Abstract

Bike rentals in Seattle have elaborate logistics behind rental site locations and bike allocation among stations. This paper used data from recent history regarding bike stations and bike transfers between stations in order to make decisions about bike distribution methodology. First, we convert the data into an directed weighted graph and use first-order measures to determine the importance of bike sharing stations. Next, we use a metric combining geographical nearness with station volatility in order to calculate the locations and respective regions of three bike distribution management centers. In the final sections, we explore the potential efficiency of bike distribution as well as the effect of introducing more variables into the analysis. We discuss the strengths, weaknesses, and potential future of our model for bike distribution in Seattle.

Metric for Bike Station Importance

We created a directed, weighted graph with the number of rides from one bike station to another acting as weights. After implementing this graph in Python and Excel using an adjacency matrix, we developed a metric for bike station importance that relied only on first-order analysis. Higher-order graph analysis methods fail in graphs which are nearly saturated with edges, since the measures become identical among nodes. This was empirically observed while calculating the Eigenvector centrality of our data, given by

$$E(x_v) = \frac{1}{\lambda} \sum_{E \in G} L_{v,t} x_t$$

where λ is an eigenvalue of the adjacency matrix, and $L_{v,t}$ is the edge weight from v to t (defined to be 0 if there is no edge.) Both this analysis technique, as well as simple visual inspection, confirmed that out graph was nearly complete. This makes sense, as nearly all bike stations had traffic between them.

Thus, only first-order graph analysis revealed valuable information about the bicycle network. Valued directed node density, given by

$$D = \frac{\sum_{i,j \in G} L_{ij}}{\frac{N!}{(N-2)!}}$$

was used to characterize the importance of each bicycle station in Seattle. The result of this analysis can be found in Figure 1, on the following page, where node importance (density) is represented by circle size and the ratio of average daily bicycle arrival versus departure is represented by each circle's internal pie chart.

Distribution Center Placement

We aimed to optimize two factors in our placement of distribution centers: geographic collocation and the number of bikes a station had in excess or in dearth, on average, at the end of every day. We used a modified version of the k-means clustering algorithm; the standard version works as follows:

1. Select three locations randomly, contained within with a region defined by the maximum and minimum of all bicycle station x and y coordinates
2. Assign each stations to nearest cluster whose mean has the least squared Euclidean distance.
3. Calculate the new means to be the centroids of the observations in the new clusters
4. The algorithm converges stations assignments no longer change. All stations are have been assigned to the optimal distribution center

However, we redefined the centroid by appropriating center of mass equations for n points in 2D space:

$$\overline{x} = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_n x_n}{m_1 + m_2 + \ldots + m_n}, \quad \overline{y} = \frac{m_1 y_1 + m_2 y_2 + \ldots + y_n x_n}{y_1 + y_2 + \ldots + y_n},$$

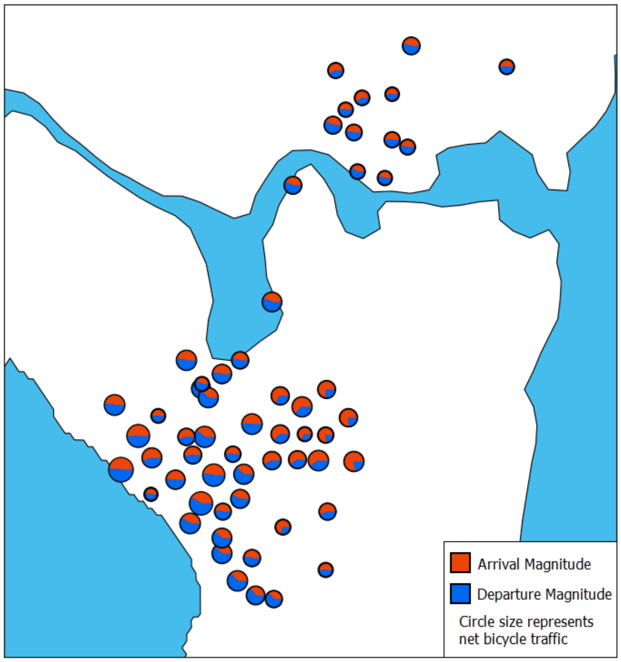

Figure 1: Station Importance and Flow Rates

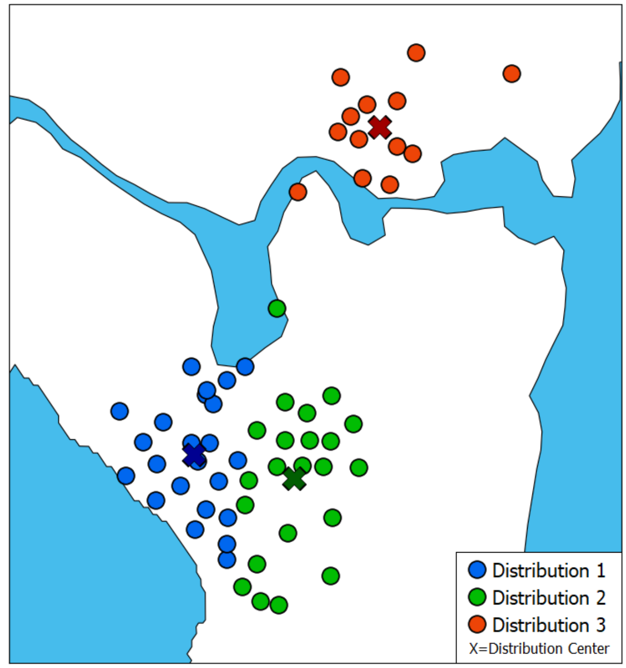

Figure 2: Distribution Centers and Attributed Stations

where mass of a node is the absolute value of its average change in bikes per day, thus making it important for distribution centers to be close to more active nodes. The result of this analysis can be seen above in Figure 2, where the three colors of circles represent bicycle stations belonging to different distribution centers. The distributions centers are also marked in their respective color; their location calculated through the aforementioned k-means partitioning. After determining the optimal locations for distribution centers in Seattle, we are able to use their location as a metric by which to evaluate any other set of distribution center locations. This can be done by evaluating their distance from the ideal locations.

Simulation and Further Investigation

Given that the distribution centers will distribute bikes among each other, a simulation can be run simply treating the distribution centers as a single entity. The goal of such a simulation is to find what conditions of daily distribution activity must be met in order to reach a steady state in bike movement, where no node ever reaches zero bikes. The solution was reached by having distributors predict and order the stations with the highest excess and highest need of bikes, and then target the first n of each list daily for redistribution. Targeting exactly the first 15 from each list produced a steady state, where numbers greater or smaller did not. This implies fine-tuning of bike allocation is certainly in order for any distribution center, although only half of all stations may need to be visited daily.

We also investigated the effect of rain on bike sharing. Using daily precipitation data from the National Oceanic and Atmospheric Administration, we were able to associate each trip in the data set with an amount of rain. Rain affected average bike traffic at every station, most notably with the Pier 66 station having nobody arriving or leaving on rainy days. This suggests a dynamic model of bike allocation taking into account weather could more efficiently distribute bikes, though time did not permit us to develop such a model.

Further work can be done by adjusting the Euclidean metric used in the k-means distribution model. This metric can be adapted to incorporate bicycle flow through nodes, thus creating 1:1 relation of input to output for each distribution center and reducing the need for interaction between these centers. Additionally, the simulation can be made more accurate by incorporating transfers between the distribution centers. These ideas have yet to be explored and leave much to be discovered regarding bicycle sharing in Seattle.