

Análisis de la Igualdad de género en la UE

Valentina Santiso, Laura Moya, Laia Flores
Máster en Data Science - DSC0222
Nuclio Digital School

Índice

01 Introducción

02 Objetivo 1 - Modelo de Clasificación y Factor Analysis

03 Objetivo 2 - Modelo de Regresión

04 Objetivo 3 - Clustering

05 Objetivo 4 - Series Temporales

06 Conclusión

Introducción

¿Qué países de la UE alcanzarán el objetivo de igualdad de género de la Agenda 2030 de la ONU?

**Agenda
2030 de
la ONU**



Datos



Country	fulltime_employment_rate	duration_working_life	employed_in_education_health_socialactivities	doing_sport_eating_healthy	no_medical_examination	no_dental_examination	year	Sex	Score
12	France	43.0	34.0	34.0	32.0	3.0	5.0	2021	W 75.50
40	France	42.0	34.0	34.0	32.0	3.0	6.0	2020	W 75.10
68	France	41.0	33.0	34.0	32.0	3.0	5.0	2019	W 74.60
96	France	41.0	33.0	35.0	32.0	4.0	7.0	2017	W 75.50
124	France	42.0	32.0	32.0	32.0	6.0	9.0	2015	W 75.50
152	France	42.0	32.0	31.0	32.0	3.5	6.0	2018	W 75.05
180	France	41.0	33.0	34.5	32.0	5.0	8.0	2016	W 75.50
208	France	41.5	32.5	33.5	32.0	5.5	8.5	2014	W 75.50
236	France	42.0	32.0	31.5	39.0	3.0	5.0	2021	M 75.50

Importación datos

Juntar datasets de hombres y mujeres

Añadir índice de igualdad

Añadimos años que faltaban



Análisis de las variables

Datasets sin nulos ni duplicados

Cambiamos outliers de risk_poverty y s20/s80 income

Visualizaciones

Graficamos los datos para su posterior análisis

Índice de igualdad



	fulltime_employment_rate		share_ministers
	duration_working_life		share_members_parliament
WORK	flexibility_at_work		share_members_regional_assemblies
	career_prospects		share_board_members_largest_companies
	employed_in_education_health_socialactivities		share_board_members_central_bank
	mean_monthly_earnings		share_board_member_research_org
MONEY	mean_equivalised_income		share_board_member_public_org
	risk_of_poverty		share_decision_making_body_olympic_org
	S20/S80_income		
	tertiary_education_graduates		good_self-perceived_health
KNOWLEDGE	people_in_education		life_expectancy
	tertiary_students_education_health_art_field		healthy_life_years
	caring_kids_eldery_people_everyday		no_smoking_nor_harmful_drinking
TIME	cooking_housework_everyday		doing_sport_eating_healthy
	doing_sport_cultural_activities		no_medical_examination
	voluntary_activities		no_dental_examination



POWER	share_ministers
	share_members_parliament
	share_members_regional_assemblies
	share_board_members_largest_companies
	share_board_members_central_bank
	share_board_member_research_org
	share_board_member_public_org
	share_decision_making_body_olympic_org
HEALTH	good_self-perceived_health
	life_expectancy
	healthy_life_years
	no_smoking_nor_harmful_drinking
	doing_sport_eating_healthy
	no_medical_examination
	no_dental_examination



01

Objetivo

¿Qué variables son las que más discriminan entre sexos?

Modelo de clasificación

ETL & Preprocessing

Análisis del Target: SEX

- 1 = Woman
- 0 = Man

Correlación entre variables

Modelización

Split, Train & Test

Modelo: XGBoost

Conclusiones

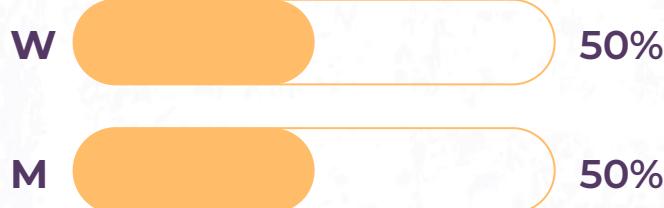
- Evaluación del resultado
- Importancia de Features
- Cálculo de *Shap Values*

Preprocessing

Info. del Dataset



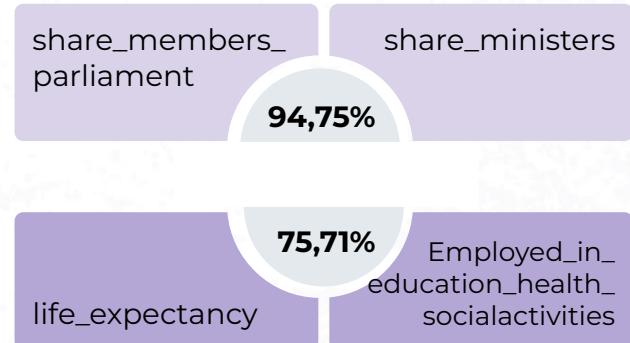
Dataset balanceado



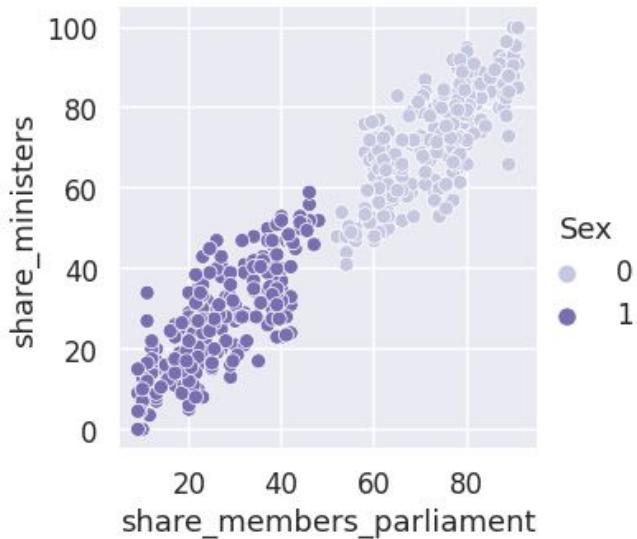
Correlaciones

Dataset con variables muy correlacionadas entre sí.

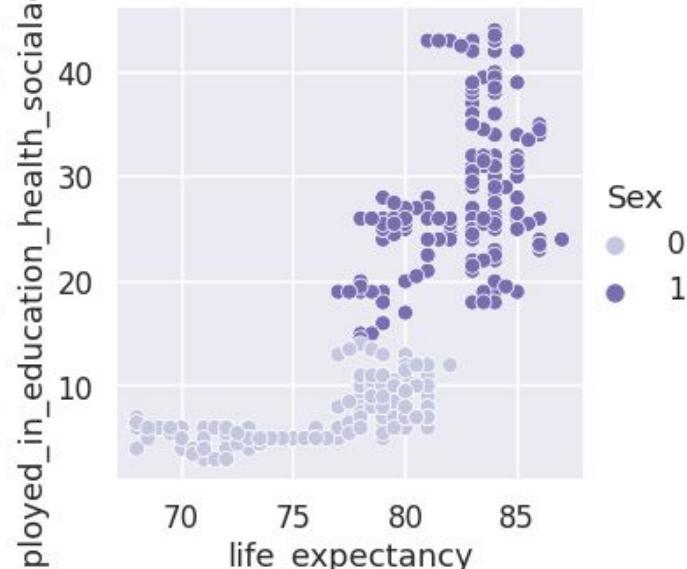
- 41 correlaciones > 75% (**corr. fuerte**)
- 15 correlaciones > 90%



**Share_ministers (%) vs.
Share_members_parliament (%)**

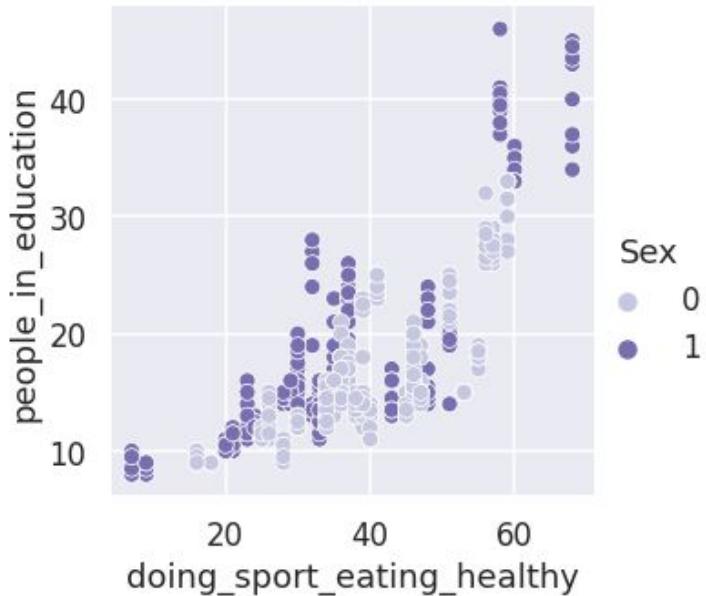


**Life_expectancy (years) vs.
Employed_in_education_health_socialactivities (%)**

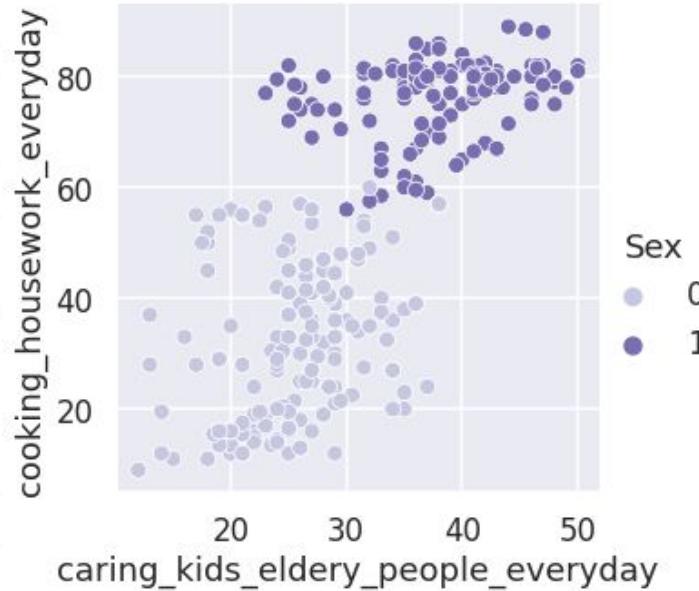


0 = Man
1 = Woman

**People_in_education (%) vs.
doing_sport_eating_healthy (%)**



**Cooking_housework_everyday (%) vs.
caring_kids_eldery_people_everyday (%)**



0 = Man
1 = Woman

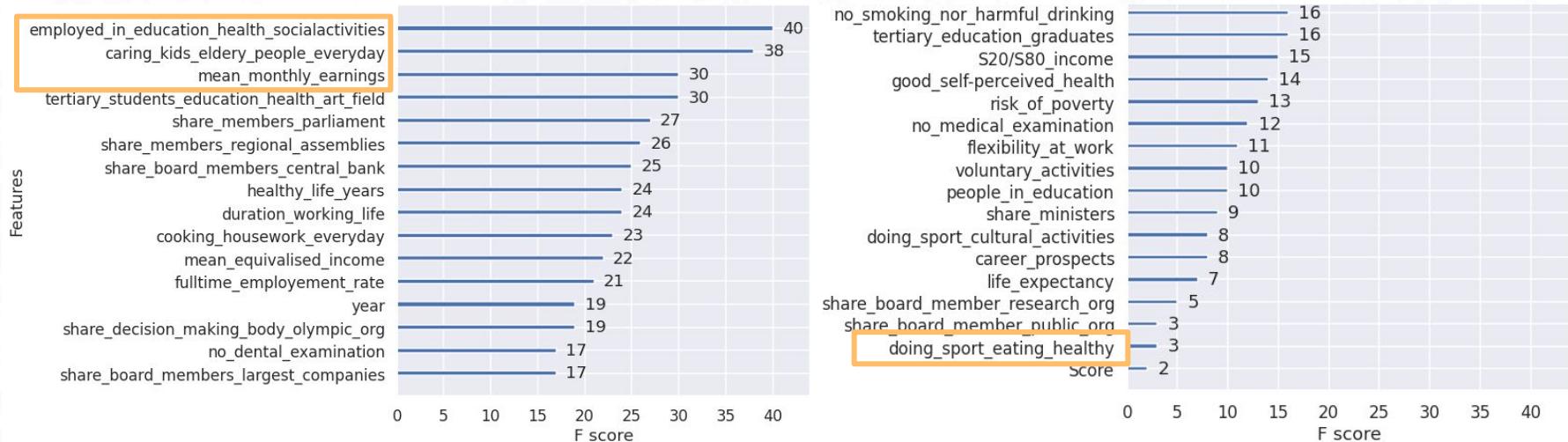


a) Evaluación del resultado

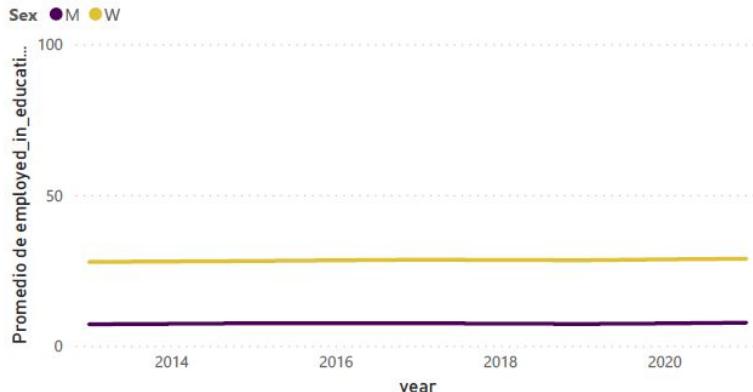
	Accuracy	Recall	Roc_Auc	Precision	F1
XGboost	1.0	1.0	1.0	1.0	1.0

Existe tal diferencia entre hombres y mujeres, que el modelo es capaz de diferenciar con un **accuracy del 100%** si es hombre o mujer.

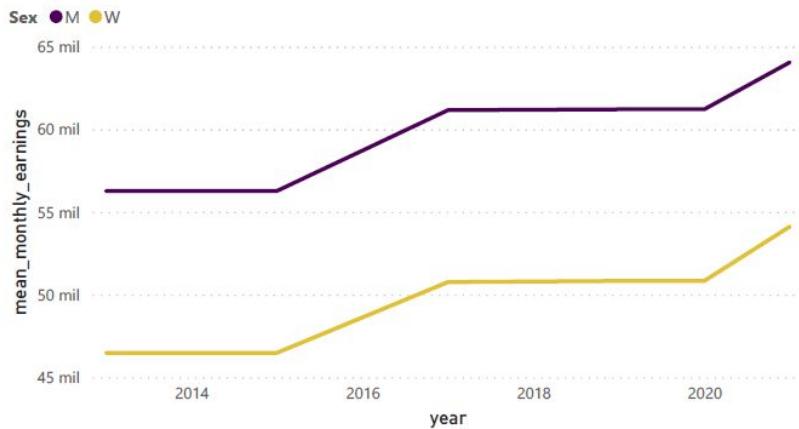
b) Feature Importance



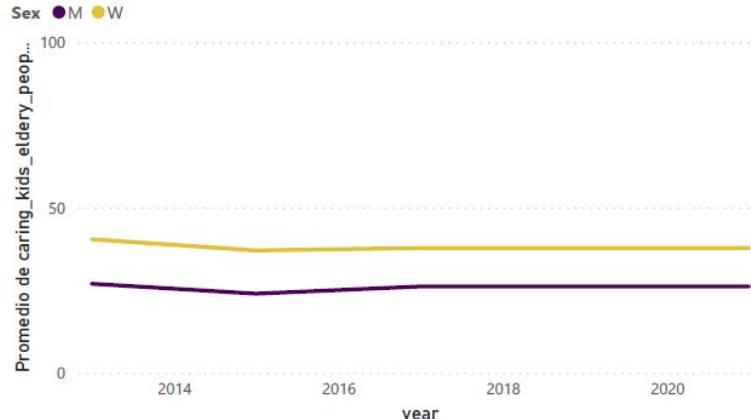
Promedio de employed_in_education_health_socialactivities por year y Sex



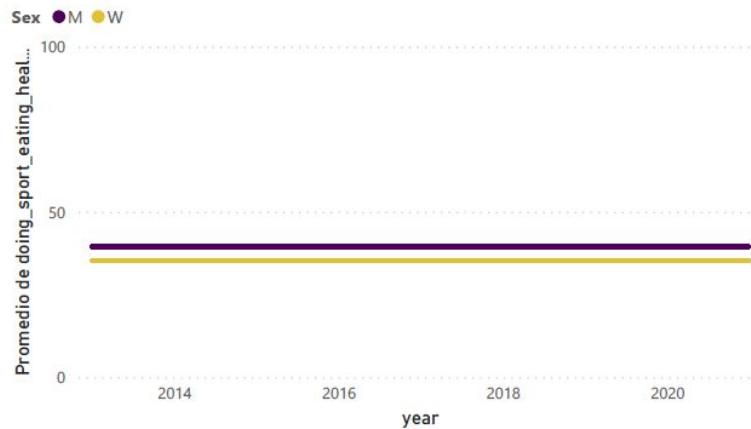
mean_monthly_earnings por year y Sex



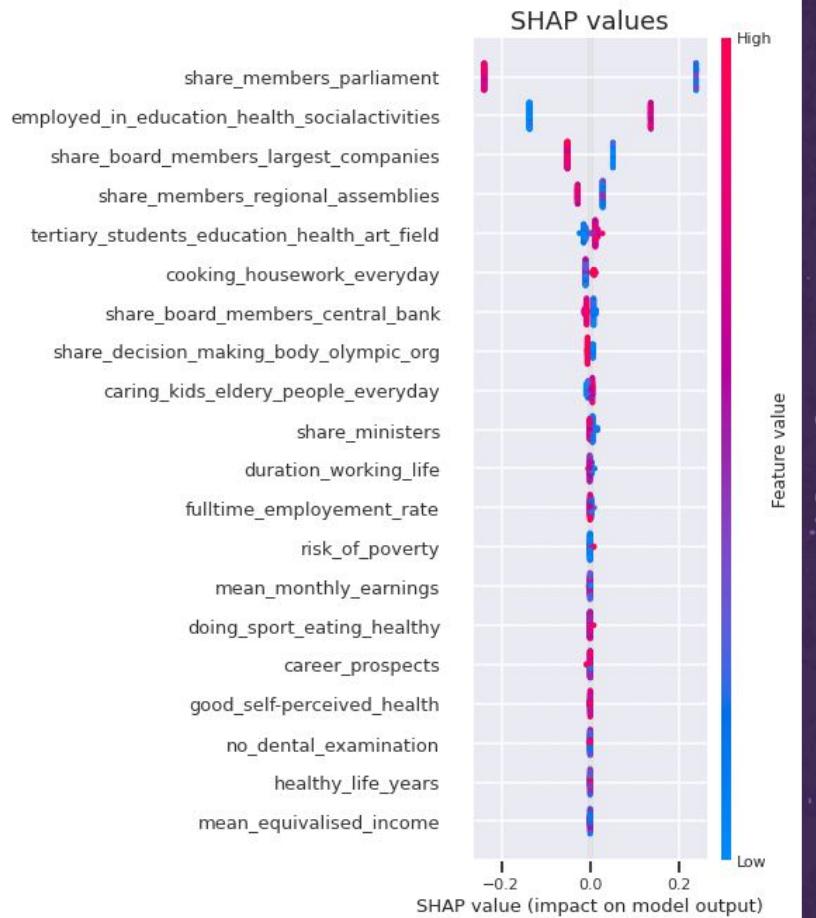
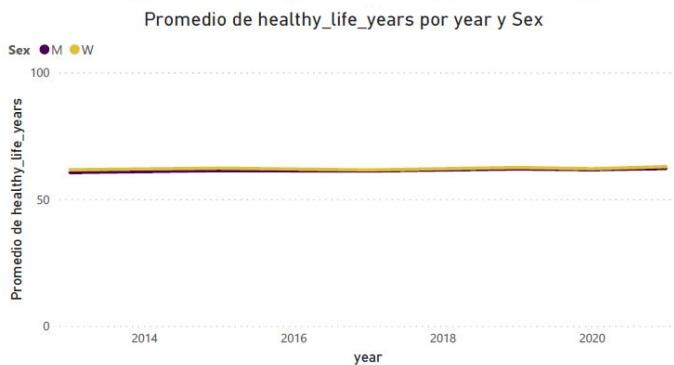
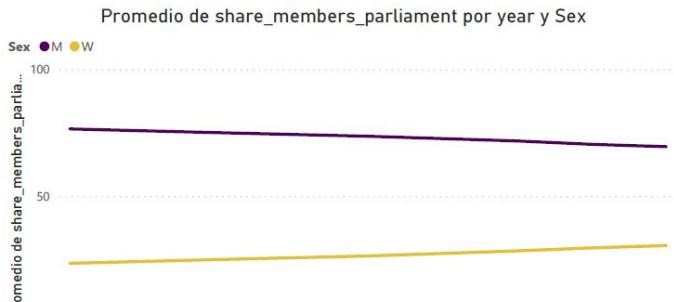
Promedio de caring_kids_eldery_people_everyday por year y Sex



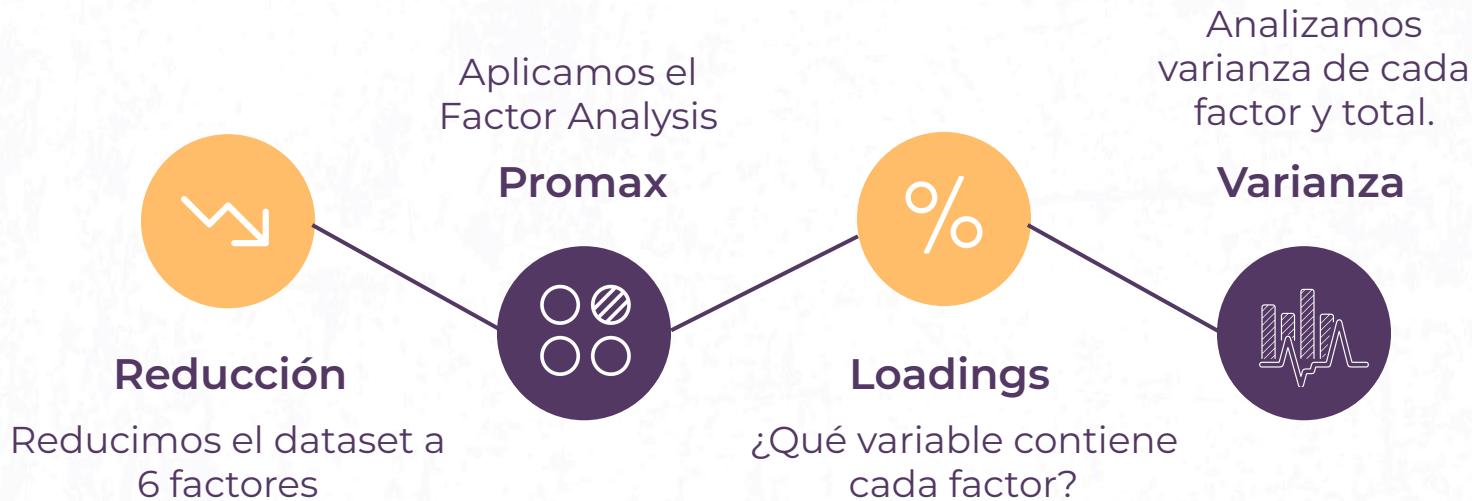
Promedio de doing_sport_eating_healthy por year y Sex



c) Shap Values



Reducción de la dimensionalidad



Factor Analysis → 6 factores

	Factor_1	Factor_2	Factor_3	Factor_4	Factor_5	Factor_6
0	-0.844425	0.585360	-0.233083	0.210231	-1.446743	0.543058
1	-0.882707	0.264490	0.181291	0.956383	-0.706597	0.489353
2	-0.677554	-1.290220	-0.820120	-0.139675	-0.067160	-0.187865
3	-1.079766	-0.416592	-0.488190	0.286456	-0.539130	1.264623
4	-1.079185	-0.459718	0.302392	-0.660288	-0.702055	-0.441754

