**Bay Wheels Bike-Sharing Station Segmentation and Trip Count Analysis**

**By**

**Sam Shojaei**

**May 2020**

# Introduction

Bike-sharing is an increasingly evolving concept across the globe, offering diverse benefits such as flexible mobility, reduced fuel use, the corollary reductions in emissions, and increased physical activity levels (*1*). Essentially, bike-sharing enables users to enjoy the advantages of biking as an active mode of transportation, without the complications of private bike ownership (i.e. purchase and regular maintenance). Appropriate station placement is recognized as a critical contributor to bike-sharing success (*2*). Literature suggests that stations in the vicinity of bicycle infrastructure, high density areas, and commercial facilities, experience higher ridership (*3*–*7*). In this line, this report sets out to deploy the location data provided by Foursquare and explore the relationship between types of venues surrounding Bay Wheels bike-sharing stations, and their average daily and morning rush hour trip counts. The insights derived from these analyses can be used for varied local purposes such as to improve service levels of stations, better manage bike-sharing rebalancing, and optimize stations siting for enhanced system performance, as well as insights for new bike-sharing systems.

# Data

Bay Wheels is a bike-sharing system operated by Lyft®, in San Francisco, East Bay, and San Jose. In this research, Foursquare location data are used to explore venues surrounding Bay Wheels bike-sharing stations. Subsequently, the publicly available anonymized trip logs of Bay Wheels (*8*) are used to investigate average monthly total and morning rush hour trips across May through August 2019.

# Methodology

In this project, I use Foursquare API to explore venues within 150 m radius from each Bay Wheels station, and use the data to segment/ cluster the stations. Guidelines generally recommend that bike-sharing stations be placed no farther apart than 1000 ft (~300 m) (*9*); in accordance, assuming the Bay Wheels stations are perfectly distanced, the 150 m radius is selected. Using the one hot encoding approach to convert the venues into binary variables, and thus computing the average frequency of each venue around each station, stations were then clustered according to their similarity in nearby venues. It must be noted different number of clusters (1 through 15) were tested using the elbow method in conjunction with the silhouette score, to find optimal number of clusters. Subsequently, anonymized trip logs of Bay Wheels bike-sharing stations were retrieved across May through August 2019 (*8*), i.e. months with favorable weather conditions for unsheltered transportation by biking .Next, average daily as well as average morning rush hour started bike-sharing trips at stations were looked into with respect to the optimal number of clusters.

# Results

## Optimal Number of Clusters

In order to find the optimal number of clusters from the range of 2 through 15, the elbow method was used in conjunction with silhouette score. In the elbow method, for each number of clusters, the Euclidean distances between all nodes, each node represented by one hot encoded venues set, and its assigned cluster centroid are summed up as the error term. The elbow method hinges on the fact that the optimal cluster number is the one at which the error term decrease slope starts to turn relatively milder. However, as there could be multiple such points, or other inclarities involved, the silhouette score is also used commonly along with the elbow method; in brief, the silhouette score lies between -1.0 and 1.0, with the score closer to 1.0 being representative of the optimal number of clusters. According to Figure ‎4‑1, optimal number of clusters is set to 10, as at this is the point of the highest relative silhouette score.

|  |  |
| --- | --- |
| 1. The Elbow Method | 1. The Silhouette Score Method |

**Figure ‎4‑1** Analysis to select the optimal number of clusters

## Examining the optimal clusters

With the optimal number of clusters, *K = 10*, here I examine to check what distinct pattern each cluster can represent. For this purpose, I explore the breakdowns of the first three most common venues whose results are can be found in ‎Appendix A. Table ‎4‑1 presents the number of stations and the most common venue types in each cluster.

