PREDICTING WINE QUALITY

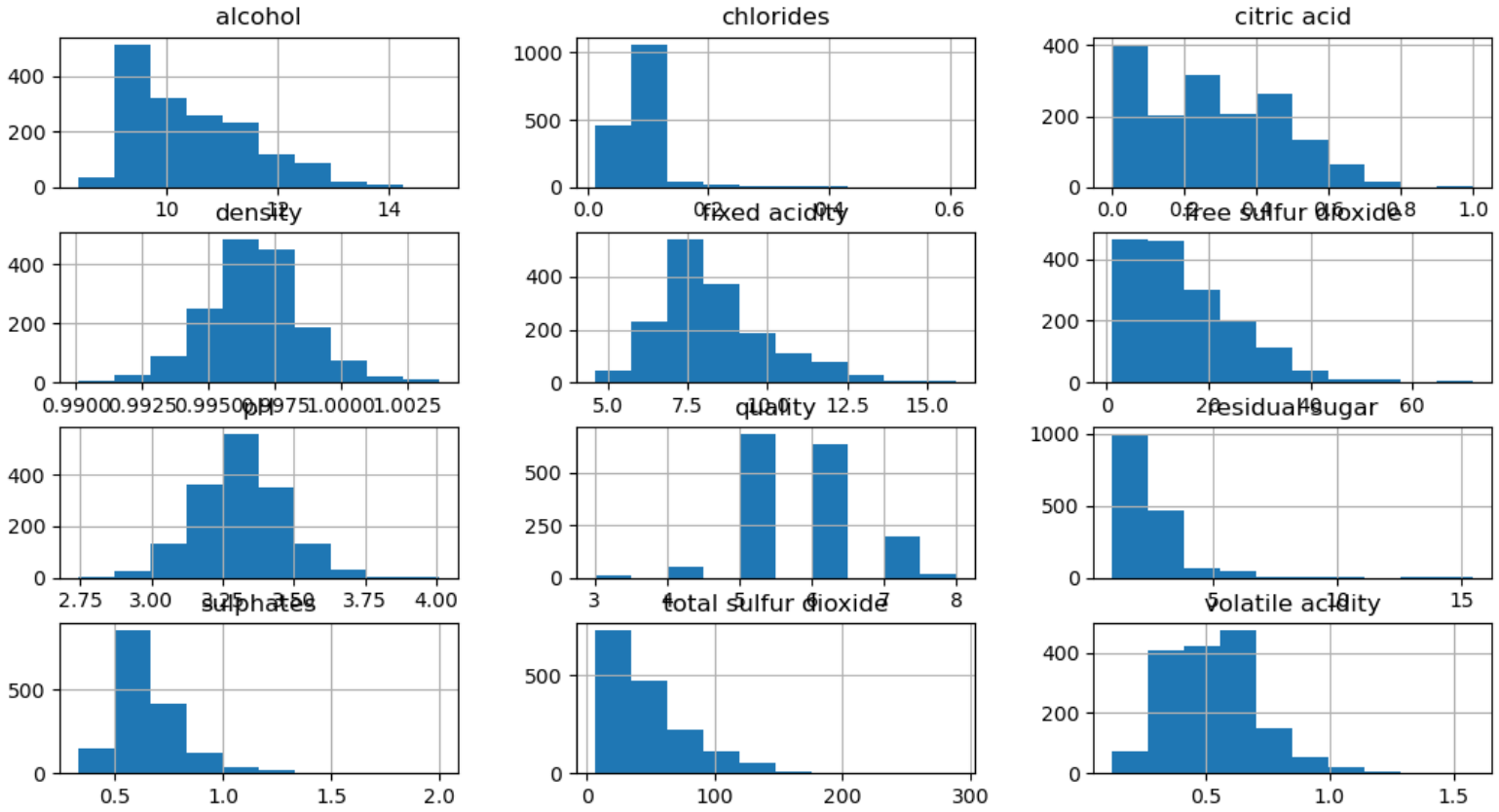
OBJECTIVE

- Our main objective here is to analyze how various features that are related to wine quality and predict quality of wine based on these features.
- Many wine websites/magazines can use this type of machine learning algorithms in giving out the best reviews.

