DSC540-T301 Project: Milestone 5

Merging the Data and Storing in a Database/Visualizing Data

Step 1. Import libraries and Load datasets

In [1]:
```
# load necessary libraries
# import pandas, numpy, and sqlite 3
import pandas as pd
import numpy as np
import sqlite3 as sq
```

In [2]:
```
# createa sqlite3 database
from pathlib import Path
Path('WHR.db').touch()
```

In [3]:
```
# connect to WHR.db in sqlite3
conn = sq.connect('WHR.db')
c = conn.cursor()
```

In [4]:
```
# load csv files in pandas
df_whr = pd.concat(map(pd.read_csv, ['WHR-Wikipedia.csv','WHRData2022.csv','WHR2017.csv']))
df_whr.columns
```

Out[4]:
```
Index(['Unnamed: 0', 'Overall_Rank', 'Country/Region', 'Score', 'GDP/capita',
       'Social_Support', 'Healthy_life_exp', 'Freedom', 'Generosity',
       'Perception_of_corrupt', 'Country', 'Happiness.Rank', 'Happiness.Score',
       'Whisker.high', 'Whisker.low', 'Dystopia.Residual',
       'Economy.GDP.per.Capita.', 'Social.support', 'Health.Life.Expectancy.',
       'Perceptions.of.corruption', 'Economy..GDP.per.Capita.', 'Family',
       'Health..Life.Expectancy.', 'Trust..Government.Corruption.'],
      dtype='object')
```

In [5]:
```
# write data to sqlite3 database
df_whr.to_sql('WHR', conn, if_exists='replace', index=False)
```

Out[5]:
```
458
```

In [6]:
```
# run select sql query and left join on country and happiness rank
c.execute('''SELECT *
             FROM WHR
             LIMIT 20;''')
df_whr.columns = [x[0] for x in c.description]
df_whr
```

Out[6]:

	Unnamed: 0	Overall_Rank	Country/Region	Score	GDP/capita	Social_Support	Healthy_life_exp	Freedom	Generosity	Perception_of_corrupt	...	Whisker.low	Dystopia.Residual	Economy.GDP.per.Capita.	Social.support	Hea
0	2.0	1.0	Finland	7.769	1.340	1.587	0.986	0.596000	0.153000	0.393	...	NaN	NaN	NaN	NaN	
1	3.0	2.0	Denmark	7.600	1.383	1.573	0.996	0.592000	0.252000	0.410	...	NaN	NaN	NaN	NaN	
2	4.0	3.0	Norway	7.554	1.488	1.582	1.028	0.603000	0.271000	0.341	...	NaN	NaN	NaN	NaN	
3	5.0	4.0	Iceland	7.494	1.380	1.624	1.026	0.591000	0.354000	0.118	...	NaN	NaN	NaN	NaN	
4	6.0	5.0	Netherlands	7.488	1.396	1.522	0.999	0.557000	0.322000	0.298	...	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
150	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.581844	0.252756	NaN	...	3.398970	0.540061	NaN	NaN	
151	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.081539	0.493664	NaN	...	3.260331	1.061574	NaN	NaN	
152	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.390018	0.354256	NaN	...	3.236570	0.621130	NaN	NaN	
153	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.059901	0.204435	NaN	...	2.735310	1.683024	NaN	NaN	
154	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.270842	0.280876	NaN	...	2.521116	2.066005	NaN	NaN	

458 rows × 24 columns

In [7]:
```
# retrieve all records
records = c.fetchall()
```

In [8]:
```
#display records
for column in records:
    print("Country: ", column[0])
    print("Happiness.Rank: ", column[1])
    print("Happiness.Score: ", column[2])
    print("Whisker.high: ", column[3])
    print("Whisker.low: ", column[4])
```

```
Country:  2.0
Happiness.Rank:  1.0
Happiness.Score:  Finland
Whisker.high:  7.769
Whisker.low:  1.34
Country:  3.0
Happiness.Rank:  2.0
Happiness.Score:  Denmark
Whisker.high:  7.6
Whisker.low:  1.383
Country:  4.0
Happiness.Rank:  3.0
Happiness.Score:  Norway
Whisker.high:  7.554
Whisker.low:  1.488
Country:  5.0
Happiness.Rank:  4.0
Happiness.Score:  Iceland
Whisker.high:  7.494
Whisker.low:  1.38
Country:  6.0
Happiness.Rank:  5.0
Happiness.Score:  Netherlands
Whisker.high:  7.488
Whisker.low:  1.396
Country:  7.0
Happiness.Rank:  6.0
Happiness.Score:  Switzerland
Whisker.high:  7.48
Whisker.low:  1.452
Country:  8.0
Happiness.Rank:  7.0
Happiness.Score:  Sweden
Whisker.high:  7.343
Whisker.low:  1.387
Country:  9.0
Happiness.Rank:  8.0
Happiness.Score:  New Zealand
Whisker.high:  7.307
Whisker.low:  1.303
Country:  10.0
Happiness.Rank:  9.0
Happiness.Score:  Canada
Whisker.high:  7.278
Whisker.low:  1.365
Country:  11.0
Happiness.Rank:  10.0
Happiness.Score:  Austria
Whisker.high:  7.246
Whisker.low:  1.376
Country:  12.0
Happiness.Rank:  11.0
Happiness.Score:  Australia
Whisker.high:  7.228
Whisker.low:  1.372
Country:  13.0
Happiness.Rank:  12.0
Happiness.Score:  Costa Rica
Whisker.high:  7.167
Whisker.low:  1.034
```

```
Country:  14.0
Happiness.Rank:  13.0
Happiness.Score:  Israel
Whisker.high:  7.139
Whisker.low:  1.276
Country:  15.0
Happiness.Rank:  14.0
Happiness.Score:  Luxembourg
Whisker.high:  7.09
Whisker.low:  1.609
Country:  16.0
Happiness.Rank:  15.0
Happiness.Score:  United Kingdom
Whisker.high:  7.054
Whisker.low:  1.333
Country:  17.0
Happiness.Rank:  16.0
Happiness.Score:  Ireland
Whisker.high:  7.021
Whisker.low:  1.499
Country:  18.0
Happiness.Rank:  17.0
Happiness.Score:  Germany
Whisker.high:  6.985
Whisker.low:  1.373
Country:  19.0
Happiness.Rank:  18.0
Happiness.Score:  Belgium
Whisker.high:  6.923
Whisker.low:  1.356
Country:  20.0
Happiness.Rank:  19.0
Happiness.Score:  United States of America
Whisker.high:  6.892
Whisker.low:  1.433
Country:  21.0
Happiness.Rank:  20.0
Happiness.Score:  Czech Republic
Whisker.high:  6.852
Whisker.low:  1.269
```

In [9]:
```
# write new csv
df_whr.to_csv('NewWHR.csv')
```