Citi Bike Data Analysis

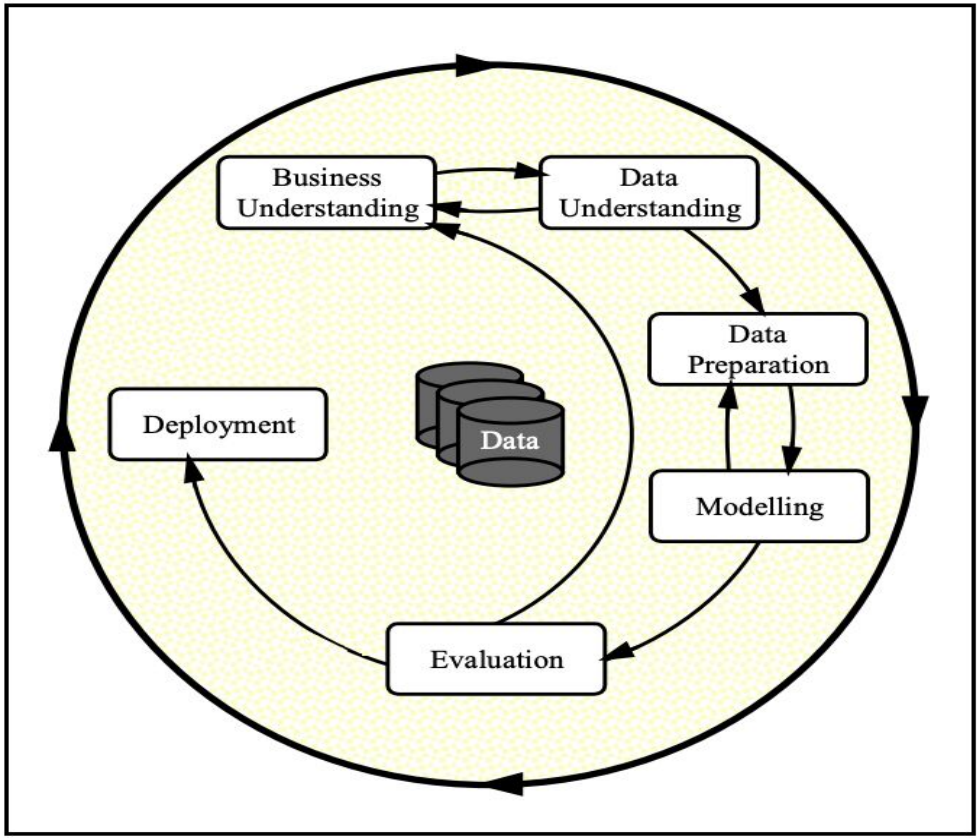
Sai Vamsy Palakollu

About Citi Bike Dataset

In this dataset, we will explore the new york city bike rider's trips. The information about the rider's trip is located here. For the capstone project, we will be taking information from January 2022 to December 2022.

Citi Bike is a bike-sharing service in New York City that allows users to rent and ride bikes from a network of self-service docking stations located throughout the city. To use Citi Bike, users must first sign up for a membership, which can be purchased online or at a Citi Bike kiosk. Once they have a membership, users can check out a bike from any Citi Bike docking station using their membership card or the Citi Bike app on their smartphone. After they have finished their ride, users can return the bike to any Citi Bike docking station. Citi Bike offers a variety of membership options, including daily, weekly, and annual memberships, as well as options for tourists and students. Citi Bike also offers electric bikes, which can be rented in the same way as standard bikes.

Framework



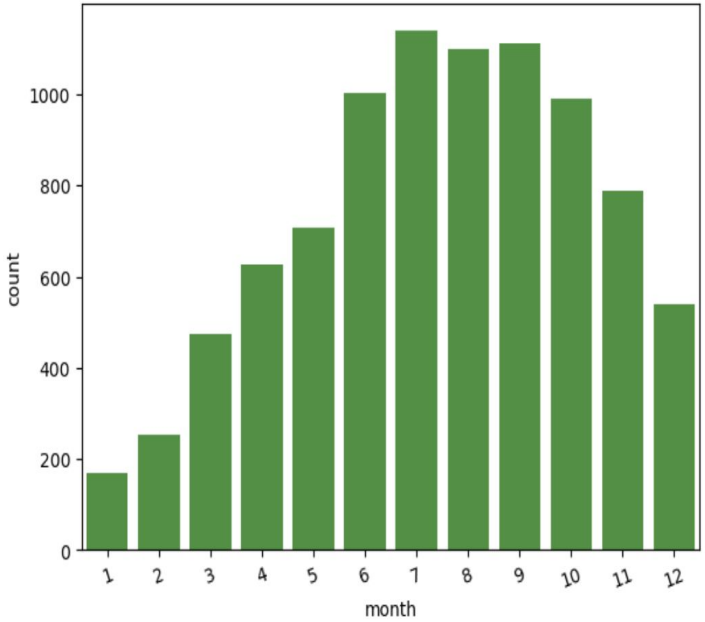
- In reference to our capstone project, we will utilize CRISP-DM, which is a widely used industry standard process for data projects.
- The framework offered by this process assists in effectively addressing a data problem

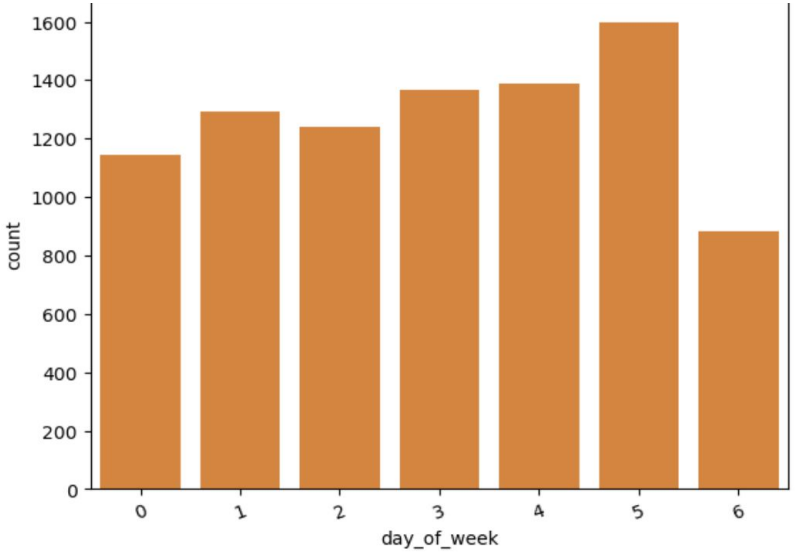
Dataset Features

The following features are part of the dataset. Below is a brief explanation of the features.

