**BIKESHARE RIDERSHIP MODEL FOR STATION-LEVEL FORECASTING**

Shraddha Sagar

**ABSTRACT**

As a relief to the current automobile centric transportation systems, many cities around the globe currently operate bike sharing system. Currently, many of the cities in the United States has bikeshare system. While much research in bikesharing is duly taking place, a significant analysis on factors affecting bikeshare ridership is required in order to maximize the utilization and benefits of bikeshare systems. This study quantifies the factors affecting bikesharing and develop a direct ridership model predicting the monthly station-level bikeshare ridership. In particular, the study investigates the effect of demographic and employment characteristics near bike share stations on bike share ridership in three US cities. A regression analysis is done to estimate how station-level bike sharing ridership correlates with a number of factors such as population density; retail job density; bike, walk, and transit commuters; median income; education level and weather. The regression model so developed can be used to predict potential levels of ridership and identify station locations for a proposed bike share system that can serve the greatest number of riders.

**INTRODUCTION**

The automobile-centric transportation system across the United States suffer from fundamental issues such as traffic congestions and lack of parking. A 2014 report of CityLabs says that Americans drove for 85 percent of their daily trips while 70 percent of them are under one-mile [1]. Mitchell et. al proposes publicly shared vehicles as a relief to the congested urban traffic network [2]. Adopting a similar idea to solve the increasing unsustainable transportation condition, the country is promoting cycling for easy travel for short distance in a city or region.

Bicycles are environment friendly and takes less space for movement and parking than automobiles. Bikeshare systems would increase the efficiency of utilization as many would share the same resources [3]. A public bike-share system is a service in which bicycles are provided for a short-term access for a price. It allows the users to borrow bikes from a station and return it to any other station in the city after the ride. Bike-share began in Europe in 1965 and it is one of the fastest way to move around with least concern about parking space. The concept of bike-sharing is prevalently increasing in the United States nowadays. A recent study reports 119 US cities with bike-share systems together having about 4800 stations [4]. With the increasing growth of bikeshare systems across the country, a number of studies have attempted to analyze the effect of them on the transportation system. A 2015 study found that in Washington D.C, 8% of the time users prefer bikeshare over car trips [5]. Additionally, a separate study on D.C showed a noticeable contribution of Capital Bikeshare in reducing traffic congestion [6]. In order to maximize the utilization and benefits of bikeshare systems, a significant analysis on factors affecting bikeshare ridership is required.

Previously, a number of studies attempted to analyze the factors influencing bike share usage and predict the bikeshare ridership at station-level as well as system-wide. These studies identified diverse correlation between demographic and environment factors with ridership. Most of such studies are based on large public bike-share network centered around major cities like New York and Washington D.C. However, there are hardly any attempts to look at how the bikeshare ridership relates to the surrounding factors in small cities and college towns. This study investigates the effects of demographic, environment, employment and climatic factors on bike share ridership at stations in three operational systems: BCycle in Boulder, Colorado; Bike Chattanooga in Chattanooga, Tennessee; and CoGo Bike Share in Columbus, Ohio. Using publicly available datasets, a direct ridership model is built to estimate the average monthly station level bikeshare ridership. This model can be utilized to forecast ridership for a proposed bikeshare system for cities with similar characteristics to the ones chosen for the modelling.

**LITERATURE REVIEW**

A number of studies have attempted to develop tools to forecast bike share ridership and identify appropriate service locations. These studies demonstrate a variety of relationships between demographic, built environment, transportation infrastructures and temporal characteristics on bike share ridership.

Weather plays a major role in bicycle usage. According to Miranda-Moreno and Nosal [7], precipitation, temperature, and humidity have significant effects on bicycle ridership. The result shows that bicycle volumes is influenced by the presence of rain in the morning and within three hours before the travel time. Further, Rose et.al [8] identified a non -linear effect in the relationship between weather variables and ridership. In 2016, Campbell et.al [9] concluded that trip distance, temperature, precipitation, and poor air quality has a strong negative impact on bikeshare demand.

Several studies demonstrated that demographic factors and build in environment contribute to the bike sharing usage [10, 11, 12]. Contrary to this, Campbell et.al [9] demonstrated that user demographics is not a strong factor on bikeshare ridership and he indicated that users are drawn across the social spectrum. However, Bachand-Marleau et.al [10] found that to have the greatest effect ton the likelihood for use of a shared bicycle system is the proximity of home to docking stations. He suggested that the potential of shared bicycle system can be maximized by increasing the number of docking stations in residential neighborhoods. Another research shows that station-to-station distance and the number of intersections with major roads have a negative impact on ridership [11]. Also, a number of studies identified that presence of restaurants and parks in the vicinity of bikeshare stations increase ridership [3, 13, 14].

This project takes into account the suggestions from previous research to builds up a direct ridership model predicting for bikeshares at station levels.

