

Transit Ridership Forecasting

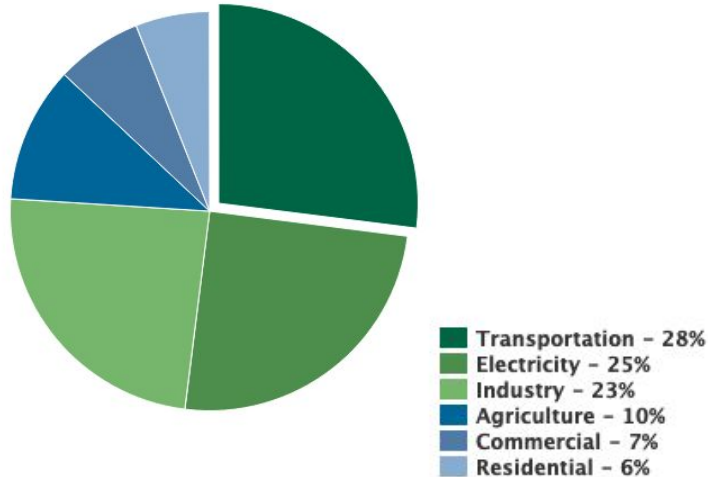
Liam Dubay & Sebastian Lopez



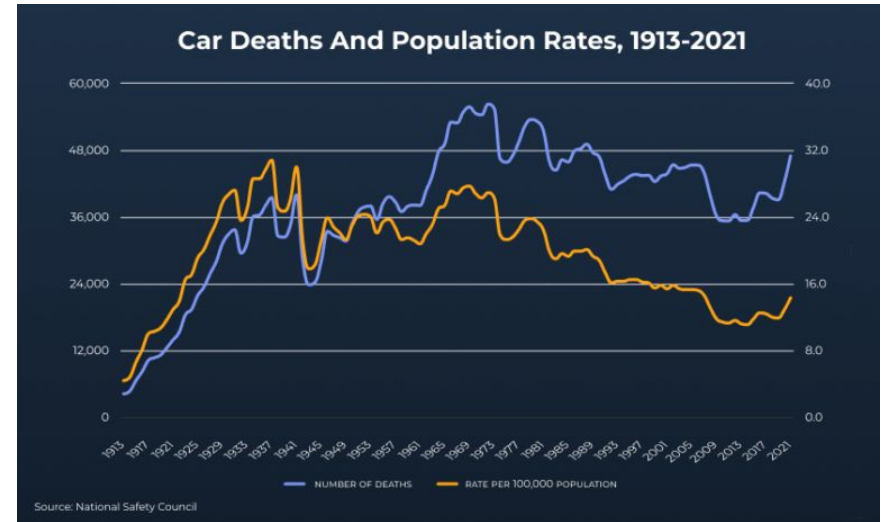
Investments in Public Transit are Crucial, but Costly

Greenhouse gas emissions

2022 U.S. GHG Emissions by Sector



Traffic deaths & injuries



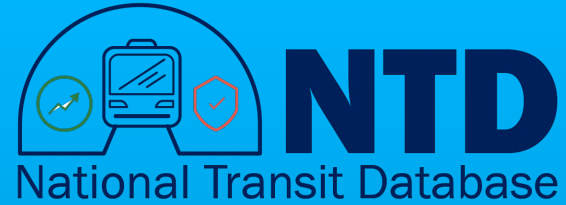
Our Goals:

- Forecast transit ridership for U.S. cities
- Identify key factors for improving ridership

