

## **GROUP 1 : SET UP FILE**

Step 1 -> Create a folder named "eda" in working directory.

Step 2 -> Create two subfolders named "2019" and "2020" in the "eda" folder.

Step 3 -> Download the data set from the GDrive from the link

<https://drive.google.com/drive/folders/1V2E5gdTdHyqLMyOP2s6-GZwmDKlQofcr?usp=sharing>

Step 4-> Make sure the data set is saved in Working Directory and under name "Taxi\_Trips.csv".

Step 5 -> Run the R code.( Use Source to run the code)(All the libraries used are mentioned in the code)

Step 6 -> Give input 1 for running the EDA part or 2 for running Modelling and evaluation

Note - As the data set is quite large it may take upto 2 hours for the whole code to run.

Note - It took us around 1:15 mins for the code to complete in IFH lab system on R version 3.4.4.