

Titanic

Exploratory data analysis



Andreea STROIA
Hrikita KATHURIA

Agenda



**Insights
on data**

**Missing
values**

**Data
visualization**

Correlations



Insights on data

The data set describes the survival status of individual passengers on the Titanic

- 12 columns
- 891 rows

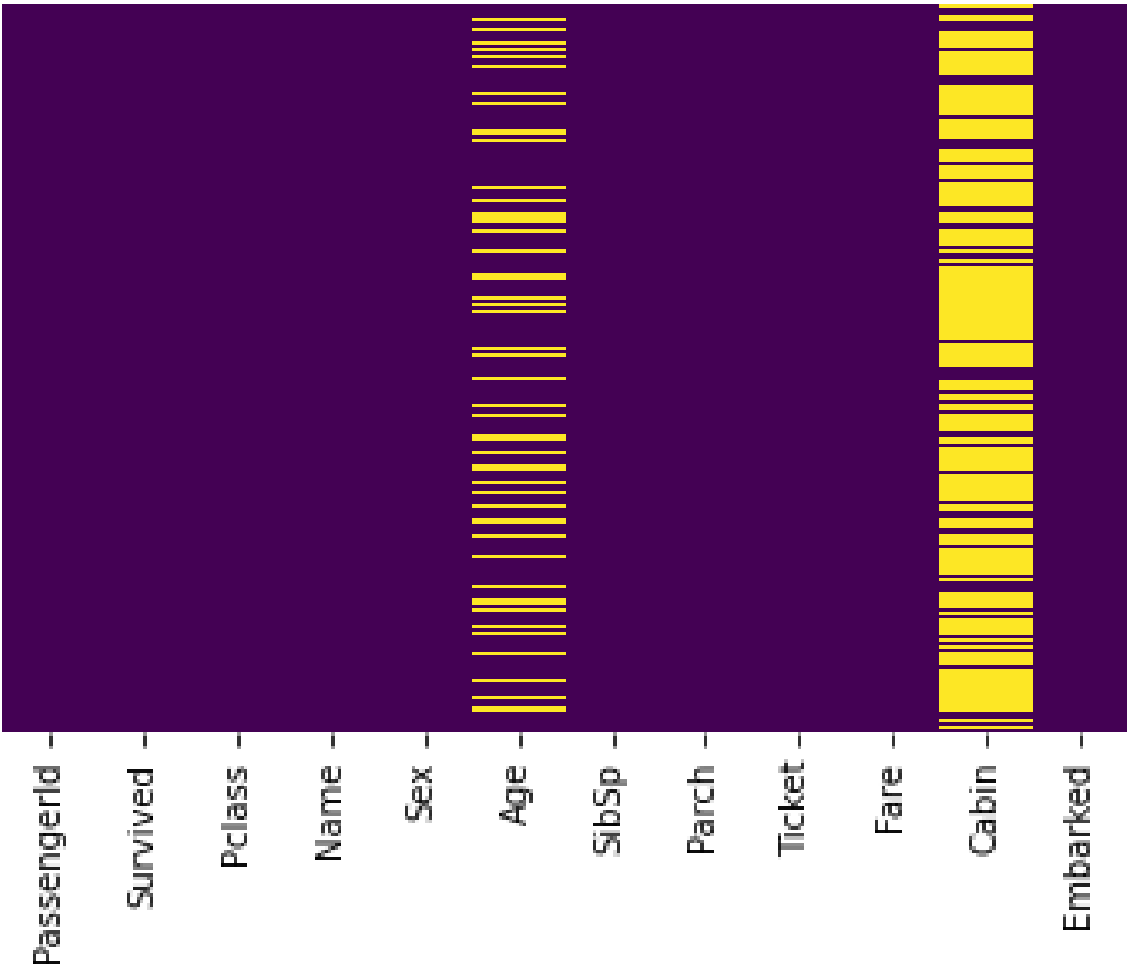
- 7 numerical variables
- 5 categorical variables

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

Missing values

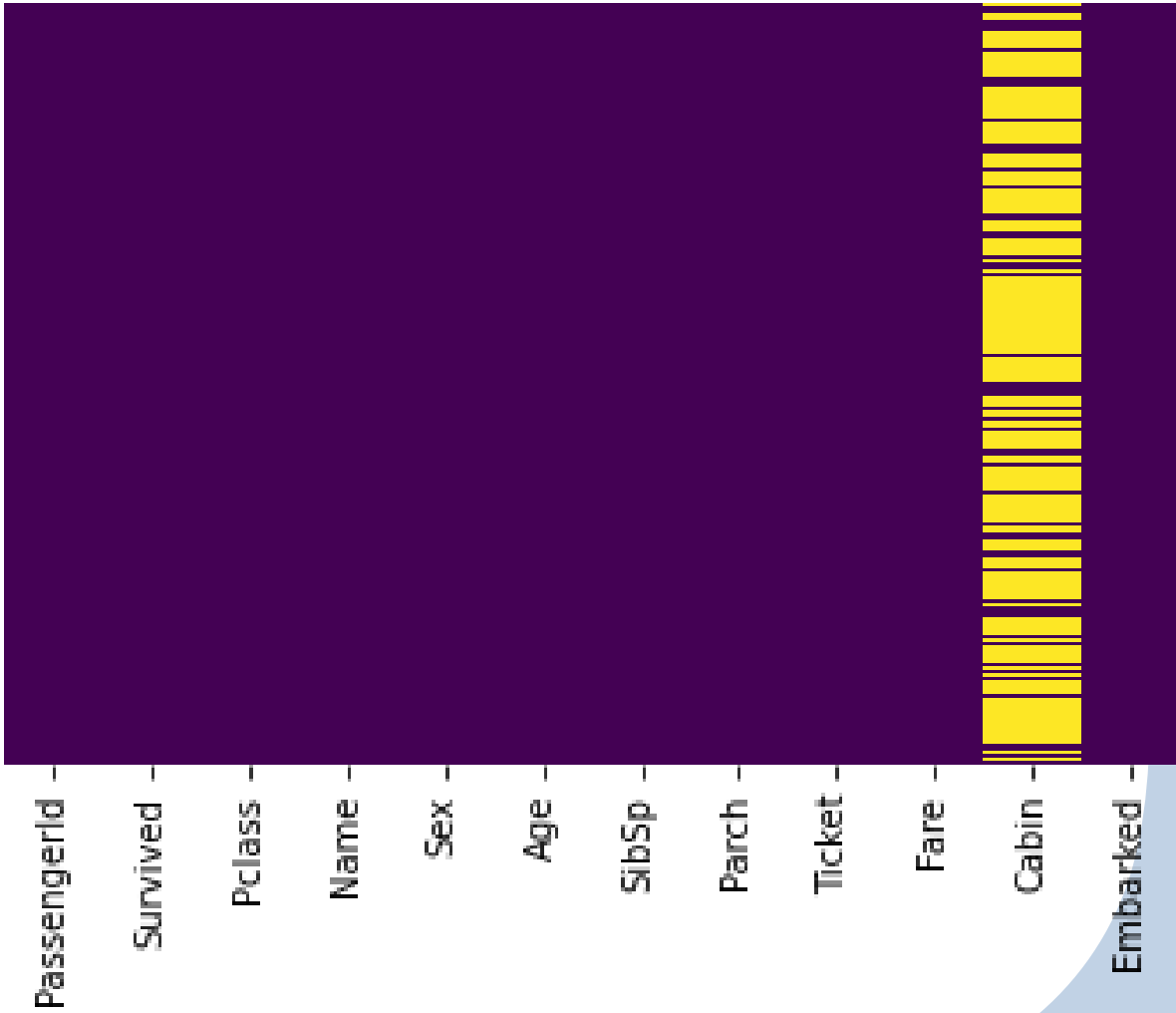
I

Checking for missing values using heatmap



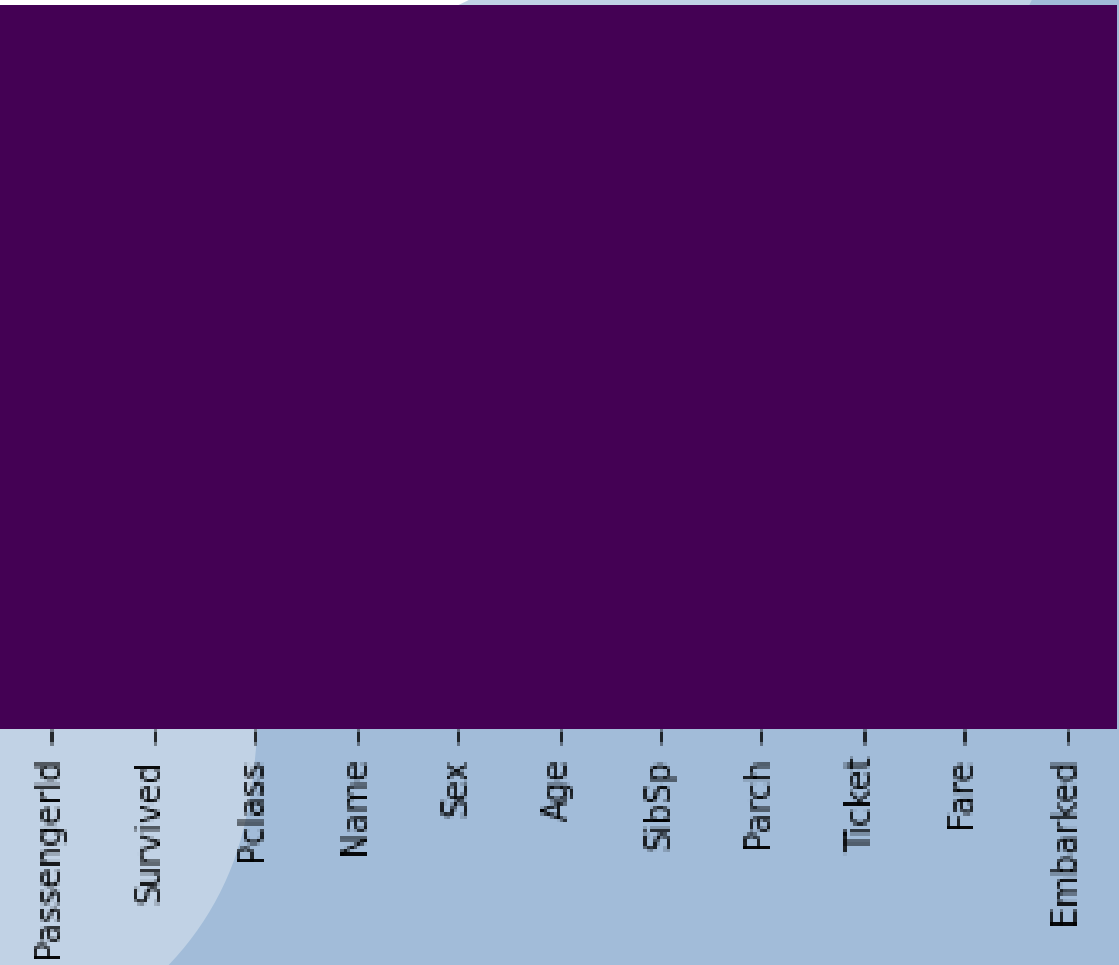
II

Replacing the mean for "Age" column
Replacing the 2 missing values from column "Embarked"



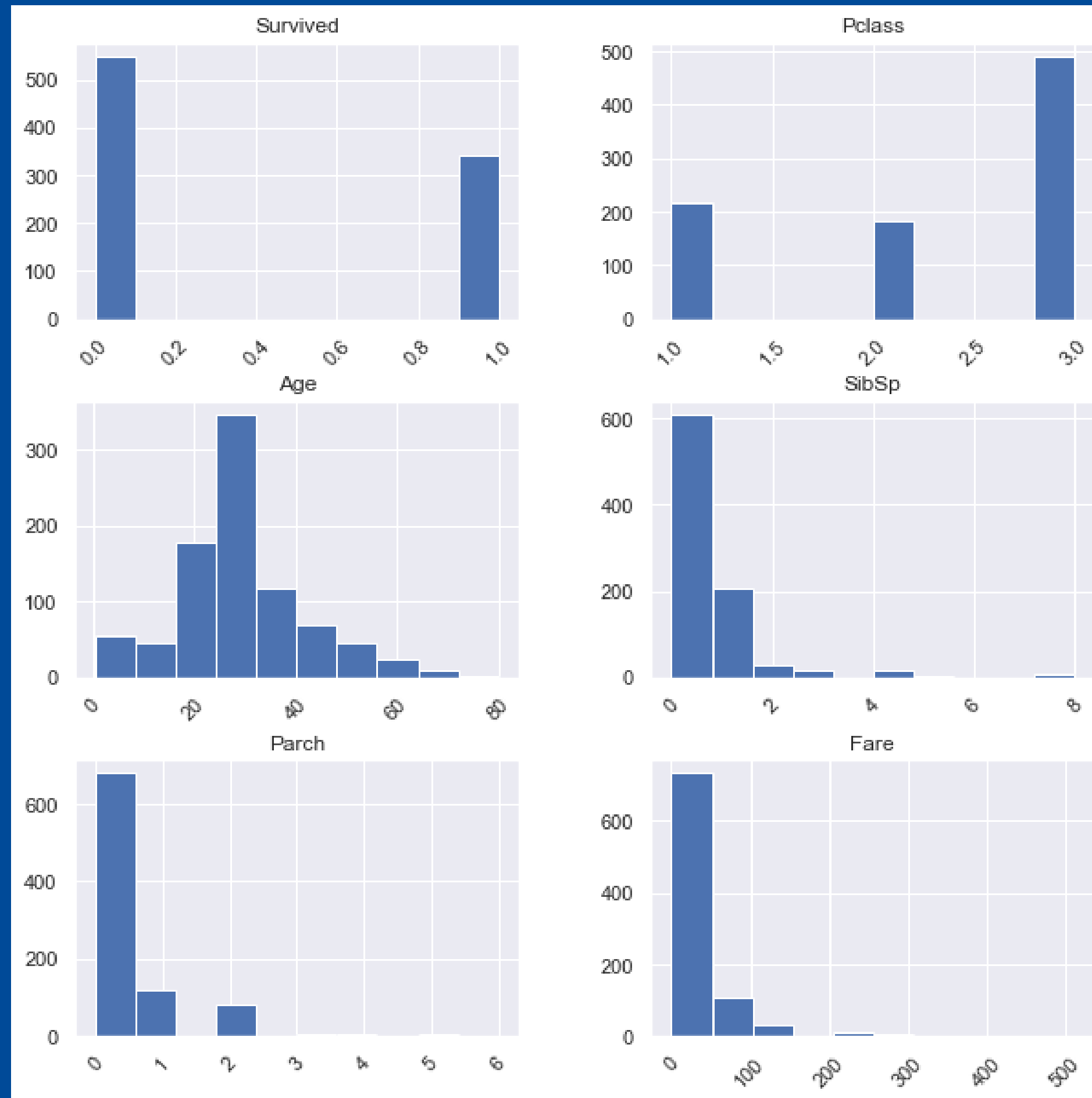
III

Dropping the column "Cabin"



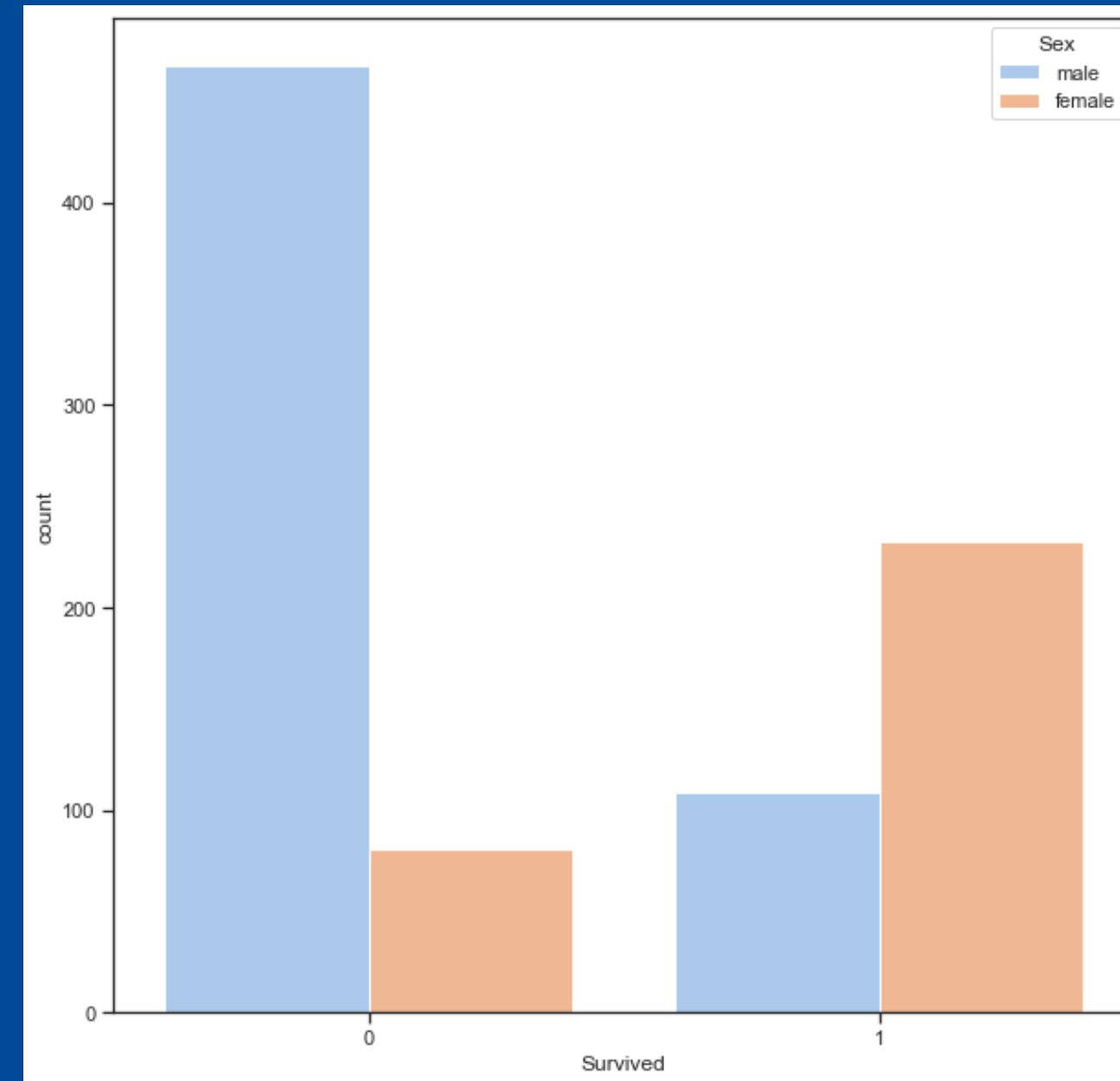
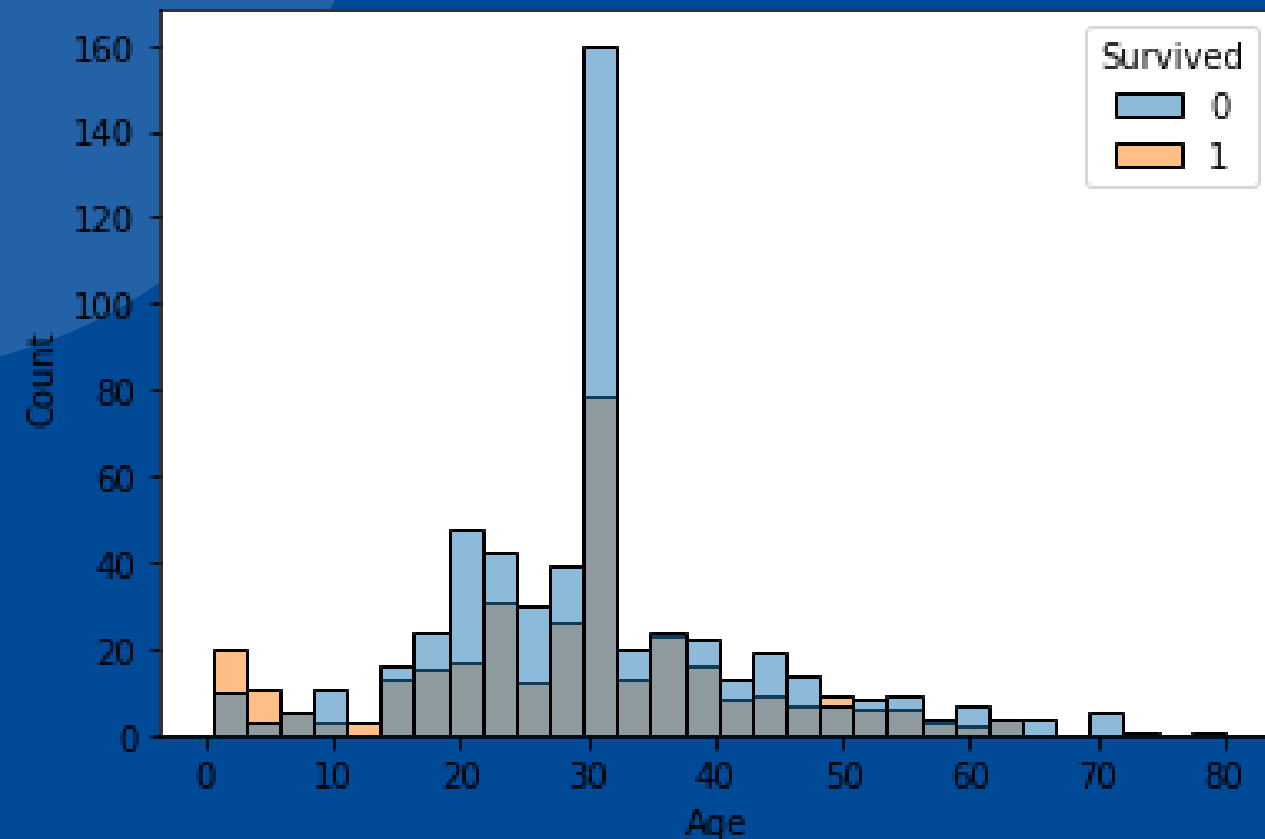
Data vizualization

Histograms



Who survived Titanic?

Let's find out..

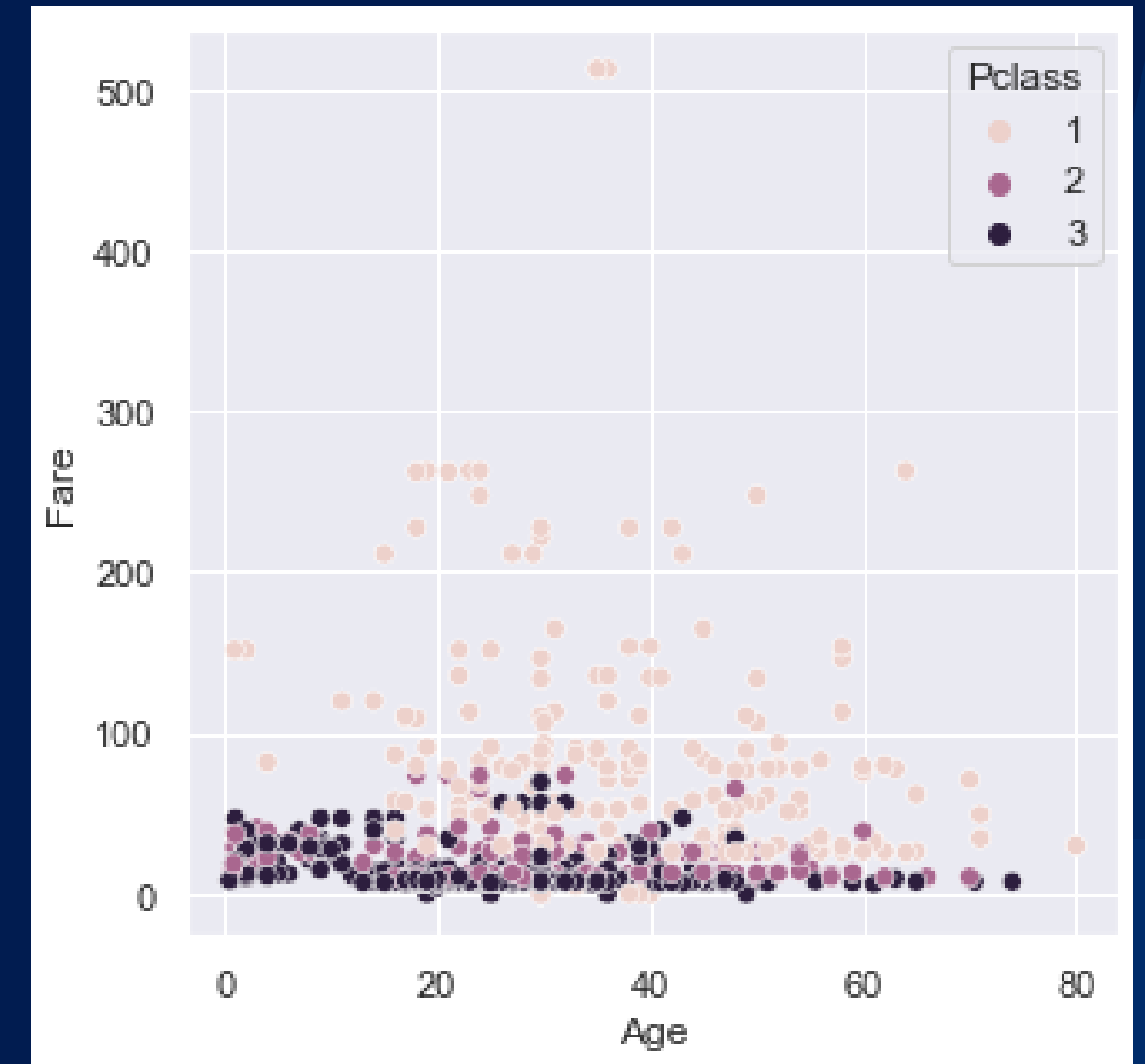
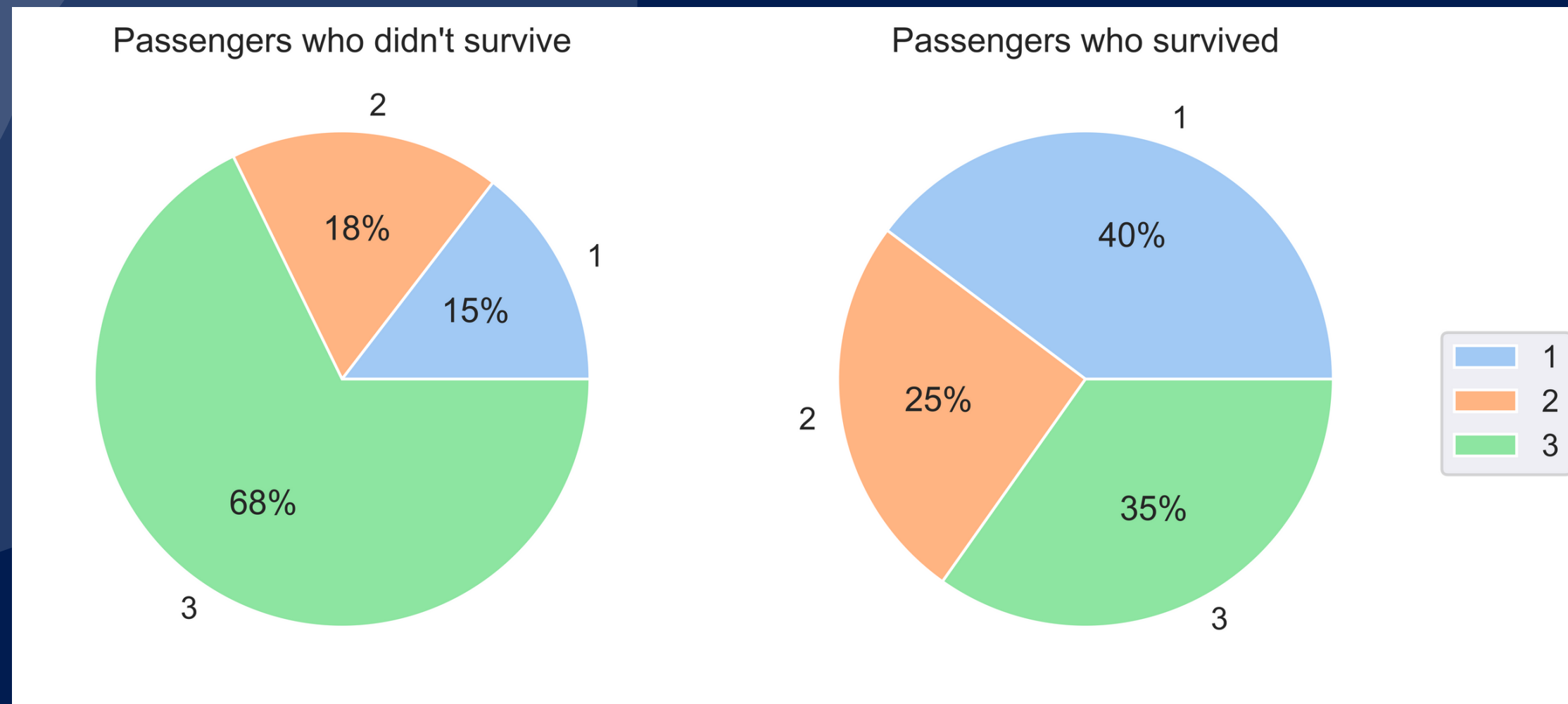


Did you know?

Only 18 % of the men on the ship survived, while more than 70% of the women survived

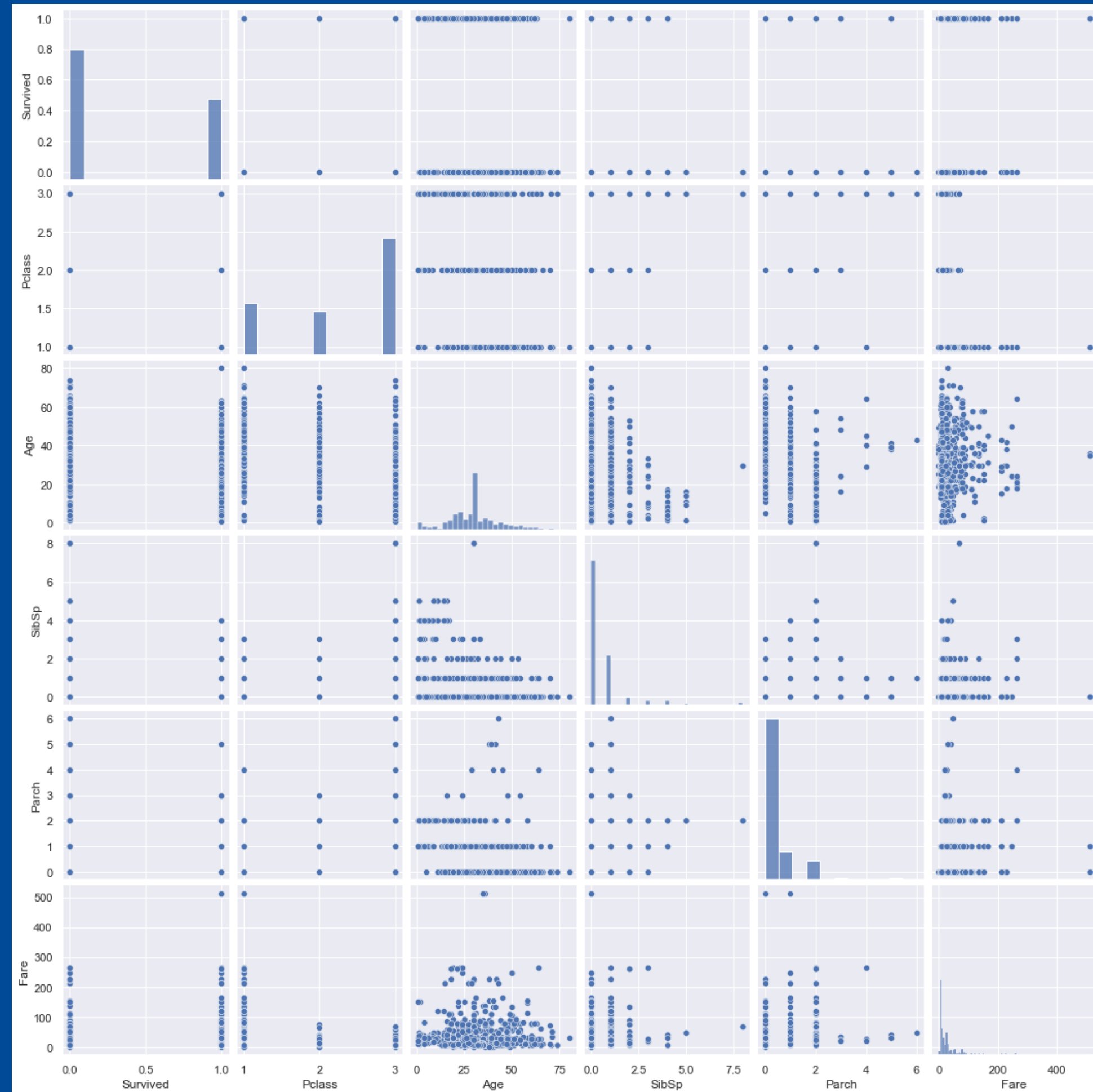
How much did the passengers spent on the ticket?

Maybe the ones who spent more, survived..
Let's see..

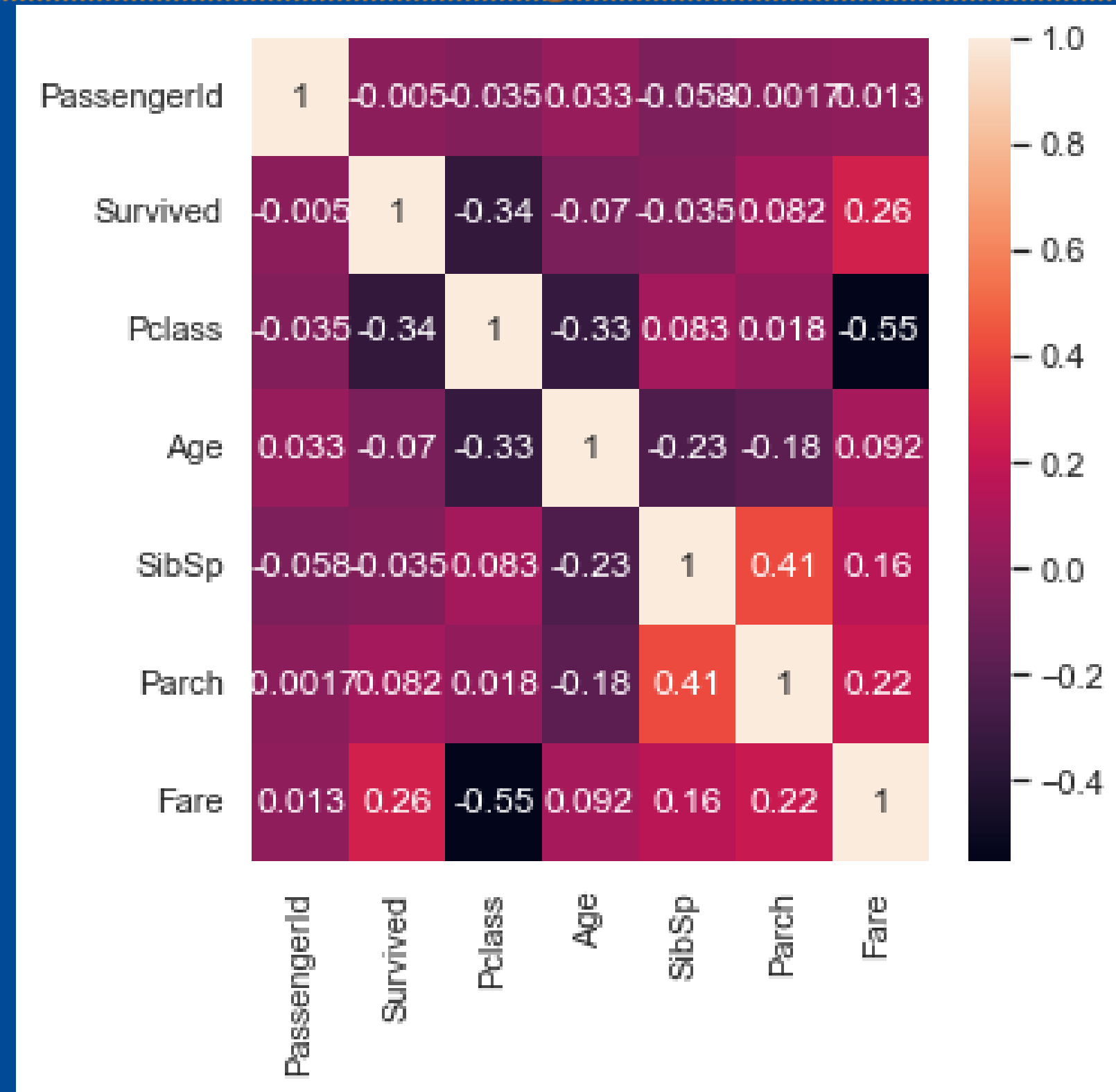


Correlation between 2 variables

Scatter plots



More correlations..



**Thank you for
your attention!**