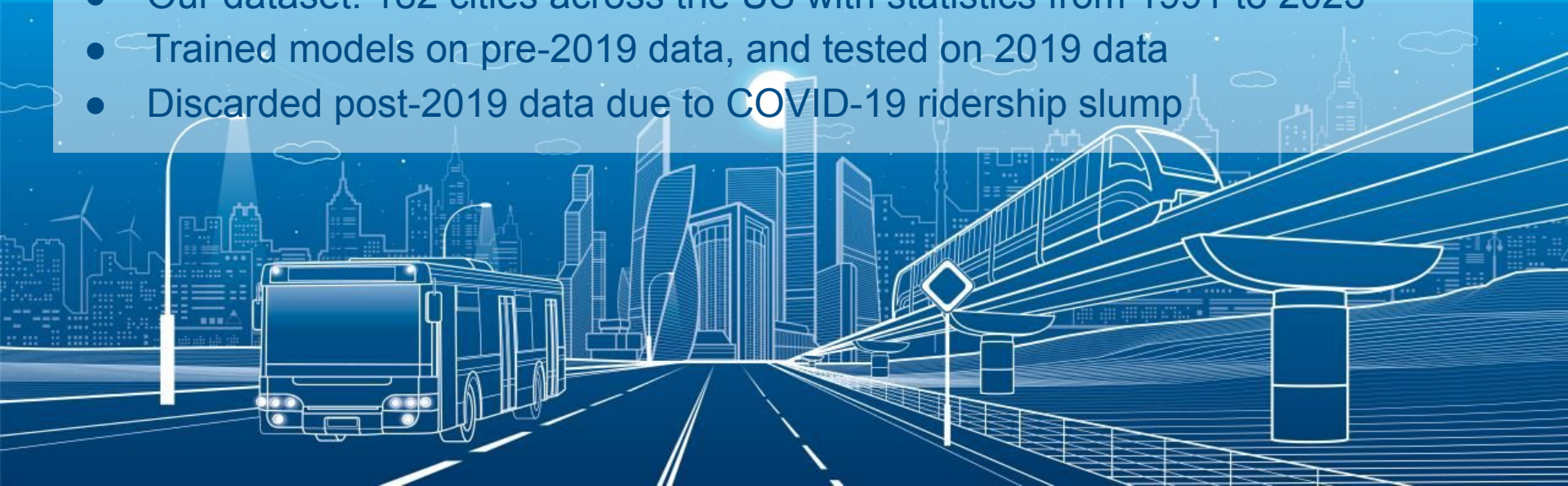


Illustration by Kazimir Iskander

The National Transit Database

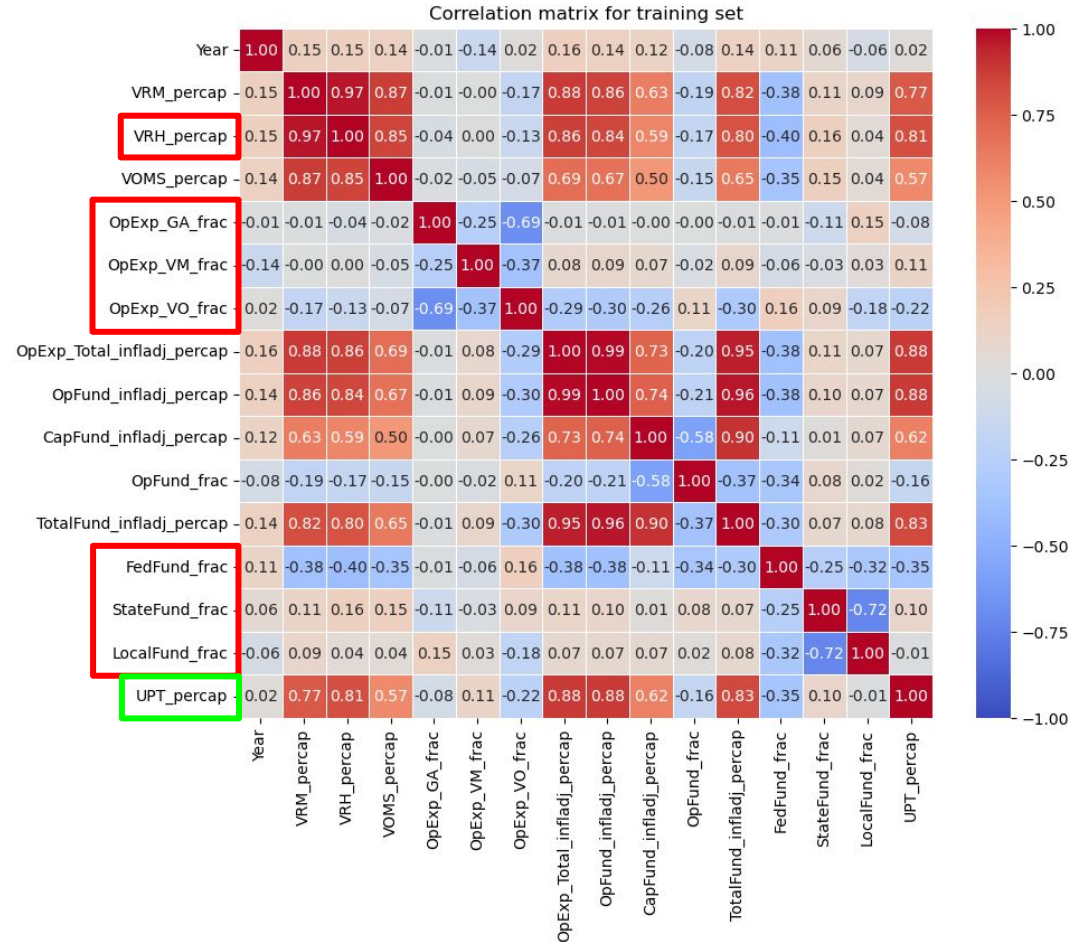


- Records the financial, operating, and asset condition of transit systems throughout the United States
- Our dataset: 182 cities across the US with statistics from 1991 to 2023
- Trained models on pre-2019 data, and tested on 2019 data
- Discarded post-2019 data due to COVID-19 ridership slump