Loadings

	Factor_1	Factor_2	Factor_3	Factor_4	Factor_5	Factor_6
fulltime_employment_rate	0.779503	0.274562	-0.065184	-0.239267	-0.046099	-0.060465
duration_working_life	0.645527	0.651580	-0.045503	-0.232768	0.037457	0.136590
employed_in_education_health_socialactivities	-0.817151	0.351073	0.056321	-0.096975	-0.031026	0.043672
flexibility_at_work	0.052906	0.570480	-0.016833	0.302958	0.122465	-0.148451
career_prospects	-0.141109	0.433731	-0.020198	-0.176636	-0.138454	-0.605146
mean_monthly_earnings	0.137451	0.562291	-0.037456	0.468363	-0.083141	-0.002083
mean_equivalised_income	-0.022824	0.563547	-0.037647	0.313709	-0.217758	0.033621
risk_of_poverty	-0.086966	-0.007843	0.837333	0.022137	0.134421	0.051620
S20/S80_income	0.002345	0.111861	1.006977	-0.007273	-0.034724	0.010910
tertiary_education_graduates	-0.267040	0.646145	-0.156711	-0.057193	0.091108	0.064477
people_in_education	-0.095906	0.986387	0.056887	-0.191953	0.103767	-0.072359
tertiary_students_education_health_art_field	-0.804934	0.253400	0.074916	0.005922	-0.091233	0.062443
caring_kids_eldery_people_everyday	-0.793270	-0.124947	-0.104414	0.125254	0.141875	0.067900



Nuevo Dataset

(+ normalización)

	Country	Power	Education	Money	Work	Health	Time	year	Sex
0	Austria	0.226983	0.544893	0.664423	0.601801	0.069483	0.537176	2021	W
1	Belgium	0.224728	0.492989	0.642974	0.710974	0.190016	0.474560	2021	W
2	Bulgaria	0.292459	0.149376	0.114091	0.477871	0.273401	0.446283	2021	W
3	Cyprus	0.152210	0.333470	0.532687	0.597987	0.169594	0.718839	2021	W
4	Czechia	0.154990	0.310745	0.825591	0.356973	0.228360	0.325296	2021	W
...
481	Portugal	0.868470	0.312730	0.400885	0.350311	0.474081	0.679498	2018	M
482	Romania	0.946917	0.132942	0.199170	0.572899	0.523032	0.237094	2018	M
483	Sweden	0.629517	0.843923	0.634198	0.681335	0.289221	0.245904	2018	M
484	Slovenia	0.867149	0.482729	0.787265	0.606089	0.235174	0.380642	2018	M
485	Slovakia	0.966590	0.246176	0.834717	0.387249	0.319286	0.307835	2018	M

486 rows × 9 columns

	0	1	2	3	4	5
Variance	12.025092	5.207969	1.851737	1.704344	1.586865	1.015058
Proportional Var	0.387906	0.167999	0.059733	0.054979	0.051189	0.032744
Cumulative Var	0.387906	0.555905	0.615639	0.670618	0.721807	0.754550

Promedio de Power por year y Sex

Sex ● M ● W

Promedio de Power

1,0
0,5
0,0

2014 2016 2018 2020

year



Promedio de Work por year y Sex

Sex ● M ● W

Promedio de Work

1,0
0,5
0,0

2014 2016 2018 2020

year



Promedio de Money por year y Sex

Sex ● M ● W

Promedio de Money

1,0

2014 2016 2018 2020

year



Promedio de Time por year y Sex

Sex ● M ● W

Promedio de Time

1,0
0,5
0,0

2014 2016 2018 2020

year



Promedio de Education por year y Sex

Sex ● M ● W

1,0

Promedio de Education

0,0

2014

2016

2018

2020

year



Promedio de Health por year y Sex

Sex ● M ● W

1,0

Promedio de Health

0,5

0,0

2014

2016

2018

2020

year



Promedio de Power, Health, Education, Money, Work y Time por Sex

● Promedio de Power ● Promedio de Health ● Promedio de Education ● Promedio de Money ● Promedio de Work ● Promedio de Time

Sex

M

W

0,0

0,2

0,4

0,6

0,8

1,0

Promedio de Power, Promedio de Health, Promedio de Education, Promedio de Money, Promedio de Work y Promedio de Time

02

Objetivo

¿Qué países presentan
unas características
similares?

Clustering - Igualdad entre países de la EU



Data Set

Creación estructura

Creación de Features

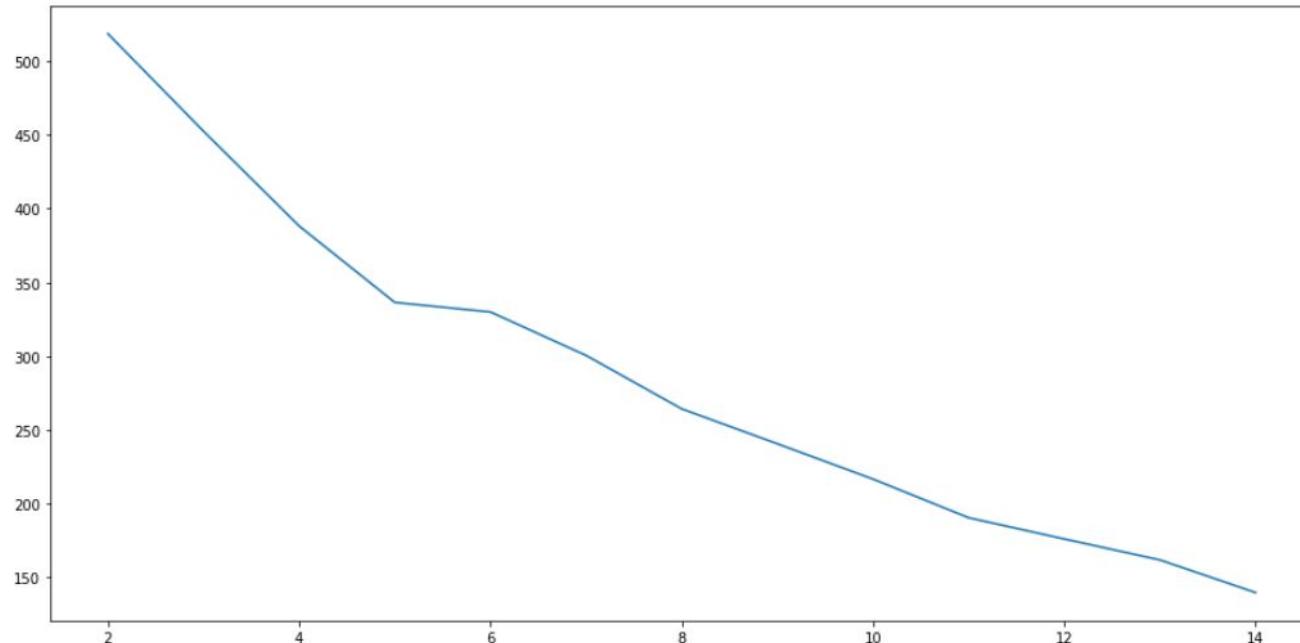
Modelo K-Means

Selección K
adecuada

Conclusiones

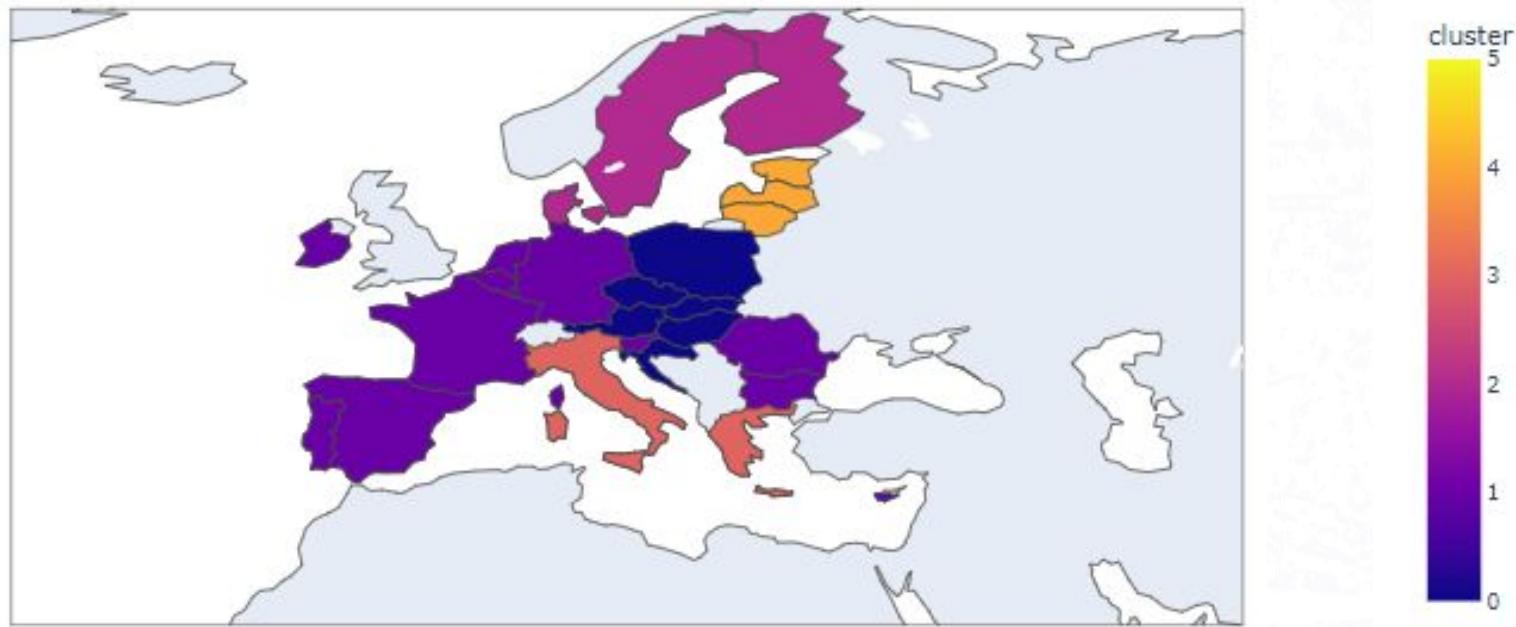
Identificación de K - Gráfica del codo:

Variación de la dispersión de los clústers en función de la k



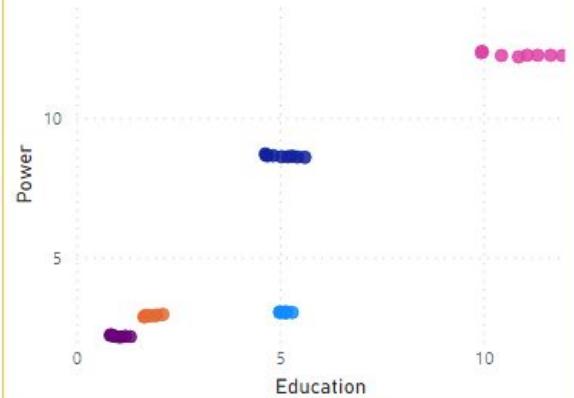
K = 5

Resultado: mapa 5 clusters



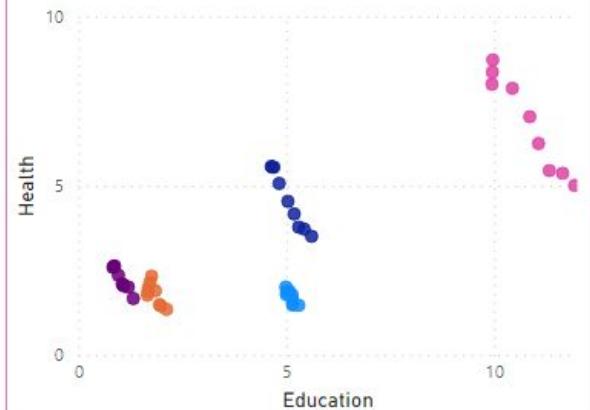
Education & Power

cluster ● 0 ● 1 ● 2 ● 3 ● 4



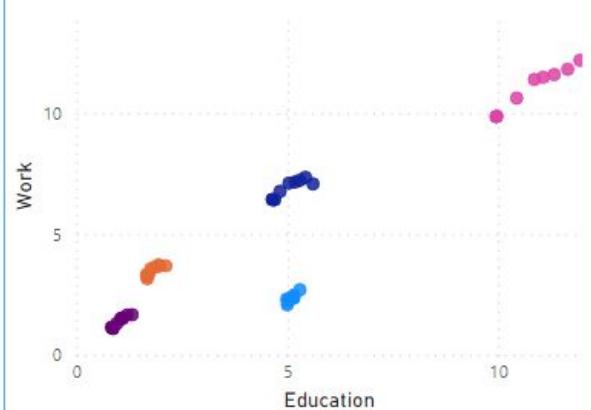
Education & Health

cluster ● 0 ● 1 ● 2 ● 3 ● 4



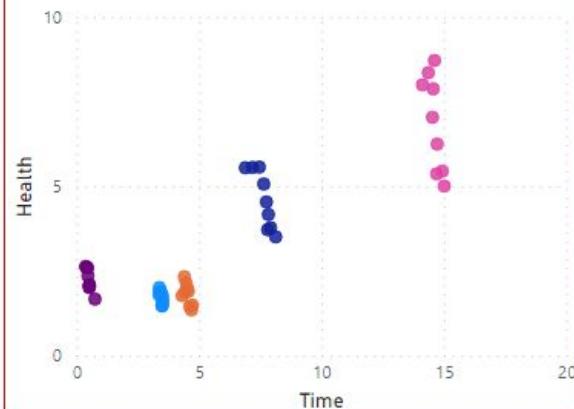
Work & Education

cluster ● 0 ● 1 ● 2 ● 3 ● 4



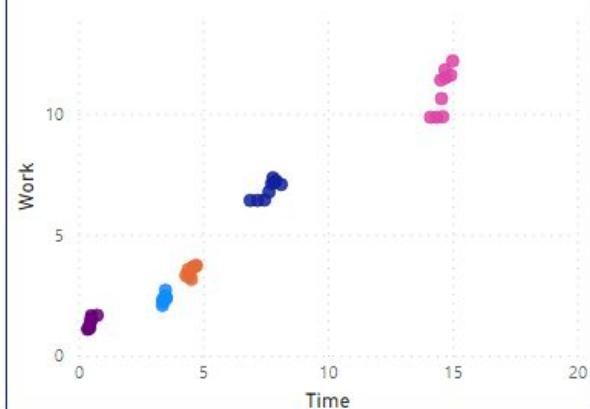
Time & Health

cluster ● 0 ● 1 ● 2 ● 3 ● 4



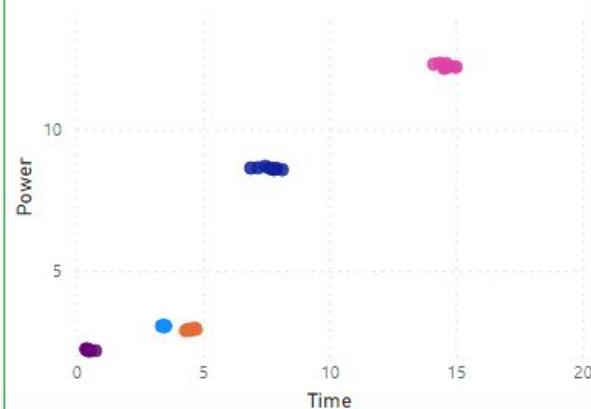
Time & Work

cluster ● 0 ● 1 ● 2 ● 3 ● 4



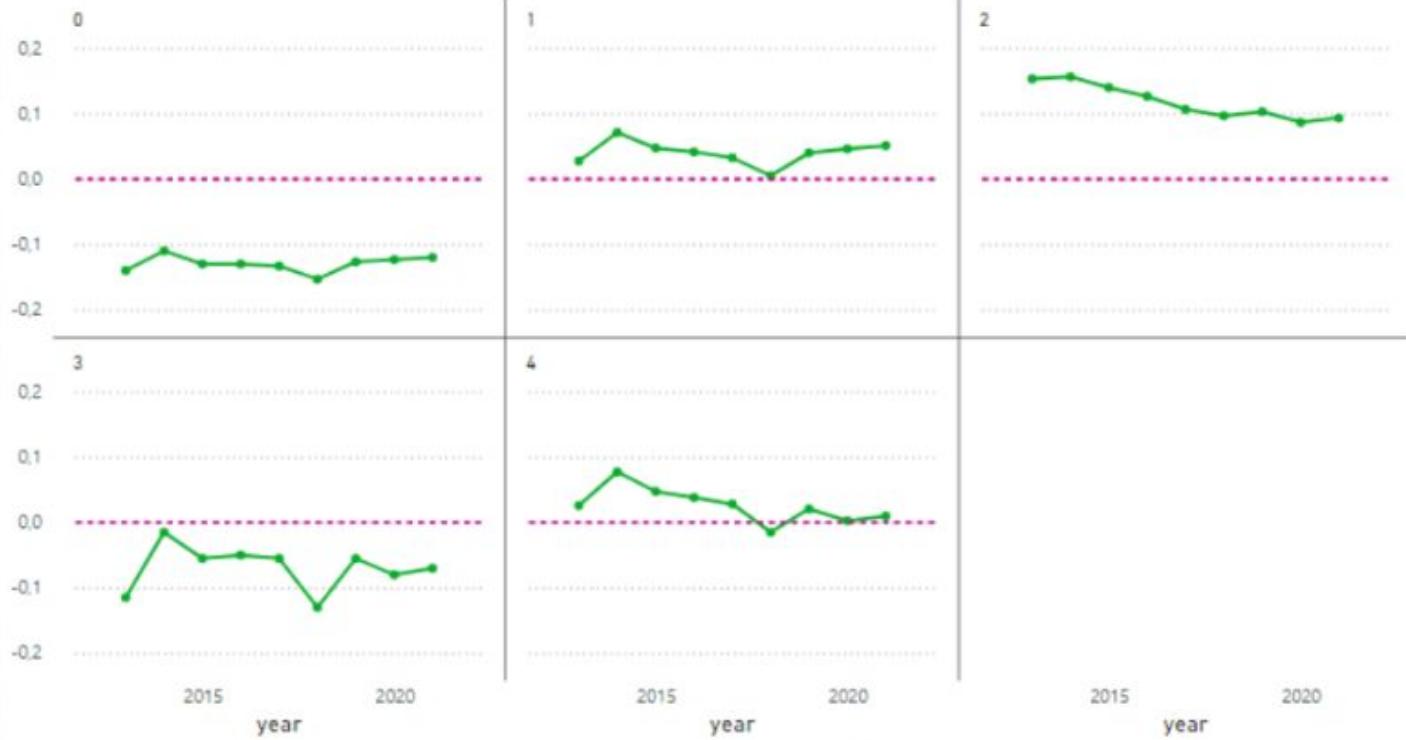
Power & Time

cluster ● 0 ● 1 ● 2 ● 3 ● 4





Knowledge



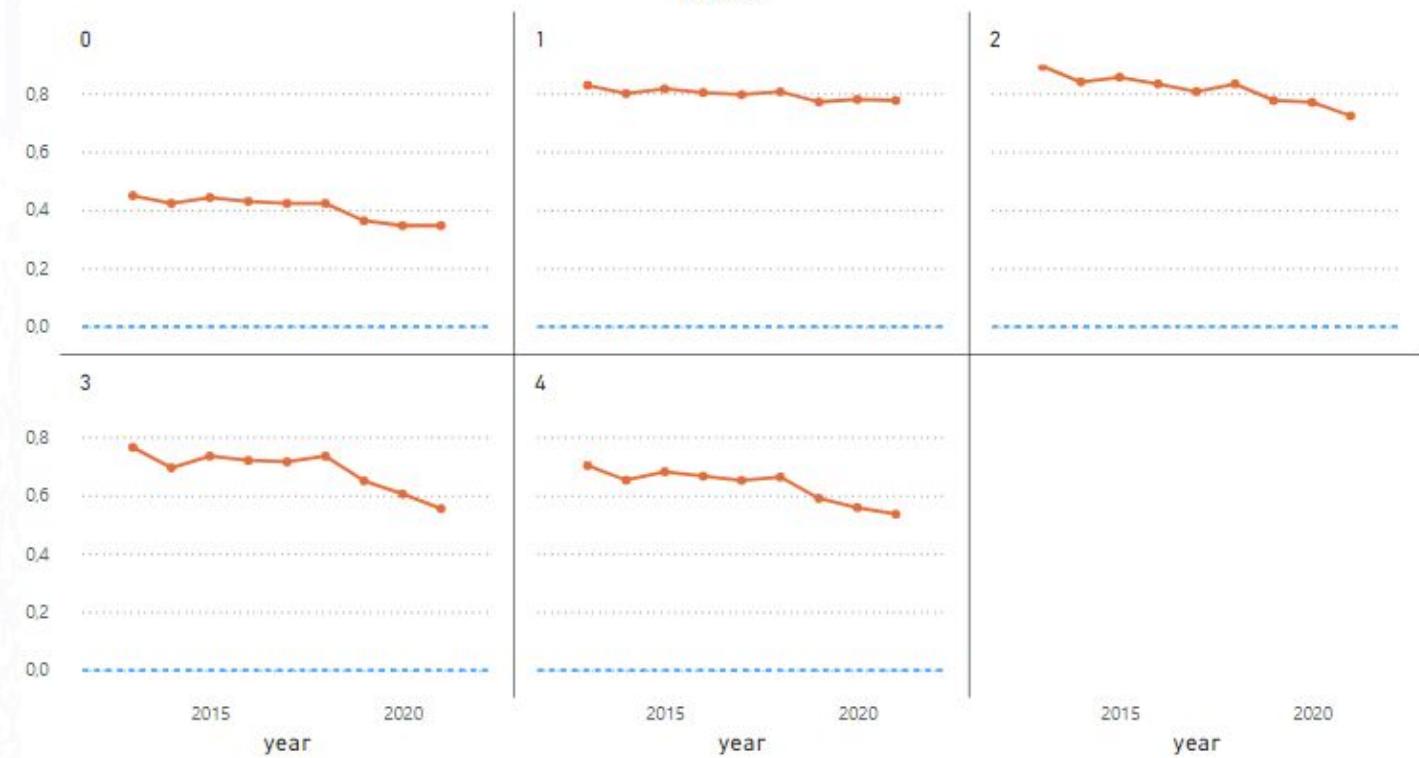
Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4

KNOWLEDGE

tertiary_education_graduates
people_in_education
tertiary_students_education_health_art_field



Power



POWER

share_ministers
share_members_parliament
share_members_regional_assemblies
Share_board_members_largest_companies

share_board_members_central_bank
share_board_member_research_org
share_board_member_public_org
share_decision_making_body_olympic_org

Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4

Women well-represented in politics at EU level

In November 2021, the Members of the European Parliament (MEPs) comprised 39.1 % women and 60.9 % men. In 10 Member States, however, at least two thirds of the MEPs were men, with highs of 84.8 % in Romania and 100 % in Cyprus.

