

## BigQuery

BigQuery, a data warehouse on Google Cloud that I use to query, filter large datasets, aggregate results, and perform complex operations. In this mini project, I will upload public data to my console; and write some simple SQL queries using SELECT, FROM, and WHERE.

- **SELECT** is the section of a query that indicates **what data** you want SQL to return to you
- **FROM** is the section of a query that indicates **which table** the desired data comes from.
- **WHERE** is the section of a query that indicates any **filters** you'd like to apply to your dataset

BigQuery:

Used dataset: London Bicycle Hires dataset from the Greater London Authority - london bicycle. -> london\_bicycles data, table: cycle\_hire

<https://console.cloud.google.com/bigquery?sq=993565623922:99242c2cba374cdeaa9d386bf205a438>

```
SELECT
    end_station_name
FROM `bigquery-public-data.london_bicycles.cycle_hire`
LIMIT
    1000;
```

```
SELECT
    COUNT(*) AS num_of_trips
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
WHERE
    duration >= 12000;
```

```
SELECT
    end_station_name
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
WHERE
    rental_id=57635395;
```

#1. From which station names has bike\_id 1710 started?

```
SELECT
    DISTINCT start_station_name
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
WHERE
    bike_id = 1710;
```

#2. How many bike\_ids have ended at "Moor Street, Soho"?

```
SELECT
    COUNT (DISTINCT bike_id) AS num_of_bikes
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
WHERE
    end_station_name = 'Moor Street, Soho';
```

#3. What is the station\_id for "Canton Street, Poplar"?

```
SELECT
    DISTINCT start_station_id
FROM
    `bigquery-public-data.london_bicycles.cycle_hire`
```

```
WHERE  
start_station_name = 'Canton Street, Poplar';
```

#4. What is the name of the station whose ID is 111?

```
SELECT  
    DISTINCT start_station_name  
FROM  
    `bigquery-public-data.london_bicycles.cycle_hire`  
WHERE  
start_station_id = 111;
```

#5. How many distinct bike\_ids had trip durations greater than 2400 seconds (or 40 minutes)?

```
SELECT  
    COUNT(DISTINCT bike_id) AS num_of_bike_trips  
FROM  
    `bigquery-public-data.london_bicycles.cycle_hire`  
WHERE  
    duration > 2400;
```