

worldbank2-1-checkpoint

March 7, 2024

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

```
[27]: df = pd.read_csv(r"C:\Users\lenovo\Downloads\PRODIGY_DS_01-Intern-Aakash-Jupyter-Lab-TASK-01-main\Aakash_WorldBank2-1-checkpoint\POP.TOTL_DS2_en_csv_v2_5871594.csv")
```

```
[20]: df
```

```
[20]:
```

	Country Name	Country Code	Indicator Name	\
0	Aruba	ABW	Population, total	
1	Africa Eastern and Southern	AFE	Population, total	
2	Afghanistan	AFG	Population, total	
3	Africa Western and Central	AFW	Population, total	
4	Angola	AGO	Population, total	
..	
261	Kosovo	XKX	Population, total	
262	Yemen, Rep.	YEM	Population, total	
263	South Africa	ZAF	Population, total	
264	Zambia	ZMB	Population, total	
265	Zimbabwe	ZWE	Population, total	

	Indicator Code	1960	1961	1962	1963	\
0	SP.POP.TOTL	54608.0	55811.0	56682.0	57475.0	
1	SP.POP.TOTL	130692579.0	134169237.0	137835590.0	141630546.0	
2	SP.POP.TOTL	8622466.0	8790140.0	8969047.0	9157465.0	
3	SP.POP.TOTL	97256290.0	99314028.0	101445032.0	103667517.0	
4	SP.POP.TOTL	5357195.0	5441333.0	5521400.0	5599827.0	
..	
261	SP.POP.TOTL	947000.0	966000.0	994000.0	1022000.0	
262	SP.POP.TOTL	5542459.0	5646668.0	5753386.0	5860197.0	
263	SP.POP.TOTL	16520441.0	16989464.0	17503133.0	18042215.0	
264	SP.POP.TOTL	3119430.0	3219451.0	3323427.0	3431381.0	
265	SP.POP.TOTL	3806310.0	3925952.0	4049778.0	4177931.0	

	1964	1965	...	2013	2014	2015	\
--	------	------	-----	------	------	------	---

0	58178.0	58782.0	...	102880.0	103594.0	104257.0
1	145605995.0	149742351.0	...	567892149.0	583651101.0	600008424.0
2	9355514.0	9565147.0	...	31541209.0	32716210.0	33753499.0
3	105959979.0	108336203.0	...	387204553.0	397855507.0	408690375.0
4	5673199.0	5736582.0	...	26147002.0	27128337.0	28127721.0
..
261	1050000.0	1078000.0	...	1818117.0	1812771.0	1788196.0
262	5973803.0	6097298.0	...	26984002.0	27753304.0	28516545.0
263	18603097.0	19187194.0	...	53873616.0	54729551.0	55876504.0
264	3542764.0	3658024.0	...	15234976.0	15737793.0	16248230.0
265	4310332.0	4447149.0	...	13555422.0	13855753.0	14154937.0

	2016	2017	2018	2019	2020 \
0	104874.0	105439.0	105962.0	106442.0	106585.0
1	616377605.0	632746570.0	649757148.0	667242986.0	685112979.0
2	34636207.0	35643418.0	36686784.0	37769499.0	38972230.0
3	419778384.0	431138704.0	442646825.0	454306063.0	466189102.0
4	29154746.0	30208628.0	31273533.0	32353588.0	33428486.0
..
261	1777557.0	1791003.0	1797085.0	1788878.0	1790133.0
262	29274002.0	30034389.0	30790513.0	31546691.0	32284046.0
263	56422274.0	56641209.0	57339635.0	58087055.0	58801927.0
264	16767761.0	17298054.0	17835893.0	18380477.0	18927715.0
265	14452704.0	14751101.0	15052184.0	15354608.0	15669666.0

	2021	2022
0	106537.0	106445.0
1	702977106.0	720839314.0
2	40099462.0	41128771.0
3	478185907.0	490330870.0
4	34503774.0	35588987.0
..
261	1786038.0	1761985.0
262	32981641.0	33696614.0
263	59392255.0	59893885.0
264	19473125.0	20017675.0
265	15993524.0	16320537.0

[266 rows x 67 columns]

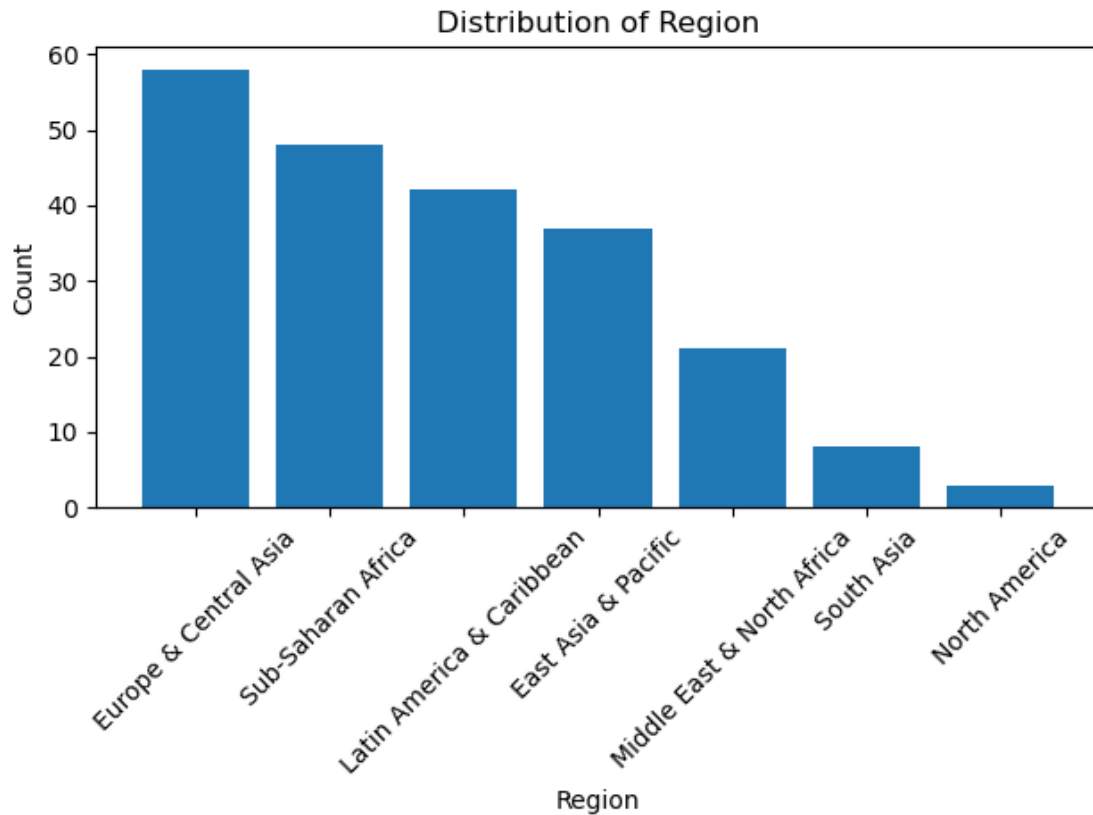
```
[28]: print(df.columns)
```

```
Index(['Country Code', 'Region', 'IncomeGroup', 'SpecialNotes', 'TableName',
      'Unnamed: 5'],
      dtype='object')
```

```
[29]: 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()
```



```
[30]: df.shape
```

```
[30]: (265, 6)
```

```
[31]: 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    127 non-null   object
4   TableName       265 non-null   object
5   Unnamed: 5      0 non-null     float64
dtypes: float64(1), object(5)
memory usage: 12.5+ KB

```

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

```

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

```

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

```

[33]: Country Code      0
      Region           48
      IncomeGroup      49
      SpecialNotes    138
      TableName        0
      Unnamed: 5      265
      dtype: int64

```

```
[34]: 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    127 non-null   object
4   TableName       265 non-null   object

```

```
5    Unnamed: 5    0 non-null    float64
dtypes: float64(1), object(5)
memory usage: 12.5+ KB
```

[]: