Phase 3 Project Presentation

Alexandra Yakovleva 2024

Seattle

Business Problem.

Stakeholder:

Seattle Police Department.

Objective:

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

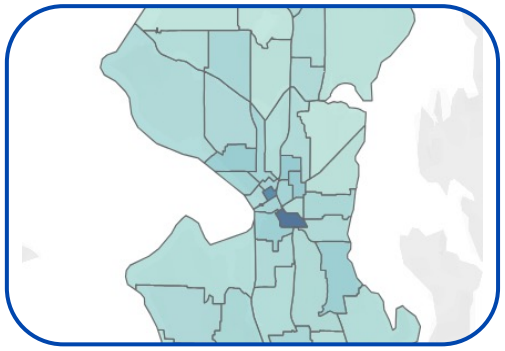
Seattle

Proposed Solution.

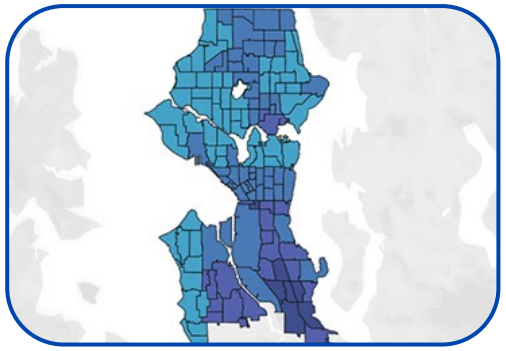
Obtain a recent city-wide public safety dataset and train a logistic regression model.

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

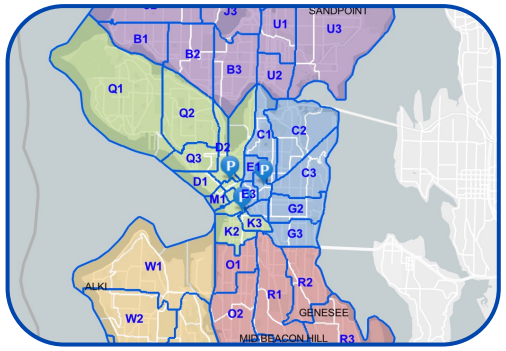
Datasets Overview.



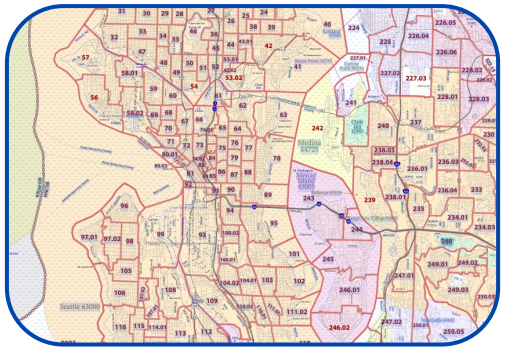
SPD’s Terry Stops Dataset (entries from 2015-2024).



Seattle’s Racial and Social Equity Index.



Geographic SPD Beats.



Geographic Seattle Census.

What is a Terry Stop?

Terry stop is a police-civilian contact that involve a stop and limited detention of an individual.

Terry stop is **authorized under law and based on an officer's reasonable suspicion.**

Terry stop **does not require a probable cause.**

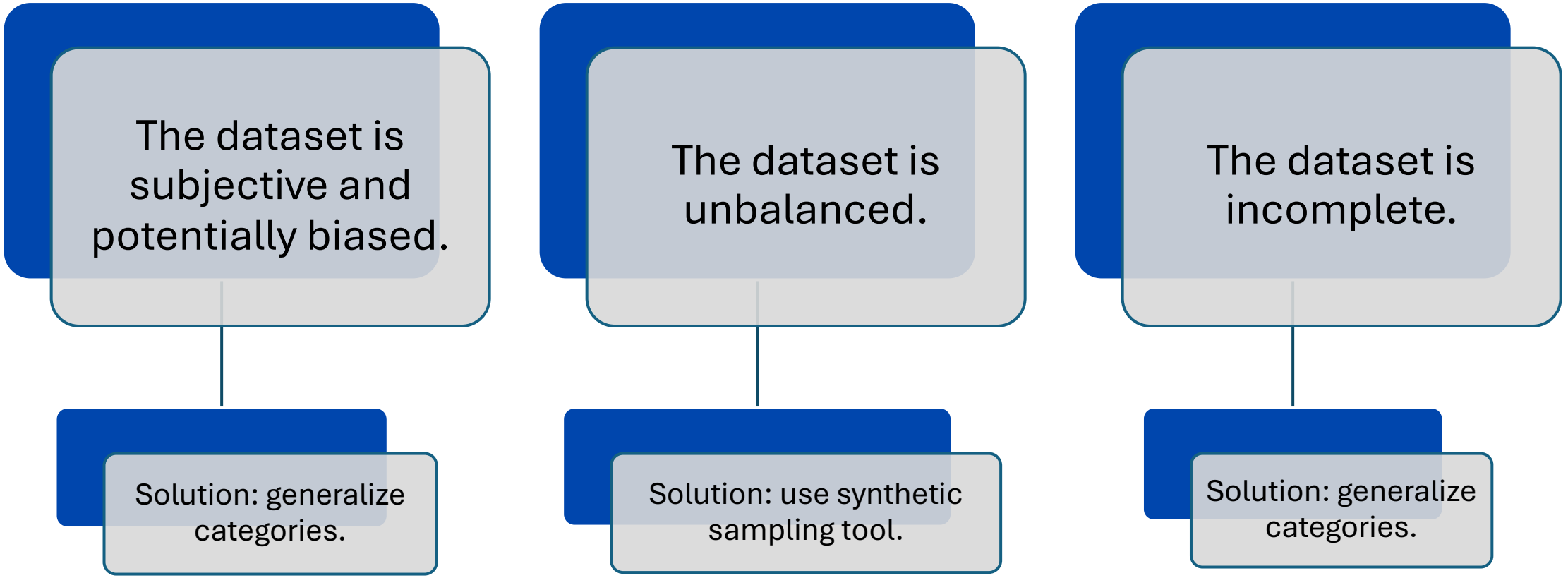
Terry Stops Dataset overview.

Over 60, 000 entries.

Each row is a unique record of a Terry Stop, *as reported by the officer conducting the stop*.

Contains 23 columns describing the officer and suspect's race, gender, age, date and time, police beat, frisk flag, weapons found, and final stop resolution.

Terry Stops Dataset Issues.



Seattle

Data Processing: Cleaning.

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

Data Processing: Classification.

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

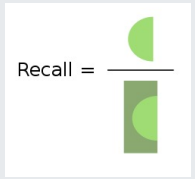
Classifier Parameters.

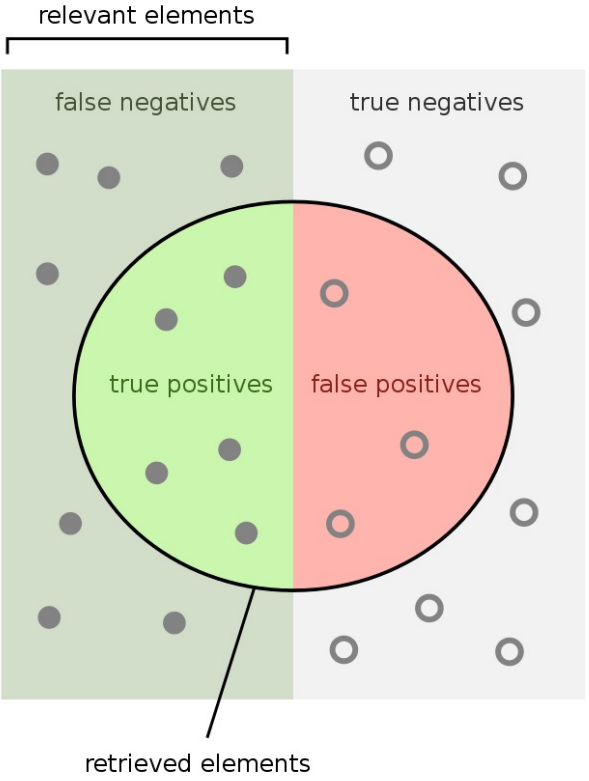
Model Type	Model Parameters.	Model Scoring.
Logistic Regression.	Dependent Variable: Arrest Flag. Independent Variables (N = 36): race, sex, gender, time.	Balance precision and recall.

Classifier Report.

Factor	Impact on Arrest Probability
Officer Age	Increases with officer’s age.
Subject Age	Peaks at 26-35 age.
Officer Race	Either increases/decreases based on race.
Subject Race	Peaks for white subjs.
Officer Sex	Increases for female officers.
Subject Sex	Increases for male subjects.

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: The chosen model might be too complex for the given data.

Feature Engineering: The features used for training might not be informative enough to allow the model to make accurate predictions.

Thank You!

- Reach out at leksea@gmail.com