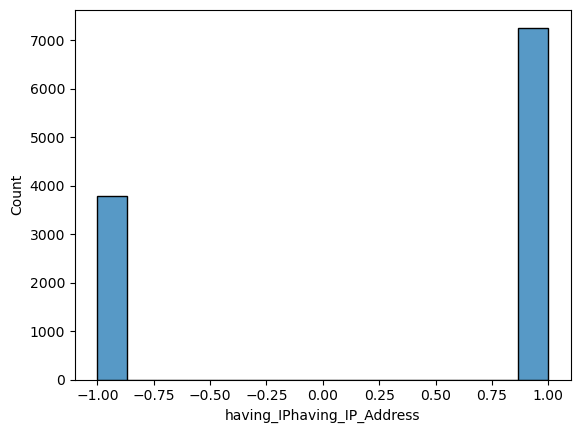
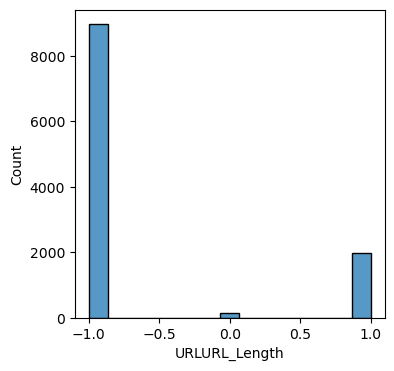
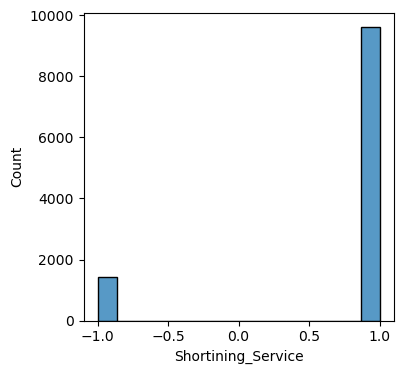
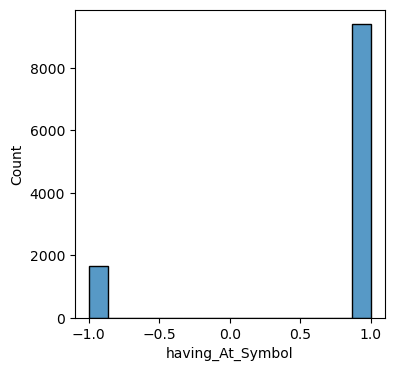
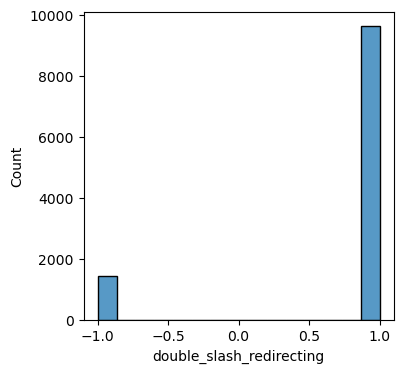
1. Each sample has 32 features ranging from -1,0,1. Explore the data using histogram, heatmaps.

