Big Data Analytics Techniques and Applications

Homework I

Due Date:

Analyzing Taxi data

You can choose any data analytic tools or library to answer following problem:

- **Questions**
  - Q1: What are the most pickups and drop offs region?
    - hint: use one of cluster algorithms and count number of data points of each cluster.
  - Q2: What is the best time to take taxi?
    - hint: count number of pickups in different hour.
  - Q3: Whether weather affects customers to take taxi or not?
  - Q4: Does long distance trip imply more tip?

- **Dataset**

Taxi data: July, August, September, October, November, December
Weather data: google drive
Description:
http://www.nyc.gov/html/tlc/downloads/pdf/data_dictionary_trip_records_yellow.pdf

- **Requirements**
  - A report (Hw1_{$student_ID}.pdf) describing:
    - Why you choose this tools
    - The results of four questions that describe above and put some plot to brief your answer.
    - Anything else worth mentioning (e.g. other valuable observations)