The current European Commission demonstrates full gender parity, allowing for the odd number of members (13 women,

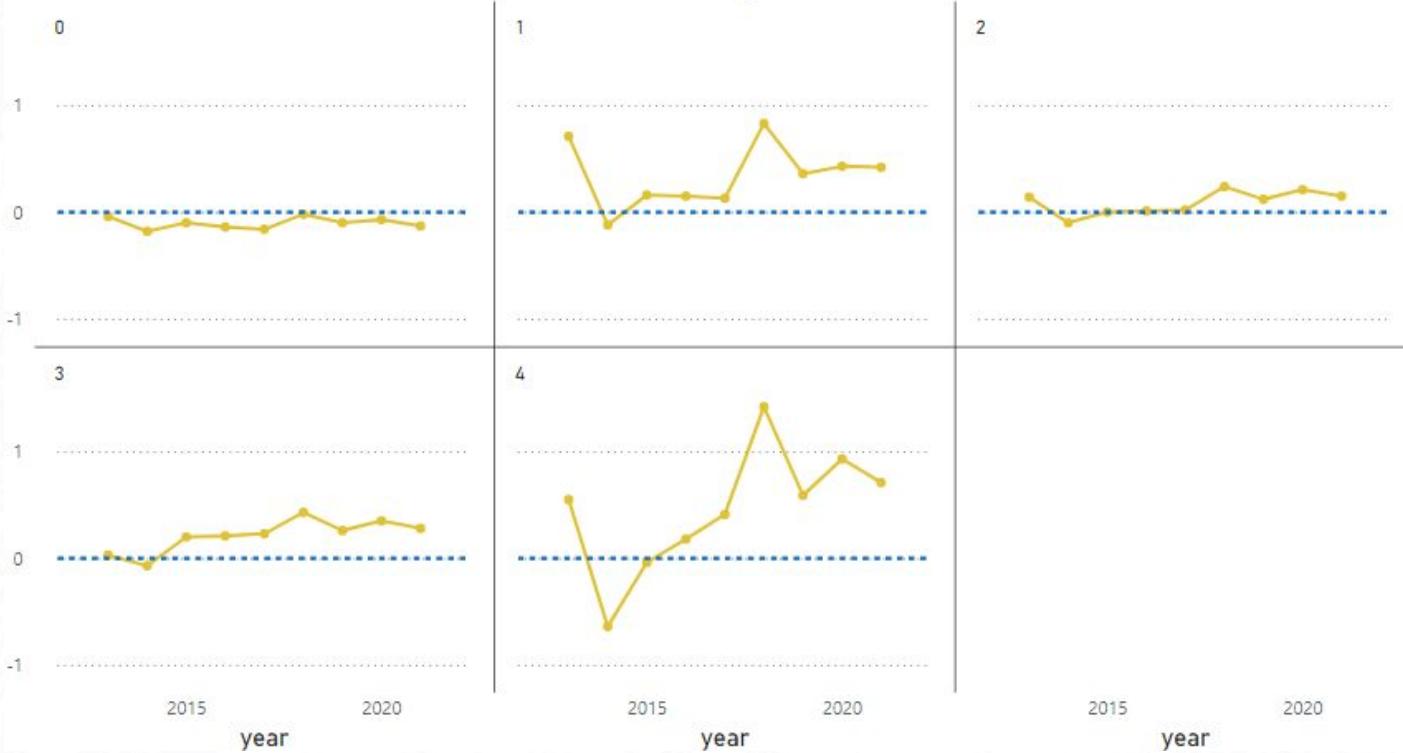
14 men). In December 2019, Ursula Von der Leyen was elected the first woman President of the Commission.

Roberta Metsola became President of the European Parliament in January 2022 and is the third woman to have been elected as leader of this key European institution. Note, however, that this change is not taken into account in EIGE's latest data, which refer to November 2021.





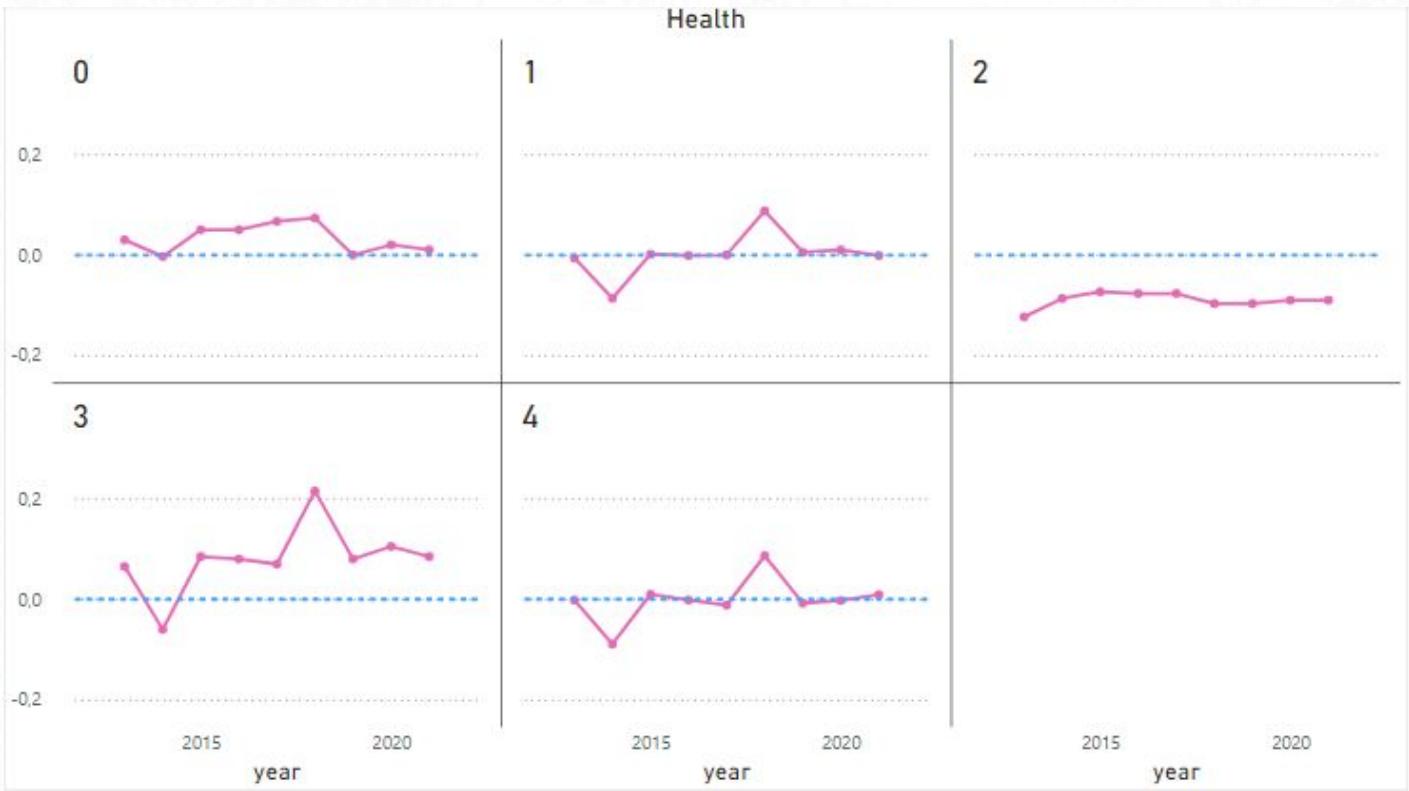
Money



MONEY

mean_monthly_earnings
mean_equivalised_income
risk_of_poverty
S20/S80_income

Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4



HEALTH

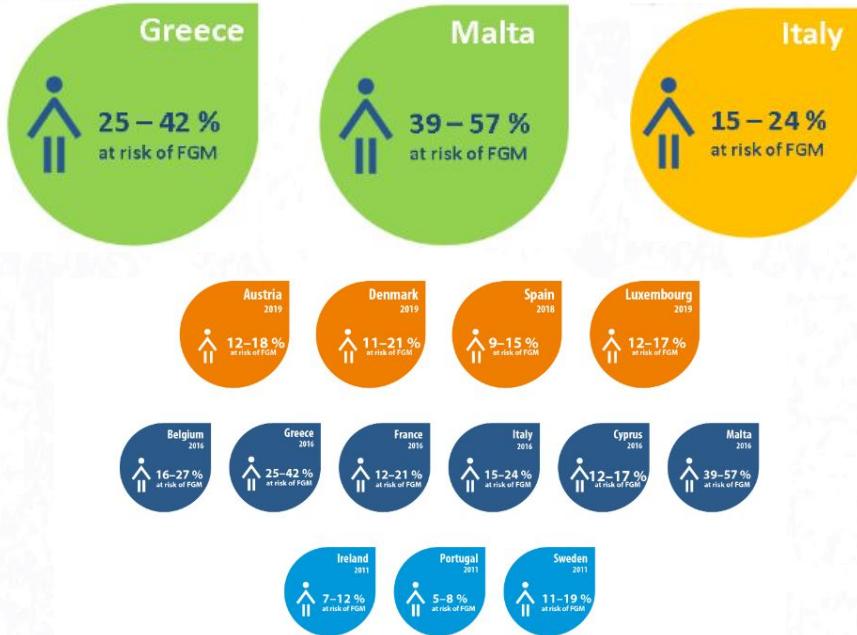
good_self-perceived_health
life_expectancy
healthy_life_years
no_smoking_nor_harm
no_dental_examination

ful_drinking
doing_sport_eating_healthy
no_medical_examination
no_dental_examination

Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4



Risk of female genital mutilation

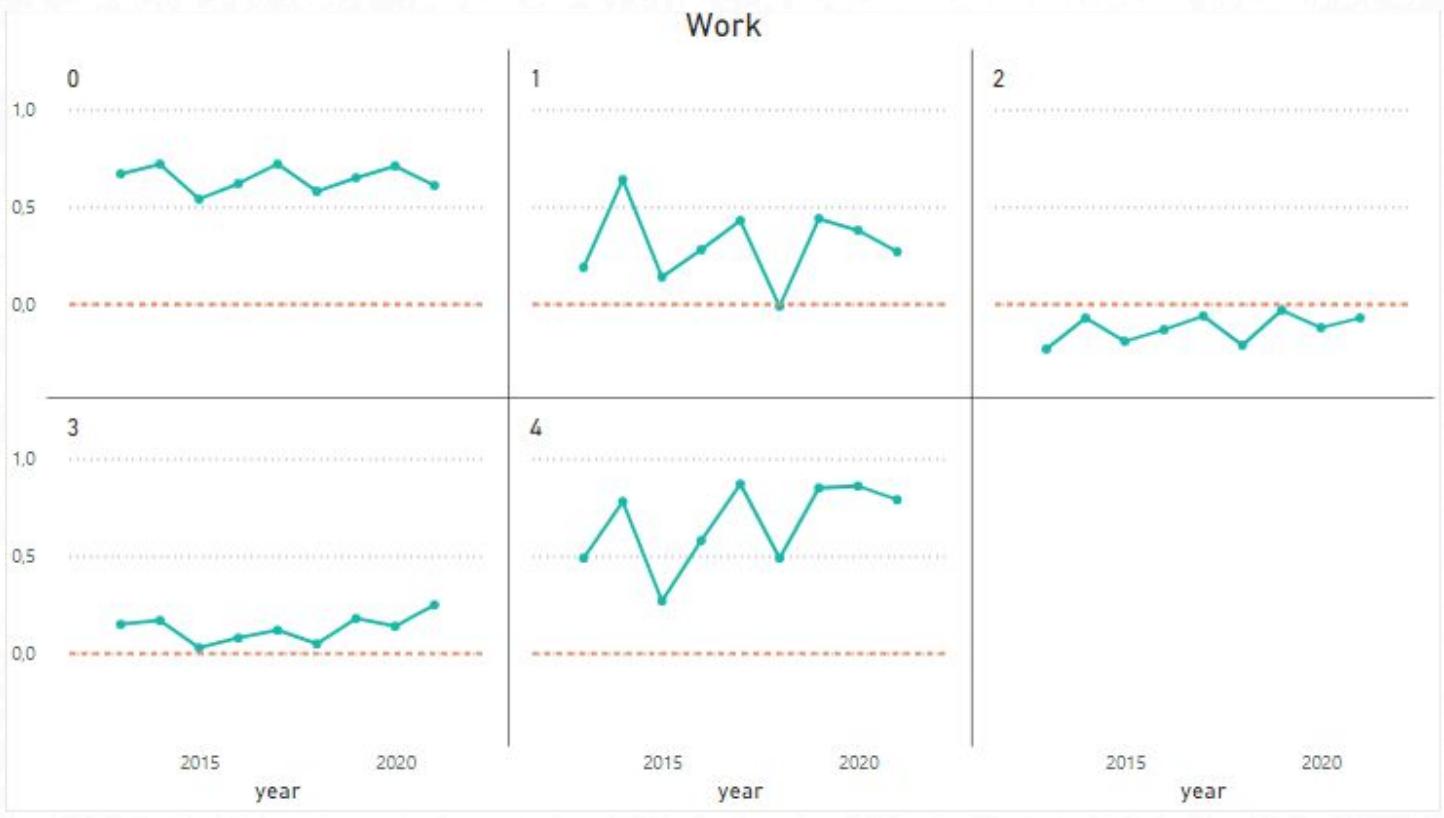


<https://eige.europa.eu/news/girls-greece-are-risk-female-genital-mutilation>

<https://eige.europa.eu/gender-based-violence/female-genital-mutilation/risk-estimations#2017>



Work



WORK

fulltime_employment_rate
duration_working_life
employed_in_education_health_socialactivities

flexibility_at_work
career_prospects

Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4

Employment rate of women and men in the EU

(as % of the population aged 20 to 64, 2018 data)



Proyecciones de carrera

Malta

MT-W	69
MT-M	67
EU-W	62
EU-M	63

Francia

FR-W	64
FR-M	67
EU-W	62
EU-M	63

Dinamarca

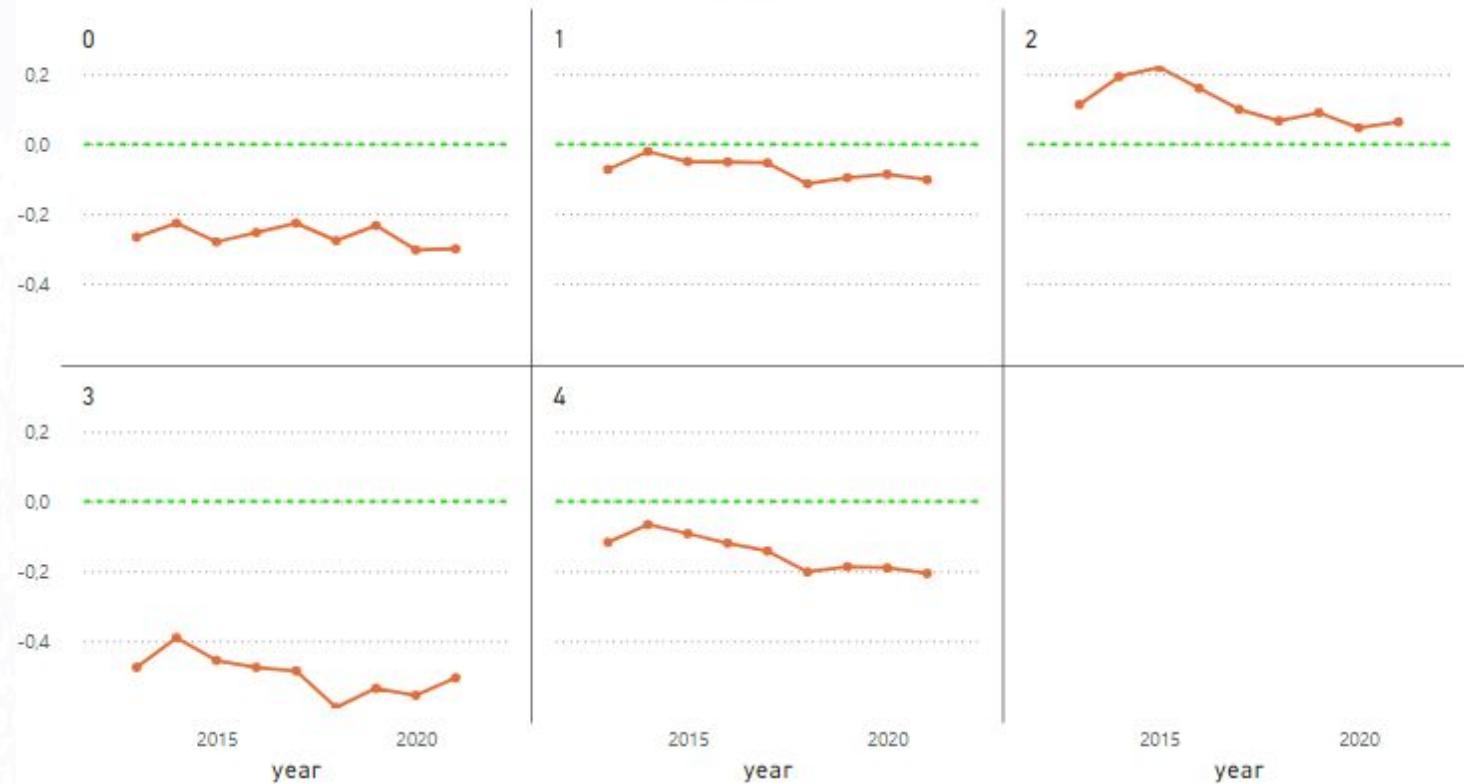
DK-W	70
DK-M	73
EU-W	62
EU-M	63

Alemania

DE-W	65
DE-M	68
EU-W	62
EU-M	63



Time



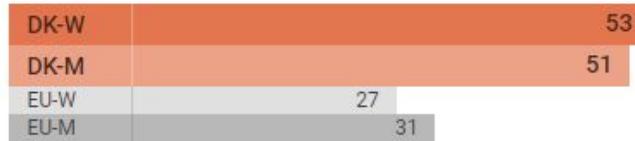
Country	cluster
Denmark	0
Finland	0
Sweden	0
Austria	1
Cyprus	1
Czechia	1
Hungary	1
Poland	1
Romania	1
Slovakia	1
Slovenia	1
Greece	2
Italy	2
Malta	2
Latvia	3
Lithuania	3
Belgium	4
Bulgaria	4
Croatia	4
Estonia	4
France	4
Germany	4
Ireland	4
Luxembourg	4
Netherlands	4
Portugal	4
Spain	4

TIME

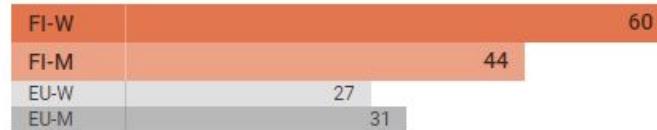
caring_kids_eldery_people_everyday
cooking_housework_everyday
doing_sport_cultural_activities
voluntary_activities

Realizar actividades culturales, deportivas y de ocio.

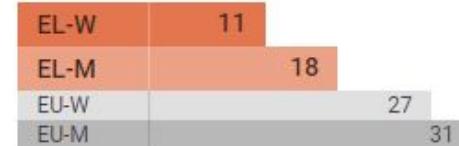
FINLANDIA



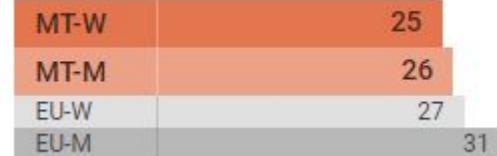
DINAMARCA



GRECIA



MALTA



ITALIA

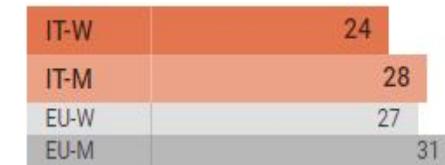
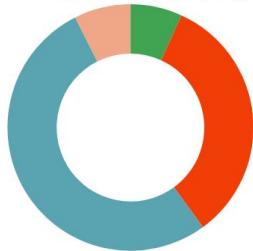
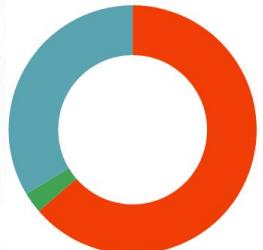


TABLA RESUMEN CLUSTERS



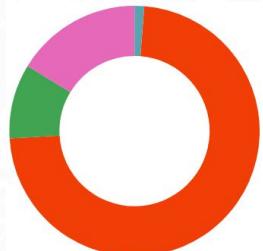
Cluster 0

Desigualdad en Work



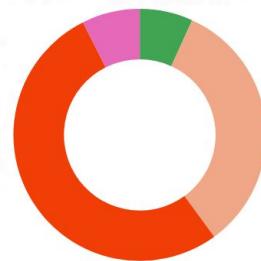
Cluster 1

Desigualdad casi nula en Health y Time



Cluster 2

MGF
Casi nula la desigualdad en Work



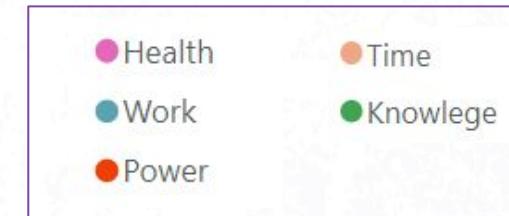
Cluster 3

Más desigualdad en mujeres en dedicación de cuidado y tareas del hogar



Cluster 4

Mayor desigualdad en Time como en Work



03

Objetivo

¿Cómo se calcula el índice de igualdad de género?

