

DAT407 Assignment 1 – Group 19

Avinash Shukla - 20 hours

Josef Rasheed - 20 hours

March 27, 2023

```
In [1]: # We import the necessary libraries under shorter aliases.
```

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Data cleaning

In this section we will inspect the data and clean it from any entries that are not useful for us. This will make it easier to avoid mistakes in the future. To avoid wasting time we will only clean data from the year 2020 since that is the latest data our datasets hold.

GDP per capita data [1]

```
In [2]: # Reads from the gdp-pc csv file and creating a dataframe from the data.
gdp_pc_df = pd.read_csv("datasets/gdp-pc.csv")

# Prints first and last five rows to get a quick overview of the data
gdp_pc_df
```

Out[2]:

	Entity	Code	Year	GDP per capita
0	Afghanistan	AFG	2002	1189.784668
1	Afghanistan	AFG	2003	1235.810059
2	Afghanistan	AFG	2004	1200.277954
3	Afghanistan	AFG	2005	1286.793701
4	Afghanistan	AFG	2006	1315.789062
...
6161	Zimbabwe	ZWE	2016	3678.217041
6162	Zimbabwe	ZWE	2017	3795.642334
6163	Zimbabwe	ZWE	2018	3923.030518
6164	Zimbabwe	ZWE	2019	3630.033936
6165	Zimbabwe	ZWE	2020	3353.411377

6166 rows × 4 columns

```
In [3]: # Get gdp pc data from 2020
gdp_pc_2020 = gdp_pc_df[gdp_pc_df.Year == 2020]

gdp_pc_2020
```

```
Out[3]:
```

	Entity	Code	Year	GDP per capita
18	Afghanistan	AFG	2020	1970.560181
49	Albania	ALB	2020	13191.952148
80	Algeria	DZA	2020	10735.382812
111	Angola	AGO	2020	6110.274414
142	Antigua and Barbuda	ATG	2020	17293.302734
...
6041	Vanuatu	VUT	2020	2854.292480
6072	Vietnam	VNM	2020	8200.332031
6103	World	OWID_WRL	2020	16185.165039
6134	Zambia	ZMB	2020	3277.816406
6165	Zimbabwe	ZWE	2020	3353.411377

205 rows × 4 columns

```
In [4]: isna_list = gdp_pc_2020[gdp_pc_2020.isna().any(axis=1)]

print(isna_list.shape)
isna_list
```

(14, 4)

```
Out[4]:
```

	Entity	Code	Year	GDP per capita
1559	East Asia and Pacific	NaN	2020	17548.126953
1802	Europe and Central Asia	NaN	2020	33480.156250
1833	European Union	NaN	2020	41713.882812
2329	High income	NaN	2020	47952.492188
2977	Latin America and Caribbean	NaN	2020	14824.330078
3165	Low and middle income	NaN	2020	10273.044922
3196	Low income	NaN	2020	1988.428223
3227	Lower middle income	NaN	2020	6764.930664
3656	Middle East and North Africa	NaN	2020	15210.993164
3687	Middle income	NaN	2020	11224.483398
4126	North America	NaN	2020	58496.777344
5174	South Asia	NaN	2020	5814.122559
5298	Sub-Saharan Africa	NaN	2020	3718.469482
5948	Upper middle income	NaN	2020	17102.671875

We are only interested in countries, and we see above that all rows containing null are not countries, thus they can be removed.

```
In [5]: # Remove all null rows
gdp_pc_2020 = gdp_pc_2020.dropna()
print(gdp_pc_2020.shape)
print(gdp_pc_2020.to_markdown())
```

(191, 4)

	Entity	Code	Year	GDP per capita
18	Afghanistan	AFG	2020	1970.56
49	Albania	ALB	2020	13192
80	Algeria	DZA	2020	10735.4
111	Angola	AGO	2020	6110.27
142	Antigua and Barbuda	ATG	2020	17293.3
173	Argentina	ARG	2020	19690.7
204	Armenia	ARM	2020	12619.9
263	Australia	AUS	2020	48678.7
294	Austria	AUT	2020	51857.7
325	Azerbaijan	AZE	2020	13726.8
356	Bahamas	BHS	2020	30846.6
387	Bahrain	BHR	2020	41480.6
418	Bangladesh	BGD	2020	4871.45
449	Barbados	BRB	2020	12655.4
480	Belarus	BLR	2020	19186.8
511	Belgium	BEL	2020	48769.7
542	Belize	BLZ	2020	6121.84
573	Benin	BEN	2020	3323.14
604	Bermuda	BMU	2020	76626.5
635	Bhutan	BTN	2020	10551.5
666	Bolivia	BOL	2020	7845.17
693	Bosnia and Herzegovina	BIH	2020	14509.3
724	Botswana	BWA	2020	14654.7
755	Brazil	BRA	2020	14064
786	Brunei	BRN	2020	62200.9
817	Bulgaria	BGR	2020	22379.1
848	Burkina Faso	BFA	2020	2155.77
879	Burundi	BDI	2020	731.063
907	Cambodia	KHM	2020	4191.57
938	Cameroon	CMR	2020	3666.41
962	Canada	CAN	2020	46064.3
993	Cape Verde	CPV	2020	6045.06
1008	Cayman Islands	CYM	2020	67871.2
1039	Central African Republic	CAF	2020	936.288
1070	Chad	TCD	2020	1519.04
1101	Chile	CHL	2020	23324.5
1132	China	CHN	2020	16315.8
1163	Colombia	COL	2020	13448.9
1194	Comoros	COM	2020	2989.17
1225	Congo	COG	2020	3433.96
1256	Costa Rica	CRI	2020	19903.1
1287	Cote d'Ivoire	CIV	2020	5181.31
1313	Croatia	HRV	2020	27077
1334	Curacao	CUW	2020	20309.1
1365	Cyprus	CYP	2020	38816.1
1396	Czechia	CZE	2020	38511.3
1427	Democratic Republic of Congo	COD	2020	1082.45
1458	Denmark	DNK	2020	55819.9
1466	Djibouti	DJI	2020	5481.11
1497	Dominica	DMA	2020	10288.9
1528	Dominican Republic	DOM	2020	17003
1590	Ecuador	ECU	2020	10329.2
1621	Egypt	EGY	2020	11951.4
1652	El Salvador	SLV	2020	7982.63
1683	Equatorial Guinea	GNQ	2020	17007.6
1709	Estonia	EST	2020	35257.2
1740	Eswatini	SWZ	2020	8404.73
1771	Ethiopia	ETH	2020	2296.83
1864	Fiji	FJI	2020	11450.8
1895	Finland	FIN	2020	47154.4
1926	France	FRA	2020	42320.5
1957	Gabon	GAB	2020	14320.8

