Covid19 Data Analysis Notebook

November 13, 2024

0.1 #Covid19 Data Analysis Notebook

0.1.1 Importing the modules

[165]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
print('Modules are imported.')
```

```
Modules are imported.
```

0.2 Task 2

0.2.1 Importing covid19 dataset

importing "Covid19_Confirmed_dataset.csv" from "./Dataset" folder.

[166]:
```
corona_dataset_csv = pd.read_csv("Datasets/covid19_Confirmed_dataset.csv")
corona_dataset_csv.head(10)
```

[166]:

	Province/State	Country/Region	Lat	Long	\
0	NaN	Afghanistan	33.0000	65.0000	
1	NaN	Albania	41.1533	20.1683	
2	NaN	Algeria	28.0339	1.6596	
3	NaN	Andorra	42.5063	1.5218	
4	NaN	Angola	-11.2027	17.8739	
5	NaN	Antigua and Barbuda	17.0608	-61.7964	
6	NaN	Argentina	-38.4161	-63.6167	
7	NaN	Armenia	40.0691	45.0382	
8	Australian Capital Territory	Australia	-35.4735	149.0124	
9	New South Wales	Australia	-33.8688	151.2093	

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	…	4/21/20	\
0	0	0	0	0	0	0	…	1092	
1	0	0	0	0	0	0	…	609	
2	0	0	0	0	0	0	…	2811	
3	0	0	0	0	0	0	…	717	
4	0	0	0	0	0	0	…	24	

```
5         0         0         0         0         0         0  …        23
6         0         0         0         0         0         0  …      3031
7         0         0         0         0         0         0  …      1401
8         0         0         0         0         0         0  …       104
9         0         0         0         0         3         4  …      2969

   4/22/20  4/23/20  4/24/20  4/25/20  4/26/20  4/27/20  4/28/20  4/29/20  \
0     1176     1279     1351     1463     1531     1703     1828     1939
1      634      663      678      712      726      736      750      766
2     2910     3007     3127     3256     3382     3517     3649     3848
3      723      723      731      738      738      743      743      743
4       25       25       25       25       26       27       27       27
5       24       24       24       24       24       24       24       24
6     3144     3435     3607     3780     3892     4003     4127     4285
7     1473     1523     1596     1677     1746     1808     1867     1932
8      104      104      105      106      106      106      106      106
9     2971     2976     2982     2994     3002     3004     3016     3016

   4/30/20
0     2171
1      773
2     4006
3      745
4       27
5       24
6     4428
7     2066
8      106
9     3025