Modelo de Regresión - Cálculo del Score

ETL & Preprocesing

Selección del Dataset
de igualdad

Target: Score

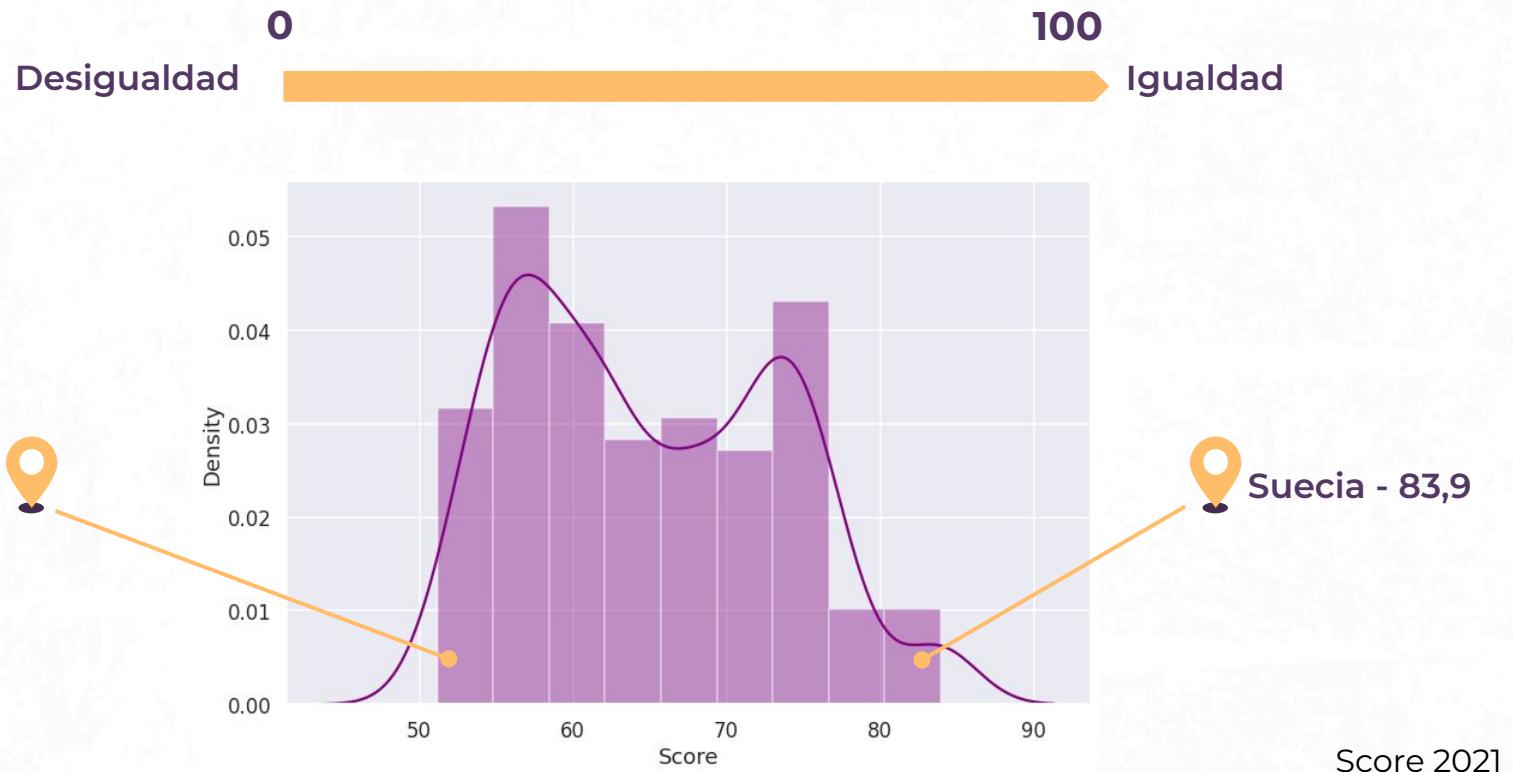
Modelización

Aplicación de
diferentes modelos
de regresión para ver
con cuál obtenemos
mejor resultado

Conclusiones

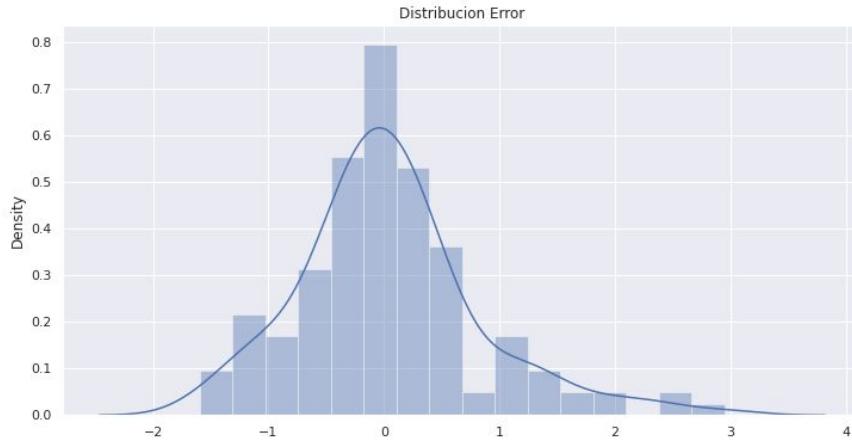
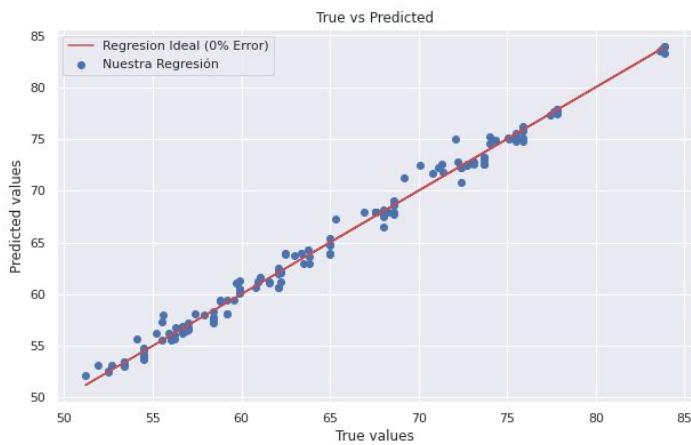
- Análisis de resultados
- Cálculo de pesos de variables

a) Análisis del target: **score igualdad**



b) Resultados de los modelos

	Model	MAE	MSE	RMSE	R2_SQUARE	Cross Validation
0	Linear Regression_inicial	2.339407	8.836586	2.972640	0.876986	0.895272
1	Linear Regression Final	0.564768	0.615009	0.784225	0.991438	0.991755
2	Lasso Regression	0.567146	0.617070	0.785538	0.991410	0.884231
3	Ridge Regression	0.564406	0.614546	0.783930	0.991445	0.987282
4	Elastic Net Regression	0.574508	0.631799	0.794858	0.991205	0.892222
5	XGBRegressor	0.564768	0.615009	0.784225	0.991438	0.991755
6	XGBRegressor2	0.564768	0.615009	0.784225	0.991438	0.991755



c) Fórmula del Score - Coefficients

	Coefficient				
Country_Sweden	2.763476	career_prospects	0.196655	people_in_education	-0.074802
Country_Denmark	1.656976	doing_sport_eating_healthy	0.140648	S20/S80_income	-0.119307
Country_Finland	1.599651	share_members_regional_assemblies	0.125721	good_self-perceived_health	-0.195190
Country_Netherlands	1.412400	voluntary_activities	0.112232	share_members_parliament	-0.219996
Country_France	1.376808	share_board_member_research_org	0.100190	year	-0.222371
Country_Spain	1.029441	cooking_housework_everyday	0.074511	doing_sport_cultural_activities	-0.236452
Country_Ireland	0.884354	fulltime_employment_rate	0.049407	risk_of_poverty	-0.245130
Country_Belgium	0.795712	caring_kids_eldery_people_everyday	0.023934	no_smoking_nor_harmful_drinking	-0.266136
Country_Luxembourg	0.774040	share_ministers	0.018466	mean_equivalised_income	-0.326140
tertiary_students_education_health_art_field	0.604438	share_board_member_public_org	0.010622	Country_Malta	-0.689565
employed_in_education_health_socialactivities	0.359681	flexibility_at_work	-0.003016	Country_Italy	-0.731811
life_expectancy	0.326680	share_board_members_central_bank	-0.006341	Country_Latvia	-0.870048
no_medical_examination	0.304666	Country_Germany	-0.007026	Country_Portugal	-0.964395
healthy_life_years	0.293961	tertiary_education_graduates	-0.024936	Sex	-1.059687
mean_monthly_earnings	0.264747	no_dental_examination	-0.054190	Country_Bulgaria	-1.282293
Country_Slovenia	0.212909	share_decision_making_body_olympic_org	-0.056212	Country_Estonia	-1.291267
career_prospects	0.196655	share_board_members_largest_companies	-0.066771	Country_Croatia	-1.550431
		duration_working_life	-0.071723	Country_Lithuania	-1.748540
				Country_Cyprus	-1.871554
				Country_Czechia	-2.113038

04

Objetivo

¿Cuál será la situación en 2030? ¿Algún país alcanzará la igualdad de género?

