

Pattern Recognition Proposal Presentation

GROUP 3

2391007 MINJEONG KIM, 2391016 SUAHN LEE,
2391017 YOONJI LEE

Table of Contents

PATTERN RECOGNITION
TEAM PROJECT

01

Data Exploration

Interesting Facts that we discovered

02

Data Visualization & Data Preprocessing

03

Model Proposal

This data is from



PORTUGAL

< Evidence why the data is from Portugal >

1. 'euribor3m'

Euribor: Euro Interbank Offered Rate

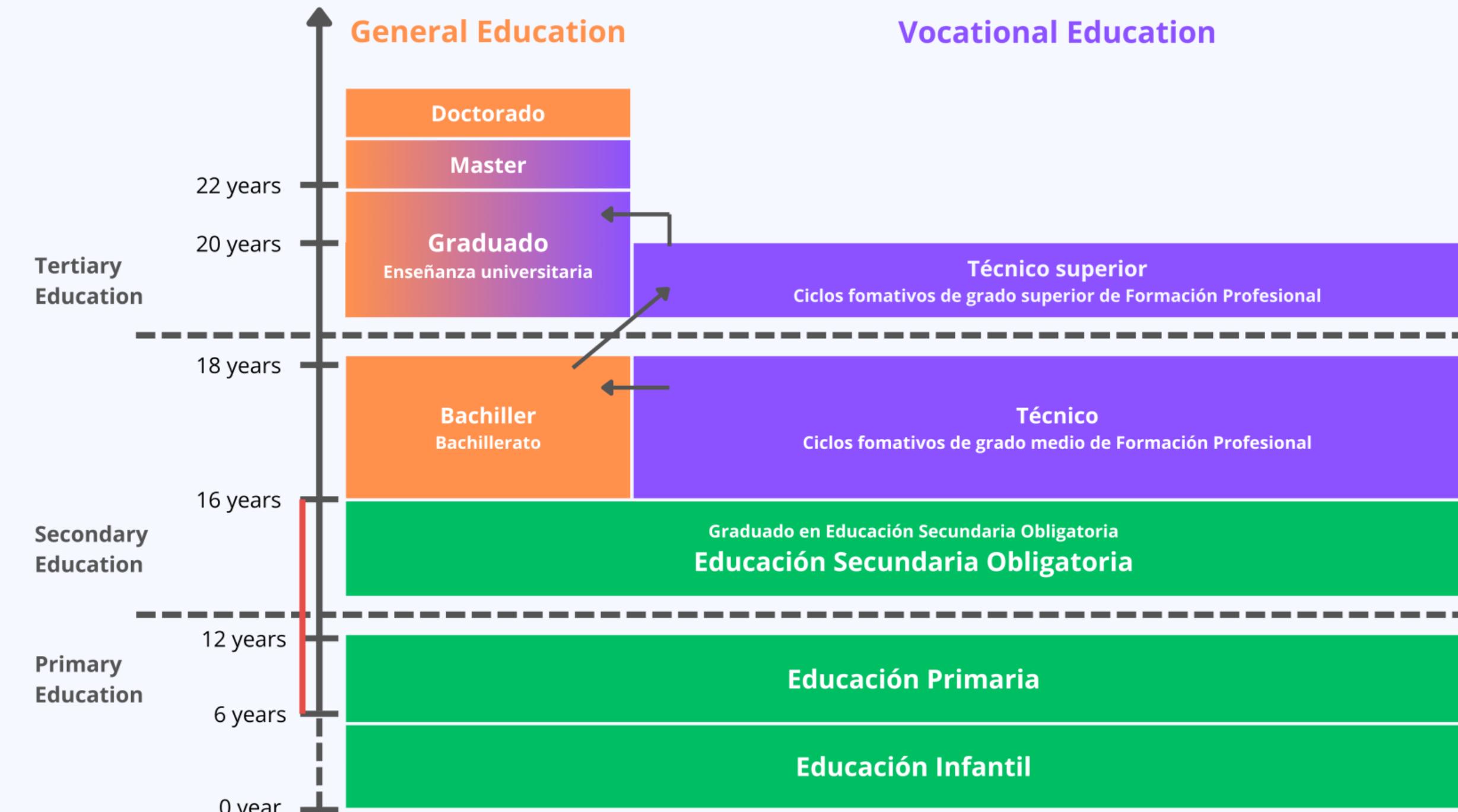
2. Education

basic 4 years, 6 years, 9 years b/f highschool

- Consumer Price Index (CPI)

