In [1]:
```python
import pandas as pd #importing pandaslibrary
import matplotlib.pyplot as plt #importing matplotlib library
import seaborn as sns #importing Seaborn
import numpy as np #importing Seaborn numpy
pd.set_option('display.max_columns', None)
```

In [4]:
```python
health = pd.read_csv("Health.csv", na_values = "-") #import health and assign
  variable
health.head() #checking if data loaded correctly
```

Out[4]:

	Continent	Country	City	Rank	Sunshine hours(City)	Cost of a bottle of water(City)	Obesity levels(Country)	ex
0	Europe	Netherlands	Amsterdam	1	1858.0	£1.92	20.40%	81
1	Oceania	Australia	Sydney	2	2636.0	£1.48	29.00%	82
2	Europe	Austria	Vienna	3	1884.0	£1.94	20.10%	81
3	Europe	Sweden	Stockholm	4	1821.0	£1.72	20.60%	81
4	Europe	Denmark	Copenhagen	5	1630.0	£2.19	19.70%	79

In [7]:
```python
health = health.dropna() # Dropping all N/A record
health = health.reset_index(drop = True) # Resetting index
health.head(20) # Checking Dataframe
```

Out[7]:

	Continent	Country	City	Rank	Sunshine hours(City)	Cost of a bottle of water(City)	Obesity levels(Country)	
0	Europe	Netherlands	Amsterdam	1	1858.0	£1.92	20.40%	8
1	Oceania	Australia	Sydney	2	2636.0	£1.48	29.00%	8
2	Europe	Austria	Vienna	3	1884.0	£1.94	20.10%	8
3	Europe	Sweden	Stockholm	4	1821.0	£1.72	20.60%	8
4	Europe	Denmark	Copenhagen	5	1630.0	£2.19	19.70%	7
5	Europe	Finland	Helsinki	6	1662.0	£1.60	22.20%	8
6	Europe	Germany	Berlin	8	1626.0	£1.55	22.30%	8
7	Europe	Spain	Barcelona	9	2591.0	£1.19	23.80%	8
8	North America	Canada	Vancouver	10	1938.0	£1.08	29.40%	8
9	Oceania	Australia	Melbourne	11	2363.0	£1.57	29.00%	8
10	North America	Canada	Toronto	15	2066.0	£1.09	29.40%	8
11	Europe	Spain	Madrid	16	2769.0	£1.30	23.80%	8
12	Asia	South Korea	Seoul	18	2066.0	£0.59	4.70%	8
13	Europe	Germany	Frankfurt	19	1662.0	£1.95	22.30%	8
14	Asia	Israel	Tel Aviv	21	3311.0	£1.63	26.10%	8
15	Asia	Turkey	Istanbul	22	2218.0	£0.15	32.10%	7
16	North America	United States	Los Angeles	25	3254.0	£1.52	36.20%	7
17	North America	United States	Boston	27	2634.0	£1.39	36.20%	7
18	Europe	Republic of Ireland	Dublin	28	1453.0	£1.40	25.30%	8
19	Asia	Japan	Tokyo	29	1877.0	£0.76	4.30%	8

In [11]:
```python
health.rename(columns = {"Outdoor activities(City)": "# of Outdoor Activities
(City)","Sunshine hours(City)":"Annual Sunshine Hours(City)" ,"Cost of a bottl
e of water(City)":"Water Bottle Cost(City)","Number of take out places(City)":
"# of Take Out Places","Cost of a monthly gym membership(City)":"Monthly Gym M
embership Cost(City)"}, inplace = True) # Renaming columns
health.head() #Checking Dataframe
```

Out[11]:

	Continent	Country	City	Rank	Annual Sunshine Hours(City)	Water Bottle Cost(City)	Obesity levels(Country)	ex
0	Europe	Netherlands	Amsterdam	1	1858.0	£1.92	20.40%	81
1	Oceania	Australia	Sydney	2	2636.0	£1.48	29.00%	82
2	Europe	Austria	Vienna	3	1884.0	£1.94	20.10%	81
3	Europe	Sweden	Stockholm	4	1821.0	£1.72	20.60%	81
4	Europe	Denmark	Copenhagen	5	1630.0	£2.19	19.70%	79

In [15]:
```python
health["Water Bottle Cost(City)"] = health["Water Bottle Cost(City)"].str.repl
ace("£","") # Removing euro sign from column
health["Obesity levels(Country)"] = health["Obesity levels(Country)"].str.repl
ace("%","") # Removing percentage sign from column
health["Monthly Gym Membership Cost(City)"] = health["Monthly Gym Membership C
ost(City)"].str.replace("£","") # Removing euro sign from column
health.head() # Checking Dataframe
```

Out[15]:

	Continent	Country	City	Rank	Annual Sunshine Hours(City)	Water Bottle Cost(City)	Obesity levels(Country)	ex
0	Europe	Netherlands	Amsterdam	1	1858.0	1.92	20.40	81
1	Oceania	Australia	Sydney	2	2636.0	1.48	29.00	82
2	Europe	Austria	Vienna	3	1884.0	1.94	20.10	81
3	Europe	Sweden	Stockholm	4	1821.0	1.72	20.60	81
4	Europe	Denmark	Copenhagen	5	1630.0	2.19	19.70	79

In [9]:
```python
dup_filter = health.duplicated() #Checking for dupes in the data
health[dup_filter] #No duplicates found, moving on to removing % signs and euro sign.
```

Out[9]:

	Continent	Country	City	Rank	Sunshine hours(City)	Cost of a bottle of water(City)	Obesity levels(Country)	expectancy(ye (Cour

In [19]:
```python
health["Monthly Sunshine Hours(City)"] = round(health["Annual Sunshine Hours(City)"] / 12,2) #Adding monhtly column
health["Monthly avg. hours worked"] =  round(health["Annual avg. hours worked"] / 12,2) #Adding monhtly column
health.head() # Checking DataFrame
```

Out[19]:

	Continent	Country	City	Rank	Annual Sunshine Hours(City)	Water Bottle Cost(City)	Obesity levels(Country)	ex
0	Europe	Netherlands	Amsterdam	1	1858.0	1.92	20.40	81
1	Oceania	Australia	Sydney	2	2636.0	1.48	29.00	82
2	Europe	Austria	Vienna	3	1884.0	1.94	20.10	81
3	Europe	Sweden	Stockholm	4	1821.0	1.72	20.60	81
4	Europe	Denmark	Copenhagen	5	1630.0	2.19	19.70	79

In [20]:
```python
health.dtypes
```

Out[20]:
```
Continent                          object
Country                            object
City                               object
Rank                                int64
Annual Sunshine Hours(City)       float64
Water Bottle Cost(City)            object
Obesity levels(Country)            object
Life expectancy(years) (Country)  float64
Pollution(Index score) (City)     float64
Annual avg. hours worked          float64
Happiness levels(Country)         float64
# of Outdoor Activities(City)       int64
# of Take Out Places                int64
Monthly Gym Membership Cost(City)  object
Monthly Sunshine Hours(City)      float64
Monthly avg. hours worked         float64
dtype: object
```

In [28]:
```python
health["Monthly Gym Membership Cost(City)"] = health["Monthly Gym Membership Cost(City)"].astype(float) #changing  column data type to float
health["Happiness levels(Country)"] = health["Happiness levels(Country)"].astype(str) #changing  column data type to string
health["Water Bottle Cost(City)"] = health["Water Bottle Cost(City)"].astype(float) #changing  column data type to float
health["Obesity levels(Country)"] = health["Obesity levels(Country)"].astype(float) #changing  column data type to float
health.dtypes
```

Out[28]:
```
Continent                            object
Country                              object
City                                 object
Rank                                  int64
Annual Sunshine Hours(City)         float64
Water Bottle Cost(City)             float64
Obesity levels(Country)             float64
Life expectancy(years) (Country)    float64
Pollution(Index score) (City)       float64
Annual avg. hours worked            float64
Happiness levels(Country)            object
# of Outdoor Activities(City)         int64
# of Take Out Places                  int64
Monthly Gym Membership Cost(City)   float64
Monthly Sunshine Hours(City)        float64
Monthly avg. hours worked           float64
dtype: object
```

In [29]: `health #Final cleansed data`

Out[29]:

	Continent	Country	City	Rank	Annual Sunshine Hours(City)	Water Bottle Cost(City)	Obesity levels(Country)	e
0	Europe	Netherlands	Amsterdam	1	1858.0	1.92	20.4	8
1	Oceania	Australia	Sydney	2	2636.0	1.48	29.0	8
2	Europe	Austria	Vienna	3	1884.0	1.94	20.1	8
3	Europe	Sweden	Stockholm	4	1821.0	1.72	20.6	8
4	Europe	Denmark	Copenhagen	5	1630.0	2.19	19.7	7
5	Europe	Finland	Helsinki	6	1662.0	1.60	22.2	8
6	Europe	Germany	Berlin	8	1626.0	1.55	22.3	8
7	Europe	Spain	Barcelona	9	2591.0	1.19	23.8	8
8	North America	Canada	Vancouver	10	1938.0	1.08	29.4	8
9	Oceania	Australia	Melbourne	11	2363.0	1.57	29.0	8
10	North America	Canada	Toronto	15	2066.0	1.09	29.4	8
11	Europe	Spain	Madrid	16	2769.0	1.30	23.8	8
12	Asia	South Korea	Seoul	18	2066.0	0.59	4.7	8
13	Europe	Germany	Frankfurt	19	1662.0	1.95	22.3	8
14	Asia	Israel	Tel Aviv	21	3311.0	1.63	26.1	8
15	Asia	Turkey	Istanbul	22	2218.0	0.15	32.1	7
16	North America	United States	Los Angeles	25	3254.0	1.52	36.2	7
17	North America	United States	Boston	27	2634.0	1.39	36.2	7
18	Europe	Republic of Ireland	Dublin	28	1453.0	1.40	25.3	8
19	Asia	Japan	Tokyo	29	1877.0	0.76	4.3	8
20	North America	United States	Chicago	30	2508.0	1.20	36.2	7
21	Europe	Belgium	Brussels	33	1546.0	2.11	22.1	8
22	North America	United States	San Francisco	34	3062.0	1.60	36.2	7
23	Europe	France	Paris	35	1662.0	1.95	21.6	8

	Continent	Country	City	Rank	Annual Sunshine Hours(City)	Water Bottle Cost(City)	Obesity levels(Country)	e
24	Europe	Switzerland	Zurich	37	1566.0	3.20	19.5	8
25	Europe	United Kingdom	London	38	1633.0	1.16	27.8	8
26	Europe	Italy	Milan	40	1915.0	1.15	19.9	8
27	North America	United States	Washington, D.C.	41	2528.0	1.45	36.2	7
28	North America	United States	New York	42	2535.0	1.32	36.2	7
29	Asia	Russia	Moscow	43	1901.0	0.41	23.1	6
30	North America	Mexico	Mexico City	44	2555.0	0.45	28.9	7