*Revolut Operations Challenge*

**TASK 1: Improving KYC**

**Contents**

[Data Understanding and Data Preprocessing 2](#_Toc30869787)

[Key Findings 2](#_Toc30869788)

[Analysis 2](#_Toc30869791)

[Other observations through data files and Veritas API: 4](#_Toc30869792)

[Conclusion: 5](#_Toc30869793)

[Solution 5](#_Toc30869794)

# **Data Understanding and Data Preprocessing**

There is exactly one observation for each KYC process in the dataset. There are 176404 observations in document and facial similarity report. There are 18, 8 variables in document and facial similarity report respectively. All variables are of character or categorical type except **'created\_at'**. The time period of the data is ranging from 2017-05-23 to 2017-10-31.

Data cleaning steps:

* Check if same user id, attempt id, created at exist in both files. (***Note***: Found same for both reports so I have dropped the attempt id and created at column of facial similarity file.)
* Rename the column names, Change the data type of created\_at variable to datetime, and other categorical variables to factors, Replace null value with "unknown".
* Add one new column "Pass " with value "Yes" or “No”.
* Check for multiple attempts by users (During this check I found few users with more than two attempts.)

# **Key Findings**



## **Analysis**

**Statistics:** There are total ***142724******unique***customers who attempted KYC process from 2017-05 to 2017-10. There total **172610**KYC process observations for up to 2 attempts.

*Table -1: Number (#) attempts Vs # of customers (2017-05 to 2017-10).*

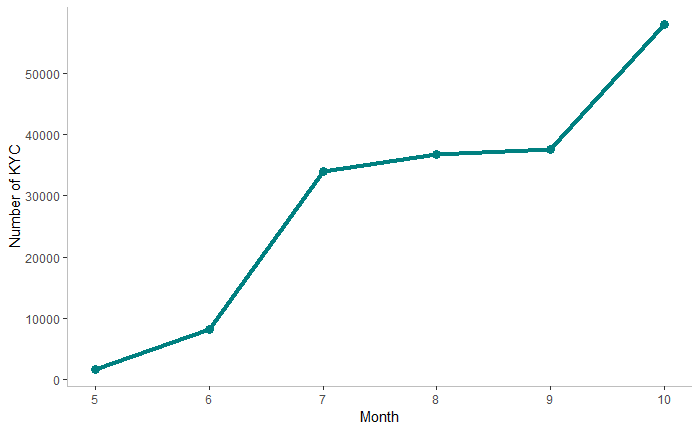
|  |  |  |  |
| --- | --- | --- | --- |
| **Attempts** | **First** | **Second** | **More than two** |
| **# of Customers** | 110374 | 31118 | 1232 |

*Table -2: Monthly Pass rate (2017-05 to 2017-10).*

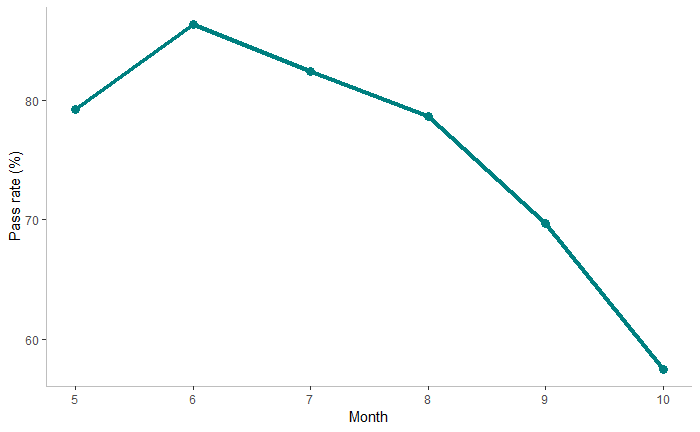
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **May** | **Jun** | **Jul** | **Aug** | **Sep** | **Oct** |
| **Pass rate (%)** | 79.24 | 86.30 | 82.38 | 78.62 | 69.69 | 57.52 |

***Note:*** *During the analysis I have found some users have taken KYC process again even if they have passed the process in first attempt. I have kept those data as it is.*

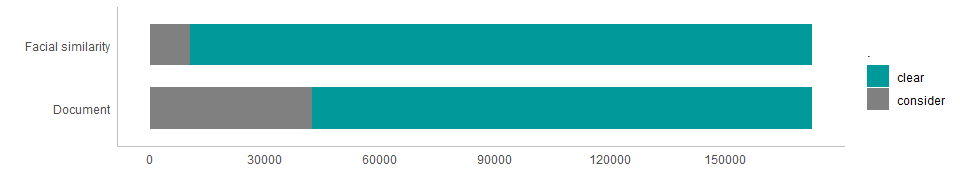
***Figure-1****: Growth in KYC attempts.*



***Figure-2:*** *Pass rate (May 17-October 17).*



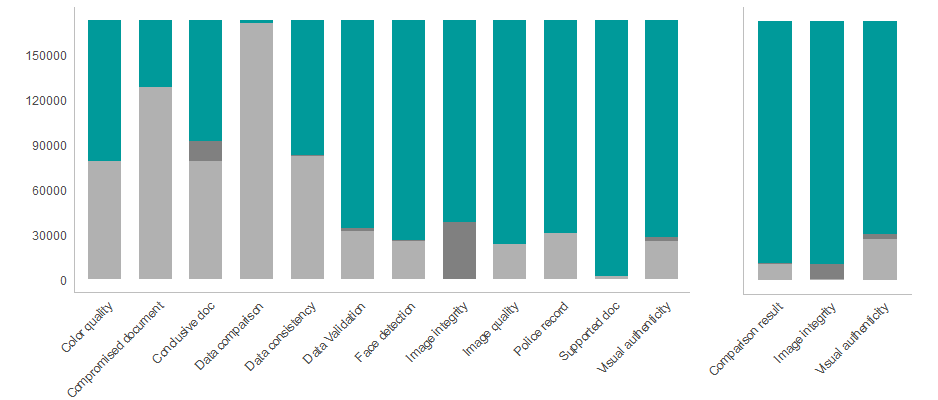
* There is a rapid growth in customer KYC attempts but the pass rate has fallen significantly (a drop of 29%) from July to October. June has highest number of pass rate which 86.30%.
* To dig into the problem of drop-in pass rate, I looked into the result of both checks in depth. First. I checked the overall result of document and facial similarity check.



***Figure-3:*** *KYC results by check*

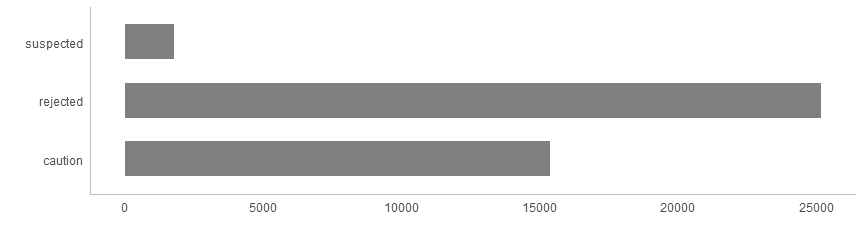
* *Figure-3* clearly shows that the most of the customers are failed to clear document check as comparing to facial similarity check. After this finding I did further detail investigation with different parameters of the both checks as shown in the *Figure-4*.
  + The overall clearance of the document check is depending on different parameters like visual authenticity, Image integrity, Face detection, Image quality, Color quality, Supported doc, Conclusive doc, Data Validation, Data consistency, Data comparison, Police record.
  + The overall clearance of the facial similarity check is depending on different parameters like Comparison result, Image integrity, Visual authenticity.

***Figure-4:*** *Document and facial similarity check parameters.*



* The main reason of decline is image integrity in both checks. Other root causes of document check failures are conclusive document quality, visual authenticity, data validation and face detection. For facial check, it is a visual authenticity.
* I also observed that there is a sub result for document check. There are categories like “caution”, “suspected”, and “rejected” in case of KYC unsuccessful.

***Figure-5:*** *Document check - sub result.*



* The sub result tells that most of the document check rejected and caution.

## **Other observations through data files and Veritas API:**

Looking at the data and Veritas API, I assume that user had been asked to upload documents and photo manually to process their KYC and also must have unique token to perform authentication. So, I assume that overall customer onboard process is time consuming.

***Note:*** *I went through Veritas API document quickly and I found Live Photo and Video section but I’m not sure of this. So, I assumed photo & document upload is manual process.* *There are two types of check - standard and express. Not sure about difference between these two.*

I also observed from data that Veritas was not considering the conclusive document quality result and color picture result till mid of August. It means the document check was started failing in mid of august because of these two parameters.

Data comparison result was captured initially till 1st week of June. I don’t think it has any impact with document check decline rate.

Most of the declined document check had been categorized as either rejected or caution.

* + **Rejected**: The reasons for rejection are poor quality or document type not supported.
  + **Caution**: If any other underlying verifications fail but they don’t necessarily point to a fraudulent document (such as the name provided by the applicant doesn’t match the one on the document). A huge amount of increase in document check failures were reported as “caution” in September and October.

I also check properties data for both check but I did not able to find any trend as I have not spent much time on analyzing the properties.

## **Conclusion:**

The main reason for pass rate decline is document check which can be categorized as “Rejected” and “Caution”.

The decline in document and facial check is because of image integrity.

* + Image integrity compromised if the quality of the uploaded files and the content contained within them is not sufficient to perform a face comparison.

Other top root cause of document check failure is conclusive document quality. It could be due to token authentication.

# **Solution**

Revolut should improve the user experience in order to improve the KYC pass rate. The workflow for the KYC should be seamlessly easy.

**For example**: Instead of asking user to upload the document &/or photo in doc, pdf, png, jpg, jpeg or any other format, introduce a new feature in existing app which will allow user to take a picture in order to upload documents and photos.

* + For document, App should ask user to take picture of the document (Front and back side).
  + For Photo, App should ask user to take a selfie of the user.
  + By implementing proper workflow for customer onboarding, Revolut can improve the problem of “Image integrity” &/or “rejected/caution” sub result. For “caution”, When the user takes a picture of the document (both sides - Front and back), App can validate immediately if the name printed on the card is same as the name of applicant. If not, then app can ask user to upload another document.
  + By allowing user to take a picture of documents and a selfie, the chances of uploading wrong documents, not supported file format will be decreased.
  + This way the conclusive document quality reports can also be improved.

Implement the AI driven verification for better and faster results of KYC. It could improve the overall time for KYC process.

To provide best experience to user, the Revolut Product/Design team should work closely with the engineers/data scientists of Veritas to improve the workflow of onboarding the customer (KYC process).

***Data analysis code file****: Revolut\_Improving\_KYC.Rmd*