Bike Shares in Urban London

Jennifer Ruiz

Bellevue University

May 28th, 2020

Introduction

The concepts and implementation of bike sharing were first seen in the 1960s in Amsterdam. However, the trend took more than 50 years to become widely accessible throughout the world. Urban areas around the globe have embraced the concept, both for its sustainability and the physical benefits for residents. Beyond creating cleaner and healthier cities, bike shares have a strong impact on other areas of the cities they serve. The economic impact of bike shares to local businesses is significant, with small businesses located near docking stations reporting a sixteen percent increase in sales after the stations were made operational (Hendrix, 2017). Cycle shares also reduce urban congestion and lessen the strain on public transportation (Nikitas, 2016).

The city of London has become a model for “cycle share” programs around the world. First establishing a program in 2010, the city saw over one million rides during the first ten weeks of operation with more than 90,000 citizens using the service (Centre for Public Impact, 2020). The program has 750 docking stations located throughout London and the surrounding boroughs with more than 11,500 bicycles available to hire (Transportation for London, 2020). The author has chosen to focus on London as a case study to better understand bike share behavior and the factors which impact its usage.

Research Question

This project endeavors to analyze the impact of multiple factors on bike share behavior for understanding and to predict usage. To achieve this end, a comprehensive statistical and visualization analysis will be conducted.

Key Research Questions include:

1. During what time periods do most bike shares occur?
2. What effect does weather have on bike shares?
3. What effect do seasonality, weekends, and holidays have on bike shares?
4. What patterns exist in the data that can shed additional light on bike sharing?
5. What factors best predict cycle shares?

Methodology

The Centre for Public Impact provides open-source access to the data through its website and Kaggle as a merged and clean dataset. The data for this project was sourced from Kaggle. The dataset included the cycle share data from Transportation for London, weather data for the 2015-2016 time period provided by freemeteo.com, and the United Kingdom Bank Holiday information provided by the UK government.

The original dataset contained 10 variables with 17, 414 observations. The data contained no missing values but there was a need to transform several variables within the dataset. The original dataset contained a variable labeled "timestamp", which contained the calendar date and hour of the cycle share transaction. To better work with the data, it made sense to split this variable into separate date and time components, thus creating two new variables. When importing the dataset, four variables were coded as numeric instead of categorical so those needed to be corrected before proceeding with analysis. Table 1 shows the structure of the dataset after cleaning and transformation was completed.

Table 1 Structure of London Dataset

```
'data.frame':   17414 obs. of  11 variables:
 $ cnt         : int  182 138 134 72 47 46 51 75 131 301 ...
 $ t1          : num  3 3 2.5 2 2 2 1 1 1.5 2 ...
 $ t2          : num  2 2.5 2.5 2 0 2 -1 -1 -1 -0.5 ...
 $ hum         : num  93 93 96.5 100 93 93 100 100 96.5 100 ...
 $ wind_speed  : num  6 5 0 0 6.5 4 7 7 8 9 ...
 $ weather_code: Factor w/ 7 levels "1","2","3","4",..: 3 1 1 1 1 1 4 4 4 3 ...
 $ is_holiday  : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
 $ is_weekend  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
 $ season      : Factor w/ 4 levels "0","1","2","3": 4 4 4 4 4 4 4 4 4 4 ...
 $ time        : chr  "00:00" "01:00" "02:00" "03:00" ...
 $ date        : chr  "01/04/2015" "01/04/2015" "01/04/2015" "01/04/2015" ...
```

Exploratory data analysis was performed on the data using a variety of plots from the ggplot package. Histograms, barplots, and scatterplots provided the best visual representation of the data and valuable insights regarding factors relating to cycle shares. Pearson's Correlational Analysis was performed on the numeric variables within the data, including temperature, number of bike shares, and weather information. A confidence interval of 95% was set for the analysis to best evaluate the p-value and determine any co-linearity between variables. The results were summarized in a correlation matrix as well as a correlation plot. Linear regression analysis was conducted to determine relationships between variables. A model was created based on these relationships to predict ride share behavior.

Analysis and Implications

The correlation analysis indicated several strong positive and negative relationships within the data. Humidity had significant negative correlations with number of cycle shares (-0.84) and the actual temperature (-0.82). The actual temperature was also moderately correlated with the number of bike shares (0.62). There was almost perfect co-linearity between the actual temperature and feels like temperature (0.998). Figure 1 shows the full correlation plot, while Table 2 shows the correlation matrix with p-values.

Figure 1 Correlation Plot for London Data

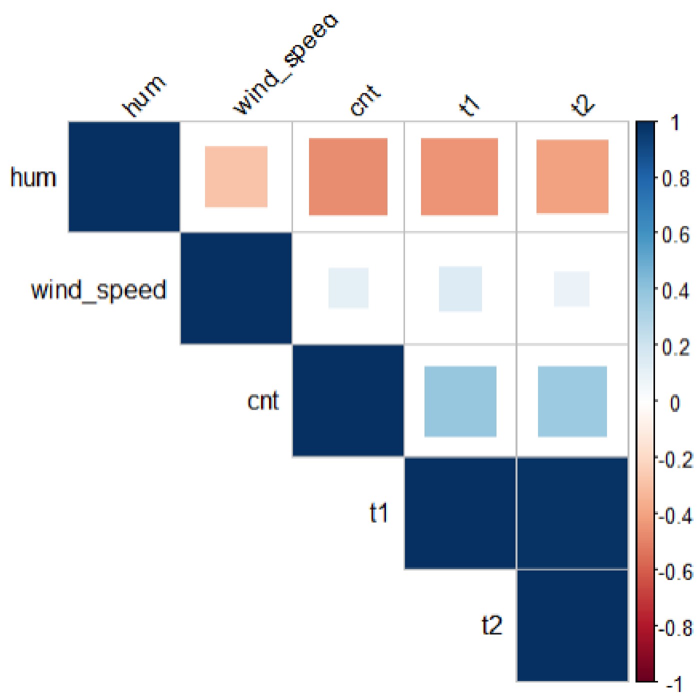


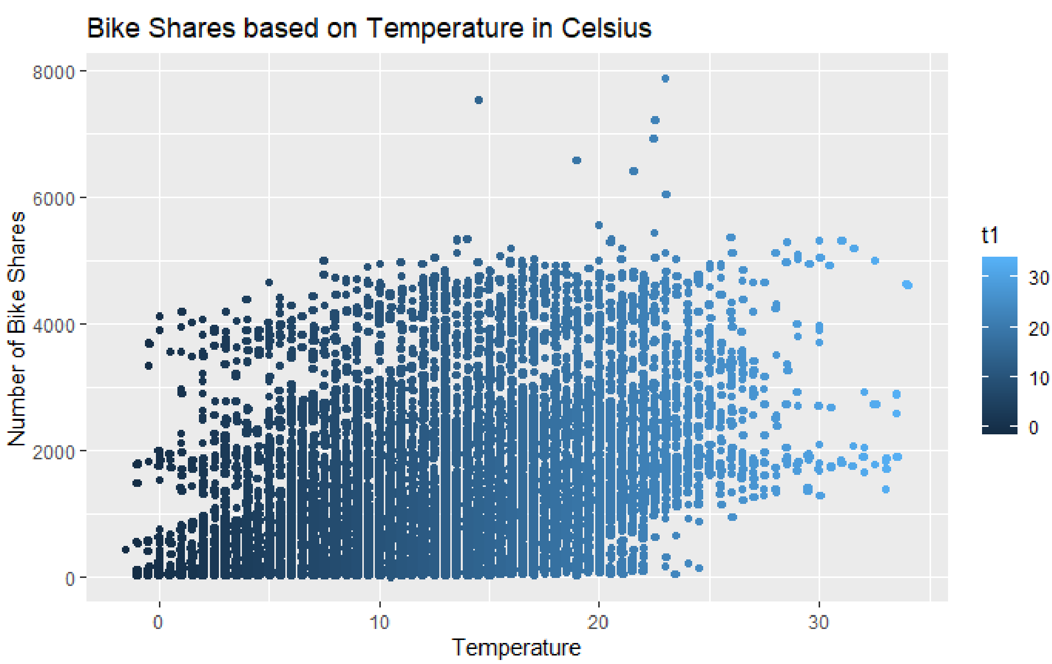
Table 2 Correlation Matrix for London Data

row <chr>	column <chr>	cor <dbl>	p <dbl>
cnt	t1	0.60563108	2.790364e-01
cnt	t2	0.58394783	3.012289e-01
t1	t2	0.99815988	9.472925e-05
cnt	hum	-0.83528686	7.823374e-02
t1	hum	-0.82367014	8.649454e-02
t2	hum	-0.79093824	1.110797e-01
cnt	wind_speed	0.15477614	8.037226e-01
t1	wind_speed	0.09639120	8.774612e-01
t2	wind_speed	0.04001319	9.490672e-01
hum	wind_speed	-0.50394852	3.866541e-01

Exploratory Data Analysis revealed some surprising insights that disproved some the assumptions of the author. There was a hypothesis that bike shares would be used more frequently on weekends for leisure outings. However, it is clear that most cycle shares (71%) occur during weekdays and between the hours of 8-9am and 5-6pm, which correspond to commuting hours. Furthermore, overwhelmingly more cycle shares occur during non-holiday hours (probability of 99.7%).

When examining weather and seasonality, the data suggests that summer and fall are the most popular seasons for bike shares. It also suggests that bike shares decrease in inclimate weather, while they remain steady during cloudy weather. Temperature is the final weather consideration examined in the data. Figure 2 shows the distribution of bike shares based on temperature in degrees celsius. As evidenced by the data, the cycle shares decrease as the temperature approaches 30 degrees Celsuis (86 degrees Fahrenheit).

Figure 2 Bike Shares based on Temperature



Following summary analysis, a model was chosen to predict bike share behaviors (cnt ~ t1 + time + is_weekend). This models uses temperature, the time of day, and whether it is a weekday or weekend to predict the number of bike shares within the city. Table 3 displays the summary statistics for the model and shows a strong linear relationship between the selected variables. The p-value's promixity to zero means that we can reject the null hypothesis and conclude that there is a positive correlation between the variables in the model and number of cycle shares.

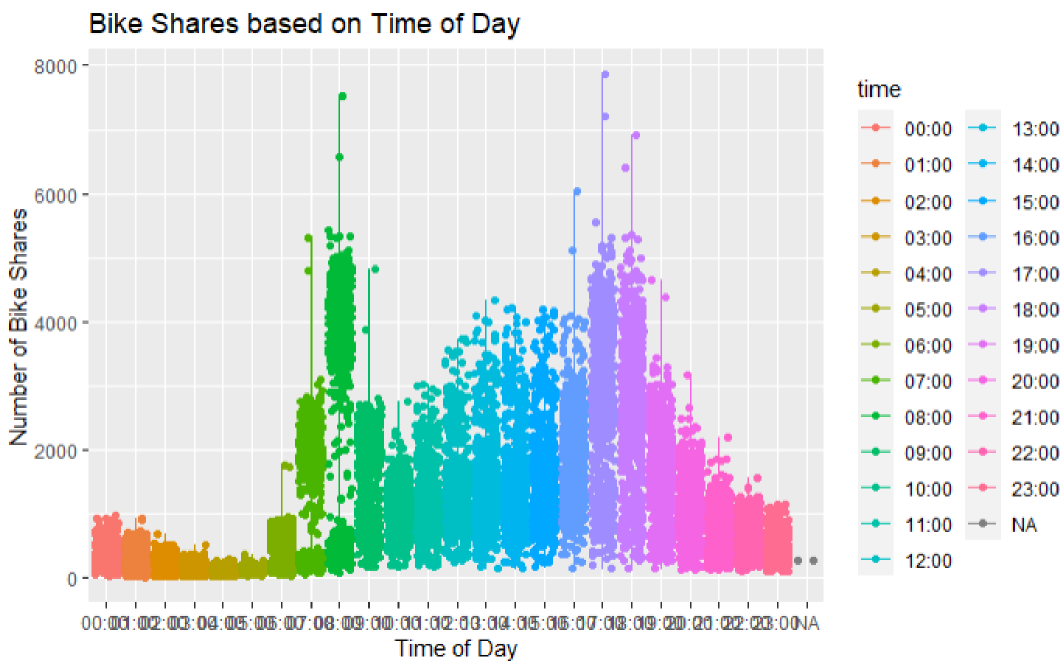
Table 3 Summary Statistics of Model

```
Call:
lm(formula = cnt ~ t1 + time + is_weekend, data = london)

Residuals:
    Min      1Q  Median      3Q     Max 
-2934.5  -274.1   -23.2   273.0  4514.3 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -226.3574    25.0640  -9.031  < 2e-16 ***
t1            51.8466     0.8743  59.303  < 2e-16 ***
time01:00    -76.3240    32.3324  -2.361   0.0183 *  
time02:00   -130.3525    32.3906  -4.024 5.74e-05 ***
time03:00   -159.7166    32.3712  -4.934 8.13e-07 ***
time04:00   -171.5881    32.3744  -5.300 1.17e-07 ***
time05:00   -128.8010    32.3767  -3.978 6.97e-05 ***
time06:00    222.7867    32.3189   6.893 5.64e-12 ***
time07:00   1204.4450    32.3124  37.275  < 2e-16 ***
time08:00   2586.1271    32.3318  79.987  < 2e-16 ***
time09:00   1312.7295    32.3091  40.630  < 2e-16 ***
time10:00    680.6213    32.3585  21.034  < 2e-16 ***
time11:00    730.8413    32.3724  22.576  < 2e-16 ***
time12:00    985.9136    32.3855  30.443  < 2e-16 ***
time13:00   1039.7388    32.4233  32.068  < 2e-16 ***
time14:00    995.0007    32.4408  30.671  < 2e-16 ***
time15:00   1086.9471    32.4308  33.516  < 2e-16 ***
time16:00   1400.7974    32.4057  43.227  < 2e-16 ***
time17:00   2379.6090    32.3997  73.445  < 2e-16 ***
time18:00   2200.5399    32.3715  67.978  < 2e-16 ***
time19:00   1248.9232    32.3538  38.602  < 2e-16 ***
time20:00    686.3624    32.3287  21.231  < 2e-16 ***
time21:00    394.0702    32.3237  12.191  < 2e-16 ***
time22:00    267.4307    32.3259   8.273  < 2e-16 ***
time23:00    132.4982    32.3552   4.095 4.24e-05 ***
is_weekend1 -231.1552    10.3239 -22.390  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 615.2 on 17386 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.6791,	Adjusted R-squared:  0.6786 
F-statistic:  1472 on 25 and 17386 DF,  p-value: < 2.2e-16
```

Insights

Several interesting insights were gained through this analysis. The data seems to show that cycle shares are more likely on weekdays during commuting hours. Figure 3 shows the distribution of bike shares based on time of day. Cloudy weather appears to best predict an increase in bike shares as does season.

Figure 3 Bike Shares based on Time



Limitations

The limitations of this project are two-fold. First, this is the author's first complete data analysis and different methods may process alternative results. Second, bike share analysis has not been analyzed extensively and new insights are being gained with every analysis. The goal of this project was to produce some new insights into the arena but not to be all encompassing. Future analysis can be strengthened by additional data for the London metropolitan area or through analysis of the multiple cities to identify trends. This was the original goal of the author, however, due to missing/incomplete data and difficulty combining large datasets that approach was beyond the scope of this work.

Concluding Remarks

Cycle Sharing is an emerging trend across urban cities in the world. The insights and recommendations for future analysis of this project are a great starting point for predicting and better understanding utilization of this service. Through future work, the author hopes to continue exploring this topic and provide additional insights to those already gained.

References

Centre for Public Impact. (2020). *London's cycle hire scheme.* Retrieved on May 26, 2020, from https://www.centreforpublicimpact.org/case-study/londons-cycle-hire-scheme/

Hendrix, A. (2017). *The economic benefits of bike sharing.* Retrieved on May 26, 2020, from https://medium.com/urbansharing/the-economic-benefits-of-bike-sharing-f69c230e5a9d

Nikitas, A. (2016). *The global bike sharing boom – why cities love a cycling scheme.* Retrieved on May 26, 2020, from https://theconversation.com/the-global-bike-sharing-boom-why-cities-love-a-cycling-scheme-53895

Transportation for London. (2020). *Santander Cycles.* Retrieved on May 26, 2020, from https://tfl.gov.uk/modes/cycling/santander-cycles