Phase 3 Project Presentation

Alexandra Yakovleva 2024

Business Problem.

Stakeholder:

Seattle Police Department.

Objective:

an independent assessment of the potential race, age, gender biases present in SPD.

Seattle

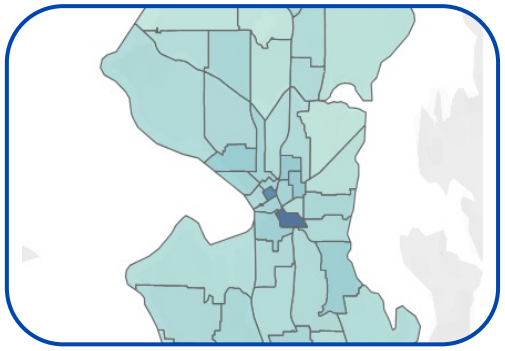
Proposed Solution.

Obtain city-wide public safety dataset.

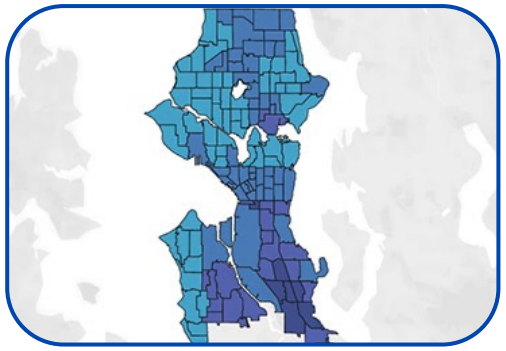
Train a binary classifier on the dataset.

Evaluate the coefficients of the classifier for potential race, age, gender biases.

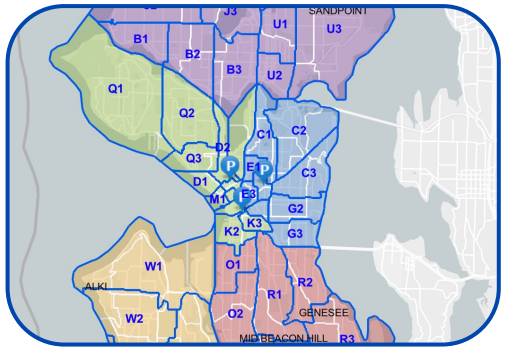
Datasets Overview.



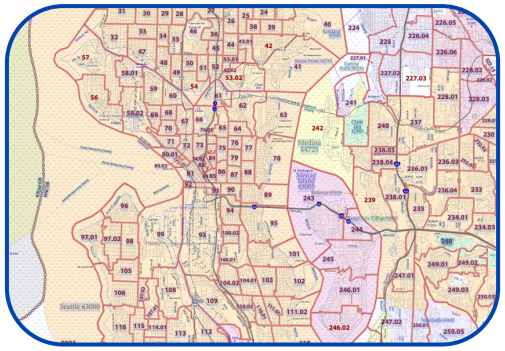
SPD’s Terry Stops Dataset (entries from 2015-2024).



Seattle’s Racial and Social Equity Index.



Geographic SPD Beats.



Geographic Seattle Census.

Seattle

Terry Stops Dataset overview.

Over 60,000 entries with 23 columns.

Each row is a unique record of a Terry Stop.

Each column is a feature of the stop.

Terry Stops Dataset Issues.

The dataset is subjective and potentially biased.

Solution: generalize categories.

The dataset is unbalanced.

Solution: use synthetic sampling tool.

The dataset has many partially filled entries.

Solution: generalize categories.

Data Processing: Cleaning.

- Removed duplicates.
- Applied categorization.
- Dropped missing values.
- Replaced SPD beats with socioeconomic index.
- Data loss: 16% entries.

Data Processing: Transforming.

- Scaled numeric data features.
- Encoded categoric data features.
- Split data into training and testing sets.
- Applied synthetic sampling to training set.

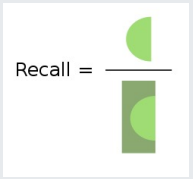
Classifier Parameters.

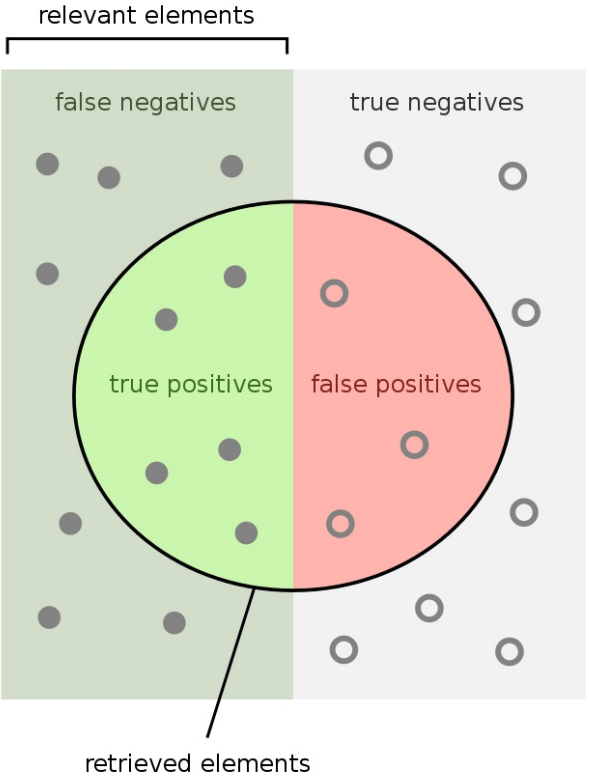
Model Type	Model Parameters.	Model Scoring.
Logistic Regression.	Dependent Variable: Arrest Flag. Independent Variables (N = 36): race, sex, gender.	Balance precision and recall.

Classifier Report.

Highest Impact Factors	Impact on Arrest Probability
Officer Age	Increases with officer's age, peak at 35-55 age.
Subject Age	Peaks at 26-35 age.
Subject Race	Highest for white subjs.
Stop Initiated by Officer	Increases when true.
Racial, Socioeconomic Score	Increases with score increase.

Classifier Performance.

Metric	Meaning	Not Arrested	Arrested
Precision		92%	14%
Recall		58%	59%
F1	Harmonic mean of precision and recall	71%	23%
Performance	-	Good	Poor



Conclusion.

Continue	Focus on	Consider
The diversity training	Officers born before 1995.	Optimizing data entry procedure for future Terry stops.

Seattle

Future Work: Improve Classifier Performance.

Imbalanced Dataset.

Model Complexity.

Feature Engineering

Thank You!

- Reach out at leksea@gmail.com