CSC-423 PROJECT SUBMISSION

Application of Linear Regression on

Kaggle Bike Sharing Dataset

Capital Bike share program,

Washington D.C., USA

(Non-Technical Report)

by

Pradeep Sathyamurthy

Under the guidance of: Prof. Nandhini Gulasingam

DePaul University

23rd November 2016

Contents

Summary

Bike sharing is an innovative transportation system ideal for short distance point-to-point trips providing users the ability to pick up a bicycle at any self-serve bike-station and return it to any other bike station located within the system's service area. Capital Bike Share Program is one such initiative in Washington D.C. As part of this project we are given the data generated by the kiosk sensors from 2011 to 2012. We will analyse this dataset to find values and insights which will help users as per the program organizers to meet the demand of bicycle for its user.

Methodology

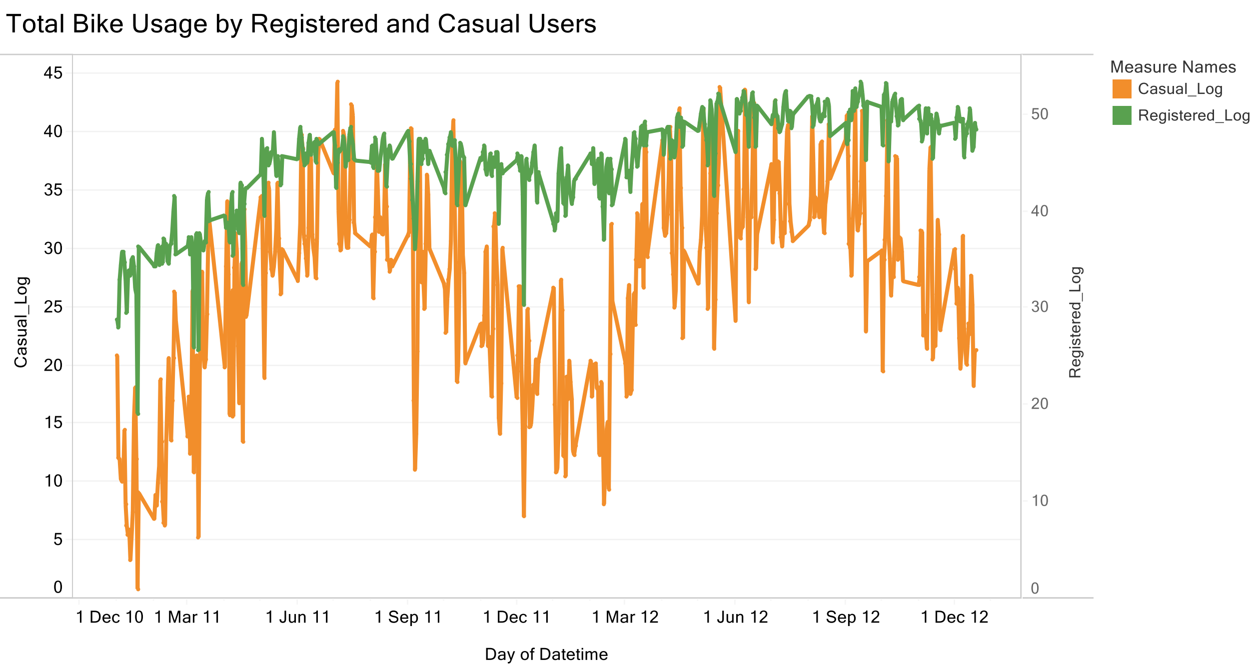
We all believe pictures speaks 1000 words, so my methodology for this analysis will be through data visualization.

Users Monthly Demands:

All these bike share program has two main users:

1. Subscribers: These are registered users who hold annual membership.
2. Customers: These are casual users who hold one-day pass.

