

# Ford GoBike System Data

## Introduction and Background

This document explores a dataset containing data from Ford GoBike, a Bay Area's bike sharing system.

You can check the definition of the variables included here:

<https://www.lyft.com/bikes/bay-wheels/system-data>

Ford GoBike has data from mid 2017 to mid 2020. For the purpose of this analysis, I will focus on data for 2019, which includes data for 2506983 trips.

Some variables included in the 2019 dataset, `bike_share_for_all_trip` and `rental_access_method`, have not been defined by Ford GoBike.

## Gathering, Assessing and Cleaning the Data

The dataset provided by Ford GoBike is available in multiple data files. In this section the data is downloaded programmatically and saved into a CSV file.

The main focus of this analysis will be on the following variables:

- `duration_sec`
- `start_time`
- `user_type`

I will also use latitude and longitude to create an interactive heat map at the end.

These variables have no null values.

Issues to solve in the cleaning process:

1. Convert `duration_sec` to integers, `start_time` to dates, and `user_type` to category.
2. Create variables for month, weekday and hour.
3. Rename `start_station_latitude` and `start_station_longitude` to 'lat' and 'lon'.
4. Remove irrelevant variables

## Exploratory Analysis and Visualizations

The cleaned data was stored in a unique data set and was ready for analysis.  
This data set includes 2084 unique rows and 12 columns.

The following questions will work as a guideline to my exploration:

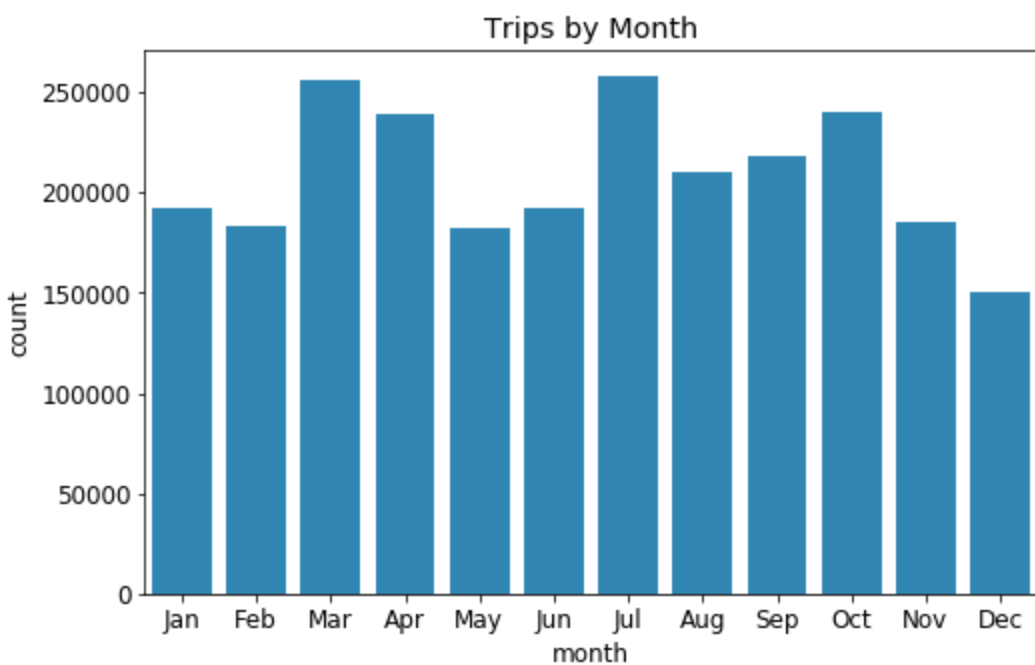
1. When are most trips taken in terms of time of day, day of the week, or month of the year?
2. How long does the average trip take?
3. Does the duration of a trip per month, weekday or hour depend on if a user is a subscriber or customer?

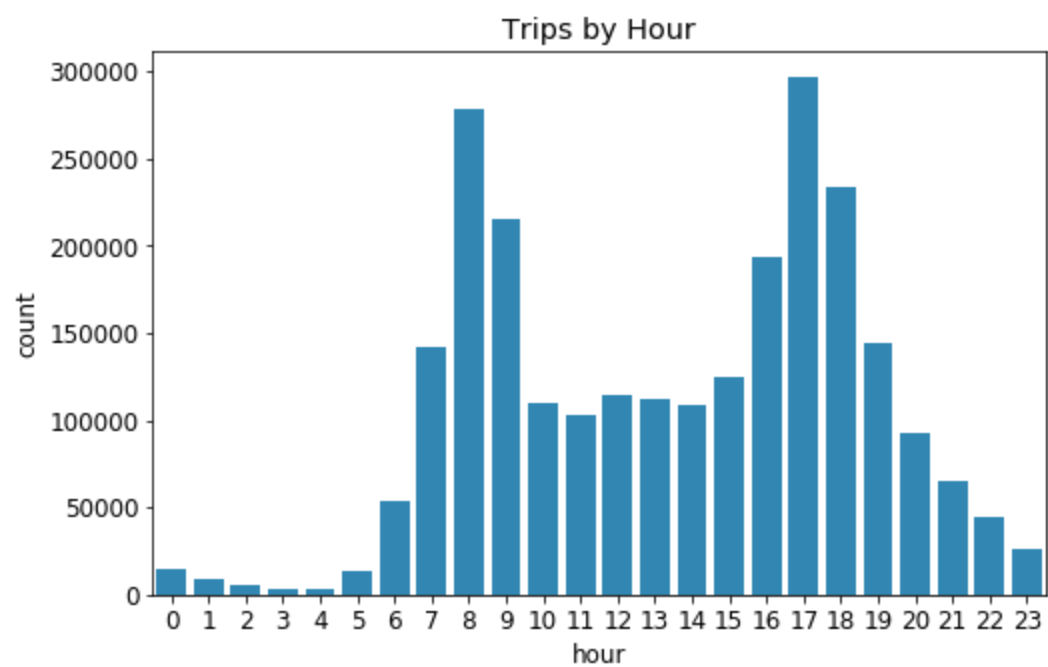
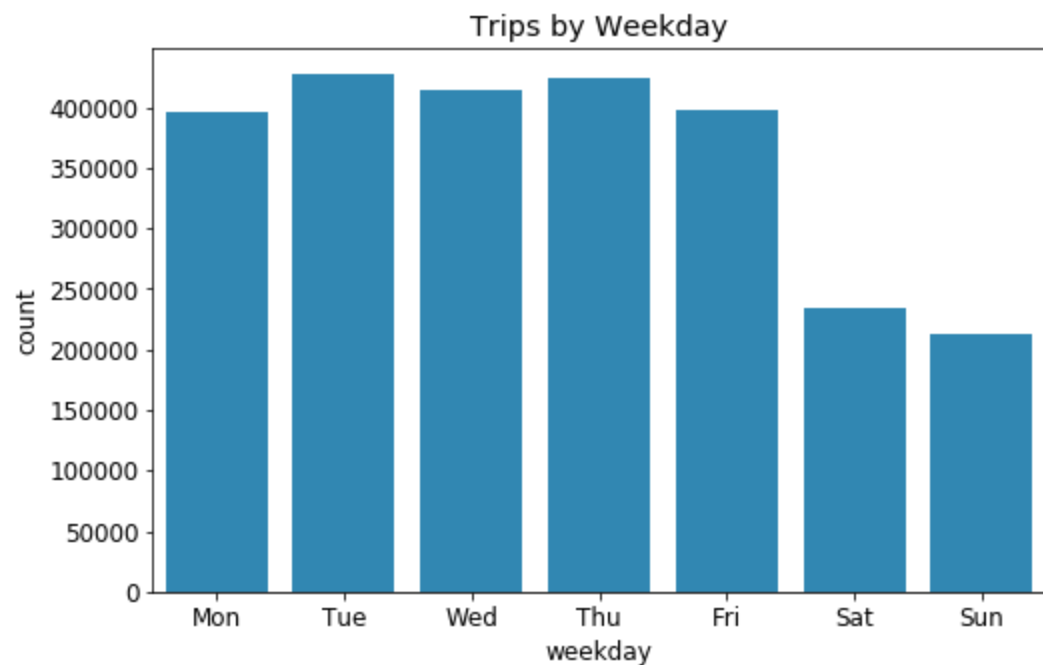
***1. When are most trips taken in terms of time of day, day of the week, or month of the year?***

Time of day: 17:00

Day of the week: Tuesday

Month of the year: July

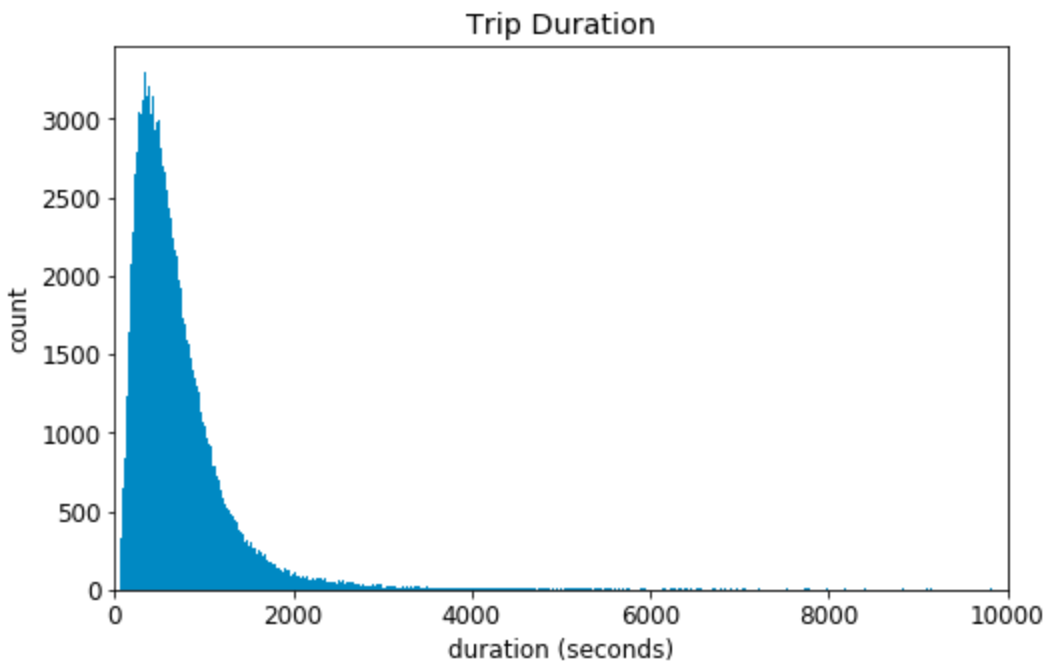


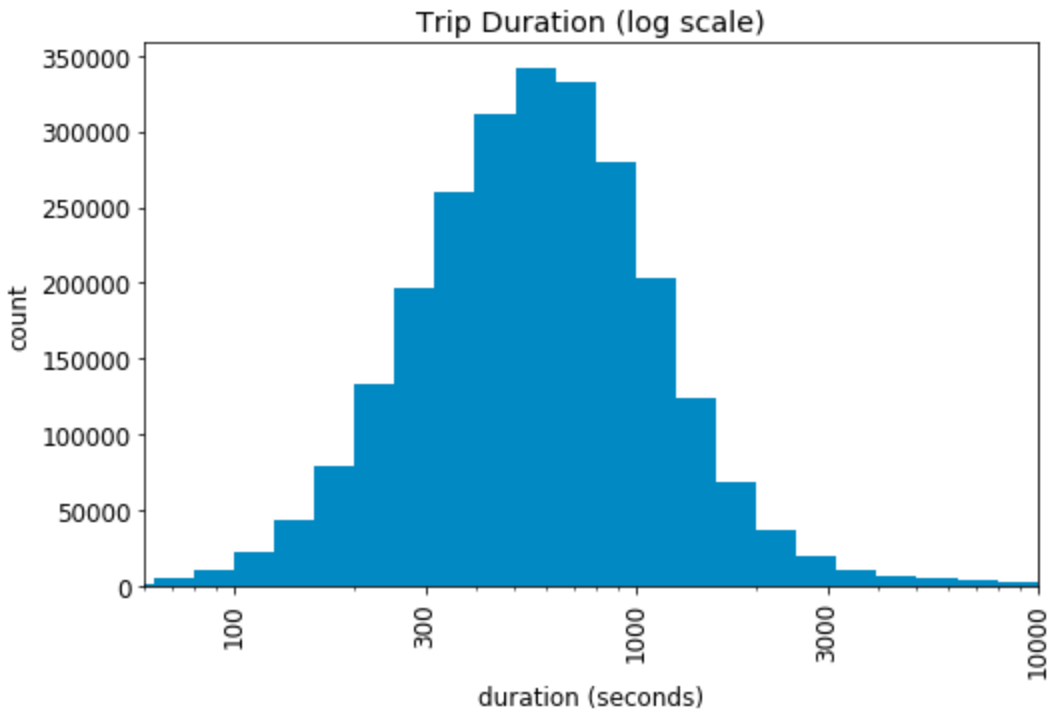


## 2. How long does the average trip take?

The average trip takes around 500 seconds. This is after wrangling the data and logarithmic transformation.

The original dataset shows an average of 800 seconds. However, this includes many, very high outliers.

