Banknote Data Analysis Project: Client Report

Purpose of the Project

The goal of this project was to use data science techniques to investigate whether it is possible to automatically identify fake banknotes based on their measurable features. By analysing data about individual banknotes, we aimed to determine whether there are clear differences between genuine and counterfeit notes and whether a computer algorithm can be trained to distinguish between them.

Description of the Data

The dataset provided include measurements from each banknote, specifically two features, called **V1** and **V2**. Think of these as scores that capture certain patterns in the images of your banknotes. Each row stands for a single note.
A few things about the data:

- Both features are just numbers, with no missing or blank spots.
- We didn't know in advance which notes were real or fake, the data didn't tell us; our job was to look for groups that "naturally" stood apart.

How We Looked at the Data

Here's what we did, step by step:

First, we got a feel for the numbers.
We looked for gaps or strange values and checked how those two features were distributed. Nothing stood out as a problem, so we moved forward.

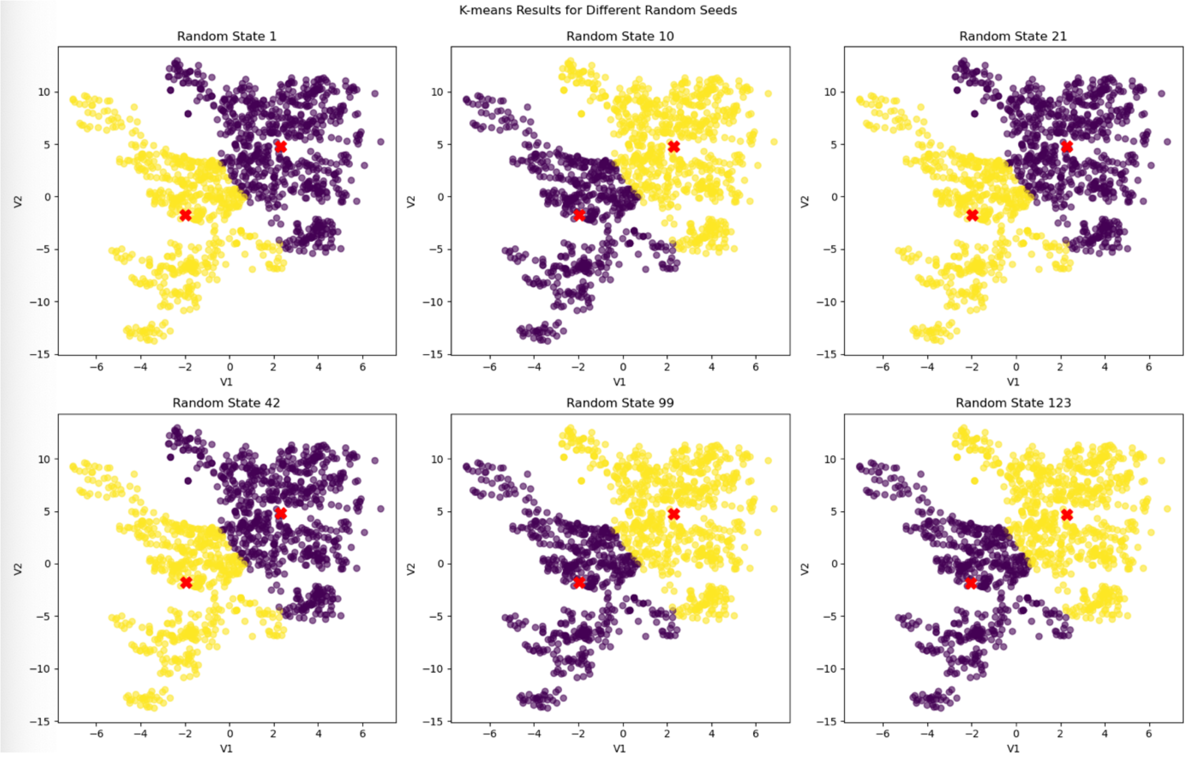
We put the two features from the banknotes (V1 and V2) on equal footing.
Since one could have bigger numbers than the other, we "standardised" them. Basically, this levels the playing field so one feature cannot overshadow the other.

We ran something called k-means clustering.
This is a way for a computer to find groups in the data without us telling it what to look for. We asked it to split the notes into two groups, just based on V1 and V2, to see if any patterns emerged.

We tested the process a bunch of times.
Because the computer starts in a different place each time, we repeated the clustering several times to make sure we weren't just seeing a fluke.

We made scatter plots to see what was going on.
These graphs show each banknote as a dot, coloured by which group it was assigned to. We also marked the “middle” of each group (the centroid) with a big X.

You can find an example of one of these scatter plots below:



What We Found

Two clear groups jumped out.
No matter how many times we ran the clustering, the notes consistently split into two main clusters. The computer kept putting the “middle” of each group (the centroid) in the same place each time.

Results were stable.
Running the analysis with different starting points barely changed the outcome. Only a handful of notes near the “border” between groups sometimes switched sides, but the vast majority were always sorted the same way.

What this tells us:
Even though we did not tell the computer which notes were real or fake, the data itself seems to hold enough information for an algorithm to sort them into two groups. That is a strong sign that the chosen features (V1 and V2) are capturing real differences between types of banknotes.

Suggestions Going Forward

You can use these features for automated screening.
Based on what we saw, it's practical to set up a system that checks these features and flags banknotes that don't fit the usual pattern.

For the tricky cases, more data could help.
The only notes that weren't always clearly sorted were those with values close to the "border" between groups. If you can add other features (for example, other measurements from your banknote images), you might be able to clear up the last few uncertainties.

Keep testing as you gather more notes.
Data is never finished. If you continue to collect measurements, you can keep updating the process, making it more accurate as new types of counterfeits appear.

Final Thoughts

This project shows the power of data science, even with a fairly simple approach. Your data already lets a computer find patterns that could help spot fake notes quickly and consistently.

With a little more information, this method could become a key part of your fraud prevention toolkit.

If you want to discuss these results further or need help with the next steps, such as setting up the screening system or gathering more features.