Data Collection

In [2]:
```
import pandas as pd
```

```
# Collected data from various sources such as kaggle to obtain a
compresenhive dataset for analysis.
```

In [3]:
```
df = pd.read_csv("D:\Data Analyst\DataSet\world_population_data.csv")
df
```

Out[3]:

	rank	cca3	country	continent	2023 population	2022 population	2020 population	2015 population	popul
0	1	IND	India	Asia	1428627663	1417173173	1396387127	1322866505	124061
1	2	CHN	China	Asia	1425671352	1425887337	1424929781	1393715448	134819
2	3	USA	United States	North America	339996563	338289857	335942003	324607776	31118
3	4	IDN	Indonesia	Asia	277534122	275501339	271857970	259091970	24401
4	5	PAK	Pakistan	Asia	240485658	235824862	227196741	210969298	19445
...	...	...	...	...	...	...	...	...	
229	230	MSR	Montserrat	North America	4386	4390	4500	5059	
230	231	FLK	Falkland Islands	South America	3791	3780	3747	3408	
231	232	NIU	Niue	Oceania	1935	1934	1942	1847	
232	233	TKL	Tokelau	Oceania	1893	1871	1827	1454	
233	234	VAT	Vatican City	Europe	518	510	520	564	

234 rows × 17 columns

Data Cleaning

```
# Renamed column's name to understand overall structure of the DataFrame.
```

In [8]:
```
df = df.rename(columns={'rank':'Rank','cca3':'Abbrev.','country':'Country','co
                        'growth rate':'Growth rate','world percentage':'World
df.columns
```

Out[8]:
```
Index(['Rank', 'Abbrev.', 'Country', 'Continent', '2023 population',
       '2022 population', '2020 population', '2015 population',
       '2010 population', '2000 population', '1990 population',
       '1980 population', '1970 population', 'Area (km²)', 'Density (km²)',
       'Growth rate', 'World percentage'],
      dtype='object')
```

```
# Meticuously cleaned the data to identify missing values from the DataFrame.
```

In [17]:
```
df = df.dropna(axis=0)
df
```

In [6]:
```
df1 = pd.read_csv("D:\Data Analyst\DataSet\world_population_data1.csv")
df1
```

Out[6]:

	Unnamed: 0	Rank	Abbrev.	Country	Continent	2023 population	2022 population	2020 population	pop
0	0	1	IND	India	Asia	1428627663	1417173173	1396387127	132
1	1	2	CHN	China	Asia	1425671352	1425887337	1424929781	139
2	2	3	USA	United States	North America	339996563	338289857	335942003	32
3	3	4	IDN	Indonesia	Asia	277534122	275501339	271857970	25
4	4	5	PAK	Pakistan	Asia	240485658	235824862	227196741	21
...	...	...	...	...	...	...	...	...	
229	229	230	MSR	Montserrat	North America	4386	4390	4500	
230	230	231	FLK	Falkland Islands	South America	3791	3780	3747	
231	231	232	NIU	Niue	Oceania	1935	1934	1942	
232	232	233	TKL	Tokelau	Oceania	1893	1871	1827	
233	233	234	VAT	Vatican City	Europe	518	510	520	

234 rows × 18 columns

```
# Removed unwanted columns from the DataFrame.
```

In [7]:
```
df1 = df1.drop(columns=['Unnamed: 0'], axis=1)
df1
```

Out[7]:

	Rank	Abbrev.	Country	Continent	2023 population	2022 population	2020 population	2015 population	p
0	1	IND	India	Asia	1428627663	1417173173	1396387127	1322866505	12
1	2	CHN	China	Asia	1425671352	1425887337	1424929781	1393715448	13
2	3	USA	United States	North America	339996563	338289857	335942003	324607776	3
3	4	IDN	Indonesia	Asia	277534122	275501339	271857970	259091970	2
4	5	PAK	Pakistan	Asia	240485658	235824862	227196741	210969298	1
...	...	...	...	...	...	...	...	...	
229	230	MSR	Montserrat	North America	4386	4390	4500	5059	
230	231	FLK	Falkland Islands	South America	3791	3780	3747	3408	
231	232	NIU	Niue	Oceania	1935	1934	1942	1847	
232	233	TKL	Tokelau	Oceania	1893	1871	1827	1454	
233	234	VAT	Vatican City	Europe	518	510	520	564	

234 rows × 17 columns

```
# Exported cleaned DataFrame to the DataFrame.
```

In [8]:
```
df1.to_csv('D:\Data Analyst\Projects\Analysis of World Population Data\world_p
```