Asylum Applications in the EU

Monday Oct 16th, 2023

Alexander Portland

Brown University

https://github.com/AlexanderPortland/migr

Asylum

The protection granted by a country to someone who has left their native country as a political refugee

How many people will get denied asylum next year? Why?

(a Regression Question)

Eurostat Asylum Decision Dataset
'Migr_asydcfstq'

- Citizenship
- Sex
- Age Range
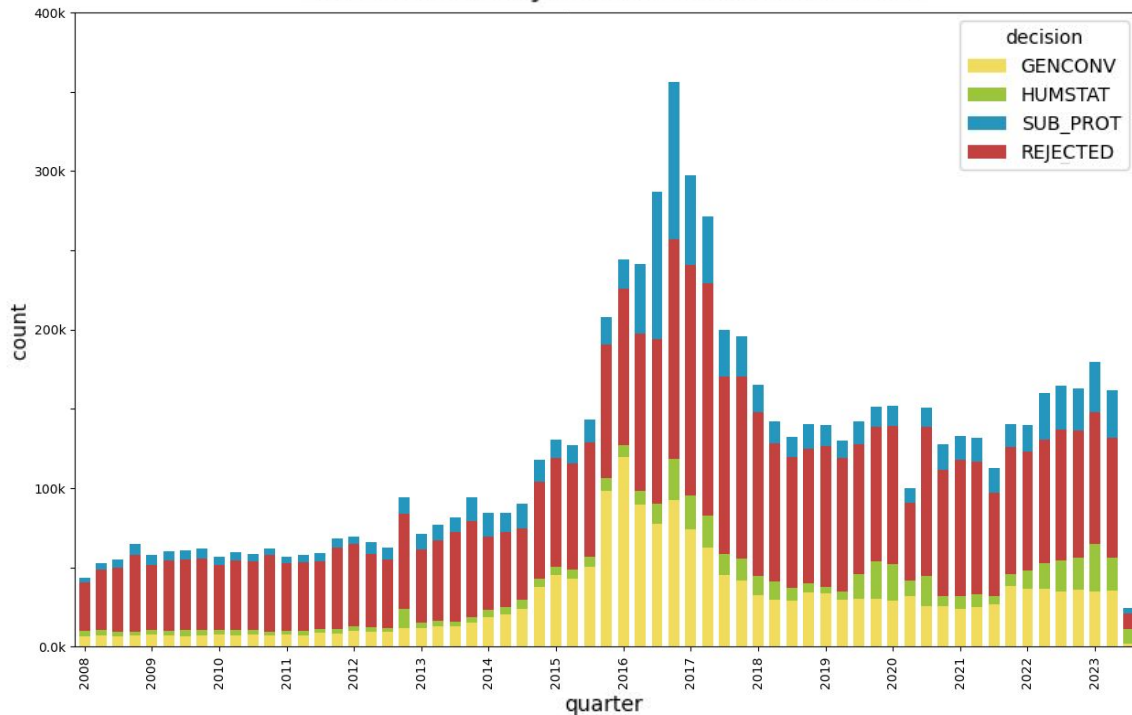- Country of Asylum
- Timing of Application (by quarter)

———

https://github.com/AlexanderPortland/migr

EDA

The Dataset

'Migr_asydcfstq'

	citizen	sex	age	decision	geo	TIME_PERIOD	OBS_VALUE	OBS_FLAG
219398	AE	F	Y_LT18	REJECTED	UK	2015-Q3	5	<NA>
233218	AE	M	Y14-17	HUMSTAT	UK	2011-Q1	5	<NA>
243119	AE	M	Y18-34	REJECTED	UK	2015-Q3	5	<NA>
247067	AE	M	Y35-64	GENCONV	UK	2014-Q1	5	<NA>
251026	AE	M	Y35-64	REJECTED	UK	2015-Q3	5	<NA>

measure of asylum outcomes over time

decision
GENCONV
HUMSTAT
SUB_PROT
REJECTED

GENCONV

Asylum granted by the 1951 Geneva Convention for war, conflict, and persecution

HUMSTAT

Humanitarian Status Asylum granted for disasters, oppression, emergency medical issues, or other similar reasons

SUB_PROT

Subsidiary Protection granted to those who don't qualify as a refugee, but would face serious harm if returned to their home country

REJECTED

Asylum was not granted

TOTAL_POS

TOTAL_APPS

https://github.com/AlexanderPortland/migr

The Dataset

'Migr_asydcfstq'

asylum outcomes by age

GENCONV

Asylum granted by the 1951 Geneva Convention for war, conflict, and persecution

HUMSTAT

Humanitarian Status Asylum granted for disasters, oppression, emergency medical issues, or other similar reasons

SUB_PROT

Subsidiary Protection granted to those who don't qualify as a refugee, but would face serious harm if returned to their home country
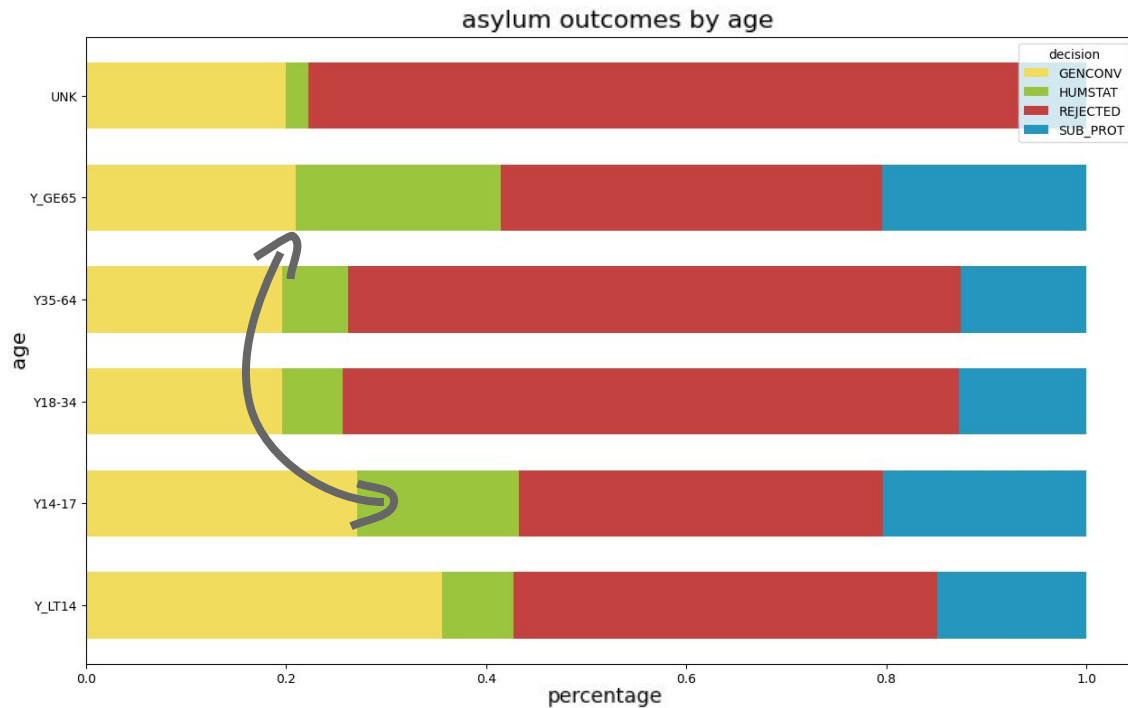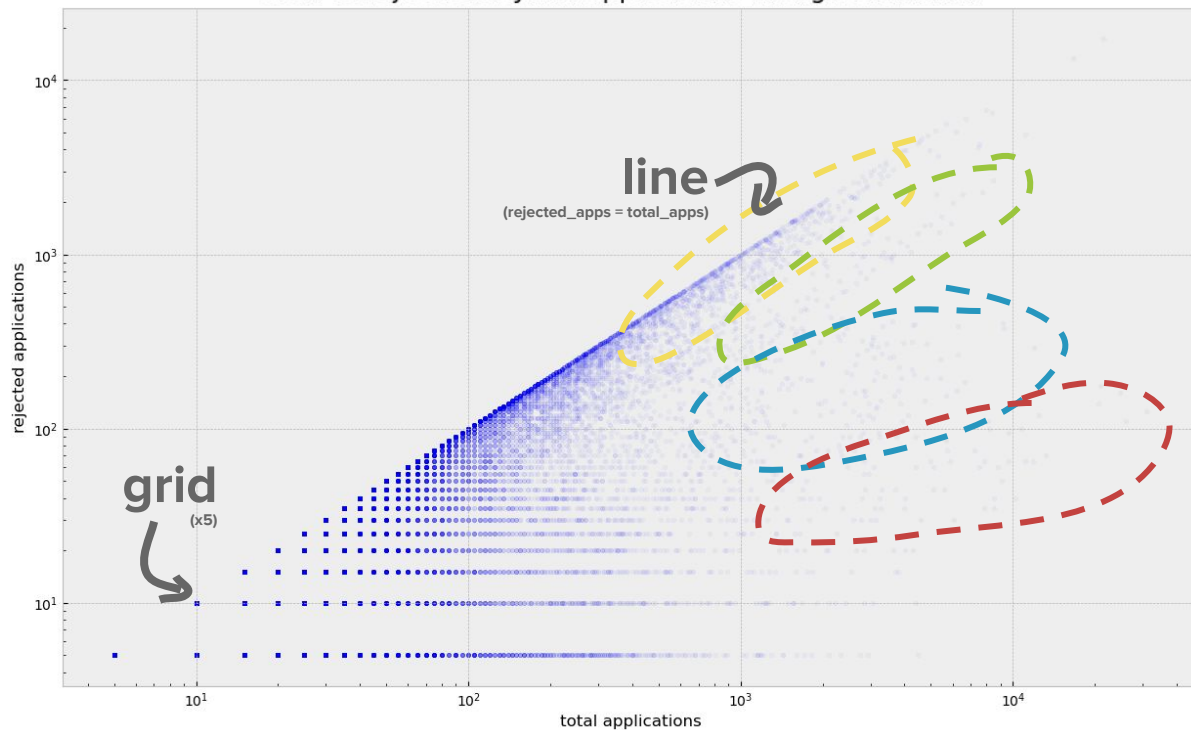
REJECTED

