

Task 1: Get to Know Your Company

1. What are the total numbers of:
 1. Bike stations?
 2. Bikes?
 3. Trips?
2. Construct a geographical plot to show the location of each bike station using the latitude and longitude provided under the Station table.
3. What is the relationship between the following columns (one to one, many to one, many to many)?
 1. bike_id (Trip table) and start_station_id (Trip table)
 2. pincode (Weather table) and station location (latitude and longitude in Station table)
 3. 8/29/2013 (date column in Weather table) and mean wind speed (Weather table)
4. Find the first and the last trip in the data.
5. What is the average duration:
 1. Of all the trips?
 2. Of trips on which customers are ending their rides at the same station from where they started?
6. Which bike has been used the most in terms of duration? (Answer with the Bike ID)
7. Plot the most suitable graph for the following:
 1. Average duration of a trip versus Number of trips
 2. Hour of start time versus No. of trips
 3. Day of the week versus No. of trips also denote subscribers and customers with different colors.

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.

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

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

$2 \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \left(1 - \sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) - \sin^2\left(\frac{\phi_2 + \phi_1}{2}\right) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)\right)}\right)$ is used

to find the distance between two points on a sphere given their longitude and latitude. (ϕ_1, λ_1) is the latitude-longitude pair for the first station, and (ϕ_2, λ_2) is the latitude-longitude pair for the second station. You can find the SQL code for this formula [here](#).

4. 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.

SQL Code link for the formula used.....

[Distance between Latitude and Longitude Coordinates in SQL \(daynebatten.com\)](https://daynebatten.com/distance-between-latitude-and-longitude-coordinates-in-sql/)

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:

1. Calculate the average number of bikes and docks available for Station 2 and Station 3 (Hint: Use the Status table.)
2. 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)
3. Plot the number of trips per hour for all the data provided in the Trip table.
4. Use the findings above to provide insights on how to optimize operations. (open ended)

Task 4: Couple Bikes? (Bonus)

5. 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?

Task 5: Presentation

6. 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.

About Data Tables:

Station Table

| Station ID | Station Name | Latitude | Longitude | Dock Count | City | Installation Date |
|------------|--------------|----------|-----------|------------|------|-------------------|
|------------|--------------|----------|-----------|------------|------|-------------------|

Status Table

| Station ID | Bikes Available | Docks Available | Time |
|------------|-----------------|-----------------|------|
|------------|-----------------|-----------------|------|

Trip Table

| Station ID | Duration | Start Date | Start Station Name | Start Station ID | End Date | End Station Name | End Station ID | Bike ID | Subscription Type |
|------------|----------|------------|--------------------|------------------|----------|------------------|----------------|---------|-------------------|
|------------|----------|------------|--------------------|------------------|----------|------------------|----------------|---------|-------------------|

Weather Table

| Temperature | Humidity | Dew Point | Date | Pincode |
|-------------|----------|-----------|------|---------|
|-------------|----------|-----------|------|---------|