BIKE SHARING ASSIGNMENT

Note: For some questions in tableau my system was not able to extract the full data. So the related answers have been given

Task 1

1.1 total number of stations:

SQL query: SELECT count(id) as total_number_of_stations from station;

Ans: 70

	total_number_of_stations
1	70

1.2 total number of bikes:

SQL query: SELECT count(DISTINCT bike_id) as toatl_number_of_bikes from trip;

Ans: 700

	toatl_number_of_bikes
1	700

1.3 total number of trip:

SQL query: SELECT count(id) as total_number_of_trips from trip;

Ans: 669959

	total_number_of_trips
1	669959

2. Construct a geographical plot to show the location of each bike station using the latitude and longitude provided under the Station table

Tableau link:

https://public.tableau.com/app/profile/lokeshwaran2939/viz/stationsacrossvariouscities/Sheet1

3.What is the relationship between the following columns (one to one, many to one, many to many)?

3.1 bike_id (Trip table) and start_station_id (Trip table):
Ans:

```
use bike_sharing_dataset;

WITH CTEA AS (
  SELECT COUNT(DISTINCT start_station_id) ac
  FROM trip
  GROUP BY bike_id
),
CTEB AS (
  SELECT COUNT(DISTINCT bike_id) bc
  FROM trip
  GROUP BY start_station_id
)
SELECT CONCAT(
        CASE WHEN MAX(bc) = 1 THEN '1' ELSE 'many' END,
        ' to ',
        CASE WHEN MAX(ac) = 1 THEN '1' ELSE 'many' END
      ) as Relationship
FROM CTEA
CROSS JOIN CTEB;
```

	Relationship
▶	many to many

Many to Many relationship is there between the columns bike_id and start_station_id in the Trip table

3.2 pincode (Weather table) and station location (latitude and longitude in Station table):
Ans: No relationship exists directly between both the tables and so there is no direct relationship between these columns

3.3 8/29/2013 (date column in Weather table) and mean wind speed (Weather table)
Ans: use bike_sharing_dataset;

```
 WITH CTE1 AS (
 SELECT COUNT(DISTINCT mean_wind_speed_mph) ac
 FROM weather
 where date= 8/29/2013
 GROUP BY date
),
CTE2 AS (
 SELECT COUNT(DISTINCT date) bc
 FROM weather
 where date= 8/29/2013
 GROUP BY mean_wind_speed_mph
)
SELECT CONCAT(
      CASE WHEN MAX(bc) = 1 THEN '1' ELSE 'many' END,
      ' to ',
      CASE WHEN MAX(ac) = 1 THEN '1' ELSE 'many' END
     ) as Relationship
FROM CTE1
CROSS JOIN CTE2;
```

	Relationship
▶	many to many

Many to many relationship exists between the date 8/29/2013 and mean wind speed as per the above code

4. Find the first and the last trip in the data.
SQL query:
SELECT id as first_trip from trip order by start_date ASC LIMIT 1
SELECT id as last_trip from trip order by start_date DESC LIMIT 1
Ans: First trip = 139545 and second trip = 444343

	first_trip
1	139545

	last_trip
1	444343

5

5.1 What is the average duration Of all the trips?

SQL query: select avg(duration) from trip;

Ans : 1107.94984618462

	average
1	1107.94984618462

5.2 What is the average duration of trips on which customers are ending their rides at the same station from where they started?

