

# **EXPLORATORY DATA ANALYSIS FOR TITATIC DATA SET**

- PRADEEP. J

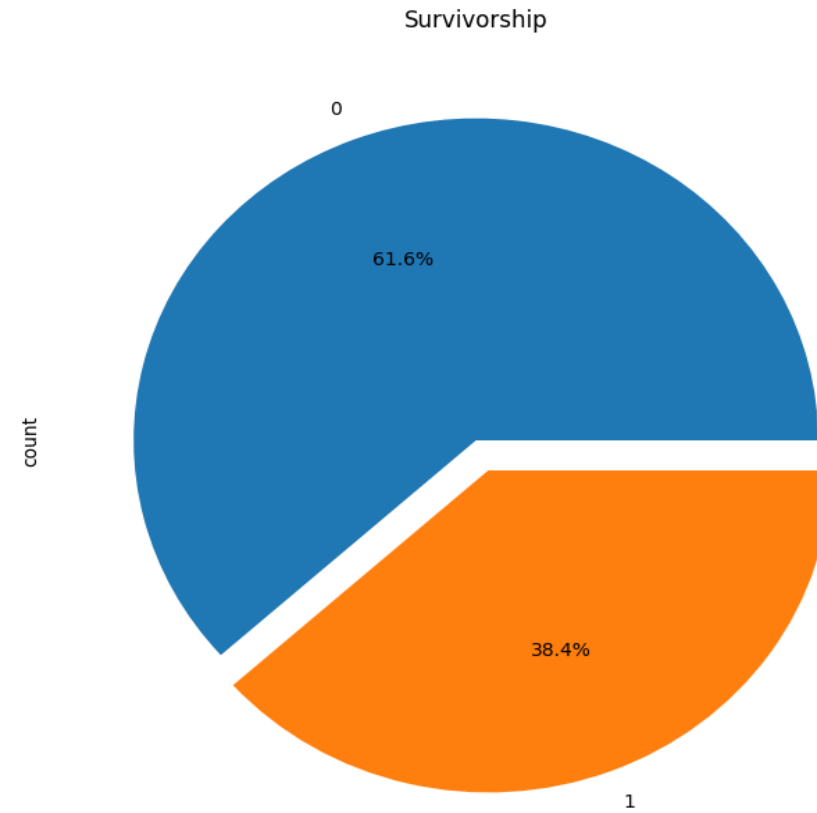
# INTRODUCTION

- In this titanic data set we have a 12 columns.
- There are a lot of continuous variable and some categorical variable in the data set. There are 177 null values in age column and 687 null values in cabin columns.

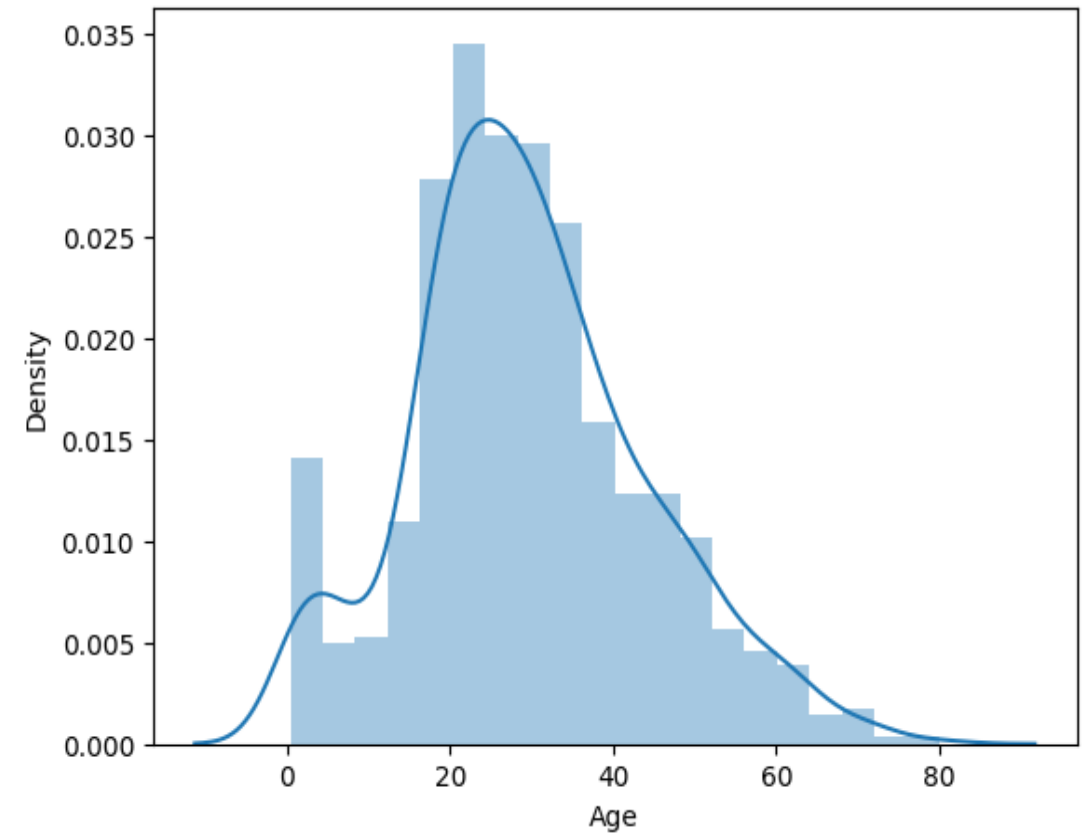
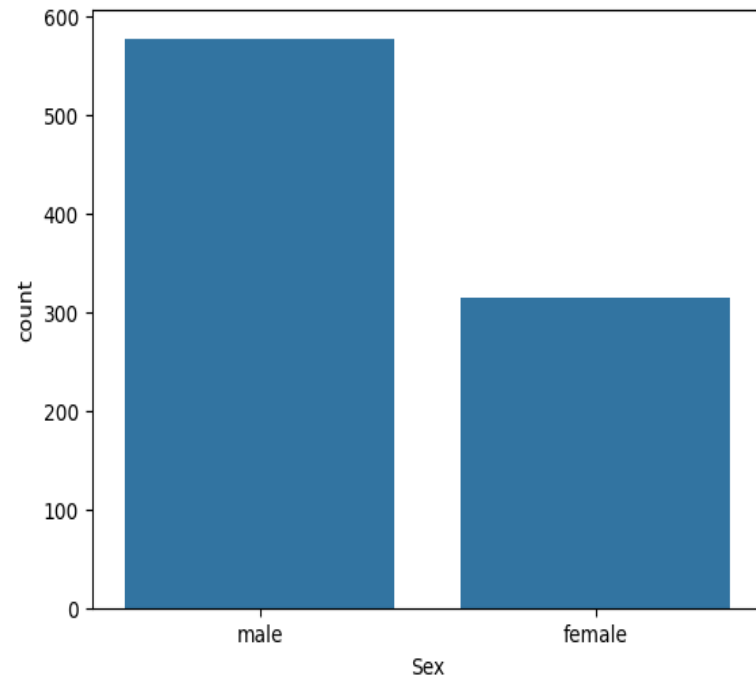
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

- The above table shows the count, mean, std, min, max, 25%, 50% and 75% values in titanic data set.

- Approximately 61.6% people are not survived in titanic ship. While 38.4 % people are survived in titanic ship.

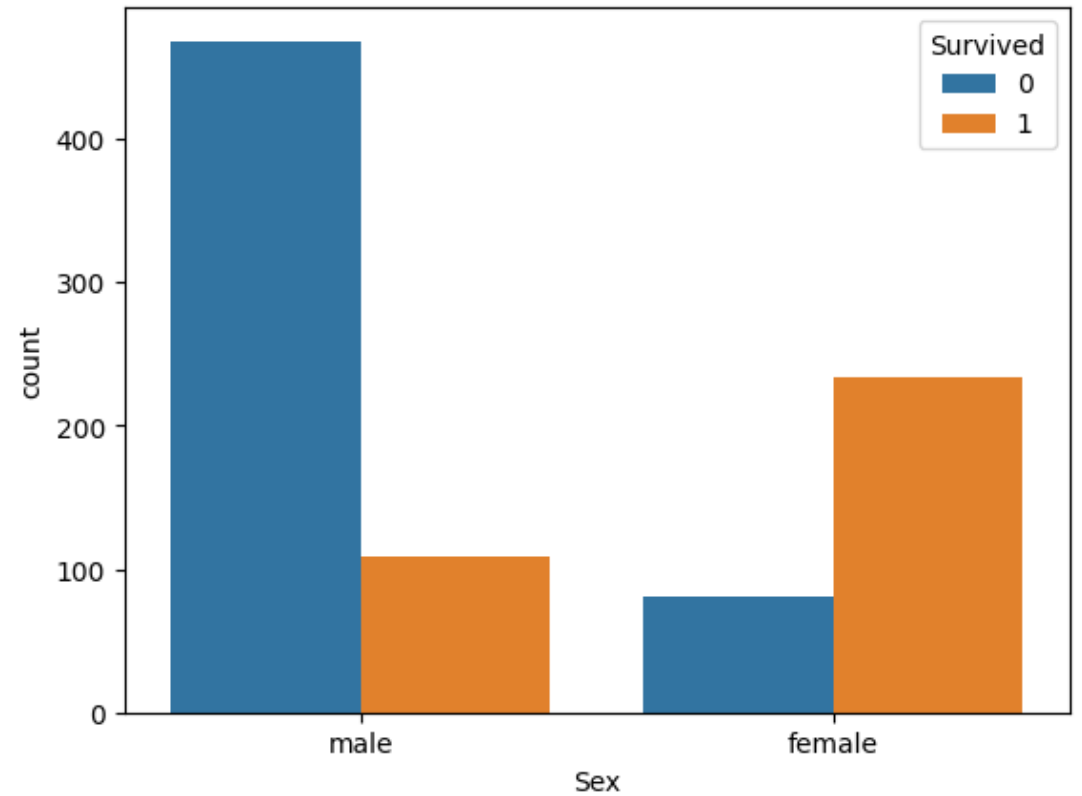


- Age Have A Normal Distribution .
- Mean=mode=median
- Most people between 20 to 40 age are traveling in the titanic ship. Few old people are travel in the ship.
- On the titanic ship, there were a total of 557 man and 314 women passenger are travelled.



- Higher percentage of women survived on the titanic compared to men.
- Specifically, 468 mans are did not survived in the ship . While unfortunately , 223 women survived in the ship.

```
.. Sex      Survived
   female  0         81
        1        233
   male   0        468
        1        109
   Name: Survived, dtype: int64
```

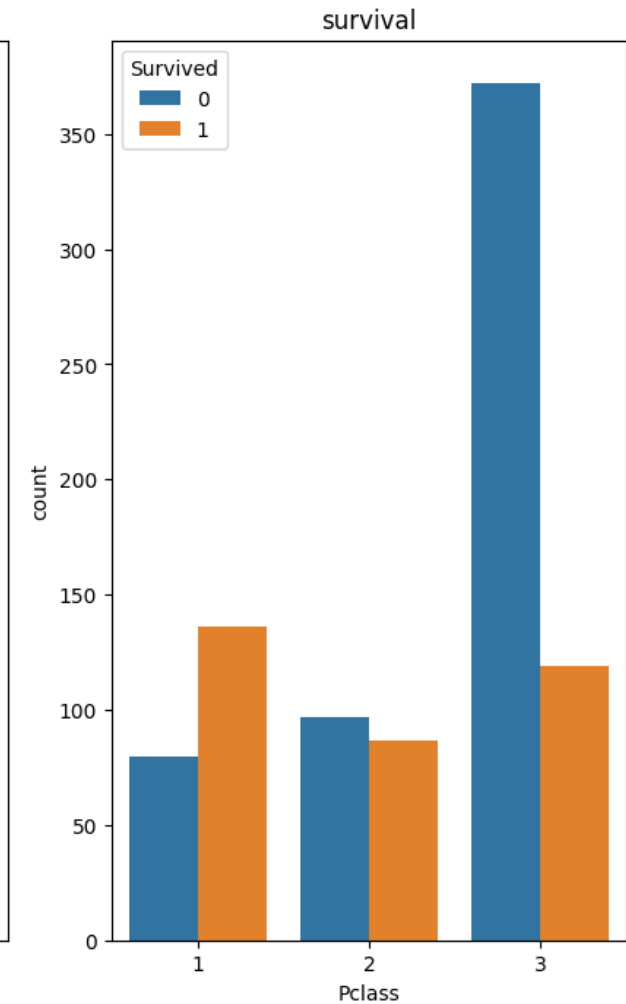
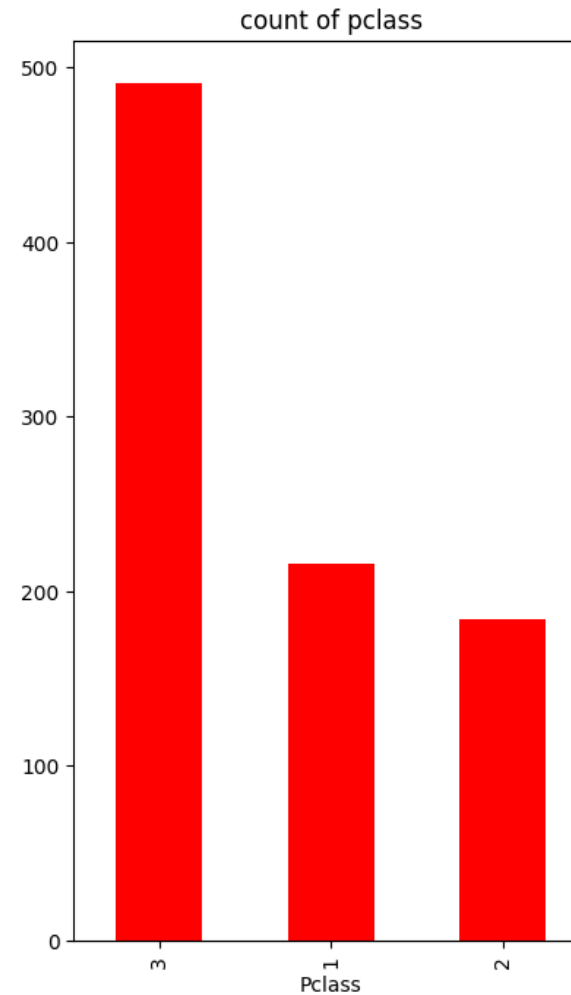


- Majority of passengers traveled in third class.so survival rate is higher in class third. Out of 372 people are survived.
- Majority of first class passenger not survived in ship compared to other class. 130 above people did not survives in the ship.
- In second class 97 passengers are survived . while unfortunately ,87 passengers did not survived.

```

Pclass  Survived
1        0         80
        1        136
2        0         97
        1         87
3        0        372
        1        119
Name: Survived, dtype: int64

```

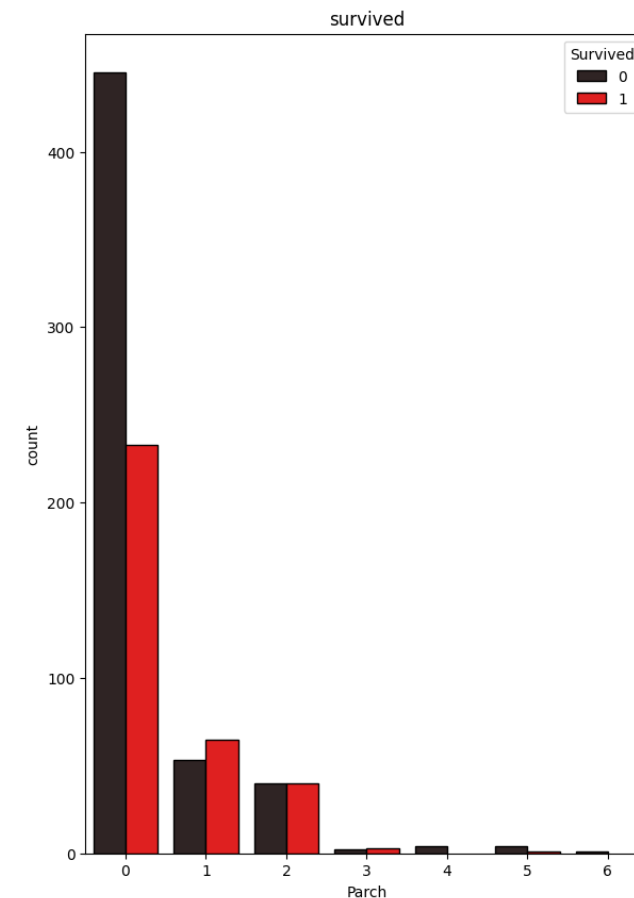
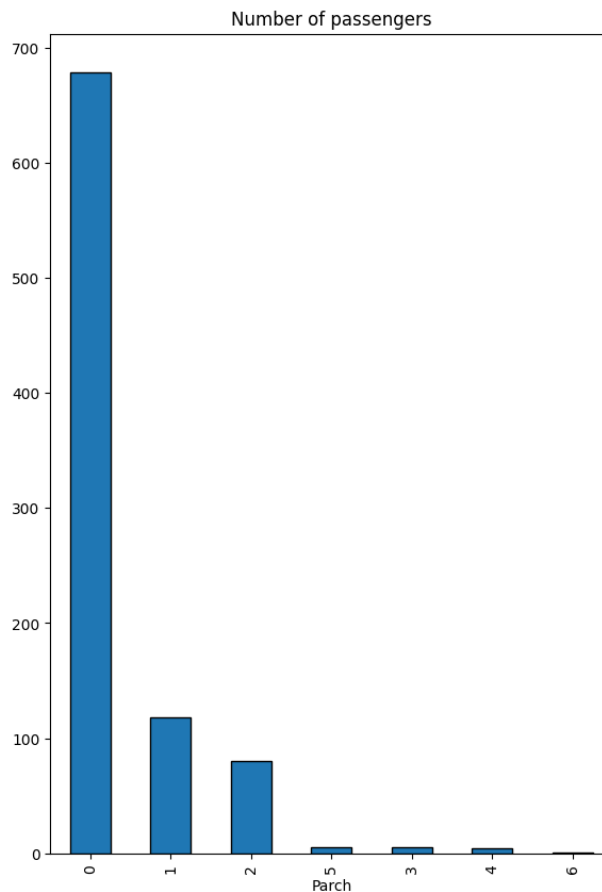


- Here to the result are quit similar . who has 0 to 3 parent on ship they has a higher survival rate.
- Majority of passengers has a 0 parent. (443 passenger survived)
- But on one survived who had four parent on the titanic ship.
- the Not survived rate was higher who had a one parent.
- the survived rate was equal in who has a two parent.(not survived = survived).

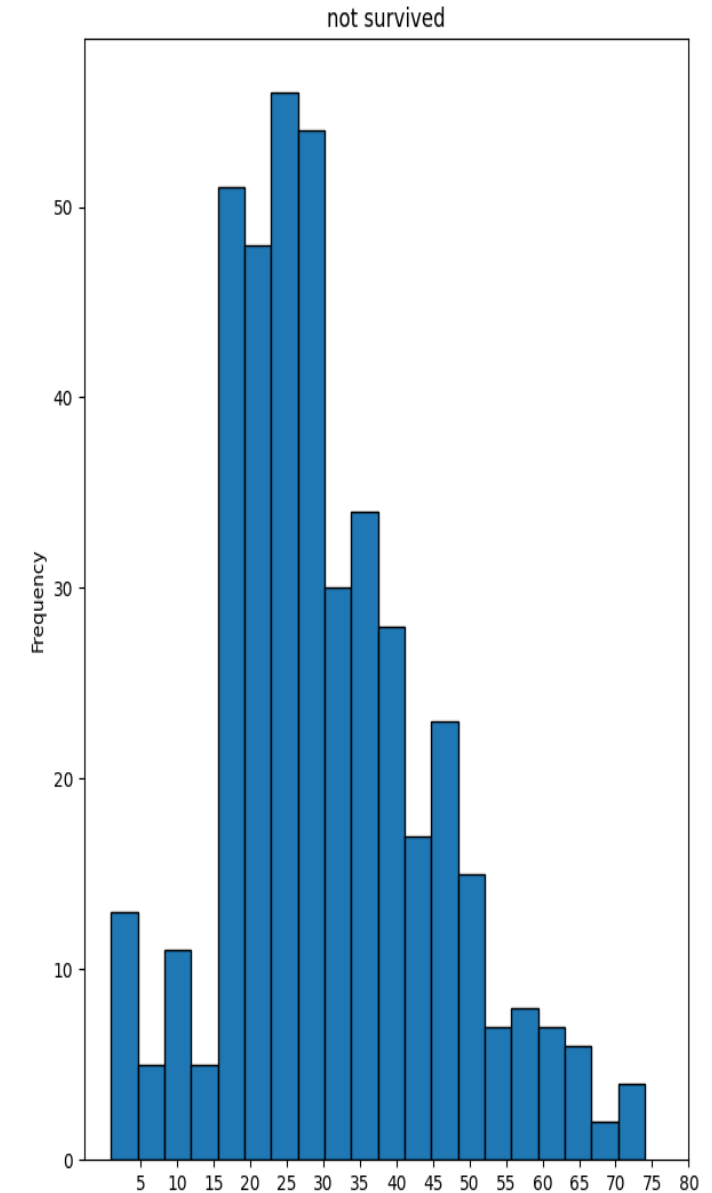
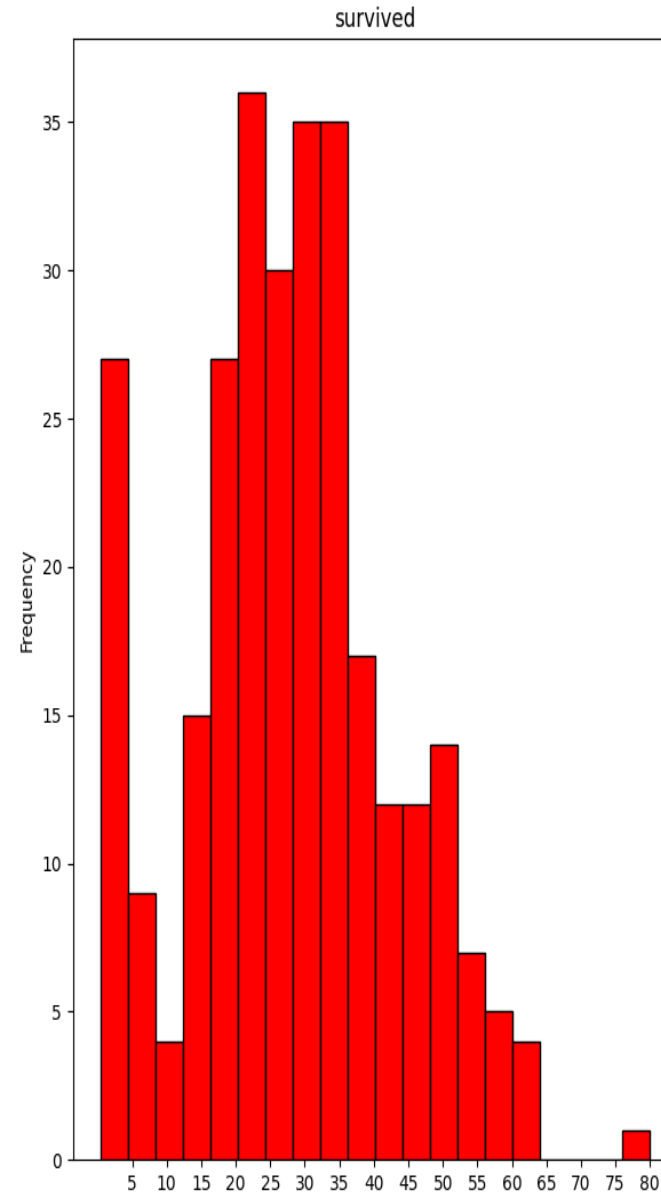
```

Parch  Survived
0      0      445
      1      233
1      0      53
      1      65
2      0      40
      1      40
3      0       2
      1       3
4      0       4
5      0       4
      1       1
6      0       1
Name: Survived, dtype: int64

```

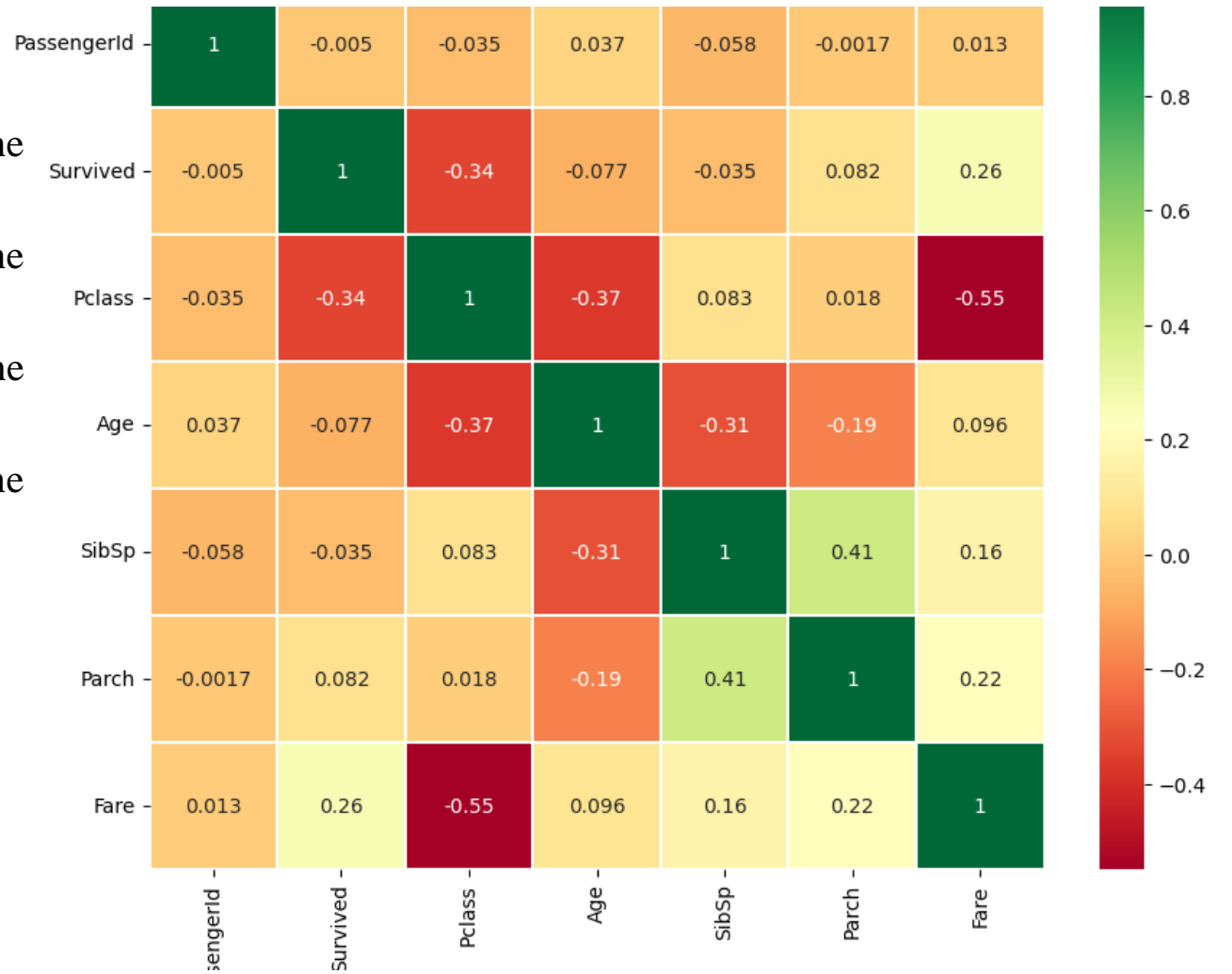


- Majority of 0-5 age passenger was survived in the ship. They give most prioritized for children and old man.
- There isn't much different in survival rate between 15 to 35 age passengers.
- Every passenger between the age of 75 to 80 survived in the ship.
- None of the passengers between the age of 65 to 75 survived in ship.





- Pclass and fare have highly negative correlation. the correlation rate is -0.55.
- Sibsp and parch have a positive correlation. the correlation rate is 0.41.
- Pclass and age have highly negative correlation. the correlation rate is -0.37.
- sibsp and pclass have highly negative correlation. the correlation rate is -0.55.



- Maximum passengers boarded from S. Majority of them being from Pclass3.
- The Passengers from C look to be lucky as a good proportion of them survived. The reason for this maybe the rescue of all the Pclass1 and Pclass2 Passengers.
- The Embark S looks to the port from where majority of the rich people boarded. Still the chances for survival is low here, that is because many passengers from Pclass3 around **81%** didn't survive.

