Project: Increasing bike-share efficiency
https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Increasing-Bike-Share-Efficiency

Riding bicycle is a eco-friendly alternative transportation option for commuting. Researchers want to add additional bike docks to improve the use of shared bike. This report provides solution by solving the following issues:
- combination of analytical methods required to optimize bike docks
- the collection and frequency of data needed for modeling
- the frequency of model re-run

1. Analysis methods

The analysis modeling used are Exponential Smoothing, Linear Regression, and Optimization Model. These are combined in the steps: Firstly, Exponential Smoothing is used to forecast the future daily bike sharing demand and filter out noise and outliers in data. Secondly, Network Analysis is used to find additional bike docks locations. Thirdly, optimization methods is used to find the optimal new bike docks locations for maximizing bike docks usage.

Exponential Smoothing is adopted as the first step as it can capture the trend, cyclical variations and seasonality in the data. It can filter out outliner data such as the high demand on special event date. It can also seasonally adjust the the data such as school holidays in the year, weekend in the week etc. Moreover, it can identify the trends in time series in order to forecast the future daily bike sharing demand. It gives insights into how the demand change over time. The smoothing factor - alpha parameter needs to be determined in order to give weight to the most recent data, there is more weight to the most recent data when the value of alpha close to 1 and there is more weight to the past data when the value of alpha close to 0. The accuracy of the model can be tested by comparing the predicted values to the actual values. Once the appropriate alpha parameter, the exponential smoothing model can predict the future demand of bike docks by using the recent actual value. As a result, the demand of shared bike at each location is calculated and will be the input for the Network Analysis and Optimization Model. An alternative modeling for predicting the demand of bike docks is logistic regression. It can analyze the relationship between demand and various factors. Logistic regression is less suitable as it is used for binary classification problems.

Network Analysis is then used to evaluate the the accessibility of potential new shared bike dock locations. The network analysis can create a nodes to represent bike dock locations and other important locations. The edge represent the bike routes between the locations. The analysis is based on the criteria such as the demand of sharing bikes, bike routes and other data detailed in the next section. The weight of the criteria factors can be identified via Principal Component Analysis (PCA). The constraints are minimum and maximum distance between bike docks, the availability of bike lanes, the maximum number of bike docks and the local government's restriction on bike docks location. As a result, new locations for shared bike docks are suggested and will be input for the Optimization Model. An alternative modeling for finding additional bike docks is KNN Clustering. It can be used to identify areas with high demand for bike sharing, which

could suggest potential locations for new bike docks. The Clustering method is not used since it does not consider the accessibility. The bike dock locations are not just data points to be clustered, these are physical objects that accessible to people. Network Analysis considers distance, connectivity, accessibility and demand.

Optimization Model is used to determine optimal locations of bike docks. It aims to maximize the efficiency of bike docks. The constraints include the capacity limitation of the number of bike docks at each location, the cost of supplying and installing new bike docks. The optimization model can be tuned and run again until the target efficiency of bike docks is achieved. The demand of the new shared bike locations can be calculated from the exponential smoothing model. Alternative modeling to find optimal locations of bike docks is Decision Trees. It can build a model to predict the optimal locations for bike docks based on the importance of factors. However, it may not provide the optimal solution. Therefore Optimization Model is able to find the best solution among a set of feasible alternatives.

2. Data collection

The time series data are required for Exponential Smoothing for forecasting the future bike sharing demand. The historical data from the bike sharing company's record: number of times the shared bike docks used, shared bike route. These time series data need to be collected daily. The users demographic such as car ownership, bike ownership, age, education and frequency of use of shared bikes. The users information can be collected via in-app surveys and only need to be collected once. The public released data as special event date, vehicular traffic and weather data and these data can be collected automatically, therefore, these data can be refreshed daily at low cost.

The data required for forecasting locations of new bike docks Network Analysis are: the number of bike docks used in the area, frequency of bike docks usage from the bike sharing company's business record. Moreover, the public released data such as the bike docks' distance from the nearby train stop or bus stop, the bike docks' distance from the popular points of interest such as shops, offices, parks and schools, route of potential bike docks to the existing bike network. The users' suggested new locations can be collected from the in-app surveys. The bike docks usage needs to be collect daily and rest of data can be collect just once.

The data for determine the optimal location of bike docks via Optimization Model are: current number of bike docks at each location, number of bikes available at each location at different times of day, the maximum use of bike docks during the day and the customer feedback regarding the availability of the bike docks at each location. These data can be collected from the bike sharing company's business record and need to be refreshed when the data are updated. Moreover, the demand of the bike docks from the result of exponential smoothing, and the potential locations of bike docks in the network analysis are needed and may need to be refreshed monthly. The data are collected monthly due to high cost of installing bike docks.

3.0 Re-run Analysis Models

The Exponential Smoothing may need to be re-run when the data are refreshed. Moreover, if the demand patterns for bike sharing changes, the model needs to be re-run to ensure that it captures these changes. In addition, the model needs to be updated when the accuracy of the model's forecasts begin to degrade.

The Network Analysis may need to be re-run to reflect the changes in infrastructure such as road network, public transport service. In addition, the frequency of re-running Network Analysis depends on the rate of change in data collected. If these factors are relatively stable, the re-run of network analysis is less frequently such as quarterly.

The Optimization Model needs to be re-run when the variable factors are updated. The first factor is the change in demand, the optimization model needs to be re-run to ensure that the bike dock locations still meet the needs of users. The new set of optimal bike dock locations need to be find when the cost of installation bike docks is changed. More optimal bike dock location may be find when the cost of installation decreases. The optimization model can be re-run monthly, as the data is collected monthly.

In summary, the budget of the modeling program and frequency of data updates will indicate how often the models need to be re-run.