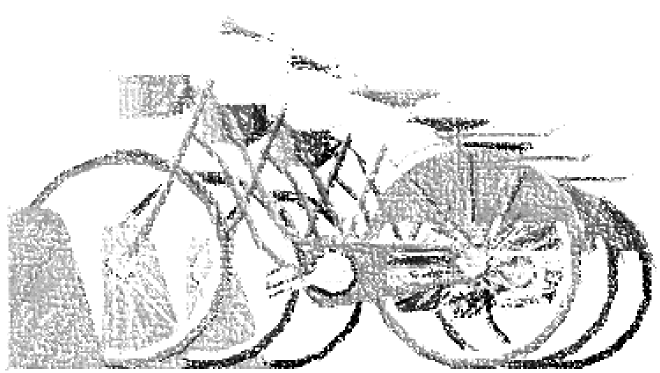
ISYE 6501

Increasing Bike-Share Efficiency

Spring 2024



Contents

I. Project Instructions

The web sites https://www.sas.com/en_us/customers.html, https://www.ibm.com/casestudies/search?search, and https://www.informs.org/Impact/O.R.-Analytics-Success-Stories (among others) contain brief overviews of some major Analytics success stories. In this course project, your job is to think carefully about what analytics models and data might have been required.

(1) Browse the short overviews of the projects. Read a bunch of them – they're really interesting. But don't try to read them all unless you have a lot of spare time; there are lots!

(2) Pick a project for which you think at least three different Analytics models might have been combined to create the solution.

(3) Think carefully and critically about what models might be used to create the solution, how they would be combined, what specific data might be needed to use the models, how it might be collected, and how often it might need to be refreshed and the models re-run. DO NOT find a description online (or elsewhere) of what the company or organization actually did. I want this project to be about your ideas, not about reading what someone else did.

(4) Write a short report describing your answers to (3).

II. Background

Case study: Increasing Bike-Share Efficiency
https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Increasing-Bike-Share-Efficiency

The case study chosen centers around increasing bike-share efficiency. It specifically focuses on implementing analytics and operations research to replenish and rebalance the bike docking stations.

Bike sharing programs are becoming increasingly popular in cities around the world. Motivate is the largest operator of these services in many major U.S. cities, including New York, Chicago, and San Francisco. In this case, they worked with researchers from Cornell University, Uber, and Lyft to develop a system for balancing the number of bikes at each docking station and creating a scheme to crowd-source the process (Informs, 2022).

This paper will approach this case by breaking down the situation and proposing various models to analyze and solve each part of the problem.

III. Defining the Problem

This problem can be split into three questions:

- What is the demand for bike sharing services at different times at each location?
- How can the placement of bike docking stations be optimized to maximize usage and accessibility?
- How can the distribution of bikes between stations be optimized to maintain balanced inventory levels so that the demand for both bikes and empty docks is met?

Each of these questions will be addressed with a different model, as described below.

IV. Demand for Bike Sharing Services

In order to investigate how to efficiently replenish and rebalance docking stations, it is essential to first explore how demand for bike sharing services varies by the time of year and day. The results of this model can later inform other models explored in this paper.

Data:

To predict the demand for bike-sharing services, the following historical usage data would be required for each station:

- Timestamp of each bike rental
- Station locations
- Bike availability
- User demographics (age, zip code)

Data on external factors would also be useful to consider:

- Seasonal weather data
- Public transport schedules
- Traffic patterns
- Large public events

Justification:

The timestamp of each bike rental would also provide the overall number of rentals. This can be combined with the other data to help forecast demand. For example, the number of rentals at each station location at different times of the day would indicate how the ridership varies throughout the day. The number of rentals and bike availability at a station would also help indicate how busy that station is throughout the day.

The user demographics data that Motivate has access to is likely limited, as little information is needed to rent a bike. Age may need to be provided to ensure the rider is over eighteen, as well as zip code in order to bill the rider.

The external factors listed above would likely have an impact on demand and should be considered along with the historical usage factors. For instance, in areas with frequent high automobile traffic congestion, bike sharing services may be in higher demand.

Model:

A time series model would be appropriate in this situation, as the aim is to forecast demand over time. The seasonal autoregressive integrated moving average model (SARIMA) would be useful in this case, as it also takes into account seasonality. It combines several simpler models, such as the autoregression model and moving average model, to create a more complex model (Peixeiro, 2023).

Specifically, this model would be used to predict demand at each docking location for bike-share services based on the station location, median age of past riders, common zip codes of past riders, the current bike availability at that station, the current weather, the current schedule of public transport, and the presence of any current major events.

This model would have to be refreshed very regularly, ideally on a daily basis, to account for changing patterns in demand due to the season, weather, events, and public transport schedules.

V. Optimizing Dock Placement

Now that demand can be predicted for bike-sharing services at each dock, it is possible to explore how to optimize dock placement in order to meet ridership demand and potentially even increase ridership.

Data:

To optimize dock placement in a city, the following data would be required at a micro level for each part (zip code or x square miles) of the city:

- Population density
- Traffic patterns
- Business locations
- Existing transportation infrastructure
- Cost of renting or acquiring dock location

Motivate's total budget for docks would be required as well.

The insights gained from the previous model could also be used to inform dock placement decisions. Specifically, the forecasted demand at each station will be used in this model.

Justification:

Areas with a relatively higher population density and higher concentration of businesses would likely require more docks within a smaller area to meet the demand at certain times of the day.

Traffic patterns were considered in the previous model, and they can also be considered in this model to inform dock placement. Areas with high pedestrian traffic and areas close to public transport hubs would be preferred dock locations. Bicycle traffic patterns can indicate common routes so that the model considers both optimal pick up and return locations.

For ease of riding and safety reasons, it is preferential to place docks in bikeable areas, ideally with existing protected bike lane infrastructure, but this has to be balanced with placing docks in highly trafficked areas that have high demand.

In addition, the ideal dock locations have to be balanced with the cost of renting or acquiring the space for a dock on the road. This would have to be optimized to stay below the company's total dock station budget while reaching the most riders.

The forecasted demand in each area that was looked at in the previous model would also be used here to optimize dock placement. The other factors listed above, however, could also induce demand, hence their inclusion in this model.

Model:

An optimization model can be used to guide dock placement decisions. The model would seek to maximize ridership, constrained by areas with a minimum standard of existing bicycle infrastructure and the total dock budget. It would take into account the data discussed above, such as population density, traffic patterns, business locations, and bike share demand by area. The resulting variables would be the number of dock stations and the geographic placement of them.

The optimization model would have to take a multiple-objective approach, as it is constrained by two variables (bicycle infrastructure and total dock budget). This avoids the model suggesting that docks be placed at every block in the city, which would likely run well over budget, as well as suggesting that docks are placed on dangerous heavily trafficked roads.

Spatial optimization methods are typically used to solve a problem of geographic nature. The greedy algorithm is an example of this. It can be used in conjunction with an optimization model to start with a feasible solution and continually test and compare other solutions until it finds a more optimal one. Although this algorithm can take some time, it usually produces effective results when applied repeatedly (Delmelle, 2010).

This model would require less frequent updates, such as once every six months, since most of the data considered in this model is slow to change. Therefore, a greedy algorithm would be appropriate. Changes in population, traffic patterns, businesses, and urban landscape in general usually take place over longer time periods. It is also not feasible to continuously change the dock locations often once they are initially set.

VI. Optimizing Bike Redistribution

The demand for bike sharing services has now been identified and the dock locations have been decided, so the redistribution of bikes among the various docks can now be optimized. This is an important part of the problem, as a lack of supply of bikes or a lack of supply of available docks to park at can significantly impact ridership. As the case study mentioned, there would need to be a crowdsourcing component to this solution.

Data:

To optimize the distribution of bikes at docking stations throughout a city, the following data would be required:

- Real-time bike and dock availability at each station
- Current usage patterns
- Predicted future demand (from the first model)
- The location of each dock (from the second model)
- Availability and location of crowdsourced riders who participate in rebalancing
- Budget for crowdsourcing incentives

Justification:

The current usage patterns and predicted future demand would indicate where bicycles are most needed at any point in time. The usage patterns would show the current demand, what routes are common, and indicate where return docks are most needed at any point in time.

The location of each dock and availability of crowdsourced riders who participate in rebalancing would be needed to figure out how to get bikes to where they are most needed.

Riders could be given incentives, such as a free ride, to participate in rebalancing. The total budget for crowdsourcing incentives would need to be considered. This budget could be variable, based on the demand at any given time.

The real-time availability of bikes and docks would be required in order to analyze much of the other data listed above.

Model:

An optimization model can again be used to rebalance and redistribute bikes across the bike sharing network. The model would have multiple objectives: it would seek to maximize bike inventory at high-demand pickup locations, while maximizing available docks at high-demand return locations, while minimizing the distance traveled by crowdsourced rebalancing participants. The constraint would be the budget for crowdsourcing incentives. The model would consider all the variables listed above.

As this is another problem of geographic nature, a spatial optimization model can be utilized. Simulated annealing would be appropriate in this situation. It is similar to the greedy algorithm but accepts solutions that are worse than the best candidate and continues to look for an optimal solution even after finding the current 'best' solution. This makes it a preferred method for real-time monitoring, as is needed in this situation (Martínez & Cao, 2018).

This model would need to be run very often, ideally every hour, to continuously adapt to changes in dock and bike availability, current usage patterns, and the availability of crowdsourced rebalancing participants.

VII. Conclusion

The three models discussed in this paper can be combined to solve the overall task of using analytics to more efficiently rebalance and redistribute bikes across Motivate's bike-share network. If run at the specified frequency, Motivate would be able to use these models to predict future demand, optimize their dock placement, and optimize bike redistribution.

VIII. References

Delmelle E. (2010). Spatial Optimization Methods. In: B. Wharf (Ed). Encyclopedia of Human Geography: 2657-2659

Informs. (2022, March 8). *Increasing bike-share efficiency*. INFORMS. https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Increasing-Bike-Share-Efficiency

Martínez, C. M., & Cao, D. (2018). 2.4.4.4 Simulated annealing. In *Science Direct*. essay, Butterworth-Heinemann.

Peixeiro, M. (2023, August 1). *The Complete Guide to Time Series models*. Built In. https://builtin.com/data-science/time-series-model