Case Study: How Does a Bike-Share Navigate Speedy Success?

Introduction

This case study project is a Google Data Analytics Certificate requirement. I will perform as a junior data analyst in a marketing team, work for a fictional company, Cyclistic, a bike-share company in Chicago, and meet different characters and team members. The marketing director believes the company's future success depends on maximizing the number of annual memberships. Therefore, our team wants to understand how casual riders and annual members use Cyclistic bikes differently. Our team will design a new marketing strategy to convert casual riders into annual members. I will follow these steps of the data analysis process: ask, prepare, process, analyze, share, and act.

Stakeholders:

Cyclistic: A bike-share program that features more than 5,800 bicycles and 600 docking stations.

Lily Moreno: The director of marketing and your manager.

Cyclistic marketing analytics team: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.

Cyclistic executive team: The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

About the company

In 2016, Cyclistic launched a successful bike-share offering. Since then, the program has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Until now, Cyclistic's marketing strategy relied on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: 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.

Cyclistic's finance analysts have concluded that annual members are much more profitable than casual riders. Although the pricing flexibility helps Cyclistic attract more customers, Moreno believes that maximizing the number of annual members will be key to future growth. Rather than creating a marketing campaign that targets all-new customers, Moreno believes there is a very good chance to convert casual riders into members. She notes that casual riders are already aware of the Cyclistic program and have chosen Cyclistic for their mobility needs.

Moreno has set a clear goal: Design marketing strategies aimed at converting casual riders into annual members. In order to do that, however, the marketing analyst team needs to better understand how annual members and casual riders differ, why casual riders would buy a membership, and how digital media could affect their marketing tactics. Moreno and her team are interested in analyzing the Cyclistic historical bike trip data to identify trends

Ask

The questions that needs to be answered are:

- How do annual members and casual riders use Cyclistic bikes differently?
- Why would casual riders buy Cyclistic annual memberships?
- How can Cyclistic use digital media to influence casual riders to become members?

Prepare

Data Source

The dataset used in this case study is Cyclistic trip data and made available by Motivate International Inc. under this license. Analyzing this case study is made using data from April 2020 to March 2021, which is the last 12 months of Cyclistic trip data. It is organized as separate files by month and year and was saved as .zip files. I downloaded the .zip files and extracted them.

Each dataset contains the following columns:

ride_id: a unique ID for each rider

rideable_type: the type of bike used

started_at: the date and time the trip was started

ended_at: the date and time the trip was ended

start_station_name: the name of the starting station

start_station_id: the unique ID of the starting station

end_station_name: the name of the ending station

end_station_id: the unique ID of the ending station

start_lat: the latitude of the starting station

start_lng: the longitude of the starting station

End_lat:the latitude of the ending station

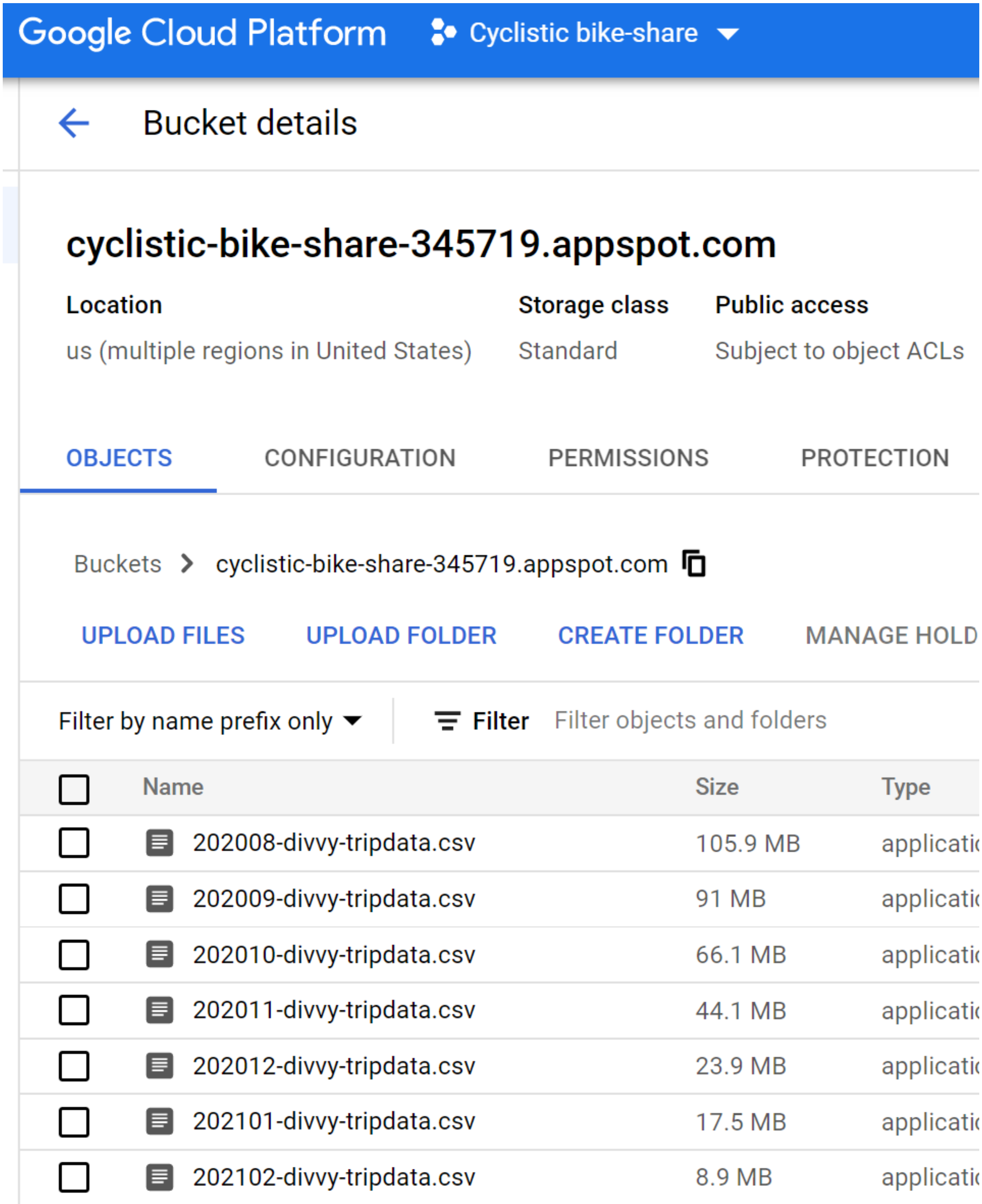
end_lng: the longitude of the ending station

