Data Mining Final Project

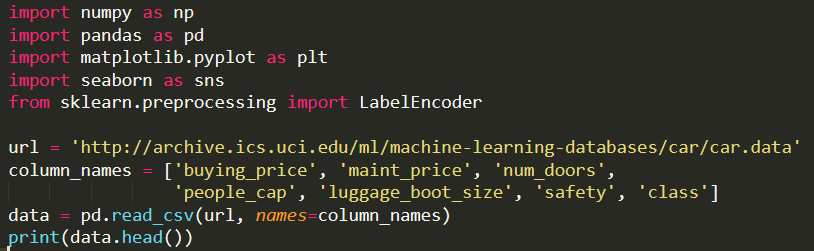
Name: Shang-Yung Hsu

ID:sh555

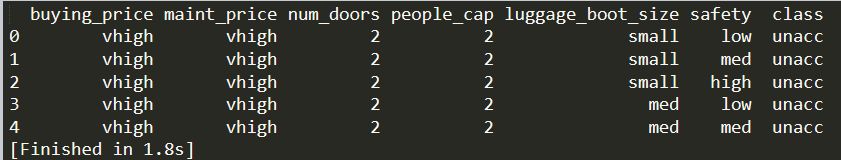
Check and Modify the Dataset

The car evaluation dataset is from <https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>.

I give the column name to each column.

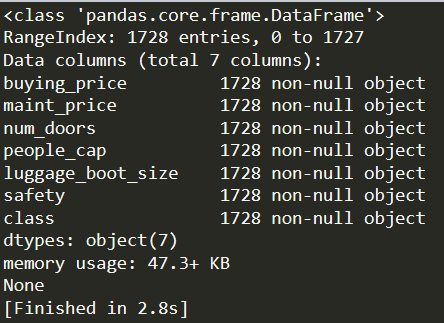




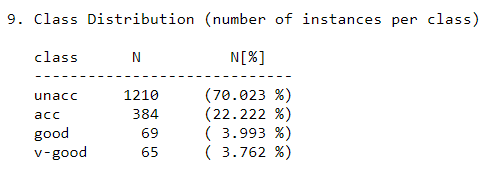


Check the dataset if there is any null value.

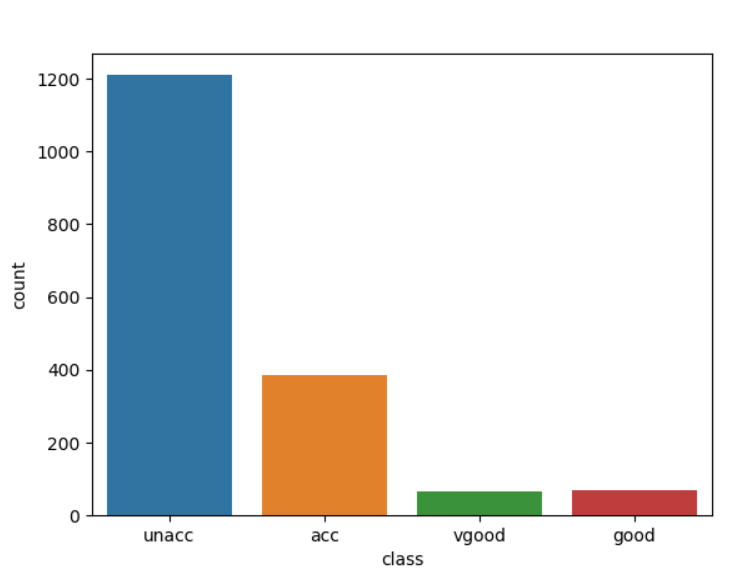




Seeing the description of the car evaluation dataset, you will find that the 'class ' distribution is really unbalance.

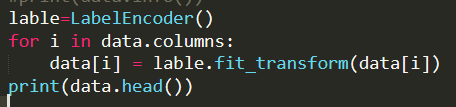


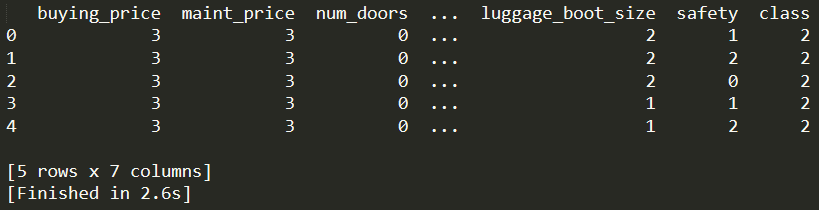




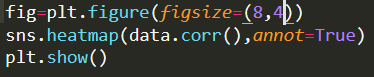
In this case, it will have an imbalanced multiclass classification problem. So that, the accuracy may not be the best stander to check this model.

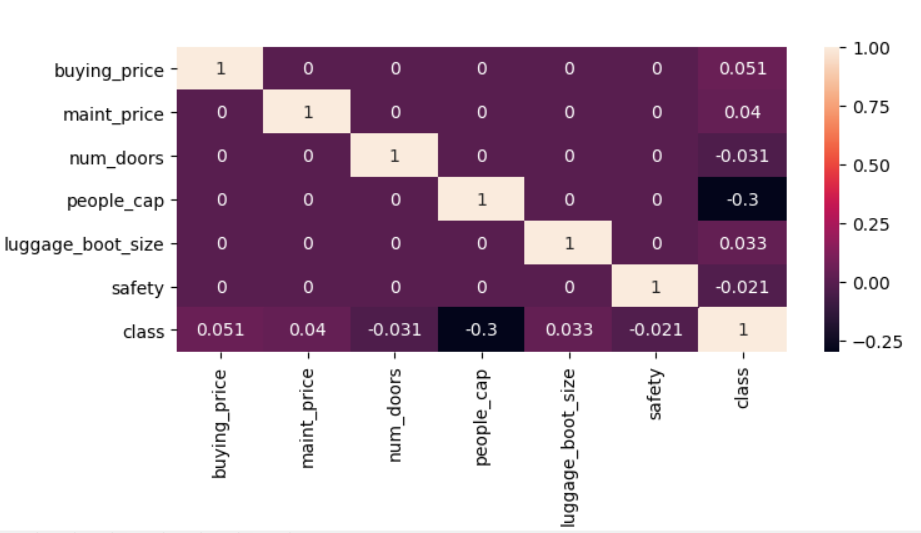
As scikit-learn algorithms do not work with string values, so I converted the string categories to integers.





Looking at the Heatmap of the columns on dataset with each other. Will find that most of the columns shows very weak correlation with 'class'. ('persons' column is the weakest). Other columns except 'class' shows no correlation with each other.





So, doing any analysis on them might not give any useful output.