Titanic Dataset EDA Report

1. About the Dataset

This dataset contains information about 891 passengers on the Titanic. It has 15 columns with a mix of numbers and categories. Our main focus is on the **survived** column, which tells us if a passenger survived (1) or didn't (0).

2. What the Data Looks Like

Some important things to note:

- There are missing values in the **Age**, **Embarked** (where passengers boarded), and **Deck** columns.

- The **Fare** column has no missing values but varies a lot—some passengers paid way more than others.

- The **Sex** and **Pclass** (passenger class) columns are complete with no missing info.

On average, passengers were about 30 years old and paid around 32 units for their ticket. Around 38% of passengers survived, while 62% did not.

3. Quick Counts

Here's a quick snapshot of the data:

- 342 people survived, 549 didn't.

- 216 passengers were in 1st class, 184 in 2nd, and 491 in 3rd.

- Most passengers were male (577), while 314 were female.

4. What We Learned

- Women and 1st class passengers had the highest survival rates.

- People who paid more for their tickets were more likely to survive.

- Most of the passengers who didn't make it were men in 3rd class.

- Kids had a slightly better chance of surviving compared to adults.

- There were some really expensive tickets, likely from luxury cabins, which skewed the fare data.

5. Looking at the Visuals

Pairplot:

- It clearly shows women survived more than men.

- Kids had better survival rates too.

- Passengers in 1st and 2nd class survived more often than those in 3rd.

- Higher ticket prices were linked to higher survival chances.

Correlation Heatmap:

- Fare has a positive connection to survival — higher fare means better chances.

- Passenger class is negatively correlated — the higher the class number (like 3rd class), the lower the survival.

- Older passengers had a slightly lower chance of survival.

- Higher class passengers paid higher fares, which makes sense.

Age Histogram:

- Most passengers were between 20 and 40 years old.

- Fewer kids and elderly people.

- Middle-aged people didn't survive as much as kids and young women.

Fare Histogram:

- Most people paid less than $100, but some paid way more — these were luxury tickets.

- The rich passengers usually survived more.

Boxplot Age by Class:

- 1st class passengers were older on average.

- 3rd class passengers were generally younger, possibly families or workers.

- 1st class had the widest range of ages.

Boxplot Fare by Survival:

- Survivors paid higher fares, meaning better accommodations and possibly easier access to lifeboats.

- Non-survivors mostly paid less, often 3rd class.

Scatterplot Age vs Fare:

- Most passengers are clustered around lower fares and ages 20-40.

- Survivors tend to be in the high-fare and younger age groups.

- Some kids with low fares survived too, which fits with the "women and children first" policy.

6. Final Thoughts

- Women had a much better chance of survival than men.

- Passengers in 1st and 2nd class, who generally paid more, also survived more often.

- Kids were more likely to survive, while older passengers had a tougher time.

- People who boarded at Cherbourg seemed to survive more, probably because more of them were in 1st class.

- In the end, a passenger's gender, age, social class, and ticket price were the biggest factors influencing survival.