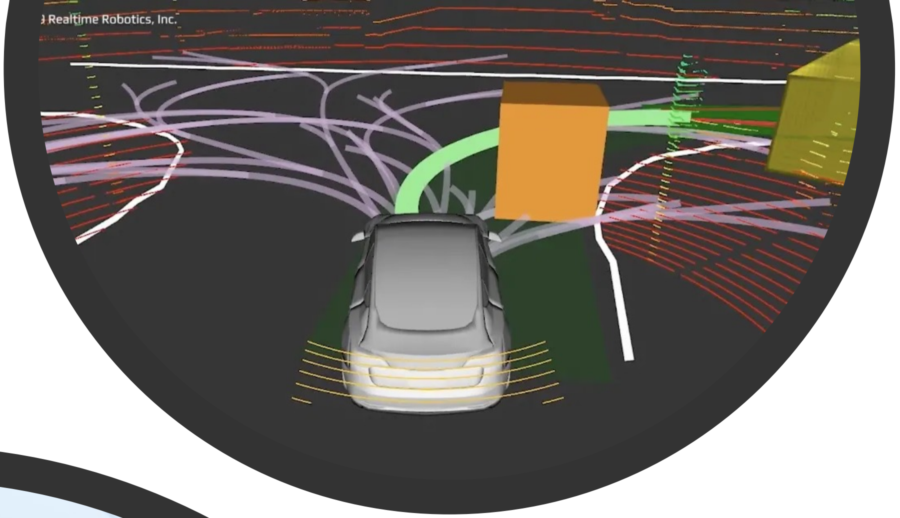
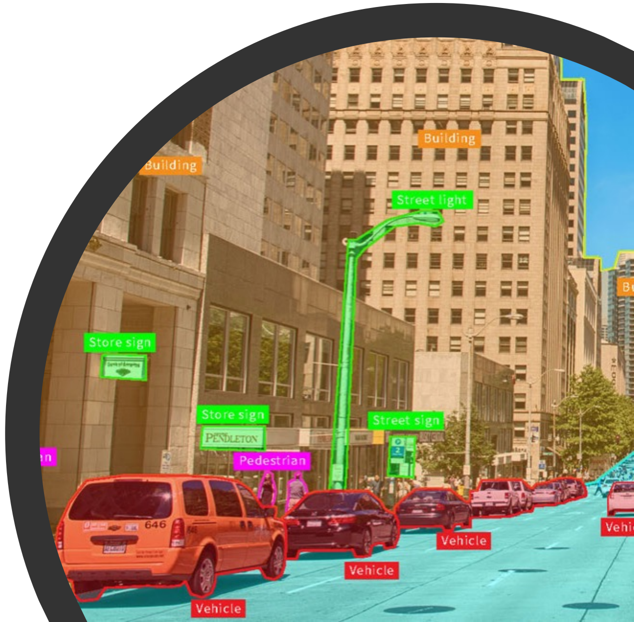
OpenSight: Leveraging Explainable AI for Targeted Congestion Reduction and Pollution Mitigation

Authors: Vedic Panda, Ish Mehta

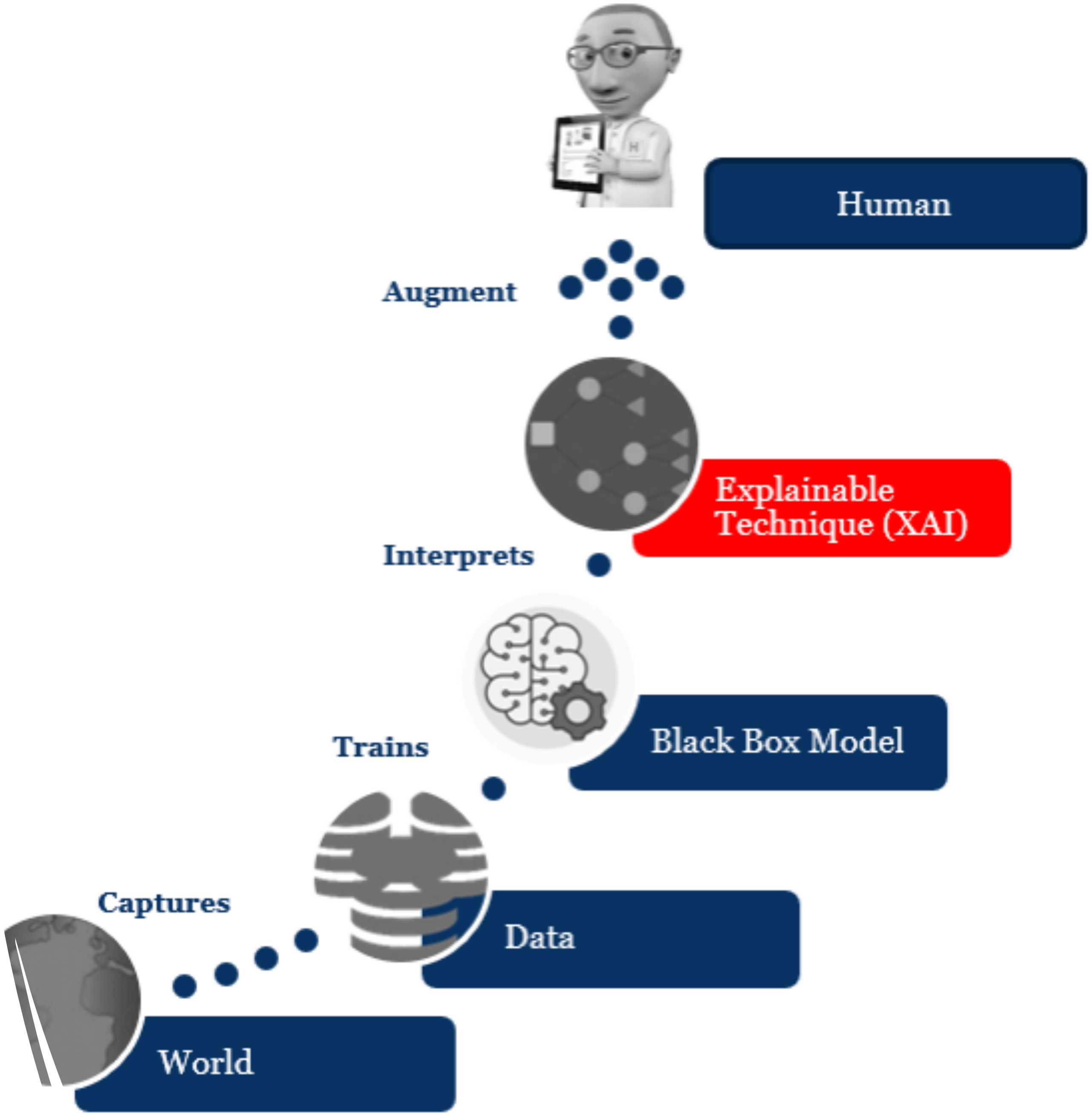
Fall 2023

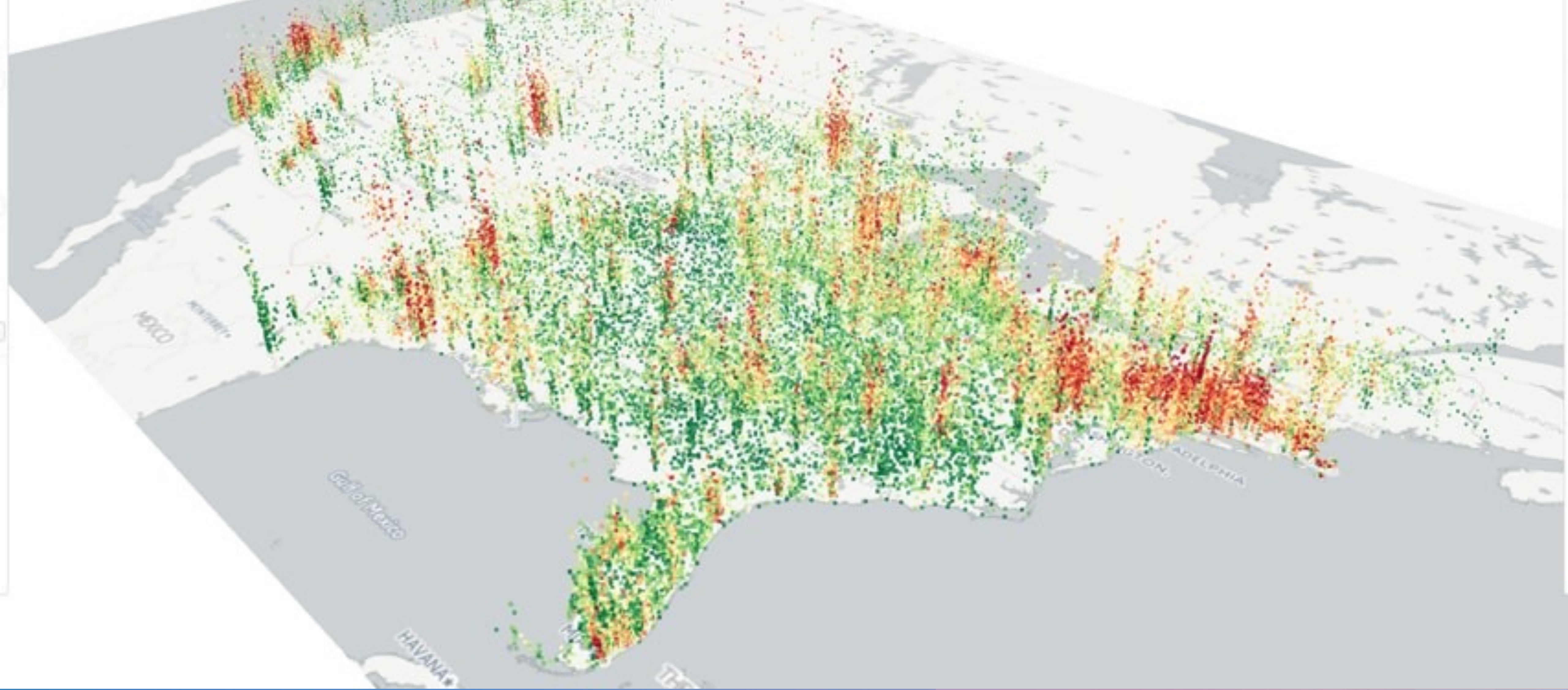