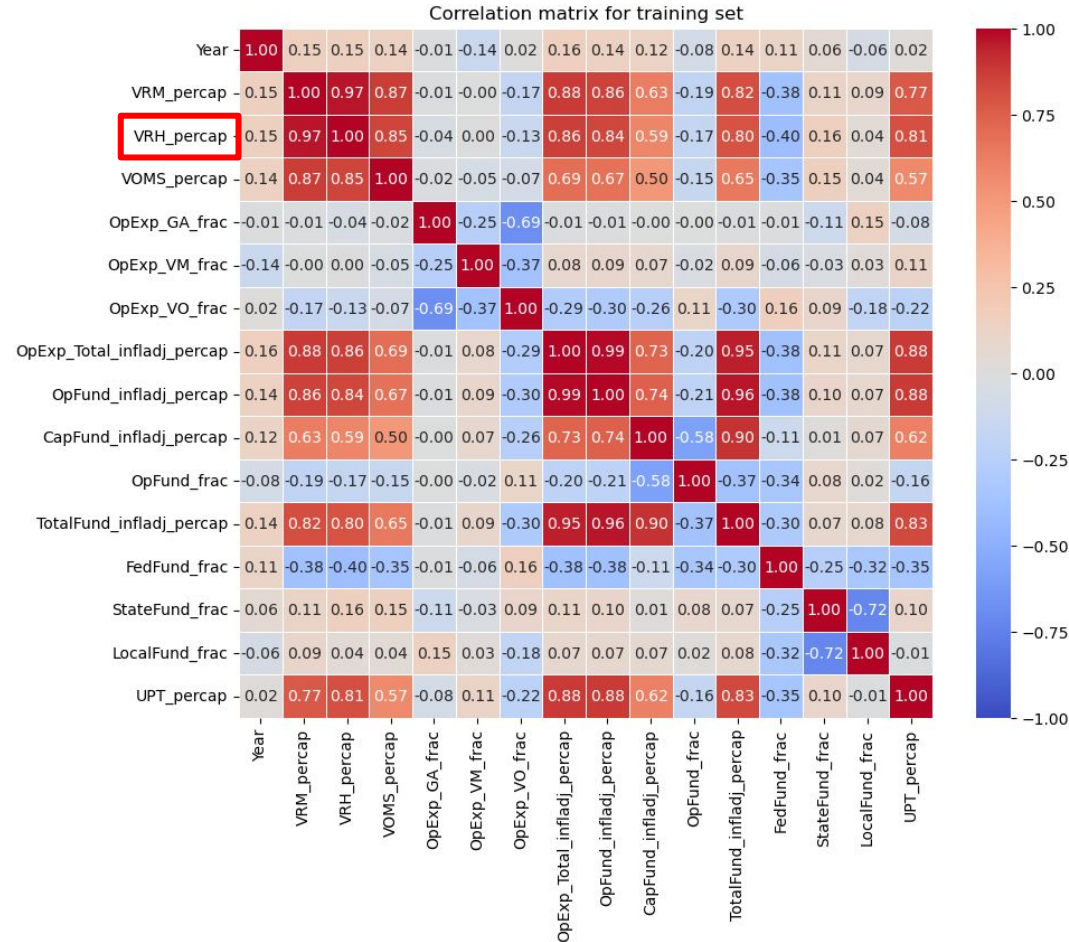
The Factors at Play

- Statistics on vehicles, operational expenses, funding sources, and ridership
