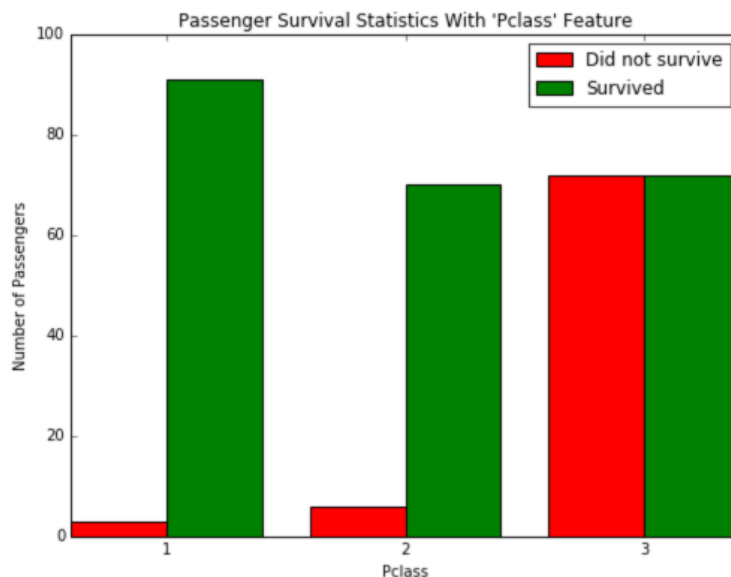


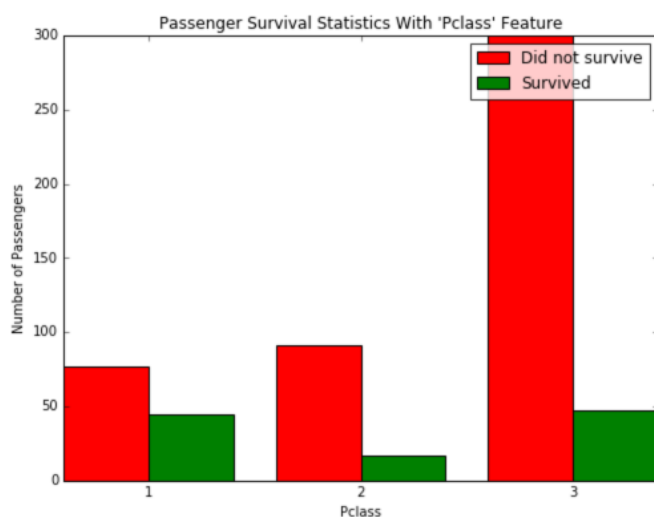
- The prediction mechanism until now didn't sub-categorise women at all, we just assumed they all survived. I hypothesise we can improve the accuracy of the algorithm by classifying some women as non-survivors.
- First I checked to see if class (Pclass) makes it less likely for a woman to survive. It does – those in class 3 are much less likely to survive, however chance is still 50% so it doesn't help with the classification since it doesn't change the default outcome of survive.

```
vs.survival_stats(data, outcomes, 'Pclass', ["Sex == 'female'"])
```



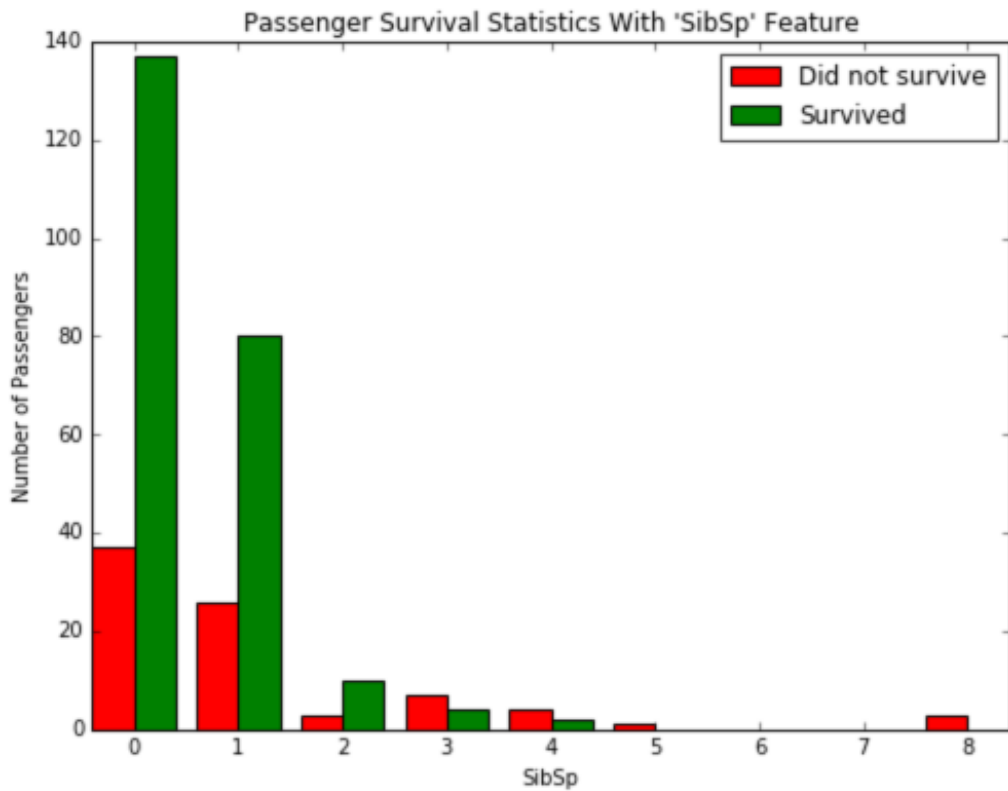
- I also checked Pclass for men. Again Pclass = 3 is much less likely to survive, but doesn't change default prediction of not survive.

```
vs.survival_stats(data, outcomes, 'Pclass', ["Sex == 'male'"])
```

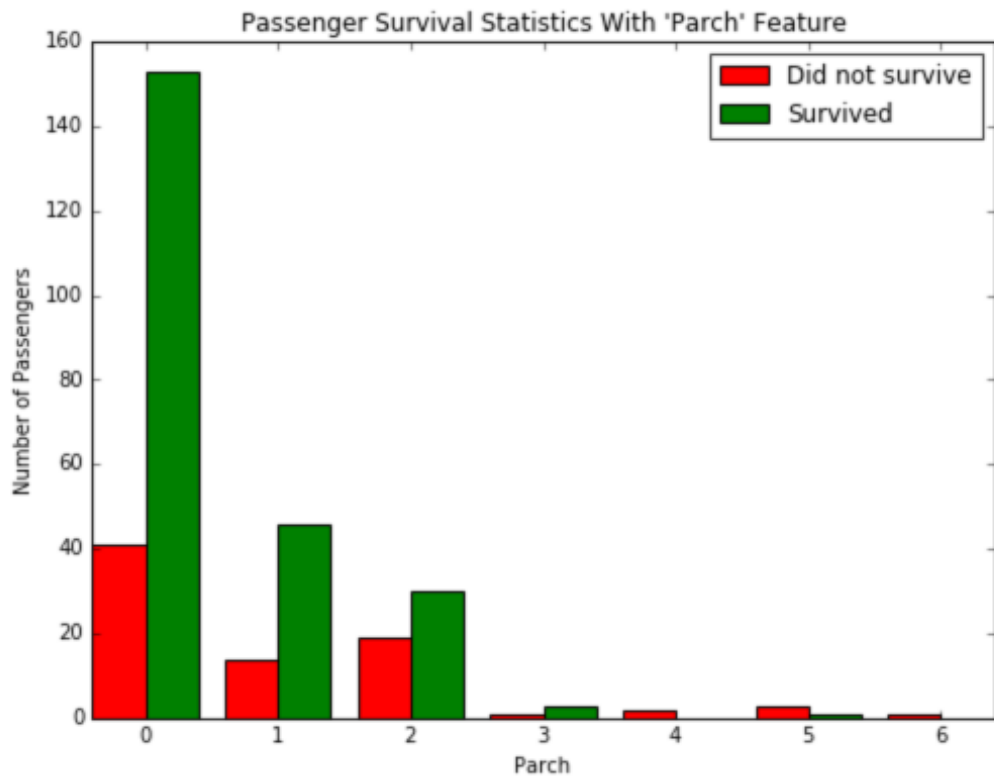


- Digging deeper for Pclass for Male > 18 also gives similar results, so it doesn't help the prediction
- For Male neither number of siblings (SibSp) nor Parch (number of parent/children) makes a difference to default outcome
- For female it appears they are less likely to survive if SibSp > 3
 - Added this to algorithm -> brings up score to 80.02%

```
vs.survival_stats(data, outcomes, 'SibSp', ["Sex == 'female'"])
```



- It also appears female less likely to survive if Parch > 4
 - Add to algorithm -> brings up score to 80.36%



- By adding these two features (Female & SibSp > 3 and Female & Parch > 4) reduced the number of misclassifications of women, and hence improved total accuracy to

80.36%, more than the target of 80%. I choose not to add any more features to avoid overfitting.