Investigation Overview

Bay Wheels (previously known as Ford GoBike) is a regional public bike sharing system in the San Francisco Bay Area, California. Bay Wheels is the first regional and large-scale bicycle sharing system deployed in California and on the West Coast of the United States with nearly 500,000 rides since the launch in 2017 and had about 10,000 annual subscribers as of January 2018. This exploration and visualization study primarily focuses on finding the bike usage pattern and customer habit characteristics from the 12-month ride trips data in 2018.

Dataset Overview

The dataset used for this exploratory analysis consists of monthly individual trip data from January 2018 to December 2018 in CSV format covering the greater San Francisco Bay area, raw data is available here. Visualizations below are created from wrangled and cleaned data to facilitate exploration analysis and help discover usage pattern and rider characteristics.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1688410 entries, 0 to 1688409
Data columns (total 26 columns):
 #   Column                  Non-Null Count    Dtype
---  ------                  --------------    -----
 0   duration_sec            1688410 non-null  int64
 1   start_time              1688410 non-null  datetime64[ns]
 2   end_time                1688410 non-null  datetime64[ns]
 3   start_station_id        1688410 non-null  object
 4   start_station_name      1688410 non-null  object
 5   start_station_latitude  1688410 non-null  float64
 6   start_station_longitude 1688410 non-null  float64
 7   end_station_id          1688410 non-null  object
 8   end_station_name        1688410 non-null  object
 9   end_station_latitude    1688410 non-null  float64
 10  end_station_longitude   1688410 non-null  float64
 11  bike_id                 1688410 non-null  object
 12  user_type               1688410 non-null  category
 13  member_birth_year       1688410 non-null  int64
 14  member_gender           1688410 non-null  category
 15  bike_share_for_all_trip 1688410 non-null  bool
 16  start_metro_area        1688410 non-null  object
 17  end_metro_area          1688410 non-null  object
 18  hour                    1688410 non-null  int64
 19  member_age              1688410 non-null  int64
 20  age_group               1688410 non-null  object
 21  duration_minute         1688410 non-null  float64
 22  start_date              1688410 non-null  object
 23  start_hourofday         1688410 non-null  int64
 24  start_dayofweek         1688410 non-null  category
 25  start_month             1688410 non-null  category
dtypes: bool(1), category(4), datetime64[ns](2), float64(5), int64(5), object(9)
memory usage: 278.6+ MB
```

User Types by Age Group

user's subscription ratio increases as the age increase. Younger user tend to use the service but do not subscribe.
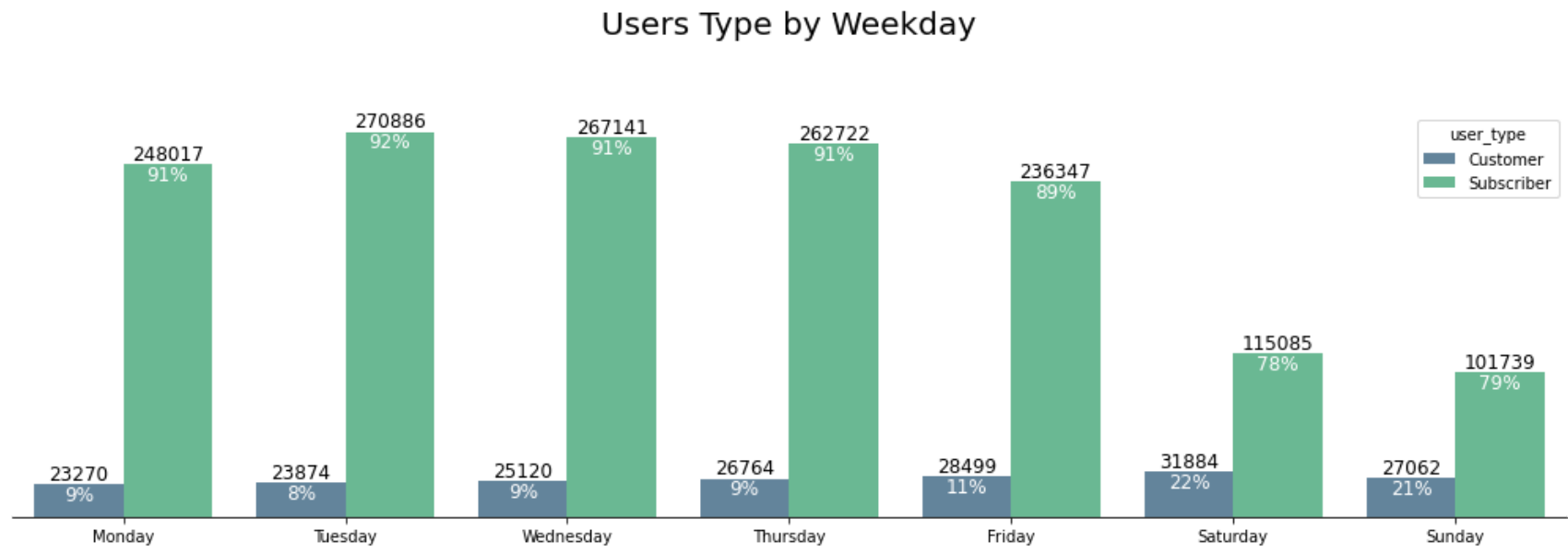
## Hourly Usage during Weekdays for Customers and Subscribers

Different usage patterns/habits between the two types of riders are clearly seen. Subscribers used the system heavily on work days i.e. Monday through Friday, whereas customers ride a lot on weekends, especially in the afternoon. Many trips concentrated around 8-9am and 17-18pm on work days for subscribers when used for work commute, yet casual customers tended to use more in the late afternoon around 17pm Monday to Friday. Other than the different usage goals, it may also have to do with the traffic situation around these rush hours in the area.

```
<Figure size 432x288 with 0 Axes>
```
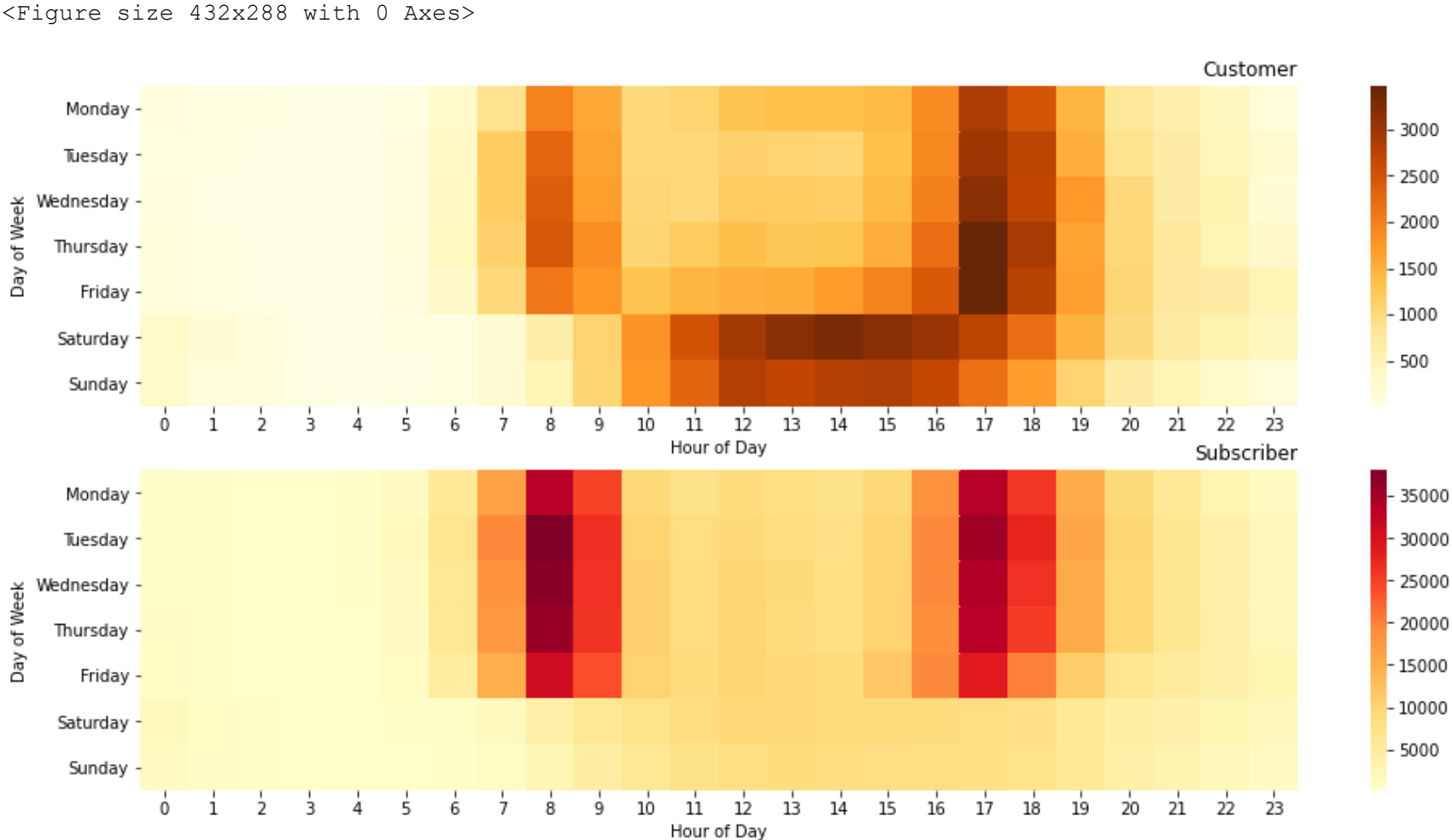

Average Weekday Trip Duration for Customers and Subscribers

Subscribers ride much shorter/quicker trips compared to customers on each day of the week. Both user types had an obvious increase of trip duration on Saturdays and Sundays (or weekends), especially casual customers. Subscriber usage was more efficient than customers overall and maintained a pretty consistent average duration Monday through Friday.

The East Bay age structure is broader than the one of San Francisco and San José has the youngest average group of users. The next plots will focus on time components of our data. What about trips per day?
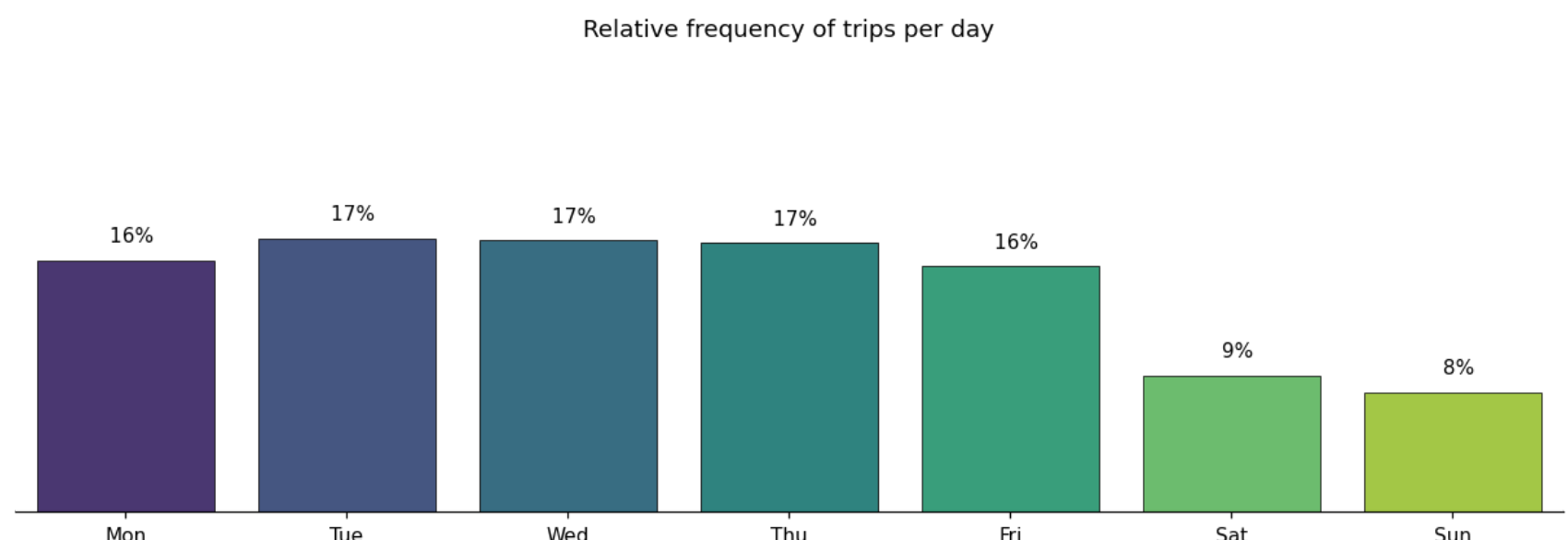
It looks like the users use the bikes more frequently during the week than during the weekend. And when do they start their trips?