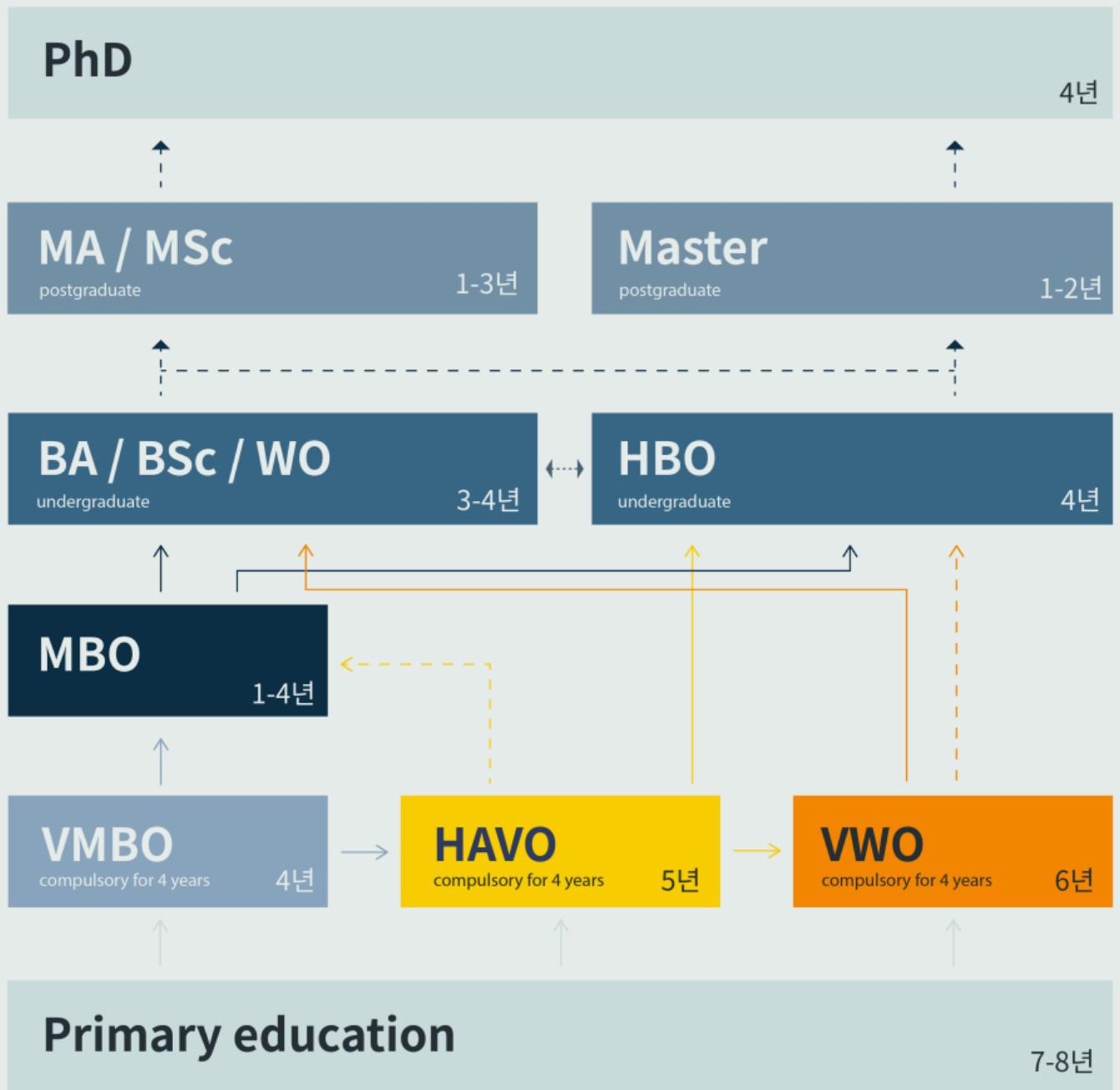


Spanish Education System

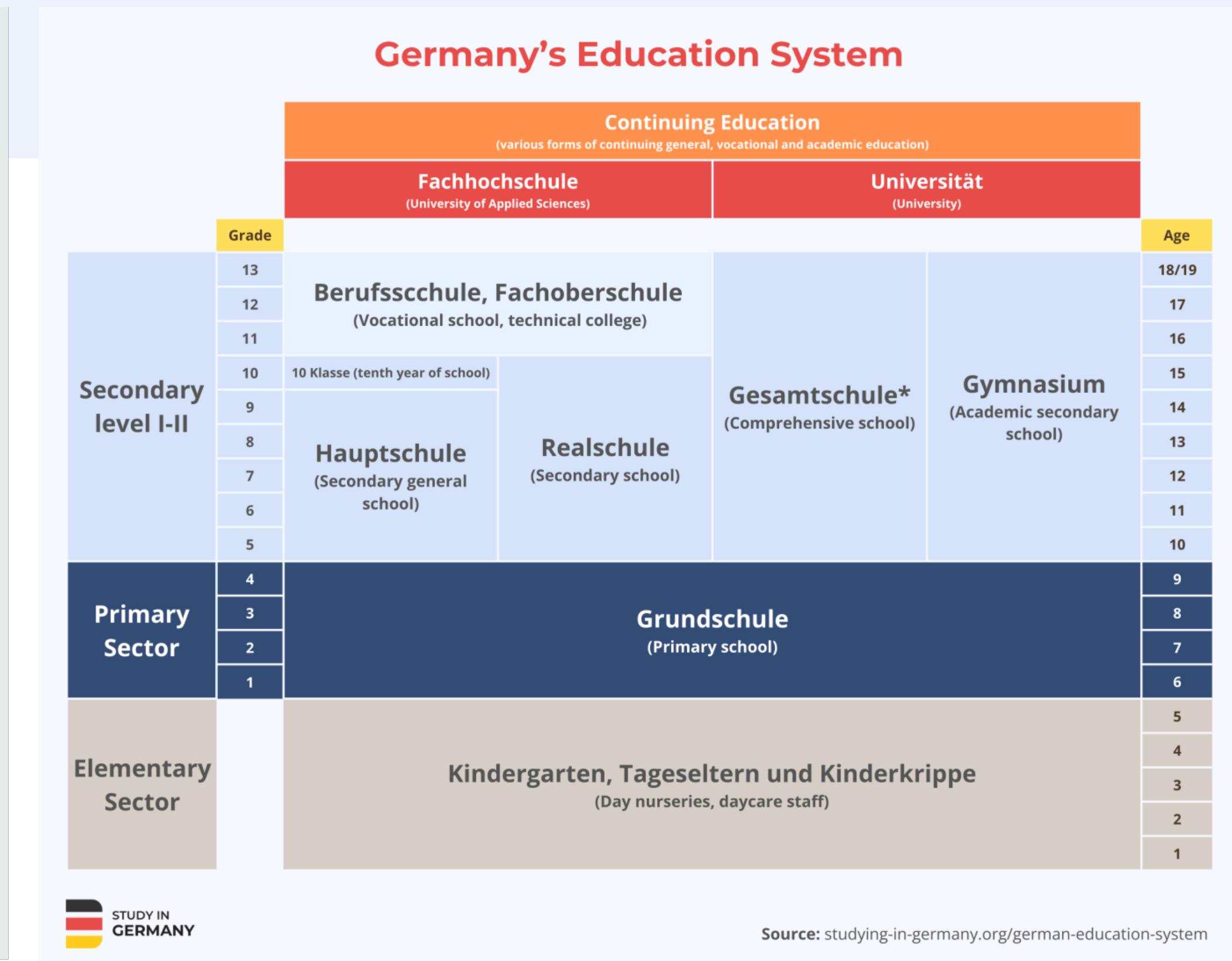


Source: studying-in-spain.com/plan-your-studies/spanish-education-system/

Germany's Education System



ExpatsAbroad



	Bac général	Bac technologique	BT	Bac pro	
Cycle terminal	Terminale générale	Terminale technologique	Terminale BT	Terminale pro	ULIS/ UPE2A
Première générale	Première technologique	Première d'adaptation	Première BT	Première pro	
Cycle de détermination	Seconde générale et technologique		Seconde BT	Seconde pro	
				CAP	
				2 ^{de} année CAP	
				1 ^{re} année CAP	

Diplôme national du brevet (DNB)		
Cycle IV (approfondissements)	Troisième	
	Quatrième	
	Cinquième	
Cycle III (consolidation)	Sixième	ULIS / UPE2A
		Sixième à troisième Segpa
Cycle III (consolidation)	Cours moyen deuxième année	
	Cours moyen première année	
Cycle II (apprentissages fondamentaux)	Cours élémentaire deuxième année	ULIS / UPE2A
	Cours élémentaire première année	
	Cours préparatoire	
Cycle I (apprentissages premiers)	Grande section	
	Moyenne section	
	Petite section (3 ans)	

BT Examen terminal

--- Le cycle III couvre la fin de l'enseignement élémentaire et la 6^e en collège

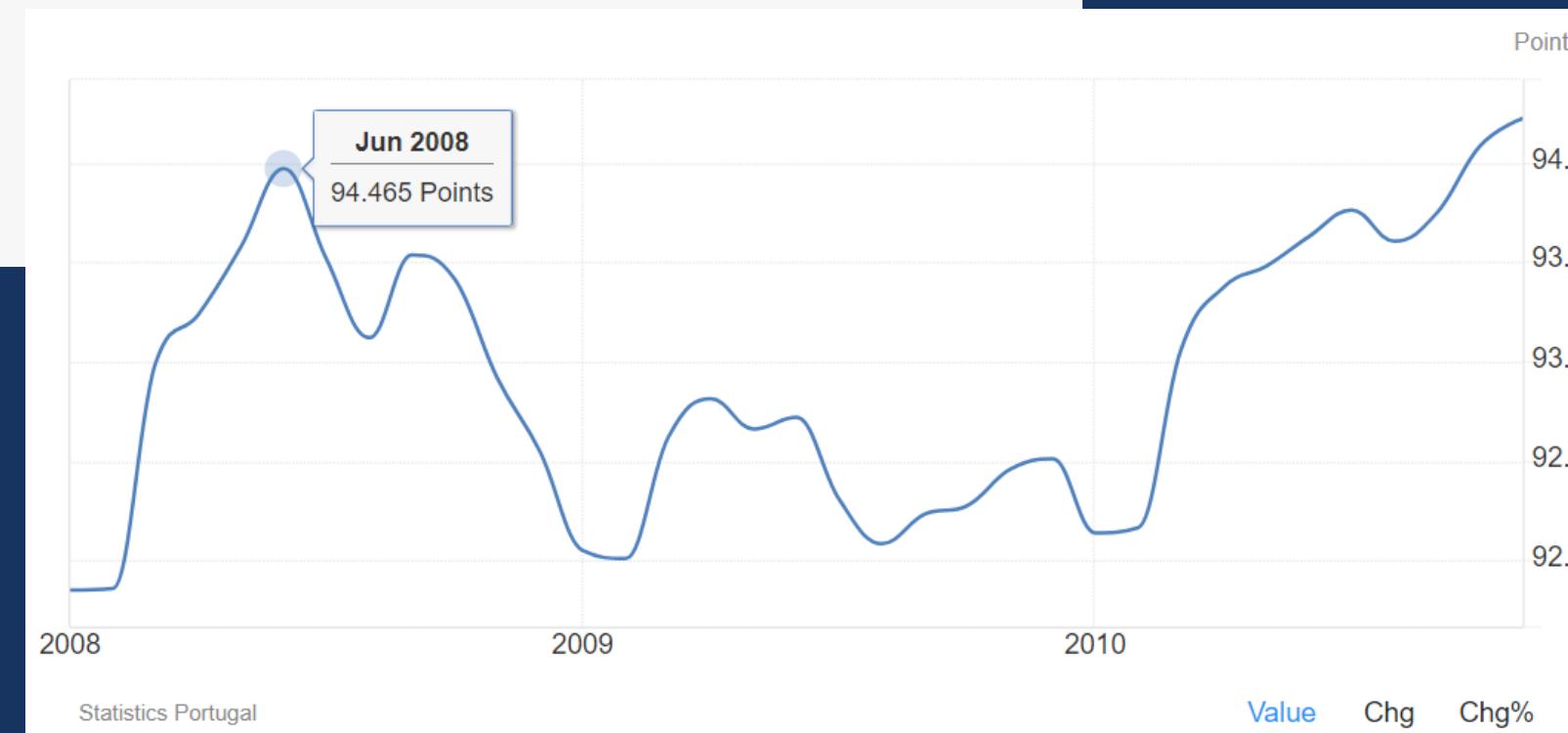
Enseignements adaptés

added Derived Variable

Before & After

new variable by using data 'cons.price.idx'

```
df_train["year"] = df_train["cons.price.idx"]
df_train["year"] = df_train["cons.price.idx"].replace({93.918 : 2008,
                                                       94.027 : 2010,
                                                       92.649 : 2009,}
```



#	Column
0	id
1	age
2	job
3	marital
4	education
5	default
6	housing
7	loan
8	contact
9	month
10	day_of_week
11	campaign
12	previous
13	poutcome
14	emp.var.rate
15	cons.price.idx
16	cons.conf.idx
17	y
18	year
19	pdays_cat

dtypes: float64(4),

Data Preprocessing

01

Data Reduction

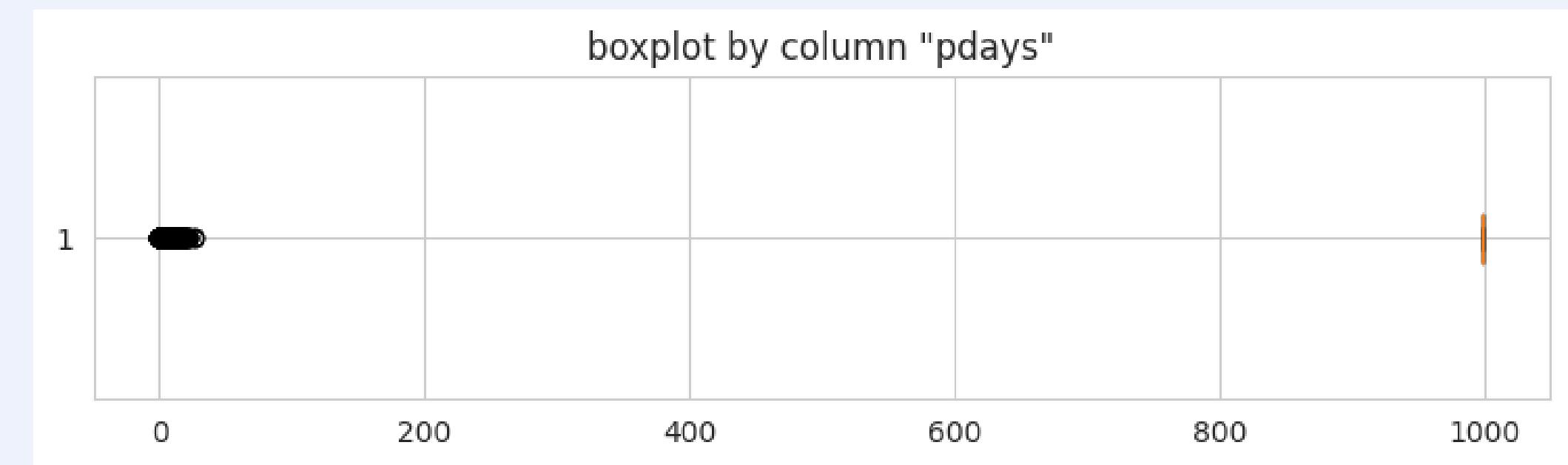
'pdays': abnormal → deleted

```
df_train_ini['pdays'].unique()  
array([999, 6, 3, 2, 13, 9, 12, 5, 7, 10, 4, 15, 1,  
       18, 14, 0, 22, 16, 11, 8, 21, 17, 26, 19, 27, 20])
```

02

Duplicate value detection

& There was no duplicate value



Data Preprocessing

03

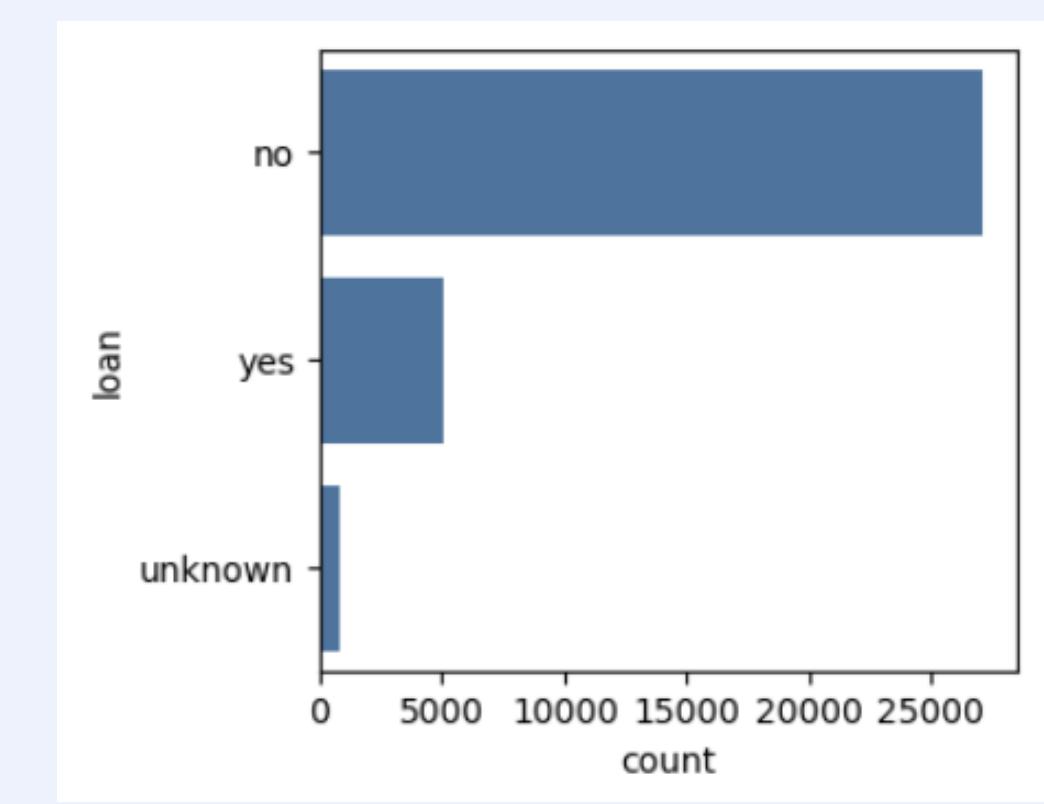
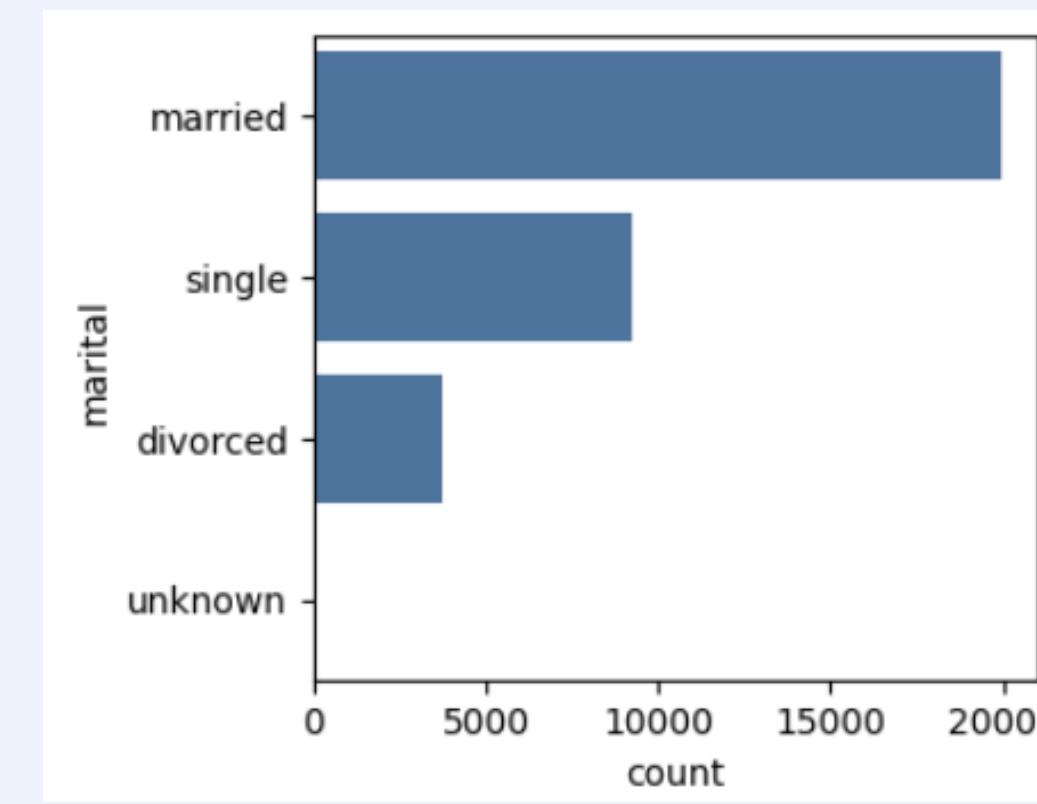
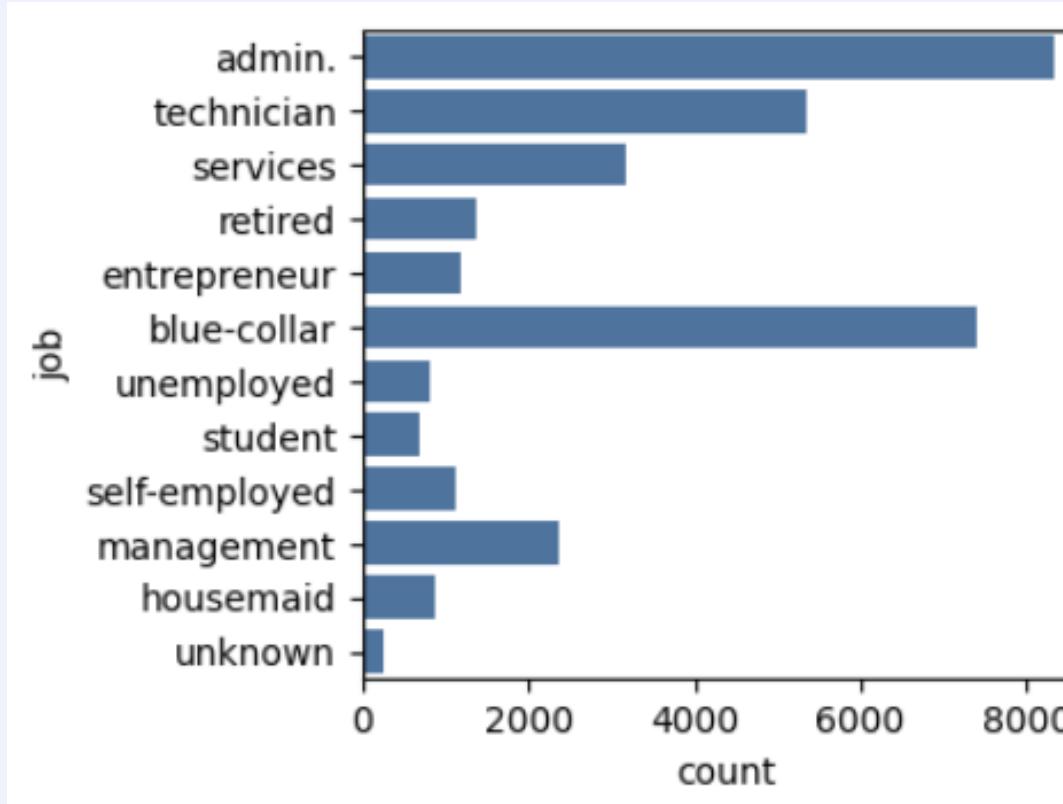
Dealing Variables

```
df_train["month"].replace({"jul": 7, "jun": 6, "aug": 8, "may": 5, "oct": 10,
```

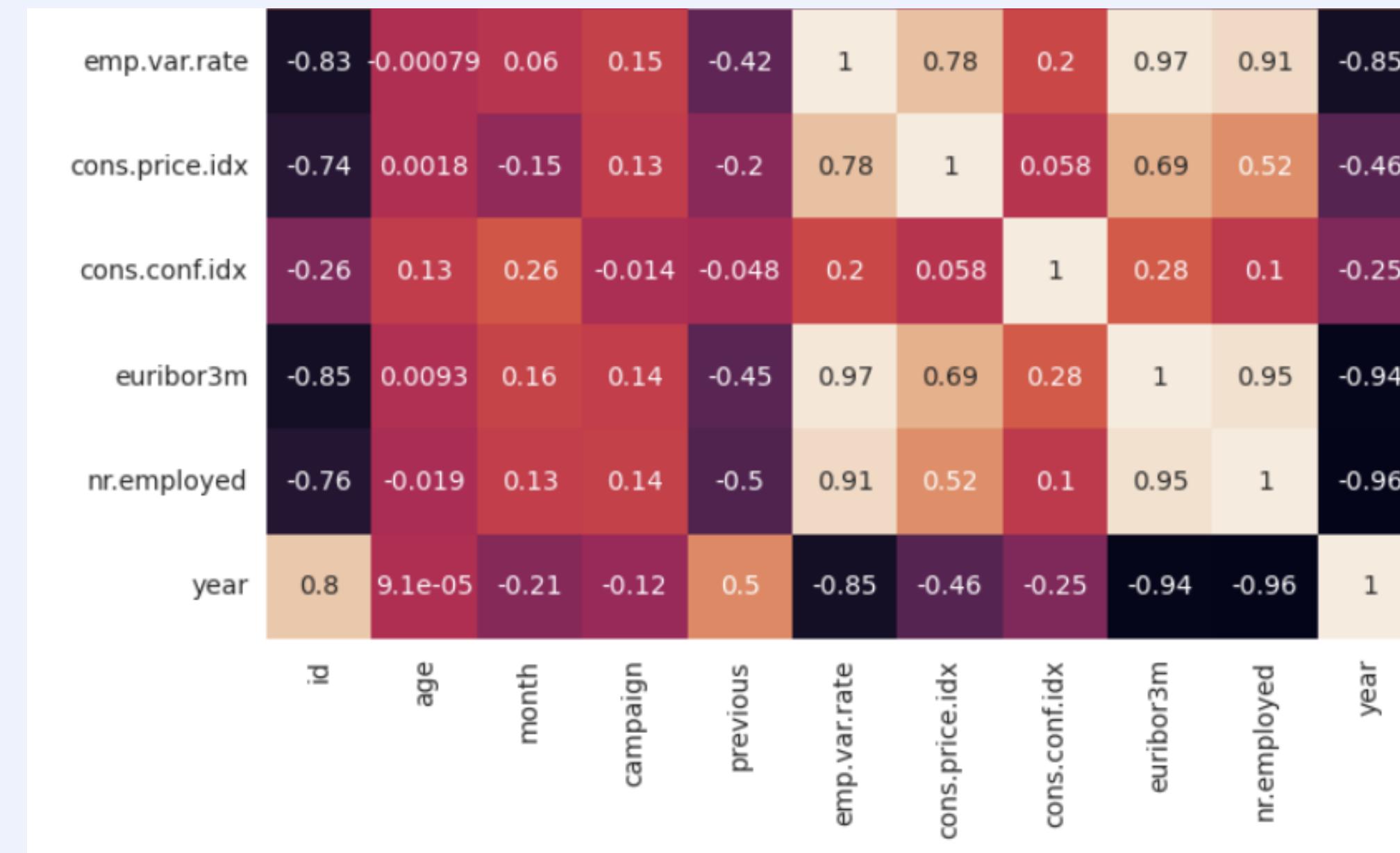
'age' >=65 : 'retired'

