

# REPORT ON TITANIC DATASET BY EXPLORATORY DATA ANALYSIS

This project performs an Exploratory Data Analysis (EDA) on the Titanic dataset from Kaggle to uncover key insights about passenger demographics, travel classes, and survival trends. The analysis uses Python, Pandas, Matplotlib, and Seaborn to explore and visualize relationships within the data.

## Load the Dataset

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

## Basic Exploration

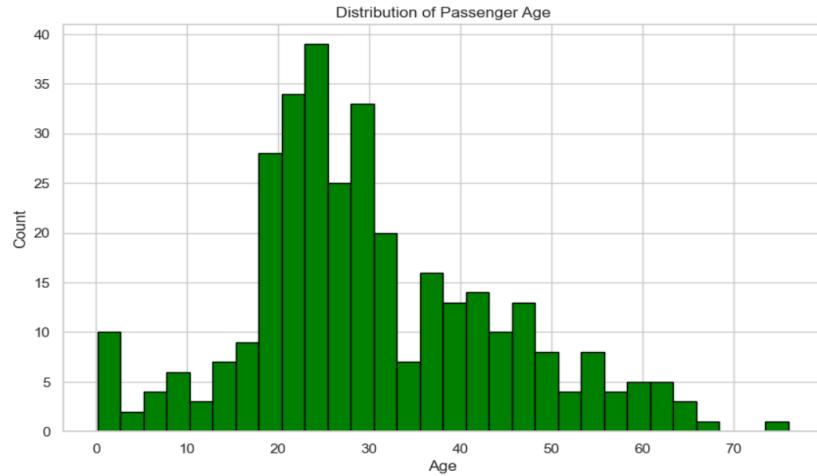
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   PassengerId     418 non-null    int64  
 1   Survived        418 non-null    int64  
 2   Pclass          418 non-null    int64  
 3   Name            418 non-null    object  
 4   Sex             418 non-null    object  
 5   Age             332 non-null    float64 
 6   SibSp          418 non-null    int64  
 7   Parch          418 non-null    int64  
 8   Ticket         418 non-null    object  
 9   Fare            417 non-null    float64 
 10  Cabin           91 non-null    object  
 11  Embarked        418 non-null    object  
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
(418, 12)
```

## Observation:

- Note missing columns (like Age, Cabin, Embarked).
- Mention that dataset contains passenger-level data with survival info.

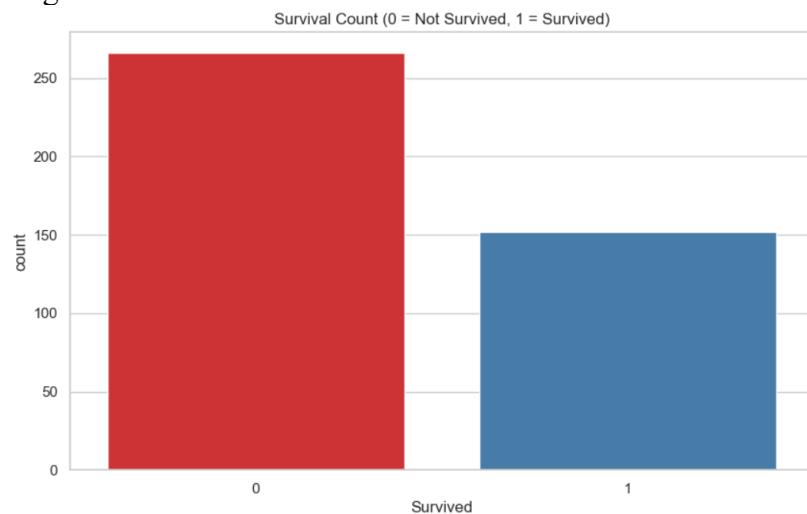
## Univariate Analysis

### a. Numerical Columns



**Observation:** Younger passengers are more common; few elderly passengers.

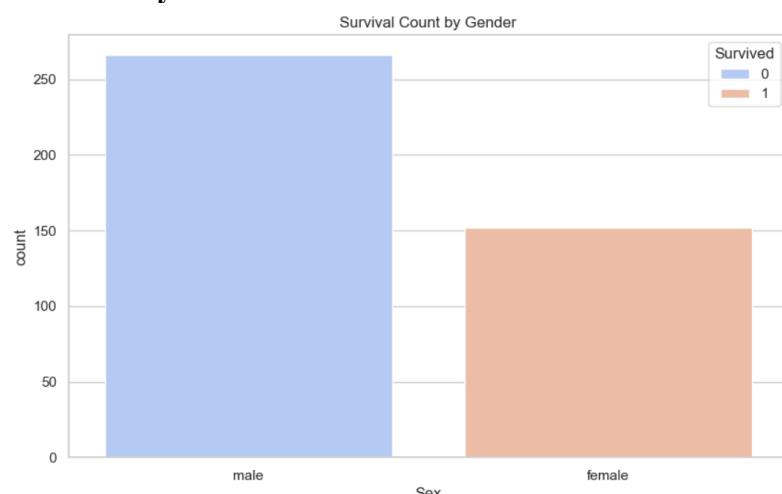
### b. Categorical Columns



**Observation:** Fewer passengers survived compared to non-survivors.

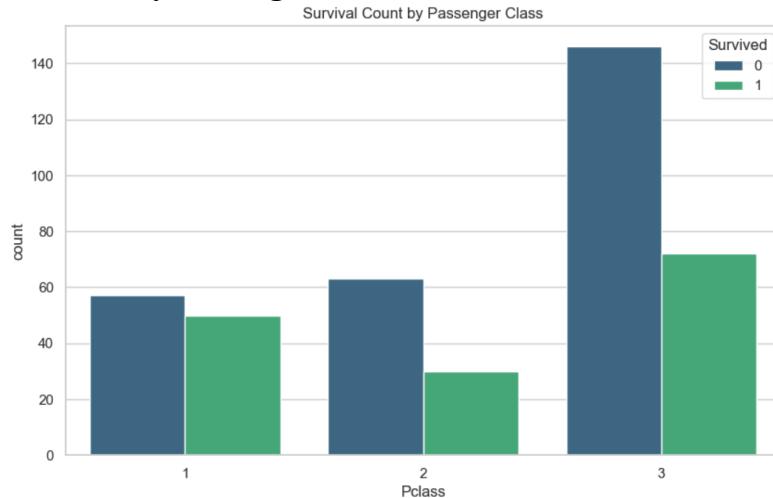
## Bivariate Analysis

### a. Survival by Gender



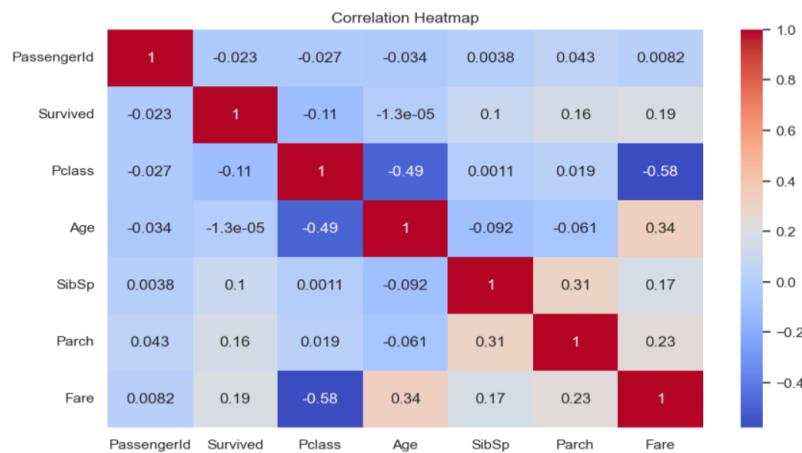
**Observation:** Female survival rate is significantly higher than male.

## b. Survival by Passenger Class



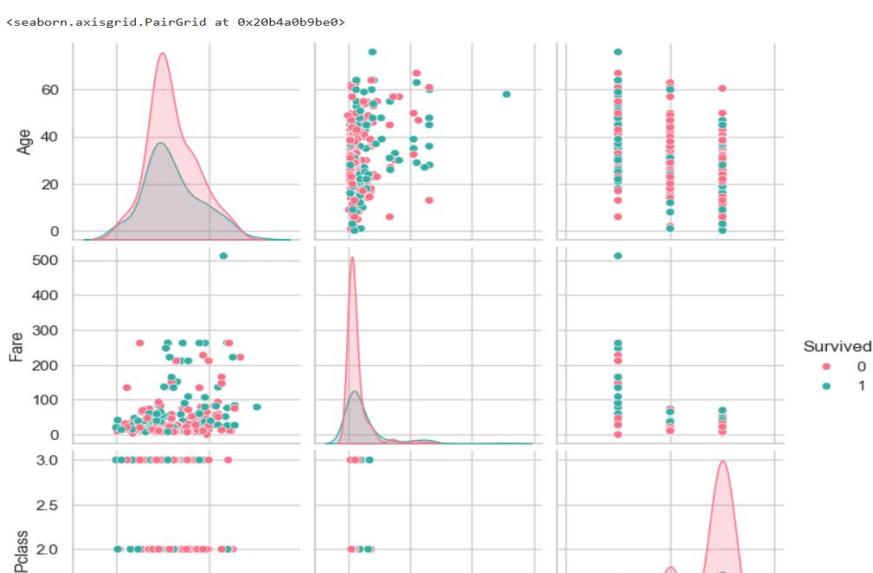
**Observation:** Passengers in 1st class had higher survival rates.

## Correlation Analysis



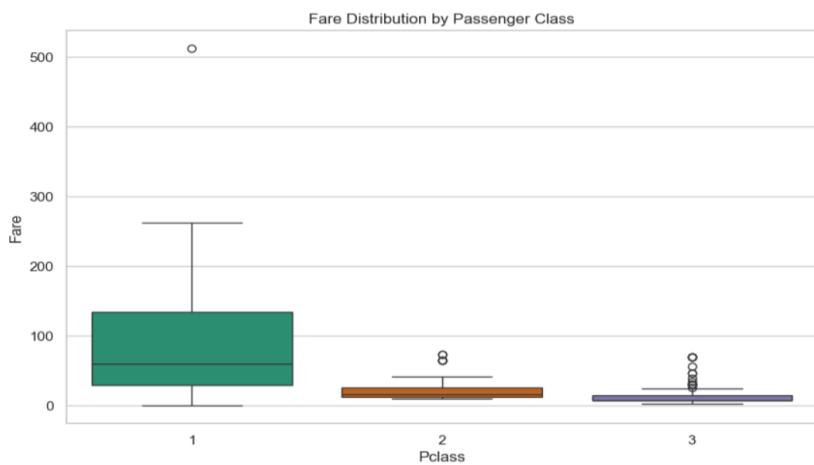
**Observation:** Fare and Pclass show moderate correlation with survival.

## Pairplot



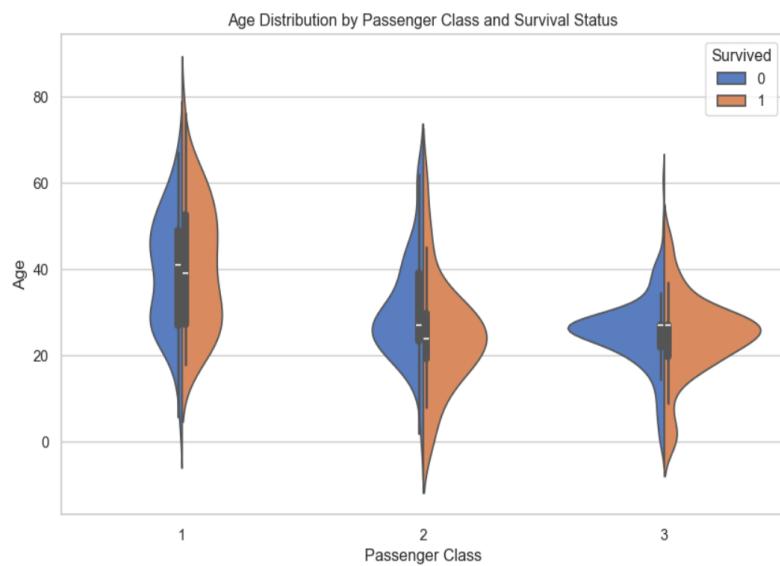
**Observation:** Higher fares often correspond to survival and 1st class

## Outlier Detection (Boxplots)



**Observation:** 1st class fares have wider range; few high-value outliers.

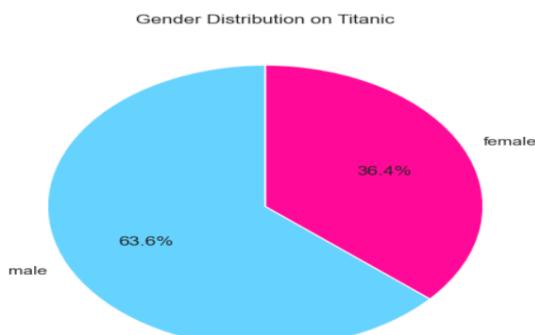
## Violin Plot (Distribution + Class Comparison)



## Observation:

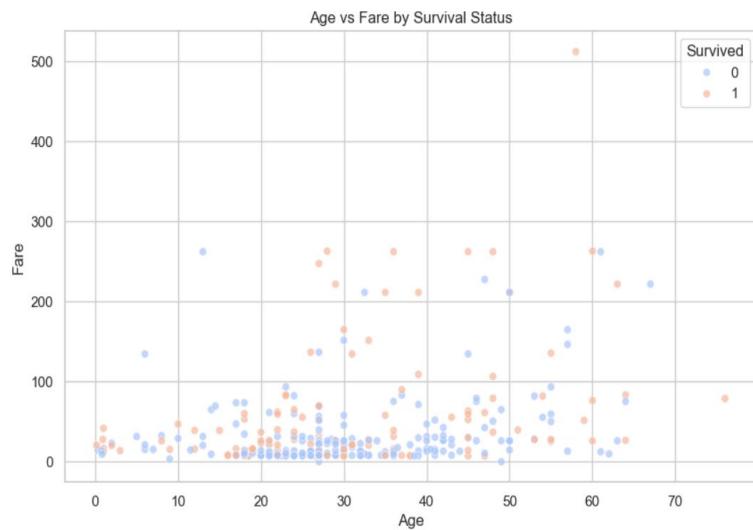
Younger passengers were mostly from 3rd class, while 1st class passengers had broader age distribution and higher survival rates.

## Pie Chart (Gender Proportion)

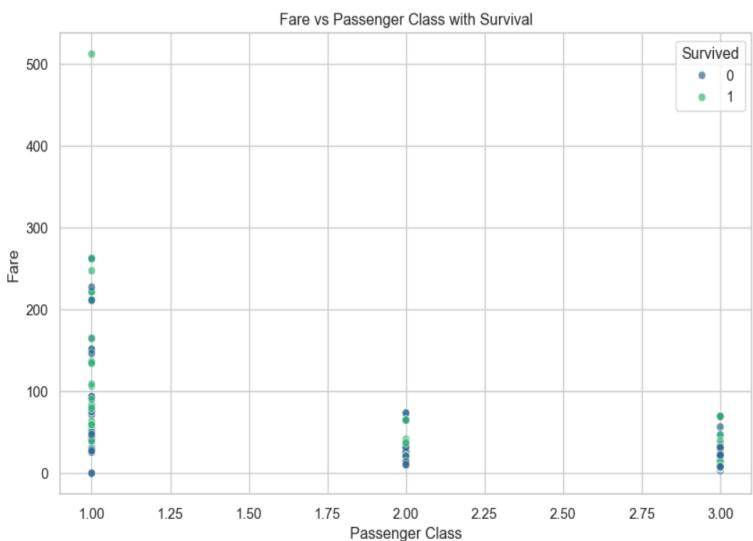


**Observation:**

Male passengers made up the majority, but females had a higher survival rate.

**Scatter Plot (Fare vs. Age, colored by Survival)****Observation:**

Higher fares (wealthier passengers) correspond to better survival chances, regardless of age.

**Scatter Plot (Fare vs Passenger Class)****Observation:**

1st class passengers paid higher fares and had higher survival rates.