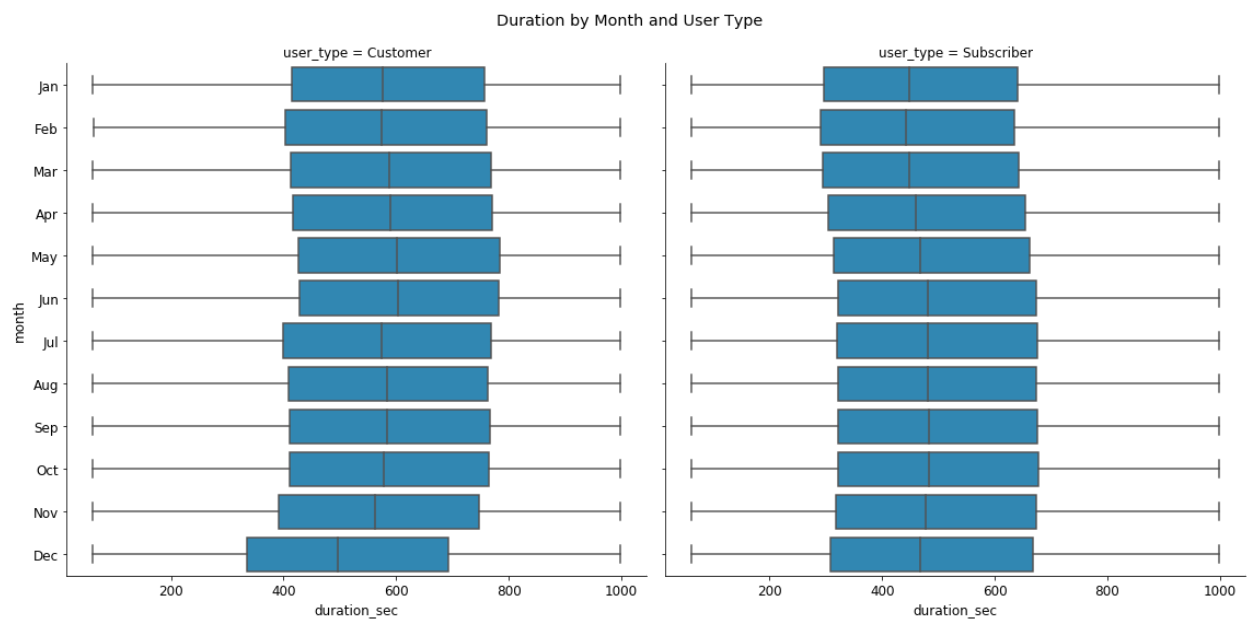
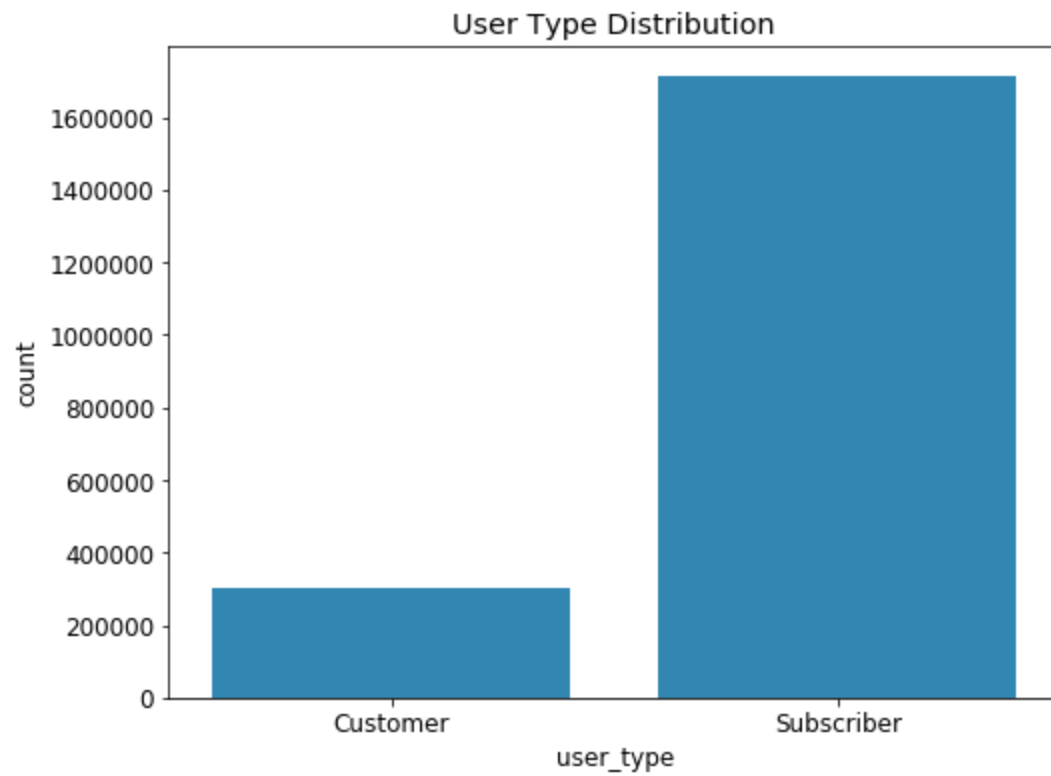


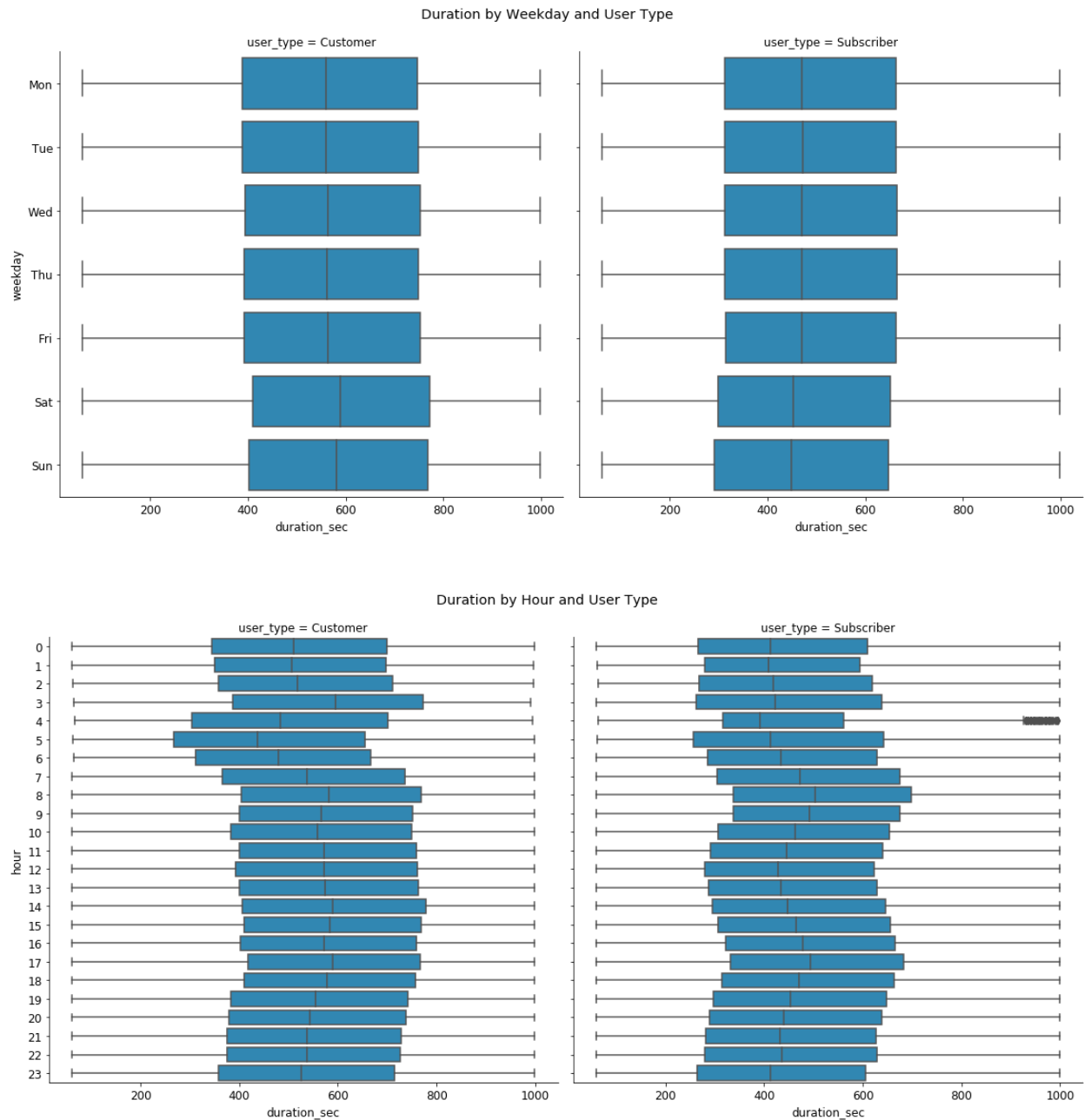


**3. Does the duration of a trip per month, weekday or hour depend on if a user is a subscriber or customer?**

Yes. The average duration for the subscriber is lower in all cases.

However, less than 20% of the users are customers.





## Explanatory Summary

- Question 1: **When are most trips taken in terms of time of day, day of the week, or month of the year?**

In terms of:

- Time of day: 17:00
- Day of the week: Tuesday

- Month of the year: July

**Other interesting insights:**

- Month:

The least popular month is December.

July and March represent more than 20% of the trips in 2019.

- Day of the week:

Weekend days have around half less trips than regular week days.

- Time of day:

3:00 is the least popular time.

There are two clear peaks between 07:00 and 10:00 and 16:00 and 19:00. This represent 54.2% of the trips in a day on average.

2. Question 2: **How long does the average trip take?**

The average trips takes around 500 seconds. This is after wrangling the data.

The original dataset shows an average of 800 seconds. However, this includes many, very high outliers.

**Other interesting insights:**

Most trips fall into the interval of 60 to 10000 seconds.

The largest amount of trips fall into the interval of 300 to 700 seconds.

3. Question 3: **Does the duration of a trip per month, weekday or hour depend on if a user is a subscriber or customer?**

Yes. The average duration for the subscriber is lower in all cases.

**Other interesting insights:**

Less than 20% of the users are customers.

- **Additional Insights**



The heat map shows that the trips' starting locations are in San Francisco, Oakland and San Jose.