'Cyclistic' Project_notes

Scenario

You are a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes di1erently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But rst, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations.

Additional details

- o Bike company 'Cyclistic'
- o 5,824 fleet
- o single-ride passes, full-day passes: **Casual members**
- o annual memberships: **Cyclistic members**
- o Profitability: **Cyclistic members> Casual members**

Business tasks

- o How do annual members and casual riders use Cyclistic bikes differently?
- o Why would casual riders buy Cyclistic annual memberships?
- o How can Cyclistic use digital media to infl1uence casual riders to become members?

Business goal

- o Maximizing Cyclistic members
- o Converting **Casual members** into **Cyclistic members**

Requirements

- o Understand how *annual members* and *casual riders* diff1er
- o Why casual riders would buy a membership
- o How digital media could aff1ect their marketing tactics

Tool used:

- o BigQuery
- o Looker

Data Integrity, Consistency

o **Checking for rows with NULL values:**

```sql
SELECT *
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
WHERE
  ride_id IS NULL
  OR rideable_type IS NULL
  OR started_at IS NULL
  OR ended_at IS NULL
  OR start_station_name IS NULL
  OR start_station_id IS NULL
  OR end_station_name IS NULL
  OR end_station_id IS NULL
  OR start_lat IS NULL
  OR start_lng IS NULL
  OR end_lat IS NULL
  OR end_lng IS NULL
  OR member_casual IS NULL
```

*** 1x stolen/abandoned bike shown on results = missing geographic end details*

o **Checking the main length of *ride_id* (*result: '16'*). Checking for rows where *ride_id* is not 16 characters:**

```sql
SELECT counter
FROM
  (
    SELECT LENGTH(ride_id) AS counter
    FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
  ) AS summary -- inner query
WHERE summary.counter!=16
```

o **Checking year's consistency:**

```sql
SELECT
  DISTINCT EXTRACT(YEAR FROM started_at)
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
```

*** repeated also for 'ended_at'*

o **Checking consistency of *start_station_name*:**

```sql
SELECT DISTINCT start_station_name
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
```

o **Checking consistency between *start_station_name* and *start_station_id*:**

```sql
SELECT
  t1.start_station_id
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020` AS t1
LEFT JOIN
  (SELECT
    DISTINCT start_station_name,
    start_station_id
  FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`) AS t2
  ON t1.start_station_id=t2.start_station_id
  WHERE t2.start_station_id IS NULL -- check only for those values that are not in
common
```

**repeated also for end_station_id

o **Checking integrity *member_casual***

```sql
SELECT
  DISTINCT member_casual
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
```

I verified integrity and consistency of the data present in the dataset.
Now it is possible to continue with processing and analyzing it in order to answer to the business tasks.

Data Processing, Analysis

o **Amount of users grouped by type of subscription (viz made on looker)**

```sql
SELECT
  member_casual,
  COUNT (member_casual) AS n_users
FROM `portfolioproject-401814.Cyclistic.trips_q1_2020`
GROUP BY member_casual
```