

London Bicycles: Data Analysis with SQL (BigQuery)

Sheethal Melnarse

27 March 2020

Objective: To diagnose the problem and offer recommendations to solve customer's complaint about bike stations being empty in London.

Diagnosis and findings:

- There are empty docks present in majority of the stations. But significant number of stations i.e. More than half of the London bicycle stations have above 50% empty docks. This is a major problem since most of these stations exist in central London where customer traffic is high throughout the day.
- Only 3.4% of the rides start and end at the same location. This might be due to visitors taking a ride for sightseeing and returning to the same station.
- Among the popular stations are King's Cross, Holborn, Hyde Park, Soho etc. Their peak usage is in the morning (7am-9am) and evening (4pm-6pm). This is because of the people commuting to workplace in the morning from train station junctions and returning in the evening. The most popular trip is between King's Cross and Holborn station.
- The peak time rides are usually taken by regular customers who work in central London and they use bicycles to commute from train station to their workplace.
- The problem arises when visitors rent the bikes around or before peak time in the evening. This creates a shortage of bikes for regular customers. The popular station for visitors is Hyde Park, Museum, King's Cross etc.
- The inventory is one of the major problems. There are only 13,705 bikes available in London whereas the number of docks is over 20,000. The **potential resolution** would be to increase the number of available bikes and install them at the tourist places and busy train junctions.

Recommendations

Offer monthly subscription to regular customers with a discount as incentive to opt-in. There can be an implementation of reservation system where if the customer used the bike in the morning then a same or different bike (depending on availability) should be reserved for them around the time they get off work in the evening. This recommendation along with the potential resolution would solve the bike stations being empty issue for customers.

Appendix

What is the size of this dataset? How many trips?

```
SELECT Count(*) FROM `bigquery-public-data.london_bicycles.cycle_hire` LIMIT 1000
```

When was the first and last trip in this dataset?

```
SELECT min(start_date), max(end_date) FROM `bigquery-public-data.london_bicycles.cycle_hire`;
```

How many bikes are there?

```
SELECT COUNT(DISTINCT bike_id) FROM `bigquery-public-data.london_bicycles.cycle_hire`;
```

How many stations are there?

```
SELECT COUNT(DISTINCT id) FROM `bigquery-public-data.london_bicycles.cycle_stations`;
```

How many trips are in the morning, and how many are in the afternoon?

```
SELECT CASE WHEN EXTRACT(hour FROM start_date) BETWEEN 6 and 11 THEN "morning"
WHEN EXTRACT(hour FROM start_date) BETWEEN 12 and 18 THEN "afternoon"
ELSE Null END as time_of_day,
COUNT(distinct rental_id) FROM `bigquery-public-data.london_bicycles.cycle_hire` GROUP BY
time_of_day;
```

Percentage of empty bikes in every station

```
SELECT nbEmptyDocks/docks_count*100 as percentage, name FROM `bigquery-public-data.london_bicycles.cycle_stations`
ORDER BY percentage desc LIMIT 1000;
```

Total number of bikes in station and total docks

```
SELECT SUM(bikes_count) as total_bikes, SUM(docks_count) as docks_count FROM `bigquery-public-data.london_bicycles.cycle_stations` LIMIT 1000;
```

Percentage of rides where start and end station is same

```
SELECT (SELECT count(rental_id) as rent FROM `bigquery-public-data.london_bicycles.cycle_hire`  
where end_station_id = start_station_id)/count(*)*100 FROM `bigquery-public-data.london_bicycles.cycle_hire` ;
```

Busiest hour in a day

```
SELECT count(*), EXTRACT(HOUR FROM start_date) as hour  
FROM `bigquery-public-data.london_bicycles.cycle_hire`  
group by hour  
order by count(*) desc;
```

Busiest start station

```
SELECT count(*), start_station_name, EXTRACT(HOUR FROM start_date) as hour  
FROM `bigquery-public-data.london_bicycles.cycle_hire`  
group by start_station_name, hour  
order by count(*) desc;
```

Customer who use bicycles for longer duration

```
SELECT start_station_name, end_station_name  
FROM `bigquery-public-data.london_bicycles.cycle_hire`  
where duration>900  
GROUP BY start_station_name,end_station_name;
```

Visitors during peak hours in the evening

```
SELECT count(*), start_station_name, end_station_name, EXTRACT(HOUR FROM start_date) as  
hour  
  
FROM `bigquery-public-data.london_bicycles.cycle_hire`  
  
where duration > 900  
  
group by start_station_name,end_station_name, hour  
  
order by count(*) desc;
```

Empty docks by station

```
SELECT nbEmptyDocks as empty , name  
  
FROM `bigquery-public-data.london_bicycles.cycle_stations`  
  
order by empty desc;
```


