

KYC

Revolut HomeTask - 1

INCENTIVE

The aim is to determine the root causes of substantial decline of 'pass rate' in the recent period and offer an optimal solution for it.

INITIALISATION

To determine the cause of an unusual behaviour we need to analyze the dataset. There are following issues in the data in its current format -

1. Data is not sorted by date.
2. Date and time is in single column which is leading to difficulty in analysis.
3. All the elements in properties column are dictionaries which is difficult to interpret.

Our first step is data preprocessing to simplify the data for easy observation and manipulation.

DATA PREPROCESSING

Sorting the rows of the sheet
by date in ascending order.

created_at
2017-06-20T23:12:57Z
2017-06-20T23:16:04Z
2017-06-20T17:59:49Z
2017-06-20T17:59:38Z
2017-06-20T18:08:09Z
2017-06-20T18:08:50Z
2017-06-20T18:11:36Z
2017-06-20T18:18:12Z
2017-06-20T18:19:39Z
2017-06-20T18:30:27Z
2017-06-20T18:28:54Z
2017-06-20T18:40:08Z
2017-06-20T18:44:37Z
2017-06-20T18:54:08Z
2017-06-20T18:54:53Z
2017-06-20T18:57:48Z
2017-06-20T19:03:32Z
2017-06-20T19:00:27Z
2017-06-20T19:00:27Z
2017-06-20T19:06:18Z
2017-06-20T19:13:17Z
2017-06-20T19:21:34Z

Previous

created_at
0 2017-05-23T15:13:02Z
0 2017-05-23T15:18:05Z
0 2017-05-23T15:22:16Z
0 2017-05-23T15:27:06Z
0 2017-05-23T15:27:56Z
0 2017-05-23T15:28:16Z
0 2017-05-23T15:29:34Z
0 2017-05-23T15:39:52Z
0 2017-05-23T15:44:00Z
0 2017-05-23T15:47:07Z
0 2017-05-23T15:48:06Z
0 2017-05-23T15:48:21Z
0 2017-05-23T16:00:00Z
0 2017-05-23T16:04:46Z
0 2017-05-23T16:04:57Z
0 2017-05-23T16:05:25Z
0 2017-05-23T16:05:25Z
0 2017-05-23T16:19:11Z
0 2017-05-23T16:22:04Z
0 2017-05-23T16:27:10Z

Now

DATA PREPROCESSING

Date Time Split

2017-05-23T15:13:02Z



[2017-05-23, 15:13:02]

	J	
id	created_at	supplier
0	2017-05-23T15:13:02Z	
0	2017-05-23T15:18:05Z	
0	2017-05-23T15:22:16Z	
0	2017-05-23T15:27:06Z	
0	2017-05-23T15:27:56Z	
0	2017-05-23T15:28:16Z	
0	2017-05-23T15:29:34Z	
0	2017-05-23T15:39:52Z	
0	2017-05-23T15:44:00Z	
0	2017-05-23T15:47:07Z	
0	2017-05-23T15:48:06Z	
0	2017-05-23T15:48:21Z	
0	2017-05-23T16:00:00Z	
0	2017-05-23T16:04:46Z	
0	2017-05-23T16:04:57Z	
0	2017-05-23T16:05:25Z	
0	2017-05-23T16:05:25Z	
0	2017-05-23T16:19:11Z	
0	2017-05-23T16:22:04Z	
0	2017-05-23T16:27:10Z	
0	2017-05-23T16:39:09Z	
0	2017-05-23T17:01:48Z	

AA	AB
created_at_date	created_at_time
5/23/2017	15:13:02
5/23/2017	15:18:05
5/23/2017	15:22:16
5/23/2017	15:27:06
5/23/2017	15:27:56
5/23/2017	15:28:16
5/23/2017	15:29:34
5/23/2017	15:39:52
5/23/2017	15:44:00
5/23/2017	15:47:07
5/23/2017	15:48:06
5/23/2017	15:48:21
5/23/2017	16:00:00
5/23/2017	16:04:46
5/23/2017	16:04:57
5/23/2017	16:05:25
5/23/2017	16:05:25
5/23/2017	16:19:11
5/23/2017	16:22:04
5/23/2017	16:27:10
5/23/2017	16:39:09
5/23/2017	17:01:48

DATA PREPROCESSING

Splitting Properties

properties

{'gender': 'Male', 'document_type': 'driving_licence', 'date_of_expiry': '2027-01-05', 'issuing_country': 'GBR'}

{'gender': 'Male', 'nationality': 'MLT', 'document_type': 'national_identity_card', 'date_of_expiry': '2026-01-30', 'issuing_country': 'MLT'}

{'document_type': 'driving_licence', 'date_of_expiry': '2024-10-21', 'issuing_country': 'ESP'}

{'gender': 'Female', 'nationality': 'PRT', 'document_type': 'national_identity_card', 'date_of_expiry': '2021-03-20', 'issuing_country': 'PRT'}

{'gender': 'Male', 'document_type': 'driving_licence', 'date_of_expiry': '2022-11-29', 'issuing_country': 'GBR'}

{'document_type': 'driving_licence', 'date_of_expiry': '2024-06-26', 'issuing_country': 'JEY'}

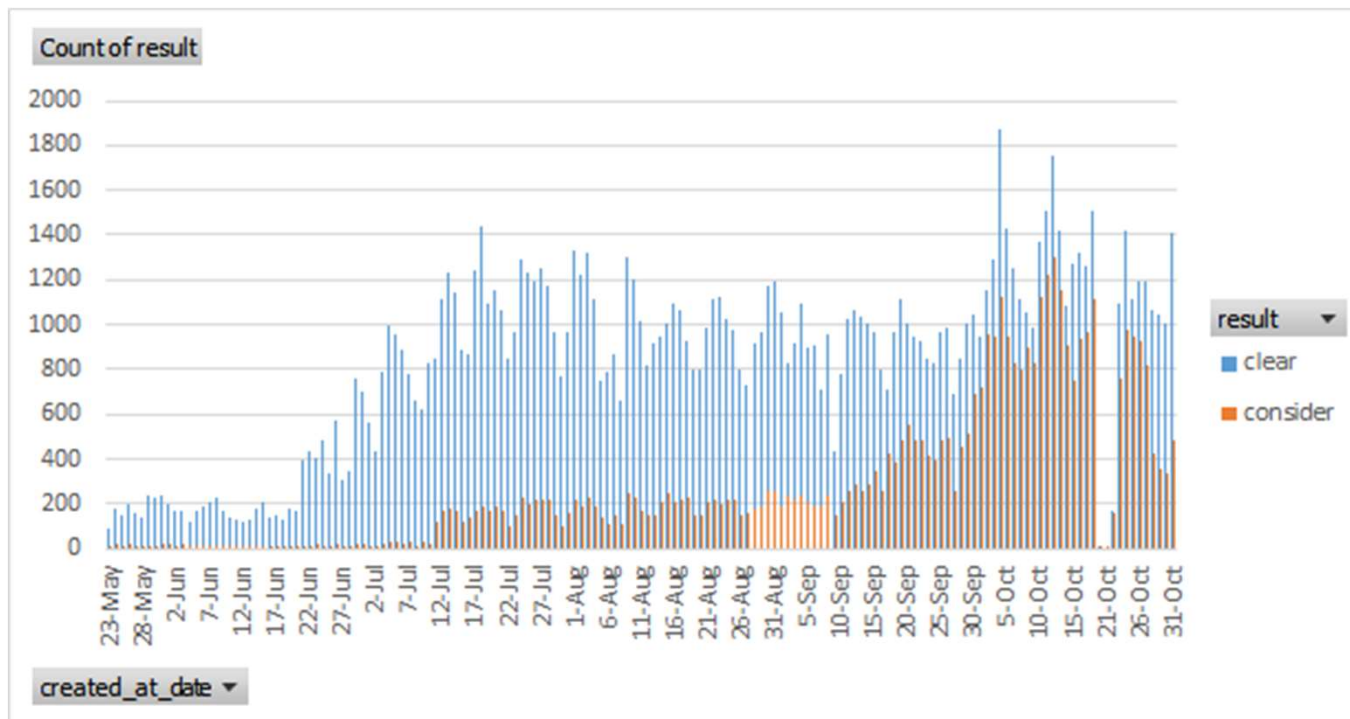
id	properties_gender	properties_nationality	properties_document_type	properties_date_of_expiry	properties_issuing_country	create
0	Male		driving_licence	1/5/2027	GBR	
0	Male	MLT	national_identity_card	1/30/2026	MLT	
0			driving_licence	10/21/2024	ESP	
1	Female	PRT	national_identity_card	3/20/2021	PRT	
0	Male		driving_licence	11/29/2022	GBR	
0			driving_licence	6/26/2024	JEY	

FINDING UNUSUAL BEHAVIOUR

Since we know that the 'pass rate' has been substantially decreased, there must be some unusual trends in the data which needs to be revealed so that we could get some clues regarding the cause of such behaviour.

