



Python Project

04.07.2023

—

Roman Alvarez

STAT 3010

Titanic Dataset

#1 How many people were on the Titanic?

2208

RangeIndex: 2208 entries, 0 to 2207

Data columns (total 10 columns):

#2 What variables were collected on the passengers?

Quantitative Variables: Age, How Much They Paid

Categorical Variables: Name, Whether they survived, Where they boarded, Man Woman or Child, Adult or Child, Sex, Type of passenger, and the class of the passenger

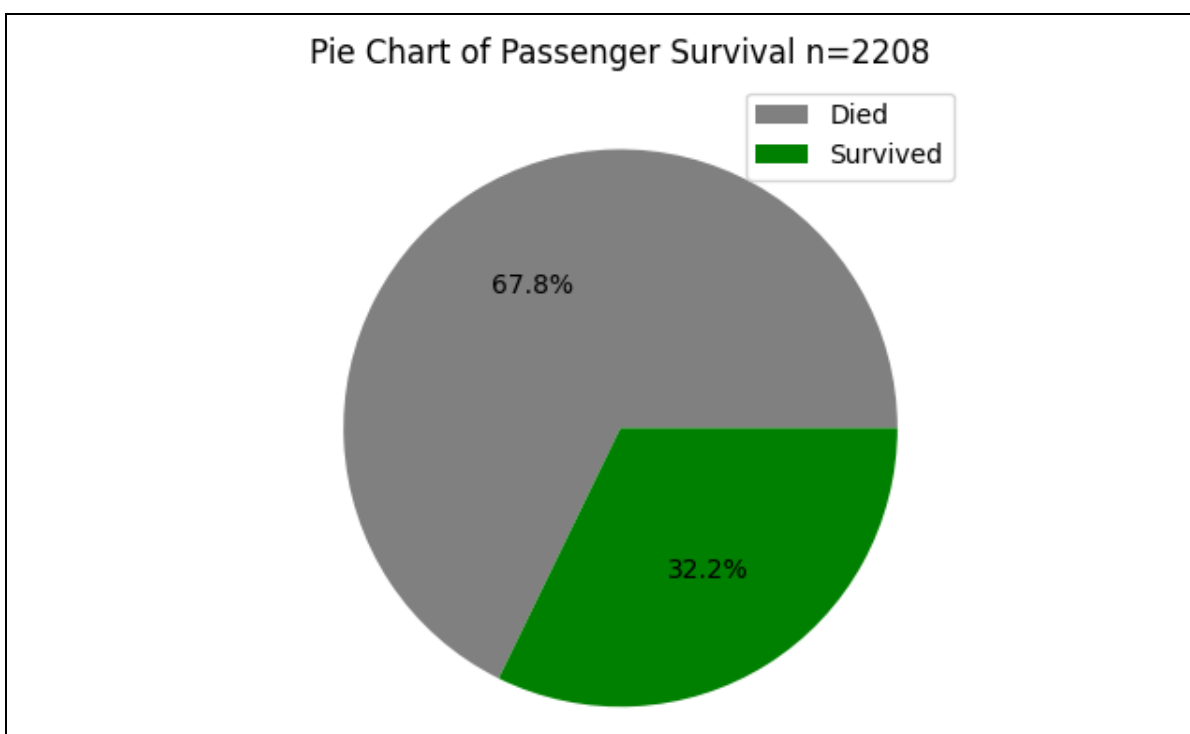
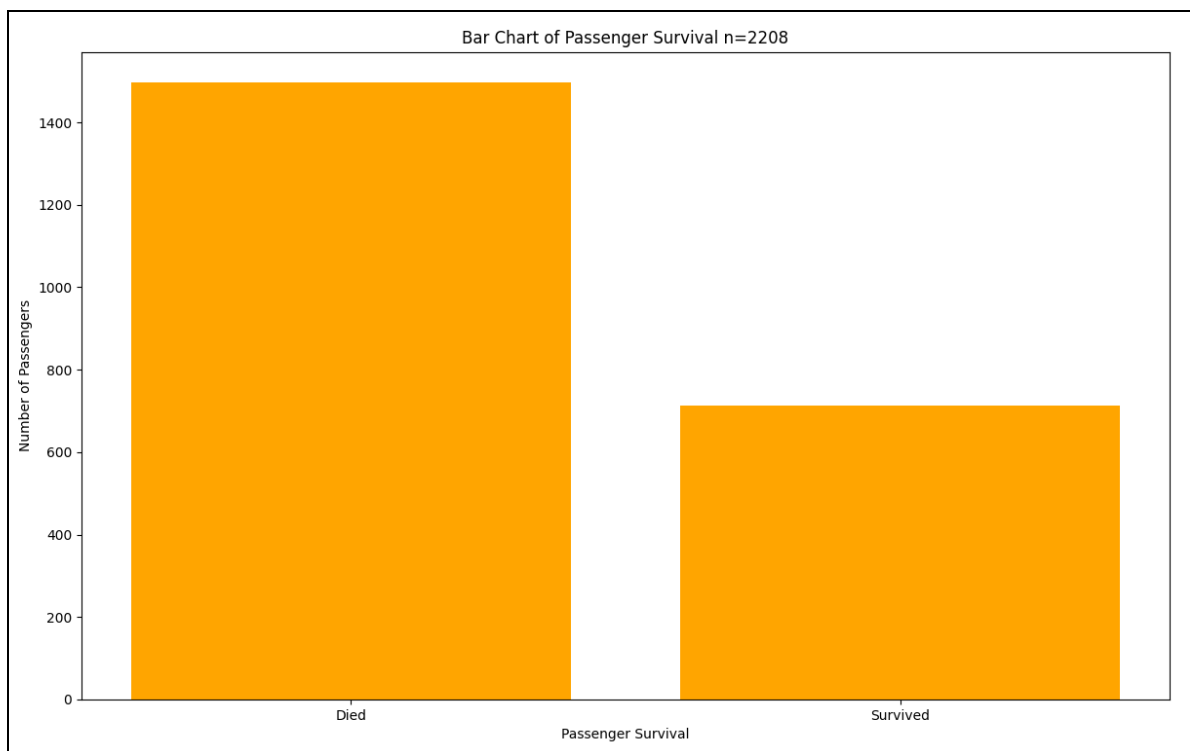
Data columns (total 10 columns):

#	Column	Non-Null Count	Dtype
0	Name	2208 non-null	object
1	Survived	2208 non-null	object
2	Boarded	2208 non-null	object
3	MWC	2208 non-null	object
4	Age	2205 non-null	float64
5	Adut_or_Chld	2208 non-null	object
6	Sex	2208 non-null	object
7	Paid	1318 non-null	float64
8	Class_Dept	2123 non-null	object
9	Class	2208 non-null	object

#3 Frequencies and proportions of passengers that survived and died

col_0	count
Survived	
Died	1496
Survived	712

col_0	count
Survived	
Died	67.75%
Survived	32.24%



#4 Frequencies and proportions of passengers' class

col_0 count

Class

1 **324**

2 **285**

3 **710**

D **66**

E **323**

R **69**

V **431**

col_0 count

Class

1 14.67%

2 12.91%

3 32.16%

D 2.99%

E 14.63%

R 3.13%

V 19.52%

1= 1st Class

2= 2nd Class

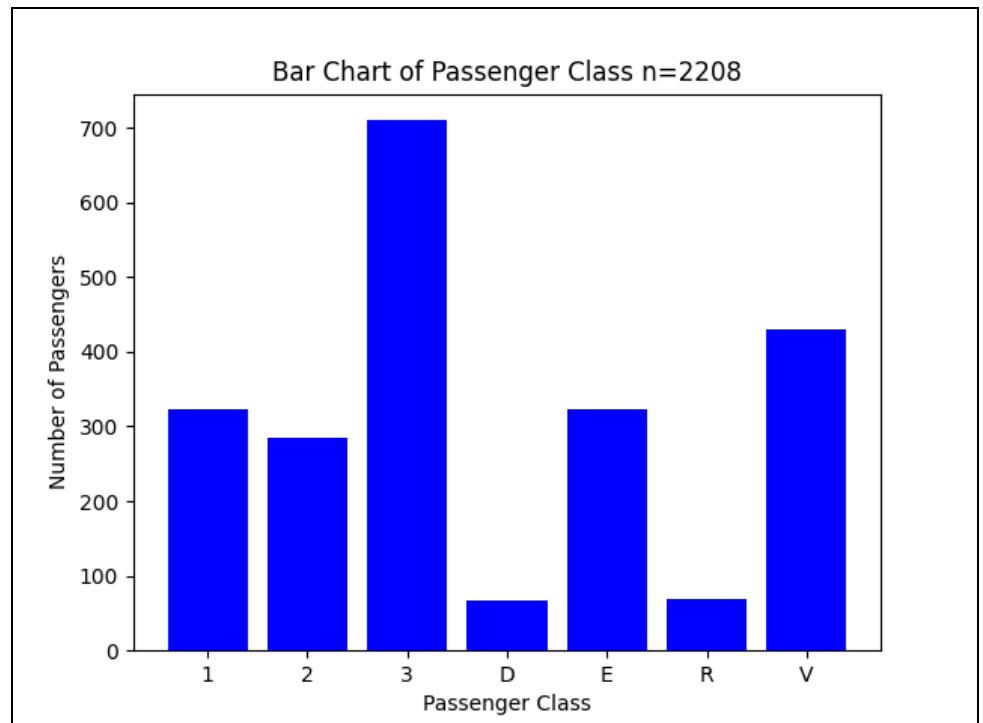
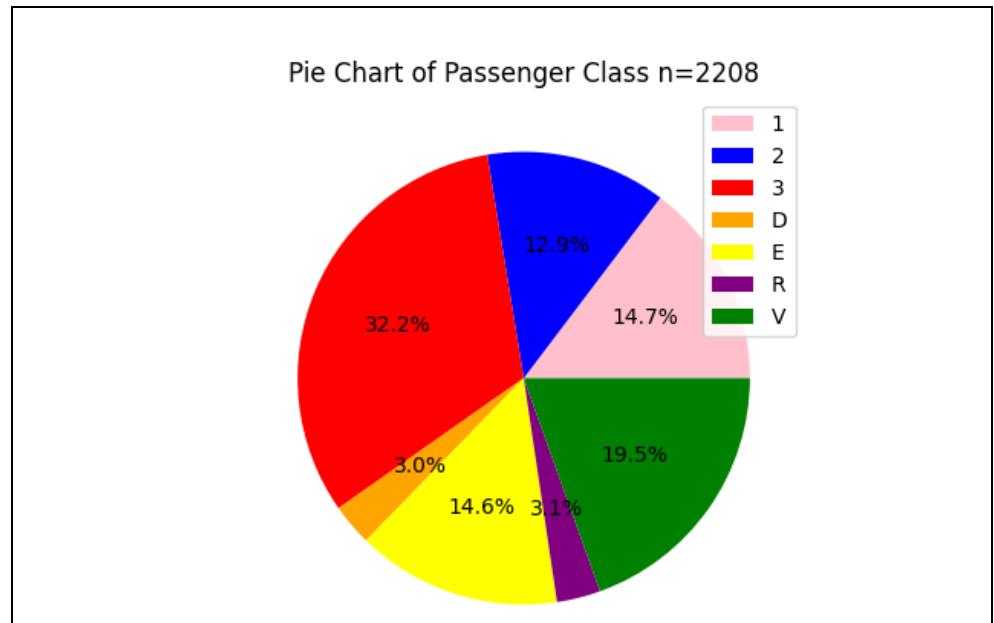
3= 3rd Class

D= Deck Staff

E= Engineering Staff

R= Restaurant Staff

V= Victualling Crew

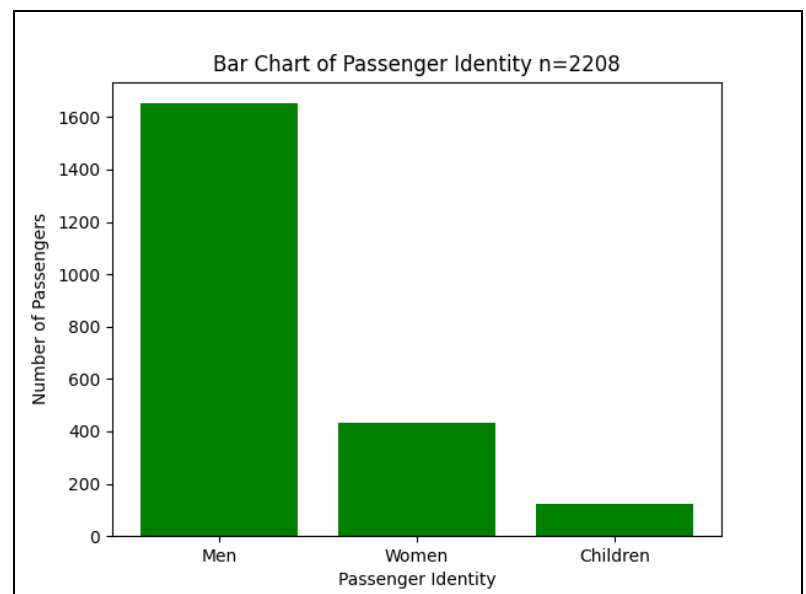
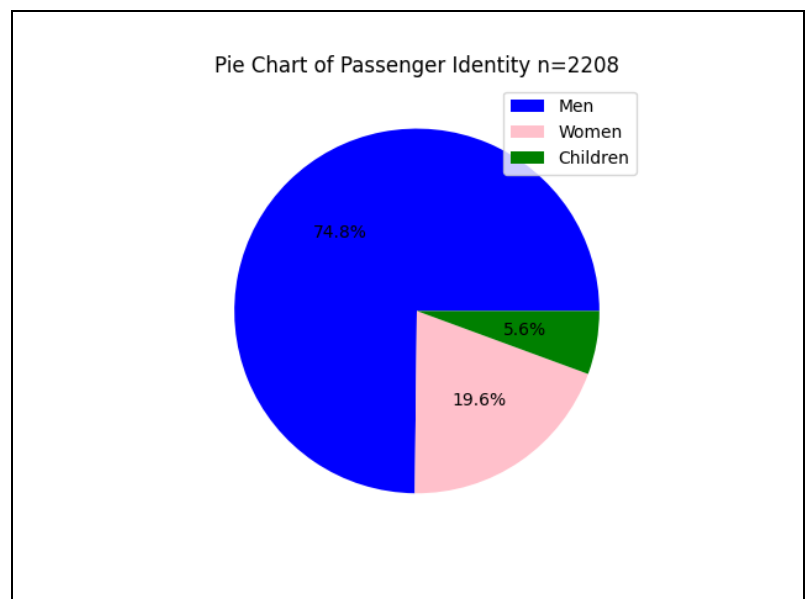


Class Code Snippet:

col_0	count
Class_Dept	
1st Class Passenger	274
1st Class PassengerH&W Guarantee Group	3
1st Class PassengerServant	36
2nd Class Passenger	252
2nd Class PassengerH&W Guarantee Group	6

#5 Frequencies and proportions of men, women, and children

col_0	count
MWC	
Child	124
Man	1652
Woman	432



#6 Frequencies of class departments

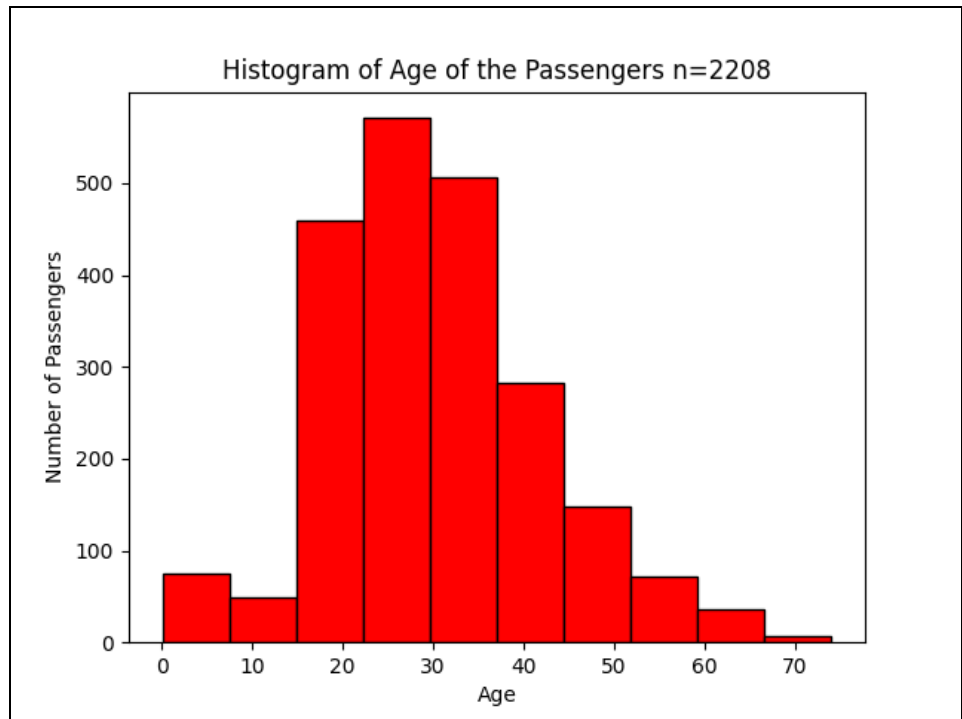
col_0	count
Class_Dept	
1st Class Passenger	274
1st Class PassengerH&W Guarantee Group	3
1st Class PassengerServant	36
2nd Class Passenger	252
2nd Class PassengerH&W Guarantee Group	6
2nd Class PassengerMusician	8
2nd Class PassengerServant	4
3rd Class Passenger	656
Deck Crew	58
Deck CrewTitanic Officers	8
Engineering Crew	320
Restaurant Staff	69
Victualling Crew	424
Victualling CrewPostal Clerk	5

- a. **The Harland and Wolff Guarantee Group was a group of nine individuals who were selected to represent the shipbuilding firm with passage on the Titanic's Maiden Voyage**
- b. **There were a total of eight musicians on the Titanic; their class department is noted as "2nd Class Passenger Musician"**

#7 How is the age of the passengers distributed

The histogram represents a symmetrical unimodal distribution with most passengers ranging in age from 15-45 years old.

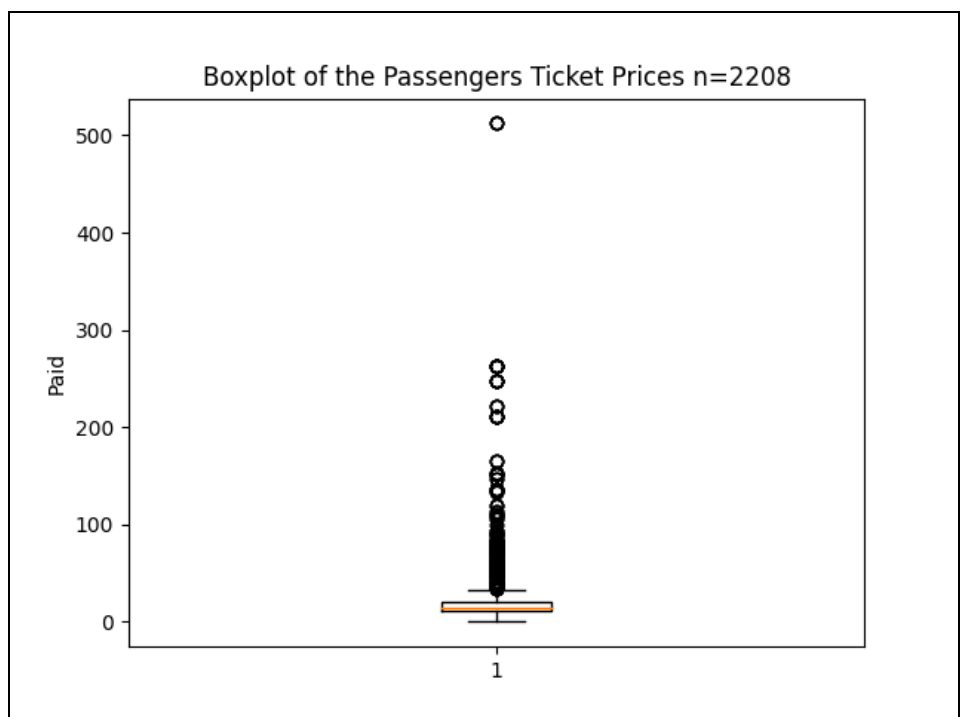
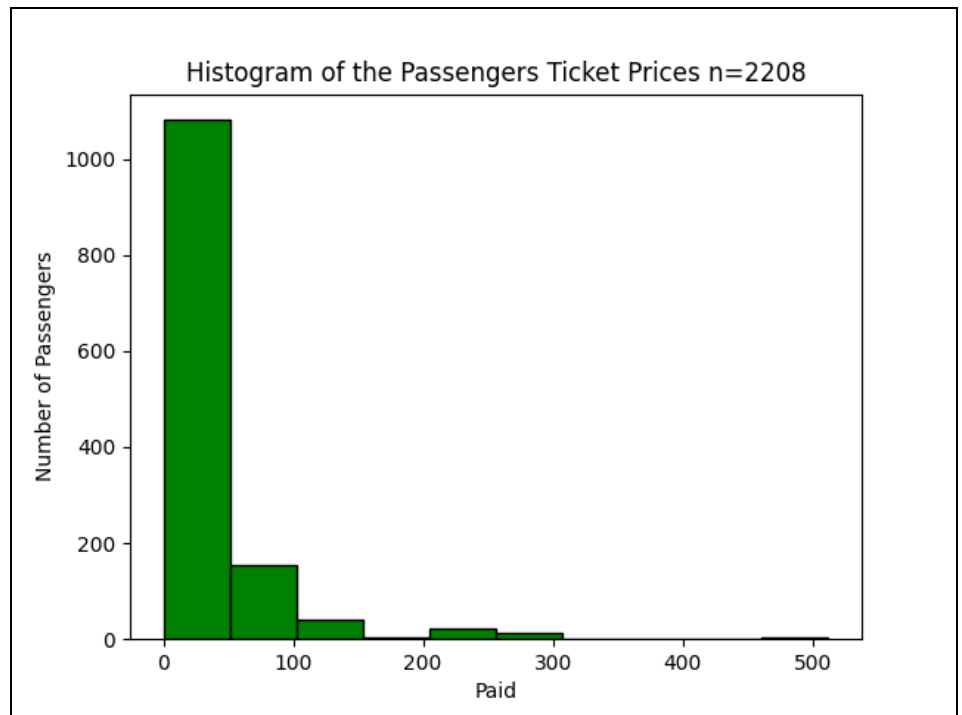
count	2205
mean	30.15
std	11.97
min	0.08
25%	22
50%	29
75%	37
max	74



#8 How are ticket prices distributed

Majority of the tickets were priced under \$50, with some outliers being priced above \$100 and up to \$500

count	1318
mean	33.01
std	51.91
min	0.00
25%	7.89
50%	14.40
75%	31.
max	512.33



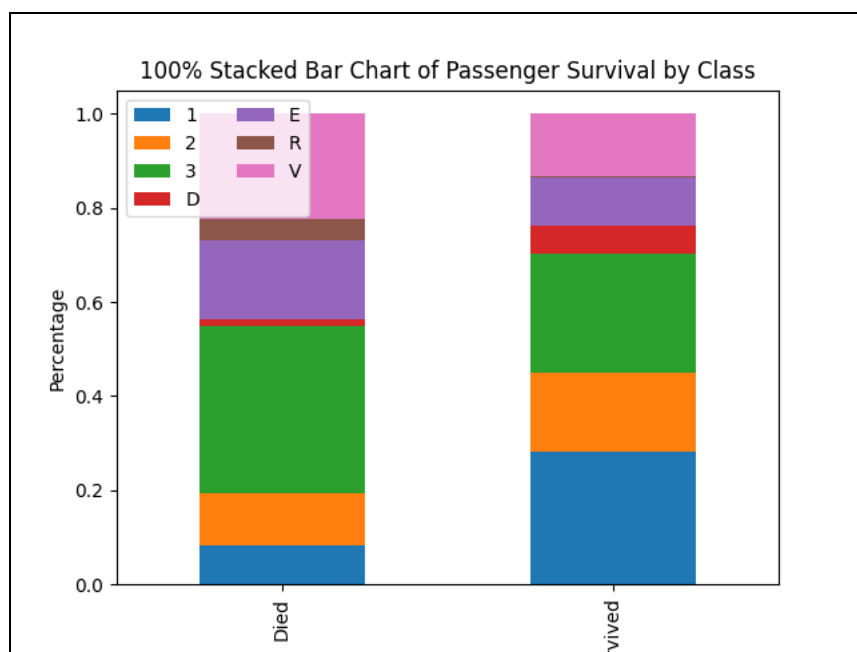
#9 Is the class a passenger was in related to whether they survived or not

Yes, the highest percentage of survivors were located in classes 1 and 2, with the majority of class 1 surviving.

Survived Died Survived

Class

1	123	201
2	166	119
3	530	180
D	23	43
E	251	72
R	66	3
V	337	94



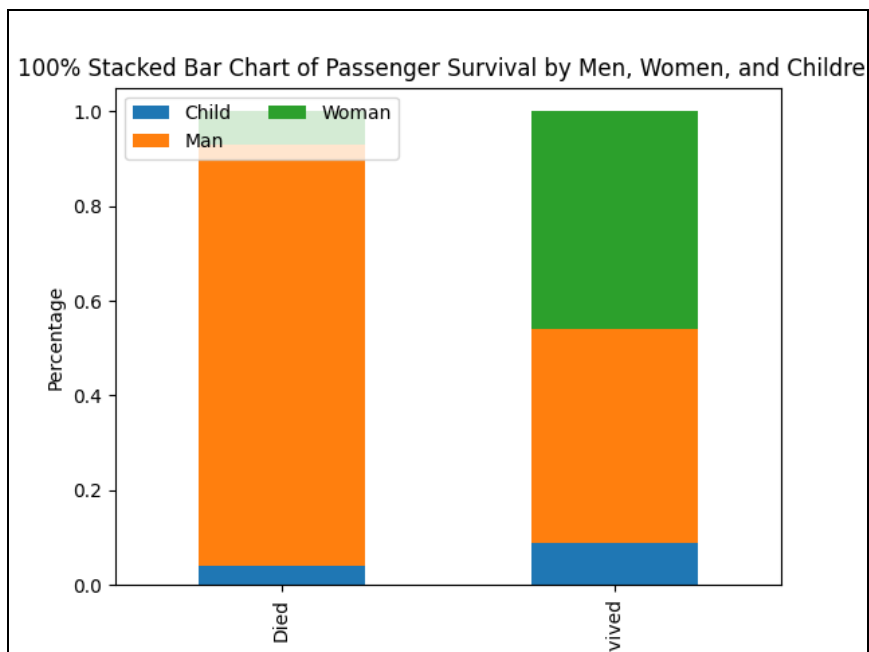
#10 Is whether the passenger was a man, woman, or child related to whether they survived or not

Yes, the identity of the passenger, Man, Woman, or Child played a role in their survival. More than half of the population of children and women survived.

Survived Died Survived

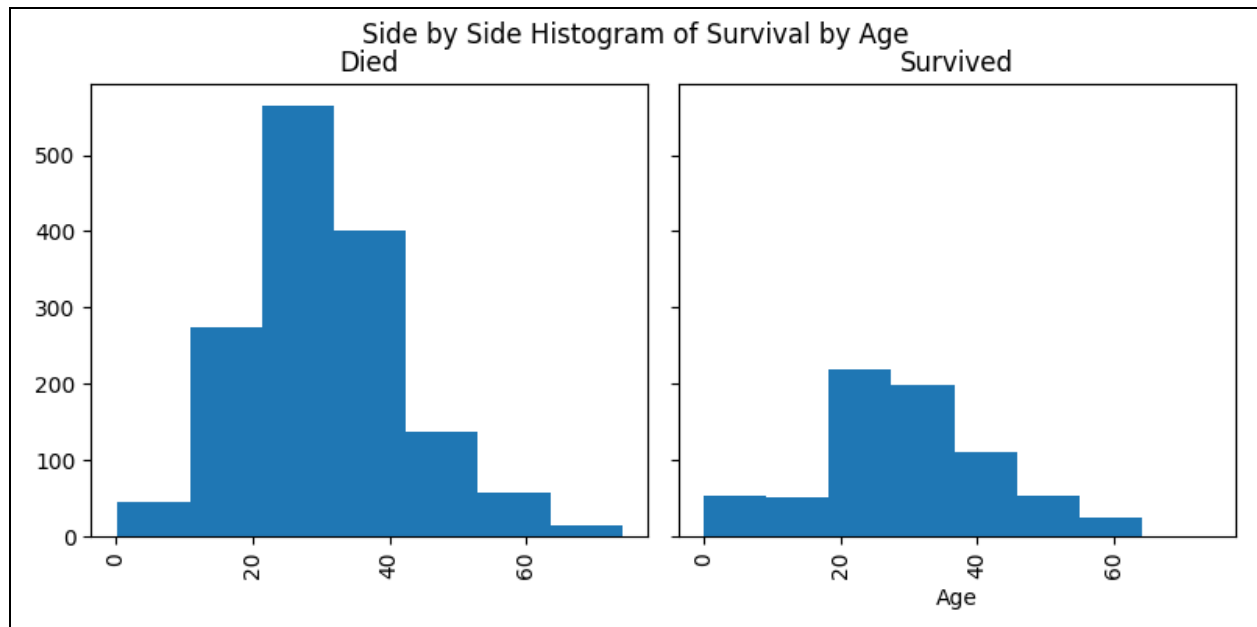
MWC

Child	60	64
Man	1331	321
Woman	105	327



#11 Distribution of age for the people who survived in comparison to those who died

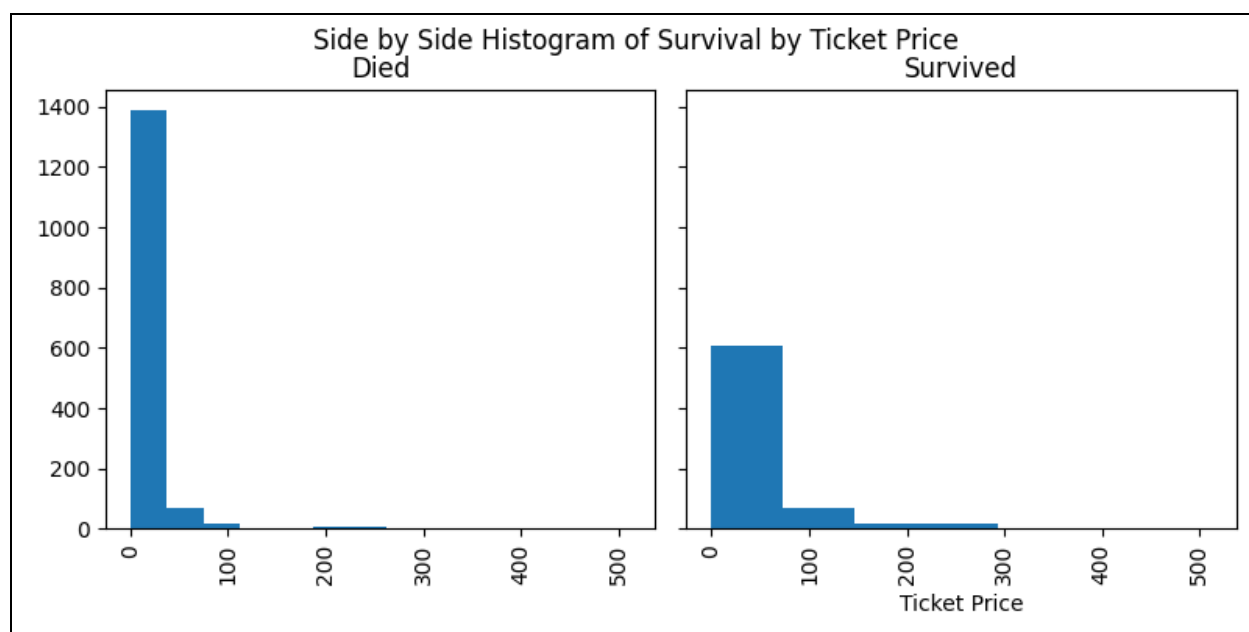
	count	mean	std	min	25%	50%	75%	max
Survived								
Died	1496	30.47	11.58	0.33	22.0	29.0	37.0	74.0
Survived	712	29.46	12.70	0.08	22.0	29.0	38.0	64.0



Based on the histogram and table of descriptive statistics, age plays little to no role in the passengers' chance of survival.

#12 Distribution of ticket prices for the people who survived in comparison to those who died

	count	mean	std	...	50%	75%	max
Survived							
Died	1496	19.06	25.53	...	14.4	14.40	263
Survived	712	39.07	60.07	...	14.4	35.81	512.33



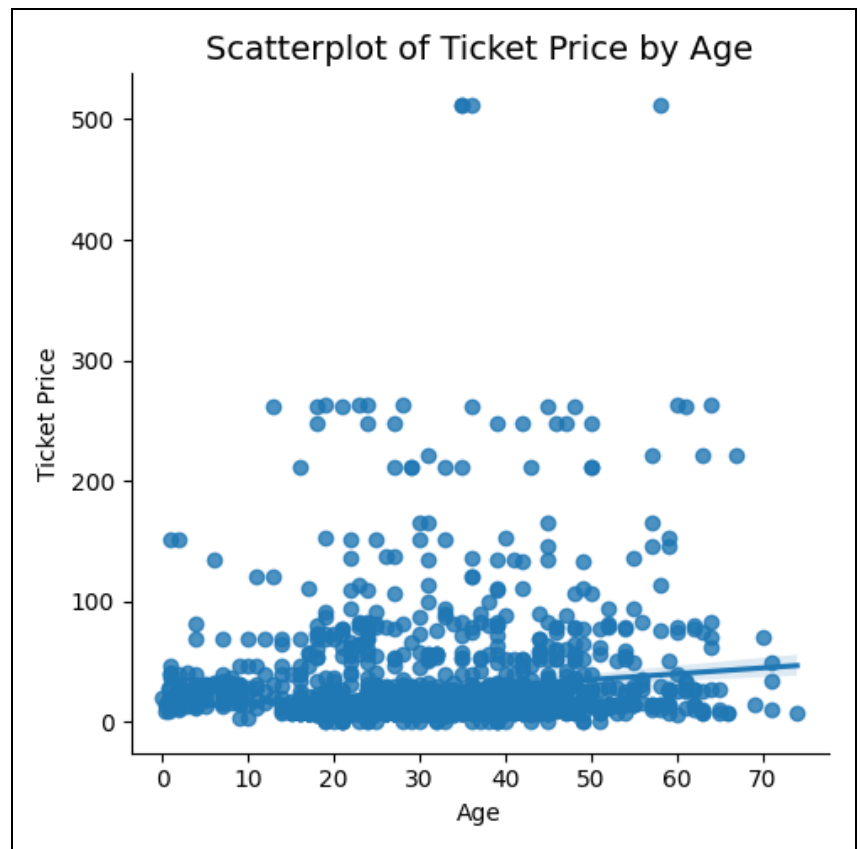
Based on the histogram and table of descriptive statistics we can make the argument that those who paid more for their ticket had a higher chance of surviving. From the survival rates of passenger class, we know over 50% of the passengers in 1st Class survived. So, ticket prices played a role in the passengers' ability to survive due to the location on the ship that the ticket was granted.

#13 Is there a relationship between age and the amount they paid

Correlation Coefficient:

0.1412

There is little to no relationship that exists between the ticket price and the age of the passenger.



#14 Frequency distribution of the new variables

Class	
Crew	889
Third	710
First	324
Second	285

#15 Frequency distribution of the new variables and their survival numbers

Age_levels

Middle Adult	695
Older Adult	614
Young Adult	465
Senior	85
Child	63
Teenager	49
Baby	43

Survived	Died	Survived
Age_levels		
Baby	16	27
Child	34	29
Middle Adult	479	216
Older Adult	417	197
Senior	61	24
Teenager	32	17
Young Adult	330	135

#16 What did I learn through my analysis

One thing I learned from the Titanic dataset was just how devastating the crash was. So many lives were lost and sacrificed for others. This is noticeable in the survival rate of the babies.

I also learned how to apply code to get a desired measurement answered. No matter how large the data set is, a couple of lines of code can help manipulate or calculate what you are looking for. Being able to pick apart such a large data set would be time-consuming on another software such as Excel. This project also helped significantly in my ability to read code. Understanding what each line's purpose is helps tremendously when figuring out error codes.