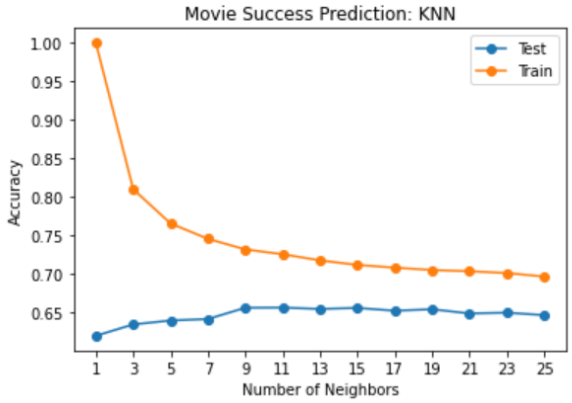
RESULTS OBTAINED - SCREENSHOTS - MOVIE SUCCESS PREDICTION

```
Logistic Regression Accuracy
Best Score = 0.7037
Best Hyper-parameters = {'C': 10, 'solver': 'newton-cg'}
```

Solver	Mean Test Score	Mean Train Score	C
sag	0.5268	0.5266	1
lbfgs	0.5266	0.5266	1
saga	0.5270	0.5268	10
newton-cg	0.7047	0.7055	1000

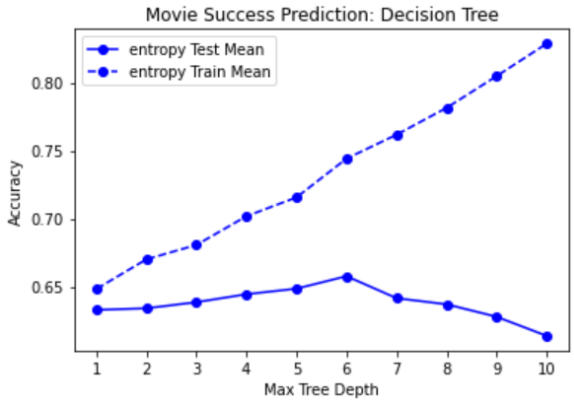
The best logistic regression model has a mean test score of 0.7047 and a mean train score of 0.7055. The difference between these values is very small, so the model is neither overfitted nor underfitted.

```
KNN Accuracy
Best Score = 0.6557
Best Hyper-parameters = {'n_neighbors': 11}
```



The graph shows that the difference in test score and training score is larger when the number of neighbors is low. This shows that the model is more overfitted when the number of neighbors is lower. But the test score and training score seem to converge as the number of neighbors increase. The best KNN model has a mean test score of 0.6557 and a mean train score of 0.7251. The hyper-parameter of this model is {n_neighbors: 11}.

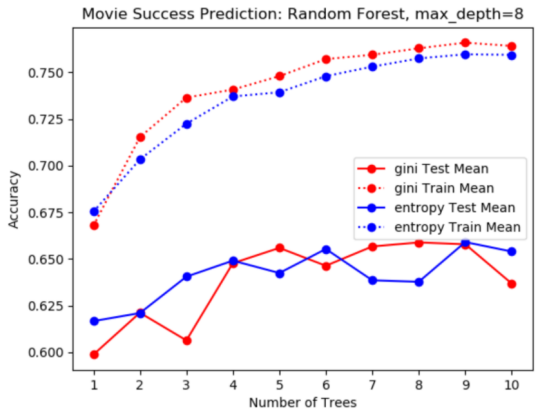
```
Decision Tree Regression Accuracy
Best Score = 0.6577
Best Hyper-parameters = {'max_depth': 6}
```



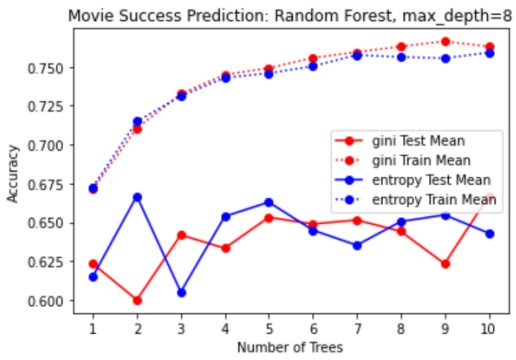
The graph shows that the test score decreases and the training score increases as the max tree depth increases. This shows that the tree is more overfitted as it grows larger. The best decision tree model has a mean test score of 0.6581 and mean training score of 0.7441. The hyper-parameters of this model are{criterion: entropy, max_depth: 6}

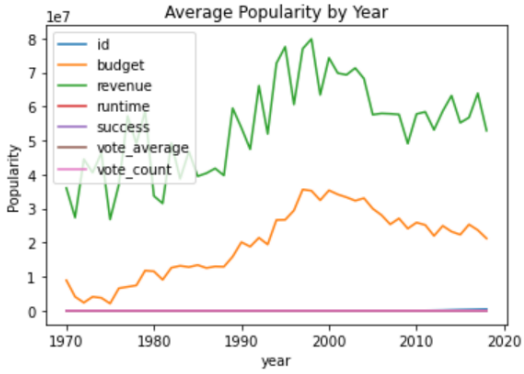
```
Random Forest Accuracy
Best Score = 0.6667
Best Hyper-parameters = {'n_estimators': 10}
```

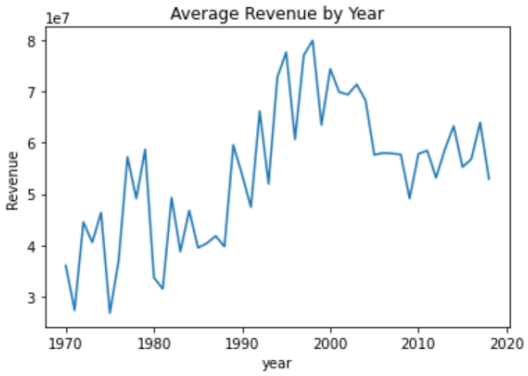
```
Random Forest Accuracy
Best Score = 0.6669
Best Hyper-parameters = {'n_estimators': 2}
```

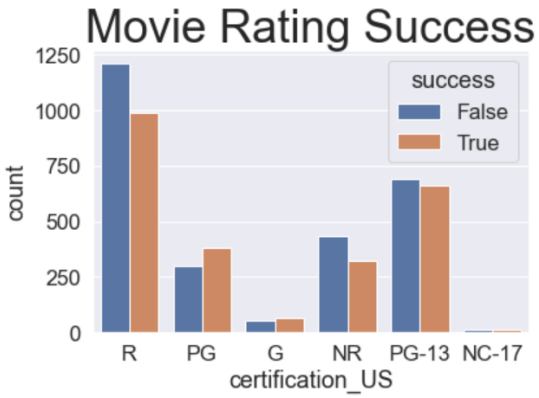


The random forest accuracy is greater than the decision tree accuracy. The best random forest model has a mean test score of 0.6735 and mean training score of 0.7558. The hyper-parameters of this model are {criterion: gini, max_depth: 8, n_estimators: 9}

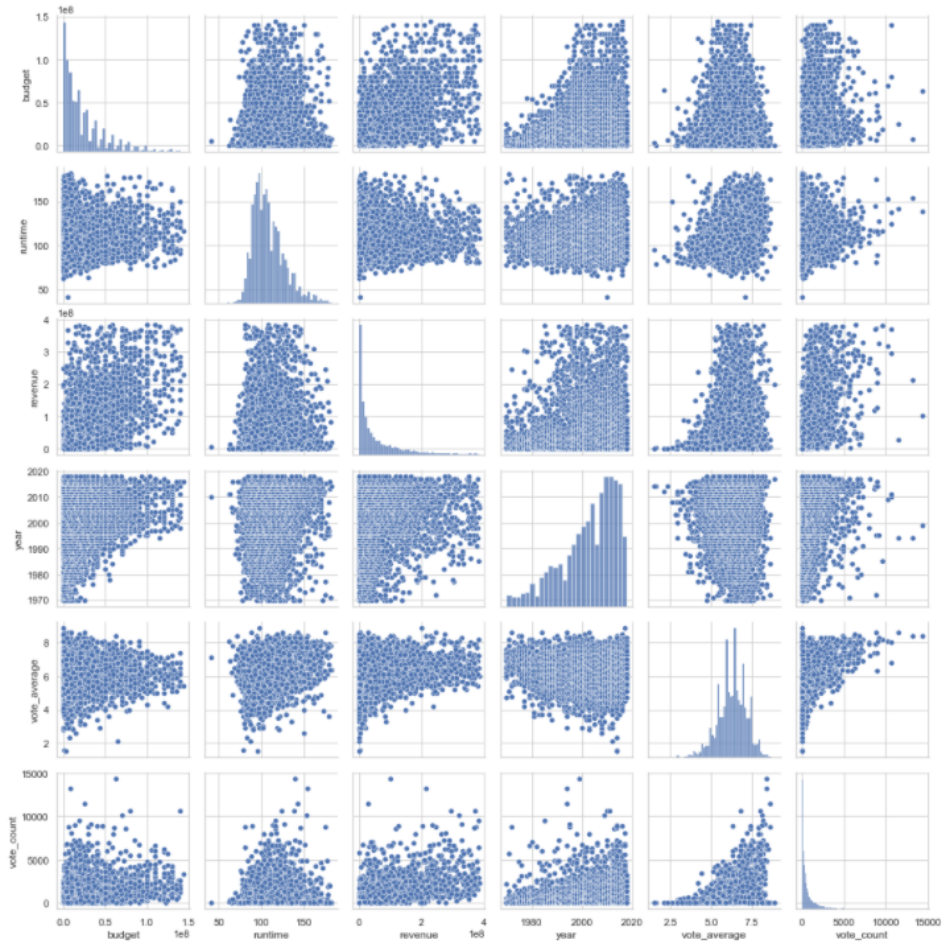








It seems that R-rated movies make up most of the dataset. But the only certification types where success is greater than failure is PG and G-rated movies, which are made for children.
Comedy, drama, and action movies make up most of the dataset. The genres with a positive success ratio are comedy, horror, and documentary.



The scatterplot matrix above shows the relationships between the continuous features.