Project Proposal

What is the problem you want to solve?

The problem I want to solve is to see if there is a correlation between car availability and booking cancellations. Often times booking cancellations might get canceled as late as hours before the trip's start time which often causes inconvenience towards the patients.

Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

The clients will be towards cab companies to see the trends of cab bookings and the cases in which cancelations occur. Based on the analysis of the data the cab companies will be able to see data regarding cases of cancellations. It is also geared towards customers to show them which times are the best to book a cab.

What data are you going to use for this? How will you acquire this data?

There is data found in Kaggle under the competition: Predicting cab booking cancellations. The data set includes the fields of package type, from address, to address, and the time in which the bookings were made.

In brief, outline your approach to solving this problem (knowing that this might change later).

The approach to this problem will be to first clean the data by adding an extra data field that determines the change in time booked and picked from the location. There is often a correlation that as the time different is shorter the chance of cancellation is greater. I will also work on breaking the data up into the package type, and whether the booking was done online or not.

What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

The deliverables for the project will code sourced on github, along with a paper report and powerpoint slide presentation.