[10 rows x 104 columns]
```

Checking the shape of the dataframe

[167]:
```
corona_dataset_csv.shape
```

[167]:
```
(266, 104)
```

0.2.2 Deleting the useless columns for the analysis

[ ]:
```

```

[168]:
```
corona_dataset_csv.drop(["Lat", "Long"], axis = 1, inplace = True)
```

[169]:
```
corona_dataset_csv.head(10)
```

[169]:

	Province/State	Country/Region	1/22/20	1/23/20	\
0	NaN	Afghanistan	0	0	
1	NaN	Albania	0	0	
2	NaN	Algeria	0	0	
3	NaN	Andorra	0	0	
4	NaN	Angola	0	0	
5	NaN	Antigua and Barbuda	0	0	
6	NaN	Argentina	0	0	
7	NaN	Armenia	0	0	
8	Australian Capital Territory	Australia	0	0	
9	New South Wales	Australia	0	0	

	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	…	4/21/20	\
0	0	0	0	0	0	0	…	1092	
1	0	0	0	0	0	0	…	609	
2	0	0	0	0	0	0	…	2811	
3	0	0	0	0	0	0	…	717	
4	0	0	0	0	0	0	…	24	
5	0	0	0	0	0	0	…	23	
6	0	0	0	0	0	0	…	3031	
7	0	0	0	0	0	0	…	1401	
8	0	0	0	0	0	0	…	104	
9	0	0	3	4	4	4	…	2969	

	4/22/20	4/23/20	4/24/20	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	\
0	1176	1279	1351	1463	1531	1703	1828	1939	
1	634	663	678	712	726	736	750	766	
2	2910	3007	3127	3256	3382	3517	3649	3848	
3	723	723	731	738	738	743	743	743	
4	25	25	25	25	26	27	27	27	
5	24	24	24	24	24	24	24	24	
6	3144	3435	3607	3780	3892	4003	4127	4285	
7	1473	1523	1596	1677	1746	1808	1867	1932	
8	104	104	105	106	106	106	106	106	
9	2971	2976	2982	2994	3002	3004	3016	3016	

	4/30/20
0	2171
1	773
2	4006
3	745
4	27
5	24
6	4428
7	2066
8	106
9	3025

[10 rows x 102 columns]

0.2.3 Aggregating the rows by the country

[170]:
```
corona_dataset_aggregated = corona_dataset_csv.groupby("Country/Region").sum()
```

[171]:
```
corona_dataset_aggregated.head()
```

[171]:

Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	\
Afghanistan	0	0	0	0	0	0	0	
Albania	0	0	0	0	0	0	0	
Algeria	0	0	0	0	0	0	0	
Andorra	0	0	0	0	0	0	0	
Angola	0	0	0	0	0	0	0	

Country/Region	1/29/20	1/30/20	1/31/20	…	4/21/20	4/22/20	4/23/20	\
				…				
Afghanistan	0	0	0	…	1092	1176	1279	
Albania	0	0	0	…	609	634	663	
Algeria	0	0	0	…	2811	2910	3007	
Andorra	0	0	0	…	717	723	723	
Angola	0	0	0	…	24	25	25	

Country/Region	4/24/20	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20
Afghanistan	1351	1463	1531	1703	1828	1939	2171
Albania	678	712	726	736	750	766	773
Algeria	3127	3256	3382	3517	3649	3848	4006
Andorra	731	738	738	743	743	743	745
Angola	25	25	26	27	27	27	27

[5 rows x 100 columns]

[172]:
```
corona_dataset_aggregated.shape
```

[172]:
```
(187, 100)
```

0.2.4 Visualizing data related to a country, for example China

visualization always helps for better understanding of our data.

[173]:
```
corona_dataset_aggregated.loc["China"]
```

[173]:
```
1/22/20      548
1/23/20      643