Artificial Intelligence



Issue: AI as a BlackBox. Solution? XAI



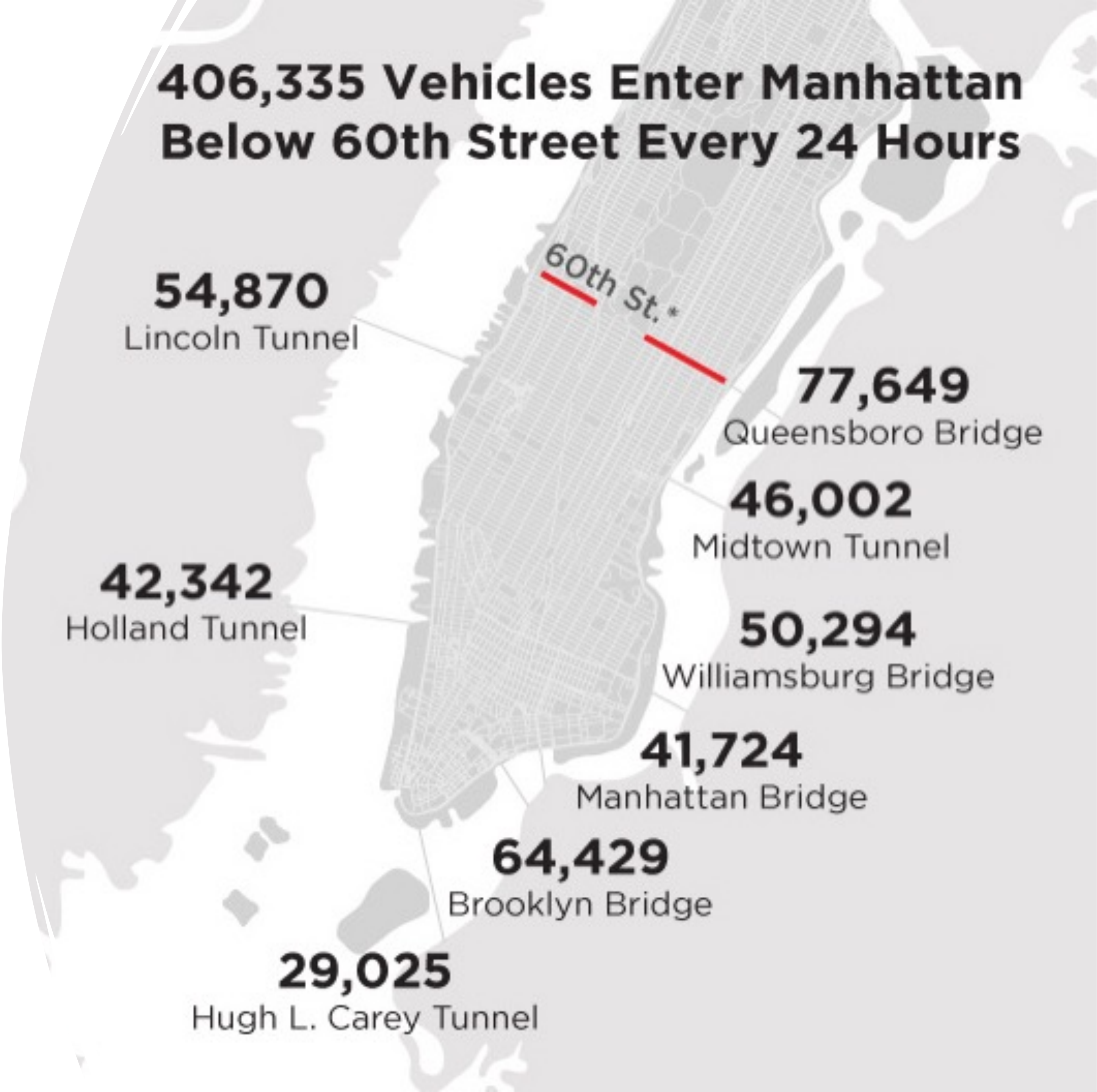


Motivation – Complex Datasets

Community needs

Our target audience:

Policy makers for NYC congestion tax, on air quality, traffic volume, and overall urban sustainability.