The first step is to plot the trend of the column 'result' with 'time' in the excel sheet doc_report.csv.

RESULT VS TIME



RESULT VS TIME

From the graph we can observe two unusual trends in 'considerable results' relative to 'clear results'.

1. Sudden increase in considerable results from 10-July
2. Unexpected rise in the month of october

So now, it can be verified from the plot that the 'pass rate' has significantly decreased with time recently.

Now, since trend in overall result has been analyzed, we have to proceed further to visualize the trend in different category of results with time to check which categories are actually affecting the result.

After visualizing each category of results, unusual behaviour in some of them were observed which are shown from the next slide.

IMAGE INTEGRITY RESULT VS TIME

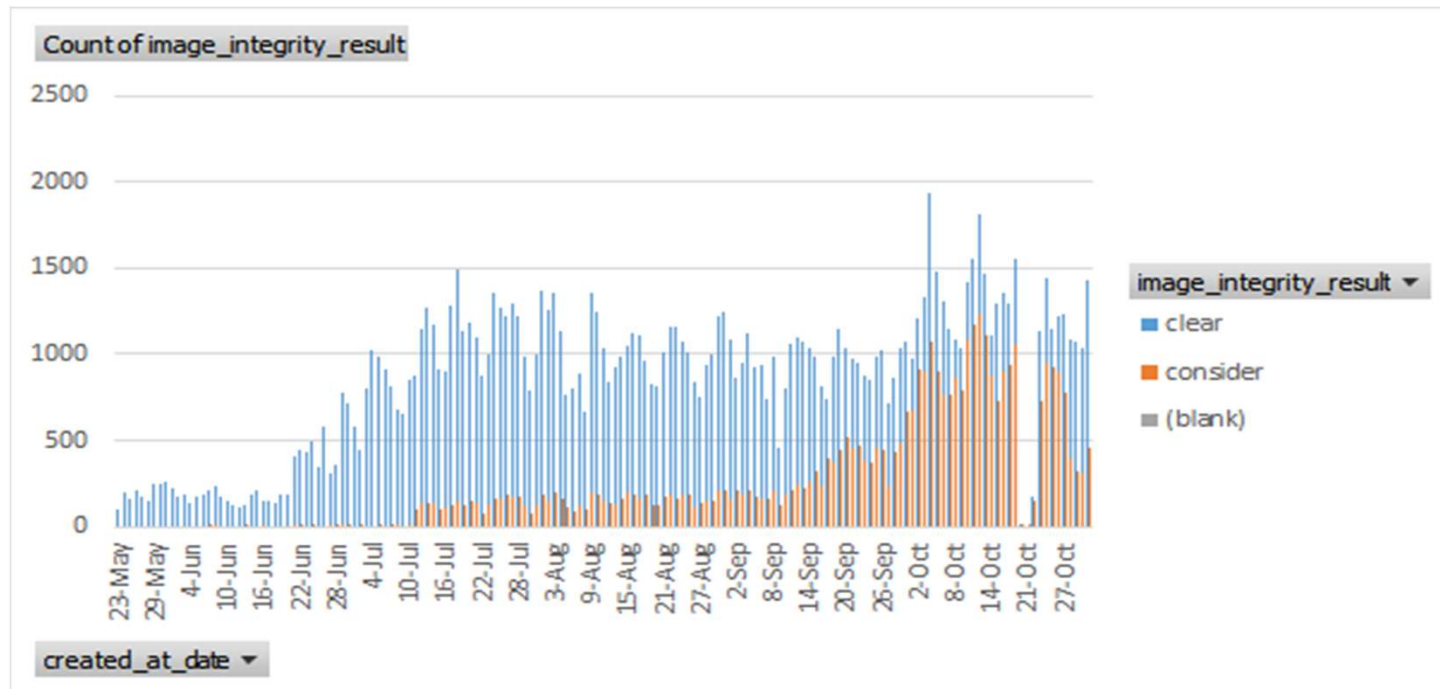
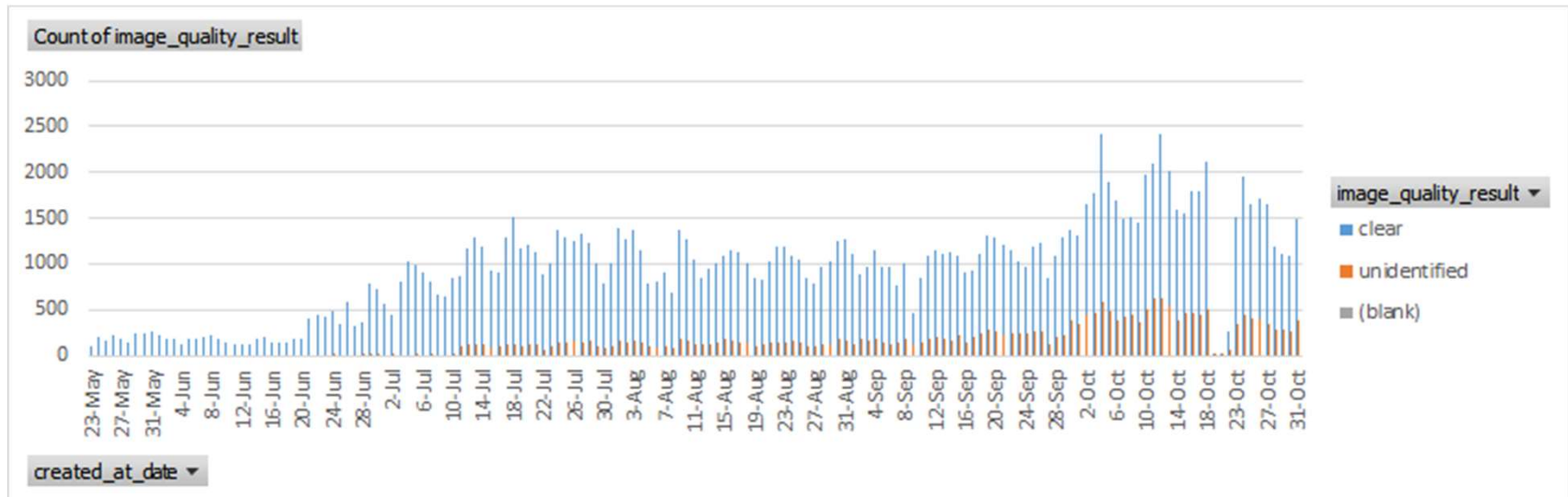
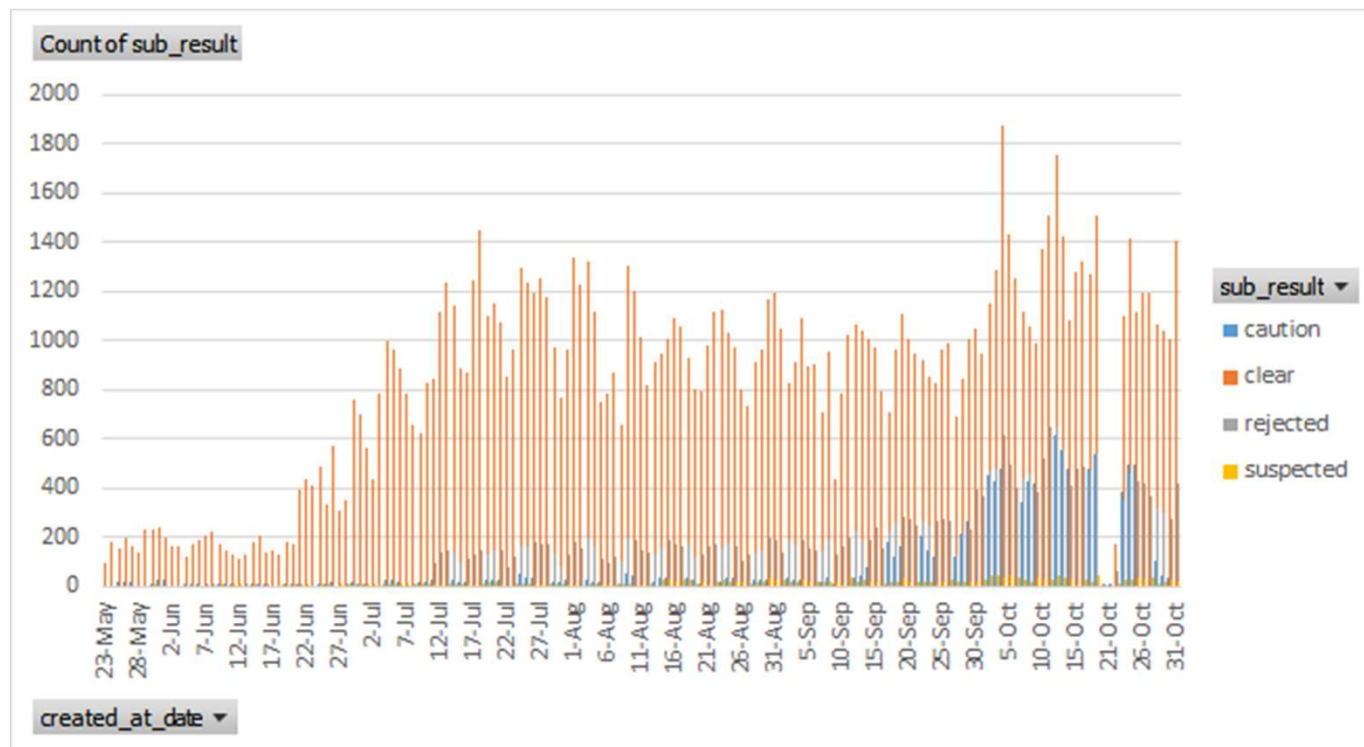


IMAGE QUALITY RESULT VS TIME



SUB RESULT VS TIME

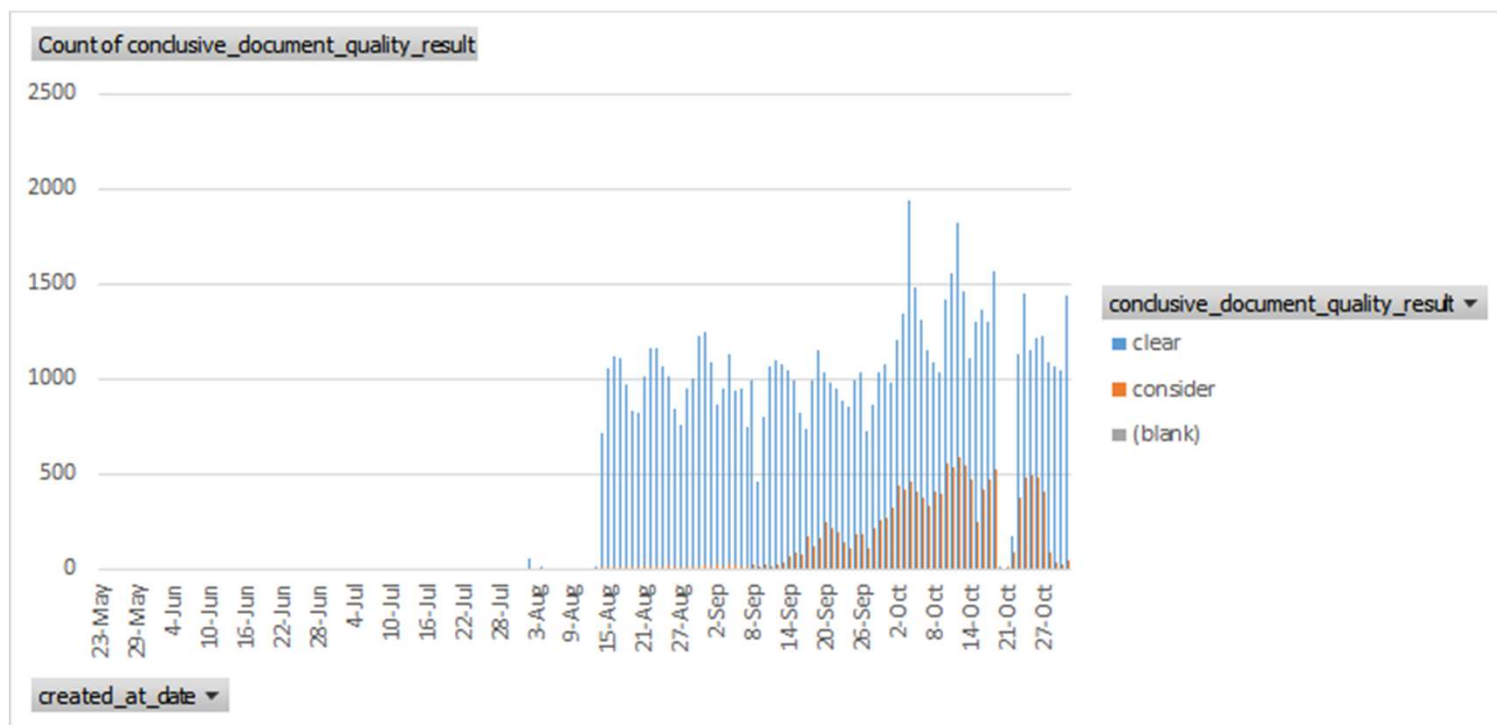


ANALYSIS

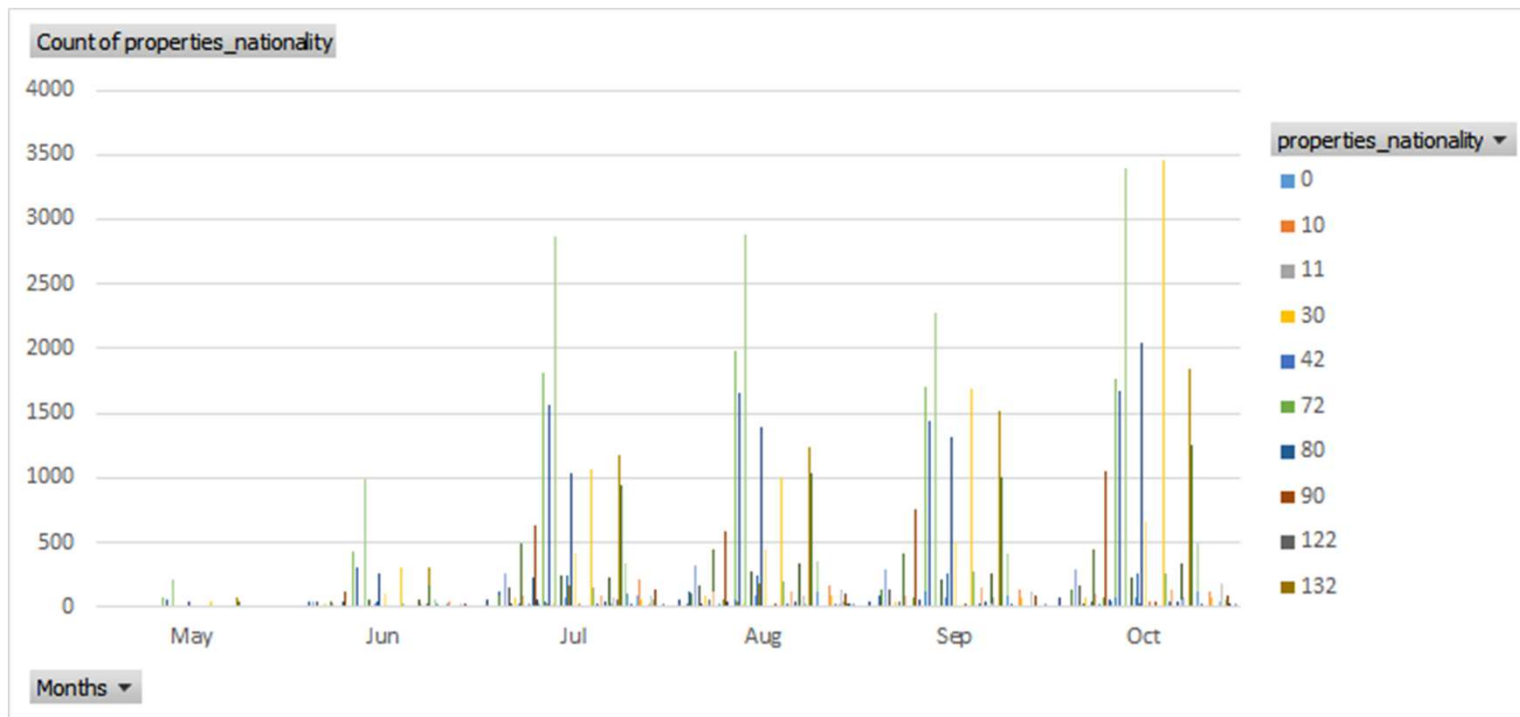
The root cause has affected the image integrity, image quality and sub results in the same way as it has affected the overall result.

THE SECOND TREND

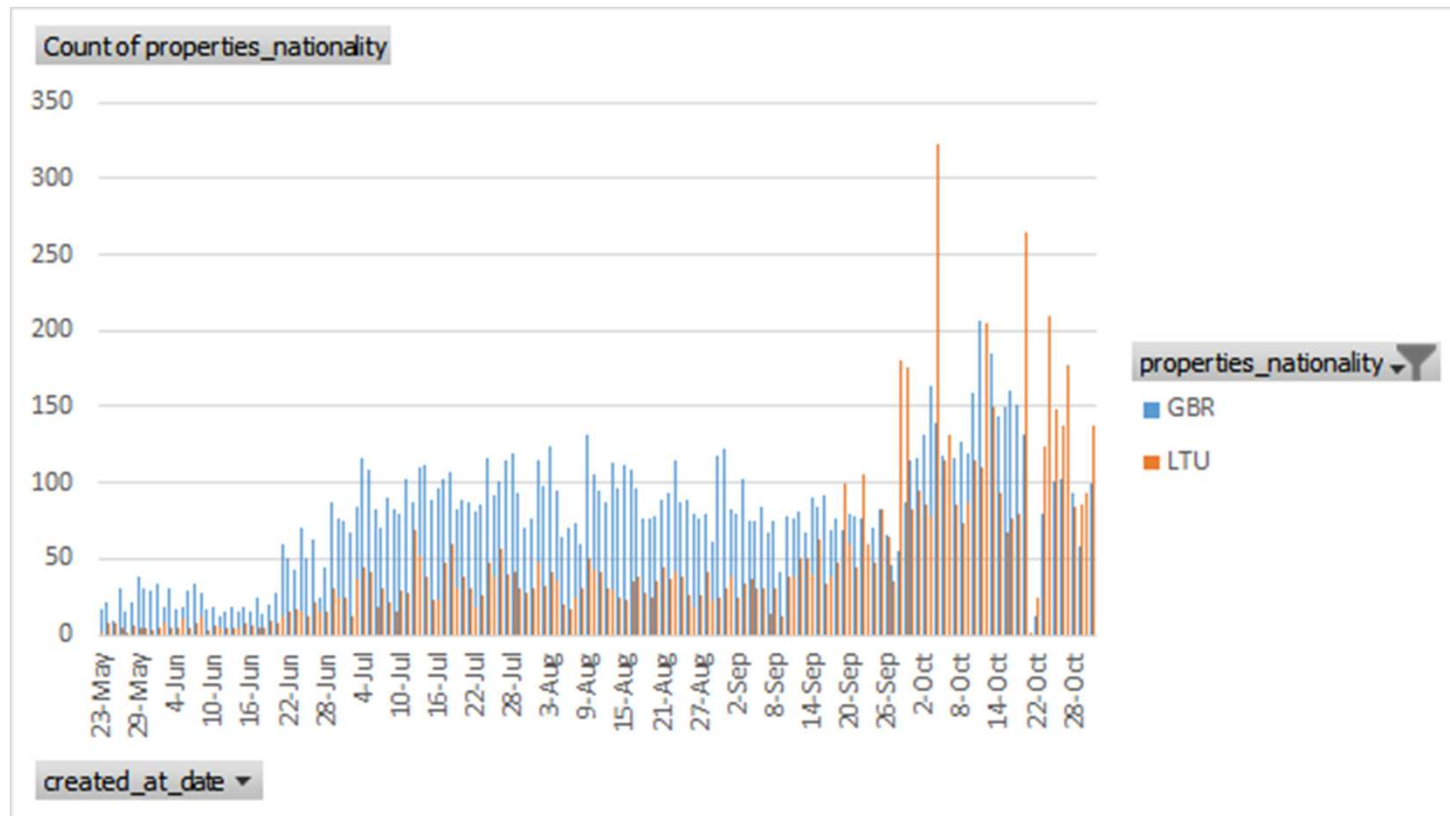
CONCLUSIVE DOCUMENT QUALITY RESULT VS TIME



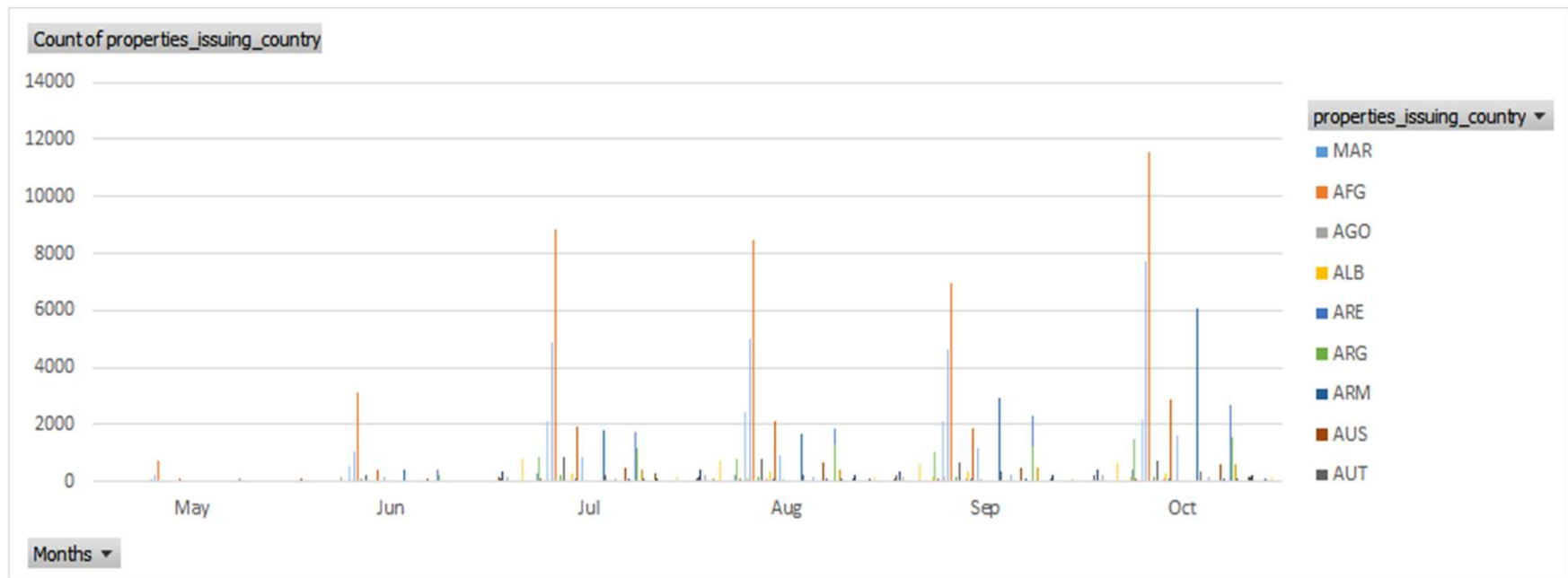
PROPERTIES NATIONALITY VS TIME



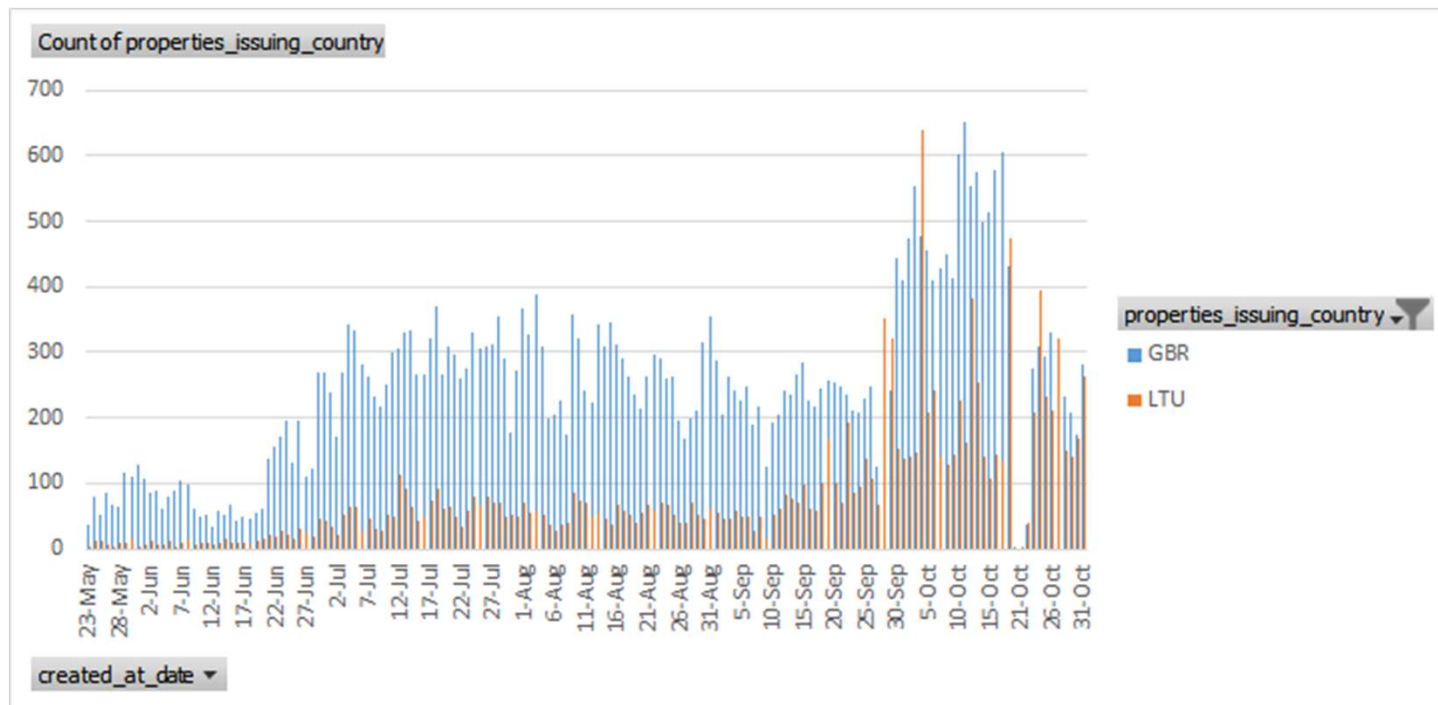
NATIONALITY (GBR & LTU) VS TIME



ISSUING COUNTRY VS TIME



ISSUING COUNTRY (GBR & LTU) VS TIME



ANALYSIS

As we know that there were two unusual trends observed in the result out of which one trend was the significant rise in the month of october. This trend has been observed in the past slides shown. From those slides following points can be concluded -

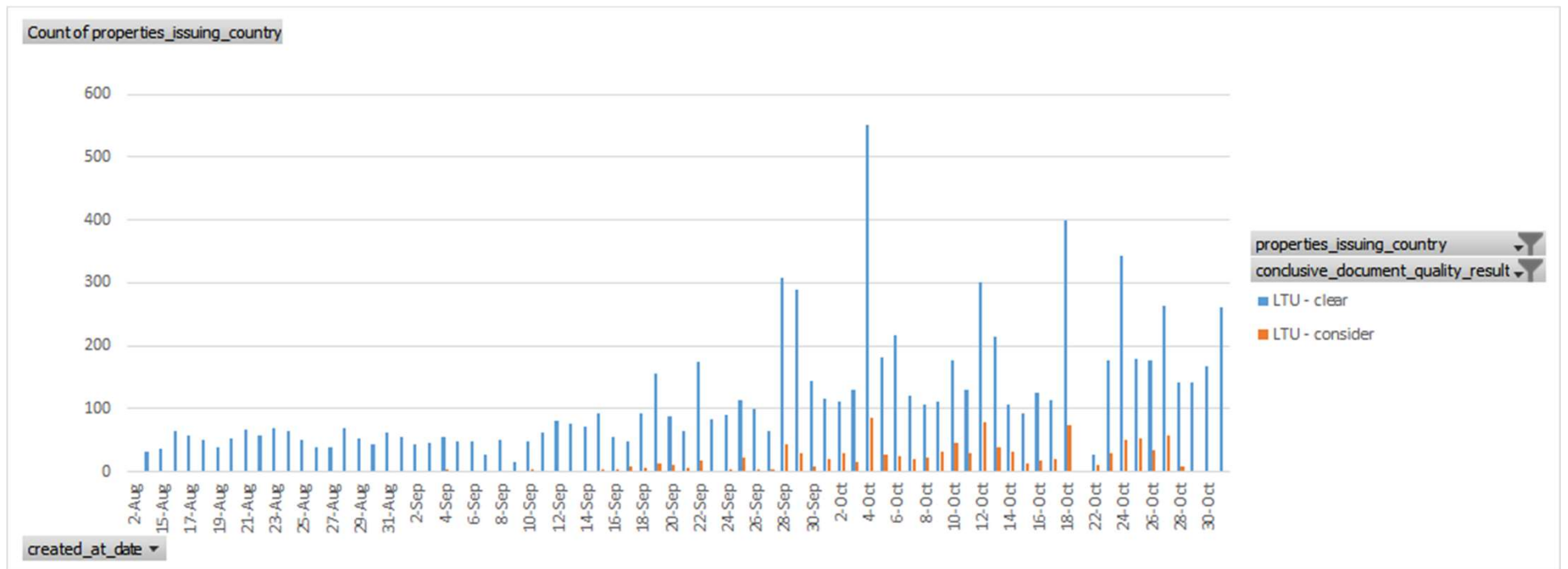
1. Conclusive document quality test has been very active since middle of the august. This test has shown significant rise in 'considerable' results in the month of october which could be the primary cause of the second trend of the overall result.
2. In both 'Issuing country' and 'Properties nationality' GBR and LTU were the only countries showing the trends equivalent to the second trend of the overall result.

ANALYSIS

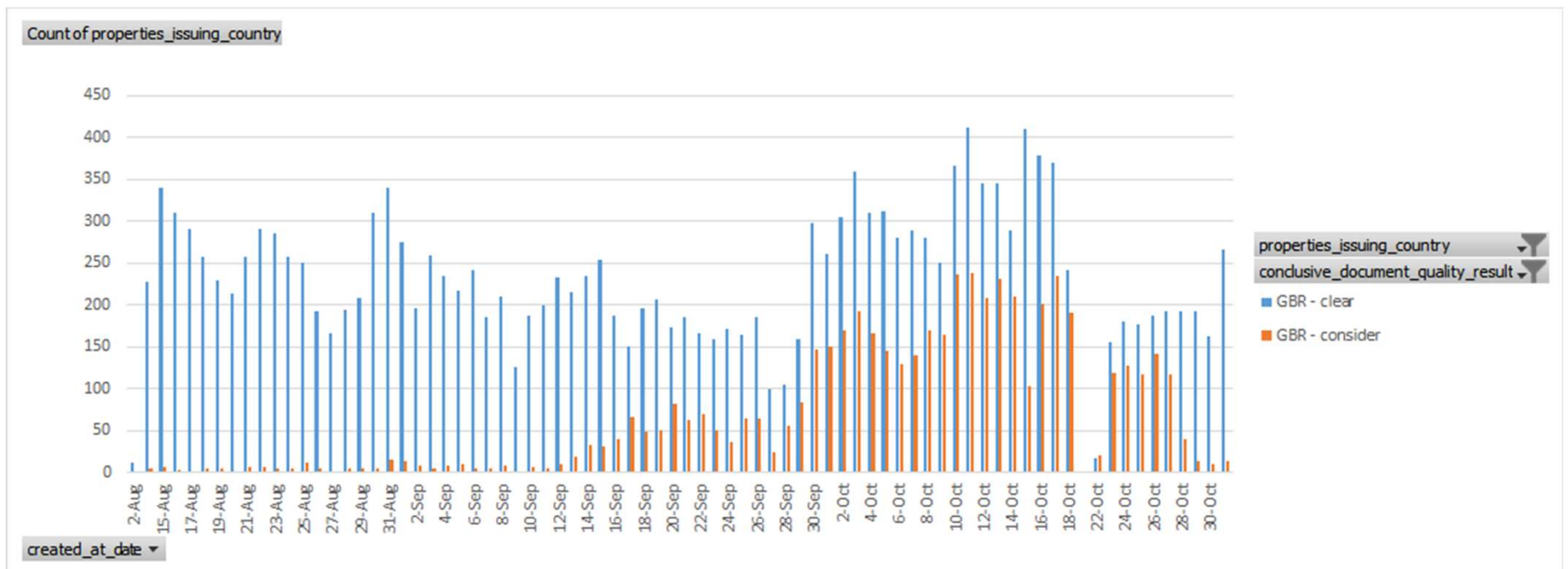
These two conclusions lead to the fact that 'GBR' and 'LTU' is failing in the 'Conclusive document quality' test which then leads to non-clear 'image integrity result'.

We have to verify this phenomenon by plotting and analyzing the trend between two countries and their Conclusive document quality results. This plot is shown in the next two slides.

CONCLUSIVE DOCUMENT RESULT OF LTU VS TIME



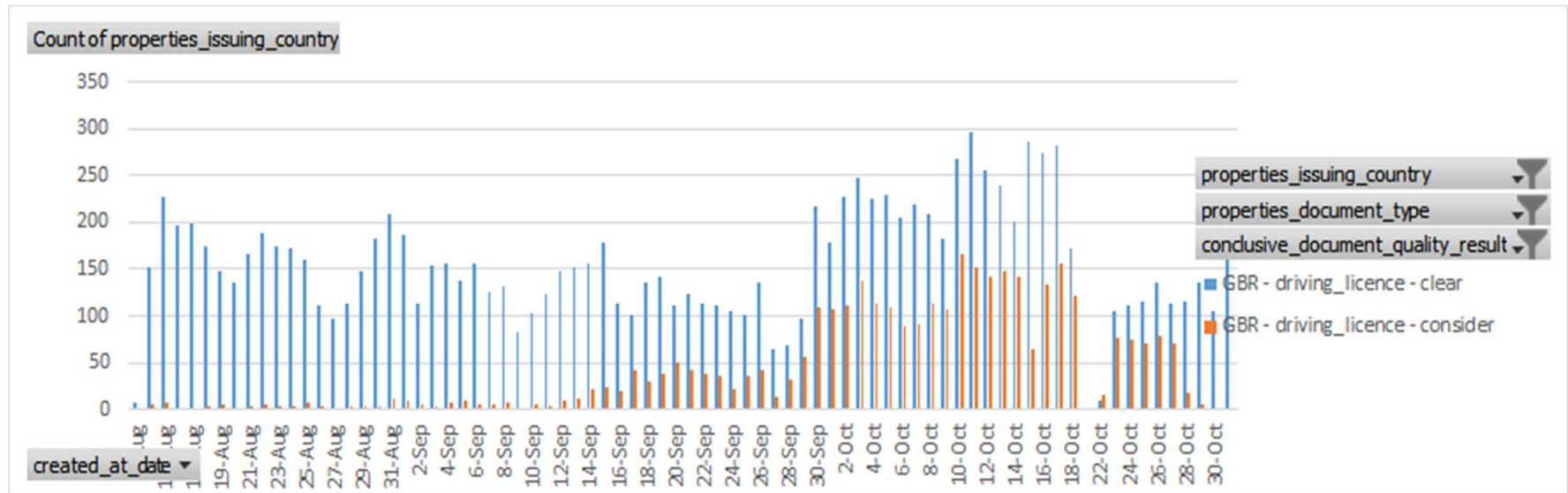
CONCLUSIVE DOCUMENT RESULT OF GBR VS TIME



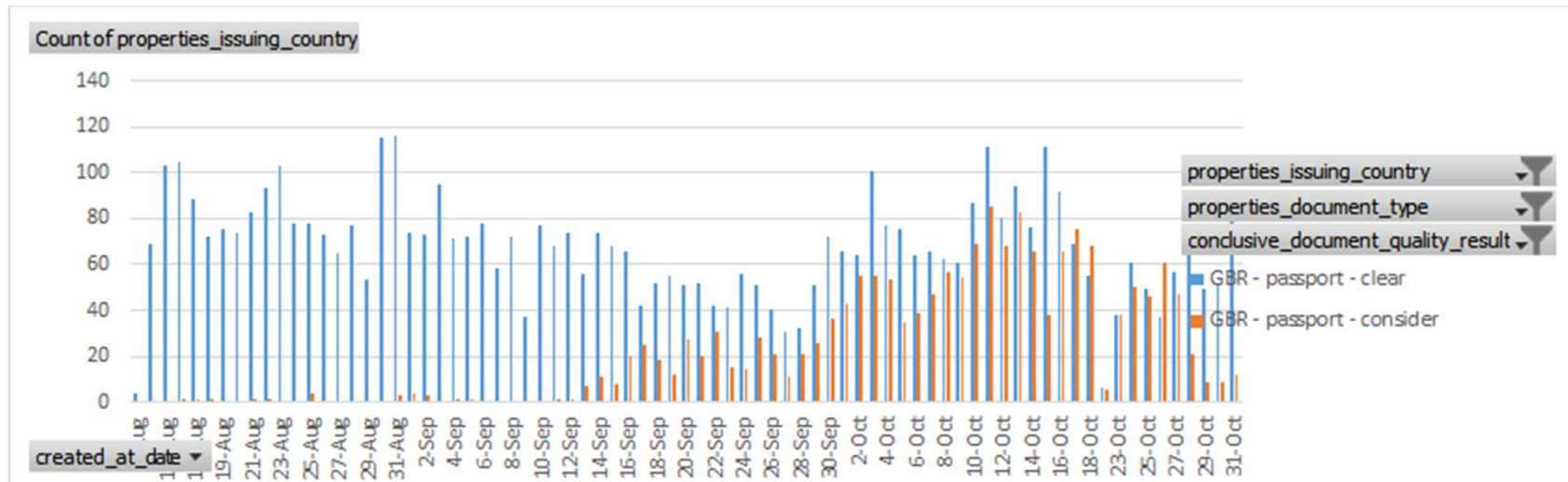
ANALYSIS

The expected trend was very strongly observed in GBR than 'LTU'. Hence, now we know that the second trend in the overall result was due to the 'document quality' issue observed in GBR which also affected the 'image quality', 'image integrity' and Sub results in them. Now all the documents used by GBR were analyzed in which some showed the expected trend. Those are plotted in the next two slides.

GBR DRIVING LICENCE QUALITY RESULT VS TIME



GBR PASSPORT QUALITY RESULT VS TIME

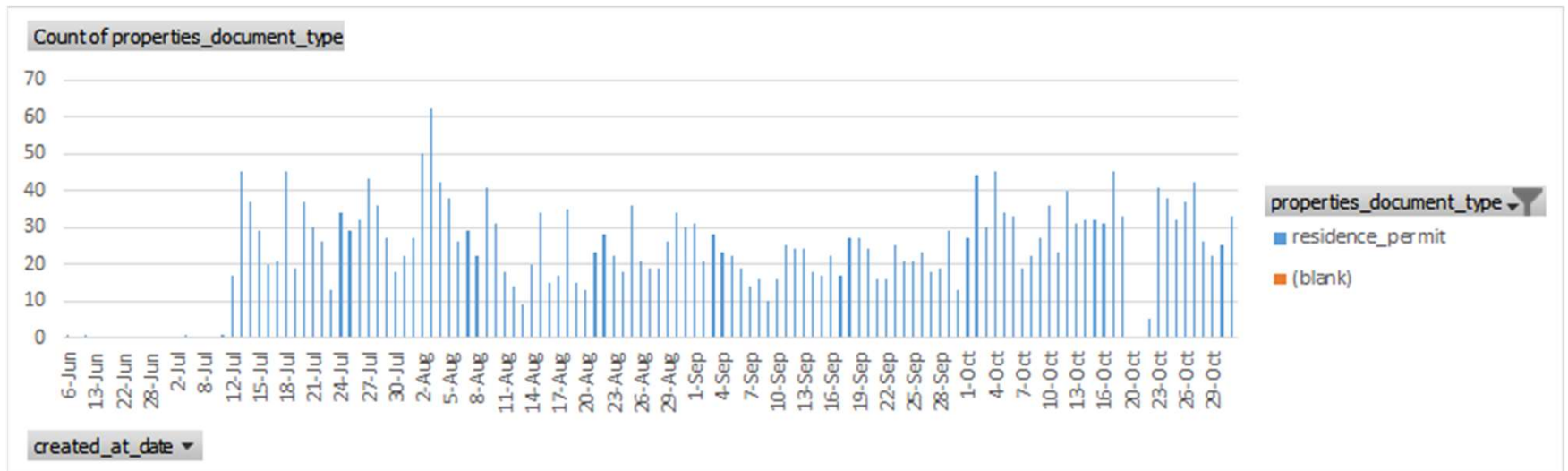


SOLUTION

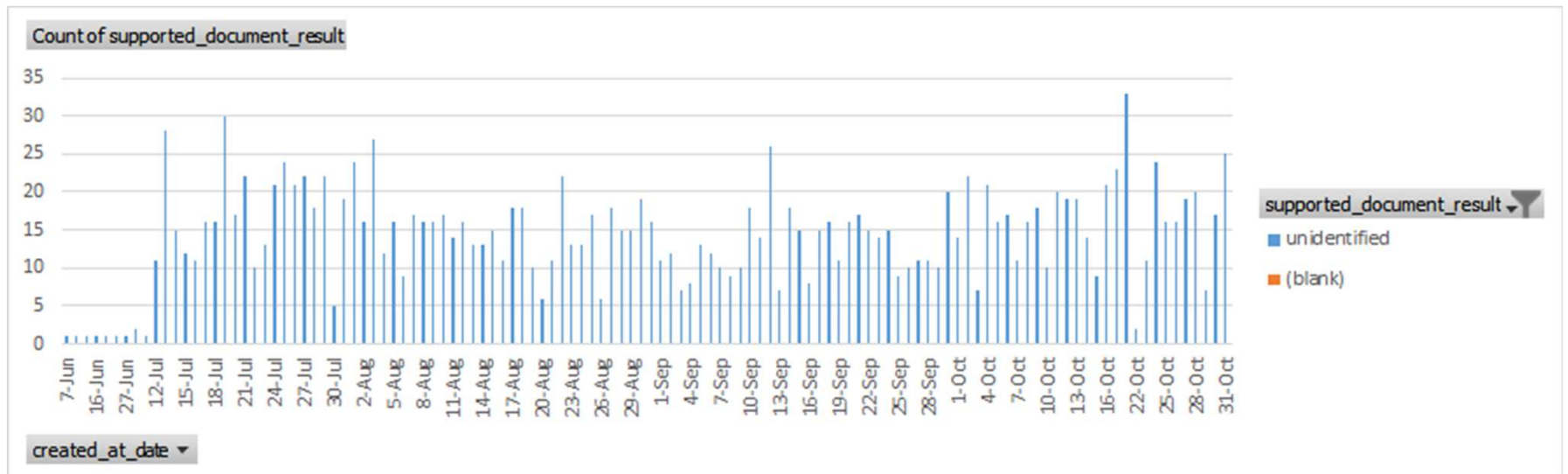
Since we saw that GBR is responsible for the recent decrease in pass rate who are using two main documents i.e. driving licence and passport. And those two documents have are suspected to have non-clear image quality. Therefore, it is suggested that the KYC applied by people in GBR should get their documents strictly verified, either manually or by a secure way.

THE FIRST TREND

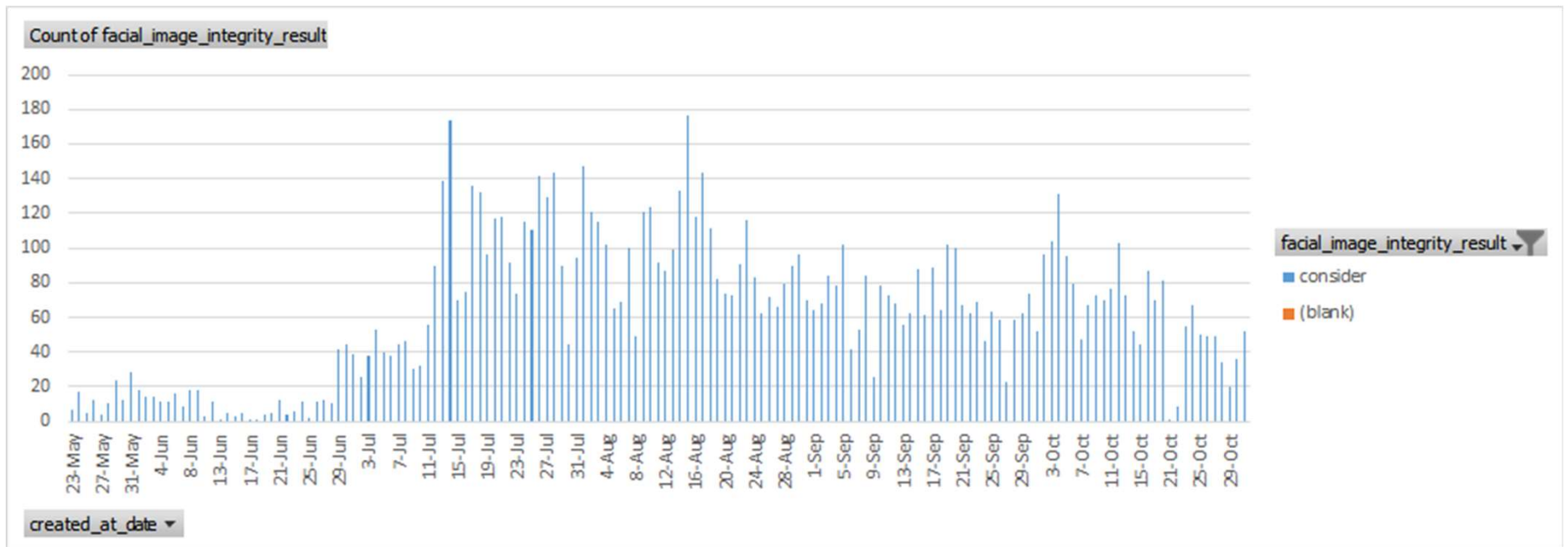
DOCUMENT TYPE VS TIME



SUPPORTED DOCUMENT RESULT VS TIME



FACIAL IMAGE INTEGRITY RESULT VS TIME

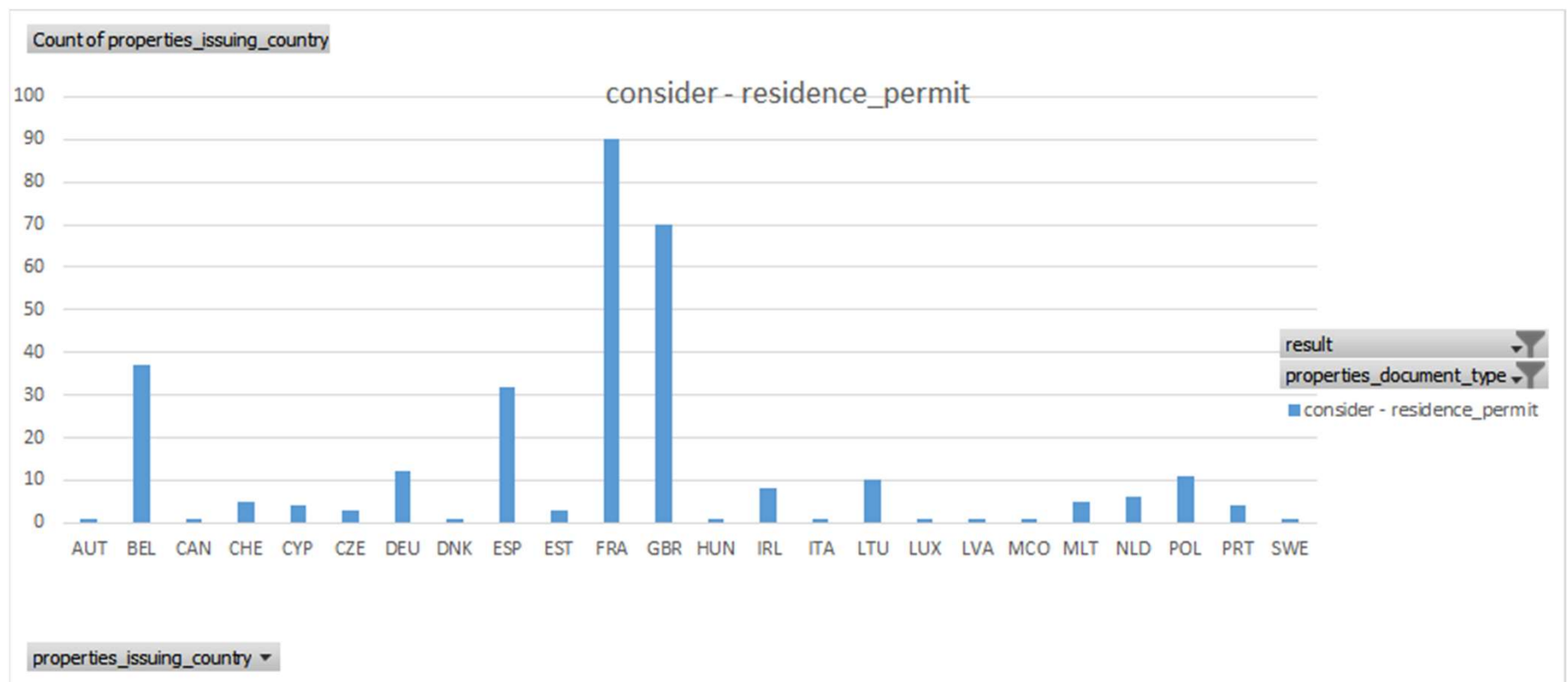


ANALYSIS

Here, we can observe the first unusual trend of the overall result which is the abrupt increase in considerable cases since 10 July. We can now say that 'residence permit' and 'supported document result' are the key factors of the abrupt change.

