📄 **Exploratory Data Analysis (EDA) Report**

1. Dataset Overview

- **Data Shape:**
  Rows: 891
  Columns: 12
- **Columns:**
  - PassengerId
  - Survived (Target Variable)
  - Pclass
  - Name
  - Sex
  - Age
  - SibSp
  - Parch
  - Ticket
  - Fare
  - Cabin
  - Embarked
- **Data Types:**
  - Numeric: PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare
  - Categorical: Sex, Embarked
  - Text: Name, Ticket, Cabin

2. Basic Statistics

- **Age:**
  - Mean: ~29.7 years
  - Min: 0.42 years
  - Max: 80 years
  - Standard deviation: ~14.5 years
- **Fare:**
  - Mean: ~32.2
  - Highly skewed with a maximum fare of 512.
- **SibSp and Parch:**
  - Most passengers traveled alone (0 siblings/spouses or parents/children).

3. Missing Values

- **Age:** 177 missing values
- **Cabin:** 687 missing values
- **Embarked:** 2 missing values

Cabin has too many missing values; it might be dropped or require heavy cleaning.

4. Univariate Analysis

Histograms:

- **Age:**
  - Normal distribution with slight right skew.
  - Majority of passengers were between 20 and 40 years old.
- **Fare:**
  - Highly right-skewed distribution.
  - Few passengers paid extremely high fares.

5. Bivariate and Multivariate Analysis

5.1 Survival Rate by Sex

- Females had a much higher survival rate than males.

5.2 Survival Rate by Passenger Class (Pclass)

- 1st Class passengers had the highest survival rate.
- 3rd Class passengers had the lowest survival rate.

5.3 Boxplots

- 1st Class fares were much higher than 2nd and 3rd Class.
- Significant number of fare outliers.

5.4 Pairplot (Selected Features: Age, Fare, SibSp, Parch)

- Younger passengers paid lower fares generally.
- Most passengers had small family sizes (0–2).
- Few passengers traveled with many family members.

5.5 Correlation Heatmap

- **Fare** and **Pclass** are negatively correlated (-0.55).
- **SibSp** and **Parch** are positively correlated (~0.41).
- **Survived** correlation:
  - Positive with Fare (0.26) → Higher fare passengers had better survival chances.
  - Negative with Pclass (-0.34) → Lower-class passengers had lower survival chances.

6. Key Insights Summary

✅ Females had a higher chance of survival than males.
✅ Passengers traveling in 1st class survived more often than those in lower classes.
✅ Higher fare-paying passengers had better survival rates.
✅ Most passengers were young adults (20–40 years).
✅ Having family members on board slightly influenced survival.

📚 Deliverables Checklist

Task	Status
Exploratory Data Analysis	✅ Done
Visualizations (Histograms, Boxplots, Pairplots, Heatmaps)	✅ Done
Observations for each visual	✅ Done
Summary of findings	✅ Done

📢 Conclusion

In this exploratory data analysis of the Titanic dataset, we identified important patterns influencing survival. Gender, passenger class, and fare amount played major roles in survival chances. The analysis highlighted the need for careful feature engineering and preprocessing for further modeling tasks. This EDA provides a strong foundation for building predictive models and deeper analytical work.