We would like to build a surge pricing mechanism and the price calculation will be dependent on the supply-demand ratio at the specific location.

- We define location as a geohash level 6 grid measuring **600M X 1200M**
- The supply and demand ratio will be an input variable to the formula to calculate the appropriate surge pricing multiplier in that location.

- On the demand side the demand measurement is based on incoming booking requests within the geohash (the booking request will have a lat lon attached, geohash has to be calculated based on the latlon)
- On the supply side the measurement is based on drivers that are within that geohash (drivers report their locations periodically (assume every 5 seconds) with their current latitude and longitude position)

We will arrive at the supply demand ratio using the counts of each stream.

Using the same dataset we would also like you to do something to estimate the traffic congestion in each geohash at any point of time using factors like travel time and trip distance.

We would like to have two versions each of the supply demand ratio calculator and traffic congestion estimator

- a realtime version that does aggregations on the two streams every 10 minutes to come up with the supply/demand ratios and traffic congestion estimates for each geohash

- and a batch version which will allow us to look back at historical supply and demand over the last few weeks.

You can use the data available from http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml as the sample dataset to generate the appriopriate datasets or data streams for completing the exercise.

You will need to write your own producers to get the streaming data (feel free to use aws kinesis for streaming or you can set up kafka on your own.)

To illustrate your results you should build some map visualizations (real time) where we can see real time changes as data is streamed in and (batch) where we can change the date / time and see the supply demand ratio and traffic congestion levels at the date and time of our choosing

For the final part of the exercise, we would like you to include hourly weather data to your data set so that we can see analysis and visualizations of how the supply,demand and congestion is affected by the weather.

You can find the historical weather data here
https://www.wunderground.com/history/airport/KJFK/2014/1/2/DailyHistory.html?MR=1

We will discuss your exercise result and methodologies during our next scheduled interview, so please make sure your networking setup is conducive for screen sharing

Feel free to drop us an email if you have questions about the exercise.