1. Introduction

Cyclistic is a bike-share company based in Chicago, Illinois. It was launched in 2016 and since then the program has grown exponentially. It has a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago.
The company has flexibility in its pricing plans with them being: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as casual riders. Customers who purchase annual memberships are Cyclistic members.

2. Scenario

The director of marketing, Lily Moreno, believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently.

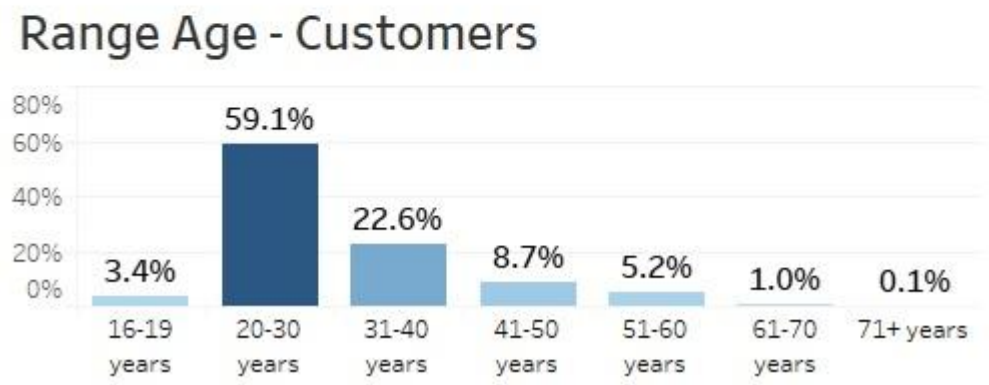
3. Objective - Business problem

The goal is to help the marketing team design marketing strategies aimed at converting casual riders into annual members. In order to do that we need to answer the question:
How do annual members and casual riders use Cyclistic bikes differently?

4. Data Analysis

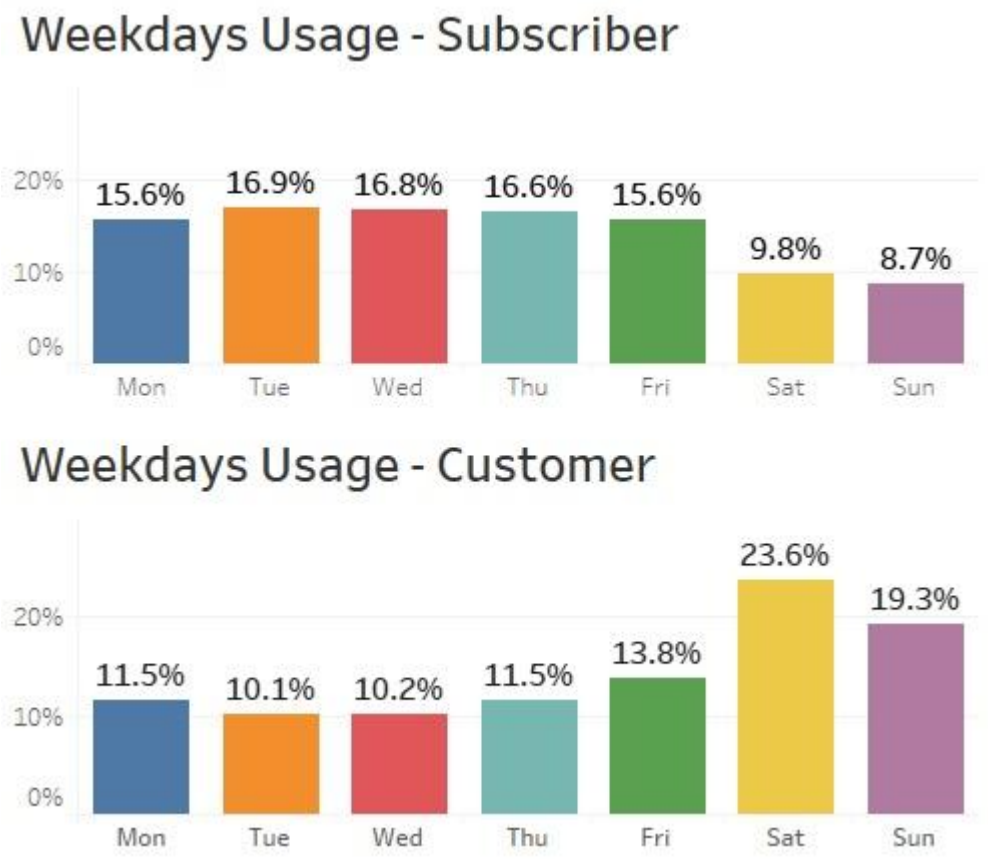
We will now analyse the data provided by the company from all the rides made in 2019. With this data we will compare how annual members(subscribers) and casual riders(customers) use Cyclistic Bikes differently:

4.1 Age Range



With the graphs above we can notice that the customers group is massively concentrated between the ages of 16 and 40 years with 85% of people being from this group. Whereas, the subscriber group has just 74% for the same range.

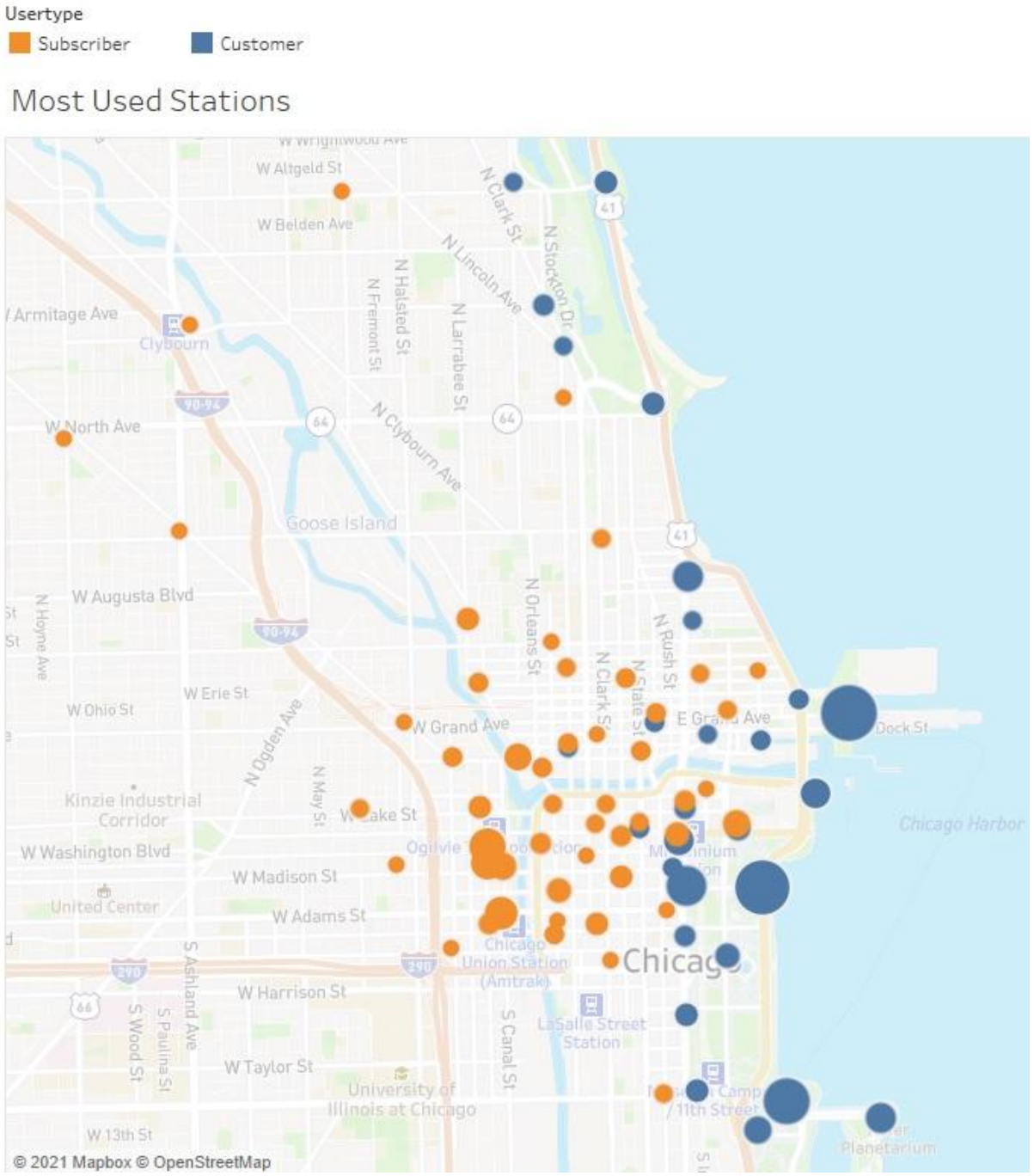
4.2 Most used days of the week



Looking at the subscribers graph it's easy to see that most of their rides happen on weekdays with just 18.5% of the rides on weekends. While the customers use the service more on weekends with over 42% of the rides happening on Saturday and Sunday.

Now with these two graphics we can start to understand a little bit more how customers and subscribers use the Cyclistic bike-share system. Customers are more concentrated between the ages of 20-30 years(60%) and they ride it more often on weekends. So, it looks like they prefer to use the system more for leisure. Whereas, the subscribers have their ages more concentrated between 20-40 years(74%) and use it more during weekdays, which indicates that they use it more for commuting.
To confirm that we can see a map of the city of Chicago with the location of the most used stations by Subscribers and Customers.

4.3 Most used stations - Map city of Chicago, IL



Above we have a map of the city of Chicago with the most used bike-stations by customers and subscribers. To be more precise we are seeing 50% of the preferred stations by each type of client. The stations are represented by circles and the size of the circle represents how much is that station used.

We can easily notice that the customers use way more of the bike-stations at the coast of the city where the main tourist attractions are located. Which indicates that they prefer to use the Cyclistic bike-share system for leisure.
When we look at the bike-stations used by the subscribers we can notice that they are located in the main part of the city, with the most used ones located next to train stations, highly indicating that the subscribers prefer to use the system for commuting.

5. Conclusion

Now we understand more about how subscribers and casual riders use Cyclistic bikes differently. We can now sum it up on the bullets points below:

- The majority of customers are between 20-30 years of age. Whereas, subscribers are more spread out between the ages of 20-40 years old;

- On one hand customers tend to use the system more on weekends with 42% of their rides happening on Saturday and Sunday. On the other hand subscribers like to use the system more often on weekdays, with just 18% of their rides being on weekends;

- We can confirm on the map that customers tend to use the system way more for leisure with their preferred bike-stations being on the coast next to the tourist spots. Also we can see that subscribers like to use the Cyclistic system more for commuting, with their top bike-stations being next to train stations.

Appendix

- Getting the percentage of rides made by different user types and gender:

```
SELECT
    gender,
    COUNT(gender) AS gender_type,
    ROUND(COUNT(gender) * 100.0 / (select count(*) from (SELECT
       gender,
       usertype

       FROM `lithe-sonar-317601.case_study_1.trips_2019`

        WHERE gender IS NOT NULL)), 2) as gender_percentage,
    usertype

FROM (SELECT
    gender,
    usertype

       FROM `lithe-sonar-317601.case_study_1.trips_2019`

       WHERE gender IS NOT NULL)

GROUP BY gender, usertype
```

- Creating age ranges for each type of user

```
SELECT
    usertype,
    CASE
        WHEN birthyear <= 2003 AND birthyear > 1999 THEN '16-19 years'
        WHEN birthyear <= 1999 AND birthyear > 1988 THEN '20-30 years'
        WHEN birthyear <= 1988 AND birthyear > 1978 THEN '31-40 years'
        WHEN birthyear <= 1978 AND birthyear > 1968 THEN '41-50 years'
        WHEN birthyear <= 1968 AND birthyear > 1958 THEN '51-60 years'
        WHEN birthyear <= 1958 AND birthyear > 1948 THEN '61-70 years'
        ELSE '71+ years'
        END AS range_birthyear,
    COUNT(CASE
        WHEN birthyear <= 2003 AND birthyear > 1999 THEN '16-19 years'
        WHEN birthyear <= 1999 AND birthyear > 1988 THEN '20-30 years'
        WHEN birthyear <= 1988 AND birthyear > 1978 THEN '31-40 years'
        WHEN birthyear <= 1978 AND birthyear > 1968 THEN '41-50 years'
        WHEN birthyear <= 1968 AND birthyear > 1958 THEN '51-60 years'
        WHEN birthyear <= 1958 AND birthyear > 1948 THEN '61-70 years'
```

```
            ELSE '71+ years'
            END) AS total_range_birthyear

    FROM `lithe-sonar-317601.case_study_1.trips_2019`

    WHERE gender is not null AND usertype = 'Subscriber'

    GROUP BY range_birthyear, usertype
    ORDER BY range_birthyear
```

- Getting the most used stations by each type of user

```
SELECT
    from_station_name,
    COUNT(from_station_name) AS trips_per_station,
    ROUND(COUNT(from_station_name) * 100.0 / (select count(*) from (SELECT
    from_station_name,
    usertype

    FROM `lithe-sonar-317601.case_study_1.trips_2019`

    WHERE usertype = 'Subscriber')), 2),
    usertype

FROM `lithe-sonar-317601.case_study_1.trips_2019`

WHERE usertype = 'Subscriber'
GROUP BY usertype, from_station_name
ORDER BY trips_per_station DESC
```

- % of use by day of the week

```
SELECT
    COUNT(trip_id) AS total_trip,
    CAST(start_time AS DATE) AS day

FROM `lithe-sonar-317601.case_study_1.trips_2019`

WHERE usertype = 'Customer'

GROUP BY day
ORDER BY day ASC
```