Project I: Traffic in U.S. Cities by Weather Data

Weather data is both intrinsically interesting, and also potentially useful when correlated with other types of data. Within itself, we can study signal processing concepts, such as filtering, Fourier transform, auto-correlation, cross-correlation, etc.

Correlating with traffic data I hope to gain more insight of how traffic in the U.S. cities are affected by the weather as well as coming up with prediction/classification models. Here, we will incorporate weather data with taxi rides data, Uber ride data, and vehicle collision data. Apart from exploratory analysis, using this combination we hope to address some questions such as

1. Is the number of taxi/Uber ride surge with cloudy/rainy weather, by how much?
2. For the same distance of travel, how long do the ride take in different weather? (Of course, this include certain assumptions)
3. Are there more collisions in specific type of weather, if so how much?
4. Do longer rides correlate with the increased number of accidents on the road?
5. How are these correlations differ in different cities of the U.S.?

After data exploration and correlations, a model will be built to

1. Predict number of Taxi/Uber ride on different time of the day/week given specific weather data.
2. Predict the increase in travel time for a given distance.
3. Predict the number of collisions.

With approximation of rates, these results can then be used to generate revenue prediction for taxi (increased/decreased in different weather of the day), etc.

Data source:
1. https://www.kaggle.com/selfishgene/historical-hourly-weather-data
2. https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city/data
3. https://www.kaggle.com/debanjanpaul/new-york-city-taxi-trip-distance-matrix/
4. https://www.kaggle.com/chicago/chicago-taxi-rides-2016/data
5. https://www.kaggle.com/chicago/chicago-taxi-trips/data
6. https://www.kaggle.com/nypd/vehicle-collisions/data
7. https://www.kaggle.com/new-york-city/nyc-parking-tickets/data