```

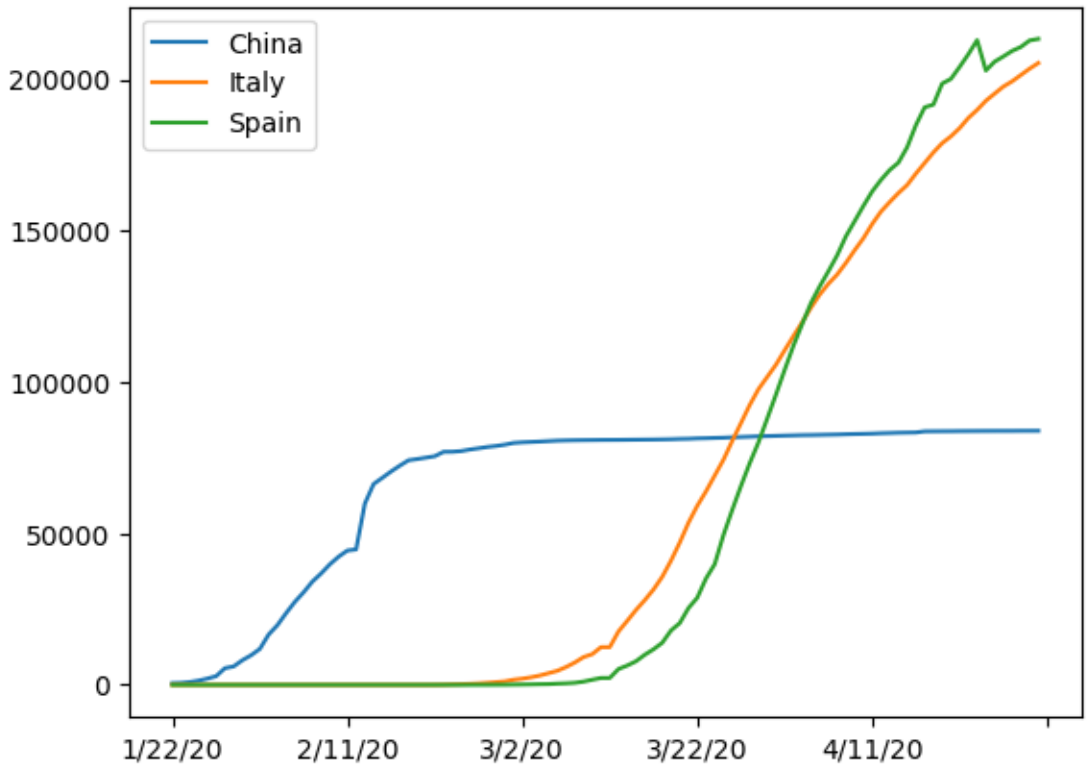
```
1/24/20       920
1/25/20      1406
1/26/20      2075
            ...
4/26/20     83912
4/27/20     83918
4/28/20     83940
4/29/20     83944
4/30/20     83956
Name: China, Length: 100, dtype: int64
```

0.2.5 Task3: Calculating a good measure

Finding a good measure as a number, describing the spread of the virus in a country.

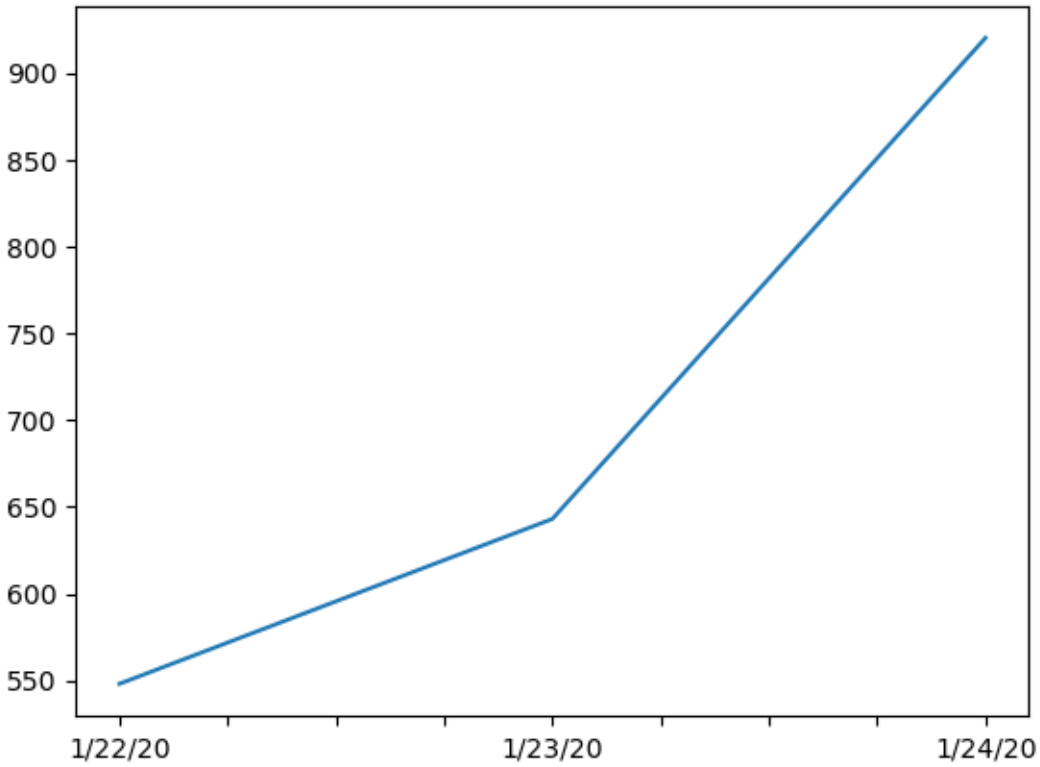
[174]:
```
corona_dataset_aggregated.loc['China'].plot()
corona_dataset_aggregated.loc['Italy'].plot()
corona_dataset_aggregated.loc['Spain'].plot()
plt.legend()
```

[174]:
```
<matplotlib.legend.Legend at 0x74c5e53643d0>
```



[175]:
```
corona_dataset_aggregated.loc['China'][:3].plot()
```

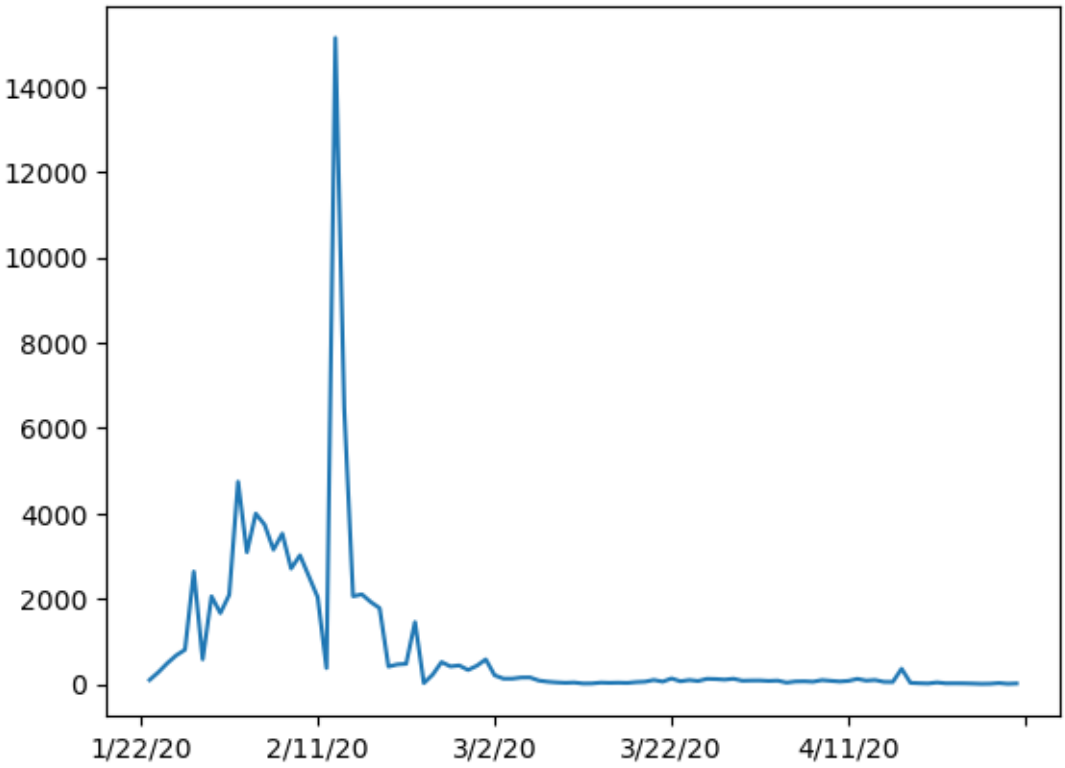
[175]: <AxesSubplot: >



0.2.6 Calculating the first derivative of the curve

[176]:
```
corona_dataset_aggregated.loc['China'].diff().plot()
```

[176]: <AxesSubplot: >



0.2.7 Finding maxmimum infection rate for China, Italy and Spain

[177]:
```
corona_dataset_aggregated.loc['China'].diff().max()
```

[177]: 15136.0

[178]:
```
corona_dataset_aggregated.loc['Italy'].diff().max()
```

[178]: 6557.0

[179]:
```
corona_dataset_aggregated.loc['Spain'].diff().max()
```

[179]: 9630.0

0.2.8 Finding maximum infection rate for all of the countries.

[180]:
```
countries = list(corona_dataset_aggregated.index)
max_infection_rates = []
for c in countries :
    max_infection_rates.append(corona_dataset_aggregated.loc[c].diff().max())
corona_dataset_aggregated["max_infection_rate"] = max_infection_rates
```

[181]:
```
corona_dataset_aggregated.head()
```

[181]:

Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	\
Afghanistan	0	0	0	0	0	0	0	
Albania	0	0	0	0	0	0	0	
Algeria	0	0	0	0	0	0	0	
Andorra	0	0	0	0	0	0	0	
Angola	0	0	0	0	0	0	0	

Country/Region	1/29/20	1/30/20	1/31/20	…	4/22/20	4/23/20	4/24/20	\
Afghanistan	0	0	0	…	1176	1279	1351	
Albania	0	0	0	…	634	663	678	
Algeria	0	0	0	…	2910	3007	3127	
Andorra	0	0	0	…	723	723	731	
Angola	0	0	0	…	25	25	25	

Country/Region	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20	\
Afghanistan	1463	1531	1703	1828	1939	2171	
Albania	712	726	736	750	766	773	
Algeria	3256	3382	3517	3649	3848	4006	
Andorra	738	738	743	743	743	745	
Angola	25	26	27	27	27	27	

Country/Region	max_infection_rate
Afghanistan	232.0
Albania	34.0
Algeria	199.0
Andorra	43.0
Angola	5.0

[5 rows x 101 columns]

0.2.9 Creating a new dataframe with only needed column

[182]:
```
corona_data = pd.DataFrame(corona_dataset_aggregated["max_infection_rate"])
```

[183]:
```
corona_data.head()
```

[183]:

Country/Region	max_infection_rate
Afghanistan	232.0
Albania	34.0
Algeria	199.0

