

# Project title: Exploratory Data Analysis on titanic dataset

## By-Sneha Maheshwari

Dataset- Used Titanic dataset

Objective of the analysis- Finding patterns ,trends, anomalies.

Tools Used:Pandas, Matplotlib,Seaborn

## Data Overview

```
titanic.sample(5)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
278	279	0	3	Rice, Master. Eric	male	7.0	4	1	382652	29.1250	NaN	Q
528	529	0	3	Salonen, Mr. Johan Werner	male	39.0	0	0	3101296	7.9250	NaN	S
141	142	1	3	Nysten, Miss. Anna Sofia	female	22.0	0	0	347081	7.7500	NaN	S
195	196	1	1	Lurette, Miss. Elise	female	58.0	0	0	PC 17569	146.5208	B80	C
137	138	0	1	Futrelle, Mr. Jacques Heath	male	37.0	1	0	113803	53.1000	C123	S

```
[3]: titanic.shape
```

```
[3]: (891, 12)
```

```
titanic.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
titanic.info()
```

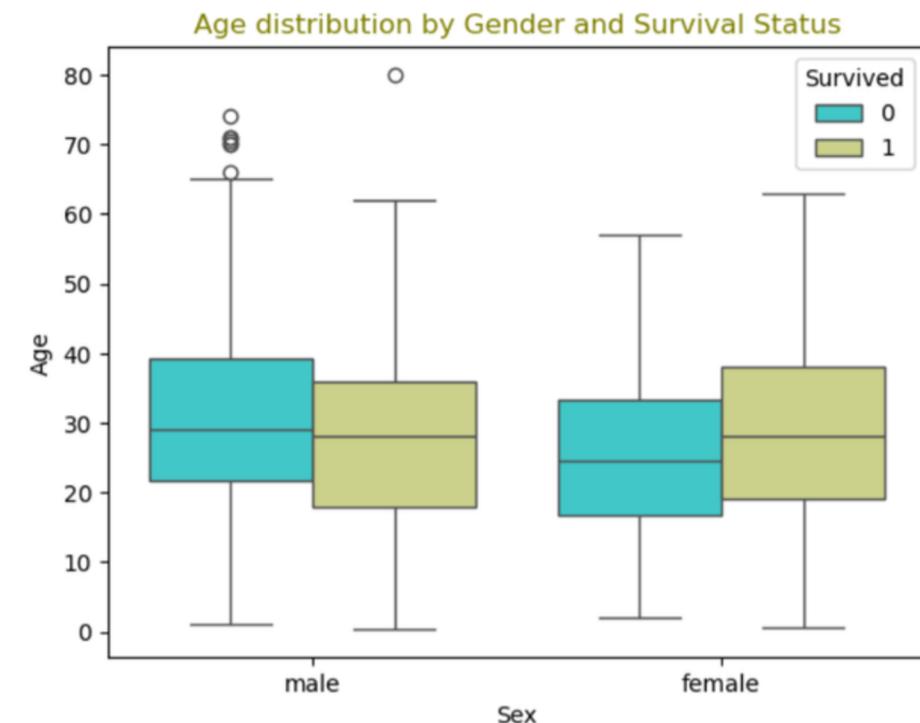
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId 891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare          891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```



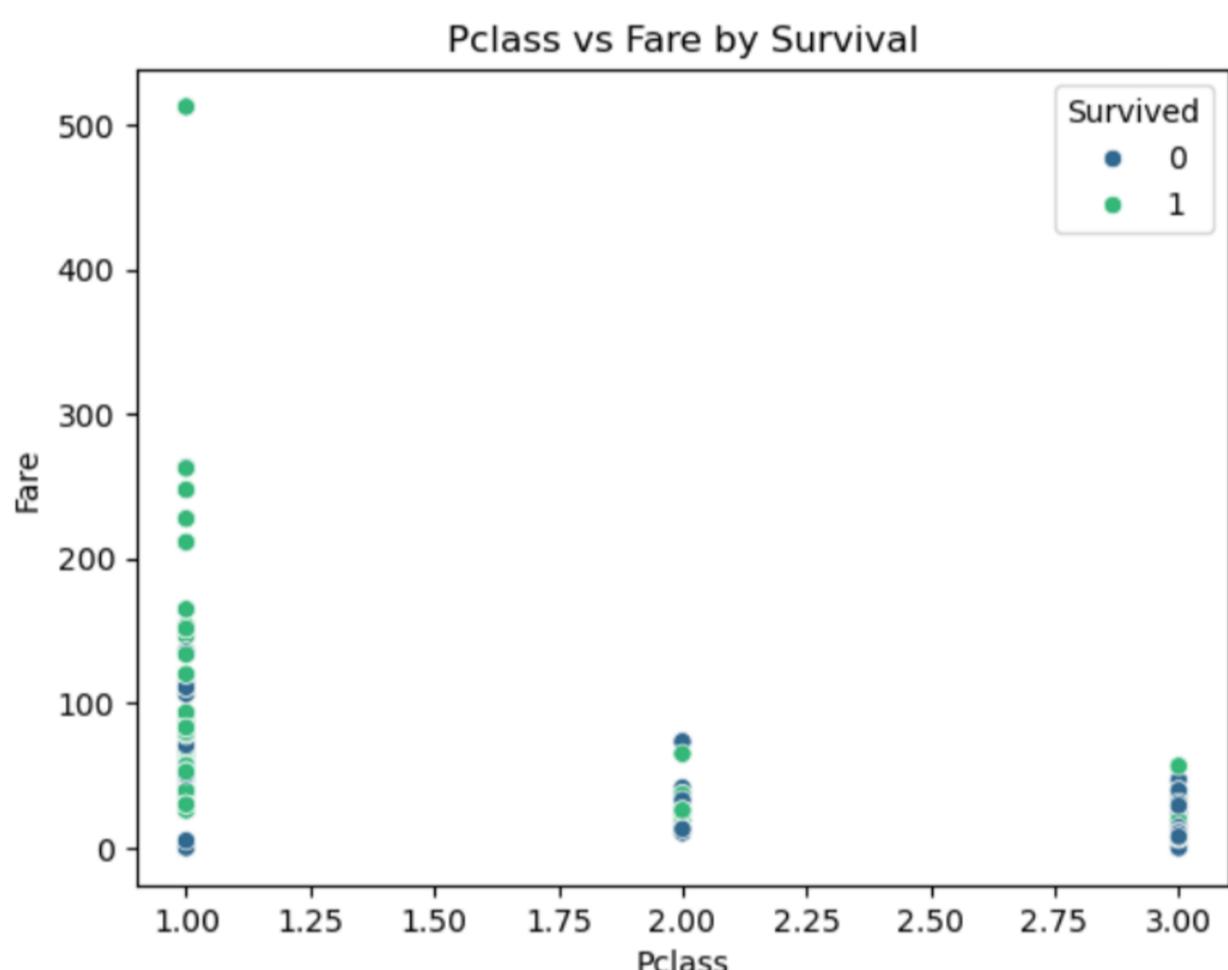
People in 1st class paid much higher fares than those in 2nd or 3rd class.

In 1st class, women paid a little more on average than men.

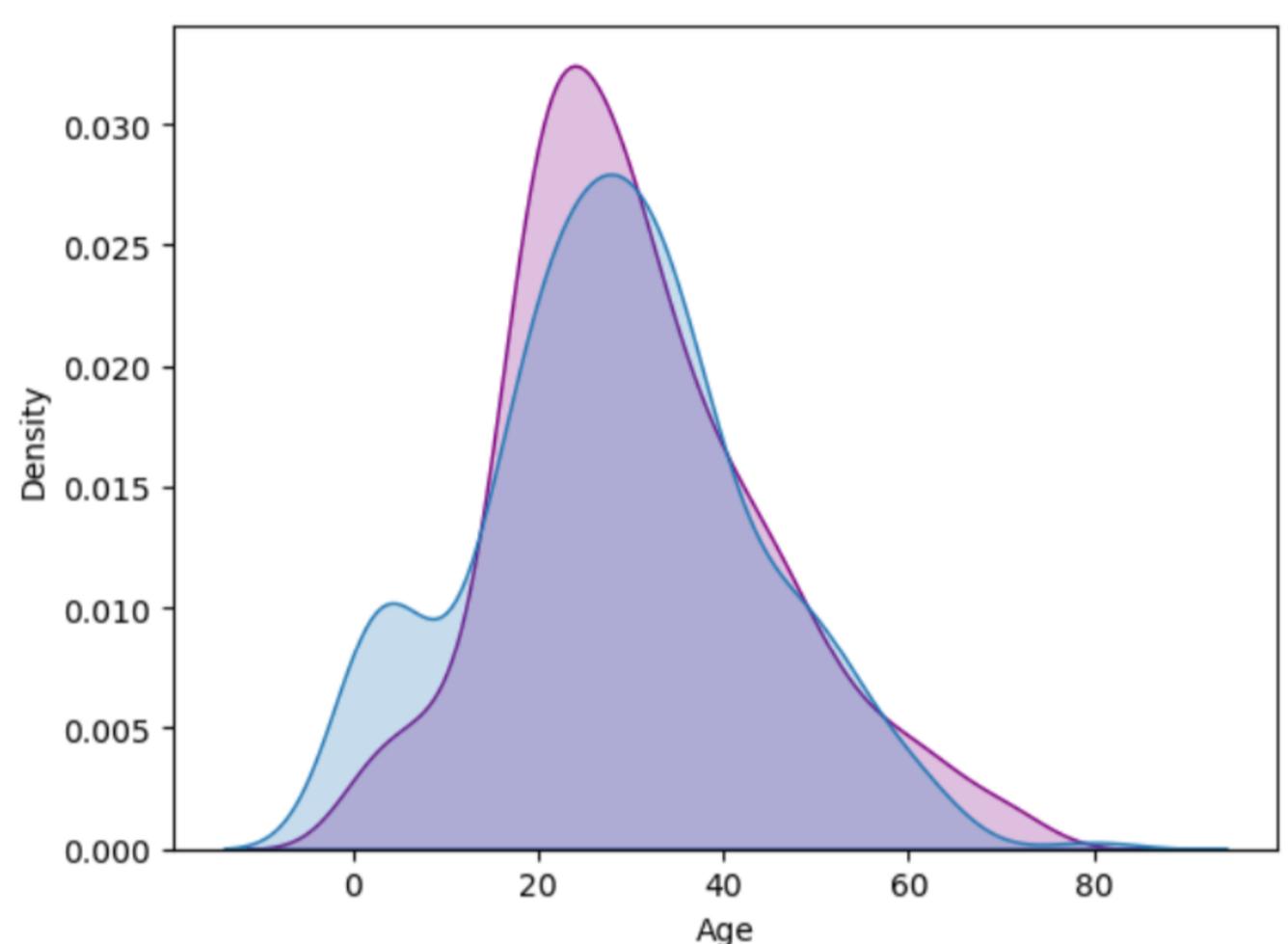
In 2nd and 3rd class, the fares for men and women were almost the same.



Females have much higher survival rate than males. Most passengers were between 20 and 40 years old, regardless of survival status. There are few outliers in males showing that there are few older passengers (around 70-80 years).



Higher fares are mostly paid by 1st class passengers, and many of them survived. Lower fares are linked to the 3rd class passengers, and survival rate seems much lower. 2nd class fares are in between and have a mixed survival rate. Overall, passengers who paid higher fare had a better chance of survival.



Most of the passengers are between the age of 20-30. The survival rate is more around the age of 0-10 showing priority had been given to the children first.

## Key Findings-

1. First class passenger had much chance of survival.
2. Younger passengers tended to survive more often.
3. About 38 % of passengers survived.
4. Non-survivors were roughly twice as many survivors.
5. Females have higher survival rate than males.