database2

April 27, 2024

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

[3]:
```
df = pd.read_csv('C:\\Users\\lenovo\\Desktop\\courses\\Task␣
↪1\\Metadata_Country_API_SP.POP.2529.FE.5Y_DS2_en_csv_v2_52627.csv', sep=',')
```

[4]:
```
df
```

[4]:

	Country Code	Region	IncomeGroup	SpecialNotes	TableName	Unnamed: 5
0	ABW	Latin America & Caribbean	High income	NaN	Aruba	NaN
1	AFE	NaN	NaN	26 countries, stretching from the Red Sea in t…		
2	AFG	South Asia	Low income	The reporting period for national accounts dat…		
3	AFW	NaN	NaN	22 countries, stretching from the westernmost …		
4	AGO	Sub-Saharan Africa	Lower middle income	The World Bank systematically assesses the app…		
..	…	…	…	…		
260	XKX	Europe & Central Asia	Upper middle income	NaN		
261	YEM	Middle East & North Africa	Low income	The World Bank systematically assesses the app…		
262	ZAF	Sub-Saharan Africa	Upper middle income	Fiscal year end: March 31; reporting period fo…		
263	ZMB	Sub-Saharan Africa	Lower middle income	National accounts data were rebased to reflect…		
264	ZWE	Sub-Saharan Africa	Lower middle income	National Accounts data are reported in Zimbabw…		

```
1    Africa Eastern and Southern         NaN
2                    Afghanistan         NaN
3     Africa Western and Central         NaN
4                         Angola         NaN
..                           …         …
260                       Kosovo         NaN
261                  Yemen, Rep.         NaN
262                 South Africa         NaN
263                       Zambia         NaN
264                     Zimbabwe         NaN

[265 rows x 6 columns]
```

[5]:
```
print("Les premières lignes du DataFrame :")
print(df.head())
```

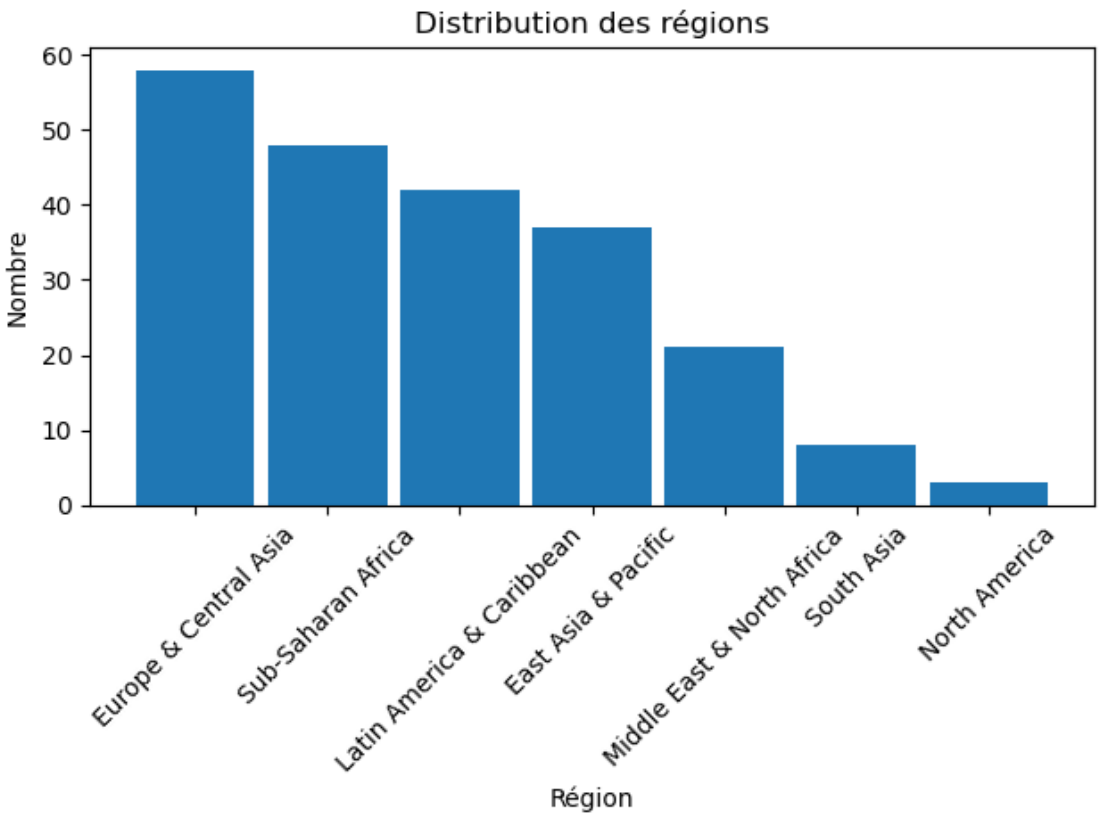
```
Les premières lignes du DataFrame :
  Country Code                     Region          IncomeGroup  \
0          ABW  Latin America & Caribbean          High income
1          AFE                        NaN                  NaN
2          AFG                 South Asia           Low income
3          AFW                        NaN                  NaN
4          AGO        Sub-Saharan Africa  Lower middle income

                                        SpecialNotes  \
0                                                NaN
1  26 countries, stretching from the Red Sea in t…
2  The reporting period for national accounts dat…
3  22 countries, stretching from the westernmost …
4  The World Bank systematically assesses the app…

                    TableName  Unnamed: 5
0                       Aruba         NaN
1  Africa Eastern and Southern         NaN
2                  Afghanistan         NaN
3   Africa Western and Central         NaN
4                       Angola         NaN
```