member_casual: the riders memberships status (member or casual)

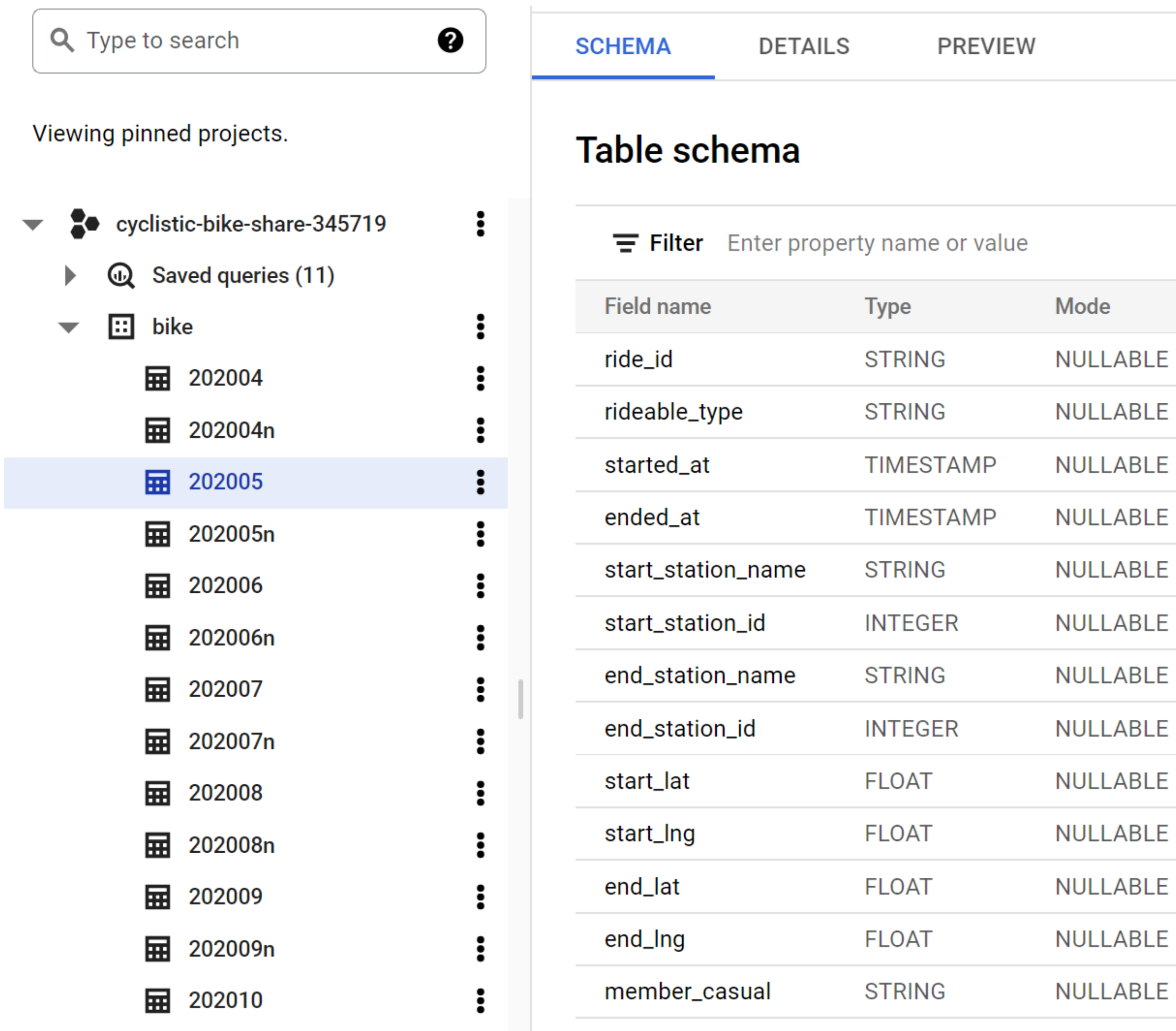
Process

I am using BigQuery from Google Sandbox for data preparation.

What I am going to do first is to upload 12 CSV data files into Google Cloud Storage.



After uploading all 12 tables into Google Cloud, then upload those data files under the “bike” dataset in BigQuery. Then I take a quick preview of each table and check its schema for consistency. Then I discovered that the data table from 202004 to 202011’s end_station_id and start_station_id’ s type is different from 202012 to 202103.



Check if there are any duplicates or errors on rideable_type and member_casual by using the SELECT DISTINCT clause; after running the query, there is no duplicate.

For all the columns to be consistent, I need to change the "start_station_id" and "end_station_id" ' s data types from integer to string in tables 202004 to 202011. I am using CREATE TABLE to create a new table that indicates to change the data type.

