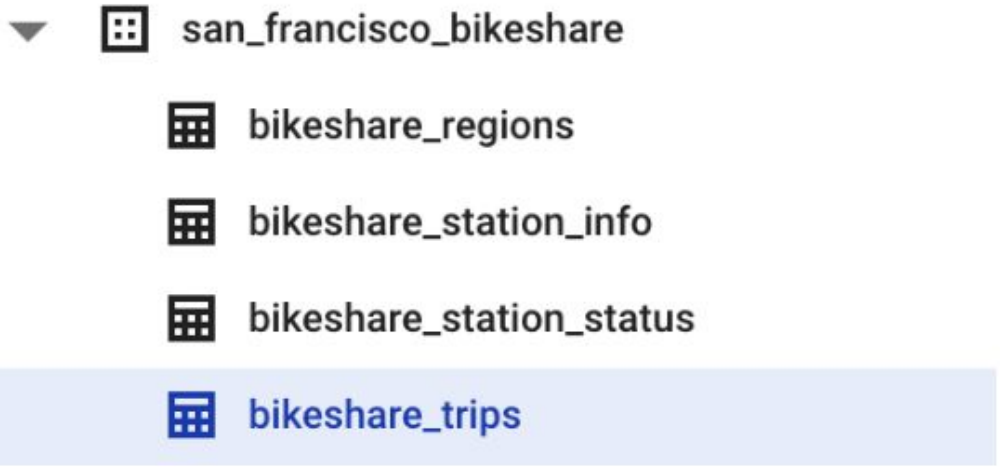
Intermediate Assignment

By Christopher Stephen

https://console.cloud.google.com/bigquery?sq=608087215512:fd6ab502176e4b8b9f0a10182e3ac031

Overview of dataset

San Francisco Ford GoBike, managed by Motivate, provides the Bay Area's bike share system. Bike share is a convenient, healthy, affordable, and fun form of transportation. It involves a fleet of specially designed bikes that are locked into a network of docking stations. Bikes can be unlocked from one station and returned to any other station in the system. People use bike share to commute to work or school, run errands, get to appointments, and more. The dataset contains trip data from 2013-2018, including start time, end time, start station, end station, and latitude/longitude for each station.



Then, try to answer the following:

1. **Create a query to get the average amount of duration (in minutes) per month (Skillset: Basic SQL & Formatting and Cleaning in SQL)**

   please use the start date from 2014-2017

   Expected output:

   - Month
   - Average (in minute)

```sql
-- Question 1
SELECT
--EXTRACT (MONTH FROM start_date) AS month,
EXTRACT (YEAR FROM start_date) AS year,
AVG(duration_sec) / 60 AS avg FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
WHERE EXTRACT (YEAR FROM start_date) BETWEEN 2014 AND 2017
GROUP BY 1
ORDER BY 1 ASC;
```

Press Alt+

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	year	avg
1	2014	18.866123315529695
2	2015	15.68234570682363
3	2016	13.816352008538596
4	2017	18.316825347957046

2. **Create a query to get total trips and total number of unique bikes grouped by region name (Skillset: Basic SQL & Joins)**

please use the start date from 2014-2017

Expected output:

- Region Name
- Total Trips
- Total Bikes

```sql
-- Question 2

SELECT
EXTRACT (YEAR FROM d.start_date) AS year,
a.name AS region_name,
COUNT(d.trip_id) AS total_trips,
COUNT(distinct c.num_bikes_available) AS total_bikes
FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_regions` AS a
JOIN `bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info` AS b on b.region_id = a.region_id
JOIN `bigquery-public-data.san_francisco_bikeshare.bikeshare_station_status` AS c on c.station_id = b.station_id
JOIN `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips` AS d on b.name = d.start_station_name
WHERE EXTRACT (YEAR FROM d.start_date) BETWEEN 2014 AND 2017
GROUP BY 1,2
ORDER BY 1 ASC;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	year	region_name	total_trips	total_bikes
1	2014	San Jose	3957	3
2	2015	San Jose	2797	3
3	2016	San Jose	1855	3
4	2016	San Francisco	69	1
5	2017	Oakland	64530	23
6	2017	Berkeley	14548	16
7	2017	Emeryville	3578	5

3. **Find the youngest and oldest age of the members, for each gender. Assume this year is 2022. (Skillset: Basic SQL & SQL CTE)**

Expected output:

- Gender
- Youngest Age
- Oldest Age

