

Evaluación

Data Science - Grupo 3

Análisis Demográfico España

Amir

Laura

Enric

## 1. Fase de extracción

Se extrae el conjunto de datos 9689.csv del sitio web del INE (Instituto Nacional de Estadística) "Población por provincia, sexo y edad".

```
df= pd.read_csv('./data/9689.csv', sep= ';')  
  
df.head(10)
```

✓ 1.9s

	Provincias	Grupo quinquenal de edad	Nacionalidad	Sexo	Periodo	Total
0	Total Nacional	Total	Total	Ambos sexos	1 de julio de 2022	47.615.033,864697
1	Total Nacional	Total	Total	Ambos sexos	1 de enero de 2022	47.432.892,943294
2	Total Nacional	Total	Total	Ambos sexos	1 de julio de 2021	47.331.301,774407
3	Total Nacional	Total	Total	Ambos sexos	1 de enero de 2021	47.398.694,945392
4	Total Nacional	Total	Total	Ambos sexos	1 de julio de 2020	47.355.684,576847
5	Total Nacional	Total	Total	Ambos sexos	1 de enero de 2020	47.332.613,88819

## 2. Fase de transformación/modelación

Eliminamos las filas con valores vacíos

```
df.dropna(inplace=True, axis = 0)  
df
```

✓ 0.3s

Transformamos campos(Periodo y Total) a de tipo object a tipo int64

```
def Eliminar_coma(x):  
    x= x.replace('.', '')  
    x = x.replace(',', '.')  
    return x
```

```
df['Total'] = df['Total'].apply(Eliminar_coma)  
df['Total'] = df['Total'].apply(lambda x : int(float (x)))
```

✓ 1.5s

#	Column	Non-Null Count	Dtype
0	Provincias	734580 non-null	object
1	Grupo quinquenal de edad	734580 non-null	object
2	Nacionalidad	734580 non-null	object
3	Sexo	734580 non-null	object
4	Periodo	734580 non-null	int64
5	Total	734580 non-null	int64

## 2.1 Análisis poblacional en periodo 2002-2022

	Media	Mediana	Des. Estándar	Max	Min
Total	45609439.57	46486621.0	1918337.51	47432892	41035270
Hombre	22441773.71	22877461.0	942764.72	23237151	20115521
Mujer	23167665.38	23623018.0	981874.45	24195741	20919749

En la tabla anterior se muestran distintas estadísticas de la población de España durante el periodo 2002-2022.

## Análisis población 2019-2022

### TVA

```
P2 = sum(df_T2019['Total'])
P1 = sum(df_T2022['Total'])

tva = ((P1-P2)/P1)*100

tva
✓ 0.0s
1.0453357977835298
```

P2 = población total en 2019 = 46937059

P1 = población total en 2022 = 47432892

TVA = tasa variación anual = 1.0453357977835298

### TDD

```
TDD_2019 = Pob_Dep_2019/Pob_Act_2019
TDD_2022 = Pob_Dep_2022/Pob_Act_2022

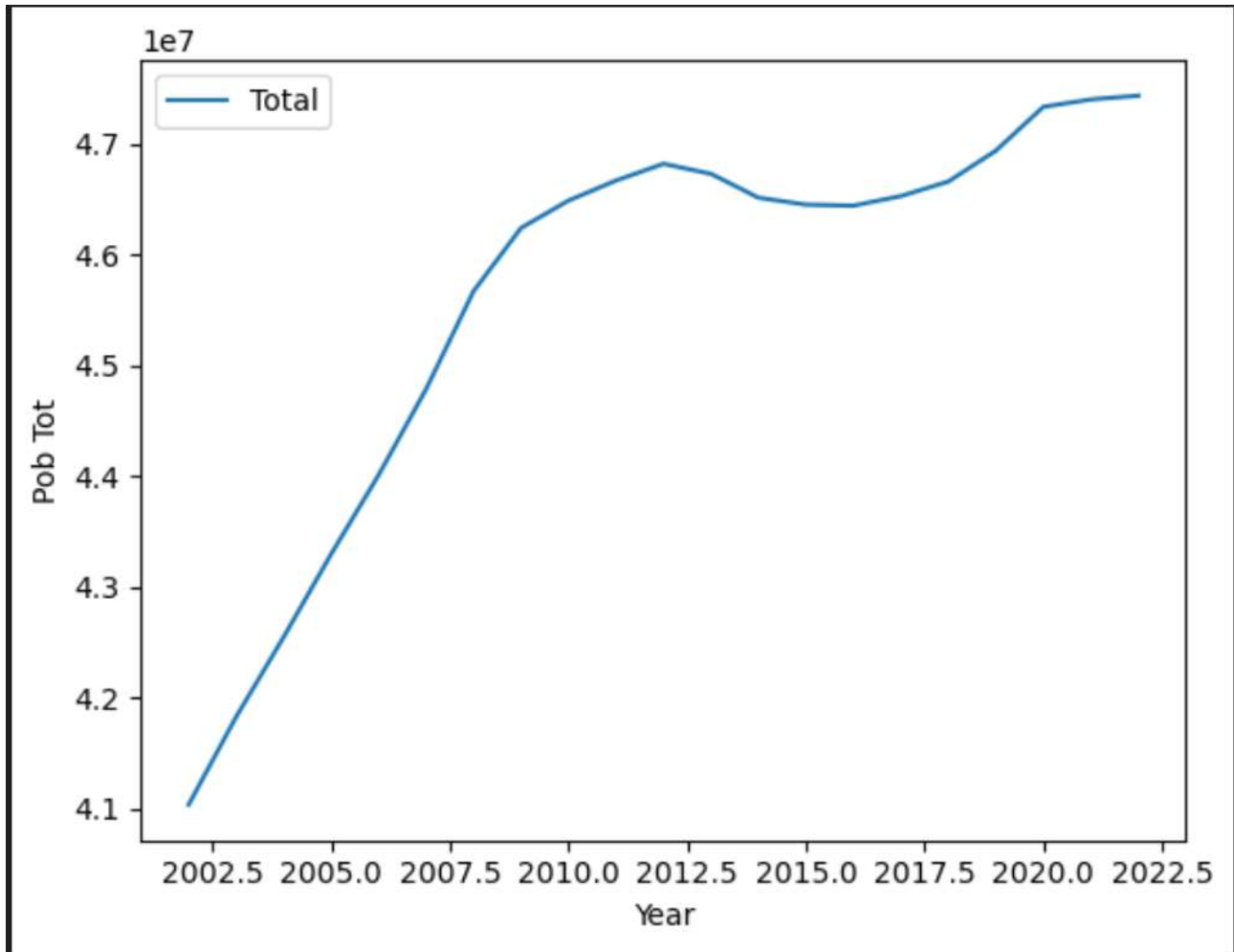
print(TDD_2019)
print(TDD_2022)
✓ 0.0s
```

TDD = Tasa de Dependencia Demográfica → 2019 = 0.51893  
→ 2022 = 0.51647

Diferencia entre TDD 2019 y TDD 2022 = 0,00246

Es decir, la TDD disminuyó un 0,00246

### 3.Fase visualización y análisis de datos



-Vemos una subida continua hasta el 2008, coincidiendo con el inicio de la crisis inmobiliaria, que desemboca en un descenso en el crecimiento de la población.

- A Partir del 2016 observamos otro crecimiento, interrumpido esta vez en el 2020 por la crisis del COVID iniciada en 2019.