PREDICTING TRAFFIC STOP OUTCOMES

General Assembly Data Science

Capstone Project

Adeniyi Harrison

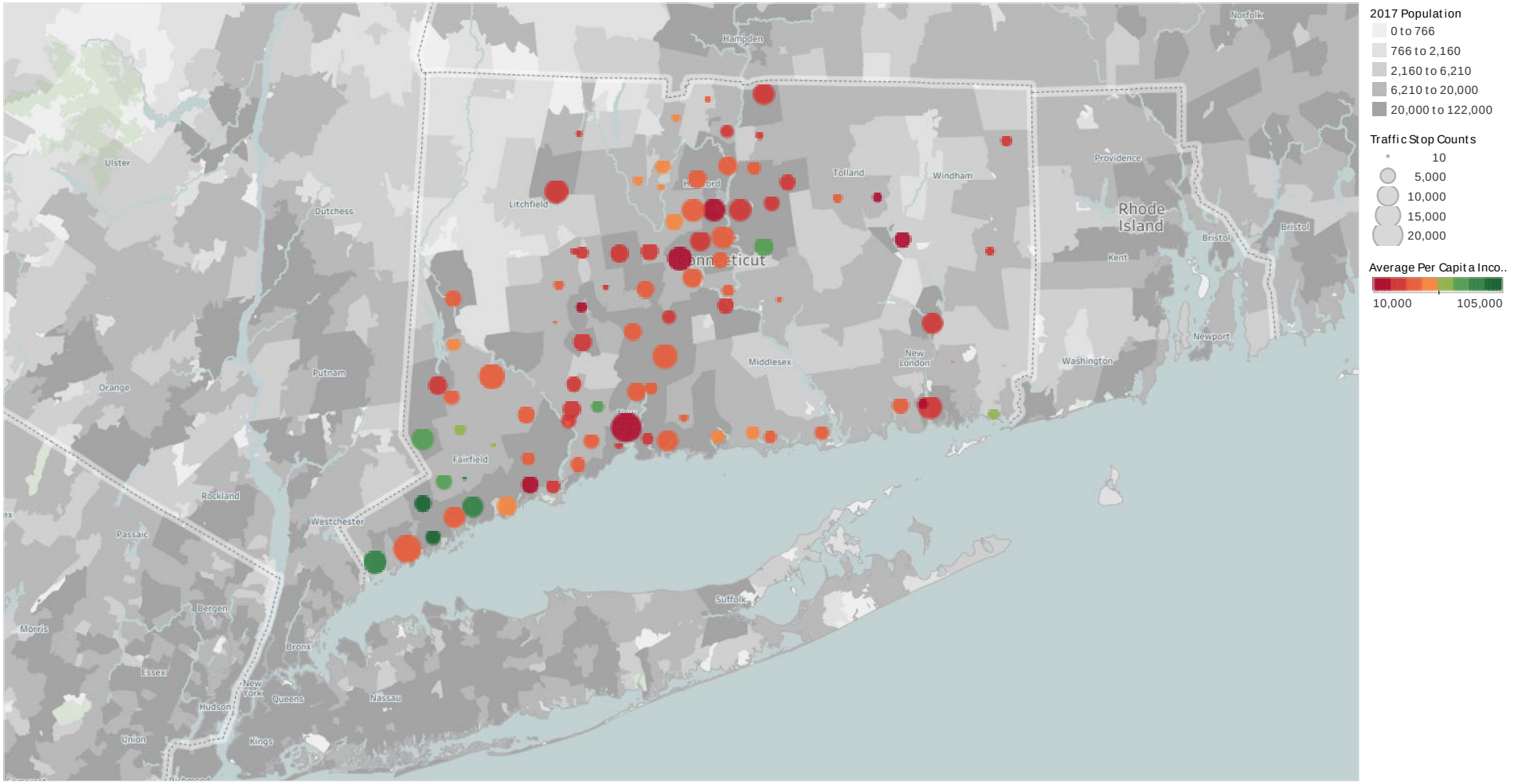
So what are we doing?

- Predicting incident outcomes of traffic stops based on driver attributes, location information and other incident meta data
- Is DWB (Driving While Black/Brown) a real thing?
- Outcome Variable: Police Intervention Disposition (Warning, Ticket or Arrested)

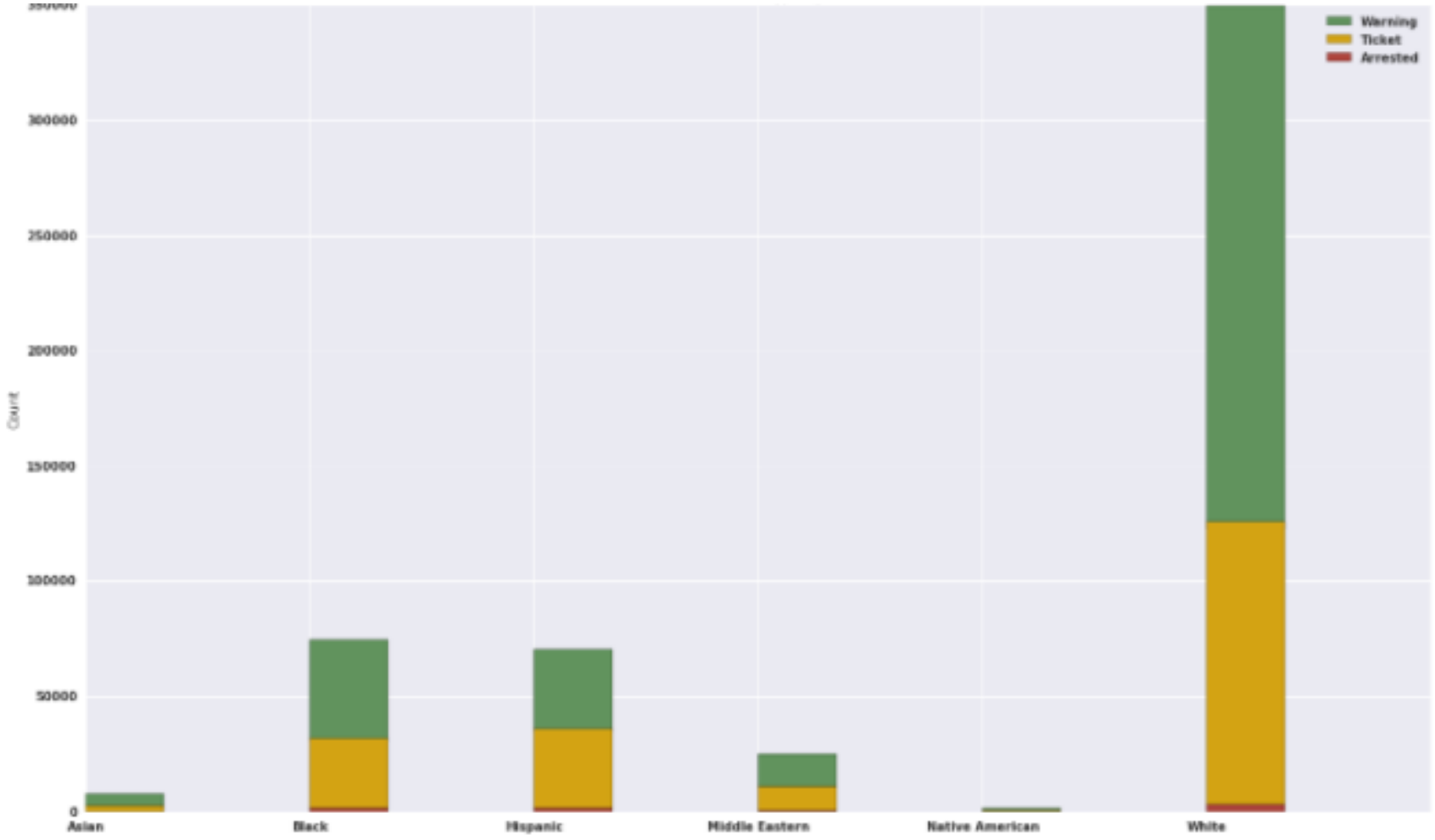
Where is the data from?

- Subscribed to “Data is Plural”
- Connecticut Data Collaborative
  - http://ctrp3.ctdata.org/
- Traffic Stop Data by each police department in Connecticut from Oct 1, 2013 to Sep 30, 2015
  - Mostly Categorical
  - Driver Information
  - Reason for Stop
  - Result of Stop

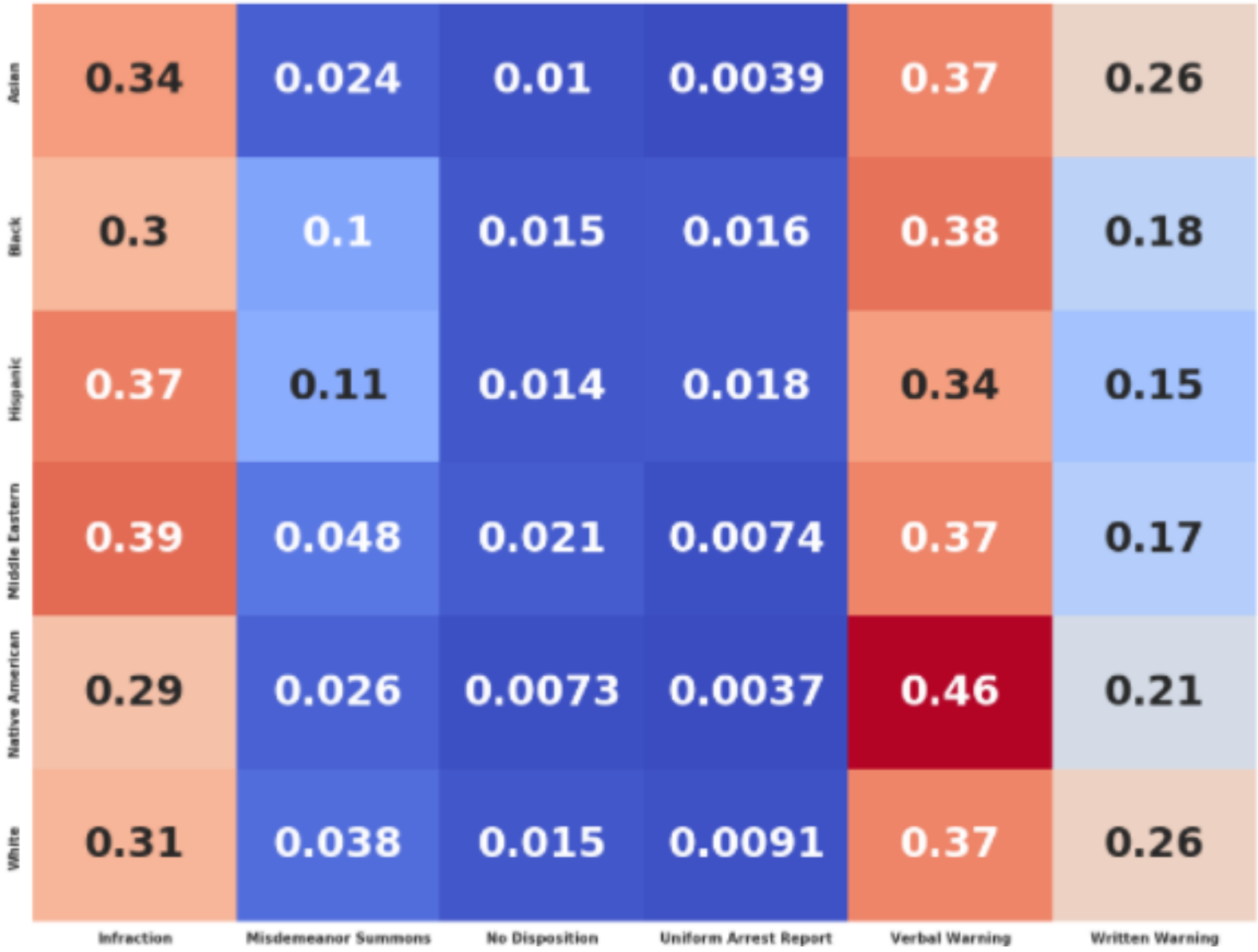
Connecticut Traffic Stops from 10/1/2013 to 9/30/2015



Count of Stops by Racial Groups



Racial Group and Intervention Result Matrix



Where did we leave off?

- Logistic Regression
- Outcome Variable: Whether a stop will result in a ticket/arrest or warning
- Feature Variables: Race, Sex and Age
- Accuracy: 54%

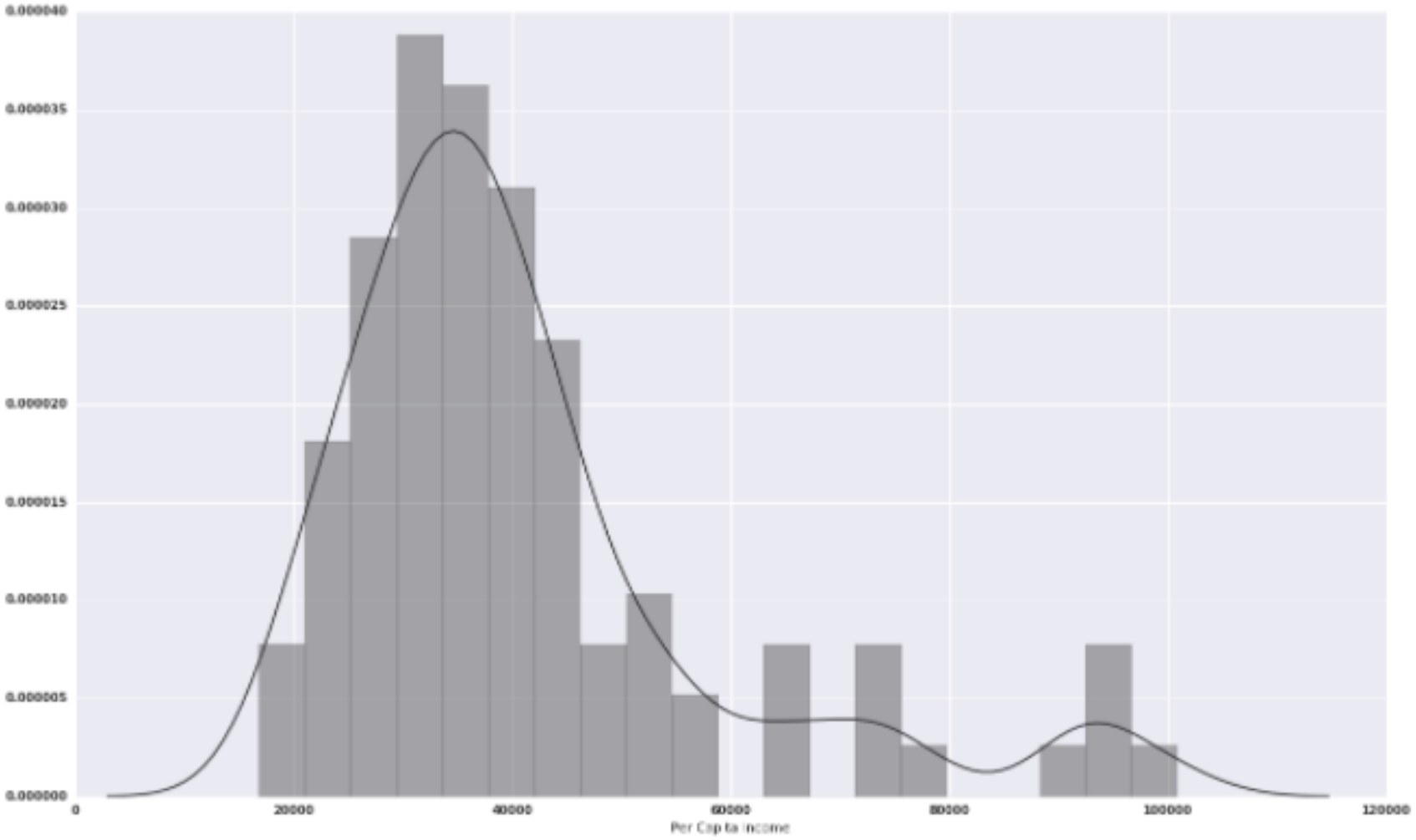
```
array([[61665, 59985],
       [38621, 95410]])
```

What have I done since?

- Google API to convert police departments into geo locations

- Scraped Wikipedia Pages to get demographic data to merge to the locations

Histogram of Per Capita Income of cities in analysis



Clustering

	Population	Per Capita Income	White	Black	Hispanic	Asian
Cluster Group						
0	89035.444444	29321.444444	65.211111	19.044444	25.700000	5.155556
1	28117.166667	59654.000000	92.866667	1.925000	4.450000	4.783333
2	23682.243902	35939.756098	88.112195	7.114634	6.104878	3.953659
3	19020.400000	91632.400000	94.840000	1.220000	2.960000	4.080000

- Kmeans Clustering
- 4 Different types of Connecticut Cities
  - Group 0: Large Population, Low Income City with Large Black and Hispanic Populations (9 Cities)
  - Group 1: Medium Population, Medium Income City with Large White Population (12 Cities)
  - Group 2: Small Population, Low Income City with Small Sized Black, Hispanic and Asian Population (41 Towns)
  - Group 3: Small, Majority White and Wealthy Cities (5 Towns)

Group 0

Bridgeport

City in Connecticut

Bridgeport is a seaport city in the U.S. state of Connecticut. It is the largest city in the state and is located in Fairfield County at the mouth of the Pequonnock River on Long Island Sound. Wikipedia

Weather: 57°F (14°C), Wind E at 9 mph (14 km/h), 44% Humidity

ZIP code: 06601, 06602, 06604, 06605, 06606, 06607, 06608, 06610, 06650, 06673, 06699

Population: 147,216 (2013)

Group 1

Trumbull

Town in Connecticut

Trumbull is a town in Fairfield County, Connecticut bordered by the towns of Monroe, Shelton, Stratford, Bridgeport, Fairfield and Easton. The population was 36,018 according to the 2010 census. Wikipedia

Weather: 58°F (14°C), Wind E at 8 mph (13 km/h), 40% Humidity

Zip code: 06611

Hotels: 3-star averaging $135, 5-star averaging $145. View hotels

Incorporated: 1797 as Trumbull

Population: 36,018 (2010)

Group 2

Vernon

Town in Connecticut

Vernon is a town in Tolland County, Connecticut, United States. The population was 29,179 at the 2010 census. Vernon contains the smaller villages of Rockville, Talcottville and Dobsonville. Wikipedia

Weather: 56°F (13°C), Wind N at 7 mph (11 km/h), 39% Humidity

Zip code: 06066

Hotels: 3-star averaging $136. View hotels

Getting there: 6 h 55 min flight, from $430. View flights

Population: 29,179 (2010)

Group 3

Darien

Town in Connecticut

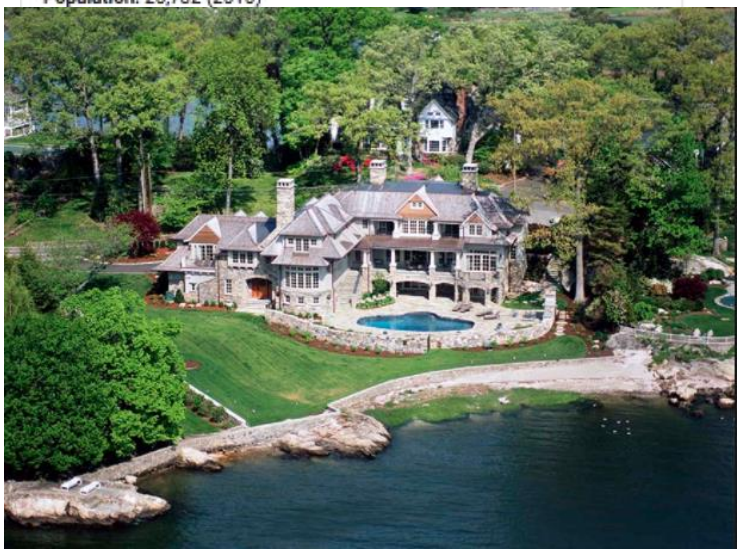
Darien is a town in Fairfield County, Connecticut, United States. Located on Connecticut's "Gold Coast," the population was 20,732 at the 2010 census. Wikipedia

Weather: 55°F (13°C), Wind E at 5 mph (8 km/h), 48% Humidity

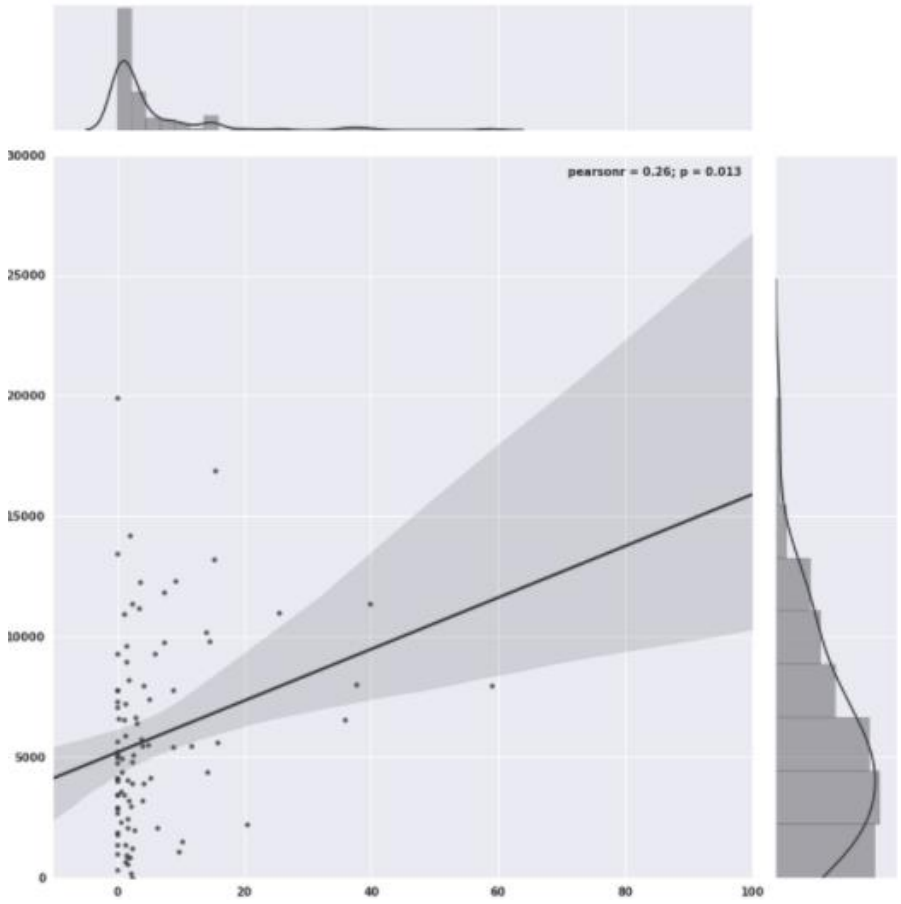
Zip code: 06820

Hotels: 3-star averaging $153. View hotels

Population: 20,732 (2010)

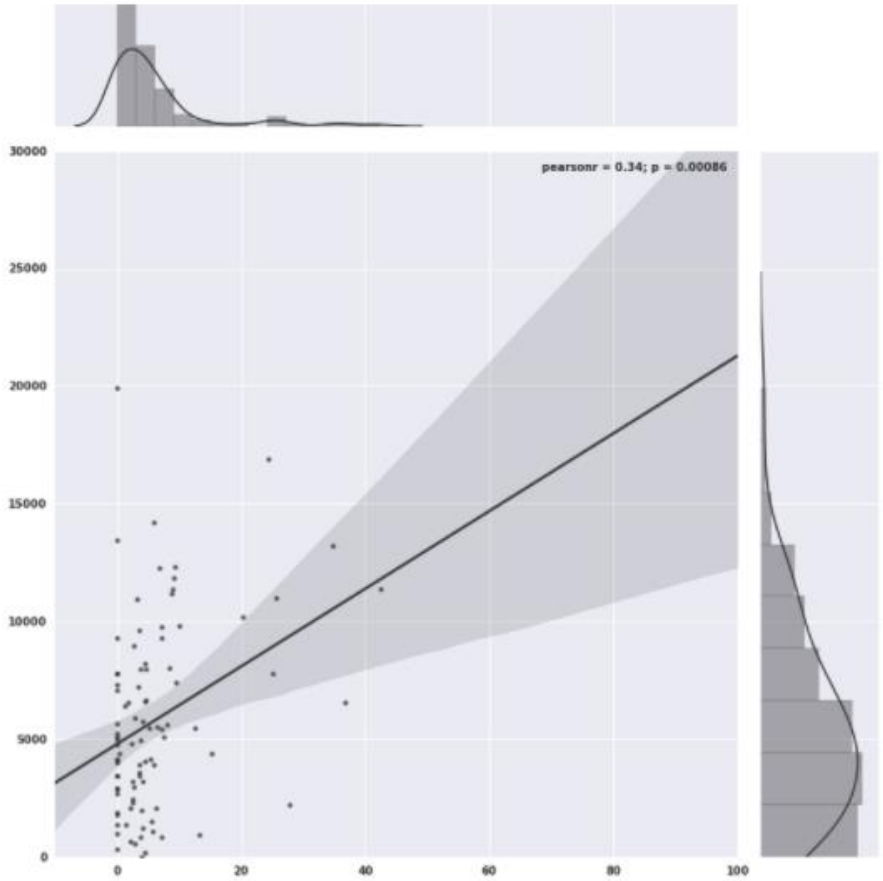


Number of Stops vs. Percentage of Black Population

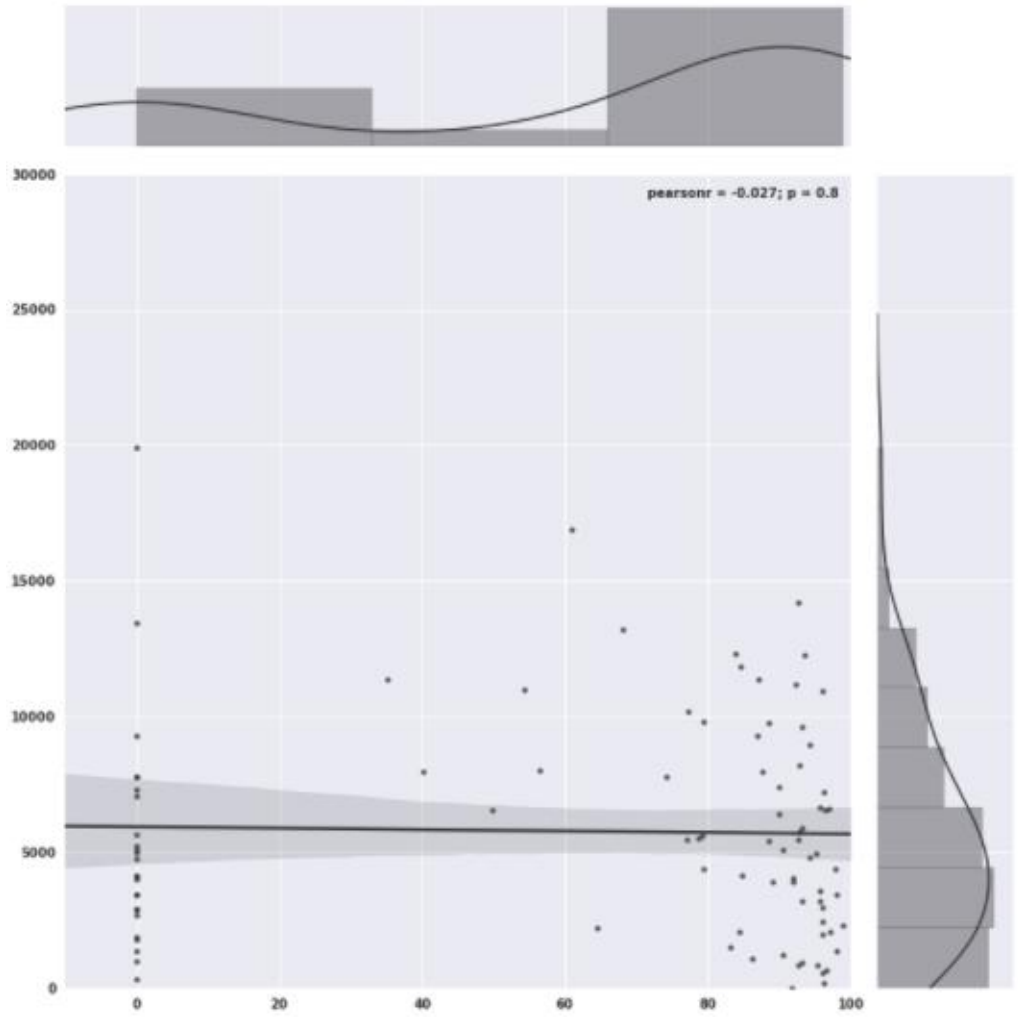


R Squared: .26

Number of Stops vs. Percentage of Hispanic Population



R Squared: .34



Number of Stops vs. Percentage of White Population

R Squared: -.03

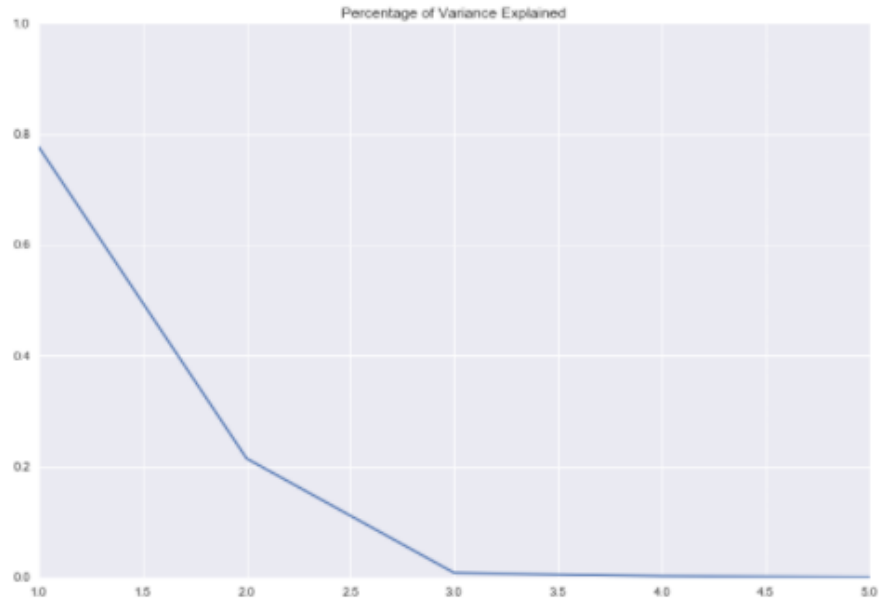
Logistic Regression

- Feature Variables: Race Group, Sex, Age, Population, Per Capita Income, Racial Population Percentages
- Predicting whether a driver will get a ticket or arrested
- Accuracy: 63%
- Coefficient Interpretation
  - Black Drivers are 2% less likely to get a ticket compared to White Drivers
  - Hispanic Drivers have 12% increased chance of getting a ticket over a White Driver
  - As a city's total black percentage population increases by 1 percentage point the chances of any driver getting a ticket increases by 1.17x
  - Each 1 point increase in Hispanic population the probability is 1.09x

Principal Component Analysis

- Feature Variables: Race, Age, Sex, Day of Week, Month ,Resident Indicator, Stop Reasoning, Stop Duration, Vehicle Searched, Vehicle Towed, Demographic data (50 Variables total)

```
Percentage of Variance Explained:  [  7.75896976e-01   2.14301526e-01   7.68517053e-03   1.93452658e-03
   1.81696400e-04]
(0, 1)
```



Principal Component Analysis

- Logistic Regression with 2 Eigenvalues
- No Improvement from original Logistic Regression
- Accuracy: 63%

```
[[220732  24248]
 [120899  30733]]

**************************

             precision    recall  f1-score   support

          0       0.65      0.90      0.75    244980
          1       0.56      0.20      0.30    151632

avg / total       0.61      0.63      0.58    396612
```

Principal Component Analysis

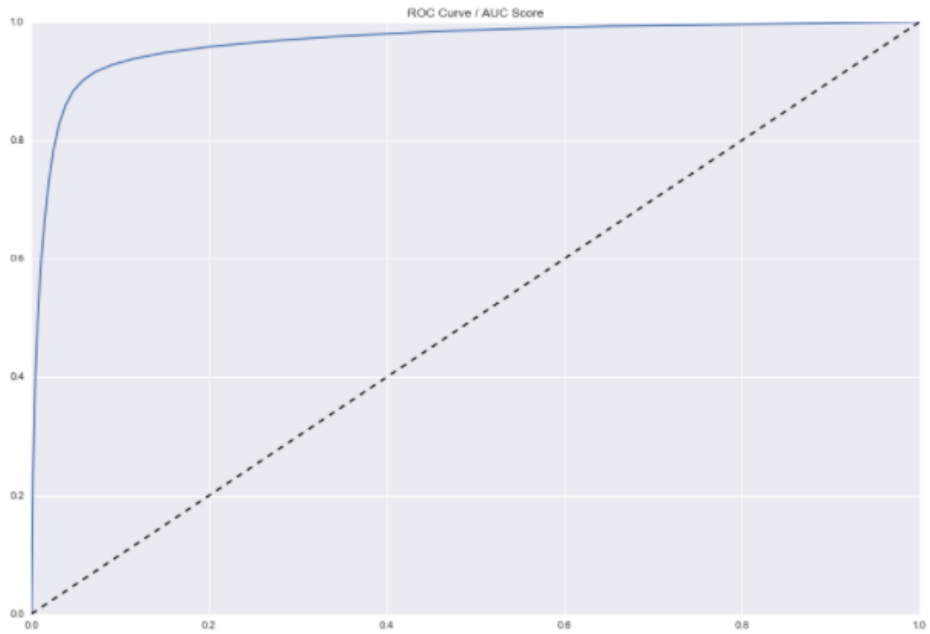
- Random Forest Classifier with 3 Eigenvalues
- Accuracy: 68%

```
[[64530  8677]
 [29815 15962]]

****************************

             precision    recall  f1-score   support

          0       0.68      0.88      0.77     73207
          1       0.65      0.35      0.45     45777

avg / total       0.67      0.68      0.65    118984
```

Random Forest Classifier

- Feature Variables: Race, Age, Sex, Day of Week, Month ,Resident Indicator, Stop Reasoning, Stop Duration, Vehicle Searched, Vehicle Towed, Demographic data (31 Variables total)
- The model is 78% accurate at predicting whether a driver will be ticketed or arrested once pulled over

	precision	recall	f1-score	support
0	0.80	0.85	0.82	73785
1	0.73	0.65	0.69	45278
avg / total	0.77	0.78	0.77	119063



End.