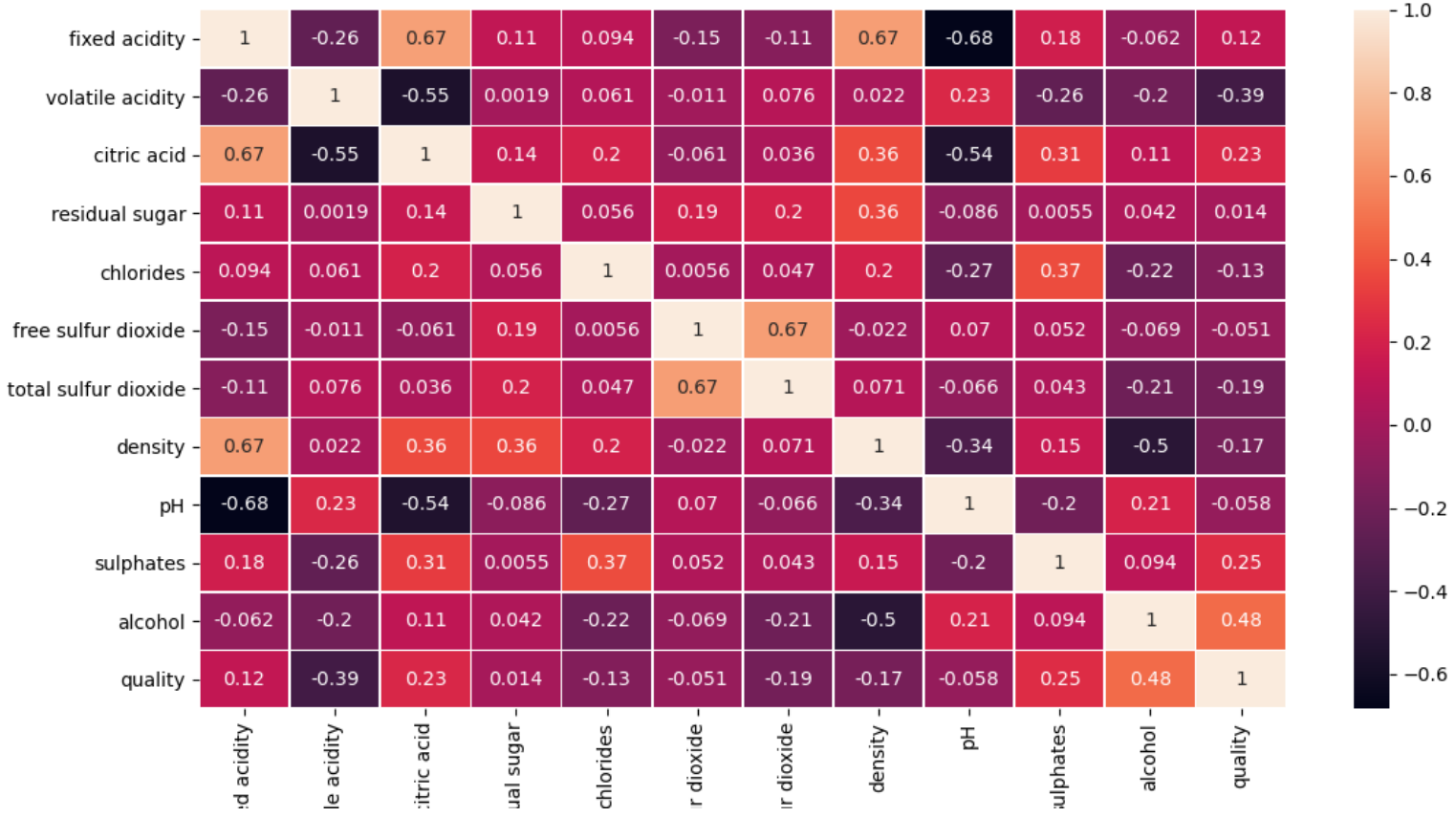
Exploratory Data Analysis

- Exploratory Data Analysis helps us to determine relationships between feature variables and target variables
- It helps us to determine various factors namely correlation, data skewness, imbalance data sets, etc.
- EDA plays an important role in feature engineering and model tuning.

Let us look at a few visualizations…

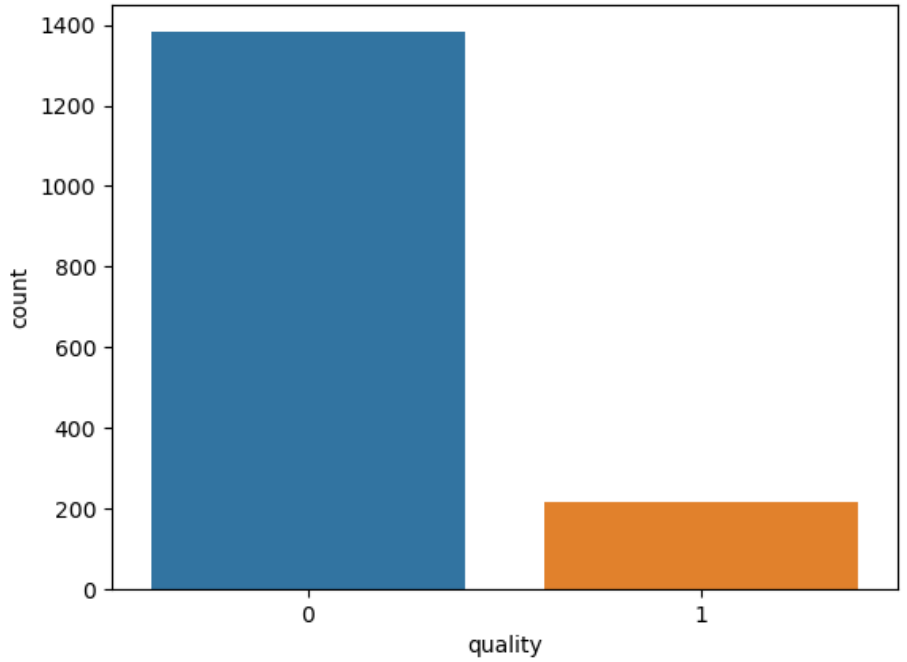


Let us look at a few visualizations…



*correlation matrix

Let us look at a few visualizations…



We can see our class 1 is much lesser than our class 0. This is an imbalance data set

Sampling Techniques

Upsample

- You randomly resample the minority class to create new data.

SMOTE

- You use the nearest neighbors of the minority observations to create new synthetic data

Downsample

- You remove some samples of the majority class

F1 Scores for each sampling

- The best F1 Score for Original data:

0.5601627894186877

- The best F1 Score for Upsampled data:

0.9719745813610599

- The best F1 Score for SMOTE data:

0.9318189772478351

- The best F1 Score for Downsampled data:

0.7971379360301338

Here upsampling gives us the best F-1 score

Logistic Regression

- Train our data for Logistic Regression model

Classification Report:

	precision	recall	f1-score	support
0	0.97	0.80	0.88	273
1	0.43	0.87	0.58	47
accuracy			0.81	320
macro avg	0.70	0.84	0.73	320
weighted avg	0.89	0.81	0.84	320

We have achieved 81% accuracy

Random Forest Classifier

- Train our data for Random Forest Classifier model

Classification Report

	precision	recall	f1-score	support
0	0.94	0.96	0.95	273
1	0.74	0.62	0.67	47
accuracy			0.91	320
macro avg	0.84	0.79	0.81	320
weighted avg	0.91	0.91	0.91	320

We achieved 91% accuracy

Conclusion

- As we can see here, Random Forest gives us the best of the two models with 91% accuracy.
- Implementing these type of machine learning models in wine reviewing magazines/website will help reduce manual effort in giving out the best reviews.
- These type of ml algorithms can also be implemented in spotting out faulty wines when mass producing.