UNDERSTANDING, VISUALIZING, AND PREDICTING TTC BUS DELAYS

Team:
PandasExpress

CME538 – Data Science
Prof. Sebastian Goodfellow

Introduction | Data Wrangling | Data Visualization | Geospatial Visualization | Predictive Model | Conclusion

10 Reasons Why You Should Use Public Transport

Transit Savings Grow as Auto Costs and Gas Prices

Transit in Toronto has never felt so broken — it's time for the TTC board of directors to demand better

By the TTC's own measure, 86 per cent of riders are getting substandard service. Just 31 of 160 tracked bus and streetcar routes met its definition of "on time."

8 Benefits of Public Transportation

Introduction | Data Wrangling | Data Visualization | Geospatial Visualization | Predictive Model | Conclusion

Data Collection

TTC Bus Stops

TTC Bus delays (2014-2022)

TTC Bus Routes shape files

Weather data

Toronto JSON

Toronto Housing Price map

Performance of Bus Service and Delays

Results

Data Analytics Visualization

GeoSpatial Visualization

Predictive machine model

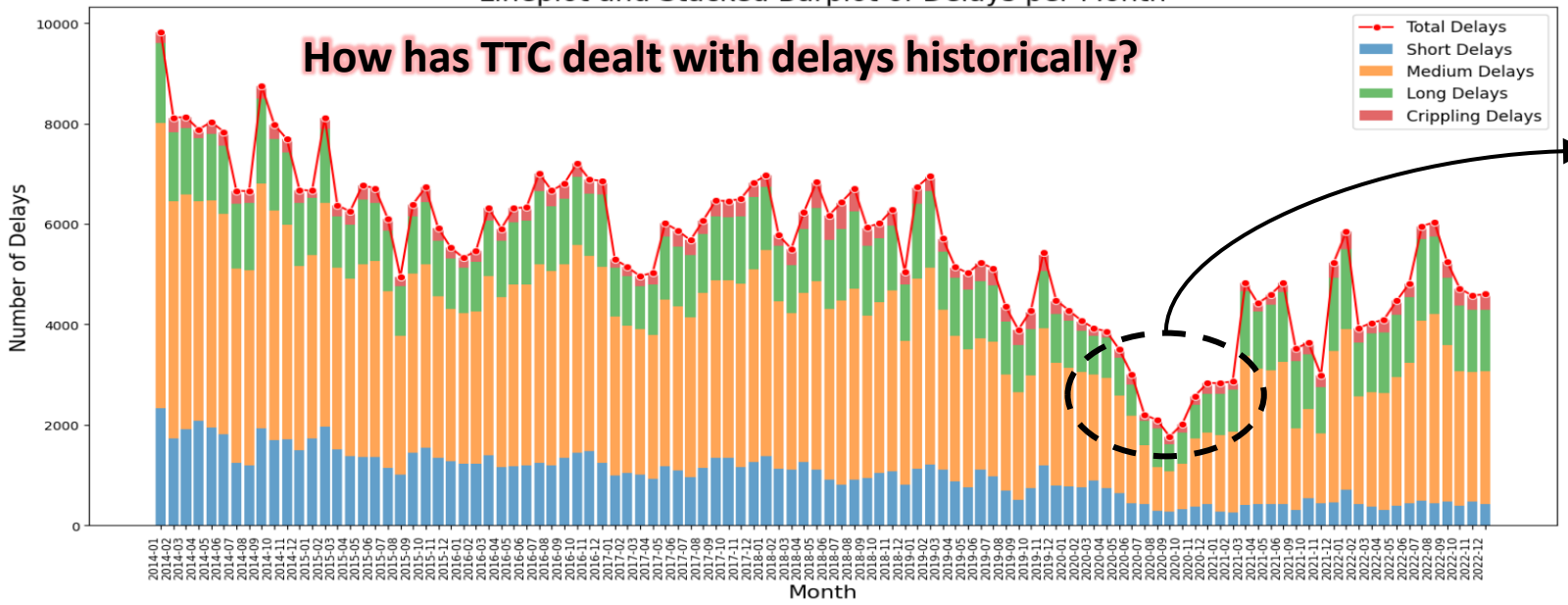
Data Processing and Visualization

GitKraken

Teamwork Enhanced by Kraken

Lineplot and Stacked Barplot of Delays per Month

How has TTC dealt with delays historically?

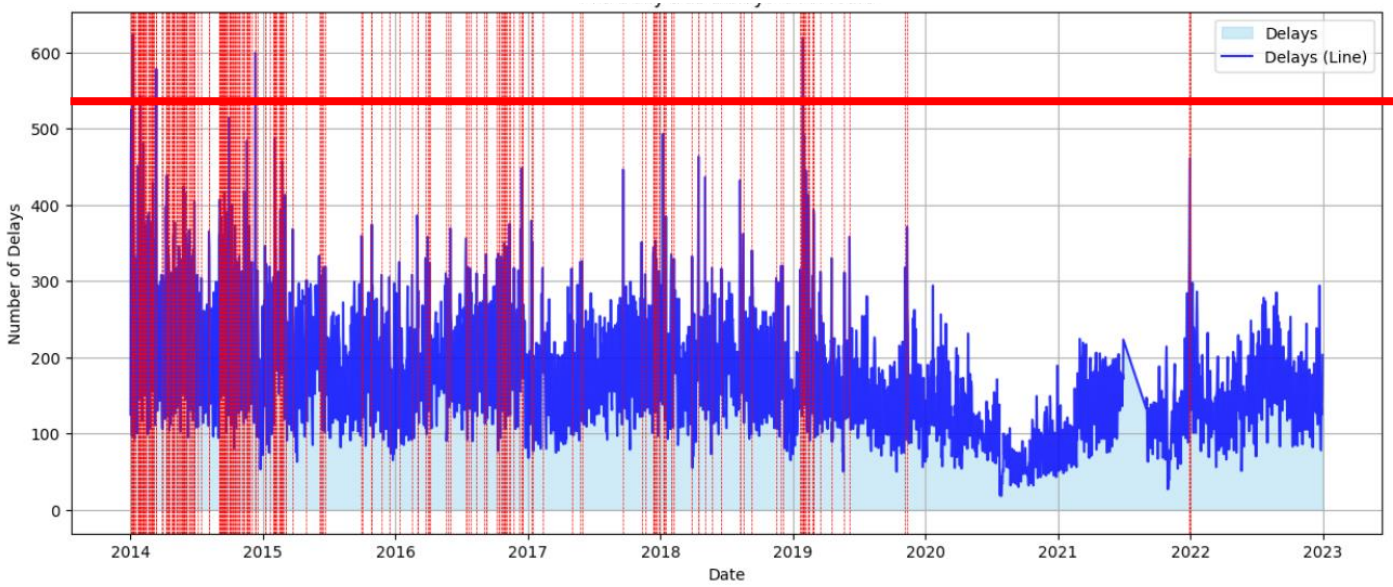


Toronto

TTC facing $92M monthly shortfall, plummeting ridership due to COVID-19

Presto taps have dropped by 86% compared to pre-pandemic levels

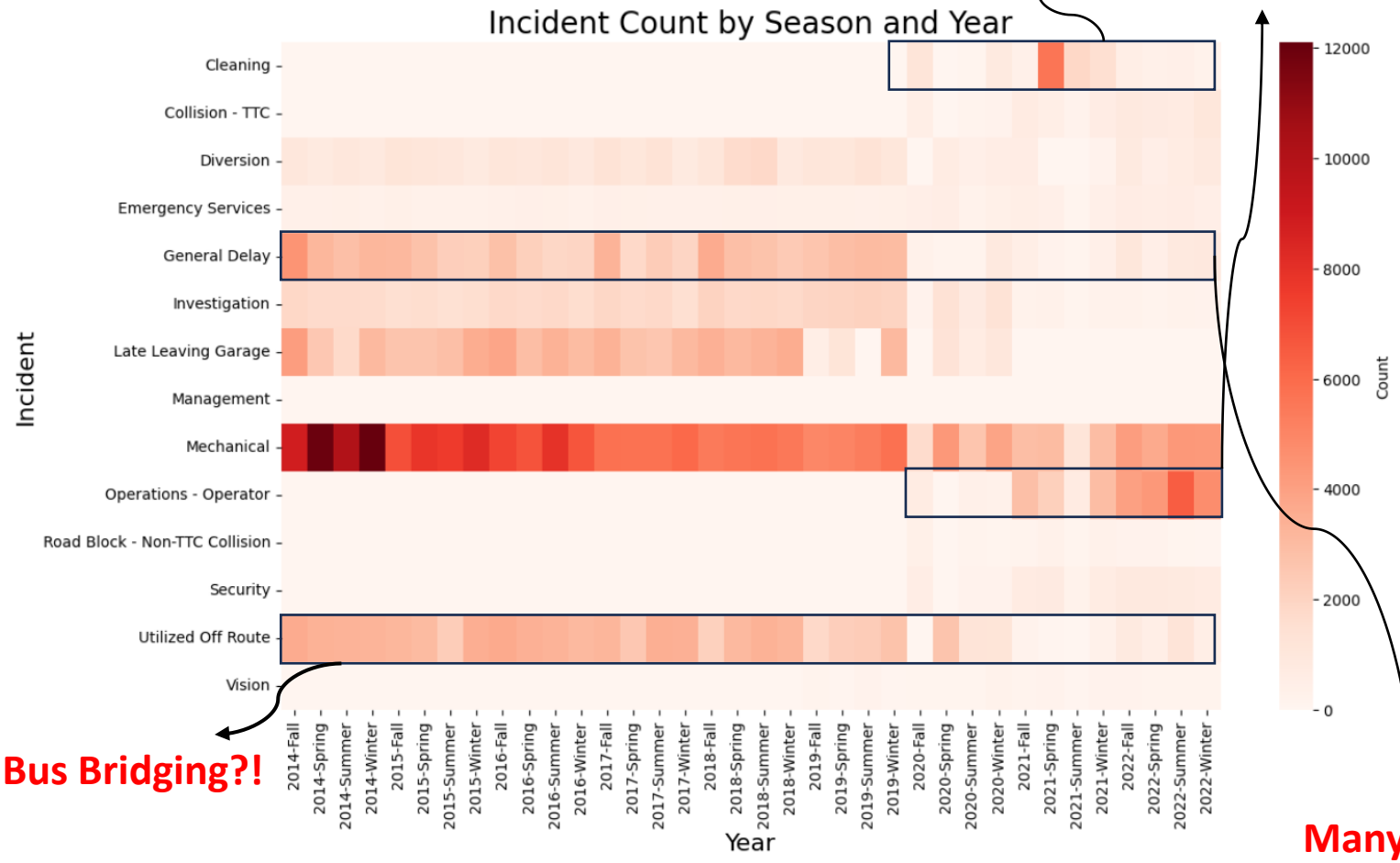
Nick Boisvert · CBC News · Posted: May 14, 2020 3:09 PM EDT | Last Updated: May 14, 2020



Red lines show days with more than 300 recorded delays in TTC bus services.

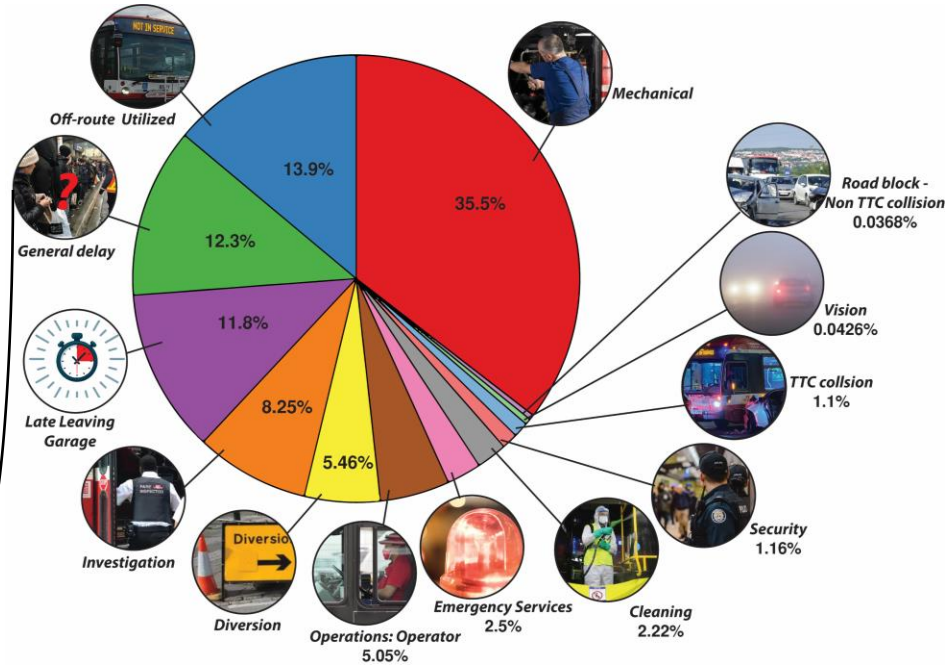
How has TTC dealt with the Delay causes?

COVID-19 Impact?!



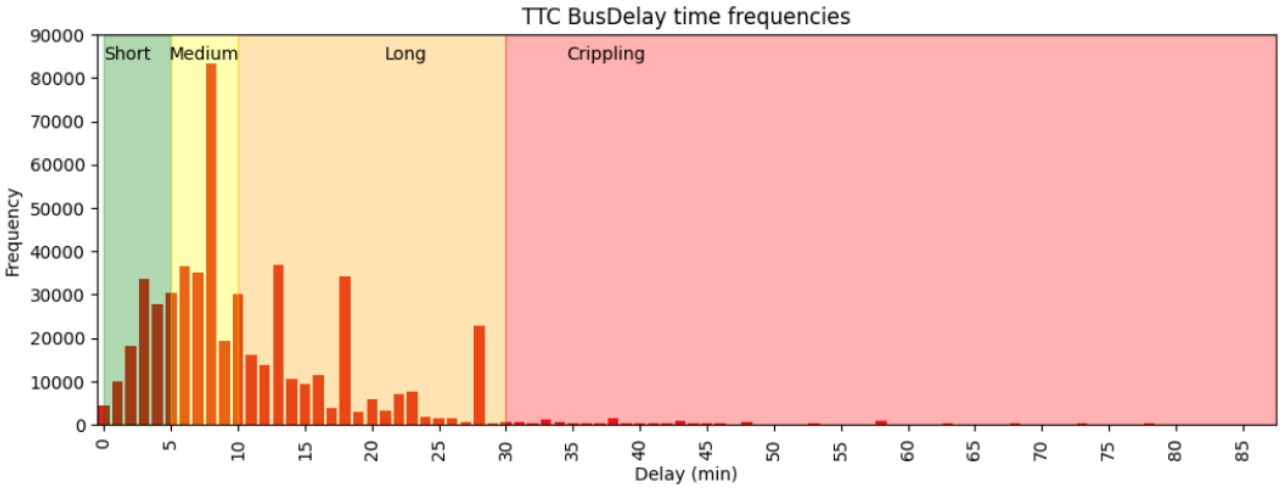
Bus Bridging?!

What Caused the Delays?!

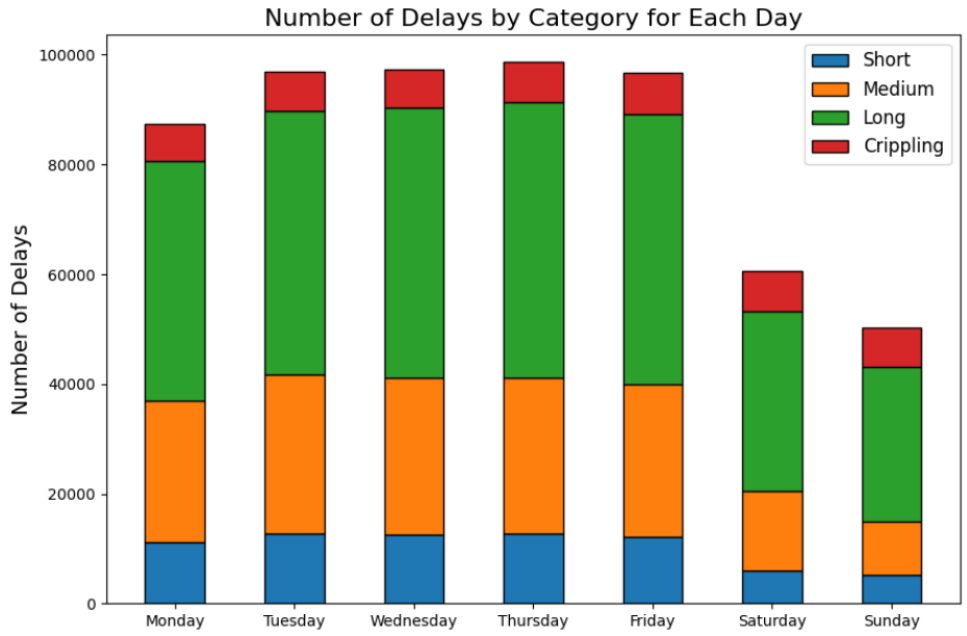


Many Incidents are not clear!

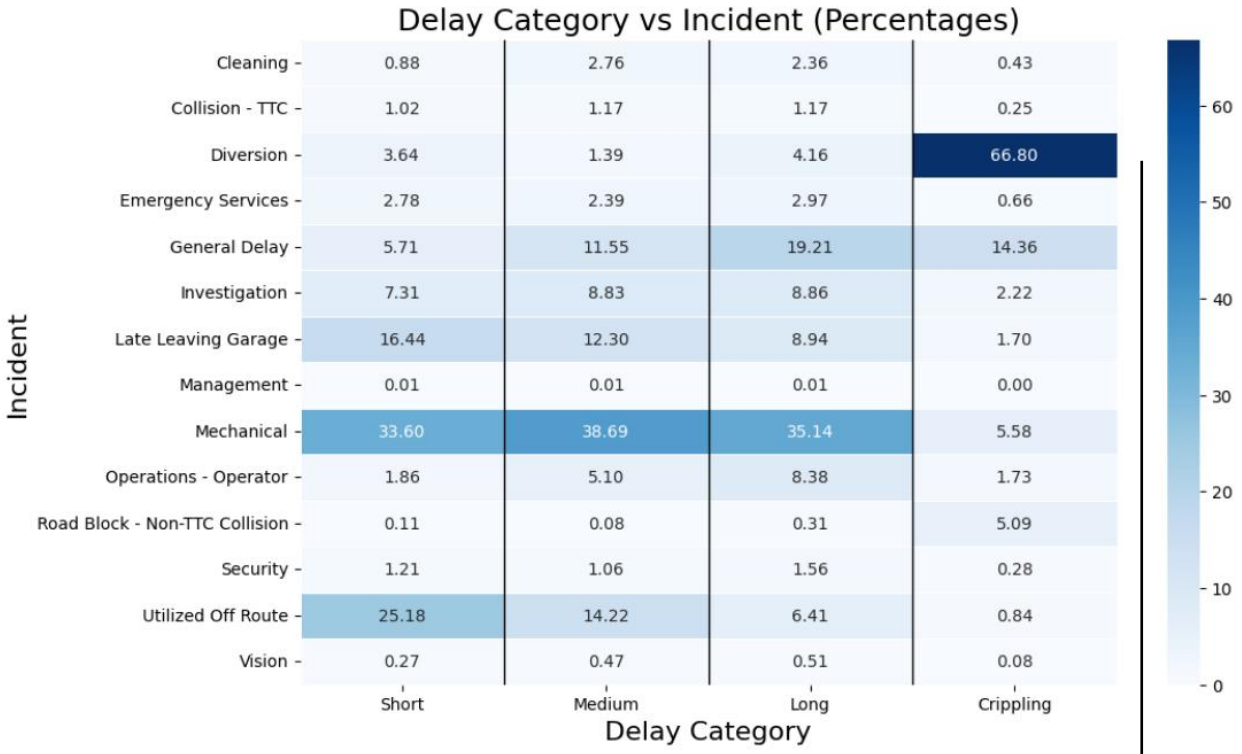
How frequent are these delays?



Delays frequency per weekdays

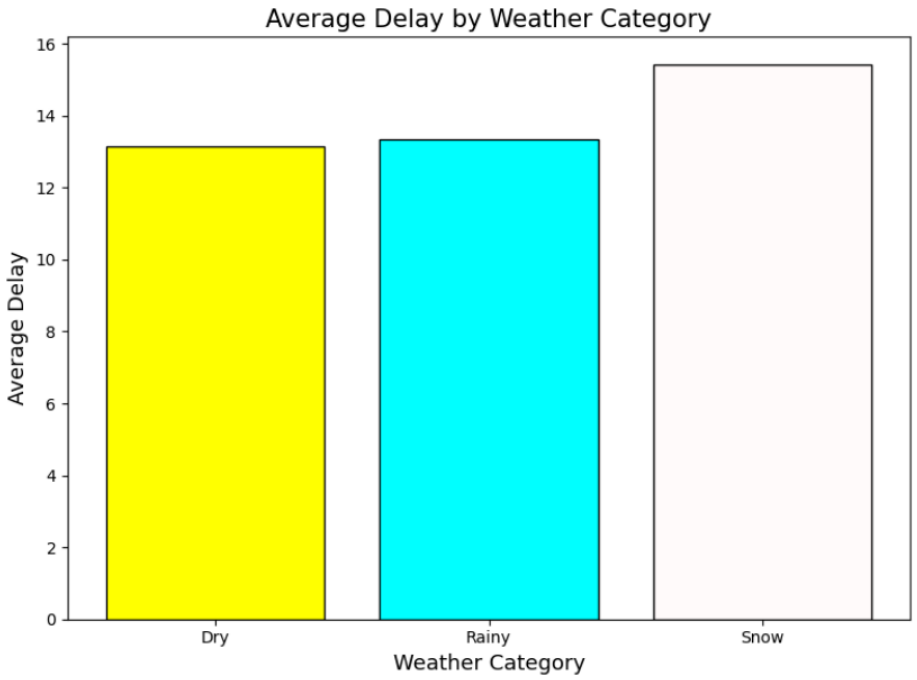


What caused each category of the delays?

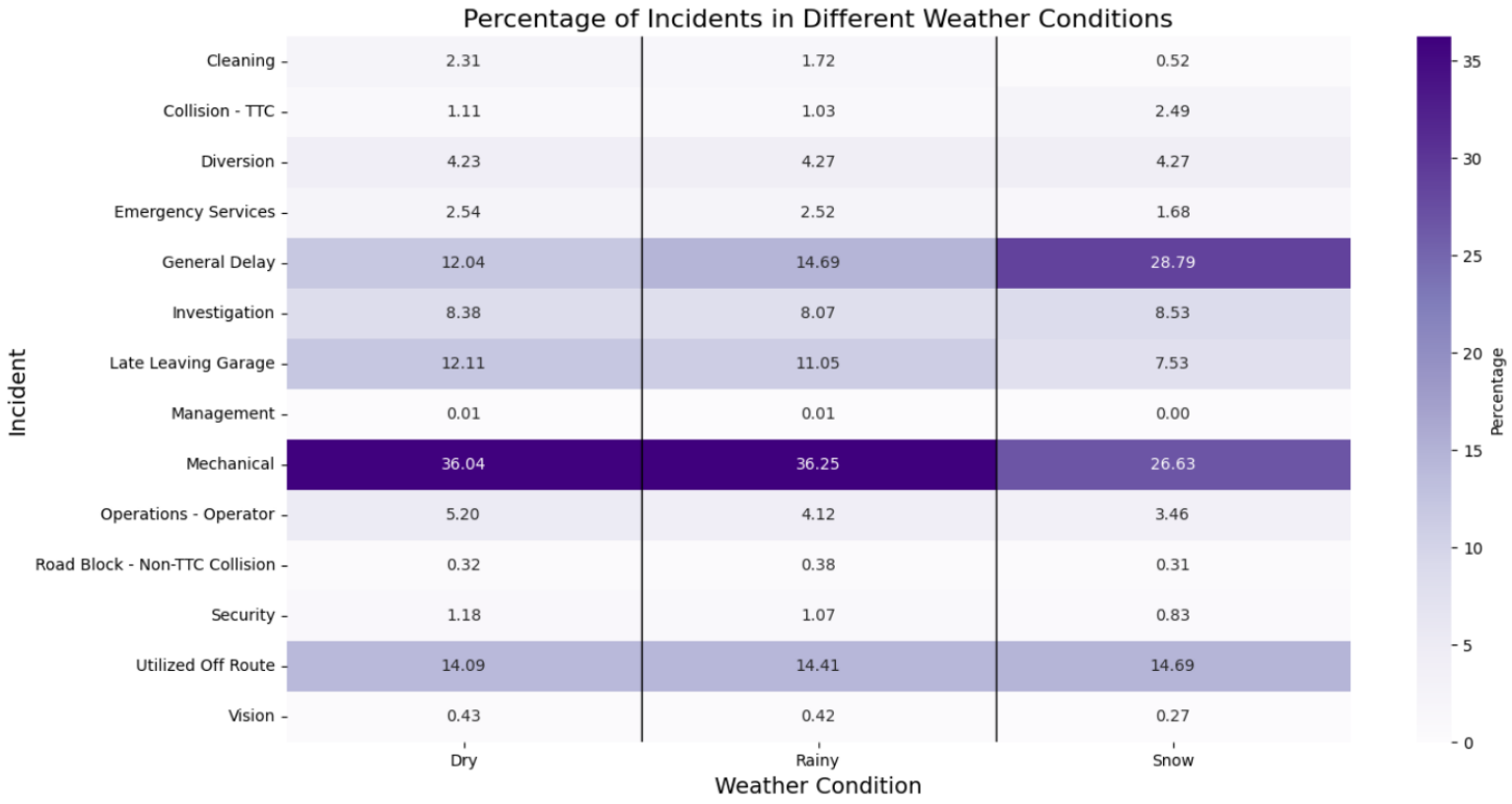


Nearly 70% of Crippling delays are caused by Diversions which were not controlled by TTC

Delays and Weather



Weather conditions and Incidents



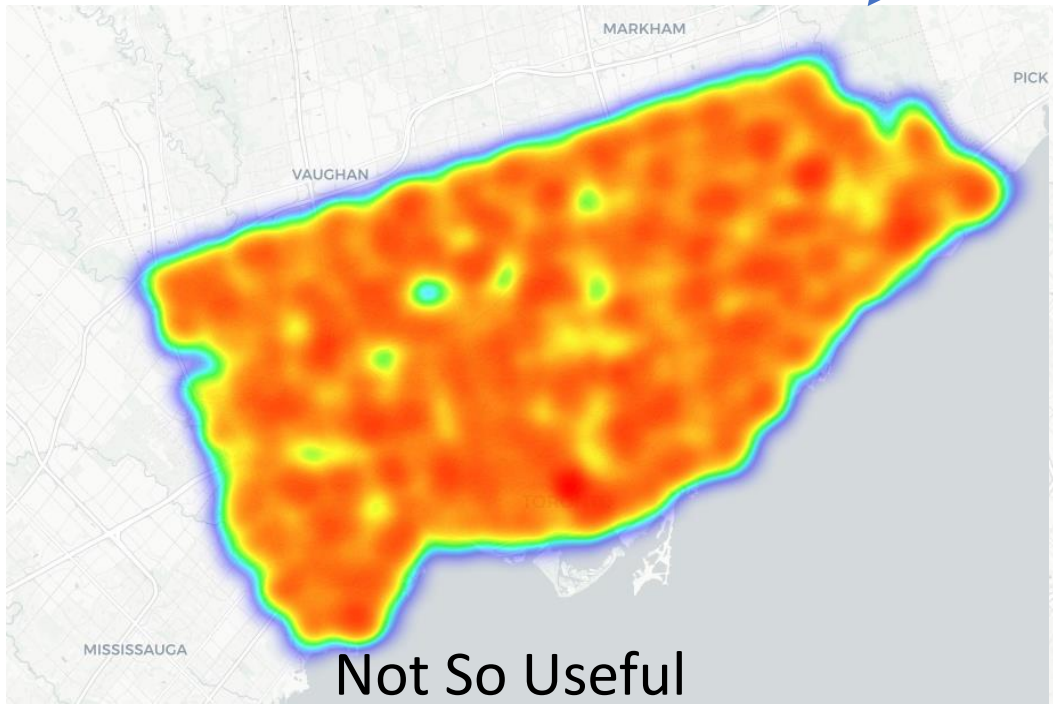
Why do we need data visualization?

ROUTE	geometry
Airport Rocket	MULTILINESTRING ((-79.60701 43.68596, -79.6070...
Alness	LINESTRING (-79.46349 43.75030, -79.46443 43.7...

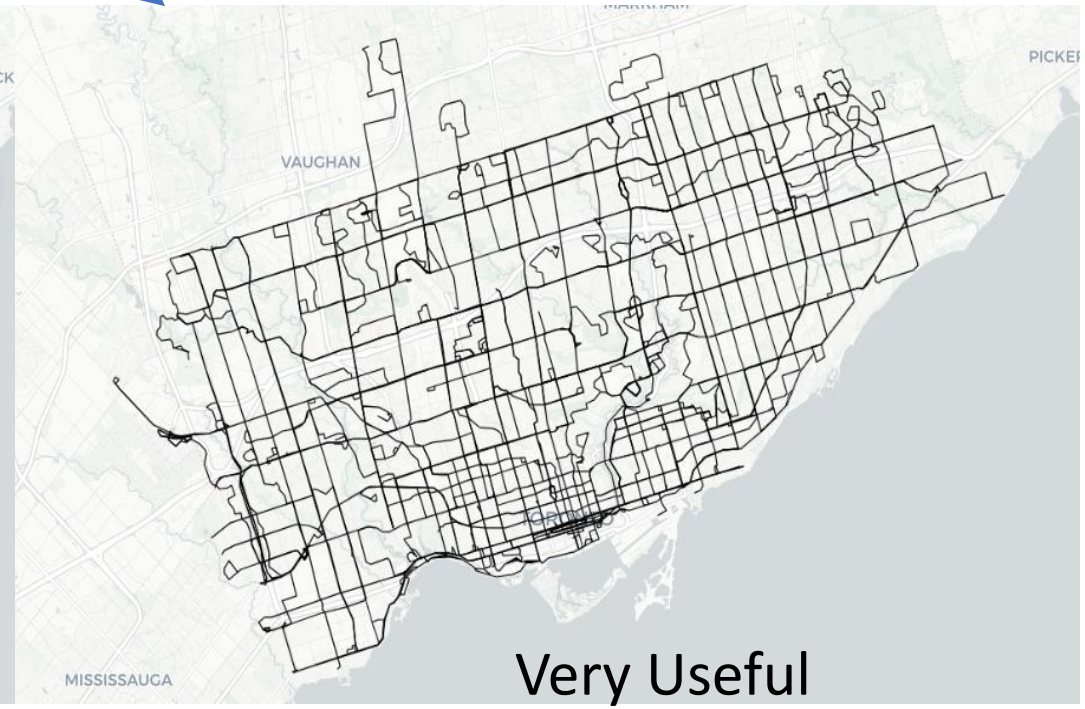
```
toronto_map = folium.Map(location=[43.6426, -79.3871],
                         tiles='cartodbpositron',
                         zoom_start=11)

min_value = min(trial3['Route'])
max_value = max(trial3['Route'])
norm = plt.Normalize(min_value, max_value)
cmap = plt.cm.viridis  # Choose any other colormap from Matplotlib
```

Gibberish to majority of TTC users

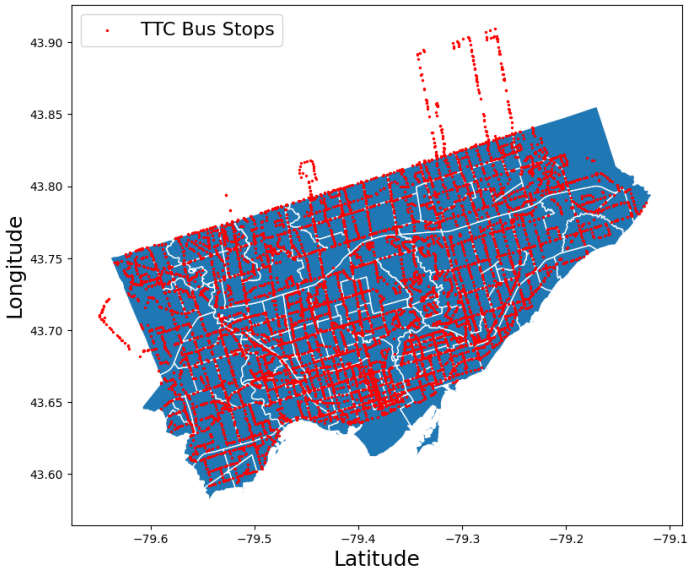


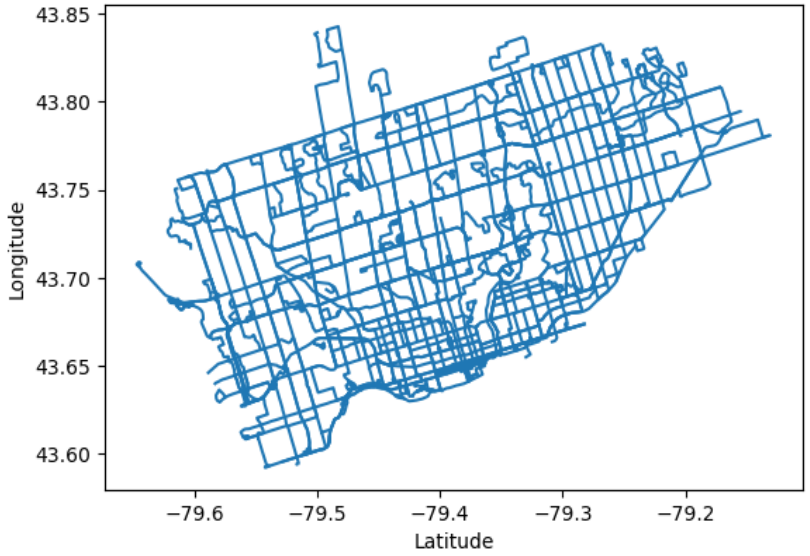
Not So Useful

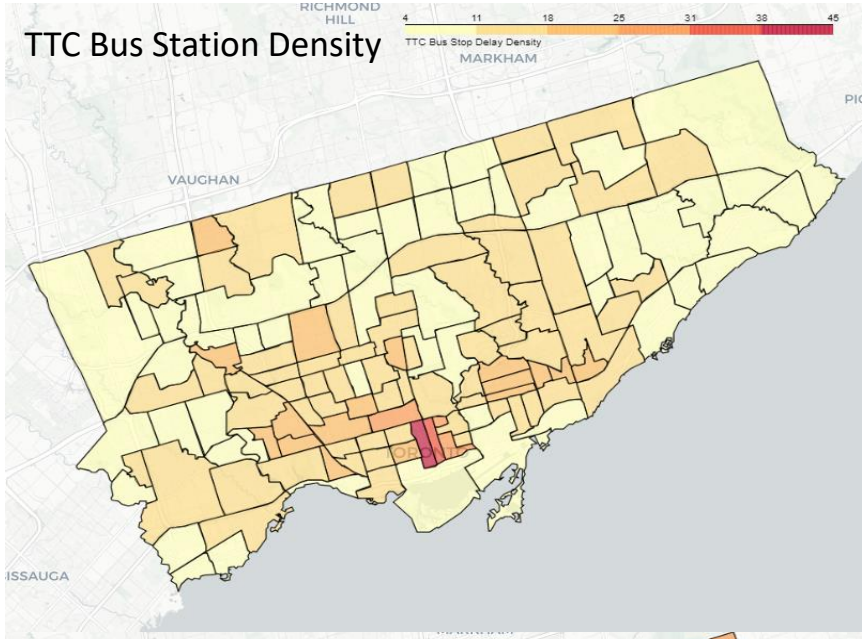


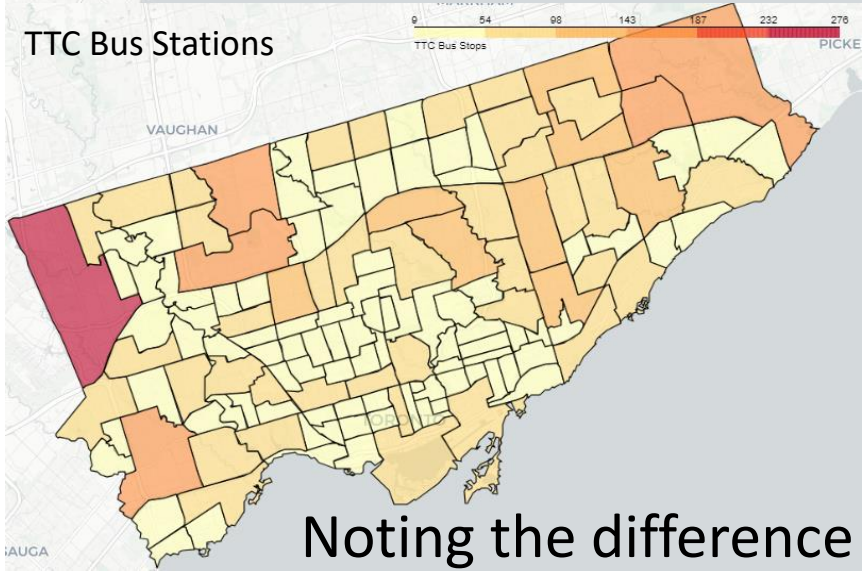
Very Useful

What can be done with geospatial visualization?









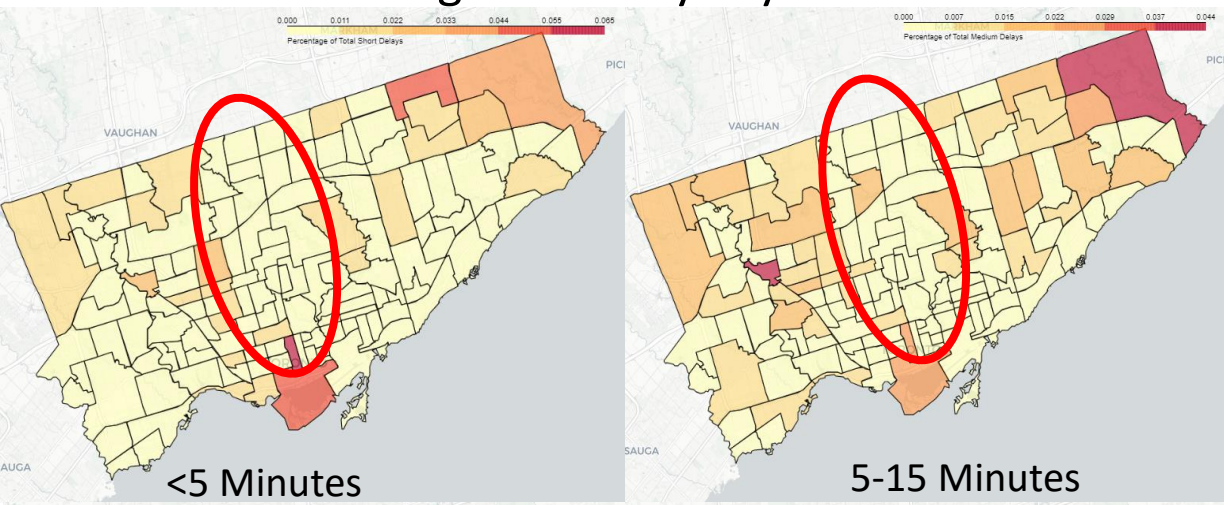
- Bus Stop data (From the TTC)
- Route Location Data (From UofT)
- Toronto areas shape File (From Git)

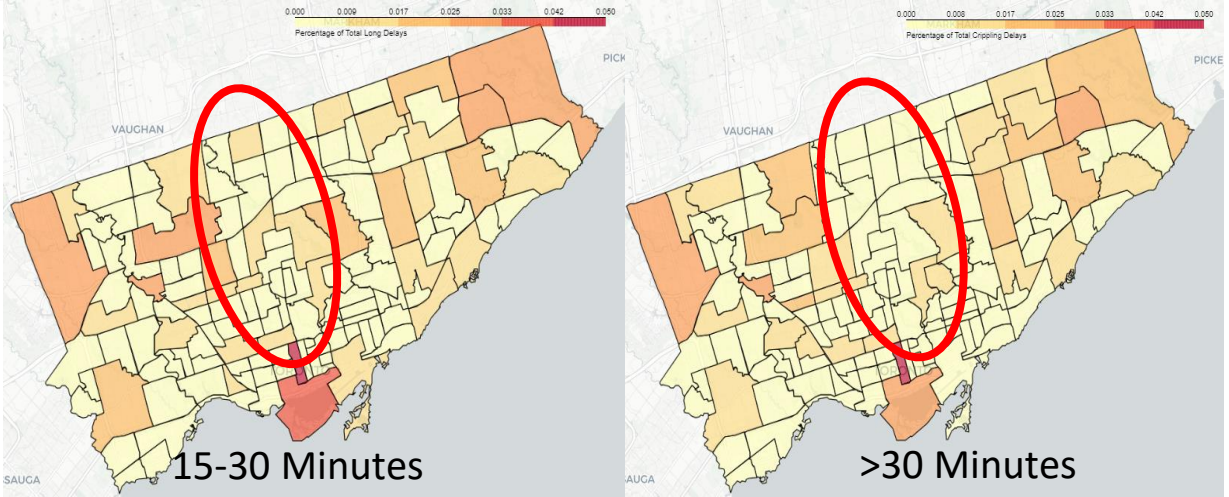
What would be useful for people to see graphically?

Noting the difference

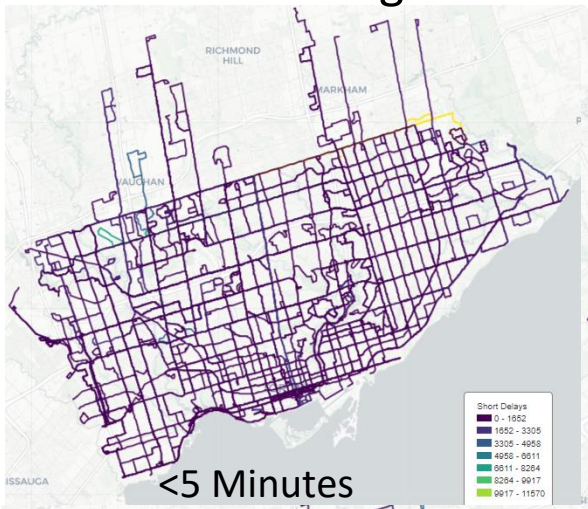
Understanding TTC Delay Locations

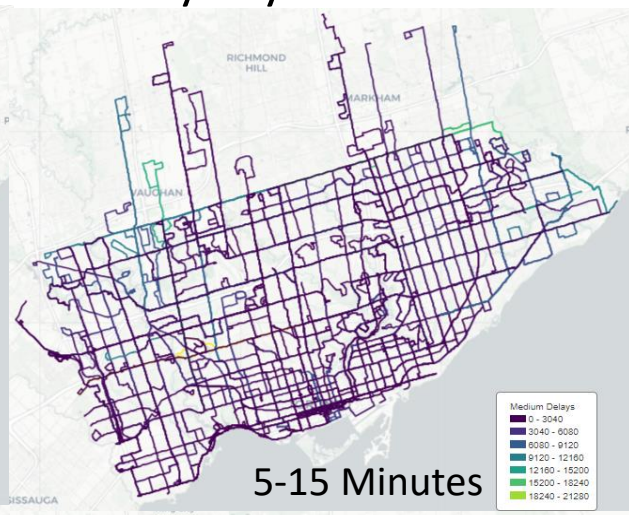
Showing % of delays by location

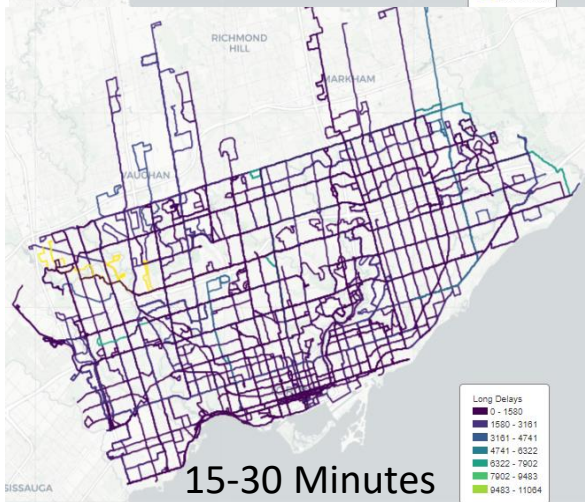


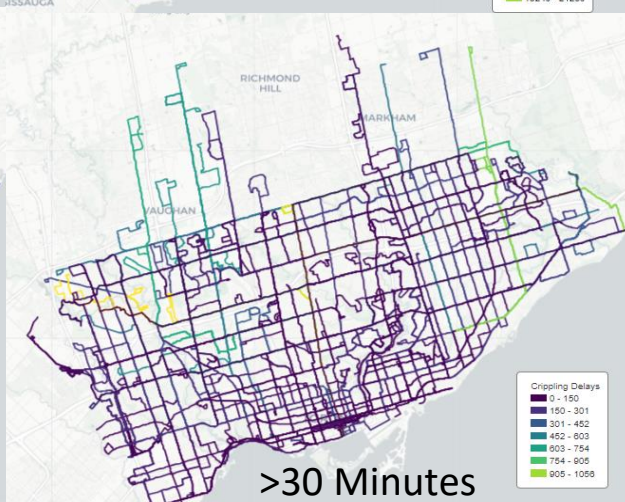


Showing number of delays by route



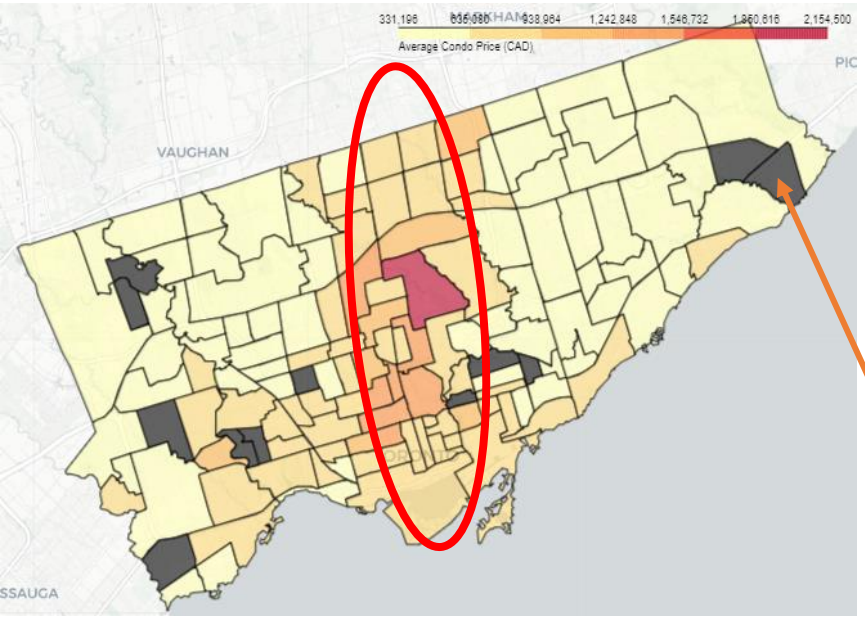




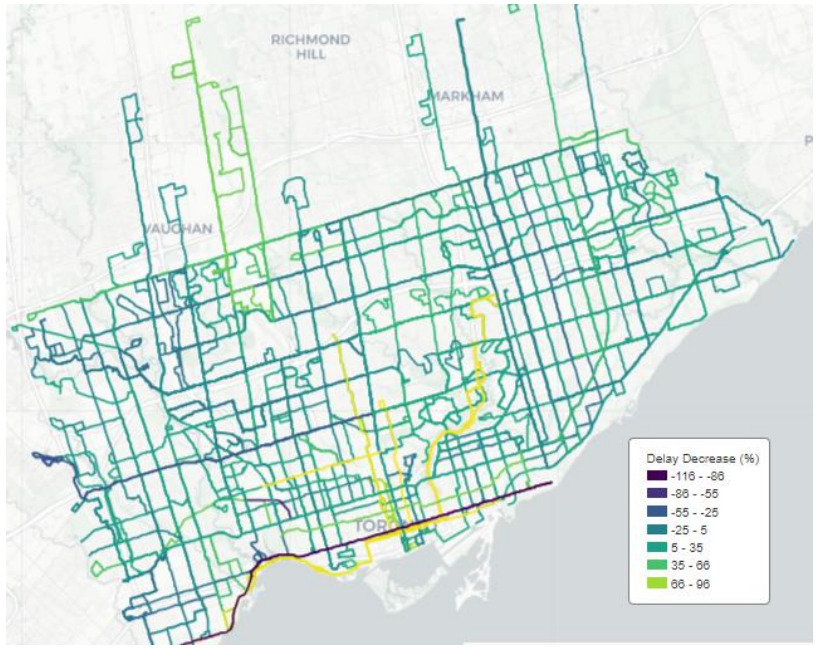


Additional Important Visualizations

Does condo price affect delay length?

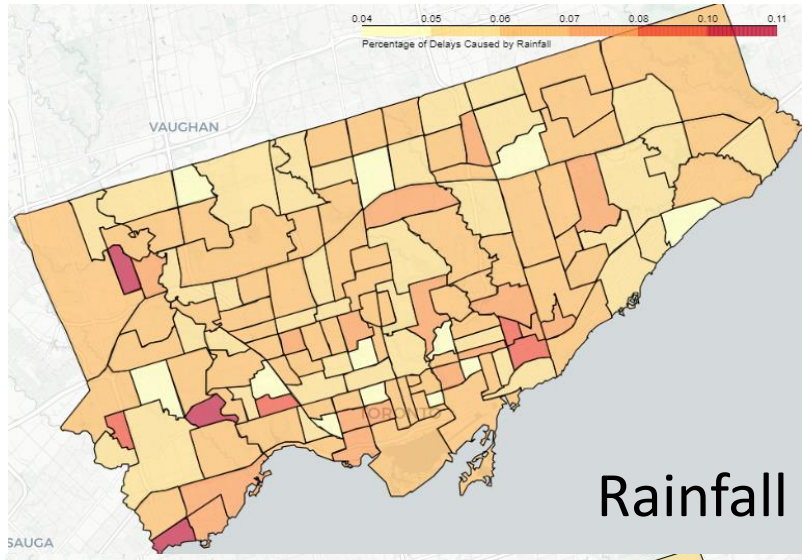


No condo price data

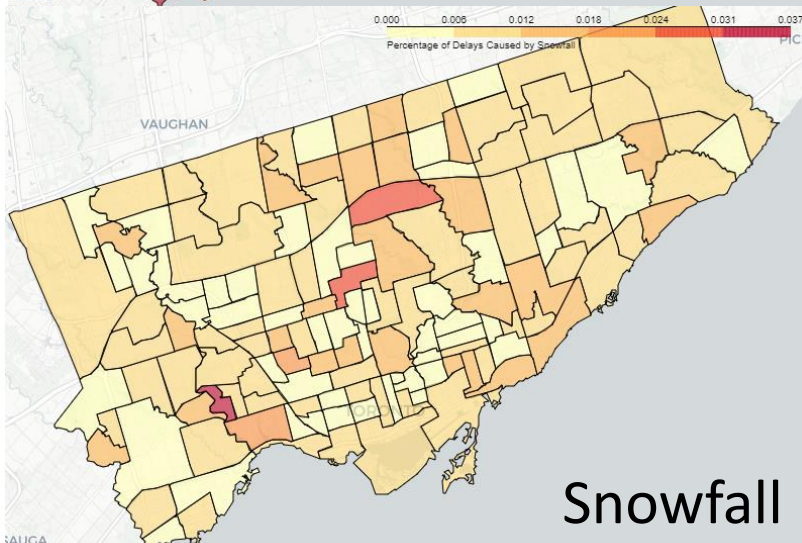


How are delays changing temporally?

What causes delays spatially?

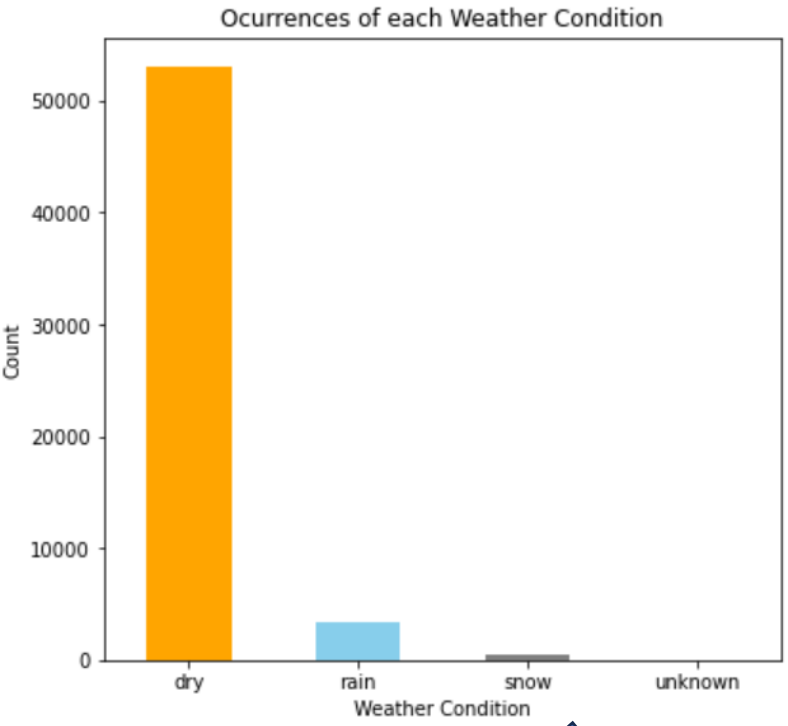


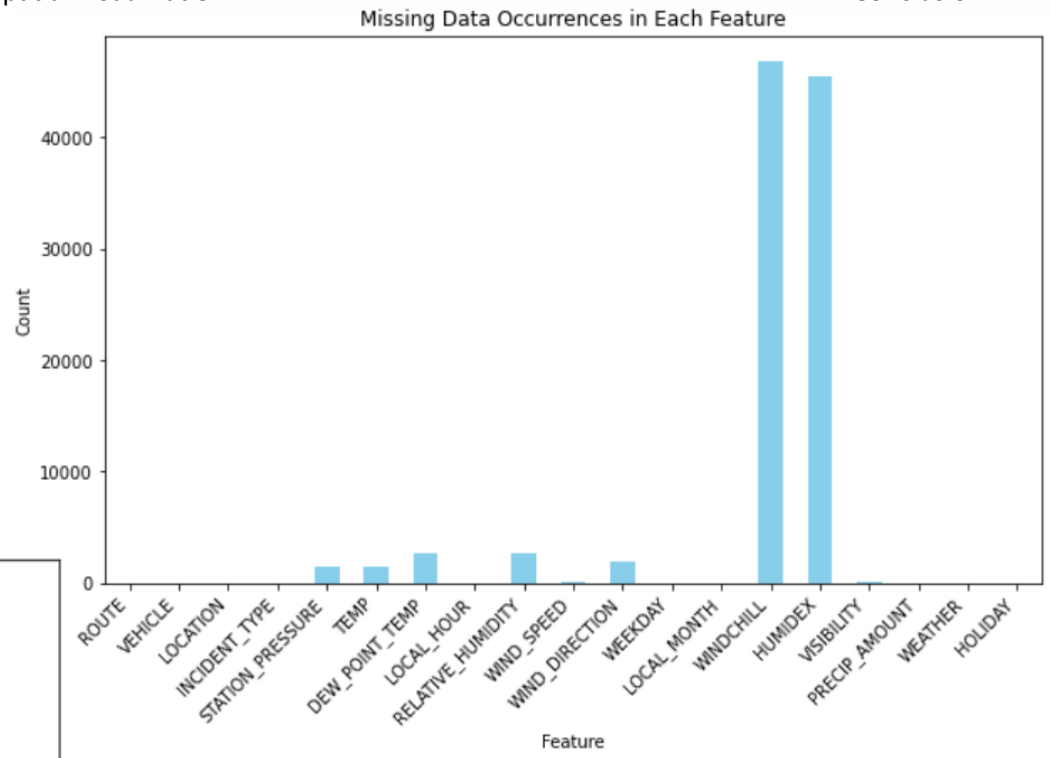
Rainfall



Snowfall

Model Creation

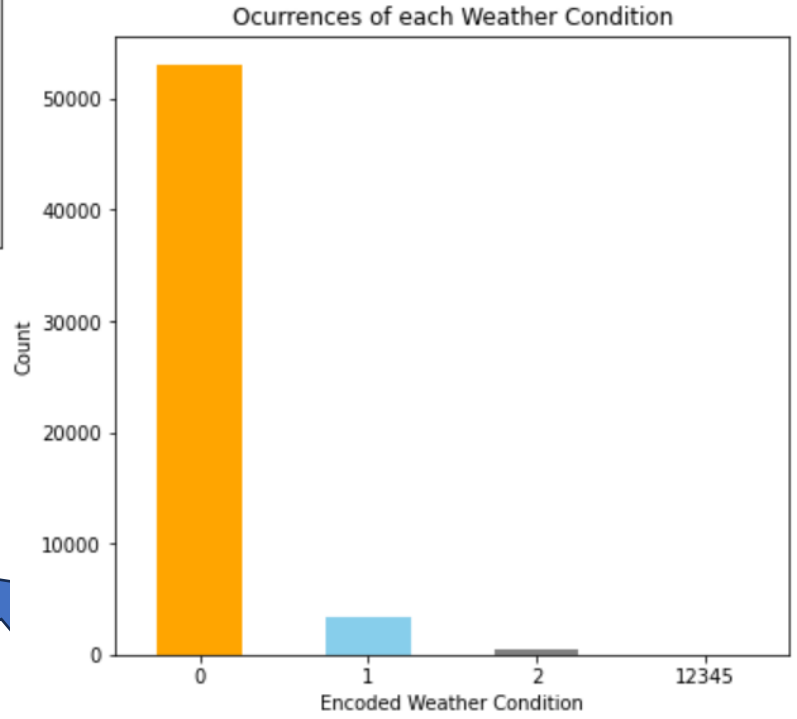




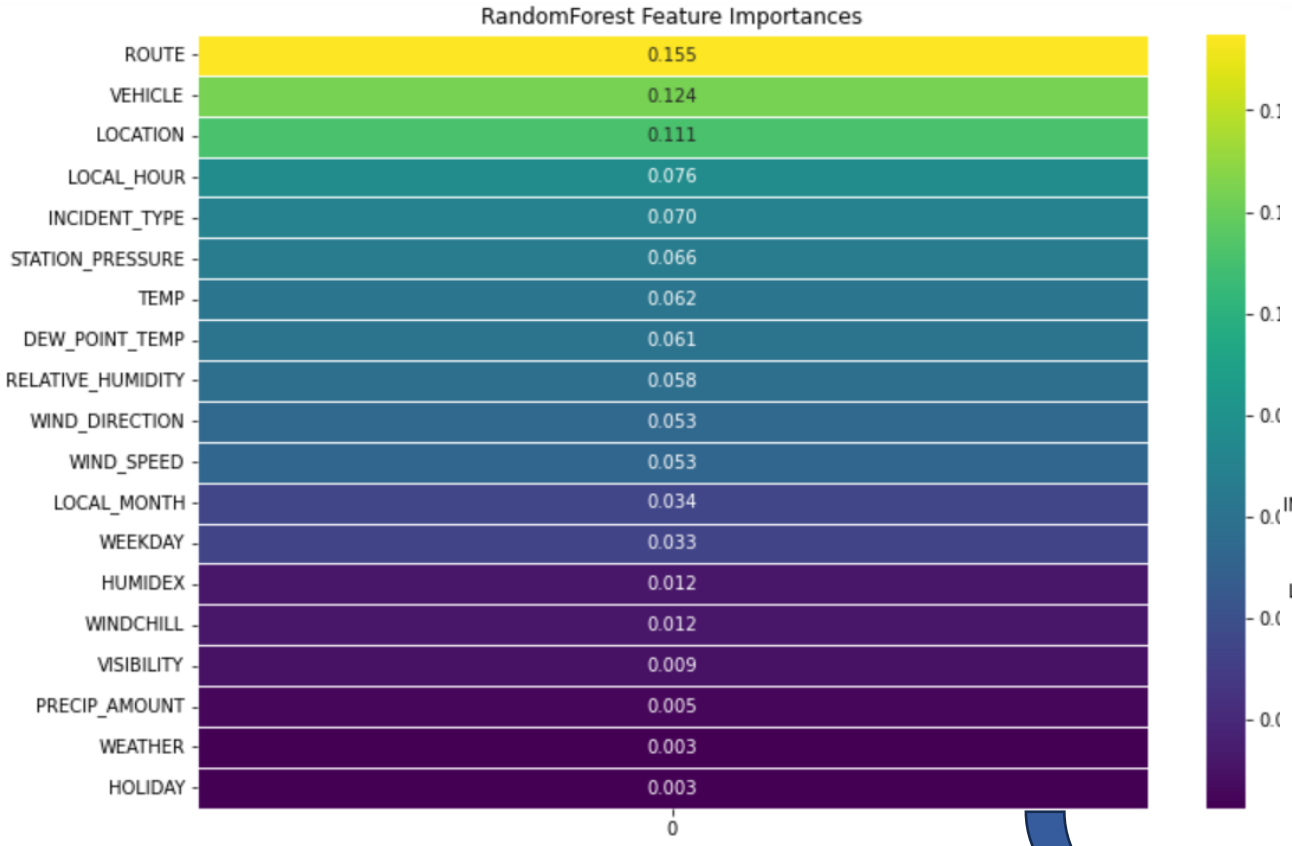
Missing Data = Feature Imbalance

Use "weighted" averaging

Speaking the language of decision trees: encoding



Feature Importance

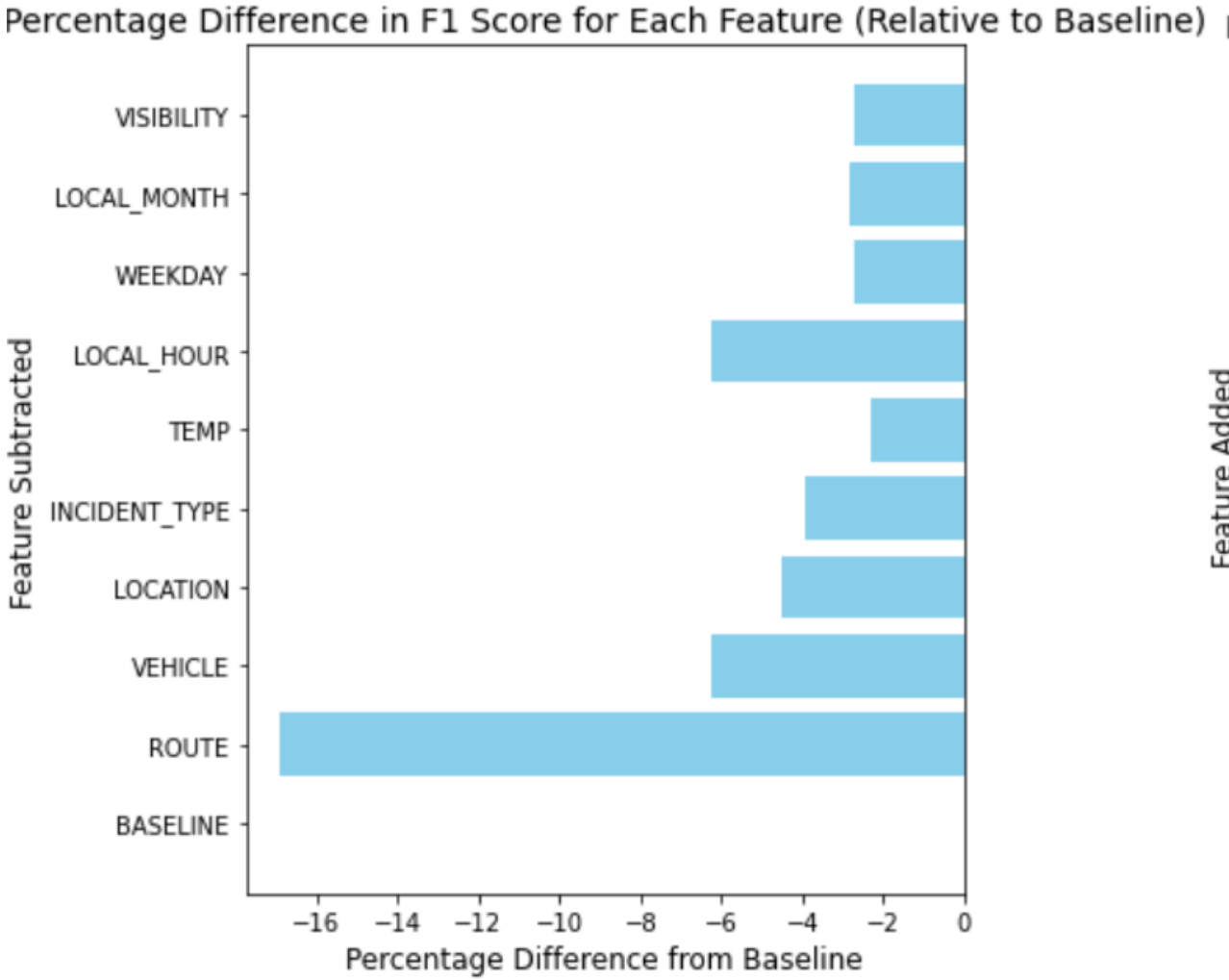


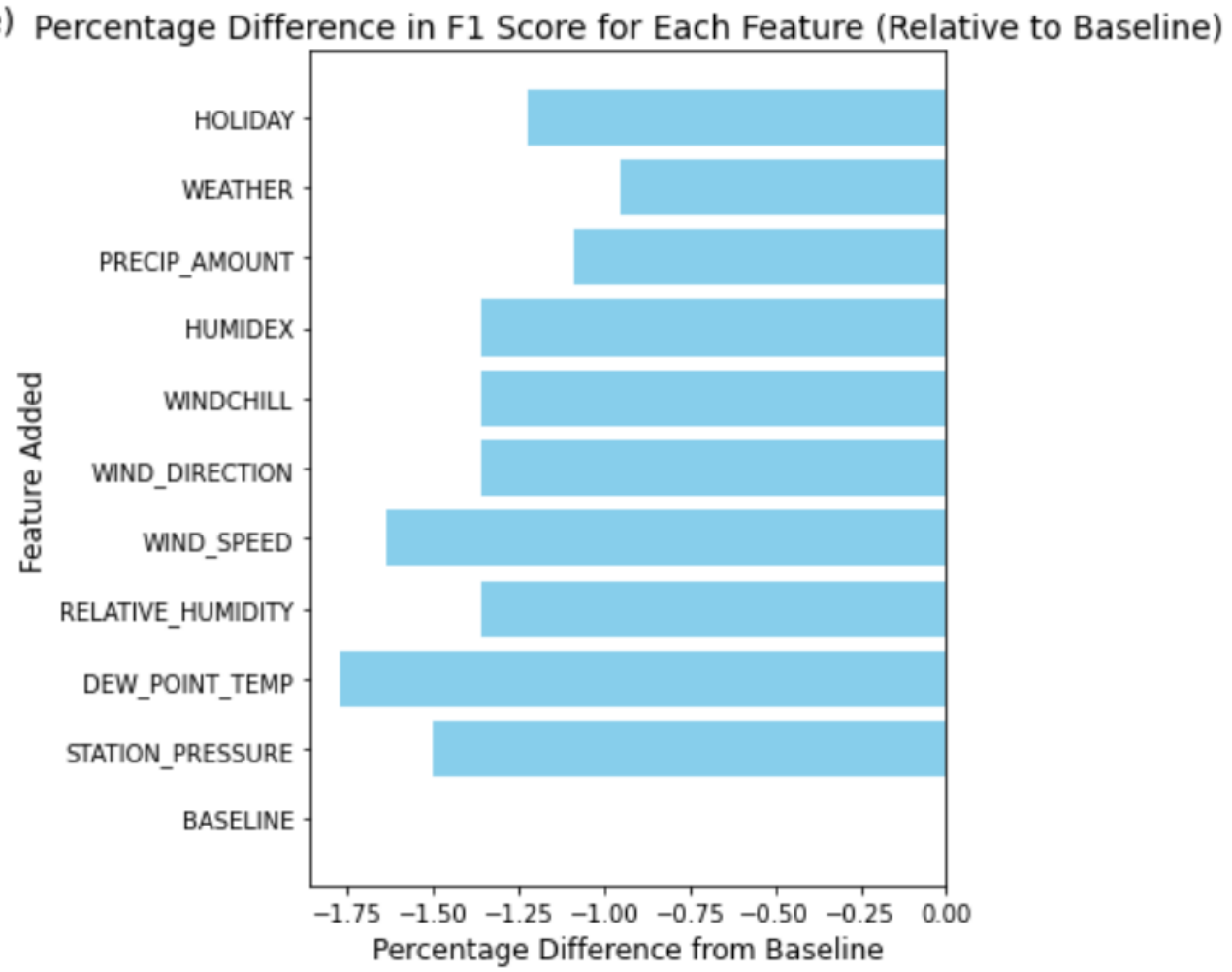
Initial F1 Score: 0.680

Improved F1 Score: 0.734

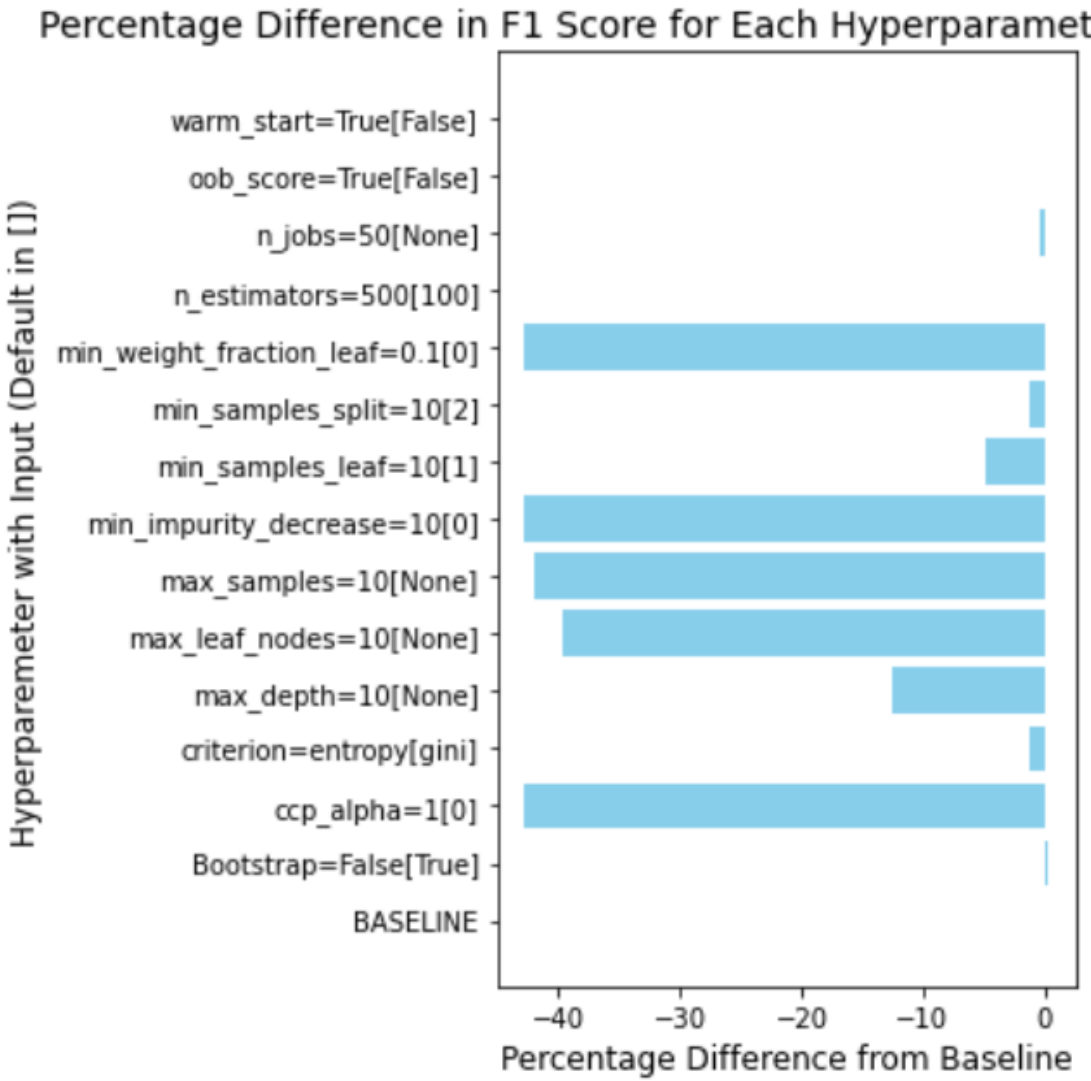
But can we do even better?

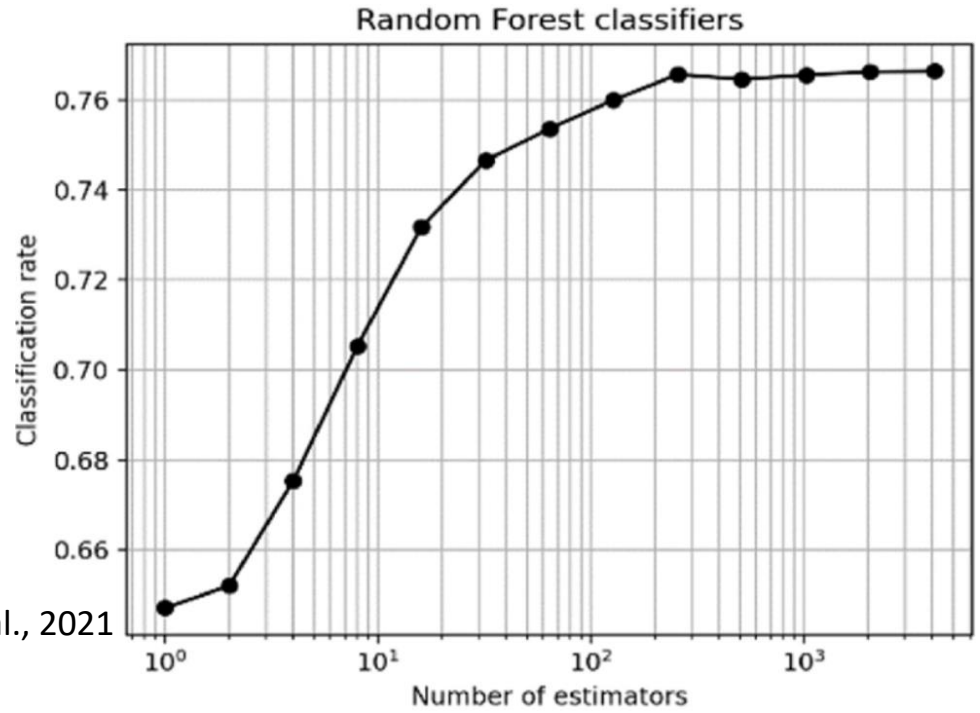
Further Feature Exploration





Hyperparameter Adjustment





Pnevmatikakis et al., 2021

Increasing n_estimators and setting bootstrap to False increased F1 score

Final F1 Score: 0.738

5-Fold Cross Validation

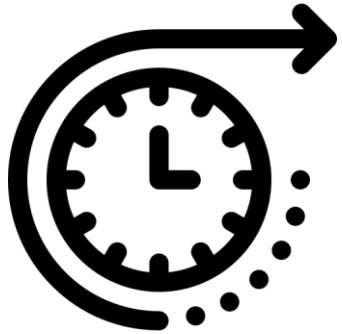
Error Margin: 0.009

Conclusions:

- TTC management has shown progress in dealing with bus delays.
- The results of this study could help TTC identify the routes where more troubles exist from different viewpoints and plan on fixing them.
- Our ML model is a helpful tool to predict the type of a delay based on incidental and weather conditions.
- We leave judgment regarding the housing prices and bus delays to the Audience.

Limitations & Future Works:

- Sentiment Analysis of Torontonian tweets. (Limitation: X is no more available for free!)
- A more comprehensive study of the Whole TTC system (Limitation: TTC is not sharing most of their gathered information)
- Comparison of TTC with Go-Transit in GTA, and other Transit companies across Canada.
- Application of more sophisticated ML methods such Neural Networks & Deep Learning.




Gracias
Thank You
Merci
Danke
Grazie