**Table ‎4‑1** Optimal clusters (k = 10) and their most common venue types

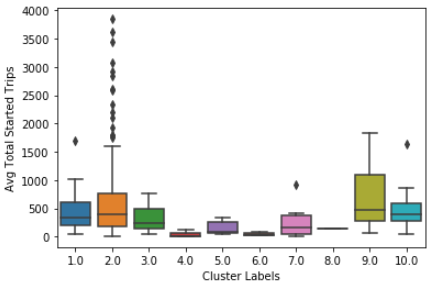
|  |  |  |
| --- | --- | --- |
| **Cluster** | **Number of Stations** | **Most Common Venue Types** |
| **1** | 21 | Park |
| **2** | 334 | Coffee shop; yoga studio |
| **3** | 4 | Art gallery; yoga studio |
| **4** | 5 | Grocery store; yoga studio |
| **5** | 5 | Gym; yoga studio |
| **6** | 12 | Mexican restaurant |
| **7** | 8 | Play ground; yoga studio |
| **8** | 1 | Food & drink shop; yoga store |
| **9** | 11 | Food truck; yoga studio |
| **10** | 14 | Café; yoga studio |

## Trip Count Analysis

Having determined the optimal clusters of stations based upon nearby venues, I set out to analyze average trip counts in relation to the clusters. For this purpose, the stations’ anonymized trip counts of May through August 2019 are used (*8*). Analysis of average monthly total and morning rush hour started trips are investigated in the following subsextions.

### Average monthly started trips

Average of total started trips at each station from May through August 2019 are used to calculate the average monthly started trips at each station; these averages are then used to visualize the boxplots of the optimal clusters to obtain an idea of how the values vary across clusters. Figure ‎4‑2 displays this visualization.

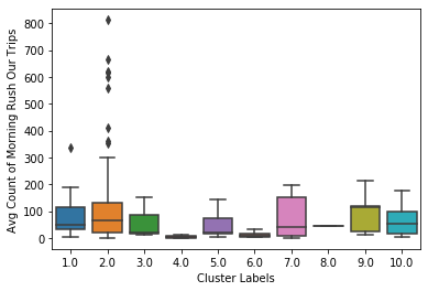


**Figure ‎4‑2** The boxplot of average monthly started trips for each of the optimal clusters

As can be seen in Figure ‎4‑2, cluster 1, 2, 7, and 10 have outliers. However, there are only 1 outlier in each of the clusters 1,7, and 10, and cluster 2 also has 15 outliers among a total of 334 stations clustered. Accordingly, analyzing average monthly started trips according to clustering/segmentation based on nearby venues indicates consistent and reliable outcome.

### Average monthly morning rush hour trips

After average monthly total started trips, average monthly morning rush hour started trips are looked into. For this analysis, I used the recorded time of started trips at the stations from the datasets of May through August 2019 and extracted the trips during morning rush hour in San Francisco, i.e. 7:00 AM – 9:00 AM. Similar to the previous subsection, the boxplot of the average monthly morning rush hour trips is plotted which is displayed in Figure ‎4‑3.



**Figure ‎4‑3** The boxplot of average monthly morning rush hour started trips for each of the optimal clusters

According to Figure ‎4‑3, clusters 1 and 2 have only 1 and 13 outliers, respectively. Thus, it can be concluded that, analyzing average monthly morning rush hour started trips on the premises of stations segmentation by nearby venue types is also consistent.

# Conclusion

In this project, Bay Wheels bike-sharing stations are clustered based on nearby venue types, which was collected using Foursquare API. After determining the optimal number of clusters according to error term and silhouette score investigation, I set out to analyze the trip counts data using publicly available anonymized trip logs of the Bay Wheels system. Specifically, I analyzed average monthly total and morning rush hour started trips based on the clusters determined. Analysis showed that the clusters can be used to reliably predict the average monthly total and morning rush hour started trips. Such predictions can be used by decision makers and operators for planning as well as managerial practices. For example, this tool can be used for informed decision on opening stations with high anticipated trips, or for managing rebalancing of fleets in the bike-sharing system. In terms of future direction, time series analysis of bike trips in the Bay Wheels system based upon the optimal clusters can also yield more detailed insights. Furthermore, incorporating other attributes such as land use and land cost, and dominant user demographics, can be used alongside the nearby venue types for station clustering.

##### Clusters of *K = 10*

###### Cluster 1

**Table A‑1** Most common venues in cluster 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Park** | 19.05 | **Park** | 33.33 | **Yoga Studio** | 23.81 |
| **Sports Bar** | 9.52 | **Chinese Restaurant** | 4.76 | **Park** | 14.29 |
| **Pool** | 9.52 | **Bus Station** | 4.76 | **Sushi Restaurant** | 9.52 |
| **Furniture / Home Store** | 9.52 | **Street Art** | 4.76 | **Electronics Store** | 4.76 |
| **Paper / Office Supplies Store** | 4.76 | **Cosmetics Shop** | 4.76 | **Pier** | 4.76 |

###### Cluster 2

**Table A-2** Most common venues in cluster 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Coffee Shop** | 9.88 | **Yoga Studio** | 6.29 | **Yoga Studio** | 4.49 |
| **Food Truck** | 3.29 | **Coffee Shop** | 5.39 | **Bar** | 3.89 |
| **Yoga Studio** | 2.99 | **Café** | 3.29 | **Grocery Store** | 2.99 |
| **Café** | 2.99 | **Food Truck** | 3.29 | **Sandwich Place** | 2.99 |
| **Pizza Place** | 2.69 | **Pizza Place** | 2.99 | **Gym** | 2.99 |

###### Cluster 3

**Table ‎4‑4** Most common venues in cluster 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Art Gallery** | 75.0 | **Yoga Studio** | 50 | **Yoga Studio** | 50 |
| **College Theater** | 25.0 | **Art Gallery** | 25 | **Flea Market** | 50 |
| **-** | - | **College Academic Building** | 25 | **-** | - |

###### Cluster 4

**Table ‎4‑5** Most common venues in cluster 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Grocery Store** | 80 | **Yoga Studio** | 40 | **Yoga Studio** | 40 |
| **Supermarket** | 20 | **Sporting Goods Shop** | 20 | **Exhibit** | 40 |
| **-** | - | **Dessert Shop** | 20 | **Skating Rink** | 20 |
|  |  | **Grocery Store** | 20 | **-** | - |

###### Cluster 5

**Table ‎4‑6** Most common venues in cluster 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Gym** | 100.0 | **Yoga Studio** | 60 | **Exhibit** | 60 |
|  |  | **Park** | 20 | **Yoga Studio** | 40 |
| **-** | - | **Middle Eastern Restaurant** | 20 | **-** | - |

###### Cluster 6

**Table ‎4‑7** Most common venues in cluster 6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Mexican Restaurant** | 58.33 | **Mexican Restaurant** | 16.67 | **Mexican Restaurant** | 25 |
| **Coffee Shop** | 8.33 | **Hot Dog Joint** | 8.33 | **Flea Market** | 25 |
| **Dance Studio** | 8.33 | **Liquor Store** | 8.33 | **Yoga Studio** | 16.67 |
| **Deli / Bodega** | 8.33 | **Vietnamese Restaurant** | 8.33 | **Chinese Restaurant** | 8.33 |
| **Locksmith** | 8.33 | **Ice Cream Shop** | 8.33 | **Latin American Restaurant** | 8.33 |

###### Cluster 7

**Table ‎4‑8** Most common venues in cluster 7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Playground** | 37.5 | **Playground** | 37.5 | **Yoga Studio** | 25 |
| **Dog Run** | 25 | **Yoga Studio** | 25 | **Fish Market** | 25 |
| **Yoga Studio** | 12.5 | **Dog Run** | 12.5 | **Southern / Soul Food Restaurant** | 12.5 |
| **Fried Chicken Joint** | 12.5 | **Plaza** | 12.5 | **Gym** | 12.5 |
| **Pool** | 12.5 | **Concert Hall** | 12.5 | **Playground** | 12.5 |

###### Cluster 8

**Table ‎4‑9** Most common venues in cluster 8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Food & Drink Shop** | 100.0 | **Yoga Studio** | 100.0 | **Fish & Chips Shop** | 100.0 |

###### Cluster 9

**Table ‎4‑10** Most common venues in cluster 9

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Food Truck** | 81.82 | **Food Truck** | 18.18 | **Fish & Chips Shop** | 18.18 |
| **Pool** | 9.09 | **Yoga Studio** | 18.18 | **Yoga Studio** | 18.18 |
| **Brewery** | 9.09 | **Non-Profit** | 9.09 | **Mexican Restaurant** | 9.09 |
| **-** | - | **Metro Station** | 9.09 | **Café** | 9.09 |
| - |  | - | - | **Art Gallery** | 9.09 |

###### Cluster 10

**Table ‎4‑11** Most common venues in cluster 10

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1st Most Common Venue** | **Percentage** | **2nd Most Common Venue** | **Percentage** | **3rd Most Common Venue** | **Percentage** |
| **Café** | 35.71 | **Food Truck** | 18.18 | **Fish & Chips Shop** | 18.18 |
| **American Restaurant** | 14.29 | **Yoga Studio** | 18.18 | **Yoga Studio** | 18.18 |
| **Pharmacy** | 14.29 | **Non-Profit** | 9.09 | **Mexican Restaurant** | 9.09 |
| **Flea Market** | 7.14 | **Metro Station** | 9.09 | **Café** | 9.09 |
| **Furniture / Home Store** | 7.14 | **Museum** | 9.09 | **Art Gallery** | 9.09 |

**REFERENCES**

1. Shaheen, S., S. Guzman, and H. Zhang. Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2143, 2010, pp. 159–167. https://doi.org/10.3141/2143-20.

2. Lin, J.-R., and T.-H. Yang. Strategic Design of Public Bicycle Sharing Systems with Service Level Constraints. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 47, No. 2, 2011, pp. 284–294. https://doi.org/10.1016/j.tre.2010.09.004.

3. Faghih-Imani, A., N. Eluru, A. M. El-Geneidy, M. Rabbat, and U. Haq. How Land-Use and Urban Form Impact Bicycle Flows: Evidence from the Bicycle-Sharing System (BIXI) in Montreal. *Journal of Transport Geography*, Vol. 41, 2014, pp. 306–314. https://doi.org/10.1016/j.jtrangeo.2014.01.013.

4. El-Assi, W., M. Salah Mahmoud, and K. Nurul Habib. Effects of Built Environment and Weather on Bike Sharing Demand: A Station Level Analysis of Commercial Bike Sharing in Toronto. *Transportation*, Vol. 44, No. 3, 2017, pp. 589–613. https://doi.org/10.1007/s11116-015-9669-z.

5. Rixey, R. A. Station-Level Forecasting of Bikesharing Ridership. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2387, No. 1, 2013, pp. 46–55. https://doi.org/10.3141/2387-06.

6. Faghih-Imani, A., and N. Eluru. Analysing Bicycle-Sharing System User Destination Choice Preferences: Chicago’s Divvy System. *Journal of Transport Geography*, Vol. 44, 2015, pp. 53–64. https://doi.org/10.1016/j.jtrangeo.2015.03.005.

7. Wang, X., G. Lindsey, J. E. Schoner, and A. Harrison. Modeling Bike Share Station Activity: Effects of Nearby Businesses and Jobs on Trips to and from Stations. *Journal of Urban Planning and Development*, Vol. 142, No. 1, 2016, p. 04015001. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000273.

8. System Data | Bay Wheels | Lyft. https://www.lyft.com/bikes/bay-wheels/system-data. Accessed May 25, 2020.

9. (No Title). https://nacto.org/wp-content/uploads/2015/09/NACTO\_Walkable-Station-Spacing-Is-Key-For-Bike-Share\_Sc.pdf. Accessed May 25, 2020.