********Data Cleaning Log********

- Past 12 month data was made available in separate csv file for each month

  The data was combined to form a single csv file and imported to SQL (pgAdmin)

  The data had many rows with null values and these were removed.

- New table created by removing rows with null values

```
CREATE TABLE tripdata AS

SELECT * FROM tripdata_master WHERE tripdata IS NOT NULL;
```

- There are rows where trip start time is greater than trip end time, and rows where they are same which seems to be irrelevant to analysis.

```
SELECT COUNT (*)

FROM tripdata

WHERE started_at > ended_at OR started_at = ended_at;
```

Since the count of such rows are very less as compared to total rows in data these rows have been removed from database and will not be considered in further analysis

```
DELETE FROM tripdata WHERE started_at > ended_at OR started_at = ended_at;
```

Total rows remaining after the cleaning process : 4159132

PS : My initial plan was to do basic cleaning using spreadsheet but after doing this for one month data I realised it would be a time consuming activity as spreadsheet could not handle entire dataset in single sheet hence decided to combine all the data in single csv file and using SQL for cleaning.