Summary of the analysis

Query 1 – determining number of trips per user type per month

```
SELECT
    DISTINCT(FORMAT_DATE("%B", start_date)) AS month,
    user_type,
    COUNT(*) AS number_of_trips
FROM `infinite-badge-322819.Cyclistic.clean_trips2019`
GROUP BY
    month,
    user_type
ORDER BY
    month DESC
```

Query 2 – determining number of trips per start time range

```
SELECT
    CASE
        WHEN start_time BETWEEN "00:00:00" AND "02:00:00"
        THEN "00AM-2AM"
        WHEN start_time BETWEEN "02:00:01" AND "04:00:00"
        THEN "2AM-4AM"
        WHEN start_time BETWEEN "04:00:01" AND "06:00:00"
        THEN "4AM-6AM"
        WHEN start_time BETWEEN "06:00:01" AND "08:00:00"
        THEN "6AM-8AM"
        WHEN start_time BETWEEN "08:00:01" AND "10:00:00"
        THEN "8AM-10AM"
        WHEN start_time BETWEEN "10:00:01" AND "12:00:00"
        THEN "10AM-12PM"
        WHEN start_time BETWEEN "12:00:01" AND "14:00:00"
        THEN "12PM-2PM"
        WHEN start_time BETWEEN "14:00:01" AND "16:00:00"
        THEN "2PM-4PM"
        WHEN start_time BETWEEN "16:00:01" AND "18:00:00"
        THEN "4PM-6PM"
        WHEN start_time BETWEEN "18:00:01" AND "20:00:00"
        THEN "6PM-8PM"
        WHEN start_time BETWEEN "20:00:01" AND "22:00:00"
        THEN "8PM-10PM"
        ELSE "10PM-12PM"
    END AS start_time_range,
    user_type,
    COUNT(*) AS number_of_trips
FROM `infinite-badge-322819.Cyclistic.clean_trips2019`
GROUP BY
    start_time_range,
    user_type
ORDER BY
    start_time_range ASC
```

Query 3 – determining number of trips and average length of the trip per day of the week

```
SELECT
```

```
    start_day,
    user_type,
    ROUND(SUM(tripduration)/60,2) AS total_duration,
    COUNT(*) AS number_of_trips,
    ROUND(AVG(tripduration)/60,2) AS mean_duration
FROM `infinite-badge-322819.Cyclistic.clean_trips2019`
GROUP BY
    start_day,
    user_type
ORDER BY
    start_day DESC
```

Query 4 – determining # of trips and their total duration per time in a day

```
SELECT
    start_day,
    user_type,
    start_time,
    COUNT(*) AS number_of_trips,
    ROUND(SUM(tripduration)/60,2) AS total_duration_min,
FROM `infinite-badge-322819.Cyclistic.clean_trips2019`
GROUP BY
    start_day,
    start_time,
    user_type
ORDER BY
    start_time ASC
```

Query 5 – finding main data qualities

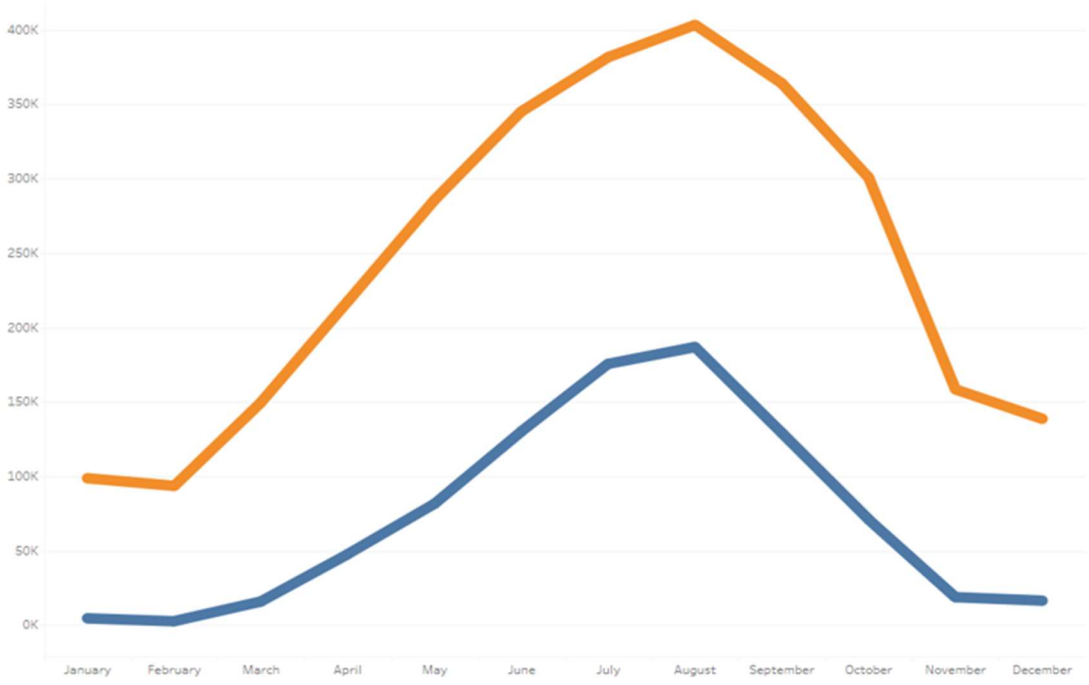
```
SELECT
    user_type,
    COUNT(*) AS num_of_trips,
    ROUND(COUNT(*)/SUM(COUNT(*)) OVER()*100,0) AS percent_of_trips,
    SUM(tripduration) AS total_trip_duration,
    ROUND(AVG(tripduration)/60,0) AS mean_duration_min,
    APPROX_TOP_COUNT(start_day, 1) AS top_weekday,
    ROUND(MAX(tripduration)/60,0) AS max_duration_min
FROM `infinite-badge-322819.Cyclistic.clean_trips2019`
GROUP BY
    user_type
```

Analysis

Main purpose of this project is to identify the differences in Cyclistic bikes usage between members and casual riders to provide Executive Team with recommendations to create marketing campaign to attract new members.

For this analysis we used monthly trip data provided by Divvy. Due to the ongoing COVID 19 pandemic we decided to analyze pre-pandemic data from 2019 to eliminate any changes in usage patterns due to the pandemic. Patterns identified in this analysis persist in 2020 as well, however it is less pronounced. 2020 data and charts will be provided in the appendix for review as well as be used to compare against 2019 findings.

We began by creating a chart displaying number of trips per month for members and casual riders. Orange line represents Members' usage and blue line represents casual riders. X-axis shows months and Y-axis shows number of trips per month. A total of 3,817,943 were completed in 2019 by members and casual riders. Whereas 77% of total number trips are attributable to members, duration wise these trips account for 46% of total ride duration. At the same time, Casual members account for 23% of total rides and 54% of total duration in a year. Finance analysts have concluded that annual members are more profitable than casual riders



The next table shows main usage indicators for 2019. We can see that Cyclistic usage by duration is evenly split between members and casual riders, even though members completed 3 times more trips than casual riders. This variance suggests that members tend to have shorter trips, which is backed by the data, average length of the member's trip is 14 minutes whereas casual rider average trip length is 56 minutes. Busiest day for members is Tuesday, while casual rode the most on Saturday.

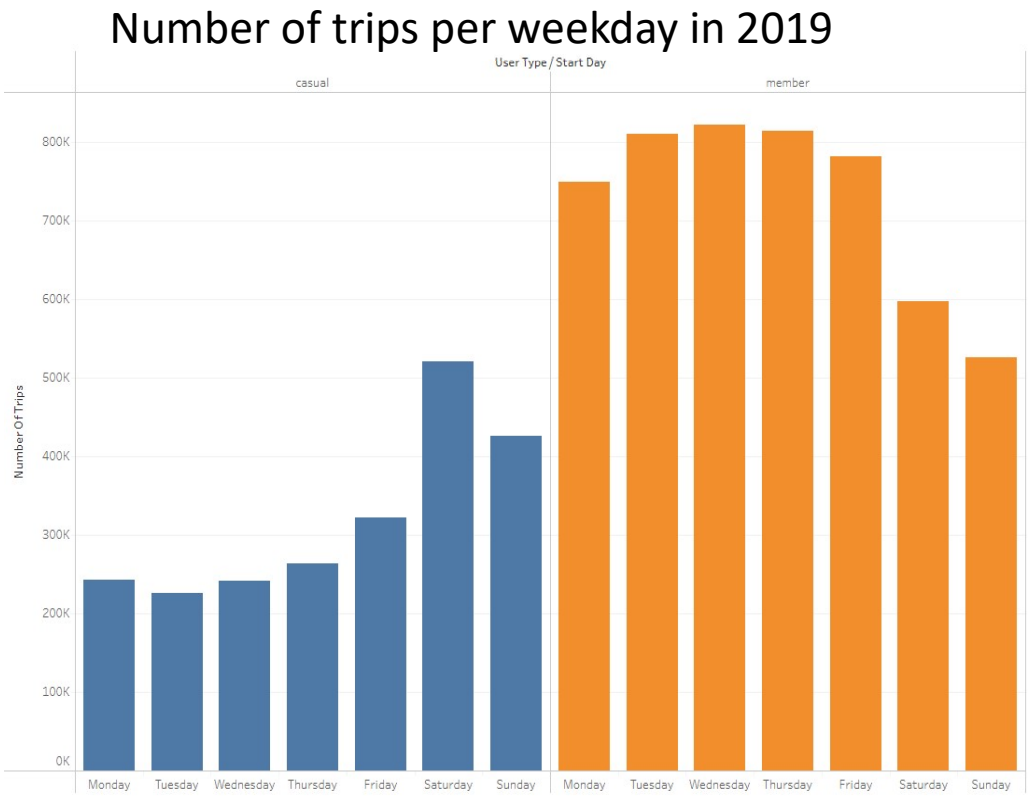
User Type	Number of trips	%	Total duration (min)	%	Mean duration (min)	Busiest day
member	2,937,350	77%	41,878,129	46%	14	Tuesday
casual	880,593	23%	48,938,763	54%	56	Saturday

With this data in mind out Hypothesis is that Members use Cyclistic bike sharing service predominantly for commute while casual riders use the service for primarily for recreation.

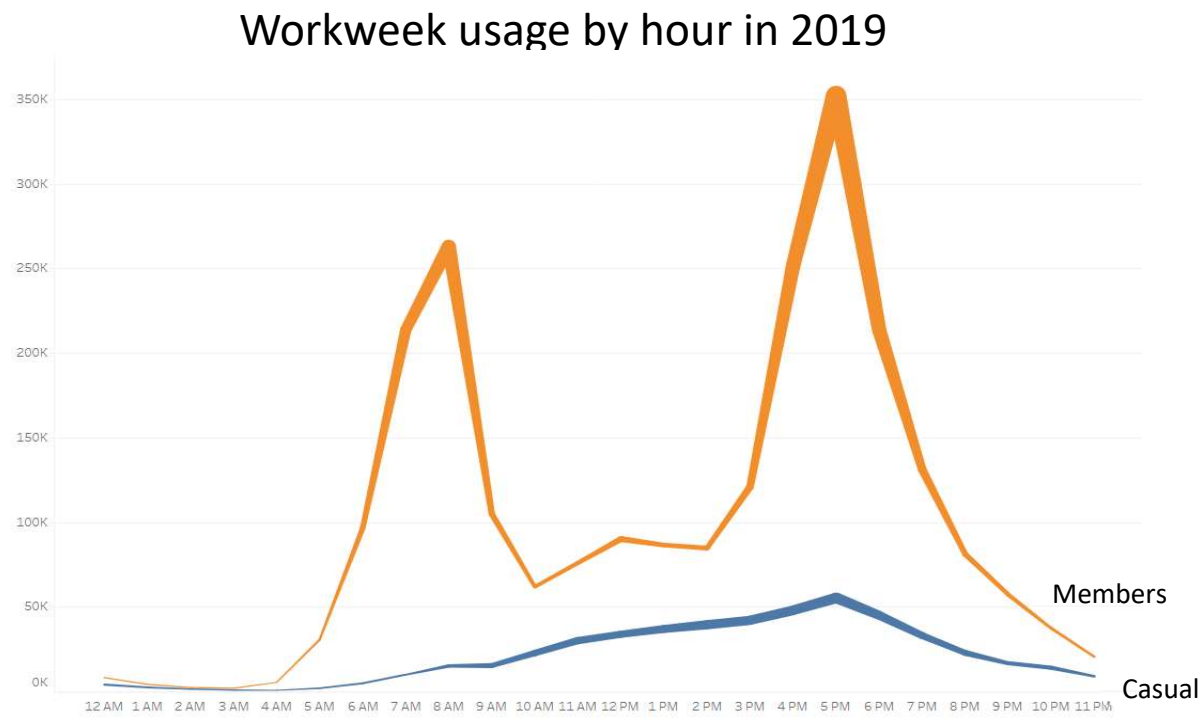
To investigate the hypothesis, we will investigate the following data points:

1. Number of trips per weekday
2. Number of trips per start time
3. Trip duration patterns
4. Correlation between time, day of the week and the number of trips
5. Trips distribution by location

We started by charting number of trips per weekday. Each bar represents the number of trips on each day of the week. Orange bars represent members and blue bars represent casual riders. Members started 82% of trips during the workweek (Monday through Friday). Number of trips started by members throughout the workweek remains stable, between 456K and 497K. However, as we approach the weekend, it decreases significantly – to 287K and 256K on Saturday and Sunday respectively. It is the opposite with casual riders. Only 57% of trips started between Monday and Friday and the rest 43% started on the weekend.

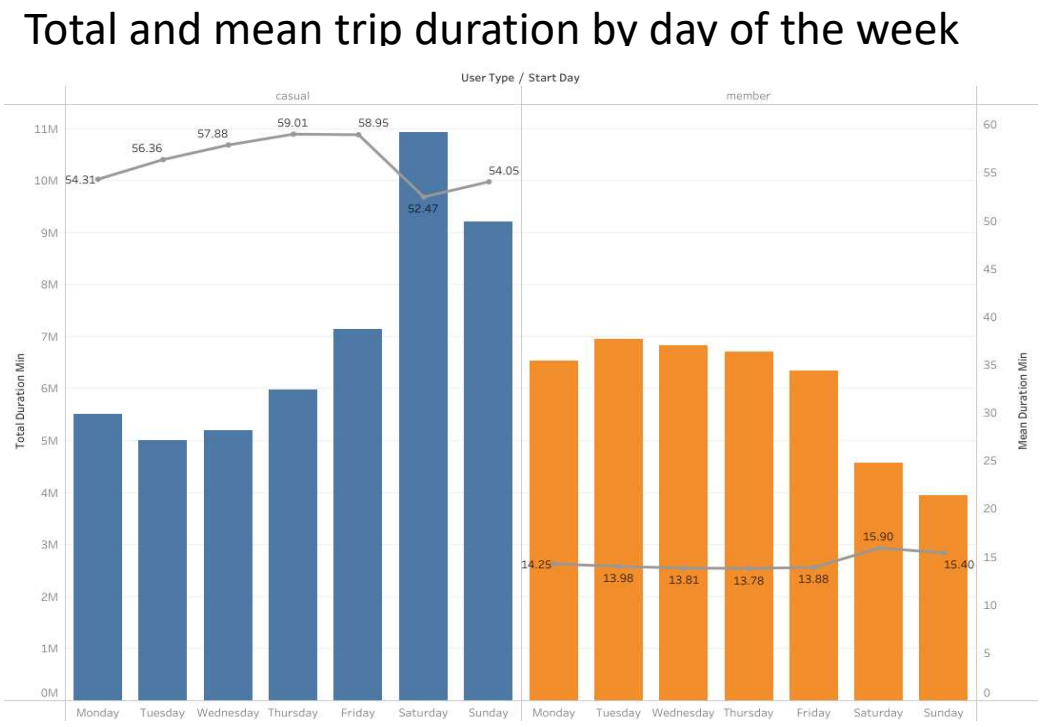


To further explore usage patterns between members and casual riders we looked at the time of the day trips are initiated during the workweek. The chart shows the number of trips per time-of-day by user type. The higher the point on the chart, the more trips were completed during that time. Time is charted on X-axis from 12 AM to 11PM. Line's thickness represents trip duration, the thicker the line, the longer trips were during that time.



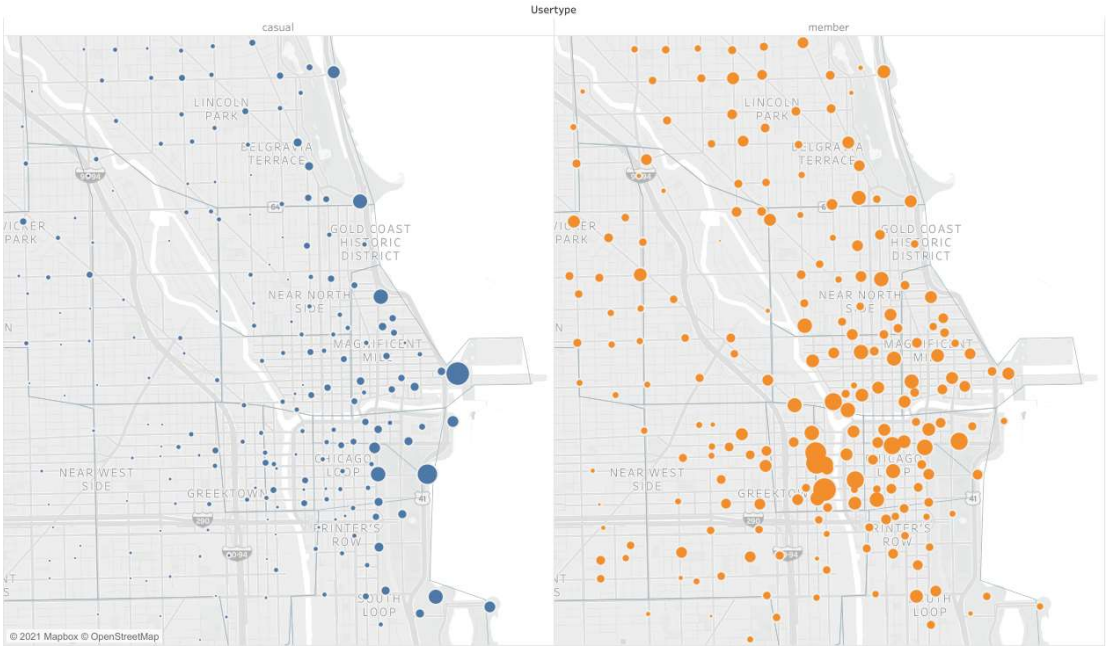
The graph shows member usage peaking at 8AM, dropping significantly by 76.5% by 10AM and then ramping back up to daily high around 5PM. Members start 37% of the trips between 7AM and 9AM and between 4PM and 6PM, corresponding to the typical workday schedule. On the other hand, casual riders' trips experience a steady increase from 5AM to 5PM and then slow down by 11PM. Weekend usage pattern is very similar between members and casual riders. The difference in usage between members and casual riders so far supports our hypothesis.

Now let's review total and mean trip duration. The next bar chart shows total trip duration by day of the week. Line chart on secondary axis displays mean trip duration for each day



Total trip length is increasing for casual riders throughout the week, peaking on Saturday. At the same time, trip length for members remains stable Monday through Friday, and decreasing during the weekend. On average members ride 42 minutes less than casual riders.

To conclude our analysis, we will explore the location of the station where most trips start. The next chart represents the number of trips started at each location throughout Chicago. Circles correspond to the location of the station and the size of the circle represents the number of trips started at that location. The bigger the circle, the more trips started at that location.



Looking at the geographical distribution of the trips, we can see that casual riders' trips gravitate towards the shoreline and the biggest number of trips started at the Streeter Dr and Grand Ave location at the Pier. Members trips are more evenly distributed throughout the downtown and concentrated more in Greektown area

To conclude, members tend to have shorter trips concentrated around the 8AM and 5PM during workweek, corresponding to the typical working schedule. Conversely, busiest time of day for casual riders is between 12PM and 3PM on weekends. Furthermore, members' start locations are distributed around the city, when casual riders choosing more scenic locations in Magnificent Mile and Pier. All this evidence supports our hypothesis that members use service primarily for work commute, whereas casual riders use it as leisure activity. This is our conclusion based on the trip data, further research is needed to confirm the findings and identify other use cases.

Based on our conclusions we recommend:

1. Conducting a survey among members and casual riders to identify their preferences and use cases for the service. This will provide executive team with additional firsthand insights into clients' preferences as well as give new data to the analysis to verify the initial hypothesis.
2. Conducting a separate analysis focusing on impact of the pandemic. Review how pandemic affected the usage, member retention rate, casual rider conversion rate and other indicators to assess how stay-at-home advisory and work from home policies affected client population
3. Providing incentives for current casual members to convert to members. Given that financial team concluded that annual members are more profitable than casual riders, our goal is to incentivize current clients to convert from casual riders to members.
4. Working with the Universities in Chicago area to subsidize discounted membership for students, faculty, and staff. This will provide further incentive for the University personnel to use bike sharing for commuting.
5. Sending monthly newsletter to current members showing how their usage of bike sharing for commuting helps the environment. This can include personalized amount of CO_2 emissions saved by using bikes instead of a car.