**Baltimore and Montgomery County, MD Crime Analysis**

Ola Sanusi [Project Coordinator] and Adali Johnson

**TOPIC**

Given its notorious crime rate, this research will investigate the times, dates, and location information of crimes in the Baltimore, Maryland area to classify violent versus non-violent crimes.

**DATA**

For this research, two datasets will be used to analyze crime in Baltimore and Montgomery County, Maryland. The Baltimore data set, found [here](https://data.baltimorecity.gov/Public-Safety/BPD-Part-1-Victim-Based-Crime-Data/wsfq-mvij) contains the following information about each crime occurrence: date, time, whether it was inside or outside, whether weapons were involved, and a description. It also includes the following information about the location of each crime: post, district, neighborhood, latitude and longitude coordinates, and premise. There are 313,634 observations amongst 16 columns between January 1, 1978 and December 31, 2019. The Montgomery County data set can be found [here](https://data.montgomerycountymd.gov/Public-Safety/Crime/icn6-v9z3) and contains 220,011 observations amongst 30 columns between January 1, 2017 and December 31, 2019. Along with descriptive columns on each crime such as Incident ID, offence code, CR number, and dispatch date/time, but most columns describe the location of the crime. These location columns include the following: city, state, zip code, place, sector, beat, address number, multiple columns of street information, and latitude and longitude coordinates.

**RESEARCH QUESTIONS**

The results of this analysis will seek to answer the question of what specific attributes of crimes can accurately determine whether a future crime will be violent or non-violent. Additionally, it will shed light on what specific information is most relevant in answering such a question whether it be dates, times, locations, etc.

**METHODOLOGY**

As a first step in data wrangling and ingestion, a column will be added to label each crime as violent or nonviolent. This labeling will be done by looking at the Description and Crime Name 2 columns in the Baltimore and Montgomery data set, respectively. The resulting column will be used as a binary target for classification modeling. Each row contains a unique crime that will be used as instances to classify the target variable. While the scope of this research may change, the classification models that will be explored include decision tree, random forest, and logistic regression.

**IMPLICATIONS**

The implications of this research could lead to an improved method of police force allocation. Specifically, if the models created can accurately predict whether crimes will be violent or non-violent in certain areas or at certain times, precincts can begin to staff those areas at those times with more experienced officers. The increased likelihood of violent crime with said demographics can then be handled more appropriately by officers with more experience or more specialized trainings in handling violent crime.