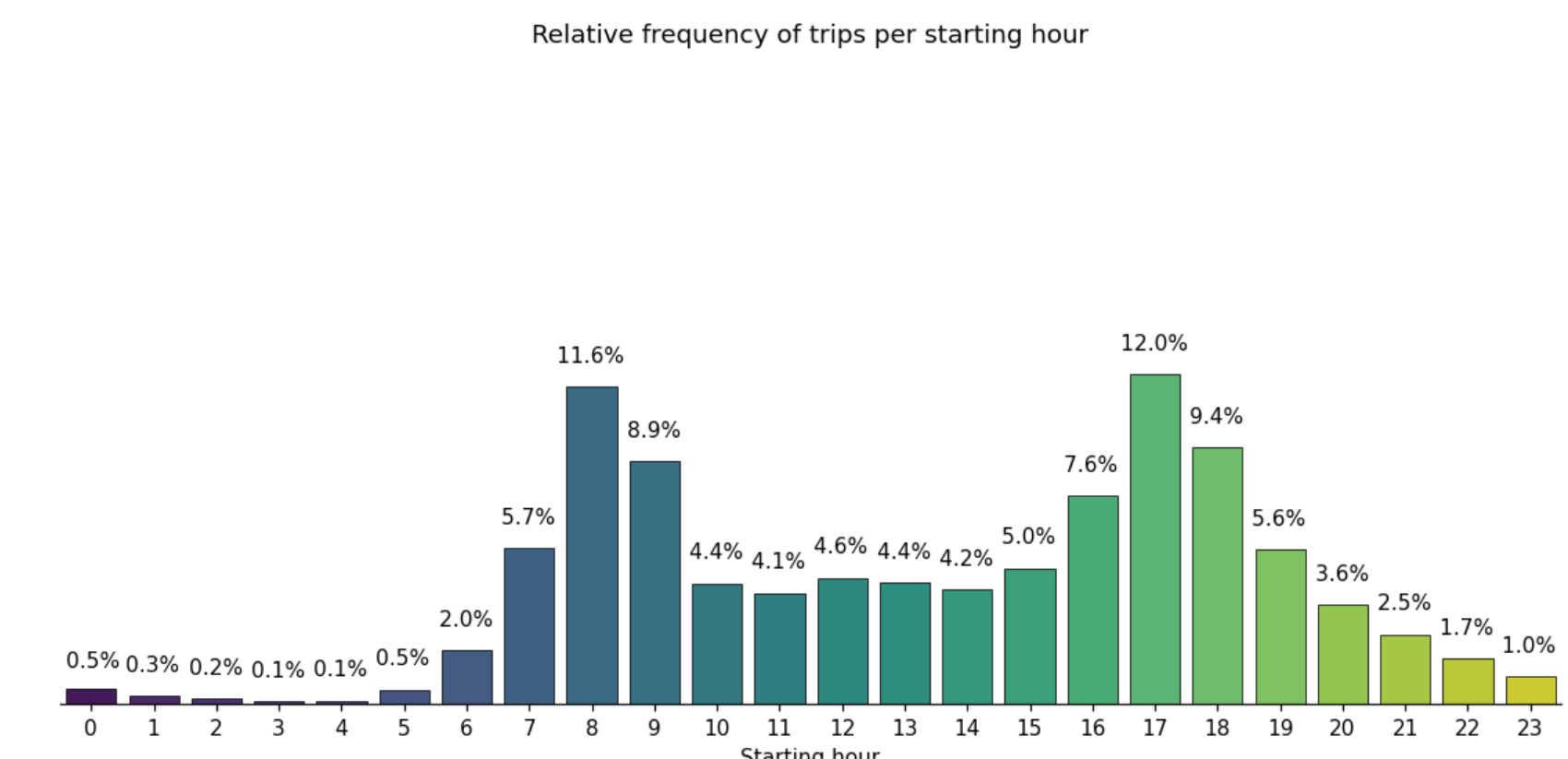
The most frequent starting hours are at 8 and at 17. Maybe people use it before and after work, which would make sense, because we have a lot of subscribers in working age in our dataset. You only subscribe to something, if you want to use it regulary. The integration into the working/study life would make sense here!

Summary

There was a lot more subscribers using the bike sharing system than casual customers overall, both of which ride the most during the summer season and the least during the winter months. Obviously different usage patterns and riding habits are observed between the two types of riders. Subscribers used the system heavily on work days concentrated around 8-9am and 17-18pm for work commute, whereas customers ride a lot over weekends and in the afternoon for leisure/touring purposes. Subscribers tended to have much shorter/quicker trips compared to customers which makes subscriber usage more efficient.