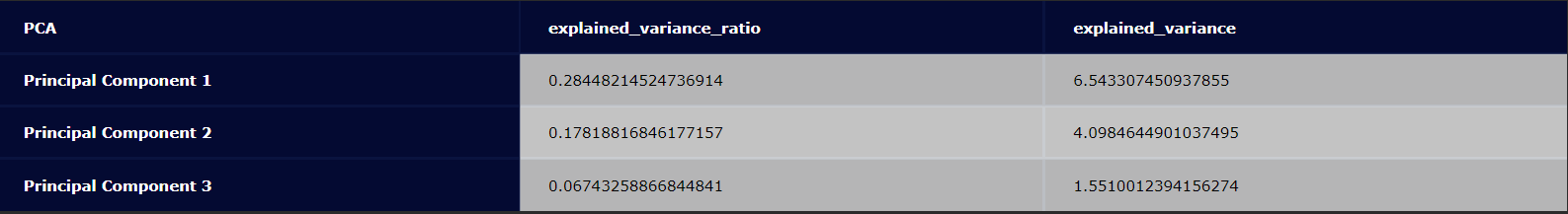
PCA is used to reduce high-dimensional data to a lower-dimensional space in a way that preserves variability.

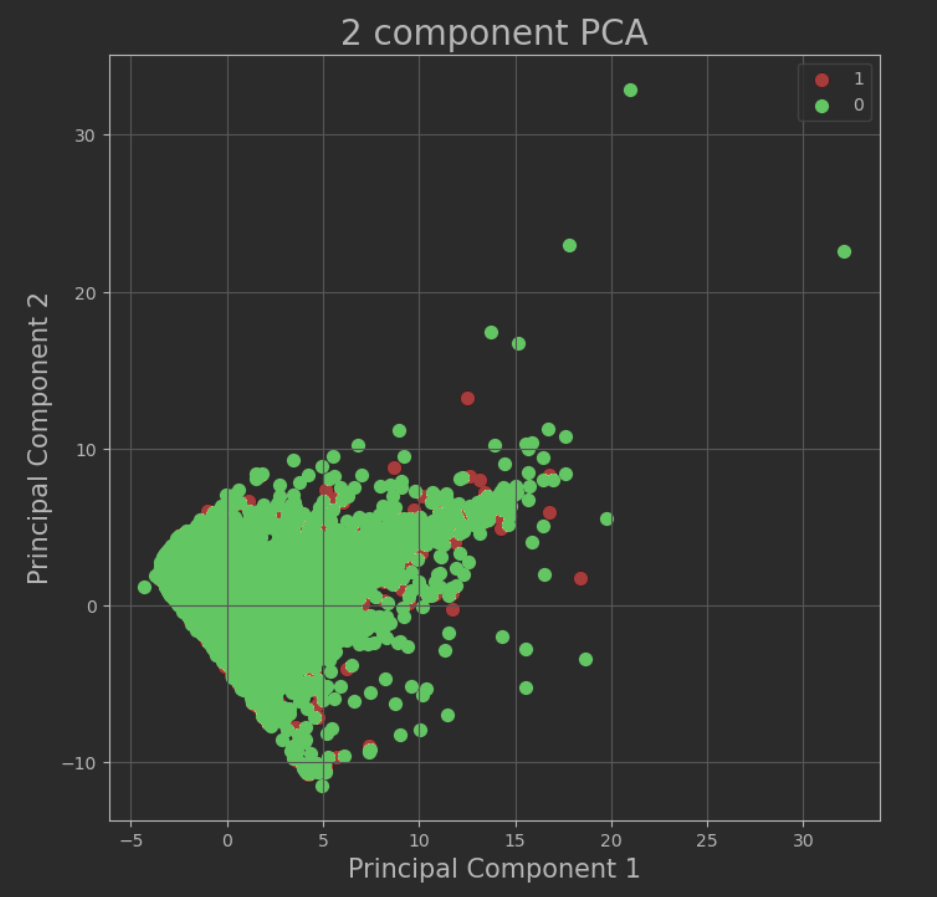
The credit card dataset has 24 columns, the final “Default payment next month“ column is the target variable and the remaining are independent variables.

Removing the ‘default.payment.next.month’ coumn, as we need to do PCA only for independent variables, so now we have 23 columns as features.

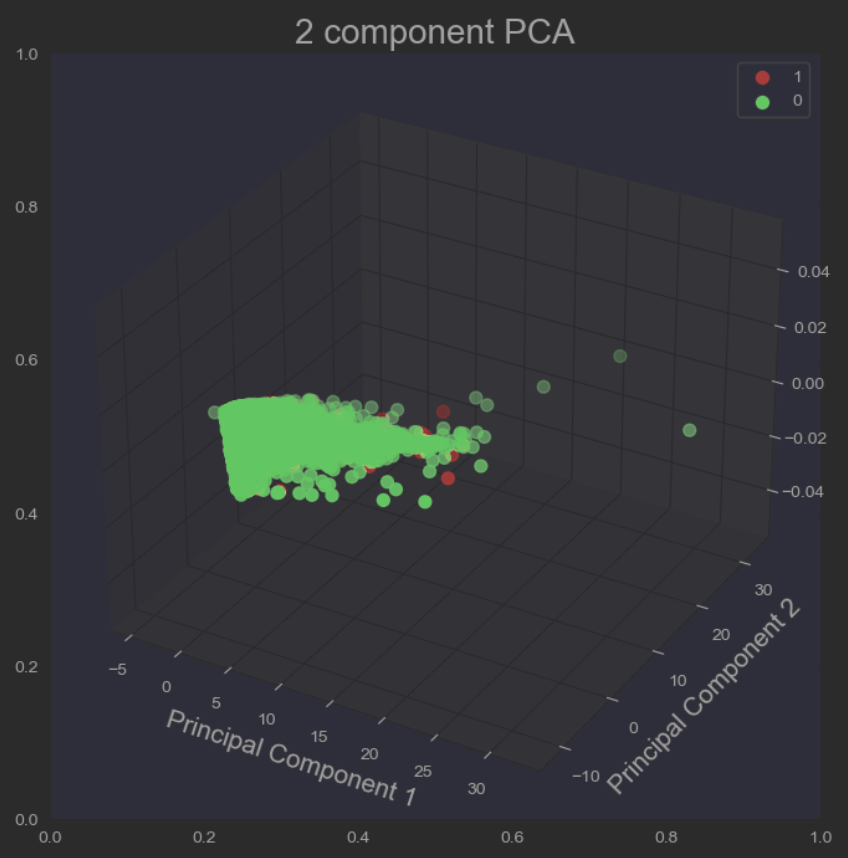
The Python library used for PCA analysis is scikit-learn. For our visualization projection we are going to use 2D and 3D plots, so we need two principal components for 2D and three principal components for 3D projections



VISUALIZE 2D PROJECTION

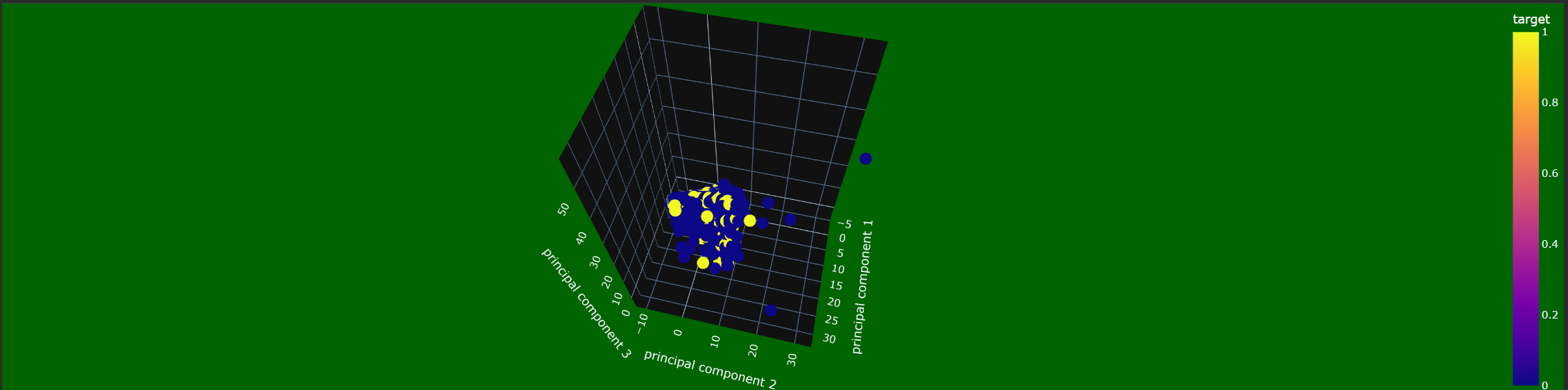


From the above plot we can see the 0 values of the target are predominant data, the Principal Components 1 and 2 explain just under 50 % the variance, for reference have a look at the above component-variance table and last Scree plots



The above plots show two principal components in 3D projection, we can also see the 0 values of the target column are predominant those two principal components explain just under 50%

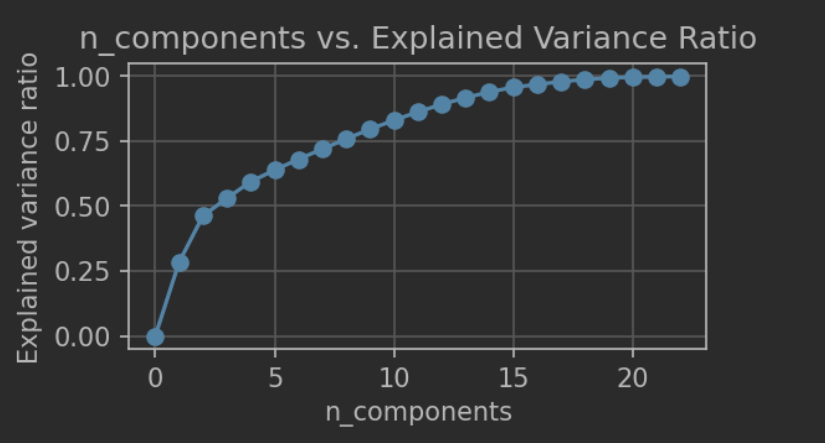
VISUALIZE 3D PROJECTION



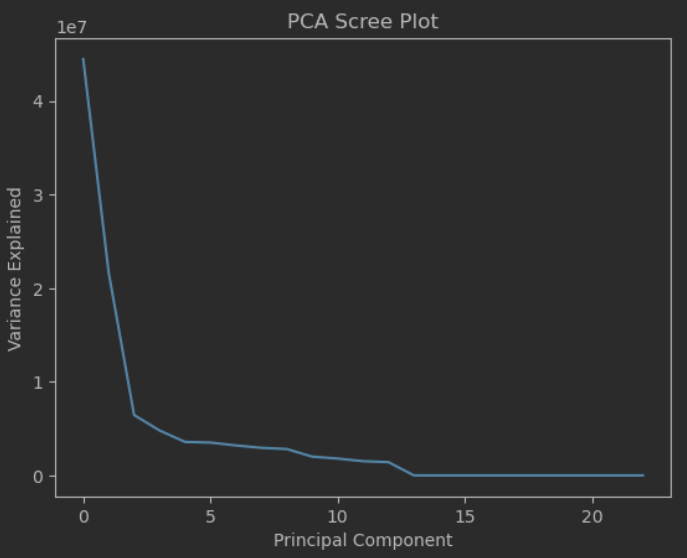
The above plot shows 3D projections of the three Principal Components 1,2 and 3 to the explained variance, the data is very clustered and is difficult to distinguish the predominant data of the target columns. Those Principal components 1,2 and 3 explain the variance just above 50% , for reference have a look at the above principal-variance table the bottom Scree Plots for all Principal Components projection.

Projection of All Principal Components to the explained variance, ordered by used python libraries.

By Scikit-Learn (PCA)



By SciPy (SVD)



Conclusion if we want to explain the variance at 80% and above, we need to 10-13 Principal Components factors.