-- data cleaned in excel.

--no duplicates were found

-- deleted null values

-- deleted unwanted columns

-- added few columns like, start\_weekday, end\_weekday, month, day\_type, tot\_days

-- corrected time in some places, where end time is more than start time

--deleted 0 usage time data

--combining all 12 months data

create table `hjprojects1.Cyclist\_capstone.ride\_data`

AS

(SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_march`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_april`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_may`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_june`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_july`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_aug`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_sept`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_oct`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_nov`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2022\_dec`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2023\_jan`

UNION ALL

SELECT rideable\_type, started\_at, ended\_at, start\_station\_name, end\_station\_name, start\_lat, start\_lng, end\_lat, end\_lng, member\_casual, start\_weekday, end\_weekday, month, day\_type, tot\_days

FROM `hjprojects1.Cyclist\_capstone.2023\_feb`)

--view table

SELECT \*

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

--total rides from march 2022 till feb 2023

SELECT COUNT(rideable\_type)

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

--total rides monthwise

select month, count(month) AS month\_ride

from `hjprojects1.Cyclist\_capstone.ride\_data`

group by month

-- count ridable typle month wise

select rideable\_type, month, count(rideable\_type) AS rideable\_monthwise

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

group by month, rideable\_type

--membership count monthwise

select member\_casual, month, count(member\_casual) AS membershi\_count

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

group by month, member\_casual

-- memebrship ride count

SELECT member\_casual, COUNT(member\_casual) AS membership\_count

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY member\_casual

--bike count based on weekday

SELECT start\_weekday, member\_casual, COUNT(member\_casual) AS tot\_rides

from `hjprojects1.Cyclist\_capstone.ride\_data`

group by start\_weekday, member\_casual

--average ride duration

SELECT

AVG(TIMESTAMP\_DIFF(ended\_at, started\_at,MINUTE)) AS avg\_duration

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

--average ride length monthwise and membership type

SELECT month, member\_casual,

AVG(TIMESTAMP\_DIFF(ended\_at, started\_at,MINUTE)) AS avg\_duration

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

Group by month, member\_casual

--polular start staion

SELECT

start\_station\_name,

rideable\_type,start\_lng,start\_lat, end\_lat, end\_lng,

COUNT(\*) AS ride\_totals

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY start\_station\_name, rideable\_type, start\_lng,start\_lat, end\_lat, end\_lng

ORDER BY ride\_totals DESC

LIMIT 20

--polular end staion

SELECT

end\_station\_name,

rideable\_type,start\_lng,start\_lat, end\_lat, end\_lng,

COUNT(\*) AS ride\_totals

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY end\_station\_name, rideable\_type, start\_lng,start\_lat, end\_lat, end\_lng

ORDER BY ride\_totals DESC

LIMIT 20

--average per membership ride duration

SELECT member\_casual,

AVG(TIMESTAMP\_DIFF(ended\_at, started\_at,MINUTE)) AS avg\_duration

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

group by member\_casual

--tot member ride based on day type

select member\_casual, day\_type, count(day\_type) AS tot\_ride

from `hjprojects1.Cyclist\_capstone.ride\_data`

group by member\_casual, day\_type

--avg ride on membership type

SELECT member\_casual,

AVG(TIMESTAMP\_DIFF(ended\_at, started\_at,MINUTE)) AS avg\_duration

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

group by member\_casual

--average distance travelled per ride

SELECT AVG(ST\_DISTANCE(ST\_GEOGPOINT(start\_lng, start\_lat), ST\_GEOGPOINT(end\_lng, end\_lat))) AS avg\_distance\_traveled\_per\_ride

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

--average distance travelled per by membership type

SELECT member\_casual, AVG(ST\_DISTANCE(ST\_GEOGPOINT(start\_lng, start\_lat), ST\_GEOGPOINT(end\_lng, end\_lat))) AS avg\_distance\_traveled\_per\_ride

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

group by member\_casual

--Most popular start and end stations

SELECT start\_station\_name, end\_station\_name,start\_lng,start\_lat, end\_lat, end\_lng, COUNT(\*) as total\_rides

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY start\_station\_name, end\_station\_name,start\_lng,start\_lat, end\_lat, end\_lng

ORDER BY total\_rides DESC

LIMIT 20

--busiest day of the week

SELECT start\_weekday, COUNT(\*) AS num\_rides

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY start\_weekday

ORDER BY num\_rides DESC

--max, min and avg hiring days

select member\_casual, max(tot\_days) AS max\_hire\_days, min(tot\_days) As min\_hire\_days, avg(tot\_days) AS avg\_hire\_days

from `hjprojects1.Cyclist\_capstone.ride\_data`

group by member\_casual

--rides per hour on weekday

SELECT EXTRACT(DAYOFWEEK FROM started\_at) AS weekday,

EXTRACT(HOUR FROM started\_at) AS hour,

COUNT(\*) AS rides\_per\_hour

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY weekday,hour

ORDER BY weekday,hour

--rides on season basis

SELECT EXTRACT(month FROM started\_at) AS trip\_month,

CASE

WHEN EXTRACT(month FROM started\_at) IN (12, 1, 2) THEN 'Winter'

WHEN EXTRACT(month FROM started\_at) IN (3, 4, 5) THEN 'Spring'

WHEN EXTRACT(month FROM started\_at) IN (6, 7, 8) THEN 'Summer'

WHEN EXTRACT(month FROM started\_at) IN (9, 10, 11) THEN 'Fall'

ELSE 'Unknown'

END AS season,

COUNT(\*) AS ride\_count

FROM `hjprojects1.Cyclist\_capstone.ride\_data`

GROUP BY season, trip\_month

--most popular bikes

SELECT rideable\_type, Count(rideable\_type) AS popular\_ride

from `hjprojects1.Cyclist\_capstone.ride\_data`

Group by rideable\_type