exploración de datos, creación de variables y procesamiento base OCDE2021

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

In [2]:
```
GEIH_2022 = pd.read_csv("GEIH_2022.csv")
SalarioMedio = pd.read_csv("baseingreso.csv")
GEIH_2022 = pd.merge(GEIH_2022, SalarioMedio, on=["DIRECTORIO",'SECUENCIA_P', 'ORDEN']
GEIH_2022
```

Out[2]:

	Unnamed: 0_x	DIRECTORIO	SECUENCIA_P	ORDEN	Sexo	JefeHogar	Etnia	CotizaEPS	Pa
0	0	5000000	1	1	Hombre	JEFEHO	Ninguno	2	
1	1	5000001	1	1	Hombre	JEFEHO	Ninguno	1	
2	2	5000001	1	2	Mujer	2	Ninguno	1	
3	3	5000001	1	3	Hombre	3	Ninguno	1	
4	4	5000002	1	1	Mujer	JEFEHO	Ninguno	2	
...	...	...	...	...	...	...	...	...	
382456	30500	7309048	1	2	Hombre	2	Ninguno	1	
382457	30501	7309048	1	3	Hombre	3	Ninguno	1	
382458	30502	7309049	1	1	Mujer	JEFEHO	Ninguno	1	
382459	30503	7309050	1	1	Hombre	JEFEHO	Ninguno	1	
382460	30504	7309050	1	2	Mujer	2	Ninguno	2	

382461 rows × 26 columns

En las siguientes celdas voy a calcular la informalidad por departamento, creando nuevos dataframes por departamento através de filtraciones y asignando a una variable, los cuales también me servirán más adelante.

In [3]:
```
Total_Expansion = GEIH_2022["FactorExpansion"].sum()
```

In [4]:
```
Antioquia = GEIH_2022["Departamento"] == "Antioquia"
Antioquia = GEIH_2022[Antioquia]
```

In [5]:
```
Ponderacion_Antioquia = Antioquia["Ponderacion"].sum()
Expansion_Antioquia = Antioquia["FactorExpansion"].sum()
```

```
AntioquiaOI = (Ponderacion_Antioquia / Expansion_Antioquia ) * 100
AntioquiaOI
```

Out[5]:
```
46.38657345146848
```

In [6]:
```
Antioquia_proporcion = (Ponderacion_Antioquia / Total_Expansion ) * 100
Antioquia_proporcion
```

Out[6]:
```
6.498087656965194
```

In [7]:
```
Antioquia_Dict1 = {"AntioquiaOI": 46.386}
Antioquia_Dict2 = {"Antioquia_proporcion": 6.49}
print(Antioquia_Dict1, Antioquia_Dict2)
```

```
{'AntioquiaOI': 46.386} {'Antioquia_proporcion': 6.49}
```

In [8]:
```
Amazonas = GEIH_2022["Departamento"] == "Amazonas"
Amazonas = GEIH_2022[Amazonas]
```

In [9]:
```
Ponderacion_Amazonas = Amazonas["Ponderacion"].sum()
Expansion_Amazonas = Amazonas["FactorExpansion"].sum()
AmazonasOI = (Ponderacion_Amazonas / Expansion_Amazonas ) * 100
AmazonasOI
```

Out[9]:
```
56.530141951917344
```

In [10]:
```
Amazonas_proporcion = (Ponderacion_Amazonas / Total_Expansion ) * 100
Amazonas_proporcion
```

Out[10]:
```
0.03564094090630834
```

In [11]:
```
Amazonas_Dict1 = {"AmazonasOI": 56.53}
Amazonas_Dict2 = {"Amazonas_proporcion": 0.035}
print(Amazonas_Dict1, Amazonas_Dict2)
```

```
{'AmazonasOI': 56.53} {'Amazonas_proporcion': 0.035}
```

In [12]:
```
Arauca = GEIH_2022["Departamento"] == "Arauca"
Arauca = GEIH_2022[Arauca]
```

In [13]:
```
Ponderacion_Arauca = Arauca["Ponderacion"].sum()
Expansion_Arauca = Arauca["FactorExpansion"].sum()
AraucaOI = (Ponderacion_Arauca / Expansion_Arauca ) * 100
AraucaOI
```

Out[13]:
```
55.438692242022
```

In [14]:
```
Arauca_proporcion = (Ponderacion_Arauca / Total_Expansion ) * 100
Arauca_proporcion
```

Out[14]:
```
0.0733639703812805
```

In [15]:
```
Arauca_Dict1 = {"AraucaOI": 55.43}
Arauca_Dict2 = {"Arauca_proporcion": 0.07}
print(Arauca_Dict1, Arauca_Dict2)
```

```
{'AraucaOI': 55.43} {'Arauca_proporcion': 0.07}
```

In [16]:
```
Atlantico = GEIH_2022["Departamento"] == "Atlantico"
Atlantico = GEIH_2022[Atlantico]
```

In [17]:
```
Ponderacion_Atlantico = Atlantico["Ponderacion"].sum()
Expansion_Atlantico = Atlantico["FactorExpansion"].sum()
AtlanticoOI = (Ponderacion_Atlantico / Expansion_Atlantico ) * 100
AtlanticoOI
```

Out[17]:
```
56.29104154232372
```

In [18]:
```
Atlantico_proporcion = (Ponderacion_Atlantico / Total_Expansion ) * 100
Atlantico_proporcion
```

Out[18]:
```
3.163019294081505
```

In [19]:
```
Atlantico_Dict1 = {"AtlanticoOI": 56.29}
Atlantico_Dict2 = {"Atlantico_proporcion": 3.16}
print(Atlantico_Dict1, Atlantico_Dict2)
```

```
{'AtlanticoOI': 56.29} {'Atlantico_proporcion': 3.16}
```

In [20]:
```
Bogota = GEIH_2022["Departamento"] == "Bogota"
Bogota = GEIH_2022[Bogota]
```

In [21]:
```
Ponderacion_Bogota = Bogota["Ponderacion"].sum()
Expansion_Bogota = Bogota["FactorExpansion"].sum()
BogotaOI = (Ponderacion_Bogota / Expansion_Bogota ) * 100
BogotaOI
```

Out[21]:
```
34.043009971725205
```

In [22]:
```
Bogota_proporcion = (Ponderacion_Bogota / Total_Expansion ) * 100
Bogota_proporcion
```

Out[22]:
```
5.9081138387870675
```

In [23]:
```
Bogota_Dict1 = {"BogotaOI": 34.04}
Bogota_Dict2 = {"Bogota_proporcion": 5.9}
print(Bogota_Dict1, Bogota_Dict2)
```

```
{'BogotaOI': 34.04} {'Bogota_proporcion': 5.9}
```

In [24]:
```
Bolivar = GEIH_2022["Departamento"] == "Bolivar"
Bolivar = GEIH_2022[Bolivar]
```

In [25]:
```
Ponderacion_Bolivar = Bolivar["Ponderacion"].sum()
Expansion_Bolivar = Bolivar["FactorExpansion"].sum()
BolivarOI = (Ponderacion_Bolivar / Expansion_Bolivar ) * 100
BolivarOI
```

Out[25]:
```
66.46884998271464
```

In [26]:
```
Bolivar_proporcion = (Ponderacion_Bolivar / Total_Expansion ) * 100
Bolivar_proporcion
```

Out[26]:
```
2.8037539161212655
```

In [27]:
```
Bolivar_Dict1 = {"BolivarOI": 66.46}
Bolivar_Dict2 = {"Bolivar_proporcion": 2.80}
print(Bolivar_Dict1, Bolivar_Dict2)
```

```
{'BolivarOI': 66.46} {'Bolivar_proporcion': 2.8}
```

In [28]:
```
Boyaca = GEIH_2022["Departamento"] == "Boyaca"
Boyaca = GEIH_2022[Boyaca]
```

In [29]:
```
Ponderacion_Boyaca = Boyaca["Ponderacion"].sum()
Expansion_Boyaca = Boyaca["FactorExpansion"].sum()
BoyacaOI = (Ponderacion_Boyaca / Expansion_Boyaca ) * 100
BoyacaOI
```

Out[29]:
```
61.63283850996487
```

In [30]:
```
Boyaca_proporcion = (Ponderacion_Boyaca / Total_Expansion ) * 100
Boyaca_proporcion
```

Out[30]:
```
1.5383176184199274
```

In [31]:
```
Boyaca_Dict1 = {"BoyacaOI": 61.63}
Boyaca_Dict2 = {"Boyaca_proporcion": 1.53}
print(Boyaca_Dict1, Boyaca_Dict2)
```

```
{'BoyacaOI': 61.63} {'Boyaca_proporcion': 1.53}
```

In [32]:
```
Caldas = GEIH_2022["Departamento"] == "Caldas"
Caldas = GEIH_2022[Caldas]
```

In [33]:
```
Ponderacion_Caldas= Caldas["Ponderacion"].sum()
Expansion_Caldas = Caldas["FactorExpansion"].sum()
CaldasOI = (Ponderacion_Caldas / Expansion_Caldas ) * 100
CaldasOI
```

Out[33]:
```
45.801086670470696
```

In [34]:
```
Caldas_proporcion = (Ponderacion_Caldas / Total_Expansion ) * 100
Caldas_proporcion
```

Out[34]:
```
0.9073280655993209
```

In [35]:
```
Caldas_Dict1 = {"CaldasOI": 45.80}
Caldas_Dict2 = {"Caldas_proporcion": 0.90}
print(Caldas_Dict1, Caldas_Dict2)
```

```
{'CaldasOI': 45.8} {'Caldas_proporcion': 0.9}
```

In [36]:
```
Caqueta = GEIH_2022["Departamento"] == "Caqueta"
Caqueta = GEIH_2022[Caqueta]
```

In [37]:
```
Ponderacion_Caqueta= Caqueta["Ponderacion"].sum()
Expansion_Caqueta = Caqueta["FactorExpansion"].sum()
CaquetaOI = (Ponderacion_Caqueta / Expansion_Caqueta) * 100
CaquetaOI
```

Out[37]:
```
67.76199436457188
```

In [38]:
```
Caqueta_proporcion = (Ponderacion_Caqueta / Total_Expansion ) * 100
Caqueta_proporcion
```

Out[38]:
```
0.4978812507998306
```

In [39]:
```
Caqueta_Dict1 = {"CaquetaOI": 67.76}
Caqueta_Dict2 = {"Caqueta_proporcion": 0.49}
print(Caqueta_Dict1, Caqueta_Dict2)
```

```
{'CaquetaOI': 67.76} {'Caqueta_proporcion': 0.49}
```

In [40]:
```
Casanare = GEIH_2022["Departamento"] == "Casanare"
Casanare = GEIH_2022[Casanare]
```

In [41]:
```
Ponderacion_Casanare = Casanare["Ponderacion"].sum()
Expansion_Casanare = Casanare["FactorExpansion"].sum()
CasanareOI = (Ponderacion_Casanare / Expansion_Casanare) * 100
CasanareOI
```

Out[41]:
```
54.79327782256384
```

In [42]:
```
Casanare_proporcion = (Ponderacion_Casanare / Total_Expansion ) * 100
Casanare_proporcion
```

Out[42]:
```
0.19371167347193727
```

In [43]:
```
Casanare_Dict1 = {"CasanareOI": 54.79}
Casanare_Dict2 = {"Casanare_proporcion": 0.19}
print(Casanare_Dict1, Casanare_Dict2)
```

```
{'CasanareOI': 54.79} {'Casanare_proporcion': 0.19}
```