- Removed features with strong multicollinearity



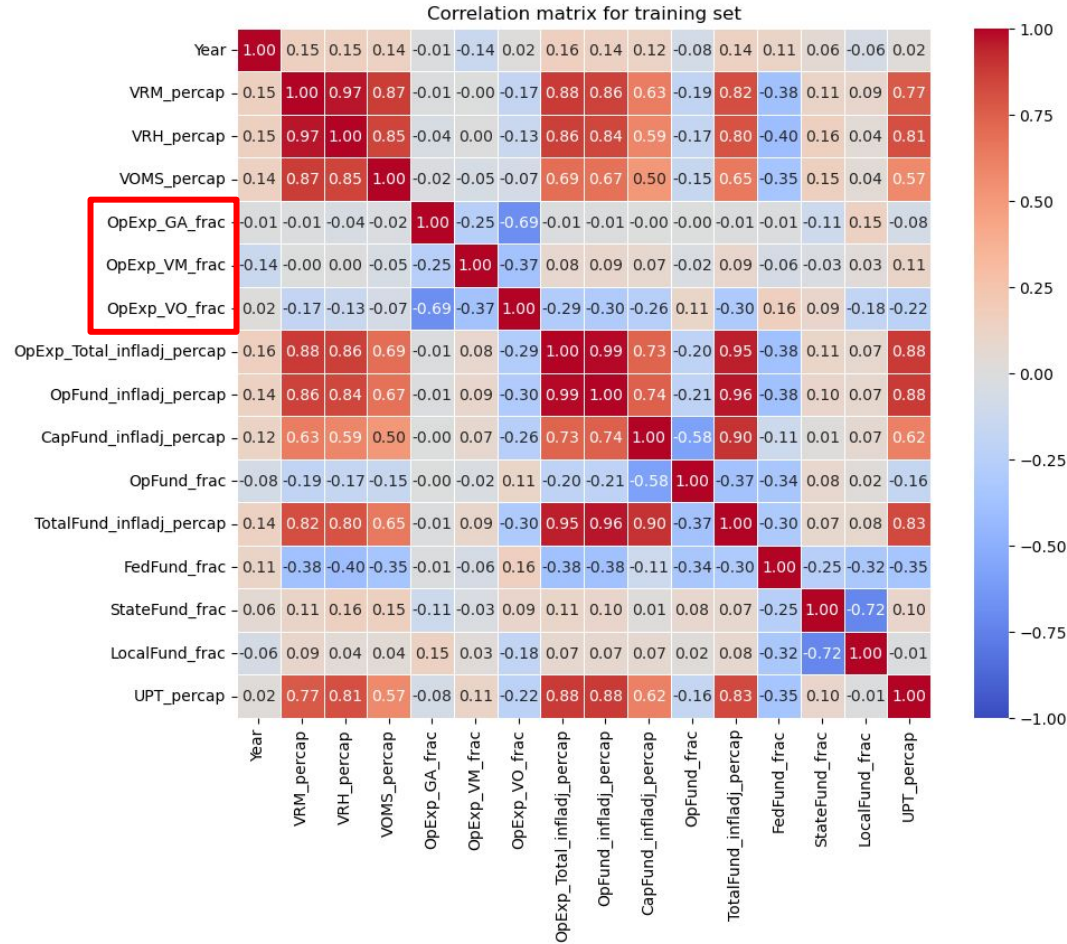
The Factors at Play

How many hours are all vehicles collecting revenue (VRH)?



