

# Exploratory Data Analysis Of Titanic Dataset by Archana..

This is an EDA on the Titanic dataset

## First 5 rows of the data

	PassengerId	Survived	Pclass	Name	Sex	Age	S
0	1	0	3	Braund, Mr. Owen Harris	male	22	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	
2	3	1	3	Heikkinen, Miss. Laina	female	26	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	
4	5	0	3	Allen, Mr. William Henry	male	35	

## First 10 rows of the data

	PassengerId	Survived	Pclass	Name	Sex	Age	S
0	1	0	3	Braund, Mr. Owen Harris	male	22	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	
2	3	1	3	Heikkinen, Miss. Laina	female	26	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	
4	5	0	3	Allen, Mr. William Henry	male	35	
5	6	0	3	Moran, Mr. James	male	None	
6	7	0	1	McCarthy, Mr. Timothy J	male	54	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	

## Statistical summary of the data

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891	891	891	714	891	891	891
mean	446	0.3838	2.3086	29.6991	0.523	0.3816	32.2042
std	257.3538	0.4866	0.8361	14.5265	1.1027	0.8061	49.6934
min	1	0	1	0.42	0	0	0
25%	223.5	0	2	20.125	0	0	7.9104
50%	446	0	3	28	0	0	14.4542
75%	668.5	1	3	38	1	0	31
max	891	1	3	80	8	6	512.3292

Missing Values.

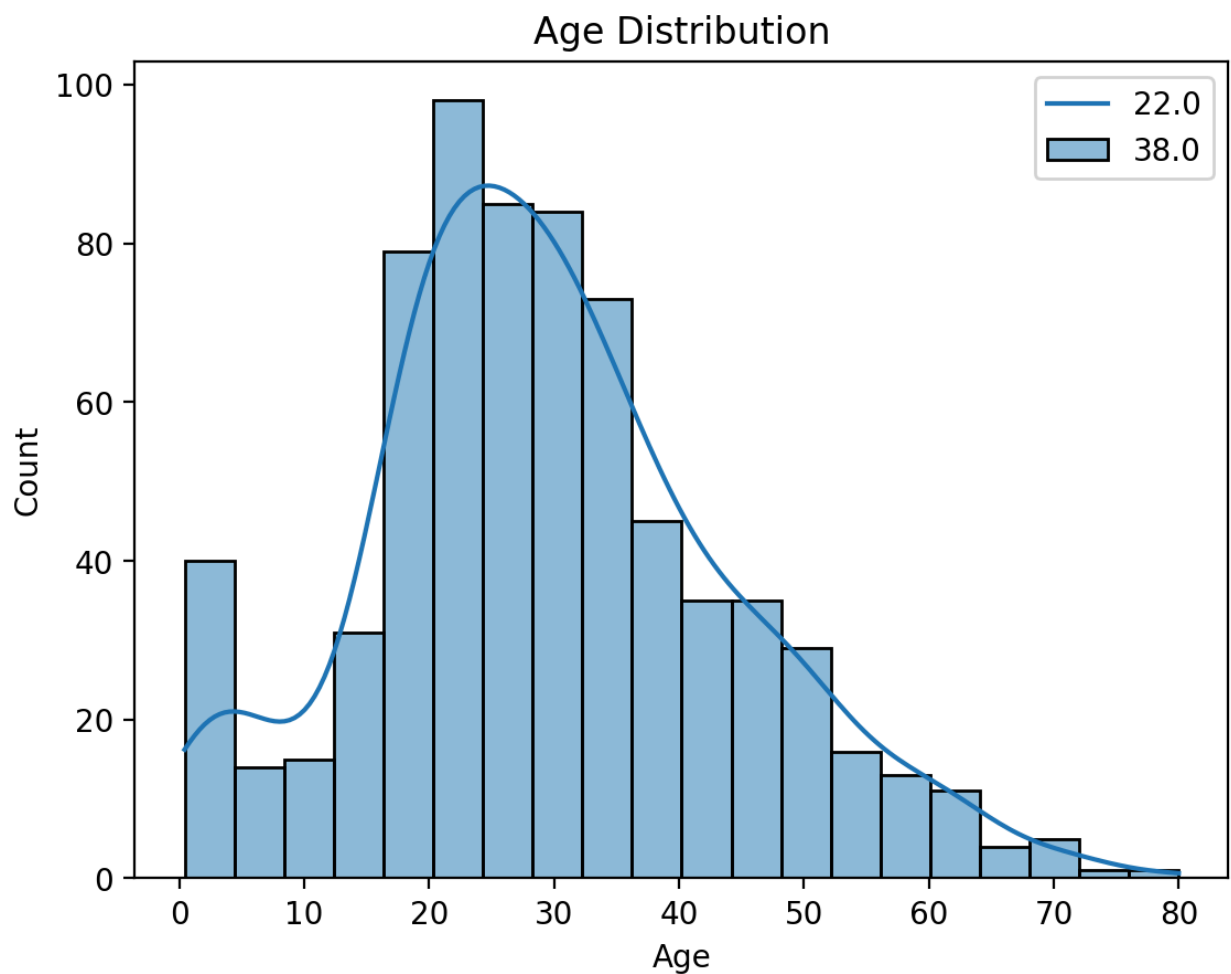
	0
PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0

- ☐ Fill the missing Age values with median
- ☐ Fill missing cabin with mode
- ☐ Drop duplicates from the data

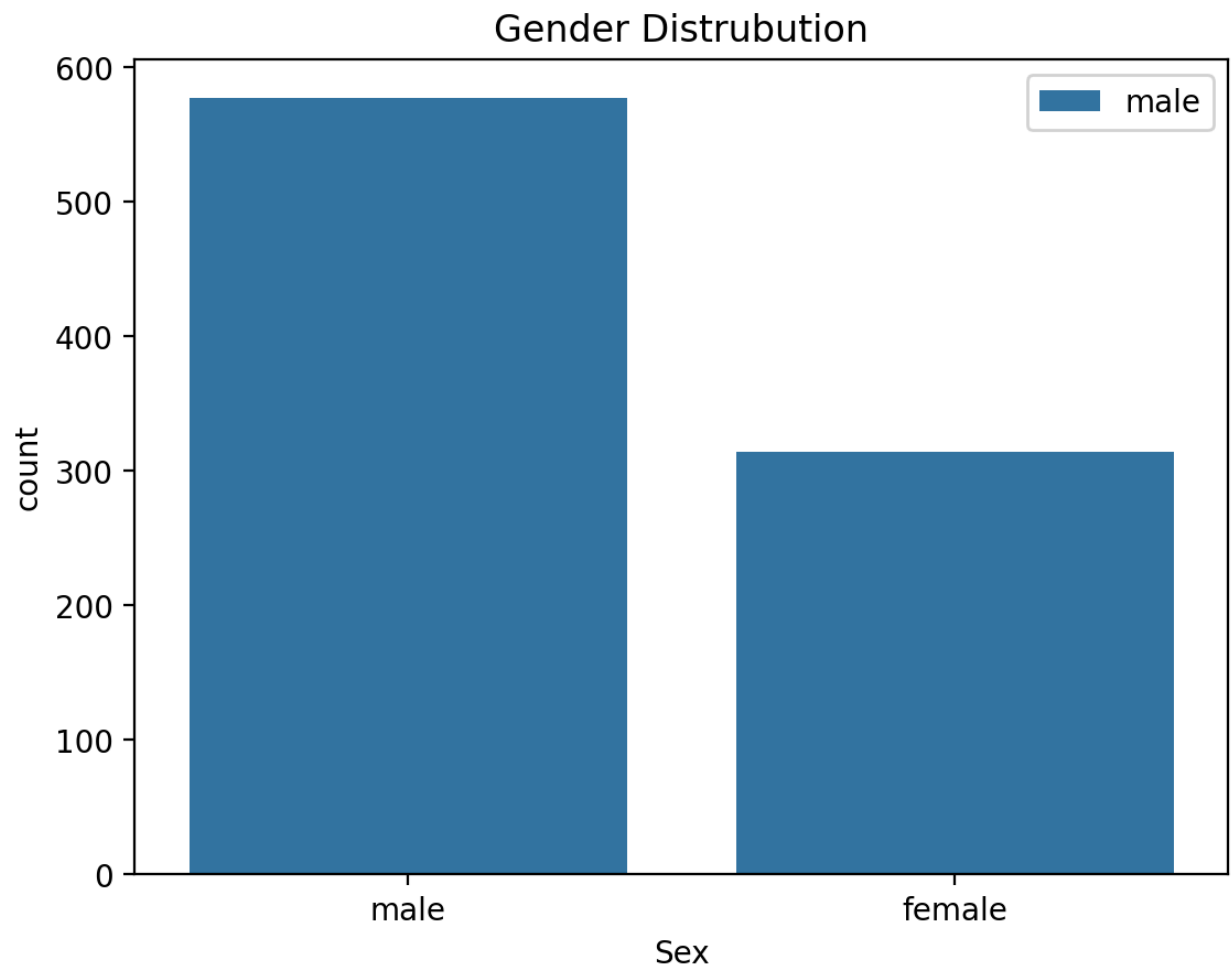
Cleaned Dataset

	PassengerId	Survived	Pclass	Name	Sex	Age	S
0	1	0	3	Braund, Mr. Owen Harris	male	22	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	
2	3	1	3	Heikkinen, Miss. Laina	female	26	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	
4	5	0	3	Allen, Mr. William Henry	male	35	
5	6	0	3	Moran, Mr. James	male	None	
6	7	0	1	McCarthy, Mr. Timothy J	male	54	
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	

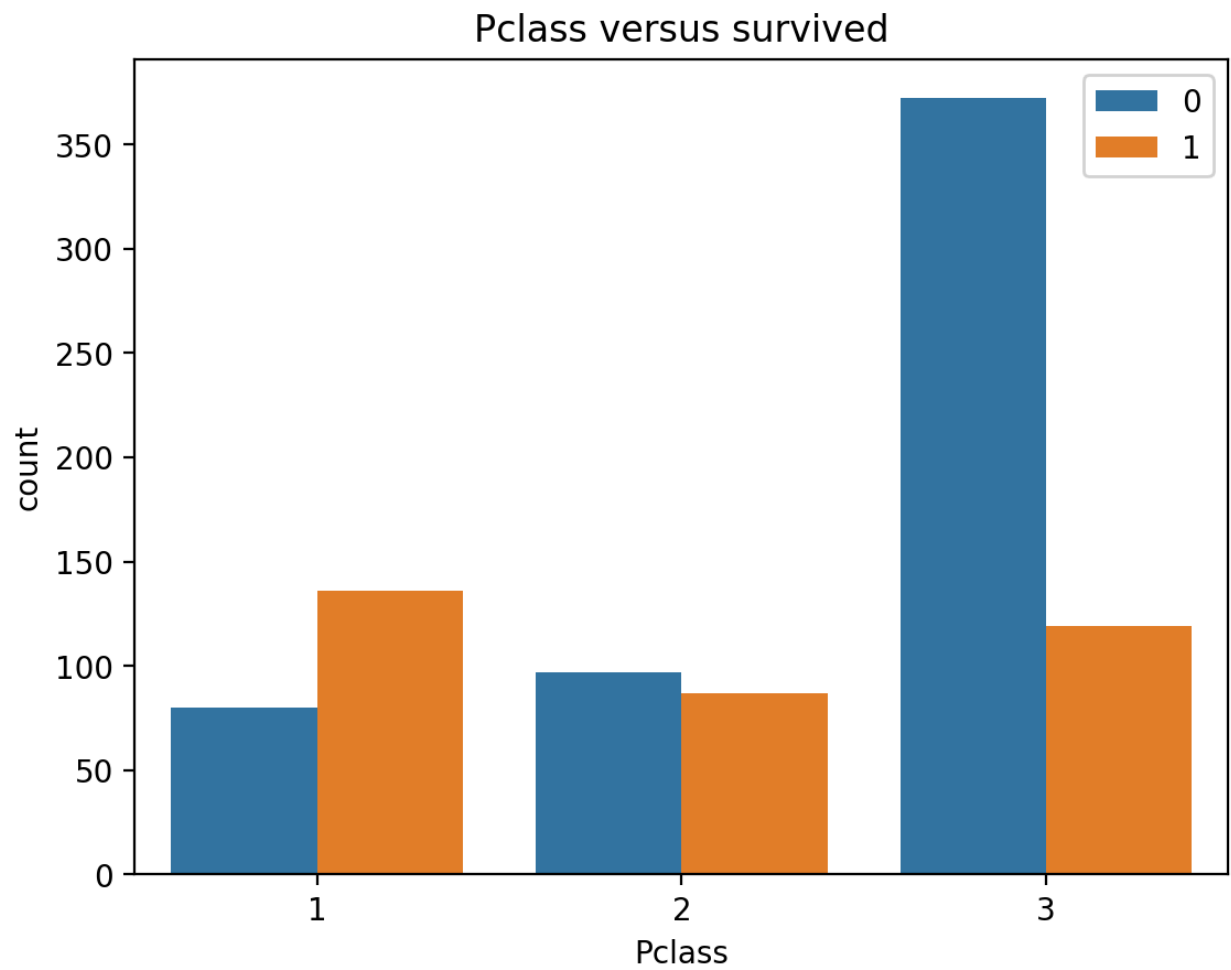
## Age Distribution



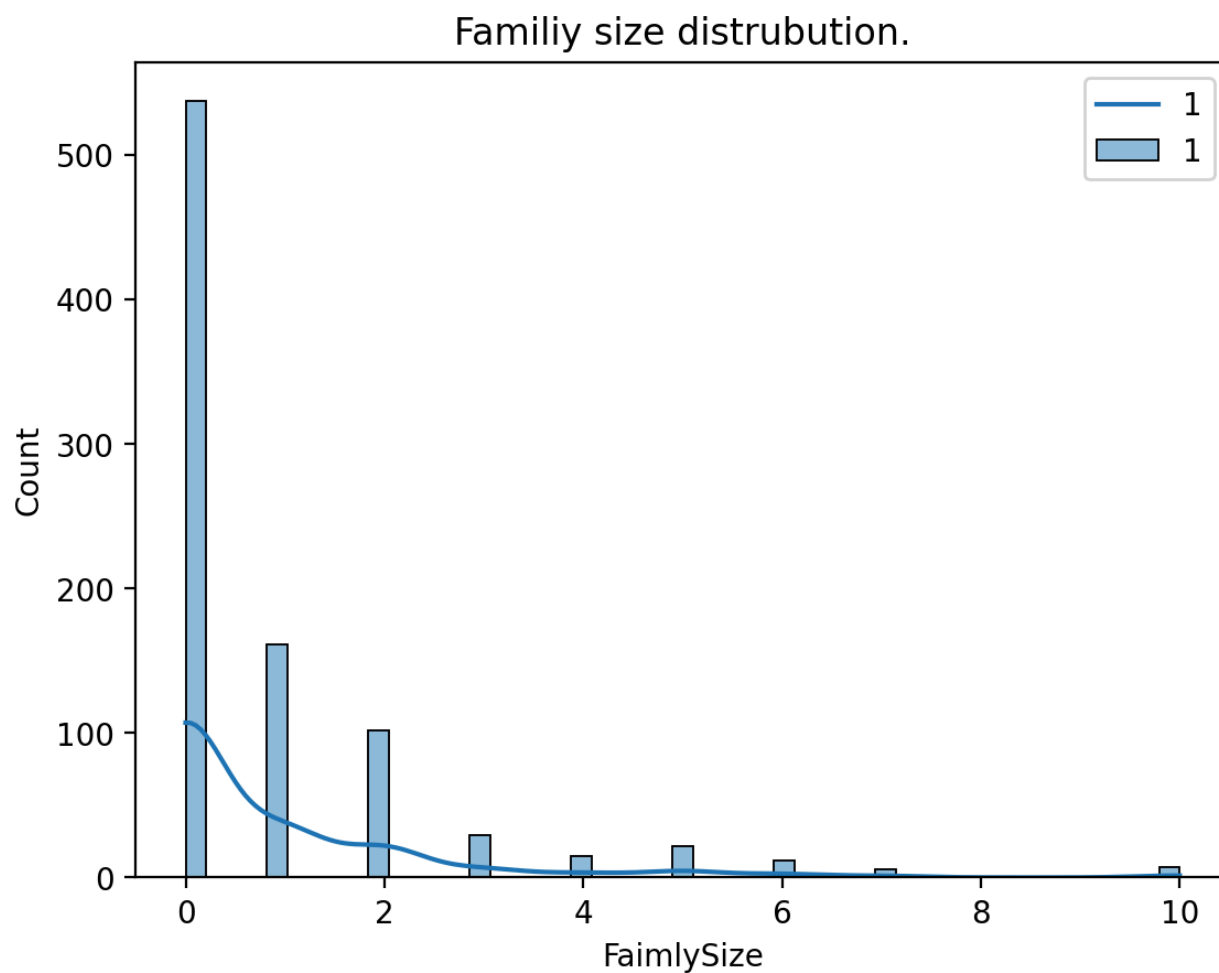
## Gender Distribution



Pclass versus survived



## Feature Engineering: Family size



## Key Insights

- Females have a higher survival rate than males.
- Passengers in 1st class had the highest survival rate.
- The majority of passengers are in Pclass 3.
- Younger passengers tended to survive more often.