Figma Mockup - Link

Approach

Data pre-procesing

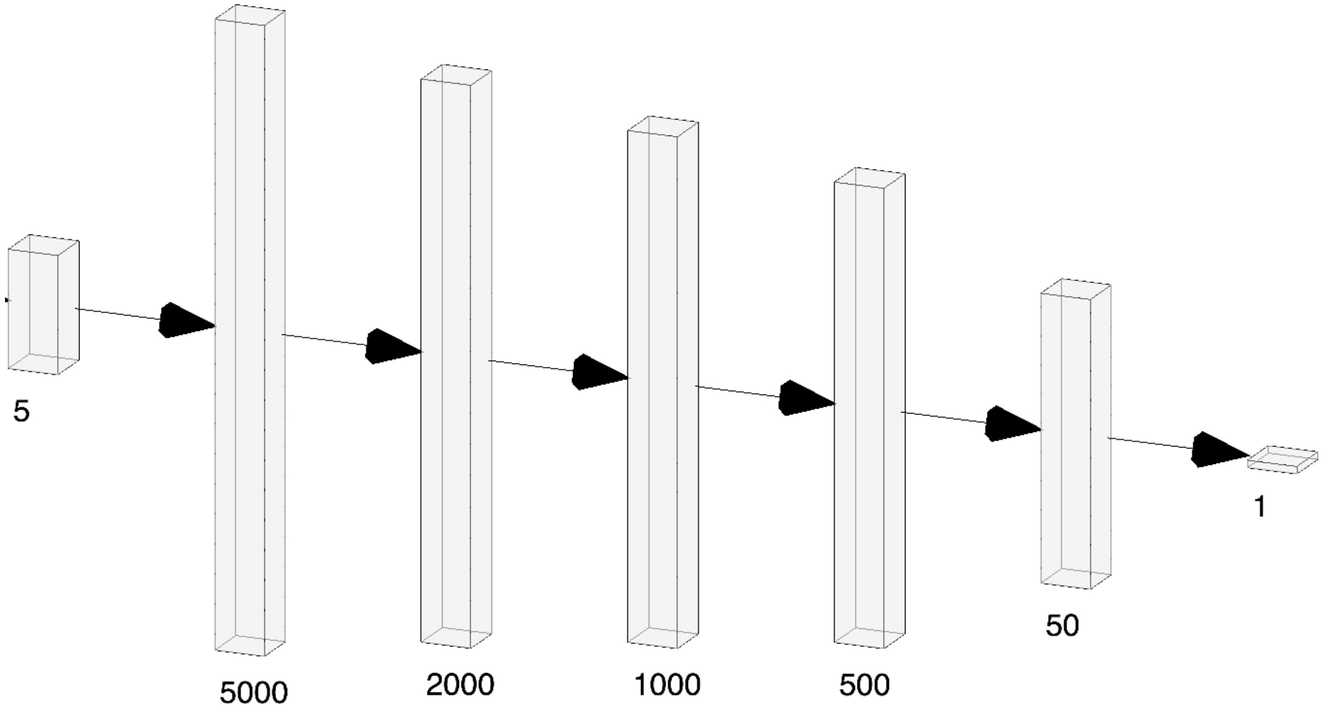
Trained with 5 features

['Annual vehicle miles traveled’,

'Annual vehicle miles travelled (cars)',
'Annual vehicle miles travelled (trucks)', 'Time Period’,

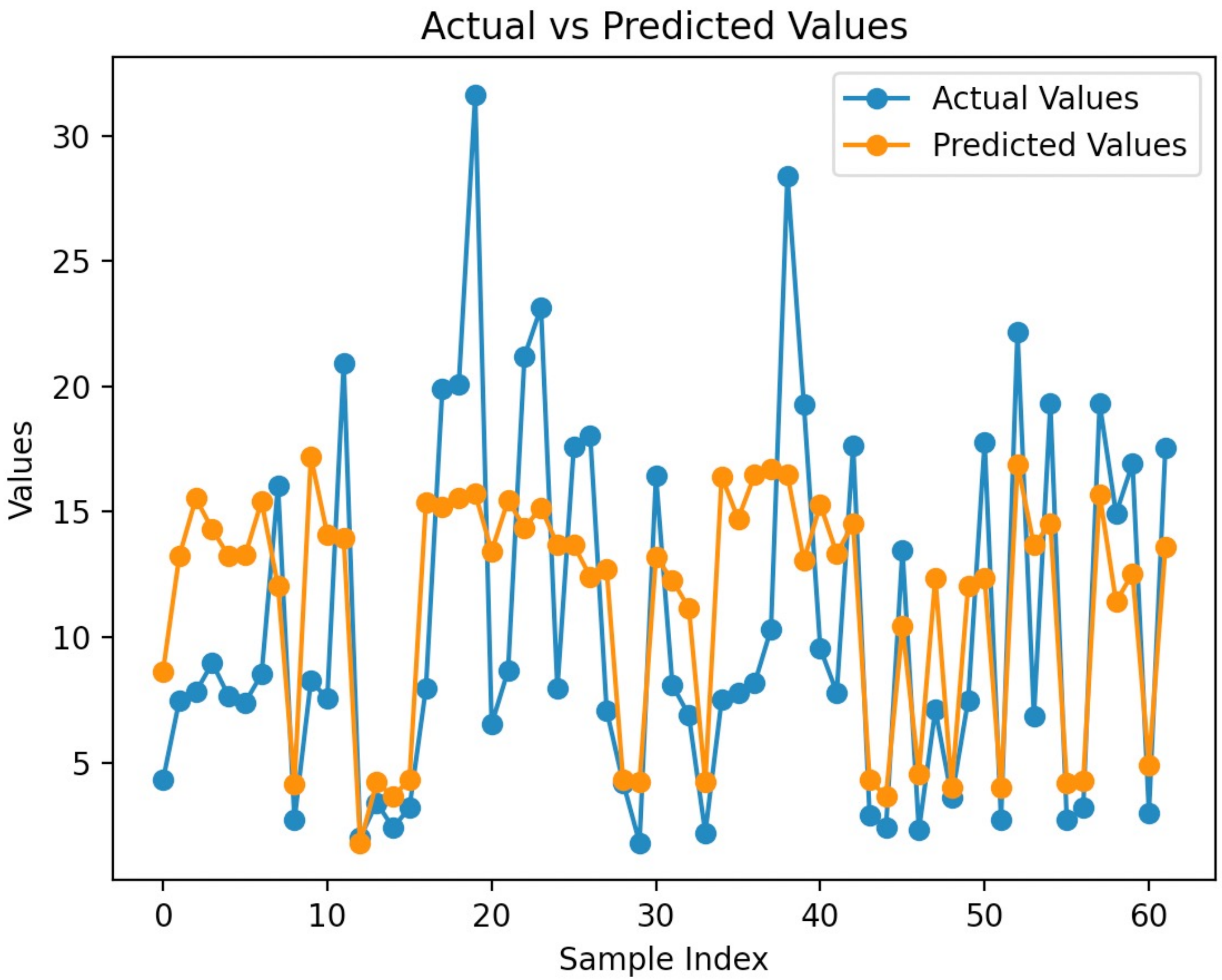
'Geo Place Name’]

SGD with Batch size of 8
Epochs: 200
Batch Normalisation after each layer
XAI Model: SHAP



Results – Predicting PM 2.5

- MSE: 31.00
- R-squared: 0.41



Pitfalls

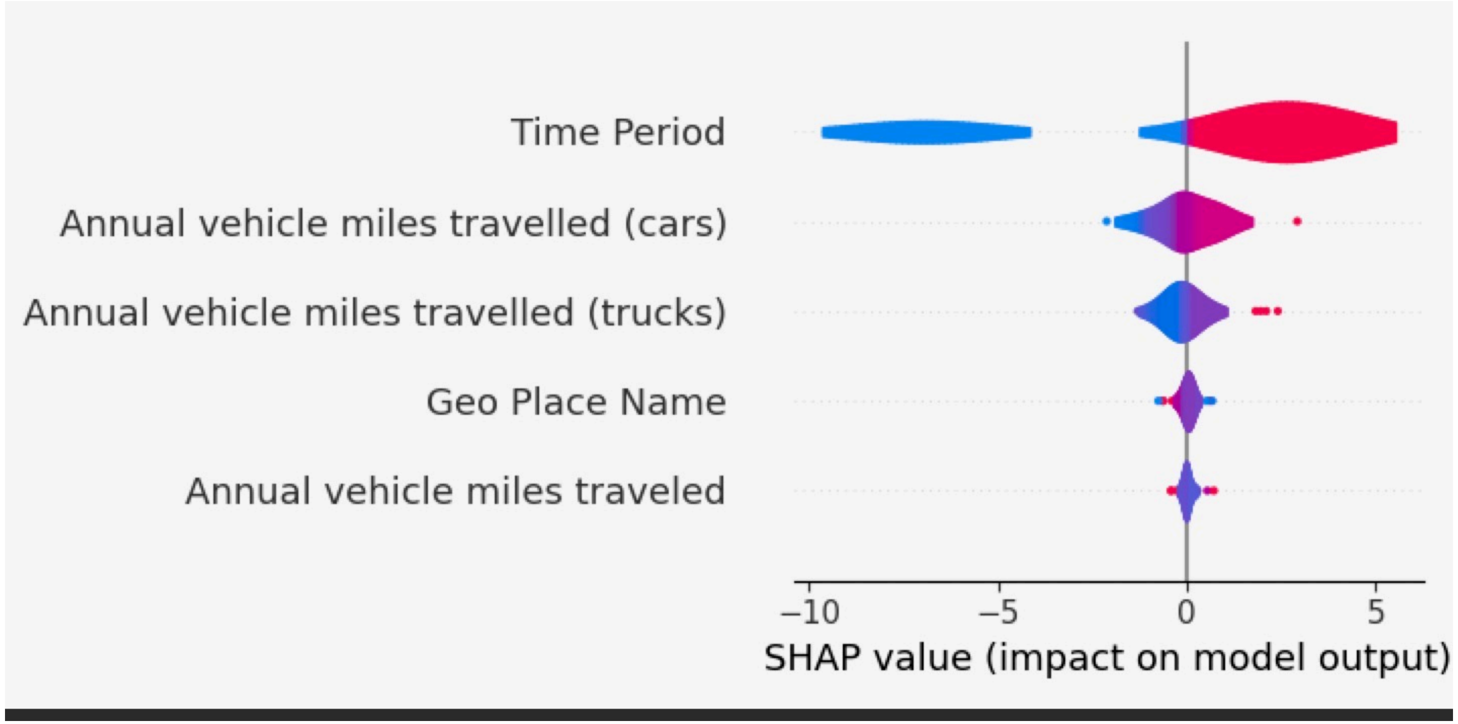
Really small and simple dataset – after preprocessing, we only had 300 datapoints

