

# Citibike Bike Share Analysis

## Introduction

As stated in the README file, the purpose of this project is to analyse the ride data of the citibike sharing company in New York/New Jersey and weather data for each day of the year 2016. This document serves as a high-level summary of the project detailing what areas are being investigated, why and what are the results.

It is best to read this document and inspect the PowerBI dashboard at the same time.

The source files for this project are:

- A .csv file containing the weather information for each day of 2016.
- Multiple .csv files where each ride of the Citibike bikes is recorded.

The areas of investigation are:

- Change in the state of the company
- The effect of time on user behavior
- The effect of location on user behavior
- The effect of weather on user behavior
- Estimation of wear and tear of the equipment

Below is a description of each area in greater detail.

## Areas of Analysis

### 1. Change in the state of the company

#### **Subject of Analysis:**

The state of the company is defined in this context as the state of the available bikes and the docking stations. This is how users interact with the business. Regarding bikes, the number of active bikes in each month is analyzed and regarding stations, the same is analyzed as well as their geographical location.

#### **Relevance:**

- Can be used to investigate any policies/intent of the company such as expansion to a new part of town.
- Can be used to accurately estimate the number of bikes and stations the company owns. This gives a measurable indicator of the size of the company which can be used for comparison with other companies/future revenue modeling or forecasting.
- Shows if/how the company responds to change.

#### **Stakeholders:**

- Company upper management
- Upper management of a competing company

**Note:** In this category it is assumed that the person carrying out the analysis is not a member of the company. The company surely has better datasets to answer the above questions and knows what its policies/intent are.

## Results:

The results are presented on the pages State of the Company and Stations Location. The number of active bikes and active stations does indeed change throughout the year with more bikes/stations added in the warmer months. In the two tables, there is a column 'Number of rides for 5<sup>th</sup> percentile' added. This is because some bikes/stations might actually have never been used in the given month and would therefore not show up in the dataset. This is due to the inherent limitation of the dataset, it describes rides taken, not the stations.

However, as the column shows, the 5% of the least used bikes are still used a couple of times, therefore the probability of a bike never being used in a given month is very low.

Interesting insights can be derived from the Stations Location page. As is found out in the data pre-processing step, riders can end their ride at any station, but they can start their rides at only a limited group of stations. When looking at the plotted map, it is obvious that all of the non-start stations are in the city of New York and that the location of these stations changes drastically throughout the year compared to stations in New Jersey whose location hardly changes.

This might mean that the company is running a pilot project in NY on what areas are the best for station placement. This is however just a speculation and more domain knowledge is needed to answer it. However, it can definitely be said that the majority of rides occurs in New Jersey.

## 2. The effect of Time on User behavior

### Subject of Analysis:

Here, it is investigated how the three metrics defining user behavior (number of rides, ride length, ride displacement) change with time. This is of course a very general question and is specified to make more sense. The following is investigated:

- The evolution of the number of rides taken by month.
- The average number of rides by day for the whole year.
- Average journey duration by day for the whole year.
- The number of rides for each hour for each day in the week (slice by month).

### Relevance:

- Knowing user behavior is crucial to any business. In this case, it might for example be beneficial for scheduling maintenance intervals.
- Knowing when users ride the bikes gives an insight into the actual userbase as tourists/employed residents will ride their bikes at different time.

### Stakeholders:

- Company upper management
- Company operations management

## Results:

The results are presented on the page User Behavior by Time in the Power BI report. The notable insights here are that during weekdays, there are two sharp peaks in the number of rides at 8:00-9:00 AM and 6:00-7:00 PM. During weekends the usage is lower with no sharp peaks but rather a gradual incline which peaks at around 1:00 PM. This trend is constant throughout the year.

Another insight that can be derived here are that the userbase are mainly working people even in the months when tourists visit New Jersey/New York.

### 3. The Effect of Location on User Behavior

#### **Subject of Analysis:**

In this section, the number of rides taken to/taken from each station is determined and visualized for each month. Viewers of this visual can therefore quickly understand which parts of the city are the busiest ones and how does this change over time.

#### **Relevance:**

- More efficient relocation of bikes around the city by the company.
- Better knowledge of where to build bike infrastructure.
- Better knowledge of where to deploy new bike stations.

#### **Stakeholders:**

- Company upper management
- Company operations management
- City representatives

#### **Results:**

The results are presented on the page User Behavior by Stations. The insights here are that almost no rides end in New York, the absolutely crushing majority of rides occurs in New Jersey (specifically in the Metroplaza area). This trend is constant throughout the year and the introduction of new stations doesn't have any effect on it.

### 4. The Effect of Weather on User Behavior

#### **Subject of analysis:**

In this section, the effect of the 3 important weather metrics (precipitation, snowfall, temperature) on the number of rides in a day is plotted.

#### **Relevance:**

- Knowing the influence of weather on the number of rides can be theoretically used to predict user behavior in the future.

#### **Stakeholders:**

- Company operations management

#### **Results:**

It is out of the scope of this project to conduct any rigorous analysis on the effect of weather on the user behavior, but viewers of the dashboard can select the timeframe of the graph so that all variables except for one are constant and investigate how does the change in the independent variable influence the general trend of bike usage.

## 5. Estimation of the wear and tear of the equipment

### **Subject of analysis:**

For each station the number of rides is determined and for each bike the number of rides as well as the total displacement throughout the year is determined.

### **Relevance:**

- Knowing this can be used to preemptively service the bikes and stations.

### **Stakeholders:**

- Company operations management

### **Results:**

The results can be viewed on the Maintenance Data page in the Power BI report.

**Note:** This is not a very useful analysis as the total displacement measured for each bike is not an entirely accurate indication of the real mileage, the number of bikes started/ended at a station is not a very accurate representation of the wear on the station and moreover, the company surely has a separate database for bikes and stations which is better suited to answer these questions.