ReadMe

**Project Overview**

In this project, I have created a Tableau storyboard consisting of two dashboards, including two visualizations each and a map to discover and analyze underlying trends of the New York City CitiBike program data.

**About the data**

New York City Bike System is the most extensive bike-sharing program in the United States. Since May 2013, the Citi Bike Program has implemented a robust infrastructure for collecting data on the program's utilization. The data consists of CitiBike monthly information from June 2023 to January 2024.

**Data Engineering**

I used Jupyter Notebook to create a single dataframe named combined df containing the aggregate eight CSV data files, one file each month. I had no duplicates but missing values that I dropped off from the start station name, end station name, and end latitude and longitude columns. Further, I transformed the combined file by adding four columns: date, hour, trip duration, and distance and dropping off records with a trip duration of less than one minute and a half. Then, I used this new dataframe to create Seasonality Trends Analysis and Station Trends Analysis dashboards, a New York City basic map, and a Tableau story in Tableau Public.

Visit the link to the story board on Tableau Public: <https://public.tableau.com/app/profile/ivette.reese/viz/SMUFeb2024NewYorkCityBikeProgram/NYCCitiBikeProgram?publish=yes>

**Visualizations**

A screenshot of a computer

Description automatically generated

The visualization Seasonality Trends Analysis shows an hourly breakdown by weekday and weekend of the number of riders for each season (except for Spring; the data set does not include Spring months) and the average trip length for selected weeks by user type. **Both users' trips have increased and decreased over time. However, two prominent total trips decrease for members happened in September and November.**Weekdays ' peak riding hours are 7 to 8 am and 5 to 6 pm. On the weekends, the peak time is 12 pm to 4 pm. The non-peak hours are 1 to 5 am. Fall may be the peak riding season because the weather is less hot than in the Summer. The peak times throughout the week are relatively steady. People ride bikes more often on weekdays because of the commute to work. Very few trips in the early morning, then settling increase by 8 in the morning, then we go down on the number of trips from 9 to 11 am, then a steady increase throughout the day (except at 2 pm) until the spike from 5 to 6 pm. The spikes seen align with commuters leaving or returning for work or college. Bimodal picking up twice at 8 am and then 6 pm around the popular commute times.

A screenshot of a graph

Description automatically generated

The Stations Trends Analysis visual shows users' trip duration average is 7 to 9 minutes on weekdays and weekends. Most of the most popular bike stations are in the middle of the city, explaining the short average bike ride trip duration. Casual users average 7 to 9 minutes on weekdays and weekends except for January 2024, with a trip duration average of 10 minutes or longer. The members take longer bike rides weekly than casual users, as expected since Members comprise 75% of the total subscribers. The bottom chart shows the top 10 stations with the count of bike rides. The number one biking riding station is Hoboken Terminal, labeled 1 on the graph, with almost 19k bike rides, while 12St & Sinatra Dr is the least popular, with 7.5k rides. Users can look at the top 3, 4, or 10 stations when selected.

New York City Basic Map

**A map of a city

Description automatically generated**

The map shows markers for all bike stations used from June 2023 and January 2024. The higher the popularity of the station, the larger the marker size. The zip codes are colored green, representing the US Census 2018 Median Household Income. The higher the income, the darker the green shade. Users can change marker data based on month and year using the Date dropdown on the legend. There is also an option to filter the data by User Type.

Most of the stations are on the east side. There is a significant trend between bike stations and the station's rank by popularity. Most of the most popular bike stations are in the middle of the city, represented by larger marker sizes, while less popular stations with smaller marker sizes are away from the city. Most of the stations' locations are in higher-income zip code neighborhoods, so the relationship between the number of stations and zip code household income is that the higher the median household income, the higher the number of stations.

75% of all users are subscribed members, over 303k compared to over 99k casual users—no pattern indicating significance between type of users and station popularity. The distribution of members and casual users is evenly distributed throughout.

**Data Sources**

<https://citibikenyc.com/system-data>

<https://www.census.gov/programs-surveys/acs/data.html>

<https://www.kaggle.com/datasets/akkithetechie/new-york-city-bike-share-dataset>