"Economics of Taxi Tips"
By: Denis Stukal, Rahul Joshi, Priyank Bhatia
Week 1

For the week 1 we started with some reasonable targets and started looking on challenges that we could face. As we were dealing with merging taxi(trip and fare) data with census data(demographics,income) we came up with following challenges:

1)**Geocoding of Taxi Data**:
As the census data for both income and population is borough based, we converted lat and long in pickup and drop off locations to boroughs and zipcodes for those lat-long using shapefile.

2)**Lat and Long for which there are no Boroughs/Zipcode in Shapefile**:
While conversion of lat and long to boroughs we realised that there are lot of points where there are no boroughs/zipcodes specified in shapefile, for this we currently put "Error" in those cases. We might be removing these cases or looking for some suggestions in this case like using imputation method to find missing values.

3**)Merging of Trip and Fare Data**:
We combined the geocoding above mentioned with the map and reduce program for joining trip and fare data. The result after this step was a file having both trip and fare data and four extra columns where each is specifying the borough on the basis of lat and long of pick up and drop off locations.

4)**Merging of Census Data with Taxi Data**:
We combined above data with income and population data on the basis of zipcodes. This data set will be used for analysis at zipcode level.

5)**Conversion of Shapefile to JSON File**:
In order to get prepared for visualizations ahead, we might be needing JSON files at zipcode level, we created JSON file. In order to create the json file, we had an input shapefile and a csv file having zipcode to borough based mapping, using these two files we created a json file at zipcode level resolution.

Github: https://github.com/pb1672/EconomicsTaxiTips

Deliverable for Week 1:

1. **Dataset having both income,population data with taxi data at zipcode and borough level.**

2. **JSON file having zipcode level resolution.**