GDP per Capita vs Life Expectancy

BY ERICH SKROTZKI

Choosing a Database

Finding a proper database was my first main goal for this project. After going through multiple different sites I decided to us the World Bank dataset.

I chose The World Bank because it was simple to use, easy to download, and had a multitude of datasets to use. These datasets ranged in categories from Agriculture Development and Aid to Infrastructure and Trade.

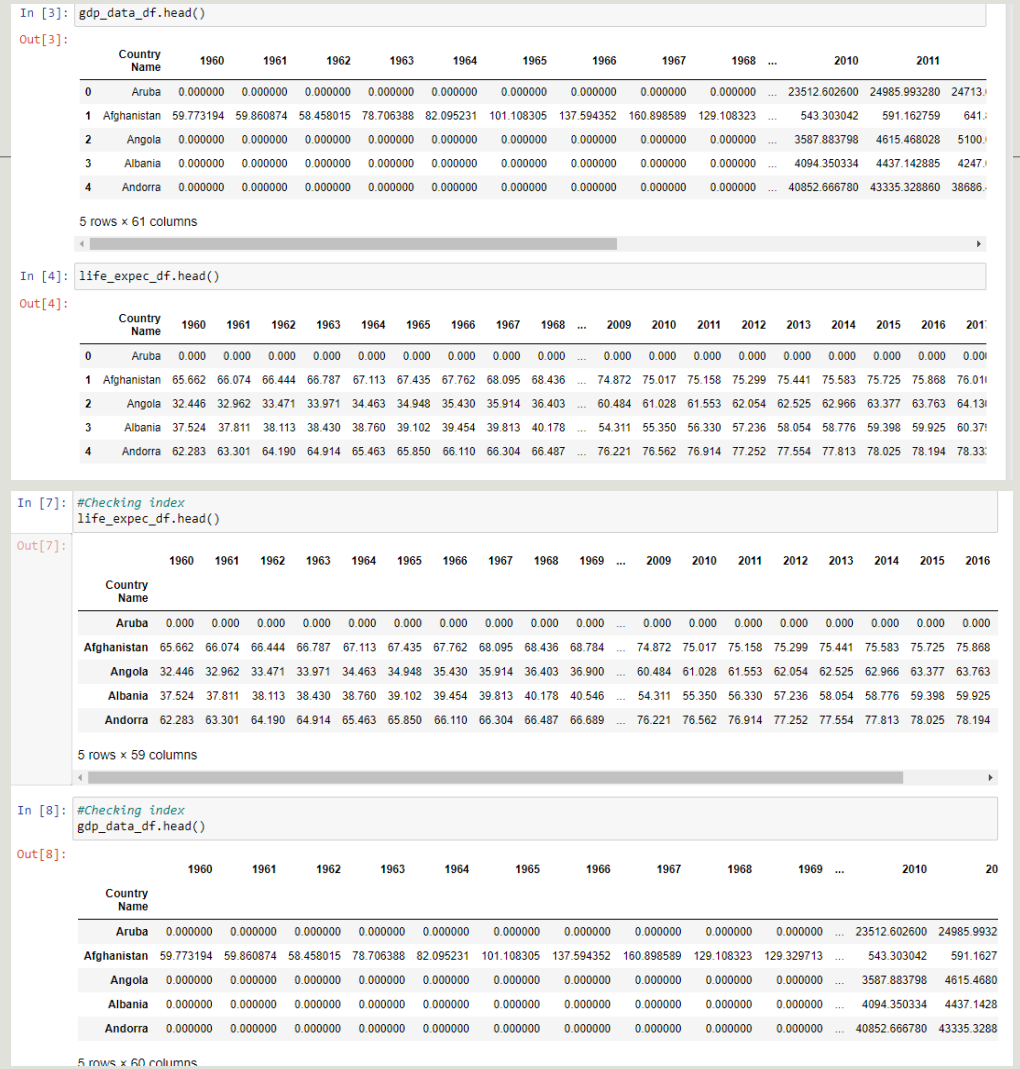
Using data from the World Bank database I decided to see if there was a correlation between GDP per Capita and Life Expectancy for the world.

Trimming the Data

The World Bank had data from 1960 to 2019 for every country.

After gathering the data, first order of business was to clean the empty cells up using fillna to fill each NaN value with a 0.

After, I set the index as the country name so that the dataset looked better and was easier to read.



```
In [3]: gdp_data_df.head()
```

Out[3]:

	Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	2010	2011	
0	Aruba	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	23512.602600	24985.993280	24713.
1	Afghanistan	59.773194	59.860874	58.458015	78.706388	82.095231	101.108305	137.594352	160.898589	129.108323	...	543.303042	591.162759	641.
2	Angola	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	3587.883798	4615.468028	5100.
3	Albania	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	4094.350334	4437.142885	4247.
4	Andorra	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	40852.666780	43335.328860	38686.

5 rows × 61 columns

```
In [4]: life_expec_df.head()
```

Out[4]:

	Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	...	2009	2010	2011	2012	2013	2014	2015	2016	201
0	Aruba	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00
1	Afghanistan	65.662	66.074	66.444	66.787	67.113	67.435	67.762	68.095	68.436	...	74.872	75.017	75.158	75.299	75.441	75.583	75.725	75.868	76.01
2	Angola	32.446	32.962	33.471	33.971	34.463	34.948	35.430	35.914	36.403	...	60.484	61.028	61.553	62.054	62.525	62.966	63.377	63.763	64.13
3	Albania	37.524	37.811	38.113	38.430	38.760	39.102	39.454	39.813	40.178	...	54.311	55.350	56.330	57.236	58.054	58.776	59.398	59.925	60.37
4	Andorra	62.283	63.301	64.190	64.914	65.463	65.850	66.110	66.304	66.487	...	76.221	76.562	76.914	77.252	77.554	77.813	78.025	78.194	78.33

```
In [7]: #Checking index
        life_expec_df.head()
```

Out[7]:

Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2009	2010	2011	2012	2013	2014	2015	2016
Aruba	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Afghanistan	65.662	66.074	66.444	66.787	67.113	67.435	67.762	68.095	68.436	68.784	...	74.872	75.017	75.158	75.299	75.441	75.583	75.725	75.868
Angola	32.446	32.962	33.471	33.971	34.463	34.948	35.430	35.914	36.403	36.900	...	60.484	61.028	61.553	62.054	62.525	62.966	63.377	63.763
Albania	37.524	37.811	38.113	38.430	38.760	39.102	39.454	39.813	40.178	40.546	...	54.311	55.350	56.330	57.236	58.054	58.776	59.398	59.925
Andorra	62.283	63.301	64.190	64.914	65.463	65.850	66.110	66.304	66.487	66.689	...	76.221	76.562	76.914	77.252	77.554	77.813	78.025	78.194

5 rows × 59 columns

```
In [8]: #Checking index
        gdp_data_df.head()
```

Out[8]:

Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2010	20
Aruba	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	23512.602600	24985.9932
Afghanistan	59.773194	59.860874	58.458015	78.706388	82.095231	101.108305	137.594352	160.898589	129.108323	129.329713	...	543.303042	591.1627
Angola	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	3587.883798	4615.4680
Albania	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	4094.350334	4437.1428
Andorra	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	40852.666780	43335.3288

5 rows × 60 columns

More Trimming

Since the data I was going for was in Years and the US Dollar ($) I decided to round both datasets to 2 decimal places and use the .applymap function to apply the accounting number format to my GDP data. This gave me the dataframe shown to the right.

```
In [13]: #Cleaning Dataframe
         gdp_data_clean_df = gdp_data_df.applymap('${:.2f}'.format)
         gdp_data_clean_df.head()
```

Out[13]:

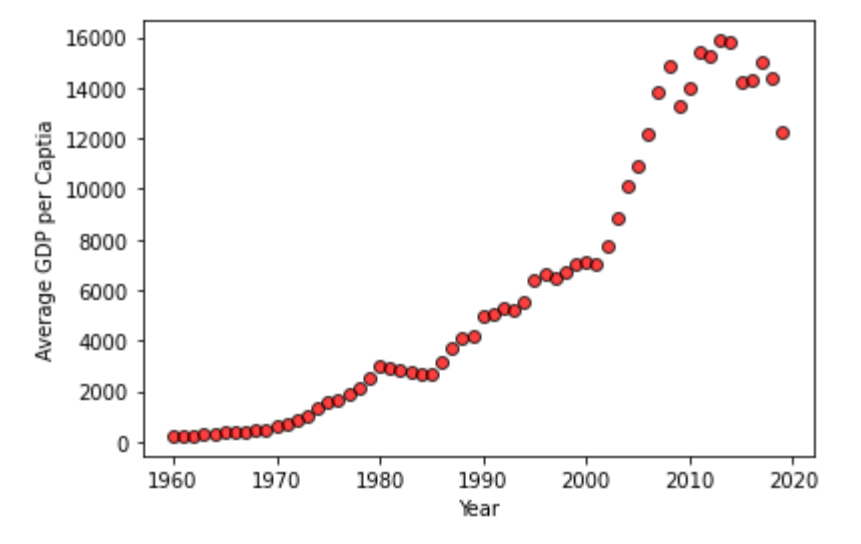
Country Name	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	...	2010	2011	2012	2013	2014	
Aruba	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$23512.60	$24985.99	$24713.70	$26189.44	$26647.94	$27
Afghanistan	$59.77	$59.86	$58.46	$78.71	$82.10	$101.11	$137.59	$160.90	$129.11	$129.33	...	$543.30	$591.16	$641.87	$637.17	$613.86	$
Angola	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$3587.88	$4615.47	$5100.10	$5254.88	$5408.41	$4
Albania	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$4094.35	$4437.14	$4247.63	$4413.06	$4578.63	$3
Andorra	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$40852.67	$43335.33	$38686.46	$39538.77	$41303.93	$35

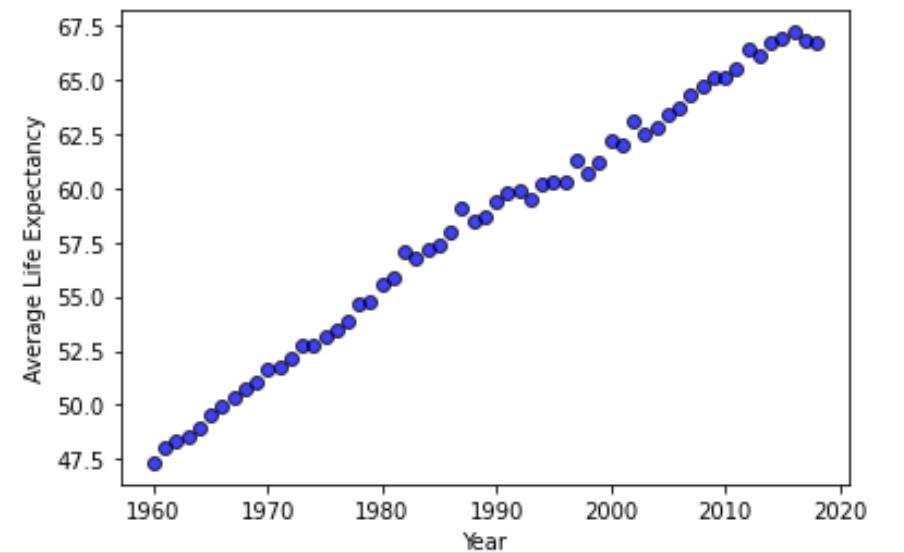
5 rows x 60 columns

Plotting the Data

The main point of the project was to see if there was a correlation between the Average Life Expectancy of the world and GDP per Capita of the world.

Below are the scatter plots for average GDP per Capita and Average Life Expectancy. Looking at the data, there seems to be a correlation between Average GDP per Capita and Life Expectancy.





Finding a Correlation

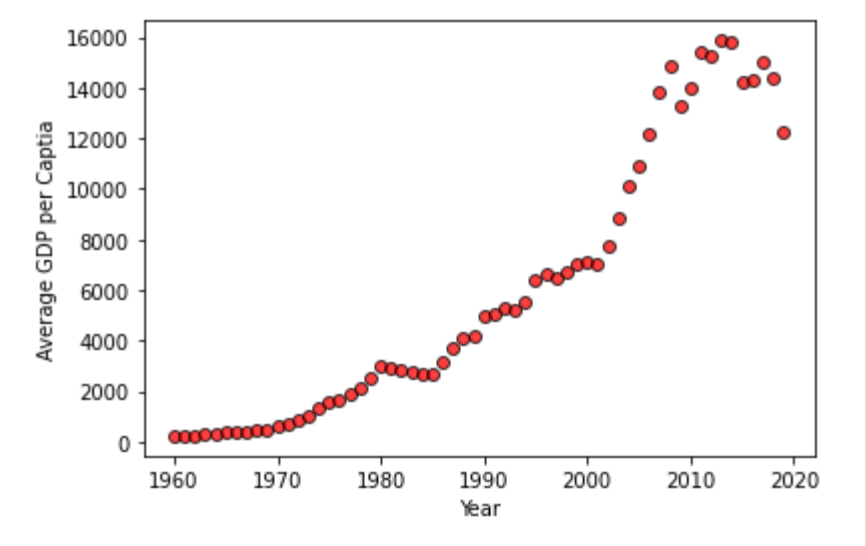
When digesting data and information, you have to also look at what brings these changes to light.

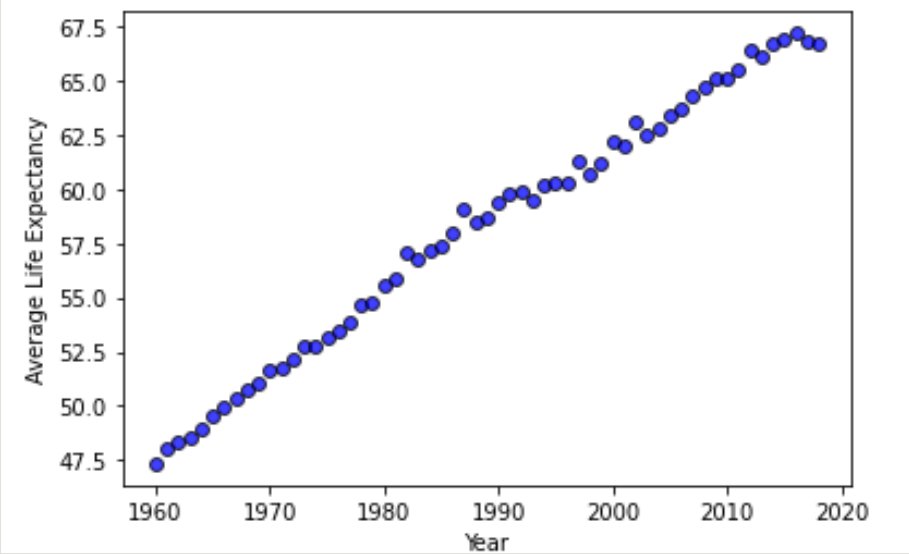
When you think about the basic jist of my project, as GDP of a country goes up, so do Government Services. Examples of these are education, roads, government subsidies and grants, and most importantly healthcare.

As the access and the improvements to healthcare in infrastructure increase people are more likely to live longer, as well the average age of the population increases.

Average GDP per Capita vs Life Expectancy

In the graphs below, as we see an increase in GDP, we also see an increase in life expectancy. This means that there is a correlation between both datasets!





Possibilities of Data

In my dataset alone, there are many possibilities of how to interpret the data. I could determine growth between certain countries, regions or continents vs other continents.

The possibilities are endless.