IMPROVING CUSTOMER CHECK PROCESS

Business Problem and Dataset

According to Financial Conduct Authority (FCA) regulations, the financial company has to verify the identity of all customers who want to open an account. The person who wants to open an account has to submit a government-issued photo ID and a facial picture to Veritas. Veritas then would perform two checks "Document Check" to verify that the photo ID is valid and authentic and "Facial Similarity Check" to verify that the face in the picture is the same as that on the submitted ID.

The customer will "pass" the checking process and get onboarded if the results of both Document and Facial Similarity checks are "clear". If the result of any check is not "clear", the customer has to submit all the photos again.

The "pass rate" is defined as the number of customers who pass both the checking process divided by the number of customers who attempt the process. Each customer has up to 2 attempts. The pass rate has decreased substantially in the recent period. This report has been written to analyze the reasons behind this decrease and to make suggestions regarding the problem.

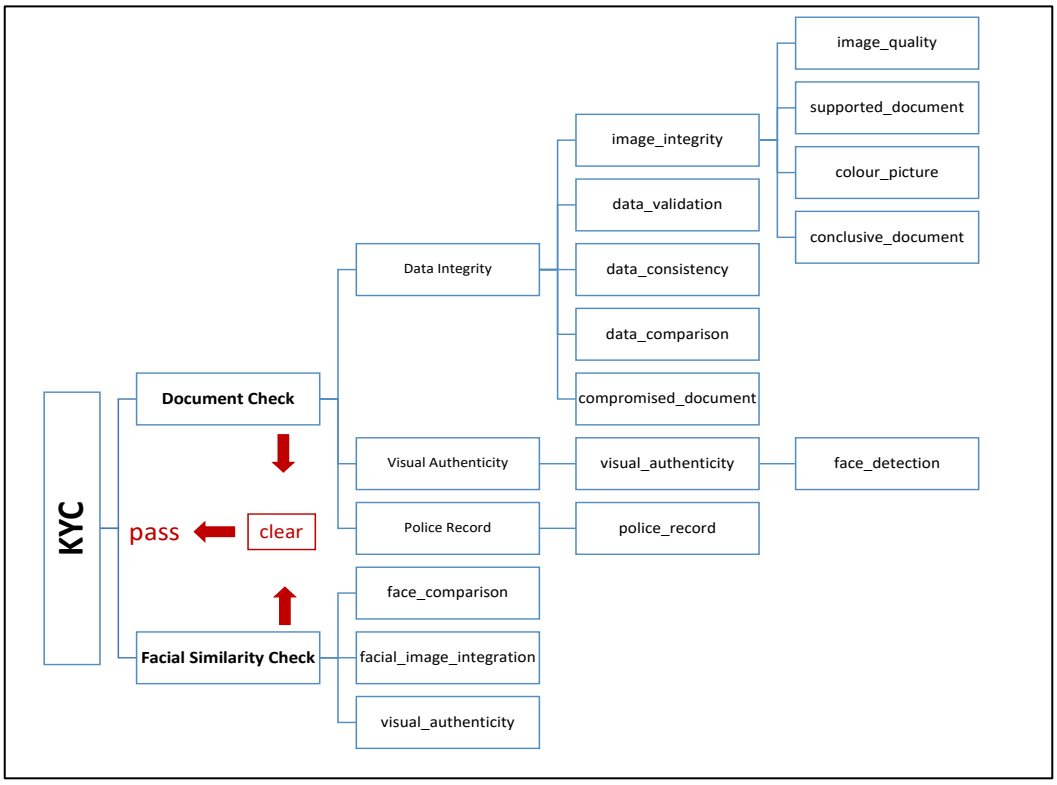

Figure-1: Dataset Diagram

The document report is composed of data integrity, visual authenticity and police record checks. It checks the internal and external consistency of the most recent identity document provided by the applicant to identify potential discrepancies. The facial similarity check will compare the most recent live photo or live video provided by the applicant to the photo in the most recent identity document provided.

The dataset includes the date between 23.05.2017 and 31.10.2017, which is approximately a five-month period. Each row refers to one prospective customer who attempts to open an account. Some "user_id"s could be repeated since each customer has a right to apply up to 2 attempts.

Data Analysis

Pass rate is the key performance indicator for this business problem. That's why it is crucial to calculate the pass rate properly. Here is the formula for the pass rate:

$$pass\ rate = \frac{\#\ of\ customers\ who\ has\ 'clear'\ result\ from\ both\ KYC\ process}{\#\ of\ customers\ who\ attempt\ to\ process}$$

The pass rate should be examined by days to be able to see the trend over time. For this purpose, the number of customers who has 'clear' result from both document check and facial similarity check and the number of distinct user_id, i.e. the number of customers who attempt to process, are calculated by date. It is obvious that there is a substantial decrease from the beginning of September. The average pass rate for the whole period is 0.85, while it is 0.90 until the end of August and 0.76 from the beginning of September. Therefore, it can be deduced that the decreasing trend has started at the beginning of September.

	Whole Period	Until the end of August	From the beginning of September
Mean	0.85	0.90	0.76
Median	0.88	0.89	0.76
Std.	0.09	0.05	0.09

Figure-2: pass_rate over time[1]

When the variables are examined, it is important to find out which check type has more impact on the decrease in the pass rate. Figure-3 and Figure-4 show the trend in results of document check i.e. 'clear' or 'consider' and the number of clear results in all attempts by prospective customers, respectively. There is an obvious increasing trend in both clear and consider results, however, Figure-4 shows that the increase in the number of consider results has higher than clear results, so that the pass rate for document check has decreased.

[1] For the 19-10-2017, data is not available in both datasets. Furthermore, in facial similarity check dataset, there is no record for 20-10-2017 and 21-10-2017.

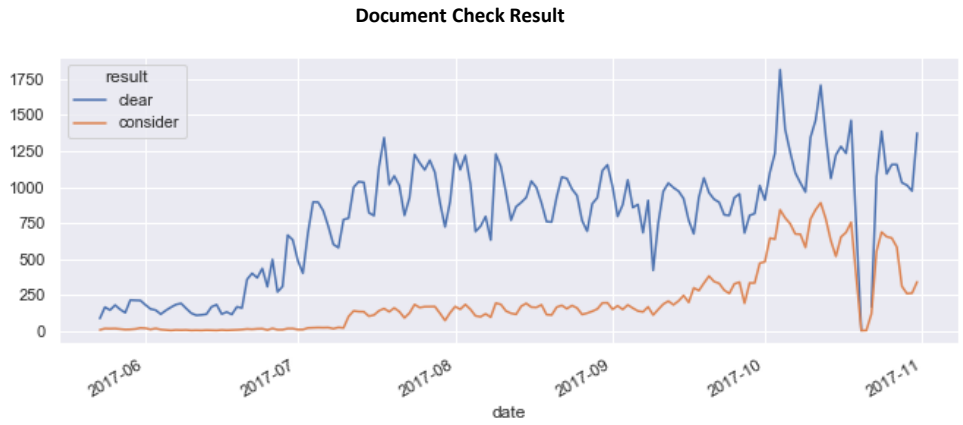

Figure-3: Trend in results of document check over time

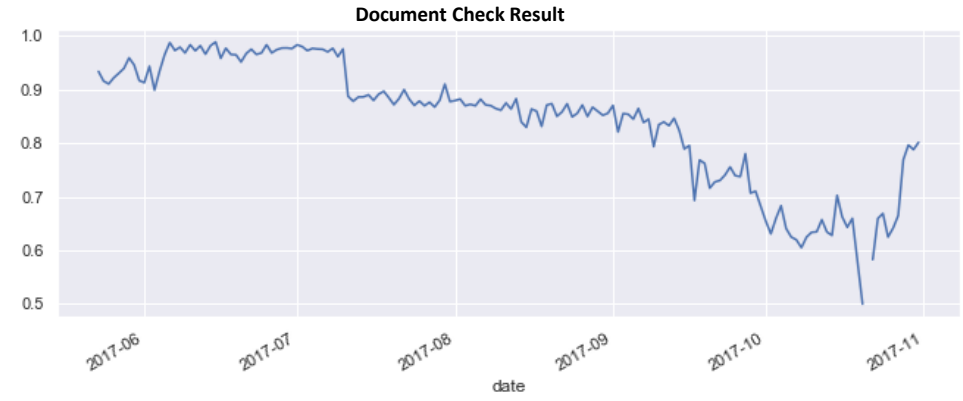

Figure-4: Clear results rate of document check

On the other hand, there is no substantial change in the result of facial similarity check as Figure-5 shows.

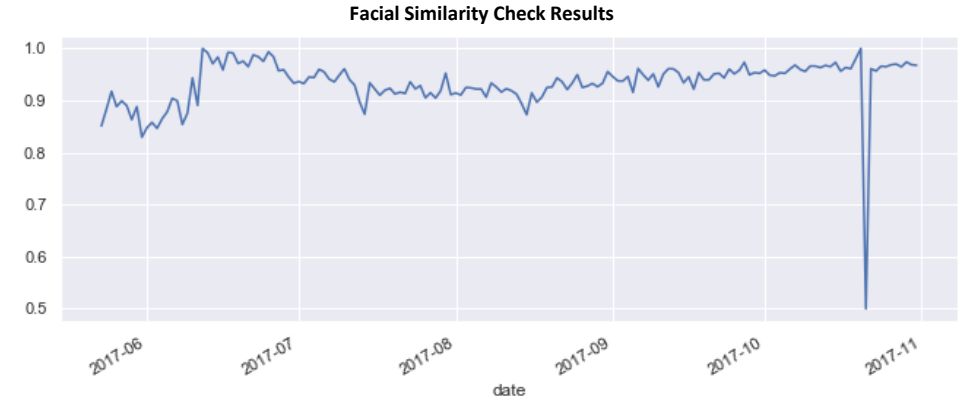

Figure-5: Trend in results of facial similarity check over time

All variables i.e. image_integrity, data_validation, data_consistency, data_comparison, compromised_ document, visual_authenticity and police_record in the first level which determine the final result for the document check were examined to reveal which variable most likely caused the decrease in the pass rate. It turns out that the number of 'consider' in image integrity results has started to increase in the middle of September and then has accelerated its increasing rate (Figure-6). There is no clue from other variables regarding the decreasing that we are digging the reasons of.

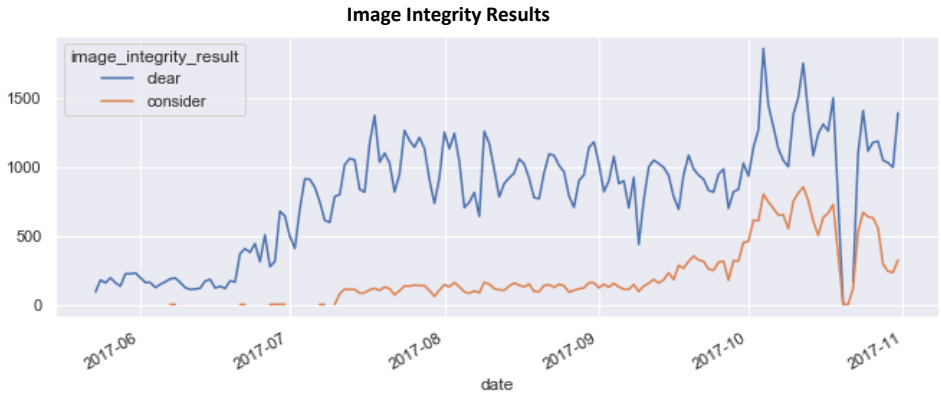

Figure-6: Trend in results of image integrity over time

The reason behind the increase in the number of 'consider's in image integrity result might be caused by its breakdowns; image quality, supported document, colour picture and conclusive document quality. The increase in the number of 'unidentified' results in image quality and 'consider' results in the conclusive document quality seem to be the most likely reason to explain the decrease in pass rate (Figure-7 and Figure-8). Moreover, the results of consider for conclusive document quality started to increase after the middle of September.