**METHEDOLOGY**

A regression analysis was performed using the 2017 trip data in the three existing bikeshare systems: BCycle in Boulder, Colorado; Bike Chattanooga in Chattanooga, Tennessee; and CoGo Bike Share in Columbus, Ohio. The bikeshare network of BCycle has 41 stations while Bike Chattanooga and CoGo Bike Share has 38 and 46 stations respectively. While BCycle and Bike Chattanooga has 300 bikes serving the city’s bikeshare, CoGo Bike Share has 365 bikes. All the three bike systems are available throughout the year with the exception of severe weather. The 2017 trip data and the station location of the three systems were collected from the open data sources available in their websites [15, 16, 17]. The natural log of monthly rentals by station is serves as the dependent variable. The natural log was selected over the direct monthly rentals to help linearize the variable which helps to improve the continuity of the discrete variable [13]. The independent variables addressing various demographic, employment and climatic factors were collected for the three cities from different public data sources available [18, 19, 20]. The variables were chosen based on review of previous bike sharing ridership estimation literatures. Table 1 shows the definition of all variables considered for the statistical analysis.

Table 1 Description of Variables

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Source** |
| **Dependent** |  |  |
| LN(Monthly Rentals) | Natural log of number of rentals at each station | Bike sharing system websites |
| **Independent** |  | |
| Demographic Factors |  | |
| Total Pop | Total Population | U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates |
| Pop Density | Population per unit area |
| Median Income | Median Household Income in past 12 months (in 2015 inflation-adjusted dollars) |
| HH units | Number of household units |
| Bachelors | Population 25 years or over with Bachelor’s degree |
| Graduate | Population 25 years or over with Graduate degree |
| Alternate Commuters | Workers 16 years and over who commute by public transport, bicycle or walking |
| Employment Factors |  | |
| Employment | Total number of jobs by Residential Area Characteristics | Longitudinal Employer Household Dynamics, 2015 |
| Edu\_Health | Number of jobs in Health Care and Educational Services by Work Area Characteristics |
| Recr\_Acc | Number of jobs in Recreation and Accommodation by Work Area Characteristics |
| Retail Jobs | Number of jobs in Retail Trade by Work Area Characteristics |
| Weather Factors |  | |
| Average Rain | Average rainfall per month in inches | World Climate |
| Average Temperature | Average temperature per month in degree Fahrenheit |

The bikeshare data collected has details on all the trip made during 2017 which included check out station location, checkout date and time. With required data management processes, number of checkouts made at every station each month was calculated. The average precipitation and temperature recorded for these cities for every month was also tagged along this data [20].

The data on demographic variables for all three cities were extracted from 5 year estimate of American Census Survey for 2015 at block group level [18]. The data was joined with the block group shapefile of these cities to add a spatial component to it [21]. This data was spatially joined with the locations of the bike stations to extract the demographic data around them.

The employment data was obtained from Longitudinal Employer-Household Dynamics at the census block level for 2015 [19]. This was spatially more detail compared to the demographic specifics. To extract the employment characteristics surrounding the stations, buffers were created to one-tenth of a mi buffer around each station. Because of the spatial detail of the LEHD data, each buffer intersects multiple census block. Therefore, the data scaled to the buffer area to aggregate the employment features around each station. With a similar approach explained previously, the data was spatially joined with the bike station locations to combine all the data together.

MODEL

A bivariate correlation analysis showed that most of the chosen variables have an expected relationship with bike sharing rentals at 5% significance level. Table 2 shows the Pearson Correlation Coefficient of the independent variables with ln(Montly Rentals)

Table 2 - Pearson Correlation Coefficient of the independent variables with ln(Montly Rentals)

|  |  |  |
| --- | --- | --- |
| **Independent Variables** | **Correlation** | **P-value** |
| Average Rain | -0.25062 | <.0001 |
| Average Temp | 0.40906 | <.0001 |
| Employment | -0.08819 | 0.0027 |
| Edu\_Health | -0.21059 | <.0001 |
| Recr\_Acc | 0.21727 | <.0001 |
| Retail Job | 0.26578 | <.0001 |
| Total Pop | 0.15626 | <.0001 |
| Pop Density | -0.00763 | 0.7955 |
| Median Income | 0.09250 | 0.0027 |
| HH units | 0.00093 | 0.9748 |
| Bachelors | 0.01905 | 0.5175 |
| Graduate | 0.15196 | <.0001 |
| Alternate Commuters | 0.33839 | <.0001 |

The variables which have appropriate correlation with the dependent variable are chosen for the final modeling to ensure a refined multivariate regression model. The correlated variables chosen were checked for possible interaction with the dependent variable. After several trials, the model with the maximum predictive power that incorporate a variety of independent variables was chosen. Table 3 shows the characteristics of the preferred model.

Table 3 – Model Characteristics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summary of Fit** | | | | |
| RSquare | 0.497742 | | | |
| RSquare Adj | 0.481583 | | | |
| Root Mean Square Error | 0.858669 | | | |
| **Analysis of Variance** | | | | |
| Source | DF | Sum of Squares | Mean Square | F Ratio |
| Model | 12 | 671.9158 | 55.993 | 75.9421 |
| Error | 1037 | 764.5924 | 0.7373 | Prob > F |
| C. Total | 1049 | 1436.5082 |  | <.0001 |
| **Parameter Estimate** | | | | |
| Term | Estimate | Std Error | t Ratio | Prob>|t| |
| Intercept | 2.802208 | 0.180202 | 15.55 | <.0001 |
| Average Rain | -0.137456 | 0.028348 | -4.85 | <.0001 |
| Average Temp | 0.036207 | 0.00188 | 19.26 | <.0001 |
| Edu\_Health | -0.000011 | 1.90E-06 | -5.75 | <.0001 |
| Employment | -3.51E-05 | 8.76E-06 | -4.01 | <.0001 |
| Recr\_Acc | 1.74E-05 | 5.89E-06 | 2.96 | 0.0032 |
| Retail Job | 7.36E-05 | 1.28E-05 | 5.78 | <.0001 |
| Total Pop | -0.000345 | 9.46E-05 | -3.65 | 0.0003 |
| Graduate | 0.0025862 | 0.000512 | 5.05 | <.0001 |
| Alternate Commuters | 0.0021627 | 0.00027 | 8.01 | <.0001 |
| Average Rain\*Alternate Commuters | -0.000839 | 0.00017 | -4.92 | <.0001 |
| Median Income | -5.30E-06 | 2.21E-06 | -2.4 | 0.0167 |
| Median Income\*Alternate Commuters | -3.96E-08 | 1.00E-08 | -3.96 | <.0001 |

LIMITATION AND NEXT STEP

With the limitations faced on the process of this study, several directions to future research are opened. As there are several bikeshare ridership systems operating in the US nowadays, the data can be expanded which is more likely to improve the predictivity of the model. Also, build environment variables (such as parks and schools) and transportation network factors (like presence of bikeways) are factors supporting ridership and can be included for a better model.

REFERENCES

1. Buehler, R. 9 Reasons the U.S. Ended Up So Much More Car-Dependent Than Europe.In, 2014.

2. Mitchell, W. J., C. E. Borroni-Bird, and L. D. Burns. *Reinventing the Automobile: Personal Urban Mobility for the 21st Century*. MIT Press, 2010.

3. Kim, D., H. Shin, H. Im, and J. Park. Factors Influencing Travel Behaviors in Bikesharing. *Transportation Research Record, Tranportation of Research Board*, 2012.

4. Malouff, D. All 119 US bikeshare systems, ranked by size.In, 2017.

5. Fishman, E. Bikeshare: A Review of Recent Literature. *Transportation Research Record, Tranportation of Research Board,* Vol. 36, 206.

6. Hamilton, T., and C. Wichman. Bicycle Infrastructure and Traffic Congestion: Evidence from DC's Capital Bikeshare. *Journal of Environmental Economics and Management,* Vol. 87, 2018, pp. 73-93.

7. Miranda-Moreno, L., and T. Nosal. Weather or Not to Cycle - Temporal Trends and Impact of Weather on Cycling in an Urban Environment. *Transportation Research Record, Tranportation of Research Board*.

8. Ahmed, F., M. Figliozzi, C. Jakob, and G. Rose. Quantifying and Comparing Effects of Weather on Bicycle Demand in Melbourne, Australia, and Portland, Oregon *Transportation Research Record, Tranportation of Research Board*, 2011.

9. Campbell, A. A., C. R. Cherry, M. S. Ryerson, and X. Yang. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. *Transportation Research Part C: Emerging Technologies,* Vol. 67, 2016, pp. 399-414.

10. Bachand-Marleau, J., B. Lee, and A. El-Geneidy. Better Understanding of Factors Influencing Likelihood of Using Shared Bicycle Systems and Frequency of Use. *Transportation Research Record, Tranportation of Research Board*.

11. El-Assi, W., M. S. Mahmoud, and K. N. Habib. Effects of Built Environment and Weather on Bike Sharing Demand: A Station Level Analysis of Commercial Bike Sharing in Toronto.

12. Krykewycz, G., C. Puchalsky, J. Rocks, B. Bonnette, and F. Jaskiewicz. Defining a Primary Market and Estimating Demand for Major Bicycle-Sharing Program in Philadelphia, Pennsylvania *Transportation Research Record, Tranportation of Research Board*.

13. Rixey, R. A. Station-Level Forecasting of Bike Sharing Ridership: Station Network Effects in Three U.S. System. *Transportation Research Record, Tranportation of Research Board*, 2013.

14. Faghih-Imani, A., RobertHampshire, L. Marlac, and N. Elurud. An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice,* Vol. 97, 2017, pp. 177-191.

15. *BCycle Boulder*. <https://boulder.bcycle.com/data-reports>.

16. *Bike Chattanooga*. <https://bikechattanooga.com/system-map/>.

17. *CoGo Bike Share*. <https://www.cogobikeshare.com/system-data>.

18. *American Census Survey*. <https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t>.

19. *Longitudinal Employer-Household Dynamics*. <https://lehd.ces.census.gov/>.

20. *WorldClimate*. <http://www.worldclimate.com/>.

21. *Topologically Integrated Geographic Encoding and Referencing(TIGER) Product*. <https://www.census.gov/geo/maps-data/data/tiger.html>.