Lack of consistency in detectors used in the dataset i.E. Pm 2.5, NO2 levels, SO2 levels, asthma levels, air toxics, cardiac deaths

Only able to express 2 years, 2005 and 2016 because we were not able to spend enough time understanding the dataset

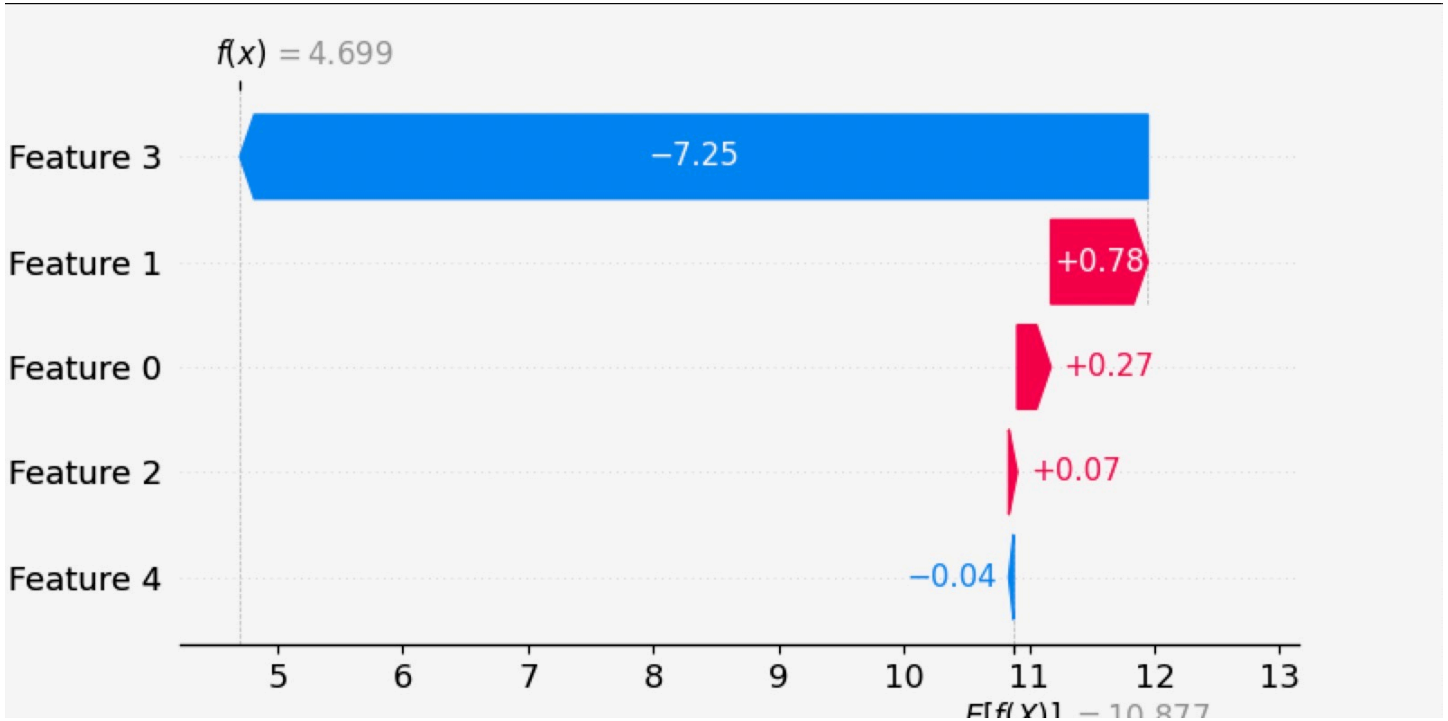
Spent a lot of time (in vain) trying to combine 2 different dataset in the hopes to achieve better data but similar issues arised

