Data Incubator Capstone project

Project Title:

Identifying the race of an individual involved in stop incident by using a statistical model

Prepared by
Aklilu z Welegebrael

Objective

➢ Constructing a model which classify the race of an individual involved in stop incident, stop and frisk incidence

➢Identifying the trend of incidents a long the years

➢To identify the distribution of incidents a long with Gender, Ethnicity group and different Districts

Introduction

☞ Washington Dc is the capital city of US which is one of the populous city.

☞ Based on 2019 population estimation, there are a bout 711,571 population living in DC.

☞ One of the highest crime rate area in US with crime rate of 60 per one thousand residents

☞ The distribution of crime is not uniformly distributed throughout the city

☞ As it is one of the populous city, the DC police has a lot of burden in maintaining the safety of the communities.

☞ Even though most officers doing a great job on maintaining the community safety, there is a claim as some officers abuse their position

➢ Data structure

✓ A bout 27,559 individuals involved in stop incidence(i.e. stop and frisk) from 2010 to 2017 in Washington dc

➢ Exploratory Data Analysis

Table 1.1: Distribution of races

Race	Observations	Percentage
Black	22589	83.4991%
white	1656	6.1213%
Asian	97	0.3586%
American Indian or Alaska Native	13	0.0481%
Native Hawaiian or Other Pacific Islander	11	0.0407%
Unknown	2687	9.9324%

❖**From Table 1.1:distribution of races**

✓A bout 83% of the individuals are black

✓A bout 13(0.0481%) are American Indian or Alaska Native

✓A bout 11(0.0407%) are Native Hawaiian or other Pacific Islander

✓ About 10% of individuals their race is not identified

Table 1.2: Distribution of incidents along Dc districts

District	Observations	Percentage
1D	4320	15.8637%
2D	2424	8.9013%
3D	3986	14.6372%
4D	3771	13.8477%
5D	3843	14.1121%
6D	4474	16.4292%
7D	4414	16.2089%

❖ From Table 1.2: Distribution of incidents a long the different Washington dc districts,

District two has a minimum number of stop and frisk incidents from 2010 to 2017, about 8.9% , however the other districts have higher incidents

Table 1.3: Distribution of incidents along gender

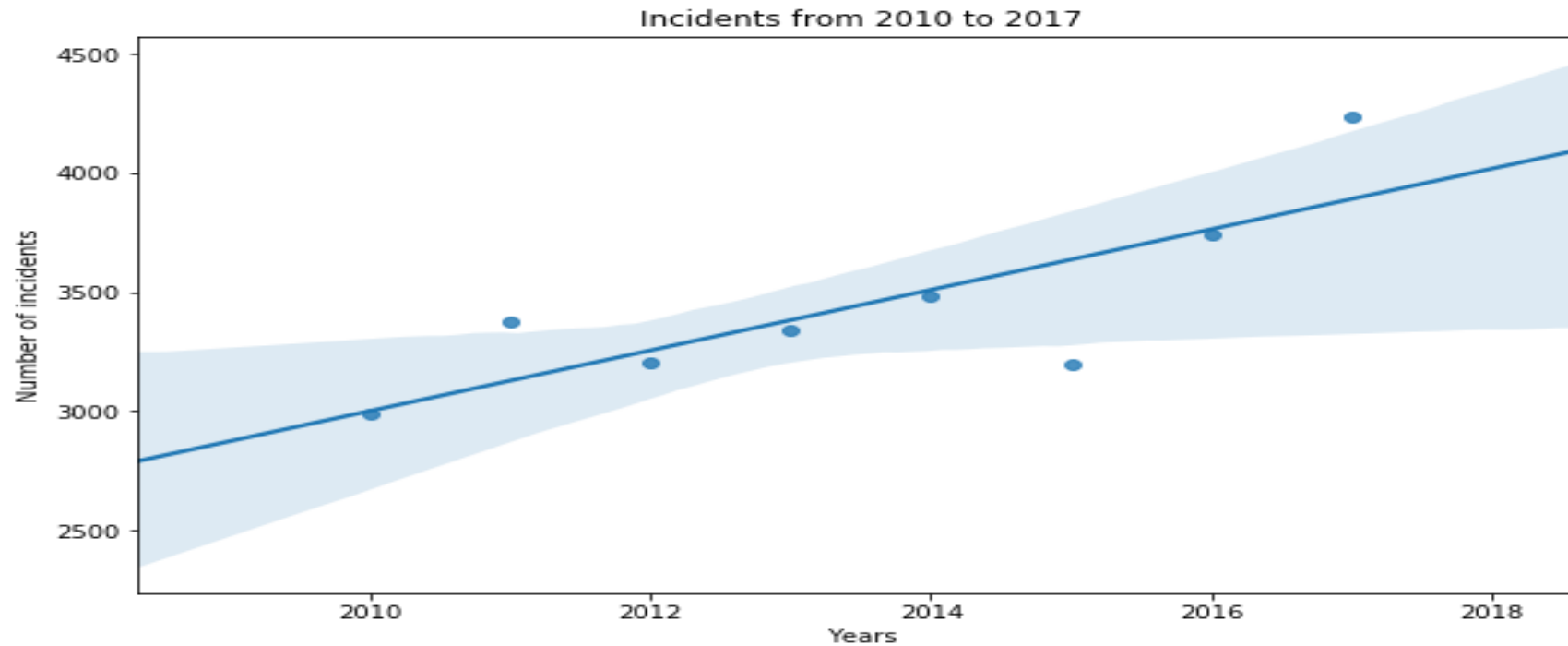| Gender	Observations	Percentage
Male	23259	85.9757%
Female	2413	8.9196%
Unknown	1381	5.1048%

✓ Most of the incidented individuals are males, about 23,259(86%).

✓ The male incidents are about ten times those of female.

Table 1.4: Distribution of incidents along Ethnicity groups

Ethnicity	Observations	Percentage
Not Hispanic or Latino	17571	64.9503%
Unknown	8058	29.7860%
Hispanic or Latino	1424	5.2637%

✓ A bout 65% of the incidents are happened on Not Hispanic or Latino

✓ About 5.00% are belong to Hipanic or Latino ethnicity group.

✓ A bout 30% of the individuals were not known their ethnicity group.

- Figure 1.0 describes the trend of the incidents from 2010 to 2017

Incidents from 2010 to 2017

✓ As we can see from the figure incident looks like increasing from year to year except in 2011(more increase) and 2015 (decrease).

✓ From the fitted line we can say that there is an increasing trend of incidents from 2010 to 2017

❖ Statistical Methods

☞Logistic regression, Decision tree and SVM will be used to classify an arbitrary incidented individual to the different race groups such as:

- ✓ white
- ✓ Black
- ✓ Asian
- ✓ Alaskan Native or American Indian
- ✓ Native Hawaiian or other Pacific islander race