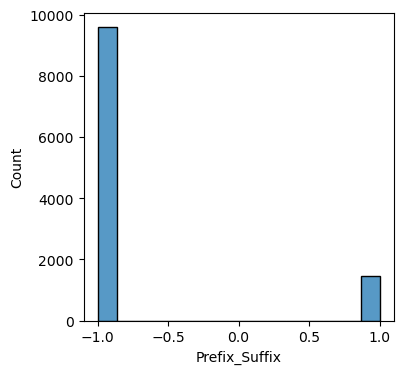


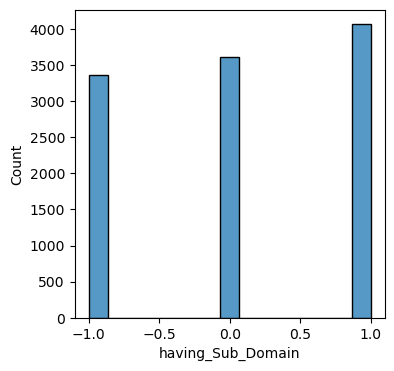


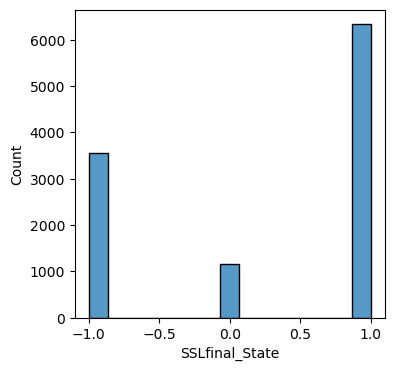


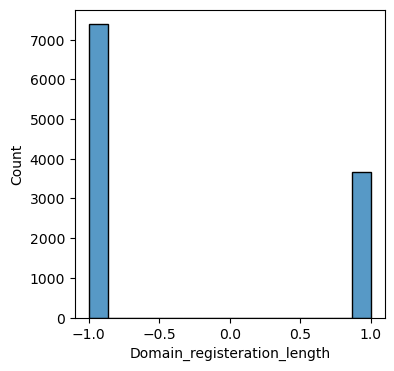


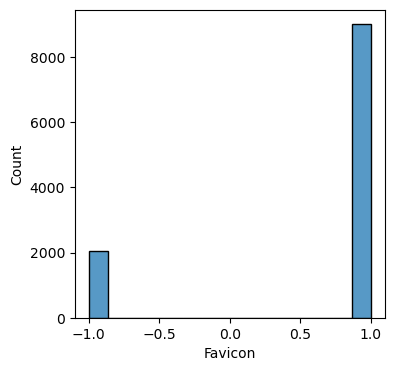


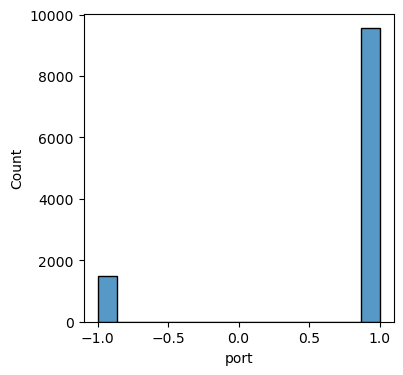


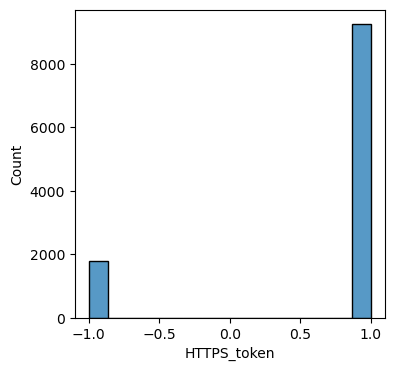


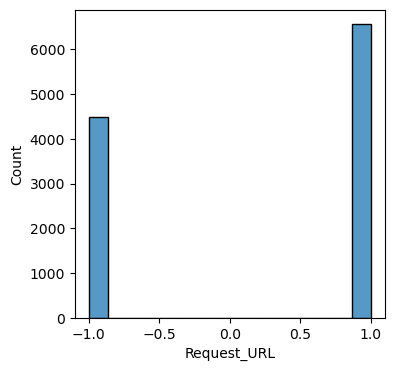


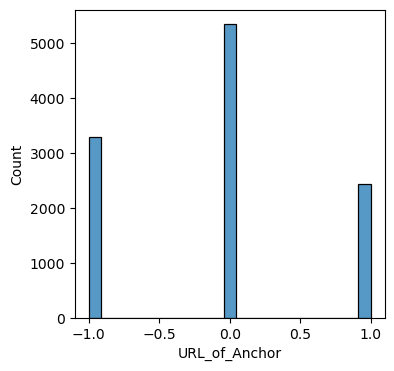


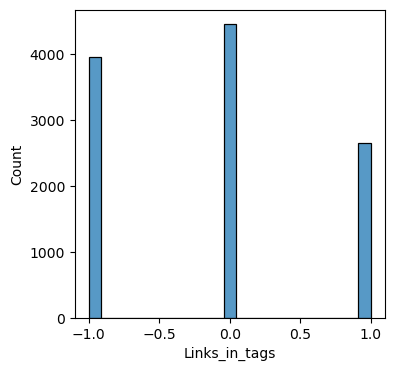


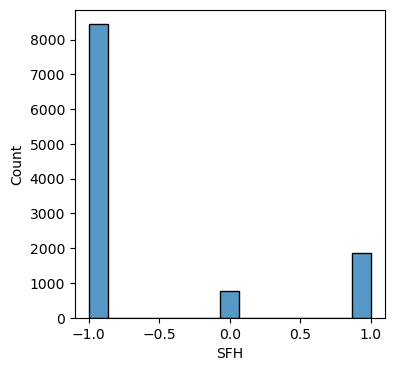


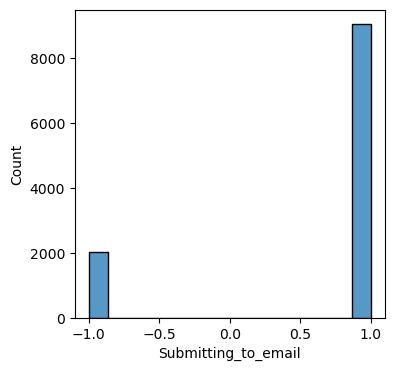


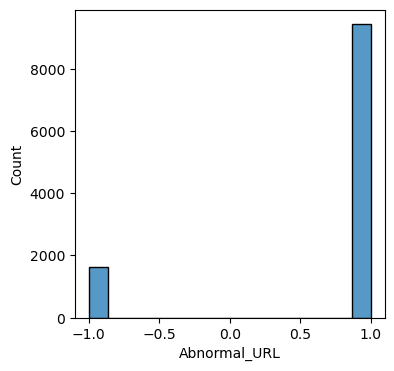


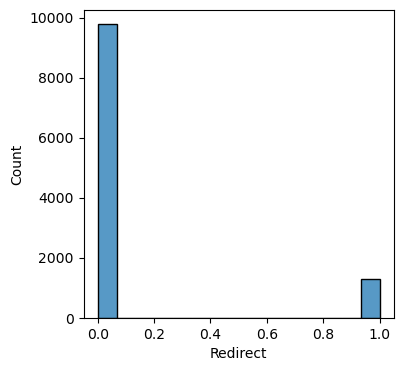


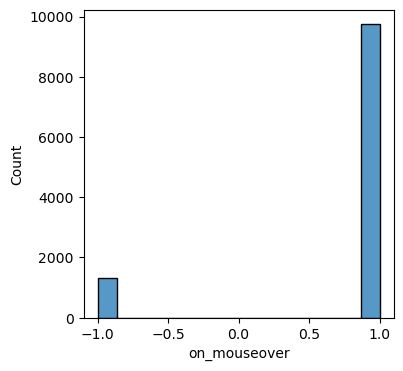


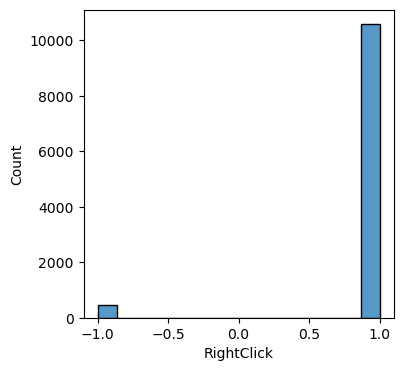


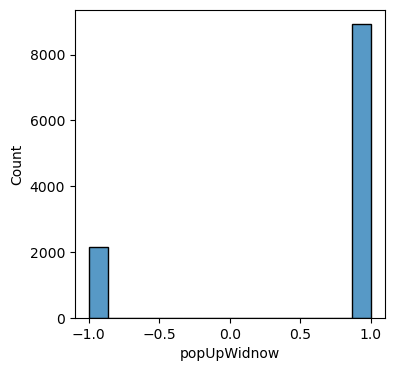


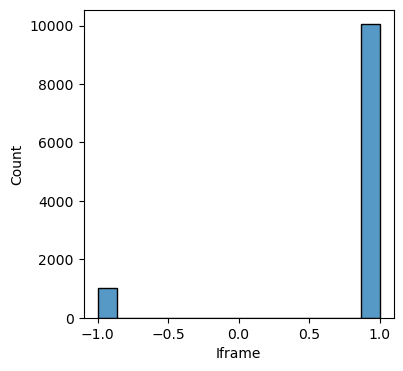


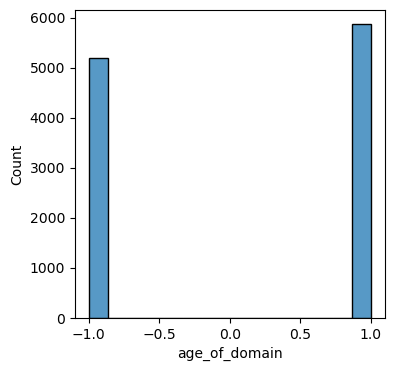


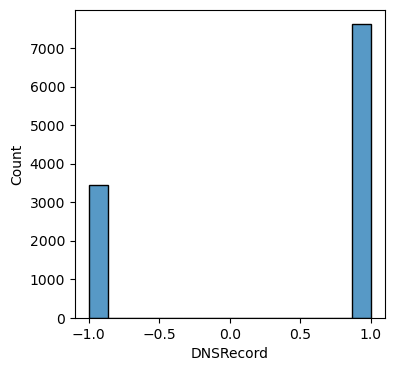


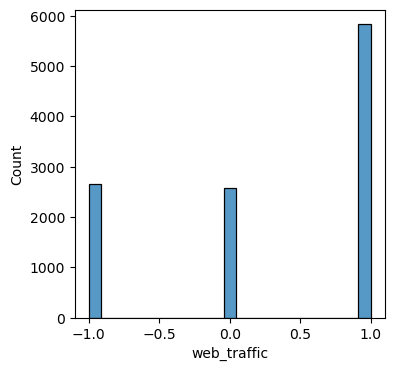


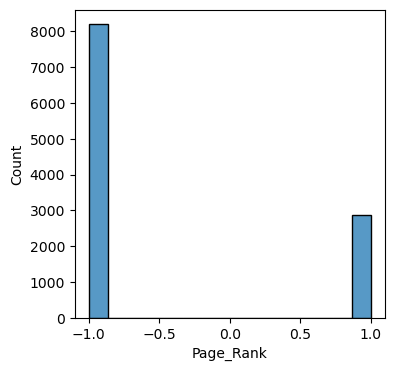


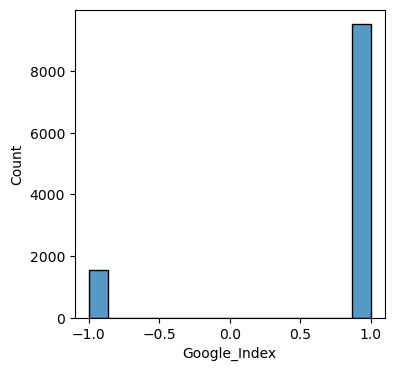


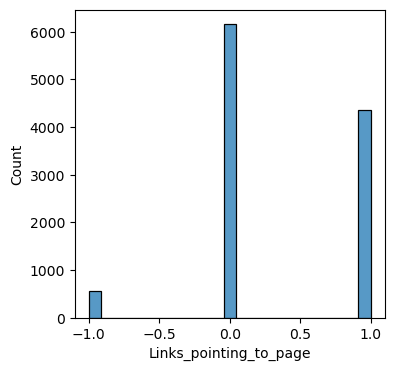


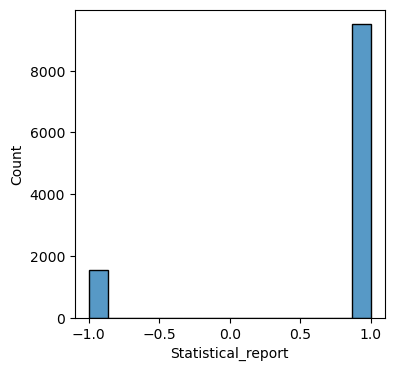


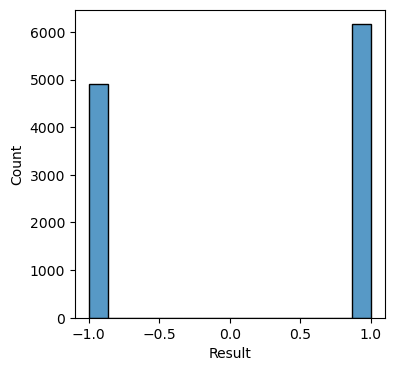


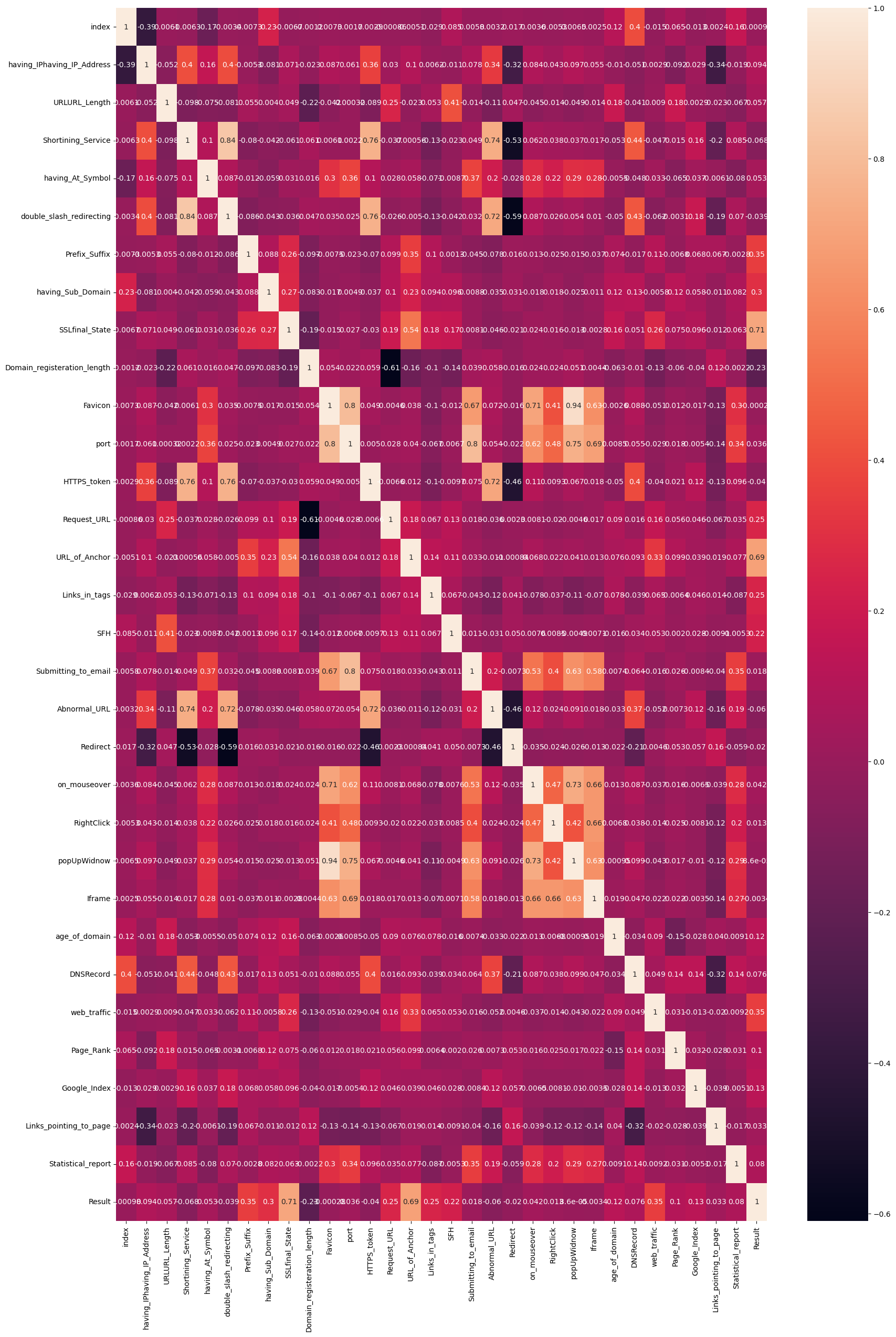


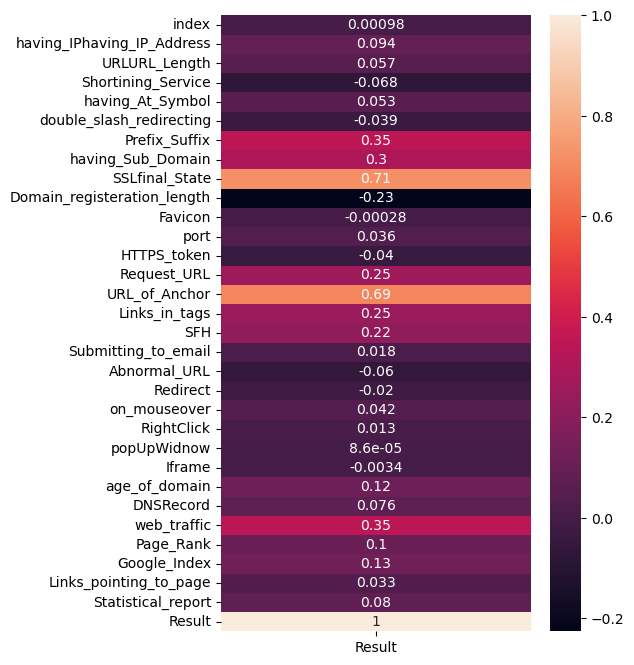




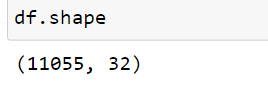


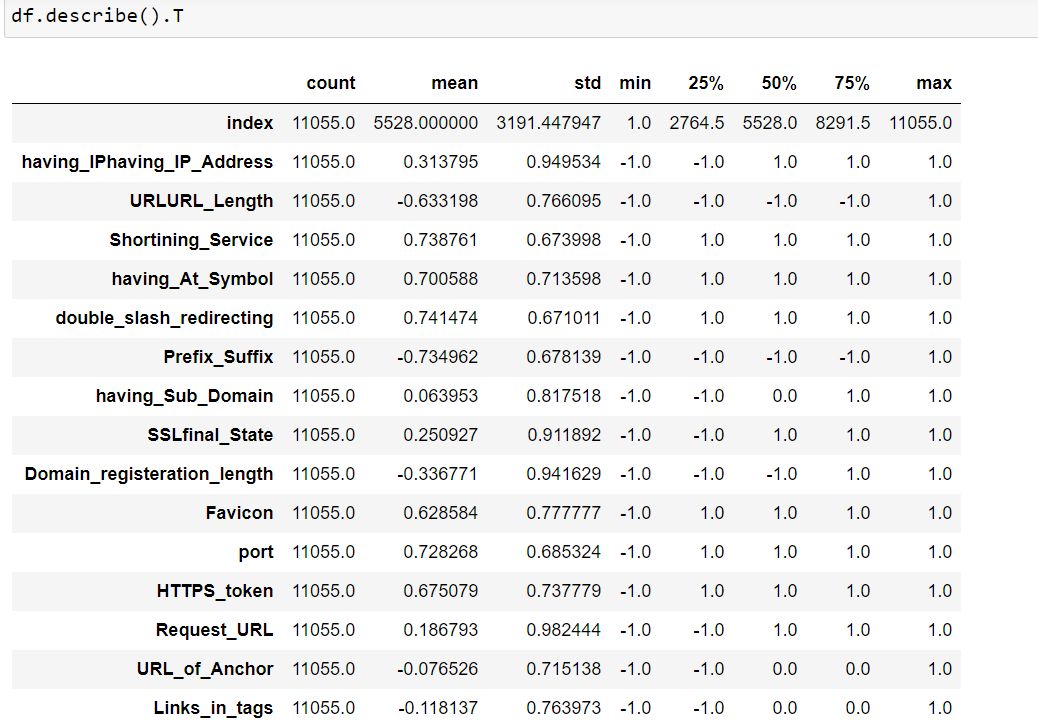


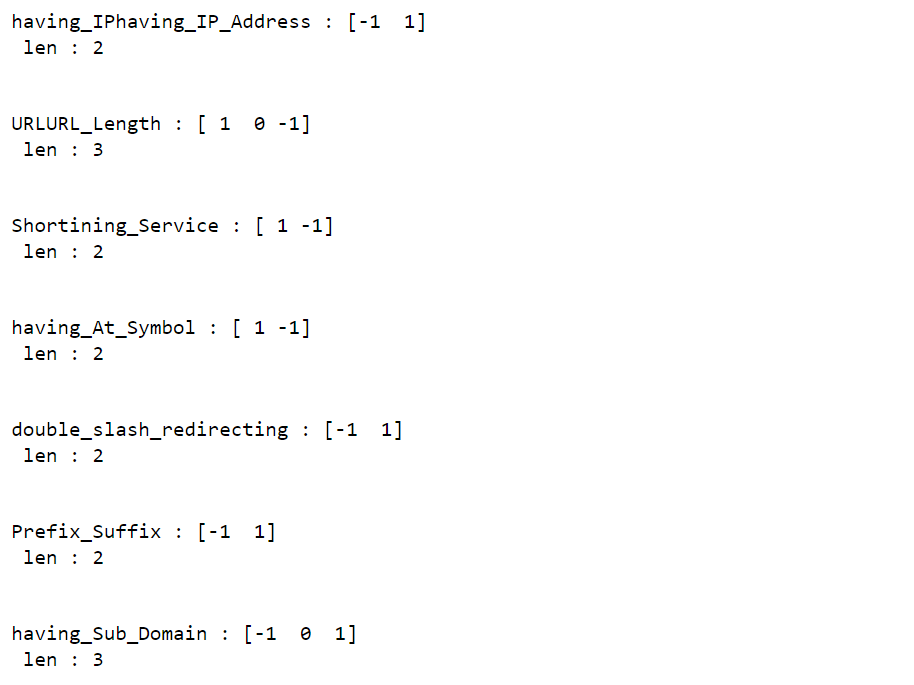


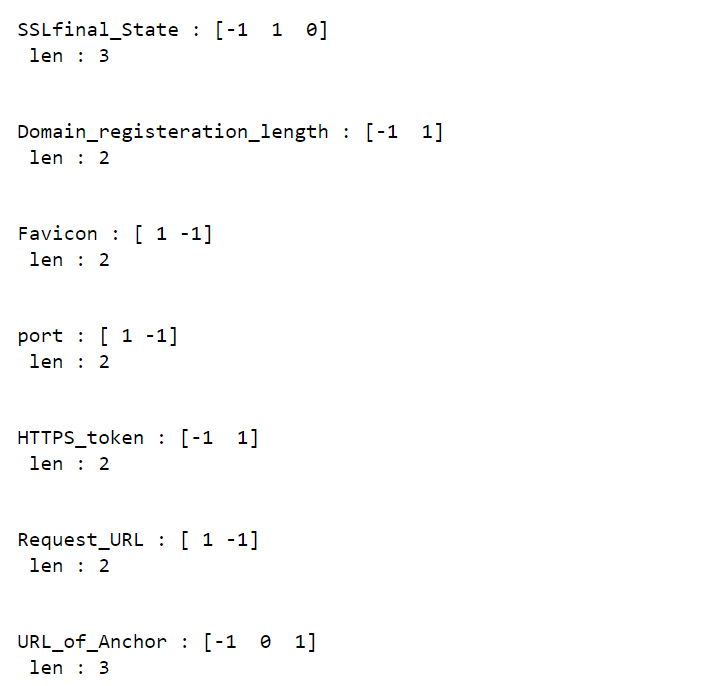


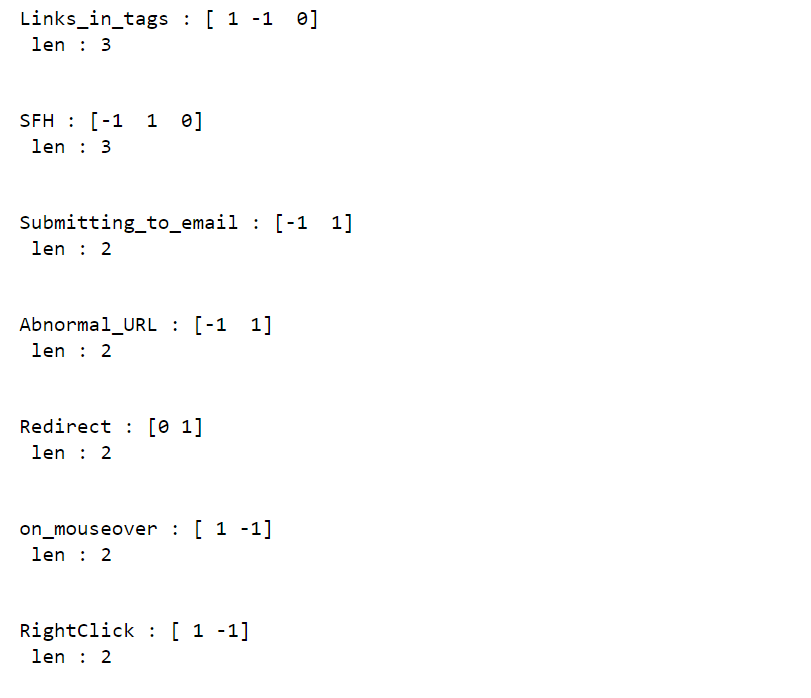
1. Determine the number of samples present in the data, unique elements in all the features.



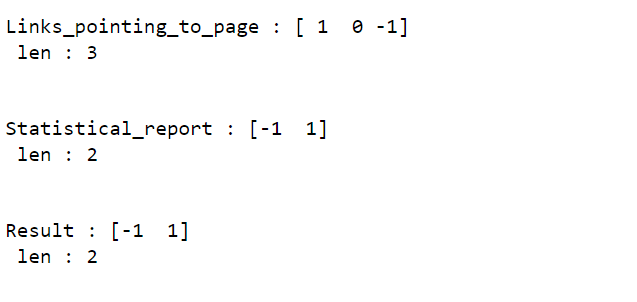


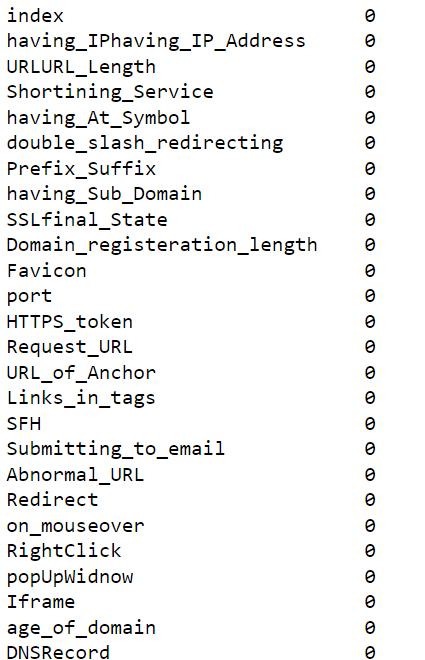


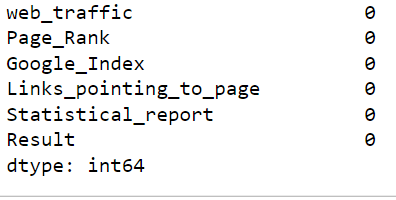




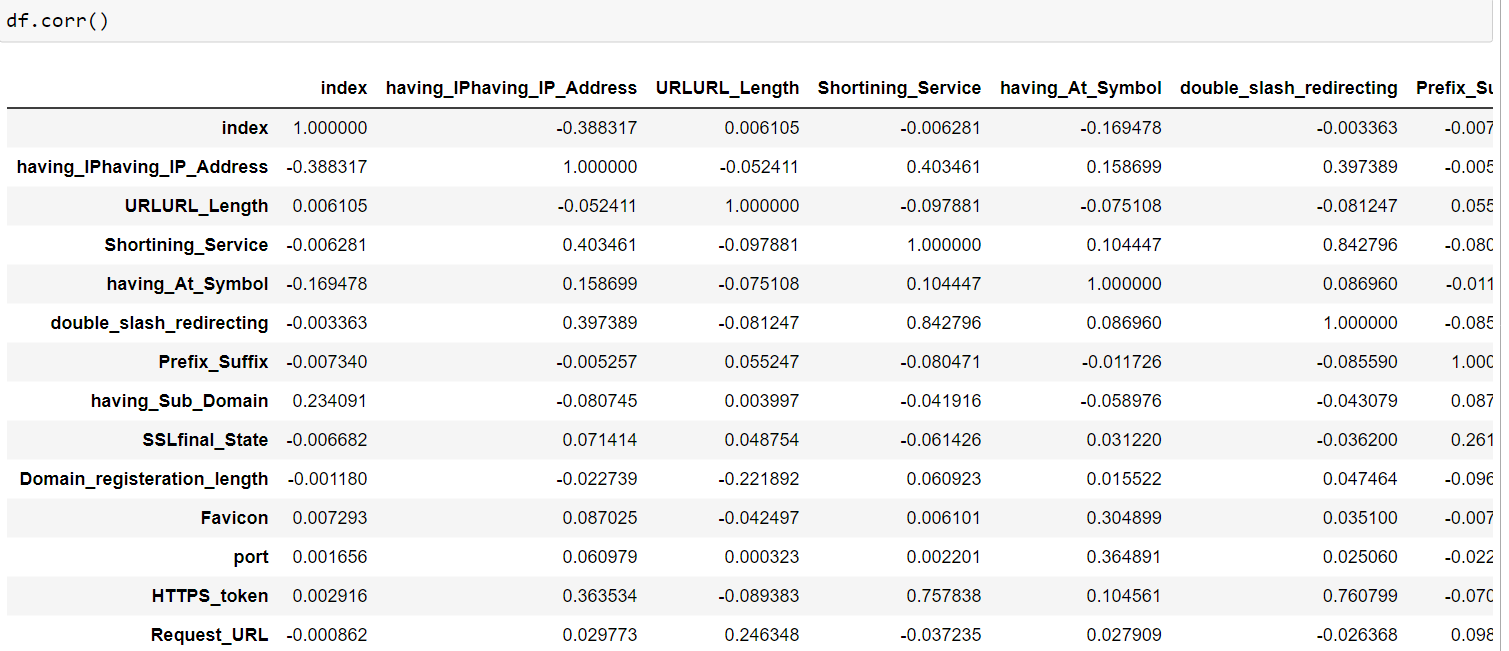


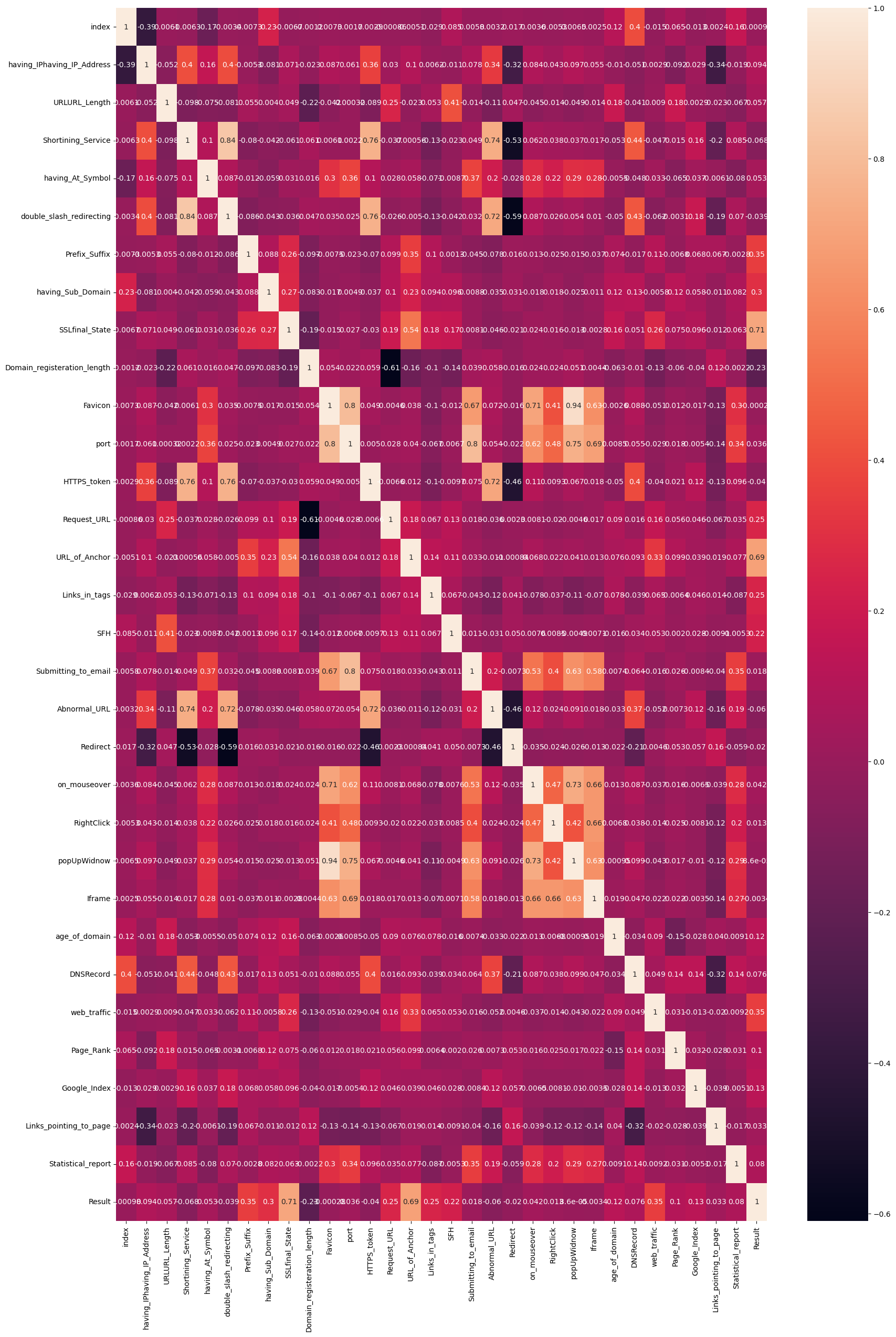
3.Check if there is any null value in any features.