xAI Integration - SHAP



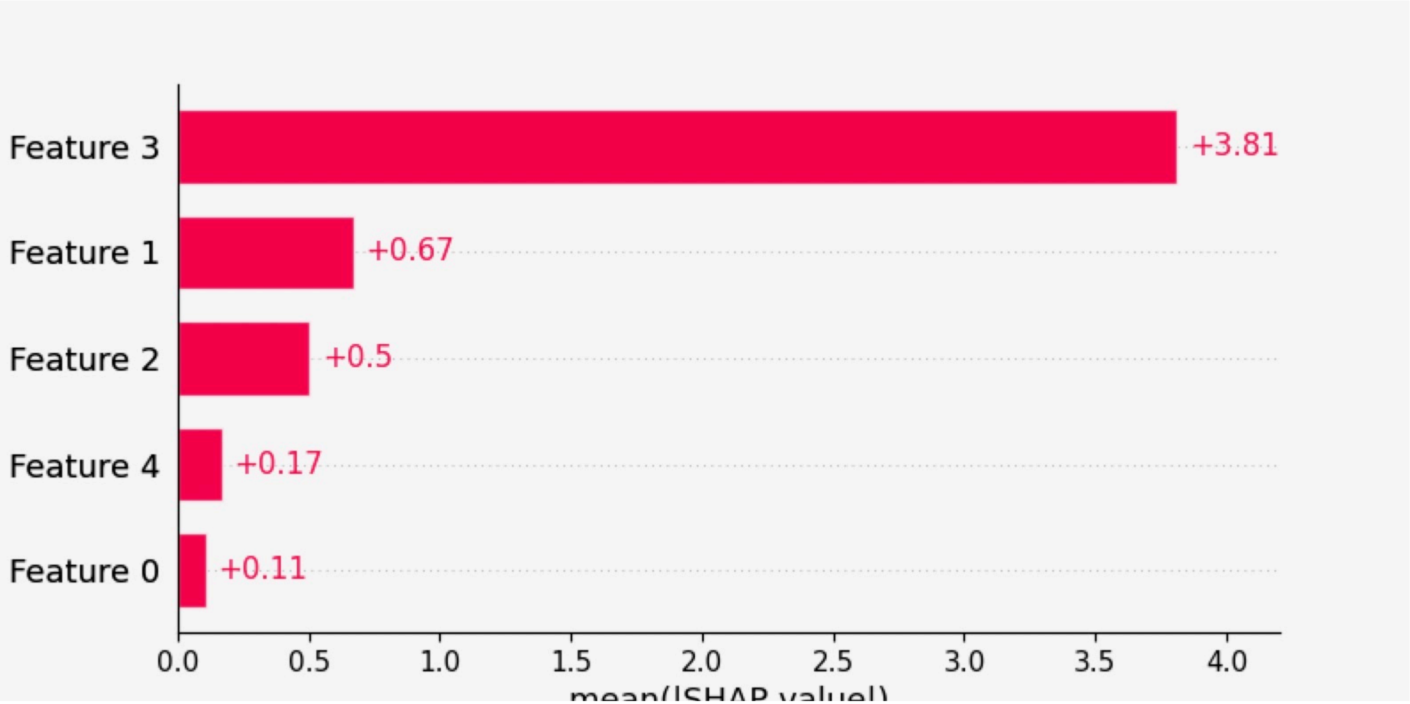
- This plot shows how higher and lower values of the feature will affect the result.
- The horizontal axis represents the SHAP value, while the color of the point shows us if that observation has a higher or a lower value, when compared to other observations.

- This plot shows us the magnitude of the SHAP value for each feature.
- The sum of all the SHAP values equals the difference between the prediction f(x) and the expected value E[f(x)].



Feature Legend:
0: Annual Vehicles Miles Travelled
1: Annual Vehicles Miles Travelled (Cars)
2: Annual Vehicles Miles Travelled (Trucks)
3: Time Period
4: Geo Place Name

- Once our SHAP values are computed, we inputted them in a global feature importance plot
- The global importance of each feature the mean absolute value for that feature, regardless of negative/positive influence
- Features are ordered based on SHAP value (model impact)



Feature Legend:
0: Annual Vehicles Miles Travelled
1: Annual Vehicles Miles Travelled (Cars)
2: Annual Vehicles Miles Travelled (Trucks)
3: Time Period
4: Geo Place Name

Thank You