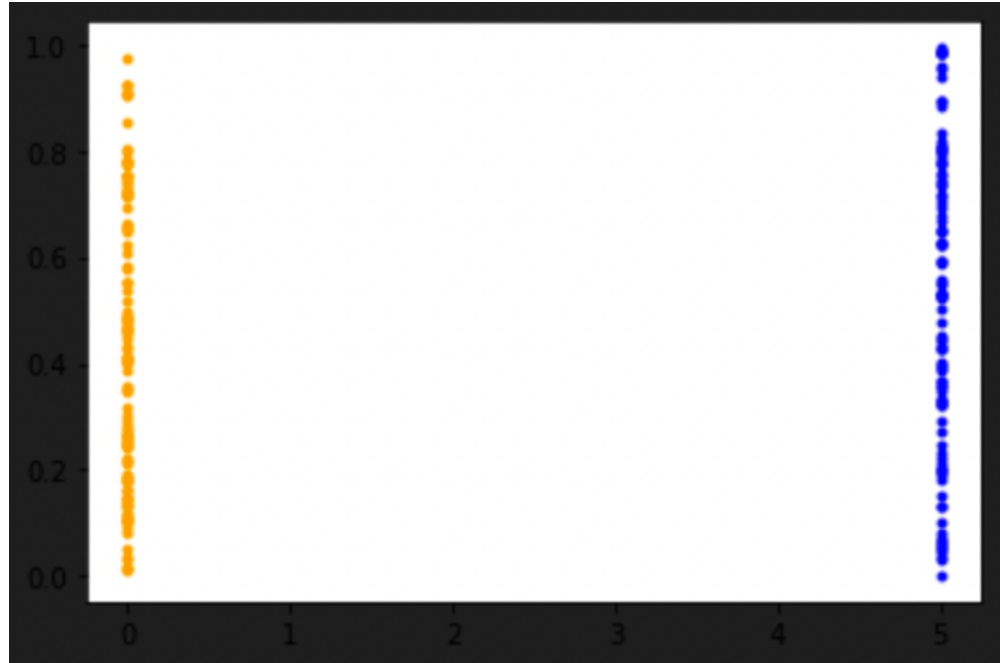
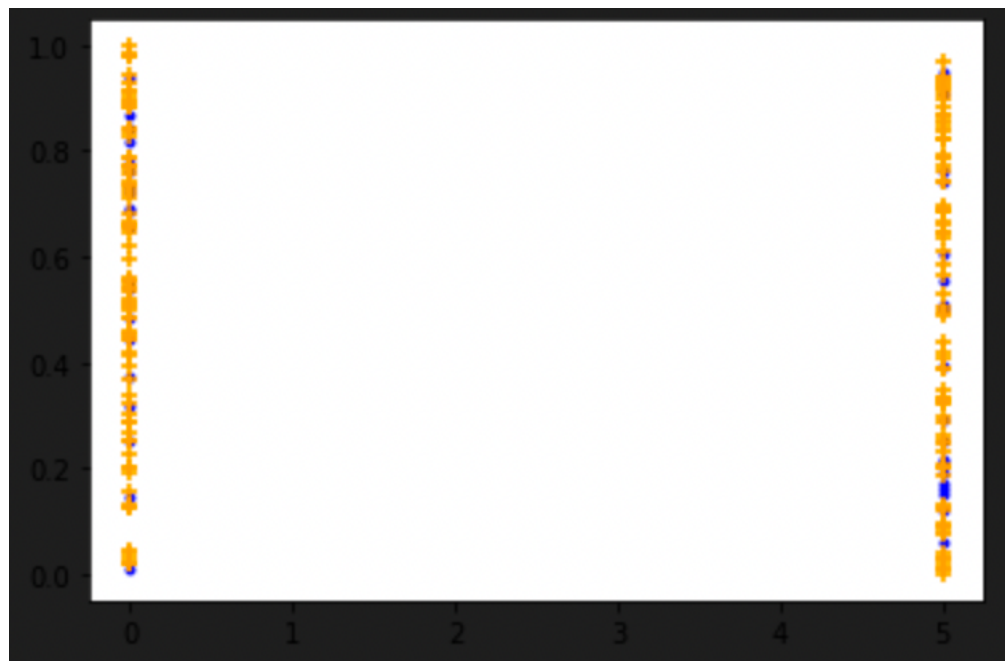


3.) a.)



3.) b.) Orange = training data  
Blue = test data



3.) c.) I predict that a 1-NN classifier will do very well in this training and test set because the single nearest neighbor for each of the blue test data points is one of the orange training data points so there should be a true positive rate of 20 and a false positive rate of 0 in the confusion matrix.

3.) d.) Confusion Matrices:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

3.) e.) Trial 1:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

Trial 2:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

Trial 3:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

Trial 4:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

Trial 5:

```
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
[[20  0]
 [ 0 20]]
```

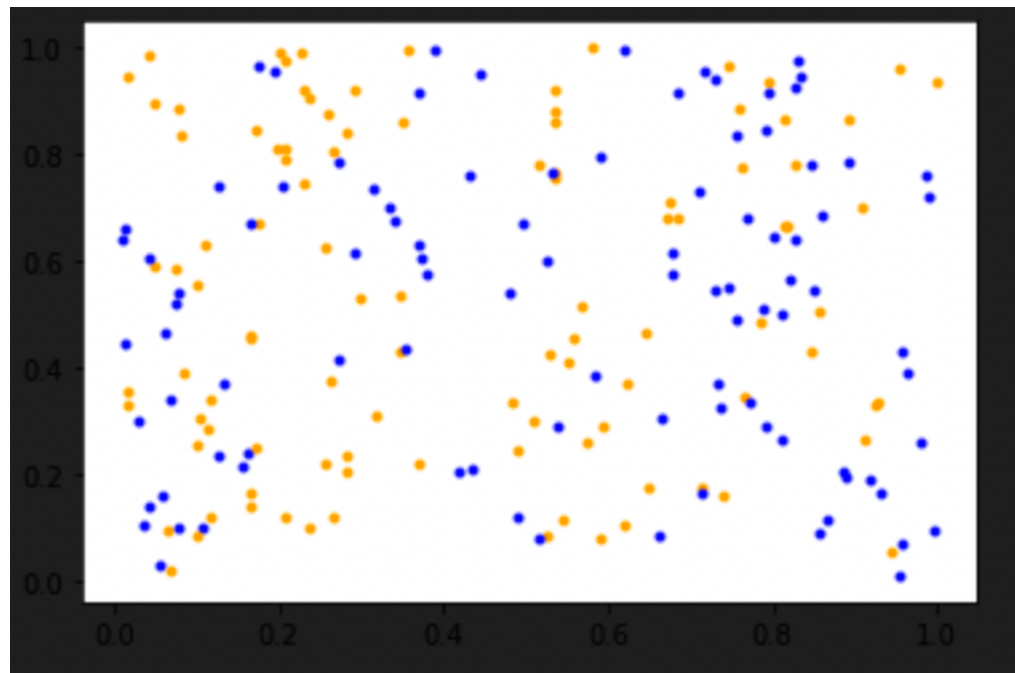
Mean of Confusion Matrices:

20	0
0	20

Standard Deviation of Confusion Matrices:

0	0
0	0

4.) a.)



4.) b.)

Confusion matrices:

```
. [[19 21]
  [16 24]]
[[15 25]
 [21 19]]
[[16 24]
 [25 15]]
[[15 25]
 [25 15]]
[[19 21]
 [24 16]]
```

4.) c.) Trial 1:

```

[ [19 21]
  [16 24]]
[ [15 25]
  [21 19]]
[ [16 24]
  [25 15]]
[ [15 25]
  [25 15]]
[ [19 21]
  [24 16]]

```

Trial 2:

```

[ [19 21]
  [18 22]]
[ [16 24]
  [25 15]]
[ [22 18]
  [22 18]]
[ [22 18]
  [21 19]]
[ [20 20]
  [19 21]]

```

Trial 3:

```

[ [21 19]
  [20 20]]
[ [15 25]
  [24 16]]
[ [20 20]
  [23 17]]
[ [16 24]
  [17 23]]
[ [17 23]
  [24 16]]

```

Trial 4:

```

[ [16 24]
  [23 17]]
[ [24 16]
  [20 20]]
[ [15 25]
  [22 18]]
[ [16 24]
  [24 16]]
[ [18 22]
  [22 18]]

```

Trial 5:

```

[ [19 21]
  [19 21]]
[ [17 23]
  [20 20]]
[ [21 19]
  [23 17]]
[ [22 18]
  [20 20]]
[ [20 20]
  [21 19]]

```

Mean of Confusion Matrices:

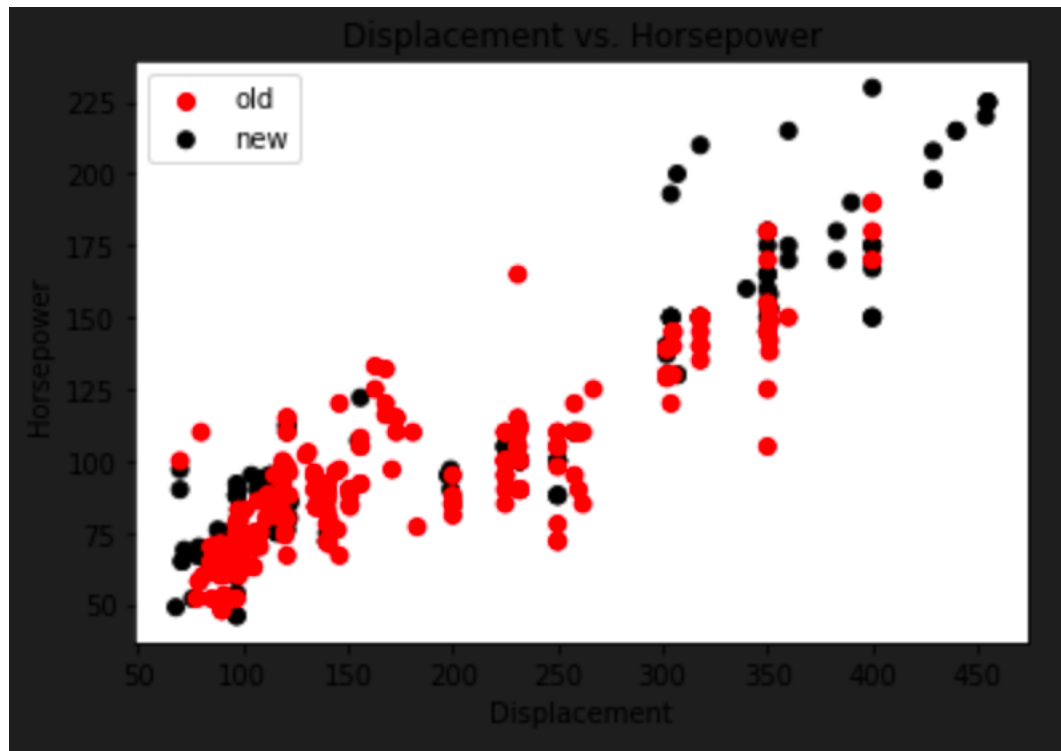
18.40	21.60
21.52	18.48

Standard Deviation of Confusion Matrices:

2.62	2.62
2.48	2.48

5.) a.) done within the code

5.) b.)



5.) c.)

```
[[ 6 24]
 [ 8 40]]
[[20 10]
 [17 31]]
[[14 16]
 [ 9 39]]
[[20 10]
 [13 35]]
[[14 16]
 [19 29]]
```

Mean of confusion matrices:

14.8	15.2
13.2	34.8

Standard deviation of confusion matrices:

5.15	5.15
4.31	4.31

5.) d.)

7-NN with 5-fold cross-validation

Confusion matrices:

Mean of confusion matrices:

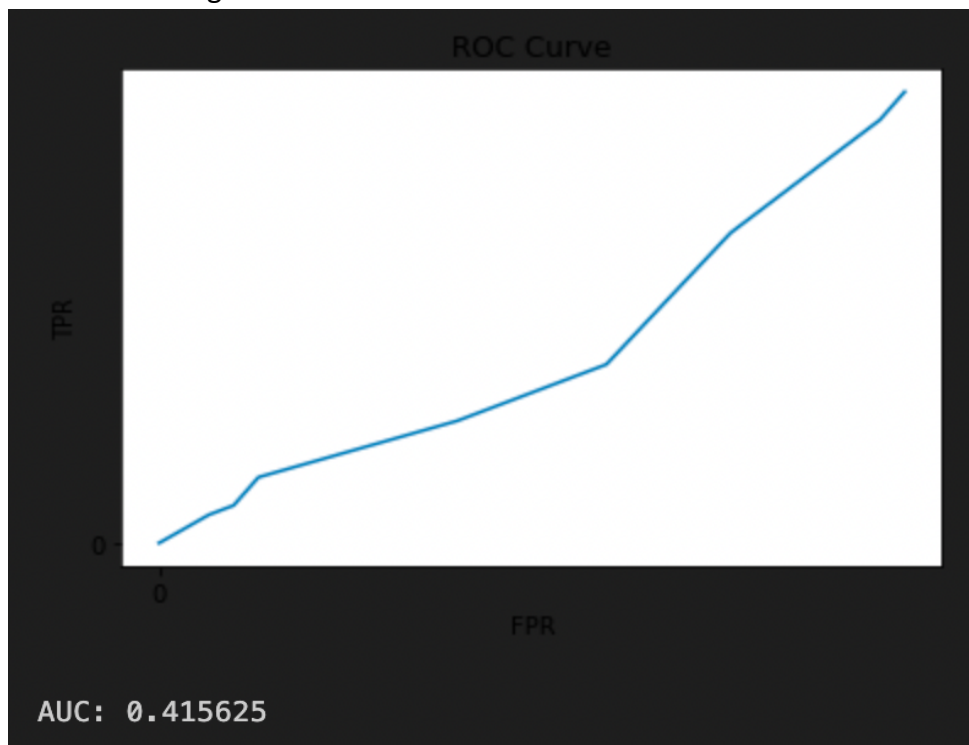
13	17
9.8	38.2

Standard Deviation of confusion matrices:

3.46	3.46
2.14	2.15

```
[[10 20]
 [ 7 41]]
[[16 14]
 [10 38]]
[[18 12]
 [ 8 40]]
[[ 9 21]
 [11 37]]
[[12 18]
 [13 35]]
```

ROC curve along with the AUC:



**Interpretation:** From this ROC curve, we have gathered that we have an AOC (Area under curve) of 0.41. The AUC is the measure of the ability of a classifier to distinguish between classes and is a representative summary statistic of the ROC curve. The higher the AUC, the better the performance of the model at distinguishing between positive and negative classes. AUC = 1 would be a perfect distinction. Therefore, with our AUC value of 0.41, we have a slightly below average ability to distinguish between positive and negative classes with our ROC curve.

**5.) e.)** The combination of predictors that seems to provide the best performance according to my code is the combination of all four possible predictors. This may be because the algorithm takes into account the largest variety of variables that help account for the most accurate True Positive and False Positive Rates because there are so many possibilities that are all accounted for. Also, the combinations of all four possible predictors outputted the highest AUC rate which was the closest to 1. Since we know that an AUC of 1 is the best possible case for an ROC curve, it is reasonable to assume that this combination of all four predictors seems to provide the best performance. For the same reasons, the combinations of singletons seemed to provide the worst performance due to the possibility that not all variables were accounted for and thus outputted relatively low AUC's ( $<0.5$ ).