# project on country GDP analysis

In [1]:
```
import pandas as pd
```

In [2]:
```
df = pd.read_csv(r'C:\Users\Gopi Reddy\Downloads\data.csv')
```

In [3]:
```
df
```

Out[3]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |
| **...** | ... | ... | ... | ... | ... |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

195 rows × 5 columns

In [4]:
```
len(df)
```

Out[4]:
```
195
```

In [5]:
```
df.shape
```

Out[5]:
```
(195, 5)
```

In [6]:
```
df.columns
```

Out[6]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [7]:
```
type(df)
```

Out[7]: pandas.core.frame.DataFrame

In [8]:
```
df
```

Out[8]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |
| **...** | ... | ... | ... | ... | ... |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

195 rows × 5 columns

In [11]:
```
pd.__version__
```

Out[11]: '2.2.2'

In [12]:
```
df
```

Out[12]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |
| **...** | ... | ... | ... | ... | ... |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

195 rows × 5 columns

In [13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 195 entries, 0 to 194
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   CountryName    195 non-null    object 
 1   CountryCode    195 non-null    object 
 2   BirthRate      195 non-null    float64
 3   InternetUsers  195 non-null    float64
 4   IncomeGroup    195 non-null    object 
dtypes: float64(2), object(3)
memory usage: 7.7+ KB
```

In [14]:
```
df.columns
```

Out[14]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [15]:
```
len(df.columns)
```

Out[15]:
```
5
```

In [16]:
```
df.head()
```

Out[16]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |

In [17]:
```
df.tail()
```

Out[17]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

In [18]:
```
df
```

Out[18]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |
| **...** | ... | ... | ... | ... | ... |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |

195 rows × 5 columns

In [19]:
```
df[::-1]
```

Out[19]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **194** | Zimbabwe | ZWE | 35.715 | 18.5 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4 | Lower middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2 | Low income |
| **191** | South Africa | ZAF | 20.850 | 46.5 | Upper middle income |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0 | Lower middle income |
| **...** | ... | ... | ... | ... | ... |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |

195 rows × 5 columns

In [20]:
```
df[:5]
```

Out[20]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |
| **1** | Afghanistan | AFG | 35.253 | 5.9 | Low income |
| **2** | Angola | AGO | 45.985 | 19.1 | Upper middle income |
| **3** | Albania | ALB | 12.877 | 57.2 | Upper middle income |
| **4** | United Arab Emirates | ARE | 11.044 | 88.0 | High income |

In [21]:
```
df[6:]
```

Out[21]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **6** | Armenia | ARM | 13.308 | 41.9000 | Lower middle income |
| **7** | Antigua and Barbuda | ATG | 16.447 | 63.4000 | High income |
| **8** | Australia | AUS | 13.200 | 83.0000 | High income |
| **9** | Austria | AUT | 9.400 | 80.6188 | High income |
| **10** | Azerbaijan | AZE | 18.300 | 58.7000 | Upper middle income |
| **...** | ... | ... | ... | ... | ... |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.0000 | Lower middle income |
| **191** | South Africa | ZAF | 20.850 | 46.5000 | Upper middle income |
| **192** | Congo, Dem. Rep. | COD | 42.394 | 2.2000 | Low income |
| **193** | Zambia | ZMB | 40.471 | 15.4000 | Lower middle income |
| **194** | Zimbabwe | ZWE | 35.715 | 18.5000 | Low income |

189 rows × 5 columns

In [22]:
```
df[0:200:10]
```

Out[22]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.900000 | High income |
| **10** | Azerbaijan | AZE | 18.300 | 58.700000 | Upper middle income |
| **20** | Belarus | BLR | 12.500 | 54.170000 | Upper middle income |
| **30** | Canada | CAN | 10.900 | 85.800000 | High income |
| **40** | Costa Rica | CRI | 15.022 | 45.960000 | Upper middle income |
| **50** | Ecuador | ECU | 21.070 | 40.353684 | Upper middle income |
| **60** | Gabon | GAB | 30.555 | 9.200000 | Upper middle income |
| **70** | Greenland | GRL | 14.500 | 65.800000 | High income |
| **80** | India | IND | 20.291 | 15.100000 | Lower middle income |
| **90** | Kazakhstan | KAZ | 22.730 | 54.000000 | Upper middle income |
| **100** | Libya | LBY | 21.425 | 16.500000 | Upper middle income |
| **110** | Moldova | MDA | 12.141 | 45.000000 | Lower middle income |
| **120** | Mozambique | MOZ | 39.705 | 5.400000 | Low income |
| **130** | Netherlands | NLD | 10.200 | 93.956400 | High income |
| **140** | Poland | POL | 9.600 | 62.849200 | High income |
| **150** | Sudan | SDN | 33.477 | 22.700000 | Lower middle income |
| **160** | Suriname | SUR | 18.455 | 37.400000 | Upper middle income |
| **170** | Tajikistan | TJK | 30.792 | 16.000000 | Lower middle income |
| **180** | Uruguay | URY | 14.374 | 57.690000 | High income |
| **190** | Yemen, Rep. | YEM | 32.947 | 20.000000 | Lower middle income |

In [26]:
```
df.describe() # it gives a descriptive statistics (print numerical data info onl
```

Out[26]:

| | BirthRate | InternetUsers |
|---|---|---|
| **count** | 195.000000 | 195.000000 |
| **mean** | 21.469928 | 42.076471 |
| **std** | 10.605467 | 29.030788 |
| **min** | 7.900000 | 0.900000 |
| **25%** | 12.120500 | 14.520000 |
| **50%** | 19.680000 | 41.000000 |
| **75%** | 29.759500 | 66.225000 |
| **max** | 49.661000 | 96.546800 |

In [27]:
```
df.describe().transpose() # tranpose convert column into rows
```

Out[27]:

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **BirthRate** | 195.0 | 21.469928 | 10.605467 | 7.9 | 12.1205 | 19.68 | 29.7595 | 49.6610 |
| **InternetUsers** | 195.0 | 42.076471 | 29.030788 | 0.9 | 14.5200 | 41.00 | 66.2250 | 96.5468 |

In [29]:
```
df.columns
```

Out[29]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [40]:
```
df.columns = ['a','b','c','d','e'] # you can change the attributes
```

In [36]:
```
df.head(1)
```

Out[36]:

| | a | b | c | d | e |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |

In [42]:
```
df.columns
```

Out[42]:
```
Index(['a', 'b', 'c', 'd', 'e'], dtype='object')
```

In [43]:
```
df.columns = ['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers', 'Incom
```

In [44]:
```
df.head(1)
```

Out[44]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | Aruba | ABW | 10.244 | 78.9 | High income |

In [45]:
```
df.columns
```

Out[45]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [46]:
```
df[['CountryName', 'CountryCode']]
```

Out[46]:

| | CountryName | CountryCode |
|---|---|---|
| **0** | Aruba | ABW |
| **1** | Afghanistan | AFG |
| **2** | Angola | AGO |
| **3** | Albania | ALB |
| **4** | United Arab Emirates | ARE |
| **...** | ... | ... |
| **190** | Yemen, Rep. | YEM |
| **191** | South Africa | ZAF |
| **192** | Congo, Dem. Rep. | COD |
| **193** | Zambia | ZMB |
| **194** | Zimbabwe | ZWE |

195 rows × 2 columns

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

Out[47]:

| | CountryName | CountryCode | BirthRate | InternetUsers | IncomeGroup |
|---|---|---|---|---|---|
| **0** | False | False | False | False | False |
| **1** | False | False | False | False | False |
| **2** | False | False | False | False | False |
| **3** | False | False | False | False | False |
| **4** | False | False | False | False | False |
| **...** | ... | ... | ... | ... | ... |
| **190** | False | False | False | False | False |
| **191** | False | False | False | False | False |
| **192** | False | False | False | False | False |
| **193** | False | False | False | False | False |
| **194** | False | False | False | False | False |

195 rows × 5 columns

In [48]:
```
df.isnull().sum()
```

Out[48]:
```
CountryName      0
CountryCode      0
BirthRate        0
InternetUsers    0
IncomeGroup      0
dtype: int64
```

In [ ]:
```
# i want to split the categorical data to numerical data
```

In [49]:
```
df.dtypes
```

Out[49]:
```
CountryName       object
CountryCode       object
BirthRate        float64
InternetUsers    float64
IncomeGroup       object
dtype: object
```

In [50]:
```
df.columns
```

Out[50]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [51]:
```
df_categorical = df[['CountryName', 'CountryCode','IncomeGroup']]
df_categorical.head()
```

Out[51]:

| | CountryName | CountryCode | IncomeGroup |
|---|---|---|---|
| **0** | Aruba | ABW | High income |
| **1** | Afghanistan | AFG | Low income |
| **2** | Angola | AGO | Upper middle income |
| **3** | Albania | ALB | Upper middle income |
| **4** | United Arab Emirates | ARE | High income |

In [52]:
```
df.describe()
```

Out[52]:

| | BirthRate | InternetUsers |
|---|---|---|
| **count** | 195.000000 | 195.000000 |
| **mean** | 21.469928 | 42.076471 |
| **std** | 10.605467 | 29.030788 |
| **min** | 7.900000 | 0.900000 |
| **25%** | 12.120500 | 14.520000 |
| **50%** | 19.680000 | 41.000000 |
| **75%** | 29.759500 | 66.225000 |
| **max** | 49.661000 | 96.546800 |

In [53]:
```
df_categorical.describe()
```

Out[53]:

| | CountryName | CountryCode | IncomeGroup |
|---|---|---|---|
| **count** | 195 | 195 | 195 |
| **unique** | 195 | 195 | 4 |
| **top** | Aruba | ABW | High income |
| **freq** | 1 | 1 | 67 |

In [56]:
```
df.columns
```

Out[56]:
```
Index(['CountryName', 'CountryCode', 'BirthRate', 'InternetUsers',
       'IncomeGroup'],
      dtype='object')
```

In [59]:
```
df_num = df[['BirthRate', 'InternetUsers']]
df_num.head()
```

Out[59]:

| | BirthRate | InternetUsers |
|---|---|---|
| **0** | 10.244 | 78.9 |
| **1** | 35.253 | 5.9 |
| **2** | 45.985 | 19.1 |
| **3** | 12.877 | 57.2 |
| **4** | 11.044 | 88.0 |

In [60]:
```
df_num.describe()
```

Out[60]:

| | BirthRate | InternetUsers |
|---|---|---|
| **count** | 195.000000 | 195.000000 |
| **mean** | 21.469928 | 42.076471 |
| **std** | 10.605467 | 29.030788 |
| **min** | 7.900000 | 0.900000 |
| **25%** | 12.120500 | 14.520000 |
| **50%** | 19.680000 | 41.000000 |
| **75%** | 29.759500 | 66.225000 |
| **max** | 49.661000 | 96.546800 |

In [ ]:
```

```