Relation between Gross Domestic Product (GDP) per capita and life satisfaction of different countries

Damian Ejlli

1 Introduction

Happiness is probably the most important thing that many people aim to achieve in life but what are the factors that contribute to it is not immediately known. In fact, happiness very likely depends on several factors that contribute to it and some of these factors are more important than the others. Obviously happiness is very subjective and there are many ways to quantify it and one of the most important quantifiers of happiness is the perceived life satisfaction of a group of people. In this article I want to study what is the relation between the Gross Domestic Product (GPD) and the life satisfaction of a given country.

2 Data file loading and overview

To study the relation between GPD and life satisfaction, first is needed to have the tabular data of the GPD and life satisfaction of different countries and second it is necessary to prepare the data for the analysis. The tabular data of the GPD and life satisfaction can be easily found on internet respectively in the (IMF website, GPD per capita in US dollars) and (OECD website, Better Life Index). Both these tabular data files are provided with the Python code for analysis. Before starting the analysis, it is very important to have a look at the data present in the files in the tabular form by using for example Google Sheets. In Fig. 1, a section of the Better Life Index aggregate data on Google Sheet is shown. The file contains different information regarding Better Life Index but I am mostly interested in some of these data. I am interest in the column A that gives the country name, column D that gives the indicator of better life which in our case is (Life satisfaction), column G that gives inequality measurement and column O that gives the numeric value of each better life index present in column D.

In Fig. 2, a section of the Better Life Index is shown where in the column D we have the desired category of index, namely that of Life satisfaction index. The first data appears in the row nr. 1815 with Australia being the first country. If we roll down little with the file, we can see that there are data of the Life satisfaction for men, women and total (women+men). Here I work only with the total data which have a corresponding average score (columns I and J). These data extend from row nr. 1815 to 1853 and will be the data that I will use in what follows. The Life satisfaction data are displayed in the score scale from 0 to 10 where the value of 10 means extremely satisfied while the score value of 0 means extremely unsatisfied.

A1 LOCATION

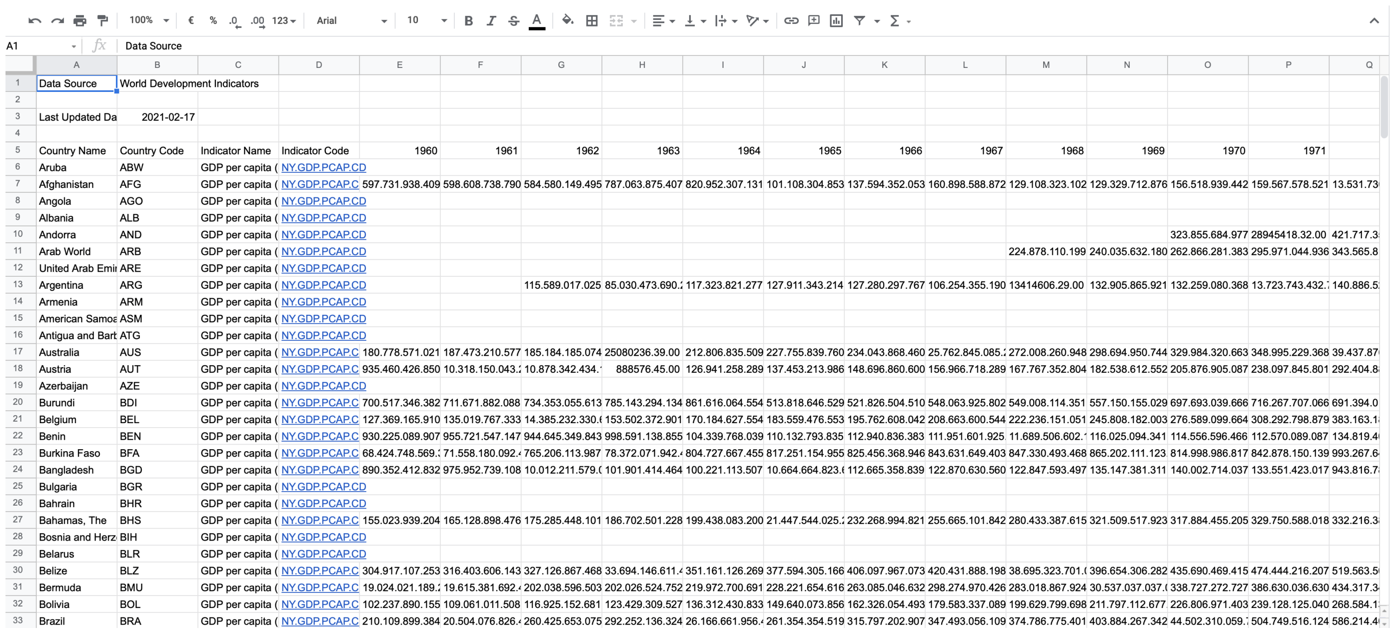
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	LOCATION	Country	INDICATOR	Indicator	MEASURE	Measure	INEQUALITY	Inequality	Unit Code	Unit	PowerCode Cod	PowerCode	Reference Perioc	Reference Perioc	Value	Flag Codes	Flags
2	AUS	Australia	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			5.4		
3	AUT	Austria	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.5		
4	BEL	Belgium	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.7		
5	CAN	Canada	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			6		
6	CZE	Czech Republic	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.1		
7	DNK	Denmark	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			4.2		
8	FIN	Finland	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.9		
9	FRA	France	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			7.6		
10	DEU	Germany	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			2.7		
11	GRC	Greece	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			29.8		
12	HUN	Hungary	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			4.7		
13	ISL	Iceland	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			0.7		
14	IRL	Ireland	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			7.8		
15	ITA	Italy	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			12.3		
16	JPN	Japan	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			1.4		
17	KOR	Korea	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			2.6		
18	LUX	Luxembourg	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			1.7		
19	MEX	Mexico	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			5.5		
20	NLD	Netherlands	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			4.8		
21	NZL	New Zealand	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			4.7		
22	POL	Poland	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			5.7		
23	PRT	Portugal	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			10		
24	SVK	Slovak Republic	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			9.9		
25	ESP	Spain	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			23.1		
26	SWE	Sweden	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.2		
27	TUR	Turkey	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			12.5		
28	GBR	United Kingdom	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			4.5		
29	USA	United States	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			7.7		
30	CHL	Chile	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			8.7		
31	EST	Estonia	JE_LMIS	Labour market in	L	Value	TOT	Total	PC	Percentage	0	Units			3.8		

Figure 1: Appearance of the first 31 rows of OECD Better Life Index aggregate data on Google Sheets. Here I am interested in the column A data (Country), column D data (Indicator), column G data (INEQUALITY) and column O data (Value).

A1 LOCATION

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1812	LVA	Latvia	HS_SFRH	Self-reported hea	L	Value	LW	Low	PC	Percentage	0	Units			28	
1813	SVN	Slovenia	HS_SFRH	Self-reported hea	L	Value	LW	Low	PC	Percentage	0	Units			53	
1814	OECD	OECD - Total	HS_SFRH	Self-reported hea	L	Value	LW	Low	PC	Percentage	0	Units			61	
1815	AUS	Australia	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.3	
1816	AUT	Austria	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.1	
1817	BEL	Belgium	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.9	
1818	CAN	Canada	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.4	
1819	CZE	Czech Republic	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.7	
1820	DNK	Denmark	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.6	
1821	FIN	Finland	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.6	
1822	FRA	France	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.5	
1823	DEU	Germany	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7	
1824	GRC	Greece	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.4	
1825	HUN	Hungary	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.6	
1826	ISL	Iceland	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.5	
1827	IRL	Ireland	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7	
1828	ITA	Italy	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6	
1829	JPN	Japan	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.9	
1830	KOR	Korea	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.9	
1831	LUX	Luxembourg	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.9	
1832	MEX	Mexico	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.5	
1833	NLD	Netherlands	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.4	
1834	NZL	New Zealand	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.3	
1835	NOR	Norway	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.6	
1836	POL	Poland	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.1	
1837	PRT	Portugal	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.4	
1838	SVK	Slovak Republic	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.2	
1839	ESP	Spain	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.3	
1840	SWE	Sweden	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.3	
1841	CHE	Switzerland	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			7.5	
1842	TUR	Turkey	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			5.5	
1843	GBR	United Kingdom	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.8	
1844	USA	United States	SW_LIFS	Life satisfaction	L	Value	TOT	Total	AVSCORE	Average score	0	Units			6.9	

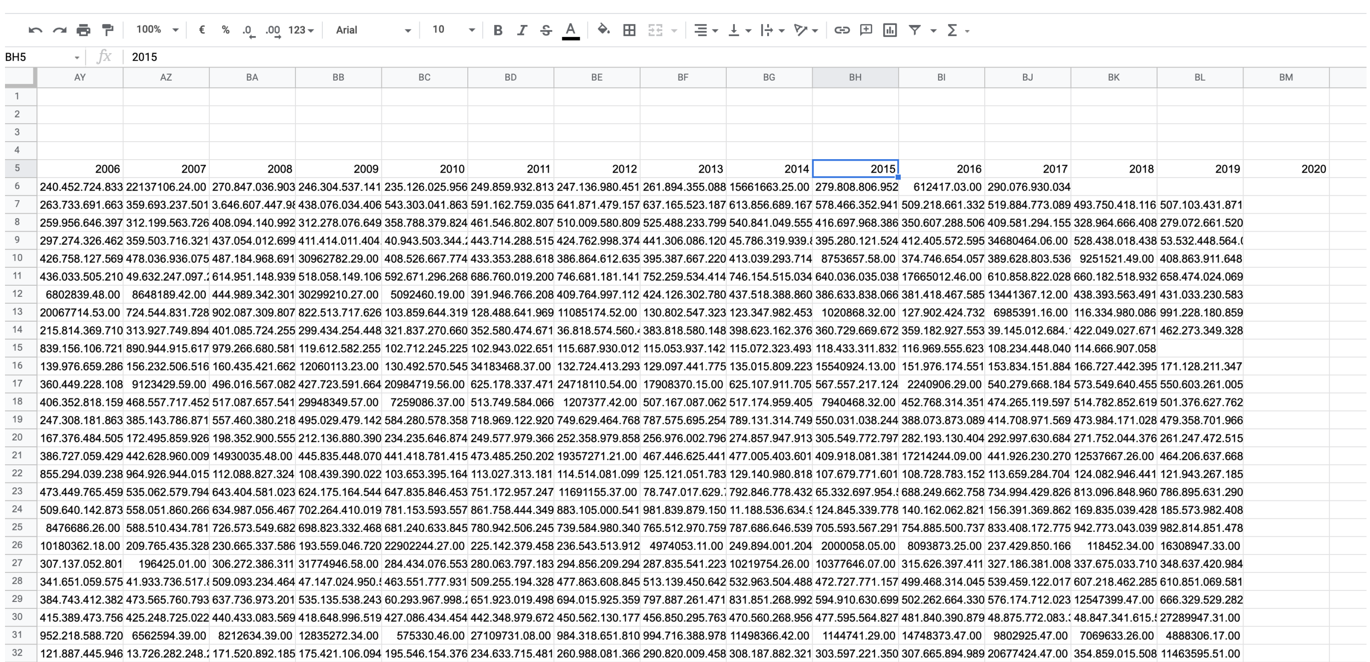
Figure 2: Google sheet section containing the (Life satisfaction) indicator in column D of the Better Life Index aggregate data file.



A1 fx Data Source

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Data Source	World Development Indicators															
2																	
3	Last Updated Da	2021-02-17															
4																	
5	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	
6	Aruba	ABW	GDP per capita (	NY.GDP.PCAP.CD													
7	Afghanistan	AFG	GDP per capita (	NY.GDP.PCAP.C	597.731.938.409	598.608.738.790	584.580.149.495	787.063.875.407	820.952.307.131	101.108.304.853	137.594.352.053	160.898.588.872	129.108.323.102	129.329.712.876	156.518.939.442	159.567.578.521	13.531.73
8	Angola	AGO	GDP per capita (	NY.GDP.PCAP.CD													
9	Albania	ALB	GDP per capita (	NY.GDP.PCAP.CD													
10	Andorra	AND	GDP per capita (	NY.GDP.PCAP.CD											323.855.684.977	28945418.32.00	421.717.3
11	Arab World	ARB	GDP per capita (	NY.GDP.PCAP.CD									224.878.110.199	240.035.632.180	262.866.281.383	295.971.044.936	343.565.8
12	United Arab Emi	ARE	GDP per capita (	NY.GDP.PCAP.CD													
13	Argentina	ARG	GDP per capita (	NY.GDP.PCAP.CD			115.589.017.025	85.030.473.690.	117.323.821.277	127.911.343.214	127.280.297.767	106.254.355.190	13414606.29.00	132.905.865.921	132.259.080.368	13.723.743.432.	140.886.5
14	Armenia	ARM	GDP per capita (	NY.GDP.PCAP.CD													
15	American Samoa	ASM	GDP per capita (	NY.GDP.PCAP.CD													
16	Antigua and Bart	ATG	GDP per capita (	NY.GDP.PCAP.CD													
17	Australia	AUS	GDP per capita (	NY.GDP.PCAP.C	180.778.571.021	187.473.210.577	185.184.185.074	25080236.39.00	212.806.835.509	227.755.839.760	234.043.868.460	25.762.845.085.	272.008.260.948	298.694.950.744	329.984.320.663	348.995.229.368	39.437.87
18	Austria	AUT	GDP per capita (	NY.GDP.PCAP.C	935.460.426.850	10.318.150.043.	10.878.342.434.	888576.45.00	126.941.258.289	137.453.213.986	148.696.860.600	156.966.718.289	167.767.352.804	182.538.612.552	205.876.905.087	238.097.845.801	292.404.8
19	Azerbaijan	AZE	GDP per capita (	NY.GDP.PCAP.CD													
20	Burundi	BDI	GDP per capita (	NY.GDP.PCAP.C	700.517.346.382	711.671.882.088	734.353.055.613	785.143.294.134	861.616.064.554	513.818.646.529	521.826.504.510	548.063.925.802	549.008.114.351	557.150.155.029	697.693.039.666	716.267.707.066	691.394.0
21	Belgium	BEL	GDP per capita (	NY.GDP.PCAP.C	127.369.165.910	135.019.767.333	14.385.232.330.	153.502.372.901	170.184.627.554	183.559.476.553	195.762.608.042	208.663.600.544	222.236.151.051	245.808.182.003	276.589.099.664	308.292.798.879	383.163.1
22	Benin	BEN	GDP per capita (	NY.GDP.PCAP.C	930.225.089.907	955.721.547.147	944.645.349.843	998.591.138.855	104.339.768.039	110.132.793.835	112.940.836.383	111.951.601.925.	11.689.506.602.	116.025.094.341	114.556.596.466	112.570.089.087	134.819.4
23	Burkina Faso	BFA	GDP per capita (	NY.GDP.PCAP.C	68.424.748.569.	71.558.180.092.	765.206.113.987	78.372.071.942.	804.727.667.455	817.251.154.955	825.456.368.946	843.631.649.403	847.330.493.468	865.202.111.123.	814.998.986.817	842.878.150.139	993.267.6
24	Bangladesh	BGD	GDP per capita (	NY.GDP.PCAP.C	890.352.412.832	975.952.739.108	10.012.211.579.	101.901.414.464	100.221.113.507	10.664.664.823.	112.665.358.839	122.870.630.560	122.847.593.497	135.147.381.311	140.002.714.037	133.551.423.017	943.816.7
25	Bulgaria	BGR	GDP per capita (	NY.GDP.PCAP.CD													
26	Bahrain	BHR	GDP per capita (	NY.GDP.PCAP.CD													
27	Bahamas, The	BHS	GDP per capita (	NY.GDP.PCAP.C	155.023.939.204	165.128.898.476	175.285.448.101	186.702.501.228	199.438.083.200	21.447.544.025.	232.268.994.821	255.665.101.842	280.433.387.615	321.509.517.923	317.884.455.205	329.750.588.018	332.216.3
28	Bosnia and Herz	BIH	GDP per capita (	NY.GDP.PCAP.CD													
29	Belarus	BLR	GDP per capita (	NY.GDP.PCAP.CD													
30	Belize	BLZ	GDP per capita (	NY.GDP.PCAP.C	304.917.107.253	316.403.606.143	327.126.867.468	33.694.146.611.	351.161.126.269	377.594.305.166	406.097.967.073	420.431.888.198	38.695.323.701.	396.654.306.282	435.690.469.415	474.444.216.207	519.563.5
31	Bermuda	BMU	GDP per capita (	NY.GDP.PCAP.C	19.024.021.189.	19.615.381.692.	202.038.596.503	202.026.524.752	219.972.700.691	228.221.654.616	263.085.046.632	298.274.970.426	283.018.867.924	30.537.037.037.	338.727.272.727	386.630.036.630	434.317.3
32	Bolivia	BOL	GDP per capita (	NY.GDP.PCAP.C	102.237.890.155	109.061.011.508	116.925.152.681	123.429.309.527	136.312.430.833	149.640.073.856	162.326.054.493	179.583.337.089	199.629.799.698	211.797.112.677.	226.806.971.403	239.128.125.040	268.584.1
33	Brazil	BRA	GDP per capita (	NY.GDP.PCAP.C	210.109.899.384	20.504.076.826.	260.425.653.075	292.252.136.324	26.166.661.956.	261.354.354.519	315.797.202.907	347.493.056.109	374.786.775.401	403.884.267.342	44.502.310.059.	504.749.516.124	586.214.4

