Sources: use data about Cyclistic's historical trip data of 9months - 9 tables from Jan 2021 to Sep 2021- with total rows are about 4.3 million.
The data has been made available by Motivate International Inc. under this license.
Data source: https://divvy-tripdata.s3.amazonaws.com/index.html

Key tasks

1. Identify the business task

- 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?

2. Key stakeholders

- Primary stakeholder:Cyclistic executive team
- Secondary stakeholder: The director of marketing, Cyclistic marketing analytics team

Upload Data to BigQuery

Clean Data

- **Check data's structure**

found that many null values of column end_station_name. After using the count function, it is about ⅛ is null data so keep using.

Calculate member of each trip:

```
Select count(*),member_casual
FROM `bike-share-case.Trip.2021*`
GROUP BY member_casual
```

Result of first 09 months in 2021

Row	f0_	member_casual
1	2261223	member
2	2095096	casual

Calculate min, max, avg of trip duration of member and Casual

```
with trip_time as (Select member_casual,
 TIMESTAMP_DIFF(ended_at, started_at,MINUTE) as trip_duration
FROM `bike-share-case.Trip.2021*`)
## calculate average trip duration
SELECT member_casual, AVG(trip_duration)as Average_trip_duration,
MAX(trip_duration)as Max_duration,Min(trip_duration)as Min_duration
FROM trip_time
Where trip_duration >=0
## to avoid value <0 for min_duration
```

```
GROUP BY member_casual
ORDER By Average_trip_duration
```

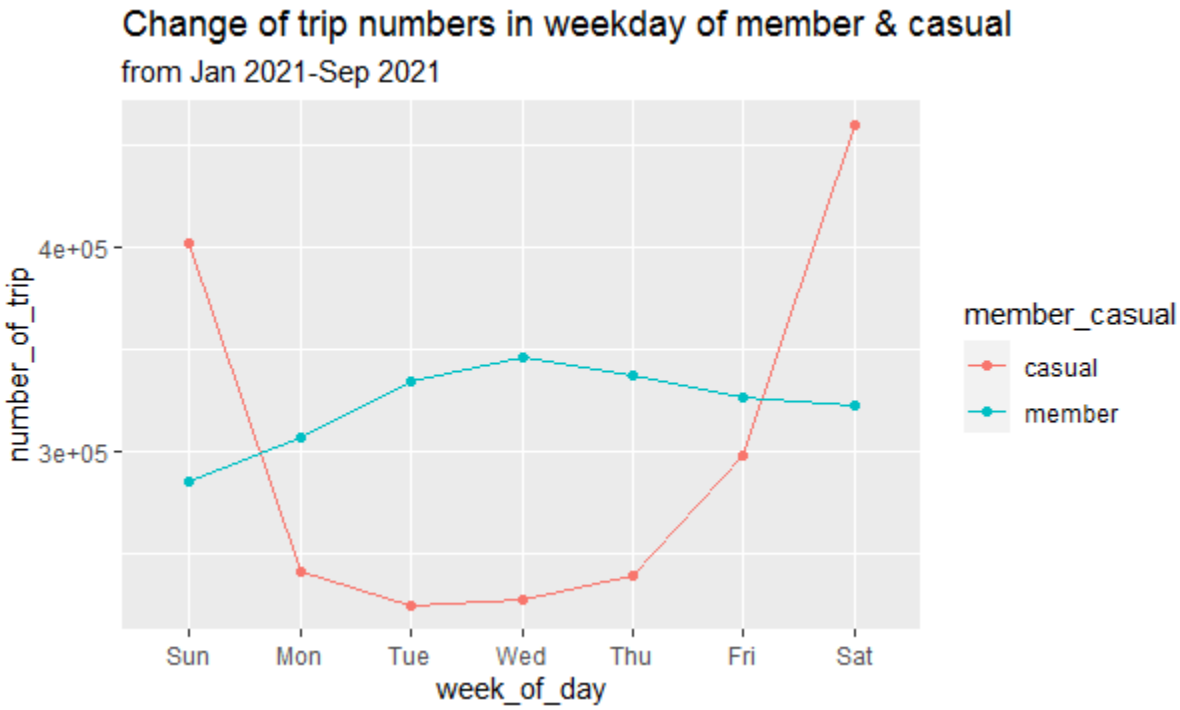
Row	member_casual	Average_trip_duration	Max_duration	Min_duration
1	member	14.029300764737199	1559	1
2	casual	33.10126283420182	55944	1

 # which day in a week have the highest number of trips

```
Select Count(ride_id)as count_weekday, member_casual, weekday
FROM
(Select Extract(DAYOFWEEK FROM started_at)as weekday, member_casual,ride_id
FROM `bike-share-case.Trip.2021*`)
# returns values in the range [1,7] with Sunday as the first day of the week
GROUP BY member_casual, weekday
ORDER BY member_casual, count_weekday DESC
```

Row	count_weekday	member_casual	weekday
1	459606	casual	7
2	402618	casual	1
3	298597	casual	6
4	241792	casual	2
5	239957	casual	5
6	227594	casual	4
7	224932	casual	3
8	346068	member	4
9	337774	member	5
10	334642	member	3
11	327037	member	6
12	322643	member	7

13	307008	member	2	
14	286051	member	1	

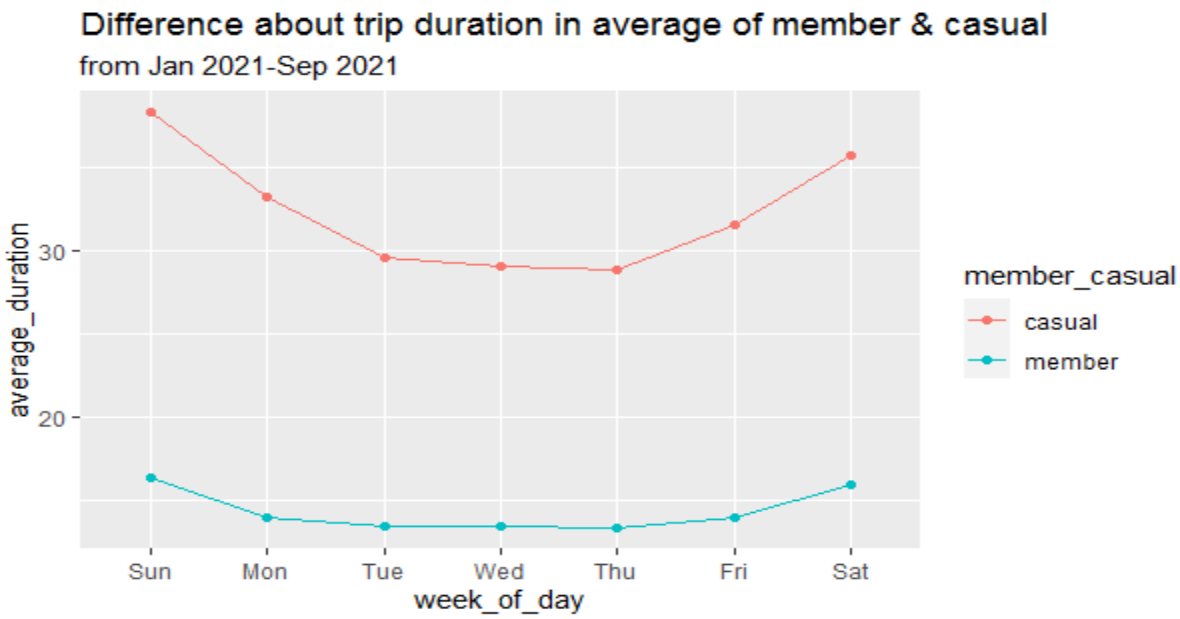


 # average trips duration of member and casual for each weekday

```
with trip_time as (Select member_casual,
 TIMESTAMP_DIFF(ended_at, started_at,MINUTE) as trip_duration,
 extract (dayofweek from started_at)as weekday,
FROM `bike-share-case.Trip.2021*`)
## calculate average trip duration,
SELECT member_casual,weekday, Count(*), AVG(trip_duration)as Average_trip_duration,
MAX(trip_duration)as Max_duration,Min(trip_duration)as Min_duration
FROM `trip_time`
Where trip_duration >=0
## to avoid value <0 for min_duration
GROUP BY member_casual,weekday
ORDER By member_casual,weekday
```

Row	member_casual	weekday	f0_	Average_trip_duration	Max_duration	Min_duration
1	casual	1	402618	37.754347793690215	53921	0

2	casual	2	241791	32.750205756210995	31681	0	
3	casual	3	224930	29.062655048237172	38922	0	
4	casual	4	227594	28.555884601527282	38963	0	
5	casual	5	239957	28.31736102718406	49107	0	
6	casual	6	298596	31.04133009149479	55691	0	
7	casual	7	459606	35.2453971445106	55944	0	
8	member	1	286051	15.846272168249733	1499	0	
9	member	2	307008	13.467349385032323	1499	0	
10	member	3	334639	12.91887974802701	1499	0	
11	member	4	346054	12.953926265842862	1499	0	
12	member	5	337774	12.866206398361	1499	0	
13	member	6	327036	13.451623062904357	1499	0	
14	member	7	322642	15.43939102782651	1559	0	



found that city have reluctant space >> TRim (city), ID station of 2 tables is not the same structure,station table was only updated upto 2017, but since total trips by city is near 3.4 million comparing to 4.3 million, it still has value to discover (80%)

```
SELECT Count(*), TRIM (station_2017.city)as New_city,trip_2021.member_casual
From `bike-share-case.Trip.2021*` as trip_2021
```

```
inner join  `bike-share-case.Station.2017_Q1Q2`as station_2017
On Trip_2021.start_station_name =station_2017.name
Group By New_city, trip_2021.member_casual
```

Row	f0_	New_city	member_casual
1	1761640	Chicago	member
2	1607169	Chicago	casual
3	10667	Evanston	casual
4	11401	Evanston	member

Response business task:

- How do annual members and casual riders use Cyclistic bikes differently?
  - + Annual members's trip duration is shorter about 2 times comparing to Casual riders
  - + Annual members intend to have more trips during weekdays than weekend, but the difference is small, within10% only
  - + Casual members is more likely use bike at weekend that seen a double of trip higher than it in week days
- Why would casual riders buy Cyclistic annual memberships?
  - + Additional databases are required like the purposes they use bike, how to encourage riders to use bike more frequently even during weekdays, updated station list...
- How can Cyclistic use digital media to influence casual riders to become members?
  - + Can focus on Chicago where most of their casual riders stay at
  - + Use more advertising during weekend