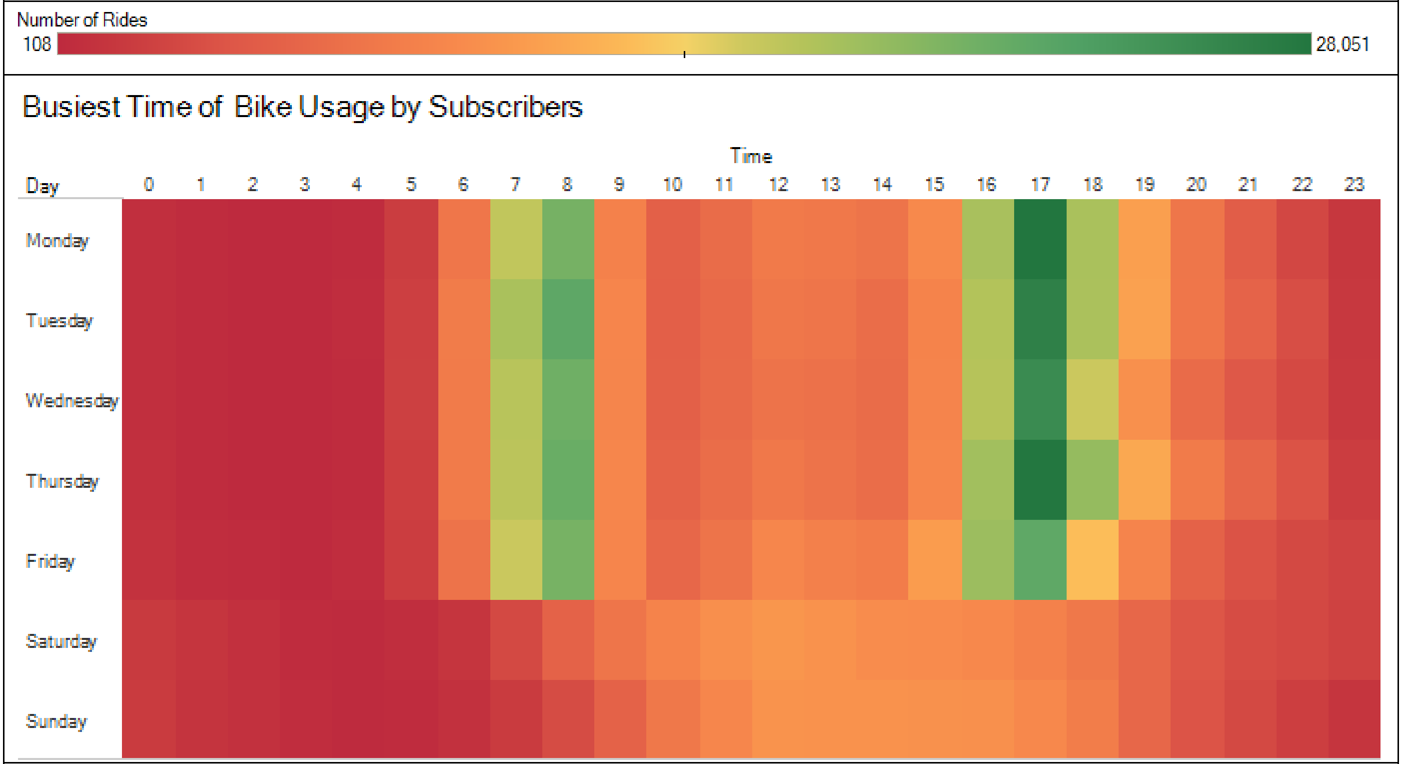
Below is the graph on total bike usage by Registered and Casual users in Capital Bike Share program:

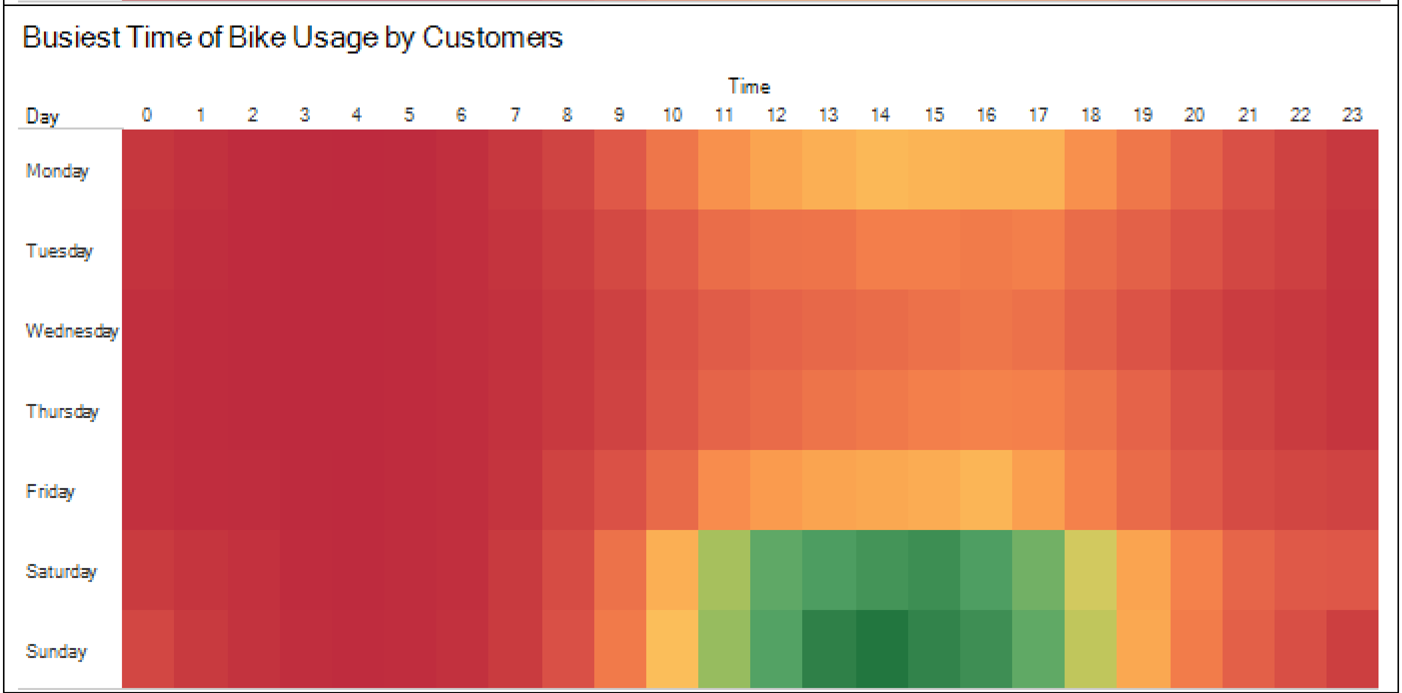


1. From above graph we could say that variation in the demand for casual user is more when compare to the registered users.
2. Usage of bikes by casual users are more during the months between Jun-Sep
3. However, usage by registered users are almost similar across the year. We see less usage in the month of December-March comparatively which may be due to heavy winter
4. From this we could infer that registered users mainly use these bike share for commuting to office/univ, while casual users are tourist who use bikes for sightseeing.
5. With this we could infer that the bikes demand pattern is based on months and months being a significant factor in deciding the demand for bikes for such program.
6. We could suggest program co-ordinator to have ***more bikes in circulation in the month of Jun-Sep to meet users demand and use rest of the months for maintenance***.

Users hourly Demand:

Bikes usage by Subscribers and Customers with x axis mapped to Hours in a day and Y axis mapped to Days in a week.

