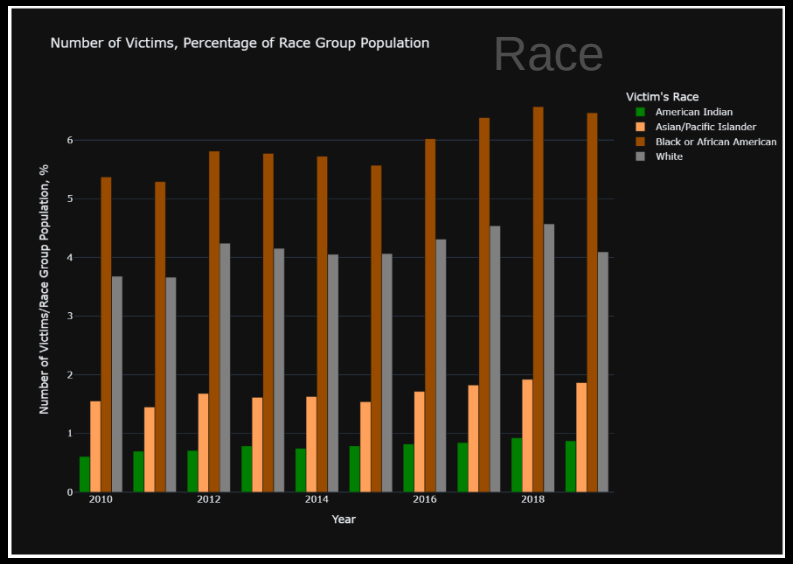


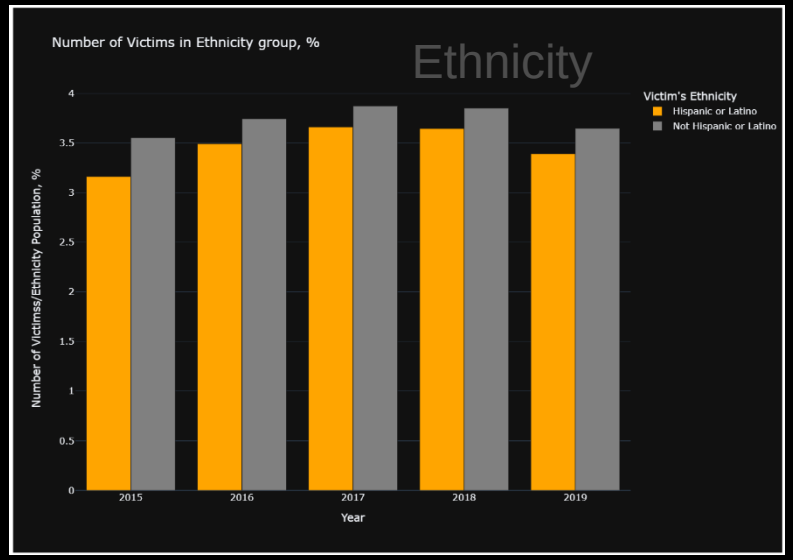




Demographics of Victims





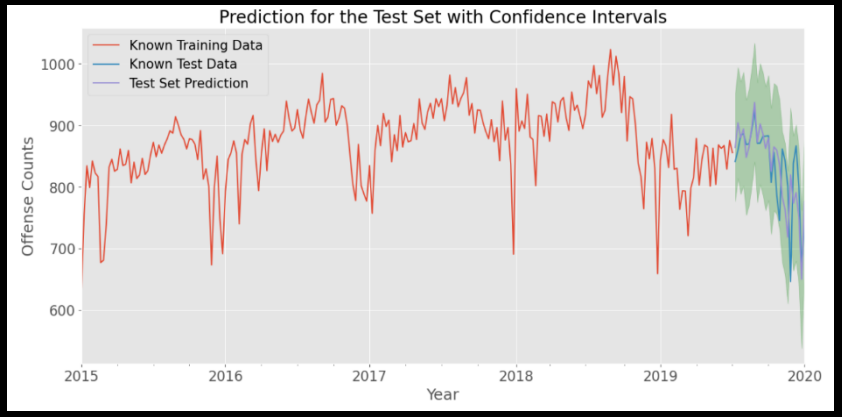


Conclusions (demographics)

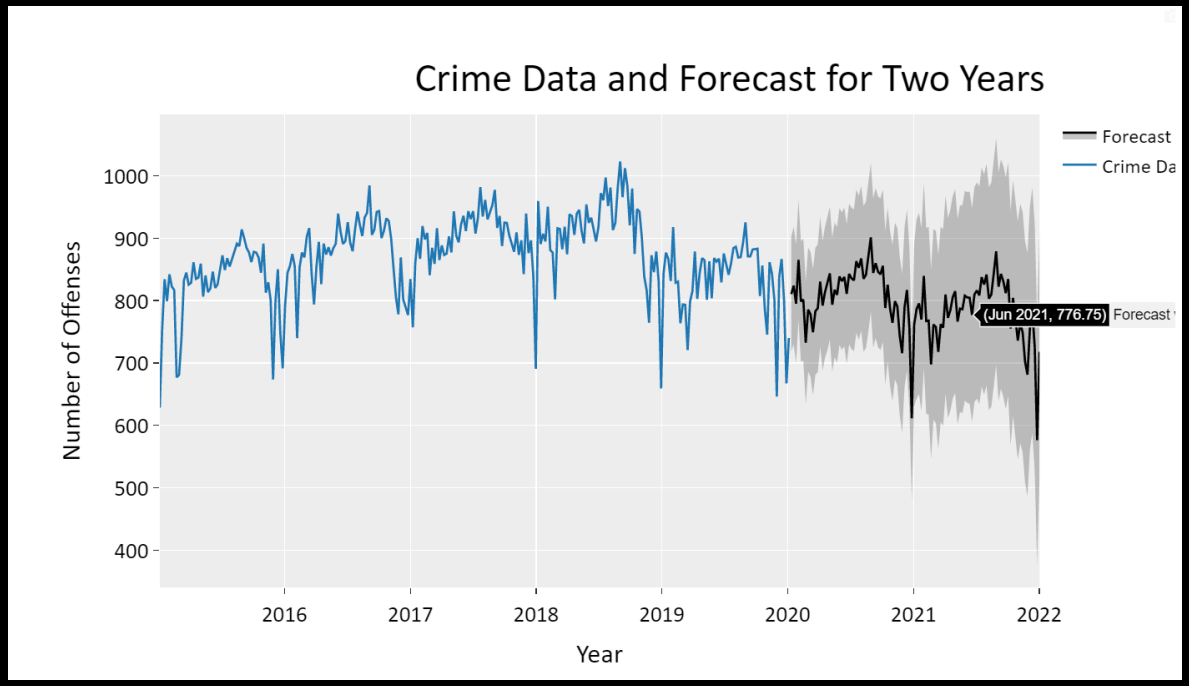
1. Demographics of victims show:
   - **Males and females are equally probable to be victims of a crime**
   - **Distribution of victims' age is skewed toward younger range with the peak at 22 years old**
   - **A victim of a crime is more probable to be Black than of any other race**
   - **A victim of a crime is more probable to be of Non-Hispanic or Latino ethnicity**
2. Demographics of offenders show:
   - **Males more than twice are probable to be an offender in a crime**
   - **Distribution of offenders' age is skewed toward younger range with several peaks at late teen years, 23-25 years and late 30s**
   - **An offender in a crime is more probable to be Black than of any other race**
   - **An offender in a crime is more probable to be of Hispanic or Latino ethnicity**

Modeling General Crime

Real Data (2015-2019)

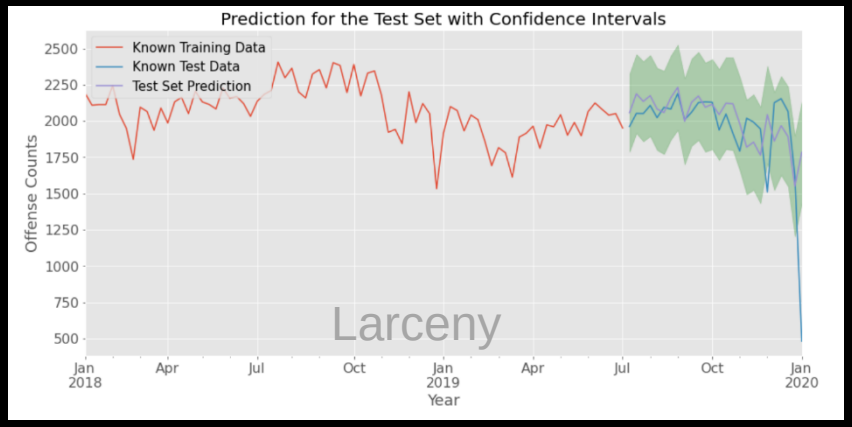


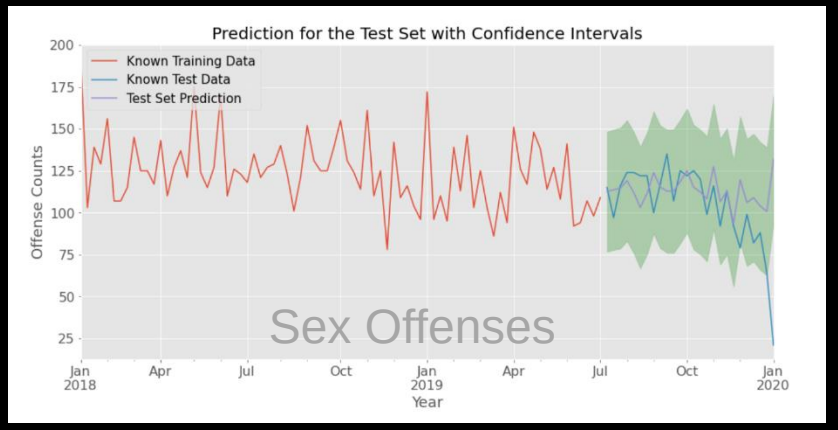
Forecasting Model (2021-2022)



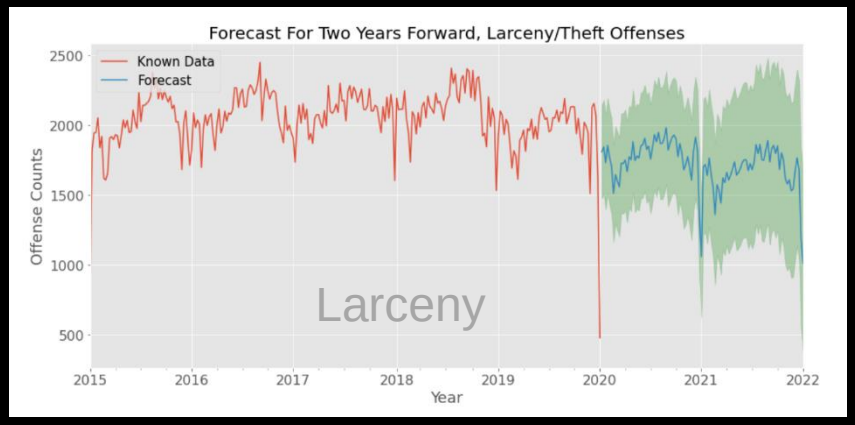
Modeling Crime Categories

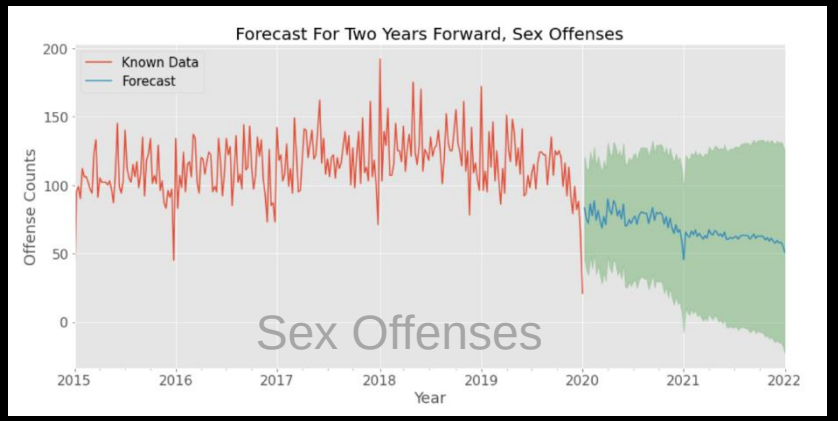
Real Data (2015-2019)





Forecasting Model (2021-2022)





Future Work:

1. Obtain current data; it isn't easy to forecast future trends with data almost two years old
2. Suppose dynamic data becomes available, build an API. This approach would be the most helpful to the general public
3. Add exogenous predictors to the time-series to improve modeling performance. The most helpful predictors:
   - **Socio-economic features of the geographic areas**
   - **Additional offenders' and victims' demographics**
     - ✓ Income
     - ✓ Education
   - **Information about local crime prevention measures and policies**
4. Add geographic locations of committed offenses

Thank You!

Email: e.v.kazakova@gmail.com

GitHub: @sealaurel

LinkedIn: https://www.linkedin.com/in/elena-v-kazakova/