Forging banknotes is a high level dilemma that needs to be addressed by central banks since no cashier, retailer or citizen actively investigates the authenticity of notes each banknote held at hand. The level of trust between citizens and banknotes is high which makes it a security gap, easily breached by scammers. Consequently, central banks need to maintain solid security measures when it comes to producing bank notes, in order to allow the economy to flourish without being hindered by fake notes that reduce small or large enterprises' income when detected by bank machines.

The purpose of this project is to create a machine learning model that classifies banknotes into two categories, forged and real, using a dataset of banknotes with extracted security features.

The dataset provided is one containing four columns representing features of bank notes and 1372 rows representing the number of bank notes observed. For simplicity, the number of features/columns has been limited to two only.

The data was initially plotted on a graph against the two features, skewness and variance, data points representing each bank note, as shown below.
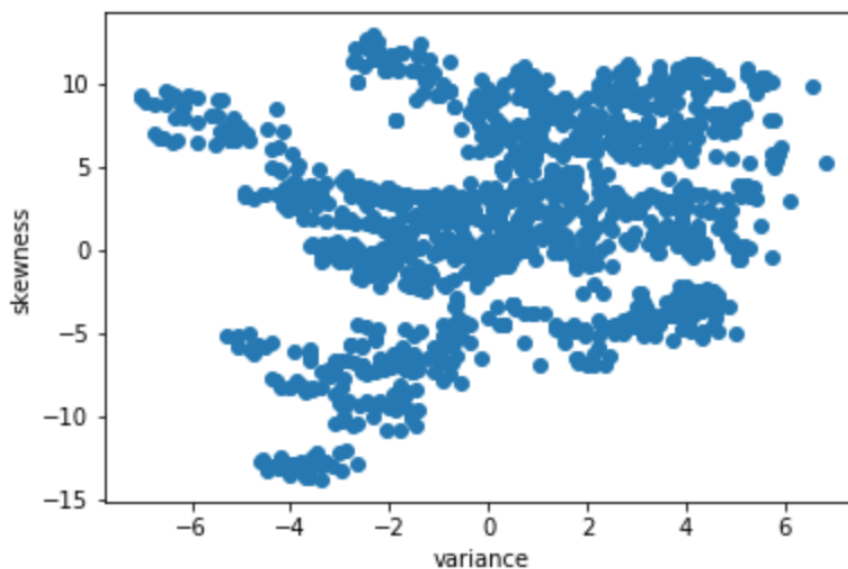
But this graph only showed the initial dataset provided, not the final one after preprocessing, analyzing and evaluating.

The method used to analyze this data is an algorithm called KMeans Clustering. It groups data close to one another together to form one group of points with similar features. If data

points are close and have similar features, then they most probably belong to the same group.
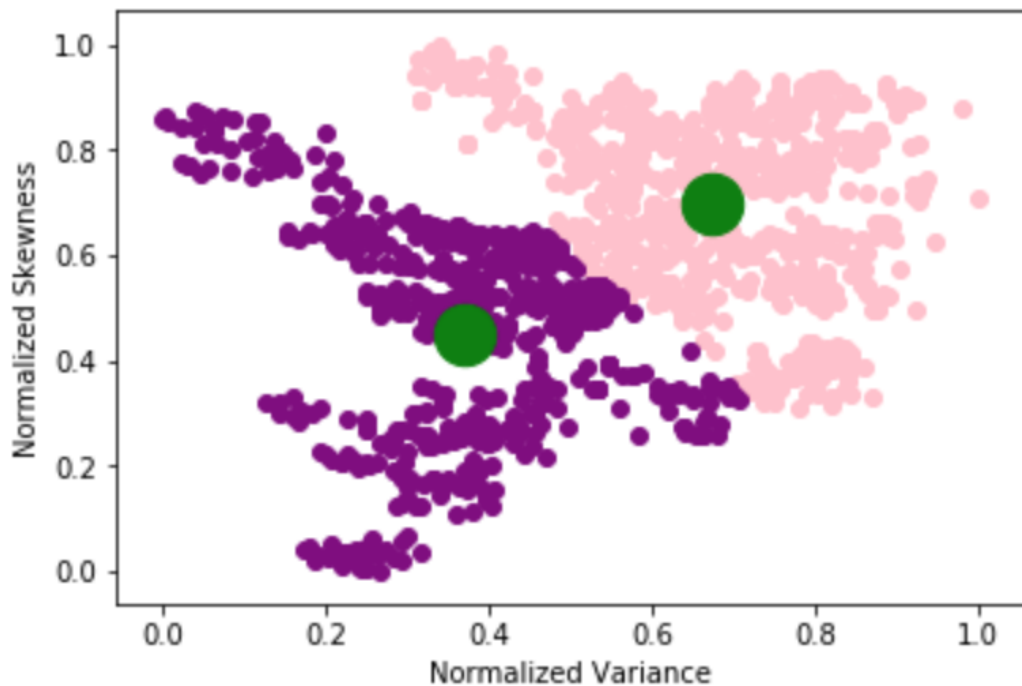
This final graph clearly distinguishes between two groups, one labeled in pink and the other labeled in purple. One group represents forged bank notes and the other represents real bank notes. The two slightly larger green circles are the centers of both groups to show balance and how each datapoint was chosen based on distance to that center point.

The results of this project were fairly consistent, given a dataset of bank notes with extracted features, specifically the two features analyzed, skewness and variance, a bank note can be automatically classified into either Forged or Real. The dataset provided is of high quality and no missing values which aided in creating a solid algorithm. It is important to note that not all features have been analyzed, only two. It could be possible that more dimensions can be of higher priority when determining the authenticity of a bank note.

I recommend creating a similar model with the same dataset but using other features and determining the clusters, later compare both results and combine. From a business point of view, automating banknote detection can reduce fraud and maintain income levels and so implementing this model can financially benefit the company and so i highly recommend it based on results from training data.