Step 1: Checked for duplicate rows

SELECT ride\_id, started\_at, ended\_at,

COUNT(\*)

FROM cyclistic\_data

GROUP BY ride\_id, started\_at, ended\_at

HAVING COUNT(\*) > 1;

Step 2: Checked for NULL values

SELECT

COUNT(\*) AS total\_rows,

COUNT(ride\_id) AS ride\_id\_count,

COUNT(rideable\_type) AS rideable\_type\_count,

COUNT(started\_at) AS started\_at\_count,

COUNT(ended\_at) AS ended\_at\_count,

COUNT(start\_station\_name) AS start\_station\_name\_count,

COUNT(start\_station\_id) AS start\_station\_id\_count,

COUNT(end\_station\_name) AS end\_station\_name\_count,

COUNT(end\_station\_id) AS end\_station\_id\_count,

COUNT(start\_lat) AS start\_lat\_count,

COUNT(start\_lng) AS start\_lng\_count,

COUNT(end\_lat) AS end\_lat\_count,

COUNT(end\_lng) AS end\_lng\_count,

COUNT(member\_casual) AS member\_casual\_count

FROM cyclistic\_data;

Step 3: Checked for inconsistencies in rideable\_type and member\_casual columns

SELECT DISTINCT rideable\_type

FROM cyclistic\_data;

SELECT DISTINCT member\_casual

FROM cyclistic\_data;

Step 4: Creation of new table

Create Table cyclistic\_data\_cleaned AS

(

Select

ride\_id,

rideable\_type AS Bike\_Types,

started\_at,

ended\_at,

start\_station\_name,

end\_station\_name,

member\_casual AS Membership\_Type,

(ended\_at - started\_at) AS Trip\_Duration

From cyclistic\_data

WHERE

(started\_at < ended\_at) AND

(ended\_at - started\_at) >= INTERVAL'1 minute' AND

(start\_station\_name IS NOT NULL AND end\_station\_name IS NOT NULL)

);

Step 5: Checked for any outliers exist in Trip\_Duration column

SELECT

\*

FROM cyclistic\_data\_cleaned

WHERE trip\_duration < Interval '1 minute'

ORDER BY trip\_duration DESC;