1988	Gambia	GMB	2020	2156.75
2019	Georgia	GEO	2020	13966.3
2050	Germany	DEU	2020	51423.2
2081	Ghana	GHA	2020	5445.73
2112	Greece	GRC	2020	27072.6
2143	Grenada	GRD	2020	14488.5
2174	Guatemala	GTM	2020	8393.28
2205	Guinea	GIN	2020	2670.82
2236	Guinea-Bissau	GNB	2020	1847.47
2267	Guyana	GUY	2020	18680
2298	Haiti	HTI	2020	2934.25
2360	Honduras	HND	2020	5138.39
2391	Hong Kong	HKG	2020	56154
2421	Hungary	HUN	2020	31098.4
2447	Iceland	ISL	2020	52375.7
2478	India	IND	2020	6165.75
2509	Indonesia	IDN	2020	11445
2540	Iran	IRN	2020	12644.4
2571	Iraq	IRQ	2020	9012.38
2602	Ireland	IRL	2020	90789.2
2628	Israel	ISR	2020	39056.4
2659	Italy	ITA	2020	39073.4
2690	Jamaica	JAM	2020	8761.02
2721	Japan	JPN	2020	40232.2
2752	Jordan	JOR	2020	9816.55
2783	Kazakhstan	KAZ	2020	25363.2
2814	Kenya	KEN	2020	4339.86
2845	Kiribati	KIR	2020	2259.5
2858	Kosovo	OWID_KOS	2020	10706.5
2884	Kuwait	KWT	2020	44847.1
2915	Kyrgyzstan	KGZ	2020	4714.94
2946	Laos	LAO	2020	7810.73
3003	Latvia	LVA	2020	30100.3
3034	Lebanon	LBN	2020	11488.4
3065	Lesotho	LSO	2020	2316.99
3086	Liberia	LBR	2020	1392
3108	Libya	LBY	2020	10282.3
3134	Lithuania	LTU	2020	37107.1
3258	Luxembourg	LUX	2020	112557
3289	Macao	MAC	2020	57729
3320	Madagascar	MDG	2020	1463.86
3351	Malawi	MWI	2020	1508.84
3382	Malaysia	MYS	2020	26471.7
3408	Maldives	MDV	2020	12744.4
3439	Mali	MLI	2020	2225.89
3470	Malta	MLT	2020	39979.7
3501	Marshall Islands	MHL	2020	3932.25
3532	Mauritania	MRT	2020	5109.81
3563	Mauritius	MUS	2020	19462.9
3594	Mexico	MEX	2020	17852.4
3625	Micronesia (country)	FSM	2020	3368.55
3713	Moldova	MDA	2020	12324.4
3744	Mongolia	MNG	2020	11723.9
3768	Montenegro	MNE	2020	18258.9
3799	Morocco	MAR	2020	6986.28
3830	Mozambique	MOZ	2020	1229.94
3861	Myanmar	MMR	2020	4857.39
3892	Namibia	NAM	2020	8814.59
3909	Nauru	NRU	2020	13594.5
3940	Nepal	NPL	2020	3800.07
3971	Netherlands	NLD	2020	54324.4
4002	New Zealand	NZL	2020	42774.7
4033	Nicaragua	NIC	2020	5280.1
4064	Niger	NER	2020	1221.27
4095	Nigeria	NGA	2020	4916.72

4157	North Macedonia	MKD	2020	15931.4
4188	Norway	NOR	2020	63548
4219	Oman	OMN	2020	29502.1
4250	Pakistan	PAK	2020	4562.62
4271	Palau	PLW	2020	15473.3
4298	Palestine	PSE	2020	5394.07
4329	Panama	PAN	2020	25389.8
4360	Papua New Guinea	PNG	2020	4063.73
4391	Paraguay	PRY	2020	12389.6
4422	Peru	PER	2020	11260.8
4453	Philippines	PHL	2020	7953.58
4484	Poland	POL	2020	32398.7
4515	Portugal	PRT	2020	31961.8
4546	Puerto Rico	PRI	2020	32551.1
4567	Qatar	QAT	2020	85290.5
4598	Romania	ROU	2020	28870.9
4629	Russia	RUS	2020	26456.4
4660	Rwanda	RWA	2020	2098.71
4691	Saint Kitts and Nevis	KNA	2020	24319.4
4722	Saint Lucia	LCA	2020	12048.9
4753	Saint Vincent and the Grenadines	VCT	2020	12044.4
4784	Samoa	WSM	2020	6416.78
4827	Sao Tome and Principe	STP	2020	4051.6
4858	Saudi Arabia	SAU	2020	44328.2
4889	Senegal	SEN	2020	3320.68
4915	Serbia	SRB	2020	18255.1
4946	Seychelles	SYC	2020	24427.7
4977	Sierra Leone	SLE	2020	1637.3
5008	Singapore	SGP	2020	93397
5047	Slovakia	SVK	2020	30509.8
5073	Slovenia	SVN	2020	37050.5
5104	Solomon Islands	SLB	2020	2482.87
5112	Somalia	SOM	2020	1180.97
5143	South Africa	ZAF	2020	12665.8
5205	South Korea	KOR	2020	42335.9
5236	Spain	ESP	2020	36210.9
5267	Sri Lanka	LKA	2020	12536.9
5329	Sudan	SDN	2020	3927.04
5360	Suriname	SUR	2020	15864.6
5391	Sweden	SWE	2020	50923.1
5422	Switzerland	CHE	2020	68755.4
5453	Tajikistan	TJK	2020	3657.57
5484	Tanzania	TZA	2020	2635.36
5515	Thailand	THA	2020	17284.7
5536	Timor	TLS	2020	3925.95
5567	Togo	TGO	2020	2107.88
5598	Tonga	TON	2020	6347
5629	Trinidad and Tobago	TTO	2020	23722.5
5660	Tunisia	TUN	2020	10260.1
5691	Turkey	TUR	2020	28393.5
5731	Turks and Caicos Islands	TCA	2020	21126.9
5762	Tuvalu	TUV	2020	4411
5793	Uganda	UGA	2020	2175.03
5824	Ukraine	UKR	2020	12375.9
5855	United Arab Emirates	ARE	2020	63299.4
5886	United Kingdom	GBR	2020	42675.8
5917	United States	USA	2020	59919.8
5979	Uruguay	URY	2020	21608.4
6010	Uzbekistan	UZB	2020	7331.86
6041	Vanuatu	VUT	2020	2854.29
6072	Vietnam	VNM	2020	8200.33
6103	World	OWID_WRL	2020	16185.2
6134	Zambia	ZMB	2020	3277.82
6165	Zimbabwe	ZWE	2020	3353.41

Upon further inspection we also identify the anomalous entries "Kosovo" with code "OWID_KOS" and "World" with code "OWID_WRL". We are not completely sure about what these entries mean but assume that the "World" entry refers to the world GDP pc, which is irrelevant for us. And that the "Kosovo" entry has something to do with that Kosovo isn't reconised as a country by some countries. Either way we will delete these entries to avoid any problems.

```
In [6]: # Remove "Kosovo" and "World"
gdp_pc_clean = gdp_pc_2020.loc[(gdp_pc_2020['Entity'] != 'Kosovo') & (gdp_pc_2020['Entity'] != 'World')]
print(gdp_pc_clean.shape)
```

(189, 4)

GDP data [2]

```
In [7]: # Reads from the gdp csv file and creating a dataframe from the data.
gdp_df = pd.read_csv("datasets/gdp.csv")

gdp_df
```

```
Out[7]:
```