In [44]:
```
Cauca = GEIH_2022["Departamento"] == "Cauca"
Cauca = GEIH_2022[Cauca]
```

In [45]:
```
Ponderacion_Cauca = Cauca["Ponderacion"].sum()
Expansion_Cauca = Cauca["FactorExpansion"].sum()
CaucaOI = (Ponderacion_Cauca / Expansion_Cauca) * 100
CaucaOI
```

Out[45]:
```
68.5320051962991
```

In [46]:
```
Cauca_proporcion = (Ponderacion_Cauca / Total_Expansion ) * 100
Cauca_proporcion
```

Out[46]:
```
2.003128938731382
```

In [47]:
```
Cauca_Dict1 = {"CaucaOI": 68.53}
Cauca_Dict2 = {"Cauca_proporcion": 2.0}
print(Cauca_Dict1, Cauca_Dict2)
```

```
{'CaucaOI': 68.53} {'Cauca_proporcion': 2.0}
```

In [48]:
```
Cesar = GEIH_2022["Departamento"] == "Cesar"
Cesar = GEIH_2022[Cesar]
```

In [49]:
```
Ponderacion_Cesar = Cesar["Ponderacion"].sum()
Expansion_Cesar = Cesar["FactorExpansion"].sum()
CesarOI = (Ponderacion_Cesar / Expansion_Cesar) * 100
CesarOI
```

Out[49]:
```
63.64707488063358
```

In [50]:
```
Cesar_proporcion = (Ponderacion_Cesar / Total_Expansion ) * 100
Cesar_proporcion
```

Out[50]:
```
1.3904930133130755
```

In [51]:
```
Cesar_Dict1 = {"CesarOI": 63.64}
Cesar_Dict2 = {"Cesar_proporcion": 1.39}
print(Cesar_Dict1, Cesar_Dict2)
```

```
{'CesarOI': 63.64} {'Cesar_proporcion': 1.39}
```

In [52]:
```
Choco = GEIH_2022["Departamento"] == "Choco"
Choco = GEIH_2022[Choco]
```

In [53]:
```
Ponderacion_Choco = Choco["Ponderacion"].sum()
Expansion_Choco = Choco["FactorExpansion"].sum()
ChocoOI = (Ponderacion_Choco/ Expansion_Choco) * 100
ChocoOI
```

Out[53]:
```
67.90588155052221
```

In [54]:
```
Choco_proporcion = (Ponderacion_Choco / Total_Expansion ) * 100
Choco_proporcion
```

Out[54]:
```
0.45284206799815857
```

In [55]:
```
Choco_Dict1 = {"ChocoOI": 67.9}
Choco_Dict2 = {"Choco_proporcion": 0.45}
print(Choco_Dict1, Choco_Dict2)
```

```
{'ChocoOI': 67.9} {'Choco_proporcion': 0.45}
```

In [56]:
```
Cordoba = GEIH_2022["Departamento"] == "Cordoba"
Cordoba = GEIH_2022[Cordoba]
```

In [57]:
```
Ponderacion_Cordoba = Cordoba["Ponderacion"].sum()
Expansion_Cordoba = Cordoba["FactorExpansion"].sum()
CordobaOI = (Ponderacion_Cordoba/ Expansion_Cordoba) * 100
CordobaOI
```

Out[57]:
```
68.90075980556057
```

In [58]:
```
Cordoba_proporcion = (Ponderacion_Cordoba / Total_Expansion ) * 100
Cordoba_proporcion
```

Out[58]:
```
2.381319335244939
```

In [59]:
```
Cordoba_Dict1 = {"CordobaOI": 68.9}
Cordoba_Dict2 = {"Cordoba_proporcion": 2.38}
print(Cordoba_Dict1, Cordoba_Dict2)
```

```
{'CordobaOI': 68.9} {'Cordoba_proporcion': 2.38}
```

In [60]:
```
Cundinamarca = GEIH_2022["Departamento"] == "Cundinamarca"
Cundinamarca = GEIH_2022[Cundinamarca]
```

In [61]:
```
Ponderacion_Cundinamarca = Cundinamarca["Ponderacion"].sum()
Expansion_Cundinamarca = Cundinamarca["FactorExpansion"].sum()
CundinamarcaOI = (Ponderacion_Cundinamarca / Expansion_Cundinamarca) * 100
CundinamarcaOI
```

Out[61]:
```
45.405352417363694
```

In [62]:
```
Cundinamarca_proporcion = (Ponderacion_Cundinamarca / Total_Expansion ) * 100
Cundinamarca_proporcion
```

Out[62]:
```
3.0819419658527965
```

In [63]:
```
Cundinamarca_Dict1 = {"CundinamarcaOI": 45.4}
Cundinamarca_Dict2 = {"Cundinamarca_proporcion": 3.08}
print(Cundinamarca_Dict1, Cundinamarca_Dict2)
```

```
{'CundinamarcaOI': 45.4} {'Cundinamarca_proporcion': 3.08}
```

In [64]:
```
Guainia = GEIH_2022["Departamento"] == "Guainia"
Guainia = GEIH_2022[Guainia]
```

In [65]:
```
Ponderacion_Guainia = Guainia["Ponderacion"].sum()
Expansion_Guainia = Guainia["FactorExpansion"].sum()
GuainiaOI = (Ponderacion_Guainia/ Expansion_Guainia) * 100
GuainiaOI
```

Out[65]:
```
67.31040237579128
```

In [66]:
```
Guainia_proporcion = (Ponderacion_Guainia / Total_Expansion ) * 100
Guainia_proporcion
```

Out[66]:
```
0.027203078415667978
```

In [67]:
```
Guainia_Dict1 = {"GuainiaOI": 67.31}
Guainia_Dict2 = {"Guainia_proporcion": 0.02}
print(Guainia_Dict1, Guainia_Dict2)
```

```
{'GuainiaOI': 67.31} {'Guainia_proporcion': 0.02}
```

In [68]:
```
Guaviare= GEIH_2022["Departamento"] == "Guaviare"
Guaviare = GEIH_2022[Guaviare]
```

In [69]:
```
Ponderacion_Guaviare = Guaviare["Ponderacion"].sum()
Expansion_Guaviare = Guaviare["FactorExpansion"].sum()
GuaviareOI = (Ponderacion_Guaviare/ Expansion_Guaviare) * 100
GuaviareOI
```

Out[69]:
```
57.96794038386547
```

In [70]:
```
Guaviare_proporcion = (Ponderacion_Guaviare / Total_Expansion ) * 100
Guaviare_proporcion
```

Out[70]:
```
0.05920769228388875
```

In [71]:
```
Guaviare_Dict1 = {"GuaviareOI": 57.96}
Guaviare_Dict2 = {"Guaviare_proporcion": 0.05}
print(Guaviare_Dict1, Guaviare_Dict2)
```

```
{'GuaviareOI': 57.96} {'Guaviare_proporcion': 0.05}
```

In [72]:
```
Huila = GEIH_2022["Departamento"] == "Huila"
Huila = GEIH_2022[Huila]
```

In [73]:
```
Ponderacion_Huila = Huila["Ponderacion"].sum()
Expansion_Huila = Huila["FactorExpansion"].sum()
HuilaOI = (Ponderacion_Huila/ Expansion_Huila) * 100
HuilaOI
```

Out[73]:
```
68.72306053091008
```

In [74]:
```
Huila_proporcion = (Ponderacion_Huila / Total_Expansion ) * 100
Huila_proporcion
```

Out[74]:
```
1.409507951100284
```

In [75]:
```
Huila_Dict1 = {"HuilaOI": 68.72}
Huila_Dict2 = {"Huila_proporcion": 1.4}
print(Huila_Dict1, Huila_Dict2)
```

```
{'HuilaOI': 68.72} {'Huila_proporcion': 1.4}
```

In [76]:
```
Guajira = GEIH_2022["Departamento"] == "Guajira"
Guajira = GEIH_2022[Guajira]
```

In [77]:
```
Ponderacion_Guajira = Guajira["Ponderacion"].sum()
Expansion_Guajira = Guajira["FactorExpansion"].sum()
GuajiraOI = (Ponderacion_Guajira/ Expansion_Guajira) * 100
GuajiraOI
```

Out[77]:
```
76.74526964252016
```

In [78]:
```
Guajira_proporcion = (Ponderacion_Guajira/ Total_Expansion ) * 100
Guajira_proporcion
```

Out[78]:
```
1.3394426608401617
```

In [79]:
```
Guajira_Dict1 = {"GuajiraOI": 76.74}
Guajira_Dict2 = {"Guajira_proporcion": 1.3}
print(Guajira_Dict1, Guajira_Dict2)
```

```
{'GuajiraOI': 76.74} {'Guajira_proporcion': 1.3}
```

In [80]:
```
Magdalena = GEIH_2022["Departamento"] == "Magdalena"
Magdalena = GEIH_2022[Magdalena]
```

In [81]:
```
Ponderacion_Magdalena = Magdalena["Ponderacion"].sum()
Expansion_Magdalena = Magdalena["FactorExpansion"].sum()
MagdalenaOI = (Ponderacion_Magdalena/ Expansion_Magdalena) * 100
MagdalenaOI
```

Out[81]:
```
70.89073108924782
```

In [82]:
```
Magdalena_proporcion = (Ponderacion_Magdalena/ Total_Expansion ) * 100
Magdalena_proporcion
```

Out[82]:
```
1.9070754326610948
```

In [83]:
```
Magdalena_Dict1 = {"MagdalenaOI": 70.89}
Magdalena_Dict2 = {"Magdalena_proporcion": 1.9}
print(Magdalena_Dict1, Magdalena_Dict2)
```

```
{'MagdalenaOI': 70.89} {'Magdalena_proporcion': 1.9}
```

In [84]:
```
Meta = GEIH_2022["Departamento"] == "Meta"
Meta = GEIH_2022[Meta]
```

In [85]:
```
Ponderacion_Meta = Meta["Ponderacion"].sum()
Expansion_Meta= Meta["FactorExpansion"].sum()
MetaOI = (Ponderacion_Meta/ Expansion_Meta) * 100
MetaOI
```

Out[85]:
```
55.631705735447824
```

In [86]:
```
Meta_proporcion = (Ponderacion_Meta/ Total_Expansion ) * 100
Meta_proporcion
```

Out[86]:
```
1.2279833131506666
```

In [87]:
```
Meta_Dict1 = {"MetaOI": 55.63}
Meta_Dict2 = {"Meta_proporcion": 1.22}
print(Meta_Dict1, Meta_Dict2)
```

```
{'MetaOI': 55.63} {'Meta_proporcion': 1.22}
```

In [88]:
```
Narino = GEIH_2022["Departamento"] == "Narino"
Narino = GEIH_2022[Narino]
```

In [89]:
```
Ponderacion_Narino = Narino["Ponderacion"].sum()
Expansion_Narino = Narino["FactorExpansion"].sum()
NarinoOI = (Ponderacion_Narino/ Expansion_Narino) * 100
NarinoOI
```

Out[89]:
```
79.11475242701762
```

In [90]:
```
Narino_proporcion = (Ponderacion_Narino/ Total_Expansion ) * 100
Narino_proporcion
```

