# **Task 1: Get to Know Your Company**

## What are the total numbers of:

### Bike stations?

select count(distinct(id)) as bike\_station from station;

Graphical user interface, application, Word

Description automatically generated

### Bikes?

select sum(bikes\_available) from status;

Graphical user interface, text, application, Word

Description automatically generated

### Trips?

select count(\*) from trip;

Graphical user interface

Description automatically generated with low confidence

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

* <https://public.tableau.com/app/profile/priya3968/viz/Assignment_2_16467134251980/bike_station_alloction?publish=yes>

Map

Description automatically generated

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

### bike\_id (Trip table) and start\_station\_id (Trip table)

* one (bike\_id) to many (Start\_station\_id)

Graphical user interface, application

Description automatically generated

### pincode (Weather table) and station location (latitude and longitude in Station table)

* no relationship in both table

Table

Description automatically generated

### 8/29/2013 (date column in Weather table) and mean wind speed (Weather table)

* Many(mean wind speed) to One (date column: 8/29/2013)

Graphical user interface, text, application

Description automatically generated

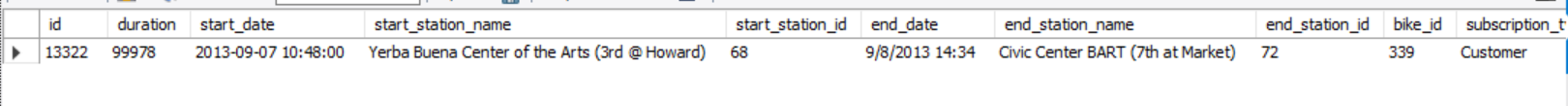
## Find the first and the last trip in the data.

select \*

from trip

order by end\_date desc

limit 1;

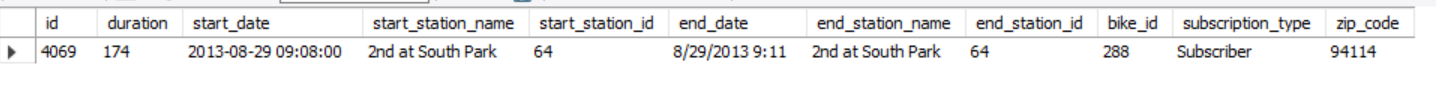


select \*

from trip

order by start\_date asc

limit 1;



## What is the average duration:

### Of all the trips?

select avg(duration)

from trip;

Graphical user interface, text, application

Description automatically generated

### Of trips on which customers are ending their rides at the same station from where they started?

select avg(duration)

from trip

where start\_station\_id = end\_station\_id;

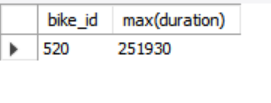
Graphical user interface

Description automatically generated with medium confidence

## Which bike has been used the most in terms of duration? (Answer with the Bike ID)

select bike\_id, max(duration)

from trip;



## Plot the most suitable graph for the following:

### The average duration of a trip versus Number of trips

Chart

Description automatically generated with medium confidence

### Hour of start time versus No. of trips

Chart, line chart

Description automatically generated

### Day of the week versus No. of trips also denote subscribers and customers with different colors.

Graphical user interface, chart, application, line chart

Description automatically generated

Link:

<https://public.tableau.com/app/profile/priya3968/viz/Assignment_21_16470777331820/Sheet1?publish=yes>

# **Task 2: Demand Prediction**

Zulip is running under a loss and has decided to shut operations for three of its stations. You have to use the data provided to help Zulip decide which three stations should be shut.

## What are the top 10 least popular stations? Hint: Find the least frequently appearing start stations from the Trip table.

Graphical user interface, application

Description automatically generated

## 2. Idle time is the duration for which a station remains inactive. You can consider this as the time for which a station has more than 3 bikes available.

### (i) Find the idle time for Station 2 on the date '2013/08/29'

No station is idle on date ‘2013/08/29’

Graphical user interface, text, application

Description automatically generated

## 3. In case two stations are nearby, it might be possible to shut one down. Find the distance between consecutive stations (between Stations 1 and 2, Stations 2 and 3, and so on). The Haversine formula 2arcsin(√sin2(ϕ2−ϕ12)+(1−sin2(ϕ2−ϕ12)−sin2(ϕ2+ϕ12).sin2(λ2−λ12)) is used to find the distance between two points on a sphere given their longitude and latitude. (ϕ1,λ1) is the latitude–longitude pair for the first station, and (ϕ2,λ2)is the latitude–longitude pair for the second station. The code can be used to find the distance between two consecutive stations.

Graphical user interface, application

Description automatically generated

Completed \_result:

[assignment%202/distance\_station.csv](https://oneharman-my.sharepoint.com/personal/priya_kumari3_harman_com/Documents/Desktop/personal%20doc/assignment%202/distance_station.csv)

## Use the findings above to recommend three stations that can be shut. (open ended) For example, if the Japantown and Ryland stations are nearby, and the Japantown is not as popular as the Ryland station, then it can be recommended to shut.

Answer:

|  |
| --- |
| 1. Ryland Park |
| 1. Castro Street and El Camino Real |
| 1. San Antonio Shopping Center |

As there are generating least revenue.

query to order the station based on usability

with least\_popular as(select start\_station\_id,start\_station\_name,count(start\_station\_id) as cn

from trip

group by start\_station\_id)

select id,name,cn,

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

left join least\_popular as t

on st.id = t.start\_station\_id

order by cn,consecutiveStationDistance;

* Find there 6 station which not generating revenue, I selected top three

Graphical user interface, application

Description automatically generated

# **Task 3: Optimizing 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:

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

Answer : average number of bikes and docks available for Station 2 -- 0.83949490

Query

select station\_id,avg(bikes\_available/docks\_available)

from status

group by station\_id;

Output:

Graphical user interface, text, application

Description automatically generated

## Plot the popularity of each station on a map for subscribers and customers. (Hint: Popular stations appear most frequently under the column start\_station\_name in the Trip table)

Chart, scatter chart

Description automatically generated

Sheet 8 of

<https://public.tableau.com/app/profile/priya3968/viz/Assignment_22_16471357336840/Sheet8?publish=yes>

## Plot the number of trips per hour for all the data provided in the Trip table.

Chart, bar chart

Description automatically generated

<https://public.tableau.com/app/profile/priya3968/viz/Assignment_23_16471358019280/Sheet4?publish=yes>

## Use the findings above to provide insights on how to optimize operations. (open ended)

Chart

Description automatically generated

<https://public.tableau.com/app/profile/priya3968/viz/Assignment_23_16471358019280/Dashboard1?publish=yes>

From dashboard we can derive below observation:

- Max Ride is used at 4:00 am in the morning

- There are peak up hour for bike ride morning(4 a.m - 9 a.m ) and evening (4-6)

- Station 2nd at Folsom is maximum used station

Based on observation we optimize operations are:

* cyclists earn points by renting or returning bikes at certain high-need stations(to attend supply chain demand)
* We can make sure that all bikes are proper functioning that peak-up hour

# **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?

Factor that I will have consider while validating the idea of couple bike:

* Both have Start station ID and Start time must have same (or difference 5 mins)
* End station id, I will decided on based of :

1. Whose ride is longer
2. Person whose ride longer are willing to drop the couple person in middle.

* Their must be rating system about rider so that other chose their partner.

# Task 5: Presentation

The senior data scientist has asked you to prepare a PowerPoint presentation for the higher management. Your task is to summarize and present your findings from Task 2 to Task 4.