

python-project-1

November 18, 2023

1 Pandemic Impact: COVID-19 and World Happiness

Submitted by:

SHIMIL SHIJO
23MSP3025

HRIDYA A
23MSP3083

1.1 AIM

Investigating the relation Between Country-Specific COVID-19 Rates and World Happiness Index Scores to Understand the Impact of the Pandemic on Societal Well-being.

1.1.1 DATASET DESCRIPTION

This project combines two datasets: the COVID-19 dataset from the WHO website and the World Happiness Report dataset from Kaggle. The former provides COVID-19 rates for countries, while the latter offers happiness scores.

[COVID-19 Dataset](#)

[World Happiness Report Dataset](#)

```
[ ]: #Importing necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
[ ]: #Reading COVID-19 dataset
covid_data = pd.read_csv('covid_dataset.csv')
covid_data.head()
```

```
[ ]: Province/State Country/Region      Lat      Long  1/22/20  1/23/20  1/24/20  \
0      NaN      Afghanistan  33.0000  65.0000      0      0      0
1      NaN      Albania    41.1533  20.1683      0      0      0
2      NaN      Algeria    28.0339   1.6596      0      0      0
3      NaN      Andorra    42.5063   1.5218      0      0      0
```

4		NaN		Angola	-11.2027	17.8739		0	0	0
---	--	-----	--	--------	----------	---------	--	---	---	---

	1/25/20	1/26/20	1/27/20	...	4/21/20	4/22/20	4/23/20	4/24/20	\
0	0	0	0	...	1092	1176	1279	1351	
1	0	0	0	...	609	634	663	678	
2	0	0	0	...	2811	2910	3007	3127	
3	0	0	0	...	717	723	723	731	
4	0	0	0	...	24	25	25	25	

	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20
0	1463	1531	1703	1828	1939	2171
1	712	726	736	750	766	773
2	3256	3382	3517	3649	3848	4006
3	738	738	743	743	743	745
4	25	26	27	27	27	27

[5 rows x 104 columns]

1.2 DATA PRE-PROCESSING

1.3 A. Pre-processing COVID-19 Dataset

```
[ ]: #checking dataframe shape
covid_data.shape
```

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

```
[ ]: #Checking column names
columns = covid_data.columns
columns
```

```
[ ]: Index(['Province/State', 'Country/Region', 'Lat', 'Long', '1/22/20', '1/23/20',
          '1/24/20', '1/25/20', '1/26/20', '1/27/20',
          ...,
          '4/21/20', '4/22/20', '4/23/20', '4/24/20', '4/25/20', '4/26/20',
          '4/27/20', '4/28/20', '4/29/20', '4/30/20'],
          dtype='object', length=104)
```

1. Checking Missing Values

```
[ ]: covid_data.isna().sum()
```

```
[ ]: Province/State    184
      Country/Region    0
      Lat              0
      Long             0
      1/22/20          0
```

```

...
4/26/20      0
4/27/20      0
4/28/20      0
4/29/20      0
4/30/20      0
Length: 104, dtype: int64

```

Interpretation : There is no missing values in the dataset.

2. Checking Duplicate Rows

```
[ ]: covid_data.duplicated().any()
```

```
[ ]: False
```

Interpretation : Dataset does not have duplicate rows

3. Remove not required columns

```
[ ]: #Remove not required columns
#Latitude and Longitude cols are irrelevant for the context.Hence this can be
    ↪removed.
covid_data.drop(['Lat','Long'],axis = 1, inplace=True)
print(covid_data.columns)
```

```

Index(['Province/State', 'Country/Region', '1/22/20', '1/23/20', '1/24/20',
      '1/25/20', '1/26/20', '1/27/20', '1/28/20', '1/29/20',
      ...
      '4/21/20', '4/22/20', '4/23/20', '4/24/20', '4/25/20', '4/26/20',
      '4/27/20', '4/28/20', '4/29/20', '4/30/20'],
      dtype='object', length=102)

```

4. Grouping the records based on country name

```
[ ]: covid_data_grouped = covid_data.groupby('Country/Region').sum()
covid_data_grouped.head()
```

```
<ipython-input-25-efa62008396f>:1: FutureWarning: The default value of
numeric_only in DataFrameGroupBy.sum is deprecated. In a future version,
numeric_only will default to False. Either specify numeric_only or select only
columns which should be valid for the function.
```

```
covid_data_grouped = covid_data.groupby('Country/Region').sum()
```

```
[ ]:
```

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	\
Country/Region								
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
--------	---	---	---	---	---	---	---

	1/29/20	1/30/20	1/31/20	...	4/21/20	4/22/20	4/23/20	\
Country/Region				...				
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	

	4/24/20	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20
Country/Region							
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]

```
[ ]: #Checking shape of grouped dataframe
covid_data_grouped.shape
```

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

5. Find total infections in all countries

```
[ ]: countries = list(covid_data_grouped.index)
infection_count = []

for c in countries:
    infection_count.append(covid_data_grouped.loc[c].sum())

infection_count
```

```
[ ]: [28462,
17864,
74325,
21893,
649,
678,
84105,
40610,
224354,
502063,
37281,
1677,
```

56608,
70829,
2326,
150475,
1082648,
429,
1246,
213,
15387,
36234,
444,
1085638,
5777,
29366,
17131,
2463,
185,
1472,
4842,
31564,
982149,
485,
832,
282177,
6686938,
107654,
1,
3516,
9281,
20853,
21985,
56068,
26458,
21935,
209985,
236274,
53404,
16408,
539,
122072,
337630,
90542,
5489,
2997,
1092,
48428,
918,

2761,
511,
112892,
4132964,
3192,
268,
11027,
4531683,
25175,
77726,
481,
7645,
16059,
1541,
1703,
1461,
305,
14128,
52370,
61442,
457945,
176940,
2783401,
48863,
389169,
393611,
6139613,
5393,
282573,
14237,
48155,
7633,
572513,
11608,
55693,
13847,
553,
23426,
24136,
2097,
1213,
3065,
37109,
116527,
278,
3414,
528,

173243,
2808,
6004,
12806,
250,
9755,
237643,
65045,
3040,
934,
8890,
75625,
1118,
569,
884,
949574,
42498,
302,
15202,
20351,
31559,
251779,
33924,
245996,
120004,
151,
5592,
418646,
175219,
251454,
579961,
167128,
237705,
1141721,
4844,
410,
508,
347,
15518,
118,
274362,
13469,
145420,
436,
1109,
202073,
31703,

```

45214,
5343,
98217,
191,
5979474,
10572,
3062,
388,
422084,
922995,
973,
16433,
15,
4738,
91657,
375,
2591,
4072,
25532,
2156983,
20606211,
2008,
140485,
181619,
3206716,
17681,
36115,
8141,
10708,
11195,
142,
31,
1948,
671]

```

5. Adding new columns of highest and lowest infection rates for all countries

```

[ ]: covid_data_grouped["Total Infections"] = infection_count
covid_data_grouped.head()

```

```

[ ]:

```

	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	\
Country/Region								
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	

	1/29/20	1/30/20	1/31/20	...	4/22/20	4/23/20	4/24/20	\
Country/Region				...				
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	