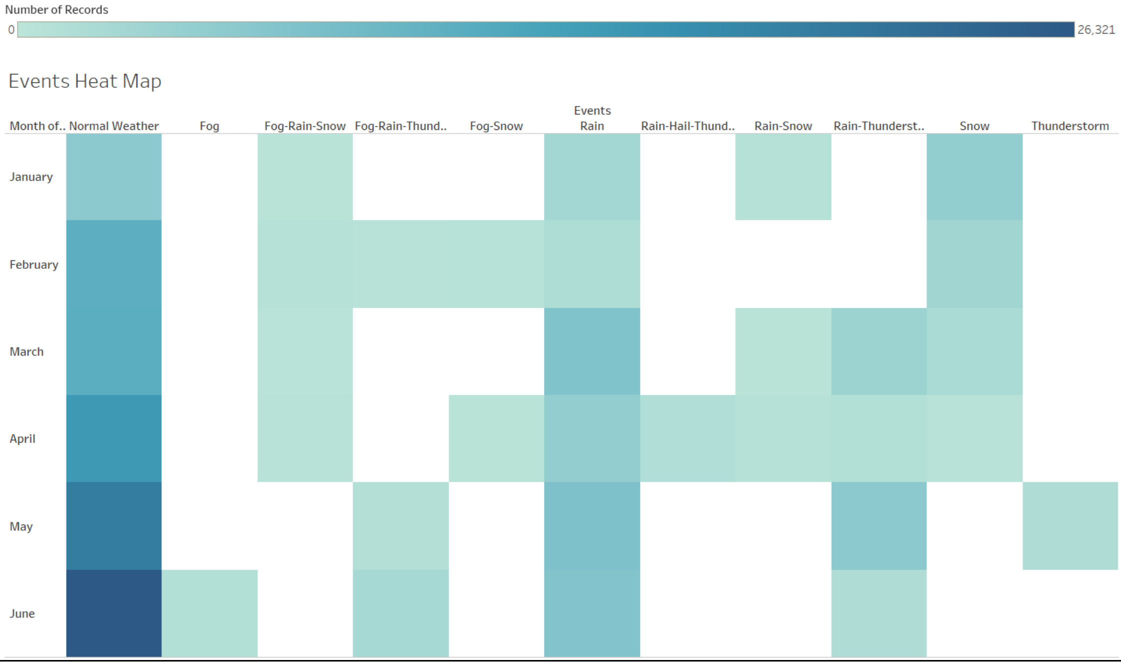




- ***Subscribers*** use frequently between ***7-8AM and 4-6PM on weekdays***, while ***customers*** use frequently during ***weekends between 11AM until 5PM***.
- Divergent color palate was used to clearly differentiate between ***High, Moderate and Low usage of Bikes*** for a particular hour by user.
- We observe a ***moderate usage of bikes*** between ***9AM until 4PM*** by customers on Monday and Friday, while other weekdays have less usage, so ***organisers can do some promotion for Tuesday – Thursday to increase the bike usage on these days***.
- With this we can infer that ***each Hours in a day also plays a very important role in deciding the bike demands*** among users.
- We suggest, organizers must ***concentrate in moving around their bicycles more actively between the working hours (6-9AM & 4-7PM)*** sighted above to meet the user demands.

Weather impact on Bike Demand:

Weather impact on Bike Usage:



- With change in temperature we could see a strong change in demand for bicycles.
- With increase in atmospheric temperature we see increase in Bike usage and with decrease in temperature we see a decrease in bike usage.
- Darker the color in above graph is more the demand of bike, we scope our visualization only for the months where we found significant demands from Graph-1[G]
- We observed high usage during normal weather, ***Snow affects the usage more than rain***

Insights

1. Usage of bikes increases every year; Time, Day, Month, season and weather have huge impact on bike demands

2. Bike usage by Subscribers is completely different when compare to Customers

3. Subscribers use bikes mostly in weekdays to commute office between 7-8 Am and 4-6Pm

4. Casual users use bikes mostly in weekends for recreations between 11Am until 5Pm

5. We suggest, organizers must concentrate in moving around their bicycles more actively between the working hours (6-9AM & 4-7PM) sighted above to meet the user demands.

6. Organizers can launch promotions on Tuesdays, Wednesdays and Thursdays as these days seem to have less than moderate usage of Bikes by customers

7. Summer is the season which has highest usage of Bikes, Winter is the season which has lowest usage of Bikes, Spring sees a moderate usage of Bikes by its users.

8. Organizers should plan for more promotions during Winter and Spring season to attract more users. However, organizers must also concentrate on safety measure in using bikes at these seasons accordingly

Timetable

Phases	Description of Work	Start and End Dates
Phase One	Obtaining Dataset from Kaggle	18-Oct to 24-Oct 2016
Phase Two	Performing EDA on Bike Sharing Dataset	25-Oct to 31-Oct-2016
Phase Three	Basic Model Building	01-Nov to 07-Nov-2016
Phase Four	Variable Binning	01-Nov to 07-Nov-2016
Phase Five	Transformations	08-Nov to 14-Nov-2016
Phase Six	Time Series Visualization	08-Nov to 14-Nov-2016
Phase Seven	Heat Map on Time and Weather	15-Nov to 20-Nov-2016
Phase Eight	Report Writing and Review	15-Nov to 20-Nov-2016
Phase Nine	Deliverable Submission	21-Nov-2016

Key Personnel

Team Member	Pradeep Sathyamurthy
Professor	Prof. Nandhini Gulasingam
Shadow Guidance	Dr. Eli T. Brown (Prof for Data Visualization course-CSC-465)
Project for	CSC-423
Target Team	DePaul CDM

Tool Used:

Tableau

Deliverables

Final Report	**Prady_CSC_423_Non_Technical_Report.pdf**	Contains final Non-Technical Report
Data Set Used	**train.csv**	Raw Dataset downloaded from Kaggle

Reference

[1] Bike Share Dataset in Kaggle Website *https://www.kaggle.com/c/bike-sharing-demand*

[2] Publicly available capital bike shares program data *http://capitalbikeshare.com/system-data*

[3] Data from which weather data was filled *http://www.freemeteo.com*

[4] Citation Request: Fanaee-T, Hadi, and Gama, Joao, Event labelling combining ensemble detectors and background knowledge, *Progress in Artificial Intelligence (2013): pp. 1-15, Springer Berlin Heidelberg.*