1. RideId - A unique identifier assigned to each bike ride in the Citi Bike rideshare dataset.
2. Rideable Type - The type of bike used for the ride, such as a classic bike, electric bike, or dockless bike.
3. Started At: The date and time when the ride started, recorded in UTC (Coordinated Universal Time).
4. Ended At: The date and time when the ride ended, recorded in UTC.
5. Start_Lat: The latitude coordinate of the starting location for the ride.
6. Start_Lng: The longitude coordinate of the starting location for the ride.
7. End_Lat: The latitude coordinate of the ending location for the ride.
8. End_Lng: The longitude coordinate of the ending location for the ride.
9. Member_Casual: Indicates whether the rider was a member (i.e., had a Citi Bike subscription) or a casual user (i.e., purchased a single ride or day pass).

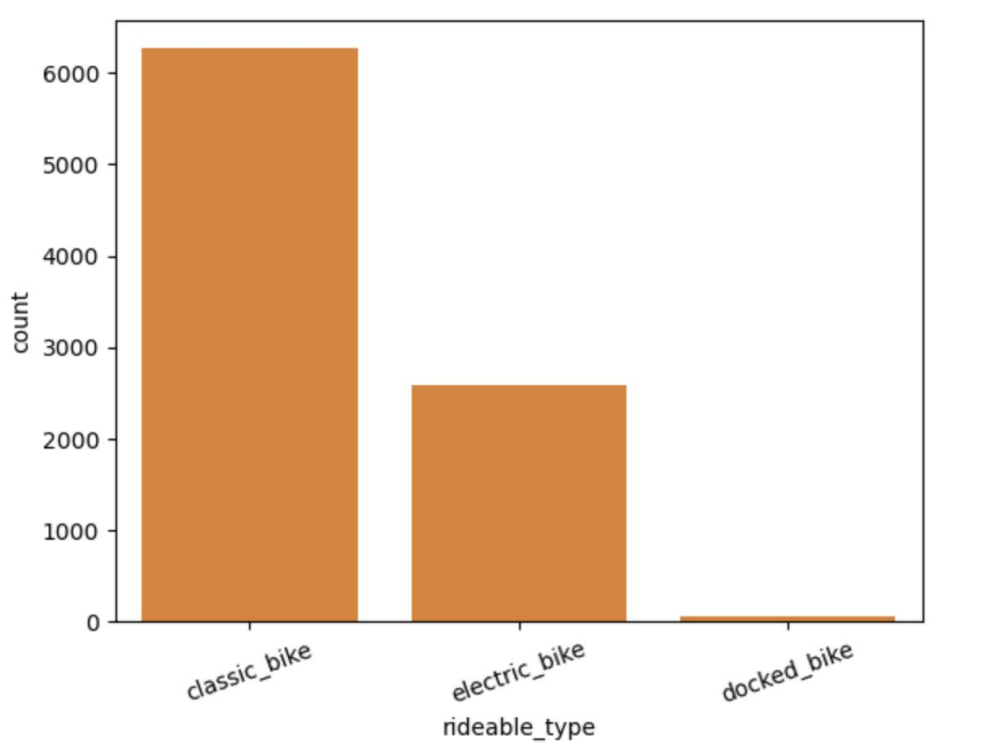
Findings - Trip Distribution





Looking in combined with the trip distribution over day of week plot, the majority of rides happened on tuesday - friday and primary usage is probably for commuting. In the 11 months in 2022, August had the most ride trips compared to the others, but overall it was the most popular during summer time (June-October), probably due to the weather in the area.

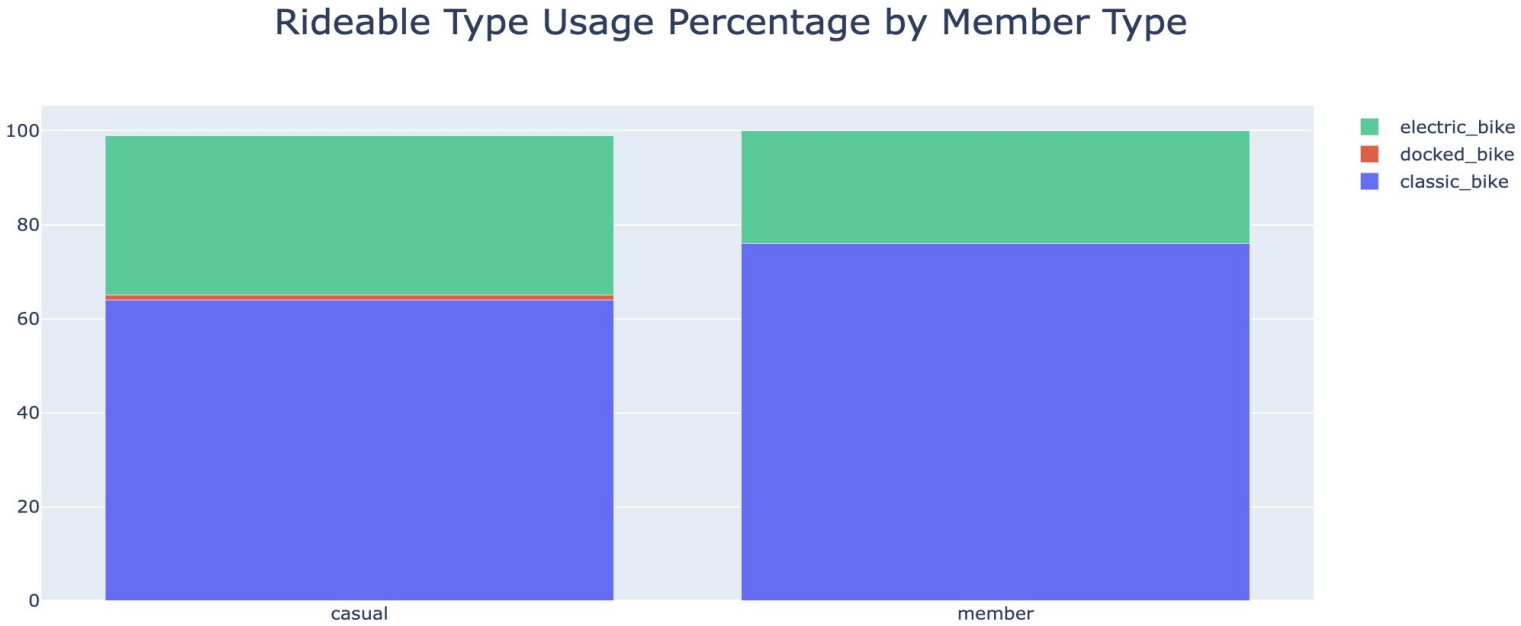
Findings - Trip Distribution Over Bike Type



The plot will display the number of occurrences for each category in the "rideable_type" column in a bar graph format, with the categories on the x-axis and the count of occurrences on the y-axis. Classic bike stands out among the bike types.

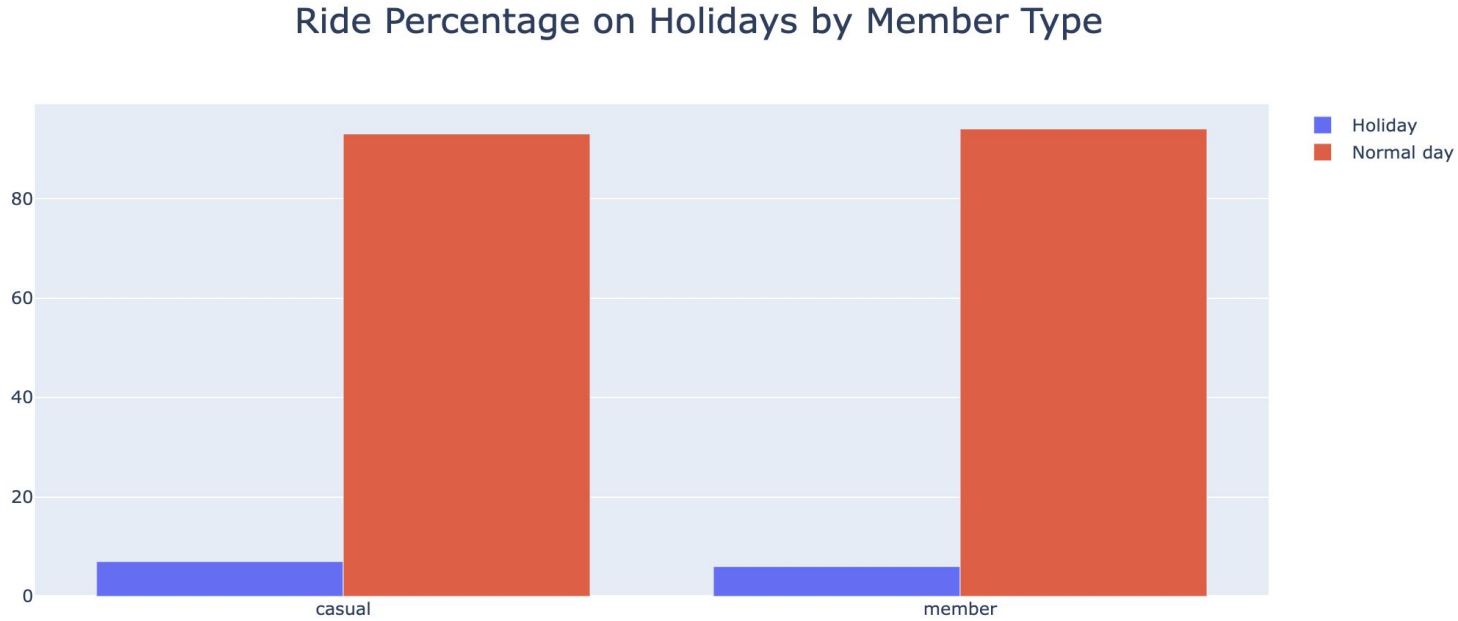
Findings - Rideable

Type Usage



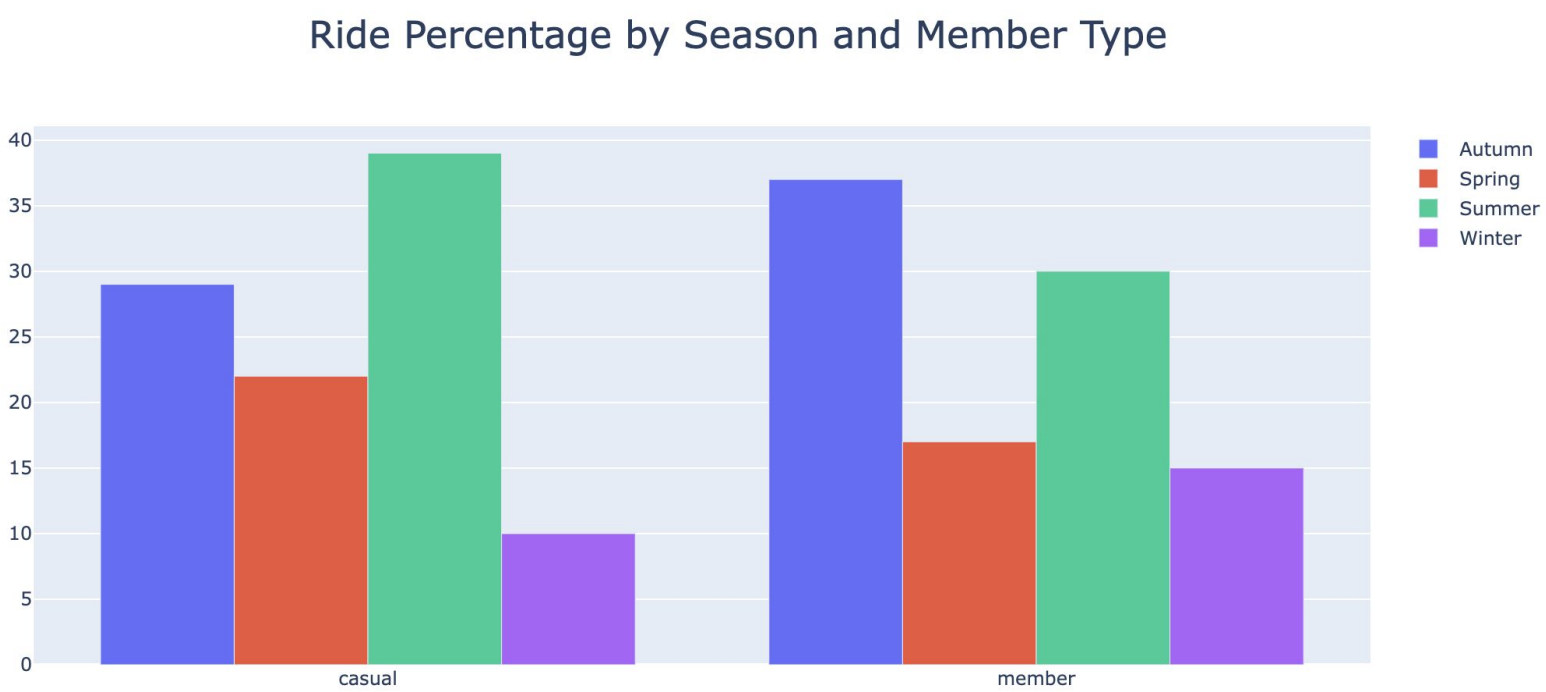
From the above, we see that there is not much difference between casual and member types. The number of rides during normal days is much greater than casual members.

