

```
In [31]: from selenium import webdriver
website="https://www.wcrf.org/cancer-trends/global-cancer-data-by-country/"
chrome_options=webdriver.ChromeOptions()
directory="E:\\ANACONDA\\scrapy_tutorials\\liver_can\\liver_can"
chrome_options.add_experimental_option('prefs',{'download.default_directory':directory})
driver=webdriver.Chrome(options=chrome_options)
driver.get(website)
```

```
In [34]: import selenium
from selenium.webdriver.common.by import By
table_button=driver.find_element(By.ID,'accordion-4-t1')
table_button.click()
```

```
In [36]: trs=driver.find_elements(By.XPATH,'//div[@id="accordion-4-c1"]/table/tbody/tr')
location=[]
number=[]
ASR=[]
for tr in trs:
    location.append(tr.find_element(By.XPATH,"./td[1]").text)
    number.append(tr.find_element(By.XPATH,"./td[2]").text)
    ASR.append(tr.find_element(By.XPATH,"./td[3]").text)
```

```
In [37]: import pandas as pd
df=pd.DataFrame({'Location':location,'no_cases':number,'ASR/10,000':ASR})
print(df)
```

	Location	no_cases	ASR/10,000
0		Number	ASR/100,000
1	World	9,894,402	100.1
2	Mongolia	4,465	175.9
3	Serbia	27,820	150.6
4	Hungary	32,637	148.1
..
182	Sri Lanka	16,520	56.6
183	Congo, Republic of	1,593	56.6
184	United Arab Emirates	1,877	55.1
185	Nepal	13,523	54.4
186	Saudi Arabia	12,986	50.9

[187 rows x 3 columns]

```
In [39]: df.to_csv('E:\\ANACONDA\\scrapy_tutorials\\liver_can\\can_liver.csv',index=False)
```

```
In [56]: df2=pd.read_csv('E:\\ANACONDA\\scrapy_tutorials\\coffee_ext\\coffee_ext\\coffee_consumpt
df2_drop=df2.reset_index(drop=True)
print(df2_drop)
df_drop=df.drop([0])
print(df_drop)
df.
```

	Rank	Country	Consumption
0	1	Finland	26.45
1	2	Norway	21.82
2	3	Iceland	19.84
3	4	Denmark	19.18
4	5	Netherlands	18.52
5	6	Sweden	18.00
6	7	Switzerland	17.42
7	8	Belgium	15.00
8	9	Luxembourg	14.33
9	10	Canada	14.33
10	11	Bosnia and Herzegovina	13.67
11	12	Austria	13.45
12	13	Italy	13.00
13	14	Brazil	12.79
14	15	Slovenia	12.79
15	16	Germany	12.13
16	17	Greece	11.90
17	18	France	11.90
18	19	Croatia	11.24
19	20	Cyprus	10.80
20	21	Lebanon	10.58
21	22	Estonia	9.92
22	23	Spain	9.92
23	24	Portugal	9.48
24	25	United States	9.26
	Location	no_cases	ASR/10,000
2	Mongolia	4,465	175.9
3	Serbia	27,820	150.6
4	Hungary	32,637	148.1
5	Montenegro	1,752	144.5
6	Slovakia	16,192	140.7
..
182	Sri Lanka	16,520	56.6
183	Congo, Republic of	1,593	56.6
184	United Arab Emirates	1,877	55.1
185	Nepal	13,523	54.4
186	Saudi Arabia	12,986	50.9

[185 rows x 3 columns]

```
In [96]: df_drop=df_drop.rename(columns={'Location':'Country','ASR/10,000':'ASR/100,000'})
#print(df2_drop)
print(df_drop.columns)
df_merged=pd.merge(df_drop,df2_drop,on='Country',how='inner')
print(df_merged)
print(df_merged.dtypes)
```

```
Index(['Country', 'no_cases', 'ASR/100,000'], dtype='object')
Country no_cases ASR/100,000 Rank Consumption
0 Croatia 14,216 132.7 19 11.24
1 Bosnia and Herzegovina 9,142 127.3 11 13.67
2 Estonia 3,980 115.8 22 9.92
3 Denmark 17,270 113.3 4 19.18
4 France 184,636 107.5 18 11.90
5 Greece 33,007 107.0 17 11.90
6 Portugal 29,951 103.4 24 9.48
7 Cyprus 2,433 102.3 20 10.80
8 Germany 250,937 102.0 16 12.13
9 Belgium 29,954 101.0 8 15.00
10 Austria 22,353 95.4 12 13.45
11 Canada 86,053 92.9 10 14.33
12 Norway 11,929 91.4 2 21.82
13 Italy 173,469 90.6 13 13.00
14 Brazil 256,954 90.2 14 12.79
15 Spain 112,335 90.0 23 9.92
16 Luxembourg 1,167 87.6 9 14.33
17 Sweden 24,380 87.0 6 18.00
18 Finland 13,340 84.0 1 26.45
19 Lebanon 6,302 84.0 21 10.58
20 Iceland 641 84.0 3 19.84
21 Switzerland 18,903 82.9 7 17.42
Country object
no_cases object
ASR/100,000 object
Rank int64
Consumption float64
dtype: object
```

```
In [97]: df_merged['no_cases']=df_merged['no_cases'].str.replace(",","").astype(int)
print(df_merged)
```

