# EDA for Regression

1. understand data

* numeric (discrete (eg: no of pax and continuous eg: weight, height),
* categorical (nominal, ranking)

2. relationship between variables

* read\_csv-> test connection -> head() Summary
* shape (# columns, #rows) eg: 1000 rows , 25 fields / columns
* info(
  + data type of all 25 fields ( numeric=int64,float64 , char/string = object)
  + breakdown : int64(5) float64(5) object (15)
  + null value count for each field

3.describe()

* 3.1 summary statistics of numerical fields by default
* 3.2 min, max, 1st q , 3rd, median, count, std

4. Find missing values using isnull().sum()

5. number of unique values in the dataset = nunique()

6. count() - count of values of all fields

7. after finding missing values, replace with mean

8. remove the column if not needed

# visualizations( matplotlib)

1. Univariate analysis (one variable)

2. Bivariate (two variable)

3. Multivariate analysis (many variables)