



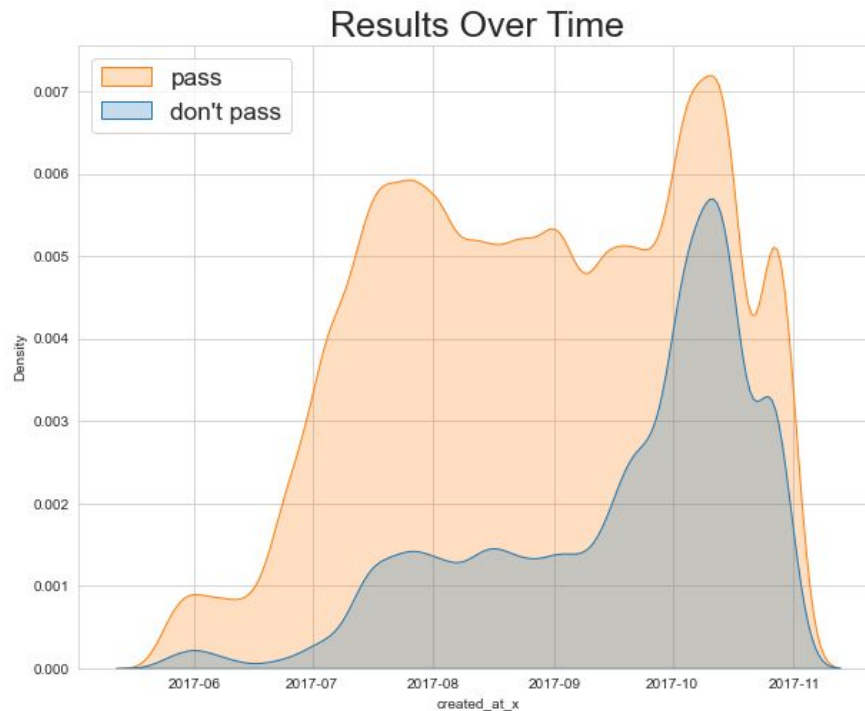
Data and Analytics Test Project

Improving KYC

The problem

Low Pass Rate

There is a significant decrease in the pass rate of Know Your Customer (KYC) process.



Analysis Road map



Step 1

Discover if the decrease in pass rate exists only in a single report or in both reports

Step 2

Find variables with the highest correlation to the not pass result

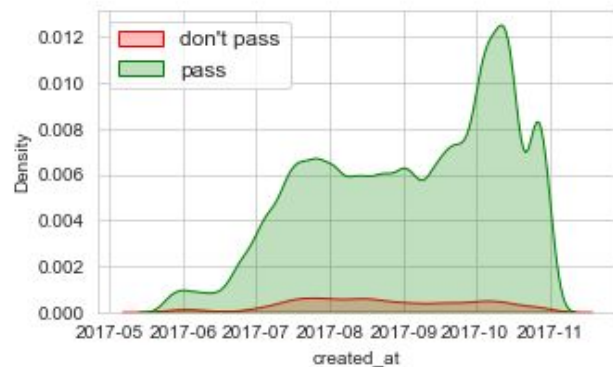
Step 3

A correlation analysis will be conducted with each of these variables to understand if there is any link between them and certain properties

Results over time in the Documents Report



Results over time in Facial Similarity Report



The decrease in pass rate is associated with the Documents Module as there is no significant decrease detected in the Facial Similarity Module



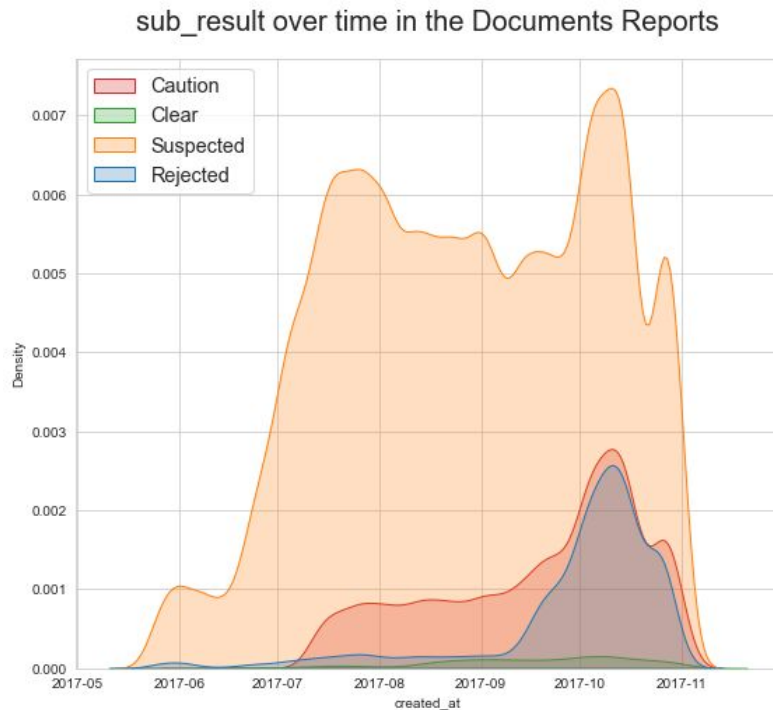
Deep-dive into Document Module

Detection of the variable associated with not pass

sub_result

It is a variable that contains 4 different values.. Clear, Caution, Suspected, and Rejected. The first will result to a pass and the rest won't.

There is an increase in the rate of cautions and rejections

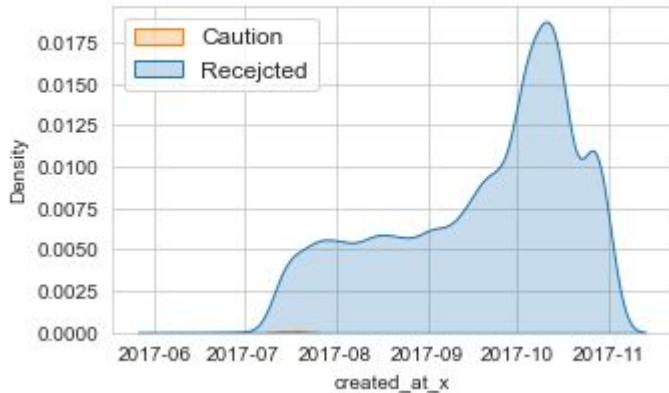


Rejections reasons

```
In [122]: 1 # Calculating the proportion of the rejection to have missing properties..  
          2 (df.query('sub_result == "rejected").properties_y == "{}").mean()
```

```
Out[122]: 0.9994633959371406
```

Properties with missing values over time



99.94% of the rejections are associated with the properties value being missing.

Proportion of cautions for every document type ..

document_type	
birth_certificate	NaN
driving_licence	0.107304
national_identity_card	0.075057
passport	0.146964
residence_permit	0.083988
tax_id	NaN
voter_id	NaN
work_permit	0.058824

Proportion of cautions
for every issuing
country that had more
than 100 visitors ..

	Country	rate
17	GBR	0.143973
29	MAR	0.143617
1	AUT	0.142287
23	IRL	0.129333
24	ITA	0.127995
30	MEX	0.123810
28	LVA	0.120607
22	IND	0.116105
12	DNK	0.113360
11	DEU	0.108706
36	PHL	0.107143
5	CAN	0.105263

Proportion of cautions
for every Nationality
that had more than
100 visitors ..

	Country	rate
1	AUT	0.184154
44	TUN	0.180952
29	MAR	0.175214
34	NOR	0.170543
12	DNK	0.167421
18	GBR	0.154442
17	FRA	0.152489
23	IRL	0.148678
16	FIN	0.142857
28	LVA	0.142544
33	NLD	0.132293
25	JPN	0.128713

Cautions reasons

After conducting multiple tests, there is no clear relation between any property to the cautions and it seems distributed between all properties. There is high chance that it is related to other properties that aren't being collected (e.g. race).

Summary

Of Findings

- The Facial Similarity module is working fine.
- The Documents module is the one with the issues and its final result is highly connected to the column of sub_results showing clear, caution, rejected, and suspected.
- There is an uptrend in the caution and rejected outputs that is causing a significant decrease in the pass rate over time.
- The rejections are highly connected to both the data and the properties being missing.
- There is a clear correlation between the cautions and the conclusive_document_quality_result column outputting consider.
- There is no clear relation between the cautions and the properties available and I suspect that there might be related to non-collected parameters (e.g. race).
- I've found that there are too many invalid country codes in the properties but it seems that it doesn't cause any harm.