```sql
WITH table1 AS (
    SELECT
    (2022 - member_birth_year) AS umur,
    member_gender
    FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
    GROUP BY 1,2
)
SELECT
MIN(umur) AS youngest_age,
MAX(umur) AS oldest_age,
member_gender
FROM table1 as a
WHERE umur is not null
GROUP BY 3;
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	youngest_age	oldest_age	member_gender
1	22	136	Male
2	22	122	Female
3	22	122	Other

4. **Get the latest departure trip in each region with detail below**

   (Skillset: Window functions, SQL CTE)

   a. trip_id
   b. duration_sec
   c. start_date
   d. start_station_name
   e. Member_gender

```sql
-- Question 4
WITH temp1 AS(
    SELECT
    a.name AS region_name,
    a.region_id AS region_id,
    b.station_id AS station_id,
    b.name AS station_name
    FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_regions` AS a
    JOIN `bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info` AS b on a.region_id = b.region_id
),
temp2 AS(
    SELECT
    trip_id,
    start_station_id,
    duration_sec,
    start_date,
    start_station_name,
    member_gender
    member_gender
    FROM`bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`


)
SELECT
--MAX(temp2.start_date) OVER (PARTITION BY temp1.region_id ORDER BY temp1.region_id) AS date,
EXTRACT(YEAR FROM temp2.start_date) AS start_date,
temp2.trip_id AS trip_id,
temp2.duration_sec AS duration_sec,
temp2.start_station_name,
temp2.member_gender,
temp1.region_name
FROM temp1 JOIN temp2 ON temp1.station_name = temp2.start_station_name
WHERE EXTRACT(YEAR FROM temp2.start_date) BETWEEN 2014 AND 2017 AND temp2.member_gender is not null
ORDER BY 1,3 ASC;
```

Query results

SAVE RESULTS ▾ EXPLORE DATA ▾

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	start_date	trip_id	duration_sec	start_station_name	member_gender	region_nar
1	2017	20171119181527.2320001166	61	Valencia St at 21st St	Male	San Franci
2	2017	20170812144918.1540001345	61	West St at 40th St	Female	Oakland
3	2017	20171006125119.3160001385	61	Frank H Ogawa Plaza	Female	Oakland
4	2017	20170826165040.1780002136	61	29th St at Church St	Female	San Franci
5	2017	20170828153818.4460002622	61	S Van Ness Ave at Market St	Male	San Franci

5. **Create a query to get Month to Date of total trips in each region, breakdown by date (Skillset: Basic SQL, Formatting and Cleaning in SQL, Window Function & SQL CTE)**

   please use timeframe from November 2017 until December 2017

   Expected Output:

   - Start Date (in date format)
   - Region Name
   - Total Trips (in cumulative)

```sql
-- Question 5
WITH temp1 AS(
    SELECT
    a.name AS region_name,
    b.station_id AS station_id,
    b.name AS station_name
    FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_regions` AS a
    JOIN `bigquery-public-data.san_francisco_bikeshare.bikeshare_station_info` AS b on a.region_id = b.region_id
),
temp2 AS(
    SELECT
    start_date,
    EXTRACT (YEAR from start_date) AS year,
    EXTRACT (MONTH from start_date) AS month,
    COUNT(trip_id) AS total_trip,
    COUNT(trip_id) AS total_trip,
    start_station_name
    FROM `bigquery-public-data.san_francisco_bikeshare.bikeshare_trips`
    GROUP BY 1,5

)
SELECT
DISTINCT EXTRACT (DATE FROM temp2.start_date) AS date,
temp2.year AS year,
temp2.month AS month,
temp1.region_name AS region_name,
COUNT(total_trip) OVER (PARTITION BY region_name ORDER BY EXTRACT (DATE FROM temp2.start_date) ASC) AS total_trips
FROM temp1 JOIN temp2 ON temp1.station_name = temp2.start_station_name
WHERE EXTRACT (YEAR FROM temp2.start_date) = 2017 AND EXTRACT (MONTH FROM temp2.start_date) BETWEEN 11 AND 12
```

Query results

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS

Row	date	year	month	region_name	total_trips
1	2017-11-01	2017	11	Berkeley	135
2	2017-11-02	2017	11	Berkeley	234
3	2017-11-03	2017	11	Berkeley	326
4	2017-11-04	2017	11	Berkeley	449
5	2017-11-05	2017	11	Berkeley	533
6	2017-11-06	2017	11	Berkeley	649