'month': categorical → discrete variable

→ 'job', 'marital', 'loan': unknown → mode



Data Preprocessing

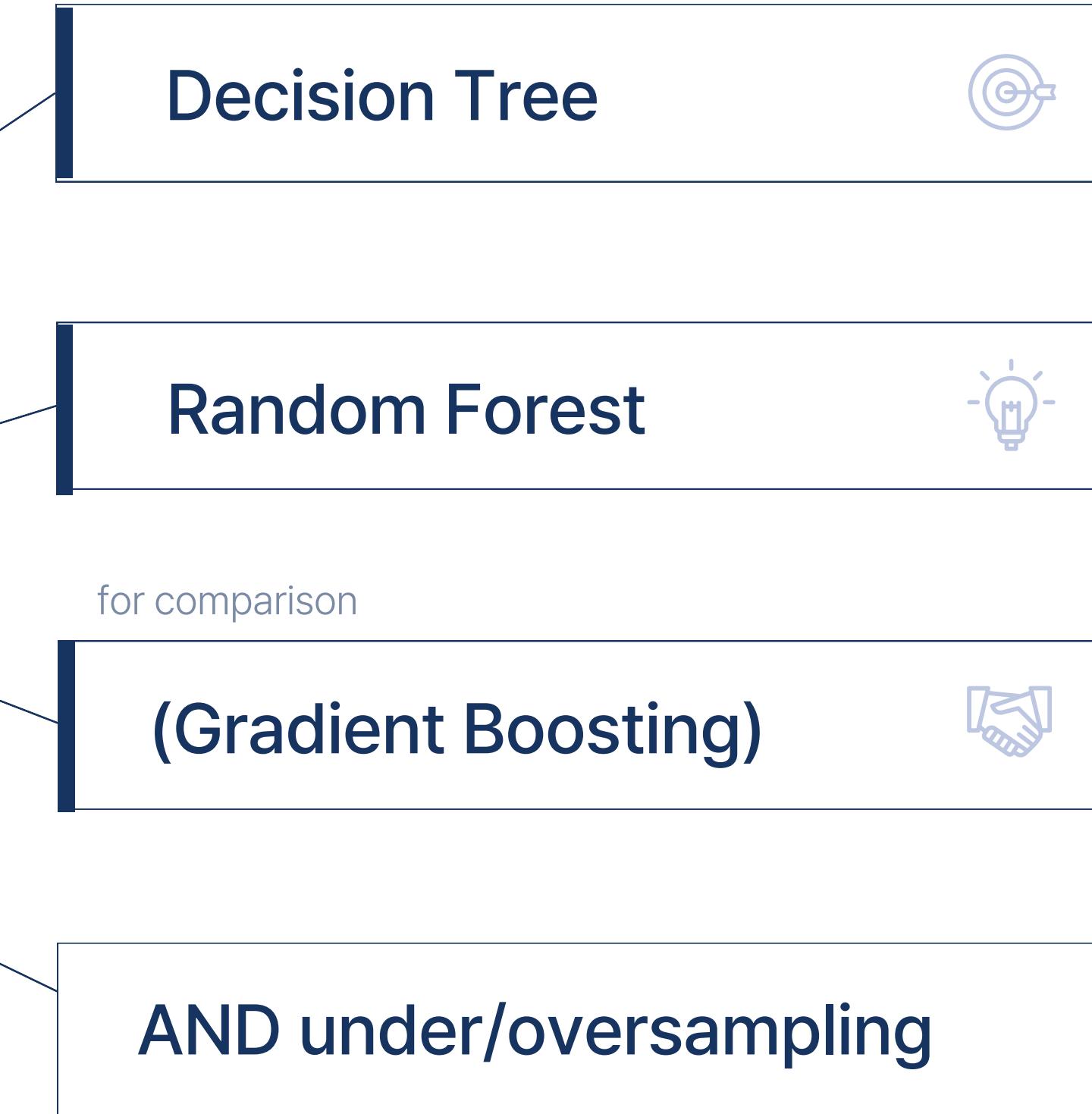
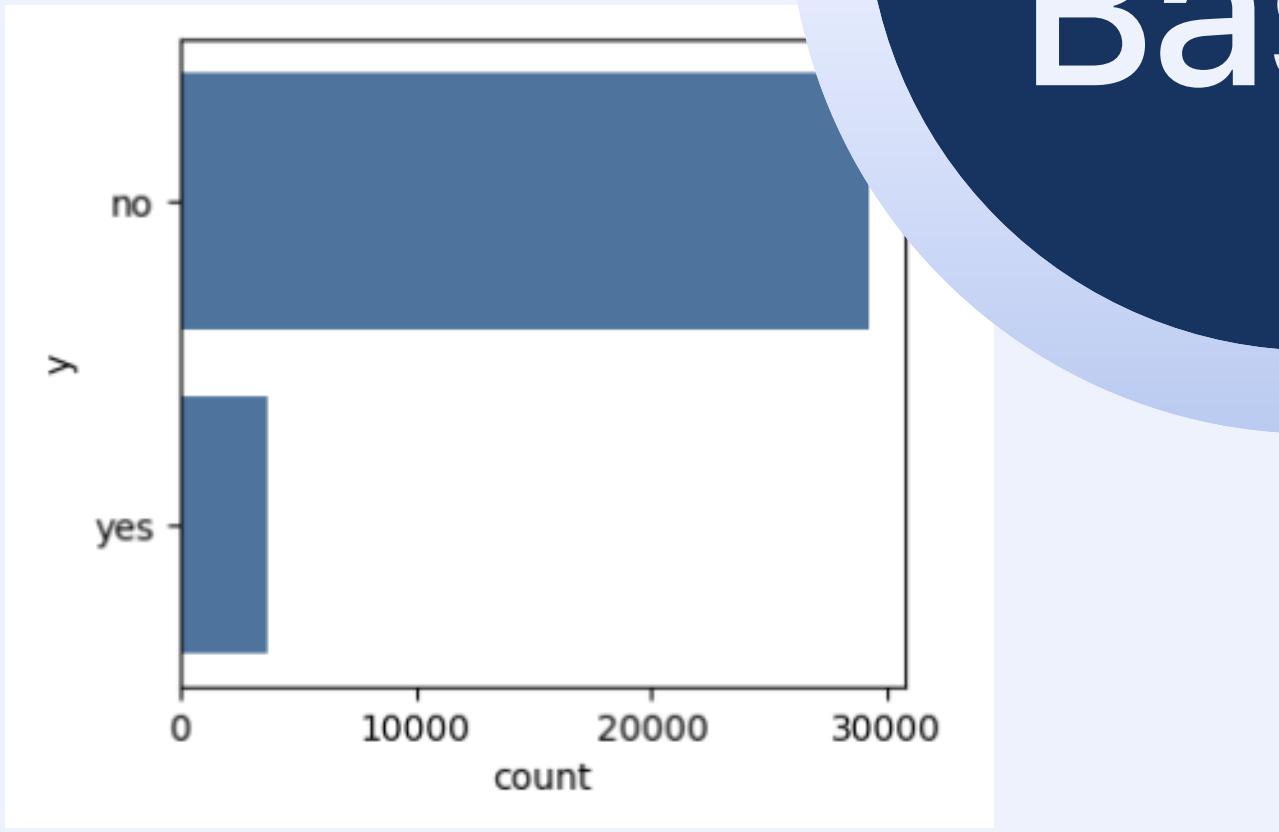


04

About Multicollinearity problem

'nr.employed' & 'euribor3m' were highly related with the "year" column we made
→ removed

Model Proposal



Thank you!

GROUP 3

2391007 MINJEONG KIM, 2391016 SUAHN LEE,
2391017 YOONJI LEE