

FINAL PROJECT

NYC CITI BIKE

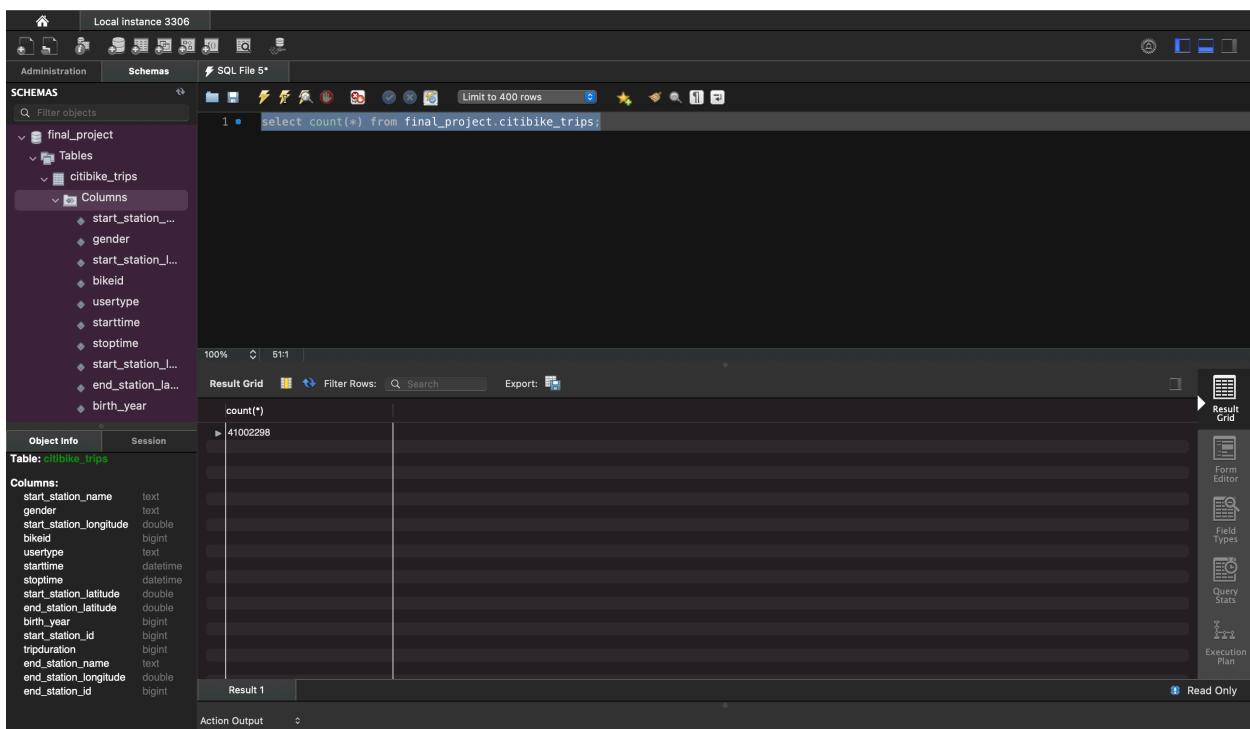
Northeastern University

Data Management and Data Design

Semester-I (Sept 2021-Dec 2021)

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

SQL Queries:



The screenshot shows the Oracle SQL Developer interface. The left sidebar displays the schema structure for the 'final_project' schema, specifically the 'citibike_trips' table, which contains 11 columns: start_station_name, gender, start_station_longitude, bikeid, usertype, starttime, stoptime, start_station_latitude, end_station_latitude, birth_year, and start_station_id. The main pane shows a SQL editor with the following query:

```
1 select count(*) from final_project.citibike_trips;
```

The result grid shows a single row with the value 41002298. The right sidebar contains various tools and tabs for managing the session.