	4/25/20	4/26/20	4/27/20	4/28/20	4/29/20	4/30/20	\
Country/Region							
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	

	Total Infections
Country/Region	
Afghanistan	28462
Albania	17864
Algeria	74325
Andorra	21893
Angola	649

[5 rows x 101 columns]

6. Creating new dataframe with only required fields

```
[ ]: country_covid_rate_df = pd.DataFrame(covid_data_grouped["Total Infections"])
country_covid_rate_df.head()
```

```
[ ]:
Country/Region
Afghanistan      28462
Albania          17864
Algeria          74325
Andorra          21893
Angola           649
```

1.4 B. Pre-processing World Happiness Index Dataset

1.4.1 Importing happiness index dataset

```
[ ]: happiness_index_data = pd.read_csv("happiness_index.csv")
happiness_index_data.head()
```

```
[ ]: 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
```

1. Checking Missing Values

```
[ ]: happiness_index_data.isnull().sum()
```

```
[ ]: Overall rank      0
      Country or region 0
      Score            0
      GDP per capita    0
      Social support    0
      Healthy life expectancy 0
      Freedom to make life choices 0
      Generosity        0
      Perceptions of corruption 0
      dtype: int64
```

Interpretation : There is no missing values in the dataset.

2. Checking Duplicate Rows

```
[ ]: happiness_index_data.duplicated().any()
```

```
[ ]: False
```

Interpretation : There is no duplicate rows in the dataset.

3. Removing not required Columns

```
[ ]: remove_columns = ["Overall rank", "Score", "Generosity"]
happiness_index_data.drop(remove_columns, axis=1, inplace=True)
happiness_index_data.head()
```

```
[ ]: 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  Perceptions of corruption
0      0.596      0.393
1      0.592      0.410
2      0.603      0.341
3      0.591      0.118
4      0.557      0.298
```

4. Changing dataframe index to country name

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

```
[ ]: 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  Perceptions of corruption
Country or region
Finland      0.596      0.393
Denmark      0.592      0.410
Norway       0.603      0.341
Iceland      0.591      0.118
Netherlands  0.557      0.298
```

1.5 C. Joining Covid Dataset and Happiness Index Dataset

```
[ ]: #COVID Dataset
country_covid_rate_df.head()
```

```
[ ]: Total Infections
Country/Region
Afghanistan      28462
```

Albania	17864
Algeria	74325
Andorra	21893
Angola	649

```
[ ]: #Shape of COVID dataset
country_covid_rate_df.shape
```

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

```
[ ]: #World Happiness index dataset
happiness_index_data.head()
```

```
[ ]:
      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  Perceptions of corruption
Country or region
Finland                        0.596                0.393
Denmark                        0.592                0.410
Norway                         0.603                0.341
Iceland                        0.591                0.118
Netherlands                    0.557                0.298
```

```
[ ]: #Shape of happiness index data
happiness_index_data.shape
```

```
[ ]: (156, 5)
```

```
[ ]: #Perform inner join to include the details of countries which appears in both
      ↪ datasets
data = country_covid_rate_df.join(happiness_index_data,
                                  how = "inner"
                                  )
data.head()
```

```
[ ]:
      Total Infections  GDP per capita  Social support \
Afghanistan           28462           0.350           0.517
Albania                17864           0.947           0.848
Algeria                74325           1.002           1.160
Argentina              84105           1.092           1.432
Armenia                40610           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

	Perceptions of corruption
Afghanistan	0.025
Albania	0.027
Algeria	0.114
Argentina	0.050
Armenia	0.064

1.6 EXPLORATORY DATA ANALYSIS

1.7 1. Correlation matrix

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

	Total Infections	GDP per capita \
Total Infections	1.000000	0.280044
GDP per capita	0.280044	1.000000
Social support	0.183900	0.759468
Healthy life expectancy	0.316972	0.863062
Freedom to make life choices	0.043221	0.394603
Perceptions of corruption	0.108712	0.311577

	Social support	Healthy life expectancy \
Total Infections	0.183900	0.316972
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
Perceptions of corruption	0.203225	0.314811

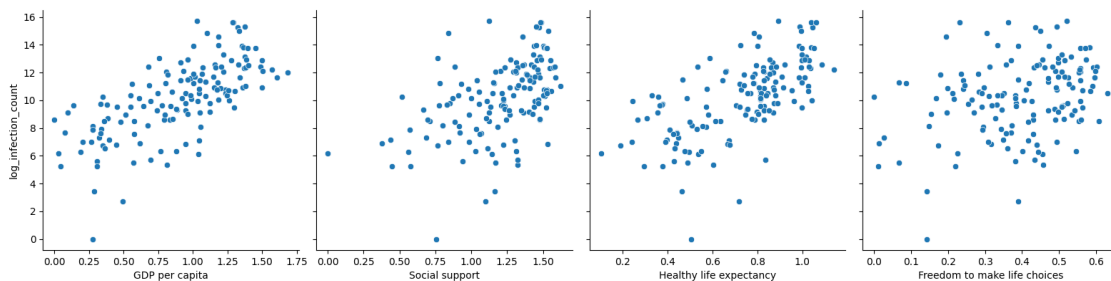
	Freedom to make life choices \
Total Infections	0.043221
GDP per capita	0.394603
Social support	0.456246
Healthy life expectancy	0.427892
Freedom to make life choices	1.000000
Perceptions of corruption	0.446677

	Perceptions of corruption
Total Infections	0.108712

GDP per capita	0.311577
Social support	0.203225
Healthy life expectancy	0.314811
Freedom to make life choices	0.446677
Perceptions of corruption	1.000000

1.8 2. Pair Plot : How infection rate is correlated with other variables

```
[ ]: #Pair Plot
data['log_infection_count'] = np.log(data['Total Infections'])
sns.pairplot(data, x_vars=['GDP per capita', 'Social support', 'Healthy life_
    ↳expectancy', 'Freedom to make life choices'], y_vars='log_infection_count',
    ↳height=4, aspect=1, kind='scatter')
plt.show()
```



Interpretation : All the graphs are weakly positively correlated with target variable.

1.9 3. Regression Plot

```
[ ]: #Regression Plot
# Set up a 1x4 grid of subplots
fig, axes = plt.subplots(1, 4, figsize=(16, 4))

# Plot each regression plot on a separate subplot
sns.regplot(x='GDP per capita', y='log_infection_count', data=data, ax=axes[0])
axes[0].set_title('GDP per capita')

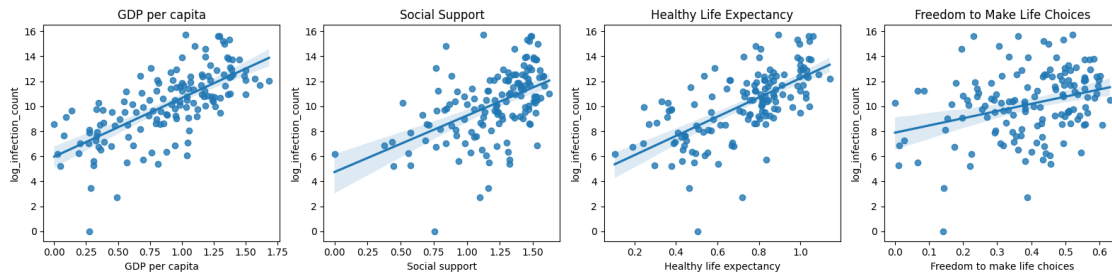
sns.regplot(x='Social support', y='log_infection_count', data=data, ax=axes[1])
axes[1].set_title('Social Support')

sns.regplot(x='Healthy life expectancy', y='log_infection_count', data=data,
    ↳ax=axes[2])
axes[2].set_title('Healthy Life Expectancy')

sns.regplot(x='Freedom to make life choices', y='log_infection_count',
    ↳data=data, ax=axes[3])
```

```
axes[3].set_title('Freedom to Make Life Choices')
```

```
# Adjust layout for better spacing
plt.tight_layout()
plt.show()
```



Interpretation : GDP per capita and Healthy life expectancy are weakly positively correlated with target variable.

1.10 4. Distribution of Target Variable

```
[ ]: # Distribution of the target variable
data['log_infection_count'] = np.log(data['Total Infections'])
sns.distplot(data['log_infection_count'])
```

<ipython-input-43-94ef63c433b6>:3: UserWarning:

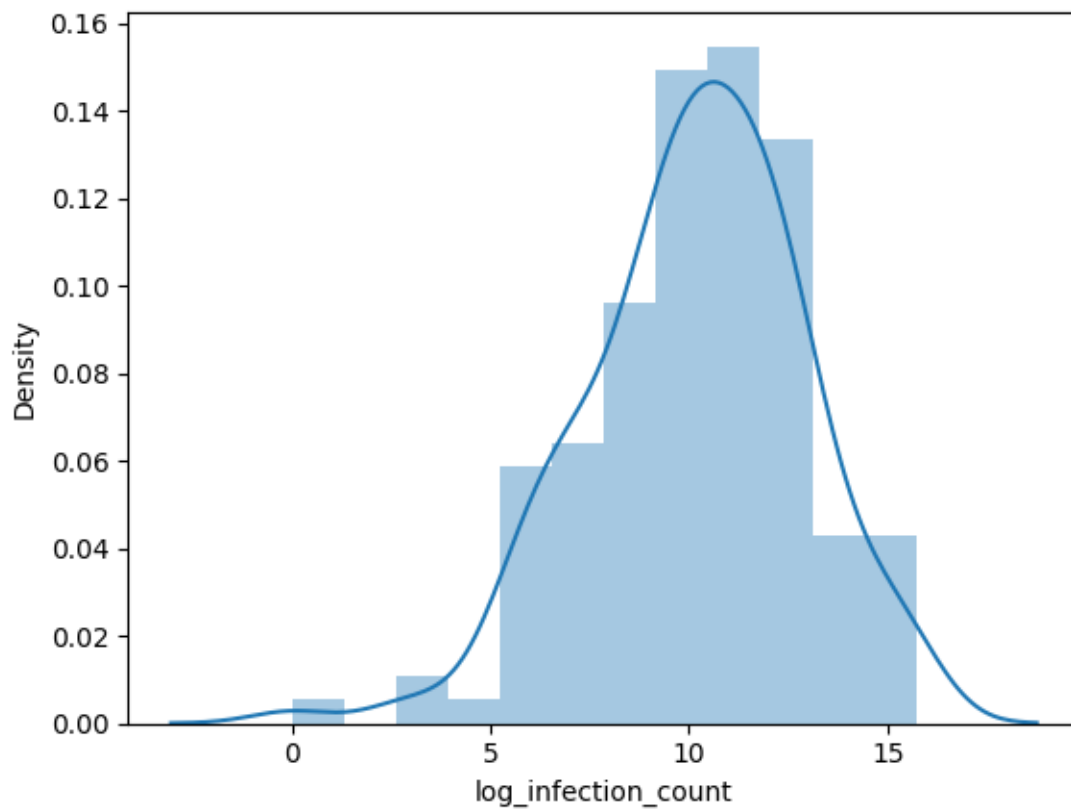
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(data['log_infection_count'])
```

```
[ ]: <Axes: xlabel='log_infection_count', ylabel='Density'>
```

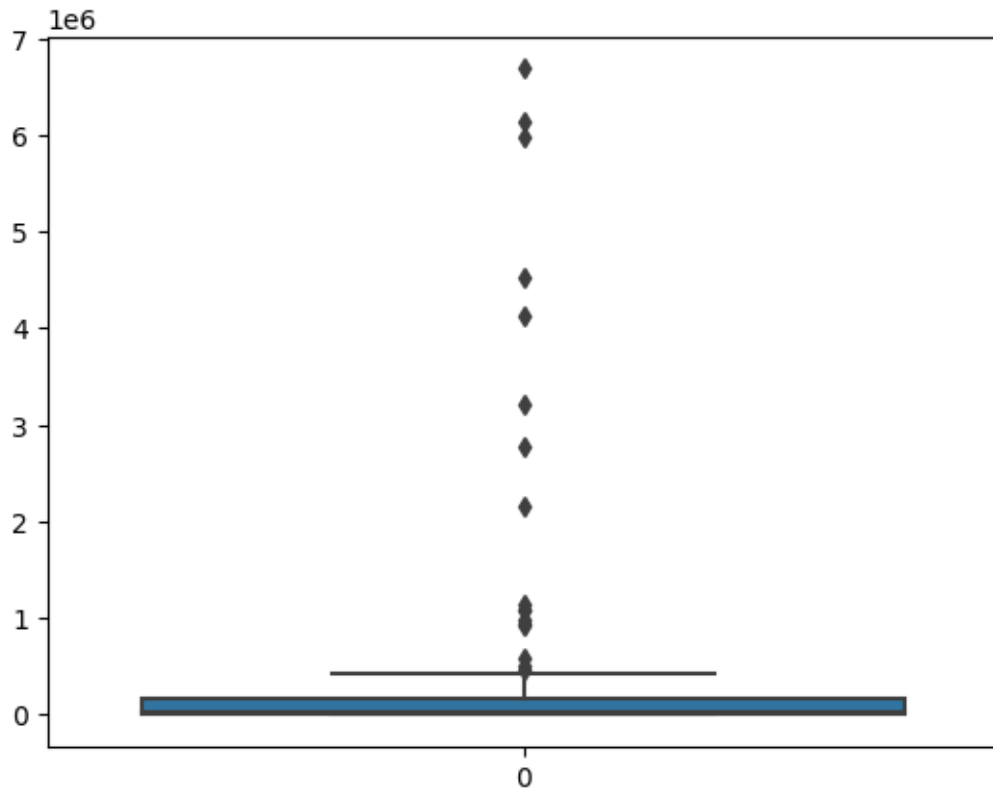


Interpretation : Distribution is slightly left skewed

1.11 5. Checking Outliers

```
[ ]: # Checking Outliers
data.head()
sns.boxplot(data['Total Infections'])
```

```
[ ]: <Axes: >
```

Interpretation : There are many outliers present in the dataset

```
[ ]: #Dropping irrelevant column
remove_columns = ["log_infection_count"]
data.drop(remove_columns, axis=1, inplace=True)
data.head()
```

```
[ ]:
```

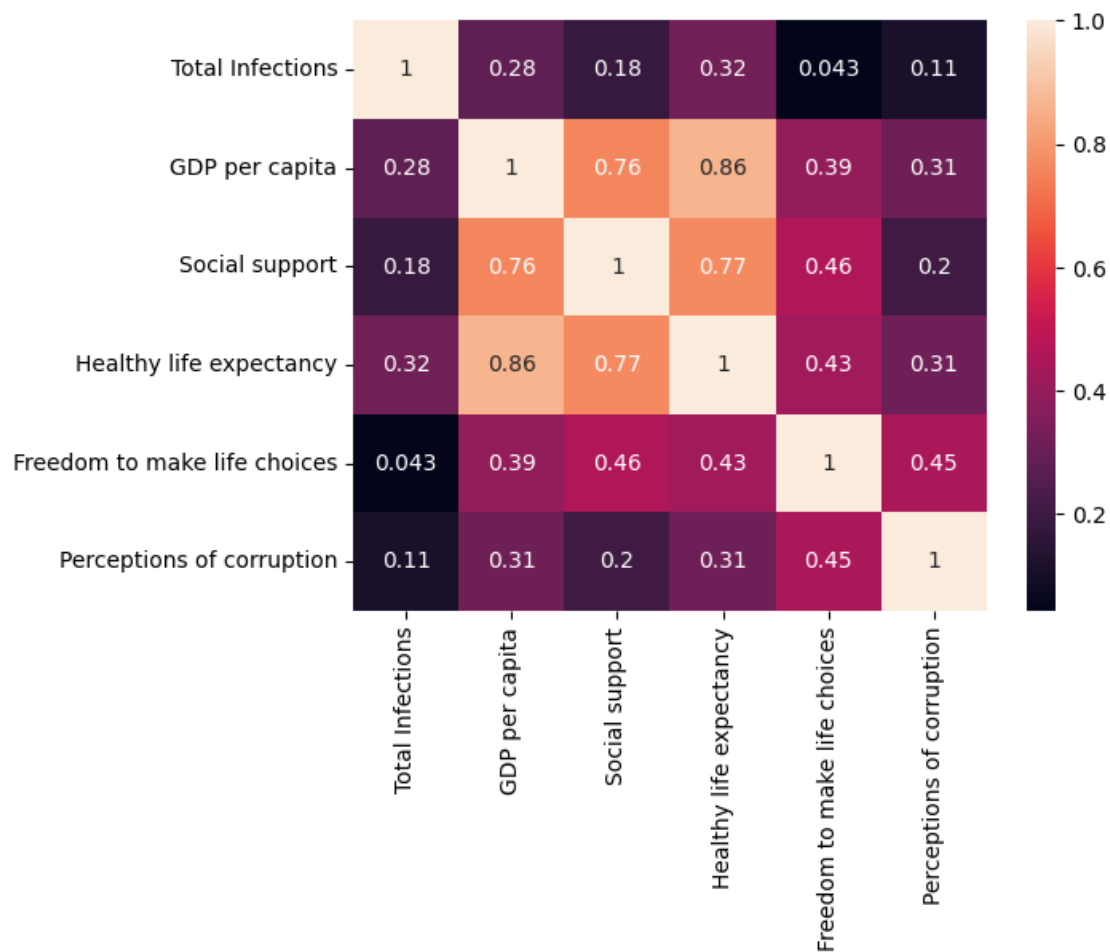
	Total Infections	GDP per capita	Social support \
Afghanistan	28462	0.350	0.517
Albania	17864	0.947	0.848
Algeria	74325	1.002	1.160
Argentina	84105	1.092	1.432
Armenia	40610	0.850	1.055

	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

	Perceptions of corruption
Afghanistan	0.025
Albania	0.027
Algeria	0.114
Argentina	0.050
Armenia	0.064

1.12 6. Heatmap

```
[ ]: #Heatmap
sns.heatmap(data.corr(), annot = True)
plt.show()
```



Interpretation: Healthy life expectancy seems to be most correlated (but weakly) with total infections compared to other independent variables.

2 STORING DATA IN DATABASE

```
[ ]: import pandas as pd
import sqlite3
```

```
[ ]: #Adding csv data into dataframe
csv_file_path1 = 'covid_dataset.csv'
csv_file_path2 = 'happiness_index.csv'
df1 = pd.read_csv(csv_file_path1)
df2= pd.read_csv(csv_file_path2)
```

```
[ ]: #Database and table creation and data population
try:
    #create a database
    covid_db = 'covid_analysis.db'

    # Create a connection
    sqliteConnection = sqlite3.connect(covid_db)
    print("Connection Successful!!")

    # COnvert dataframe into table
    df1.to_sql('covid_table', sqliteConnection, index=False, if_exists='replace')
    df2.to_sql('happiness_table', sqliteConnection, index=False,
    ↪if_exists='replace')
    print("Tables created")

except sqlite3.Error as error:
    print("Error while creating table",error)

finally:
    #finally block will be executed always
    if sqliteConnection:
        sqliteConnection.close()
        print("Connection is closed")
```

Connection Successful!!

Tables created

Connection is closed

```
[ ]: # READ covid TABLE
def readTable():
    try:
        sqliteConnection = sqlite3.connect(covid_db)
        cursor = sqliteConnection.cursor()
        print("Connection Successful!!")
        sqlite_create_table_query = '''SELECT * FROM covid_table;'''
        cursor.execute(sqlite_create_table_query)
```

```
readTable()
```

length of records: 266

```
24, 24, 24, 24, 24), (None, 'Argentina', -38.4161, -63.6167, 0, 0, 0, 0, 0, 0,
```

[illegible]

[illegible]

[illegible]

