Data Incubator Project: Plot 1

Applicant Name: Nayana Charwad

Date : April 13, 2015

Problem Statement

Analyze data patterns for bike rental demand in Bay Area, Washington DC and New York City based on station locations and area. Compare bike rental demands across these three regions to comprehend bike rental needs across multiple cities. Propose installation/removal of stations based on bike rental demand in specific areas.

Dataset

Bay Area Bike Share Data (http://www.bayareabikeshare.com/datachallenge)
New York City Bike Share Data (http://www.citibikenyc.com/system-data)
Washington DC Bike Share Data (https://www.capitalbikeshare.com/trip-history-data)

Utilization for business

Proposed project can be used by bike rental service providers to forecast bike rental demands in various bike station. This will help service providers in multiple ways as follows:

- Decisions to install new stations/ remove existing less popular stations can be taken based on bike rental demand in particular area. This will augment business revenue in popular regions and reduce loss due to unused bikes on less popular stations.
- Service providers can analyze data from other regions to comprehend business needs and can think of opening new businesses in cities with greater demand.
- Some stations in cities with less popular bike rental needs can be moved to cities with higher bike rental demands.
- Possibility of across regions bike pick up/drop off can be considered to motivate people travelling across various cities to rent bike in origin city and drop it at destination city.
- Seasonal changes in rental demands can be used to move bikes across regions so as to balance supply and demand in various regions.

Preliminary Data Analysis

- Initial analysis of bike stations data shows that some stations are completely unused (number of bikes available is extremely high and number of available docks is near to zero) or some stations are highly used (number of bikes available is near to zero and number of available docks is extremely high). In both cases, data analysis should be done so as to find out real cause behind the extreme situations.
- Analysis of trip data shows that some stations are highly used as entry point stations whereas some are rarely used as entry point stations.
- After reviewing data from stations file it can be seen that bike rentals are popular in some areas as compare to bike stations in other areas.
- If bike rental needs across cities can be analyzed based on locations of stations there is possibility of getting interesting data patterns in various locations.

Technologies

MySQL, R, Python/Java, Data Visualization (if time permits)