Schemas

```
1 • select * from final_project.citibike_trips;
```

start_station_name	gender	start_station_longitude	bikeid	usertype	starttime	stoptime	start_station_latitude	end_station_latitude	birth_year	tripduration	end_st...
W 52 St & 11 Ave	unknown	-73.99392888	15331	Customer	2019-01-19 19:11:13	2019-01-19 19:41:18	40.76727216	40.73921553	1969	72	1805
W 52 St & 11 Ave	unknown	-73.99392888	28673	Customer	2019-01-19 19:11:13	2019-01-19 19:41:18	40.76727216	40.73921553	1969	72	1805
Franklin St & W Broadway	unknown	-74.00666661	2838	Customer	2019-01-20 11:29:31	2019-01-20 11:54:49	40.71911582	40.7285887	1969	79	1518
St James Pl & Pearl St	unknown	-74.00016545	31344	Customer	2019-01-19 14:14:16	2019-01-19 14:22:25	40.71117416	40.71117416	1969	82	488
St James Pl & Pearl St	unknown	-74.00016545	27013	Customer	2019-01-19 14:22:49	2019-01-19 14:50:42	40.71117416	40.7143724085228	1969	82	1672
St James Pl & Pearl St	unknown	-74.00016545	31345	Customer	2019-01-19 14:50:17	2019-01-19 14:50:42	40.71117416	40.7143724085228	1969	82	2073
St James Pl & Pearl St	unknown	-74.00016545	18030	Customer	2019-01-19 14:14:55	2019-01-19 14:50:42	40.71117416	40.7143724085228	1969	82	2317
Barrow St & Hudson St	unknown	-74.00674436	26758	Customer	2019-01-18 18:50:58	2019-01-18 19:05:57	40.73172428	40.73172428	1969	127	1748
Barrow St & Hudson St	unknown	-74.00674436	19528	Customer	2019-01-19 16:16:00:24	2019-01-19 16:16:41	40.73172428	40.7149787	1969	127	976
Barrow St & Hudson St	unknown	-74.00674436	28604	Customer	2019-01-19 15:13:03	2019-01-19 15:38:32	40.73172428	40.75929124	1969	127	1528
Barrow St & Hudson St	unknown	-74.00674436	17236	Customer	2019-01-19 15:13:17	2019-01-19 15:47:39	40.73172428	40.75929124	1969	127	2062
MacDougal St & Prince St	unknown	-74.00297088	21444	Customer	2019-01-19 11:52:04	2019-01-19 12:58:18	40.72710258	40.728846	1969	128	3973
MacDougal St & Prince St	unknown	-74.00297088	28759	Customer	2019-01-23 15:04:40	2019-01-23 15:23:19	40.72710258	40.74096374	1969	128	1118
MacDougal St & Prince St	unknown	-74.00297088	3274	Customer	2019-01-22 08:17:42	2019-01-22 08:30:36	40.72710258	40.75466591	1969	128	1313
Nassau St & Navy St	unknown	-73.98068914	28616	Customer	2019-01-21 12:57:30	2019-01-21 13:07:45	40.69839895	40.69839895	1969	144	615
Nassau St & Navy St	unknown	-73.98068914	16246	Customer	2019-01-22 23:23:14	2019-01-22 23:34:33	40.69839895	40.69839895	1969	144	678
Nassau St & Navy St	unknown	-73.98068914	16150	Customer	2019-01-19 04:37:36	2019-01-19 04:44:14	40.69839895	40.69610226	1969	144	398
Nassau St & Navy St	unknown	-73.98068914	34307	Customer	2019-01-22 07:51:17	2019-01-22 08:39:04	40.69839895	40.76344058	1969	144	2867
Nassau St & Navy St	unknown	-73.98068914	32982	Customer	2019-01-19 11:52:43	2019-01-19 12:18:22	40.69839895	40.72608	1969	144	1538

1.Seasonality

select case when month(starttime) in (06,07,08) then 'Summer'

when month(starttime) in (09,10,11) then 'Autumn'

when month(starttime) in (12,01,02) then 'Winter'

else 'Spring' end as Season , count(*) as total_rides_in_season from final_project.citibike_trips

group by Season;

```

24
25 • ② select case when month(starttime) in (06,07,08) then 'Summer'
26   when month(starttime) in (09,10,11) then 'Autumn'
27   when month(starttime) in (12,01,02) then 'Winter'
28   else 'Spring' end as Season , count(*) as total_rides_in_season from final_project.citibike_trips
29   group by Season;
30

```

Result Grid

Season	total_rides_in_season
Winter	7286340
Spring	8257726
Summer	12688253
Autumn	12485979

Action Output

Time	Action	Response	Duration / Fetch Time
35 01:16:40	select case when month(starttime) in (06,07,08) then 'Summer' when month(starttime) in (0...	2000 row(s) returned	0.0046 sec / 0.0014...
36 01:18:34	select case when month(starttime) in (06,07,08) then 'Summer' when month(starttime) in (0...	Error Code: 1140. In aggregated query without GROUP BY,...	0.0021 sec
37 01:19:34	select case when month(starttime) in (06,07,08) then 'Summer' when month(starttime) in (0...	4 row(s) returned	28.659 sec / 0.0005...

Query Completed

2. Month-wise trend

SELECT

month (starttime) AS month ,COUNT(*) as Trip_count

FROM final_project.citibike_trips

group by month;

```

30
31 • ② SELECT
32   month (starttime) AS month ,COUNT(*) as Trip_count
33   FROM final_project.citibike_trips
34   group by month;
35

```

Result Grid

month	Trip_count
1	2207883
2	5534318
3	2530417
4	2448856
5	3412453
6	4007643
7	4286872
8	4673738
9	4933125
10	4341442
11	3215412
12	2044139

Action Output

Time	Action	Response	Duration / Fetch Time
38 01:27:06	SELECT month (starttime) AS month ,COUNT(*) as Trip_count FROM final_project.citibike_trips LIMIT 0, 2000	Error Code: 1140. In aggregated query without GROUP BY,...	0.0016 sec
39 01:27:07	SELECT month (starttime) AS month ,COUNT(*) as Trip_count FROM final_project.citibike_trips group by month LIMIT 0, 12	12 row(s) returned	21.851 sec / 0.00051...

3. Trend over time

- calendar year

SELECT

month (starttime) AS month ,COUNT(*) as Trip_count

FROM final_project.citibike_trips

WHERE year(starttime) = 2019

GROUP BY month

HAVING month is NOT NULL

ORDER BY month;

The screenshot shows the MySQL Workbench interface. On the left, the object browser displays the schema of the 'final_project' database, specifically the 'citibike_trips' table, which contains columns like start_station_name, gender, start_station_longitude, bikeid, usertype, starttime, and stoptime. The main area shows a query window with the following SQL code:

```
1 • SELECT
2   month (starttime) AS month ,COUNT(*) as Trip_count
3   FROM final_project.citibike_trips
4   WHERE year(starttime) = 2019
5   GROUP BY month
6   HAVING month is NOT NULL
7   ORDER BY month;
```

Below the query window is a 'Result Grid' showing the output of the query. The results are as follows:

month	Trip_count
1	967287
2	1887488
3	1327960
4	1766094
5	1924563
6	2125370
7	2181064
8	2344224
9	2444900
10	2092573
11	1478708
12	1955210

At the bottom, the 'Action Output' section shows two log entries:

Action	Time	Response	Duration / Fetch Time
24	00:41:26	SELECT month (starttime) AS month ,COUNT(*) as Trip_count FROM final_project.citibike_trips	15.489 sec / 0.00046...
25	00:42:06	SELECT month (starttime) AS month ,COUNT(*) as Trip_count FROM final_project.citibike_trips	19.439 sec / 0.00005...

▪ calendar month & year

SELECT

day (starttime) AS day ,COUNT(*) as Trip_count

FROM final_project.citibike_trips

WHERE month(starttime) = 09 and year(starttime) = 2019

GROUP BY day

HAVING day is NOT NULL

ORDER BY day;

The screenshot shows the MySQL Workbench interface with a query editor and a results grid.

Query Editor:

```

7    ORDER BY month;
8
9 • SELECT
10   day (starttime) AS day ,COUNT(*) as Trip_count
11   FROM final_project.citibike_trips
12   WHERE month(starttime) = 09 and year(starttime) = 2019
13   GROUP BY day
14   HAVING day is NOT NULL
15   ORDER BY day;

```

Results Grid:

day	Trip_count
1	62667
2	37791
3	83717
4	84833
5	89656
6	55227
7	78220
8	72515
9	84809
10	87584
11	90448
12	70159
13	87636
...	...

Action Output:

Time	Action	Response	Duration / Fetch Time
20 00:49:30	SELECT day (starttime) AS day ,COUNT(*) as trip_count FROM final_project.citibike_trips...	Error Code: 1054. Unknown column 'month' in 'group...	0.0021 sec
27 00:50:09	SELECT day (starttime) AS day ,COUNT(*) as Trip_count FROM final_project.citibike_trips...	30 row(s) returned	16.068 sec / 0.00007 ...

▪ calendar week

SELECT

dayofweek (starttime) AS dayofweek ,COUNT(*) as Trip_count

FROM final_project.citibike_trips

WHERE week(starttime) = 36 and year(starttime) = 2019

GROUP BY dayofweek

HAVING dayofweek is NOT NULL

ORDER BY dayofweek;

Table: citibike_trips

```

    ORDER BY month;
8   SELECT
9    day (starttime) AS day ,COUNT(*) as Trip_count
10   FROM final_project.citibike_trips
11  WHERE month(starttime) = 09 and year(starttime) = 2019
12  GROUP BY day
13  HAVING day is NOT NULL;
14  ORDER BY day;
15
100%  ◇ 1:8

```

Result Grid | Filter Rows: Search Export: Result Grid | Form Editor | Field Types | Read Only

day	Trip_count
1	62667
2	37791
3	83717
4	84833
5	89656
6	55227
7	78220
8	72515
9	84809
10	87584
11	90448
12	70159
13	87636
...	...

Action Output ◇

Time	Action	Response	Duration / Fetch Time
26 00:49:30	SELECT day (starttime) AS day ,COUNT(*) as Trip_count FROM final_project.citibike_trips...	Error Code: 1054. Unknown column 'month' in 'group...	0.0021 sec
27 00:50:09	SELECT day (starttime) AS day ,COUNT(*) as Trip_count FROM final_project.citibike_trips...	30 row(s) returned	16.068 sec / 0.00007...

Examining trips:-

Table: citibike_trips

```

    num_of_round_trips usertype
35   SELECT COUNT(bikeid) as num_of_round_trips,usertype
36   FROM final_project.citibike_trips
37
38  WHERE start_station_name = end_station_name
39  group by usertype
40
100%  ◇ 1:35

```

Result Grid | Filter Rows: Search Export: Result Grid | Form Editor | Field Types | Read Only

num_of_round_trips	usertype
529196	Customer
856462	Subscriber
...	...

Action Output ◇

Time	Action	Response	Duration / Fetch Time
39 01:27:07	SELECT month (starttime) AS month,COUNT(*) as Trip_count FROM final_project.citibike_trips...	12 row(s) returned	21.851 sec / 0.00051...
40 01:54:58	SELECT COUNT(bikeid) as num_of_round_trips,usertype FROM final_project.citibike_trips...	2 row(s) returned	19.715 sec / 0.00077...

starttime stoptime

```

    average_round
45   SELECT avg(tripduration) as average_round
46   FROM final_project.citibike_trips
47
48  WHERE start_station_name = end_station_name;
49
100%  ◇ 45:48

```

Result Grid | Filter Rows: Search Export: Result Grid | Form Editor | Field Types | Read Only

average_round
2003.5114
...

Action Output ◇

Time	Action	Response	Duration / Fetch Time
42 02:01:37	SELECT COUNT(bikeid) as num_of_non_round_trips,usertype FROM final_project.citibike_trips...	Error Code: 2013. Lost connection to MySQL server d...	30.006 sec
43 02:04:02	SELECT avg(tripduration) as average_round FROM final_project.citibike_trips WHERE start_...	1 row(s) returned	17.961 sec / 0.00069...

Query Completed

4. Analysis of length of trip

Object Info Session

Table: citibike_trips

Columns:

- start_station_name text
- gender text
- start_station_longitude double
- bikeid bigint
- usertype text
- starttime datetime
- stoptime datetime
- start_station_latitude double
- end_station_latitude double
- birth_year bigint
- start_station_id bigint
- tripduration bigint
- end_station_name text
- end_station_longitude double
- end_station_id bigint

```

49
50 • select avg(tripduration) as avgtrip from final_project.citibike_trips;
51
52 • select avg(tripduration) as avgtrip, usertype from final_project.citibike_trips
100% ◇ 34:47

```

Result Grid Filter Rows: Search Export: Result Grid

avgtrip

	avgtrip
▶	1132.4765

Result 12

Action Output Time Action Response Duration / Fetch Time

✖ 44	02:09:33	select avg(tripduration) as avgtrip, usertype from final_project.citibike_trips group by 2 LIMIT 0, 1000	Error Code: 2013. Lost connection to MySQL server d... 30.008 sec
✔ 45	02:11:16	select avg(tripduration) as avgtrip from final_project.citibike_trips LIMIT 0, 1000	1 row(s) returned 15.221 sec / 0.00002...

Query Completed

1. Trips less than 15 minutes

select Count(tripduration) as tripslessthan15 from final_project.citibike_trips

where tripduration <900;

Object Info Session

Table: citibike_trips

Columns:

- start_station_name text
- gender text
- start_station_longitude double
- bikeid bigint
- usertype text
- starttime datetime
- stoptime datetime
- start_station_latitude double
- end_station_latitude double
- birth_year bigint
- start_station_id bigint
- tripduration bigint
- end_station_name text
- end_station_longitude double
- end_station_id bigint

```

54
55 • select Count(tripduration) as tripslessthan15 from final_project.citibike_trips
56 where tripduration <900;
100% ◇ 25:56

```

Result Grid Filter Rows: Search Export: Result Grid

tripslessthan15

	tripslessthan15
▶	25580598

Result 13

Action Output Time Action Response Duration / Fetch Time

✔ 45	02:11:16	select avg(tripduration) as avgtrip from final_project.citibike_trips LIMIT 0, 1000	1 row(s) returned 15.221 sec / 0.00002...
✔ 46	02:21:16	select Count(tripduration) as tripslessthan15 from final_project.citibike_trips where tripdura... 1 row(s) returned	14.895 sec / 0.00059...

2. Trip duration greater than 15 minutes and less than an hour

select Count(tripduration) as tripduration from final_project.citibike_trips

where tripduration >900 and tripduration <3600;

Object Info Session

Table: citibike_trips

Columns:

- start_station_name text
- gender text
- start_station_longitude double
- bikeid bigint
- usertype text
- starttime datetime
- stoptime datetime
- start_station_latitude double
- end_station_latitude double
- birth_year bigint
- start_station_id bigint
- tripduration bigint
- end_station_name text
- end_station_longitude double
- end_station_id bigint

```

57
58 • select Count(tripduration) as tripduration from final_project.citibike_trips
59 where tripduration >900 and tripduration <3600;
100% ◇ 48:59

```

Result Grid Filter Rows: Search Export: Result Grid

tripduration

	tripduration
▶	14758173

Result 14

Action Output Time Action Response Duration / Fetch Time

✔ 46	02:21:16	select Count(tripduration) as tripslessthan15 from final_project.citibike_trips where tripdura... 1 row(s) returned	14.895 sec / 0.00059...
✔ 47	02:23:55	select Count(tripduration) as tripduration from final_project.citibike_trips where tripduration... 1 row(s) returned	15.400 sec / 0.00006...

Query Completed

3. Trip duration more than 1 hour

```
select Count(tripduration) as tripdurationmorethan1 from final_project.citibike_trips
where tripduration >3600;
```

The screenshot shows the MySQL Workbench interface with the following details:

- Object Info:** Shows the session and the table being used: `citibike_trips`.
- Session:** Shows the current session information.
- Query:**

```
60
61 • select Count(tripduration) as tripdurationmorethan1 from final_project.citibike_trips
62 where tripduration >3600;
```
- Result Grid:** Displays the result of the query, which is a single row: `643753`.
- Action Output:** Shows the execution history with two entries:
 - Entry 48: Time 02:28:56, Action: select Count(tripduration) as tripdurationmorethan1 from 'healthy-genre330203.damg_dats...', Response: Error Code: 1146. Table 'anshita.healthy-genre33020...', Duration / Fetch Time: 0.014 sec
 - Entry 49: Time 02:29:13, Action: select Count(tripduration) as tripdurationmorethan1 from final_project.citibike_trips where ..., Response: 1 row(s) returned, Duration / Fetch Time: 14.238 sec / 0.00040...
- Status:** Query Completed.

5. Analysis of gender

1. year over year trend

```
select year(starttime) as year, gender
from final_project.citibike_trips
where gender in ('male','female');
```

The screenshot shows the MySQL Workbench interface with the following details:

- Object Info:** Shows the session and the table being used: `citibike_trips`.
- Session:** Shows the current session information.
- Query:**

```
75 • select year(starttime) as year, gender
76   from final_project.citibike_trips
77  where gender in ('male','female');
```
- Result Grid:** Displays the result of the query, showing the count of trips for each gender per year:

year	gender	Count
2019	male	10
2019	male	10
2019	female	10
2019	male	10
2019	male	10
2019	male	10
2019	female	10
2019	female	10
2019	male	10
- Action Output:** Shows the execution history with two entries:
 - Entry 57: Time 02:45:22, Action: select gender,usertype, avg(tripduration) as avgduration from final_project.citibike_trips where gender in ('male','female')..., Response: Error Code: 2013. Lost connection to MySQL server during query., Duration / Fetch Time: 30.005 sec
 - Entry 58: Time 02:50:10, Action: select year(starttime) as year, gender from final_project.citibike_trips where gender in ('male','female') LIMIT 0,1000, Response: 1000 row(s) returned, Duration / Fetch Time: 0.0065 sec /