[6]:
```
print("\nInformations sur le DataFrame :")
print(df.info())
```

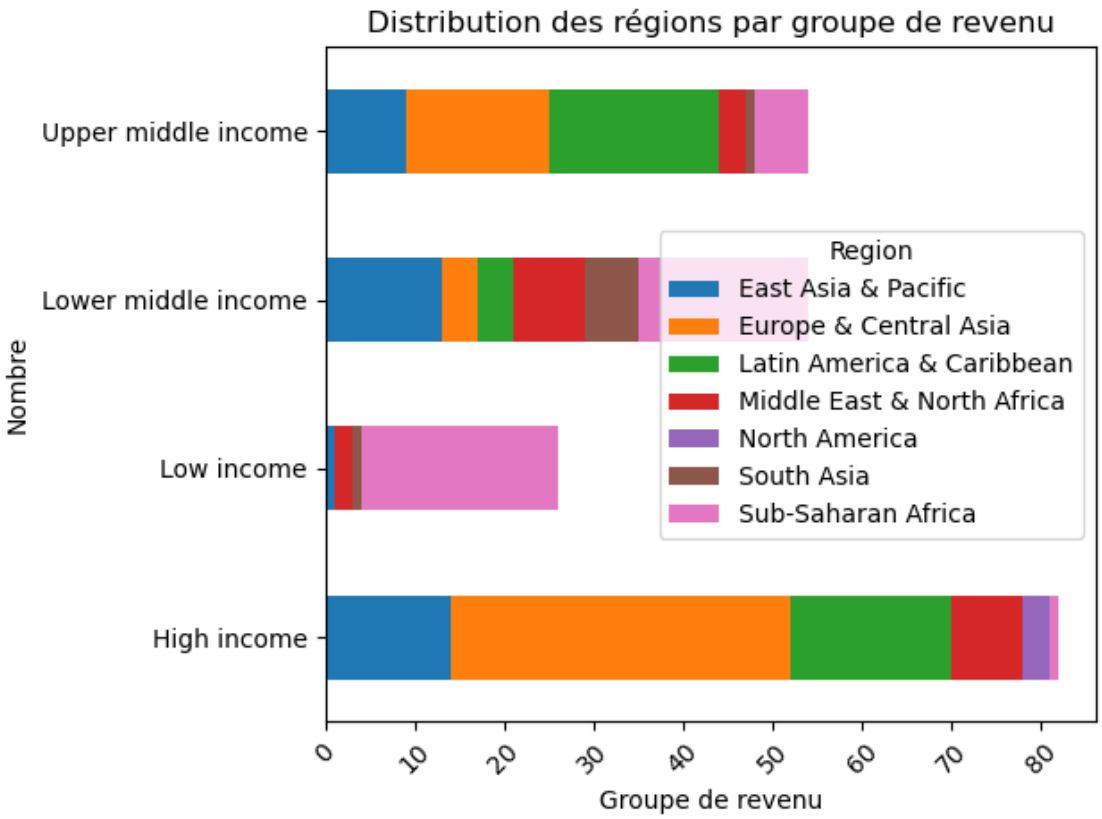
```
Informations sur le DataFrame :
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 265 entries, 0 to 264
Data columns (total 6 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
```

```
 0   Country Code  265 non-null    object
 1   Region        217 non-null    object
 2   IncomeGroup   216 non-null    object
 3   SpecialNotes  126 non-null    object
 4   TableName     265 non-null    object
 5   Unnamed: 5    0 non-null      float64
dtypes: float64(1), object(5)
memory usage: 12.6+ KB
None
```

