Operation Challenge Analysis

Problem

Task 1: Improving KYC

As a financial institution regulated by the FCA, XYZ has the obligation to verify the identity of all customers who want to open a XYZ account. Each prospective customer has to go through a Know Your Customer (KYC) process by submitting a government-issued photo ID and a facial picture of themself to our partner, ABC. ABC then would perform 2 checks:

- Document check: To verify that the photo ID is valid and authentic;
- 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 KYC 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 KYC 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.

Please write a report that outlines the root causes and solutions.

Relevant files:

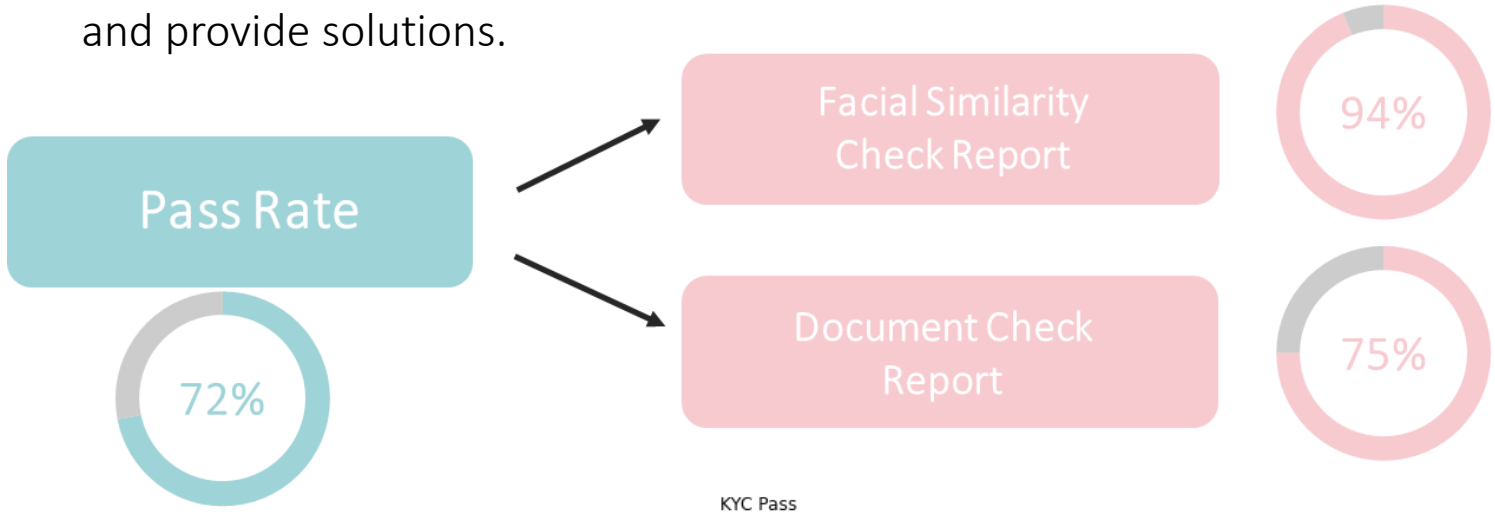
- facial_similarity_reports.csv - Reports of all Facial Similarity checks
- doc_reports.csv - Reports of all Document checks
- ABC.html - The API documentation of ABC explaining some terms used in the reports (you might need to download it to your computer and open it with a browser)

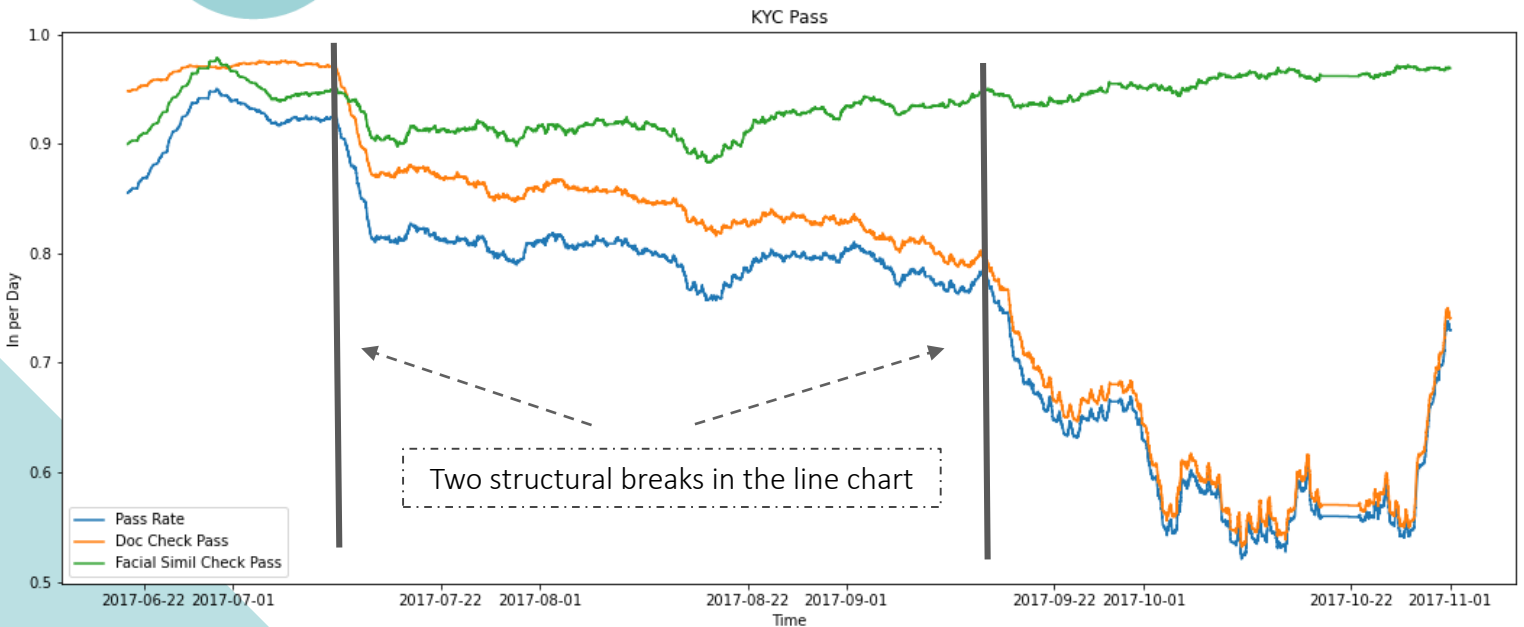
The candidate is free to use Excel or any scripting language to parse and analyse the data.