```
Andorra                          43.0
Angola                            5.0
```

0.2.10 Task4:

- Importing the WorldHappinessReport.csv dataset
- selecting needed columns for our analysis
- join the datasets
- calculate the correlations as the result of our analysis

0.2.11 Importing the dataset

[184]:
```
happiness_report_csv = pd.read_csv("Datasets/worldwide_happiness_report.csv")
```

[185]:
```
happiness_report_csv.head()
```

[185]:
```
   Overall rank Country or region  Score  GDP per capita  Social support  \
0             1           Finland  7.769           1.340           1.587
1             2           Denmark  7.600           1.383           1.573
2             3            Norway  7.554           1.488           1.582
3             4           Iceland  7.494           1.380           1.624
4             5       Netherlands  7.488           1.396           1.522

   Healthy life expectancy  Freedom to make life choices  Generosity  \
0                    0.986                         0.596       0.153
1                    0.996                         0.592       0.252
2                    1.028                         0.603       0.271
3                    1.026                         0.591       0.354
4                    0.999                         0.557       0.322

   Perceptions of corruption
0                      0.393
1                      0.410
2                      0.341
3                      0.118
4                      0.298
```

0.2.12 Droping the useless columns

[186]:
```
useless_cols = ["Overall rank", "Score", "Generosity", "Perceptions of␣
↪corruption"]
```

[187]:
```
happiness_report_csv.drop(useless_cols, axis = 1, inplace = True)
happiness_report_csv.head()
```

[187]:
```
  Country or region  GDP per capita  Social support  Healthy life expectancy  \
0           Finland           1.340           1.587                    0.986
```

```
1           Denmark           1.383           1.573                     0.996
2            Norway           1.488           1.582                     1.028
3           Iceland           1.380           1.624                     1.026
4       Netherlands           1.396           1.522                     0.999

   Freedom to make life choices
0                         0.596
1                         0.592
2                         0.603
3                         0.591
4                         0.557
```

0.2.13 Changing the indices of the dataframe

[188]:
```
happiness_report_csv.set_index("Country or region", inplace = True)
```

[189]:
```
happiness_report_csv.head()
```

[189]:
```
                   GDP per capita  Social support  Healthy life expectancy  \
Country or region
Finland                     1.340           1.587                    0.986
Denmark                     1.383           1.573                    0.996
Norway                      1.488           1.582                    1.028
Iceland                     1.380           1.624                    1.026
Netherlands                 1.396           1.522                    0.999

                   Freedom to make life choices
Country or region
Finland                                   0.596
Denmark                                   0.592
Norway                                    0.603
Iceland                                   0.591
Netherlands                               0.557
```

0.2.14 Joining the two datasets prepared

Corona Dataset :

[190]:
```
corona_data.head()
```

[190]:
```
                max_infection_rate
Country/Region
Afghanistan                  232.0
Albania                       34.0
Algeria                      199.0
Andorra                       43.0
Angola                         5.0
```

[191]:
```
corona_data.shape
```

[191]:
```
(187, 1)
```

World happiness report Dataset :

[192]:
```
happiness_report_csv.head()
```

[192]:

Country or region	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Finland	1.340	1.587	0.986	0.596
Denmark	1.383	1.573	0.996	0.592
Norway	1.488	1.582	1.028	0.603
Iceland	1.380	1.624	1.026	0.591
Netherlands	1.396	1.522	0.999	0.557

[193]:
```
happiness_report_csv.shape
```

[193]:
```
(156, 4)
```

[194]:
```
data = corona_data.join(happiness_report_csv, how = "inner")
data.head()
```

[194]:

	max_infection_rate	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Afghanistan	232.0	0.350	0.517	0.361	0.000
Albania	34.0	0.947	0.848	0.874	0.383
Algeria	199.0	1.002	1.160	0.785	0.086
Argentina	291.0	1.092	1.432	0.881	0.471
Armenia	134.0	0.850	1.055	0.815	0.283

0.2.15 Correlation matrix

[195]:
```
data.corr()
```

[195]:
```
                                max_infection_rate  GDP per capita  \
max_infection_rate                        1.000000        0.250118
GDP per capita                            0.250118        1.000000
Social support                            0.191958        0.759468
Healthy life expectancy                   0.289263        0.863062
Freedom to make life choices              0.078196        0.394603

                                Social support  Healthy life expectancy  \
max_infection_rate                    0.191958                 0.289263
GDP per capita                        0.759468                 0.863062
Social support                        1.000000                 0.765286
Healthy life expectancy               0.765286                 1.000000
Freedom to make life choices          0.456246                 0.427892

                                Freedom to make life choices
max_infection_rate                                  0.078196
GDP per capita                                      0.394603
Social support                                      0.456246
Healthy life expectancy                             0.427892
Freedom to make life choices                        1.000000
```

0.2.16 Task 5: Visualization of the results

[197]:
```
data.head()
```

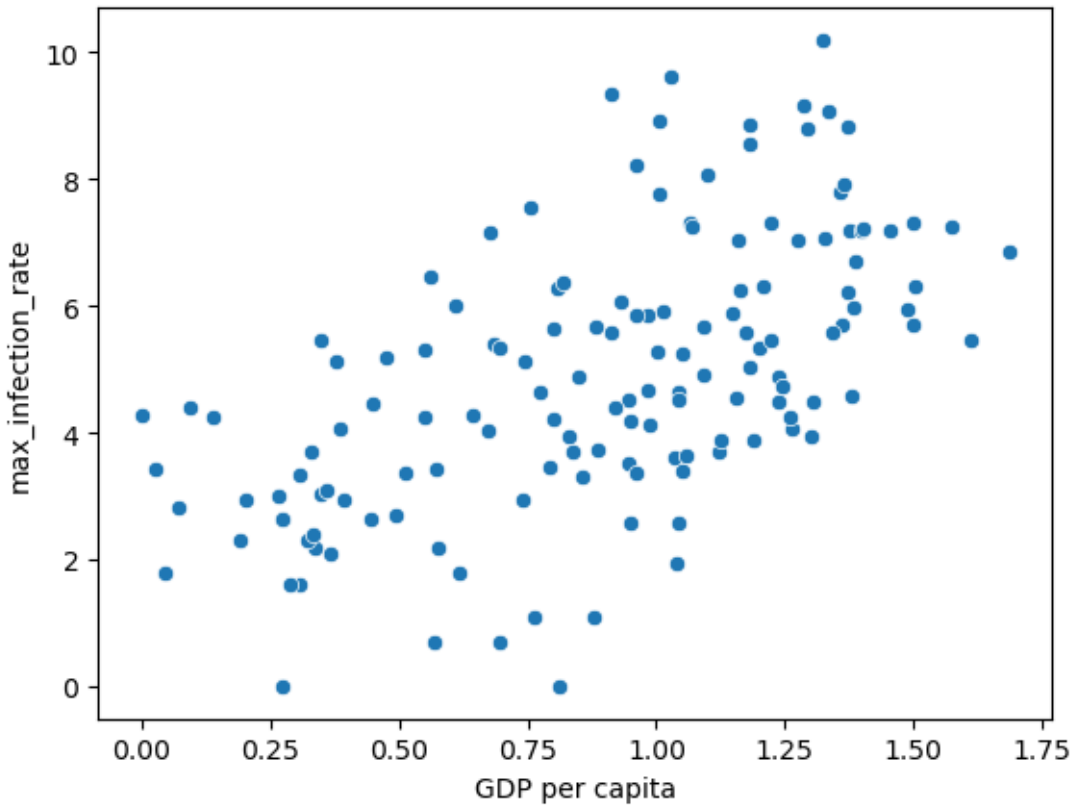
[197]:
```
             max_infection_rate  GDP per capita  Social support  \
Afghanistan               232.0           0.350           0.517
Albania                    34.0           0.947           0.848
Algeria                   199.0           1.002           1.160
Argentina                 291.0           1.092           1.432
Armenia                   134.0           0.850           1.055

             Healthy life expectancy  Freedom to make life choices
Afghanistan                    0.361                         0.000
Albania                        0.874                         0.383
Algeria                        0.785                         0.086
Argentina                      0.881                         0.471
Armenia                        0.815                         0.283
```

0.2.17 Plotting GDP vs maximum Infection rate

[199]:
```
x = data["GDP per capita"]
y = data["max_infection_rate"]
```

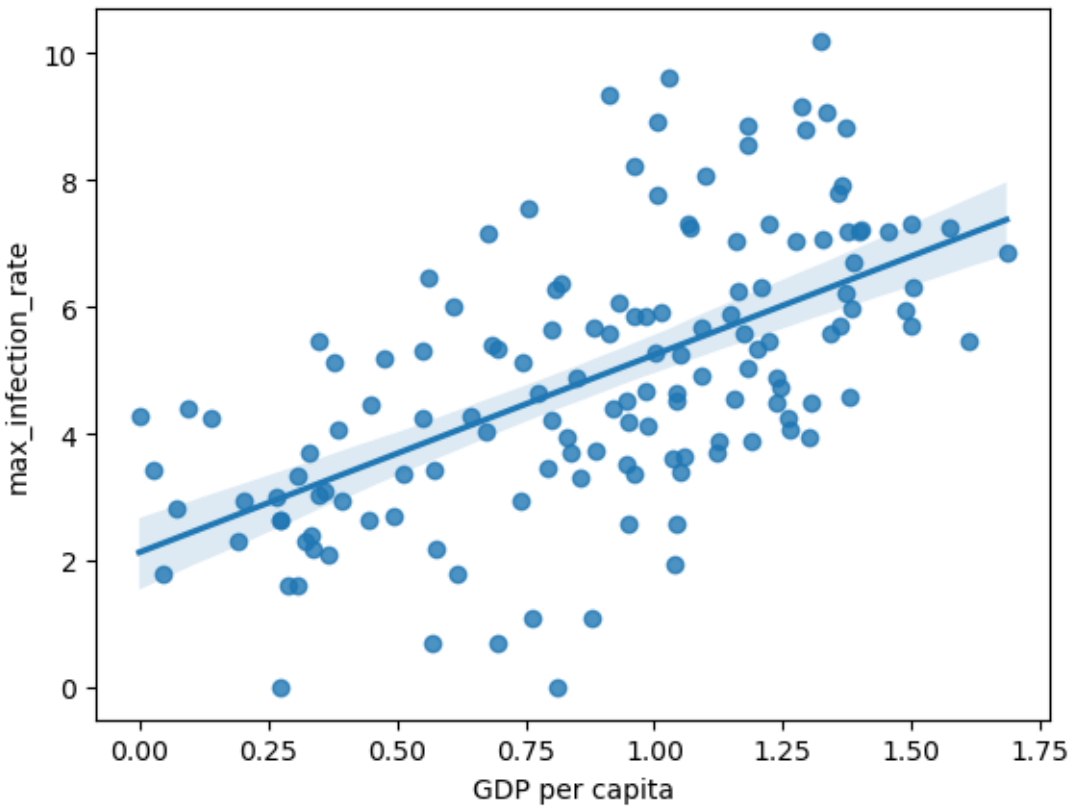
[211]:
```
sns.scatterplot(x = data["GDP per capita"], y = np.
↪log(data["max_infection_rate"]))
```

[211]: <AxesSubplot: xlabel='GDP per capita', ylabel='max_infection_rate'>



[212]:
```
sns.regplot(x = data["GDP per capita"], y = np.log(data["max_infection_rate"]))
```

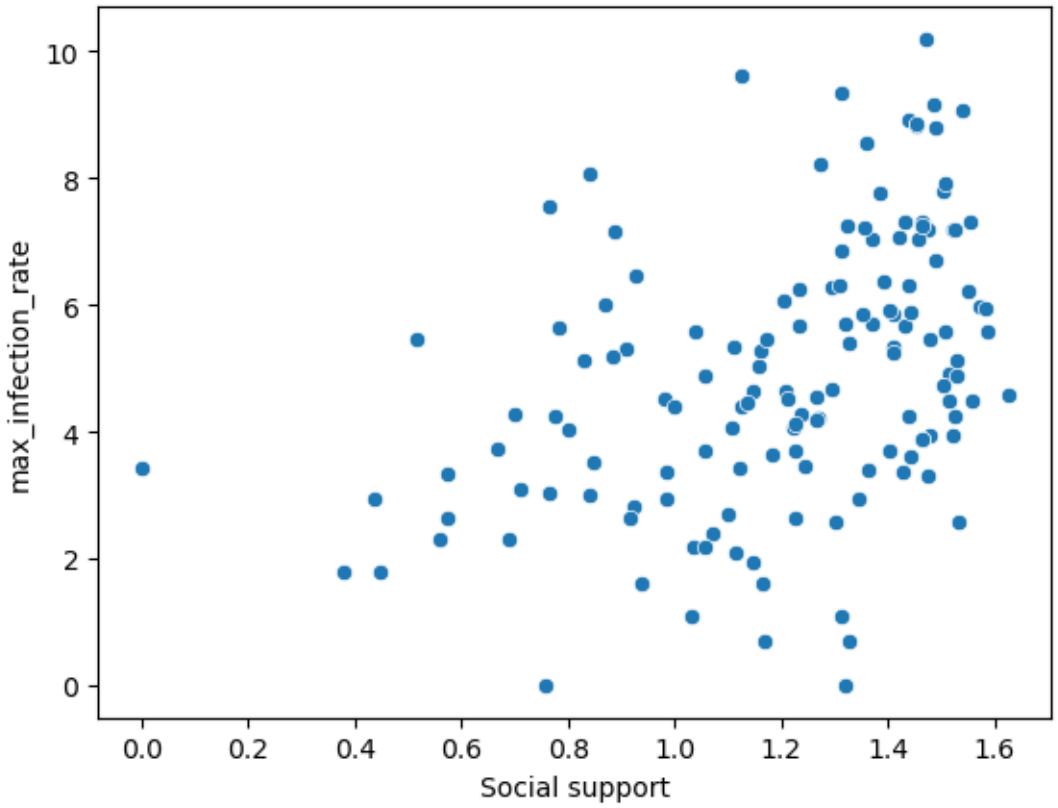
[212]: <AxesSubplot: xlabel='GDP per capita', ylabel='max_infection_rate'>



0.2.18 Plotting Social support vs maximum Infection rate

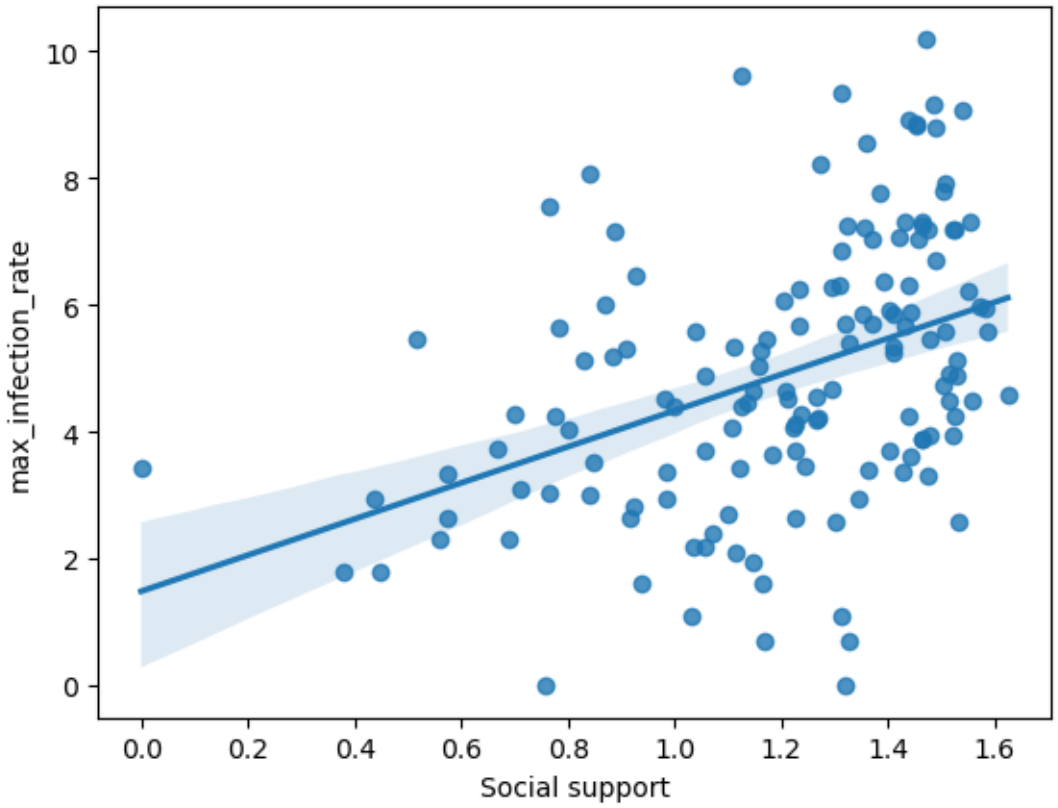
[213]:
```
sns.scatterplot(x = data["Social support"], y = np.
↪log(data["max_infection_rate"]))
```

[213]: <AxesSubplot: xlabel='Social support', ylabel='max_infection_rate'>



[214]:
```
sns.regplot(x = data["Social support"], y = np.log(data["max_infection_rate"]))
```

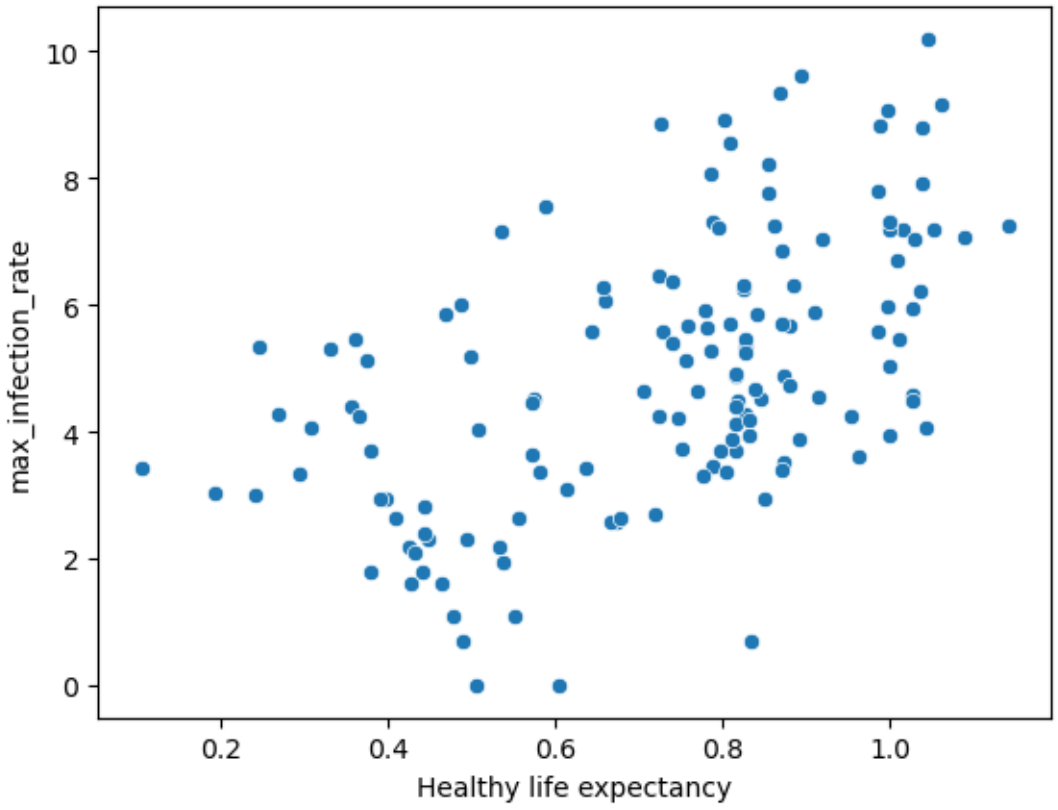
[214]:
```
<AxesSubplot: xlabel='Social support', ylabel='max_infection_rate'>
```



0.2.19 Plotting Healthy life expectancy vs maximum Infection rate

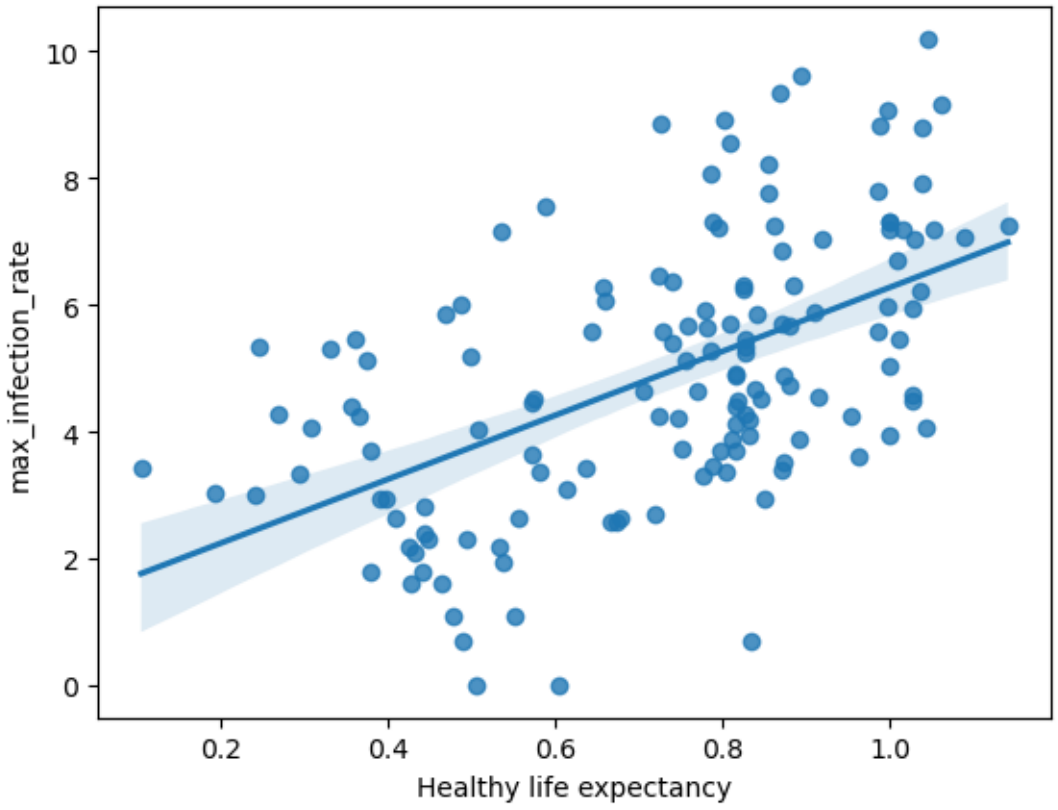
[215]:
```
sns.scatterplot(x = data["Healthy life expectancy"], y = np.
↪log(data["max_infection_rate"]))
```

[215]: <AxesSubplot: xlabel='Healthy life expectancy', ylabel='max_infection_rate'>



[216]:
```
sns.regplot(x = data["Healthy life expectancy"], y = np.
↪log(data["max_infection_rate"]))
```

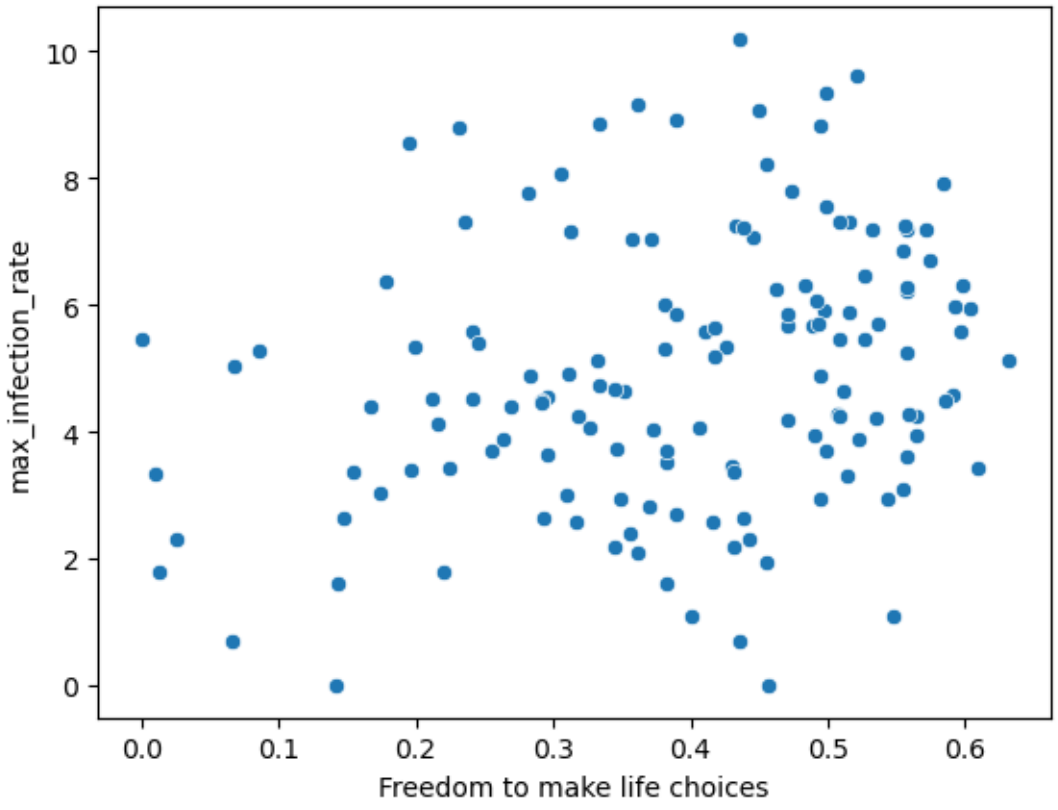
[216]: <AxesSubplot: xlabel='Healthy life expectancy', ylabel='max_infection_rate'>



0.2.20 Plotting Freedom to make life choices vs maximum Infection rate

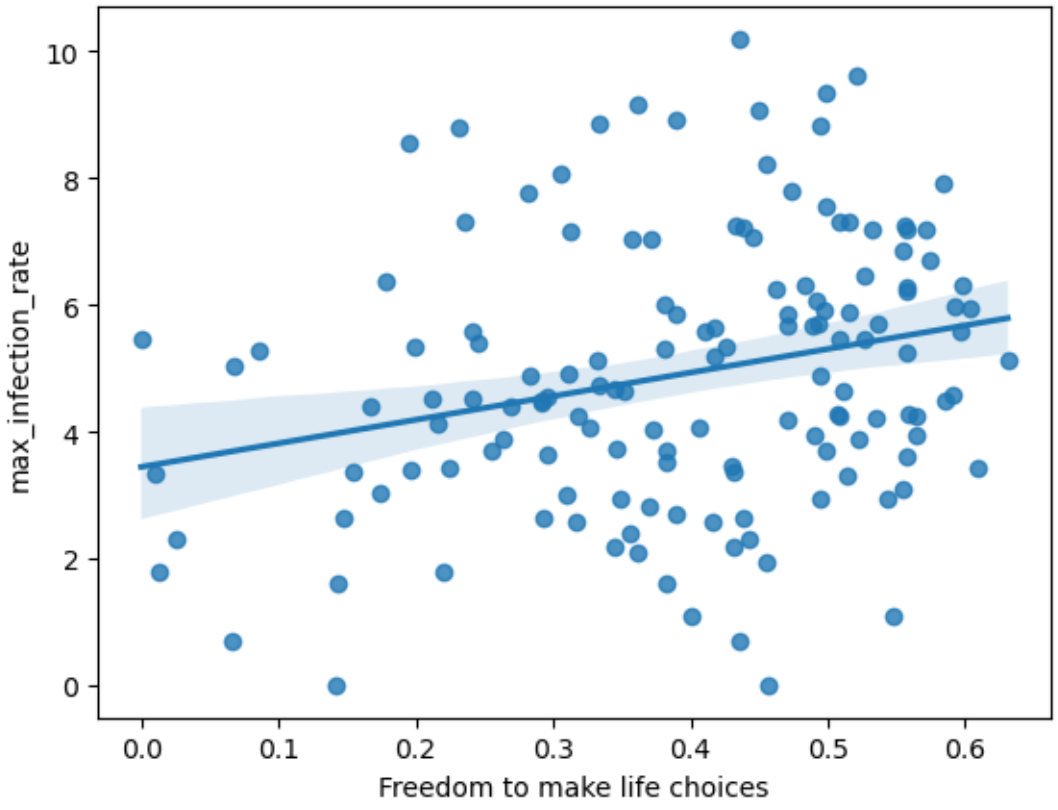
[217]:
```
sns.scatterplot(x = data["Freedom to make life choices"], y = np.
↪log(data["max_infection_rate"]))
```

[217]:
```
<AxesSubplot: xlabel='Freedom to make life choices',
ylabel='max_infection_rate'>
```



[218]:
```
sns.regplot(x = data["Freedom to make life choices"], y = np.
↪log(data["max_infection_rate"]))
```

[218]:
```
<AxesSubplot: xlabel='Freedom to make life choices',
ylabel='max_infection_rate'>
```



0.2.21 Conclusion. People who are living in developed countries are more prone to getting the infection of COVID19 with compare of less developed countries.