Group Members : Andy Hwang (solo project)

Guaranteed Datasets:

- Toronto Bike Share Ridership 2016-2024: https://open.toronto.ca/dataset/bike-share-toronto-ridership-data/
- Toronto City Centre Hourly Weather Data: https://climate.weather.gc.ca/climate_data/hourly_data_e.html?hlyRange=2009-12-10%7C2025-09-24&dlyRange=2010-02-02%7C2025-09-24&mlyRange=%7C&StationID=48549&Prov=ON&urlExtension=_e.html&searchType=stnProx&optLimit=specDate&Month=1&Day=1&StartYear=1840&EndYear=2016&Year=2024&selRowPerPage=25&Line=1&txtRadius=25&optProxType=navLink&txtLatDecDeg=43.666666666667&txtLongDecDeg=-79.4&timeframe=1

Potential Supplementary Datasets:

- Toronto Bicycle Traffic Counters: https://open.toronto.ca/dataset/permanent-bicycle-counters/
- Toronto Police Annual Statistical Report - Traffic Collisions: https://open.toronto.ca/dataset/police-annual-statistical-report-traffic-collisions/
- Toronto Subway Delay Data: https://open.toronto.ca/dataset/ttc-subway-delay-data/
- Toronto Streetcar Delay Data: https://open.toronto.ca/dataset/ttc-streetcar-delay-data/
- Toronto Bus Delay Data: https://open.toronto.ca/dataset/ttc-bus-delay-data/
- TTC Routes and Schedules: https://open.toronto.ca/dataset/ttc-routes-and-schedules/

The goal of this project will be to create a model that predicts Toronto bike share usage of each station based on a variety of factors. The first iteration will look primarily look at time (day of week, month, season) and weather (temperature, wind, precipitation) to predict bike ridership across the city for any given hour/day. Then, expanding on this, we'll attempt to predict bike ridership on a per station basis. As a part of this, we'll look to explore the datasets to see if we can quantify how much ridership changes with weather events and profile stations by their activity.

Depending on the results of the initial exploration and modeling, we'll look to bring in some of the potential supplementary datasets. While it is still not decided, which one's I will incorporate as this will likely depend on how the first phase goes.

The traffic collisions dataset could be used to predict the periods/locations of higher risk for cycling accidents, depends on factors such as time, weather, and location. This will look to answer the question of when and where is more dangerous for cyclists on the road. We can also look to predict the number of collisions per day/hour using these factors.

The subway, streetcar, and bus delays datasets could be used to add an additional layer to the bike share demand prediction models to see the impact of public transportation delays. This will look to answer the question of whether bike share serves as a backup transit option.