FINAL PROJECT

NYC CITI BIKE

Northeastern University

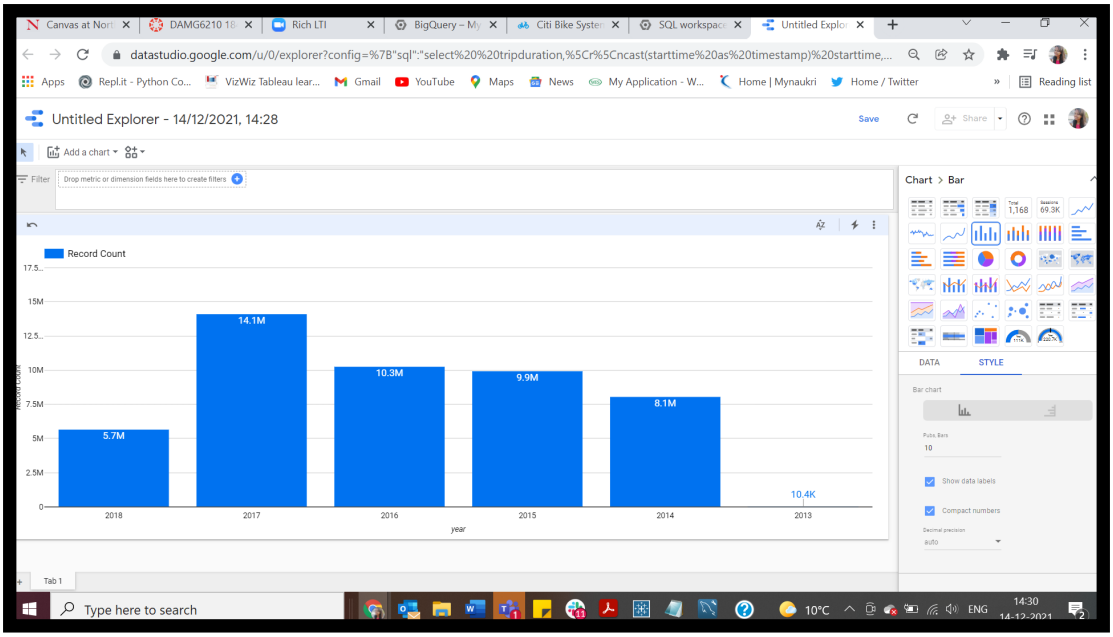
Data Management and Data Design

Semester-I (Sept 2021-Dec 2021)

Target: Analysis of Business Requirement Questions on public NYC Citi Bike Dataset

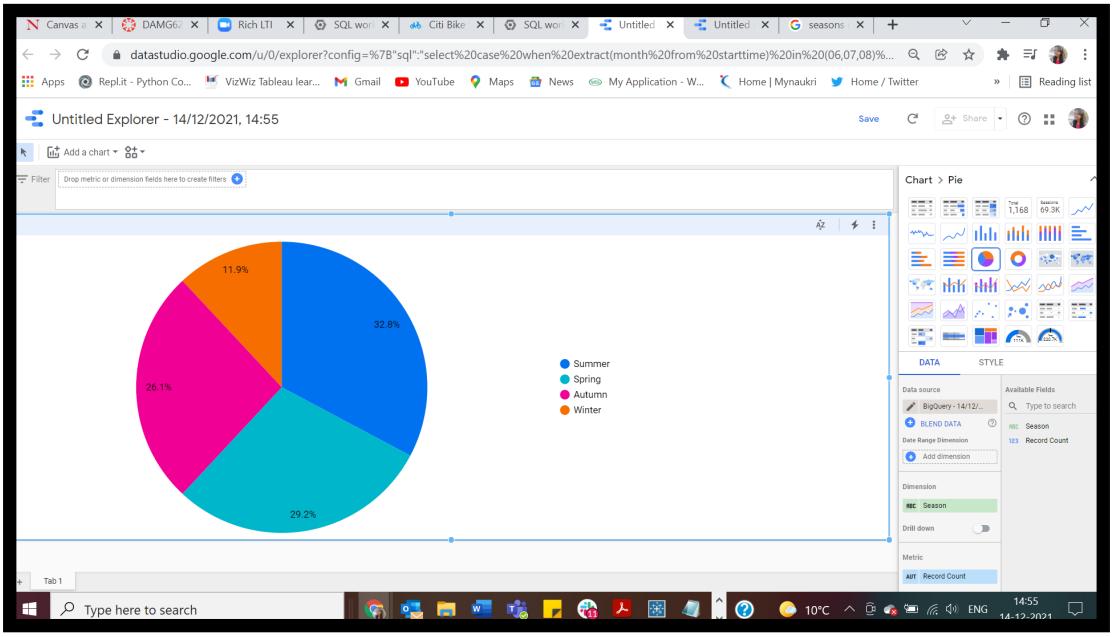
1. Trend year over year

```
select tripduration,
cast(starttime as timestamp) starttime,
cast(stoptime as timestamp) stoptime,
extract(year from starttime) as year,
start_station_id,start_station_name, start_station_latitude,start_station_lo
ngitude,
end_station_id, end_station_name, end_station_latitude,end_station_longitude
,
bikeid, usertype, birth_year, gender, customer_plan from `bigquery-public-
data.new_york_citibike.citibike_trips`
where starttime > '2013-12-31'
order by starttime desc
```



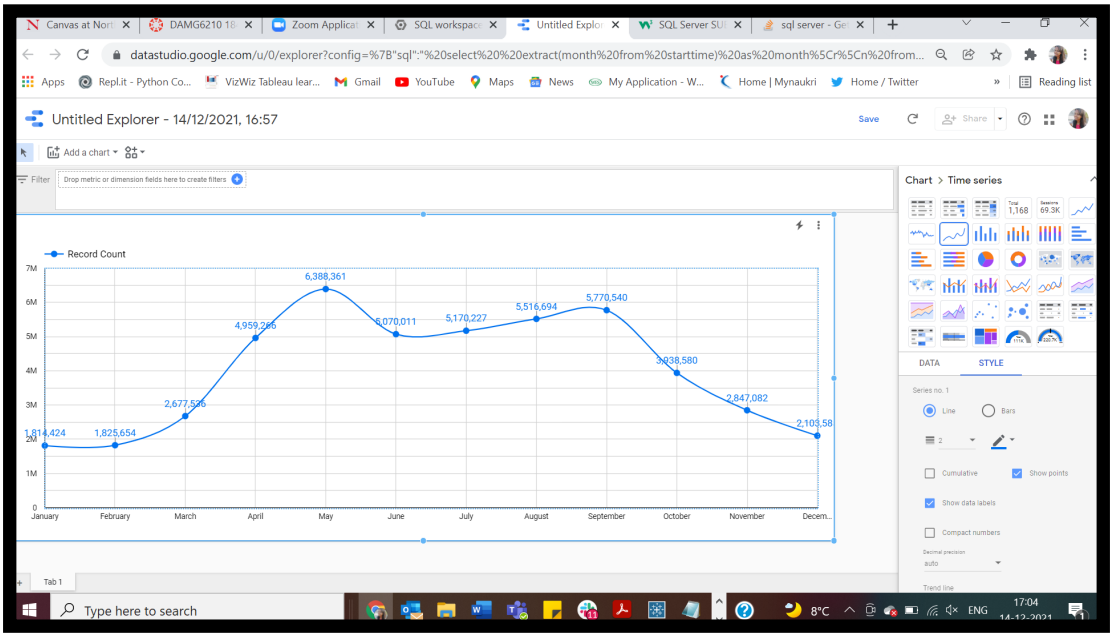
2. Seasonality

```
select case when extract(month from starttime) in (06,07,08) then 'Summer'
when extract(month from starttime) in (09,10,11) then 'Autumn'
when extract (month from starttime) in (12,01,02) then 'Winter'
else 'Spring' end as Season from `healthy-
genre-330203.damg_datset.citibike_trips`
```



3. Month-wise trend

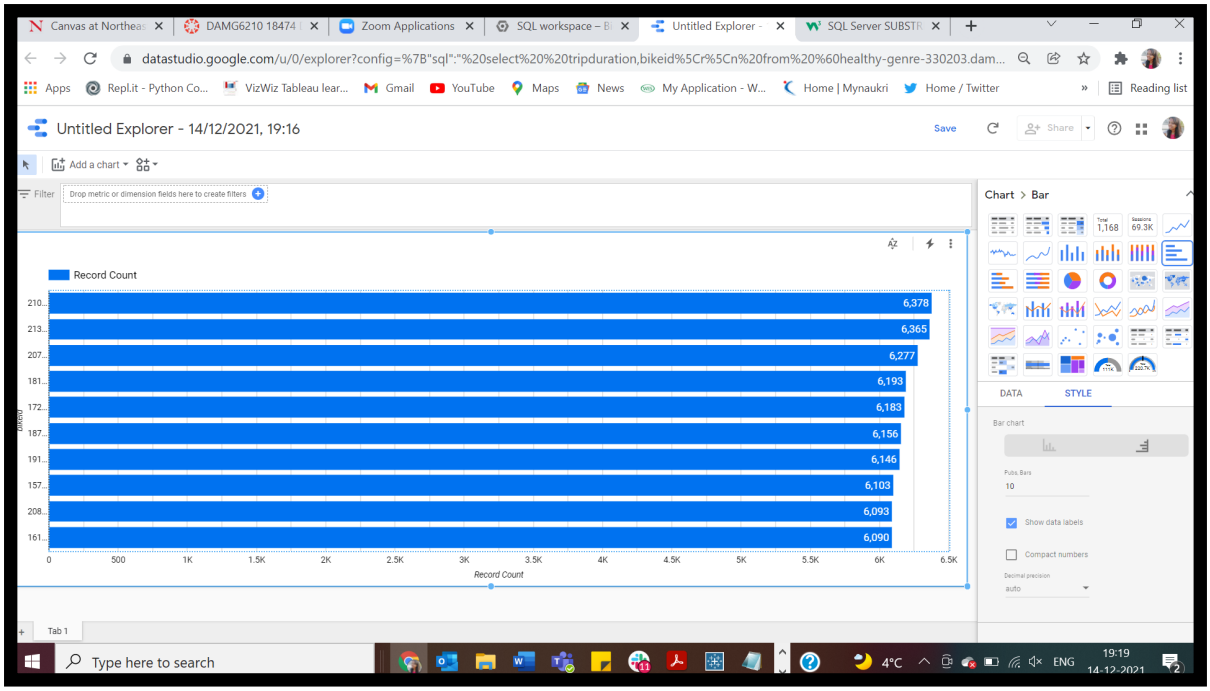
```
select  extract(month from starttime) as month
from `healthy-genre-330203.damg_datset.citibike_trips`
```



Bike Analysis:

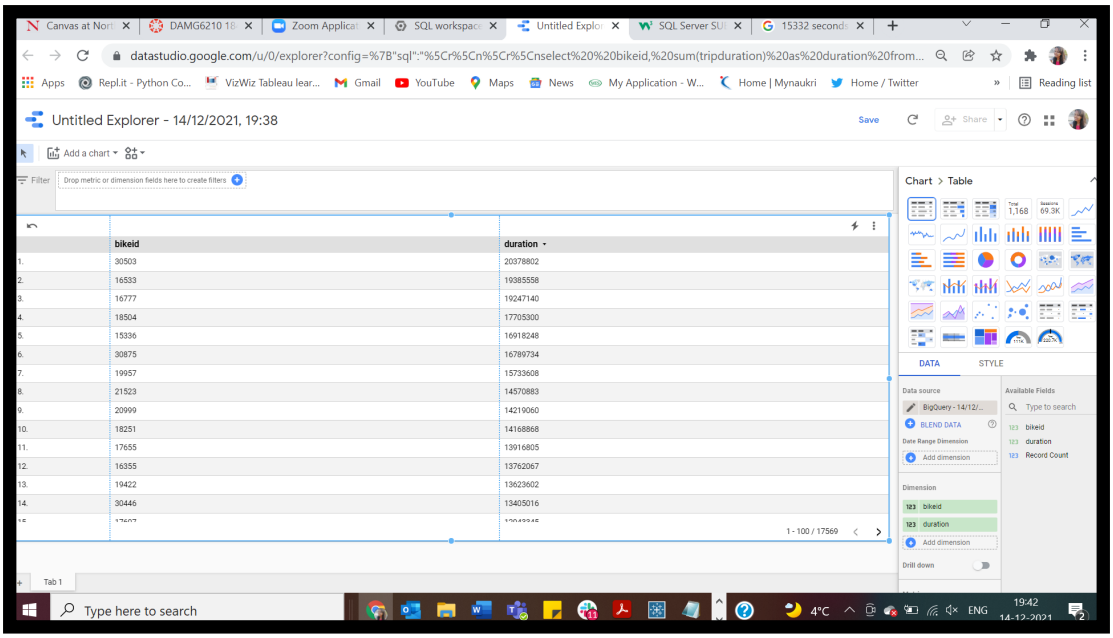
1. Top 5 bike ID by trips

```
select  tripduration,bikeid
from `healthy-genre-330203.damg_datset.citibike_trips`
```



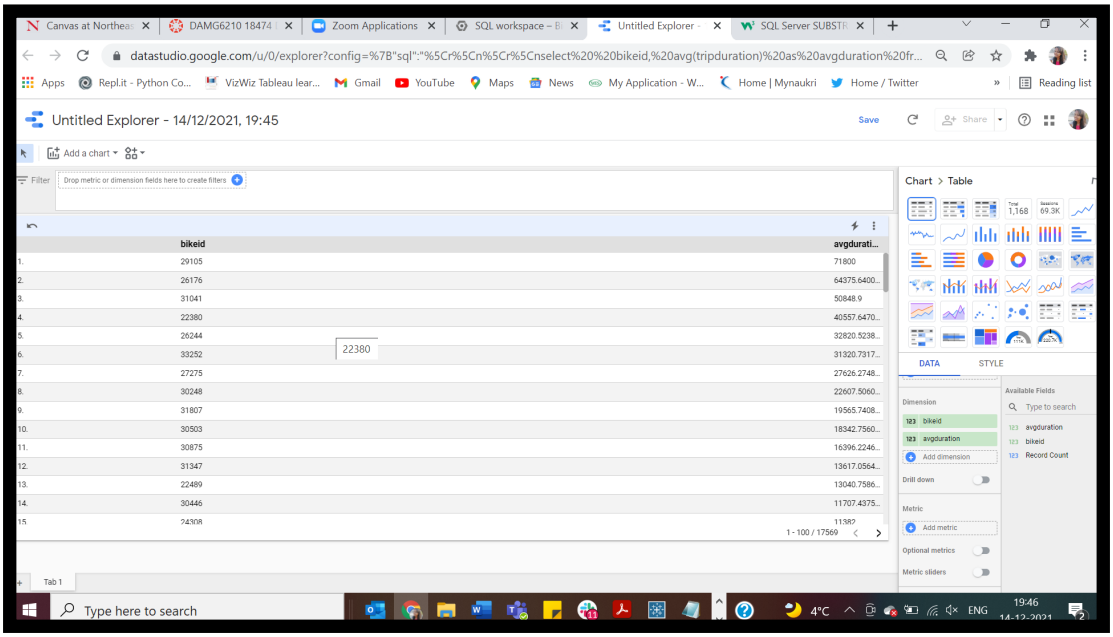
2. Top 5 bikes by time

```
select  bikeid, sum(tripduration) as duration from `healthy-genre-330203.damg_datset.citibike_trips`
group by 1
```



3. Average trip time

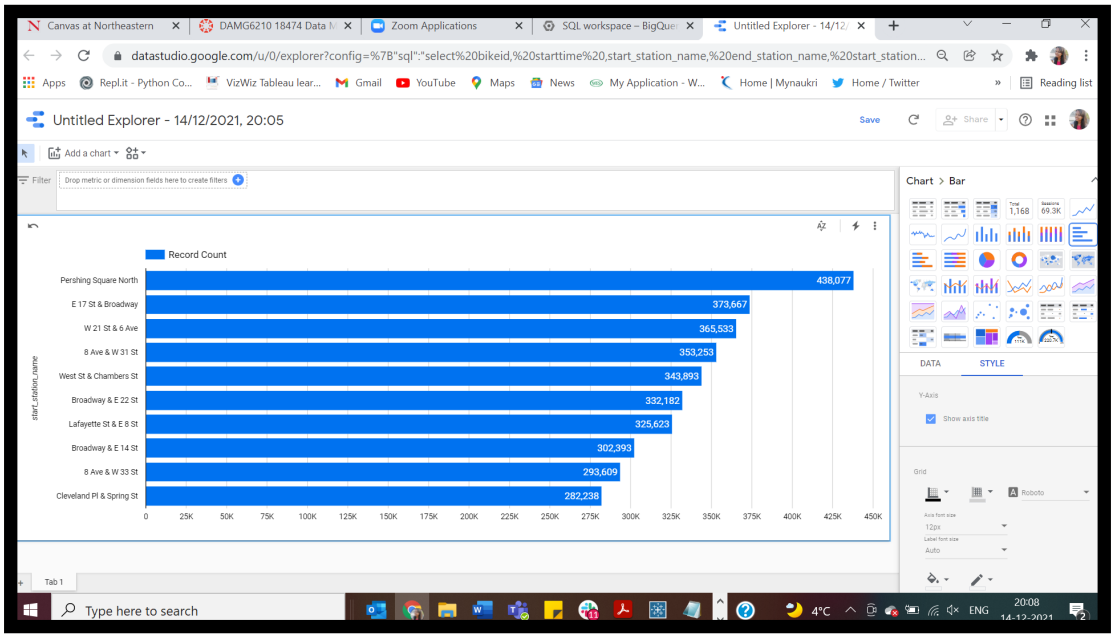
```
select  bikeid, avg(tripduration) as avgduration from `healthy-genre-330203.damg_datset.citibike_trips`
group by 1
```



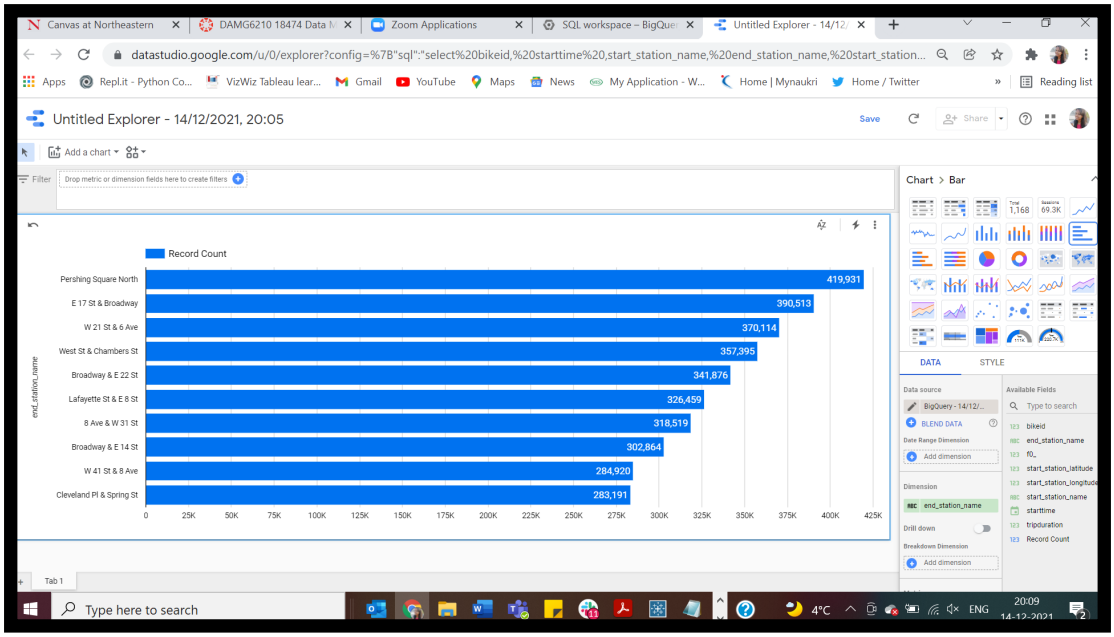
4. Top Bike

```
select bikeid, starttime ,start_station_name, end_station_name, start_station_latitude, start_station_longitude, tripduration, count(tripduration)
from `healthy-genre-330203.damg_datset.citibike_trips`
group by 1,2,3,4,5,6,7
```

5. Route- start location



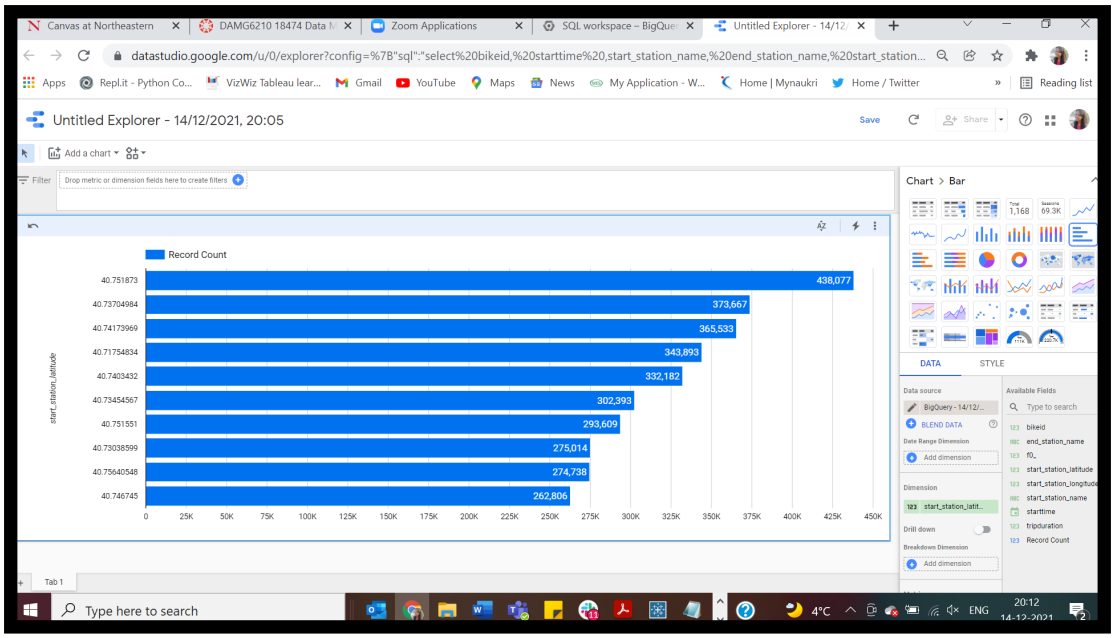
6. Route- End location



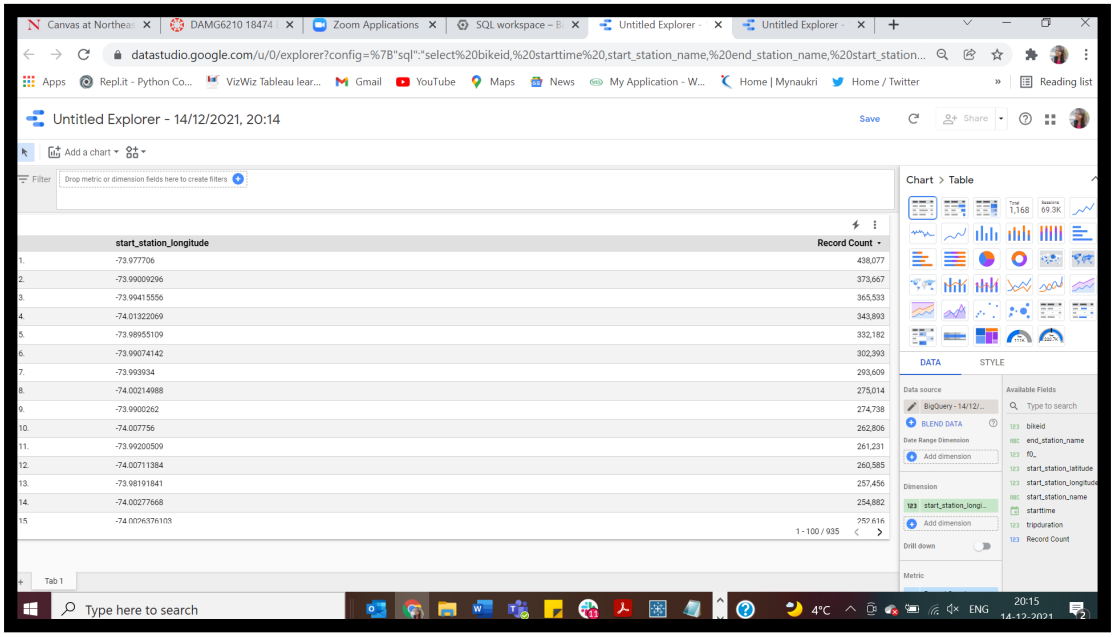
7. Top latitude

```
select bikeid, starttime ,start_station_name, end_station_name, start_statio
n_latitude, start_station_longitude, tripduration, count(bikeid)
from `healthy-genre-330203.damg_datset.citibike_trips`
```

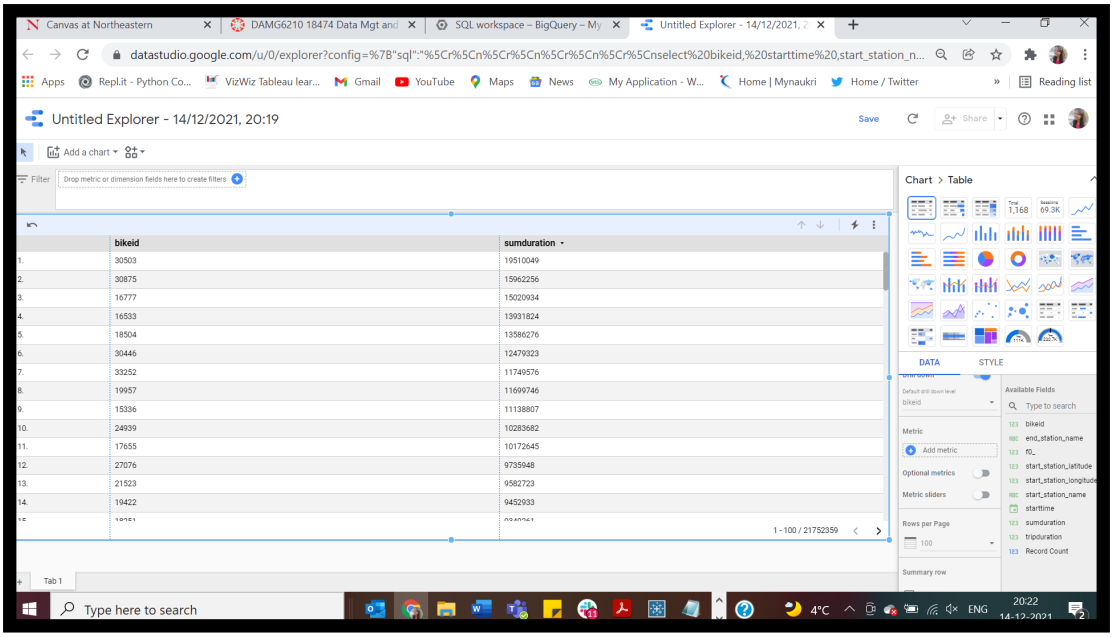
```
group by 1,2,3,4,5,6,7
```



8. Top longitude



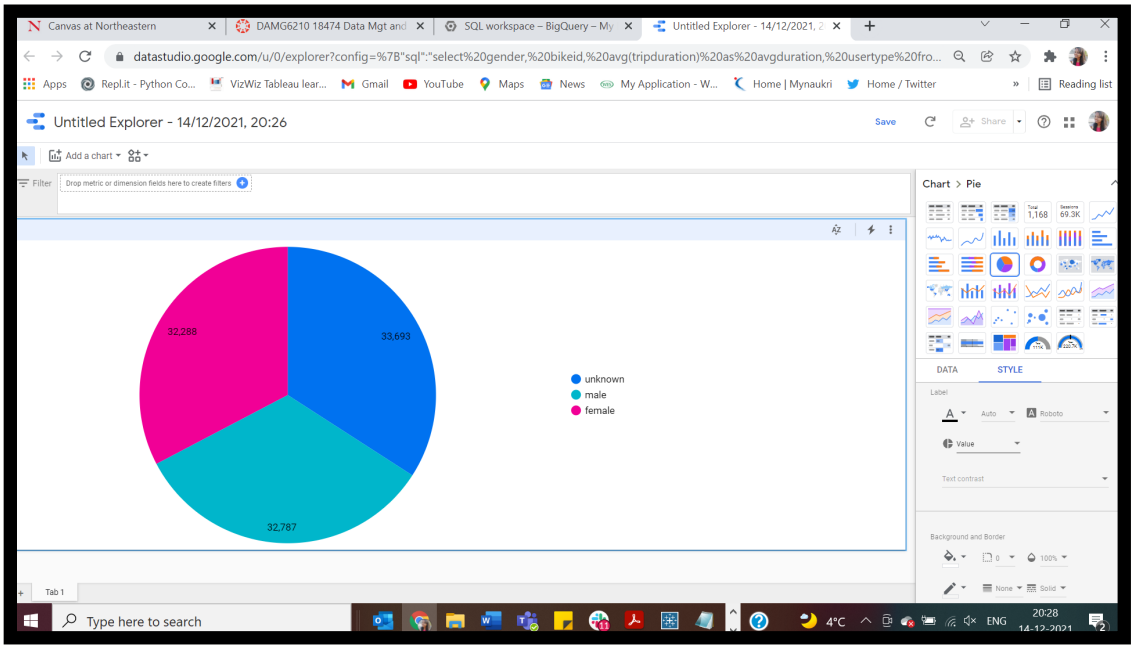
9. Duration



Gender Analysis:

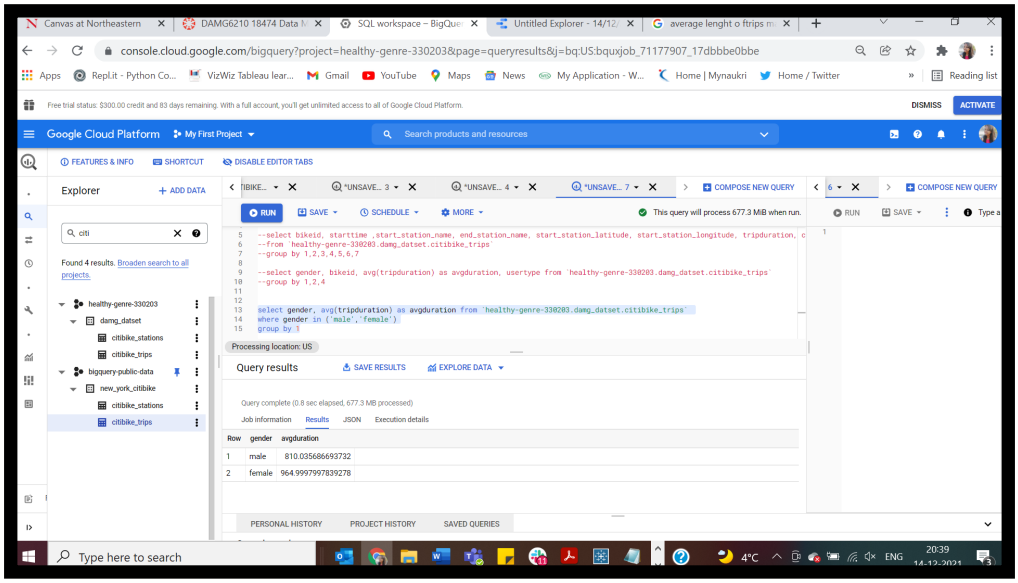
1. Male Vs Female

```
select gender, bikeid, avg(tripduration) as avgduration, usertype from `h
ealthy-genre-330203.damg_datset.citibike_trips`
group by 1,2,4
```



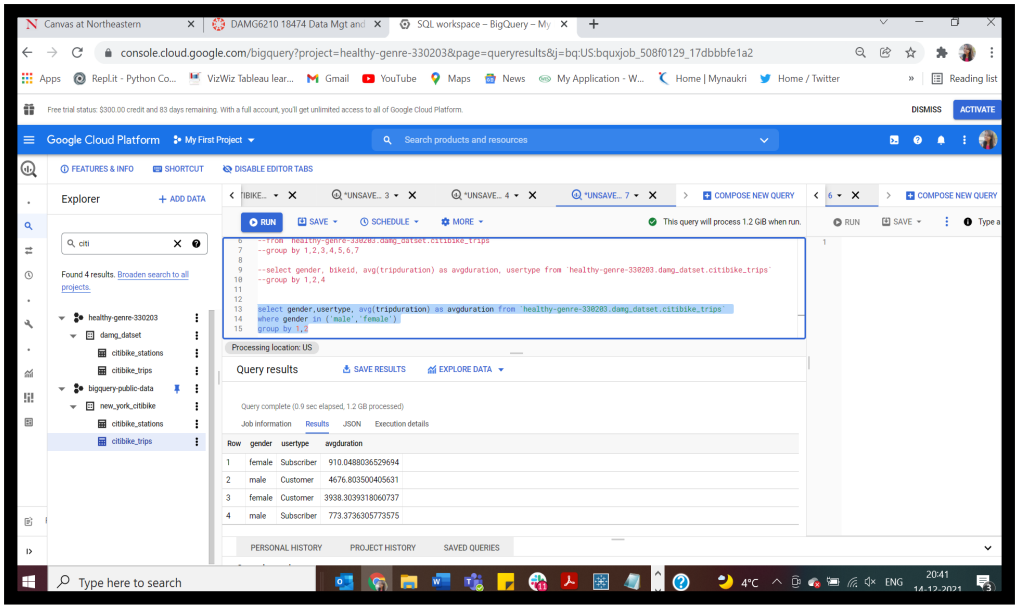
2. Average trip male vs female

```
select gender, avg(tripduration) as avgduration from `healthy-
genre-330203.damg_datset.citibike_trips`
where gender in ('male','female')
group by 1
```



3. Average trip duration by gender

```
select gender,usertype, avg(tripduration) as avgduration from `healthy
-genre-330203.damg_datset.citibike_trips`
where gender in ('male','female')
group by 1,2
```



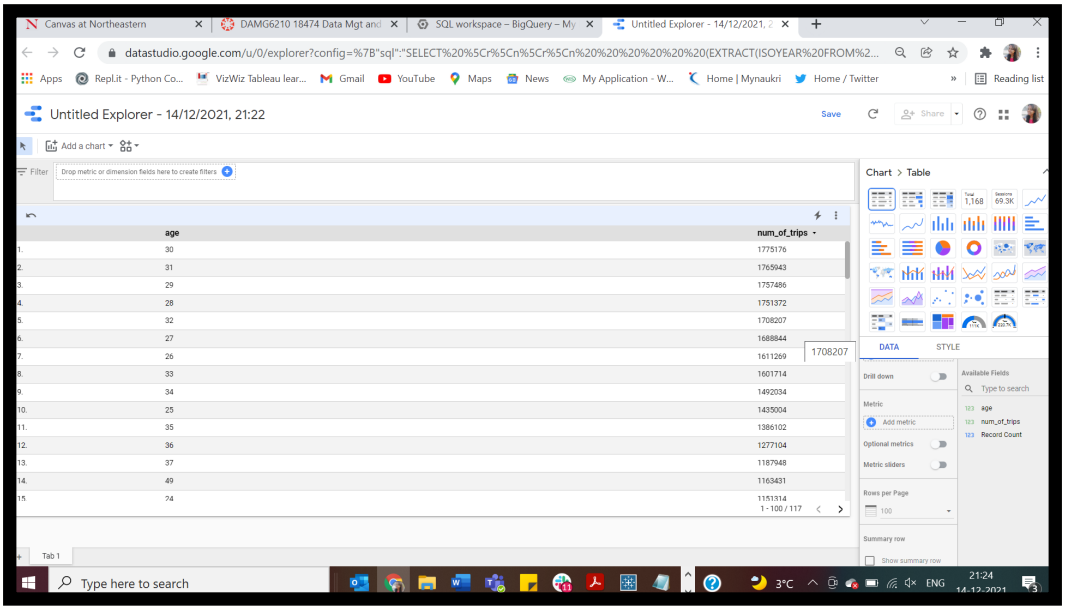
4. Year over year trend

```
select  extract(year from starttime) as year, gender
 from `healthy-genre-330203.damg_datset.citibike_trips`
 where gender in ('male','female')
```

Age Analysis

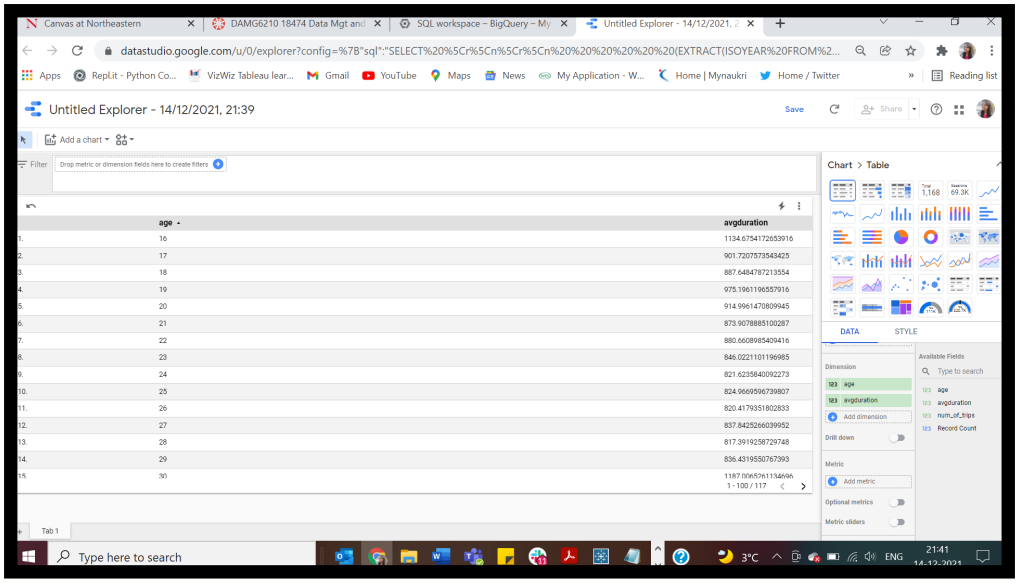
1. SELECT

```
      (EXTRACT(ISOYEAR FROM starttime) -
 (cast(birth_year as INT64))) as age,

      COUNT(bikeid) num_of_trips, FROM `healthy-
genre-330203.damg_datset.citibike_trips`

WHERE CAST(birth_year as string) != "null"

GROUP BY age

ORDER BY age
```

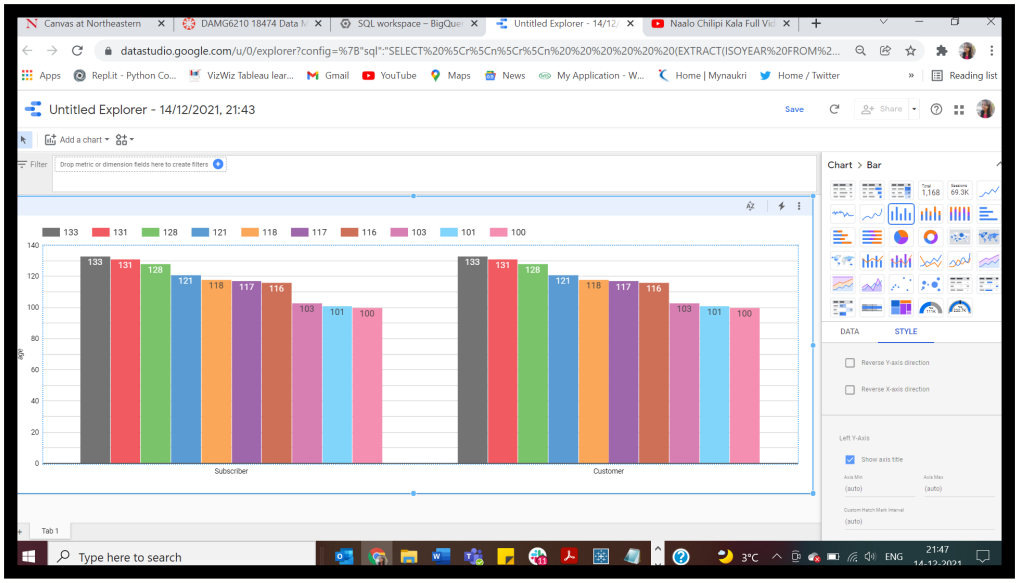


2. SELECT

```
      (EXTRACT(ISOYEAR FROM starttime) -
 (cast(birth_year as INT64))) as age,

      COUNT(bikeid) num_of_trips,avg(tripduration) as avgduration

FROM `healthy-genre-330203.damg_datset.citibike_trips`

WHERE CAST(birth_year as string) != "null"

GROUP BY age

ORDER BY age
```

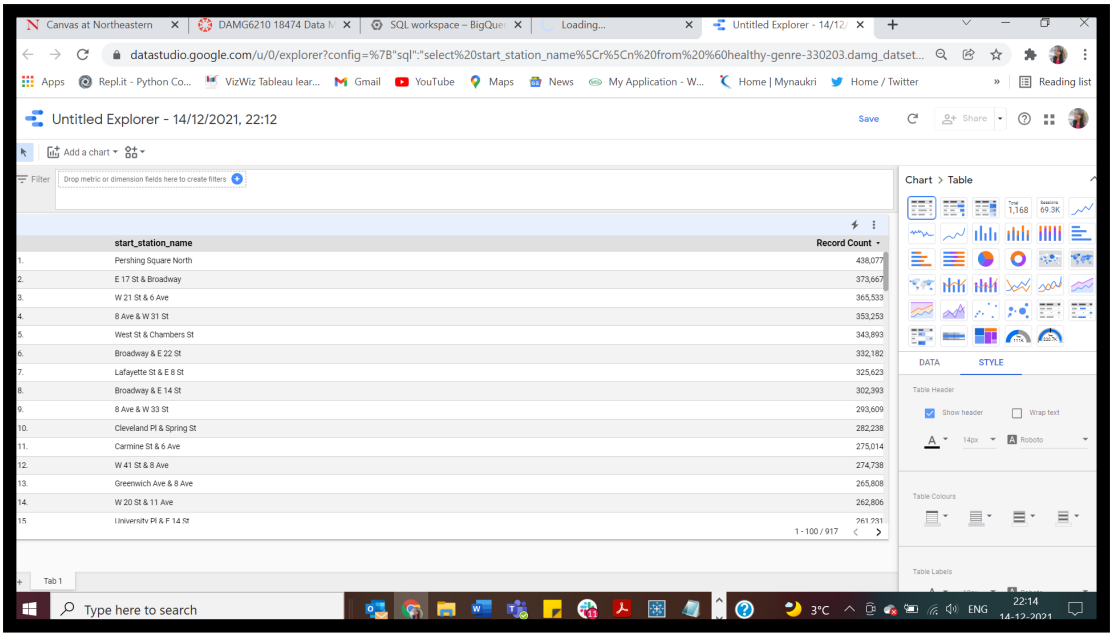


3.
```
SELECT (EXTRACT(ISOYEAR FROM starttime) -
 (cast(birth_year as INT64))) as
age, COUNT(bikeid) num_of_trips,usertype

FROM `healthy-genre-330203.damg_datset.citibike_trips`

WHERE CAST(birth_year as string) != "null"

GROUP BY age, usertype ORDER BY age
```

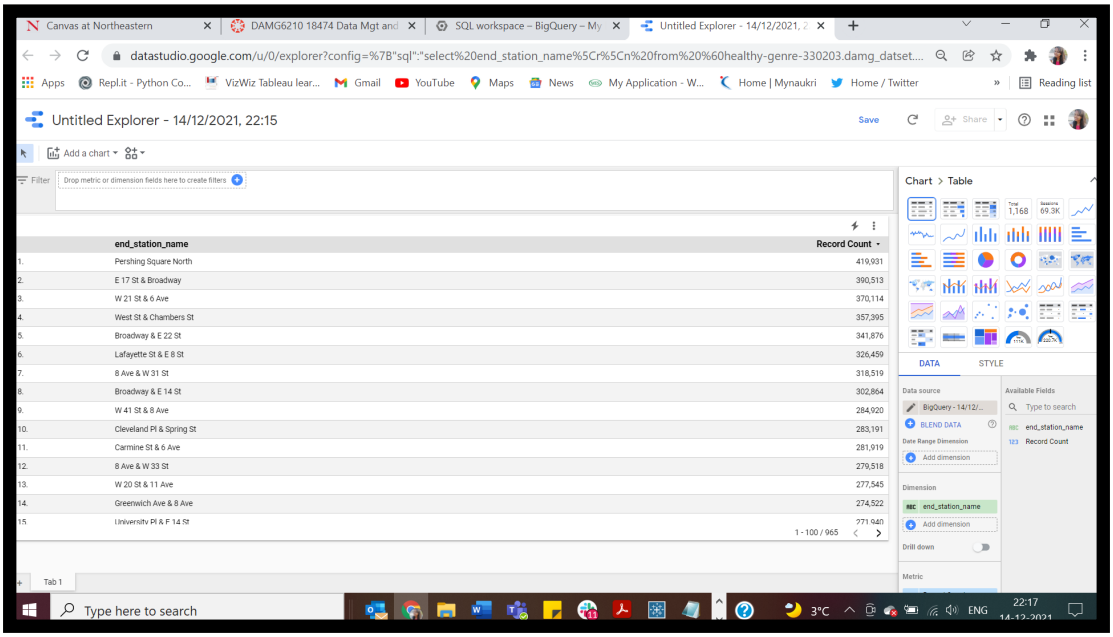


Most Popular Stations:

1.
```
select start_station_name
 from `healthy-genre-330203.damg_datset.citibike_trips`
```

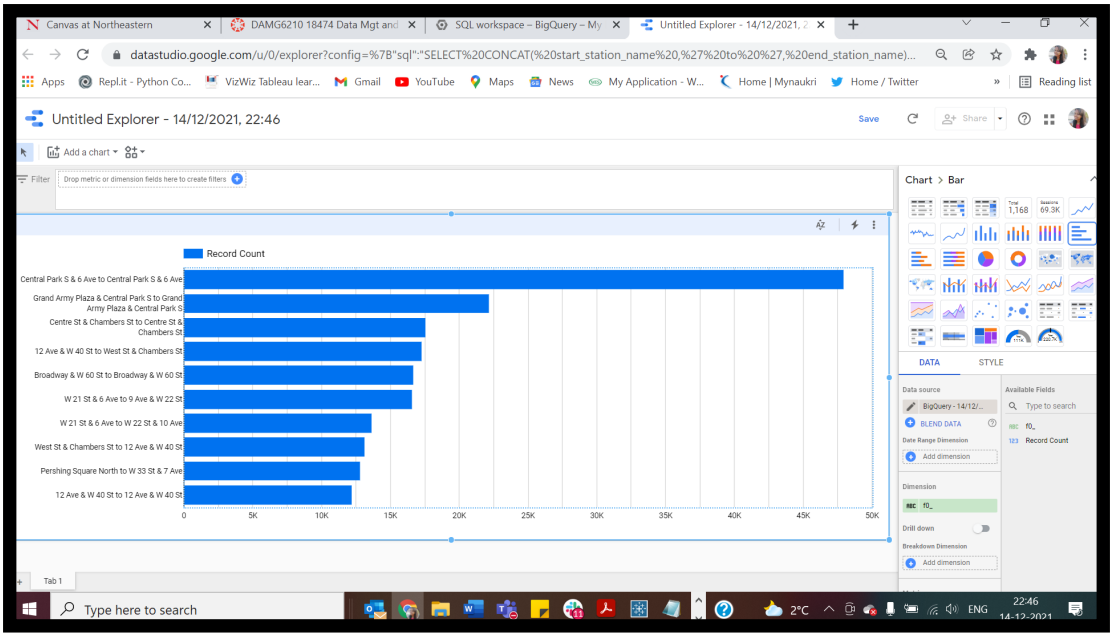


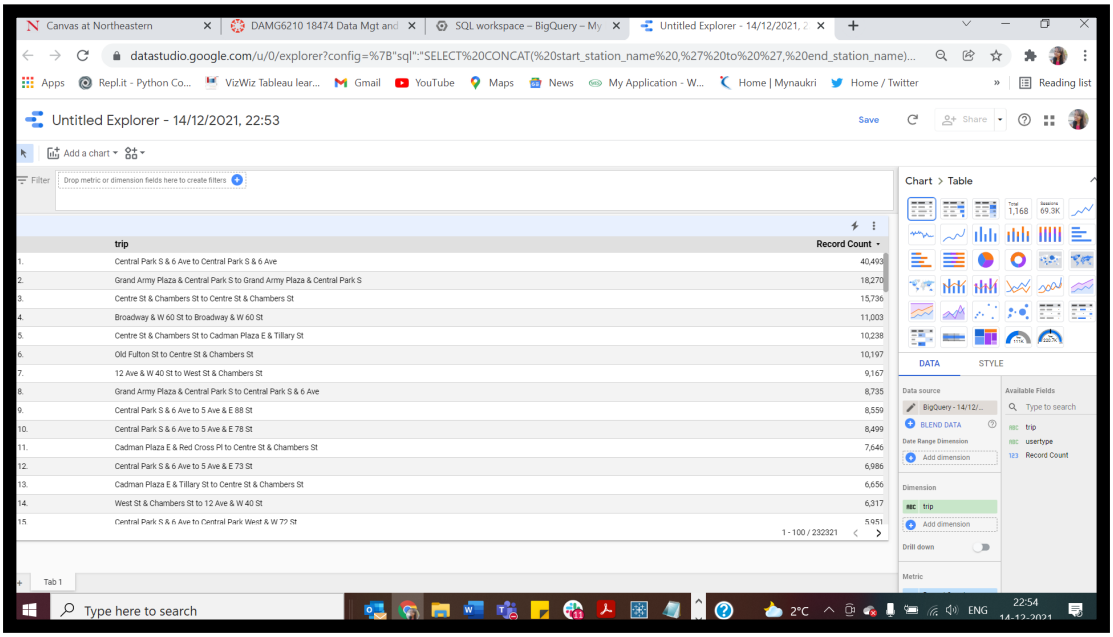
2. End station highest

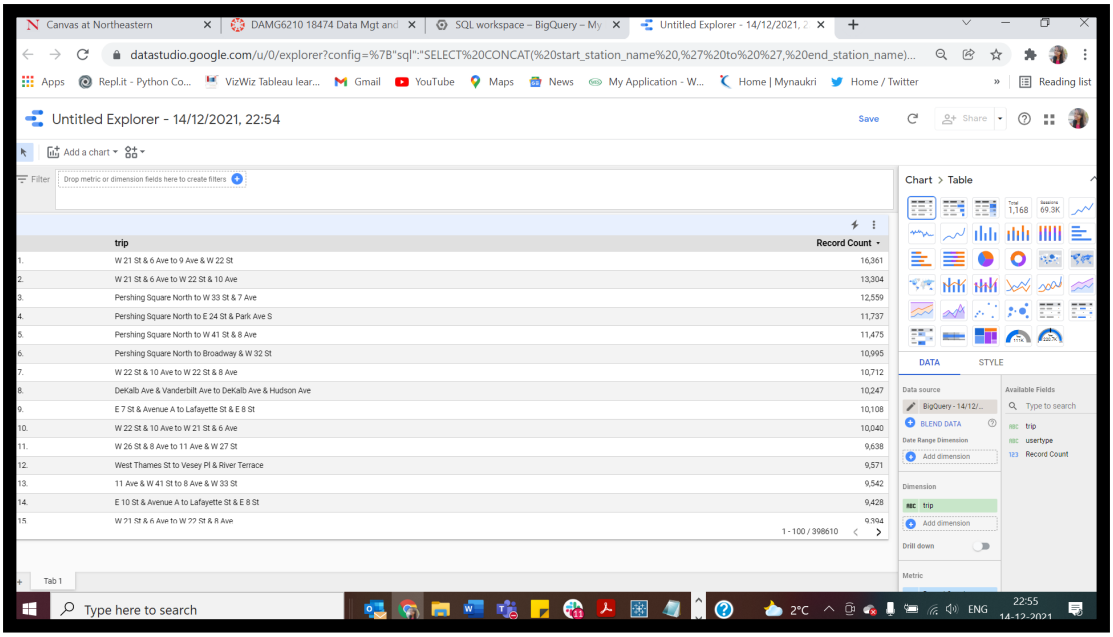


3. Both start and end trip

```
SELECT CONCAT( start_station_name ,' to ', end_station_name) from `healthy-genre-330203.damg_datset.citibike_trips` ;
```


Untitled Explorer - 14/12/2021, 22:46
Record Count
Central Park S & 6 Ave to Central Park S & 6 Ave
Grand Army Plaza & Central Park S to Grand Army Plaza & Central Park S
Centre St & Chambers St to Centre St & Chambers St
12 Ave & W 40 St to West St & Chambers St
Broadway & W 60 St to Broadway & W 60 St
W 21 St & 6 Ave to 9 Ave & W 22 St
W 21 St & 6 Ave to W 22 St & 10 Ave
West St & Chambers St to 12 Ave & W 40 St
Pershing Square North to W 33 St & 7 Ave
12 Ave & W 40 St to 12 Ave & W 40 St
0
5K
10K
15K
20K
25K
30K
35K
40K
45K
50K
Chart > Bar
DATA
STYLE

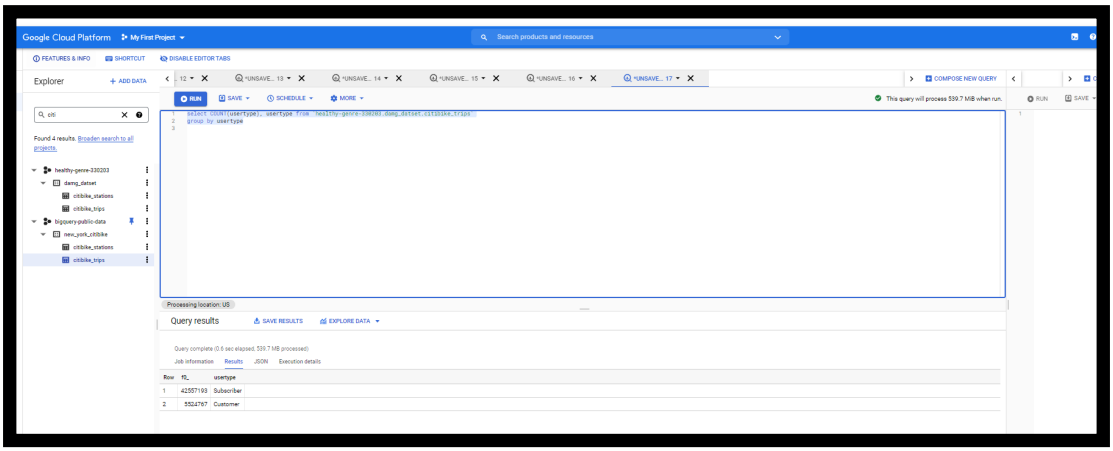

Untitled Explorer - 14/12/2021, 22:53
Chart > Table
DATA
STYLE
trip
Record Count
1 Central Park S & 6 Ave to Central Park S & 6 Ave 40,403
2 Grand Army Plaza & Central Park S to Grand Army Plaza & Central Park S 18,270
3 Centre St & Chambers St to Centre St & Chambers St 15,736
4 Broadway & W 60 St to Broadway & W 60 St 11,003
5 Centre St & Chambers St to Cadman Plaza E & Tillary St 10,238
6 Old Fulton St to Centre St & Chambers St 10,197
7 12 Ave & W 40 St to West St & Chambers St 9,167
8 Grand Army Plaza & Central Park S to Central Park S & 6 Ave 8,735
9 Central Park S & 6 Ave to 5 Ave & E 88 St 8,559
10 Central Park S & 6 Ave to 5 Ave & E 78 St 8,499
11 Cadman Plaza E & Red Cross Pl to Centre St & Chambers St 7,646
12 Central Park S & 6 Ave to 5 Ave & E 73 St 6,986
13 Cadman Plaza E & Tillary St to Centre St & Chambers St 6,656
14 West St & Chambers St to 12 Ave & W 40 St 6,317
15 Central Park S & 6 Ave to Central Park West & W 72 St 5,951
1 - 100 / 232321

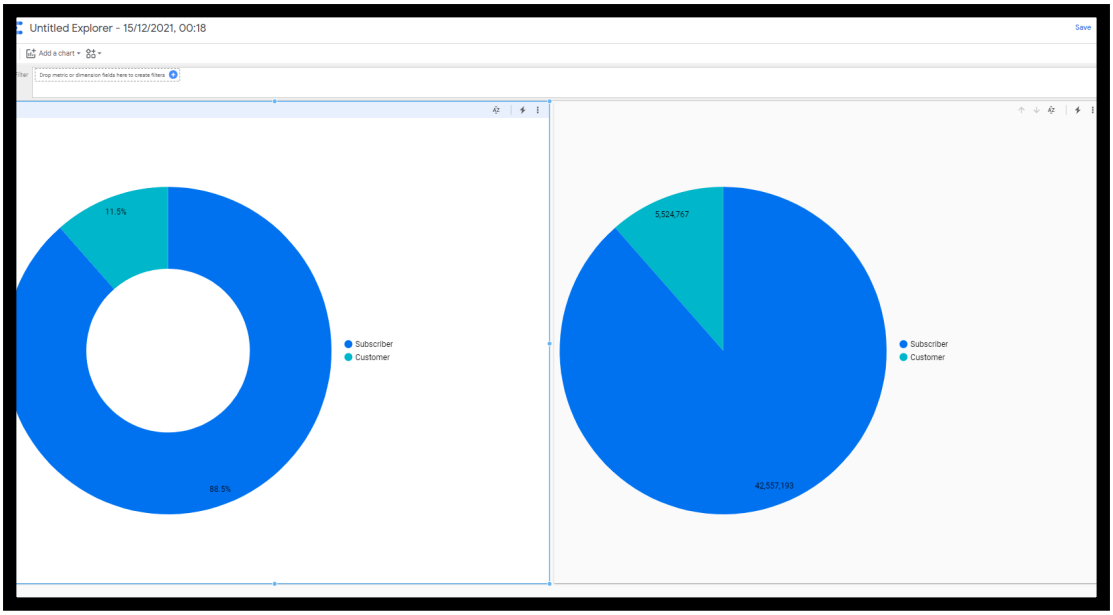


Trip Analysis:

1. Subscriber Vs Customer

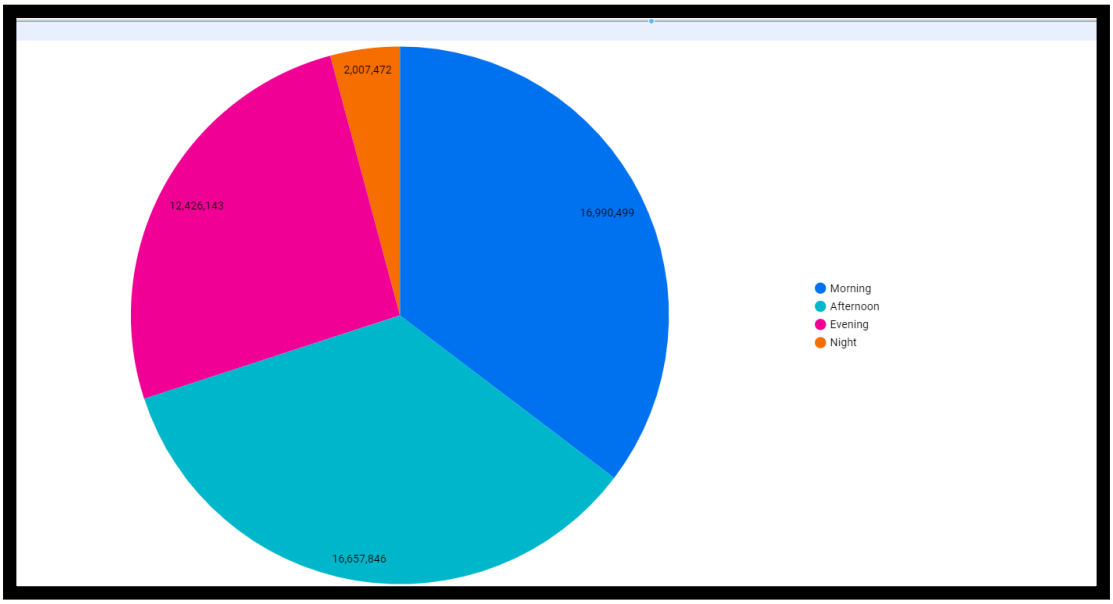
```
select COUNT(usertype), usertype from `healthy-genre-330203.damg_datset.citibike_trips`
group by usertype
```





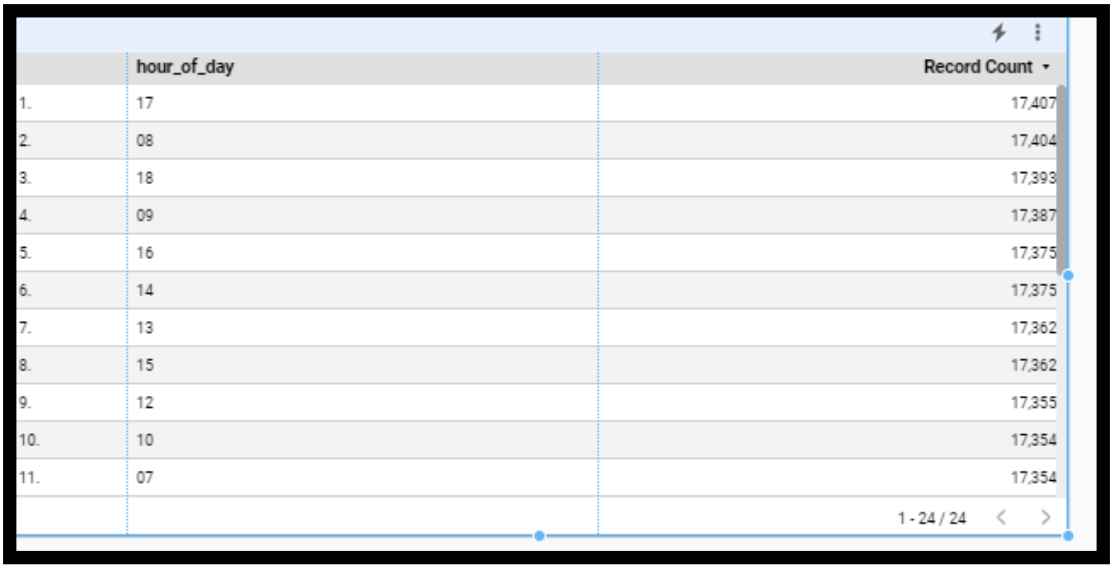
2. Bike trips during what period of the day

```
select  case when extract(hour from starttime) in (06,07,08, 09, 10, 11, 12) then '
Morning'
when extract(hour from starttime) in (13,14,15, 16, 17) then 'Afternoon'
when extract (hour from starttime) in (18, 19, 20, 21, 22) then 'Evening'
else 'Night' end as timeday from `healthy-genre-330203.damg_datset.citibike_trips`
```



3. Bike trips for what hour

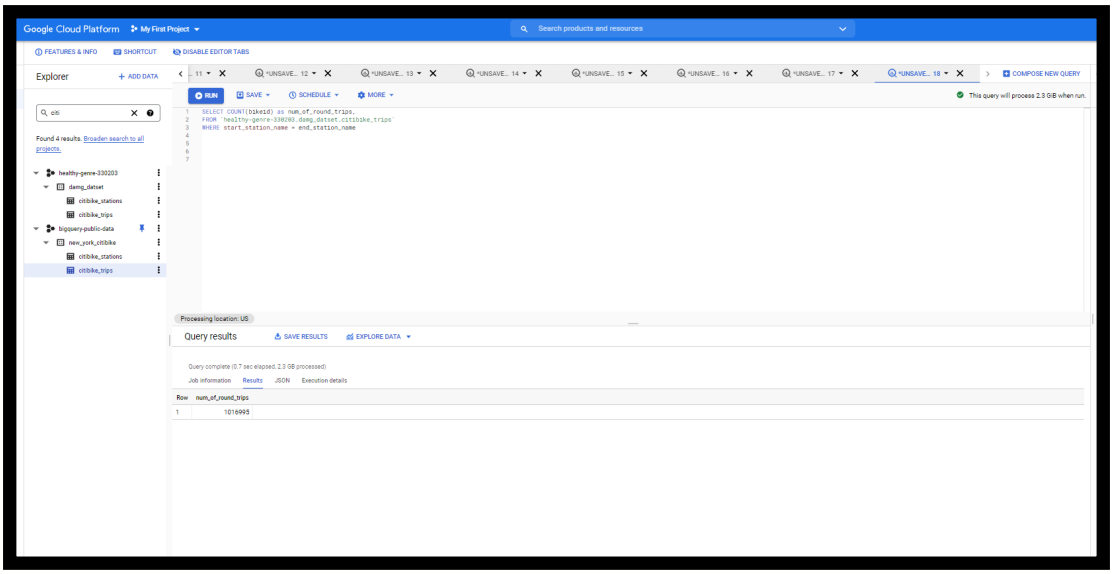
```sql
select count(bikeid) as number_of_trips, extract(hour from starttime) as hou
r_of_day
 from `healthy-genre-330203.damg_datset.citibike_trips`
 group by bikeid, 2
```

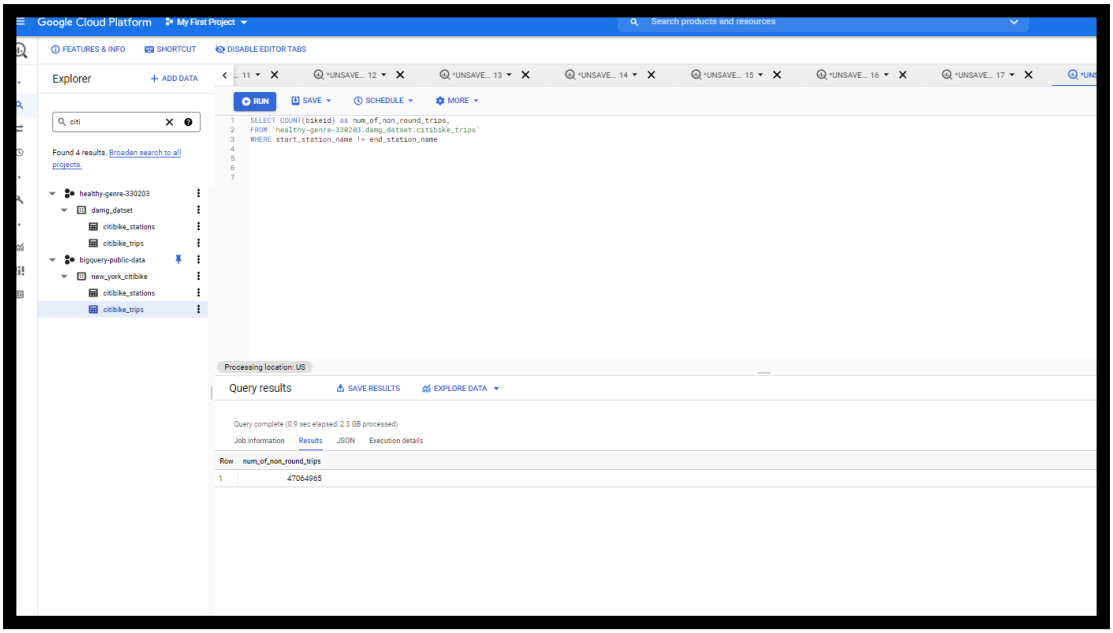


	hour_of_day	Record Count
1.	17	17,407
2.	08	17,404
3.	18	17,393
4.	09	17,387
5.	16	17,375
6.	14	17,375
7.	13	17,362
8.	15	17,362
9.	12	17,355
10.	10	17,354
11.	07	17,354

1 - 24 / 24

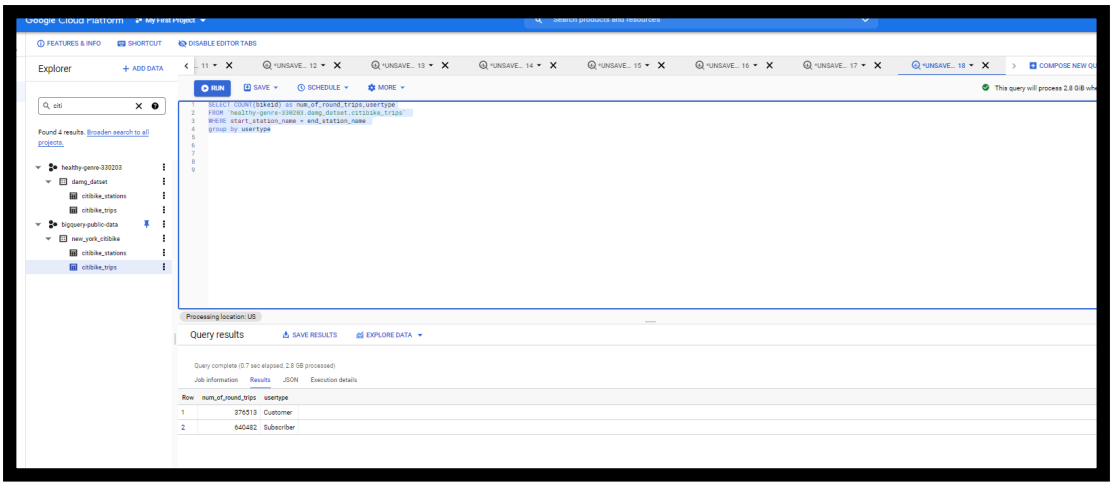
4. Round trips





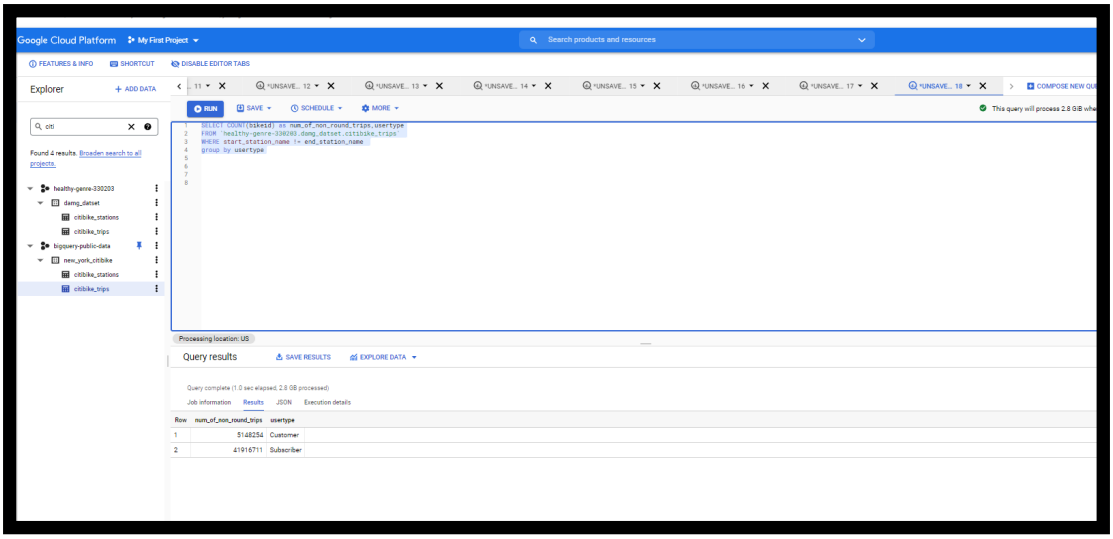
5. Round trips customer vs subscriber

```
SELECT COUNT(bikeid) as num_of_round_trips,usertype
FROM `healthy-genre-330203.damg_datset.citibike_trips`
WHERE start_station_name = end_station_name
group by usertype
```



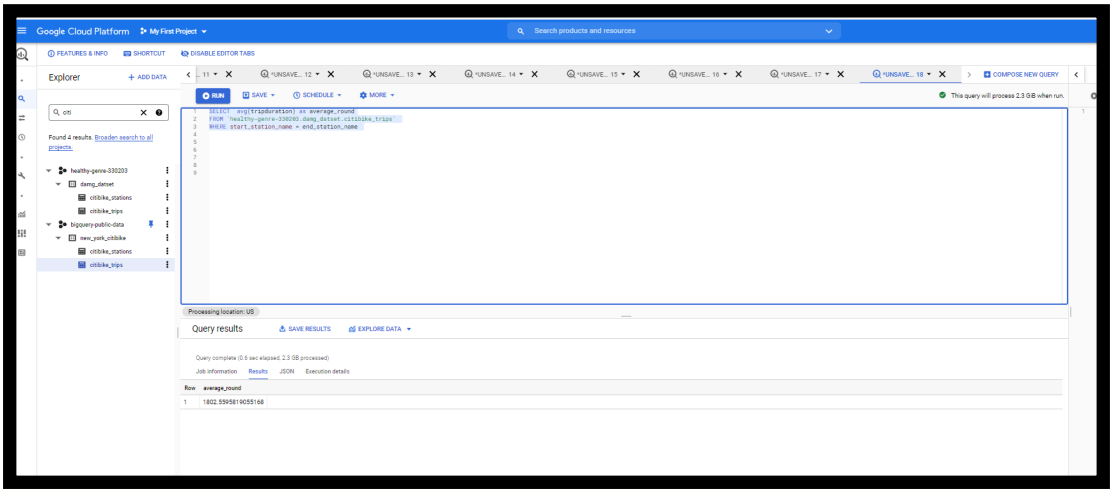
6.

```
SELECT COUNT(bikeid) as num_of_non_round_trips,usertype
FROM `healthy-genre-330203.damg_datset.citibike_trips`
WHERE start_station_name != end_station_name
group by usertype
```



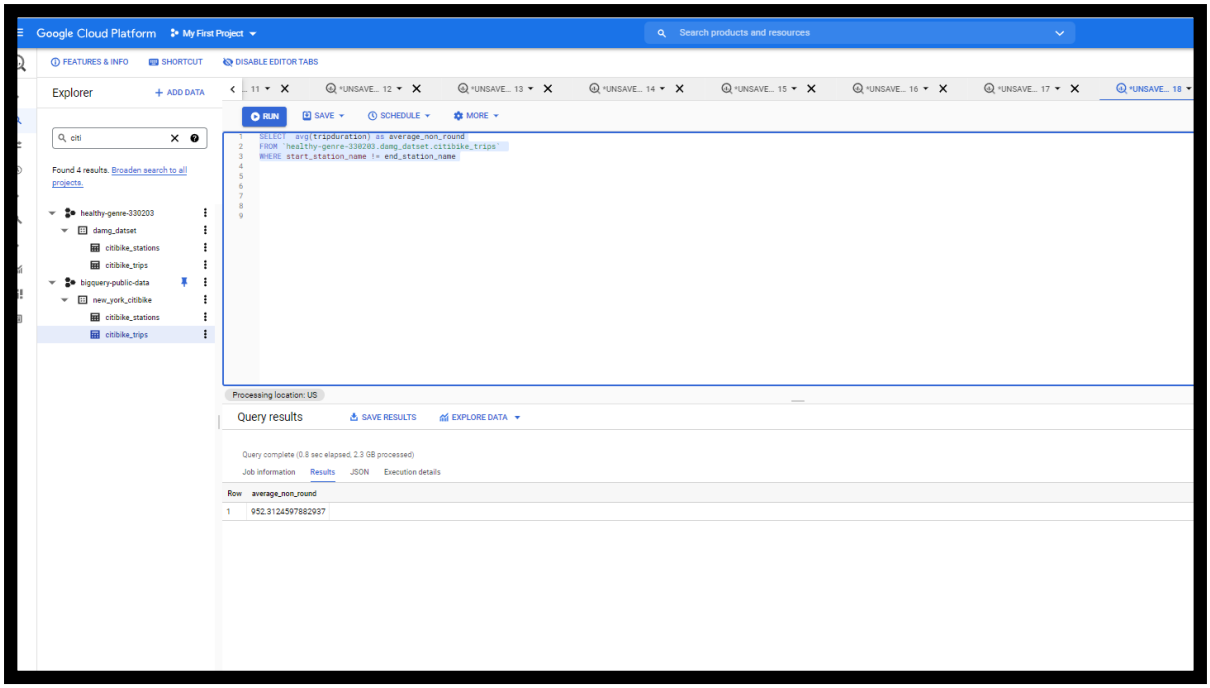
7. Round trips average length of trip

```
SELECT  avg(tripduration) as average_round
FROM `healthy-genre-330203.damg_datset.citibike_trips`
WHERE start_station_name = end_station_name
```



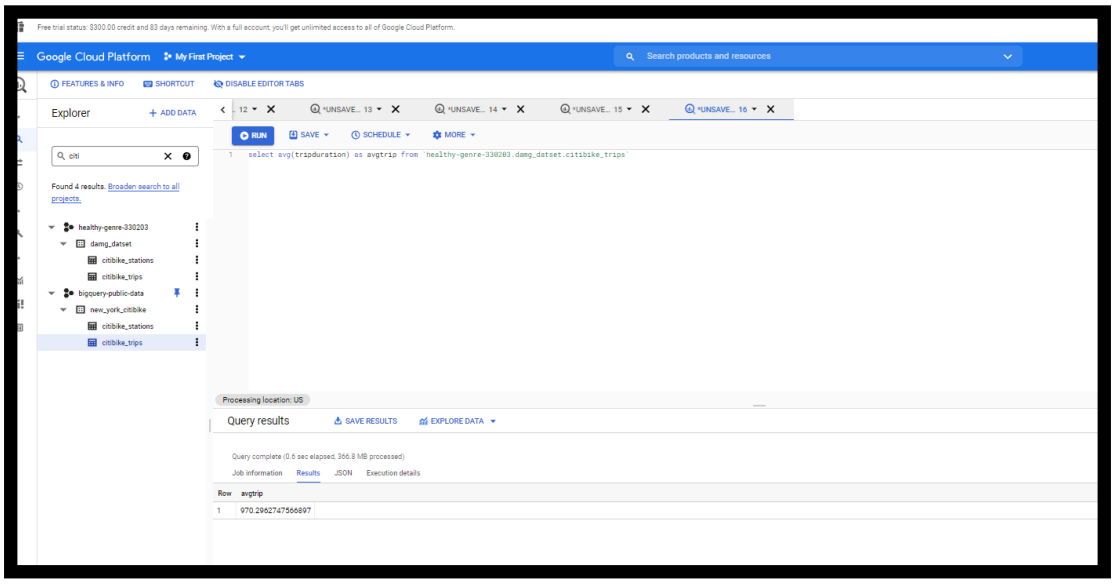
8.

```
SELECT  avg(tripduration) as average_non_round
FROM `healthy-genre-330203.damg_datset.citibike_trips`
WHERE start_station_name != end_station_name
```



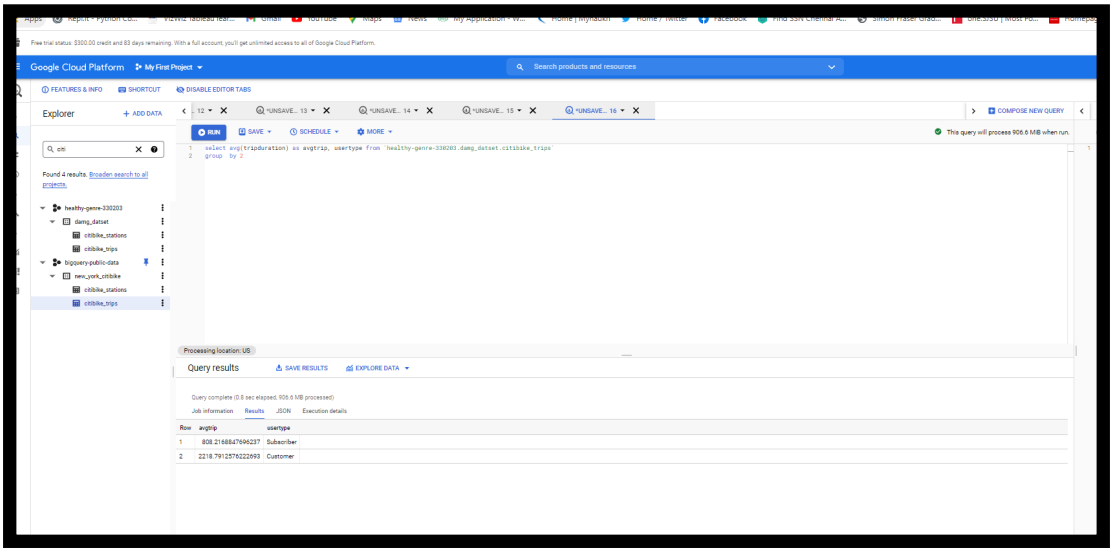
Analysis of length of trip

1. Average length of trip



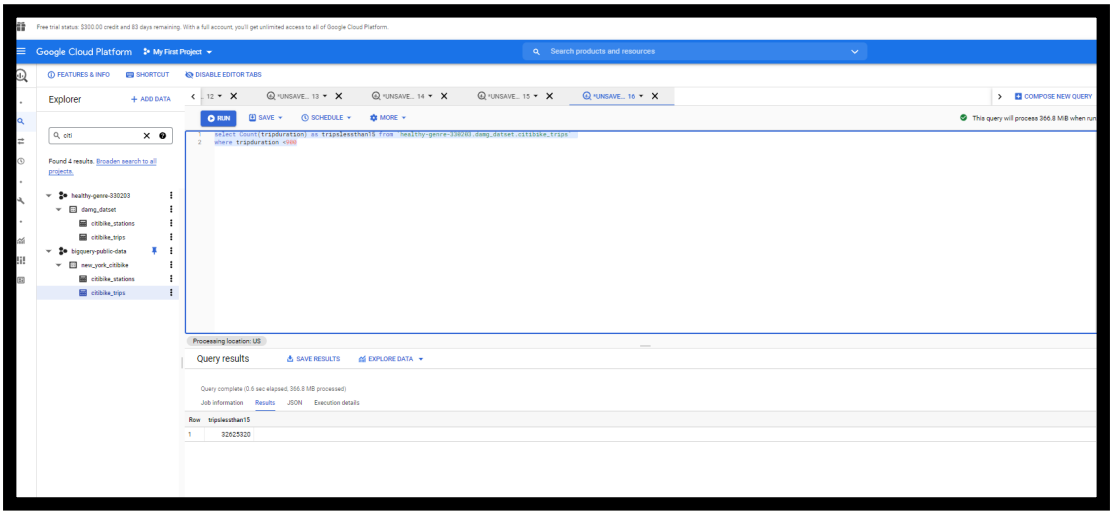
2. Average length of trip by customer vs Subscriber

```
select avg(tripduration) as avgtrip, usertype from `healthy-
genre-330203.damg_datset.citibike_trips`
group  by 2
```



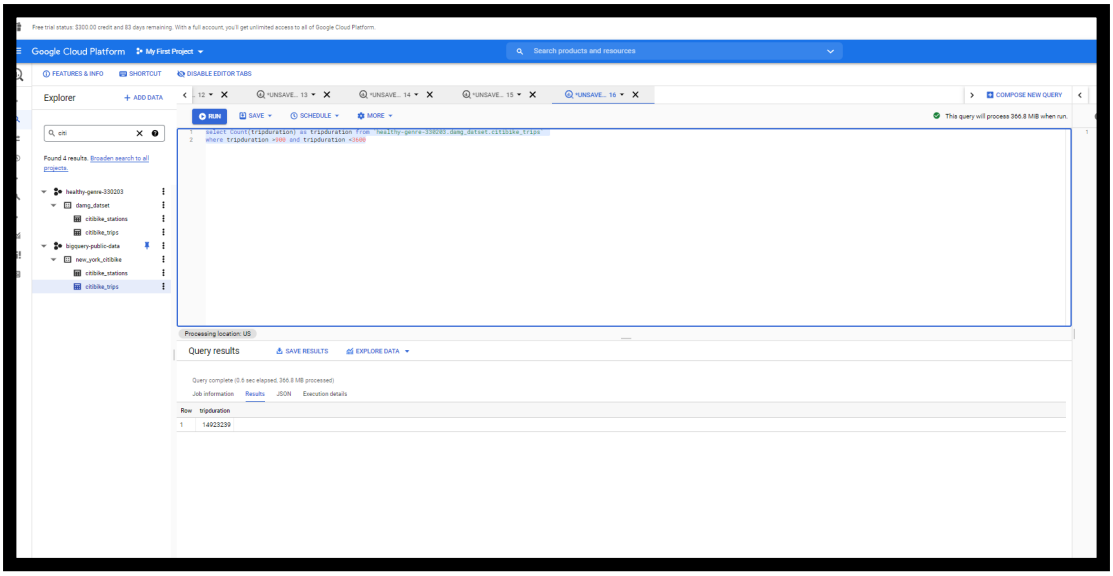
3. Trips less than 15 minutes

```
select Count(tripduration) as tripslessthan15 from `healthy-
genre-330203.damg_datset.citibike_trips`
where tripduration <900
```



4. Trip duration less than an hour

```
select Count(tripduration) as tripduration from `healthy-
genre-330203.damg_datset.citibike_trips`
where tripduration >900 and tripduration <3600
```



5. Trip duration more than 1 hour

```
select Count(tripduration) as tripdurationmorethan1 from `healthy-genre-330203.damg_datset.citibike_trips`
where  tripduration >3600
```