Data and Results

\*\*\*Add words before graph even if it’s a few sentences.

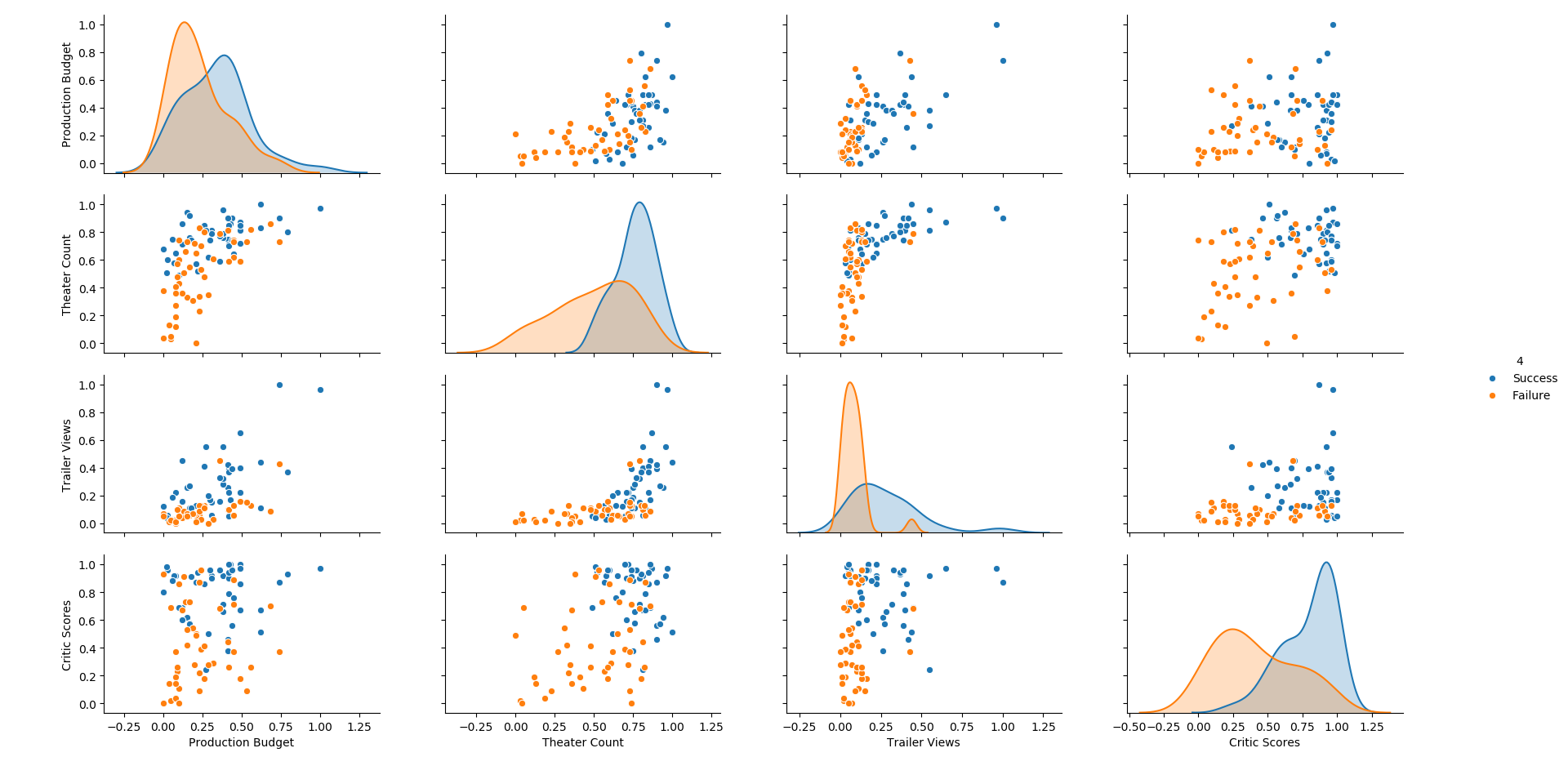


Figure 1

The results of the input data represented in scatter plots for two input variables compared to one another, along with histograms of the input variables themselves.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Movie Title | Production Budget | Opening Theater Count | YouTube Trailer count (m) | Rotten Tomatoes Critic scores | Box Office Success or Failure |
| *Avengers: Endgame (2019)* | $400,000,000 | 4,662 | 233.4 | 94% | Success |
| *Joker (2019)* | $55,000,000 | 4,374 | 112.9 | 69% | Success |
| *Dark Phoenix (2019)* | $200,0000,000 | 3,721 | 41.6 | 23% | Failure |
| *Maleficent: Mistress of Evil (2019)* | $185,000,000 | 3,790 | 18.8 | 40% | Failure |
| *Zombieland: Double Tap (2019)* | $48,000,000 | 3,468 | 15.5 | 69% | Success |

Figure 1 is a collection of scatter plots and histograms of input variable compared to each other to find possible trends in the data for the program to train and test with. The input variables are (in ascending order) movie production budget, number of theaters shown, amount of YouTube trailer views before the opening weekend, and Rotten Tomatoes critic scores. The 4th column is the possible outputs of the opening weekend, which is either a success or failure. The term success or failure for the opening weekend relates to how much money was expected to be made and if the movie reached those expectations or not. \*(Go in procedures)

[caption]

\*Explain ranges I used for grid searches

From the grid search which was used on the program, the number of recommended hidden layers and nodes were shown through running multiple tests. The number of hidden layers recommended was 1 with 3 nodes within that layer.

\*Wording

\*Put some of this in procedures

\*Change this

\*\*Table maybe

Using a neural network, The best train and test score were then gained from cross validating the data. The train score resulted in a 84% accuracy from training with 60% of the data, while the test score resulted in a 90% accuracy using the remaining 40% of the data.

\*Lead to conclusion, don’t go all the way though.