Implementing Machine Learning (ML) or Deep Learning (DL) models without a predefined Exploratory Data Analysis (EDA), Feature selection and engineering is not the best strategy to implement. EDA enables data scientists to get a preliminary understanding of the data's structure, anomalies, and patterns. By examining data distributions, relationships, and variances, EDA helps uncover hidden insights that might not be immediately apparent.Feature selection is the process of selecting the most relevant features from the original features set to be used in model building, by removing the redundant, irrelevant, or noisy features.Wrapper techniques are techniques assess subsets of features based on their performance with the selected machine learning model. By iteratively adding or removing features aiming to find the combination of features that leads to the best model performance.

Let's dive into what has been done, First DataPreprocessing and EDA, the dataset is explored for the columns values, dataFrame shape, duplicates, null values and outliers. Column values are all numerical of type float64 except the feature "Class" is numerical of type int64, which means we don't need any of one hot encoding or labeler. The DataFrame is of shape (284807, 31) the 31 columns are splitted into v1..v28 which are PCA features due to data confidentiality and the rest are "Time", "Amount" and the target columns "Class", After investigating the duplicated rows they are 1081 rows, after removing them the DataFrame became of shape 283726 rows × 31 columns. Null values are checked and resulted in no Null Values. Outliers, due to our data distribution outliers are the instances from the minority class with 492 outliers, resulting in the fact that we cannot remove the outliers. In addition, using a data augmentation tool like "smot" to rebalance the data.

[NoteBook Link]

EDA are done using profilers, "ydata_profiling" and "sweetviz profiler" showing very insightful graphs exploring the data and features relations. YData Profiling stands as a powerful tool designed to simplify and enhance the data exploration experience, catering specifically to the needs of statisticians and data scientists.

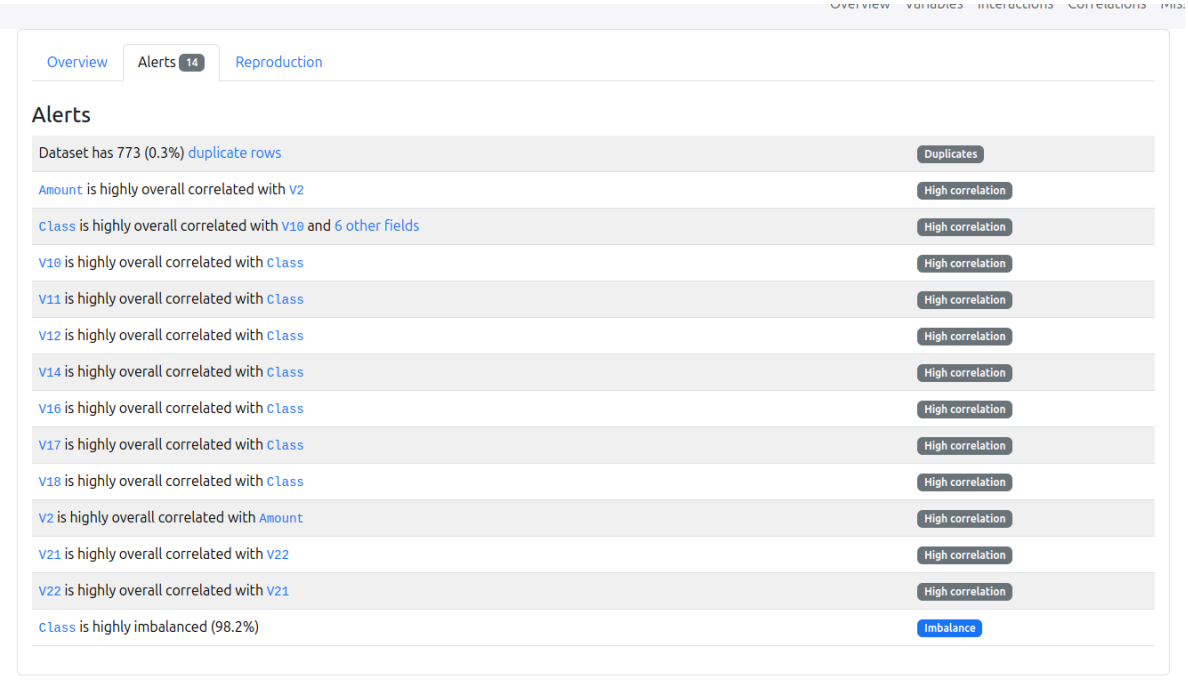


Figure [1]

Figure [1] represents a snap from the YData profilers, puts an alert on the imbalance data and the highly correlated features. Also applying univariate/ multivariate analysis on the whole data resulting in a dynamic feature to interpret the data and its distribution. In addition to normal EDA without using profilers, applying univariate analysis, Correlation Coefficient with the target and Anova test. Anova test is a statistical test used to analyze the difference between the means of more than two groups.

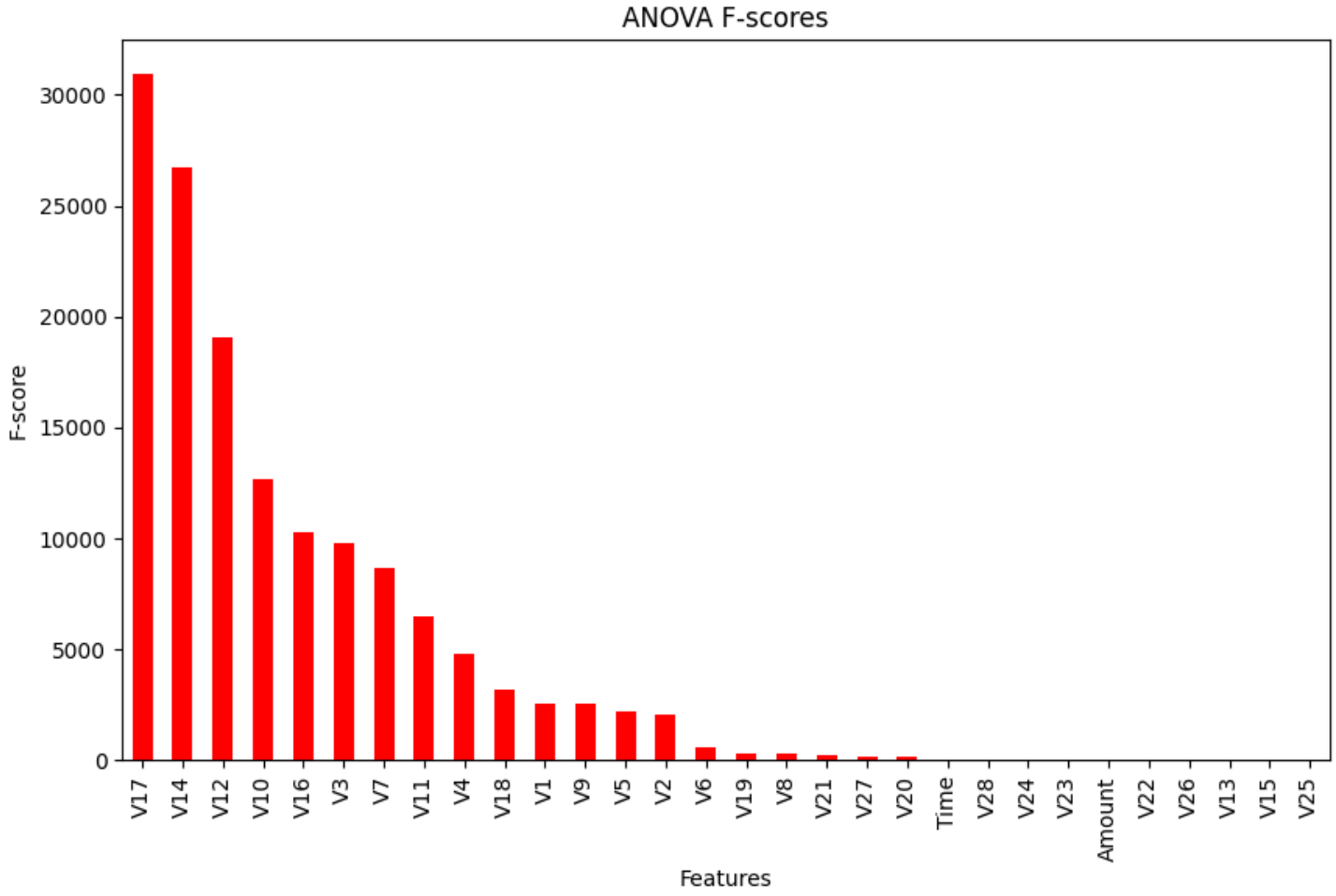


Figure [2]

Higher F-score means Big difference between class means thus more predictive power, More Relevant Feature such as V17,V14 and V12 while on contrast lower F-score means Weak or No Relevance Features like V23, V24, V25, V13, and V15 have very low F-scores, suggesting they don't contribute much to the target prediction.

A correlation coefficient is a number between -1 and 1 that tells you the strength and direction of a relationship between variables

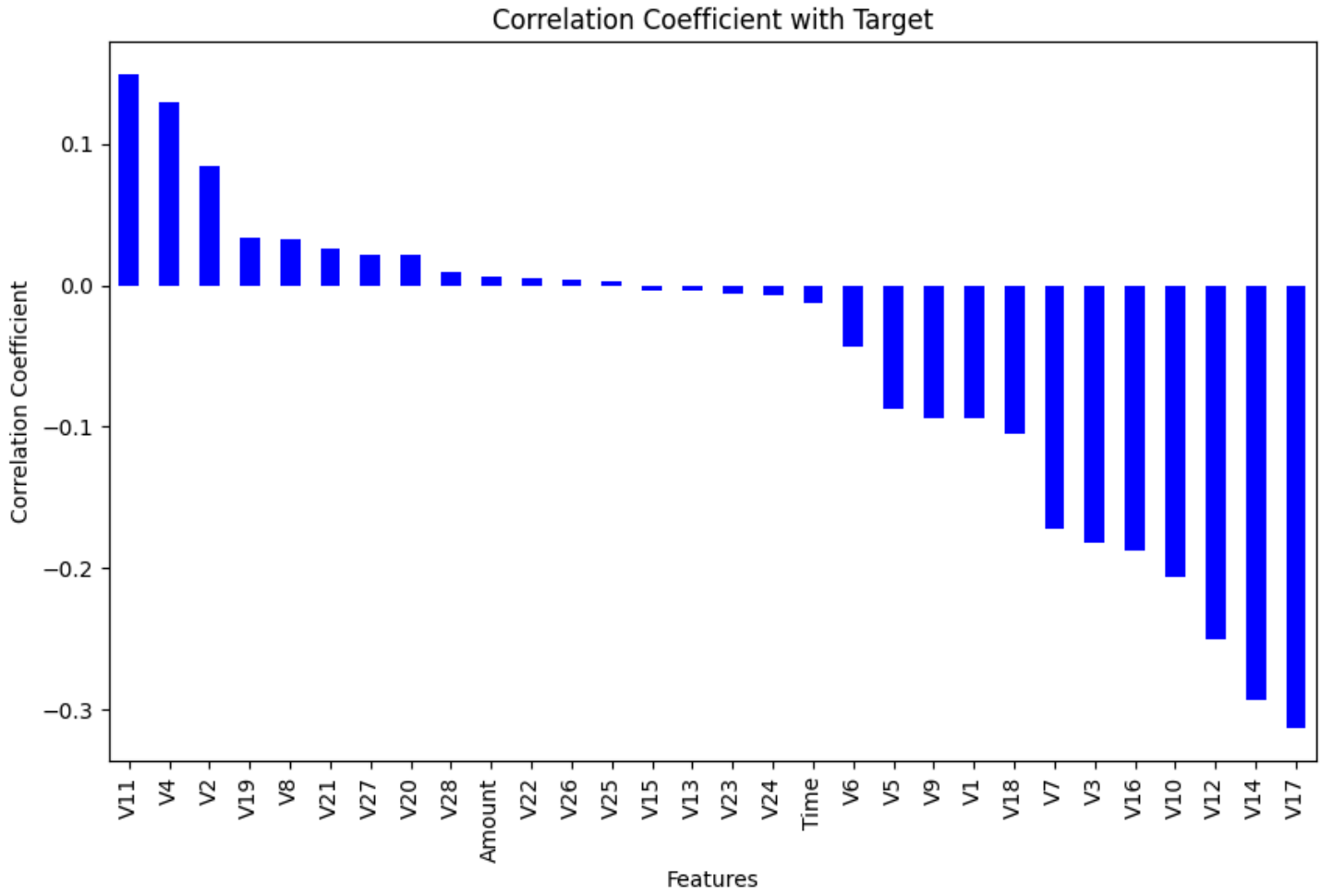


Figure [3]

It can be seen from the figure [3] the correlation coefficients, as mostly negative coefficients like v17,14,12 are more related with the positive class of the target feature, on the other hand V11, V4, V2 show positive correlations.

Comparing ANOVA test with Correlation Coefficients Features like V17, V14, and V12 showed high F-scores in the ANOVA plot and also showed strong negative correlation here. Confirms they're very informative for predicting the target.

Feature selection is a critical step in the data preprocessing phase of machine learning. It involves selecting a subset of relevant features (variables, predictors) for use in model construction. Effective feature selection improves model performance, reduces overfitting, and enhances interpretability. Wrapper techniques are used to evaluate the select features, using forward selection and backward elimination, using an logistic regression
model the forward selection choose the following 10 features :
Forward Feature Selection Features: ['V2', 'V3', 'V4', 'V9', 'V10', 'V11', 'V12', 'V14', 'V16', 'V17']

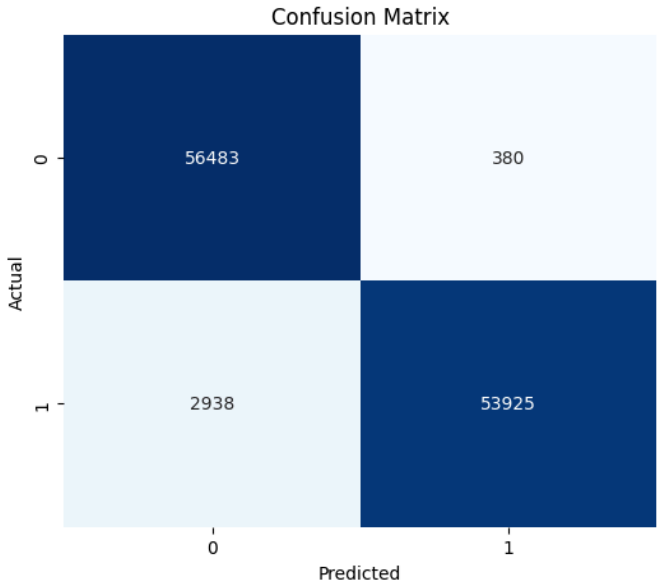
```
Precision: 0.99
Recall: 0.95
F1 Score: 0.97

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97     56863
           1       0.99      0.95      0.97     56863

    accuracy                           0.97    113726
   macro avg       0.97      0.97      0.97    113726
weighted avg       0.97      0.97      0.97    113726
```

	Feature	Importance
7	V14	0.370497
4	V10	0.200511
6	V12	0.138685
9	V17	0.116796
2	V4	0.105681
5	V11	0.027514
1	V3	0.017889
8	V16	0.015950
3	V9	0.004244
0	V2	0.002233



Results in 97% accuracy and accepted confusion matrix. The top 10 features are chosen by the FR model, also the RF trees are shown in the code, explaining each decision and how it was taken based on a simple visual to trace the trees and voting for the predicted class.

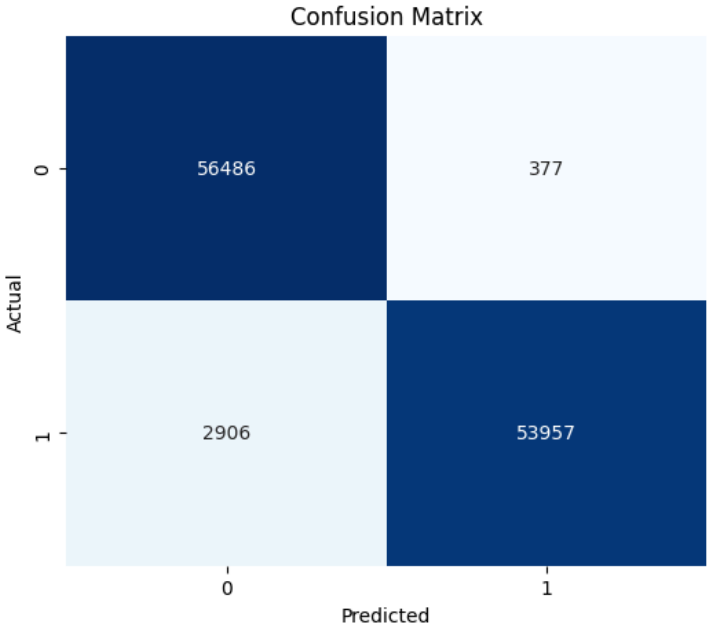
On the other hand Backward elimination technique choose the following features :
Recursive Feature Elimination Features: ['V4', 'V10', 'V11', 'V12', 'V13', 'V14', 'V16', 'V17', 'V27', 'V28']

```
Precision: 0.99
Recall: 0.95
F1 Score: 0.97

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97     56863
           1       0.99      0.95      0.97     56863

    accuracy                           0.97    113726
   macro avg       0.97      0.97      0.97    113726
weighted avg       0.97      0.97      0.97    113726
```



	Feature	Importance
5	V14	0.338776
1	V10	0.198203
3	V12	0.161231
7	V17	0.119362
0	V4	0.103578
2	V11	0.045706
6	V16	0.023222
8	V27	0.003949
9	V28	0.003502
4	V13	0.002471

Results in 97% accuracy and accepted confusion matrix. The top 10 features are chosen by the FR model, also the RF trees are shown in the code, explaining each decision and how it was taken based on a simple visual to trace the trees and voting for the predicted class.

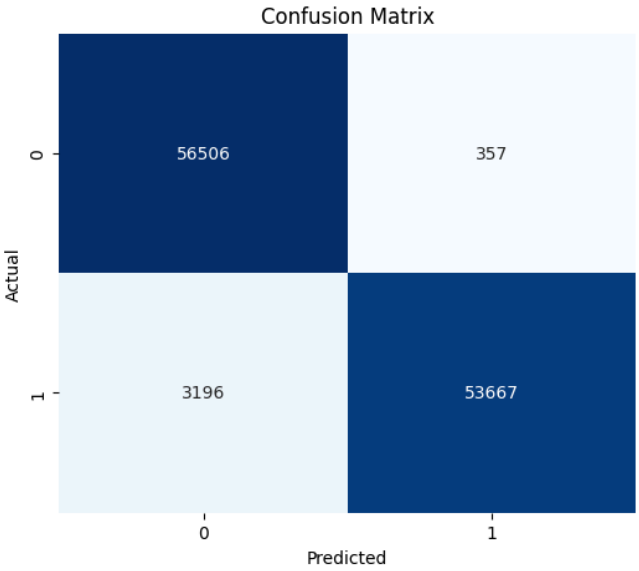
And an employed Random forest to choose the feature from model choose the following features :
Selected Features by RF + SelectFromModel: ['V1', 'V2', 'V3', 'V4', 'V6', 'V7', 'V9', 'V10', 'V11', 'V12', 'V14', 'V16', 'V17', 'V18', 'V21']

```
Precision: 0.99
Recall: 0.94
F1 Score: 0.97

Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.99      0.97     56863
           1       0.99      0.94      0.97     56863

    accuracy                           0.97    113726
   macro avg       0.97      0.97      0.97    113726
weighted avg       0.97      0.97      0.97    113726
```



	Feature	Importance
10	V14	0.271480
9	V12	0.153779
7	V10	0.142381
12	V17	0.108982
3	V4	0.095568
8	V11	0.068336
2	V3	0.059594
5	V7	0.027720
1	V2	0.025991
11	V16	0.023777

Results in 97% accuracy and accepted confusion matrix. The top 10 features are chosen by the FR model, also the RF trees are shown in the code, explaining each decision and how it was taken based on a simple visual to trace the trees and voting for the predicted class.

In comparison to the three feature selection techniques there is no significant difference in accuracy between them but it can be seen that the feature importance intersects for the three features selection technique, explaining the feature selection importance. Concluding that the three feature selection techniques performed very well.