# Report on Automated Detection of Forged Banknotes

This project has the objective of assisting the bank system in automating the detection of forged banknotes. Automating this process with data science techniques such as Machine Learning can provide the bank with a fast and reliable method to identify counterfeit notes.

---

## Description of the Data

The dataset used in this project contains features extracted from imagens of banknotes, with each entry characterized by features labeled as V1 and V2.

**V1:** A continuous variable representing a transformed component from the image.

**V2:** Another transformed feature likely related to skewness or entropy.

An example of the data is shown below:

|      | V1        | V2        |
| ---- | --------- | --------- |
| 1300 | -0.971746 | -0.082540 |
| 823  | -0.849039 | 0.222902  |
| 1118 | -0.834927 | -0.059955 |
| 1196 | -0.861672 | 0.300848  |
| 620  | 1.065854  | -1.085982 |
| 1307 | -1.081855 | -1.080340 |
| 1349 | -0.710426 | -0.695179 |
| 543  | -0.652925 | 1.658769  |
| 92   | 1.362786  | 1.071169  |
| 767  | -0.955101 | -0.379862 |

The values for V1 and V2 exhibit variability, suggesting differences between genuine and forged banknotes.

## Methods of Analysis

The data were analyzed using K-Means clustering algorithm, a popular unsupervised learning technique. K-Means groups data points into clusters based on similarity, which can help identify distinct categories, like genuine or forged. The analysis process was repeated with different *random_state* values (**1, 2, 3, 4)** to evaluate the stability and consistency of the clustering results.

**Key steps of the process:**

- Data Processing: The data were standardized to ensure features contributed equally to the clustering process.
- Clustering: K-means was applied to partition the data into two clusters
- Evaluation: The impact of the *random_state* parameter on cluster formation was assessed to determine the reliability of the results.

## Summary

- The data was successfully partined into distincts clusters by the K-means algorithm, indicating potential separation between genuine and forged banknotes.
- Varying the *random_state* value led to minor differences in cluster assignments, but the overall structure of the clusters remained consistent.
- The features V1 and V2 appear to be effective discriminators, as their values show clear separation in the clusters.



Comparison of K-means Executions with Different random_state Values
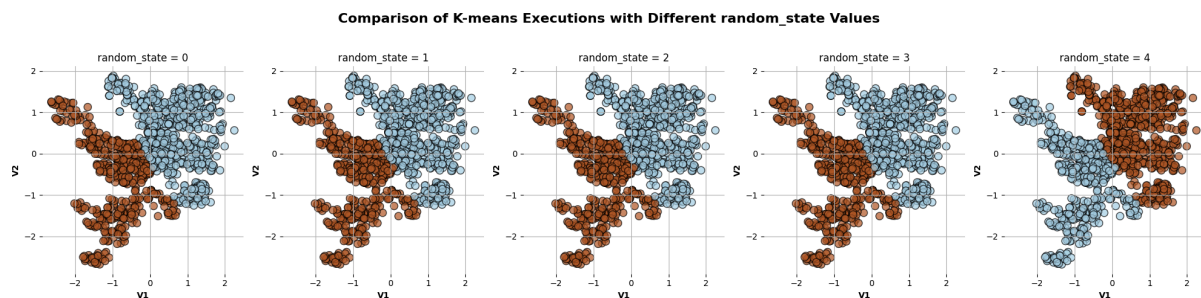
## Recommendations for the Bank

Based on the findings, the following recommendations are proposed:

**Implement Automated Pre-Screening:**

- Deploy an automated system based on clustering (or supervised models if labels become available) to perform a first-pass check on banknotes.

- This can flag suspicious notes for further manual verification, significantly reducing workload.

**System Integration:**

- Integrate this detection model into the bank's cash handling machines or teller systems.

- Provide real-time feedback to staff when a note is likely forged.

**Monitor and Update Models:**

- Continuously monitor the system's performance and retrain the model with new data to adapt to changes in counterfeit patterns.