Applied Project Proposal – Uber Trips Analysis
Stat 695 – BDA
David Arthur & Krishnan Raman

Description: Uber is a popular ride sharing service used by millions of people across the United States each day. Amongst many concerns that Uber drivers have, one of the most important deals with how they can make the most money in any given day. This primary aim of this project is to try and answer that question, along with other related questions. The data set that will be used for this analysis consists of 30 million Uber trips in New York City made between September 2014 and September 2015. Each trip details a starting and ending location along with the duration and distance of the trip. This information can be used to compute the money made on each trip as well as the number of cars at each location at a given time point.

Proposed Methods: This project will consist of four parts. First, an initial exploratory data analysis will be carried out to understand important aspects of the data such as average trip length, average number of trips, average cost of a trip, etc. Second, a Bayesian time series model will be used to identify periods (such as weeks of the year or days of the week) in which the most money can be made. Third, a Bayesian spatio-temporal model will be used to model the average number of trips in a day. This model will provide insights into both the locations where the most money is made and times of day when the most money is made. Finally, we shall devise an algorithm to find the optimal route for making the most money in a day based on insights from the Bayesian spatio-temporal model.

Division of Labor: While the group members will collaborate equally & extensively on models, distributions, project report & presentation, David will focus on the Gaussian processes and time series model, while Krishnan will be responsible for the circuit optimization and exploratory data analysis.