	Entity	Code	Year	GDP
0	Afghanistan	AFG	2002	7228792320
1	Afghanistan	AFG	2003	7867259392
2	Afghanistan	AFG	2004	7978511360
3	Afghanistan	AFG	2005	8874475520
4	Afghanistan	AFG	2006	9349916672
...
10452	Zimbabwe	ZWE	2016	20114016256
10453	Zimbabwe	ZWE	2017	21061283840
10454	Zimbabwe	ZWE	2018	22077325312
10455	Zimbabwe	ZWE	2019	20720840704
10456	Zimbabwe	ZWE	2020	19426048000

10457 rows × 4 columns

```
In [8]: # Get gdp data from 2020
gdp_2020 = gdp_df[gdp_df.Year == 2020]
gdp_2020
```

```
Out[8]:
```

	Entity	Code	Year	GDP
18	Afghanistan	AFG	2020	20621946880
59	Albania	ALB	2020	12457862144
120	Algeria	DZA	2020	168144142336
139	American Samoa	ASM	2020	653477888
190	Andorra	AND	2020	2649303808
...
10207	Vanuatu	VUT	2020	810535808
10244	Vietnam	VNM	2020	258508668928
10305	World	OWID_WRL	2020	81830090899456
10395	Zambia	ZMB	2020	23418945536
10456	Zimbabwe	ZWE	2020	19426048000

210 rows × 4 columns

```
In [9]: isna_list = gdp_2020[gdp_2020.isna().any(axis=1)]

print(isna_list.shape)
isna_list
```

(14, 4)

```
Out[9]:
```

	Entity	Code	Year	GDP
2667	East Asia and Pacific	NaN	2020	26267552317440
3074	Europe and Central Asia	NaN	2020	21122173108224
3125	European Union	NaN	2020	13890496757760
4097	High income	NaN	2020	49843141607424
5163	Latin America and Caribbean	NaN	2020	5220250157056
5414	Low and middle income	NaN	2020	31704873959424
5454	Low income	NaN	2020	479579078656
5515	Lower middle income	NaN	2020	7530226909184
6216	Middle East and North Africa	NaN	2020	3243739447296
6277	Middle income	NaN	2020	31226517782528
7043	North America	NaN	2020	20908301352960
8711	South Asia	NaN	2020	3235660169216
8962	Sub-Saharan Africa	NaN	2020	1797727715328
10070	Upper middle income	NaN	2020	23696597057536

None of the above are countries so we can get rid of them.


```
In [10]: # Remove all null rows
gdp_2020 = gdp_2020.dropna()
print(gdp_2020.shape)
print(gdp_2020.to_markdown())
```

(196, 4)

	Entity	Code	Year	GDP
18	Afghanistan	AFG	2020	20621946880
59	Albania	ALB	2020	12457862144
120	Algeria	DZA	2020	168144142336
139	American Samoa	ASM	2020	653477888
190	Andorra	AND	2020	2649303808
231	Angola	AGO	2020	104128684032
275	Antigua and Barbuda	ATG	2020	1301036672
336	Argentina	ARG	2020	514772402176
367	Armenia	ARM	2020	11915300864
460	Australia	AUS	2020	1490967855104
521	Austria	AUT	2020	386514223104
552	Azerbaijan	AZE	2020	51307184128
613	Bahamas	BHS	2020	10700067840
654	Bahrain	BHR	2020	33206179840
715	Bangladesh	BGD	2020	270695546880
762	Barbados	BRB	2020	3906818560
793	Belarus	BLR	2020	58482352128
854	Belgium	BEL	2020	466836652032
915	Belize	BLZ	2020	1577955712
976	Benin	BEN	2020	14725558272
1037	Bermuda	BMU	2020	6416620032
1078	Bhutan	BTN	2020	2221963776
1139	Bolivia	BOL	2020	34820960256
1166	Bosnia and Herzegovina	BIH	2020	17825165312
1227	Botswana	BWA	2020	14813376512
1288	Brazil	BRA	2020	1749104721920
1335	Brunei	BRN	2020	13429361664
1376	Bulgaria	BGR	2020	54923718656
1437	Burkina Faso	BFA	2020	15291208704
1498	Burundi	BDI	2020	3218732800
1526	Cambodia	KHM	2020	23012192256
1587	Cameroon	CMR	2020	37686546432
1611	Canada	CAN	2020	1607402389504
1652	Cape Verde	CPV	2020	1632003328
1667	Cayman Islands	CYM	2020	5060362240
1728	Central African Republic	CAF	2020	2001437824
1789	Chad	TCD	2020	10426343424
1850	Chile	CHL	2020	247639113728
1911	China	CHN	2020	14631844184064
1972	Colombia	COL	2020	299826446336
2013	Comoros	COM	2020	1091379968
2074	Congo	COG	2020	8877391872
2135	Costa Rica	CRI	2020	61668978688
2196	Cote d'Ivoire	CIV	2020	61033877504
2222	Croatia	HRV	2020	52557893632
2273	Cuba	CUB	2020	81054474240
2294	Curacao	CUW	2020	2364843008
2340	Cyprus	CYP	2020	23579791360
2371	Czechia	CZE	2020	203094949888
2432	Democratic Republic of Congo	COD	2020	45259706368
2493	Denmark	DNK	2020	327737540608
2501	Djibouti	DJI	2020	3207792384
2545	Dominica	DMA	2020	473932832
2606	Dominican Republic	DOM	2020	83287064576
2728	Ecuador	ECU	2020	93820100608
2789	Egypt	EGY	2020	412246048768
2845	El Salvador	SLV	2020	23560826880
2886	Equatorial Guinea	GNQ	2020	9509201920
2932	Estonia	EST	2020	26279915520
2983	Eswatini	SWZ	2020	4319633920
3023	Ethiopia	ETH	2020	95069175808
3187	Fiji	FJI	2020	4402511360

