London Bicycle Dataset

By: Noah Dunn and Scott Harris

The London Bicycle dataset seeks to analyze the usage of London’s public bicycle offerings that can be rented, used, and dropped off at different stations city-wide, as shown in [this](https://www.visitlondon.com/traveller-information/getting-around-london/london-cycle-hire-scheme) article. The dataset, london\_bicycles contains two tables. The first, cycle\_hire, takes up 2.59 GB in 24369201 rows and focuses on gathering the duration of bike rides, and the corresponding pickup and drop off stations for each ride. The second, cycle\_stations, takes up 76.74KB in 782 rows in focuses more on the stations, providing the details of installation dates, GPS positioning, and the number of available bike docs at each station.

**Dataset:** london\_bicycles

**Table:** cycle\_hire

**Columns:** rental\_id

duration

bike\_id

end\_date

end\_station\_id

end\_station\_name

start\_date

start\_station\_id

start\_station\_name

end\_station\_logical\_terminal

start\_station\_logical\_terminal

end\_station\_priority\_id

**Table:** cycle\_stations

**Columns:** id (foreign key to cycle\_hire’s start/end\_station\_id)

installed

latitude

locked

longitude

name

bikes\_count

docks\_count

nbEmptyDocks

temporary

terminal\_name

install\_date

removal\_date

The specified dataset can be found [here](https://console.cloud.google.com/marketplace/details/greater-london-authority/london-bicycles) as well as [here.](https://data.london.gov.uk/dataset/number-bicycle-hires)

**Query Analysis:**

**Query 1:**

Find the Top 10 start stations in London based on the number of docks. The query should return the start station name, as well as the number of docks.

**SELECT** **DISTINCT** CH.start\_station\_name, CS.docks\_count

**FROM** `bigquery-public-data.london\_bicycles.cycle\_hire` **as** CH

**JOIN** `bigquery-public-data.london\_bicycles.cycle\_stations` **as** CS

**ON** CH.start\_station\_id = CS.id

**ORDER** **BY** CS.docks\_count

**DESC**

**LIMIT** 10



**Size:** 860.37 MB Processed.

**Runtime:** 1.8 seconds

**Query 2:**

Determine the most common start and end location pairs, as well as the number of trips that occured between those two points. The query should return the start station name, the end station name, and the total number of trips.

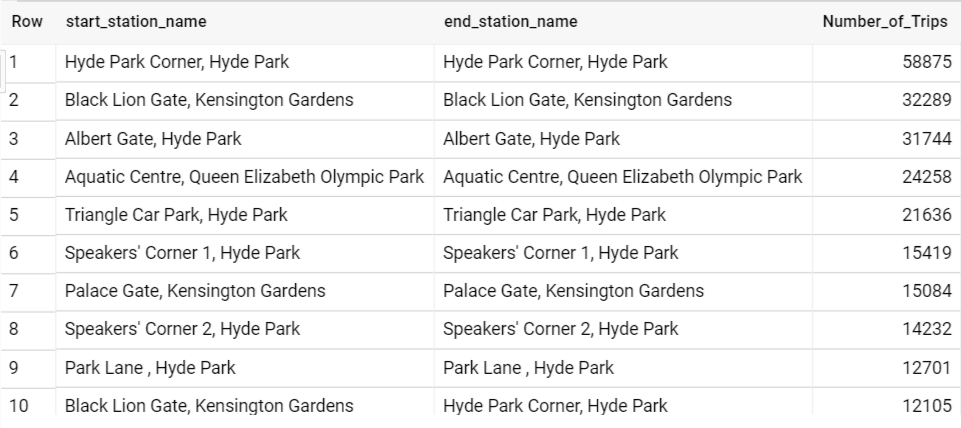
**SELECT** start\_station\_name, end\_station\_name, **COUNT**(end\_station\_name) as Number\_of\_Trips

**FROM** `bigquery-public-data.london\_bicycles.cycle\_hire` **as** CH

**GROUP** **BY** start\_station\_name, end\_station\_name

**ORDER** **BY** **COUNT**(end\_station\_name) **DESC**

**LIMIT** 10



**Size:** 1.3 GB Processed.

**Runtime:** 7.4 seconds

**Query 3:**

Determine the date bicycle stations were installed, alongside the average and maximum ride length of trips from that station in minutes. The query should return the station id, station name, station install date, the average ride length in minutes and the longest ride length in minutes.

**SELECT** h.start\_station\_id **AS** Station\_Id,

h.start\_station\_name **AS** Station\_Name,

**s**.install\_date **AS** Station\_Install\_Date,

**AVG**(h.duration) / 60 **AS** Average\_Ride\_Duration\_In\_Minutes,

**MAX**(h.duration) / 60 **AS** Longest\_Ride\_Duration\_In\_Minutes

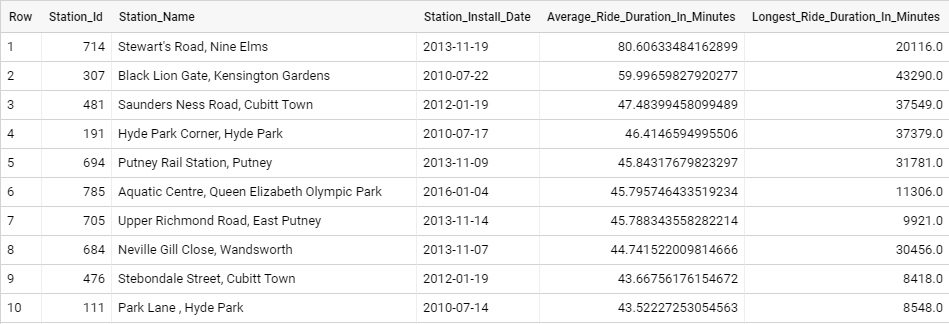
**FROM** `bigquery-**public**-**data**.london\_bicycles.cycle\_hire` h,

`bigquery-**public**-**data**.london\_bicycles.cycle\_stations` **s**

**WHERE** h.start\_station\_id = **s**.**id**

**GROUP** **BY** h.start\_station\_id, h.start\_station\_name, **s**.install\_date

**ORDER** **BY** **AVG**(h.duration) **DESC**;



**Size:** 1 GB Processed.

**Runtime:** 2.5 seconds

(After researching London bike rental prices, I calculated that the longest bike ride taken would cost almost £3000.)