Series temporales - Predicción 2030

Dataprocessing

Groupby del score y los países

Modelización

- Aplicación prophet model para la predicción
- Cálculo del error

Conclusiones

- RMSE
- Ranking 2030
- Pocos datos

Score

01



02



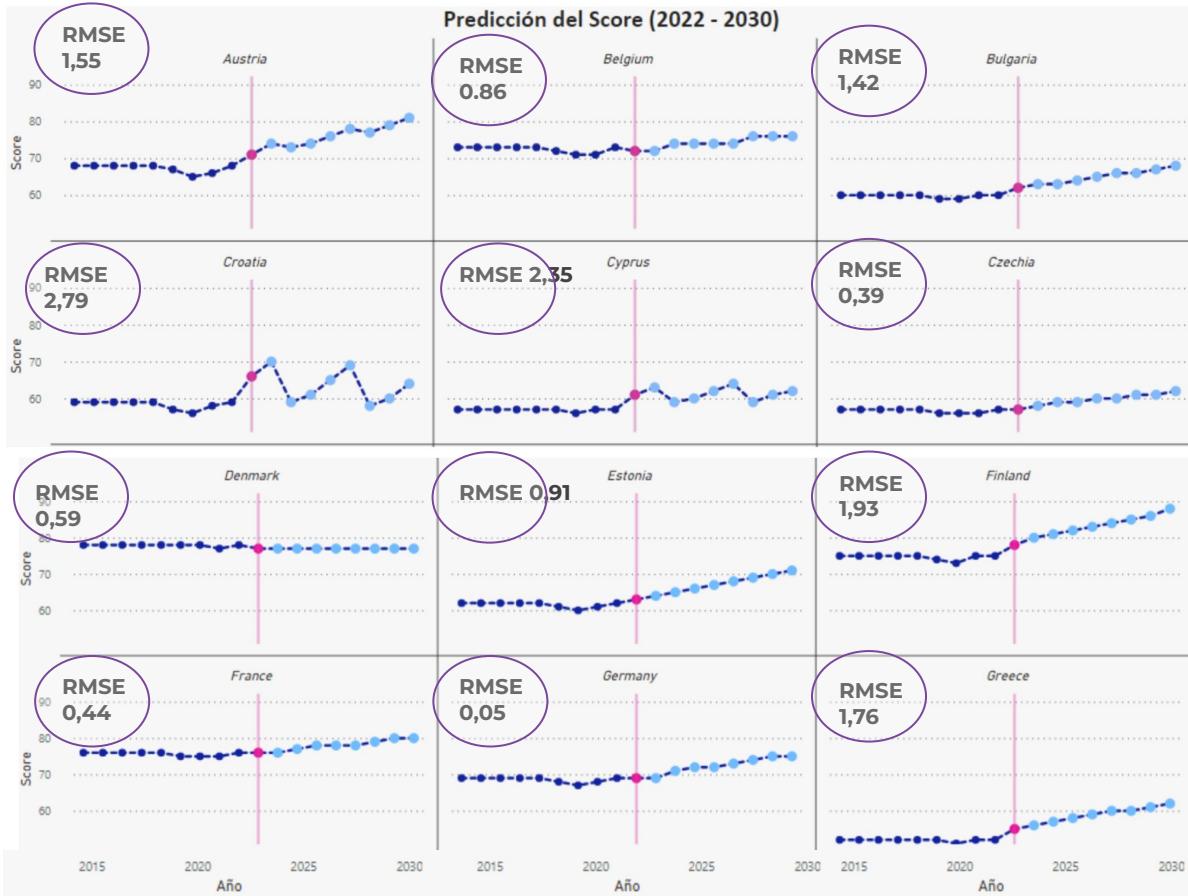
03



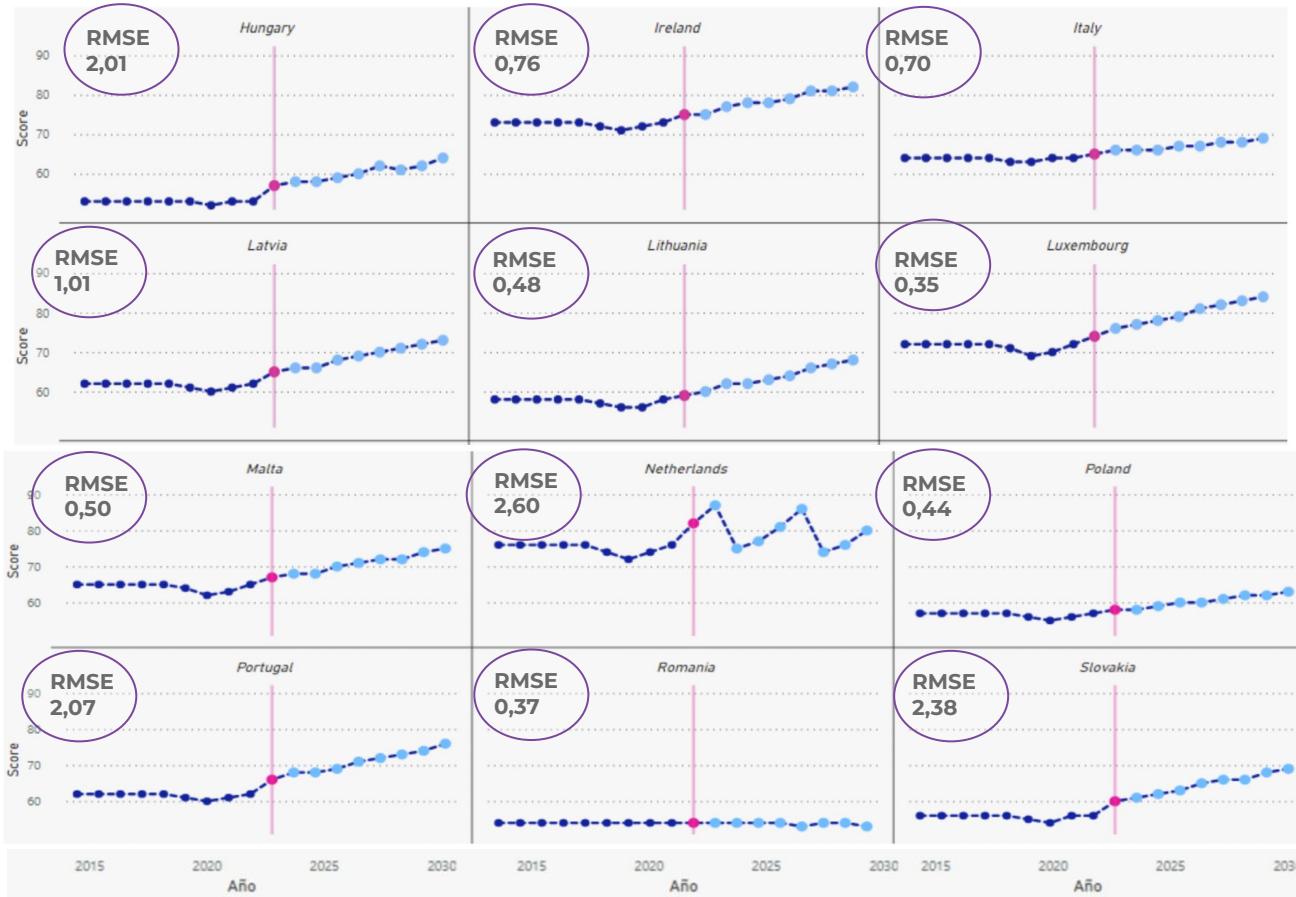
Score real 2021

Country	Score	Score_pred	Country	Score	Score_pred
Sweden	83.9	84.088550	Bulgaria	59.9	61.328692
Denmark	77.8	77.202148	Croatia	59.2	61.995978
Netherlands	75.9	78.505685	Lithuania	58.4	57.916361
France	75.5	75.949730	Cyprus	57.0	59.350397
Finland	75.3	77.238792	Czechia	56.7	57.092990
Spain	73.7	75.965502	Poland	56.6	57.045592
Ireland	73.1	73.863579	Slovakia	56.0	58.381729
Belgium	72.7	71.837349	Romania	54.5	54.128771
Luxembourg	72.4	72.752057	Hungary	53.4	55.413424
Germany	68.6	68.549340	Greece	52.5	54.264981
Austria	68.0	69.552416			
Slovenia	67.6	66.102663			
Malta	65.0	65.509764			
Italy	63.8	64.509275			
Portugal	62.2	64.279037			
Latvia	62.1	63.203141			
Estonia	61.6	62.510133			

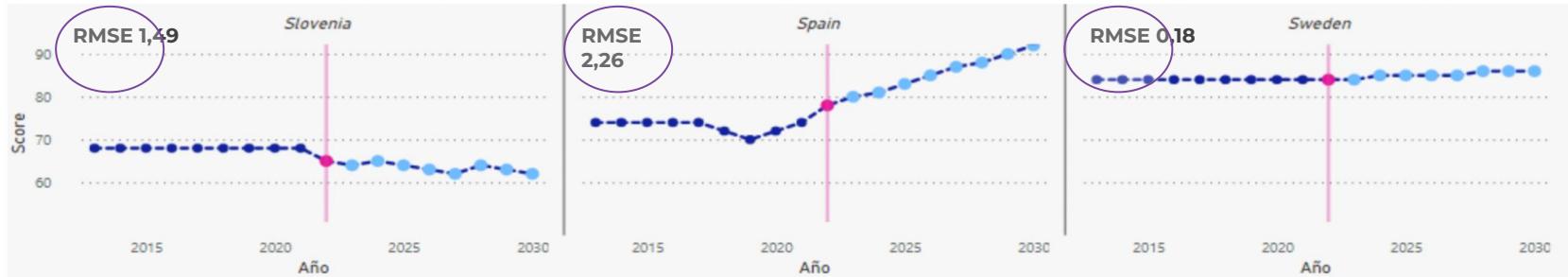
Resultados de la predicción



Resultados de la predicción



Resultados de la predicción



Mean RMSE

1.22

Mean Score

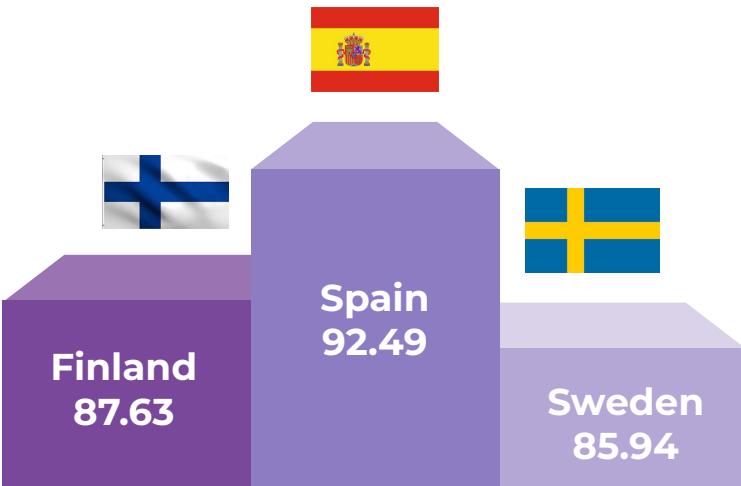
65.31



RMSE/Mean Score =

1.86%

Ranking final 2030



Country Score_2030

Spain	92.485616	▲ 5
Finland	87.632994	▲ 4
Sweden	85.942764	▼ 2
Luxembourg	84.379417	■ 0
Ireland	82.225853	▲ 2
Austria	80.796391	▲ 5
Netherlands	80.108178	▼ 4
France	80.090448	▼ 4
Denmark	77.292045	▼ 7
Belgium	76.491185	▼ 2
Portugal	75.708694	▲ 4
Germany	75.159033	▼ 2
Malta	74.828847	■ 0
Latvia	73.306258	▲ 2
Estonia	70.596034	▲ 2
Slovakia	69.284041	▲ 8
Italy	68.583280	▼ 2
Lithuania	67.897479	▲ 2
Bulgaria	67.590011	▼ 1
Hungary	63.787941	▲ 6
Croatia	63.715712	▼ 2
Poland	62.870568	▲ 1
Cyprus	62.436822	▼ 2
Greece	61.942113	▲ 3
Slovenia	61.852895	▼ 13
Czechia	61.605680	▼ 4
Romania	53.294456	▼ 2



Impacto Covid-19



Más estrés laboral y carga familiar



Pérdidas de trabajo



Más mujeres en sectores esenciales



Igualdad
de género

Conclusiones

01

OBJETIVO: Feature Importance

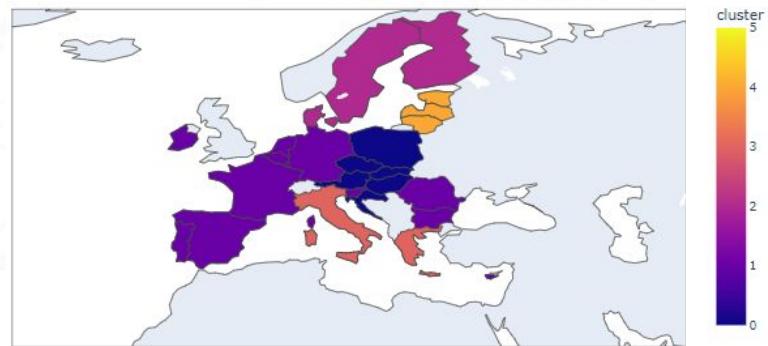
- Employed_education_health_socialactivities
- Caring_kids_eldery_people_everyday
- Mean_monthly_earnings
- Share_ministers



02

OBJETIVO: Clustering

5 clusters diferenciados.



Conclusiones

03

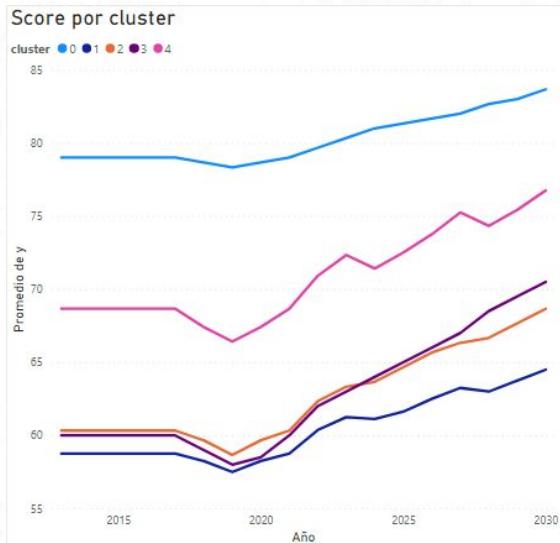
OBJETIVO: Fórmula del Score Igualdad

- Ridge Regression
- Importancia del peso de los países.

04

OBJETIVO: Predicción 2030

- Ningún país alcanza la igualdad de género.
- Necesidad de aplicar políticas que reduzcan la brecha entre sexos.



Work



Full-time employment rate is higher for men

Money



Knowledge



Fewer young men than women take part in adult learning

Time



Women are more likely to do regular housework compared to employed men

Power



of board members among the largest EU companies are women

Health



Women aged over 65 have greater unmet needs for medical examinations than men the same age

Gracias

