**Day 3 Prompts:**

Extract the following data in google drive: **cars.csv**

Find the number of duplicate rows in the data

Remove duplicate rows in that data and later check the number of duplicate rows into it

Extract the following data in google drive: **Indian\_districts.xlsx**

Remove first row

Make first row to be column header

Show column headers

Extract the following data in google drive: **airline1.xlsx**

Select 2nd sheet of the excel file

Extract the data in following link:

<https://github.com/RSRemya/Feature_Engineering/blob/main/weather_dataset.csv>

Check the datatype of each column

Change the datatype of column : day to date and time type and then show datatypes of all columns

Now make day column to be index column and show the dataset

Remove rows having only nulls values and show the dataset

Conduct interpolation through method=time to deal with missing values in the data

Fill nulls in column event with mode of the column

Extract the data in following link:

<https://raw.githubusercontent.com/SheepShaun/Bank-customer-churn-prediction/main/bank%20customer%20churn%20dataset.csv>

Now remove following columns: RowNumber, CustomerId and Surname and show the dataset

Extract the data in following link:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

Find number of null values in each column

Fill missing values in column Embarked with mode value of it

Remove the column : Cabin

Create boxplot for column Age

Fill missing values in column Age with its median

Remove outliers in all columns in dataset by replacing them with maximum and minimum values

Create boxplot for column Age

Create an age group column by using binning