[9]:
```
region_counts = df['Region'].value_counts()

bar_width = 0.9


plt.bar(region_counts.index, region_counts.values, width=bar_width)

plt.xlabel('Région')
plt.ylabel('Nombre')
plt.title('Distribution des régions')

plt.xticks(rotation=45)


plt.tight_layout()
plt.show()
```



[10]:
```
region_income_counts = df.groupby(['IncomeGroup', 'Region']).size()

region_income_counts.unstack().plot(kind='barh', stacked=True)

plt.xlabel('Groupe de revenu')
plt.ylabel('Nombre')
plt.title('Distribution des régions par groupe de revenu')

plt.xticks(rotation=45)

plt.legend(title='Region')

plt.tight_layout()
plt.show()
```

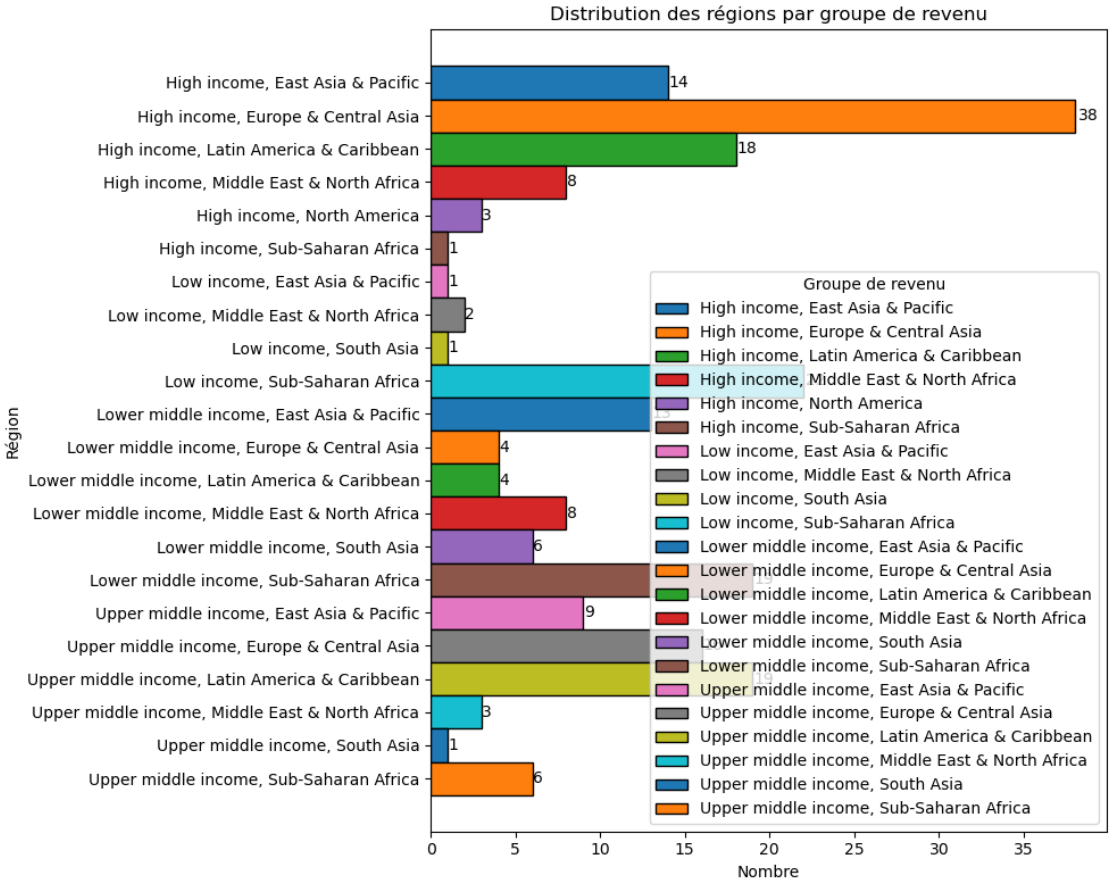


[14]:
```
print(region_income_counts)
```

```
IncomeGroup           Region
High income           East Asia & Pacific           14
                      Europe & Central Asia         38
                      Latin America & Caribbean     18
                      Middle East & North Africa     8
                      North America                  3
                      Sub-Saharan Africa             1
Low income            East Asia & Pacific            1
                      Middle East & North Africa     2
                      South Asia                     1
                      Sub-Saharan Africa            22
Lower middle income   East Asia & Pacific           13
                      Europe & Central Asia          4
                      Latin America & Caribbean      4
                      Middle East & North Africa     8
                      South Asia                     6
                      Sub-Saharan Africa            19
Upper middle income   East Asia & Pacific            9
                      Europe & Central Asia         16
```

```
                        Latin America & Caribbean     19
                        Middle East & North Africa     3