Findings - Ride Percentage on Holidays



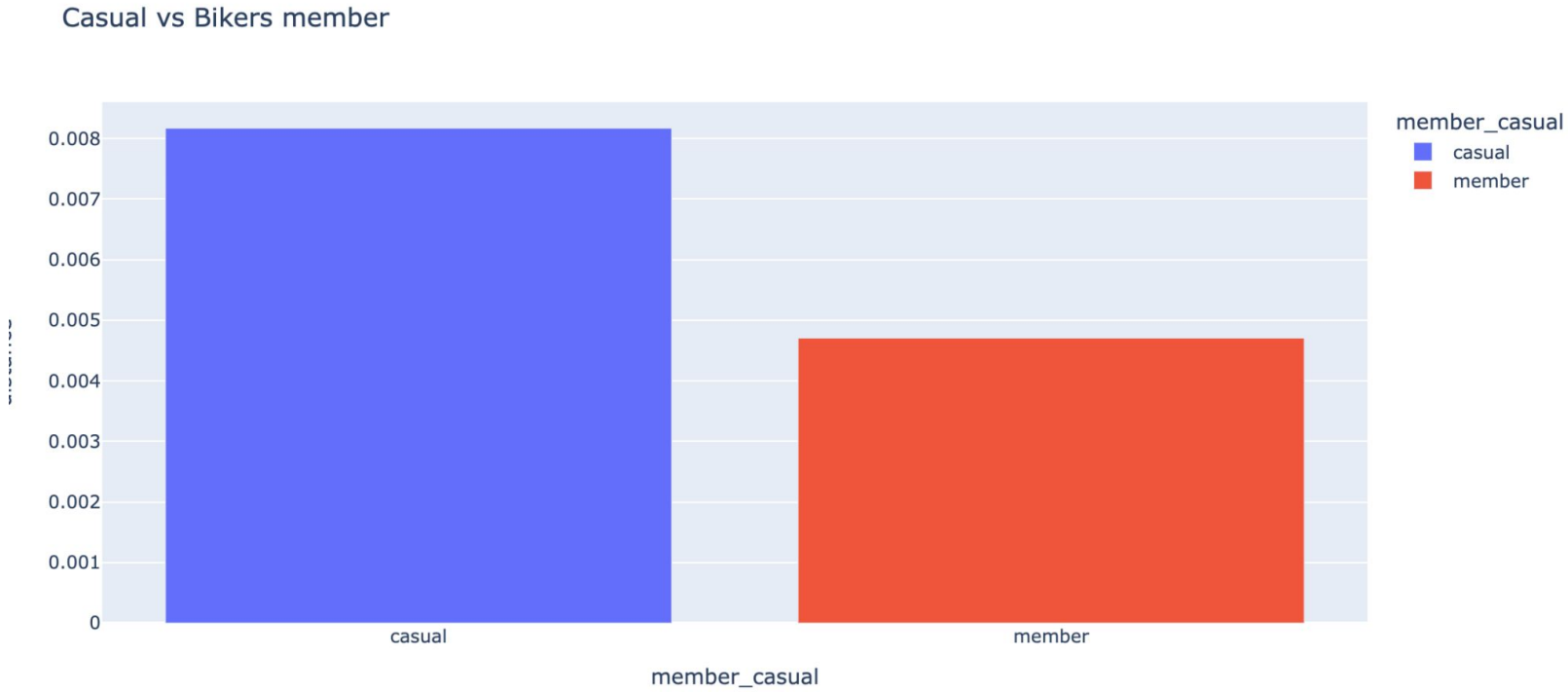
From the above, we see that there is not much difference between casual and member types. The number of rides during normal days is much greater than casual members.

Findings -RideablePercentage By Season



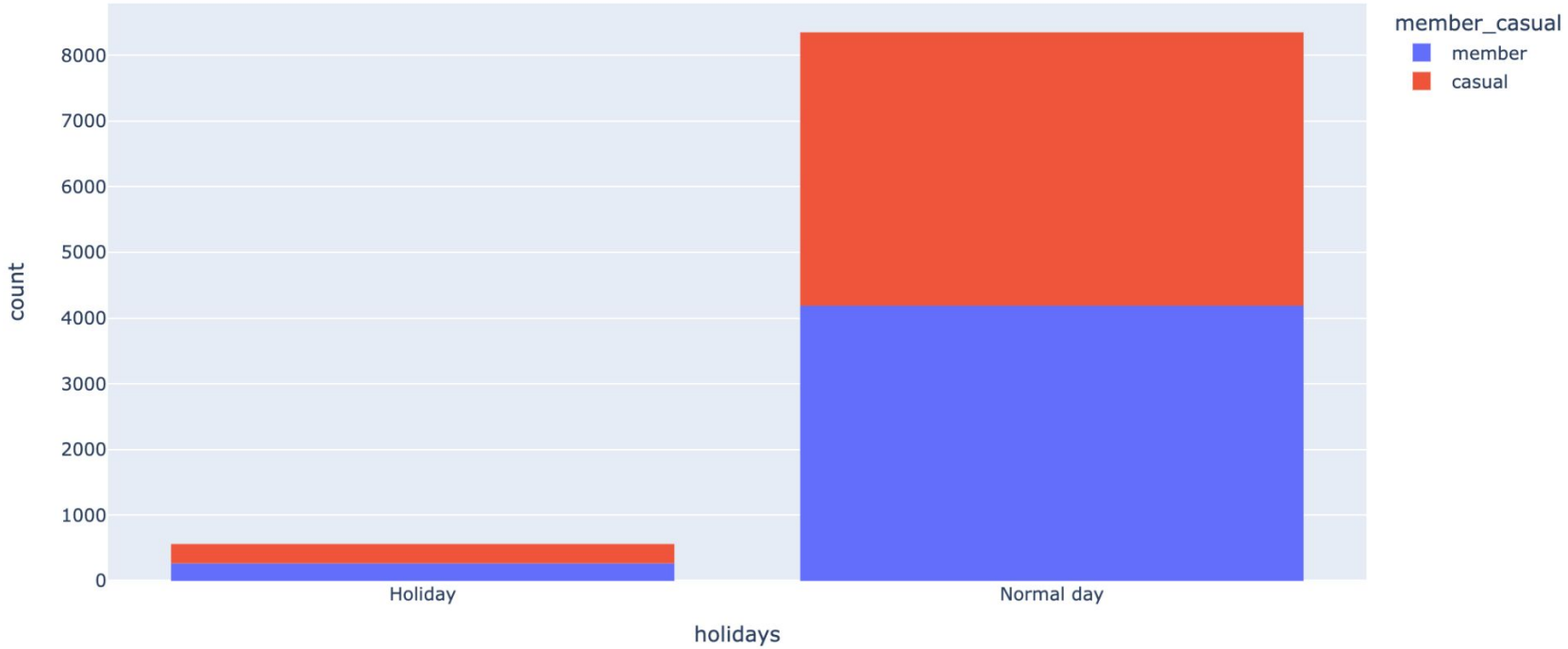
Based on the above plot, we see that Summer season is best to casual members and Autumn is best for member types.