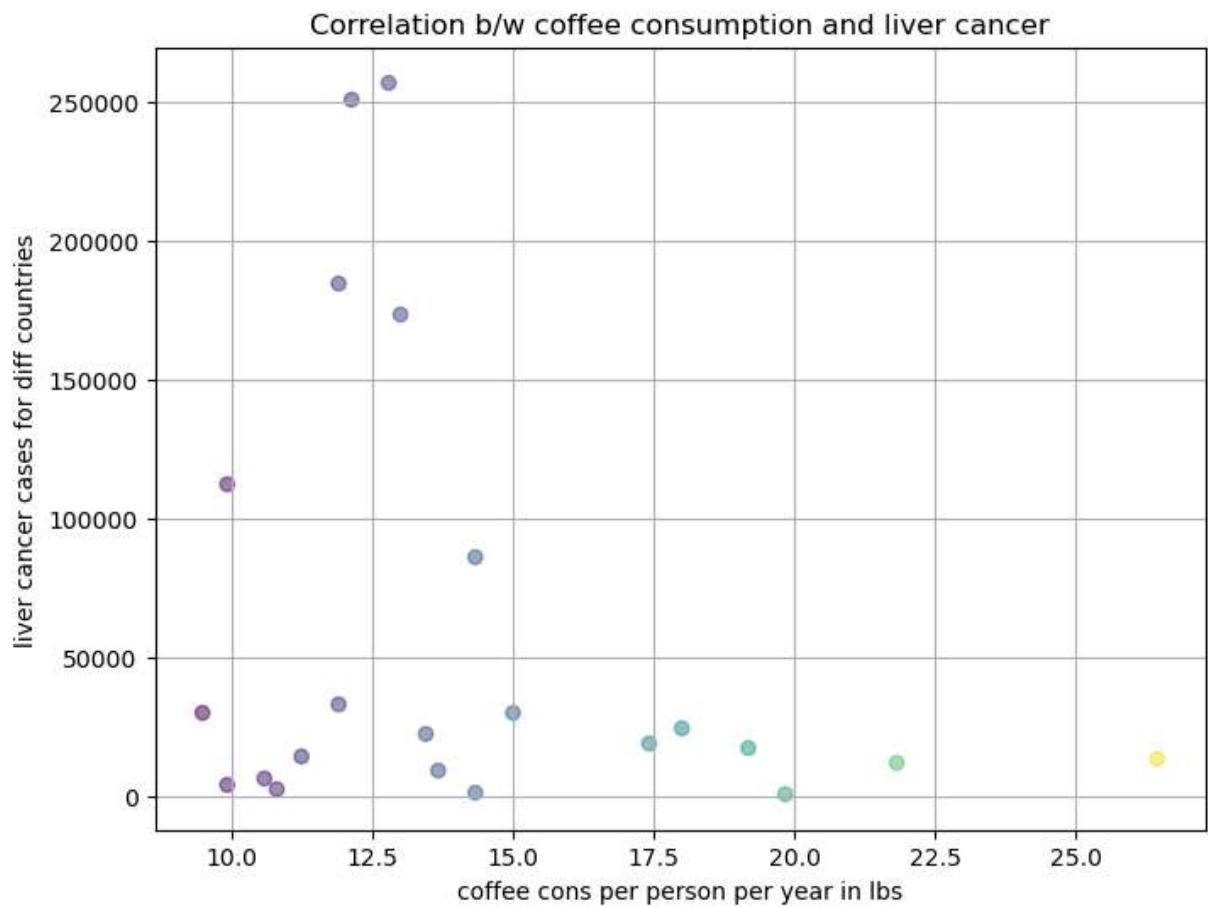
	Country	no_cases	ASR/100,000	Rank	Consumption
0	Croatia	14216	132.7	19	11.24
1	Bosnia and Herzegovina	9142	127.3	11	13.67
2	Estonia	3980	115.8	22	9.92
3	Denmark	17270	113.3	4	19.18
4	France	184636	107.5	18	11.90
5	Greece	33007	107.0	17	11.90
6	Portugal	29951	103.4	24	9.48
7	Cyprus	2433	102.3	20	10.80
8	Germany	250937	102.0	16	12.13
9	Belgium	29954	101.0	8	15.00
10	Austria	22353	95.4	12	13.45
11	Canada	86053	92.9	10	14.33
12	Norway	11929	91.4	2	21.82
13	Italy	173469	90.6	13	13.00
14	Brazil	256954	90.2	14	12.79
15	Spain	112335	90.0	23	9.92
16	Luxembourg	1167	87.6	9	14.33
17	Sweden	24380	87.0	6	18.00
18	Finland	13340	84.0	1	26.45
19	Lebanon	6302	84.0	21	10.58
20	Iceland	641	84.0	3	19.84
21	Switzerland	18903	82.9	7	17.42

```
In [100... import matplotlib.pyplot as plt
import numpy as np

correlation=df_merged['no_cases'].corr(df_merged['Consumption'])
print("correlation b/w coffee consumption and liver cancer is:",correlation)
```

correlation b/w coffee consumption and liver cancer is: -0.2755345033083868

```
In [108... x_points=df_merged['Consumption']
y_points=df_merged['no_cases']
colors=df_merged['Consumption']
plt.figure(figsize=(8,6))
plt.scatter(x_points,y_points,c=colors, cmap='viridis',alpha=0.5)
plt.title("Correlation b/w coffee consumption and liver cancer")
plt.xlabel("coffee cons per person per year in lbs")
plt.ylabel("liver cancer cases for diff countries")
plt.grid(True)
plt.show()
```



In []: