

PRODIGY_DS_01

Task #1: Create a barchart or histogram to visualize the distribution of a categorical or continuous variable, such as the distribution of ages or gender in the population

```
In [1]: ▶ #IMPORTING LIBRARIES  
import pandas as pd  
import numpy as np  
import matplotlib as mpl  
import matplotlib.pyplot as pyp  
import seaborn as sns  
import matplotlib.pyplot as plt
```

```
In [41]: ▶ #IMPORTING DATASET  
df=pd.read_csv("C:/Users/HP/Desktop/Internship/Prodigy/world_population
```

Overview Of the given Dataset

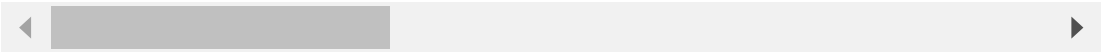
In [42]:

df

Out[42]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962
0	India	IND	Population, total	SP.POP.TOTL	445954579.0	456351876.0	467024193
1	China	CHN	Population, total	SP.POP.TOTL	667070000.0	660330000.0	665770000
2	Sub-Saharan Africa	SSF	Population, total	SP.POP.TOTL	227948869.0	233483265.0	239280622
3	Sub-Saharan Africa (IDA & IBRD countries)	TSS	Population, total	SP.POP.TOTL	227948869.0	233483265.0	239280622
4	Sub-Saharan Africa (excluding high income)	SSA	Population, total	SP.POP.TOTL	227907169.0	233440376.0	239236580
...
220	St. Martin (French part)	MAF	Population, total	SP.POP.TOTL	4135.0	4258.0	4388
221	British Virgin Islands	VGB	Population, total	SP.POP.TOTL	7850.0	7885.0	7902
222	Palau	PLW	Population, total	SP.POP.TOTL	9446.0	9639.0	9851
223	Nauru	NRU	Population, total	SP.POP.TOTL	4582.0	4753.0	4950
224	Tuvalu	TUV	Population, total	SP.POP.TOTL	5404.0	5436.0	5471

225 rows × 67 columns

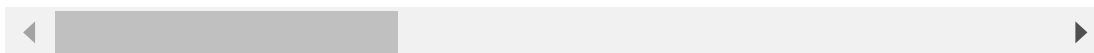


In [43]: `df.head()`

Out[43]:

	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962
0	India	IND	Population, total	SP.POP.TOTL	445954579.0	456351876.0	467024193.0
1	China	CHN	Population, total	SP.POP.TOTL	667070000.0	660330000.0	665770000.0
2	Sub-Saharan Africa	SSF	Population, total	SP.POP.TOTL	227948869.0	233483265.0	239280622.0
3	Sub-Saharan Africa (IDA & IBRD countries)	TSS	Population, total	SP.POP.TOTL	227948869.0	233483265.0	239280622.0
4	Sub-Saharan Africa (excluding high income)	SSA	Population, total	SP.POP.TOTL	227907169.0	233440376.0	239236580.0

5 rows × 67 columns



In [44]: `df.shape`

Out[44]: (225, 67)

In [45]: `df.dtypes`

```
Out[45]: Country Name      object
Country Code      object
Indicator Name     object
Indicator Code     object
1960              float64
...
2018              int64
2019              int64
2020              int64
2021              int64
2022              int64
Length: 67, dtype: object
```

In [46]: ▶ `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 225 entries, 0 to 224
```

```
Data columns (total 67 columns):
```

#	Column	Non-Null Count	Dtype
0	Country Name	225 non-null	object
1	Country Code	225 non-null	object
2	Indicator Name	225 non-null	object
3	Indicator Code	225 non-null	object
4	1960	224 non-null	float64
5	1961	224 non-null	float64
6	1962	224 non-null	float64
7	1963	224 non-null	float64
8	1964	224 non-null	float64
9	1965	224 non-null	float64
10	1966	224 non-null	float64
11	1967	224 non-null	float64
12	1968	224 non-null	float64
13	1969	224 non-null	float64
14	1970	224 non-null	float64
15	1971	224 non-null	float64
16	1972	224 non-null	float64
17	1973	224 non-null	float64
18	1974	224 non-null	float64
19	1975	224 non-null	float64
20	1976	224 non-null	float64
21	1977	224 non-null	float64
22	1978	224 non-null	float64
23	1979	224 non-null	float64
24	1980	224 non-null	float64
25	1981	224 non-null	float64
26	1982	224 non-null	float64
27	1983	224 non-null	float64
28	1984	224 non-null	float64
29	1985	224 non-null	float64
30	1986	224 non-null	float64
31	1987	224 non-null	float64
32	1988	224 non-null	float64
33	1989	224 non-null	float64
34	1990	225 non-null	int64
35	1991	225 non-null	int64
36	1992	225 non-null	int64
37	1993	225 non-null	int64
38	1994	225 non-null	int64
39	1995	225 non-null	int64
40	1996	225 non-null	int64
41	1997	225 non-null	int64
42	1998	225 non-null	int64
43	1999	225 non-null	int64
44	2000	225 non-null	int64
45	2001	225 non-null	int64
46	2002	225 non-null	int64
47	2003	225 non-null	int64
48	2004	225 non-null	int64
49	2005	225 non-null	int64
50	2006	225 non-null	int64
51	2007	225 non-null	int64
52	2008	225 non-null	int64
53	2009	225 non-null	int64
54	2010	225 non-null	int64
55	2011	225 non-null	int64

```

56 2012      225 non-null    int64
57 2013      225 non-null    int64
58 2014      225 non-null    int64
59 2015      225 non-null    int64
60 2016      225 non-null    int64
61 2017      225 non-null    int64
62 2018      225 non-null    int64
63 2019      225 non-null    int64
64 2020      225 non-null    int64
65 2021      225 non-null    int64
66 2022      225 non-null    int64
dtypes: float64(30), int64(33), object(4)
memory usage: 117.9+ KB

```

In [47]: `df.isnull().sum()`

```

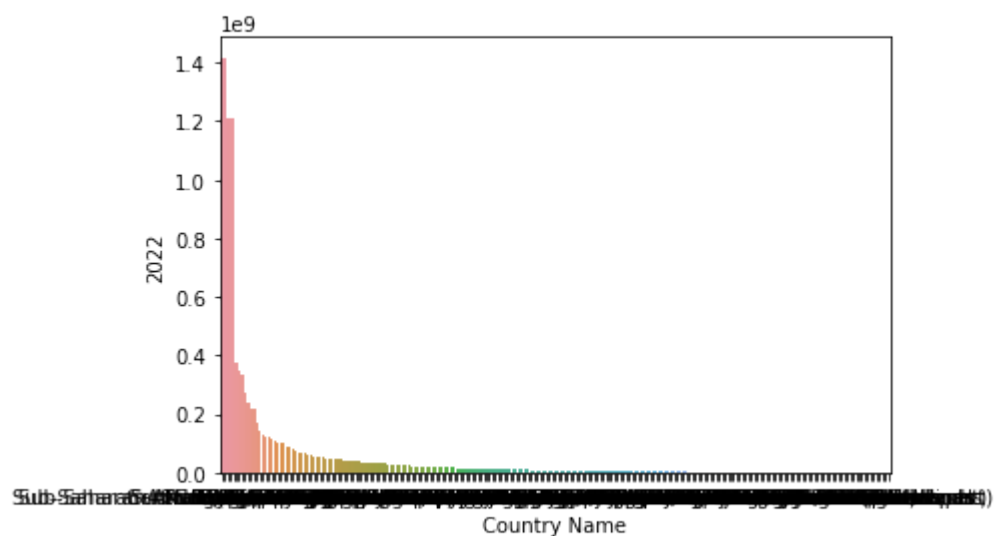
Out[47]: Country Name      0
Country Code      0
Indicator Name      0
Indicator Code      0
1960              1
..
2018              0
2019              0
2020              0
2021              0
2022              0
Length: 67, dtype: int64

```

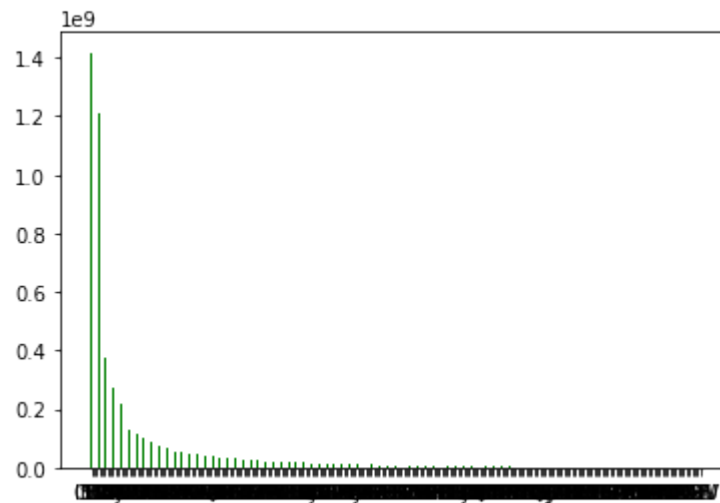
Categorization and Visualisation Using Seaborn and Matplotlib

In [48]: `sns.barplot(x='Country Name', y='2022', data=df)`

Out[48]: `<matplotlib.axes._subplots.AxesSubplot at 0x2153f301508>`



```
In [49]: ▶ df = pd.DataFrame(df)
Continent = df['Country Code']
pop = df['2022']
pyp.bar(Continent, pop, width=0.25, color="green")
pyp.show()
```



From the above graph we see the diagrammatic representation of population for the year of 2022. We see that India has the highest population growth for the year 2022.

```
In [ ]: ▶
```