An Outcomes Based Analysis into Racial Bias for Traffic Stops in Nashville & Hartford

Alex Conroy

Importance of study

- Public Perception of Policing
- Outcomes based analysis is an underexplored side of this equation

The Dataset

Stanford Dataset

City Level Data for 2 cities

1 year of data

Data: Categorical

- 11 common variables
  - Geographic
  - Demographic
  - Outcome
- Combined Census Data: ZCTA & Location Demographic
- GeoPy and Google Maps location used
- All treated as categorical

Approach

Reproduce model accuracy with full and lean variable sets to determine if ethnicity variables are critical.

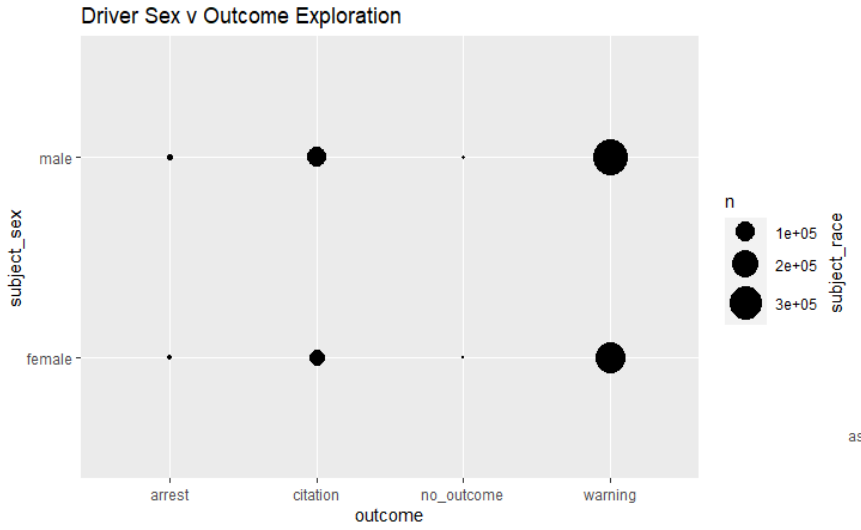
Using Multinomial Logarithmic Regression

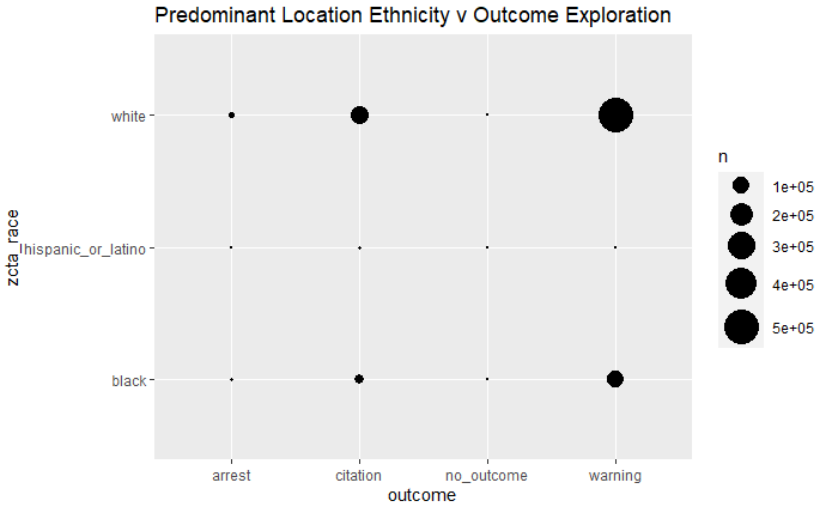
Cities targeted based on Ethnical make up compared to US average, and size.

Data missing critical demographic, location, and outcome variables was removed.

Exploration

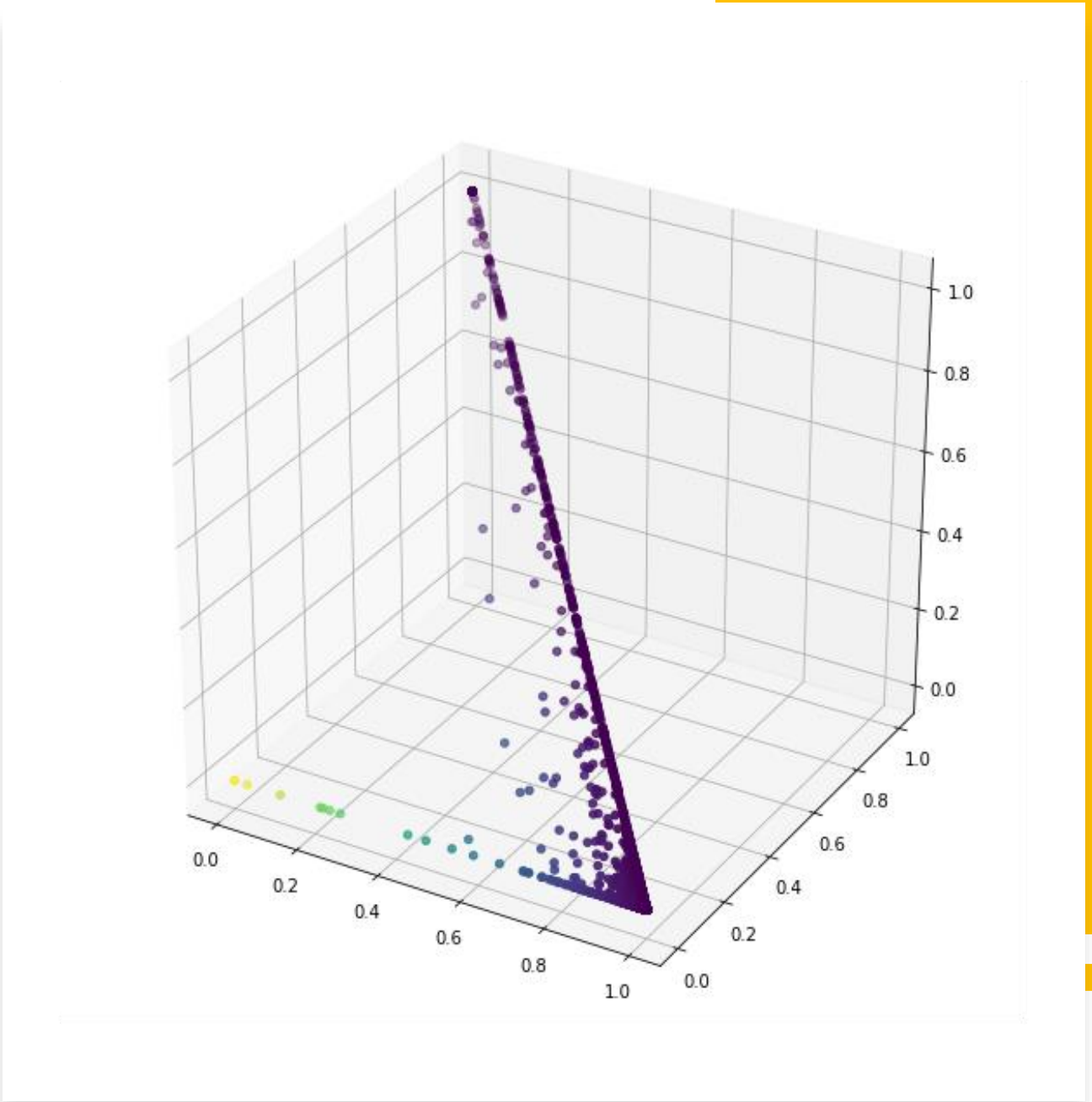
- Focused on understanding the spread of outcome to combined demographic and geographic variables
- Correlation
  - No significant values, but indication of above normal relationships with ethnicity of driver and location





Exploration

ZCTA Ethnicity Check



Model

Run separately on Nashville and Hartford

- Common variables

Run as a combination

- Not separated as Test/Train due to different definitions within the data

Combined City Dataset chosen as main model

- Was able to maintain accuracy on Test data using no demographic or geographic variables
- With comparable AIC, Deviation, McFadden, and R^2 scores

Best AIC and Fit scores recorded using the Hartford only dataset.

- Used Driver Ethnicity Variable
- Lower Accuracy

Model Results

Model	Variables	AIC	McFadden	R^2ML	R^2CU	Accuracy	Error	McNemar's Test
Full	date, time, zcta_race, subject_age, subject_race, subject_sex, outcome, contraband_found, search_conducted, search_basis, reason_for_stop	621,603	0.15	0.17	0.24	79.53%	20.47%	< 2.2e-16
Lean	reason_for_stop, contraband_found, search_basis	621,933	0.15	0.17	0.24	79.56%	20.43%	< 2.2e-16

Findings

Evidence to suggest the model choice is not the best for the data

However, reasonable high accuracy was achieved

Evidence to show that Ethnicity doesn’t affect stop outcome

More modelling required

Limitations

Cities differed in both population and ethnical proportions

Low number of consistent variables between cities

Model Fit

Self Reported Data

ZCTA inaccuracies

Future Work

Find more suitable cities to add to analysis

Investigate more appropriate models

Questions