CMPE 188 HW1 Submitted

1. Answer is 3
2. 0.25 and 0.375
   1. Normalization for each row based on median value from each column.
3. 6 samples missing values
4. All the attributes of the weather dataset are numerical.
   1. False
5. Because the missing values in each column are not necessarily in the same row
6. The youngest passenger was an infant and the oldest was 80 years old.
7. The most strong positive correlation to the survival attribute exists between this and the fare.
8. Gender

|  |  |
| --- | --- |
|  | Turning the categorical variables into numerical, which of the following is a good strategy (take as examples the ones provided in the tutorial)?  Create as many numerical variables for each categorical one. |
|  |  |