                        South Asia                     1
                        Sub-Saharan Africa             6
dtype: int64
```

[17]:
```
region_income_counts = df.groupby(['IncomeGroup', 'Region']).size()

region_income_counts.index = region_income_counts.index.map(lambda x: ', '.
 ↪join(x))

fig, ax = plt.subplots(figsize=(10, 8))

for i, (income_group, region_counts) in enumerate(region_income_counts.
 ↪groupby(level=0)):
    ax.barh(region_counts.index, region_counts.values, label=income_group,␣
 ↪height=1.0, edgecolor='black')

plt.xlabel('Nombre')
plt.ylabel('Région')
plt.title('Distribution des régions par groupe de revenu')

for p in ax.patches:
    ax.annotate(str(p.get_width()), (p.get_width() * 1.005, p.get_y() + p.
 ↪get_height() / 2), va='center')

plt.legend(title='Groupe de revenu')

plt.gca().invert_yaxis()

plt.tight_layout()
plt.show()
```


Distribution des régions par groupe de revenu
Région
Nombre
High income, East Asia & Pacific
High income, Europe & Central Asia
High income, Latin America & Caribbean
High income, Middle East & North Africa
High income, North America
High income, Sub-Saharan Africa
Low income, East Asia & Pacific
Low income, Middle East & North Africa
Low income, South Asia
Low income, Sub-Saharan Africa
Lower middle income, East Asia & Pacific
Lower middle income, Europe & Central Asia
Lower middle income, Latin America & Caribbean
Lower middle income, Middle East & North Africa
Lower middle income, South Asia
Lower middle income, Sub-Saharan Africa
Upper middle income, East Asia & Pacific
Upper middle income, Europe & Central Asia
Upper middle income, Latin America & Caribbean
Upper middle income, Middle East & North Africa
Upper middle income, South Asia
Upper middle income, Sub-Saharan Africa
14
38
18
8
3
1
1
2
1
4
4
8
6
9
3
1
6
0
5
10
15
20
25
30
35

[ ]: