Data Incubator Project: Plot 2

Applicant Name: Nayana Charwad

Date: April 13, 2015

Problem Statement

Forecast bike rental demand in Bay Area, Washington DC and New York City based on existing bike share dataset. Analyze correlation between bike rental needs based on time of day and weather information. Also analyze if bike rental demands show specific patterns across various seasons in year. If time permits, online application can be developed to fetch current location of users and show them nearby bike stations, availability of bikes, docks and alternate stations in case of unavailability of bikes/docks.

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:

- Based on forecast of seasonal demand, service providers can increase bikes and docks at most popular stations and remove them from less popular stations.
- Bikes can be moved from one station to another if some stations are busy during particular time of day and are rarely used at some other time interval.
- Correlation between weather and bike rental demand can be used to decide bike demand based on weather forecast. (Weather data is currently available in Bay area dataset but we can try and get it for New York and Washington D.C from publicly available weather datasets)
- If demand for bikes is low at majority of stations then bike rental promotions events such as free bike for a day / discount on bike rentals can be organized to increase usage of rental bikes.

User application will help users to find nearest bike stations easily and check alternate bike stations in case of unavailability of bikes/docks.

Preliminary Data Analysis

- Initial analysis of bike stations data shows that during specific time interval of the day some stations are completely unused (number of bikes available is extremely high and number of available docks is near to zero) and at some other times, 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.
- Weather information file can be combined with bike stations data so as to analyze effect of weather on bike rental demand.
- If analyzed carefully interesting data patterns can be found out at various stations based on duration and time of day and weather information.

Technologies MySQL, R, Python/Java, Data Visualization