**Proposal for Capstone Project – Rideshare Revenue Prediction in Chicago**

**Statement of Problem:**

Since November of 2018, Transportation Network Providers (rideshare companies) operating in Chicago have been required to report all trips to the City of Chicago. A data set of these trips has been compiled with each row being an individual trip with 21 attributes. There are currently 45.3 million trips in this data set. The attributes of the trip data include the locations of pick up and drop off, the times of pick up and drop off, trip duration, fare amount and tip amount.

From this data, it is proposed that two machine learning models be developed:

1. A machine learning model to predict expected revenue for a driver who works for a set number of hours starting at a specific time of the day.
2. A machine learning model to predict expected revenue so that the rideshare company dispatcher can decide which trip requests should be given the greatest acceptance priority.

Both of these models will allow for more informed decision making rather than simply relying on intuition.

**Description of Client:**

For the first machine learning model, the client is the individual driver who is using the rideshare app. This model will allow a driver to decide the time when they should work and for how long so that their earnings are maximized. Given that drivers have the flexibility to choose when they wish to work and for how long, this model would be of great interest to the drivers.

For the second machine learning model, the client is the rideshare company itself. This model will allow the rideshare company dispatcher to decide which trips they should accept first so that earnings are maximized. This model will be useful at times when the demand for rides is greater than the number of cars available as it will allow the dispatcher to make better decisions about which ride requests should be accepted first.

**Data:**

The source of data for these models is the Chicago Data Portal website at <https://data.cityofchicago.org/> The data set itself is called “Transportation Network Providers – Trips”. This data set contains 45.3 million rideshare trips from November 1, 2018 to March 31, 2019. Each line in this data set is an individual trip with 21 attributes. The 21 attributes are as follows:

* Trip ID, start time, end time, duration, miles, fare, tip, additional charges, Trip total.
* Pickup census tract, drop off census tract, pick up community area, drop off community area, pick up centroid longitude, pick up centroid latitude, drop off centroid longitude, drop off centroid latitude.

A data set which maps the Census Tracts to Community Areas is available in the Chicago Data Portal and is called “Boundaries - Census Tracts – 2010”. A geographic map of the community areas is also available in the Chicago Data Portal.

**Approach:**

The approach used to execute this capstone project will be as follows:

1. Perform Exploratory Data Analysis: this work will involve preparing summary statistics and visualizations in order to understand the data and look for patterns.

1. Preparation of Project Data Set: this work will involve cleaning and manipulating the data to develop a working data set for solving the two proposed problems.
2. Model Development and Evaluation: this work will involve the development and testing of machine learning models using the techniques learned in this course to solve the two proposed problems.

**Deliverables:**

The deliverables for this capstone project are as follows:

1. A report documenting the project.
2. A Power Point presentation.
3. All Python code used for data exploration, for data manipulation and for machine learning model development and evaluation.