Out[90]:
```
3.2266659601665713
```

In [91]:
```
Narino_Dict1 = {"NarinoOI": 79.11}
Narino_Dict2 = {"Narino_proporcion": 3.22}
print(Narino_Dict1, Narino_Dict2)
```

```
{'NarinoOI': 79.11} {'Narino_proporcion': 3.22}
```

In [92]:
```
NorteSantander= GEIH_2022["Departamento"] == "NorteSantander"
NorteSantander = GEIH_2022[NorteSantander]
```

In [93]:
```
Ponderacion_NorteSantander = NorteSantander["Ponderacion"].sum()
Expansion_NorteSantander = NorteSantander["FactorExpansion"].sum()
NorteSantanderOI = (Ponderacion_NorteSantander/ Expansion_NorteSantander) * 100
NorteSantanderOI
```

Out[93]:
```
70.90460743652932
```

In [94]:
```
NorteSantander_proporcion = (Ponderacion_NorteSantander/ Total_Expansion ) * 100
NorteSantander_proporcion
```

Out[94]:
```
2.1724216490707415
```

In [95]:
```
NorteSantander_Dict1 = {"NorteSantanderOI": 70.90}
NorteSantander_Dict2 = {"NorteSanntander_proporcion": 2.17}
print(NorteSantander_Dict1, NorteSantander_Dict2)
```

```
{'NorteSantanderOI': 70.9} {'NorteSanntander_proporcion': 2.17}
```

In [96]:
```
Putumayo = GEIH_2022["Departamento"] == "Putumayo"
Putumayo = GEIH_2022[Putumayo]
```

In [97]:
```
Ponderacion_Putumayo = Putumayo["Ponderacion"].sum()
Expansion_Putumayo = Putumayo["FactorExpansion"].sum()
PutumayoOI = (Ponderacion_Putumayo/ Expansion_Putumayo) * 100
PutumayoOI
```

Out[97]:
```
44.86040590956706
```

In [98]:
```
Putumayo_proporcion = (Ponderacion_Putumayo/ Total_Expansion ) * 100
Putumayo_proporcion
```

Out[98]:
```
0.030551788668497008
```

In [99]:
```
Putumayo_Dict1 = {"PutumayoOI": 44.86}
Putumayo_Dict2 = {"Putumayo_proporcion": 0.03}
print(Putumayo_Dict1, Putumayo_Dict2)
```

```
{'PutumayoOI': 44.86} {'Putumayo_proporcion': 0.03}
```

In [100…
```
Quindio= GEIH_2022["Departamento"] == "Quindio"
Quindio = GEIH_2022[Quindio]
```

In [101…
```
Ponderacion_Quindio = Quindio["Ponderacion"].sum()
Expansion_Quindio = Quindio["FactorExpansion"].sum()
QuindioOI = (Ponderacion_Quindio/ Expansion_Quindio) * 100
QuindioOI
```

Out[101]:
```
50.14380893227116
```

In [102…
```
Quindio_proporcion = (Ponderacion_Quindio/ Total_Expansion ) * 100
Quindio_proporcion
```

Out[102]:
```
0.5295470704575407
```

In [103…
```
Quindio_Dict1 = {"QuindioOI": 50.14}
Quindio_Dict2 = {"Quindio_proporcion": 0.52}
print(Quindio_Dict1, Quindio_Dict2)
```

```
{'QuindioOI': 50.14} {'Quindio_proporcion': 0.52}
```

In [104…
```
Risaralda = GEIH_2022["Departamento"] == "Risaralda"
Risaralda = GEIH_2022[Risaralda]
```

In [105…
```
Ponderacion_Risaralda = Risaralda["Ponderacion"].sum()
Expansion_Risaralda = Risaralda["FactorExpansion"].sum()
RisaraldaOI = (Ponderacion_Risaralda/ Expansion_Quindio) * 100
RisaraldaOI
```

Out[105]:
```
80.49049770386942
```

In [106…
```
Risaralda_proporcion = (Ponderacion_Risaralda/ Total_Expansion ) * 100
Risaralda_proporcion
```

Out[106]:
```
0.8500253205001775
```

In [107…
```
Risaralda_Dict1 = {"RisaraldaOI": 80.49}
Risaralda_Dict2 = {"Risaralda_proporcion": 0.85}
print(Risaralda_Dict1, Risaralda_Dict2)
```

```
{'RisaraldaOI': 80.49} {'Risaralda_proporcion': 0.85}
```

In [108…
```
SanAndres = GEIH_2022["Departamento"] == "SAI"
SanAndres = GEIH_2022[SanAndres]
```

In [109…
```
Ponderacion_SanAndres = SanAndres["Ponderacion"].sum()
Expansion_SanAndres = SanAndres["FactorExpansion"].sum()
SanAndresOI = (Ponderacion_SanAndres/ Expansion_SanAndres) * 100
SanAndresOI
```

Out[109]:
```
40.99390678052223
```

In [110…
```
SanAndres_proporcion = (Ponderacion_SanAndres/ Total_Expansion ) * 100
SanAndres_proporcion
```

Out[110]:
```
0.039638156251945424
```

In [111…
```
SanAndres_Dict1 = {"SanAndresOI": 40.99}
SanAndres_Dict2 = {"SanAndres_proporcion": 0.03}
print(SanAndres_Dict1, SanAndres_Dict2)
```

```
{'SanAndresOI': 40.99} {'SanAndres_proporcion': 0.03}
```

In [112…
```
Santander = GEIH_2022["Departamento"] == "Santander"
Santander = GEIH_2022[Santander]
```

In [113…
```
Ponderacion_Santander = Santander["Ponderacion"].sum()
Expansion_Santander = Santander["FactorExpansion"].sum()
SantanderOI = (Ponderacion_Santander/ Expansion_Santander) * 100
SantanderOI
```

Out[113]:
```
54.0418020630734
```

In [114…
```
Santander_proporcion = (Ponderacion_Santander/ Total_Expansion ) * 100
Santander_proporcion
```

Out[114]:
```
2.485914792217203
```

In [115…
```
Santander_Dict1 = {"SantanderOI": 54.04}
Santander_Dict2 = {"Santander_proporcion": 2.48}
print(Santander_Dict1, Santander_Dict2)
```

```
{'SantanderOI': 54.04} {'Santander_proporcion': 2.48}
```

In [116…
```
Sucre = GEIH_2022["Departamento"] == "Sucre"
Sucre = GEIH_2022[Sucre]
```

In [117…
```
Ponderacion_Sucre = Sucre["Ponderacion"].sum()
Expansion_Sucre = Sucre["FactorExpansion"].sum()
SucreOI = (Ponderacion_Sucre/ Expansion_Sucre) * 100
SucreOI
```

Out[117]:
```
74.52802288324078
```

In [118…
```
Sucre_proporcion = (Ponderacion_Sucre/ Total_Expansion ) * 100
Sucre_proporcion
```

Out[118]:
```
1.3208177701843011
```

In [119…
```
Sucre_Dict1 = {"SucreOI": 74.52}
Sucre_Dict2 = {"Sucre_proporcion": 1.32}
print(Sucre_Dict1, Sucre_Dict2)
```

```
{'SucreOI': 74.52} {'Sucre_proporcion': 1.32}
```

In [120…
```
Tolima = GEIH_2022["Departamento"] == "Tolima"
Tolima = GEIH_2022[Tolima]
```

In [121…
```
Ponderacion_Tolima = Tolima["Ponderacion"].sum()
Expansion_Tolima= Tolima["FactorExpansion"].sum()
TolimaOI = (Ponderacion_Tolima/ Expansion_Tolima) * 100
TolimaOI
```

Out[121]:
```
60.240769038624606
```

In [122…
```
Tolima_proporcion = (Ponderacion_Tolima/ Total_Expansion ) * 100
Tolima_proporcion
```

Out[122]:
```
1.4359200050681658
```

In [123…
```
Tolima_Dict1 = {"TolimaOI": 60.24}
Tolima_Dict2 = {"Tolima_proporcion": 1.43}
print(Tolima_Dict1, Tolima_Dict2)
```

```
{'TolimaOI': 60.24} {'Tolima_proporcion': 1.43}
```

In [124…
```
Valle = GEIH_2022["Departamento"] == "Valle"
Valle = GEIH_2022[Valle]
```

In [125…
```
Ponderacion_Valle = Valle["Ponderacion"].sum()
Expansion_Valle= Valle["FactorExpansion"].sum()
ValleOI = (Ponderacion_Valle/ Expansion_Valle) * 100
ValleOI
```

Out[125]:
```
48.90236195779351
```

In [126…
```
Valle_proporcion = (Ponderacion_Valle/ Total_Expansion ) * 100
Valle_proporcion
```

Out[126]:
```
4.461335908725409
```

In [127…
```
Valle_Dict1 = {"ValleOI": 48.90}
Valle_Dict2 = {"Valle_proporcion": 4.46}
print(Valle_Dict1, Valle_Dict2)
```

```
{'ValleOI': 48.9} {'Valle_proporcion': 4.46}
```

In [128…
```
Vaupes = GEIH_2022["Departamento"] == "Vaupes"
Vaupes = GEIH_2022[Vaupes]
```

In [129…
```
Ponderacion_Vaupes = Vaupes["Ponderacion"].sum()
Expansion_Vaupes= Vaupes["FactorExpansion"].sum()
VaupesOI = (Ponderacion_Vaupes/ Expansion_Vaupes) * 100
VaupesOI
```

Out[129]:
```
37.06786867701845
```

In [130…
```
Vaupes_proporcion = (Ponderacion_Vaupes/ Total_Expansion ) * 100
Vaupes_proporcion
```

Out[130]:
```
0.007501439105261037
```

In [131…
```
Vaupes_Dict1 = {"VaupesOI": 37.06}
Vaupes_Dict2 = {"Vaupes_proporcion": 0.007}
print(Vaupes_Dict1, Vaupes_Dict2)
```

```
{'VaupesOI': 37.06} {'Vaupes_proporcion': 0.007}
```

In [132…
```
Vichada = GEIH_2022["Departamento"] == "Vichada"
Vichada = GEIH_2022[Vichada]
```

In [133…
```
Ponderacion_Vichada = Vichada["Ponderacion"].sum()
Expansion_Vichada= Vichada["FactorExpansion"].sum()
VichadaOI = (Ponderacion_Vichada/ Expansion_Vichada) * 100
VichadaOI
```

Out[133]:
```
53.92285329402272
```

In [134…
```
Vichada_proporcion = (Ponderacion_Vichada/ Total_Expansion ) * 100
Vichada_proporcion
```

Out[134]:
```
0.01718799498399812
```

In [135…
```
Vichada_Dict1 = {"VichadaOI": 53.92}
Vichada_Dict2 = {"Vichada_proporcion": 0.01}
print(Vichada_Dict1, Vichada_Dict2)
```

```
{'VichadaOI': 53.92} {'Vichada_proporcion': 0.01}
```

In [136…
```
Mujer = GEIH_2022["Sexo"] == "Mujer"
Mujer = GEIH_2022[Mujer]
```

In [137…
```
Ponderacion_Mujer = Mujer["Ponderacion"].sum()
Expansion_Mujer= Mujer["FactorExpansion"].sum()
MujerOI = (Ponderacion_Mujer/ Expansion_Mujer) * 100
MujerOI
```

