

```
In [1]: !pip install opendatasets
```

```
Requirement already satisfied: opendatasets in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (0.1.22)
Requirement already satisfied: tqdm in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from opendatasets) (4.66.4)
Requirement already satisfied: kaggle in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from opendatasets) (1.6.14)
Requirement already satisfied: click in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from opendatasets) (8.1.7)
Requirement already satisfied: colorama in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from click->opendatasets) (0.4.6)
Requirement already satisfied: six>=1.10 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (1.16.0)
Requirement already satisfied: certifi>=2023.7.22 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (2024.7.4)
Requirement already satisfied: python-dateutil in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (2.9.0.post0)
Requirement already satisfied: requests in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (2.32.3)
Requirement already satisfied: python-slugify in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (8.0.4)
Requirement already satisfied: urllib3 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (2.2.2)
Requirement already satisfied: bleach in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from kaggle->opendatasets) (6.1.0)
Requirement already satisfied: webencodings in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from bleach->kaggle->opendatasets) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from python-slugify->kaggle->opendatasets) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from requests->kaggle->opendatasets) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\users\yogit\appdata\local\programs\python\python312\lib\site-packages (from requests->kaggle->opendatasets) (3.7)

[notice] A new release of pip is available: 24.0 -> 24.1.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
In [2]: import opendatasets as od
```

```
In [3]: dataset = 'https://data.worldbank.org/indicator/SP.POP.TOTL'
```

```
In [4]: od.download(dataset)
```

Using downloaded and verified file: .\SP.POP.TOTL

```
In [5]: import os
```

```
In [6]: data_dir = './SP.POP.TOTL'
```

```
In [7]: import pandas as pd
```

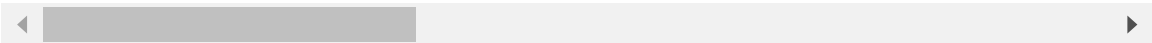
```
In [8]: df = pd.read_csv('API_SP.POP.TOTL_DS2_en_csv_v2_589802.csv')
```

```
In [9]: df
```

Out[9]:

	Data Source	World Development Indicators	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnam
0	NaN	NaN	NaN	NaN	NaN	NaN	N
1	Last Updated Date	28-06-2024	NaN	NaN	NaN	NaN	N
2	NaN	NaN	NaN	NaN	NaN	NaN	N
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	196
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	566
...
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	103861
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	575338
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	1750313
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	332342
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	404977

270 rows × 68 columns

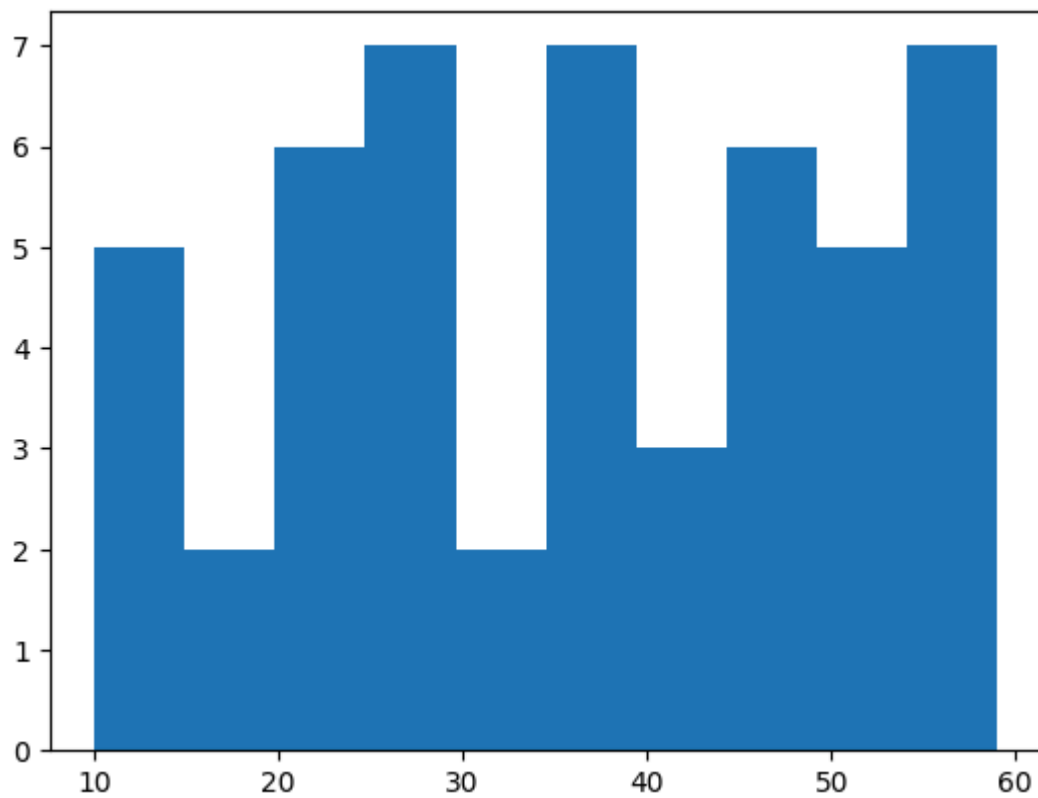


```
In [10]: import matplotlib.pyplot as plt
%matplotlib inline
from matplotlib import style
import pandas as pd
import numpy as np
import random
x=np.random.randint(10,60,(50))
print(x)
```

[55 29 19 58 54 40 59 29 44 59 38 19 12 16 35 48 30 12 28 22 25 20 33 11
29 49 48 52 13 28 59 43 44 46 34 19 54 30 45 15 47 59 44 44 40 41 28 24
18 51]

```
In [11]: no=[22,48,22,51,58,24,38,29,55,49,11,24,41,57,12,31,45,26,59,53,26,37,29,36,30,1
plt.hist(no)

plt.show()
```



```
In [12]: df.isnull().sum()
```

```
Out[12]: Data Source                2
World Development Indicators      2
Unnamed: 2                        3
Unnamed: 3                        3
Unnamed: 4                        5
..
Unnamed: 63                      4
Unnamed: 64                      4
Unnamed: 65                      4
Unnamed: 66                      4
Unnamed: 67                      4
Length: 68, dtype: int64

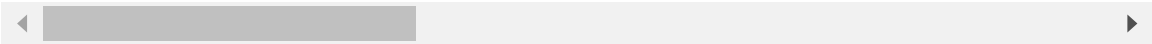
df.isnull().sum().sum()
```

```
In [13]: df.fillna(value = 0)
```

Out[13]:

	Data Source	World Development Indicators	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnam
0	0	0	0	0	0.0	0.0	
1	Last Updated Date	28-06-2024	0	0	0.0	0.0	
2	0	0	0	0	0.0	0.0	
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	196
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	566
...	
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	103861
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	575338
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	1750313
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	332342
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	404977

270 rows × 68 columns



```
In [14]: df.isnull().sum().sum()
```

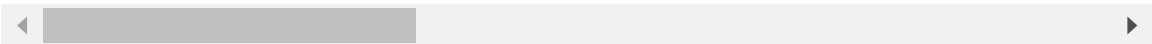
Out[14]: np.int64(296)

```
In [15]: df
```

Out[15]:

	Data Source	World Development Indicators	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnam
0	NaN	NaN	NaN	NaN	NaN	NaN	N
1	Last Updated Date	28-06-2024	NaN	NaN	NaN	NaN	N
2	NaN	NaN	NaN	NaN	NaN	NaN	N
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	196
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	566
...
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	103861
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	575338
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	1750313
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	332342
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	404977

270 rows × 68 columns

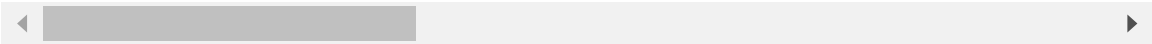


In [16]: df.fillna(value = 0)

Out[16]:

	Data Source	World Development Indicators	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnam
0	0	0	0	0	0.0	0.0	
1	Last Updated Date	28-06-2024	0	0	0.0	0.0	
2	0	0	0	0	0.0	0.0	
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	196
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	566
...	
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	103861
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	575338
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	1750313
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	332342
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	404977

270 rows × 68 columns



In [17]:

```
df = df.rename(columns={'Unnamed: 2':'indicator name'})
df
```

Out[17]:

	Data Source	World Development Indicators	indicator name	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnam
0	NaN	NaN	NaN	NaN	NaN	NaN	N
1	Last Updated Date	28-06-2024	NaN	NaN	NaN	NaN	N
2	NaN	NaN	NaN	NaN	NaN	NaN	N
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	196
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	566
...
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	103861
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	575338
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	1750313
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	332342
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	404977

270 rows × 68 columns



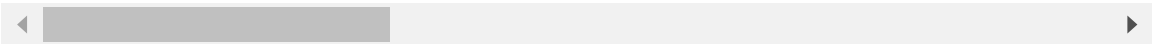
In [18]:

```
df.dropna(axis = 0)
```

Out[18]:

	Data Source	World Development Indicators	indicator name	Unnamed: 3	Unnamed: 4	Unnamed: 5	Un
3	Country Name	Country Code	Indicator Name	Indicator Code	1960.0	1961.0	
4	Aruba	ABW	Population, total	SP.POP.TOTL	54608.0	55811.0	
5	Africa Eastern and Southern	AFE	Population, total	SP.POP.TOTL	130692579.0	134169237.0	1371
6	Afghanistan	AFG	Population, total	SP.POP.TOTL	8622466.0	8790140.0	89
7	Africa Western and Central	AFW	Population, total	SP.POP.TOTL	97256290.0	99314028.0	1014
...	
265	Kosovo	XKX	Population, total	SP.POP.TOTL	990150.0	1014211.0	10
266	Yemen, Rep.	YEM	Population, total	SP.POP.TOTL	5542459.0	5646668.0	57
267	South Africa	ZAF	Population, total	SP.POP.TOTL	16520441.0	16989464.0	171
268	Zambia	ZMB	Population, total	SP.POP.TOTL	3119430.0	3219451.0	33
269	Zimbabwe	ZWE	Population, total	SP.POP.TOTL	3806310.0	3925952.0	40

265 rows × 68 columns



```
import matplotlib.pyplot as plt # Replace [Unnamed: 4] with the actual population data for the top 10 rows Unnamed: 4 =
[number_1, number_17, number_33, number_49, number_65, number_81, number_97, number_113, number_129] # Replace
[country_names] with the actual names of the countries for the top 10 rows country_names = ['Aruba', 'Africa Eastern and
Southern', 'Afghanistan', 'Africa Western and Central', 'Kosovo', 'Yemen, Rep.', 'South Africa', 'Zambia', 'Zimbabwe']
plt.figure(figsize=(10, 6)) plt.bar(country_names, population_data, color='skyblue') plt.xlabel('Country')
plt.ylabel('Population') plt.title('Top 10 Countries by Population') plt.xticks(rotation=45) plt.show()
```

```
In [19]: import matplotlib.pyplot as plt

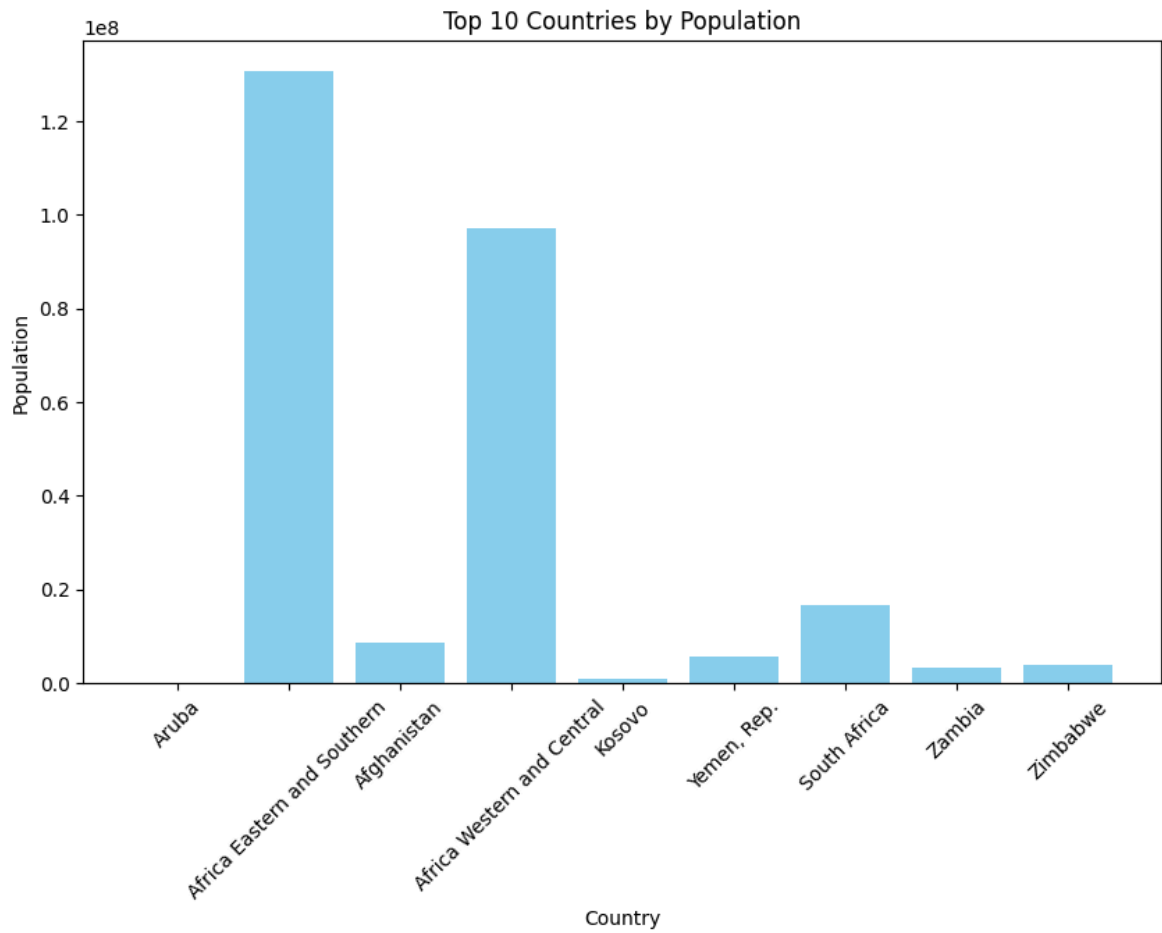
# Replace [Unnamed: 4] with the actual population data for the top 10 rows
population = [54608,130692579,8622466,97256290,947000,5542459,16520441,3119430,3

# Replace [country_names] with the actual names of the countries for the top 10
country_names = ['Aruba', 'Africa Eastern and Southern', 'Afghanistan', 'Africa

plt.figure(figsize=(10, 6))
plt.bar(country_names,population, color='skyblue')
plt.xlabel('Country')
plt.ylabel('Population')
plt.title('Top 10 Countries by Population')
```

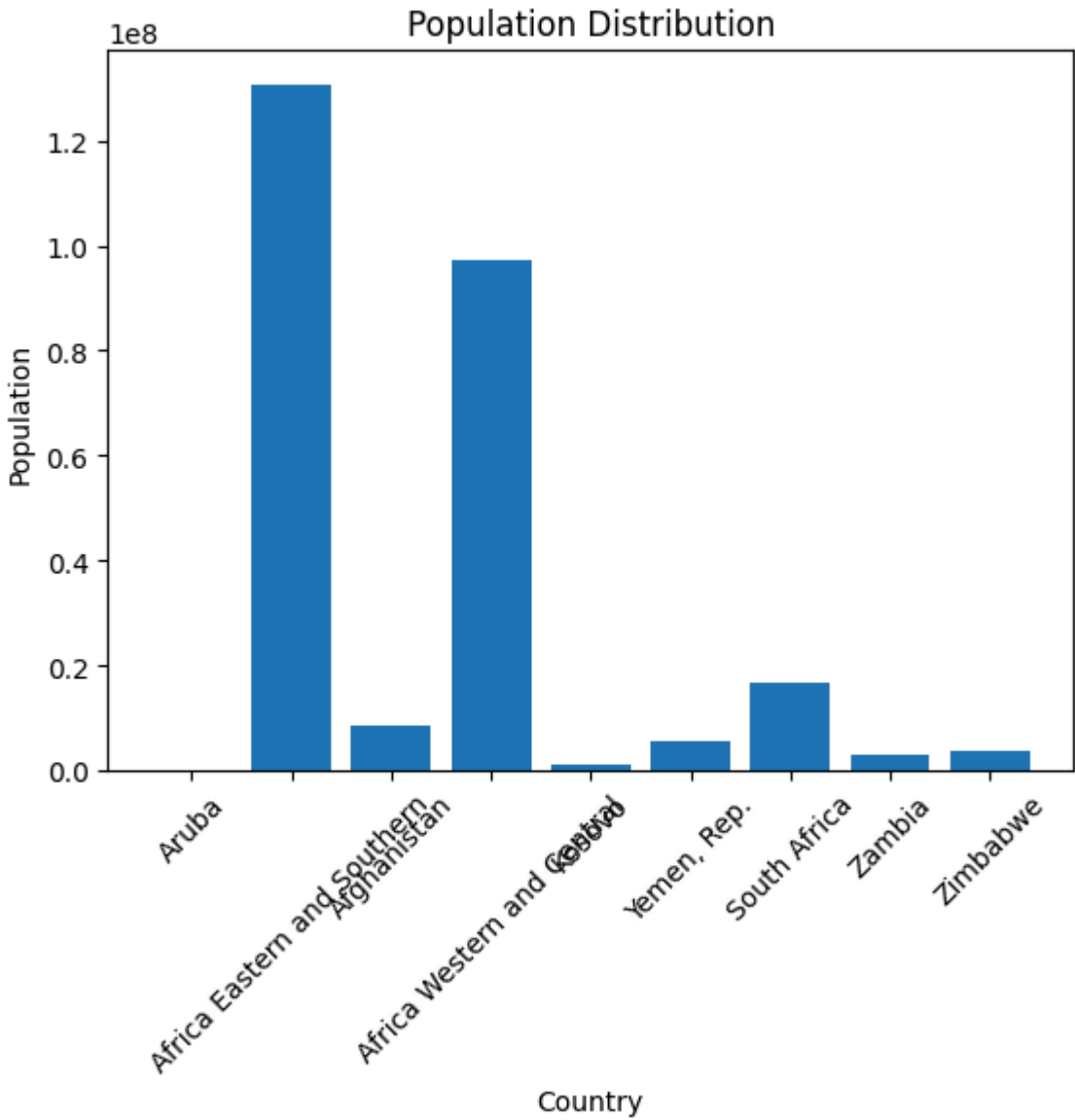


```
plt.xticks(rotation=45)
plt.show()
```



```
In [21]: population = [54608,130692579,8622466,97256290,947000,5542459,16520441,3119430,3

# Plotting the histogram
plt.bar(country_names, population)
plt.xlabel('Country')
plt.ylabel('Population')
plt.title('Population Distribution')
plt.xticks(rotation=45)
plt.show()
```



```
In [ ]:
```