Now, it is required that we see the countries which used 'residence permit' as their document and analyse them.

RESIDENCE PERMIT RESULT VS ISSUING COUNTRY

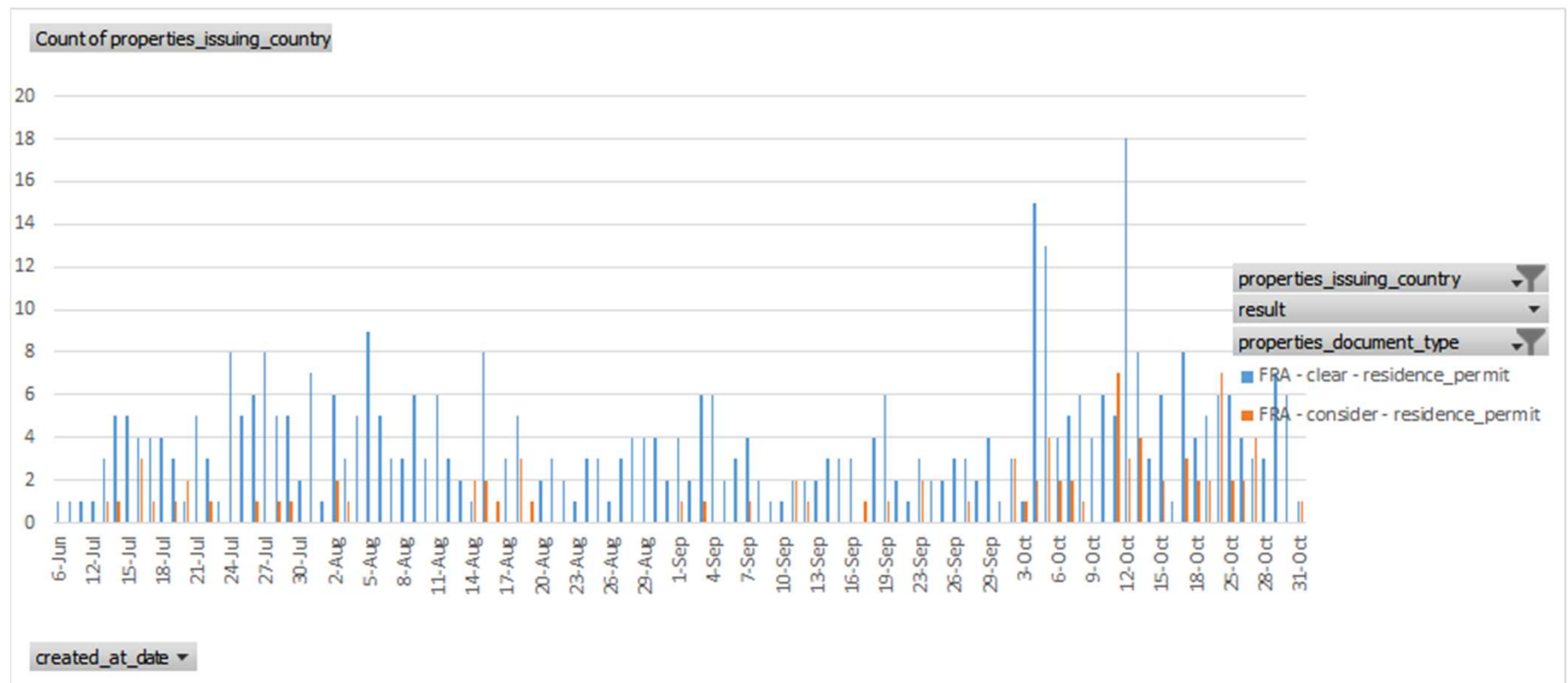


ANALYSIS

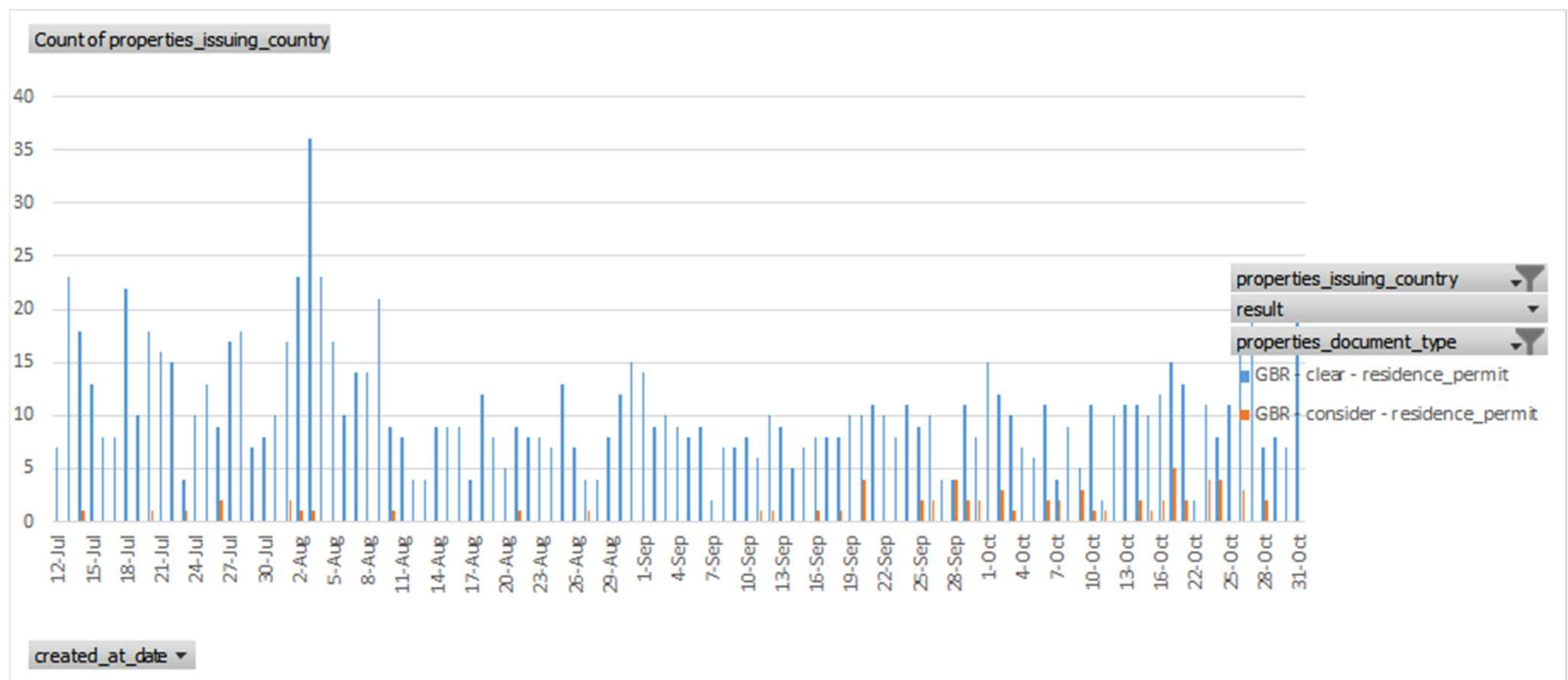
From the graph, we can observe that the residence permit was used primarily in two countries i.e. GBR and FRA. Also, we have seen how GBR was responsible for second trend in the overall result. This makes our prediction more stronger.

Now it plausible to plot the trend of these two countries using 'residence permit' as document. This has been done in the next two slices.

FRA RESIDENCE PERMIT RESULT VS TIME



GBR RESIDENCE PERMIT RESULT VS TIME



ANALYSIS

Therefore, from these trends we can conclude that GBR and FRA people who use 'residence permit' as the document have considerably unclear facial image integrity which is creating the first trend in the overall result of the KYC process.

SOLUTION

Therefore, it is advised that KYC applied using 'residence permit' should be strictly checked for image integrity either manually or with state-of-the-art ways especially by people from GBR and FRA to avoid any further issues.

THANK YOU