

Overview

This project seeks to create a business plan for next year by understanding the customers, stations, time and conditions that drives growth

The Problem

Explore how our customers are using their bikes:

- Identify the customer at different station location
- Data insights to grow station vertically or horizontally

The Solution

Enhance overall accuracy in predicting trip demand, expand station coverage, foster community cooperation, and optimize stock replenishment/rebalancing using data analysis and machine learning techniques.

Details

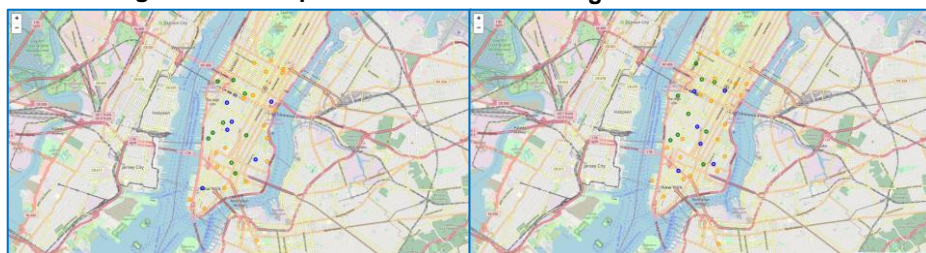
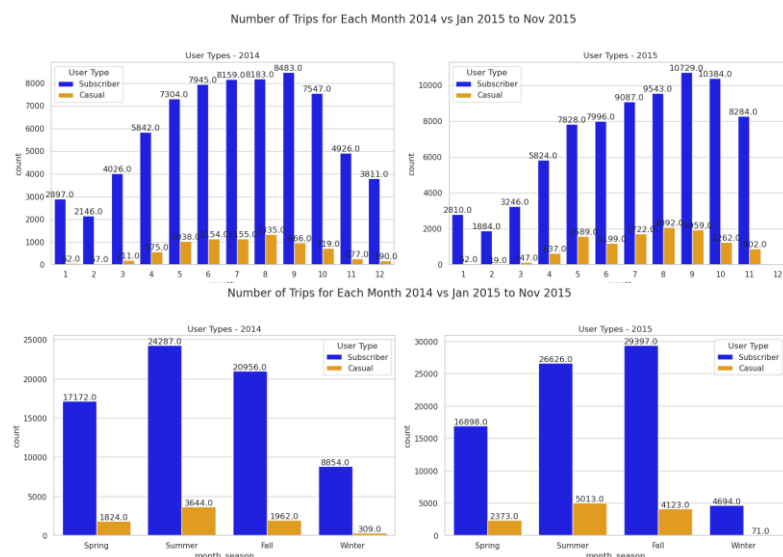
Keys to success

- Data must be acquired in a timely manner from all customers, trips and time, and date.
- Creating Business plans that are driven by customer insights.
- Identifying customer demand at different station locations.
- Include key data points that are summarized and aggregated
- Develop a dashboard that summarizes key insights

NYC Top Pareto Neighborhood Departure

NYC Top Pareto Neighborhood Destination

NYC Citibike Jan 2014 to Nov 2015



Results Summary

- Expand to the east with priority near the top trip neighborhood stations and explore the character of the top locations. The Pareto analysis reveals that the top 80% of trip neighborhoods constitute around 50% of all neighborhoods in 2015.
- The growth in 2015 exhibits seasonality, with a steady increase from summer, peaking in the fall in September, followed by a gradual decrease. This trend enables us to focus on medium to large maintenance, planning, or asset purchases during the slower seasons.

Results Summary (continued)

- Allocate resources for repair and maintenance on-site, with a particular focus on quality checks in top trip areas during off-peak hours.
- Establish initiatives for safe cycling, including safety training and education, distribution of safety equipment, and related endeavors.
- Foster community partnerships and programs, such as bike ambassadors, local businesses, youth and family engagement, grants for low-income individuals, etc.
- Hire both seasonal and permanent staff members who can actively engage with communities and contribute to the organization's objectives.
- Implement marketing and promotional awareness campaigns, such as #Bike4Youth and bike health events, as well as prominently displaying the NYCHA residents rate \$5 per month at stations.

The feature selection and engineering:

- The dataset includes string features for both neighborhood start and neighborhood end, encompassing 70 neighborhoods each. However, incorporating these features adds a substantial number of columns to the dataset, thereby increasing model complexity and the risk of overfitting. As a result, establishing priorities which hopefully are the most obvious ones were generated during the first round of modeling. Feature retention is essential for both improving processing time and mitigating overfitting. Consequently, we have made the decision to exclude certain location-related details such as borough, station name, and zip code. These exclusions are based on their similarity in purpose and intent to other location points within the dataset.
- Our augmentation encompasses the inclusion of several new columns: hours classification, am rush, day time, pm rush, and night time. Additionally, we've introduced age group, month season, day and service overtime columns.
- Other avenues of feature engineering remain unexplored. For instance, we could contemplate the creation of three novel columns, categorizing trip distances as short, medium, or far. Also, we can distinct the bike id into e-bike and reg-bike columns. Moreover, leveraging Pareto data, we can stratify trips into quartiles: the top 25% as high trips, the middle 50% as medium trips, the subsequent 80% as low trips, and any beyond the 80% threshold as developmental trips. This approach unlocks further potential for feature refinement.
- We can also consider creating location characteristics based on proximity to subway stations, residential areas, commercial zones, recreational spots, places of interest for social media (Instagrammable spots), and tourist attractions.

Reflections/ Next Steps

- The NYC citibike has set an internal goal to increase the number of trips for the predictive model of at least 90% accuracy.
- We have prediction system to improve by demand accuracy and stock accuracy for business planning.
- The goal is to share prediction modeling and data.
- Areas for large, medium and small demand and investment are now identified.
- Future results should continue the distance of each station affecting demand to choose new location.