SQL query: select avg(duration) from trip where start_station_name = end_station_name

Ans : 6357.40110921146

	avg(duration)
1	6357.40110921146

6. Which bike has been used the most in terms of duration?

SQL query: SELECT bike_id, sum(duration) as dur from trip GROUP by bike_id order by dur DESC LIMIT 1

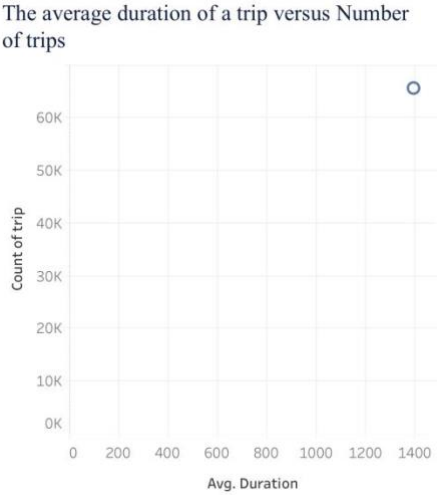
Ans : 535

	bike_id	dur
1	535	18611693

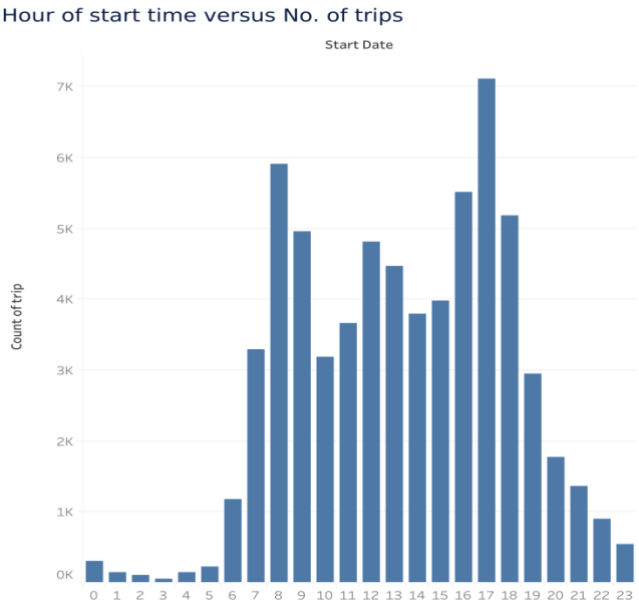
7. Plot the most suitable graph for the following:

7.1 The average duration of a trip versus Number of trips

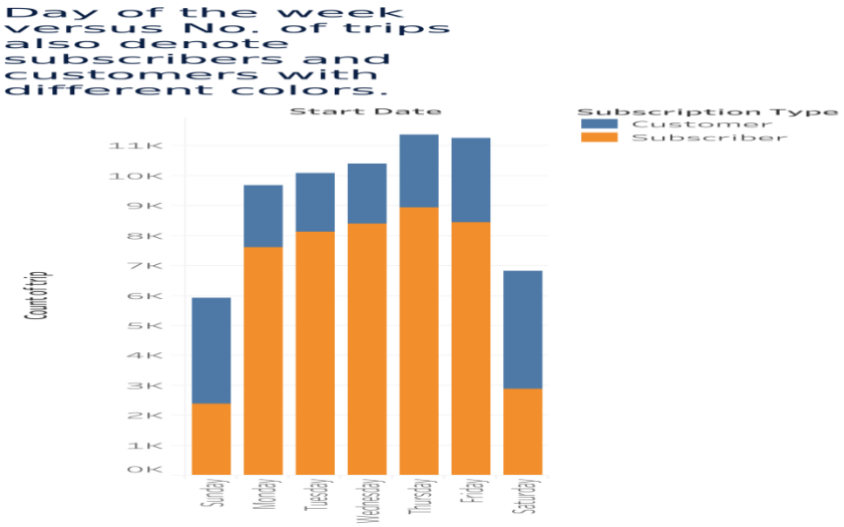
Tableau link: https://public.tableau.com/app/profile/lokeshwaran2939/viz/Task1-qus7_1/Sheet1



7.2 Hour of start time versus No. of trips
Tableau link: https://public.tableau.com/app/profile/lokeshwaran2939/viz/Task1-qus7_2/Sheet2



7.3 Day of the week versus No. of trips also denote subscribers and customers with different colors.
Tableau link: https://public.tableau.com/app/profile/lokeshwaran2939/viz/Task1-qus7_3/Sheet3



----------end of task 1----------

Task 2:

1. What are the top 10 least popular stations?

SQL query: select start_station_name as least_popular_stations,count(start_station_name) as frequency
from trip group by start_station_name order by frequency LIMIT 10;
Ans :
Top 10 least popular stations
1. San Jose Government Center
2. Broadway at Main
3. Redwood City Public Library
4. Franklin at Maple
5. San Mateo County Center
6. Redwood City Medical Center
7. Mezes Park
8. Stanford in Redwood City
9. Park at Olive
10.Santa Clara County Civic Center

	least_popular_stations	frequency
1	San Jose Government Center	23
2	Broadway at Main	67
3	Redwood City Public Library	213
4	Franklin at Maple	224
5	San Mateo County Center	287
6	Redwood City Medical Center	311
7	Mezes Park	341
8	Stanford in Redwood City	436
9	Park at Olive	750
10	Santa Clara County Civic Center	840

2. Find the idle time for Station 2 on the date '2013/08/29'
SQL query: select time as idle_time from status where station_id = 2 and bikes_available > 2 and time BETWEEN '2013/08/29 00:00:00' and '2013/08/29 23:59:59';
Ans :
2013/08/29 18:00:02
2013/08/29 19:00:02
2013/08/29 23:00:01

	idle_time
1	2013/08/29 18:00:02
2	2013/08/29 19:00:02
3	2013/08/29 23:00:01

3.Find the distance between consecutive stations

SQL query:

```
select *,
acos(
cos(radians( st.lat ))
* cos(radians( st.lead_lat ))
* cos(radians( st.long ) - radians( st.lead_long ))
+ sin(radians( st.lat ))
* sin(radians( st.lead_lat ))
) AS consecutiveStationDistance from (select *,
LEAD(station.lat) OVER(ORDER BY station.id) as lead_lat,
LEAD(station.long) OVER(ORDER BY station.id ) as lead_long
from station) AS st;
```

Ans :

id	name	lat	long	dock_c	city	installation	lead_lat	lead_long	consecutiveStationDistance
2	San Jose Diridon Caltrain Station	37.329732	-121.90178...	27	San Jose	8/6/2013	37.330698	-121.888979	0.00017847881130444583
3	San Jose Civic Center	37.330698	-121.888979	15	San Jose	8/5/2013	37.333988	-121.894902	0.00010026772911804018
4	Santa Clara at Almaden	37.333988	-121.894902	11	San Jose	8/6/2013	37.331415	-121.8932	0.00005074008181264947 5
5	Adobe on Almaden	37.331415	-121.8932	19	San Jose	8/5/2013	37.3367...	-121.894074	0.00009339805110226917
6	San Pedro Square	37.33672...	-121.894074	15	San Jose	8/7/2013	37.333798	-121.8869...	0.00011133402167923283
7	Paseo de San Antonio	37.333798	-121.88694...	15	San Jose	8/7/2013	37.330165	-121.8858...	0.00006525870274853827
8	San Salvador at 1st	37.330165	-121.88583...	15	San Jose	8/5/2013	37.348742	-121.8947...	0.00034687505274667 67
9	Japantown	37.348742	-121.89471...	15	San Jose	8/5/2013	37.337391	-121.886995	0.00022521819800875196
10	San Jose City Hall	37.337391	-121.886995	15	San Jose	8/6/2013	37.335885	-121.8856...	0.00003215714087476411
11	MLK Library	37.335885	-121.88566...	19	San Jose	8/6/2013	37.332808	-121.8838...	0.00005904867487122153 7
12	SJSU 4th at San Carlos	37.332808	-121.88389...	19	San Jose	8/7/2013	37.339301	-121.8899...	0.00014100220149727307
13	St James Park	37.339301	-121.88993...	15	San Jose	8/6/2013	37.332692	-121.900084	0.00018202399531240104
14	Arena Green / SAP Center	37.332692	-121.900084	19	San Jose	8/5/2013	37.3339...	-121.877349	0.00031627357873038 14

4.Use the findings above to recommend three stations that can be shut.

Ans :

Based on my findings from the above data I would recommend the San Jose Government Center, Broadway at Main and Redwood City Public Library I am recommending the San Jose Government center and Broadway at Main stations as they are the least popular and less frequently used compared to other stations. And In case of Redwood City Public Library though it is one of least used station it is also near to another least popular station Redwood City Medical Center. So having two least popular stations nearby is not advisable. So its better to shut the least popular of two i.e Redwood City Public Library

----------end of task 2----------

Task 3: Optimising Operations

Throughout the day, bikes keep moving around the city due to the trips. Zulip has to find out how to effectively move bikes around to ensure the demand is met with adequate supply. This is to ensure that at any time, there are sufficient bikes available at a given station. Here are some points that you will have to consider while deciding on the transportation of bikes from one place to another:

1. Calculate the average number of bikes and docks available for Station 2. (Hint: Use the Status table.)

SQL query:

```
select avg(bikes_available) as average_number_of_bikes_available, avg(docks_available) as
average_number_of_docks_available
from status where station_id=2;
```

Ans :

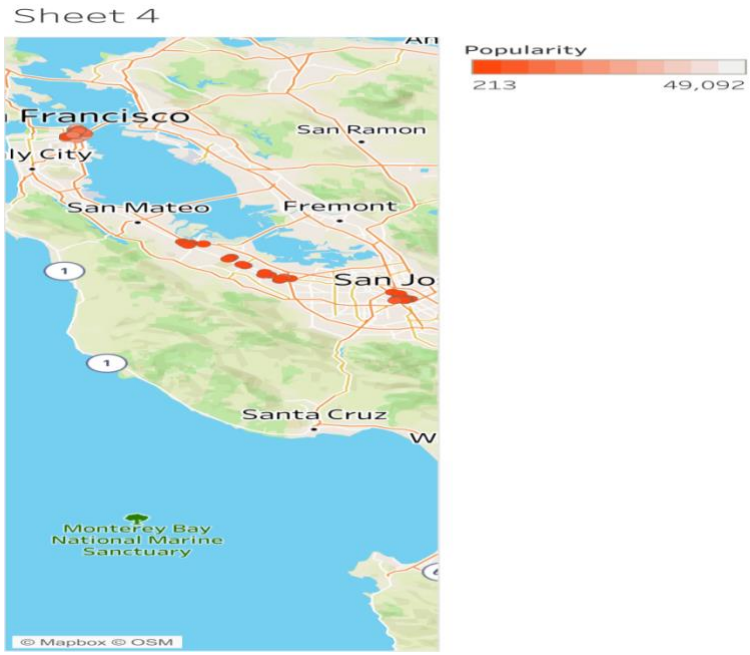
average_number_of_bikes_available : 10.831

