SQL Bicycles

Validating/Cleaning Data

checked ride_id unique no duplicates

```
SELECT
    ride_id, COUNT(ride_id)
FROM
    202301_divvy_tripdata
GROUP BY
    ride_id
HAVING
    COUNT(ride_id) > 1
```

Validating/Cleaning Data

Checked only 2 types of members

SELECT member_casual, COUNT(*) FROM 202301_divvy_tripdata GROUP BY member_casual

Check start station name and start lat and start long start id all same:

SELECT start_station_name, start_station_id, start_lat, start_lng FROM `202301_divvy_tripdata`

Validating/Cleaning Data

Check end station name matched end_lat and end_lng

SELECT end_station_name, end_station_id, end_lat, end_lng from 202301_divvy_tripdata WHERE start_station_name IN (SELECT end_station_name FROM 202301_divvy_tripdata GROUP BY end_station_name HAVING COUNT(distinct end_station_id) > 1)

Validating/Cleaning Data

Check to see all the rides that were under 1 minute assume they were a mistake and not count toward analysis

6188 total

SELECT started_at, ended_at, TIMESTAMPDIFF(MINUTE, started_at , ended_at) AS Minute_Differnce, member_casual
FROM 202301_divvy_tripdata WHERE TIMESTAMPDIFF(MINUTE, started_at , ended_at) < 1

Validating/Cleaning Data

Adding missing data records:

UPDATE `202301_divvy_tripdata`

SET start_station_name = "Pulaski Rd & 51st St",

start_station_id= 343

WHERE start_lat = 41.8;

Validating/Cleaning Data

If not enough data was given to complete records:

DELETE FROM `202301_divvy_tripdata` where end_lat is NULL or end_lat=""

How many casuals vs members

```
SELECT member_casual, avg(TIMESTAMPDIFF(MINUTE,started_at, ended_at)) as minutes FROM 202301_divvy_tripdata where member_casual = "casual"
```

☐ Show all | Number of rows: 25 ⌄   Filter rows: Search this table

+ Options

member_casual	minutes
casual	13.1005

Average session members per ride

Showing rows 0 - 0 (1 total, Query took 0.2070 seconds.)

SELECT member_casual, avg(TIMESTAMPDIFF(MINUTE,started_at, ended_at)) as minutes FROM 202301_divvy_tripdata where member_casual = "member"

Show all | Number of rows: 25 ▾   Filter rows: Search this table

+ Options

member_casual	minutes
member	9.6378

How many different bicycles were used by casuals

```sql
select COUNT(*), rideable_type, member_casual from 202301_divvy_tripdata where member_casual = "casual" GROUP by rideable_type
```

Show all | Number of rows: 25 ⌄   Filter rows: Search this table

+ Options

COUNT(*)	rideable_type	member_casual
13862	classic_bike	casual
1682	docked_bike	casual
24361	electric_bike	casual

Show all | Number of rows: 25 ⌄   Filter rows: Search this table

Time of day riding casuals

```
select count(*), hour(started_at) as hour
from 202301_divvy_tripdata

where (
CASE
    WHEN hour(started_at) > 20 THEN 'Evening ride'
    WHEN hour(started_at) > 16 THEN 'afternoon ride'
    WHEN hour(started_at)  > 12 THEN 'midday ride'
    WHEN hour(started_at) < 12 THEN 'morning ride'
    ELSE 'time is off'
END
) IN ('morning ride')

AND member_casual = "casual"
```

Which day did casuals ride most

Casual

select *, dayofweek(started_at) as day, COUNT(dayofweek(started_at))
from 202301_divvy_tripdata
where member_casual ="casual"
group by dayofweek(started_at)
ORDER BY `COUNT(dayofweek(started_at))` DESC