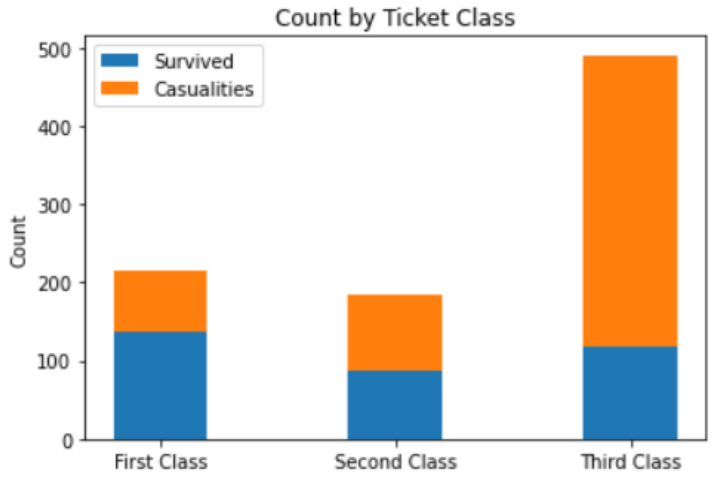
## Assignment 3 – Summary Report

## GitHub Repo link: https://github.com/KajalMittal01/Titanic_A3

## Determine if the survival rate is associated to the class of passenger:

```
First class passengers percentage =  24.242424242424242  | Count:  216
Second class passengers percentage =  20.65095398428732  | Count:  184
Third class passengers percentage =  55.106621773288445  | Count:  491
Survival rates =  Pclass
1    39.766082
2    25.438596
3    34.795322
Name: Survived, dtype: float64  | Count:  Pclass
1    136
2     87
3    119
Name: Survived, dtype: int64
```




First Class represents approximate 24% passengers and approx. 40% of survivors.

Second Class represents approx. 20% passengers and ~25% of survivors.

Third Class represents ~ 55% passengers and ~ 34% of survivors.

We can see that there is a clear correlation between death and ticket class: the higher the ticket class, the higher the chances of survival.

**Chi-Square Analysis Result:**

```
Power_divergenceResult(statistic=1592.1666666666665, pvalue=2.0399395484622444e-42)
```

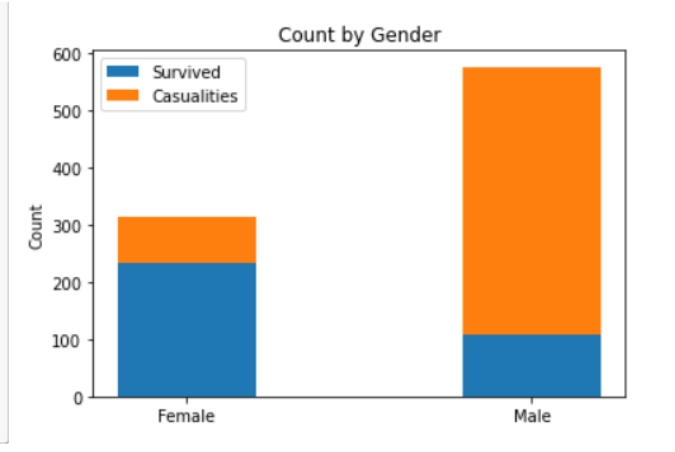
P-value is less than 0.05, hence we have to reject the null hypothesis, which means there is a positive relationship between Survival and Pclass variables.

**Determine if the survival rate is associated to the gender:**

```
TOTAL COUNT BY GENDER
Females count: 314
Males count: 577

SURVIVAL COUNT BY GENDER
Female survival count:  233
Male survival count:  109

CASUALTIES COUNT BY GENDER
Female death count:  81
Male death count:  468
```




According to above analysis it is clear that women had a higher chance of survival than men. Let's verify this by using Spearman's correlation between Survival and Gender.

| | Sex | Survived |
|---|---|---|
| **Sex** | 1.000000 | 0.543351 |
| **Survived** | 0.543351 | 1.000000 |

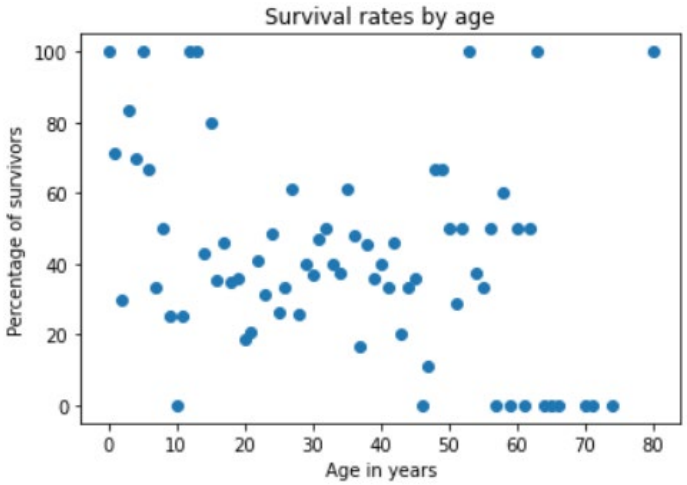
Co-relation value is coming as "0.54" which is also suggesting that there is a positive and moderate correlation between Sec and Survival.

**Chi-Square Analysis Result:**

```
(260.71702016732104,
 1.1973570627755645e-58,
 1,
 array([[193.47474747, 355.52525253],
        [120.52525253, 221.47474747]]))
```

P-value is less than 0.05, hence we have to reject the null hypothesis, which means there is a positive relationship between Survival and Gender variables.

## Determine the survival rate is associated to the age:




We can interpret from the scatterplot above that younger people have more chances to survive than older because higher Y values have lower X values.