# Using Union All merged all trips data from the year 2024

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_01

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_02

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_03

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_04

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_05

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_06

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_07

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_08

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_09

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_10

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_11

union all

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_12

# After merging the table, I saved the results naming my table as Trips\_2024\_All

# Find total rows

select \* from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_All

# results: total rows = 5860568

#Create new column ride\_length

UPDATE data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_All

SET ride\_length = date\_diff(ended\_at, started\_at, minute)

where true

# I’m extratcting ride\_length in minutes

# Get duplicates

select D\_Trips.\*

from (select Trips.\*, count(\*) over (partition by ride\_id) as cnt

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_All as Trips

) D\_Trips

where cnt > 1

# results: total duplicate rows: 211

# Get unique records

with new\_table as

(select trips.\*, Row\_Number() over (partition by ride\_id order by ride\_id) as cnt

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_All as trips )

select \* except(cnt) from new\_table

where cnt = 1

#results: total unique rows = 5860357

# after getting clean file, I saved the results naming my table as Trips\_2024\_unique\_records

# Mean, min, max of ride\_length in minutes:

select min(ride\_length) as minimum, max(ride\_length) as maximum, round(avg(ride\_length),2) as mean

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

# min = -2748 , max = 1559, mean = 16.82

# While analyzing the data, I noticed that the minimum ride length is negative, likely due to errors in the started\_at and ended\_at columns. After investigating started\_at & ended\_at columns, it appears that in about 95 records, the values may have been accidentally swapped—either the date or time entries are incorrect. However, this needs to be confirmed before proceeding. As this couldn’t be verified in the fictional dataset, I removed these 95 records to ensure the further analysis is based on clean and reliable data.

delete

FROM `data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records`

where ride\_length < 0

#results: total rows comes down to 5860262

# Count of casual riders and members:

select member\_casual, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

group by 1

#reults: member= 3708778 , casual = 2151484

# Rider’s use of bike types

select rideable\_type, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

group by 1

order by 2 desc

#results: electric\_bike = 2980429 , classic\_bike = 2735496 & electric\_scooter = 144337

# By member

select rideable\_type, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'member'

group by 1

order by 2 desc

#results: electric\_bike = 1889026 , classic\_bike = 1760630 & electric\_scooter = 59122

# By casual

select rideable\_type, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'casual'

group by 1

order by 2 desc

#results: electric\_bike = 1091403 , classic\_bike = 974866 & electric\_scooter = 85215

# Added new column day\_of week to see the trend over the weekdays and weekends:

select ride\_id, started\_at, extract(DAYOFWEEK FROM started\_at) as day\_of\_week

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

order by 1

(UPDATE data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

SET day\_of\_week = extract(DAYOFWEEK FROM started\_at)

where true)

# Count of rides by casual riders by day\_of\_week:

select day\_of\_week, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'casual'

group by 1

order by 2 desc

#results: top day\_of\_week are 7, 1, 6

# Count of rides by member riders by day\_of\_week:

select day\_of\_week, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'member'

group by 1

order by 2 desc

#results: top day\_of\_week are 4,3,5

# mean of ride\_length of members & casuals

select member\_casual, round(avg(ride\_length),2)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

group by 1

#results: member = 12.28 , casual = 24.63

# count of riders by ride\_duration

# By Casual

select ride\_length, count(\*) from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'casual'

group by 1

order by 1 desc

#results: of top 4 ride\_length riders: ride\_length 1559 min has 13 riders , 1509 min has 1 rider , 1500 min has 6 riders and 1499 min has 5990 riders

#By Member

select ride\_length, count(\*) from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'member'

group by 1

order by 1 desc

#results: of top 4 ride\_length riders: ride\_length 1559 min has 2 riders , 1500 min has 1 rider, 1499 min has 1393 riders and 1497 has 2 riders

# Peak Time of day

#By Member

select extract(hour FROM started\_at) as start\_hour, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'member'

group by 1

order by 2 desc

#results: peak hours for members here are 17, 16 , 18 and 8.

# By Casual

select extract(hour FROM started\_at) as start\_hour, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'casual'

group by 1

order by 2 desc

#results: peak hours for members here are 17, 16 , 18 and 15.

#station crowd

#By Member

select start\_station\_name, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'member'

group by 1

order by 2 desc

#results: are top 4 : null = 662769, Kingsbury St & Kinzie St = 29522, Clinton St & Washington Blvd = 27745, Clinton St & Madison St = 24893

#By Casual

select start\_station\_name, count(\*)

from data-analysis-cyclistic.Cyclistic\_trips.Trips\_2024\_unique\_records

where member\_casual = 'casual'

group by 1

order by 2 desc

#results: are top 4 : null = 411106, Streeter Dr & Grand Ave = 51049, DuSable Lake Shore Dr & Monroe = 34102, Michigan Ave & Oak St = 25137