

# Exploratory DataAnalysis for Titanic Dataset

This page covers Data analysis captured for Titanic Dataset.

Glimpse of the dataset to understand features/attributes.

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542	None
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16	None
16	17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125	None
17	18	1	2	Williams, Mr. Charles Eugene	male	None	0	0	244373	13	None
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoorte	female	31	1	0	345763	18	None
19	20	1	3	Masselmani, Mrs. Fatima	female	None	0	0	2649	7.225	None
20	21	0	2	Fynney, Mr. Joseph J	male	35	0	0	239865	26	None
21	22	1	2	Beesley, Mr. Lawrence	male	34	0	0	248698	13	D56
22	23	1	3	McGowan, Miss. Anna "Annie"	female	15	0	0	330923	8.0292	None
23	24	1	1	Sloper, Mr. William Thompson	male	28	0	0	113788	35.5	A6
24	25	0	3	Palsson, Miss. Torborg Danira	female	8	3	1	349909	21.075	None

## Let us start by checking for Null/Empty values in the Dataset.

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

## Let us fix the Null values and Empty values.

Choose the way to full empty values of Age Attribute

- ☒ Fill Missing age using Average of other ages
- ☐ Fill Missing age using Median/Middle of other ages
- ☐ Fill Missing age using the most common of all ages

Choose the way to fill empty CABIN information

- ☐ Fill Cabin information with the most common value of the Cabin Feature set
- ☒ Fill Cabin information with default value Unknown Cabin

Choose the way to fill empty embarked information

- ☒ Fill embarked information with the most common value of the embarked Feature set
- ☐ Fill embarked information with the next row value

☐ Fill embarked information with the previous row value

Clean Data Set

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25	Unknow
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925	Unknow
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05	Unknow
5	6	0	3	Moran, Mr. James	male	30	0	0	330877	8.4583	Unknow
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075	Unknow
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333	Unknow
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708	Unknow

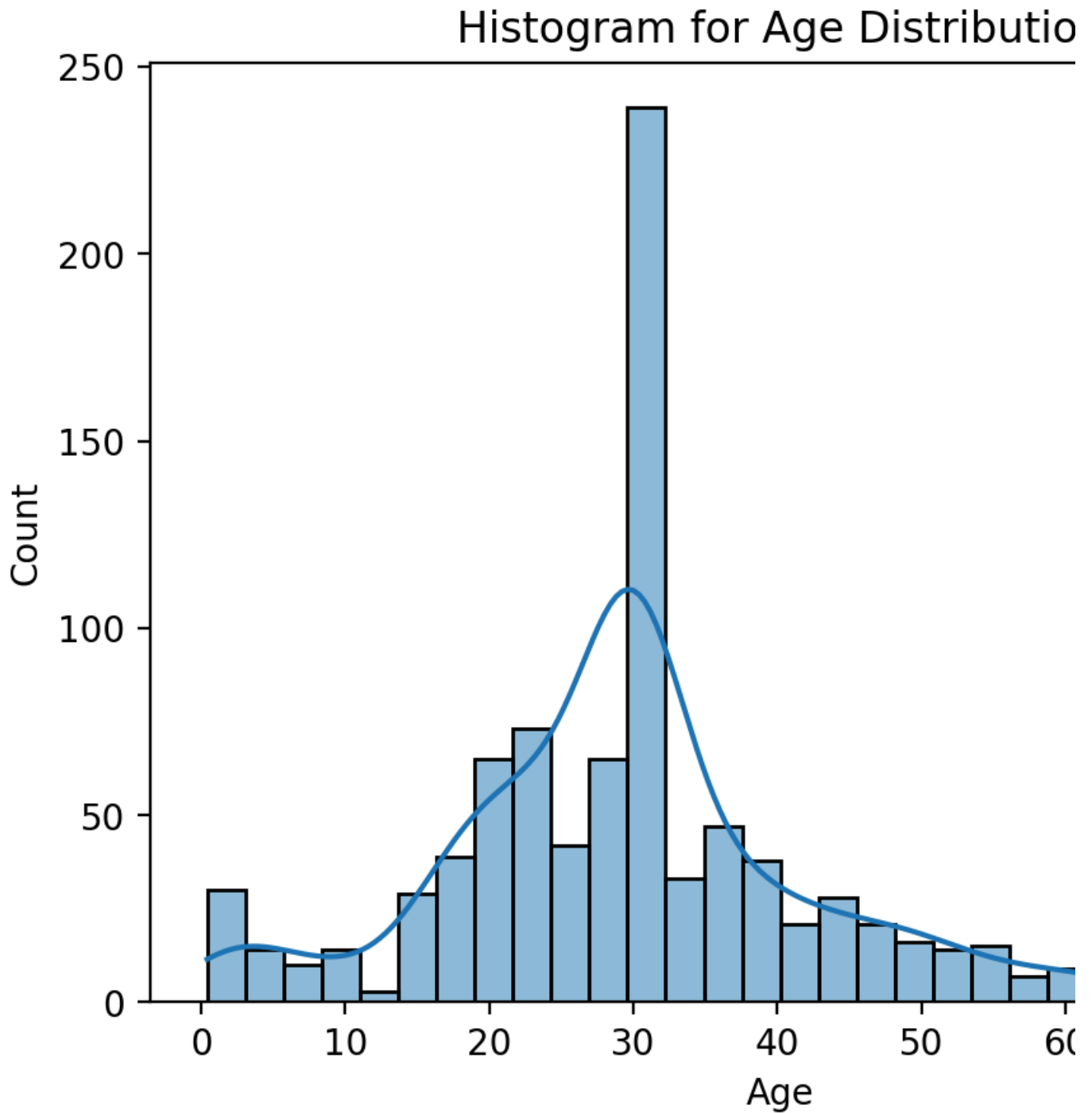
Listing Statistical summary for the Data set

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891	891	891	891	891	891	891
mean	446	0.3838	2.3086	29.7589	0.523	0.3816	32.2042
std	257.3538	0.4866	0.8361	13.0026	1.1027	0.8061	49.6934
min	1	0	1	0.42	0	0	0
25%	223.5	0	2	22	0	0	7.9104
50%	446	0	3	30	0	0	14.4542
75%	668.5	1	3	35	1	0	31
max	891	1	3	80	8	6	512.3292

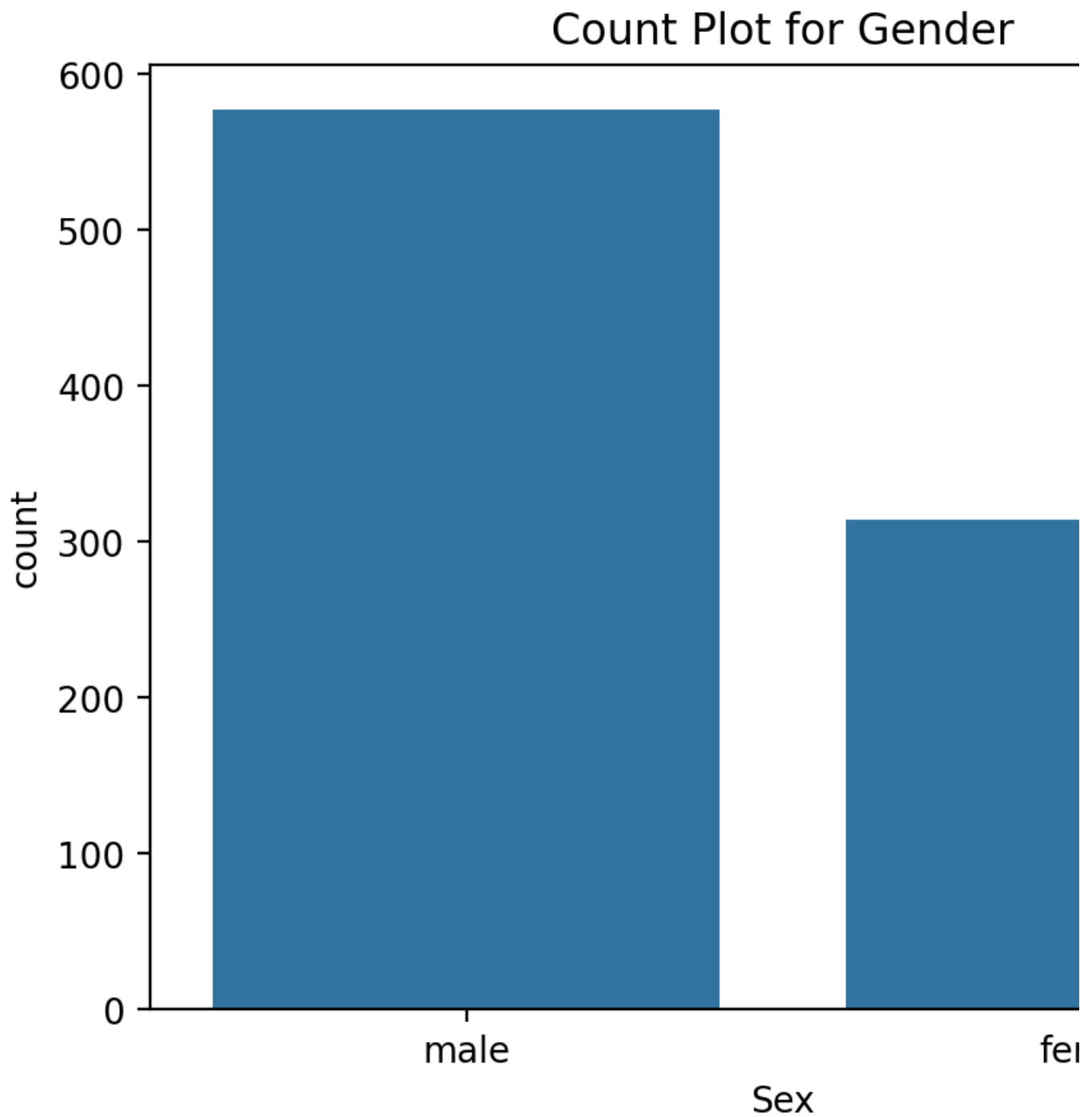
Age Distribution

Choose KDE Curve for Age Distibution Histogram?

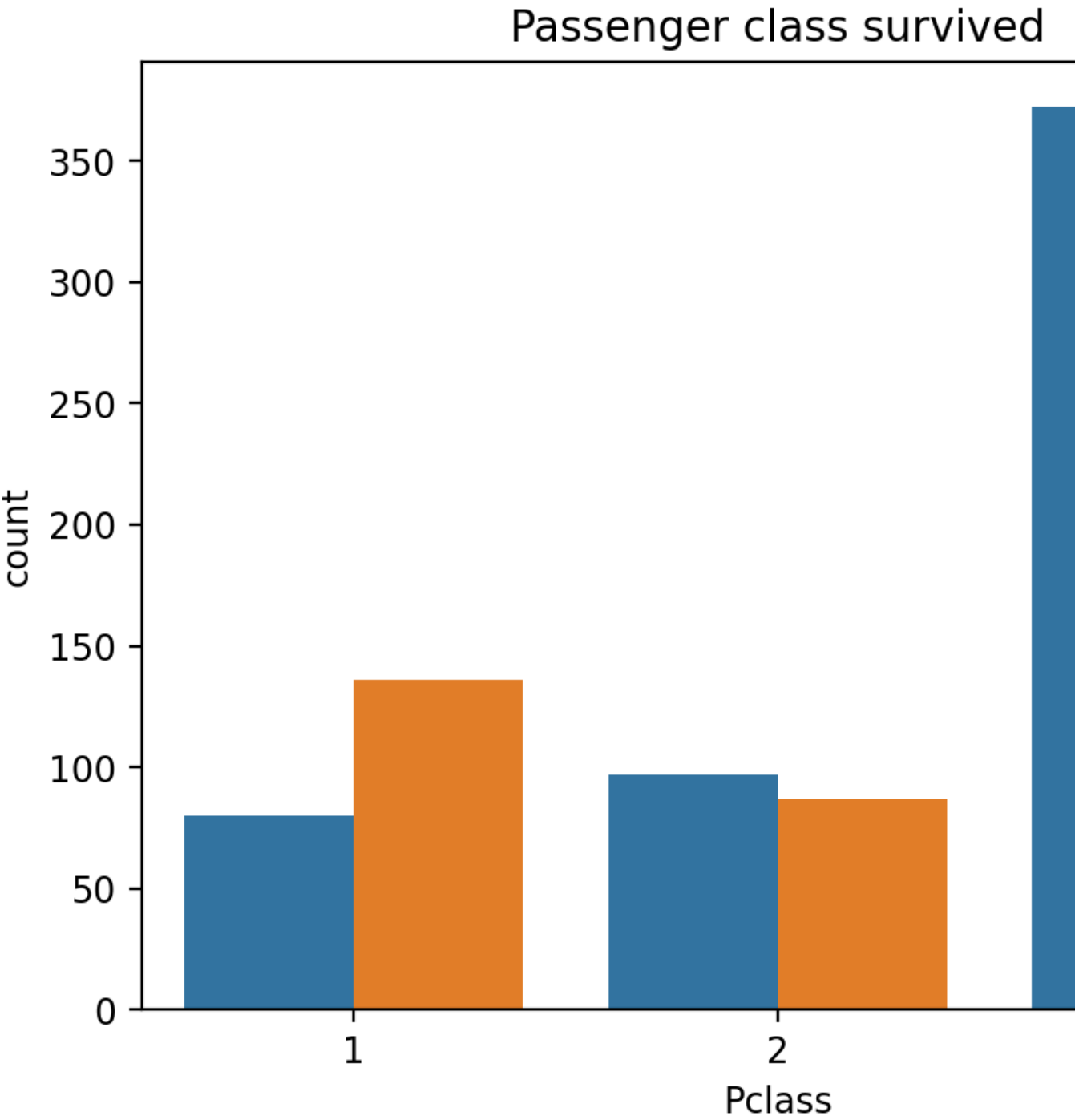
- ☒ Yes
- ☐ No



Gender Distribution

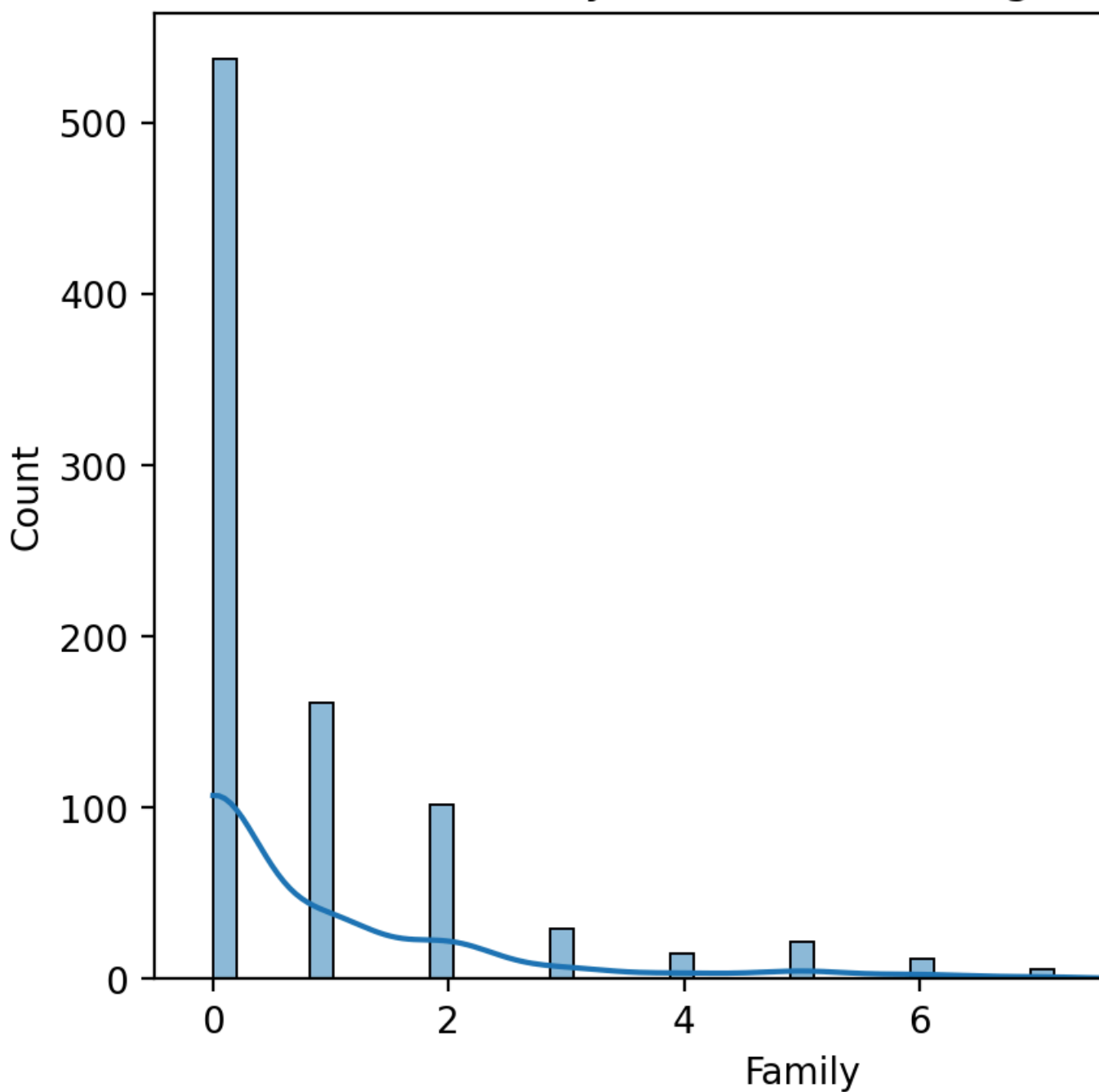


Passenger Class Survived.



Family Distribution

## Family Distribution Histogram



### Insights collected from the Above Visualizations are :-

- Age distribution is mostly around ages 30-35 years
- Male Passengers onboarded are more than Females
- Class 3 Survival is more
- Family members are mostly 2 or less in a given family.