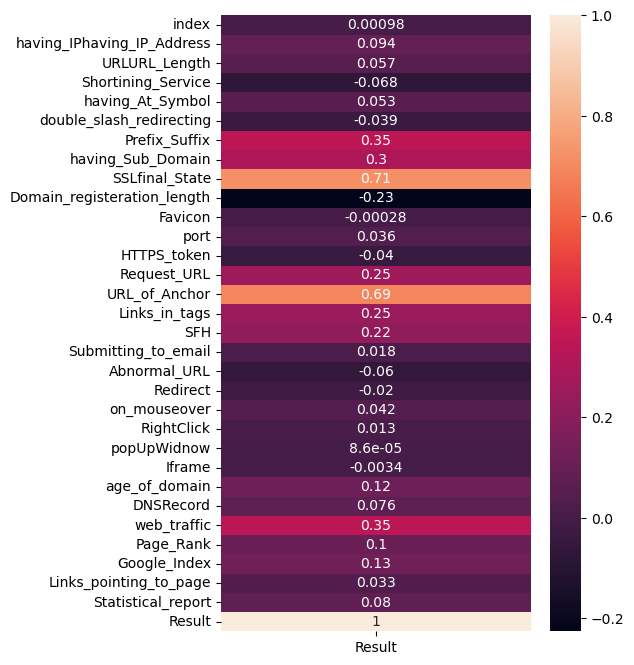




4.Next, we have to find if there are any correlated features present in the data. Remove the feature which might be correlated with some threshold.

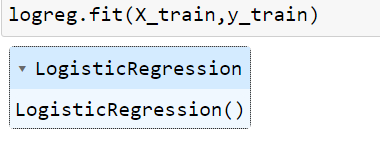


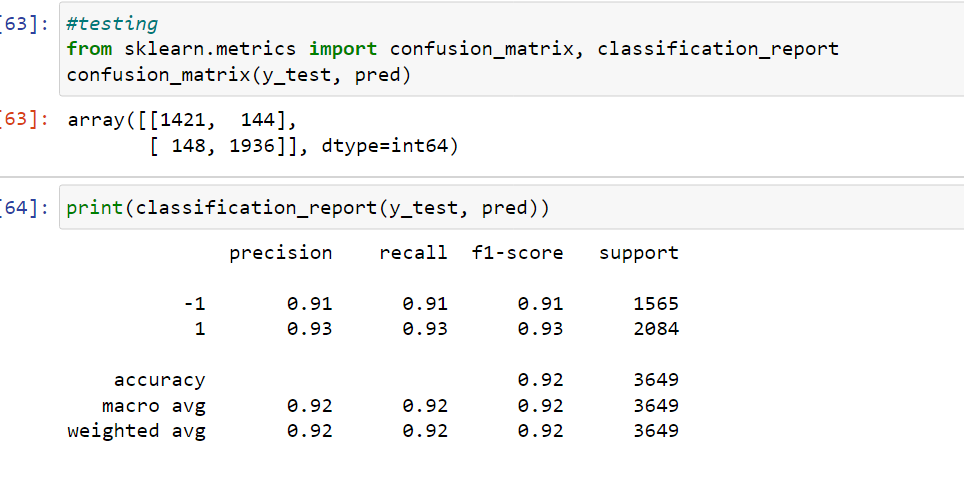




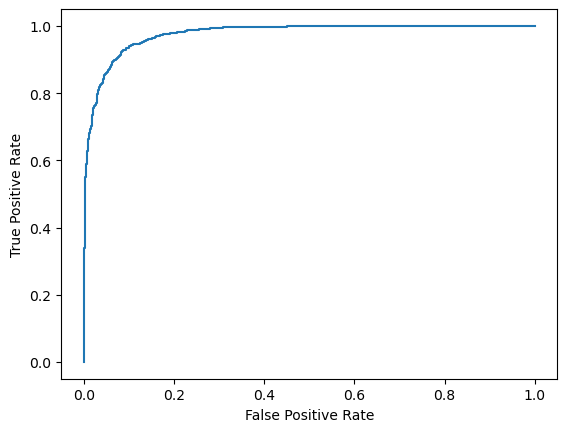
5.Finally, build a robust classification system that classifies whether the URL sample is a phishing site or not.

* Build classification models using a binary classifier to detect malicious or phishing URLs.

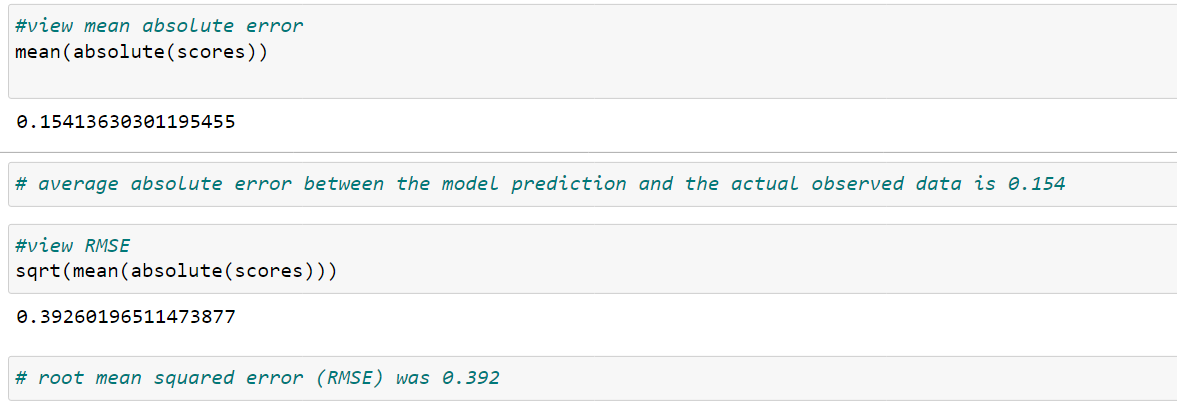




* Illustrate the diagnostic ability of this binary classifier by plotting the ROC curve.



* Validate the accuracy of data by the K-Fold cross-validation technique.



* The final output consists of the model, which will give maximum accuracy on the validation dataset with selected attributes.