Out[137]:
```
48.98320558150915
```

In [138…
```
Mujer_proporcion = (Ponderacion_Mujer/ Total_Expansion ) * 100
Mujer_proporcion
```

Out[138]:
```
19.992332191872684
```

In [139…
```
Mujer_Dict1 = {"MujerOI": 48.98}
Mujer_Dict2 = {"Mujer_proporcion": 19.99}
print(Mujer_Dict1, Mujer_Dict2)
```

```
{'MujerOI': 48.98} {'Mujer_proporcion': 19.99}
```

In [140…
```
Hombre= GEIH_2022["Sexo"] == "Hombre"
Hombre = GEIH_2022[Hombre]
```

In [141…
```
Ponderacion_Hombre = Hombre["Ponderacion"].sum()
Expansion_Hombre= Hombre["FactorExpansion"].sum()
HombreOI = (Ponderacion_Hombre/ Expansion_Hombre) * 100
HombreOI
```

Out[141]:
```
56.5757723766198
```

In [142…
```
Hombre_proporcion = (Ponderacion_Hombre/ Total_Expansion ) * 100
Hombre_proporcion
```

Out[142]:
```
33.48455933865288
```

In [143…
```
Hombre_Dict1 = {"HombreOI": 56.57}
Hombre_Dict2 = {"Hombre_proporcion": 33.4}
print(Hombre_Dict1, Hombre_Dict2)
```

```
{'HombreOI': 56.57} {'Hombre_proporcion': 33.4}
```

In [144…
```
Ninguno = GEIH_2022["Etnia"] == "Ninguno"
Ninguno = GEIH_2022[Ninguno]
```

In [145…
```
Ponderacion_Ninguno = Ninguno["Ponderacion"].sum()
Ninguno_proporcion = (Ponderacion_Ninguno/ Total_Expansion ) * 100
Ninguno_proporcion
```

Out[145]:
```
46.35430557659932
```

In [146…
```
Expansion_Ninguno= Ninguno["FactorExpansion"].sum()
NingunoOI = (Ponderacion_Ninguno/ Expansion_Ninguno) * 100
NingunoOI
```

Out[146]:
```
51.95624255732781
```

In [147…
```
Ninguno_Dict = {"Ninguno_proporcion": 46.35}
Ninguno_Dict2 = {"NingunoOI": 51.95}
print(Ninguno_Dict,Ninguno_Dict2)
```

```
{'Ninguno_proporcion': 46.35} {'NingunoOI': 51.95}
```

In [148…
```
Indigena = GEIH_2022["Etnia"] == "Indígena"
Indigena = GEIH_2022[Indigena]
```

In [149…
```
Ponderacion_Indigena = Indigena["Ponderacion"].sum()
Indigena_proporcion = (Ponderacion_Indigena/ Total_Expansion ) * 100
Indigena_proporcion
```

Out[149]:
```
2.7589362535725663
```

In [150…
```
Expansion_Indigena= Indigena["FactorExpansion"].sum()
IndigenaOI = (Ponderacion_Indigena/ Expansion_Indigena) * 100
IndigenaOI
```

Out[150]:
```
73.87891580185237
```

In [151…
```
Indigena_Dict = {"Indigena_proporcion": 2.75}
Indigena_Dict2 = {"IndigenaOI": 73.87}
print(Indigena_Dict,Indigena_Dict2)
```

```
{'Indigena_proporcion': 2.75} {'IndigenaOI': 73.87}
```

In [152…
```
Afros_Mulatos = GEIH_2022["Etnia"] == "Negro_mulato_afrod_afroc"
Afros_Mulatos = GEIH_2022[Afros_Mulatos]
```

In [153…
```
Ponderacion_Afros_Mulatos = Afros_Mulatos["Ponderacion"].sum()
Afros_Mulatos_proporcion = (Ponderacion_Afros_Mulatos / Total_Expansion ) * 100
Afros_Mulatos_proporcion
```

Out[153]:
```
4.3178153437973785
```

In [154…
```
Expansion_Afros_Mulatos= Afros_Mulatos["FactorExpansion"].sum()
Afros_MulatosOI = (Ponderacion_Afros_Mulatos/ Expansion_Afros_Mulatos) * 100
Afros_MulatosOI
```

Out[154]:
```
61.973973184321395
```

In [155…
```
Afros_Mulatos_Dict = {"Afros_Mulatos_proporcion": 4.31}
Afros_Mulatos_Dict2 = {"Afros_MulatosOI": 61.97}
print(Afros_Mulatos_Dict,Afros_Mulatos_Dict2)
```

```
{'Afros_Mulatos_proporcion': 4.31} {'Afros_MulatosOI': 61.97}
```

In [156…
```
Gitano =  GEIH_2022["Etnia"] == "Gitano"
Gitano = GEIH_2022[Gitano]
```

In [157…
```
Ponderacion_Gitano = Gitano["Ponderacion"].sum()
Gitano_proporcion = (Ponderacion_Gitano/ Total_Expansion ) * 100
Gitano_proporcion
```

Out[157]:
```
0.007186057774803632
```

In [158…
```
Gitano_Dict = {"Gitano_proporcion": 0.007}
Gitano_Dict
```

Out[158]:
```
{'Gitano_proporcion': 0.007}
```

In [159…
```
Palenquero =  GEIH_2022["Etnia"] == "Palenquero"
Palenquero = GEIH_2022[Palenquero]
```

In [160…
```
Ponderacion_Palenquero = Palenquero["Ponderacion"].sum()
Palenquero_proporcion = (Ponderacion_Palenquero/ Total_Expansion ) * 100
Palenquero_proporcion
```

Out[160]:
```
0.01946446505261181
```

In [161…
```
Palenquero_Dict = {"Palenquero_proporcion": 0.01}
Palenquero_Dict
```

Out[161]:
```
{'Palenquero_proporcion': 0.01}
```

In [162…
```
Raizal_SAI=  GEIH_2022["Etnia"] == "Raizal_SAI"
Raizal_SAI = GEIH_2022[Raizal_SAI]
```

In [163…
```
Ponderacion_Raizal_SAI = Raizal_SAI["Ponderacion"].sum()
Raizal_SAI_proporcion = (Ponderacion_Raizal_SAI/ Total_Expansion ) * 100
Raizal_SAI_proporcion
```

Out[163]:
```
0.01918383372887771
```

In [164…
```
Raizal_SAI_Dict = {"Raizal_SAI_proporcion": 0.01}
Raizal_SAI_Dict
```

Out[164]:
```
{'Raizal_SAI_proporcion': 0.01}
```

In [165…
```
print([Antioquia_Dict1, Amazonas_Dict1, Arauca_Dict1, Atlantico_Dict1, Bogota_Dict1, B
                                        Caldas_Dict1, Caqueta_Dict1, Casanare_Dict1,
                                        Choco_Dict1, Cordoba_Dict1, Guainia_Dict1, G
                                        Meta_Dict1, Narino_Dict1, NorteSantander_Dic
                                        Santander_Dict1, Sucre_Dict1, Tolima_Dict1,
```

```
[{'AntioquiaOI': 46.386}, {'AmazonasOI': 56.53}, {'AraucaOI': 55.43}, {'AtlanticoOI':
56.29}, {'BogotaOI': 34.04}, {'BolivarOI': 66.46}, {'BoyacaOI': 61.63}, {'CaldasOI':
45.8}, {'CaquetaOI': 67.76}, {'CasanareOI': 54.79}, {'CaucaOI': 68.53}, {'CesarOI': 6
3.64}, {'CundinamarcaOI': 45.4}, {'ChocoOI': 67.9}, {'CordobaOI': 68.9}, {'GuainiaO
I': 67.31}, {'GuaviareOI': 57.96}, {'HuilaOI': 68.72}, {'GuajiraOI': 76.74}, {'Magdal
enaOI': 70.89}, {'MetaOI': 55.63}, {'NarinoOI': 79.11}, {'NorteSantanderOI': 70.9},
{'PutumayoOI': 44.86}, {'QuindioOI': 50.14}, {'RisaraldaOI': 80.49}, {'SanAndresOI':
40.99}, {'SantanderOI': 54.04}, {'SucreOI': 74.52}, {'TolimaOI': 60.24}, {'ValleOI':
48.9}, {'VaupesOI': 37.06}, {'VichadaOI': 53.92}]
```

In [166…

```
print([Antioquia_Dict2, Amazonas_Dict2, Arauca_Dict2, Atlantico_Dict2, Bogota_Dict2, B
                                        Caldas_Dict2, Caqueta_Dict2, Casanare_Dict2,
                                        Choco_Dict2, Cordoba_Dict2, Guainia_Dict2, G
                                        Meta_Dict2, Narino_Dict2, NorteSantander_Dic
                                        Santander_Dict2, Sucre_Dict2, Tolima_Dict2,
```

```
[{'Antioquia_proporcion': 6.49}, {'Amazonas_proporcion': 0.035}, {'Arauca_proporcio
n': 0.07}, {'Atlantico_proporcion': 3.16}, {'Bogota_proporcion': 5.9}, {'Bolivar_prop
orcion': 2.8}, {'Boyaca_proporcion': 1.53}, {'Caldas_proporcion': 0.9}, {'Caqueta_pro
porcion': 0.49}, {'Casanare_proporcion': 0.19}, {'Cauca_proporcion': 2.0}, {'Cesar_pr
oporcion': 1.39}, {'Cundinamarca_proporcion': 3.08}, {'Choco_proporcion': 0.45}, {'Co
rdoba_proporcion': 2.38}, {'Guainia_proporcion': 0.02}, {'Guaviare_proporcion': 0.0
5}, {'Huila_proporcion': 1.4}, {'Guajira_proporcion': 1.3}, {'Magdalena_proporcion':
1.9}, {'Meta_proporcion': 1.22}, {'Narino_proporcion': 3.22}, {'NorteSanntander_propo
rcion': 2.17}, {'Putumayo_proporcion': 0.03}, {'Quindio_proporcion': 0.52}, {'Risaral
da_proporcion': 0.85}, {'SanAndres_proporcion': 0.03}, {'Santander_proporcion': 2.4
8}, {'Sucre_proporcion': 1.32}, {'Tolima_proporcion': 1.43}, {'Valle_proporcion': 4.4
6}, {'Vaupes_proporcion': 0.007}, {'Vichada_proporcion': 0.01}]
```

In [167…

```
Dptos_informalidad = {'Antioquia': 46.386, 'Amazonas': 56.53, 'Arauca': 55.43, 'Atlant
```

In [168…

```
Informalidad_Departamentos = pd.DataFrame(list(Dptos_informalidad.items()), columns=['
Informalidad_Departamentos
```

Out[168]:

	Departamento	%informalidad
0	Antioquia	46.386
1	Amazonas	56.530
2	Arauca	55.430
3	Atlantico	56.290
4	Bogota	34.040
5	Bolivar	66.460
6	Boyaca	61.630
7	Caldas	45.800
8	Caqueta	67.760
9	Casanare	54.790
10	Cauca	68.530
11	Cesar	63.640
12	Cundinamarca	45.400
13	Choco	67.900
14	Cordoba	68.900
15	Guainia	67.310
16	Guaviare	57.960
17	Huila	68.720
18	Guajira	76.740
19	Magdalena	70.890
20	Meta	55.630
21	Narino	79.110
22	NorteSantander	70.900
23	Putumayo	44.860
24	Quindio	50.140
25	Risaralda	80.490
26	SanAndres	40.990
27	Santander	54.040
28	Sucre	74.520
29	Tolima	60.240
30	Valle	48.900
31	Vaupes	37.060
32	Vichada	53.920

In [169…
```
Dptos_proporciones = {'Antioquia': 6.49, 'Amazonas': 0.035, 'Arauca': 0.07, 'Atlantic
```

In [170…
```
Proporciones_Departamentos = pd.DataFrame(list(Dptos_proporciones.items()), columns=['
Proporciones_Departamentos
```

Out[170]:

	Departamento	proporcion_informalidad_nacional
0	Antioquia	6.490
1	Amazonas	0.035
2	Arauca	0.070
3	Atlantico	3.160
4	Bogota	5.900
5	Bolivar	2.800
6	Boyaca	1.530
7	Caldas	0.900
8	Caqueta	0.490
9	Casanare	0.190
10	Cauca	2.000
11	Cesar	1.390
12	Cundinamarca	3.080
13	Choco	0.450
14	Cordoba	2.380
15	Guainia	0.020
16	Guaviare	0.050
17	Huila	1.400
18	Guajira	1.300
19	Magdalena	1.900
20	Meta	1.220
21	Narino	3.220
22	NorteSantander	2.170
23	Putumayo	0.030
24	Quindio	0.520
25	Risaralda	0.850
26	SanAndres	0.030
27	Santander	2.480
28	Sucre	1.320
29	Tolima	1.430
30	Valle	4.460
31	Vaupes	0.007
32	Vichada	0.010

In [171…

```
Departamentos = pd.merge(Informalidad_Departamentos,Proporciones_Departamentos, on=["D
Departamentos.to_csv("Departamentos.csv")
Departamentos
```

Out[171]:

	Departamento	%informalidad	proporcion_informalidad_nacional
0	Antioquia	46.386	6.490
1	Amazonas	56.530	0.035
2	Arauca	55.430	0.070
3	Atlantico	56.290	3.160
4	Bogota	34.040	5.900
5	Bolivar	66.460	2.800
6	Boyaca	61.630	1.530
7	Caldas	45.800	0.900
8	Caqueta	67.760	0.490
9	Casanare	54.790	0.190
10	Cauca	68.530	2.000
11	Cesar	63.640	1.390
12	Cundinamarca	45.400	3.080
13	Choco	67.900	0.450
14	Cordoba	68.900	2.380
15	Guainia	67.310	0.020
16	Guaviare	57.960	0.050
17	Huila	68.720	1.400
18	Guajira	76.740	1.300
19	Magdalena	70.890	1.900
20	Meta	55.630	1.220
21	Narino	79.110	3.220
22	NorteSantander	70.900	2.170
23	Putumayo	44.860	0.030
24	Quindio	50.140	0.520
25	Risaralda	80.490	0.850
26	SanAndres	40.990	0.030
27	Santander	54.040	2.480
28	Sucre	74.520	1.320
29	Tolima	60.240	1.430
30	Valle	48.900	4.460
31	Vaupes	37.060	0.007
32	Vichada	53.920	0.010

In [172…
```
Informalidad_Sexo = {'Mujer': 48.98, 'Hombre': 56.57}
Informalidad_proporcion = {'Mujer': 19.99, "Hombre": 33.4}
```

In [173…
```
Informalidad_Sexo = pd.DataFrame(list(Informalidad_Sexo.items()), columns=["Sexo","%Ir
Informalidad_Sexo
```

Out[173]:

	Sexo	%Informalidad
0	Mujer	48.98
1	Hombre	56.57

In [174…
```
Informalidad_proporcion = pd.DataFrame(list(Informalidad_proporcion.items()), columns=
Informalidad_proporcion
```

Out[174]:

	Sexo	Proporcion
0	Mujer	19.99
1	Hombre	33.40

In [175…
```
Sexo = pd.merge(Informalidad_Sexo,Informalidad_proporcion, on=["Sexo"])
Sexo.to_csv("Sexo.csv")
Sexo
```

Out[175]:

	Sexo	%Informalidad	Proporcion
0	Mujer	48.98	19.99
1	Hombre	56.57	33.40

In [176…
```
Informalidad_Etnias = {"Afros_Mulatos": 61.97, "Indigena": 73.87, 'Ninguno': 51.95}
Informalidad_Etnias = pd.DataFrame(list(Informalidad_Etnias.items()), columns=["Etnia'
Informalidad_Etnias.to_csv("Etnias.csv")
Informalidad_Etnias
```

Out[176]:

	Etnia	%informalidad
0	Afros_Mulatos	61.97
1	Indigena	73.87
2	Ninguno	51.95

In [177…
```
GEIH_2022["NivelEducativo"]=GEIH_2022["NivelEducativo"].replace(["Ninguno", "Preescola
                                                                ["Ninguno", "Secunda
```

In [178…
```
Ninguno = GEIH_2022["NivelEducativo"] == "Ninguno"
Ninguno = GEIH_2022[Ninguno]
```

In [179…
```
Ponderacion_Ninguno = Ninguno["Ponderacion"].sum()
Expansion_Ninguno = Ninguno["FactorExpansion"].sum()
NingunoOI = (Ponderacion_Ninguno/ Expansion_Ninguno) * 100
NingunoOI
```

Out[179]:
```
79.85645168395871
```

In [180…
```
Secundaria_menos= GEIH_2022["NivelEducativo"] == "Secundaria≤"
Secundaria_menos = GEIH_2022[Secundaria_menos]
```

In [181…
```
Ponderacion_Secundaria_menos = Secundaria_menos["Ponderacion"].sum()
Expansion_Secundaria_menos = Secundaria_menos["FactorExpansion"].sum()
Secundaria_menosOI = (Ponderacion_Secundaria_menos/ Expansion_Secundaria_menos) * 100
Secundaria_menosOI
```

Out[181]:
```
73.3179252854153
```

In [182…
```
Bachillerato= GEIH_2022["NivelEducativo"] == "Bachillerato"
Bachillerato = GEIH_2022[Bachillerato]
```

In [183…
```
Ponderacion_Bachillerato = Bachillerato["Ponderacion"].sum()
Expansion_Bachillerato = Bachillerato["FactorExpansion"].sum()
BachilleratoOI = (Ponderacion_Bachillerato/ Expansion_Bachillerato) * 100
BachilleratoOI
```

Out[183]:
```
55.244577292549565
```

In [184…
```
Superior = GEIH_2022["NivelEducativo"] == "Superior"
Superior = GEIH_2022[Superior]
```

In [185…
```
Ponderacion_Superior= Superior["Ponderacion"].sum()
Expansion_Superior = Superior["FactorExpansion"].sum()
SuperiorOI = (Ponderacion_Superior/ Expansion_Superior) * 100
SuperiorOI
```

Out[185]:
```
30.71889909869147
```

In [186…
```
SuperiorT = GEIH_2022["NivelEducativo"] == "SuperiorT"
SuperiorT = GEIH_2022[SuperiorT]
```

In [187…
```
Ponderacion_SuperiorT= SuperiorT["Ponderacion"].sum()
Expansion_SuperiorT = SuperiorT["FactorExpansion"].sum()
SuperiorTOI = (Ponderacion_SuperiorT/ Expansion_SuperiorT) * 100
SuperiorTOI
```

Out[187]:
```
37.08281166136646
```

In [188…
```
Superior2 = GEIH_2022["NivelEducativo"] == "Superior+"
Superior2 = GEIH_2022[Superior2]
```

In [189…
```
Ponderacion_Superior2= Superior2["Ponderacion"].sum()
Expansion_Superior2 = Superior2["FactorExpansion"].sum()
Superior2OI = (Ponderacion_Superior2/ Expansion_Superior2) * 100
Superior2OI
```

Out[189]:
```
13.732699514916652
```

In [190…
```
NivelesEducativos = {"Ninguno":79.85, "Secundaria≤": 73.31, "Bachillerato":55.24, "Sup
```

In [191…
```
Informalidad_NivelEducacion = pd.DataFrame(list(NivelesEducativos.items()), columns=['
Informalidad_NivelEducacion.to_csv("NivelesEducativos.csv")
Informalidad_NivelEducacion
```

Out[191]:

	Nivel	%Informalidad
0	Ninguno	79.85
1	Secundaria≤	73.31
2	Bachillerato	55.24
3	SuperiorT	37.08
4	Superior	30.70
5	Superior+	13.73

In [192…
```
enero_personas = pd.read_csv("Enero/Características generales, seguridad social en sal
enero_ocupados = pd.read_csv("Enero/Ocupados.csv", sep=",", encoding = "latin-1", usec
enero = pd.merge(enero_personas,enero_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P
enero
```

Out[192]:

	DIRECTORIO	SECUENCIA_P	ORDEN	DPTO	P6500	P3051
0	5000000	1	1	8	768000.0	NaN
1	5000001	1	1	8	NaN	NaN
2	5000001	1	2	8	NaN	NaN
3	5000001	1	3	8	NaN	NaN
4	5000002	1	1	8	360000.0	NaN
…	…	…	…	…	…	…
31814	5032617	1	1	91	200000.0	NaN
31815	5032618	1	1	91	600000.0	NaN
31816	5032619	1	2	91	1000000.0	NaN
31817	5032620	1	1	91	2300000.0	NaN
31818	5032620	1	2	91	2200000.0	NaN

31819 rows × 6 columns

In [193…
```
febrero_personas = pd.read_csv("Febrero/Características generales, seguridad social er
febrero_ocupados = pd.read_csv("Febrero/Ocupados.csv", sep=";", encoding = "latin-1",
febrero = pd.merge(febrero_personas,febrero_ocupados, on=["DIRECTORIO", "ORDEN","SECUE
febrero
```

Out[193]:

	DIRECTORIO	SECUENCIA_P	ORDEN	DPTO	P6500	P3051
0	5032621	1	1	19	NaN	NaN
1	5032621	1	2	19	1200000.0	NaN
2	5032622	1	1	19	NaN	3600000.0
3	5032627	1	1	19	NaN	2400000.0
4	5032628	1	2	19	2180000.0	NaN
...	...	...	...	...	...	...
32974	5063808	1	1	76	NaN	NaN
32975	5063809	1	1	66	NaN	NaN
32976	5063810	1	2	63	1000000.0	NaN
32977	5063810	1	3	63	800000.0	NaN
32978	5063810	1	4	63	2000000.0	NaN

32979 rows × 6 columns

In [194…
```
marzo_personas = pd.read_csv("Marzo/Características generales, seguridad social en sal
marzo_ocupados = pd.read_csv("Marzo/Ocupados.csv", sep=";", encoding = "latin-1", usec
marzo = pd.merge(marzo_personas,marzo_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P
```

In [195…
```
abril_personas = pd.read_csv("Abril/Características generales, seguridad social en sal
abril_ocupados = pd.read_csv("Abril/Ocupados.csv", sep=";", encoding = "latin-1", usec
abril = pd.merge(abril_personas,abril_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P
```

In [196…
```
mayo_personas = pd.read_csv("Mayo/Características generales, seguridad social en saluc
mayo_ocupados = pd.read_csv("Mayo/Ocupados.csv", sep=";", encoding = "latin-1", usecol
mayo = pd.merge(mayo_personas,mayo_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P"])
```

