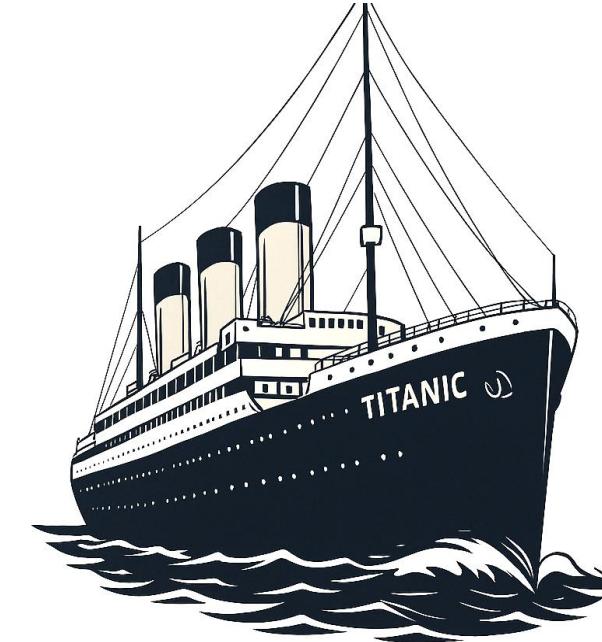


PYTHON – FINAL PROJECT

TITANIC

DATA ANALYST - TECHOF

Filipa Tiago, Hugo Varela e Joana Correia



INTRODUCTION



Titanic Passengers Data:

- Name, Age, Gender
- Family on board
- Class, Ticket fare and number
- Local embarked
- Binary survival variable
 - (0 – if not survived; 1 – if survived)

DataFrame:

- 891 records
- 12 columns

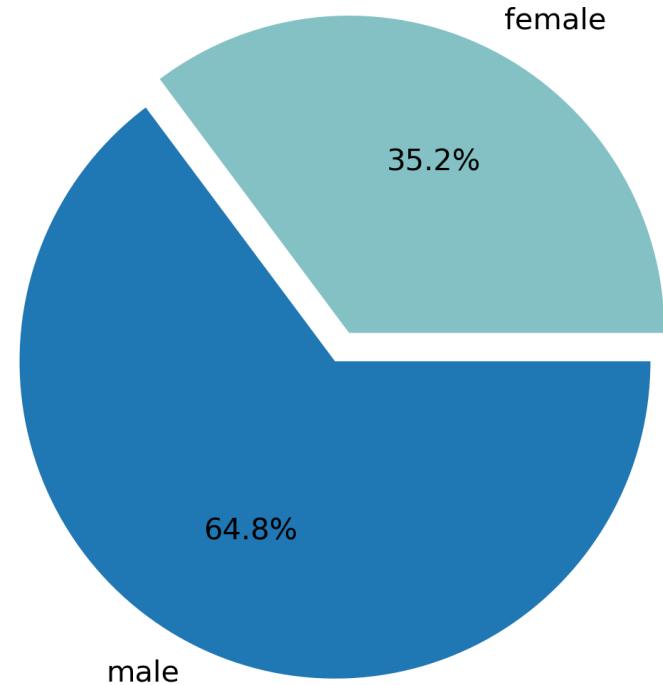
DATA TREATMENT

1. Check for null values
2. Index: *PassengerID*
3. Delete columns: *Cabin*
 - o 687 null values
4. Null values in column *Age*:
 - o Replaced with median values for men (29) and women (27)
5. Add new columns:
 - o Sex boolean
 - 0 – Female, 1 – Male
 - o Family size
 - Sum of siblings, parents, children and spouses aboard
 - o Group by family size
 - Alone, Small family and Large family
 - o Group by age
 - Passengers grouped by ages
 - o Survived non boolean
 - For chart analysis
6. Null values in *Embarked* column
 - o Changed to the most common value 'S'

OVERALL STATISTICS

DISTRIBUTION OF PASSENGERS BY GENDER

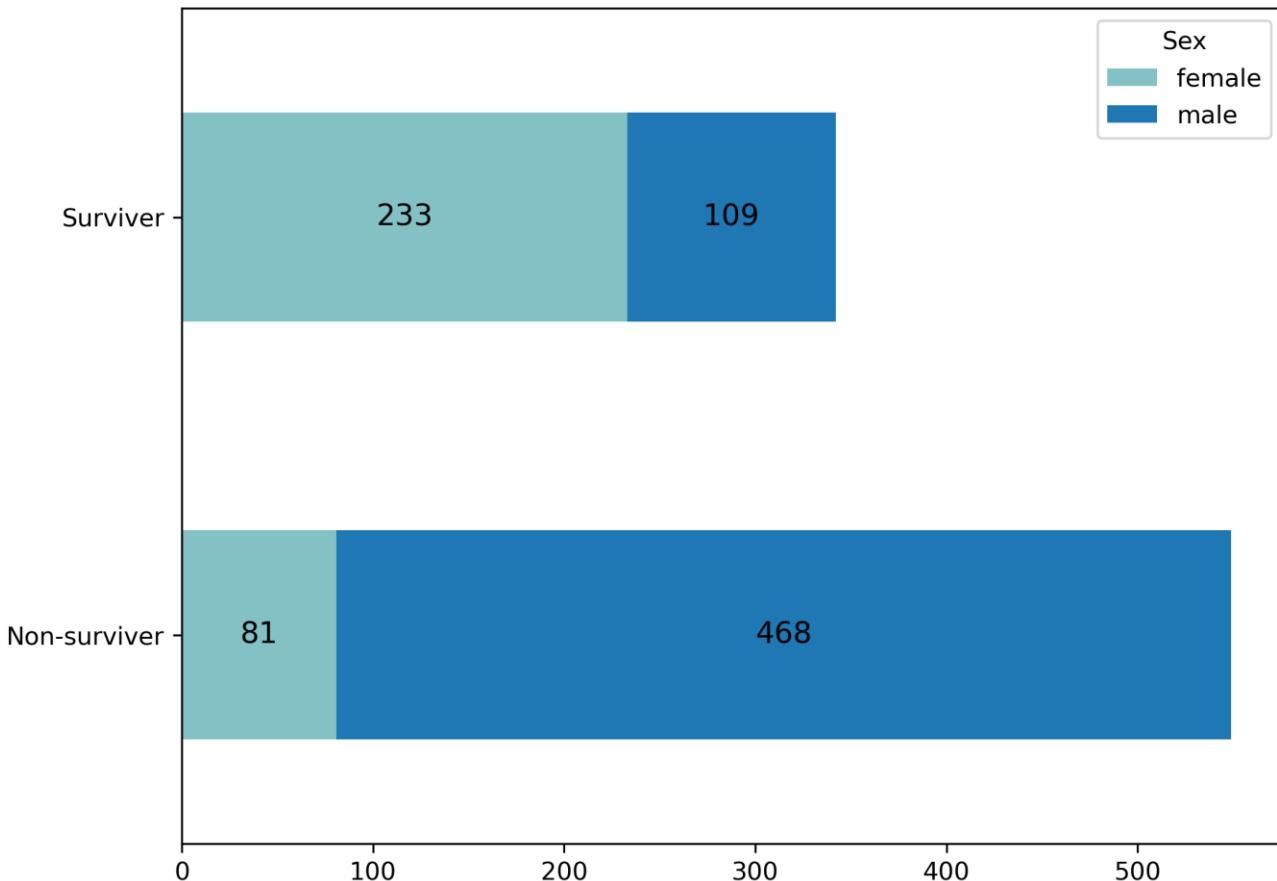
Distribution of passengers by gender



Total Passengers: 891

Men: 577
Women: 314

Survivors Distribution by Gender

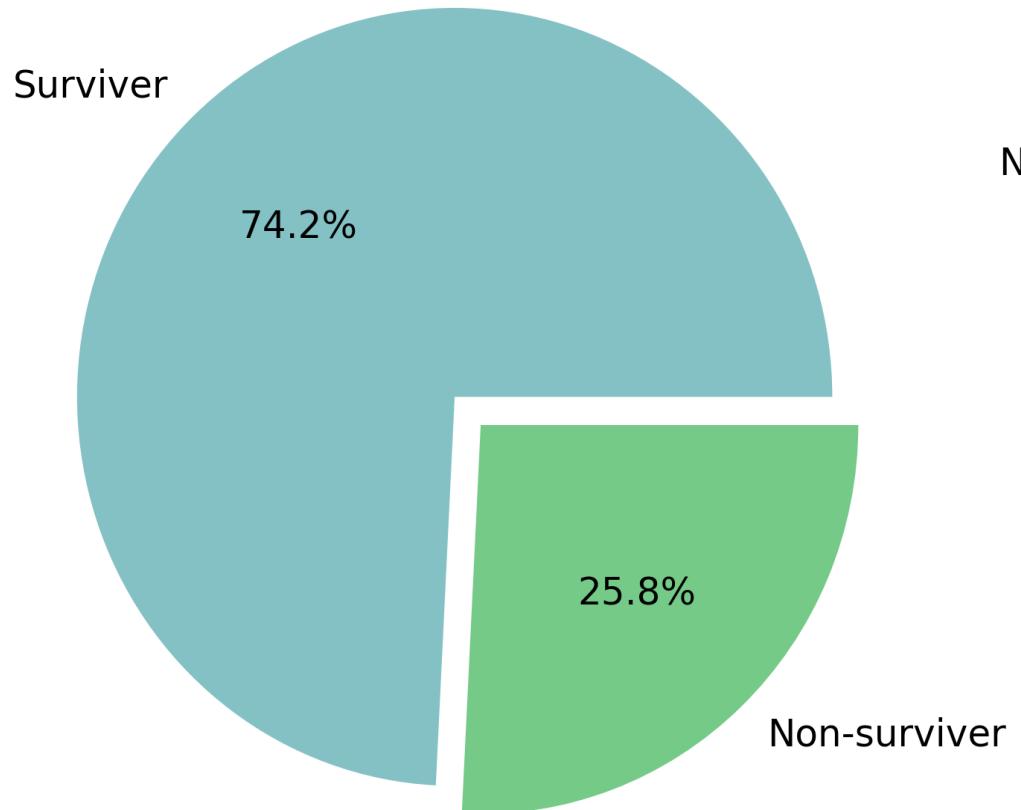


Survivors: 342

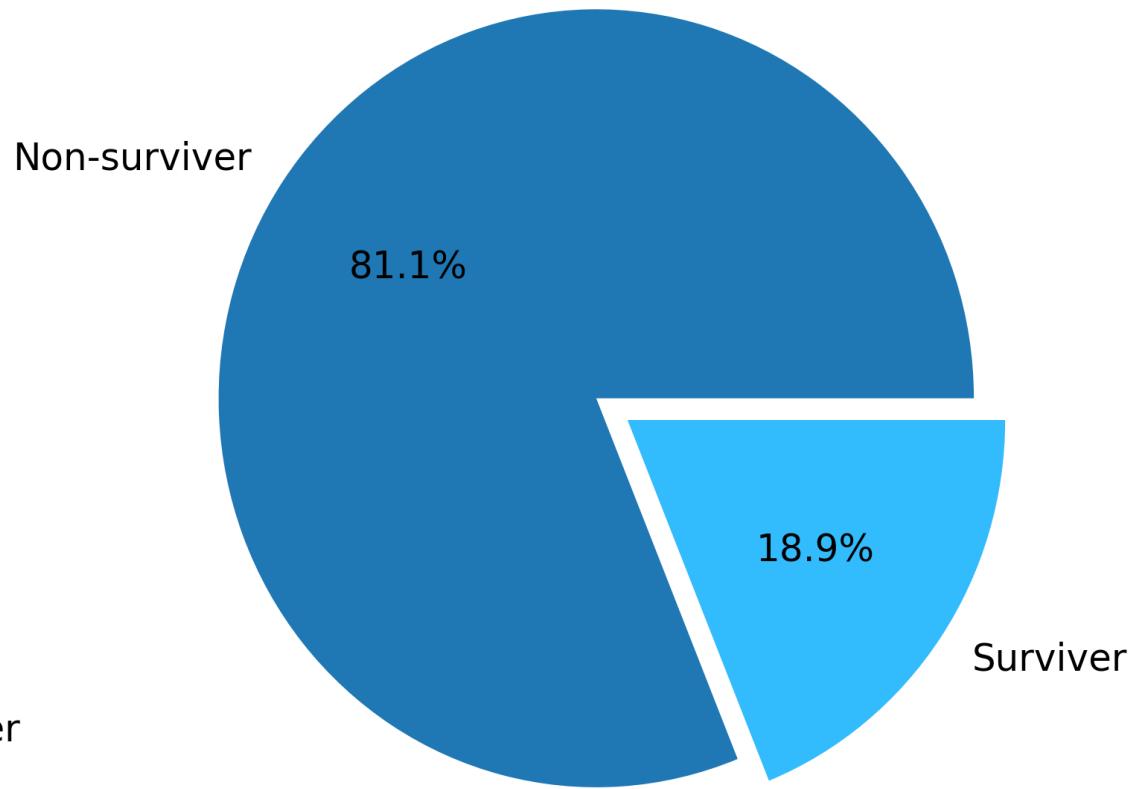
Non-survivors: 549

Survival Distribution by Gender

Female



Male

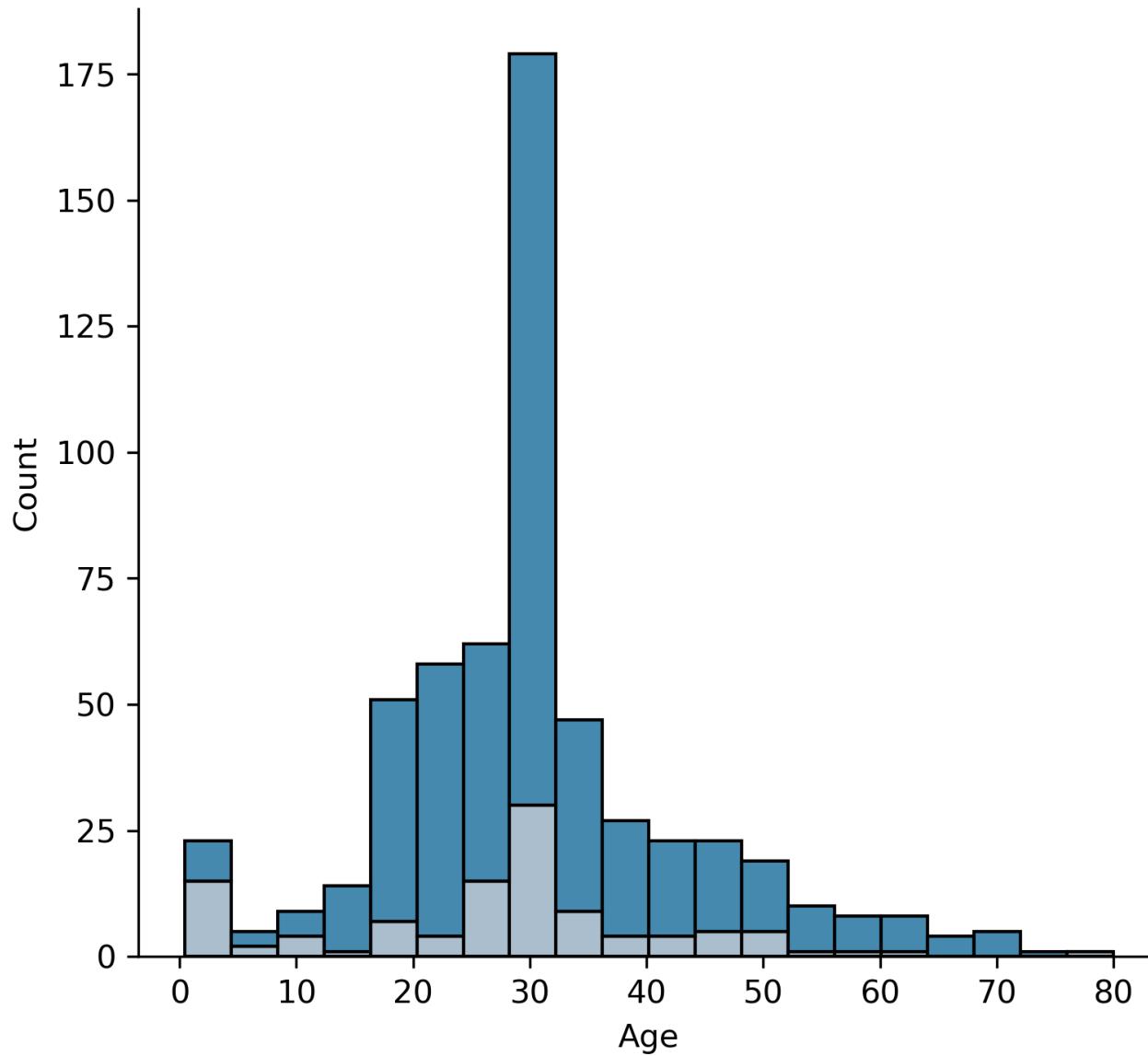


GENDER COMPARISON OF SURVIVERS AND NON-SURVIVERS

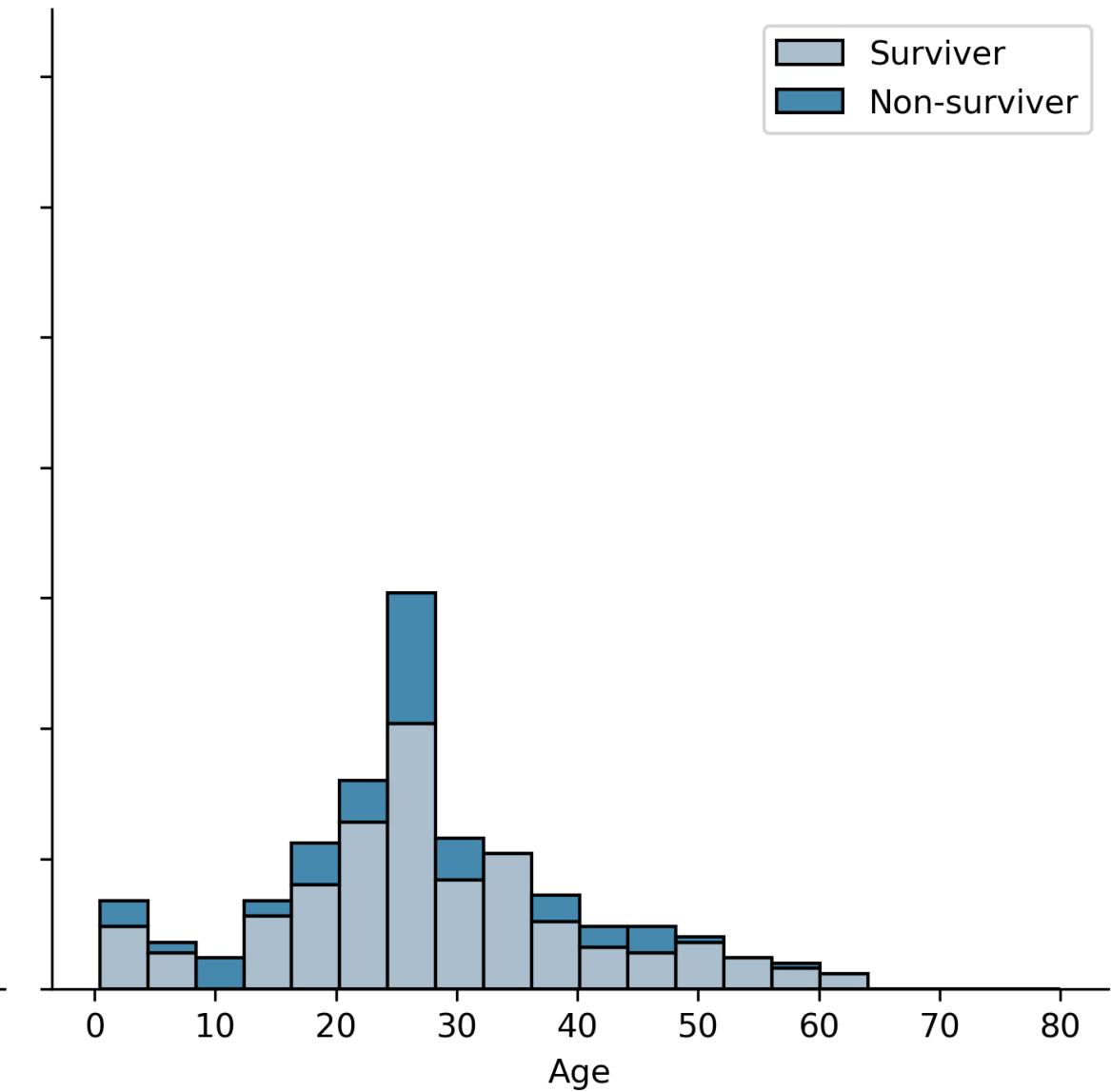
DISTRIBUTION OF SURVIVERS BY GENDER AND AGE

Average Age: 29

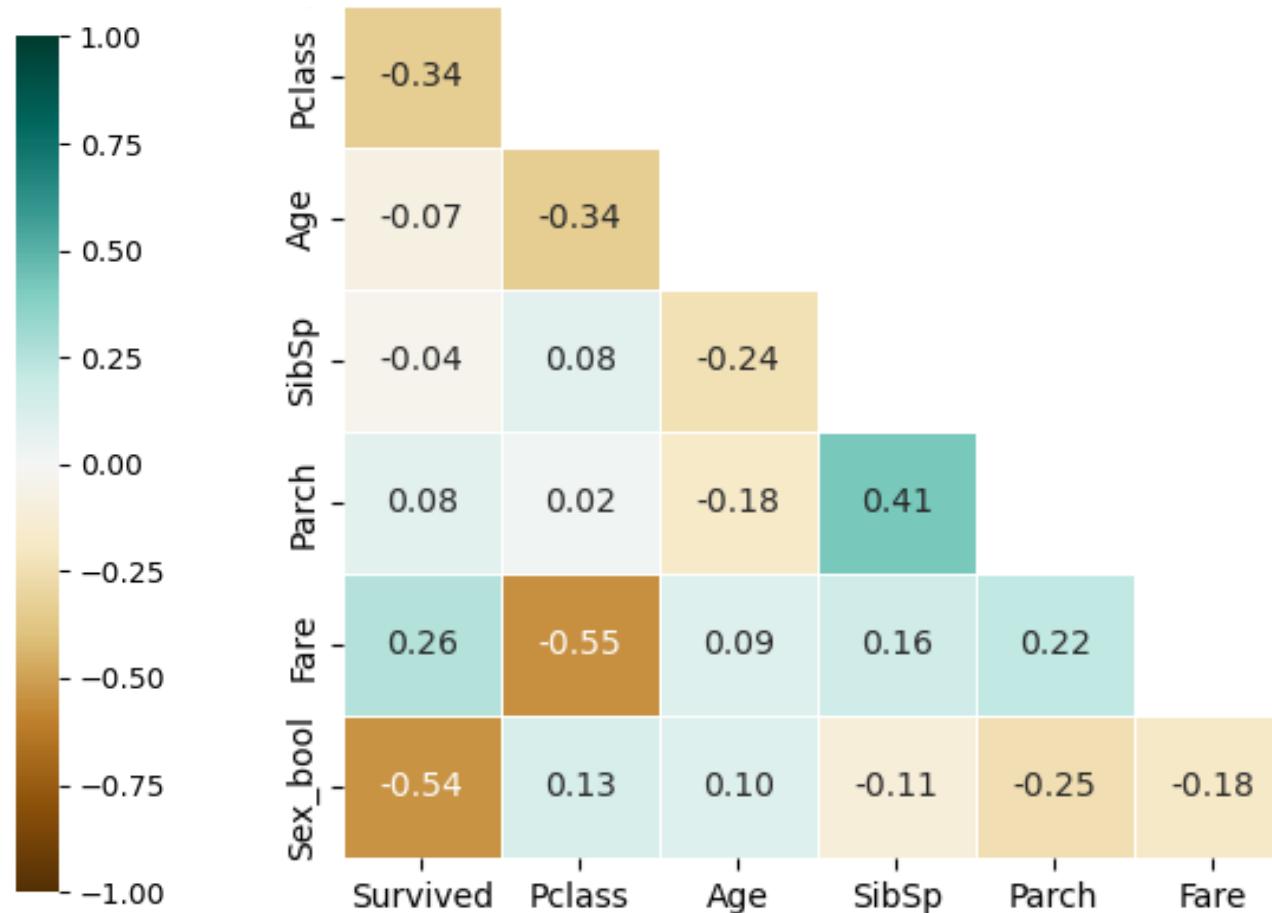
Sex = male



Sex = female



CORRELATION MATRIX - HEATMAP



Strongest Correlations:

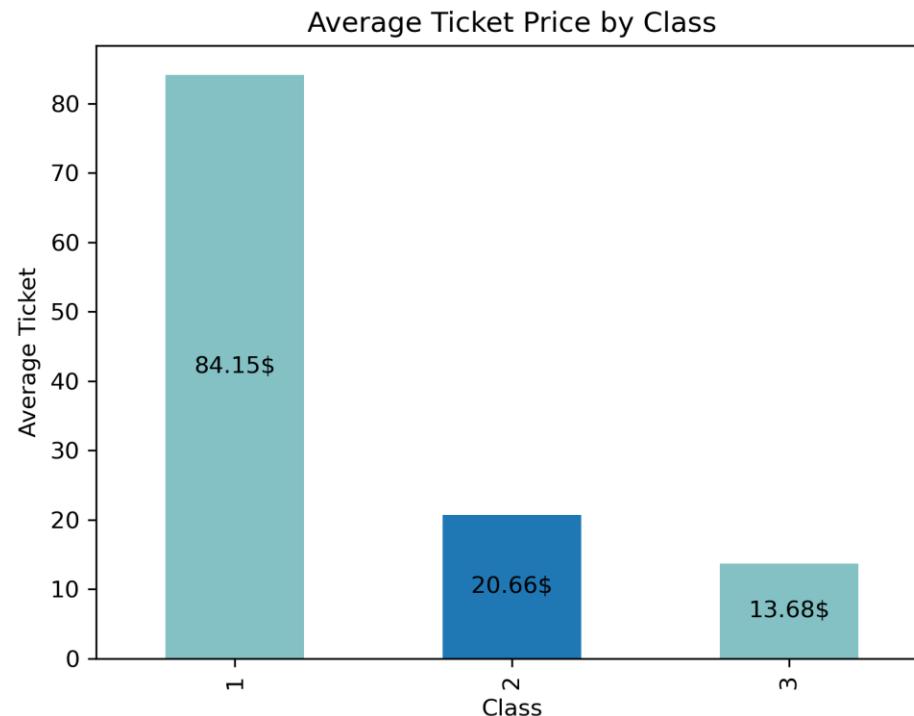
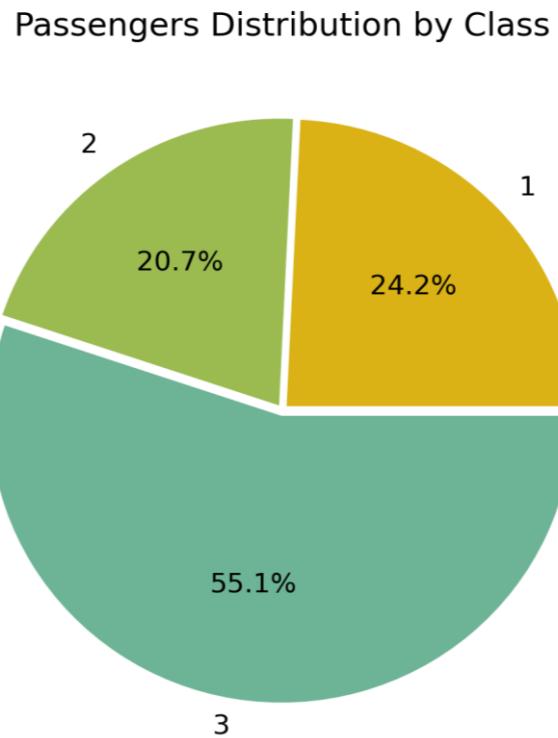
1. Ticket price vs Class
2. Gender vs Survival
3. Sibling/Spouses vs Parents/Children

Most meaningful correlations:

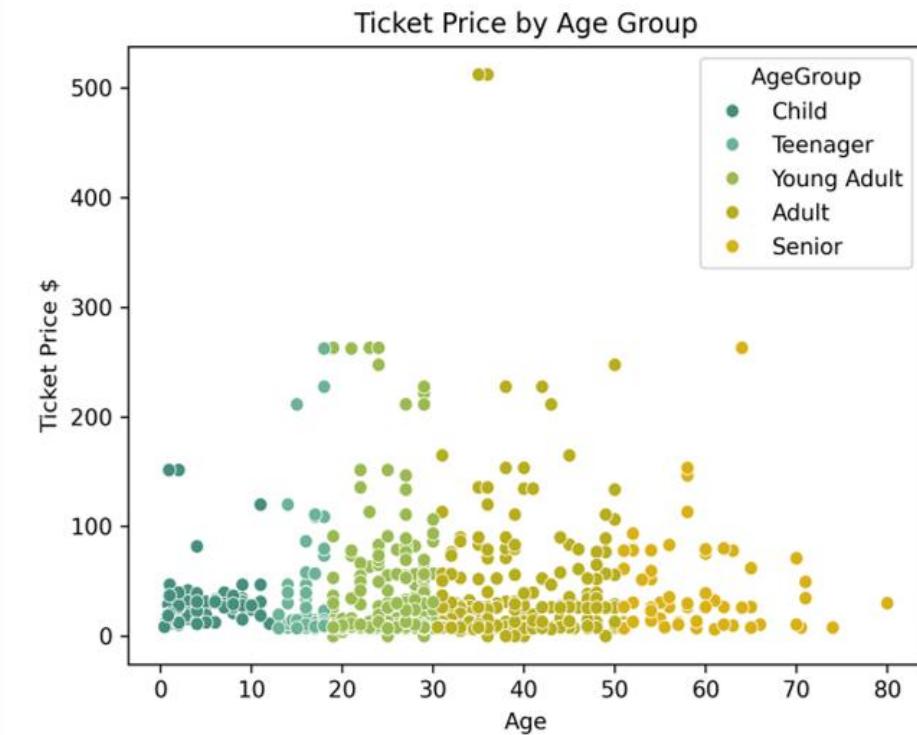
- Survival vs Gender
- Survival vs Class
- Survival vs Ticket price

STUDY QUESTIONS

HOW DOES THE SURVIVAL RATE VARY ACROSS PASSENGERS CLASS?



WHAT IS THE AVERAGE FARE ACROSS PASSENGERS CLASS AND HOW IS IT RELATED WITH AGE ?

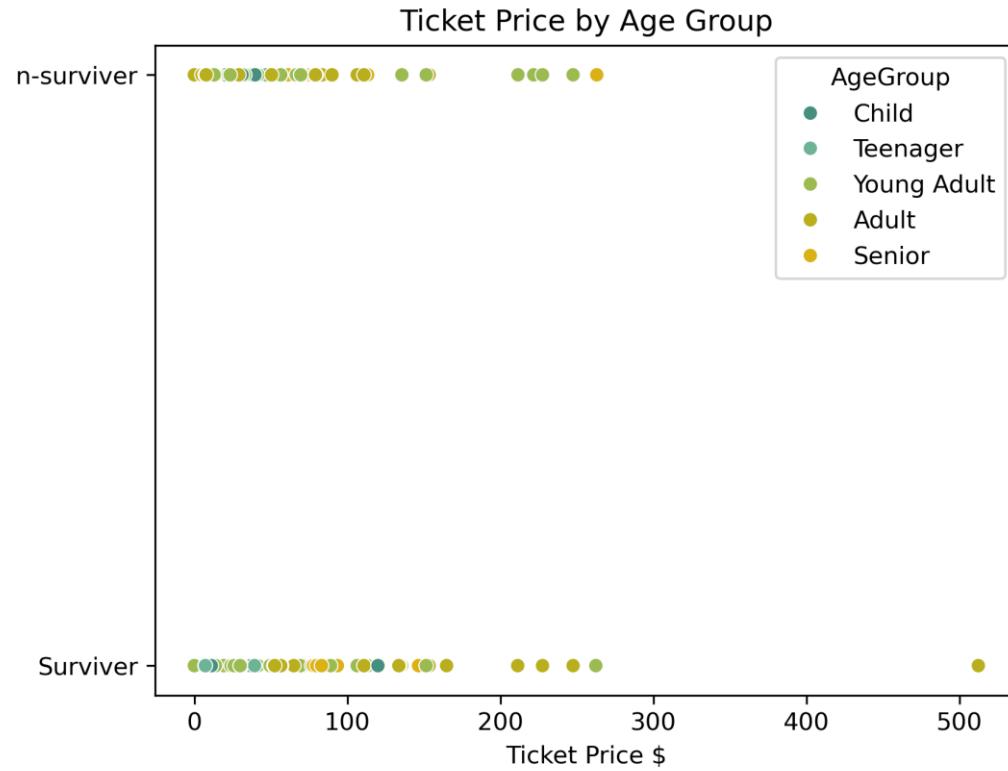
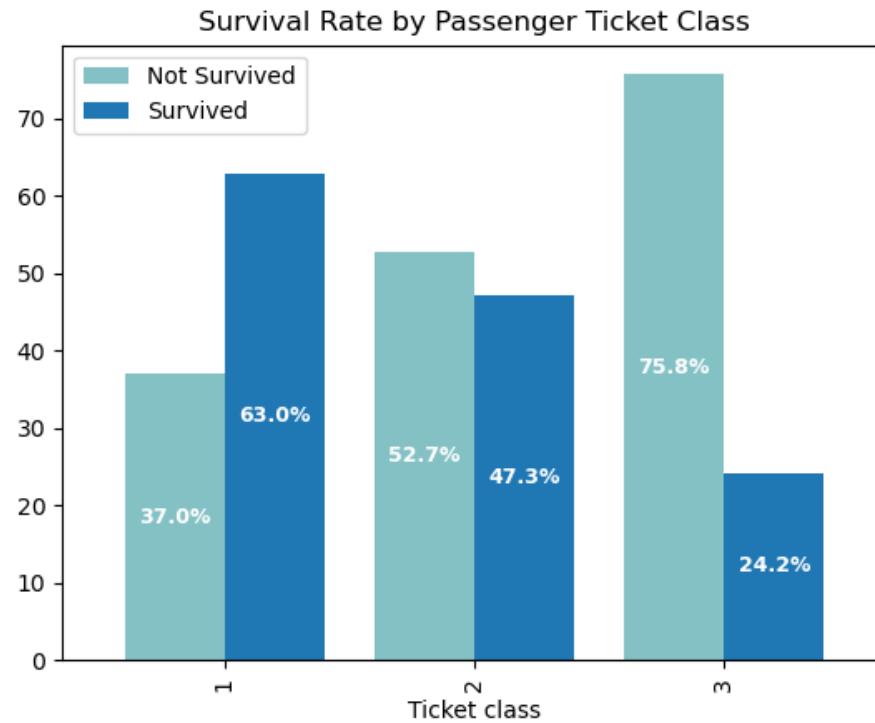


Average Ticket Price 32\$

Most expensive Ticket: 512\$

HOW DOES THE SURVIVAL RATE VARY ACROSS PASSENGERS CLASS?

IS THERE A CORRELATION BETWEEN TICKET PRICE AND SURVIVAL RATE?



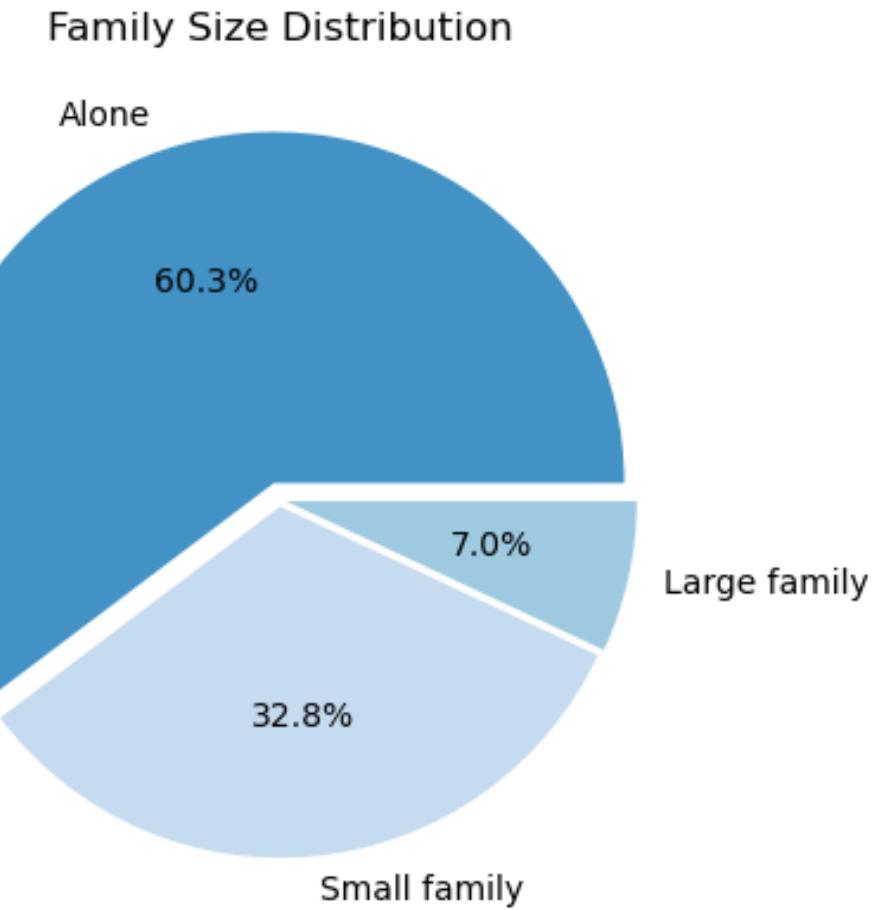
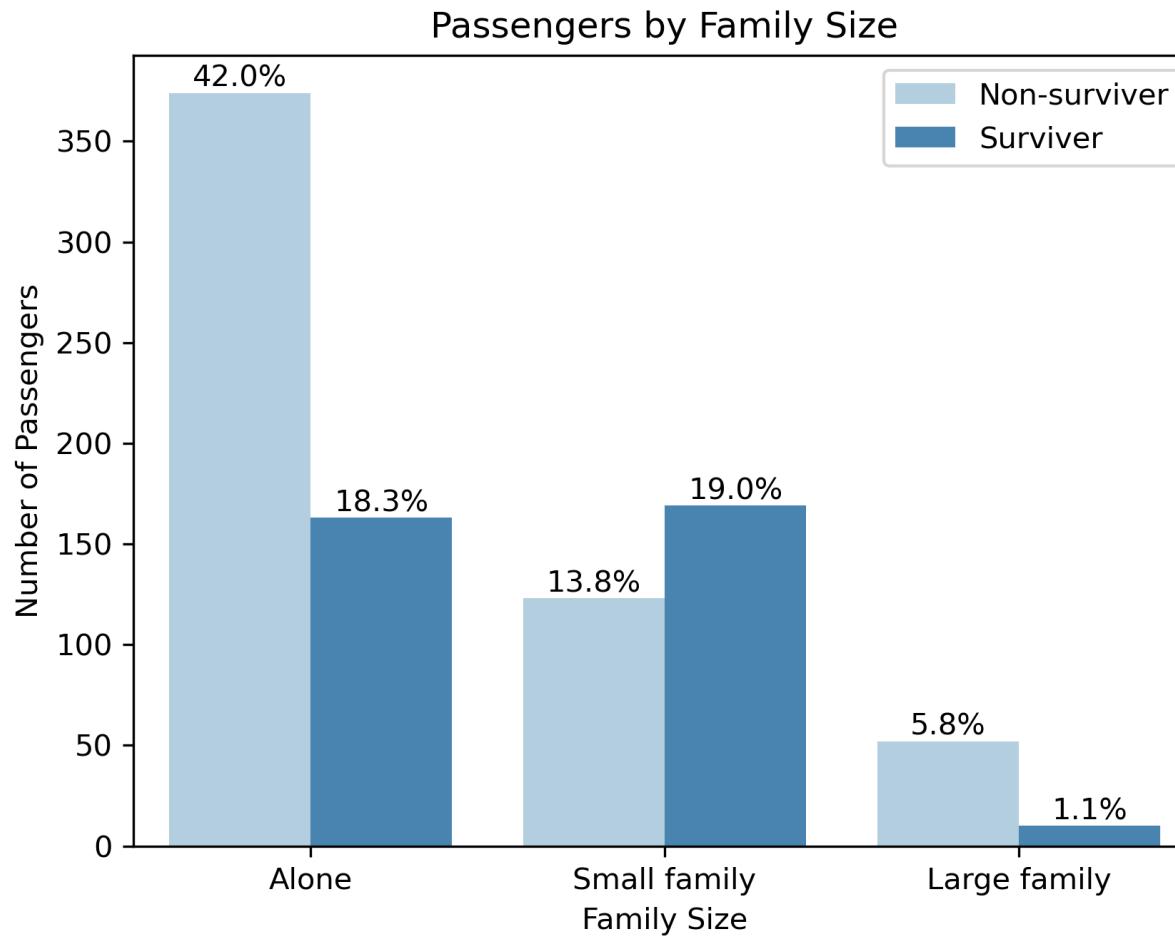
Death rate by class:

Class 1 – 37% Deaths; Class 2 – 53% Deaths; Class 3 – 76% Deaths

IS THERE A RELATION BETWEEN SURVIVAL AND FAMILY NUMBER?

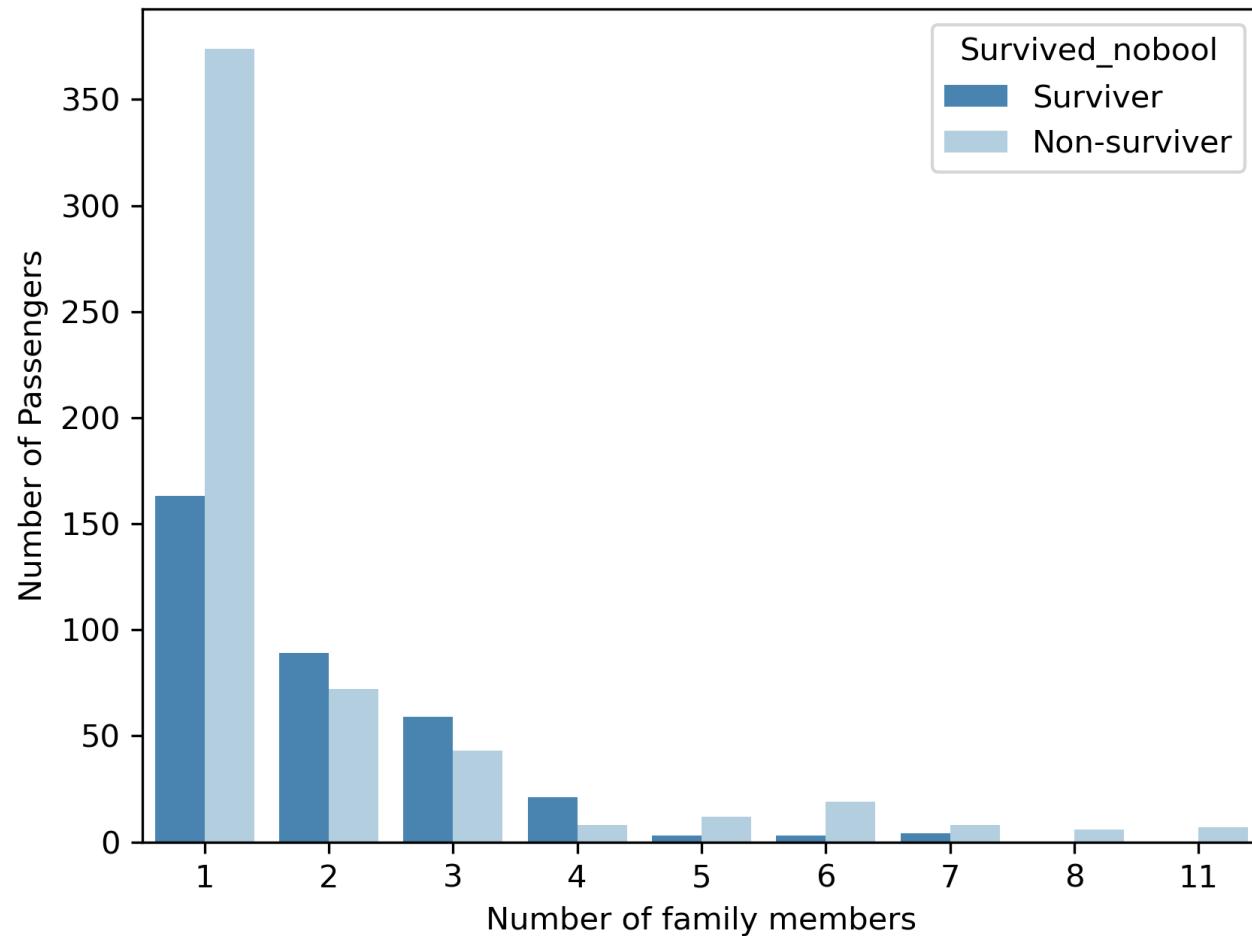
Family Size:

Alone	1
Small family	2 - 4
Large family	>5

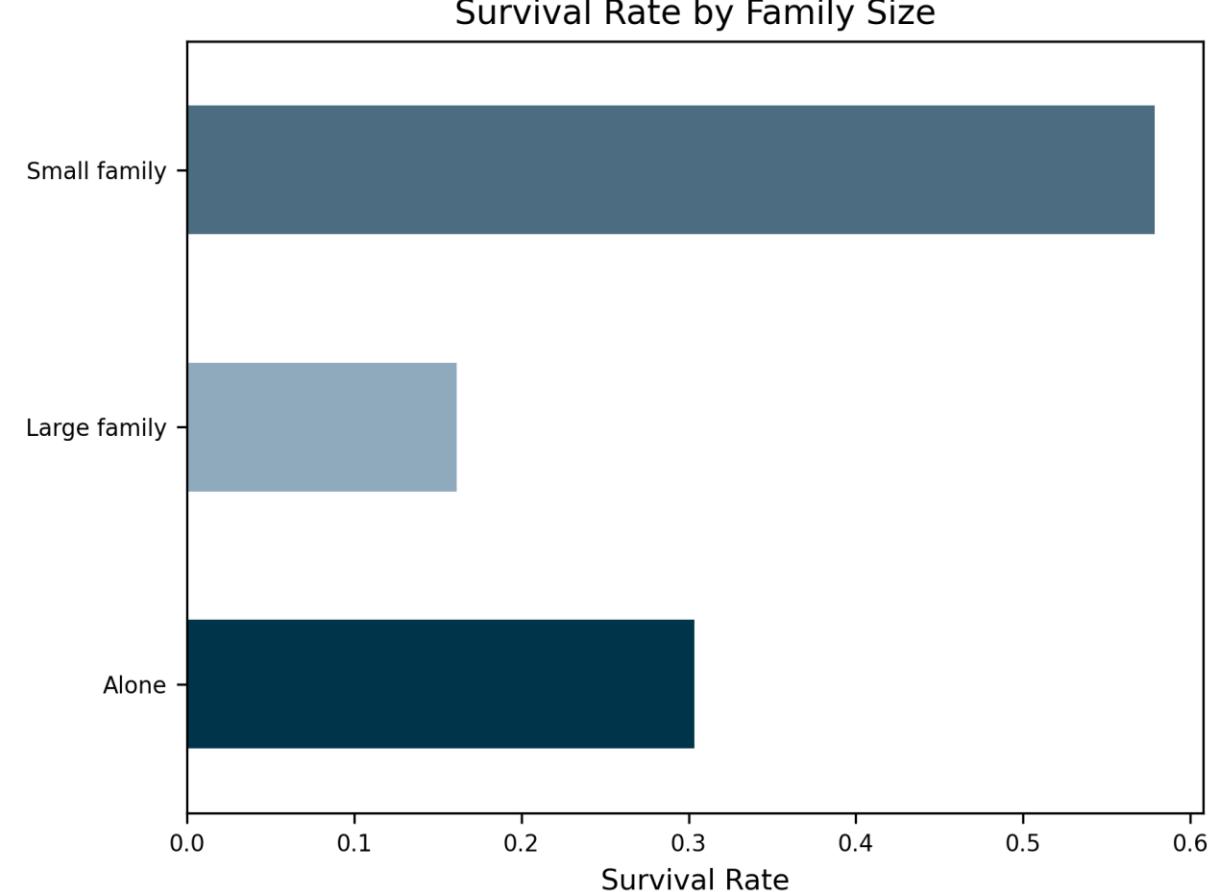


IS THERE A RELATION BETWEEN SURVIVAL AND FAMILY NUMBER?

Passengers by Number of family members



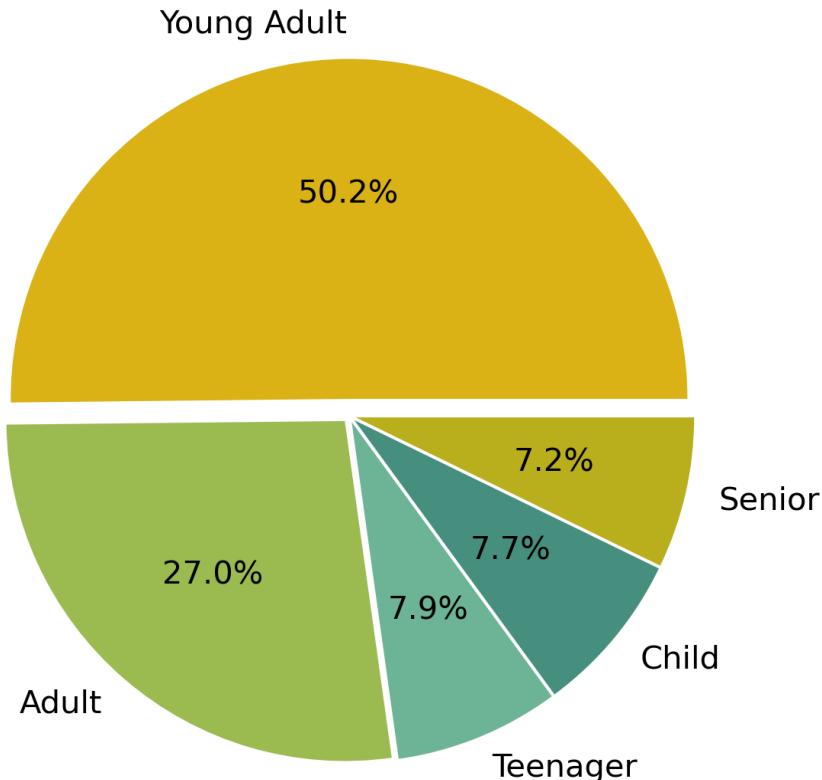
Survival Rate by Family Size



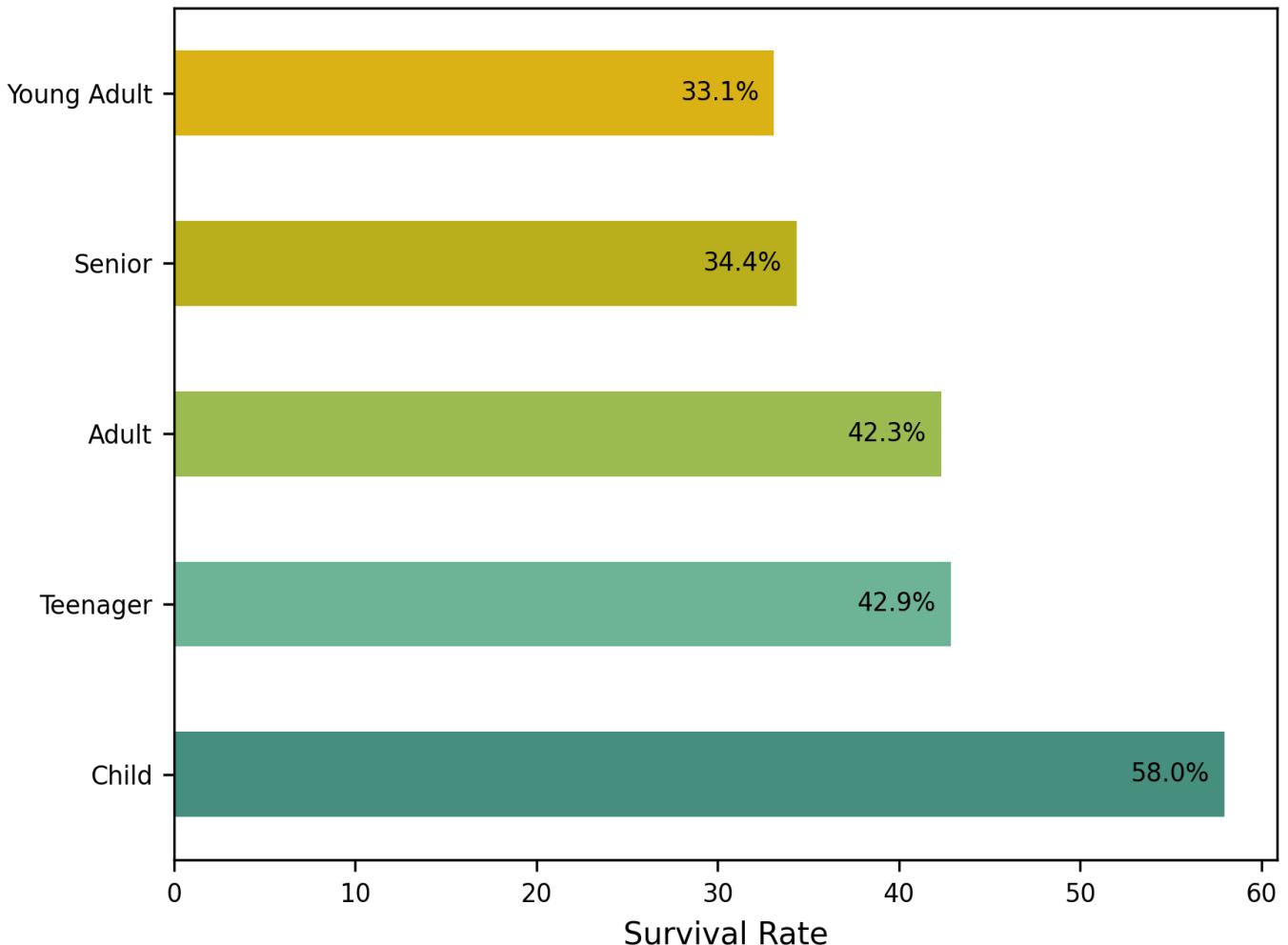
WHAT IS THE SURVIVAL RATE BY AGE GROUP?

Child	< 12
Teenager	13 - 18
Young Adult	19 - 30
Adult	31 - 50
Senior	> 51

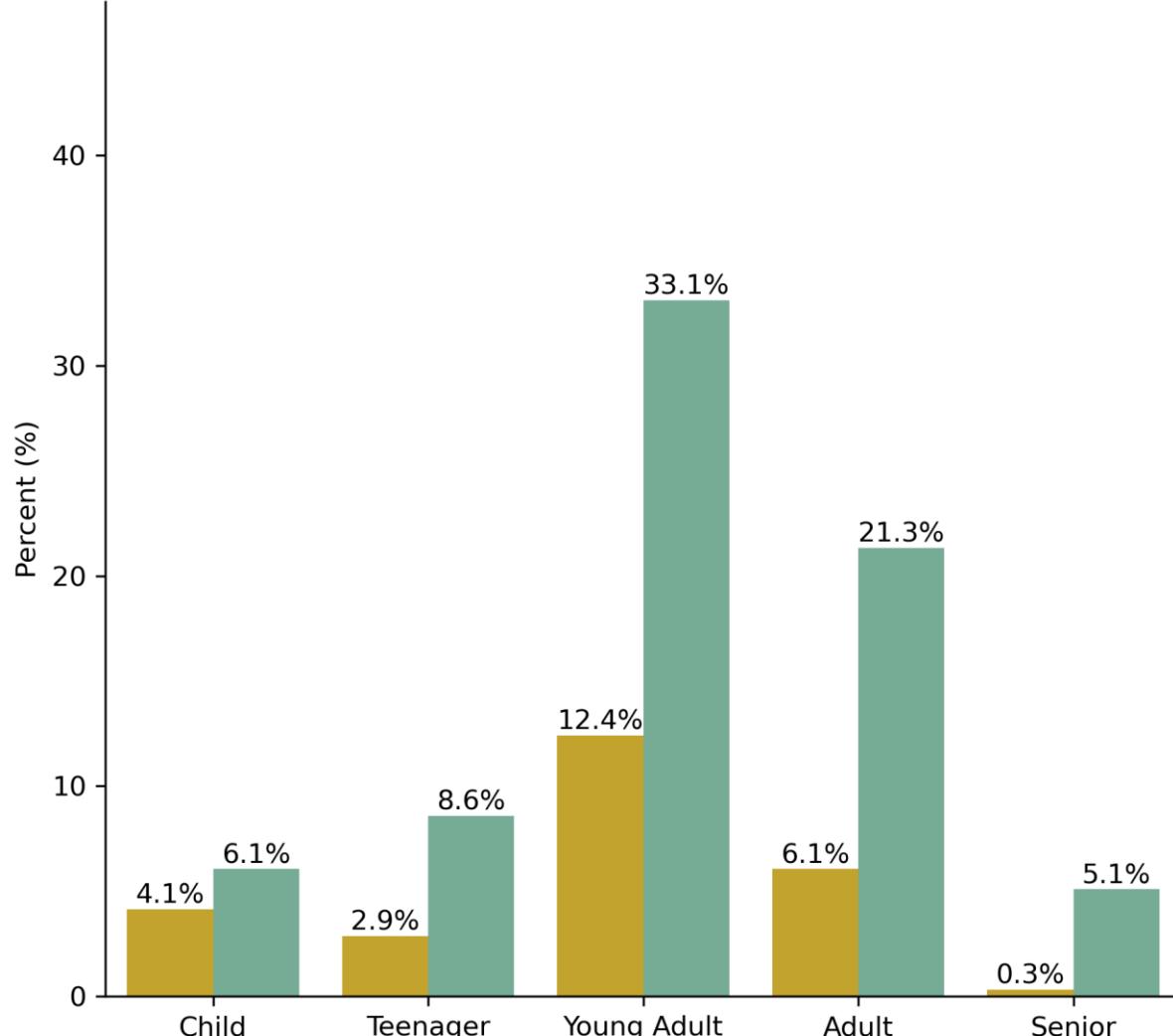
Passengers Distribution by Age Group



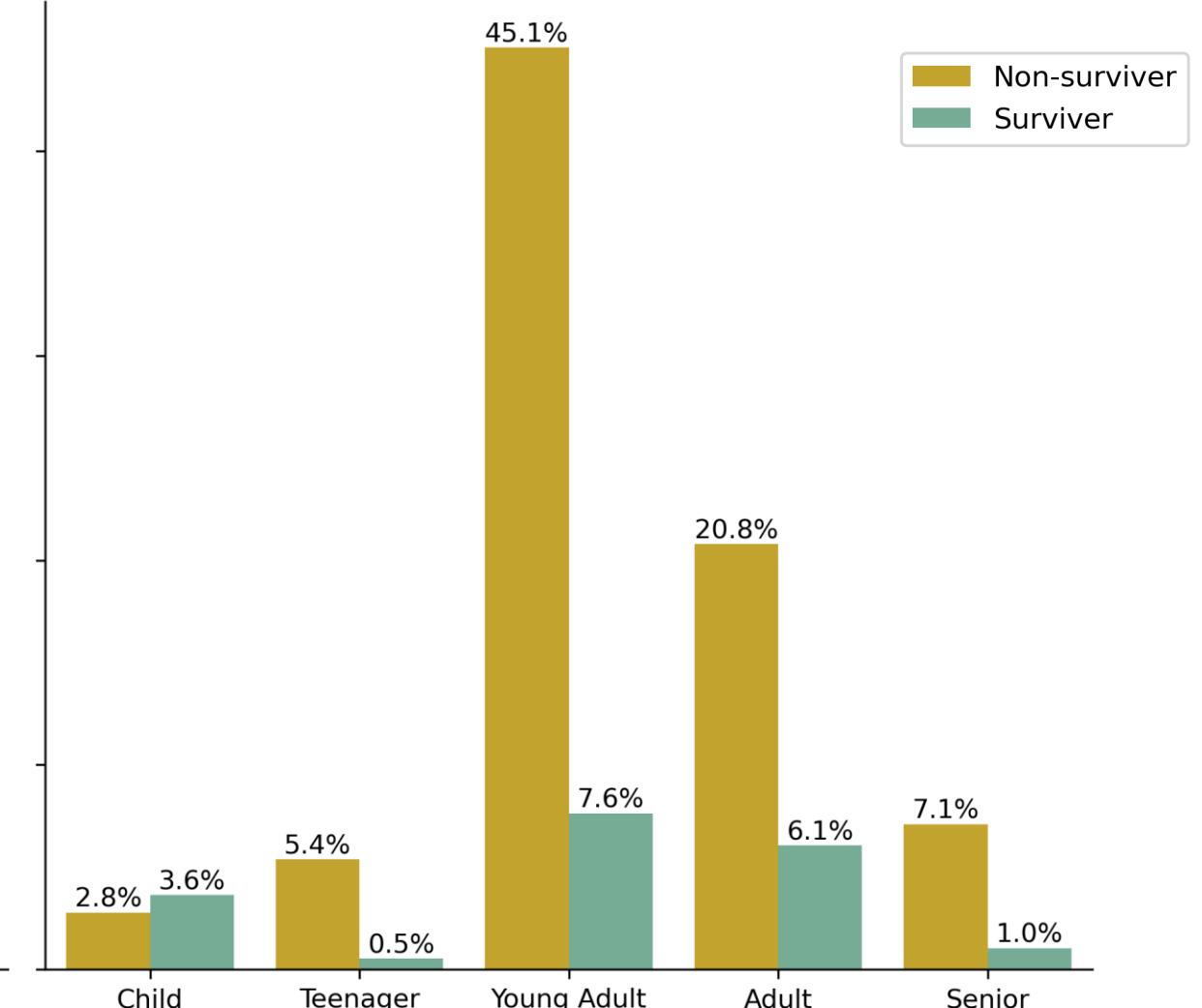
Survival Rate by Age Group



Sex = female



Sex = male



WHAT IS THE SURVIVAL RATE BY AGE GROUP?

Age Groups:

Child	< 12
Teenager	13 - 18
Young Adult	19 - 30
Adult	31 - 50
Senior	> 51

CONCLUSIONS

- Of a total 891 passengers, 62% died
- Survival rate:
 - 1/5 men survived
 - 3/4 women survived
- "*Women and children first*" principle:
 - 58% children saved
 - 56% adult women saved
- Socioeconomics status had impact:
 - High mortality rate in Class 3 (76%)
- Family structure:
 - Smaller families had higher survival rate

TITANIC SURVIVAL KIT



Be a child...
or a woman!



Travel with family,
preferably a small one...



Buy a Class 1 Ticket,
for better position view.

Survival rate not guaranteed



QUESTIONS