Please show all your work (including your code if applicable) and assumptions.

KYC Challenge

Problem Statement: The pass rate (no. of customer who pass both KYC process) has decreased substantially in the recent period. Find out why and provide solutions.





Before Mid July

Pass Rate	Facial Check	Doc Check
89%	94%	95%

Mid July-Mid Sept

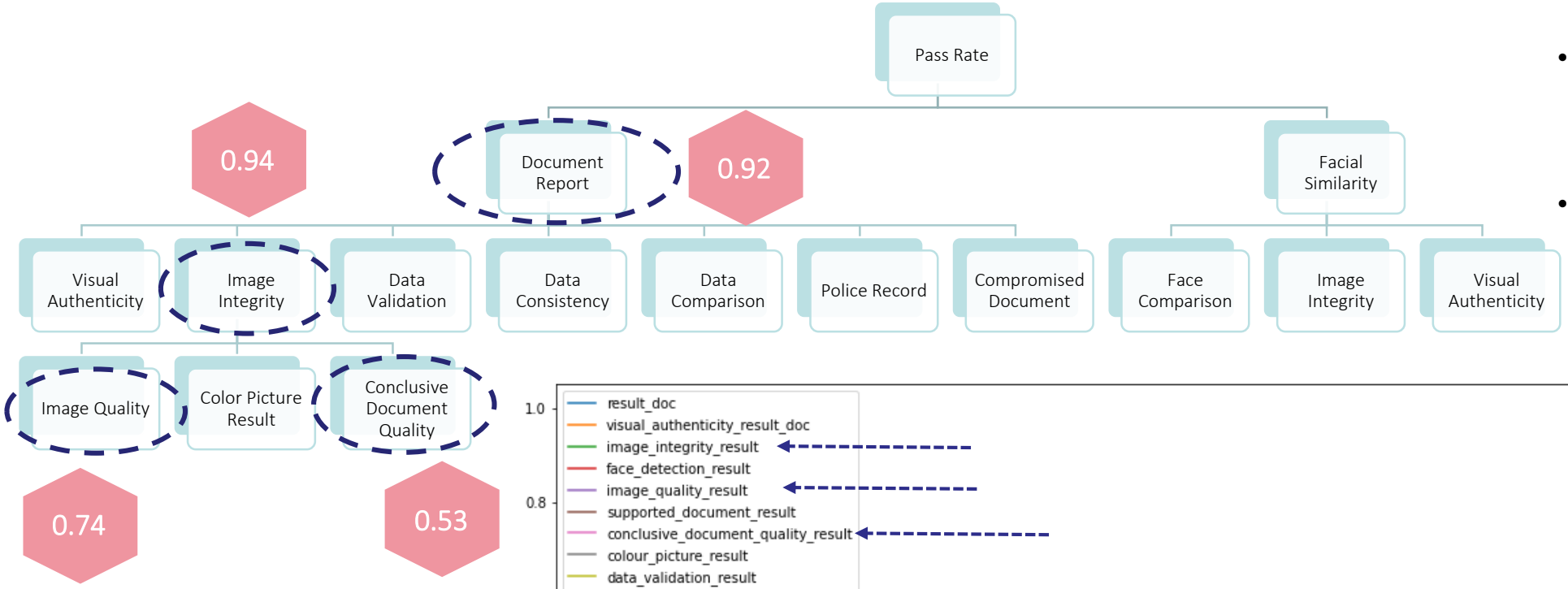
80%	92%	84%

After Mid Sept

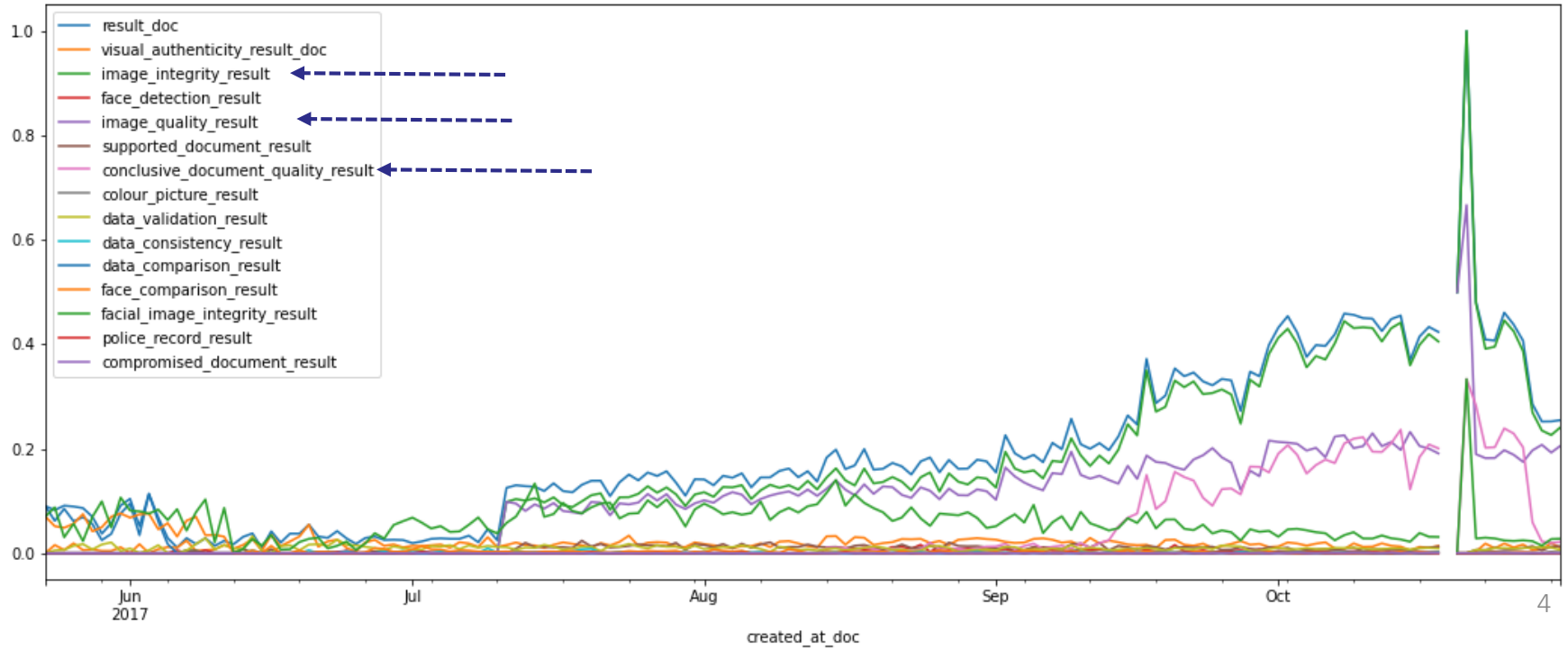
61%	96%	62%

- Customer passes KYC checks when he passes both:
  - Facial Similarity Check
  - Document Check
- From the adjacent graph we can find two structural breaks in Pass rate
  - From Mid July 2017
  - From Mid September 2017
- Also, it can be inferred that there is high correlation in pass rate and document check pass rate

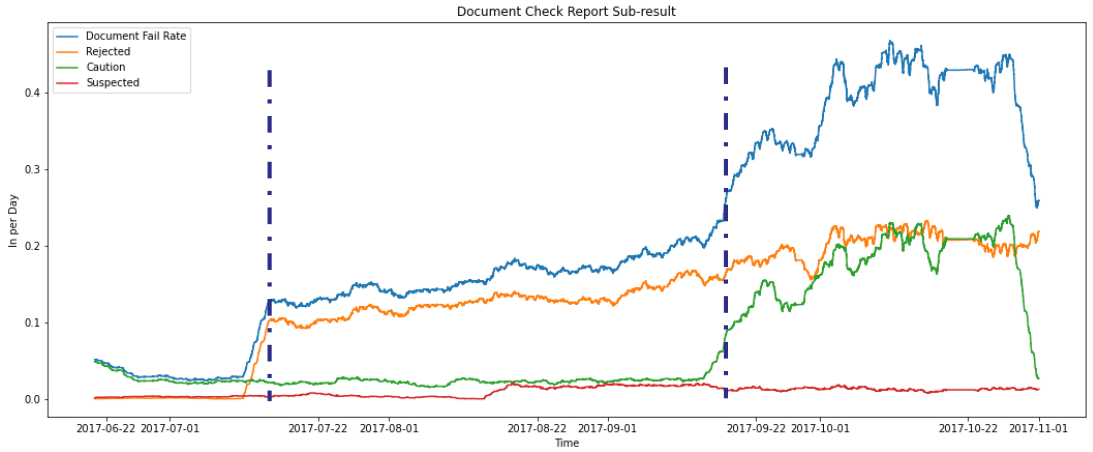
Correlation ANALYSIS

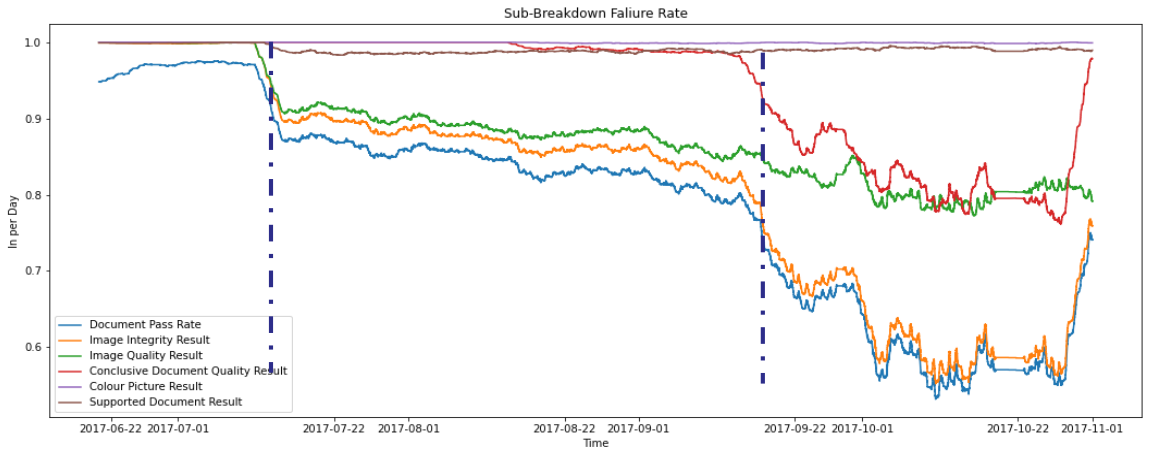


- result_doc has very high correlation to image_integrity_result 0.953
- Conclusive Document Quality & image_Quality_result is directly correlated to the decrease in the Document Pass Rate,
- The second drop coincides with the appearance of the Conclusive Document Result and Colour Picture Results starting to register results in August.

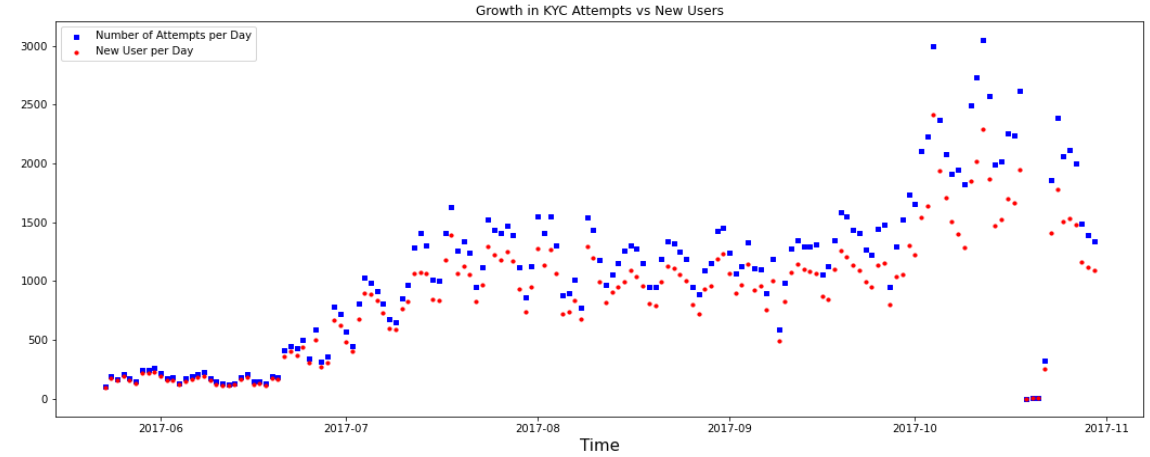


Drill Down on Breakdowns

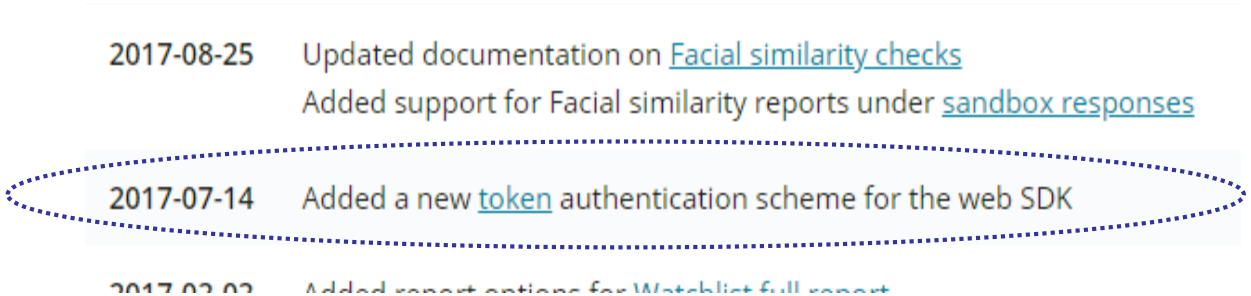




- As Document Fail rate increased between 1st-22nd July, 2017 so did **'Rejected'** increased but in mid September **'Caution'** results produced large increase in failure rate
- The same trend can be seen from fig: 'Sub-Breakdown Failure Rate', from mid-July to mid-Sept **'Image Quality Result'** but from mid-Sept **'Conclusive Document Quality Result'** has driven the failure rate
- Cause of two failures could be:
  - External
    - Increases in Fraud Attempts – Proved wrong by figure 3
  - Internal
    - Property Attribute Specific Issue – No correlation (proof)
    - Specific Events
      - Increased check parameters
      - New updates



RCA



2017-08-25	Updated documentation on Facial similarity checks Added support for Facial similarity reports under sandbox responses
2017-07-14	Added a new token authentication scheme for the web SDK

1

- The first increase in failure rate in overall Pass rate coincides directly with the "Addition of a new token authentication scheme for the web SDK"
- This may have **caused problem downstream**
- The same **changes may not** have been done to **XYZ web application**
- The new token addition is to **only one platform** as it hasn't resulted in catastrophic faliures

A result of `clear` in the `conclusive_document_quality` breakdown of `image_integrity` 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 (in cases such as parts of the document are not visible).

2

- The Image Integrity reports' main reason for failure from early to mid Sept is because of **'Consider'** result in Conclusive Doc Qual. Report
- 'Conclusive Document Quality Report' started appearing from **mid August**, and immediately next month we see a **high correlation** in Image Integrity failure rate and Conclusive Doc. Qual. Report faliure
- This seems to be **additional layer of check introduced**

Solution

- For the **first problem** following steps could be taken:
  - Establish **direct contact between engineering teams** of ABC and XYZ
  - Will keep **both teams updated** about new updates and will give **time to create action plan**
  - Will **decrease time** taken in identifying problems
  - Both teams could **collaborate** and improve the **overall integration**

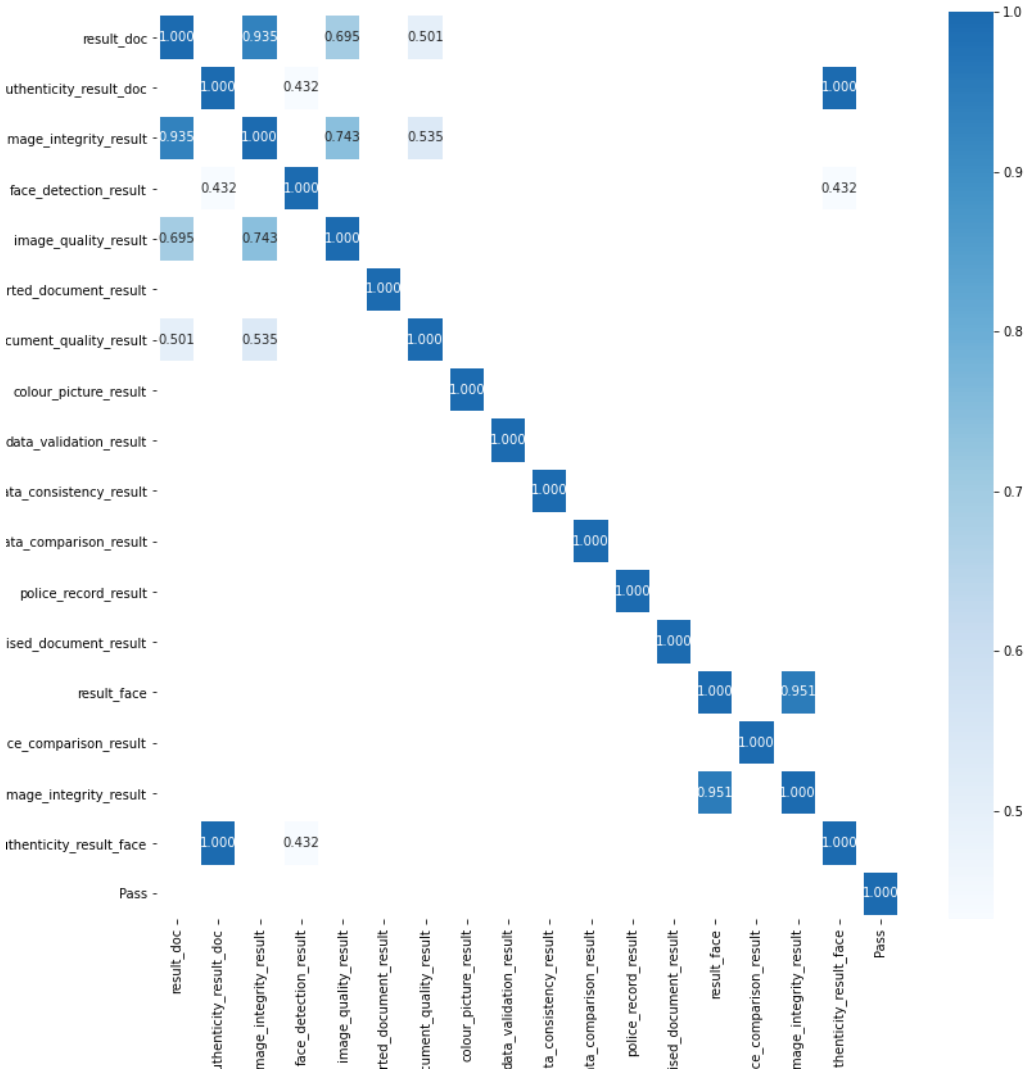
- For the **second problem:**
  - We can refer the reports with 'Caution' result for **Expert review**
  - As the API Documentation of ABC tells us: *"Expert review is required when we encounter images that use sophisticated counterfeiting techniques, or the image is of poor quality (blurred, low resolution, obscured, cropped, or held at an unreadable angle)."*
  - Additionally, more information is required towards the need for CDQ report.
    - Is it a regulatory requirement?
    - Why the sudden change?
    - What is CDQ report checking for? Alternate ways to do the same task?
    - Is it because of poor scanning capability of user? Can we create our own scanner in app to ensure that this doesn't happen (like Google PhotoScan)

Assumption & Caution: The above analysis and solutions are based on the limited data (which would be a sample of larger database). This Depends on how representative the data was of actual database.

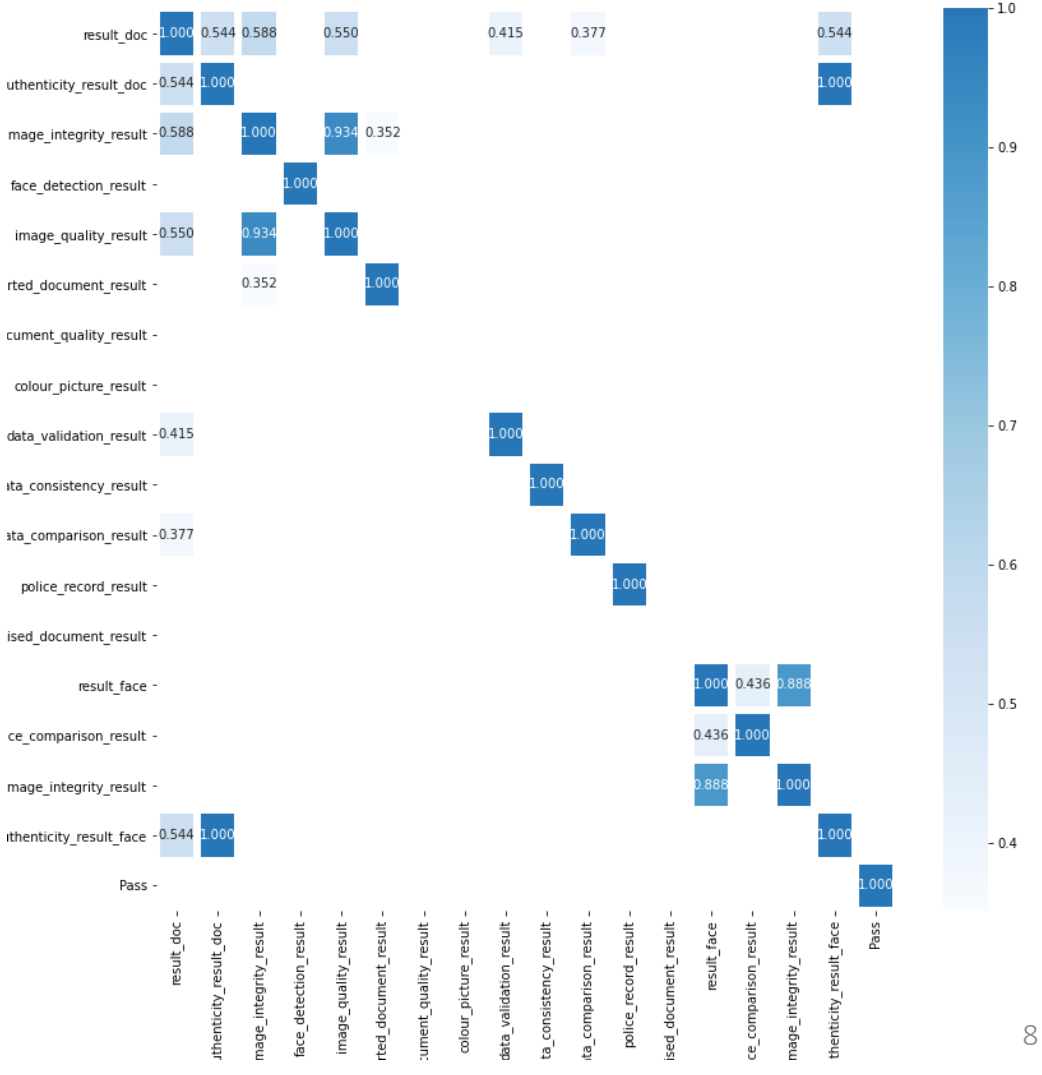
APPENDIX

Correlation Matrix - 1

Overall

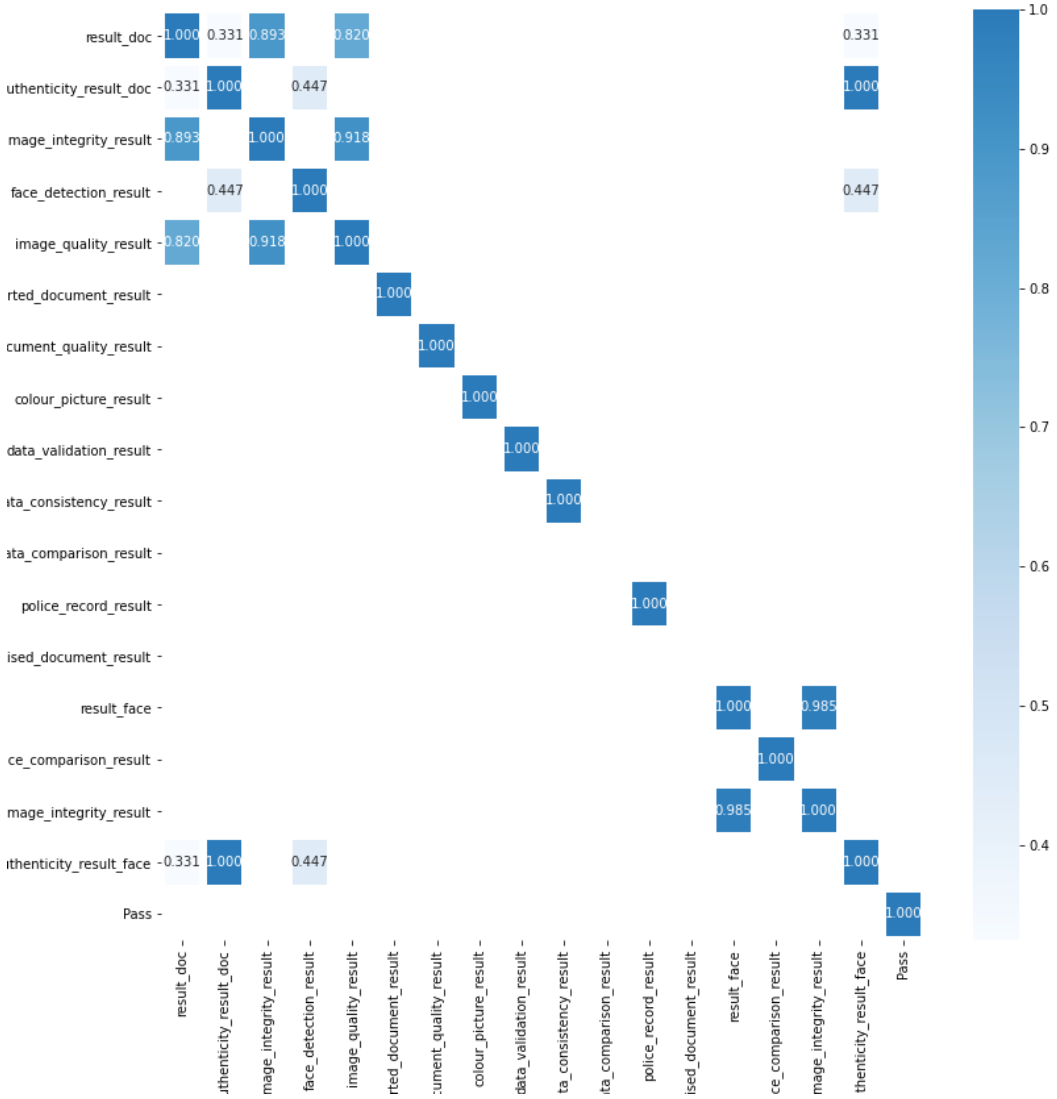


Before Mid July

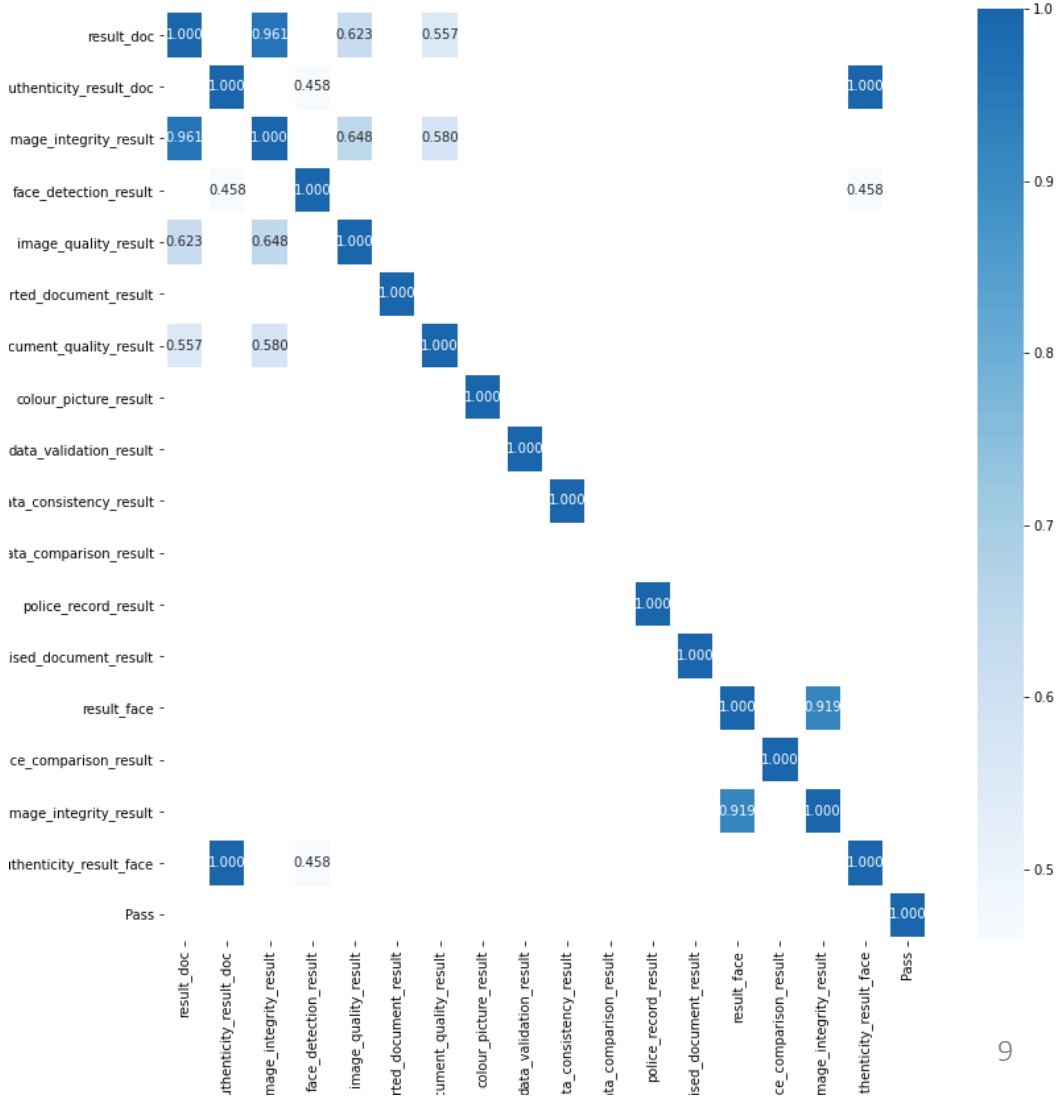


Correlation Matrix - 2

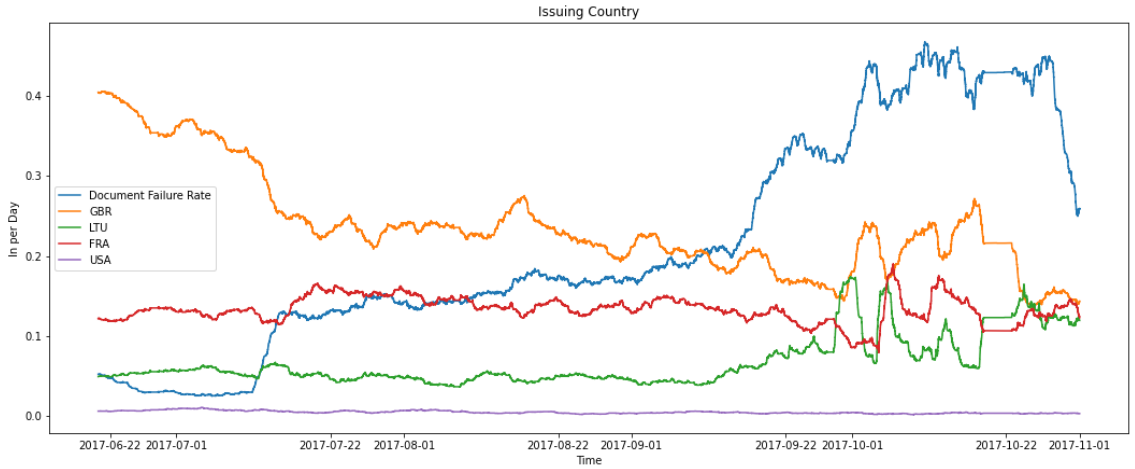
Mid July – Mid Sept

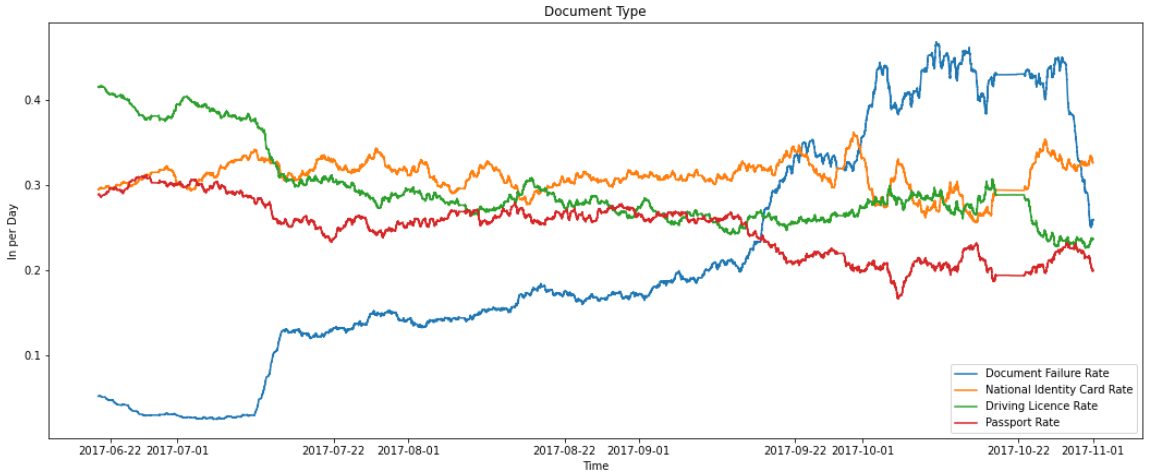


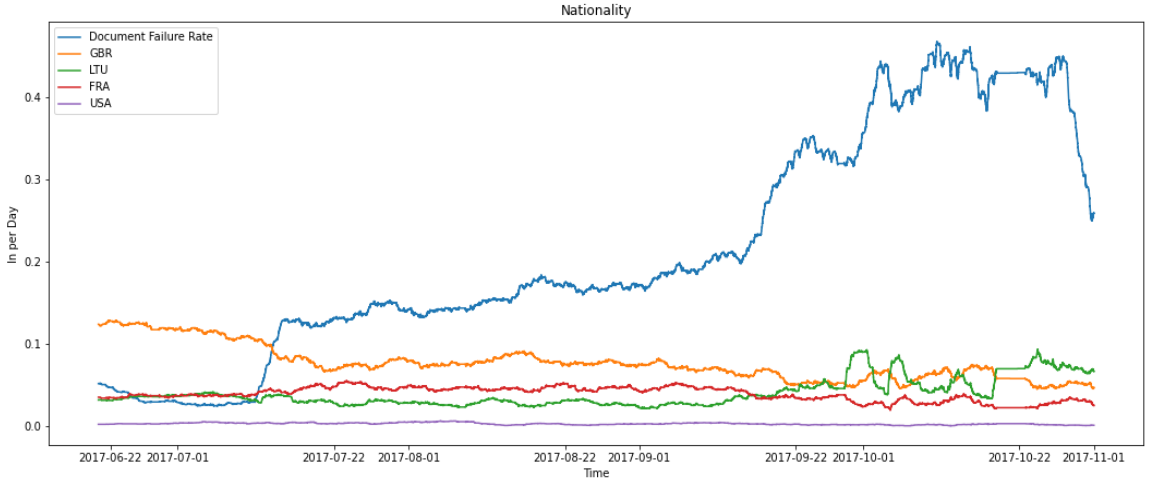
After Mid Sept



Properties Attribute







THANK YOU