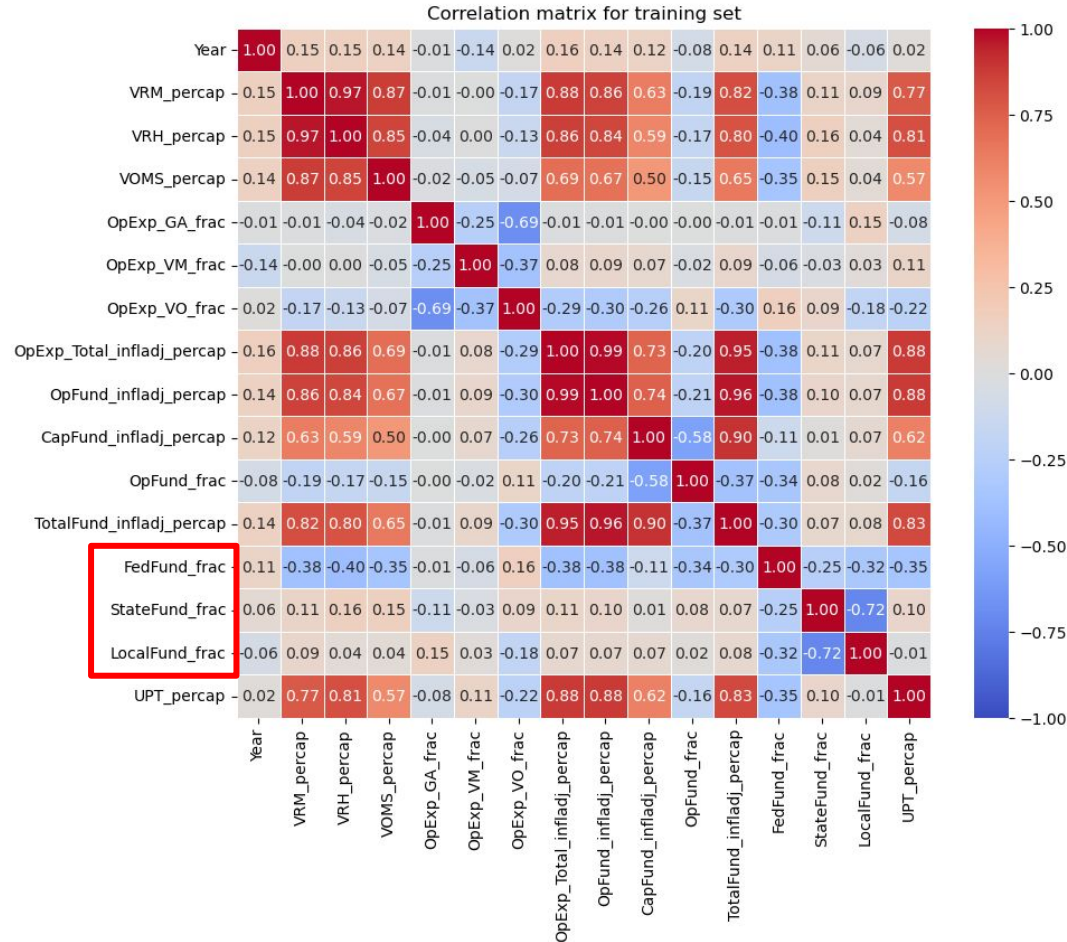
The Factors at Play

What fraction of funds are spent on administration, vehicle maintenance, & vehicle operation?



The Factors at Play

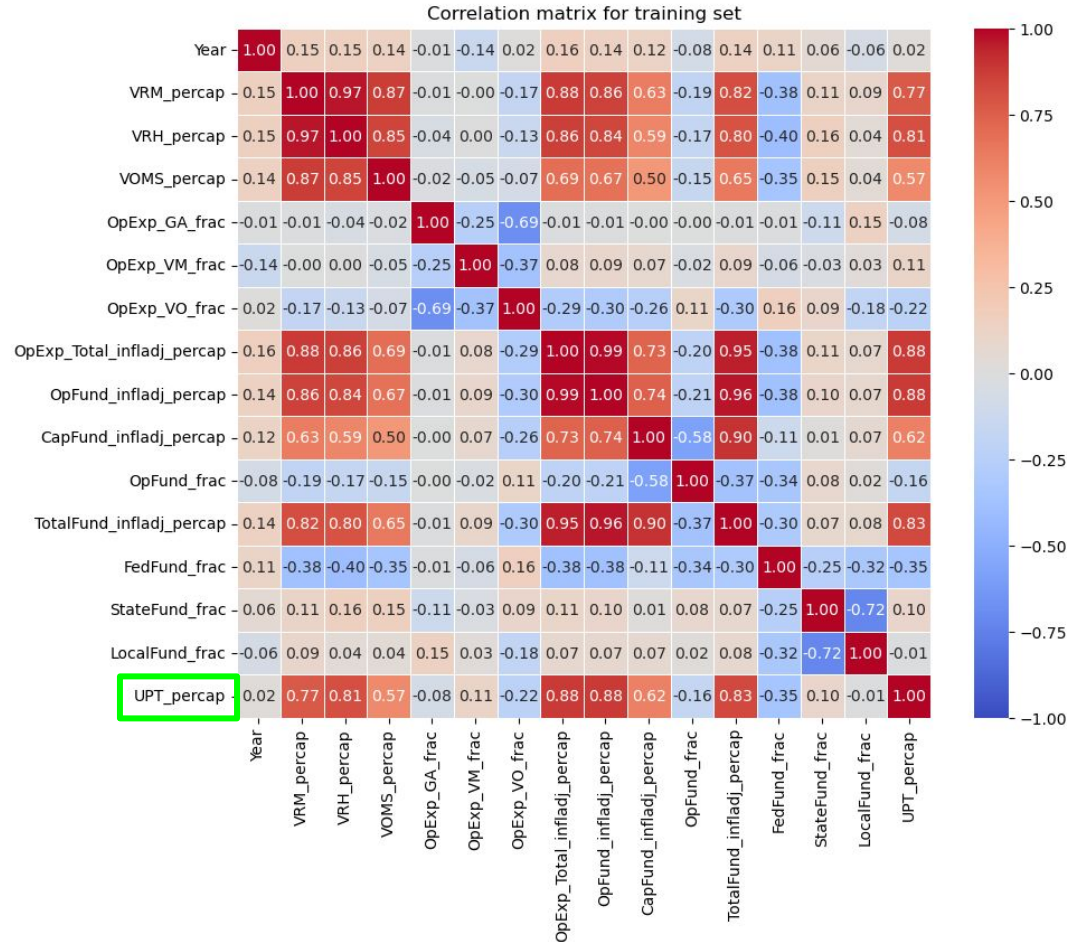
What fraction of funding comes from federal, state, and local governments?



The Factors at Play

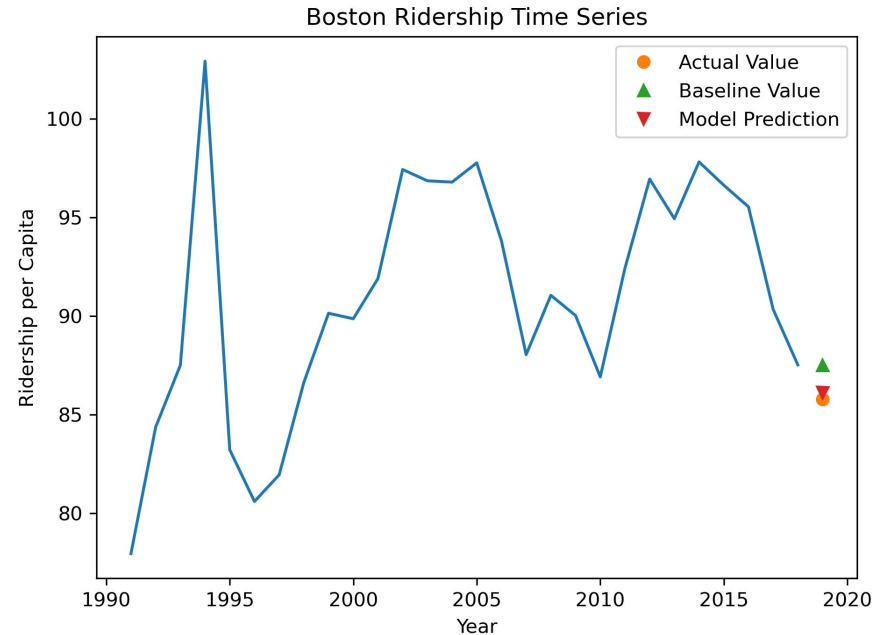
Forecast variable:

Ridership is measured by annual
Unlinked Passenger Trips (UPT)
per capita



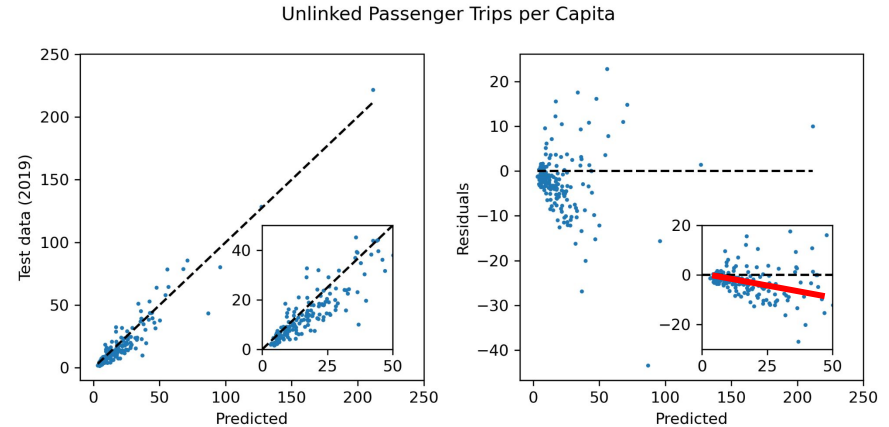
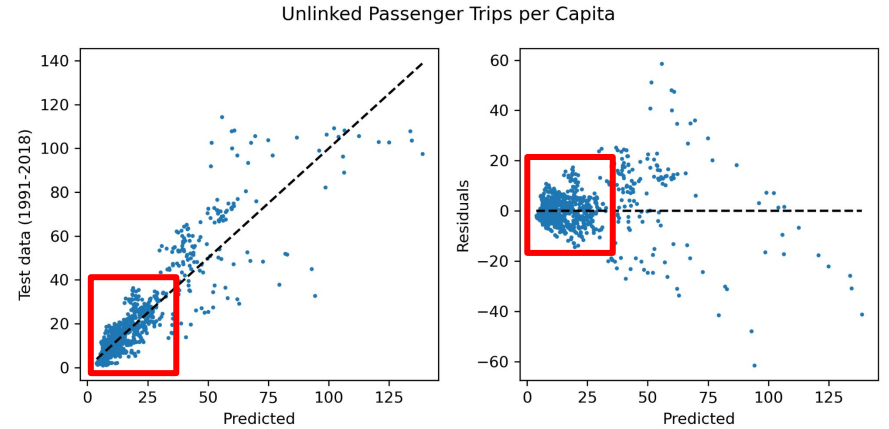
A City by City Model

- Unique model trained for each city
- Lasso regression eliminates extraneous features
- Example: Boston, MA
 - Features with non-zero coefficients were Vehicle Maintenance Expenses, Federal, and State Funding.
 - Our model provides a better prediction than the baseline
- Bespoke model out-performs a naive forecast for only 36/182 cities



All-Cities Model

- Trained on 1991-2018 data for 80% of cities
- Cross-validation: XGBoost out-performed baseline and multiple linear regression
- Tests well on pre-2019 data
 - Except high-ridership cities
- Residuals of 2019 data reveal un-modeled trend

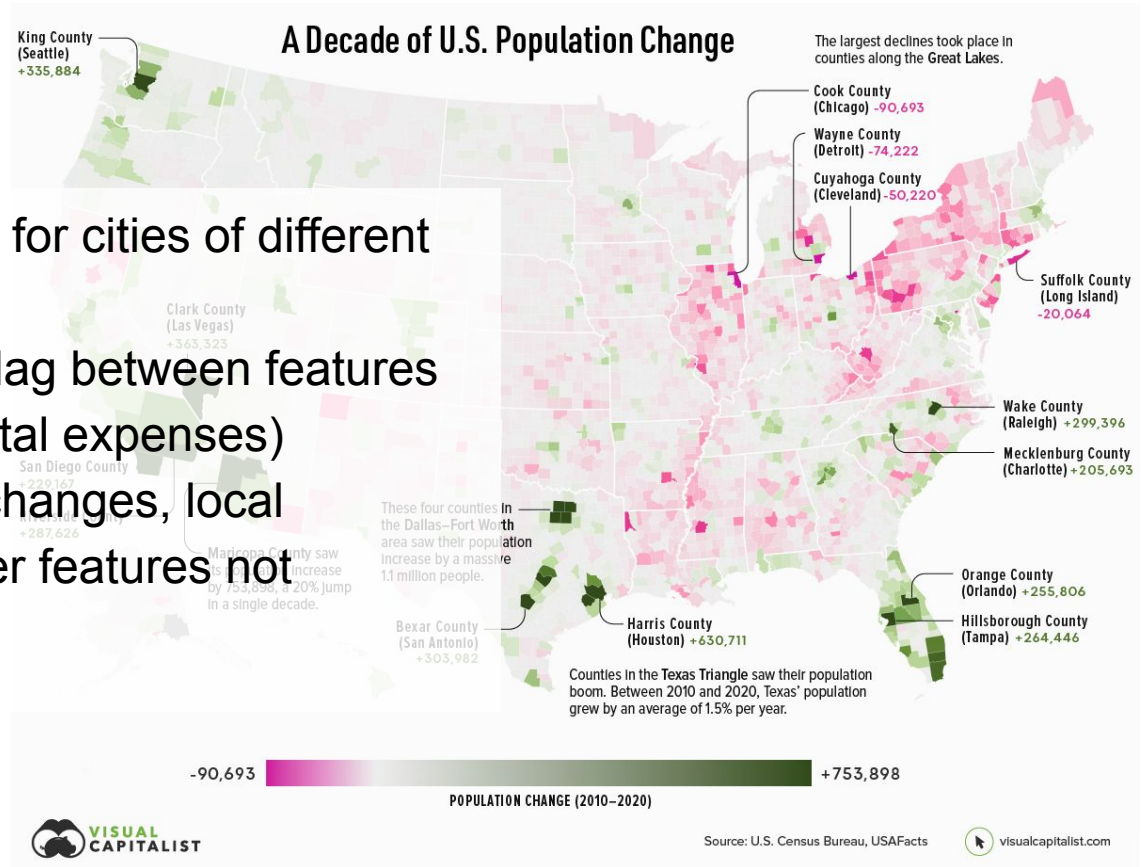


Insights and Shortcomings

- The City by City Model provided useful insights into where transit agency can dedicate resources.
- However, it performed worse than a naive forecast for most cities.
- We may be missing important factors for understanding ridership.
- The All-City Model performed better due to a much larger training set.
- Additionally, it showed trends across different cities.
- However, large residuals exist for some larger cities.

Future Work

- Develop unique models for cities of different size categories
- Better account for time lag between features and ridership (e.g., capital expenses)
- Investigate population changes, local developments, and other features not included in the NTD



Thank You!

