

## **Executive Summary**

Citi Bike is New York City's bike share system, and the largest and most successful in the nation. Since its launch in 2013, Citi Bike has expanded to 12,000 bikes at more than 750 stations and has become a fun, affordable, and integral part of New York's public transportation network, with more than 70 million rides taken to date. Citi Bike consists of a fleet of specially designed, durable bikes that are locked into a network of stations in more than sixty neighborhoods across three boroughs. Customers can download the Citi Bike app to unlock bikes, find real-time station information, and see individualized statistics including miles traveled and calories burned.

This project will be done assuming that I am a Lead Data Analyst, responsible for overseeing the New York Citi Bike Program. The goal of this project is to provide thoughtful analysis for dictating programmatic changes and build visualizations to be included in press releases. The data found in the Citi Bike Trip History Logs will be aggregated to build a set of data reports for city officials looking to publicize and improve the program.

## **Business Case and Objectives:**

At a time when our subway system is in crisis, we need to invest in transportation alternatives that help people get around reliably, efficiently, and affordably. Doing so will help reduce our carbon footprint and make our city more livable for everyone. Lyft is planning to invest \$100 million to improve and expand Citi Bike. This is true public-private partnership between Department of Transportation, Citi Bike and Lyft.

This project will create visualizations that answer following questions and more:

- How many trips have been recorded total during the chosen period?
- By what percentage has total ridership grown?
- How has the proportion of short-term customers and annual subscribers changed?
- What are the peak hours in which bikes are used during summer months?
- What are the peak hours in which bikes are used during winter months?
- Today, what are the top 10 stations in the city for starting a journey? (Based on data, why?)
- Today, what are the top 10 stations in the city for ending a journey? (Based on data, why?)
- Today, what are the bottom 10 stations in the city for starting a journey? (Based on data, why?)
- Today, what are the bottom 10 stations in the city for ending a journey? (Based on data, why?)
- Today, what is the gender breakdown of active participants (Male v. Female)?
- How effective has gender outreach been in increasing female ridership over the timespan?
- How does the average trip duration change by age?
- What is the average distance in miles that a bike is ridden?
- Which bikes (by ID) are most likely due for repair or inspection in the timespan?
- How variable is the utilization by bike ID?

## **Stakeholders:**

The audience will be city officials, public administrators, and heads of New York City departments. Hence the data and analysis need to be presented in a way that is focused, concise, easy-to-understand, and visually compelling.

## **Dataset:**

Since 2013, the Citi Bike Program has implemented a robust infrastructure for collecting data on the program's utilization. Through the team's efforts, each month bike data is collected, organized, and made public on the [Citi Bike Data](#) webpage. The data is in CSV format and for this project 5 CSV files will be used for the years 2015 – 2019.

This data has been processed to remove trips that are taken by staff as they service and inspect the system, and any trips that were below 60 seconds in length (potentially false starts or users trying to re-dock a bike to ensure it's secure).

## **How?**

Different kinds of graphs and charts will be used along with Tableau Story Points to better understand Citi Bike Ridership by creating an operating report for 2015-2019 for New York City.

A dynamic map will be created that shows how each station's popularity changes over time (by month and year).

This project will also attempt to find any unexpected phenomena in the data and provide a visualization or analysis to document their presence.

## **Challenges:**

It will be crucial to ensure that the data is consistent and clean throughout the analysis. I will particularly keep a close eye for obvious outliers or false data as there could be people who sign up for the program and not provide honest answers. This project is attempting to answer a lot of questions and I foresee a possibility of cramming in everything hence It will be important to achieve the right balance between answering all the questions and using medium that is most effective.

**Persona: Bil de Blasio**



“New York City is one of the world’s great biking cities – and it’s about to get even better,”.

- **Bil de Blasio**

**Role: Mayor of New York City**

**Goals:**

- Delivering affordable and reliable transportation options that improve quality of life
- Promote, publicize and improve the Citi Bike program
- Work with communities across the city to make this expansion a success
- Establish geographic boundaries of the Citi Bike program expansion
- Continue improving bike infrastructure

**Challenges:**

He will most likely want to use the analysis to promote and publicize the program hence visualizations will need to be colorful enough to be included in press releases.