

## # Summary of findings:

1. We found that the mean age in the dataset is 30 based on summary.
2. We can conclude that data columns have data type as integer, float and object.
3. There are 86 null values in age column as well 327 in column Cabin.
4. Pair plot:
  - The variable 'pclass' is discrete variable and thus points appear discrete in the plot.
  - The variable 'age' appears to have right skewed distribution.
  - 'SibSp' and Parch variables show low or zero numbers mostly.
  - 'Fare' variables show right skewed distribution.
5. Heat map: We check correlation between variables using heat map
  - We observe weak correlation between Age and fare.
  - We observe SibSp and Parch to have weak correlation.
  - We observe Fare and Parch to have weak correlation.
  - We observe Age and Pclass to have high correlation.
  - We observe Fare and Pclass to have high correlation.
6. Histogram:
  - The variable 'pclass' is discrete variable and thus points appear discrete in the plot.
  - The variable 'age' appears to have right skewed distribution.
  - 'SibSp' and Parch variables show low or zero numbers mostly.
  - 'Fare' variables show right skewed distribution.
7. Boxplot: It is used to check data distribution and identify outliers in data.
  - We observe significant number of outliers in age.
  - We observe few outliers in SibSp.
  - We observe significant number of outliers in Pitch and Fare.
  - There are no outliers in Cabin.
  - We observe two outliers Embarked.