Liz Aharonian – 316584960
Ori Ben Zaken - 311492110

Final Project – Report 1

Milestone #1 : Find relevant data

- After examining different taxi trips datasets we decided to take our data from 2014_Uber_Data, can be found [here].
- Our dataset contains records of ~360K uber taxis trips. Each record contains the following details: pickup date and time, pickup latitude, pickup longitude and region code.

Milestone #2: Data cleansing

- We performed full data cleansing that includes: removing records with Null values and removing columns that more than half to their values are Null or empty. In our case, no column was removed.
- Our model target is to predict the demand for taxi per time interval and area. In order to do so we added the "demand" column. We counted all of the taxi trips that occurred in a certain region and certain 10-minutes time interval. The result was added to each of relevant trip records.
- Added more features in order to help the prediction. For each record we added the features: is_Holiday, is_Weekend according to the trip date.
- In addition, in order to prepare the data for machine learning model we converted all the string values to numerical values.

Milestone #3: Training ML model

- We chose randomly 100K trip records as our dataset for the model. Splitting of course the dataset to training set and validation set.
- We tried many regression models, reached so far to the best results with Random Forest Regression model.
- These days we are also examining neural models to see if you can get even better results.

Reporting hours:

Our working day on the project is Sunday. We sat every Sunday, usually from 9am-7pm, sometimes with our mentors at Gigaspaces, Hertzeliya.

- Tutorials for AWS tools
    - 28/10/2018
    - 4/11/2018
    - 11/11/2018
- Find relevant data
    - 18/11/2018
    - 25/11/2018

Liz Aharonian – 316584960
Ori Ben Zaken - 311492110

- Data cleansing
    - 2/12/2018
    - 9/12/2018
    - 16/12/2018
- Training ML model
    - 16/12/2018
    - 23/12/2018
    - 30/12/2018

Since we started our work on the project, each of us spent 100 hours.