Course Project – ISYE6501

Background: This project aims to analyze possible models that can be utilized in the “Increasing bike-share efficiency” case study. The background of this case is for bike-share companies to optimize the allocation of bikes for each docking station in various areas. Factors that need to be considered for household include willingness, accessibility and the availability to bike ridership, and factors critical to businesses include supply, demand, rebalancing cost given limited number of bikes, docking stations, trucks and operators. This project would focus on estimating future demand and optimizing the allocation of bikes without digging into pricing mechanism that can affect household incentive to biking. To approach the analysis, the idea is to segment the regions into different clusters and take action based on the characteristics of the target groups.

Step 0: Working with data. The target for this model is to optimize the allocation of bikes in the available docking stations, so ideally maximize bike ridership rate with present resources available. Data required to perform the analyses include current bike usage data and community demographic information which can affect the number of bikes assigned to docking stations in various locations. To accurately classify stations into groups, sufficient amount of data would be needed to reflect the demographic information of the region. Second piece of data needed would be the current bike utilization information and the cost of bike rebalance, including various labor and travel cost to relocate bikes. Given the estimated demand for bike and the present supply in inventory, the goal is to build optimization model to minimize rebalance cost subject to target allocation of inventory.

Step 1: Classify docking stations into groups based on demographic and cyclical information for location analysis. The crucial factor for bike-share ridership would be the density of community since we want to allocate more bikes to higher demand stations at the right time to increase ridership rate. Having limited amount of resources, the number of docking stations and bikes assigned would be based on the features of the community in various cycles.

Given: Docking station id, population density per square mile, number of stations per square mile, average daily temperature, day-of-week

Use: Random forest. While support vector machine is capable to achieve classification, random forest is more suited for multiclass scenario with given data which has a mixture of numerical and categorical features.

To: Classify community into multiple location priorities so we can proceed to estimate the target inventory need for each station, with higher population density in suitable weather condition being given higher priority.

Step 2: Estimate the target amount of bike inventory for each station based on bike demand priority from previous step, bike ridership rate which can be derived from information such as bike pickup and drop-off time and trip records.

Given: Docking station id, location priority, number of bike rentals per day, pickup/drop-off time, bike inventory, station capacity

Use: ARIMA. ARIMA is preferred over simple linear regression in this case since time sequence can be a factor that affect bike ridership rate.

To: Predict target amount of bike inventory for each station on the next cycle date

Step 3: Optimize bike rebalancing route based on cost for rebalance and inventory information in nearby docking stations.

Given: Docking station id, target inventory for each station, distance to nearby docking station, cost per rebalance job, number of trucks for rebalancing, bike inventory for each station, station capacity

Use: Optimization model with an objective function to minimize rebalancing distance from fulfilling target inventory. Model constraints would be labor cost and number of trucks available

To: Sort docking stations by the efforts to rebalance inventory

Step 4: Analyze the length of cycle for re-optimization. When the inventory of docking stations become near full or empty, optimization model would need to be performed again to avoid system out-of-service. This step can be performed after sufficient data on bike utilization rate for each station has been collected.

Given: Docking station id, length of time the inventory is full/empty for each station, bike utilization rate, distance covered by rebalancing trucks per week, number of rebalance job for each station per week

Use: Exponential smoothing. While ARIMA can be used for forecast, it works better with stationary data points since it aims to describe correlations among data, whereas exponential smoothing can account for non-stationary data points and seasonality since it gives more weights to recent observations.

To: Estimate the length of re-optimization cycle for bike allocation