Download this dataset from this link :

Brief overview about the dataset:

- This dataset is about titanic **Dataset** describing the survival status of individual passengers on the **Titanic**.
- It has a total of 12 columns and ~ 900 rows

Tasks:

1. Find out if there are any null values or not in the dataset. If yes calculate at a column level how many total null values are there
2. How will you handle the null values? Do you wish to drop or impute them?
3. What is the total number of people on the ship who survived v/s not survived
4. How is the survival distribution spread across male and female categories?
5. Does their Age factor a major contribution to their survival factor or not?
6. Are there any outliers for Age?
7. Are there any outliers in Fare?
8. What does the overall Distribution For fare look like?
9. What is the min, mean and max age in the entire dataset looks like?
10. Identify the important features/columns from the dataset which gives most of the information in the entire dataset
11. Is there any specific preference/relation between age and fare price that they opted for?
12. How does Passenger Id and Name contribute to this data and the analysis we are drawing?
13. How does the distribution of Age look like for both the male and female category?