KYC Document Identification and Facial Image Analysis Report

Version 1.1

**Manoj C**

# Executive Summary

This report aims to delivers an explanation of the roots causes affecting the decreasing pass rate of on-boarding new customers as part of KYC joining process as well as provide solutions to these issues.

The report will outline supporting tools used, such as Python and Excel to assist with the data analysis. The Python code and the Excel sheets will accompany this document.  
  
The findings suggests improving the quality of the uploaded document for identification is the single most reason for the decreasing pass rate for client on-boarding.

Other lesser factors include a link between the issuing country of the document and the attributable “Score” value, and the potential fraudulent nature of the documentation submitted from such countries, is itself a legitimate reasons for failing validation.

# Heat Map Analysis Python scripts used to generate the heat maps are accompanied with this document. The heat maps highlight attributes that positively or negatively correlate with each other and identify patterns in the data for further investigation. The first heat map shows the overall correlation between passing and non-passing verification-determining attributes, the second, displays the attributes contributing to document Id validation failure rate and the last the attributes contributing to the facial image recognition failures.

### Overall heat map of all passing and non-passing verification of clients

# This overall heatmap shows the complete correlation of the underlying contributing attributes for both documentation Id and facial image matching verification. Facial matching compares the image of the user against image on the document Id. Figure 1.

# Where there is high correlation (more green), this would indicate certain attributes positively contributing to the success or failure of the overall document validation process. Low correlation (more red) boxes would indicate an opposite negative correlation identifying an inverse relationship. By looking at the highest and lowest correlations it should be possible identify the most likely contributing attributes affecting the low on-boarding pass rate of new clients.

### Document Id heat map of non-passing verification of clients

The following heat map identifies the correlation between attributes that caused the validation of the document Id to fail.

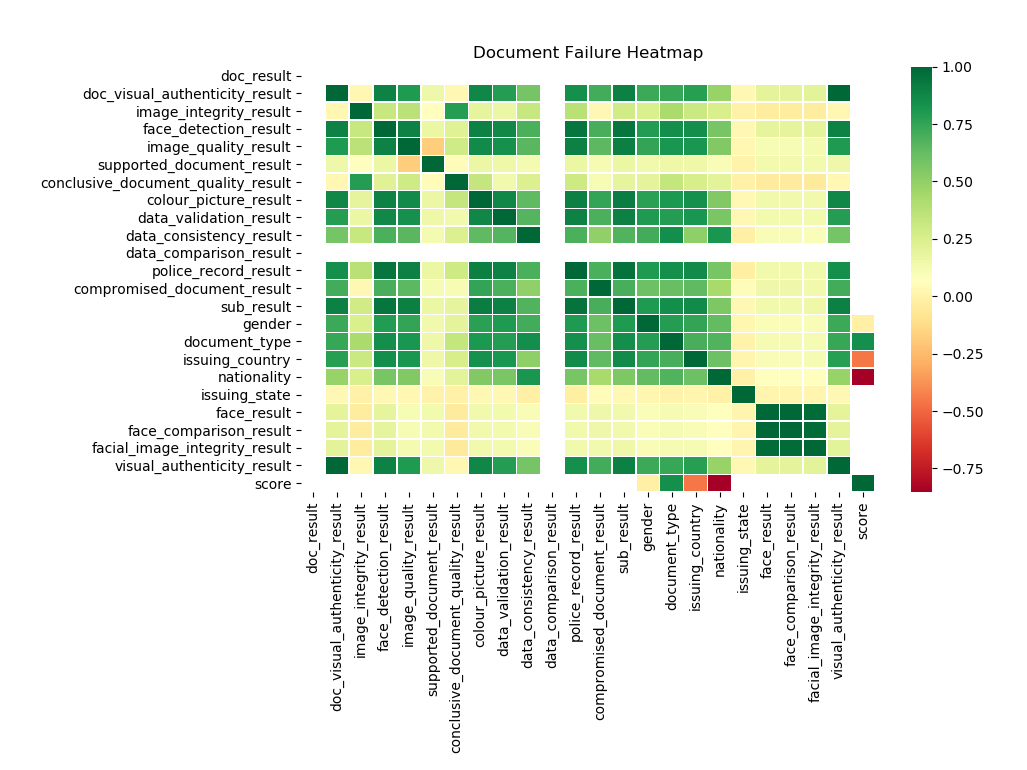
There is a clear positive correlation between the document’s “visual authenticity” and the “image integrity” attribute – which is based on the “face detection” and the quality of the facial image.  
  
There is also a positive correlation between the “Sub Result” and the documents underlying “visual authenticity”, “face detection” and “image quality” attributes. The documents’ “image integrity” is also a factor but data analysis suggests the most predominant cause is the documents’ poor “image quality” attribute.

Figure 2.

### Overall Heat Map of all passing and non-passing verification of clients

The following heat map identifies the correlation between attributes that caused the validation of the facial image processing to fail.

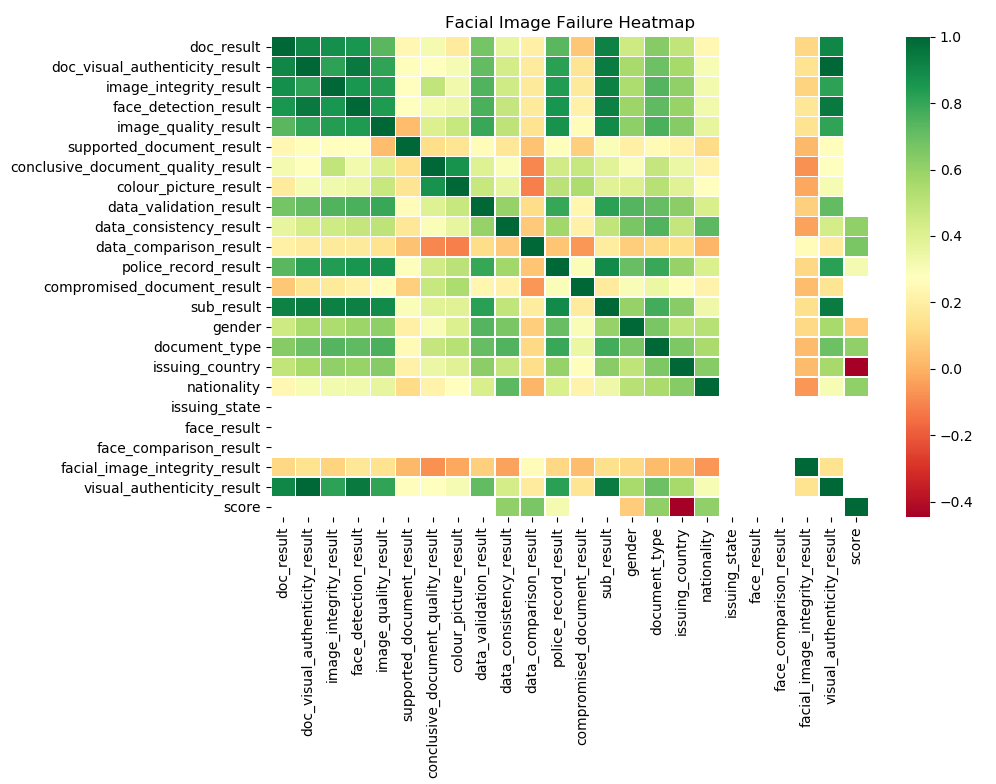


Figure 3.

# Analysis of Key Findings and Solutions Here are findings from the analysis as outlined below and a proposed solution to the each one.

## **Failure to identify an individual from their photo or video**

### Analysis

The largest reason for decreasing passing rate of document Id verification is due to the failure of identifying the individuals from their photo or video based on the facial “Visual Authenticity” attribute. 834 verifications failed because of this attribute and is the largest causing issue affecting client on-boarding.  
  
Subsequently, the “Visual Authenticity” attribute affects whether or not the individual will be searched on databases, defining whether the “sub\_result” field will be populated with an “unidentified” value. Failure to identify the individual, prevents further police records from being checked - affecting 184 client cases to skip verification.

### Solution

Before submitting the document, provide a tool that can image or ocr scan the document to ensure it is readable. Where its not readable, make the user upload the document again, before processing it until it is readable. They could even upload a pre-scanned document, sitting on their phone file system or cloud, which they’d previously scanned from a printer as an example.

## **Document Id verification from certain groups of countries more likely to fail**

### Analysis

# There appears to be a set of countries whose nationals have a low “Score” or have blank “Score” values. The “Score” indicates whether the facial image check was manually done or failed the “image integrity” checks with a low score. There is a strong negative correlation between the “Score”, the clients “Nationality”, the document “Issuing Country” and document “Type”, indicating that certain states are providing documents that aren’t of the necessary quality to be verified easily. Here is a table breakdown of which countries and their documents are causing the most document Id validations to fail:

|  |  |  |
| --- | --- | --- |
| **Nationality** | **Count of Document Types affected** | **Total Failed Documentation Id Count** |
| BEL | National Id Card 9 | 9 |
| ESP | Passport=6, National Id Card=12 | 18 |
| FRA | Passport=36 | 36 |
| GB | Passport=67 | 67 |
| IRL | Passport=36 | 36 |
| LTU | Passport=10, National Id Card=18 | 28 |
| POL | Passport=2, National Id Card=14 | 16 |
| PRT | Passport=8, National Id Card=9 | 17 |
| DEU | Passport=6, National Id Card=6 | 12 |
|  | **Total** | **239** |

### Solution

Request government authorities who have these issues to improve the quality of their documents.  
  
Again, put in place an OCR scanning process to validate the document before submitting it for verification.   
  
Develop with phone manufacturers a technical solution to improve digitisation of documents when they are taken from a phone camera.   
  
Find a technical solution to that allows registering your documents officially with your local government and using a government provided password key and reference to pass a one-time generated code on your phone to the 3rd party you need to divulge the information to.  
Since if the information linked to your registered mobile phone it could be the single point to authenticate government documents in the future.

## **Improve and automate the verification rate of “Suspected clients**

### Analysis

# There are 56 cases where individuals’ status was set to “Suspected” because their documentation Id might be fraudulent. 40 of these cases had their client documents’ “Visual Authenticity” verification result set as unclear.

# These 56 case belong males, mainly from the following countries: ITA, CHE, LTU, IRL ,BGD, LVA, CHN, GBR, SRB, ESP, DEU, FRA, POL, BEL, SVK, PRT and GRC who’ve primarily used their passports, followed by their national id cards to apply for a bank account.

### Solution

If we can improve the image quality as outlined earlier, this would increase the “visual authenticity” of the document verification based on the data analysis done so far. Clearly we want to filter fraudulent Ids, but not at the expense of poor image quality.   
  
We could again ask the governments to improve their documentation Ids with better digitally secure fonts. Improved digital layout of these documents so they are more camera friendly to read, having security in different places on the document.  
  
We could investigate introducing a system that utilises a robust one-time-based mobile phone network transmitted password and key to government accessible store of this information digitally.