

worldbank-dataset2

April 20, 2024

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

```
[5]: df = pd.read_csv('C:\\Users\\EL HASSANI_
↳SAFAA\\Desktop\\Task1_Datasets\\Metadata_Country_API_SP.POP.
↳TOTL_DS2_en_csv_v2_84031.csv')
```

```
[6]: df
```

```
[6]:   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
..      ...                  ...          ...
260     XKX  Europe & Central Asia  Upper middle income
261     YEM  Middle East & North Africa  Low income
262     ZAF  Sub-Saharan Africa  Upper middle income
263     ZMB  Sub-Saharan Africa  Lower middle income
264     ZWE  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...
..                                      ...
260                                     NaN
261  The World Bank systematically assesses the app...
262  Fiscal year end: March 31; reporting period fo...
263  National accounts data were rebased to reflect...
264  National Accounts data are reported in Zimbabw...
```

TableName Unnamed: 5

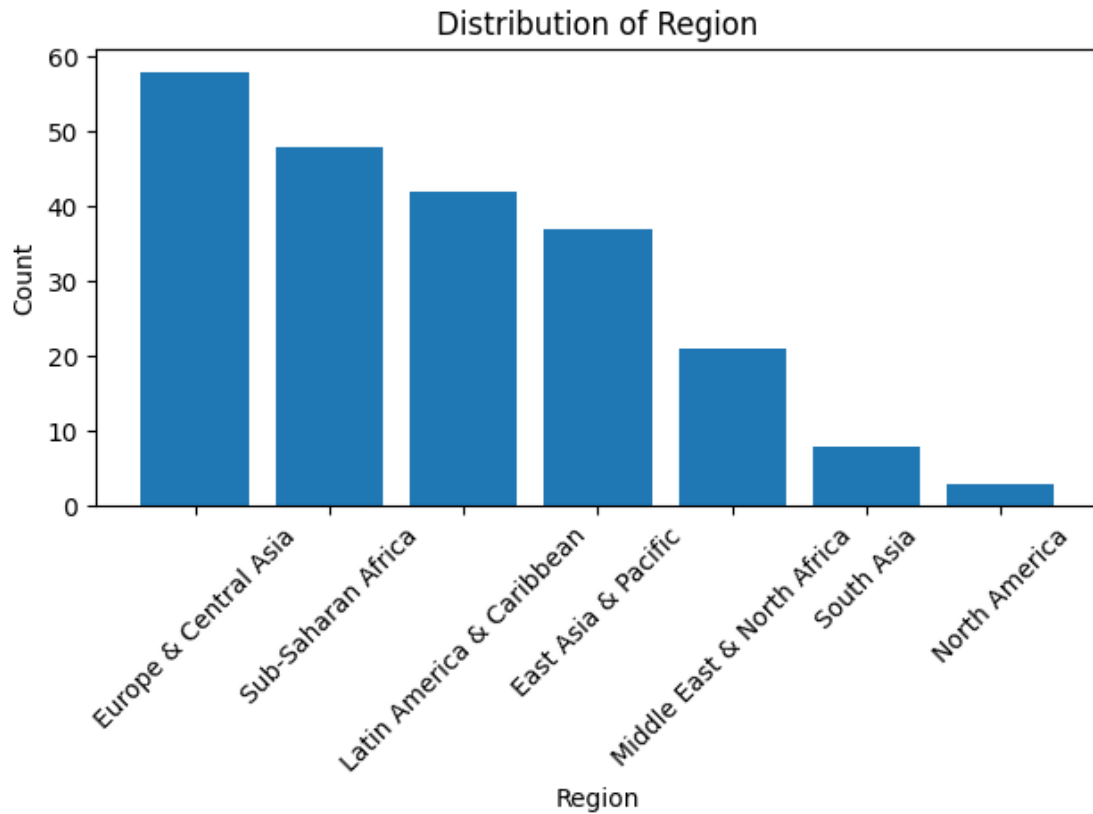
0	Aruba	NaN
1	Africa Eastern and Southern	NaN
2	Afghanistan	NaN
3	Africa Western and Central	NaN
4	Angola	NaN
..
260	Kosovo	NaN
261	Yemen, Rep.	NaN
262	South Africa	NaN
263	Zambia	NaN
264	Zimbabwe	NaN

[265 rows x 6 columns]

```
[7]: gender_counts = df['Region'].value_counts()
bar_width = 0.9
x=range(len(gender_counts.index))

plt.bar(gender_counts.index,gender_counts.values)
plt.xlabel('Region')
plt.ylabel('Count')
plt.title('Distribution of Region')

plt.xticks(x,gender_counts.index,rotation=45)
plt.tight_layout()
plt.show()
```



```
[8]: df.shape
```

```
[8]: (265, 6)
```

```
[9]: df.info()
```

```
<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
```

```
[10]: df.describe()
```

```
[10]:      Unnamed: 5
      count      0.0
      mean      NaN
      std       NaN
      min       NaN
      25%       NaN
      50%       NaN
      75%       NaN
      max       NaN
```

```
[11]: df.isnull().sum()
```

```
[11]: Country Code      0
      Region          48
      IncomeGroup      49
      SpecialNotes    139
      TableName        0
      Unnamed: 5      265
      dtype: int64
```

```
[12]: df.info()
```

```
<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
```

```
[ ]:
```