I have taken **“Titanic Dataset”** from Kaggle. On that dataset I performed Data Pre-processing. Also along with that I used **.describe(), .info(), .value_counts().** After that I had done some visualizations like :

Histogram:

1) Histogram for **Age-Distribution**
   In that I found that there are **mostly younger people** whose age lies in 20-30years.

2) Histogram for **Fare-Distribution**
   The Fare distribution is **right-skewed** with a small number of passengers paying much higher fares.

Boxplot:

These are used to detect the outliers.And we can see that

1) The Age boxplot shows a large **range of ages**, with most passengers falling between 20 and 40 years.
2) The Fare boxplot has significant **outliers** on the upper side, indicating the presence of passengers who paid significantly higher fares.

Scatterplot:

Scatterplots are used to identify the correlations in the data. It reveals the two variables are positively, negatively , or there is no correlation.

1) **1)Age vs Fare Scatterplot**
   There is **no clear linear relationship** between Age and Fare, suggesting that the fare paid was not dependent on age. The scatterplot reveals a mix of social classes across all ages, with younger passengers paying both low and high fares, indicating no direct correlation.
2) **Fare vs Survived**
   You will likely see that passengers who paid higher fares tended to survive more, as 1st-class passengers (with higher fares) had a better survival rate.

Barcharts:

It shows relationship between variables in form of bars. Let’s have some barchat visualizations.

1) **Survival vs Sex**
   Females survived at a much higher rate than males**.**

2) **Survival vs Pclass (Ticket Class)**
   1st class passengers had much better survival than 2nd and 3rd class.

3) **Age vs Survival**
   Children (younger age) survived more; older passengers less.

4) **Fare vs Survival**
Higher fare-paying passengers survived more (they were richer → better access to lifeboats).

5) **Family Size (SibSp + Parch) vs Survival**
Small families (1-3 members) had better survival; alone people or very large families struggled.

Pairplot:

The **pairplot()** is great for visualizing pairwise relationships between features in a dataset. It will generate scatterplots for each pair of continuous variables and histograms on the diagonal.

- **Trend**: Pairplot shows the **relationships between numerical variables** (Age, Fare, SibSp, Parch, etc.) and how they vary with Survived.
- **Key Observations**:
  - **Higher Fare passengers** are more likely to have survived (clear separation visible).
  - **Age** does not show a strong direct separation for survival (but many children survived).
  - **SibSp** and **Parch** values around 0–1 are more associated with survival (large families seem to have lower survival).
- **Overall**: Passengers with **higher Fare**, **lower SibSp**, and **lower Parch** values had better chances of survival.

Heatmap:

A **correlation heatmap** is useful to understand the strength of relationships between numeric features. The heatmap() function will show a color-coded matrix of correlations between the variables.

- **Trend**: Heatmap shows **correlation strength** between numerical variables.
- **Key Observations**:
  - **Fare** has a **positive correlation** with **Survived** (higher Fare → higher survival).
  - **Pclass** has a **negative correlation** with **Survived** (lower class number (1st class) → higher survival).
  - **SibSp** and **Parch** are **positively correlated** with each other (makes sense — families often traveled together).
  - **Age** has a weak negative correlation with Survived (younger passengers slightly more likely to survive).
- **Overall**: **Fare** and **Pclass** are the most important features influencing survival.