

Please do the following steps:

1. Read the file and display columns.
2. Calculate some basic statistics from this data (count, mean, std, etc), examine it and state your observations on what you find (any inferences you can make are bonus points).
3. Select the columns that will be important in your opinion to predict “crew” size. Explain why you removed any columns
4. Use one-hot encoding for categorical features.
5. Create training and testing sets (use 60% of the data for the training and the rest for testing).
6. Build a machine learning model to predict the ‘crew’ size.
7. Calculate the Pearson correlation coefficient for the training set and testing data sets.
8. Describe hyper-parameters in your model and how you would change them to improve the performance of the model.
9. What is regularization? What is the regularization parameter in your model?
10. Plot regularization parameter value vs Pearson correlation for the test and training sets, and see whether your model has a bias problem or a variance problem.