Figure 3: Google sheet section containing the GDP per capita in (US dollars) for different countries and years.



BH5 fx 2015

	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM
1															
2															
3															
4															
5	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
6	240.452.724.833	22137106.24.00	270.847.036.903	246.304.537.141	235.126.025.956	249.859.932.813	247.136.980.451	261.894.355.088	15661663.25.00	279.808.806.952	612417.03.00	290.076.930.034			
7	263.733.691.663	359.693.237.501	3.646.607.447.9	438.076.034.406	543.303.041.863	591.162.759.035	641.871.479.157	637.165.523.187	613.856.689.167	578.466.352.941	509.218.661.332	519.884.773.089	493.750.418.116	507.103.431.871	
8	259.956.646.397	312.199.563.726	408.094.140.992	312.278.076.649	358.788.379.824	461.546.802.807	510.009.580.809	525.488.233.799	540.841.049.555	416.697.968.386	350.607.288.506	409.581.294.155	328.964.666.408	279.072.661.520	
9	297.274.326.462	359.503.716.321	437.054.012.699	411.414.011.404.	40.943.503.344.	443.714.288.515	424.762.998.374	441.306.086.120	45.786.319.939.	395.280.121.524	412.405.572.595	34680464.06.00	528.438.018.438	53.532.448.564.	
10	426.758.127.569	478.036.936.075	487.184.968.691	30962782.29.00	408.526.667.774	433.353.288.618	386.864.612.635	395.387.667.220	413.039.293.714	8753657.58.00	374.746.654.057	389.628.803.536	9251521.49.00	408.863.911.648	
11	436.033.505.210	49.632.247.097.	614.951.148.939	518.058.149.106	592.671.296.268	686.760.019.200	746.681.181.141	752.259.534.414	746.154.515.034	640.036.035.038	17665012.46.00	610.858.822.028	660.182.518.932	658.474.024.069	
12	6802839.48.00	8648189.42.00	444.989.342.301	30299210.27.00	5092460.19.00	391.946.766.208	409.764.997.112	424.126.302.780	437.518.388.860	386.633.838.066	381.418.467.585	13441367.12.00	438.393.563.491	431.033.230.583	
13	20067714.53.00	724.544.831.728	902.087.309.807	822.513.717.626	103.859.644.319	128.488.641.969	11085174.52.00	130.802.547.323	123.347.982.453	1020868.32.00	127.902.424.732	6985391.16.00	116.334.980.086	991.228.180.859	
14	215.814.369.710	313.927.749.894	401.085.724.255	299.434.254.448	321.837.270.660	352.580.474.671	36.818.574.560.	383.818.580.148	398.623.162.376	360.729.669.672	359.182.927.553	39.145.012.684.	422.049.027.671	462.273.349.328	
15	839.156.106.721	890.944.915.617	979.266.680.581	119.612.582.255	102.712.245.225	102.943.022.651	115.687.930.012	115.053.937.142	115.072.323.493	118.433.311.832.	116.969.555.623	108.234.448.040	114.666.907.058		
16	139.976.659.286	156.232.506.516	160.435.421.662	12060113.23.00	130.492.570.545	34183468.37.00	132.724.413.293	129.097.441.775	135.015.809.223	15540924.13.00	151.976.174.551	153.834.151.884	166.727.442.395	171.128.211.347	
17	360.449.228.108	9123429.59.00	496.016.567.082	427.723.591.664	20984719.56.00	625.178.337.471	24718110.54.00	17908370.15.00	625.107.911.705	567.557.217.124	2240906.29.00	540.279.668.184	573.549.640.455	550.603.261.005	
18	406.352.818.159	468.557.717.452	517.087.657.541	29948349.57.00	7259086.37.00	513.749.584.066	1207377.42.00	507.167.087.062	517.174.959.405	7940468.32.00	452.768.314.351	474.265.119.597	514.782.852.619	501.376.627.762	
19	247.308.181.863	385.143.786.871	557.460.380.218	495.029.479.142	584.280.578.358	718.969.122.920	749.629.464.768	787.575.695.254	789.131.314.749	550.031.038.244	388.073.873.089	414.708.971.569	473.984.171.028	479.358.701.966	
20	167.376.484.505	172.495.859.926	198.352.900.555	212.136.880.390	234.235.646.874	249.577.979.366	252.358.979.858	256.976.002.796	274.857.947.913	305.549.772.797	282.193.130.404	292.997.630.684	271.752.044.376	261.247.472.515	
21	386.727.059.429	442.628.960.009	14930035.48.00	445.835.448.070	441.418.781.415	473.485.250.202	19357271.21.00	467.446.625.441	477.005.403.601	409.918.081.381	17214244.09.00	441.926.230.270	12537667.26.00	464.206.637.668	
22	855.294.039.238	964.926.944.015	112.088.827.324	108.439.390.022	103.653.395.164	113.027.313.181	114.514.081.099	125.121.051.783	129.140.980.818	107.679.771.601	108.728.783.152	113.659.284.704	124.082.946.441	121.943.267.185	
23	473.449.765.459	535.062.579.794	643.404.581.023	624.175.164.544	647.835.846.453	751.172.957.247	11691155.37.00	78.747.017.629.	792.846.778.432	65.332.697.954.	688.249.662.758	734.994.429.826	813.096.848.960	786.895.631.290	
24	509.640.142.873	558.051.860.266	634.987.056.467	702.264.410.019	781.153.593.557	861.758.444.349	883.105.000.541	981.839.879.150	11.188.536.634.	124.845.339.778	140.162.062.821	156.391.369.862	169.835.039.428	185.573.982.408	
25	8476686.26.00	588.510.434.781	726.573.549.682	698.823.332.468	681.240.633.845	780.942.506.245	739.584.980.340	765.512.970.759	787.686.646.539	705.593.567.291	754.885.500.737	833.408.172.775	942.773.043.039	982.814.851.478	
26	10180362.18.00	209.765.435.328	230.665.337.586	193.559.046.720	22902244.27.00	225.142.379.458	236.543.513.912	4974053.11.00	249.894.001.204	2000058.05.00	8093873.25.00	237.429.850.166	118452.34.00	16308947.33.00	
27	307.137.052.801	196425.01.00	306.272.386.311	31774946.58.00	284.434.076.553	280.063.797.183	294.856.209.294	287.835.541.223	10219754.26.00	10377646.07.00	315.626.397.411	327.186.381.008	337.675.033.710	348.637.420.984	
28	341.651.059.575	41.933.736.517.	509.093.234.464	47.147.024.950.	463.551.777.931	509.255.194.328	477.863.608.845	513.139.450.642	532.963.504.488	472.727.771.157	499.468.314.045	539.459.122.017	607.218.462.285	610.851.069.581	
29	384.743.412.382	473.565.760.793	637.736.973.201	535.135.538.243	60.293.967.998.	651.923.019.498	694.015.925.359	797.887.261.471	831.851.268.992	594.910.630.699	502.262.664.330	576.174.712.023	12547399.47.00	666.329.529.282	
30	415.389.473.756	425.248.725.022	440.433.083.569	418.648.996.519	427.086.434.454	442.348.979.672	450.562.130.177	456.850.295.763	470.560.268.956	477.595.564.827	481.840.390.879	48.875.772.083.	48.847.341.615.	27289947.31.00	
31	952.218.588.720	6562594.39.00	8212634.39.00	12835272.34.00	575330.46.00	27109731.08.00	984.318.651.810	994.716.388.978	11498366.42.00	1144741.29.00	14748373.47.00	9802925.47.00	7069633.26.00	4888306.17.00	
32	121.887.445.946	13.726.282.248.	171.520.892.185	175.421.106.094	195.546.154.376	234.633.715.481	260.988.081.366	290.820.009.458	308.187.882.321	303.597.221.350	307.665.894.989	20677424.47.00	354.859.015.508	11463595.51.00	

Figure 4: Google sheet section containing the GDP per capita in (US dollars) where I select as a matter of example the column corresponding to the year 2015 for my analysis.

In Fig. 3 a section of the data file containing the information of the GPD per capita is (US dollars) for different countries and years in Google sheets is shown. We can see that the data appear quite messy and they overlap with each other. The data sheet contains different data and I will be interested only in some of them. The heading of the GPD per capita file starts at the row nr. 1 in Fig. 3 "Data Source" and "World Development Indicators" appear. In what follows, I will be interested in the data with heading in row nr. 5 which among them are the "Country Name", "Indicator Name" and years (in numerical values). In Fig. 4 a section of the GPD per capita file corresponding to the years from 2006 to 2019 is shown for different countries. As a matter of example, in my analysis I choose[1] the year 2015 which is the selected column in blue in Fig. 4.

3 Data wrangling

In sec. 2, I showed some sections of the Better Life Index and GPD per capita files as they appear when visualized with Google Sheets. In this section, I show how to clean and prepare the data for the analysis by using Python 3. It is important to set since know the goal of my analysis and the way to reach that goal. The goal of my analysis is that to collect and tabulate the data of the life satisfaction index and GDP per capita for each country name and find the relationship (if there is one) between the data.

I start first by importing the libraries and/or modules that I use in my python analysis by using the "Jupyter Notebook":

- ***import*** numpy ***as*** np
- ***import*** matplotlib.pyplot ***as*** plt
- ***import*** pandas ***as*** pd
- ***import*** scipy ***as*** sp
- ***import*** sklearn.neighbors

Here I have imported the NumPy, Matplotlib, Pandas, Scipy and Sklearn modules/libraries and have used the standard abbreviations to call them. As I discussed in sec. 2 both data files contain many data categories that are not needed for this analysis and I select only those data that are necessary to reach my goal. I start first by loading the files in Jupyter Notebook by using Pandas library

- GDP=pd.read_csv("gdp per capita.csv", delimiter=",", header=2)
- LS=pd.read_csv("better life index.csv")

The next step is that of data wrangling by using the Pandas library. In the GDP per capita file I set the "Country name" as the index in the Pandas dataframe and the value column being the " GDP per capita 2015 (USD)". Here I use Pandas to rename some columns and set as dataframe index the column that I am interested in of the original tabular data. Details of these operation are presented in the accompanying Jupyter Notebook file. I use similar data wrangling for the Better Life Index file where

[1] One is free to choose another year for the analysis if wishes so.

I rename some columns and set the country column as index of the Pandas dataframe and column value being the “Life Satisfaction Value”. In Fig. 5 sections of the Pandas dataframes are shown. On the left a section of the “GDP per capita 2015 (USD)” dataframe is shown and on the right only the first ten entries of the “Life Satisfaction Value” dataframe is shown. After data wrangling the “GDP per capita 2015 (USD)” dataframe has 264 country entry values of the GPD per capita as can be seen from the left dataframe in Fig. 5 while the “Life Satisfaction Value” dataframe has in total 40 country value entries of the life satisfaction values.

	GPD per capita 2015 (USD)
Country Name	
Aruba	27980.880695
Afghanistan	578.466353
Angola	4166.979684
Albania	3952.801215
Andorra	35762.523074
...	...
Kosovo	3603.025501
Yemen, Rep.	1602.037841
South Africa	5734.633629
Zambia	1337.795586
Zimbabwe	1445.071062

264 rows × 1 columns

	Life Satisfaction Value
Country Name	
Australia	7.3
Austria	7.1
Belgium	6.9
Canada	7.4
Czech Republic	6.7
Denmark	7.6
Finland	7.6
France	6.5
Germany	7.0
Greece	5.4

Figure 5: Pandas sections containing the “GDP per capita 2015 (USD)” and “Life Satisfaction Value” for different countries is shown.

The next step of the analysis is to make a final Pandas joint dataframe where are selected only those countries that have available values of both “GDP per capita 2015 (USD)” and “Life Satisfaction Value”. This can be done by using the Pandas join function as shown in details in the accompanying Jupyter Notebook file. At the end we obtain the final Pandas joint data frame as shown in Fig. 6 where only a section of the whole dataframe is shown. The whole joint Pandas dataframe has 38 country entries for the “GDP per capita 2015 (USD)” and “Life Satisfaction Value” values where the country names do not necessarily appear in alphabetic order.

4 Data analysis and statistical/machine learning methods

After collecting the data in one dataframe as shown in Fig. 6, the first step to proceed with the analysis is to make a scatter plot and see if there is relation between the data. A scatter plot of the data present in the dataframe of Fig. 6 is shown in Fig. 7 where the feature data (X), namely “ GDP per capita 2015 (USD)” is plotted versus the output data ($Y = f(X)$), namely “Life Satisfaction value”.

4.1 Simple linear regression method (supervised learning)

By looking at the scatter plot in Fig. 7 it seems that the data might follow a linear relationship between them. Indeed, by using the correlation function of the Pandas dataframe, “pd.corr(...)”, a Pearson corre-

Country Name	Life Satisfaction Value	GPD per capita 2015 (USD)
Australia	7.3	56755.721712
Austria	7.1	44178.047378
Belgium	6.9	40991.808138
Canada	7.4	43585.511982
Czech Republic	6.7	17829.698322
Denmark	7.6	53254.856370
Finland	7.6	42784.698362
France	6.5	36638.184929
Germany	7.0	41086.729674
Greece	5.4	18167.773727
Hungary	5.6	12706.891215
Iceland	7.5	52564.429179
Ireland	7.0	61995.422803
Italy	6.0	30230.226302
Japan	5.9	34524.469861
Luxembourg	6.9	101376.496574
Mexico	6.5	9616.645006
Netherlands	7.4	45175.231893
New Zealand	7.3	38615.995185
Norway	7.6	74355.515858

Figure 6: Pandas section of the joint "GDP per capita 2015 (USD)" and "Life Satisfaction Value" dataframes for different countries is shown.

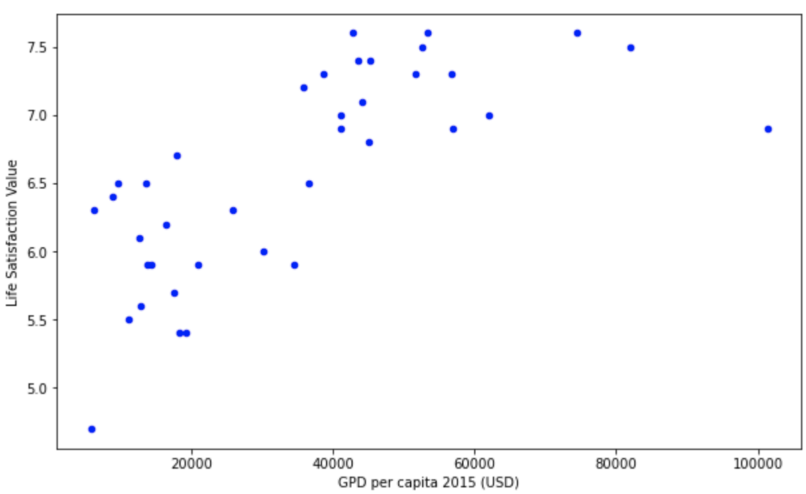


Figure 7: Scatter plot of the "Life Satisfaction Value" vs. " GDP per capita 2015 (USD)" for 38 different countries of the dataframe in Fig. 6 is shown.

lation coefficient of the value $r \simeq 0.72$ is found. Clearly such a value of r indicates that the data might be modelled by a linear relationship of the type $Y = \beta_0 + \beta_1 X$ where β_0 is the intercept value and β_1 is the slope value of the linear equation.

In order to find a possible linear relationship, I make use of the "stats" module of the "SciPy" library[2] and use the linear regression built in method:

- result = sp.stats.linregress(X, Y)
- ***print***(result)
  LinregressResult(slope=2.399629982572962e-05, intercept=5.741754353755319, rvalue=0.7202871953226535, pvalue=3.426556470065171e-07, stderr=3.851624914535906e-06, intercept_stderr=0.15853194959552191)

The linear regression method used gives the values of the intercept $\hat{\beta}_0 \simeq 5.74$ and slope $\hat{\beta}_1 \simeq 2.39 \times 10^{-5}$ with their respective standard errors $s_{\hat{\beta}_0} \simeq 0.15$ and $s_{\hat{\beta}_1} \simeq 3.35 \times 10^{-6}$. The equation for the linear relationship between the data is thus given by the equation line

$$Y(X) \simeq 5.74 + 2.39 \times 10^{-5} X. \tag{1}$$

In addition to the equation line (1) it is also very useful to have the confidence intervals (CIs) for the regression coefficients $\beta_{0,1}$ which are give by

$$\beta_{0,1} \in [\hat{\beta}_{0,1} - s_{\hat{\beta}_{0,1}} t_{1-\alpha/2,n-2}, \hat{\beta}_{0,1} + s_{\hat{\beta}_{0,1}} t_{1-\alpha/2,n-2}],$$

where $t_{1-\alpha/2,n-2}$ is the $1-\alpha/2$ pecentile (or t-score) of the random T variable that enters the Student-T distribution function with $n-2$ degrees of freedom and α is the level of significance. If we ask a test statistic for the T variable at the level of significance of $\alpha = 0.05$ or confidence level (CL) of 95%, we get for $n-2 = 36$ a t-score of $t_{0.975} \simeq 2.028$. Thus at the 95% CL, we get the following CIs for the intercept β_0 and slope β_1 of the OLS linear regression method

$$\beta_0 \in [5.43, 6.04], \qquad \beta_1 \in [1.71 \times 10^{-5}, 3.06 \times 10^{-5}].$$

Other two important parameters derived from the linear regression method are the Pearson correlation coefficient, $r \simeq 0.72$, that fits the data relatively well and the P-value of the parameter β_1 under the null Hypothesis, $H_0 : \beta_1 = 0$. The null Hypothesis is accepted if $P(\beta_1) \geq 1-\alpha$ otherwise it is rejected. Since, $P(\beta_1) \simeq 3.42 \times 10^{-7} \ll 1-\alpha = 0.95$, we thus reject the null hypothesis and consequently there is a relation between the data, a fact that is also confirmed from the value of $r \simeq 0.72$. In Fig. 8 the OLS fit line equation (1) and the original data of Fig. 7 are shown.

With the help of the linear relationship in (1) we can make predictions and inference. For example, in the year 2015 it is not known the life satisfaction value of Albania where the country had a GDP per capita of 3952.8 (USD), see the left dataframe in Fig. 5. By using equation (1) and the GDP per capita of Albania, $X_{\text{Albania}} = 3952.8$ (USD), one can easily find the life satisfaction of Albanians, $Y_{\text{Albania}} \simeq 5.83$. On the other hand, we can use equation (1) for inference purposes. For example, it is not known the GDP per capita in 2015 of the countries of Korea and Russia but are known the values of the life satisfaction. Korea in 2015 had a life satisfaction value of $Y_{\text{Korea}} = 5.9$ while Russia had a life satisfaction value of $Y_{\text{Russia}} = 5.8$. By using equation (1), one can find $X_{\text{Korea}} \simeq 6694.5$ (USD) and $X_{\text{Russia}} \simeq 2510.4$ (USD).

[2] In alternative to the "stats" module one can also use the "sklearn.linear_model" module of the Scikit-learn library.

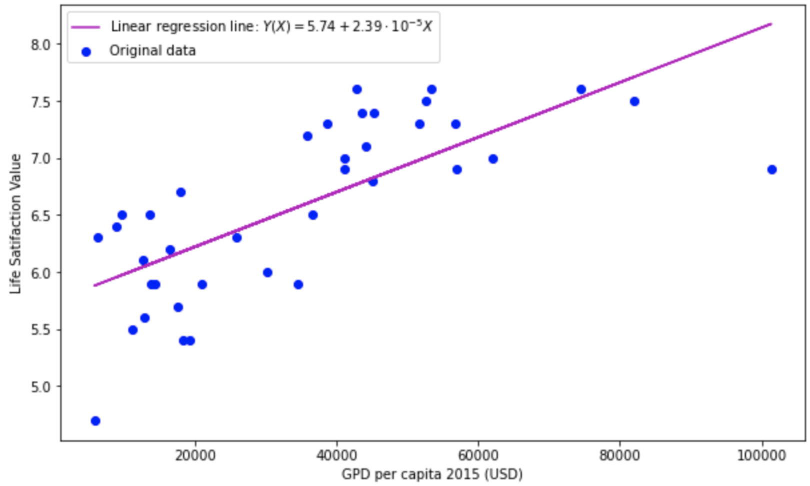


Figure 8: Scatter plot of the "Life Satisfaction Value" vs. " GDP per capita 2015 (USD)" for 38 different countries of the dataframe in Fig. 6 is shown.

4.2 KNN regression method (supervised learning)

In alternative to the simple linear regression method, one can also use the K-nearest neighbour (KNN) regression method. From the dataset of Fig. 6, we can see that many countries have close values of the "Life Satisfaction Value". For example, Australia, Canada, Denmark, Finland, Iceland, Netherland, New Zeland, Norway have respectively life satisfaction values of (7.3, 7.4, 7.6, 7.6, 7.5, 7.4, 7.3, 7.6) and the average value of life satisfaction of these countries is 7.4. The KNN regression method can be implemented in python as follows where the default value of regression neighbors is $K = 5$.

- model=sklearn.neighbors.KNeighborsRegressor()
- model.fit(X, y)
- X_{new}=[[(input new predictor numerical value)]]
- ***print***(model.predict(X_{new}))

Now I can make predictions for new life satisfaction values given new GDP per capita values. For example in our training method, the countries of Albania, United Arab Emirates (UAE) and Armenia were not included in our analysis. These countries had respectively a GDP per capita in the year 2015 (see the whole dataframe on the left in Fig. 5) in USD: $X_{\text{Albania}} \simeq 3952.8$ (USD), $X_{\text{UAE}} \simeq 38663.38$ (USD) and $X_{\text{Armenia}} \simeq 3607.29$ (USD). The KNN regression model predicts the following life satisfaction values: $Y_{\text{Albania}} = 5.88$, $Y_{\text{UAE}} = 6.98$ and $Y_{\text{Armenia}} = 5.88$.

5 Conclusions

In this article I studied the relationship between the GDP per capita of a given country and the perceived life satisfaction of that country. After analysing the data by using two different regression methods, one is able to predict and make inferences for new data not present in the training set. The simple linear regression method, gives a value of the linear correlation coefficient of $r \simeq 0.72$ and a generalized correlation coefficient of $R^2 = r^2 \simeq 0.51$, where R^2 is the generalized correlation coefficient. On the other hand, the KNN regression method gives a generalized correlation coefficient of $R^2 \simeq 0.69$. A comparison of R^2 between the two models, would suggest that the KNN regression model by having a higher value of R^2 would give better fit and predictions for the analyzed data.