My Google Capstone Queries

Tyler Tucker 12/15/2023

Raw Data

My raw data was in the format of 11 CSV files. Each file was for an individual month of 2023. After uploading each of them to a BigQuery project I appended them all into one table, after ensuring they all followed the same formatting. The data totalled 5,495,804 rows. Below is an example of the table format.

```
[{
  "ride_id": "08D63AE5147A8A12",
  "rideable_type": "docked_bike",
  "started_at": "2023-01-09 14:20:41.000000 UTC",
  "ended_at": "2023-01-15 04:19:09.000000 UTC",
  "start_station_name": "Indiana Ave \u0026 Roosevelt Rd",
  "start_station_id": "SL-005",
  "end_station_name": null,
  "end_station_id": null,
  "start_lat": "41.867888",
  "start_lng": "-87.623041",
  "end_lat": null,
  "end_lng": null,
  "member_casual": "casual"
}]
```

Query 1

Query 1 was used to produce a table of 5,495,791 rows to analyze trip data. I allowed null values for start_station_name and end_station_name I wanted to analyze trips without focusing on stations. "OH Charging Stx - Test" proved to be an outlier with only a hand full of rows with no valuable trip data.

```
--Create a table that separates started_at and ended_at into dates, hours, and trip lengt
SELECT
  ride_id,
  rideable_type,
  member_casual,
  date(started_at) as start_date,
  date(ended_at) as end_date,
  FORMAT_TIMESTAMP('%H:00:00', started_at) AS start_time,
```

```
    FORMAT_TIMESTAMP('%H:00:00', ended_at) AS end_time,
    timestamp_diff(ended_at, started_at, MINUTE) as trip_length

  FROM `sonic-charmer-407302.bike_share_data.2023_trip_combined_data`

  --Remove outlying rows that don't benefit the analysis. Keeps nulls to focus on trip anal
  WHERE
  end_station_id <> "OH Charging Stx - Test" or end_station_id is null
```

Below is the resulting table format.

```
[{
  "ride_id": "B5D8E42062B31CC9",
  "rideable_type": "electric_bike",
  "member_casual": "member",
  "start_date": "2023-07-20",
  "end_date": "2023-07-20",
  "start_time": "17:00:00",
  "end_time": "17:00:00",
  "trip_length": "6"
}]
```

Query 2

Query 2 is where I began to focus on station information. After investigating the station data using a map visual in Power BI I noticed that there were stations that shared the same name and general location with only small differences. Some had "Public Rack" at the front of the name and others had "Temp" at the end of their name. This query was written to find the stations that exhibited this behavior.

```
  -- Appends end_station_name values into one column that share an end_station_id
  SELECT
      end_station_id,
      STRING_AGG(DISTINCT end_station_name ORDER BY end_station_name) AS station_names

  FROM
      `bike_share_data.2023_trip_combined_data`

  -- Removes null rows for end_station_name
  WHERE
      end_station_name IS NOT NULL

  -- Aggregates the data by end_station_id
  GROUP BY
```

```
        end_station_id

  -- Produces only the rows where end_station_name has more than 1 distinct value
  HAVING
        COUNT(DISTINCT end_station_name) > 1

  ORDER BY
        end_station_id;
```

Below is an example of the table **Query 2** produced. Station names were comma delimited.

```
[{
  "end_station_id": "13259",
  "station_names": "California Ave \u0026 Francis Pl,California Ave \u0026 Francis Pl (Te
}]
```

Query 3

There turned out to be 85 stations that were sharing a station i.d. but with slightly different station names. Since start stations and end stations were physically the same set of stations, this query was written to produce a cleaned dataset of trips focusing on station information. Value names of some fields were changed so that they would present better in Power BI visuals and keep Power BI from having to transform the formats. The query produced 4,164,667 rows. I also did side by side comparisons with visuals to the trip data that contained nulls to ensure no large trend differences were present. They were nearly identical.

```
SELECT
  ride_id,
    CASE
      WHEN rideable_type = 'electric_bike' THEN 'Electric'
      WHEN rideable_type = 'classic_bike' THEN 'Classic'
      WHEN rideable_type = 'docked_bike' THEN 'Docked'
      ELSE rideable_type
  END AS bike_type,
    CASE
      WHEN member_casual = 'member' THEN 'Member'
      WHEN member_casual = 'casual' THEN 'Casual'
      ELSE member_casual
  END AS rider_status,
  start_station_id,
    CASE
      WHEN start_station_id IS NULL AND start_station_name = 'Elizabeth St & Randolph St'
      WHEN start_station_id IS NULL AND start_station_name = 'Stony Island Ave & 63rd St'
      ELSE start_station_id
```

```sql
        END AS start_station_id,
    CASE
        WHEN start_station_id IS NULL AND start_station_name = 'Elizabeth St & Randolph S
        WHEN start_station_id IS NULL AND start_station_name = 'Stony Island Ave & 63rd S
        WHEN start_station_id = '13259' THEN 'California Ave & Francis Pl'
        WHEN start_station_id = '13290' THEN 'Noble St & Milwaukee Ave'
        WHEN start_station_id = '15541' THEN 'Buckingham Fountain'
        WHEN start_station_id = '15541.1.1' THEN 'Buckingham Fountain'
        WHEN start_station_id = '15623' THEN 'Campbell Ave & Montrose Ave'
        WHEN start_station_id = '20133' THEN 'Woodlawn & 103rd'
        WHEN start_station_id = '303' THEN 'Kildare Ave & Division St'
        WHEN start_station_id = '335.0' THEN 'Komensky Ave & 31st St'
        WHEN start_station_id = '390' THEN 'California Ave & Marquette Rd'
        WHEN start_station_id = '410' THEN 'Campbell Ave & Augusta Blvd'
        WHEN start_station_id = '514' THEN 'Ridge Blvd & Howard St'
        WHEN start_station_id = '517' THEN 'Clark St & Jarvis Ave'
        WHEN start_station_id = '518' THEN 'Conservatory Dr & Lake St'
        WHEN start_station_id = '519' THEN 'Wolcott Ave & Fargo Ave'
        WHEN start_station_id = '520' THEN 'Greenview Ave & Jarvis Ave'
        WHEN start_station_id = '523' THEN 'Eastlake Ter & Howard St'
        WHEN start_station_id = '525' THEN 'Glenwood Ave & Touhy Ave'
        WHEN start_station_id = '528' THEN 'Pulaski Rd & Lake St'
        WHEN start_station_id = '534' THEN 'Karlov Ave & Madison St'
        WHEN start_station_id = '535' THEN 'Pulaski Rd & Congress Pkwy'
        WHEN start_station_id = '536' THEN 'Kostner Ave & Lake St'
        WHEN start_station_id = '537' THEN 'Kenton Ave & Madison St'
        WHEN start_station_id = '543' THEN 'Laramie Ave & Gladys Ave'
        WHEN start_station_id = '545' THEN 'Kostner Ave & Adams St'
        WHEN start_station_id = '546' THEN 'Damen Ave & Pershing Rd'
        WHEN start_station_id = '549' THEN 'Marshfield Ave & 44th St'
        WHEN start_station_id = '553' THEN 'Elizabeth St & 47th St'
        WHEN start_station_id = '554' THEN 'Damen Ave & 51st St'
        WHEN start_station_id = '559' THEN 'Racine Ave & Garfield Blvd'
        WHEN start_station_id = '560' THEN 'Marshfield Ave & 59th St'
        WHEN start_station_id = '561' THEN 'Damen Ave & 59th St'
        WHEN start_station_id = '562' THEN 'Racine Ave & 61st St'
        WHEN start_station_id = '564' THEN 'Racine Ave & 65th St'
        WHEN start_station_id = '567' THEN 'May St & 69th St'
        WHEN start_station_id = '569' THEN 'Woodlawn Ave & 75th St'
        WHEN start_station_id = '570' THEN 'Evans Ave & 75th St'
        WHEN start_station_id = '571' THEN 'Vernon Ave & 75th St'
        WHEN start_station_id = '572' THEN 'State St & 76th St'
        WHEN start_station_id = '573' THEN 'State St & 79th St'
        WHEN start_station_id = '574' THEN 'Vernon Ave & 79th St'
        WHEN start_station_id = '575' THEN 'Cottage Grove Ave & 78th St'
        WHEN start_station_id = '577' THEN 'Stony Island Ave & South Chicago Ave'
        WHEN start_station_id = '579' THEN 'Phillips Ave & 79th St'
        WHEN start_station_id = '580' THEN 'Exchange Ave & 79th St'
        WHEN start_station_id = '583' THEN 'Stony Island Ave & 82nd St'
```

```sql
        WHEN start_station_id = '584' THEN 'Ellis Ave & 83rd St'
        WHEN start_station_id = '585' THEN 'Cottage Grove Ave & 83rd St'
        WHEN start_station_id = '586' THEN 'MLK Jr Dr & 83rd St'
        WHEN start_station_id = '588' THEN 'South Chicago Ave & 83rd St'
        WHEN start_station_id = '592' THEN 'Knox Ave & Montrose Ave'
        WHEN start_station_id = '593' THEN 'Halsted St & 59th St'
        WHEN start_station_id = '594' THEN 'Western Blvd & 48th Pl'
        WHEN start_station_id = '595' THEN 'Wabash Ave & 87th St'
        WHEN start_station_id = '596' THEN 'Benson Ave & Church St'
        WHEN start_station_id = '599' THEN 'Valli Produce – Evanston Plaza'
        WHEN start_station_id = '604' THEN 'Sheridan Rd & Noyes St (NU)'
        WHEN start_station_id = '620' THEN 'Orleans St & Chestnut St (NEXT Apts)'
        WHEN start_station_id = '623' THEN 'Michigan Ave & 8th St'
        WHEN start_station_id = '624' THEN 'Dearborn St & Van Buren St'
        WHEN start_station_id = '631' THEN 'Malcolm X College'
        WHEN start_station_id = '636' THEN 'Orleans St & Hubbard St'
        WHEN start_station_id = '637' THEN 'Wood St & Chicago Ave'
        WHEN start_station_id = '638' THEN 'Clinton St & Jackson Blvd'
        WHEN start_station_id = '639' THEN 'Lakefront Trail & Wilson Ave'
        WHEN start_station_id = '642' THEN 'Latrobe Ave & Chicago Ave'
        WHEN start_station_id = '643' THEN 'Smith Park'
        WHEN start_station_id = '644' THEN 'Western Ave & Fillmore St'
        WHEN start_station_id = '646' THEN 'State St & 54th St'
        WHEN start_station_id = '647' THEN 'Elizabeth St & 59th St'
        WHEN start_station_id = '649' THEN 'Stewart Ave & 63rd St'
        WHEN start_station_id = '650' THEN 'Eggleston Ave & 69th St'
        WHEN start_station_id = '651' THEN 'Michigan Ave & 71st St'
        WHEN start_station_id = '654' THEN 'Racine Ave & Washington Blvd'
        WHEN start_station_id = '657' THEN 'Wood St & Augusta Blvd'
        WHEN start_station_id = '658' THEN 'Leavitt St & Division St'
        WHEN start_station_id = '660' THEN 'Sheridan Rd & Columbia Ave'
        WHEN start_station_id = '661' THEN 'Evanston Civic Center'
        WHEN start_station_id = '662' THEN 'Dodge Ave & Mulford St'
        WHEN start_station_id = '665' THEN 'South Chicago Ave & Elliot Ave'
        WHEN start_station_id = 'KA1504000086' THEN 'Kedzie Ave & 24th St'
        WHEN start_station_id = 'KA1504000141' THEN 'Pulaski Rd & Eddy St'
        WHEN start_station_id = 'TA1306000026' THEN 'Racine Ave & Fullerton Ave'
        WHEN start_station_id = 'TA1308000026' THEN 'Wentworth Ave & 24th St'
        WHEN start_station_id = 'TA1309000042' THEN 'Lincoln Ave & Melrose St'
        ELSE start_station_name
    END AS start_station,
    start_lng,
    start_lat,
      CASE
        WHEN end_station_id IS NULL AND end_station_name = 'Elizabeth St & Randolph St' T
        WHEN end_station_id IS NULL AND end_station_name = 'Stony Island Ave & 63rd St' T
        ELSE end_station_id
    END AS end_station_id,
      CASE
```

```
        WHEN end_station_id IS NULL AND end_station_name = 'Elizabeth St & Randolph St' T
        WHEN end_station_id IS NULL AND end_station_name = 'Stony Island Ave & 63rd St' T
        WHEN end_station_id = '13259' THEN 'California Ave & Francis Pl'
        WHEN end_station_id = '13290' THEN 'Noble St & Milwaukee Ave'
        WHEN end_station_id = '15541' THEN 'Buckingham Fountain'
        WHEN end_station_id = '15541.1.1' THEN 'Buckingham Fountain'
        WHEN end_station_id = '15623' THEN 'Campbell Ave & Montrose Ave'
        WHEN end_station_id = '20133' THEN 'Woodlawn & 103rd'
        WHEN end_station_id = '303' THEN 'Kildare Ave & Division St'
        WHEN end_station_id = '335.0' THEN 'Komensky Ave & 31st St'
        WHEN end_station_id = '390' THEN 'California Ave & Marquette Rd'
        WHEN end_station_id = '410' THEN 'Campbell Ave & Augusta Blvd'
        WHEN end_station_id = '514' THEN 'Ridge Blvd & Howard St'
        WHEN end_station_id = '515' THEN 'Paulina St & Howard St'
        WHEN end_station_id = '517' THEN 'Clark St & Jarvis Ave'
        WHEN end_station_id = '518' THEN 'Conservatory Dr & Lake St'
        WHEN end_station_id = '519' THEN 'Wolcott Ave & Fargo Ave'
        WHEN end_station_id = '520' THEN 'Greenview Ave & Jarvis Ave'
        WHEN end_station_id = '523' THEN 'Eastlake Ter & Howard St'
        WHEN end_station_id = '525' THEN 'Glenwood Ave & Touhy Ave'
        WHEN end_station_id = '527' THEN 'Western Ave & Howard St'
        WHEN end_station_id = '528' THEN 'Pulaski Rd & Lake St'
        WHEN end_station_id = '530' THEN 'Laramie Ave & Kinzie St'
        WHEN end_station_id = '534' THEN 'Karlov Ave & Madison St'
        WHEN end_station_id = '535' THEN 'Pulaski Rd & Congress Pkwy'
        WHEN end_station_id = '536' THEN 'Kostner Ave & Lake St'
        WHEN end_station_id = '537' THEN 'Kenton Ave & Madison St'
        WHEN end_station_id = '540' THEN 'Laramie Ave & Madison St'
        WHEN end_station_id = '543' THEN 'Laramie Ave & Gladys Ave'
        WHEN end_station_id = '545' THEN 'Kostner Ave & Adams St'
        WHEN end_station_id = '546' THEN 'Damen Ave & Pershing Rd'
        WHEN end_station_id = '549' THEN 'Marshfield Ave & 44th St'
        WHEN end_station_id = '553' THEN 'Elizabeth St & 47th St'
        WHEN end_station_id = '554' THEN 'Damen Ave & 51st St'
        WHEN end_station_id = '559' THEN 'Racine Ave & Garfield Blvd'
        WHEN end_station_id = '560' THEN 'Marshfield Ave & 59th St'
        WHEN end_station_id = '561' THEN 'Damen Ave & 59th St'
        WHEN end_station_id = '562' THEN 'Racine Ave & 61st St'
        WHEN end_station_id = '564' THEN 'Racine Ave & 65th St'
        WHEN end_station_id = '567' THEN 'May St & 69th St'
        WHEN end_station_id = '569' THEN 'Woodlawn Ave & 75th St'
        WHEN end_station_id = '570' THEN 'Evans Ave & 75th St'
        WHEN end_station_id = '571' THEN 'Vernon Ave & 75th St'
        WHEN end_station_id = '572' THEN 'State St & 76th St'
        WHEN end_station_id = '573' THEN 'State St & 79th St'
        WHEN end_station_id = '574' THEN 'Vernon Ave & 79th St'
        WHEN end_station_id = '576' THEN 'Greenwood Ave & 79th St'
        WHEN end_station_id = '577' THEN 'Stony Island Ave & South Chicago Ave'
        WHEN end_station_id = '579' THEN 'Phillips Ave & 79th St'
```

```sql
        WHEN end_station_id = '580' THEN 'Exchange Ave & 79th St'
        WHEN end_station_id = '583' THEN 'Stony Island Ave & 82nd St'
        WHEN end_station_id = '584' THEN 'Ellis Ave & 83rd St'
        WHEN end_station_id = '585' THEN 'Cottage Grove Ave & 83rd St'
        WHEN end_station_id = '586' THEN 'MLK Jr Dr & 83rd St'
        WHEN end_station_id = '587' THEN 'Wabash Ave & 83rd St'
        WHEN end_station_id = '588' THEN 'South Chicago Ave & 83rd St'
        WHEN end_station_id = '592' THEN 'Knox Ave & Montrose Ave'
        WHEN end_station_id = '594' THEN 'Western Blvd & 48th Pl'
        WHEN end_station_id = '595' THEN 'Wabash Ave & 87th St'
        WHEN end_station_id = '596' THEN 'Benson Ave & Church St'
        WHEN end_station_id = '604' THEN 'Sheridan Rd & Noyes St (NU)'
        WHEN end_station_id = '605' THEN 'University Library (NU)'
        WHEN end_station_id = '620' THEN 'Orleans St & Chestnut St (NEXT Apts)'
        WHEN end_station_id = '623' THEN 'Michigan Ave & 8th St'
        WHEN end_station_id = '624' THEN 'Dearborn St & Van Buren St'
        WHEN end_station_id = '631' THEN 'Malcolm X College'
        WHEN end_station_id = '636' THEN 'Orleans St & Hubbard St'
        WHEN end_station_id = '637' THEN 'Wood St & Chicago Ave'
        WHEN end_station_id = '638' THEN 'Clinton St & Jackson Blvd'
        WHEN end_station_id = '639' THEN 'Lakefront Trail & Wilson Ave'
        WHEN end_station_id = '642' THEN 'Latrobe Ave & Chicago Ave'
        WHEN end_station_id = '643' THEN 'Smith Park'
        WHEN end_station_id = '644' THEN 'Western Ave & Fillmore St'
        WHEN end_station_id = '646' THEN 'State St & 54th St'
        WHEN end_station_id = '647' THEN 'Elizabeth St & 59th St'
        WHEN end_station_id = '649' THEN 'Stewart Ave & 63rd St'
        WHEN end_station_id = '650' THEN 'Eggleston Ave & 69th St'
        WHEN end_station_id = '651' THEN 'Michigan Ave & 71st St'
        WHEN end_station_id = '654' THEN 'Racine Ave & Washington Blvd'
        WHEN end_station_id = '657' THEN 'Wood St & Augusta Blvd'
        WHEN end_station_id = '658' THEN 'Leavitt St & Division St'
        WHEN end_station_id = '661' THEN 'Evanston Civic Center'
        WHEN end_station_id = '662' THEN 'Dodge Ave & Mulford St'
        WHEN end_station_id = 'KA1504000086' THEN 'Kedzie Ave & 24th St'
        WHEN end_station_id = 'KA1504000141' THEN 'Pulaski Rd & Eddy St'
        WHEN end_station_id = 'TA1306000026' THEN 'Racine Ave & Fullerton Ave'
        WHEN end_station_id = 'TA1308000026' THEN 'Wentworth Ave & 24th St'
        WHEN end_station_id = 'TA1309000042' THEN 'Lincoln Ave & Melrose St'
        ELSE end_station_name
    END AS end_station,
    end_lng,
    end_lat,
    date(started_at) as trip_date,
    FORMAT_TIMESTAMP('%H:00:00', started_at) AS start_hour,
    timestamp_diff(ended_at, started_at, MINUTE) as trip_length
  FROM `sonic-charmer-407302.bike_share_data.2023_trip_combined_data`
  WHERE
    start_station_id is not null
```

```
    and
    end_station_id is not null
    and
    end_station_id <> "OH Charging Stx - Test";
```

Below is an example of the resulting data.

```json
[{
  "ride_id": "64F97952FC88F460",
  "bike_type": "Electric",
  "rider_status": "Member",
  "start_station_id": "302",
  "start_station": "Lamon Ave \u0026 Chicago Ave",
  "start_lng": "-87.748221",
  "start_lat": "41.895192833333333",
  "end_station_id": "18069",
  "end_station": "Cherry Ave \u0026 Blackhawk St",
  "end_lng": "-87.655618",
  "end_lat": "41.907221",
  "trip_date": "2023-07-13",
  "start_hour": "08:00:00",
  "trip_length": "32"
}]
```