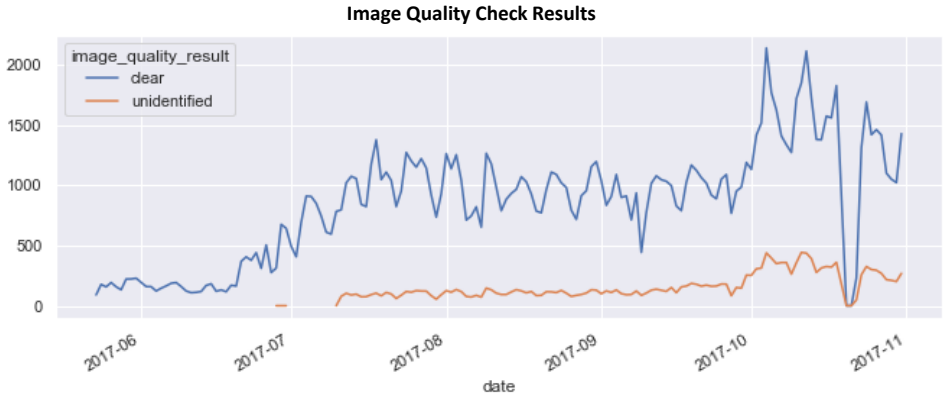

Figure-7: Trend in results of image quality over time

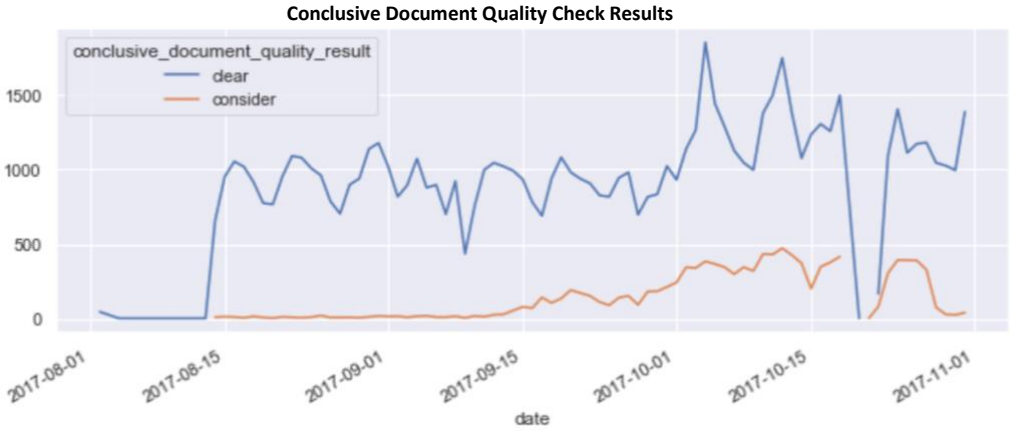

Figure-8: Trend in results of image quality over time

As a further step, a statistical approach has been used to approve inferences performed with plots. To this end, the covariance matrix has been constructed to reveal the relationship between the pass rate and all other variables available on both documents and facial similarity checks. Figure-9 shows the covariance matrix as a heat map. The covariance matrix suggests that the document check results (dc),

image integrity results (ir), image quality results (iqr) and conclusive document quality results (cdqr) are the most related variables with the pass rate (Figure-9).

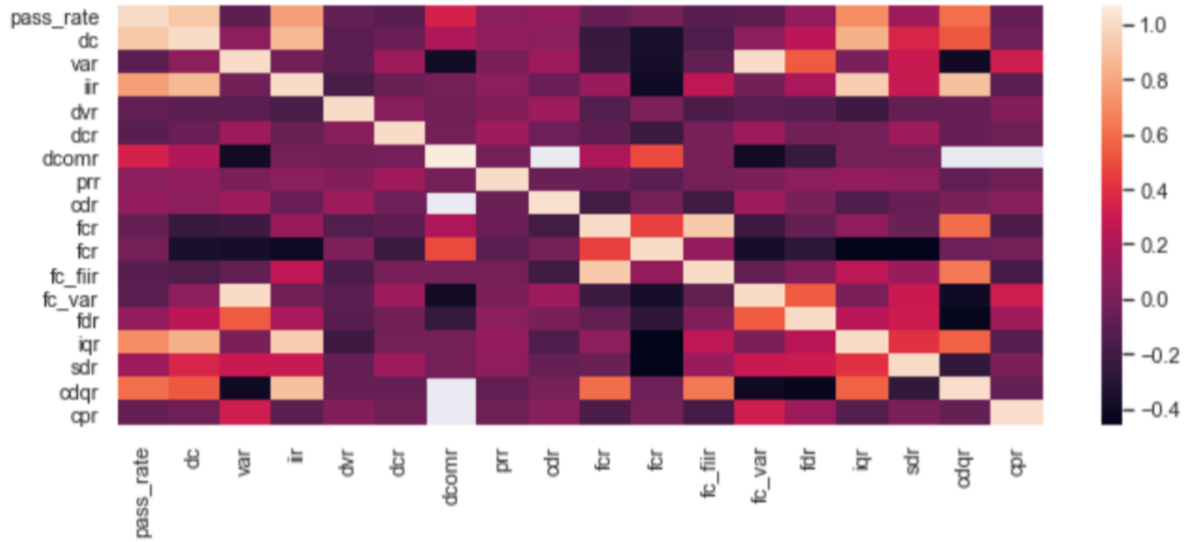

Figure-9: Heatmap for covariance matrix

Interpretation of Results

Based on the result of the analysis, it is found out that the increase in 'unidentified' results in image quality and 'consider' results in conclusive document quality are the reasons of the decrease in the pass rate over time. There are two possible scenarios related to the decrease in the pass rate over time:

- In the API documentation of Veritas, it is stated that a result of 'clear' in the conclusive document quality will assert if the document was of enough quality to be able to perform a fraud inspection. A result of 'consider' will mean that even if sub breakdowns of visual authenticity fail, we cannot positively say the document is fraudulent or not. The conclusive document quality data is available after '14-08-2017', the middle of August and the number of consider result has increased substantially after the middle of September. This means that before having conclusive document quality data, it could be harder to be aware of some fraud documents by only checking visual authenticity and its breakdown. The availability of conclusive document quality data helped to catch fraudulent attempts and improve costumer checking process.
- Regarding the image quality process, a new technology might have been started to be applied by Veritas to check the quality of images from late October. Therefore, the probability of detecting fraudulent documents' images might have increased.

The decrease in pass rate over time could be interpreted as the fraudulent attempts have been caught recently. On the other hand, if the process with document and facial similarity checking overestimate the fraudulent situations i.e. returning 'consider' result while the applicants' files are actually clear, this will cause a decrease in the number of customers. Minimizing the fraud risk and maximizing customer numbers in a healthy balance is one of the most challenging topics in digital banking. For this purpose, it is crucial to figure out the reasons behind the results by diving deep into cases by keeping touch with the related teams.