SQL analysis

Data analyst: Abraham Cedeño

Date: 9th October 2022

In this document, I will use BIGQUERY to answer the S.M.A.R.T questions I created when I started the data analysis.

**Data category 1 - Bike preferences**

**SMART Question one:** “What was the percentual distribution of bike preferences among casual and member riders in the last 12 months?”

**Answer:** I performed the following queries to answer my question, then I made a visualization to better share my findings.

SELECT COUNT(rideable\_type) AS electric\_bike\_preference\_among\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'casual' AND rideable\_type = 'electric\_bike'

Graphical user interface, text, application, chat or text message

Description automatically generated

SELECT COUNT(rideable\_type) AS classic\_bike\_preference\_among\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'casual' AND rideable\_type = 'classic\_bike'

Graphical user interface, text, application

Description automatically generated

SELECT COUNT(rideable\_type) AS docked\_bike\_preference\_among\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'casual' AND rideable\_type = 'docked\_bike'

Graphical user interface, text, application

Description automatically generated

SELECT COUNT(rideable\_type) AS electric\_bike\_preference\_among\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'member' AND rideable\_type = 'electric\_bike'

Graphical user interface, text, application, Teams

Description automatically generated

SELECT COUNT(rideable\_type) AS classic\_bike\_preference\_among\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'member' AND rideable\_type = 'classic\_bike'

Graphical user interface, application, Teams

Description automatically generated

SELECT COUNT(rideable\_type) AS docked\_bike\_preference\_among\_member\_riders

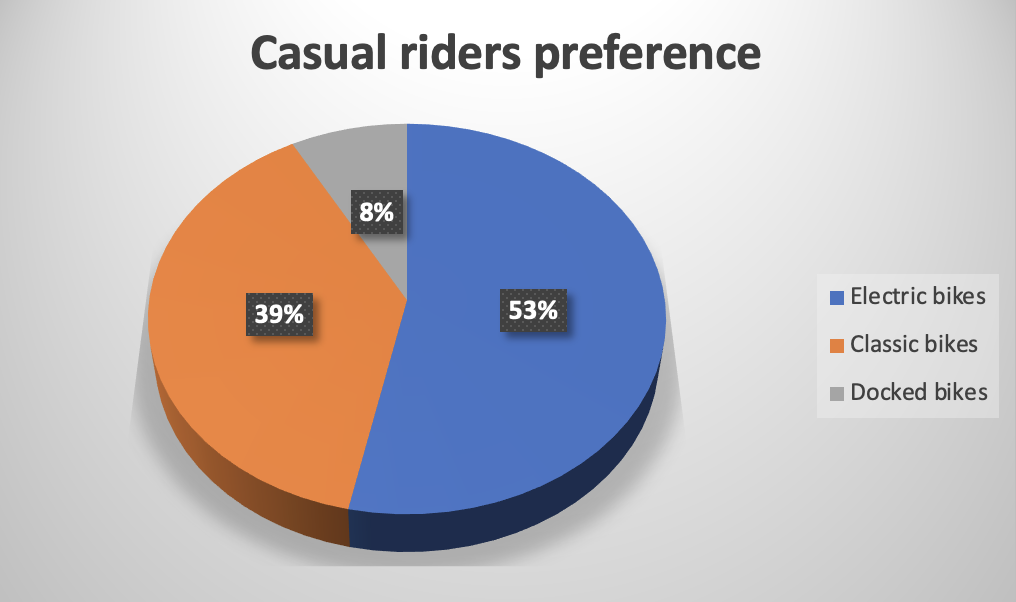
FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'member' AND rideable\_type = 'docked\_bike'

Graphical user interface, application

Description automatically generated

**Visualizations**



Chart, pie chart

Description automatically generated

Chart, pie chart

Description automatically generated

**Data category 2 - Trips start day and time**

**SMART Question two:** “On which days of the week have our members used more often the services in the last 12 months?”

**Answer:** I performed the following queries to answer my question, then I made a visualization to better share my findings.

SELECT COUNT (day\_of\_week) AS Trips\_per\_day\_casual\_riders, day\_of\_week

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

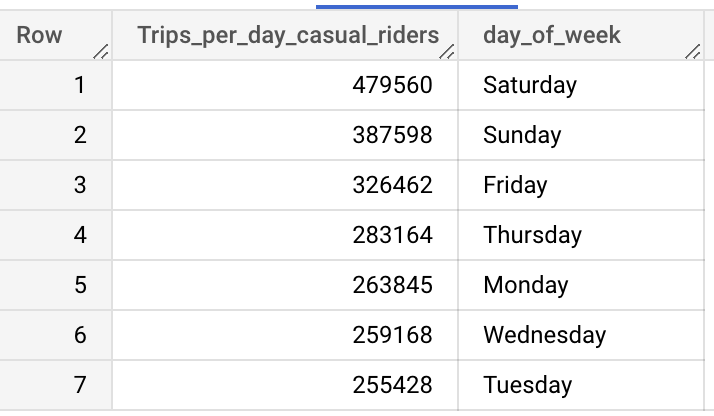
WHERE member\_casual = 'casual'

GROUP BY

day\_of\_week

Order by

Trips\_per\_day\_casual\_riders DESC



SELECT COUNT (day\_of\_week) AS Trips\_per\_day\_member\_riders, day\_of\_week

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = "member"

GROUP BY

day\_of\_week

Order by

Trips\_per\_day\_member\_riders DESC

Table

Description automatically generated

SELECT COUNT (day\_of\_week) AS Trips\_per\_day\_both\_riders, day\_of\_week

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

GROUP BY

day\_of\_week

Order by

Trips\_per\_day\_both\_riders DESC

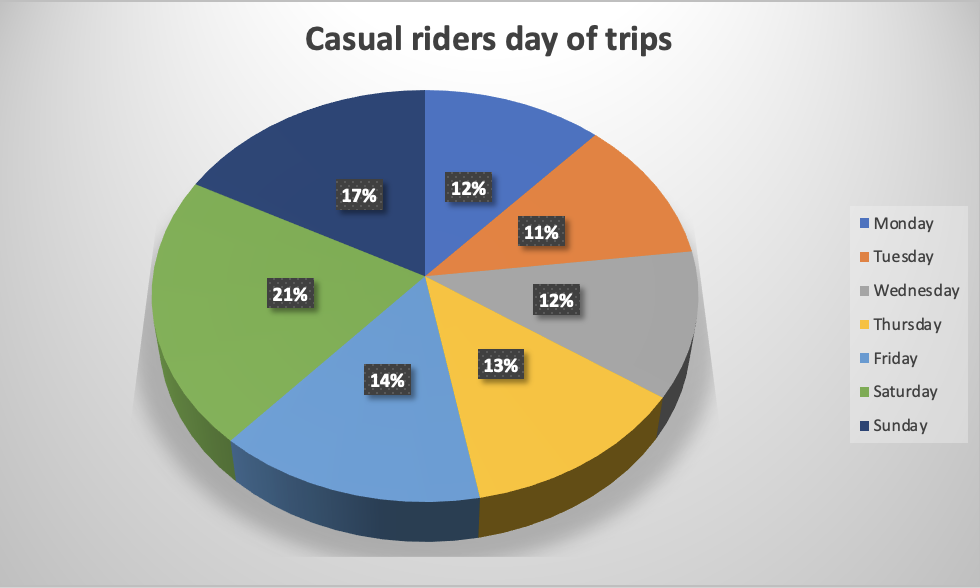
Table

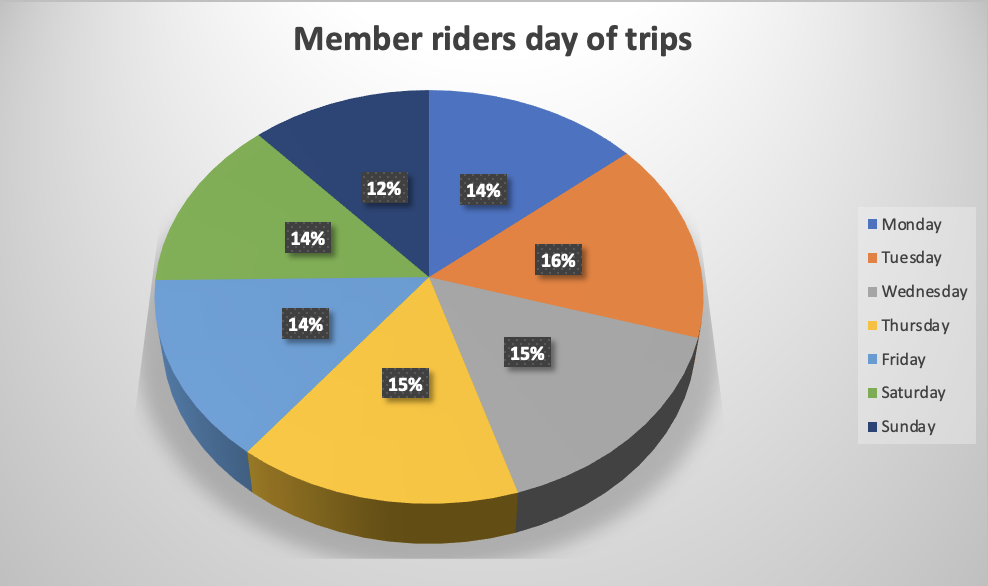
Description automatically generated

**Visualizations**

Chart, bar chart

Description automatically generated





Chart, pie chart

Description automatically generated

**SMART Question three:** “At which times of the day have our members used more often the services in the last 12 months?”

**Answer:** Although with SQL I can count the trips at any time, unfortunately, because I’m using a free account, I cannot export that much data from BIGQUERY, therefore I wouldn’t be able to visualize the data and make some findings. I will try to answer this question with R later.

**Data category 3 - Trips end day and time**

**SMART Question four:** “What was the average trip duration of our members in the last 12 months?

**SMART Question five:** “What was the average trip duration of our casual riders in the last 12 months?”

**Answer:** To answer both questions I used the following queries. Then I made a visualization to better share my findings.

SELECT

TIME\_ADD(TIME '00:00:00', INTERVAL CAST(AVG(TIME\_DIFF(ride\_length, TIME '00:00:00', SECOND)) AS INT64) SECOND) average\_trip\_duration\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'casual'

Graphical user interface, text, application

Description automatically generated

SELECT

TIME\_ADD(TIME '00:00:00', INTERVAL CAST(AVG(TIME\_DIFF(ride\_length, TIME '00:00:00', SECOND)) AS INT64) SECOND) average\_trip\_duration\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

WHERE member\_casual = 'member'

Graphical user interface, text, application, chat or text message

Description automatically generated

SELECT

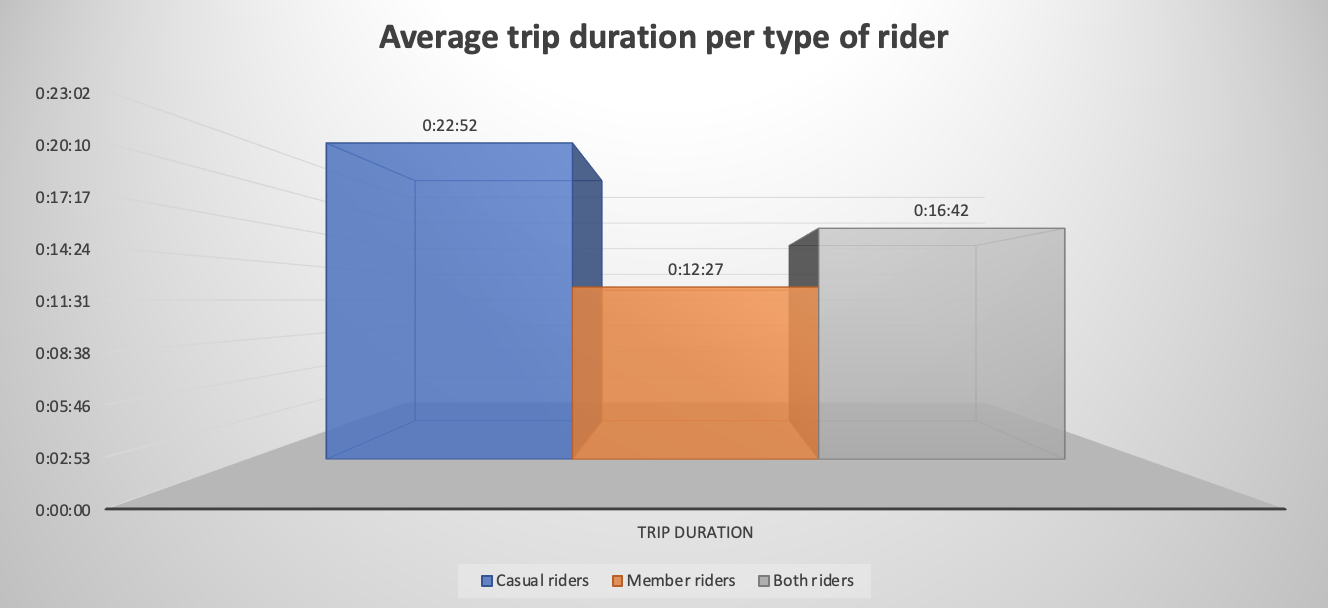
TIME\_ADD(TIME '00:00:00', INTERVAL CAST(AVG(TIME\_DIFF(ride\_length, TIME '00:00:00', SECOND)) AS INT64) SECOND) average\_trip\_duration\_both\_riders

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

Graphical user interface, text, application, Word

Description automatically generated

**Visualizations**



**Data category 4 – Start stations**

**SMART Question six:** “On which stations did the member riders start their rides more often in the last 12 months?

**Answer:** When I started to do some queries I found out that data need more cleaning because I have two start stations named the same but one with a capital letter.

Table

Description automatically generated

To clean the data I used the next query.

SELECT \* except(start\_station\_name),

CASE

WHEN start\_station\_name = 'Start\_station\_none\_given' THEN 'start\_station\_none\_given'

ELSE start\_station\_name

END AS cleaned\_start\_station\_name

FROM `capstone-1-abraham.Trip\_data\_v2.BikeTrips\_todos`

An then I saved this new table in a new dataset called Trip\_data\_v3

Graphical user interface, text, application

Description automatically generated

Then I did the following query to find the top 11 started stations of member and casual rider trips. After that, I make some visualizations.

SELECT COUNT (cleaned\_start\_station\_name) AS Number\_of\_trips, cleaned\_start\_station\_name AS start\_station\_of\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v1`

WHERE member\_casual = 'casual'

GROUP BY

cleaned\_start\_station\_name

Order by

COUNT (cleaned\_start\_station\_name) DESC

LIMIT 11

Table

Description automatically generated

SELECT COUNT (cleaned\_start\_station\_name) AS Number\_of\_trips, cleaned\_start\_station\_name AS start\_station\_of\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v1`

WHERE member\_casual = 'member'

GROUP BY

cleaned\_start\_station\_name

Order by

COUNT (cleaned\_start\_station\_name) DESC

LIMIT 11

Table

Description automatically generated

SELECT COUNT (cleaned\_start\_station\_name) AS Number\_of\_trips, cleaned\_start\_station\_name AS start\_station\_of\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v1`

GROUP BY

cleaned\_start\_station\_name

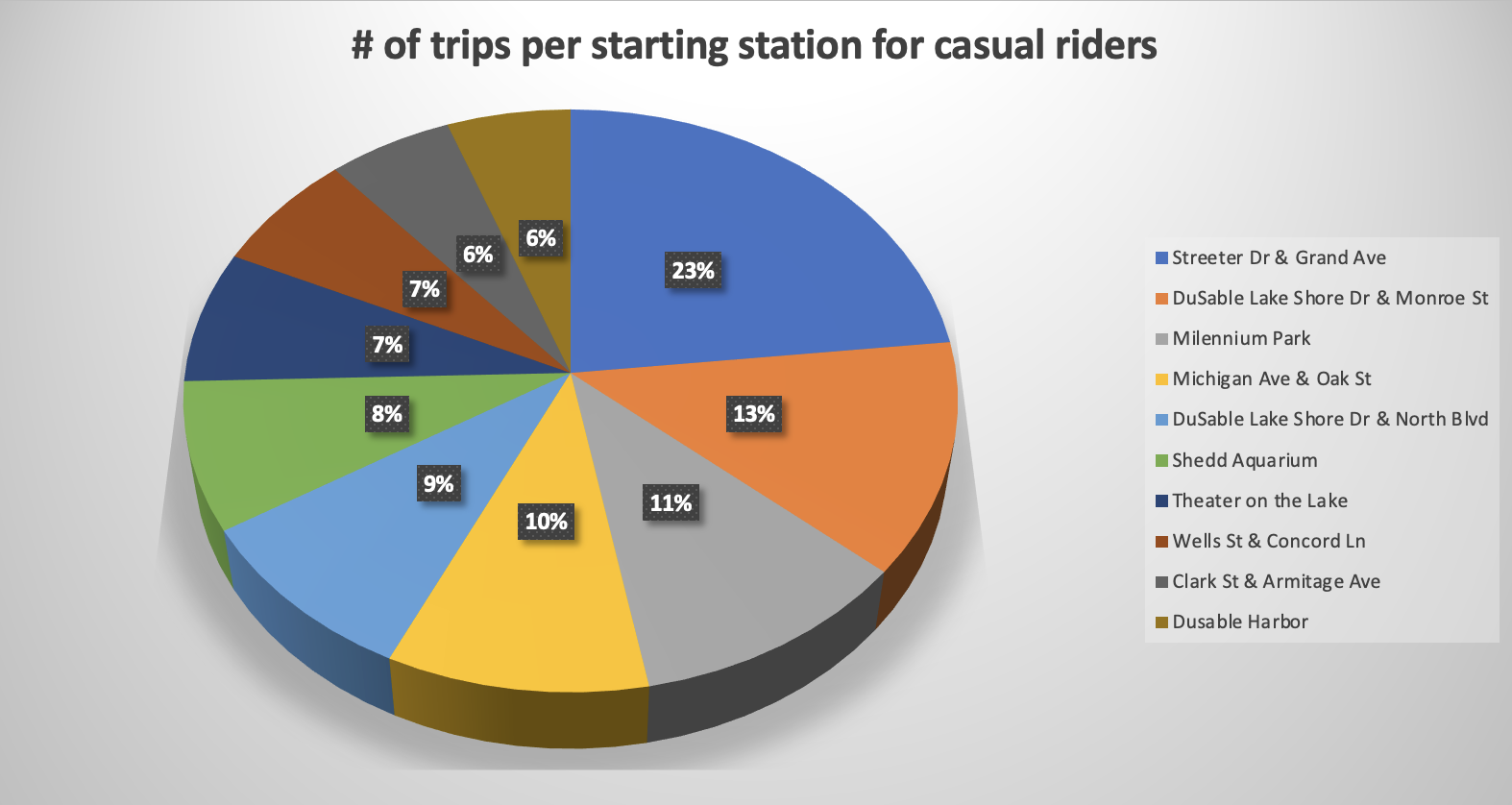
Order by

COUNT (cleaned\_start\_station\_name) DESC

LIMIT 11

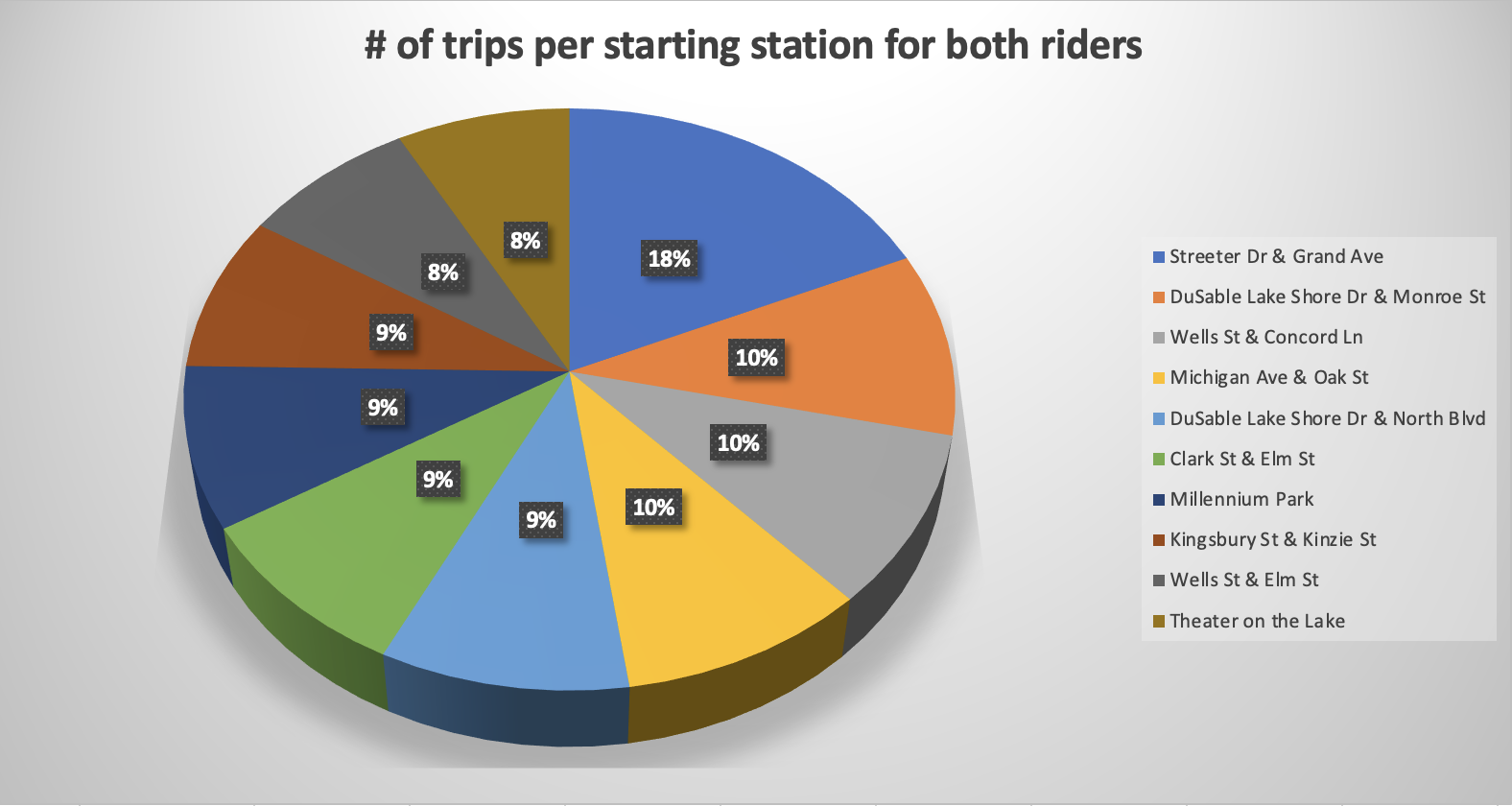


**Visualizations**



Chart, pie chart

Description automatically generated



**Data category 5 – End stations**

**SMART Question seven:** “On which stations did the member riders finish their rides more often in the last 12 months?

**Answer:** I did the following query to find the top 11 end stations of member and casual rider trips. After that, I make some visualizations.

SELECT COUNT (cleaned\_end\_station\_name) AS Number\_of\_trips, cleaned\_end\_station\_name AS end\_station\_of\_casual\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v2`

WHERE member\_casual = 'casual'

GROUP BY

cleaned\_end\_station\_name

Order by

COUNT (cleaned\_end\_station\_name) DESC

LIMIT 11

Table

Description automatically generated

SELECT COUNT (cleaned\_end\_station\_name) AS Number\_of\_trips, cleaned\_end\_station\_name AS end\_station\_of\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v2`

WHERE member\_casual = 'member'

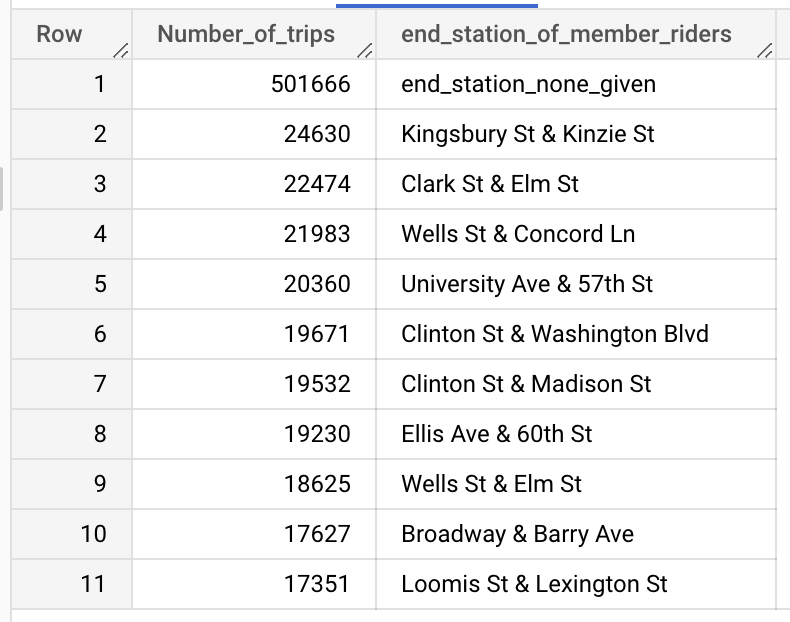
GROUP BY

cleaned\_end\_station\_name

Order by

COUNT (cleaned\_end\_station\_name) DESC

LIMIT 11



SELECT COUNT (cleaned\_end\_station\_name) AS Number\_of\_trips, cleaned\_end\_station\_name AS end\_station\_of\_member\_riders

FROM `capstone-1-abraham.Trip\_data\_v3.BikeTrips\_todos\_v2`

GROUP BY

cleaned\_end\_station\_name

Order by

COUNT (cleaned\_end\_station\_name) DESC

LIMIT 11

Table

Description automatically generated

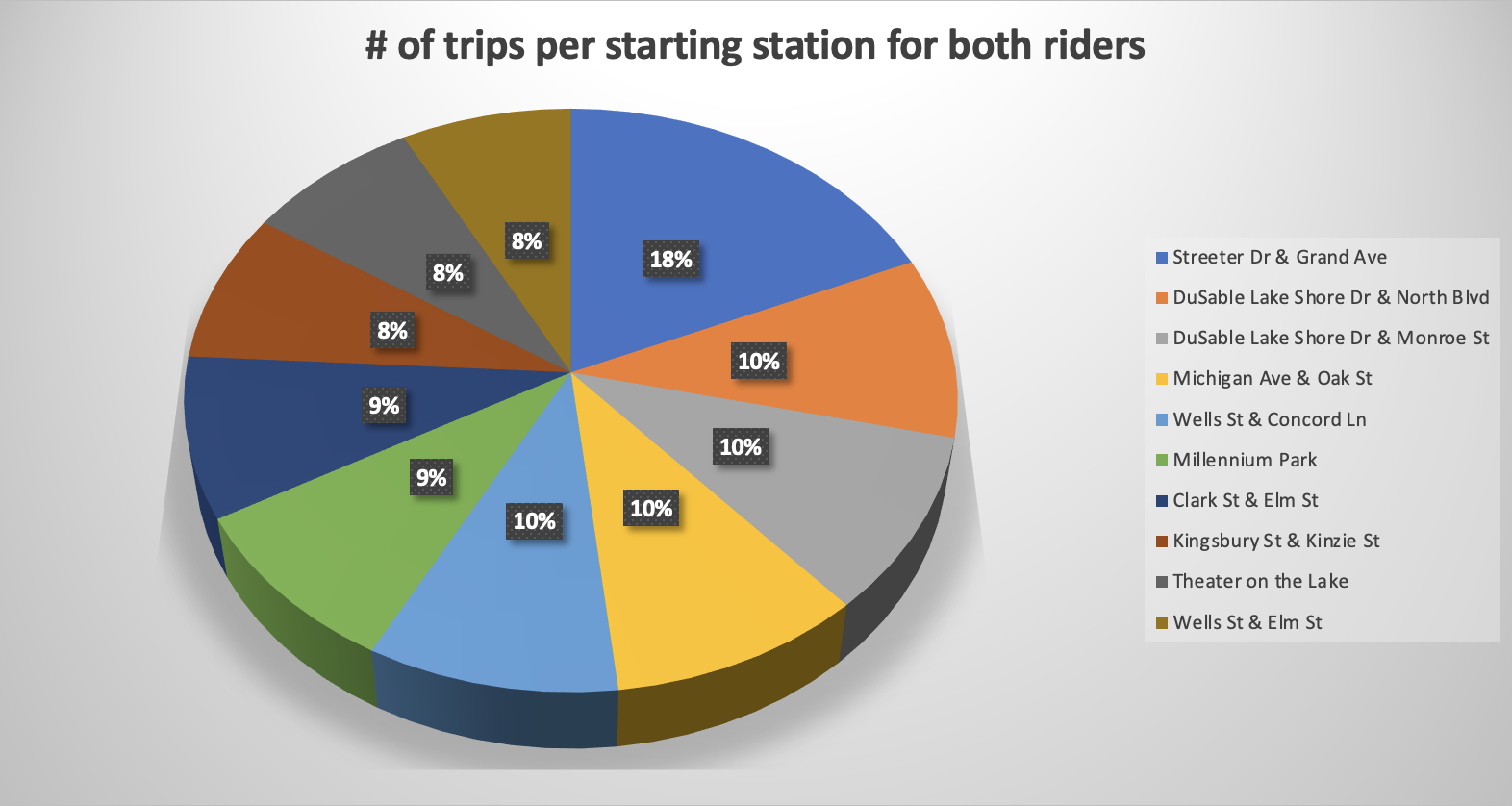
**Visualizations**

Chart, pie chart

Description automatically generated

Chart, pie chart

Description automatically generated



**Data category 6 – Start latitude and longitude**

**SMART Question eight:** “What was the distribution of rides starting point per longitude for both casual and member riders in the last 12 months?”

**SMART Question nine:** “What was the distribution of rides starting point per latitude for both casual and member riders in the last 12 months?”

**Answer:** Although with SQL I can count the trips at any time, unfortunately, because I’m using a free account, I cannot export that much data from BIGQUERY, therefore I wouldn’t be able to visualize the data and make some findings. I will try to answer this question with R later.

**Data category 7 – End latitude and longitude**

**SMART Question ten:** “What was the distribution of rides ending point per longitude for both casual and member riders in the last 12 months?”

**SMART Question eleven: “**What was the distribution of rides ending point per latitude for both casual and member riders in the last 12 months?”

**Answer:** Although with SQL I can count the trips at any time, unfortunately, because I’m using a free account, I cannot export that much data from BIGQUERY, therefore I wouldn’t be able to visualize the data and make some findings. I will try to answer this question with R later.