Information on test set vs training set

In this chart, I have the numbers of passengers for pclass and sex to see if the two sets match up. I put the ratio between male and female for each class since we know that the split between male and female will be important. I also put the percent of total for each class.

I don’t particularly see anything that jumps out at me that is a problem. If anyone else has other ideas, please let me know.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Test set |  |  |  |  | Training set |  |  |  |  |
| Pclass | Sex | Number | ratio | % of total | Pclass | Sex | Number | ratio | % of total |
| 1 | M | 57 | 1.14 | 13.6 | 1 | M | 122 | 1.3 | 13.7 |
| 1 | F | 50 |  | 12 | 1 | F | 94 |  | 10.5 |
| 2 | M | 63 | 1.05 | 15 | 2 | M | 108 | 1.42 | 12.1 |
| 2 | F | 30 |  | 7.2 | 2 | F | 76 |  | 8.5 |
| 3 | M | 146 | 2 | 35 | 3 | M | 347 | 2.4 | 38.9 |
| 3 | F | 72 |  | 17 | 3 | F | 144 |  | 16.6 |