

Project 4 Proposal

Team Members: Carlos Delarosa, Raj Agrawal, Ann Ly, John Banowsky

Subject: Homicide Over the Past Decade

Dataset: [Homicides](#) | [Kaggle](#)

About The Dataset:

The Washington Post collected data on more than 52,000 criminal homicides over the past decade in 50 of the largest American cities. The data included the location of the killing, whether an arrest was made and, in most cases, basic demographic information about each victim. Reporters received data in many formats, including paper, and worked for months to clean and standardize it, comparing homicide counts and aggregate closure rates with FBI data to ensure the records were as accurate as possible.

Inspiration:

- Carlos – My inspiration is to analyzing a homicides dataset to investigate how victim age influences arrest outcomes in cases
- Raj – To learn more about homicides dataset and related investigation, and used acquired knowledge of the data science bootcamp for the data analysis. Eventually, participate in the volunteer program to help community to serve better.
- Ann – I am interested in knowing more about the shared factors among all the homicide cases.
- John – Homicides are never a good thing. Having a better understanding of which populations are affected will help us better address homicide to protect these demographics.

Goals:

Our main purpose is to use a standard dataset and utilize various tools that we learn so far i.e. – machine learning, Tableau

- Ability to connect to a file set
- Ability to fetch data and organize them with various useable data frame
- Perform supervised model / predictions
- To use correlation and create various Bar, Donut, line charts using Tableau

Breakdown of tasks:

John- data cleaning
Carlos – Website
Ann – Machine Learning
Raj – Tableau