990, 990, 990, 990, 990, 990, 990, 990, 991, 991, 991, 991, 991, 991, 991, 991, 991), ('Beijing',
991, 991, 991, 991, 991, 991, 991, 991, 991, 991, 991, 991, 991), ('China', 40.1824, 116.4142, 14, 22, 36, 41, 68, 80, 91, 111, 114, 139, 168, 191,
212, 228, 253, 274, 297, 315, 326, 337, 342, 352, 366, 372, 375, 380, 381, 387,
393, 395, 396, 399, 399, 399, 400, 400, 410, 410, 411, 413, 414, 414, 418, 418,
422, 426, 428, 428, 429, 435, 435, 436, 437, 442, 452, 456, 469, 480, 491, 504,
522, 537, 558, 561, 566, 569, 573, 577, 577, 580, 580, 582, 584, 585, 586, 587,
587, 588, 588, 588, 589, 589, 589, 589, 590, 593, 593, 593, 593, 593, 593, 593,
593, 593, 593, 593, 593, 593, 593, 593), ('Chongqing', 'China', 30.0572,
107.874, 6, 9, 27, 57, 75, 110, 132, 147, 182, 211, 247, 300, 337, 366, 389,
411, 426, 428, 468, 486, 505, 518, 529, 537, 544, 551, 553, 555, 560, 567, 572,
573, 575, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576,
576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 576, 577, 578, 578,
578, 578, 578, 578, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579,
579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579, 579,
579, 579, 579, 579, 579), ('Fujian', 'China', 26.0789, 117.9874, 1, 5, 10, 18,
35, 59, 80, 84, 101, 120, 144, 159, 179, 194, 205, 215, 224, 239, 250, 261, 267,
272, 279, 281, 285, 287, 290, 292, 293, 293, 293, 293, 293, 293, 294, 294, 296,
296, 296, 296, 296, 296, 296, 296, 296, 296, 296, 296, 296, 296, 296, 296,
296, 296, 296, 296, 296, 299, 303, 313, 313, 318, 322, 328, 331, 337, 338, 340,
343, 345, 345, 349, 350, 350, 350, 351, 351, 351, 351, 351, 352, 352, 353, 353,
353, 354, 355, 355, 355, 355, 355, 355, 355, 355, 355, 355, 355, 355, 355),
('Gansu', 'China', 37.8099, 101.0583, 0, 2, 2, 4, 7, 14, 19, 24, 26, 29, 40, 51,
55, 57, 62, 62, 67, 79, 83, 83, 86, 87, 90, 90, 90, 90, 91, 91, 91, 91, 91, 91,
91, 91, 91, 91, 91, 91, 91, 91, 91, 91, 91, 102, 119, 120, 124, 124, 125, 127,
127, 127, 129, 133, 133, 133, 133, 134, 134, 134, 136, 136, 136, 136, 136, 136,
136, 138, 138, 138, 138, 138, 138, 138, 138, 138, 139, 139, 139, 139, 139, 139,
139, 139, 139, 139, 139, 139, 139, 139, 139, 139, 139, 139, 139, 139, 139,
139, 139), ('Guangdong', 'China', 23.3417, 113.4244, 26, 32, 53, 78, 111, 151,
207, 277, 354, 436, 535, 632, 725, 813, 895, 970, 1034, 1095, 1131, 1159, 1177,
1219, 1241, 1261, 1294, 1316, 1322, 1328, 1331, 1332, 1333, 1339, 1342, 1345,
1347, 1347, 1347, 1348, 1349, 1349, 1350, 1350, 1350, 1351, 1352, 1352, 1352,
1352, 1353, 1356, 1356, 1356, 1356, 1360, 1361, 1364, 1370, 1378, 1395, 1400,
1413, 1415, 1428, 1433, 1448, 1456, 1467, 1475, 1484, 1494, 1501, 1507, 1514,
1516, 1524, 1532, 1533, 1536, 1539, 1544, 1548, 1552, 1555, 1564, 1566, 1571,
1577, 1579, 1580, 1581, 1582, 1582, 1585, 1585, 1586, 1587, 1587, 1588, 1588,
1588), ('Guangxi', 'China', 23.8298, 108.7881, 2, 5, 23, 23, 36, 46, 51, 58, 78,
87, 100, 111, 127, 139, 150, 168, 172, 183, 195, 210, 215, 222, 222, 226, 235,
237, 238, 242, 244, 245, 246, 249, 249, 251, 252, 252, 252, 252, 252, 252,
252, 252, 252, 252, 252, 252, 252, 252, 252, 252, 252, 252, 253, 253,
253, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254,
254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254,
254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254, 254), ('Guizhou', 'China',
26.8154, 106.8748, 1, 3, 3, 4, 5, 7, 9, 9, 12, 29, 29, 38, 46, 58, 64, 71, 81,
89, 99, 109, 127, 133, 135, 140, 143, 144, 146, 146, 146, 146, 146, 146,
146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146,
146, 146, 146, 146, 146, 146, 147, 146, 146, 146, 146, 146, 146, 146, 146,
146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146, 146,

146, 146, 146, 146, 146, 146, 146, 147, 147, 147, 147, 147, 147, 147, 147, 147,
 147, 147, 147), ('Hainan', 'China', 19.1959, 109.7453, 4, 5, 8, 19, 22, 33, 40,
 43, 46, 52, 62, 64, 72, 80, 99, 106, 117, 124, 131, 138, 144, 157, 157, 159,
 162, 162, 163, 163, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168,
 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168,
 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168,
 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168,
 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168, 168), ('Hebei', 'China',
 39.549, 116.1306, 1, 1, 2, 8, 13, 18, 33, 48, 65, 82, 96, 104, 113, 126, 135,
 157, 172, 195, 206, 218, 239, 251, 265, 283, 291, 300, 301, 306, 306, 307, 308,
 309, 311, 311, 311, 312, 317, 318, 318, 318, 318, 318, 318, 318, 318, 318,
 318, 318, 318, 318, 318, 318, 318, 318, 318, 318, 318, 318, 319, 319, 319,
 319, 319, 319, 319, 319, 321, 321, 323, 325, 326, 326, 327, 327, 327, 327, 327,
 327, 327, 327, 327, 327, 327, 328, 328, 328, 328, 328, 328, 328, 328, 328,
 328, 328, 328, 328, 328), ('Heilongjiang', 'China', 47.862, 127.7615, 0, 2, 4,
 9, 15, 21, 33, 38, 44, 59, 80, 95, 121, 155, 190, 227, 277, 295, 307, 331, 360,
 378, 395, 419, 425, 445, 457, 464, 470, 476, 479, 479, 480, 480, 480, 480, 480,
 480, 480, 480, 480, 480, 480, 481, 481, 481, 481, 481, 481, 482, 482, 482, 482,
 482, 482, 482, 482, 483, 484, 484, 484, 484, 484, 484, 484, 484, 484, 484,
 484, 484, 488, 489, 491, 504, 524, 544, 569, 609, 638, 661, 684, 740, 819, 841,
 861, 872, 892, 898, 905, 913, 921, 928, 930, 935, 936, 939, 939, 939, 944),
 ('Henan', 'China', 33.882, 113.614, 5, 5, 9, 32, 83, 128, 168, 206, 278, 352,
 422, 493, 566, 675, 764, 851, 914, 981, 1033, 1073, 1105, 1135, 1169, 1184,
 1212, 1231, 1246, 1257, 1262, 1265, 1267, 1270, 1271, 1271, 1271, 1271, 1272,
 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1272, 1273,
 1273, 1273, 1273, 1273, 1273, 1273, 1273, 1273, 1273, 1273, 1274, 1274, 1274,
 1274, 1275, 1275, 1275, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276,
 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276,
 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276, 1276), ('Hong Kong',
 'China', 22.3, 114.2, 0, 2, 2, 5, 8, 8, 8, 10, 10, 12, 13, 15, 15, 17, 21, 24,
 25, 26, 29, 38, 49, 50, 53, 56, 56, 57, 60, 62, 63, 68, 68, 69, 74, 79, 84, 91,
 92, 94, 95, 96, 100, 100, 105, 105, 107, 108, 114, 115, 120, 126, 129, 134, 140,
 145, 155, 162, 181, 208, 256, 273, 317, 356, 386, 410, 453, 519, 561, 641, 682,
 714, 765, 802, 845, 862, 890, 914, 935, 960, 973, 989, 1000, 1004, 1009, 1012,
 1017, 1017, 1021, 1024, 1025, 1025, 1029, 1033, 1035, 1035, 1037, 1037, 1037,
 1037, 1037, 1037), ('Hubei', 'China', 30.9756, 112.2707, 444, 444, 549, 761,
 1058, 1423, 3554, 3554, 4903, 5806, 7153, 11177, 13522, 16678, 19665, 22112,
 24953, 27100, 29631, 31728, 33366, 33366, 48206, 54406, 56249, 58182, 59989,
 61682, 62031, 62442, 62662, 64084, 64084, 64287, 64786, 65187, 65596, 65914,
 66337, 66907, 67103, 67217, 67332, 67466, 67592, 67666, 67707, 67743, 67760,
 67773, 67781, 67786, 67790, 67794, 67798, 67799, 67800, 67800, 67800, 67800,
 67800, 67800, 67801, 67801, 67801, 67801, 67801, 67801, 67801, 67801, 67802,
 67802, 67802, 67803, 67803, 67803, 67803, 67803, 67803, 67803, 67803, 67803,
 67803, 67803, 67803, 67803, 68128, 68128, 68128, 68128, 68128, 68128, 68128,
 68128, 68128, 68128, 68128, 68128, 68128, 68128), ('Hunan', 'China', 27.6104,
 111.7088, 4, 9, 24, 43, 69, 100, 143, 221, 277, 332, 389, 463, 521, 593, 661,
 711, 772, 803, 838, 879, 912, 946, 968, 988, 1001, 1004, 1006, 1007, 1008, 1010,
 1011, 1013, 1016, 1016, 1016, 1016, 1017, 1017, 1018, 1018, 1018, 1018, 1018,

[illegible]

[illegible]

114, 116, 116, 117, 122, 123, 124, 126, 130, 131), (None, 'Fiji', -17.7134,
 178.065, 0,
 0,
 0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 3, 4, 5, 5, 5, 5, 5, 5, 5, 5, 7, 7, 12, 12, 14,
 15, 15, 15, 16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 18, 18, 18, 18, 18, 18, 18,
 18, 18, 18, 18), (None, 'Finland', 64.0, 26.0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
 1, 2, 2, 2,
 3, 6, 6, 6, 6, 12, 15, 15, 23, 30, 40, 59, 59, 155, 225, 244, 277, 321, 336,
 400, 450, 523, 626, 700, 792, 880, 958, 1041, 1167, 1240, 1352, 1418, 1446,
 1518, 1615, 1882, 1927, 2176, 2308, 2487, 2605, 2769, 2905, 2974, 3064, 3161,
 3237, 3369, 3489, 3681, 3783, 3868, 4014, 4129, 4284, 4395, 4475, 4576, 4695,
 4740, 4906, 4995), ('French Guiana', 'France', 3.9339, -53.1258, 0, 0, 0, 0, 0,
 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 5, 5, 5, 5, 5, 5, 5, 7, 11, 11, 11,
 11, 15, 18, 18, 20, 23, 28, 28, 28, 28, 28, 43, 43, 51, 51, 57, 61, 61, 72, 72,
 77, 83, 83, 83, 86, 86, 86, 86, 86, 96, 96, 96, 97, 97, 97, 97, 107, 111, 111, 111,
 111, 125, 125, 126), ('French Polynesia', 'France', -17.6797, 149.4068, 0, 0, 0,
 0,
 0, 3, 3, 3, 3, 3, 3,
 6, 11, 15, 18, 18, 25, 25, 30, 30, 30, 30, 36, 36, 37, 37, 39, 40, 41, 42, 47,
 51, 51, 51, 51, 53, 55, 55, 55, 55, 55, 55, 55, 56, 56, 57, 57, 57, 57, 57, 57,
 58, 58, 58), ('Guadeloupe', 'France', 16.25, -61.5833, 0, 0, 0, 0, 0, 0, 0, 0,
 0,
 0, 1, 1,
 1, 3, 3, 6, 7, 11, 24, 36, 36, 36, 50, 63, 63, 82, 94, 94, 116, 128, 134, 147,
 147, 171, 171, 184, 191, 196, 196, 207, 217, 217, 233, 245, 254, 271, 271, 311,
 326, 326, 354, 380, 401, 401, 460, 460, 539), ('New Caledonia', 'France',
 -20.9043, 165.618, 0,
 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 4, 4, 8, 10, 14, 14, 15, 15, 15, 15, 16, 16,
 18, 18, 17, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18, 18,
 18, 18, 18, 18, 18, 18, 18, 18, 18), ('Reunion', 'France', -21.1351, 55.2471, 0,
 0,
 0, 1, 1,
 1, 3, 3, 6, 7, 11, 24, 36, 36, 36, 50, 63, 63, 82, 94, 94, 116, 128, 134, 147,
 147, 171, 171, 184, 191, 196, 196, 207, 217, 217, 233, 245, 254, 271, 271, 311,
 326, 326, 354, 380, 401, 401, 460, 460, 539), ('Saint Barthelemy',
 'France', 17.9, -62.8333, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 3, 3, 3,
 3, 3, 1, 1, 1, 1, 1, 1, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 5, 5, 5, 6, 6, 6, 6,
 6,
 6), ('St Martin', 'France', 18.0708, -63.0501, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0,

0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 4, 4, 4, 5, 8, 8, 11,
 11, 11, 11, 11, 15, 15, 15, 22, 22, 24, 32, 32, 32, 32, 32, 32, 32, 32, 32, 32,
 35, 35, 35, 37, 37, 37, 37, 38, 38, 38, 38, 38, 38, 38, 38, 38), ('Martinique',
 'France', 14.6415, -61.0242, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0,
 0, 2, 2, 2, 2, 3, 3, 3, 9, 9, 15, 16, 19, 23, 32, 32, 44, 53, 57, 66, 66, 81,
 93, 93, 93, 128, 135, 138, 143, 145, 149, 151, 152, 154, 154, 155, 155, 155,
 157, 157, 158, 158, 158, 158, 163, 163, 163, 164, 164, 170, 175, 175, 175, 175,
 175, 178), (None, 'France', 46.2276, 2.2137, 0, 0, 2, 3, 3, 3, 4, 5, 5, 5, 6, 6,
 6, 6, 6, 6, 6, 11, 11, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12, 12,
 12, 14, 18, 38, 57, 100, 130, 191, 204, 285, 377, 653, 949, 1126, 1209, 1784,
 2281, 2281, 3661, 4469, 4499, 6633, 7652, 9043, 10871, 12612, 14282, 16018,
 19856, 22304, 25233, 29155, 32964, 37575, 40174, 44550, 52128, 56989, 59105,
 64338, 68605, 70478, 74390, 78167, 82048, 86334, 90676, 93790, 120633, 124298,
 129257, 132473, 144944, 146923, 146906, 151808, 154188, 156921, 154715, 157026,
 158636, 160292, 160847, 164589, 167605, 165093, 165764), (None, 'Gabon',
 -0.8037, 11.6094, 0,
 0,
 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 3, 4, 5, 5, 6, 6, 7, 7, 7, 7, 7, 16, 18, 21, 21,
 21, 21, 24, 30, 34, 44, 44, 46, 49, 57, 57, 80, 80, 108, 108, 109, 120, 156,
 166, 167, 172, 176, 176, 211, 238, 276, 276), (None, 'Gambia', 13.4432,
 -15.3101, 0,
 0,
 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4,
 4, 4, 4, 9, 9, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10,
 11), (None, 'Georgia', 42.3154, 43.3569, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 1, 1, 1, 1, 3,
 3, 3, 3, 4, 4, 4, 13, 15, 15, 24, 24, 25, 30, 33, 33, 34, 38, 40, 43, 49, 54,
 61, 70, 75, 79, 83, 90, 91, 103, 110, 117, 134, 155, 162, 174, 188, 196, 211,
 218, 234, 242, 257, 272, 300, 306, 348, 370, 388, 394, 402, 408, 416, 425, 444,
 456, 486, 497, 511, 517, 539), (None, 'Germany', 51.0, 9.0, 0, 0, 0, 0, 0, 1, 4,
 4, 4, 5, 8, 10, 12, 12, 12, 12, 13, 13, 14, 14, 16, 16, 16, 16, 16, 16, 16, 16,
 16, 16, 16, 16, 16, 16, 17, 27, 46, 48, 79, 130, 159, 196, 262, 482, 670, 799,
 1040, 1176, 1457, 1908, 2078, 3675, 4585, 5795, 7272, 9257, 12327, 15320, 19848,
 22213, 24873, 29056, 32986, 37323, 43938, 50871, 57695, 62095, 66885, 71808,
 77872, 84794, 91159, 96092, 100123, 103374, 107663, 113296, 118181, 122171,
 124908, 127854, 130072, 131359, 134753, 137698, 141397, 143342, 145184, 147065,
 148291, 150648, 153129, 154999, 156513, 157770, 158758, 159912, 161539, 163009),
 (None, 'Ghana', 7.9465, -1.0232, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 6, 6, 7, 7, 11, 16, 19, 23, 27, 53, 93, 132, 137,
 141, 152, 152, 161, 195, 204, 205, 205, 214, 214, 287, 313, 378, 378, 408, 566,
 566, 636, 636, 641, 641, 834, 1042, 1042, 1042, 1154, 1154, 1279, 1279, 1550,
 1550, 1671, 1671, 2074), (None, 'Greece', 39.0742, 21.8243, 0, 0, 0, 0, 0, 0, 0,
 0,
 0, 1, 3, 4, 4, 7, 7, 7, 9, 31, 45, 46, 73, 73, 89, 99, 99, 190, 228, 331, 331,
 387, 418, 418, 495, 530, 624, 695, 743, 821, 892, 966, 1061, 1156, 1212, 1314,
 1415, 1544, 1613, 1673, 1735, 1755, 1832, 1884, 1955, 2011, 2081, 2114, 2145,

[illegible]

[illegible]

205, 400, 598, 702, 996, 1090, 1221, 1333, 1463, 1550, 1746, 1914, 2118, 2385,
 2621, 2863, 3084, 3369, 3755, 4015, 4284, 4445, 4641, 4863, 5147, 5370, 5550,
 5687, 5865, 6086, 6086, 6211, 6314, 6409, 6525, 6603, 6623, 6740, 6896, 6937,
 7036, 7078, 7156, 7191, 7338, 7401, 7463, 7499, 7527, 7599, 7660, 7710, 7738),
 (None, 'Oman', 21.0, 57.0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 4, 4, 4, 6, 6, 6, 12, 15, 16,
 16, 16, 16, 16, 18, 18, 18, 19, 19, 22, 22, 24, 39, 48, 48, 52, 55, 66, 84, 99,
 109, 131, 152, 167, 179, 192, 210, 231, 252, 277, 298, 331, 371, 419, 457, 484,
 546, 599, 727, 813, 910, 1019, 1069, 1180, 1266, 1410, 1508, 1614, 1716, 1790,
 1905, 1998, 2049, 2131, 2274, 2348), (None, 'Pakistan', 30.3753, 69.3451, 0, 0,
 0,
 0, 0, 0, 0, 0, 0, 2, 2, 2, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 16, 19, 20, 28, 31, 53,
 136, 236, 299, 454, 501, 730, 776, 875, 972, 1063, 1201, 1373, 1495, 1597, 1717,
 1938, 2118, 2421, 2686, 2818, 3157, 3766, 4035, 4263, 4489, 4695, 5011, 5230,
 5496, 5837, 6383, 6919, 7025, 7638, 8348, 8418, 9565, 10076, 11155, 11940,
 12723, 13328, 13915, 14612, 15525, 16817), (None, 'Panama', 8.538, -80.7821, 0,
 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 8, 11, 27, 36,
 43, 55, 69, 86, 109, 137, 200, 313, 345, 345, 443, 558, 674, 786, 901, 989,
 1181, 1181, 1317, 1475, 1673, 1801, 1988, 2100, 2249, 2528, 2752, 2974, 3234,
 3400, 3472, 3574, 3751, 4016, 4210, 4273, 4467, 4658, 4821, 5166, 5338, 5538,
 5779, 6021, 6021, 6378, 6532), (None, 'Papua New Guinea', -6.315, 143.9555, 0,
 0,
 0,
 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2,
 2, 2, 2, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8), (None, 'Paraguay',
 -23.4425, -58.4438, 0,
 0, 1,
 1, 1, 5, 5, 6, 6, 6, 8, 9, 11, 11, 13, 18, 22, 22, 27, 37, 41, 52, 56, 59, 64,
 65, 69, 77, 92, 96, 104, 113, 115, 119, 124, 129, 133, 134, 147, 159, 161, 174,
 199, 202, 206, 208, 208, 213, 213, 223, 228, 228, 228, 239, 239, 266), (None,
 'Peru', -9.19, -75.0152, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 1,
 1, 6, 7, 11, 11, 15, 28, 38, 43, 86, 117, 145, 234, 234, 318, 363, 395, 416,
 480, 580, 635, 671, 852, 950, 1065, 1323, 1414, 1595, 1746, 2281, 2561, 2954,
 4342, 5256, 5897, 6848, 7519, 9784, 10303, 11475, 12491, 13489, 14420, 15628,
 16325, 17837, 19250, 20914, 21648, 25331, 27517, 28699, 31190, 33931, 36976),
 (None, 'Philippines', 13.0, 122.0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2,
 2, 3,
 3, 3, 5, 6, 10, 20, 33, 49, 52, 64, 111, 140, 142, 187, 202, 217, 230, 307, 380,
 462, 552, 636, 707, 803, 1075, 1418, 1546, 2084, 2311, 2633, 3018, 3094, 3246,
 3660, 3764, 3870, 4076, 4195, 4428, 4648, 4932, 5223, 5453, 5660, 5878, 6087,
 6259, 6459, 6599, 6710, 6981, 7192, 7294, 7579, 7777, 7958, 8212, 8488), (None,
 'Poland', 51.9194, 19.1451, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 1, 1,
 5, 5, 11, 16, 22, 31, 49, 68, 103, 119, 177, 238, 251, 355, 425, 536, 634, 749,
 901, 1051, 1221, 1389, 1638, 1862, 2055, 2311, 2554, 2946, 3383, 3627, 4102,
 4413, 4848, 5205, 5575, 5955, 6356, 6674, 6934, 7202, 7582, 7918, 8379, 8742,

[illegible]

[illegible]

[illegible]

3 BUILDING PREDICTIVE MODEL AND TESTING : MULTIPLE LINEAR REGRESSION

```
[ ]: #Import required libraries
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

3.1 Object Oriented Programming for Predictive Modelling

```
[ ]: class LinearRegressionModel:
    def __init__(self, data):
        self.data = data
        self.x = data[['GDP per capita', 'Social support', 'Healthy life_
↪expectancy', 'Freedom to make life choices']]
        self.y = data['Total Infections']
        self.x_train, self.x_test, self.y_train, self.y_test =
↪train_test_split(self.x, self.y, test_size=0.2, random_state=100)
        self.mlr = LinearRegression()

    def train_model(self):
        self.mlr.fit(self.x_train, self.y_train)

    def print_coefficients(self):
        print("Intercept Value:", self.mlr.intercept_)
        # pair the feature names with the coefficients
        print("Other Coefficients:", list(zip(self.x.columns, self.mlr.coef_)))

    def predict_test_set(self):
        y_pred_mlr = self.mlr.predict(self.x_test)
        print("Prediction for test set:", y_pred_mlr)

    def print_actual_vs_predicted(self):
        mlr_diff = pd.DataFrame({'Actual value': self.y_test, 'Predicted value':
↪ self.mlr.predict(self.x_test)})
        print("\n-----Actual vs Predicted Value-----")
        print(mlr_diff)

    def predict_for_values(self, values):
        pred_vals = self.mlr.predict([values])
        print(f'R Predicted Value: {pred_vals}')

    def print_r_squared_value(self):
        r_squared = self.mlr.score(self.x, self.y) * 100
        print('R squared value of the model: {:.2f}'.format(r_squared))
```

```

def print_error_metrics(self):
    y_pred_mlr = self.mlr.predict(self.x_test)
    mean_absolute_error = metrics.mean_absolute_error(self.y_test,
    ↪y_pred_mlr)
    mean_squared_error = metrics.mean_squared_error(self.y_test, y_pred_mlr)
    root_mean_squared_error = np.sqrt(mean_squared_error)

    print('Mean Absolute Error:', mean_absolute_error)
    print('Mean Square Error:', mean_squared_error)
    print('Root Mean Square Error:', root_mean_squared_error)

linear_reg_model = LinearRegressionModel(data)
linear_reg_model.train_model()
linear_reg_model.print_coefficients()
linear_reg_model.predict_test_set()
linear_reg_model.print_actual_vs_predicted()
linear_reg_model.predict_for_values([0.9, 0.6, 0.55, 0.78])
linear_reg_model.print_r_squared_value()
linear_reg_model.print_error_metrics()

```

Intercept Value: -134524.67717574624

Other Coefficients: [('GDP per capita', 272329.25260819617), ('Social support', -732709.7723702246), ('Healthy life expectancy', 1724646.3974207772), ('Freedom to make life choices', -306889.91130496183)]

Prediction for test set: [516627.33696103 549742.57482286 627047.76705019 519714.3196811

-280412.86639126 655140.23551356 382076.80776309 406788.21951885
 422413.86373857 688100.71152579 330893.86359747 -215106.22503947
 290206.00985003 731025.51202748 681229.93838944 -93601.014
 485980.81944443 809524.51378545 848613.27182111 314134.25214808
 539529.81085346 212800.30364683 79281.35781691 52911.35873529
 414019.42131276 496686.27698456 737916.99081618 576726.12682345
 699436.27084722]

-----Actual vs Predicted Value-----

	Actual value	Predicted value
Poland	251454	516627.336961
Kuwait	55693	549742.574823
Bosnia and Herzegovina	36234	627047.767050
Slovakia	31703	519714.319681
Mozambique	1118	-280412.866391
Chile	282177	655140.235514
Vietnam	10708	382076.807763
Colombia	107654	406788.219519
Venezuela	8141	422413.863739
Belgium	1082648	688100.711526

Rwanda	4844	330893.863597
Zimbabwe	671	-215106.225039
Ukraine	140485	290206.009850
Morocco	75625	731025.512027
United Kingdom	3206716	681229.938389
Guinea	16059	-93601.014000
Sri Lanka	10572	485980.819444
Georgia	11027	809524.513785
Italy	6139613	848613.271821
Tajikistan	15	314134.252148
Mexico	237643	539529.810853
Comoros	1	212800.303647
Tanzania	4738	79281.357817
Kyrgyzstan	13847	52911.358735
Lithuania	37109	414019.421313
Ecuador	337630	496686.276985
Tunisia	25532	737916.990816
Lebanon	24136	576726.126823
Norway	251779	699436.270847

R Predicted Value: [380127.17451305]

R squared value of the model: 11.24

Mean Absolute Error: 610685.4550839054

Mean Square Error: 1343126647727.9663

Root Mean Square Error: 1158933.4095313528

```
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does
not have valid feature names, but LinearRegression was fitted with feature names
warnings.warn(
```

3.2 Interpretation:

- Intercept: The intercept represents the predicted value of the dependent variable when all predictor variables are zero. In your case, it's -134524.68.
- GDP per capita coefficient: A one-unit increase in GDP per capita is associated with an increase of approximately 272329.25 in the dependent variable.
- Social support coefficient: A one-unit increase in social support is associated with a decrease of approximately 732709.77 in the dependent variable.
- Healthy life expectancy coefficient: A one-unit increase in healthy life expectancy is associated with an increase of approximately 1724646.40 in the dependent variable.
- Freedom to make life choices coefficient: A one-unit increase in freedom to make life choices is associated with a decrease of approximately 306889.91 in the dependent variable.
- R Predicted Value: The predicted value of the dependent variable for the entire test set is 380127.17.
- R squared value: The R squared value of 11.24% indicates that approximately 11.24% of the variance in the dependent variable is explained by the independent variables in your model.

- Mean Absolute Error (MAE): The average absolute difference between the predicted and actual values is 610685.46.
- Mean Square Error (MSE): The average squared difference between the predicted and actual values is 1343126647727.97.
- Root Mean Square Error (RMSE): The square root of MSE is 1158933.41, representing the average magnitude of the errors in the predicted values.

4 CONCLUSION

In summary, the study on how COVID-19 rates relate to World happiness index shows that the response variable is very slightly positively dependent on chosen predictor variables and hence the model doesn't explain much (low R squared at 11.24%). The model's predictions differ a lot from the actual values, suggesting it might not be the best fit for our data. This could be because:

- Model not capturing genuine predictor relationships.
- Predictions stray, compromising accurate dependent variable forecasts.
- Negative predictions, a sign of model instability.
- High coefficient magnitudes suggest potential multicollinearity issues.
- Outliers significantly disrupt regression coefficients and prediction accuracy.