Asylum was not granted

TOTAL_POS

TOTAL_APPS

The Target Variable
Rejected Applications

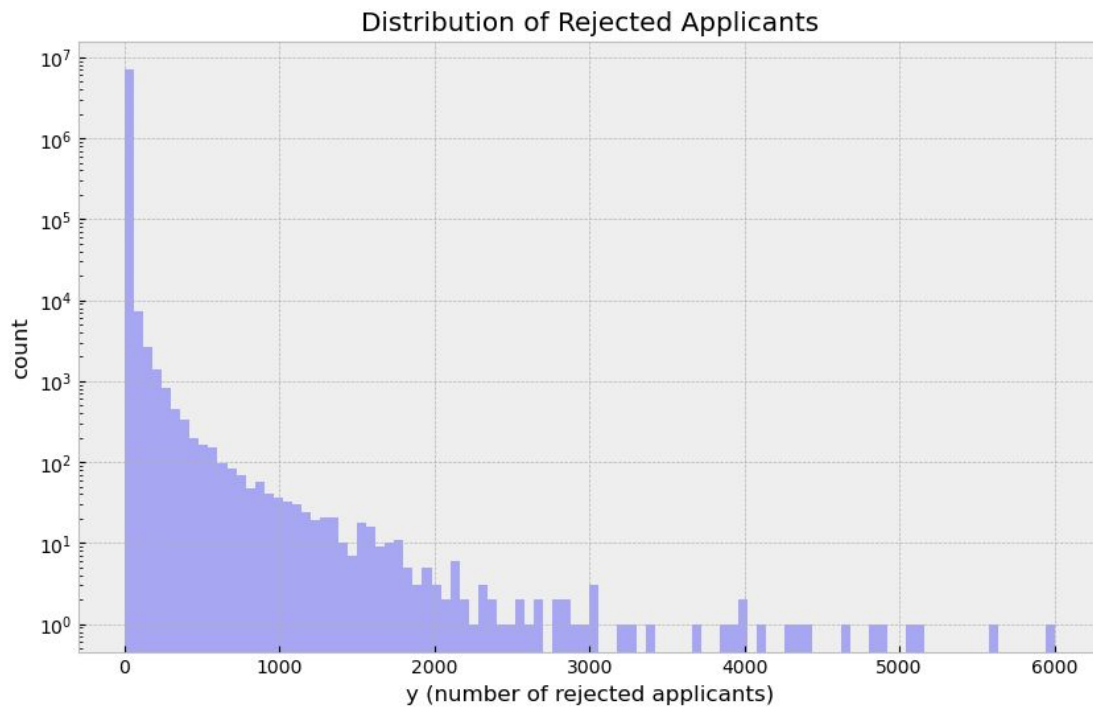total vs rejected asylum applications in a given month

Splitting

How to Split?