RUN SAVE SHARE SCHEDULE MORE

```
CREATE TABLE bike.202006n AS
SELECT ride_id, rideable_type,started_at,ended_at,start_station_name,
CAST(start_station_id AS STRING) AS start_station_id, end_station_name,
CAST( end_station_id AS STRING) AS end_station_id,start_lat, start_lng,
end_lat, end_lng,member_casual
FROM bike.202006
```

Combine 12 tables into one new table by using UNION ALL.

RUN SAVE SHARE SCHEDULE

```
CREATE TABLE
  bike.total AS
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202004n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202005n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202006n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202007n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202008n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202009n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202010n`
UNION ALL
SELECT
  *
FROM
  `cyclistic-bike-share-345719.bike.202011n`
UNION ALL
SELECT
  *
FROM
```

All datasets are consistent with 12 columns, same date frames, and consistent column names. Many data sets are missing the start and end station names and IDs, LNG and LAT columns, but that won't change the analysis since there is information on station coordinates to calculate ride distances.

For example, by using SELECT… FROM… WHERE… clause to find out what is missing on this table.

```
SELECT
  ride_id,end_station_id
FROM
  bike.total
WHERE
  end_station_id IS NULL;
```

And on the right corner of the page, it shows there is a total number of 143703 missing values of end_station_id.

Then, delete the date and time of the trip that is started (started_at), which is larger or equal to the date and time of the trip that is ended (ended_at) by using the DELETE function.

```
DELETE FROM bike.total WHERE datetime(started_at) >= datetime(ended_at)
```

I am using the DATETIME_DIFF function to calculate the time difference between the started_at and ended_at fields to know the duration of trips.

```
SELECT
  ride_id,
  DATETIME_DIFF(ended_at,
    started_at,
    second) AS trip_length,
FROM
  bike.total;
```

I am using a subquery to find out the maximum time difference between the started_at and ended_at fields.

```
SELECT
  *
FROM
  bike.total
WHERE
  DATETIME_DIFF(ended_at,
    started_at,
    minute) IN (
  SELECT
    MAX(DATETIME_DIFF(ended_at,
        started_at,
        minute))
  FROM
    bike.total);
```

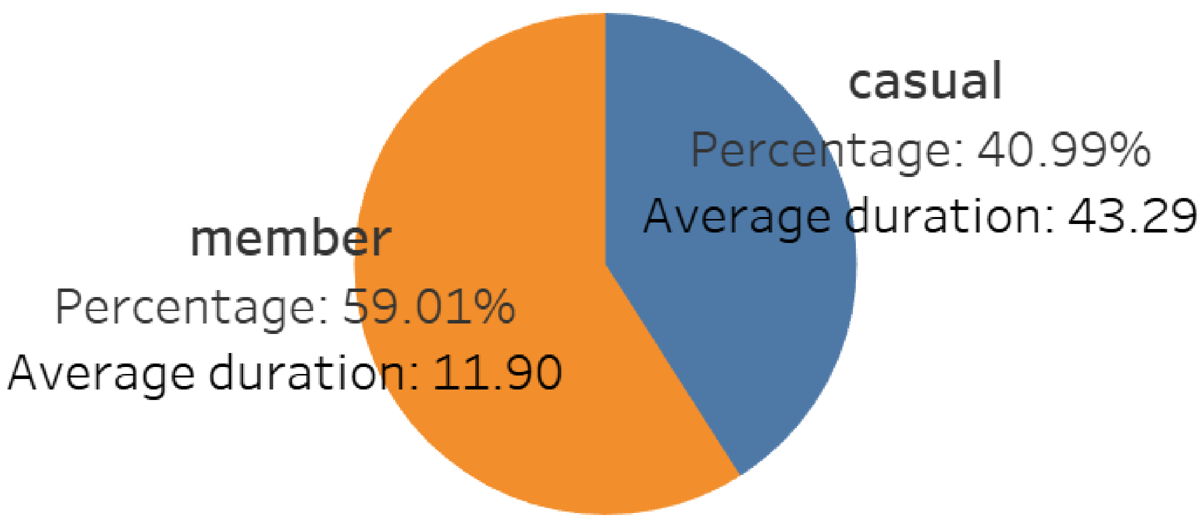
I am using a subquery to calculate the average bike renting time for members or casual riders.

```
SELECT
  member_casual,
  COUNT(*) AS num_ride,
  AVG(trip_length) AS average_ridetime,
  MAX(trip_length) AS max_ridetime,
FROM (
  SELECT
    ride_id,
    member_casual,
    DATETIME_DIFF(ended_at,
      started_at,
      minute) AS trip_length
  FROM
    bike.total )
GROUP BY
  member_casual;
```

member_casual	num_ride	average_ridetime	max_ridetime
member	2059372	11.415756356792004	58720
casual	1430376	42.799977768083679	55683

