Name : **Omkar Shinde**

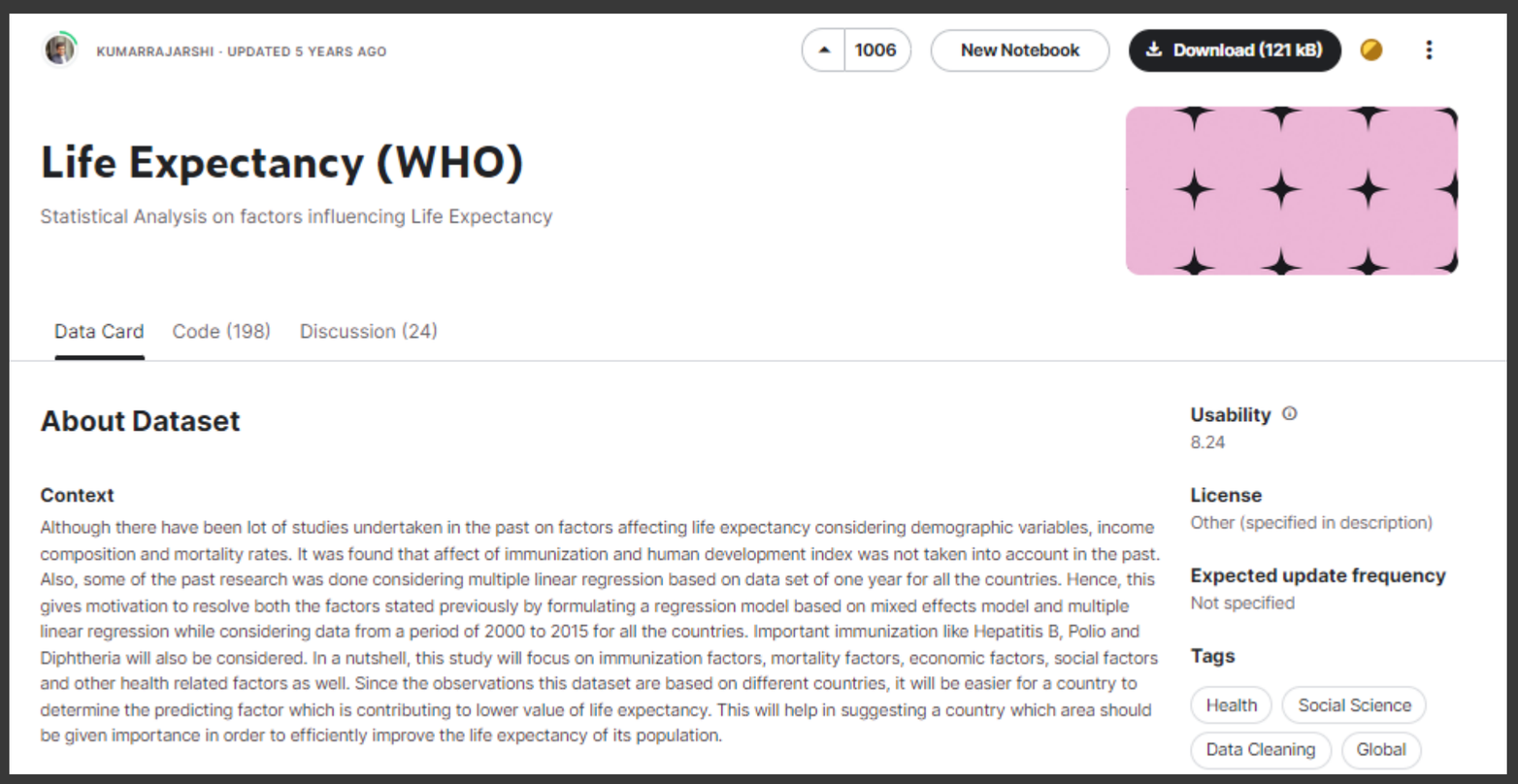
Div : **E**

Roll No : **582**

Assignment **"4"** : **Oprations on Dataset with Pandas**

PRN : **202201040190**

DataSet :



Dataset Source : https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

```
import pandas as pd

df = pd.read_csv("Life Expectancy Data.csv")
```

```
print(df.columns)
```

```
Index(['Country', 'Year', 'Status', 'Life expectancy', 'Adult Mortality',
       'infant deaths', 'Alcohol', 'percentage expenditure', 'Hepatitis B',
       'Measles ', ' BMI ', 'under-five deaths ', 'Polio', 'Total expenditure',
       'Diphtheria ', ' HIV/AIDS', 'GDP', 'Population',
       ' thinness  1-19 years', ' thinness 5-9 years',
       'Income composition of resources', 'Schooling'],
      dtype='object')
```

1.How many unique countries are present in the dataset?

```
unique_countries = df['Country'].nunique()
print("Number of unique countries:", unique_countries)
```

```
Number of unique countries: 193
```

2.What is the range of years covered in the dataset?

```
min_year = df['Year'].min()
max_year = df['Year'].max()
print("Range of years:", min_year, "-", max_year)
```

```
Range of years: 2000 - 2015
```

3.How many records are there for each country?

```
records_per_country = df['Country'].value_counts()
print("Records per country:")
print(records_per_country)
```

```
Records per country:
Afghanistan             16
Peru                    16
Nicaragua               16
Niger                   16
Nigeria                 16
                        ..
Niue                     1
San Marino               1
Nauru                    1
Saint Kitts and Nevis    1
Dominica                 1
Name: Country, Length: 193, dtype: int64
```

4.What is the average life expectancy across all countries and years?

```
average_life_expectancy = df['Life expectancy'].mean()
print("Average life expectancy:", average_life_expectancy)
```

```
Average life expectancy: 69.22493169398908
```

5.Which country has the highest life expectancy, and in which year was it recorded?

```
max_life_expectancy = df.loc[df['Life expectancy'].idxmax()]
country = max_life_expectancy['Country']
year = max_life_expectancy['Year']
print("Country with the highest life expectancy:", country)
print("Year recorded:", year)
```

```
Country with the highest life expectancy: Belgium
Year recorded: 2014
```

6.Which country has the lowest life expectancy, and in which year was it recorded?

```
min_life_expectancy = df.loc[df['Life expectancy'].idxmin()]
country = min_life_expectancy['Country']
year = min_life_expectancy['Year']
print("Country with the lowest life expectancy:", country)
print("Year recorded:", year)
```

```
Country with the lowest life expectancy: Haiti
Year recorded: 2010
```

7.What is the correlation between life expectancy and GDP?

```
correlation = df['Life expectancy'].corr(df['GDP'])
print("Correlation between life expectancy and GDP:", correlation)
```

```
Correlation between life expectancy and GDP: 0.46145519262073803
```

8.How does the average life expectancy differ between developed and developing countries?

```
avg_life_expectancy_by_status = df.groupby('Status')['Life expectancy'].mean()
print("Average life expectancy by country status:")
print(avg_life_expectancy_by_status)
```

```
Average life expectancy by country status:
Status
Developed     79.197852
Developing    67.111465
Name: Life expectancy, dtype: float64
```

9.What is the distribution of alcohol consumption across the countries?

```
alcohol_distribution = df.groupby('Country')['Alcohol'].mean()
print("Distribution of alcohol consumption:")
print(alcohol_distribution)
```

```
Distribution of alcohol consumption:
Country
Afghanistan                         0.014375
Albania                             4.848750
Algeria                             0.406667
Angola                              5.740667
Antigua and Barbuda                 7.949333
                                        ...
Venezuela (Bolivarian Republic of)  7.420000
Viet Nam                            3.087333
Yemen                               0.047333
Zambia                              2.239333
Zimbabwe                            4.482000
Name: Alcohol, Length: 193, dtype: float64
```

10.How does schooling affect life expectancy?

```
correlation = df['Schooling'].corr(df['Life expectancy'])
print("Correlation between schooling and life expectancy:", correlation)
```

```
Correlation between schooling and life expectancy: 0.7519754627366979
```

11.How many babies deaths were recorded for each country?

```
infant_deaths_per_country = df.groupby('Country')['infant deaths'].sum()
print("Total infant deaths per country:")
print(infant_deaths_per_country)
```

```
Total infant deaths per country:
Country
Afghanistan                          1252
Albania                                11
Algeria                               325
Angola                               1340
Antigua and Barbuda                     0
                                      ...
Venezuela (Bolivarian Republic of)    150
Viet Nam                              467
Yemen                                 630
Zambia                                535
Zimbabwe                              425
Name: infant deaths, Length: 193, dtype: int64
```

12.What is the average percentage expenditure on healthcare across all countries and years?

```
average_healthcare_expenditure = df['percentage expenditure'].mean()
print("Average percentage expenditure on healthcare:", average_healthcare_expenditure)
```

```
Average percentage expenditure on healthcare: 738.2512954533831
```

13.Relationship between measles incidence and hepatitis B coverage?

```
correlation = df['Measles '].corr(df['Hepatitis B'])
print("Correlation between measles incidence and hepatitis B coverage:", correlation)
```

```
Correlation between measles incidence and hepatitis B coverage: -0.12052937169376281
```

14.What is the distribution of BMI (Body Mass Index) values across the countries?

```
bmi_distribution = df.groupby('Country')[' BMI '].mean()
print("Distribution of BMI values:")
print(bmi_distribution)
```

```
Distribution of BMI values:
Country
Afghanistan                          15.51875
Albania                              49.06875
Algeria                              48.74375
Angola                               18.01875
Antigua and Barbuda                  38.42500
                                       ...
Venezuela (Bolivarian Republic of)   54.48750
Viet Nam                             11.18750
Yemen                                33.48750
Zambia                               17.45000
Zimbabwe                             25.13750
Name:  BMI , Length: 193, dtype: float64
```

15.How many countries have a life expectancy greater than 75?

```
countries_gt_75 = df[df['Life expectancy'] > 75]['Country'].nunique()
print("Number of countries with life expectancy greater than 75:", countries_gt_75)
```

```
Number of countries with life expectancy greater than 75: 90
```

16.What is the maximum alcohol consumption recorded among all countries?

```
max_alcohol_consumption = df['Alcohol'].max()
print("Maximum alcohol consumption:", max_alcohol_consumption)
```

```
Maximum alcohol consumption: 17.87
```

17.What is the median GDP per capita?

```
median_gdp = df['GDP'].median()
print("Median GDP per capita:", median_gdp)
```

```
Median GDP per capita: 1766.947595
```

18.What is the average schooling level across all countries?

```
average_schooling = df['Schooling'].mean()
print("Average schooling level:", average_schooling)
```

```
Average schooling level: 11.992792792792793
```

19.What is the average life expectancy for developed and developing countries?

```
average_life_expectancy_by_status = df.groupby('Status')['Life expectancy'].mean()
print("Average life expectancy by country status:")
print(average_life_expectancy_by_status)
```

```
Average life expectancy by country status:
Status
Developed     79.197852
Developing    67.111465
Name: Life expectancy, dtype: float64
```

20.Which country has the highest adult mortality rate?

```
country_highest_adult_mortality = df['Country'].loc[df['Adult Mortality'].idxmax()]
print("Country with the highest adult mortality rate:", country_highest_adult_mortality)
```

```
Country with the highest adult mortality rate: Zimbabwe
```

End of **Code**

✓ 0s completed at 12:59 AM