average_number_of_docks_available : 16.169

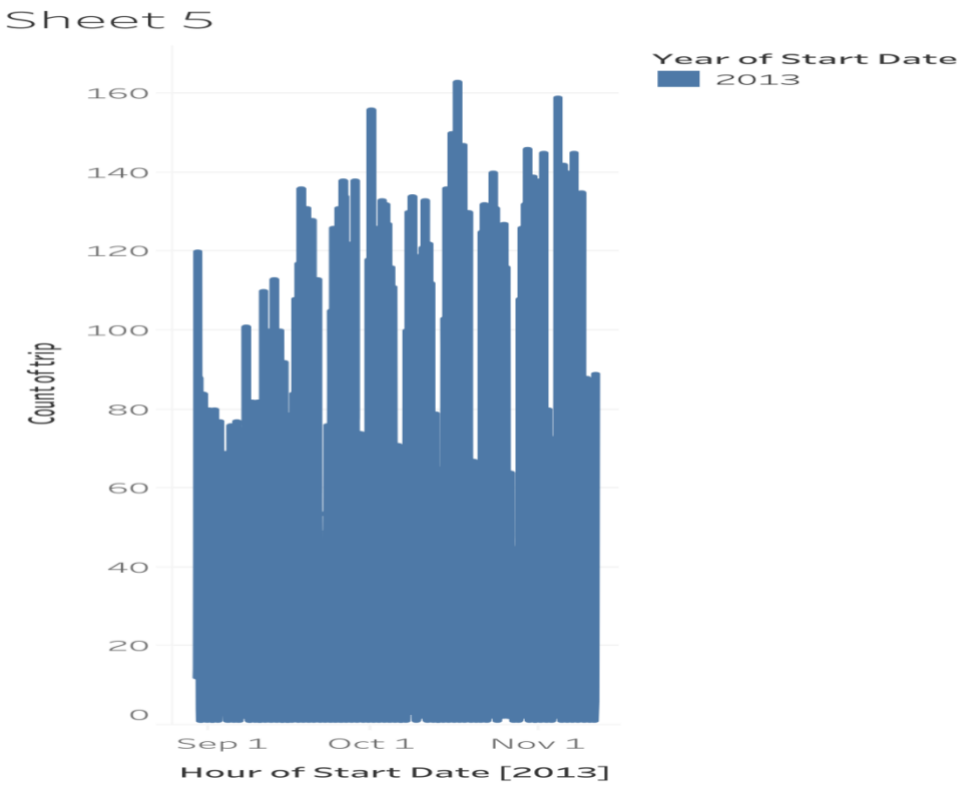
	average_number_of_bikes_available	average_number_of_docks_available
1	10.831	16.169

2.Plot the popularity of each station on a map for subscribers and customers.

Tableau link: https://public.tableau.com/app/profile/lokeshwaran2939/viz/Task3-qus2/Sheet4



3. Plot the number of trips per hour for all the data provided in the Trip table.
Tableau link: https://public.tableau.com/app/profile/lokeshwaran2939/viz/Task3-qus3/Sheet5



4.Use the findings above to provide insights on how to optimize operations. (open ended)
Ans: From the findings above we can find the average no. of bikes and docks available. This shows that there is irregular amount of bikes and docks available across each station. This might happen in cases where a traveller takes a bike from a popular station to a less popular one. Now what happens is the number of bikes available at the lesser popular station will be higher than it's required availability on the other hand the bike availability of the popular station might get affected. So to optimise this we can implement the above findings to those popular stations along with a data of the no.of trips made between that station and other stations. This give a better understanding of what's the bike availability of a station and from which station it can be fullfilled.

----------end of task 3----------

Task 4: Couple Bikes? (Bonus)
Zulip has decided to start a new product line called Couple Bikes. This will enable two persons to travel from one station to another at the same time. What are some of the factors that you will have to consider while validating the idea of couple bikes?

Ans:
The idea of couple bikes is really an amazing one. It actually helps two person to share a bike, to Travel along the same route. This actually helps the travellers to travel at nearly an half rate as it is as the bike as been shared between two of them. Similarly it will reduce the operation cost for zulip and provides more available bikes and docks for customers. The most important factor I will consider is safety of both the traveller.In some cases two of them might be strangers, so to avoid any uncertain events between those travellers I suggest to provide this feature only to the subscribers and not all customers. As incase of subscribers we have much data of them so that we can avoid conflicts. Also In terms of business, providing an offer like, that subscribers can travel for nearly half the cost will obviously increase the numbers of subscribers. The second factor that I consider is the price per travel, As we are reducing the number of bikes per travel it should not reduce the profit. So the price fixed can be explained with an example as follows. Consider that previously a person pays 2$/km out of 1-1.5$ will be used for fuel and maintenance , the remaining 0.5$ is the profit per km. So now as the travellers share the bike the should be 1.2$/Km per traveller. Now the over all amount will 2.4$ out which 1-1.5$ goes for fuel and maintenance , the remaining 1.5-1$ will be the profit per bike. So implementing this gives the profit of two bikes in one bike and there will also be more number of subscribers.

----------The End----------