**Surge Prediction Application**

Objectives; Expected Outcome; Why are these research / toolkits important?

Our objective is to develop an easily accessible and reliable way of predicting surge prices for a given time frame.

One of the major problems we are facing these days with apps like Uber and Lyft is that the surge prices significantly increases our expense on transportations and there is almost no way of knowing when the surge will end. Uber, for example, will indicate that surge may change in 2 mins – which, sadly enough, is never the case.

Our application, however, is expected to generate results like when the surge will end and predict what the surge price will be for a given time.

The importance for this transportation wise application is quite obvious – with our application, we no longer have to undergo the excruciating pain of waiting the surge to end indefinitely. All we need to do is simply check the application and figure our when the surge will end or 'where' the surge will end (where there are no surges), and act accordingly – either enjoy a tasteful afternoon at starbucks or walk few blocks down the road and catch a uber – free of surge!

What will be the dataset you use? How did you get that? If the dataset is public, please **definitely** also submit its description via this page:

We will be using Uber/Lyft price data combined with location and time information collected to develop our application.

Since Uber/Lyf does not provide historical data, we will have to gather price data ourselves through Uber/Lyf API. We are planning to gather data from entire NYC area every day for the next three weeks (hopefully). The expected dataset will be around 3 GB with 30 million records per company.

We will also be incorporating data like weather conditions, ongoing large events as well as tourist numbers.

Which computer languages and platforms will you use?

Python; Spark; Mahout (to be determined)

Which analytics, algorithms, visualization will you implement?

* Provided algorithms, analytics and visualization are a superset of our potential candidates. Detailed usage will be explored once we have obtained the necessary amount of actual dataset.

Statistical Modeling, such as regression (linear, regression, lasso (least absolute shrinkage and selection operator); we will also be using artificial intelligence algorithms BP neural network and svm. The core concept is developing a robust model that facilitates surge prediction.

As for visualization, we will adopt application level visualization tools (D3.js bootstrap.js System G etc.) Typically D3.js provides enough data visualization capabilities and we will adopt other third party libraries if necessary.