Distribution of Rejected Applicants

Stratified Split
With respect to rejected applicants

Logarithmic Bins

Distribution of Rejected Applicants in Split Data
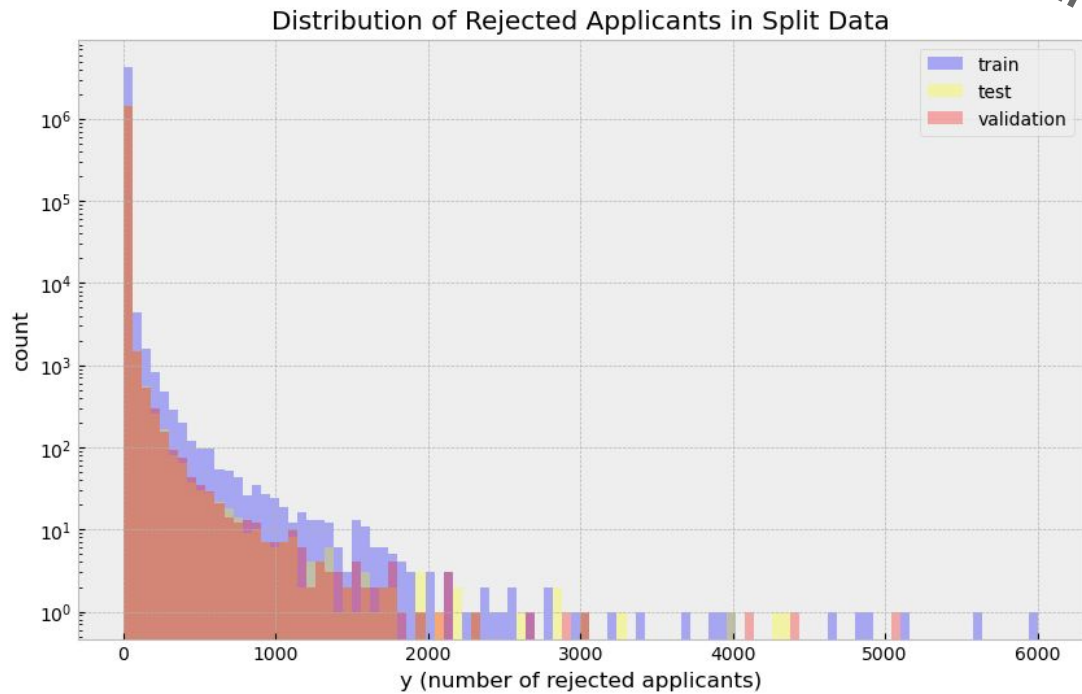
ONE-HOT ENCODED

Country of Citizenship (203 possible values)

Sex [Male, Female, Unknown]

Country of Asylum (33 possible values)

ORDINALLY ENCODED

Time Period (quarterly 2008-Q1 to 2023-Q3)

Age [<14, 14-17, 18-34, 35-64, >=65]

Preprocessing

AFTER

241 features

still 4.3 million data points

	citizen	sex	age	geo	TIME_PERIOD	TOTAL_APPS
3812586	LY	F	Y18-34	MT	2011-Q2	0
3982589	ME	UNK	Y_GE65	ME	2022-Q1	0
5359784	RU	M	Y_LT14	FI	2019-Q3	25
2694593	HU	UNK	Y14-17	CY	2013-Q3	0
1770258	DZ	UNK	Y14-17	IS	2022-Q1	0
...	...	...	...	...	...	...

BEFORE

6 features

4.3 million data points for training

1.4 for test & validation

Preprocessing

Missing Values??

yes :(almost 25% of rows have one

Missing Values by Column:
0 REJECTED
3k GENCONV
1.5m HUMSTAT
3k SUBPROT

	citizen	sex	age	geo	TIME_PERIOD	GENCONV	HUMSTAT	SUB_PROT	REJECTED
62	AD	F	UNK	BE	2008-Q1	0	\<NA\>	0	0
63	AD	F	UNK	BE	2008-Q2	0	\<NA\>	0	0
64	AD	F	UNK	BE	2008-Q3	0	\<NA\>	0	0
65	AD	F	UNK	BE	2008-Q4	0	\<NA\>	0	0
66	AD	F	UNK	BE	2009-Q1	0	\<NA\>	0	0
...	...	...	...	...	...	...	...	...	...
7220996	ZW	UNK	Y_LT14	SI	2022-Q2	0	\<NA\>	0	0
7220997	ZW	UNK	Y_LT14	SI	2022-Q3	0	\<NA\>	0	0
7220998	ZW	UNK	Y_LT14	SI	2022-Q4	0	\<NA\>	0	0
7220999	ZW	UNK	Y_LT14	SI	2023-Q1	0	\<NA\>	0	0
7221000	ZW	UNK	Y_LT14	SI	2023-Q2	0	\<NA\>	0	0

1584264 rows × 12 columns

https://github.com/AlexanderPortland/migr

Thank You