In [197…
```
junio_personas = pd.read_csv("Junio/Características generales, seguridad social en sal
junio_ocupados = pd.read_csv("Junio/Ocupados.csv", sep=";", encoding = "latin-1", usec
junio = pd.merge(junio_personas,junio_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P
```

In [198…
```
julio_personas = pd.read_csv("Julio/Características generales, seguridad social en sal
julio_ocupados = pd.read_csv("Julio/Ocupados.csv", sep=";", encoding = "latin-1", usec
julio = pd.merge(julio_personas,julio_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCIA_P
```

In [199…
```
agosto_personas = pd.read_csv("Agosto/Características generales, seguridad social en s
agosto_ocupados = pd.read_csv("Agosto/Ocupados.csv", sep=";", encoding = "latin-1", us
agosto = pd.merge(agosto_personas,agosto_ocupados, on=["DIRECTORIO", "ORDEN","SECUENCI
```

In [200…
```
septiembre_personas = pd.read_csv("Septiembre/Características generales, seguridad soc
septiembre_ocupados = pd.read_csv("Septiembre/Ocupados.csv", sep=";", encoding = "lati
septiembre = pd.merge(septiembre_personas,septiembre_ocupados, on=["DIRECTORIO", "ORDE
```

In [201…
```
octubre_personas = pd.read_csv("Octubre/Características generales, seguridad social er
octubre_ocupados = pd.read_csv("Octubre/Ocupados.csv", sep=";", encoding = "latin-1",
octubre = pd.merge(octubre_personas,octubre_ocupados, on=["DIRECTORIO", "ORDEN","SECUE
```

In [202…
```
noviembre_personas = pd.read_csv("Noviembre/Características generales, seguridad socia
noviembre_ocupados = pd.read_csv("Noviembre/Ocupados.csv", sep=";", encoding = "latin-
noviembre = pd.merge(noviembre_personas,noviembre_ocupados, on=["DIRECTORIO", "ORDEN",
```

In [203…
```
diciembre_personas = pd.read_csv("Diciembre/Características generales, seguridad socia
diciembre_ocupados = pd.read_csv("Diciembre/Ocupados.csv", sep=";", encoding = "latin-
diciembre = pd.merge(diciembre_personas,diciembre_ocupados, on=["DIRECTORIO", "ORDEN",
```

In [204…
```
GEIH2022_ingreso = pd.concat([enero,febrero,marzo,abril, mayo, junio, julio, agosto, s
GEIH2022_ingreso= GEIH2022_ingreso.fillna(0)
```

In [205…
```
ceros = GEIH_2022[GEIH_2022["Salario"] == 0.0].index
GEIH_2022.drop(ceros, inplace=True)
GEIH_2022.to_csv("SalarioMedio.csv")
```

In [206…
```
AraucaSalario = Arauca["Salario"].mean()
AraucaSalario
```

Out[206]:
```
633877.0978441128
```

In [207…
```
AmazonasSalario = Amazonas["Salario"].mean()
AmazonasSalario
```

Out[207]:
```
617707.3541666666
```

In [208…
```
AntioquiaSalario = Antioquia["Salario"].mean()
AntioquiaSalario
```

Out[208]:
```
1203229.33237544
```

In [209…
```
AtlanticoSalario = Atlantico["Salario"].mean()
AtlanticoSalario
```

Out[209]:
```
732442.4489531743
```

In [210…
```
BogotaSalario = Bogota["Salario"].mean()
BogotaSalario
```

Out[210]:
```
1720549.6882770192
```

In [211…
```
BolivarSalario = Bolivar["Salario"].mean()
BolivarSalario
```

Out[211]:
```
554065.0003359838
```

In [212…
```
BoyacaSalario = Boyaca["Salario"].mean()
BoyacaSalario
```

Out[212]:
```
1133030.9251118926
```

In [213…
```
CaldasSalario = Caldas["Salario"].mean()
CaldasSalario
```

Out[213]:
```
1104061.364039409
```

In [214…
```
CaquetaSalario = Caqueta["Salario"].mean()
CaquetaSalario
```

Out[214]:
```
795871.9770142754
```

In [215…
```
CasanareSalario = Casanare["Salario"].mean()
CasanareSalario
```

Out[215]:
```
1221574.5320537428
```

In [216…
```
CaucaSalario = Cauca["Salario"].mean()
CaucaSalario
```

Out[216]:
```
994267.742861316
```

In [217…
```
CesarSalario = Cesar["Salario"].mean()
CesarSalario
```

Out[217]:
```
640610.1699346405
```

In [218…
```
ChocoSalario = Choco["Salario"].mean()
ChocoSalario
```

Out[218]:
```
851723.2008396798
```

In [219…
```
CordobaSalario = Cordoba["Salario"].mean()
CordobaSalario
```

Out[219]:
```
683984.5622045915
```

In [220…
```
CundinamarcaSalario = Cundinamarca["Salario"].mean()
CundinamarcaSalario
```

Out[220]:
```
866788.837367419
```

In [221…
```
GuainiaSalario = Guainia["Salario"].mean()
GuainiaSalario
```

Out[221]:
```
708864.5466188524
```

In [222…
```
GuaviareSalario = Guaviare["Salario"].mean()
GuaviareSalario
```

Out[222]:
```
928895.6039370078
```

In [223…
```
HuilaSalario = Huila["Salario"].mean()
HuilaSalario
```

Out[223]:
```
809459.4146000155
```

In [224…
```
GuajiraSalario = Guajira["Salario"].mean()
GuajiraSalario
```

Out[224]:
```
589259.5788213486
```

In [225…
```
MagdalenaSalario = Magdalena["Salario"].mean()
MagdalenaSalario
```
Out[225]: 566829.9244483784

In [226…
```
MetaSalario = Meta["Salario"].mean()
MetaSalario
```
Out[226]: 977752.4235704415

In [227…
```
NarinoSalario = Narino["Salario"].mean()
NarinoSalario
```
Out[227]: 767985.1958308388

In [228…
```
NorteSSalario = NorteSantander["Salario"].mean()
NorteSSalario
```
Out[228]: 561033.3804897525

In [229…
```
PutumayoSalario = Putumayo["Salario"].mean()
PutumayoSalario
```
Out[229]: 467234.4602425876

In [230…
```
GuaviareSalario = Guaviare["Salario"].mean()
GuaviareSalario
```
Out[230]: 928895.6039370078

In [231…
```
HuilaSalario = Huila["Salario"].mean()
HuilaSalario
```
Out[231]: 809459.4146000155

In [232…
```
GuajiraSalario = Guajira["Salario"].mean()
GuajiraSalario
```
Out[232]: 589259.5788213486

In [233…
```
MagdalenaSalario = Magdalena["Salario"].mean()
MagdalenaSalario
```
Out[233]: 566829.9244483784

In [234…
```
MetaSalario = Meta["Salario"].mean()
MetaSalario
```
Out[234]: 977752.4235704415

In [235…
```
NarinoSalario = Narino["Salario"].mean()
NarinoSalario
```
Out[235]: 767985.1958308388

In [236…
```
NorteSSalario = NorteSantander["Salario"].mean()
NorteSSalario
```
Out[236]: 561033.3804897525

In [237…
```
PutumayoSalario = Putumayo["Salario"].mean()
PutumayoSalario
```
Out[237]: 467234.4602425876

In [238…
```
GuaviareSalario = Guaviare["Salario"].mean()
GuaviareSalario
```
Out[238]: 928895.6039370078

In [239…
```
HuilaSalario = Huila["Salario"].mean()
HuilaSalario
```
Out[239]: 809459.4146000155

In [240…
```
GuajiraSalario = Guajira["Salario"].mean()
GuajiraSalario
```
Out[240]: 589259.5788213486

In [241…
```
MagdalenaSalario = Magdalena["Salario"].mean()
MagdalenaSalario
```
Out[241]: 566829.9244483784

In [242…
```
MetaSalario = Meta["Salario"].mean()
MetaSalario
```
Out[242]: 977752.4235704415

In [243…
```
NarinoSalario = Narino["Salario"].mean()
NarinoSalario
```
Out[243]: 767985.1958308388

In [244…
```
NorteSSalario = NorteSantander["Salario"].mean()
NorteSSalario
```
Out[244]: 561033.3804897525

In [245…
```
PutumayoSalario = Putumayo["Salario"].mean()
PutumayoSalario
```
Out[245]: 467234.4602425876

In [246…
```
NorteSSalario = NorteSantander["Salario"].mean()
NorteSSalario
```
Out[246]: 561033.3804897525

In [247…
```
PutumayoSalario = Putumayo["Salario"].mean()
PutumayoSalario
```

Out[247]:
```
467234.4602425876
```

In [248…
```
RisaraldaSalario = Risaralda["Salario"].mean()
RisaraldaSalario
```

Out[248]:
```
1093172.5531060959
```

In [249…
```
SAISalario = SanAndres["Salario"].mean()
SAISalario
```

Out[249]:
```
1096731.8397580192
```

In [250…
```
SantanderSalario = Santander["Salario"].mean()
SantanderSalario
```

Out[250]:
```
964342.9661658199
```

In [251…
```
SucreSalario = Sucre["Salario"].mean()
SucreSalario
```

Out[251]:
```
546925.7494062564
```

In [252…
```
TolimaSalario = Tolima["Salario"].mean()
TolimaSalario
```

Out[252]:
```
853075.6514701915
```

In [253…
```
ValleSalario = Valle["Salario"].mean()
ValleSalario
```

Out[253]:
```
1001836.6507301646
```

In [254…
```
Vaupesalario = Vaupes["Salario"].mean()
Vaupesalario
```

Out[254]:
```
1534246.379705401
```

In [255…
```
VichadaSalario = Vichada["Salario"].mean()
VichadaSalario
```

Out[255]:
```
1048821.2568690097
```

In [256…
```
SM_Dptos= {"Arauca":633877,
"Amazonas" : 617707,
"Antioquia": 1203229,
"Atlantico": 732442,
"Bogota": 1720549,
"Bolivar":554065,
    "Boyaca":1133030,
"Caldas":1104061,
    "Caqueta":795871,
        "Casanare":1221574,
```

```
        "Cauca":994267,
        "Cesar": 640610,
        "Choco":851723,
        "Cordoba": 683984,
        "Cundinamarca":866788,
        "Guainia":708864,
        "Guaviare":928895,
        "Huila":809459,
        "Guajira":589259,
        "Magdalena": 566829,
        "Meta": 977752,
        "Narino": 767985,
        "NorteSantander": 561033,
        "Putumayo":467234,
        "Quindio":964620,
        "Risaralda":1093172,
        "SanAndres":1096731,
        "Santander": 964342,
        "Tolima":853075,
        "Valle":1001836,
        "Vaupes":1534246,
        "Vichada":1048821,}
```

In [257…

```
SM_dptos = pd.DataFrame(list(SM_Dptos.items()), columns=["Departamento","SalarioMedio"
SM_dptos.to_csv("SMdptos.csv")
SM_dptos.sort_values(by=["SalarioMedio"])
```

Out[257]:

	Departamento	SalarioMedio
23	Putumayo	467234
5	Bolivar	554065
22	NorteSantander	561033
19	Magdalena	566829
18	Guajira	589259
1	Amazonas	617707
0	Arauca	633877
11	Cesar	640610
13	Cordoba	683984
15	Guainia	708864
3	Atlantico	732442
21	Narino	767985
8	Caqueta	795871
17	Huila	809459
12	Choco	851723
28	Tolima	853075
14	Cundinamarca	866788
16	Guaviare	928895
27	Santander	964342
24	Quindio	964620
20	Meta	977752
10	Cauca	994267
29	Valle	1001836
31	Vichada	1048821
25	Risaralda	1093172
26	SanAndres	1096731
7	Caldas	1104061
6	Boyaca	1133030
2	Antioquia	1203229
9	Casanare	1221574
30	Vaupes	1534246
4	Bogota	1720549

In [258…
```
Informalidad_Departamentos.sort_values(by=["%informalidad"], ascending=False, inplace=
Informalidad_Departamentos.to_csv("inforsort.csv")
```

PROCESAMIENTO BASES OCDE

In [259…
```
Informalidad_OCDE = pd.read_csv("Informalidad_2021.csv")
Informalidad_OCDE.columns
```

Out[259]:
```
Index(['LOCATION', 'Country', 'SUBJECT', 'Subject', 'FREQUENCY', 'Frequency',
       'TIME', 'Time', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Value', 'Flag Codes',
       'Flags'],
      dtype='object')
```

In [260…
```
Informalidad_OCDE= Informalidad_OCDE.drop(columns=['LOCATION', 'SUBJECT', 'Subject', '
       'TIME', 'Time', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Flag Codes',
       'Flags'])
```

In [261…
```
Informalidad_OCDE= Informalidad_OCDE.rename(columns={"Value":"%informalidad"})
Informalidad_OCDE
```

Out[261]:

	Country	%informalidad
0	Australia	9.562654
1	Austria	11.910980
2	Belgium	14.140240
3	Canada	7.633201
4	Czechia	15.936410
5	Denmark	8.839274
6	Finland	14.572440
7	France	12.606570
8	Germany	8.931904
9	Greece	31.824720
10	Hungary	12.507940
11	Ireland	14.063140
12	Italy	21.831650
13	Japan	9.832524
14	Korea	23.906210
15	Luxembourg	10.225700
16	Mexico	31.826360
17	Netherlands	15.772440
18	New Zealand	19.857520
19	Norway	4.713736
20	Poland	19.726530
21	Portugal	15.484280
22	Spain	15.843730
23	Sweden	10.595570
24	Switzerland	15.300250
25	United States	6.595855
26	Brazil	33.252250
27	Chile	24.820500
28	Israel	12.447800
29	Slovenia	14.005970
30	Colombia	53.067870
31	Euro area (19 countries)	14.252470
32	Latvia	12.981970
33	Lithuania	11.628460

	Country	%informalidad
34	Costa Rica	27.440160
35	European Union – 27 countries (from 01/02/2020)	14.515310

In [262…

```
Informalidad_OCDE.drop([26,31,35], axis=0, inplace=True)
Informalidad_OCDE
```

Out[262]:

	Country	%informalidad
0	Australia	9.562654
1	Austria	11.910980
2	Belgium	14.140240
3	Canada	7.633201
4	Czechia	15.936410
5	Denmark	8.839274
6	Finland	14.572440
7	France	12.606570
8	Germany	8.931904
9	Greece	31.824720
10	Hungary	12.507940
11	Ireland	14.063140
12	Italy	21.831650
13	Japan	9.832524
14	Korea	23.906210
15	Luxembourg	10.225700
16	Mexico	31.826360
17	Netherlands	15.772440
18	New Zealand	19.857520
19	Norway	4.713736
20	Poland	19.726530
21	Portugal	15.484280
22	Spain	15.843730
23	Sweden	10.595570
24	Switzerland	15.300250
25	United States	6.595855
27	Chile	24.820500
28	Israel	12.447800
29	Slovenia	14.005970
30	Colombia	53.067870
32	Latvia	12.981970
33	Lithuania	11.628460
34	Costa Rica	27.440160

In [263…
```
Impoempresas_OCDE = pd.read_csv("Corporate_tax.csv", sep= ",")
SalarioMedio_OCDE = pd.read_csv("Wage_median.csv", sep= ",")
SalarioMinimo_OCDE = pd.read_csv("RMW.csv",sep= ",")
PPA = pd.read_csv("PPA_OCDE.csv", sep= ",")
```

In [264…
```
SalarioMedio_OCDE
```

Out[264]:

	COUNTRY	Country	SERIES	Series	TIME	Time	Unit Code	Unit	PowerCode Code	PowerCode
0	AUS	Australia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
1	BEL	Belgium	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
2	CAN	Canada	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
3	CZE	Czechia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
4	FRA	France	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
5	GRC	Greece	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
6	HUN	Hungary	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
7	IRL	Ireland	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
8	JPN	Japan	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
9	KOR	Korea	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
10	LUX	Luxembourg	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
11	MEX	Mexico	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
12	NLD	Netherlands	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
13	NZL	New Zealand	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
14	POL	Poland	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
15	PRT	Portugal	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
16	SVK	Slovak Republic	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
17	ESP	Spain	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
18	TUR	Türkiye	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
19	GBR	United Kingdom	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
20	USA	United States	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
21	SVN	Slovenia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
22	EST	Estonia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
23	LTU	Lithuania	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
24	ROU	Romania	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
25	LVA	Latvia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
26	ISR	Israel	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
27	CHL	Chile	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
28	DEU	Germany	MEDIAN	Median	2021	2021	PC	Percentage	0	Units
29	COL	Colombia	MEDIAN	Median	2021	2021	PC	Percentage	0	Units

	COUNTRY	Country	SERIES	Series	TIME	Time	Unit Code	Unit	PowerCode Code	PowerCode
30	CRI	Costa Rica	MEDIAN	Median	2021	2021	PC	Percentage	0	Units

In [265…
```
SalarioMedio_OCDE.columns
```

Out[265]:
```
Index(['COUNTRY', 'Country', 'SERIES', 'Series', 'TIME', 'Time', 'Unit Code',
       'Unit', 'PowerCode Code', 'PowerCode', 'Reference Period Code',
       'Reference Period', 'Value', 'Flag Codes', 'Flags'],
      dtype='object')
```

In [266…
```
SM_OCDE = SalarioMedio_OCDE.drop(columns = ['COUNTRY', 'SERIES', 'Series', 'TIME', 'Ti
       'Unit', 'PowerCode Code', 'PowerCode', 'Reference Period Code',
       'Reference Period', 'Flag Codes', 'Flags'])
SM_OCDE = SM_OCDE.rename(columns = {"Value":"%SM"})
SM_OCDE = SM_OCDE.round(1)
SM_OCDE = SM_OCDE.sort_values(by=["%SM"])
SM_OCDE
```

Out[266]:

	Country	%SM
20	United States	29.1
1	Belgium	40.1
25	Latvia	42.3
22	Estonia	42.6
3	Czechia	43.2
8	Japan	44.9
6	Hungary	45.5
31	Croatia	46.1
7	Ireland	46.2
12	Netherlands	46.3
23	Lithuania	46.7
17	Spain	48.5
2	Canada	48.9
28	Germany	49.5
5	Greece	49.8
0	Australia	51.5
16	Slovak Republic	52.4
26	Israel	53.0
10	Luxembourg	54.9
24	Romania	54.9
14	Poland	55.1
11	Mexico	56.7
19	United Kingdom	57.0
21	Slovenia	60.4
4	France	61.0
9	Korea	61.4
15	Portugal	66.3
13	New Zealand	67.6
18	Türkiye	70.4
27	Chile	72.0
30	Costa Rica	88.3
29	Colombia	90.9

In [267…

```
Impoempresas_OCDE
```

Out[267]:

	COU	Country	CORP_TAX	Corporate income tax rate	YEA	Year	Unit Code	Unit	PowerCode Code	PowerCode	R
0	AUS	Australia	CIT_RATE	Corporate income tax rate	2000	2000	PC	Percentage	0	Units	
1	AUS	Australia	CIT_RATE	Corporate income tax rate	2001	2001	PC	Percentage	0	Units	
2	AUS	Australia	CIT_RATE	Corporate income tax rate	2002	2002	PC	Percentage	0	Units	
3	AUS	Australia	CIT_RATE	Corporate income tax rate	2003	2003	PC	Percentage	0	Units	
4	AUS	Australia	CIT_RATE	Corporate income tax rate	2004	2004	PC	Percentage	0	Units	
...	...	...	...	...	...	...	...	...	...	...	
907	CRI	Costa Rica	CIT_RATE	Corporate income tax rate	2019	2019	PC	Percentage	0	Units	
908	CRI	Costa Rica	CIT_RATE	Corporate income tax rate	2020	2020	PC	Percentage	0	Units	
909	CRI	Costa Rica	CIT_RATE	Corporate income tax rate	2021	2021	PC	Percentage	0	Units	
910	CRI	Costa Rica	CIT_RATE	Corporate income tax rate	2022	2022	PC	Percentage	0	Units	
911	CRI	Costa Rica	CIT_RATE	Corporate income tax rate	2023	2023	PC	Percentage	0	Units	

912 rows × 15 columns

In [268…
```
Impoempresas_OCDE.columns
```

Out[268]:
```
Index(['COU', 'Country', 'CORP_TAX', 'Corporate income tax rate', 'YEA',
       'Year', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Value', 'Flag Codes',
       'Flags'],
      dtype='object')
```

In [269…
```
Impoempresas_OCDE= Impoempresas_OCDE[Impoempresas_OCDE['YEA'] == 2021]
```

In [270…
```
Impoempresas_OCDE = Impoempresas_OCDE.drop(columns = ['COU', 'CORP_TAX', 'Corporate ir
       'Year', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
```

```
       'Reference Period Code', 'Reference Period', 'Flag Codes',
       'Flags'])
Impoempresas_OCDE = Impoempresas_OCDE.rename(columns = {"Value":"%tax"})
Impoempresas_OCDE
```

Out[270]:

	Country	%tax
21	Australia	30.0000
45	Austria	25.0000
69	Belgium	25.0000
93	Canada	15.0000
117	Czechia	19.0000
141	Denmark	22.0000
165	Finland	20.0000
189	France	28.4075
213	Germany	15.8250
237	Greece	22.0000
261	Hungary	9.0000
285	Iceland	20.0000
309	Ireland	12.5000
333	Italy	24.0000
357	Japan	23.2000
381	Korea	25.0000
405	Luxembourg	18.1900
429	Mexico	30.0000
453	Netherlands	25.0000
477	New Zealand	28.0000
501	Norway	22.0000
525	Poland	19.0000
549	Portugal	30.0000
573	Slovak Republic	21.0000
597	Spain	25.0000
621	Sweden	20.6000
645	Switzerland	8.5000
669	Türkiye	25.0000
693	United Kingdom	19.0000
717	United States	21.0000
741	Chile	27.0000
765	Estonia	20.0000
789	Israel	23.0000
813	Latvia	20.0000

	Country	%tax
837	Slovenia	19.0000
861	Lithuania	15.0000
885	Colombia	31.0000
909	Costa Rica	30.0000

In [ ]:

In [271…

```
Impoempresas_OCDE=Impoempresas_OCDE.round(1)
Impoempresas_OCDE
```

Out[271]:

	Country	%tax
21	Australia	30.0
45	Austria	25.0
69	Belgium	25.0
93	Canada	15.0
117	Czechia	19.0
141	Denmark	22.0
165	Finland	20.0
189	France	28.4
213	Germany	15.8
237	Greece	22.0
261	Hungary	9.0
285	Iceland	20.0
309	Ireland	12.5
333	Italy	24.0
357	Japan	23.2
381	Korea	25.0
405	Luxembourg	18.2
429	Mexico	30.0
453	Netherlands	25.0
477	New Zealand	28.0
501	Norway	22.0
525	Poland	19.0
549	Portugal	30.0
573	Slovak Republic	21.0
597	Spain	25.0
621	Sweden	20.6
645	Switzerland	8.5
669	Türkiye	25.0
693	United Kingdom	19.0
717	United States	21.0
741	Chile	27.0
765	Estonia	20.0
789	Israel	23.0
813	Latvia	20.0

	Country	%tax
837	Slovenia	19.0
861	Lithuania	15.0
885	Colombia	31.0
909	Costa Rica	30.0

In [272…
```
SM = SalarioMinimo_OCDE
```

In [273…
```
SM = SM.drop(columns = ['COUNTRY', 'SERIES', 'Series', 'PERIOD', 'Pay period',
       'TIME', 'Time', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Flag Codes',
       'Flags'])
SM = SM.rename(columns={"Value":"SM"})
SM=SM.round()
```

In [274…
```
SM=SM.sort_values(by=["SM"])
SM
```

Out[274]:

	Country	SM
7	Mexico	3384.0
28	Colombia	8708.0
19	Chile	8847.0
29	Costa Rica	11692.0
27	Latvia	11825.0
11	Hungary	12258.0
4	Estonia	12404.0
21	Slovak Republic	12874.0
24	Czechia	13701.0
20	Türkiye	14901.0
2	Portugal	15658.0
16	Greece	15895.0
9	Israel	16149.0
3	United States	16287.0
26	Lithuania	16505.0
12	Japan	17559.0
8	Poland	18702.0
17	Slovenia	20715.0
23	Spain	20933.0
0	Ireland	21905.0
5	Canada	23680.0
18	Korea	23739.0
15	United Kingdom	25120.0
6	France	25137.0
10	Belgium	26278.0
25	Germany	26341.0
13	Australia	27605.0
14	New Zealand	27716.0
1	Luxembourg	28963.0
22	Netherlands	28989.0

In [275…

```
OCDE= pd.merge(Informalidad_OCDE, Impoempresas_OCDE, on=["Country"], how= "left")
OCDE= pd.merge(OCDE, SM_OCDE, on=["Country"], how= "left")
OCDE= pd.merge(OCDE, SM, on=["Country"], how= "left")
OCDE=OCDE.fillna(0)
OCDE.to_csv("BaseOCDE.csv")
```

```
OCDE
#Ni islandia ni Turquia tienen tasa informalidad de 2021 registrada en OCDE
```

Out[275]:

	Country	%informalidad	%tax	%SM	SM
0	Australia	9.562654	30.0	51.5	27605.0
1	Austria	11.910980	25.0	0.0	0.0
2	Belgium	14.140240	25.0	40.1	26278.0
3	Canada	7.633201	15.0	48.9	23680.0
4	Czechia	15.936410	19.0	43.2	13701.0
5	Denmark	8.839274	22.0	0.0	0.0
6	Finland	14.572440	20.0	0.0	0.0
7	France	12.606570	28.4	61.0	25137.0
8	Germany	8.931904	15.8	49.5	26341.0
9	Greece	31.824720	22.0	49.8	15895.0
10	Hungary	12.507940	9.0	45.5	12258.0
11	Ireland	14.063140	12.5	46.2	21905.0
12	Italy	21.831650	24.0	0.0	0.0
13	Japan	9.832524	23.2	44.9	17559.0
14	Korea	23.906210	25.0	61.4	23739.0
15	Luxembourg	10.225700	18.2	54.9	28963.0
16	Mexico	31.826360	30.0	56.7	3384.0
17	Netherlands	15.772440	25.0	46.3	28989.0
18	New Zealand	19.857520	28.0	67.6	27716.0
19	Norway	4.713736	22.0	0.0	0.0
20	Poland	19.726530	19.0	55.1	18702.0
21	Portugal	15.484280	30.0	66.3	15658.0
22	Spain	15.843730	25.0	48.5	20933.0
23	Sweden	10.595570	20.6	0.0	0.0
24	Switzerland	15.300250	8.5	0.0	0.0
25	United States	6.595855	21.0	29.1	16287.0
26	Chile	24.820500	27.0	72.0	8847.0
27	Israel	12.447800	23.0	53.0	16149.0
28	Slovenia	14.005970	19.0	60.4	20715.0
29	Colombia	53.067870	31.0	90.9	8708.0
30	Latvia	12.981970	20.0	42.3	11825.0
31	Lithuania	11.628460	15.0	46.7	16505.0
32	Costa Rica	27.440160	30.0	88.3	11692.0

In [276…
```
Historico_imporenta = pd.read_csv("Corporate_tax.csv", sep=",")
```

In [277…
```
Historico_imporenta = Historico_imporenta[Historico_imporenta["Country"]== "Colombia"]
Historico_imporenta.columns
```

Out[277]:
```
Index(['COU', 'Country', 'CORP_TAX', 'Corporate income tax rate', 'YEA',
       'Year', 'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Value', 'Flag Codes',
       'Flags'],
      dtype='object')
```

In [278…
```
Historico_imporenta = Historico_imporenta.drop(columns=["Country", 'CORP_TAX', 'Corpor
        'Unit Code', 'Unit', 'PowerCode Code', 'PowerCode',
       'Reference Period Code', 'Reference Period', 'Flag Codes',
       'Flags', "COU"])

Historico_imporenta= Historico_imporenta.rename(columns={"Value":"%tax"})
```

In [279…
```
Salarios=[1000000,
908526,
877803,
828116 ,
781242 ,
737717 ,
689455 ,
644350 ,
616000 ,
589500 ,
566700 ,
535600 ,
515000 ,
496900 ,
461500 ,
433700 ,
408000 ,
381500 ,
358000 ,
332000 ,
309000 ,
286000 ,
260100 ]
```

In [280…
```
Periodos = [x for x in range(2000,2023)]
Periodos = Periodos[::-1]
print(Periodos)
```

```
[2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009,
2008, 2007, 2006, 2005, 2004, 2003, 2002, 2001, 2000]
```

In [281…
```
Historicos_SM = pd.DataFrame({"Year":Periodos,
                              "SM":Salarios})
Historicos_SM= Historicos_SM.astype(int)
Historicos_SM
```

Out[281]:

	Year	SM
0	2022	1000000
1	2021	908526
2	2020	877803
3	2019	828116
4	2018	781242
5	2017	737717
6	2016	689455
7	2015	644350
8	2014	616000
9	2013	589500
10	2012	566700
11	2011	535600
12	2010	515000
13	2009	496900
14	2008	461500
15	2007	433700
16	2006	408000
17	2005	381500
18	2004	358000
19	2003	332000
20	2002	309000
21	2001	286000
22	2000	260100

In [282…

```
Historicos_Colombia = pd.merge(Historicos_SM,Historico_imporenta, on=["Year"])
Historicos_Colombia.to_csv("SM_Imporenta.csv")
Historicos_Colombia
```

Out[282]:

	Year	SM	%tax
0	2022	1000000	35.00
1	2021	908526	31.00
2	2020	877803	32.00
3	2019	828116	33.00
4	2018	781242	37.00
5	2017	737717	40.00
6	2016	689455	40.00
7	2015	644350	39.00
8	2014	616000	34.00
9	2013	589500	34.00
10	2012	566700	33.00
11	2011	535600	33.00
12	2010	515000	33.00
13	2009	496900	33.00
14	2008	461500	33.00
15	2007	433700	34.00
16	2006	408000	38.50
17	2005	381500	38.50
18	2004	358000	38.50
19	2003	332000	36.75
20	2002	309000	35.00
21	2001	286000	35.00
22	2000	260100	35.00

In [283…

```
Historico_Informalidad = pd.read_csv("HistInformalidad.csv",sep=",")
Historico_Informalidad
```

Out[283]:

	LOCATION	INDICATOR	SUBJECT	MEASURE	FREQUENCY	TIME	Value	Flag Codes
0	COL	SELFEMP	TOT	PC_EMP	A	2007	47.30726	NaN
1	COL	SELFEMP	TOT	PC_EMP	A	2008	52.05827	NaN
2	COL	SELFEMP	TOT	PC_EMP	A	2009	54.37944	NaN
3	COL	SELFEMP	TOT	PC_EMP	A	2010	55.44683	NaN
4	COL	SELFEMP	TOT	PC_EMP	A	2011	56.02482	NaN
5	COL	SELFEMP	TOT	PC_EMP	A	2012	55.27148	NaN
6	COL	SELFEMP	TOT	PC_EMP	A	2013	54.33998	NaN
7	COL	SELFEMP	TOT	PC_EMP	A	2014	53.49324	NaN
8	COL	SELFEMP	TOT	PC_EMP	A	2015	53.01191	NaN
9	COL	SELFEMP	TOT	PC_EMP	A	2016	52.88547	NaN
10	COL	SELFEMP	TOT	PC_EMP	A	2017	52.76831	NaN
11	COL	SELFEMP	TOT	PC_EMP	A	2018	52.59024	NaN
12	COL	SELFEMP	TOT	PC_EMP	A	2019	49.95089	NaN
13	COL	SELFEMP	TOT	PC_EMP	A	2020	50.63253	NaN
14	COL	SELFEMP	TOT	PC_EMP	A	2021	53.06787	NaN

In [284…

```
Historico_Informalidad.columns
```

Out[284]:

```
Index(['LOCATION', 'INDICATOR', 'SUBJECT', 'MEASURE', 'FREQUENCY', 'TIME',
       'Value', 'Flag Codes'],
      dtype='object')
```

In [285…

```
Historico_Informalidad.drop(columns=['LOCATION', 'INDICATOR', 'SUBJECT', 'MEASURE', 'F
Historico_Informalidad.rename(columns={"TIME":"Year", "Value":"%Informalidad"}, inplac
Historico_Informalidad
```

Out[285]:

	Year	%Informalidad
0	2007	47.30726
1	2008	52.05827
2	2009	54.37944
3	2010	55.44683
4	2011	56.02482
5	2012	55.27148
6	2013	54.33998
7	2014	53.49324
8	2015	53.01191
9	2016	52.88547
10	2017	52.76831
11	2018	52.59024
12	2019	49.95089
13	2020	50.63253
14	2021	53.06787

In [286…

```
Historicos_Colombia = pd.merge(Historicos_Colombia,Historico_Informalidad, on=["Year"]
Historicos_Colombia
```

Out[286]:

	Year	SM	%tax	%Informalidad
0	2021	908526	31.0	53.06787
1	2020	877803	32.0	50.63253
2	2019	828116	33.0	49.95089
3	2018	781242	37.0	52.59024
4	2017	737717	40.0	52.76831
5	2016	689455	40.0	52.88547
6	2015	644350	39.0	53.01191
7	2014	616000	34.0	53.49324
8	2013	589500	34.0	54.33998
9	2012	566700	33.0	55.27148
10	2011	535600	33.0	56.02482
11	2010	515000	33.0	55.44683
12	2009	496900	33.0	54.37944
13	2008	461500	33.0	52.05827
14	2007	433700	34.0	47.30726

In [ ]: