

## Introduction

The goal of this project is to compare bikeshare data from two cities, San Francisco and New York City, to understand how user behavior differs between the two locations. We'll also be merging this information with weather data to better understand how weather factors influence ridership in each of the cities.

As a first step, we created two dataframes, one for each city. Each dataframe includes trip and weather information. We leveraged NYC bikeshare trip data and merged it with weather data collected from the government using date as a key. Since the weather data was collected through various weather stations, we decided to use the Central Park station data in our analysis. However, we used trip and weather data provided by the bikeshare organization in San Francisco, instead of merging in government weather data since the weather was readily available. To do this, we had to merge the trip data, which included station ID's, to the station data, which included station zip codes. We used Google to map these zip codes to actual city names (San Francisco, Mountain View, etc.) and then mapped those city names to the weather data.

After creating the dataframes for each city, we analyzed and mapped fields across the dataframes, standardizing the header names and contents for overlapping fields. This was particularly difficult for the weather data, since we had to understand weather metrics to accurately map the data and ensure we were doing an apples-to-apples comparison. The reason for standardizing the header names and contents is to enable us to leverage the same code when we analyze the data simply by changing the csv files loaded.

Here is a summary of the the dataframe contents:

| Field Name         | NYC Source    | SF Source    |
|--------------------|---------------|--------------|
| tripduration       | NYC Trip Data | SF Trip Data |
| starttime          | NYC Trip Data | SF Trip Data |
| stoptime           | NYC Trip Data | SF Trip Data |
| start station id   | NYC Trip Data | SF Trip Data |
| start station name | NYC Trip Data | SF Trip Data |
| end station id     | NYC Trip Data | SF Trip Data |
| end station name   | NYC Trip Data | SF Trip Data |
| bikeid             | NYC Trip Data | SF Trip Data |
| usertype           | NYC Trip Data | SF Trip Data |
| city               | Added by Team | Station Data |
| weather key        | date          | date + zip   |

|               |                             |  |
|---------------|-----------------------------|--|
| Precipitation | From Gov Weather File, PRCP | From weather file, Precipitation In    |
| max temp      | From Gov Weather File, TMAX | From weather file, Max TemperatureF    |
| min temp      | From Gov Weather File, TMIN | From weather File, Min TemperatureF    |
| avg wind      | From Gov Weather File, AWND | From weather File, Mean Wind Speed MPH |

## Final Report Outline

For the final report, we plan on having eight major sections, with each about one to two pages:

1. Introduction
2. Overall Ridership
3. Rider Types
4. Trip Times
5. Time of Day Analysis
6. Impact of Weather
7. Conclusion

## Questions & Column Names

The table below describes the questions we would like to answer along with the column names we plan on using for the analysis:

| Section                     | Questions  | Column Names  | Frequency |
|-----------------------------|--|---|-----------|
| <b>Introduction</b>         | N/A  | N/A   | N/A       |
| <b>Overall Ridership</b>    | What is the overall trend in ridership by city?                              | Trips   | Per month |
|                             | Within each city, what is the average number of trips per station per month? | Trips per start station, trips per end station, or trips for each start-end station combination | Per month |
| <b>Rider Types</b>          | What is the distribution of rider types?                                     | User / subscription type  | Per month |
| <b>Trip Times</b>           | What is the average trip time by start/stop station?                         | Trip duration   | Per month |
|                             | What is the median trip time?  | Trip duration   | Per month |
|                             | What is the range of trip times?   | Trip duration   | Per month |
|                             | What is the standard deviation of trip times?                                | Trip duration   | Per month |
| <b>Time of Day Analysis</b> | Are there more riders on the weekdays vs. weekends?                          | Start date / end date   | Per week  |

|                          |  |  |            |
|--------------------------|--|--|------------|
|                          | Does the composition of riders change on the weekends (i.e., more leisure riders)? | Start date, end date, and user / subscription type | Per week   |
|                          | What are the most popular and least popular time blocks (by hour)?                 | Start date / end date                              | Per day    |
| <b>Impact of Weather</b> | How do seasons affect ridership?   | Start date, end date                               | Per season |
|                          | How does temperature affect ridership?   | TMAX, TMIN   | Per day    |
|                          | How does precipitation affect ridership?   | Precipitation                                      | Per day    |
|                          | How do extreme weather events affect ridership?                                    | WT01-WT06, WT08-WT11, events                       | Per day    |
|                          | How does wind affect ridership?  | Average daily wind speed                           | Per day    |
| <b>Conclusion</b>        | N/A  | N/A  | N/A        |

### Initial Plots and Graphs



