1. Files split on mac terminal with split command in order to make them small enough to upload on BgiQuery SandBox.
2. Files uploaded on BigQuery.
3. Relevant files merged with "UNION ALL" into one table as "2013-2021" in order to run summary statistics. Excluded multiple datatype station IDs.

```
SELECT * EXCEPT(start_station_id, end_station_id)
FROM table_name
UNION ALL
SELECT * EXCEPT(start_station_id, end_station_id)
FROM table_name
```

4. CSV files' column names were altered in order to ensure standardization.

```
ALTER TABLE table_name
RENAME COLUMN column_old to column_new
```

5. The trip_duration column was added to all of the datasets as a time interval.
6. User types have been standardized as "Subscriber" and "Customer".
7. Dependents assumed as customers (casual users).
8. Trips that lasted more than 120 seconds are processed in order to eliminate test data and nontypical trips.
9. Summary statistics were produced including average trip durations, the total number of trips, and total trip durations yearly and monthly in order to analyze trends.

```
#creating a main table in order to run summary statistics from all the data since DML
is not supported in Bigquery Sandbox and this tem_table later saved as 2013to2021
WITH temp_tab AS (
  SELECT
    new_user_type,
    start_station_name,
    end_station_name,
    EXTRACT(year FROM start_time) AS year,
    EXTRACT(month FROM start_time) AS month,
    EXTRACT(hour FROM trip_duration) AS trip_hours,
    EXTRACT(minute FROM trip_duration) AS trip_minutes,
    EXTRACT(second FROM trip_duration) AS trip_seconds,
  FROM `sunny-lightning-361217.cyclistic.2013-2021`)
SELECT
  year,
  month,
  new_user_type,
  start_station_name,
  end_station_name,
  trip_hours * 3600 + trip_minutes * 60 + trip_seconds AS tirp_total_seconds,
FROM
```

```
  temp_tab
WHERE
  trip_hours * 3600 + trip_minutes * 60 + trip_seconds >= 120

# for total trips
SELECT
  year,
  AVG(trip_total_seconds) AS average_trip_seconds,
  SUM(trip_total_seconds) AS total_trip_duration_seconds
FROM
  cyclistic.2013to2021
GROUP BY
  year
ORDER BY
  year

 # for Subscribers
SELECT
  year,
  AVG(trip_total_seconds) AS average_trip_seconds,
  SUM(trip_total_seconds) AS total_trip_duration_seconds
FROM
  cyclistic.2013to2021
WHERE
  new_user_type = "Subscriber"
GROUP BY
  year
ORDER BY
  year

# for Customers
SELECT
  year,
  AVG(trip_total_seconds) AS average_trip_seconds,
  SUM(trip_total_seconds) AS total_trip_duration_seconds
FROM
  cyclistic.2013to2021
WHERE
  new_user_type = "Customer"
GROUP BY
  year
ORDER BY
  year
```

10. Weekdays of the trips were extracted from the start time in order to analyze daily trends.

```
SELECT
  *,
  FORMAT_DATE('%A', start_time) AS trip_day
```

```
FROM
  cyclistic.table_name
# Aggregate trip_day data
SELECT
  new_user_type,
  trip_day,
  AVG(tirp_total_seconds) AS average_trip_seconds,
  SUM(tirp_total_seconds) AS total_trip_duration_seconds,
  COUNT(trip_day) AS number_of_trips
FROM
  `sunny-lightning-361217.cyclistic.2013to2021`
GROUP BY
  new_user_type,
  trip_day
```

11. The number of different consumer groups aggregated to inspect the change started in 2019.

```
SELECT
  year,
  COUNT(new_user_type) AS type
FROM
  cyclistic.2013to2021
WHERE
  new_user_type = 'type'
GROUP BY
  year
ORDER BY
  year
```