Findings - Casual vs Members Distance



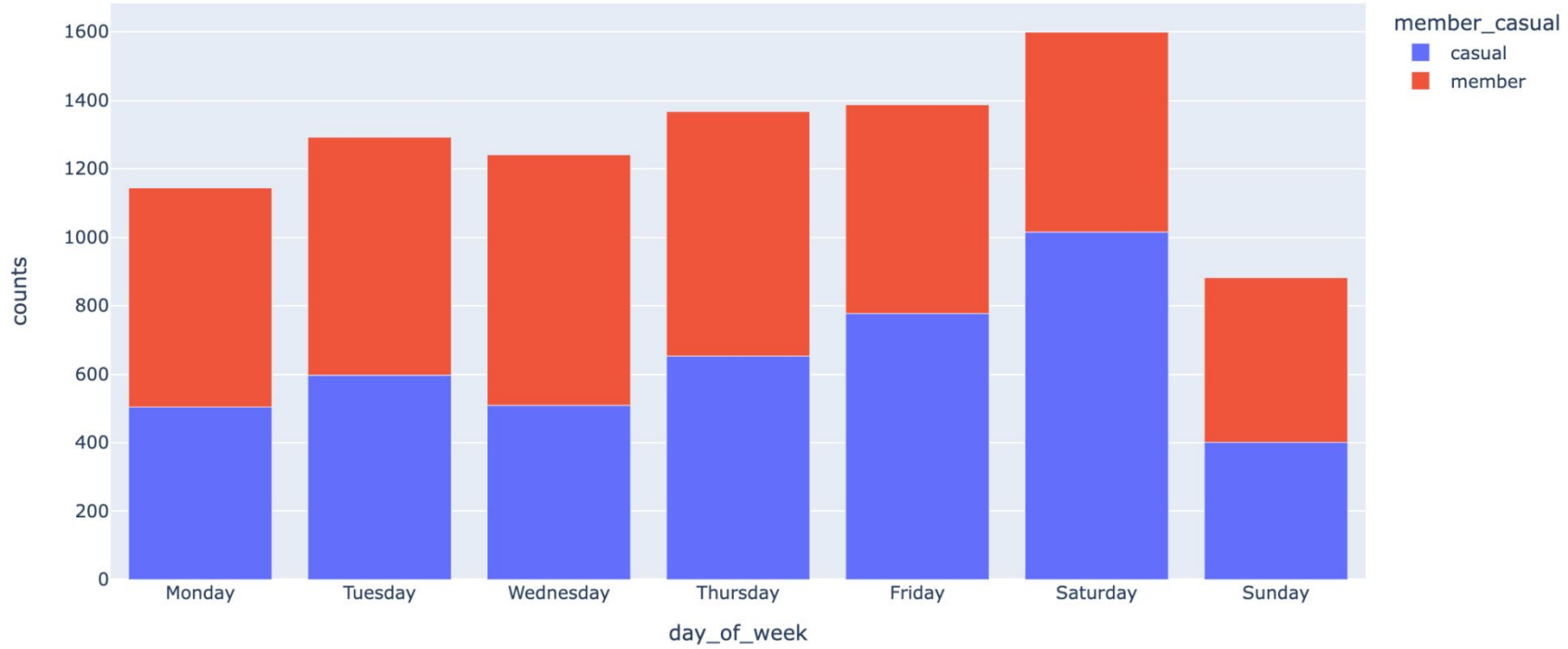
As you can see from the above, casual bike users travel greater distance than member bike users.

Findings - Do casual riders ride more on holidays?



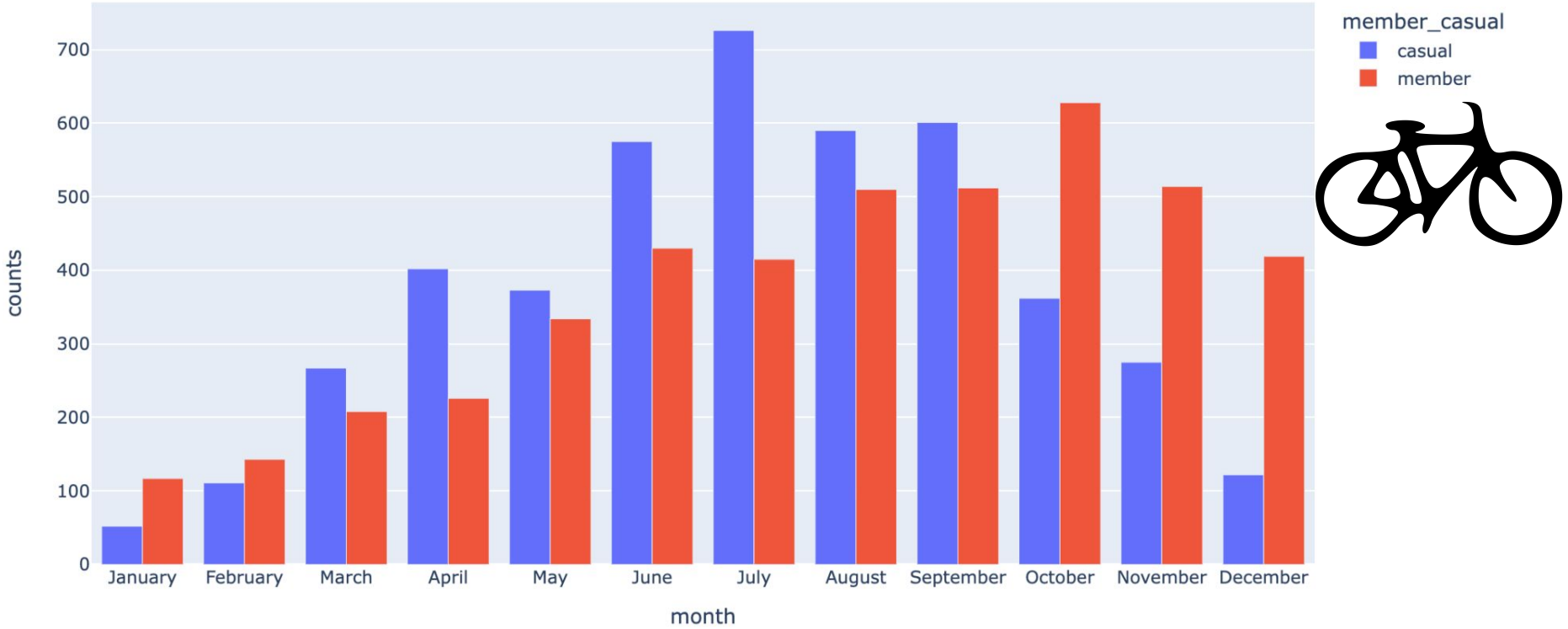
As we can see above, casual bikers ride more on a holiday and normal day. However, the total number of bikers are relatively smaller on a holiday.

Findings - Do member bikers ride more on weekdays?



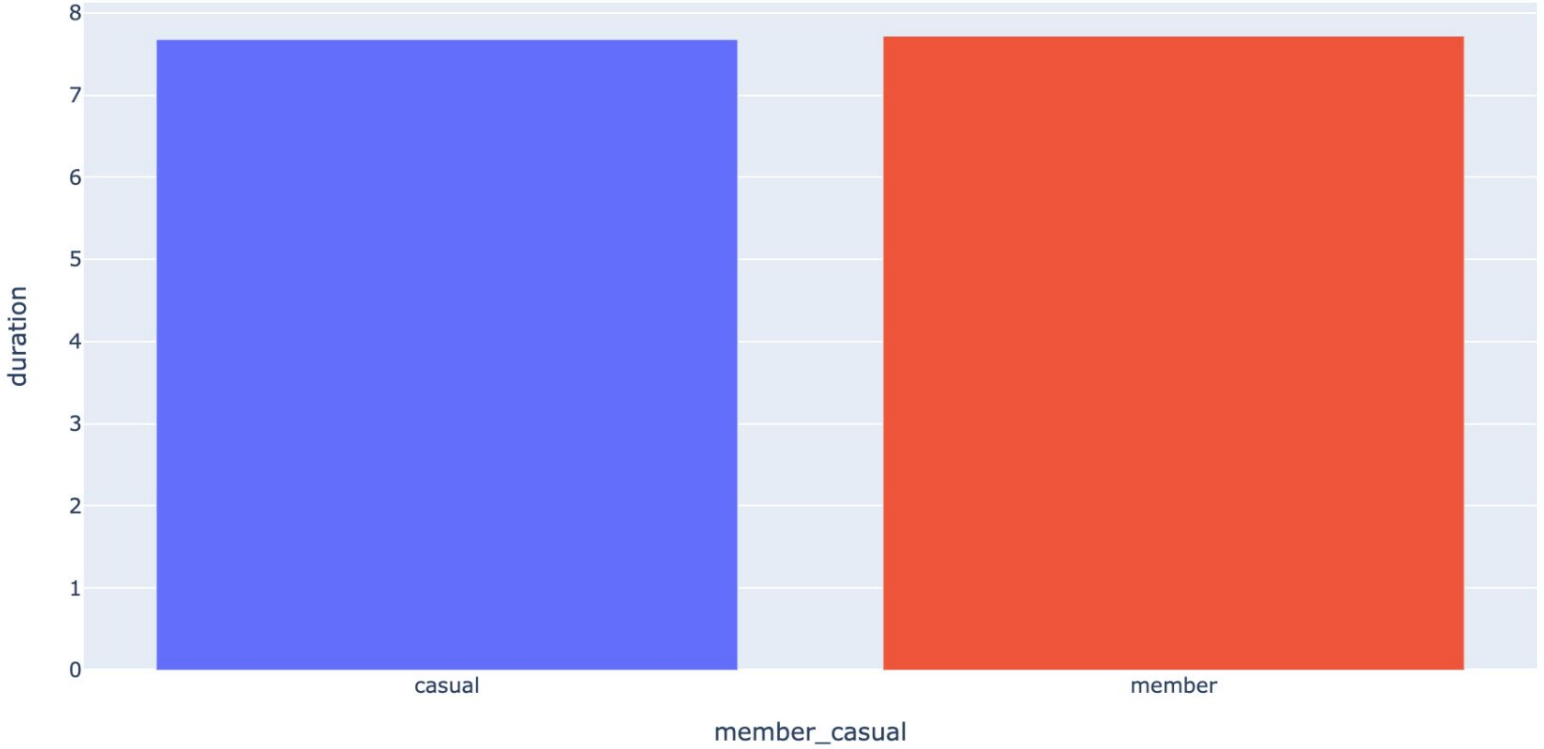
Member bikers do ride more of weekdays compared to weekends.

Findings - Do member bikers ride consistently over the months?



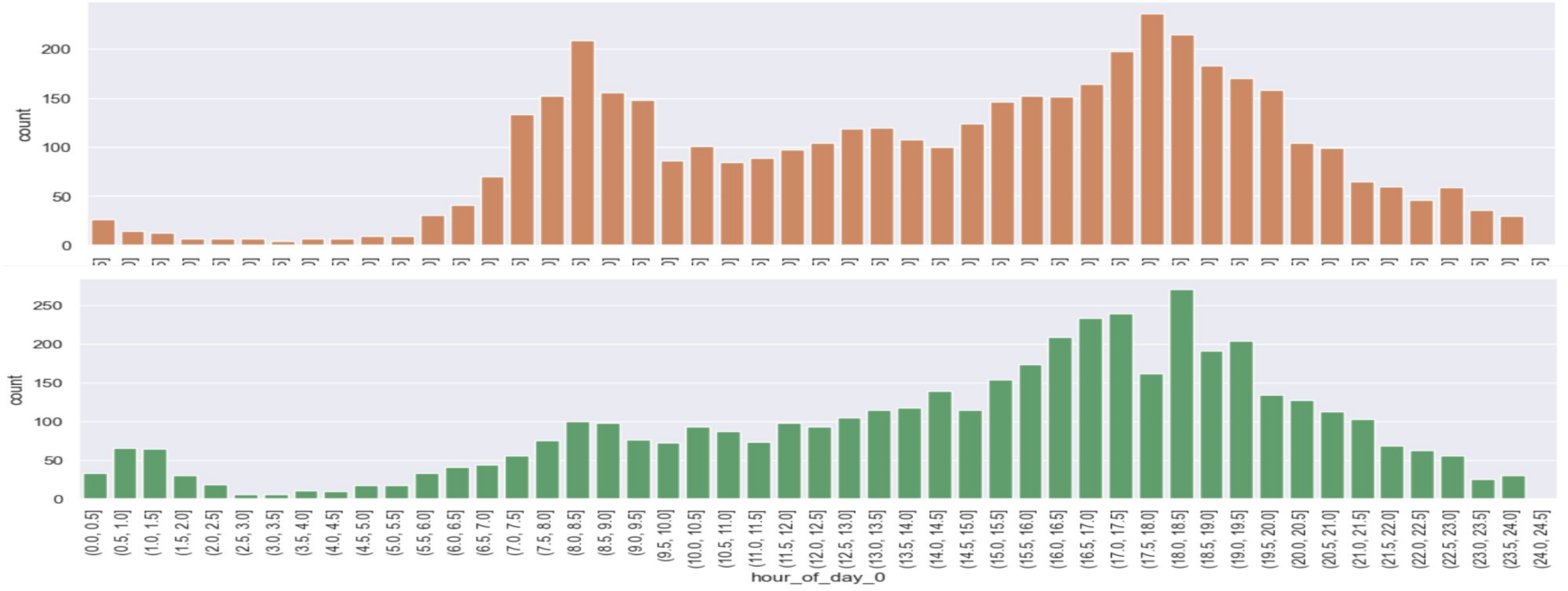
Member riders have been consistent over a period of time with respect to casual users. Overall usage of bike riding members have incresae in the months starting June to November.

Findings - Do casual bikers travel for long period of time?



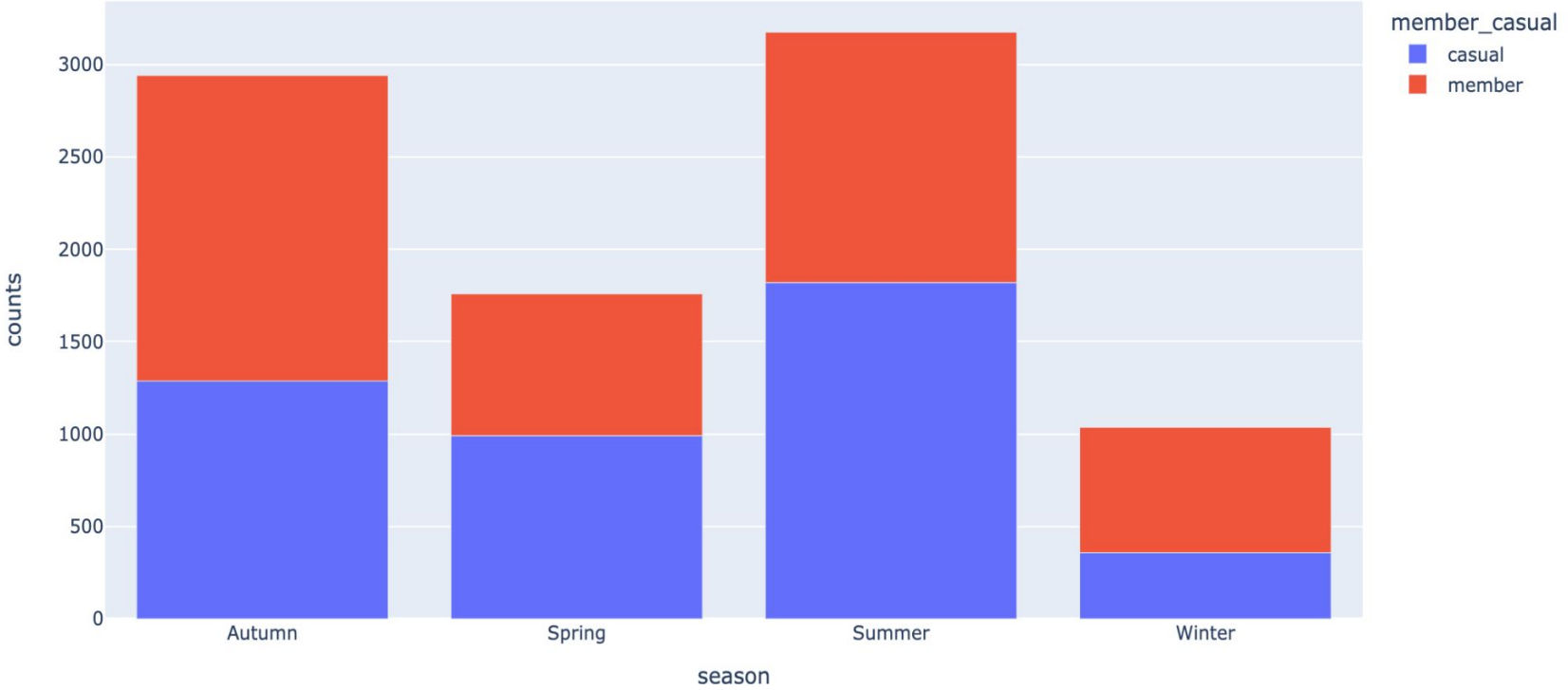
Casual bikers and member bikers ride for the same amount of time

Findings - Do member bikers use more in the morning or late afternoon?



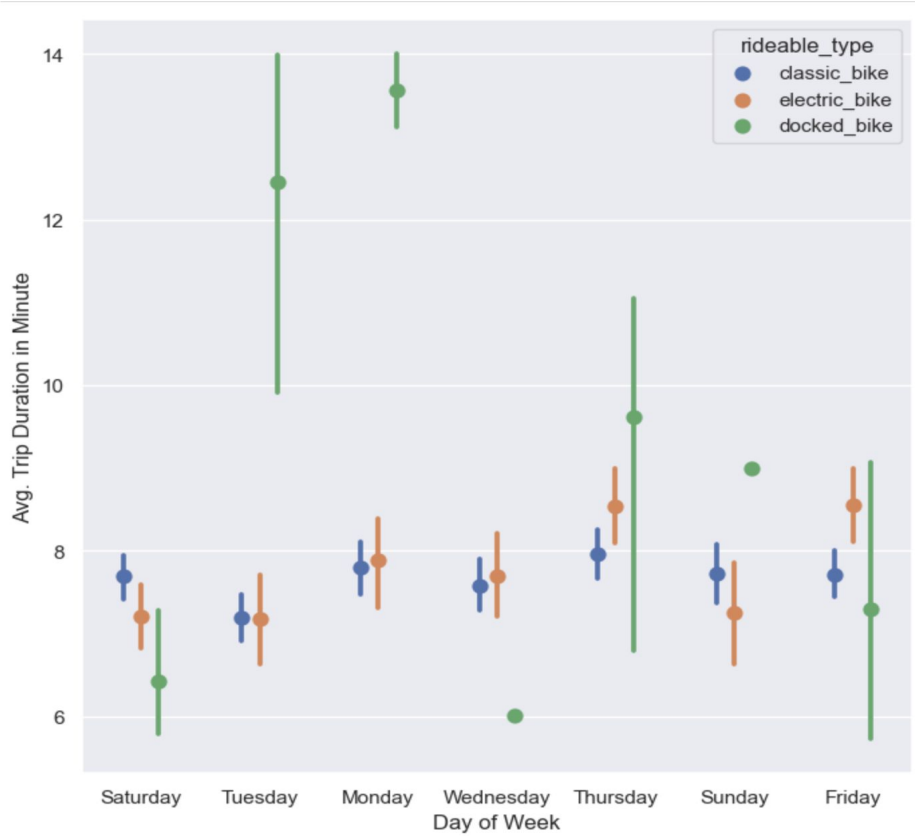
From the above, we can clearly see that the member bikers ride more during afternoon compared to morning.

Findings - Do casual bikers ride more in the summer compared to other seasons?



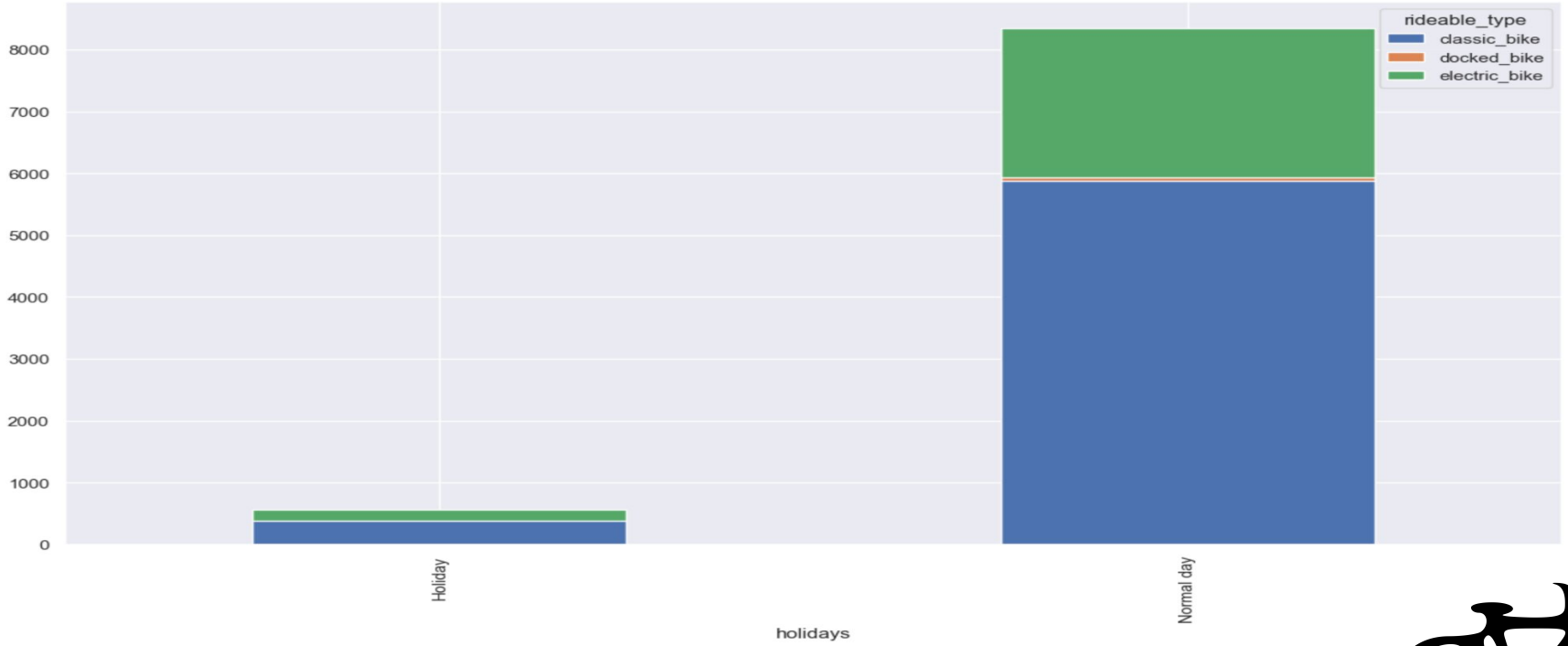
From the above, it is pretty clear that casual bike riders ride more during the summer season compared to other seasons.

Findings - How do the rideable types vary in weekdays?



From the above, it is pretty clear that casual bike riders ride more during the summer season compared to other seasons.

Findings - Compare the bike type usage on normal day with a holiday.



From the above, we can see that the classic bike and electric bikes are driven more compared to docked bike.

Findings

The findings from the coefficients above suggest that the latitude and longitude of both the start and end locations of the bike ride are positively correlated with the number of bike rides taken. The distance of the bike ride and the duration of the ride are also positively correlated with the number of bike rides, but the correlations are weaker compared to the latitude and longitude of the start and end locations.

The type of bike used also affects the number of bike rides taken. It appears that electric bikes are the most popular, followed by classic bikes and docked bikes. Additionally, there is a positive correlation between the number of bike rides taken and holidays, suggesting that more bike rides are taken on holidays.

Overall, these findings provide useful information for Citi Bike and stakeholders in the bike-sharing industry to understand the usage patterns and preferences of users, and to make informed decisions regarding the deployment of bikes and docking stations, marketing strategies, and other business operations.

Findings

Usage patterns:

- Casual bike users travel greater distances compared to member bike users.
- Casual bikers are more likely to ride on holidays and normal days, but the total number of bikers is relatively smaller on holidays.
- Member bikers are more likely to ride on weekdays compared to weekends.
- Member riders have been consistent over time, with an overall increase in usage from June to November.
- The busy hours for Citi Bike are between 6 am and 8:30 pm.
- Both member and casual bikers are more likely to ride in the afternoon compared to the morning.
- Casual bike riders are more likely to ride during the summer season compared to other seasons.
- The docked bike is more likely to be used on Wednesdays, Thursdays, and Fridays.
- The classic bike and electric bikes are more popular compared to the docked bike.
- The average distance travelled by bikers is less than one mile.
- The two most popular stations are Grove St Path and South Waterfront Walkway - Sintara Dr & 1st.