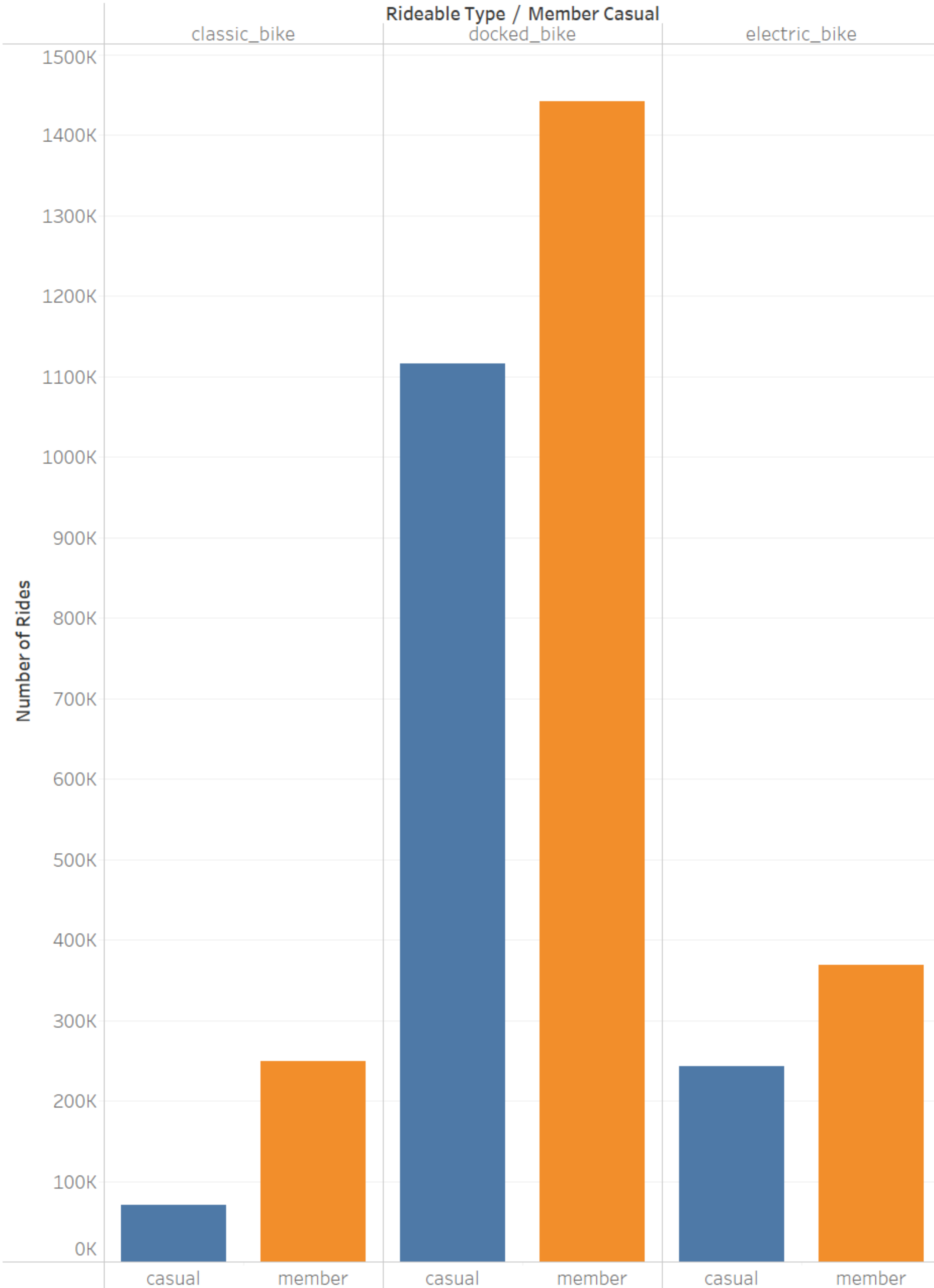
Analyze

Percentile for Member and Casual



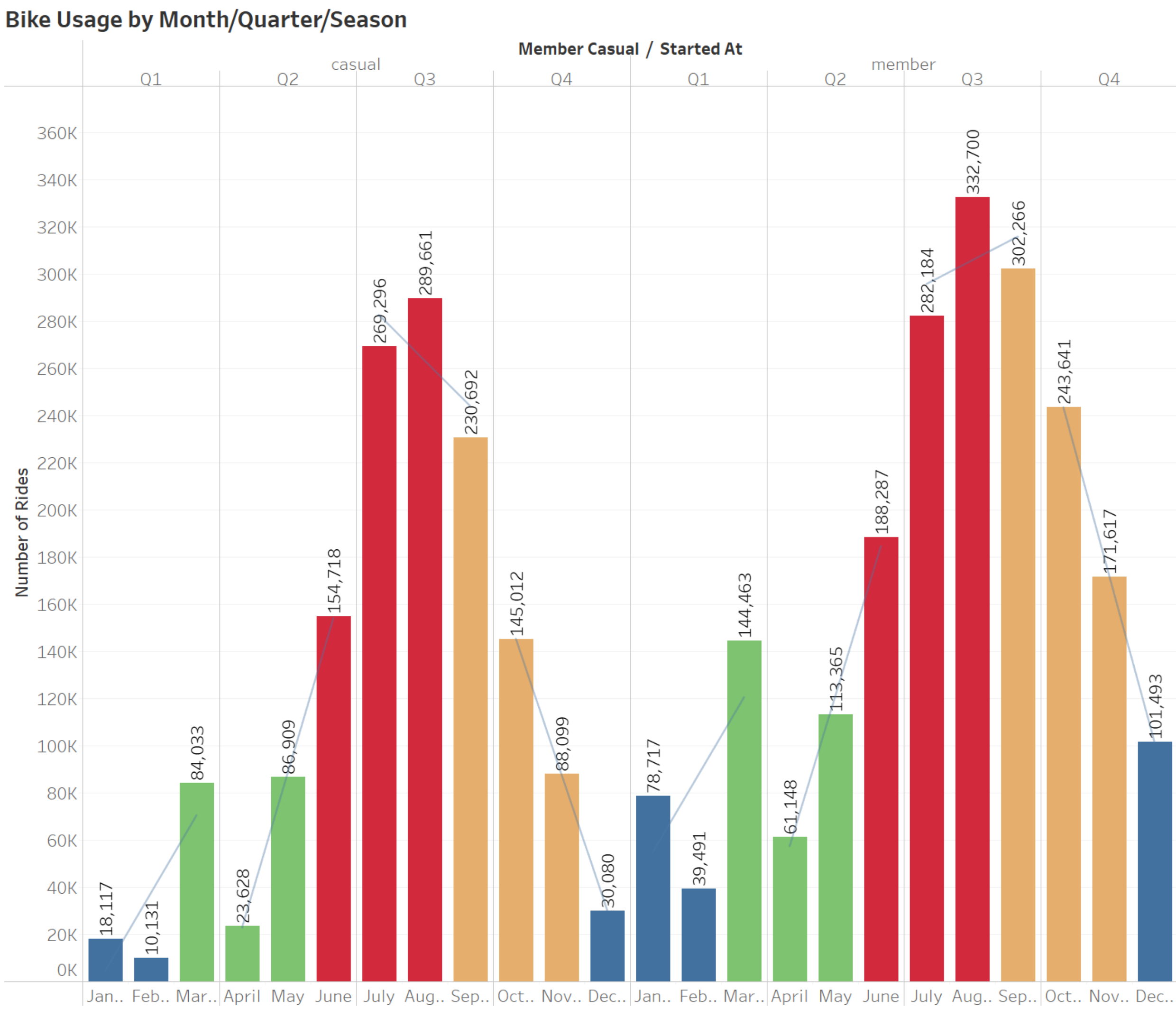
Casual riders represent 40.99% and member riders 59.01% of the total rides. The average casual ride duration is 43.29 minutes and 11.9 minutes for members. Members ride a bike a little bit more than casual, and for the average duration of every time riding, casual riders spend more time than member riders.

Bike Type

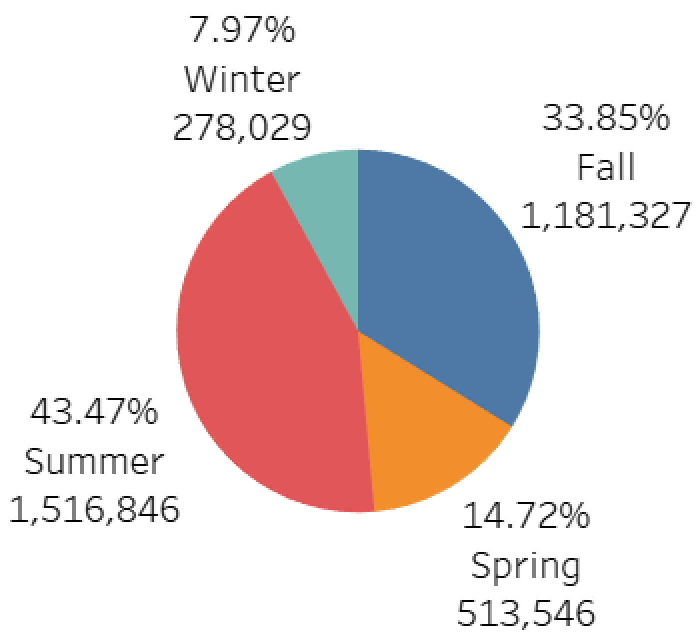


Rideable Type / Member Casual
classic_bike
docked_bike
electric_bike
1500K
1400K
1300K
1200K
1100K
1000K
900K
800K
700K
600K
500K
400K
300K
200K
100K
0K
Number of Rides
casual
member
casual
member
casual
member

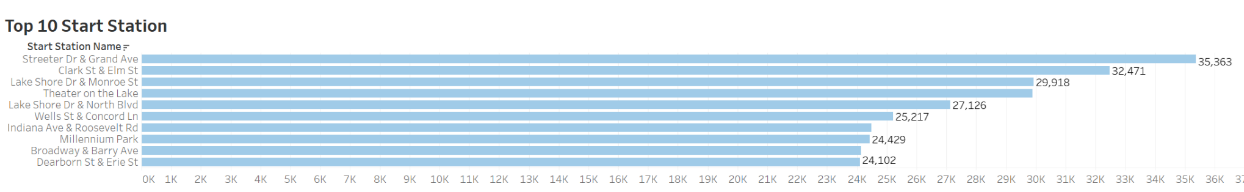
Cyclistic has three kinds of bike types for rides to choose, which is classic bike, docked bike, and electric bike. Riders like to use docked bikes the most. The classic bike is the least used.



Bike Usage by Season

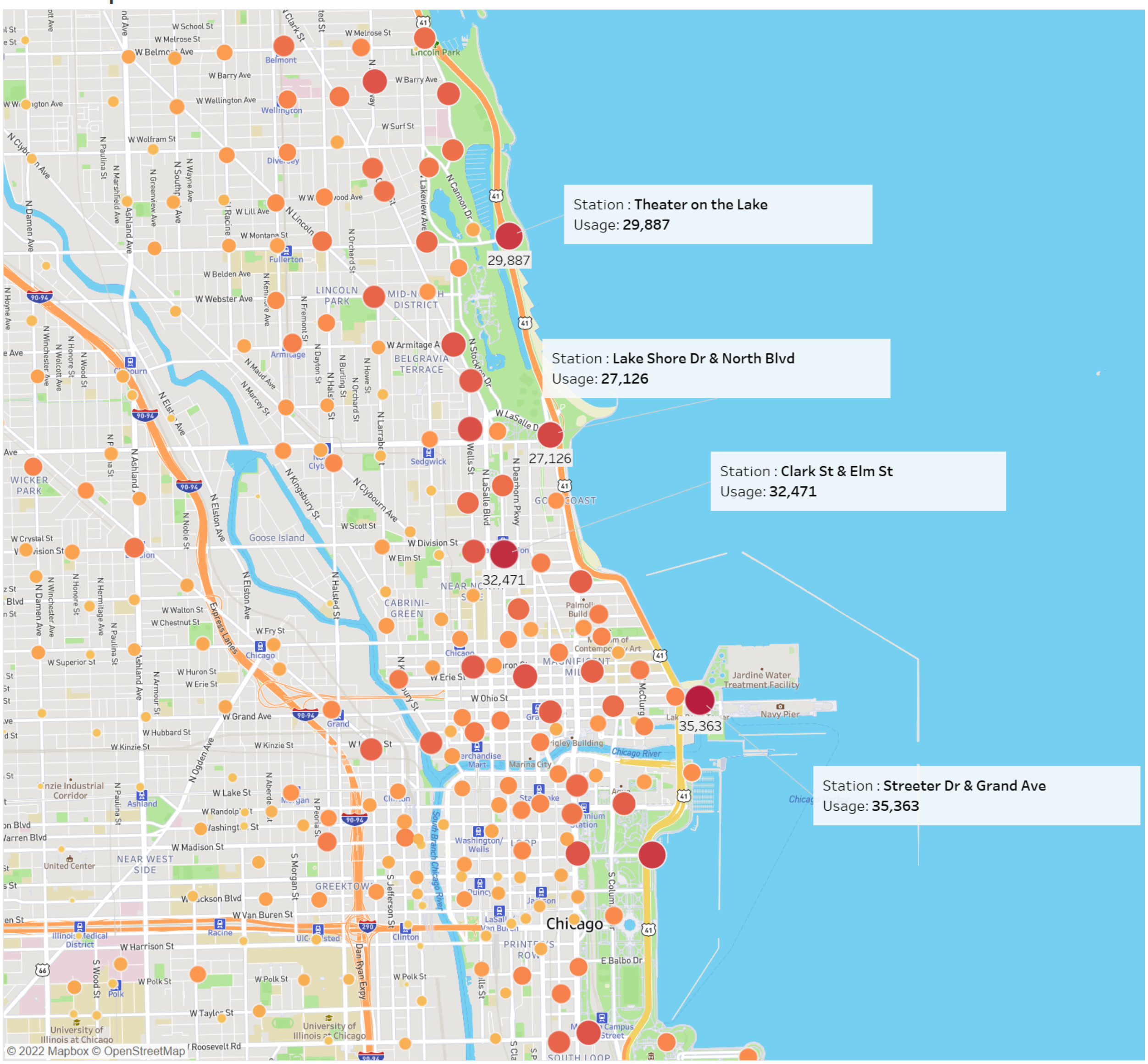


Based on bike usage by month/ quarter/ season and bike usage by season, those two charts up, August is the most bike usage month in a year, and 43.47% of riders choose to take a trip during summer. The bike usage by month/ quarter/ season chart shows that both members and casual riders have a similar trend, with more trips made in summer(Q3) and fewer trips during winter(Q1).



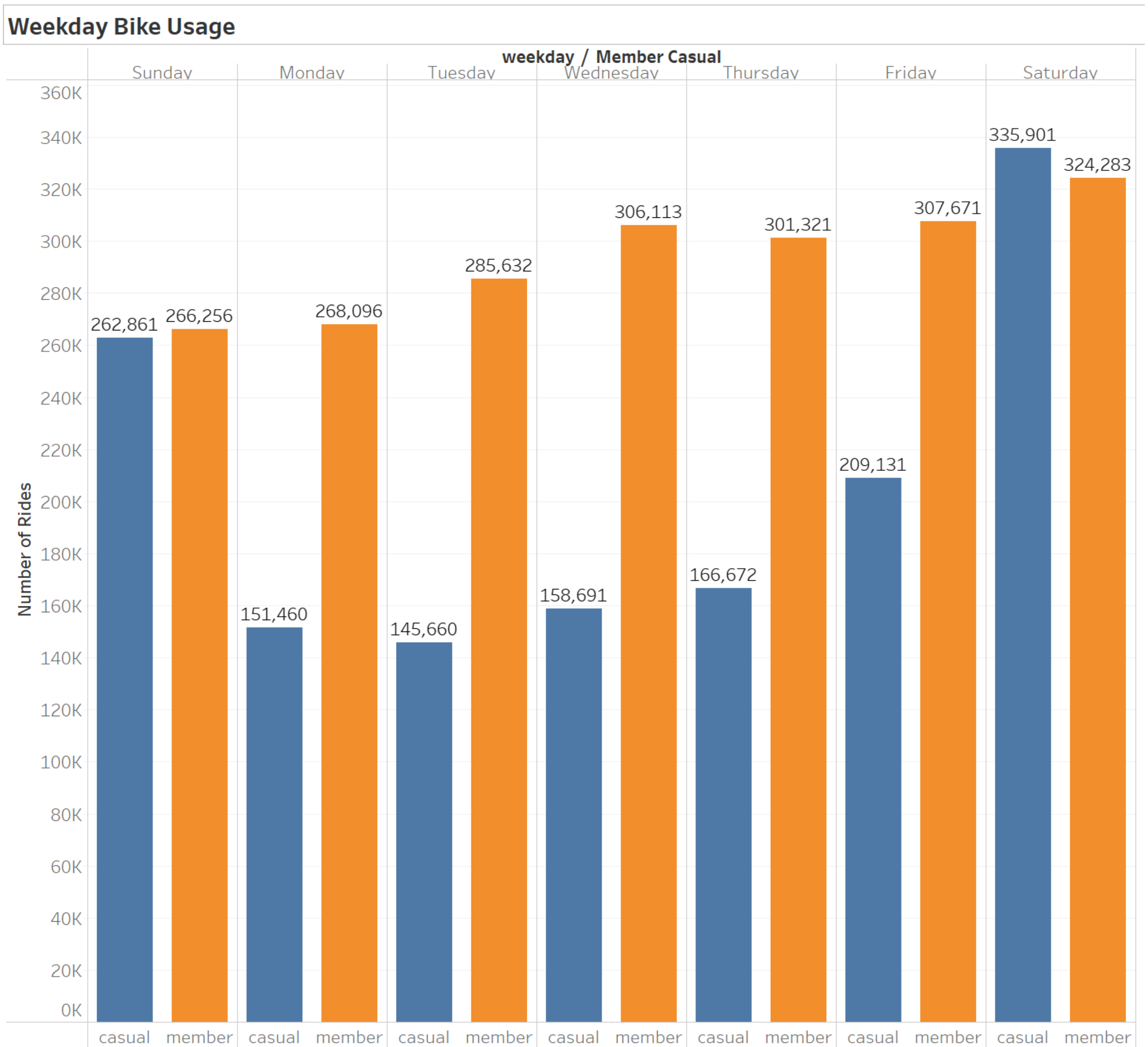
This Top 10 start station chart shows the most popular starting station in Cyclistic's network. Streeter Dr & Grand Ave, and Clark st& Elm St station are significantly more prevalent among riders.

Street Map

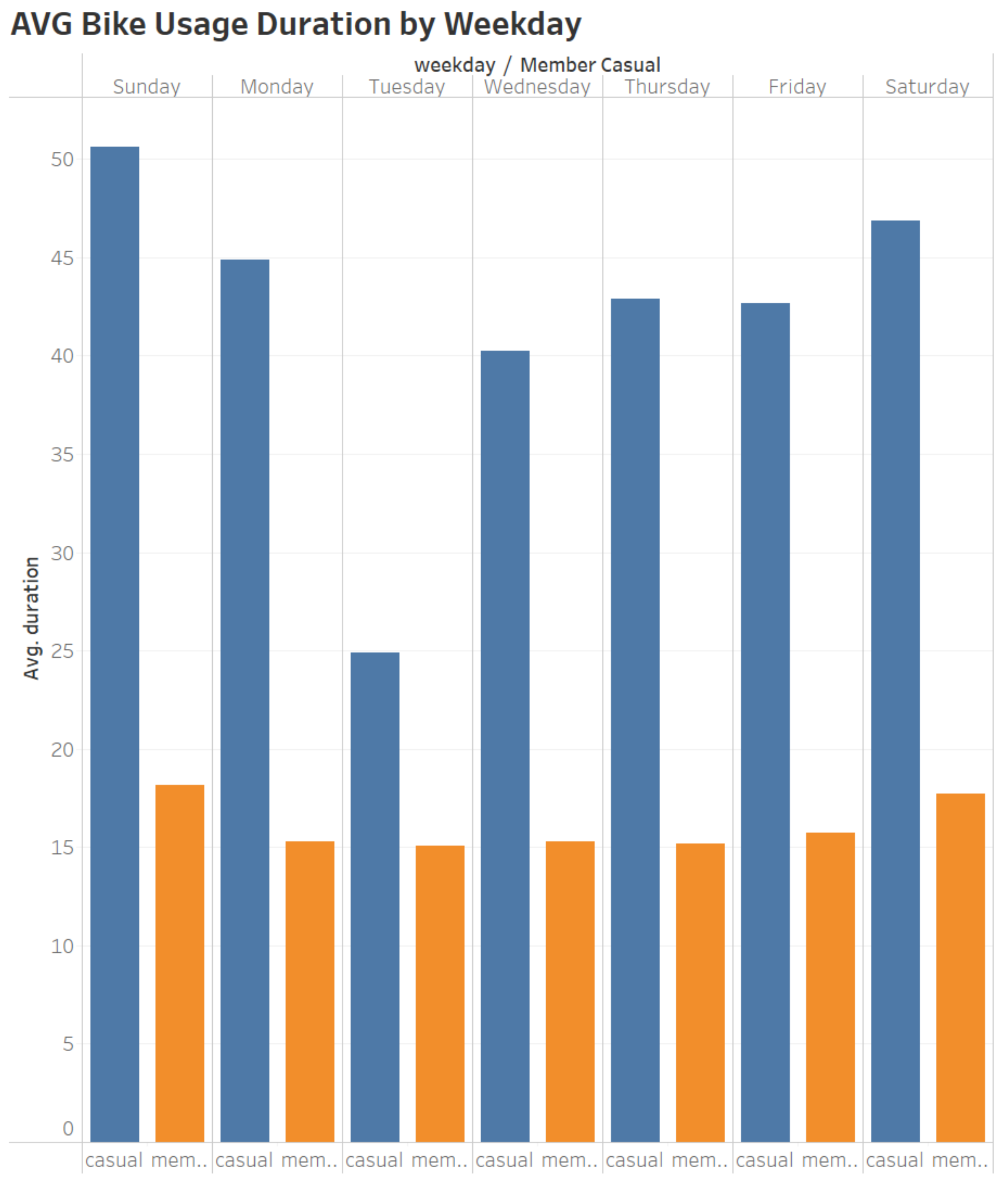


This is a street map that could clearly show the starting station on the map. The busiest station color is redder and uses light gold to represent less used stations.

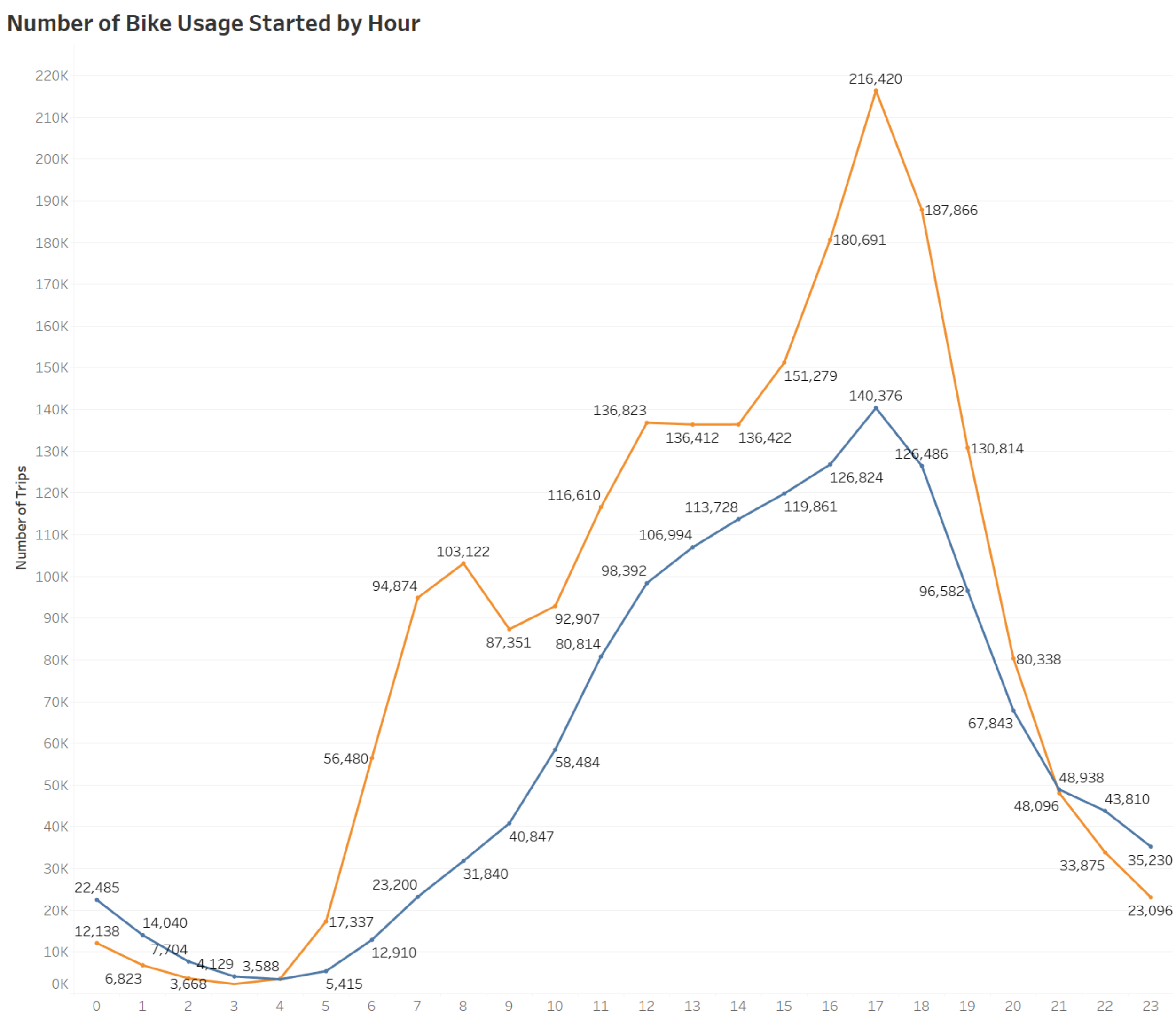
I pointed out the top 4 start stations on the map. Three of them are located in a tourist area overlooking Lake Michigan, and one is located in a shopping area.



The weekday bike usage chart shows that members' usage trend remains consistent throughout the week. However, casual riders use the bikes more during the weekend, with the number of trips on Saturday and Sunday even surpassing members'.



The average bike usage duration chat shows that members' average trip duration does not change much throughout the week. Their daily average trip duration difference is within 5 minutes. However, casual riders use bikes longer than members, and the usage peaks on weekends and is low on Tuesdays. Casuals spend more time per trip on the bike than members.



This chart shows the number of bike usage start times. Here we can see that members' usage has two peaks, the first is around eight o'clock, and the second is around 17 o'clock, corresponding with the start and end of the workday. Casual riders like to start riding bikes from noon to sunset.

Share

All charts for this case study are on my Tableau Public profile [here].

Act

How do annual members and casual riders use Cyclistic bikes differently?

1. Casual riders represent 40.99% and member riders 59.01% of the total rides.
2. The average casual ride duration is about four times that of members.
3. Casuals spend more time per travel on the bike than members.
4. Members' average trip duration does not change much throughout the week, and casuals use bikes more on weekends.
5. The peak hours for members are 8 and 17, and the casual riders ride bikes from noon to sunset.

Why would casual riders buy Cyclistic annual memberships?

1. Offer discounts for weekend rides since casual riders will most likely ride on weekends.
2. Offer discount rate for first-time members. For example,

   a). Two weeks free trial for first-time members.

   b). Pay a one-year membership fee and get 15 months of use, which is three months free.
3. We could hold a marketing campaign during summer along Lakefront Trail to reach more casuals and potential customers because more than 43% of bike riders are traveling in the summer season.

4. We could create some weekend events because riders are willing to travel on weekends more than weekdays.
5. We could provide a particular service for only members. Services might include complimentary water, ice cream, and a Chicago tour guide, or some popular Chicago hot dog restaurant or deep dish pizza coupon.

How can Cyclistic use digital media to influence casual riders to become members?

1. Put advertisements on Health, fitness, and travel apps.
2. We are sending promotions through email or text messages.
3. We are holding a “refer a friend” promotion on apps. Such as sharing a link or code with a friend. When they use it for their first bike ride, they get $10 off, and so do you.
4. Before holding the Cyslistic events, we post those events on Twitter, Reddit, Instagram, Facebook, and TikTok, and after the event, encourage them to #Cyclistic on social media.