3248	Finland	FIN	2020	247606657024
3309	France	FRA	2020	2411255037952
3370	Gabon	GAB	2020	15176119296
3425	Gambia	GMB	2020	1672833024
3481	Georgia	GEO	2020	16557390848
3532	Germany	DEU	2020	3435817336832
3593	Ghana	GHA	2020	62724595712
3654	Greece	GRC	2020	185374539776
3748	Grenada	GRD	2020	979357760
3767	Guam	GUM	2020	5259729920
3828	Guatemala	GTM	2020	69560942592
3863	Guinea	GIN	2020	12922776576
3914	Guinea-Bissau	GNB	2020	1218759808
3975	Guyana	GUY	2020	7275908096
4036	Haiti	HTI	2020	14956794880
4158	Honduras	HND	2020	22022361088
4218	Hong Kong	HKG	2020	311574462464
4248	Hungary	HUN	2020	140096897024
4274	Iceland	ISL	2020	19491448832
4335	India	IND	2020	2500132077568
4396	Indonesia	IDN	2020	1027602841600
4457	Iran	IRN	2020	410188054528
4510	Iraq	IRQ	2020	170857725952
4561	Ireland	IRL	2020	392534851584
4623	Israel	ISR	2020	345459851264
4684	Italy	ITA	2020	1745433788416
4739	Jamaica	JAM	2020	13440715776
4800	Japan	JPN	2020	4380756541440
4845	Jordan	JOR	2020	41108074496
4876	Kazakhstan	KAZ	2020	205829292032
4937	Kenya	KEN	2020	83859111936
4988	Kiribati	KIR	2020	191947344
5001	Kosovo	OWID_KOS	2020	7144367616
5030	Kuwait	KWT	2020	104326529024
5065	Kyrgyzstan	KGZ	2020	7238139392
5102	Laos	LAO	2020	18584864768
5189	Latvia	LVA	2020	29616465920
5222	Lebanon	LBN	2020	36736831488
5283	Lesotho	LSO	2020	2105770368
5304	Liberia	LBR	2020	3115542528
5326	Libya	LBY	2020	27805589504
5353	Lithuania	LTU	2020	48110632960
5576	Luxembourg	LUX	2020	66117877760
5615	Macao	MAC	2020	23488188416
5676	Madagascar	MDG	2020	12244773888
5737	Malawi	MWI	2020	7537262592
5798	Malaysia	MYS	2020	344099356672
5824	Maldives	MDV	2020	3600460032
5878	Mali	MLI	2020	15830714368
5929	Malta	MLT	2020	12886269952
5969	Marshall Islands	MHL	2020	207699808
6029	Mauritania	MRT	2020	7206774272
6074	Mauritius	MUS	2020	11465334784
6135	Mexico	MEX	2020	1148749217792
6170	Micronesia (country)	FSM	2020	326589248
6303	Moldova	MDA	2020	8517409792
6354	Monaco	MCO	2020	6248493056
6394	Mongolia	MNG	2020	13528524800
6418	Montenegro	MNE	2020	4046328320
6473	Morocco	MAR	2020	105726173184
6514	Mozambique	MOZ	2020	17959221248
6574	Myanmar	MMR	2020	86343024640
6615	Namibia	NAM	2020	10285278208
6632	Nauru	NRU	2020	90650120
6693	Nepal	NPL	2020	29966137344

6754	Netherlands	NLD	2020	808332361728
6799	New Zealand	NZL	2020	204482363392
6860	Nicaragua	NIC	2020	12734737408
6921	Niger	NER	2020	12649313280
6982	Nigeria	NGA	2020	493917962240
7074	North Macedonia	MKD	2020	10602899456
7153	Norway	NOR	2020	403552927744
7209	Oman	OMN	2020	80394698752
7270	Pakistan	PAK	2020	319589679104
7291	Palau	PLW	2020	239735280
7318	Palestine	PSE	2020	14015399936
7379	Panama	PAN	2020	52520697856
7440	Papua New Guinea	PNG	2020	23854663680
7501	Paraguay	PRY	2020	40446808064
7562	Peru	PER	2020	190979129344
7623	Philippines	PHL	2020	358294061056
7654	Poland	POL	2020	555630460928
7715	Portugal	PRT	2020	203589533696
7776	Puerto Rico	PRI	2020	92660080640
7797	Qatar	QAT	2020	161416822784
7828	Romania	ROU	2020	208838852608
7860	Russia	RUS	2020	1416124432384
7921	Rwanda	RWA	2020	10807188480
7965	Saint Kitts and Nevis	KNA	2020	973116736
8009	Saint Lucia	LCA	2020	1588638592
8070	Saint Vincent and the Grenadines	VCT	2020	772496640
8109	Samoa	WSM	2020	864486016
8152	Sao Tome and Principe	STP	2020	370868352
8205	Saudi Arabia	SAU	2020	650714611712
8266	Senegal	SEN	2020	22852755456
8292	Serbia	SRB	2020	45184790528
8353	Seychelles	SYC	2020	1386262400
8414	Sierra Leone	SLE	2020	4976754176
8475	Singapore	SGP	2020	330099785728
8514	Slovakia	SVK	2020	94769119232
8540	Slovenia	SVN	2020	48144052224
8581	Solomon Islands	SLB	2020	1467403264
8589	Somalia	SOM	2020	7069120000
8650	South Africa	ZAF	2020	335640166400
8772	South Korea	KOR	2020	1623895048192
8841	Spain	ESP	2020	1181205135360
8901	Sri Lanka	LKA	2020	88832163840
9023	Sudan	SDN	2020	49449889792
9084	Suriname	SUR	2020	4417127424
9145	Sweden	SWE	2020	533611839488
9186	Switzerland	CHE	2020	740026220544
9282	Tajikistan	TJK	2020	11436204032
9315	Tanzania	TZA	2020	61522710528
9376	Thailand	THA	2020	432649175040
9397	Timor	TLS	2020	2093497600
9458	Togo	TGO	2020	5187639296
9498	Tonga	TON	2020	489364576
9559	Trinidad and Tobago	TTO	2020	20813758464
9615	Tunisia	TUN	2020	44681502720
9676	Turkey	TUR	2020	1015326638080
9719	Turks and Caicos Islands	TCA	2020	802388544
9750	Tuvalu	TUV	2020	44708880
9789	Uganda	UGA	2020	40768765952
9823	Ukraine	UKR	2020	97866342400
9869	United Arab Emirates	ARE	2020	370866192384
9930	United Kingdom	GBR	2020	2891615567872
9991	United States	USA	2020	19294482071552
10131	Uruguay	URY	2020	52260970496
10165	Uzbekistan	UZB	2020	107981856768
10207	Vanuatu	VUT	2020	810535808

10244	Vietnam	VNM	2020	258508668928
10305	World	OWID_WRL	2020	81830090899456
10395	Zambia	ZMB	2020	23418945536
10456	Zimbabwe	ZWE	2020	19426048000

Just like the other data, we will remove "Kosovo" and "World".

```
In [11]: # Remove "Kosovo" and "World"
gdp_clean = gdp_2020.loc[(gdp_2020['Entity'] != 'Kosovo') & (gdp_2020['Entity'] != 'World')]
print(gdp_clean.shape)

(194, 4)
```

Life expectancy data [3]

```
In [12]: # Reads from the life-expectancy csv file and creating a dataframe from the data.
le_df = pd.read_csv("datasets/life-expectancy.csv")

le_df
```

```
Out[12]:
```

	Entity	Code	Year	Life expectancy
0	Afghanistan	AFG	1950	27.7
1	Afghanistan	AFG	1951	28.0
2	Afghanistan	AFG	1952	28.4
3	Afghanistan	AFG	1953	28.9
4	Afghanistan	AFG	1954	29.2
...
20444	Zimbabwe	ZWE	2017	60.7
20445	Zimbabwe	ZWE	2018	61.4
20446	Zimbabwe	ZWE	2019	61.3
20447	Zimbabwe	ZWE	2020	61.1
20448	Zimbabwe	ZWE	2021	59.3

20449 rows × 4 columns

```
In [13]: # Get Life expectancy data from 2020
le_2020 = le_df[le_df.Year == 2020]
le_2020
```

```
Out[13]:
```

	Entity	Code	Year	Life expectancy
70	Afghanistan	AFG	2020	62.6
144	Africa	NaN	2020	62.2
216	Albania	ALB	2020	77.0
291	Algeria	DZA	2020	74.5
363	American Samoa	ASM	2020	72.4
...
20152	Western Sahara	ESH	2020	70.5
20231	World	OWID_WRL	2020	72.0
20303	Yemen	YEM	2020	64.7
20375	Zambia	ZMB	2020	62.4
20447	Zimbabwe	ZWE	2020	61.1

256 rows × 4 columns

```
In [14]: isna_list = le_2020[le_2020.isna().any(axis=1)]

print(isna_list.shape)
isna_list
```

(19, 4)

```
Out[14]:
```

	Entity	Code	Year	Life expectancy
144	Africa	NaN	2020	62.20
441	Americas	NaN	2020	74.85
1037	Asia	NaN	2020	73.70
5567	Europe	NaN	2020	77.70
7620	High-income countries	NaN	2020	80.40
9596	Land-locked Developing Countries (LLDC)	NaN	2020	64.10
9740	Latin America and the Caribbean	NaN	2020	73.10
9887	Least developed countries	NaN	2020	64.50
10103	Less developed regions	NaN	2020	70.60
10175	Less developed regions, excluding China	NaN	2020	68.60
10247	Less developed regions, excluding least develo...	NaN	2020	71.90
10609	Low-income countries	NaN	2020	62.90
10681	Lower-middle-income countries	NaN	2020	68.10
12294	More developed regions	NaN	2020	78.60
13562	Northern America	NaN	2020	77.90
13885	Oceania	NaN	2020	79.50
15233	Saint Barthlemy	NaN	2020	80.20
16608	Small Island Developing States (SIDS)	NaN	2020	72.20
19566	Upper-middle-income countries	NaN	2020	75.70

None of the above are countries so we can get rid of them.

```
In [15]: # Remove all null rows
le_2020 = le_2020.dropna()
print(le_2020.shape)
print(le_2020.to_markdown())
```


(237, 4)

	Entity	Code	Year	Life expectancy
70	Afghanistan	AFG	2020	62.6
216	Albania	ALB	2020	77
291	Algeria	DZA	2020	74.5
363	American Samoa	ASM	2020	72.4
513	Andorra	AND	2020	79
586	Angola	AGO	2020	62.3
658	Anguilla	AIA	2020	76.9
730	Antigua and Barbuda	ATG	2020	78.8
817	Argentina	ARG	2020	75.9
889	Armenia	ARM	2020	72.2
961	Aruba	ABW	2020	75.7
1141	Australia	AUS	2020	84.3
1223	Austria	AUT	2020	81.5
1295	Azerbaijan	AZE	2020	66.9
1367	Bahamas	BHS	2020	72.7
1439	Bahrain	BHR	2020	79.2
1516	Bangladesh	BGD	2020	72
1588	Barbados	BRB	2020	77.4
1661	Belarus	BLR	2020	72.5
1837	Belgium	BEL	2020	80.8
1909	Belize	BLZ	2020	72.9
1981	Benin	BEN	2020	60.1
2053	Bermuda	BMU	2020	81.1
2125	Bhutan	BTN	2020	71.6
2203	Bolivia	BOL	2020	64.5
2275	Bonaire Sint Eustatius and Saba	BES	2020	75.9
2347	Bosnia and Herzegovina	BIH	2020	76.2
2419	Botswana	BWA	2020	65.6
2497	Brazil	BRA	2020	74
2569	British Virgin Islands	VGB	2020	75.8
2641	Brunei	BRN	2020	74.8
2723	Bulgaria	BGR	2020	73.6
2795	Burkina Faso	BFA	2020	59.7
2867	Burundi	BDI	2020	61.6
2940	Cambodia	KHM	2020	70.4
3013	Cameroon	CMR	2020	60.8
3123	Canada	CAN	2020	82
3195	Cape Verde	CPV	2020	74.8
3267	Cayman Islands	CYM	2020	75.1
3339	Central African Republic	CAF	2020	54.6
3411	Chad	TCO	2020	52.8
3498	Chile	CHL	2020	79.4
3575	China	CHN	2020	78.1
3655	Colombia	COL	2020	74.8
3727	Comoros	COM	2020	64.2
3799	Congo	COG	2020	63.8
3871	Cook Islands	COK	2020	74.7
3949	Costa Rica	CRI	2020	79.3
4021	Cote d'Ivoire	CIV	2020	59
4093	Croatia	HRV	2020	78
4176	Cuba	CUB	2020	77.6
4248	Curacao	CUW	2020	76.1
4326	Cyprus	CYP	2020	81.4
4403	Czechia	CZE	2020	78.6
4475	Democratic Republic of Congo	COD	2020	59.7
4663	Denmark	DNK	2020	81.5
4735	Djibouti	DJI	2020	62.7
4807	Dominica	DMA	2020	73.6
4882	Dominican Republic	DOM	2020	72.9
4954	Ecuador	ECU	2020	72.2
5049	Egypt	EGY	2020	71
5124	El Salvador	SLV	2020	71.1

5196	Equatorial Guinea	GNQ	2020	60.7
5268	Eritrea	ERI	2020	67.1
5344	Estonia	EST	2020	78.3
5416	Eswatini	SWZ	2020	59.7
5488	Ethiopia	ETH	2020	65.4
5639	Faeroe Islands	FRO	2020	79.4
5711	Falkland Islands	FLK	2020	78.4
5783	Fiji	FJI	2020	67.9
5940	Finland	FIN	2020	81.9
6146	France	FRA	2020	82.2
6218	French Guiana	GUF	2020	77
6290	French Polynesia	PYF	2020	82.2
6362	Gabon	GAB	2020	66.5
6434	Gambia	GMB	2020	62.6
6506	Georgia	GEO	2020	72.8
6587	Germany	DEU	2020	81.1
6660	Ghana	GHA	2020	64.1
6732	Gibraltar	GIB	2020	82.2
6819	Greece	GRC	2020	80.9
6891	Greenland	GRL	2020	72.3
6963	Grenada	GRD	2020	74.9
7035	Guadeloupe	GLP	2020	82.4
7107	Guam	GUM	2020	76.6
7185	Guatemala	GTM	2020	71.8
7257	Guernsey	GGY	2020	82.2
7329	Guinea	GIN	2020	59.3
7401	Guinea-Bissau	GNB	2020	60
7476	Guyana	GUY	2020	68.5
7548	Haiti	HTI	2020	64.1
7695	Honduras	HND	2020	71.5
7767	Hong Kong	HKG	2020	85.2
7845	Hungary	HUN	2020	75.7
8029	Iceland	ISL	2020	82.6
8112	India	IND	2020	70.1
8189	Indonesia	IDN	2020	68.8
8261	Iran	IRN	2020	74.8
8333	Iraq	IRQ	2020	69.1
8411	Ireland	IRL	2020	82.5
8483	Isle of Man	IMN	2020	80.6
8555	Israel	ISR	2020	82.4
8705	Italy	ITA	2020	82.4
8846	Jamaica	JAM	2020	71.9
8937	Japan	JPN	2020	84.7
9009	Jersey	JEY	2020	80.7
9081	Jordan	JOR	2020	75.2
9160	Kazakhstan	KAZ	2020	70
9235	Kenya	KEN	2020	62.7
9307	Kiribati	KIR	2020	67.3
9379	Kosovo	OWID_KOS	2020	76.6
9452	Kuwait	KWT	2020	76.9
9524	Kyrgyzstan	KGZ	2020	69.6
9668	Laos	LAO	2020	68.5
9815	Latvia	LVA	2020	75.5
9959	Lebanon	LBN	2020	77.8
10031	Lesotho	LSO	2020	54.7
10319	Liberia	LBR	2020	60.9
10391	Libya	LBY	2020	72.5
10463	Liechtenstein	LIE	2020	82.8
10537	Lithuania	LTU	2020	75.1
10802	Luxembourg	LUX	2020	81.4
10874	Macao	MAC	2020	85.2
10946	Madagascar	MDG	2020	65.2
11018	Malawi	MWI	2020	63.7
11090	Malaysia	MYS	2020	75.9
11162	Maldives	MDV	2020	79.9

11234	Mali	MLI	2020	58.6
11306	Malta	MLT	2020	83.4
11378	Marshall Islands	MHL	2020	65
11450	Martinique	MTQ	2020	83
11522	Mauritania	MRT	2020	64.5
11599	Mauritius	MUS	2020	74.3
11671	Mayotte	MYT	2020	75.4
11790	Mexico	MEX	2020	70.1
11862	Micronesia (country)	FSM	2020	70.7
11934	Moldova	MDA	2020	70.2
12006	Monaco	MCO	2020	86.5
12078	Mongolia	MNG	2020	72.1
12150	Montenegro	MNE	2020	76.3
12222	Montserrat	MSR	2020	75.3
12366	Morocco	MAR	2020	73.9
12438	Mozambique	MOZ	2020	61.2
12511	Myanmar	MMR	2020	66.8
12583	Namibia	NAM	2020	62.8
12655	Nauru	NRU	2020	63.4
12727	Nepal	NPL	2020	69.2
12899	Netherlands	NLD	2020	81.6
12971	New Caledonia	NCL	2020	80.8
13045	New Zealand	NZL	2020	82.7
13120	Nicaragua	NIC	2020	71.8
13194	Niger	NER	2020	61.5
13266	Nigeria	NGA	2020	52.9
13338	Niue	NIU	2020	70.8
13418	North Korea	PRK	2020	73.3
13490	North Macedonia	MKD	2020	75.2
13634	Northern Mariana Islands	MNP	2020	77.2
13810	Norway	NOR	2020	83.2
13957	Oman	OMN	2020	74.8
14032	Pakistan	PAK	2020	66.3
14104	Palau	PLW	2020	65.3
14176	Palestine	PSE	2020	74.4
14250	Panama	PAN	2020	76.7
14323	Papua New Guinea	PNG	2020	65.8
14401	Paraguay	PRY	2020	73.2
14474	Peru	PER	2020	73.7
14547	Philippines	PHL	2020	72.1
14620	Poland	POL	2020	76.9
14702	Portugal	PRT	2020	81.1
14774	Puerto Rico	PRI	2020	78
14846	Qatar	QAT	2020	79.1
14918	Reunion	REU	2020	82.1
14991	Romania	ROU	2020	75.3
15089	Russia	RUS	2020	71.3
15161	Rwanda	RWA	2020	66.8
15305	Saint Helena	SHN	2020	76.9
15377	Saint Kitts and Nevis	KNA	2020	71.6
15449	Saint Lucia	LCA	2020	73.4
15521	Saint Martin (French part)	MAF	2020	80.1
15593	Saint Pierre and Miquelon	SPM	2020	77.1
15665	Saint Vincent and the Grenadines	VCT	2020	72.1
15737	Samoa	WSM	2020	72.8
15809	San Marino	SMR	2020	79.6
15881	Sao Tome and Principe	STP	2020	67.8
15953	Saudi Arabia	SAU	2020	76.2
16028	Senegal	SEN	2020	68
16100	Serbia	SRB	2020	75.4
16172	Seychelles	SYC	2020	73.5
16245	Sierra Leone	SLE	2020	59.8
16317	Singapore	SGP	2020	82.9
16389	Sint Maarten (Dutch part)	SXM	2020	74.6
16464	Slovakia	SVK	2020	77

16536	Slovenia	SVN	2020	80.4
16680	Solomon Islands	SLB	2020	70.2
16752	Somalia	SOM	2020	56
16824	South Africa	ZAF	2020	65.3
16904	South Korea	KOR	2020	83.6
16976	South Sudan	SSD	2020	55.5
17093	Spain	ESP	2020	82.3
17173	Sri Lanka	LKA	2020	76.4
17245	Sudan	SDN	2020	65.6
17317	Suriname	SUR	2020	72.6
17588	Sweden	SWE	2020	82.4
17734	Switzerland	CHE	2020	83.1
17806	Syria	SYR	2020	72.1
17878	Taiwan	TWN	2020	80.9
17950	Tajikistan	TJK	2020	68
18022	Tanzania	TZA	2020	66.4
18098	Thailand	THA	2020	79.3
18170	Timor	TLS	2020	68.5
18242	Togo	TGO	2020	61
18314	Tokelau	TKL	2020	75.2
18386	Tonga	TON	2020	70.9
18461	Trinidad and Tobago	TTO	2020	74.4
18536	Tunisia	TUN	2020	75.3
18611	Turkey	TUR	2020	75.8
18683	Turkmenistan	TKM	2020	68.7
18755	Turks and Caicos Islands	TCA	2020	75
18827	Tuvalu	TUV	2020	64.4
18902	Uganda	UGA	2020	62.9
18987	Ukraine	UKR	2020	72.6
19059	United Arab Emirates	ARE	2020	78.9
19299	United Kingdom	GBR	2020	80.4
19422	United States	USA	2020	77.4
19494	United States Virgin Islands	VIR	2020	74.9
19643	Uruguay	URY	2020	78.4
19715	Uzbekistan	UZB	2020	70.3
19787	Vanuatu	VUT	2020	70.3
19859	Vatican	VAT	2020	83
19936	Venezuela	VEN	2020	71.1
20008	Vietnam	VNM	2020	75.4
20080	Wallis and Futuna	WLF	2020	78.8
20152	Western Sahara	ESH	2020	70.5
20231	World	OWID_WRL	2020	72
20303	Yemen	YEM	2020	64.7
20375	Zambia	ZMB	2020	62.4
20447	Zimbabwe	ZWE	2020	61.1

Just like the GDP pc data, we will remove "Kosovo" and "World".

```
In [16]: # Remove "Kosovo" and "World"
le_clean = le_2020.loc[(le_2020['Entity'] != 'Kosovo') & (le_2020['Entity'] != 'World')]

print(le_clean.shape)

(235, 4)
```

Now we will combine our clean datasets for answering the questions below.

```
In [17]: # Merges gdp pc, gdp and life expectancy data from 2020 into one dataframe,
# then removes all null/NaN data,
# then sorts by life expectancy
merged = pd.merge(gdp_pc_clean, gdp_clean, how='inner', on=['Entity'])
merged = (pd.merge(merged, le_clean, how='inner', on=['Entity'])
          .dropna()
          .sort_values(['Life expectancy']))
merged = merged[['Entity', 'GDP per capita', 'GDP', 'Life expectancy']]
merged
```

```
Out[17]:
```

	Entity	GDP per capita	GDP	Life expectancy
34	Chad	1519.036743	10426343424	52.8
125	Nigeria	4916.721191	493917962240	52.9
33	Central African Republic	936.287598	2001437824	54.6
95	Lesotho	2316.987793	2105770368	54.7
158	Somalia	1180.965942	7069120000	56.0
...
160	South Korea	42335.867188	1623895048192	83.6
7	Australia	48678.683594	1490967855104	84.3
85	Japan	40232.210938	4380756541440	84.7
100	Macao	57728.988281	23488188416	85.2
74	Hong Kong	56153.972656	311574462464	85.2

189 rows × 4 columns

Question A

Write a Python program that draws a scatter plot of GDP per capita vs life expectancy. State any assumptions and motivate decisions that you make when selecting data to be plotted, and in combining data.

```
In [18]: # Defining the axis
xValues = merged['Life expectancy']
yValues = merged['GDP per capita']

# Plots the values in a scatter plot
plt.scatter(xValues, yValues)

# Formating the figure to make it nicer
plt.title('GDP per capita vs life expectancy')

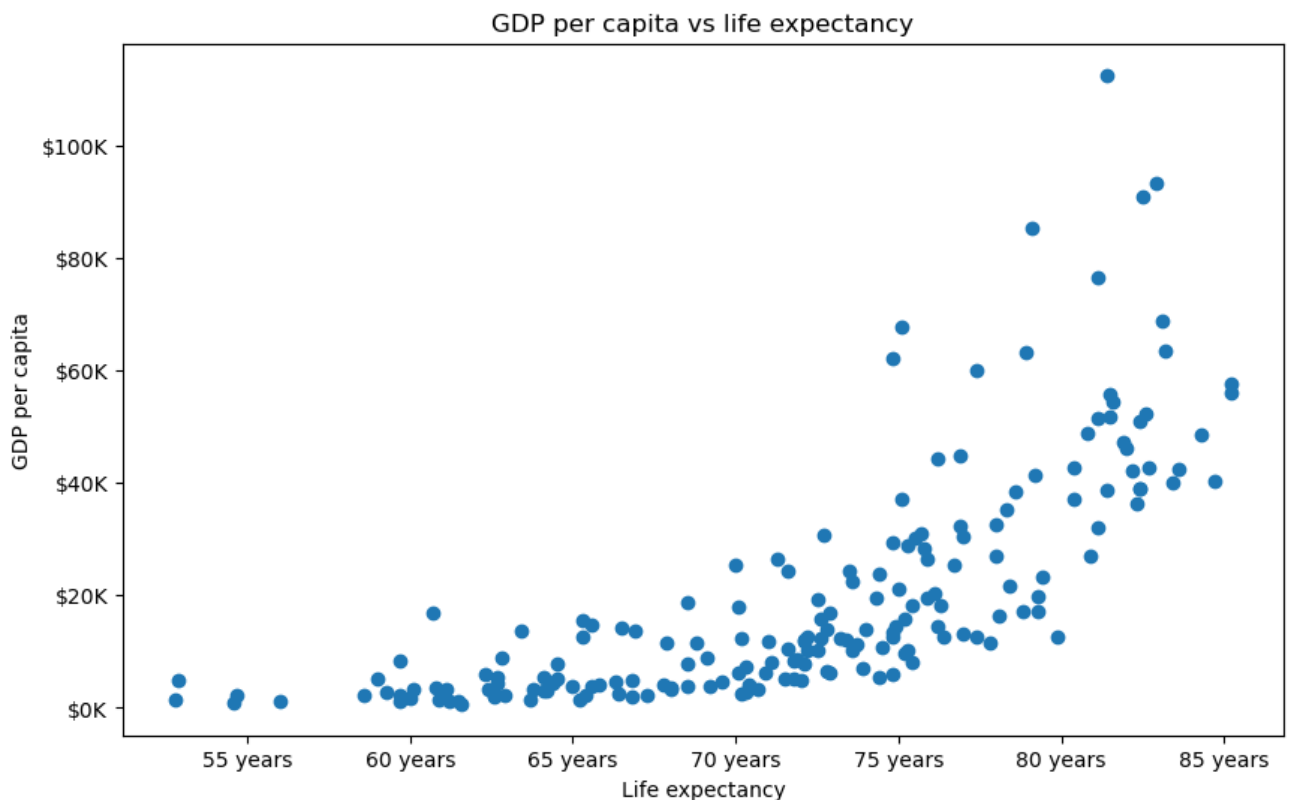
plt.xlabel('Life expectancy')
plt.ylabel('GDP per capita')

# This format function will make the extremely large numbers more readable by displaying
def num_formatter(value, indx):
    if value >= 1_000_000_000:
        formatter = '${:1.0f}B'.format(value*0.000_000_001)
    elif value >= 1_000_000:
        formatter = '${:1.0f}M'.format(value*0.000_001)
    else:
        formatter = '${:1.0f}K'.format(value*0.001)
    return formatter

plt.gca().xaxis.set_major_formatter('{x:1.0f} years')
plt.gca().yaxis.set_major_formatter(num_formatter)

plt.gcf().set_size_inches(10,6)

plt.show()
```



We see a very clear trend with wealthier countries having a higher life expectancy.

Question B

Which countries have a life expectancy higher than one standard deviation above the mean?

```
In [19]: # Get the mean of Life expectancy
le_mean = merged['Life expectancy'].mean()
le_mean
```

Out[19]: 71.93809523809524

```
In [20]: # Get the standard deviation of 2020 Life expectancy
le_std = merged['Life expectancy'].std()
le_std
```

Out[20]: 7.472220081630909

```
In [21]: # Calculate "one standard deviation above the mean".
std_over_mean = le_mean + le_std
std_over_mean
```

Out[21]: 79.41031531972615

```
In [22]: # Get a dataframe with Life expectancy one standard deviation above the mean
std_over_mean_df = (merged[merged['Life expectancy'] > std_over_mean] [['Entity', 'Life
                        .sort_values('Life expectancy', ascending = False))
std_over_mean_df
```

```
Out[22]:
```

	Entity	Life expectancy
74	Hong Kong	85.2
100	Macao	85.2
85	Japan	84.7
7	Australia	84.3
160	South Korea	83.6
106	Malta	83.4
127	Norway	83.2
166	Switzerland	83.1
154	Singapore	82.9
122	New Zealand	82.7
76	Iceland	82.6
81	Ireland	82.5
83	Italy	82.4
82	Israel	82.4
165	Sweden	82.4
161	Spain	82.3
60	France	82.2
30	Canada	82.0
59	Finland	81.9
121	Netherlands	81.6
47	Denmark	81.5
8	Austria	81.5
99	Luxembourg	81.4
44	Cyprus	81.4
64	Germany	81.1
138	Portugal	81.1
18	Bermuda	81.1
66	Greece	80.9
15	Belgium	80.8
156	Slovenia	80.4
181	United Kingdom	80.4
104	Maldives	79.9

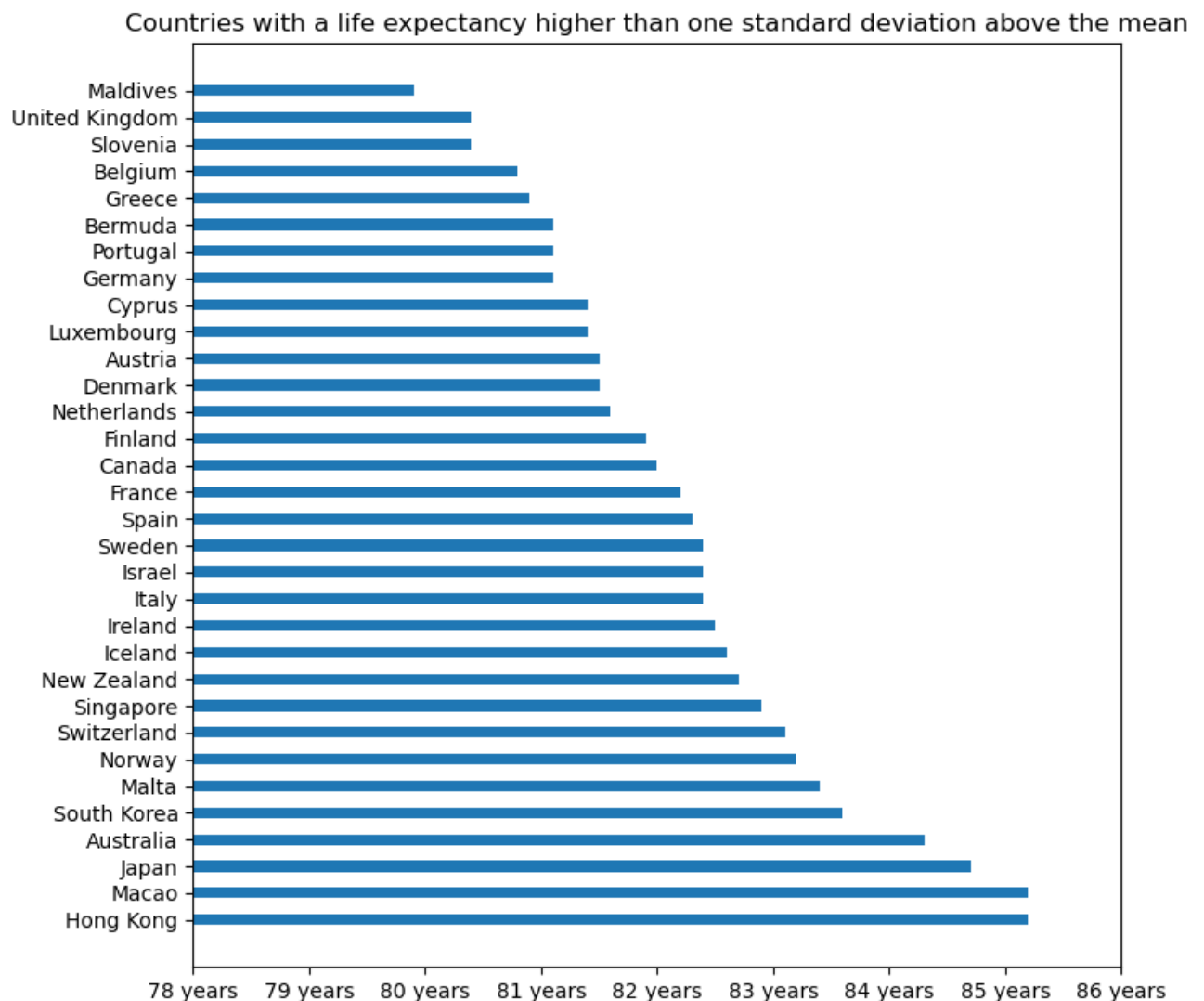

```
In [23]: # Clear the previous plot
plt.clf()

# Defining the axis
xValues = std_over_mean_df['Life expectancy']
yValues = std_over_mean_df['Entity']

# Plots the values in a horizontal bar graph
plt.barh(yValues, xValues, height=0.4)

# Formating the figure to make it nicer
plt.title("Countries with a life expectancy higher than one standard deviation above the mean")
plt.gca().xaxis.set_major_formatter('{x:1.0f} years')
plt.gcf().set_size_inches(8, 8)
plt.xlim([78,86])

plt.show()
```



Question C

Which countries have high life expectancy but have low GDP? (note: GDP and not GDP per capita in question c and d) Motivate how you have chosen to define “high” and “low”.

We will simply define high life expectancy as being life expectancy above the mean and low GDP as being GDP below the mean.

```
In [24]: # Get the GDP mean
gdp_mean = gdp_clean['GDP'].mean()

gdp_mean
```

Out[24]: 416179121628.6598

```
In [25]: # Get the Life expectancy mean
le_mean = le_clean['Life expectancy'].mean()

le_mean
```

Out[25]: 72.83531914893618

```
In [26]: high_le_low_gdp = (merged[(merged['Life expectancy'] > le_mean) & (merged['GDP'] < gdp_m
        .sort_values('Life expectancy', ascending = True))
high_le_low_gdp = high_le_low_gdp[['Entity', 'GDP', 'Life expectancy']]

print(high_le_low_gdp.to_markdown())
```

	Entity	GDP	Life expectancy
-----:	-----:	-----:	-----:
50	Dominican Republic	83287064576	72.9
16	Belize	1577955712	72.9
134	Paraguay	40446808064	73.2
145	Saint Lucia	1588638592	73.4
152	Seychelles	1386262400	73.5
25	Bulgaria	54923718656	73.6
49	Dominica	473932832	73.6
135	Peru	190979129344	73.7
115	Morocco	105726173184	73.9
109	Mauritius	11465334784	74.3
131	Palestine	14015399936	74.4
173	Trinidad and Tobago	20813758464	74.4
2	Algeria	168144142336	74.5
31	Cape Verde	1632003328	74.8
79	Iran	410188054528	74.8
37	Colombia	299826446336	74.8
128	Oman	80394698752	74.8
24	Brunei	13429361664	74.8
67	Grenada	979357760	74.9
176	Turks and Caicos Islands	802388544	75
98	Lithuania	48110632960	75.1
32	Cayman Islands	5060362240	75.1
86	Jordan	41108074496	75.2
126	North Macedonia	10602899456	75.2
141	Romania	208838852608	75.3
174	Tunisia	44681502720	75.3
151	Serbia	45184790528	75.4
186	Vietnam	258508668928	75.4
93	Latvia	29616465920	75.5
75	Hungary	140096897024	75.7
103	Malaysia	344099356672	75.9
43	Curacao	2364843008	76.1
21	Bosnia and Herzegovina	17825165312	76.2
114	Montenegro	4046328320	76.3
162	Sri Lanka	88832163840	76.4
132	Panama	52520697856	76.7
90	Kuwait	104326529024	76.9
155	Slovakia	94769119232	77
1	Albania	12457862144	77
13	Barbados	3906818560	77.4
94	Lebanon	36736831488	77.8
42	Croatia	52557893632	78
139	Puerto Rico	92660080640	78
55	Estonia	26279915520	78.3
183	Uruguay	52260970496	78.4
45	Czechia	203094949888	78.6
4	Antigua and Barbuda	1301036672	78.8
180	United Arab Emirates	370866192384	78.9
140	Qatar	161416822784	79.1
11	Bahrain	33206179840	79.2
40	Costa Rica	61668978688	79.3
35	Chile	247639113728	79.4
104	Maldives	3600460032	79.9
156	Slovenia	48144052224	80.4
66	Greece	185374539776	80.9
18	Bermuda	6416620032	81.1
138	Portugal	203589533696	81.1
44	Cyprus	23579791360	81.4
99	Luxembourg	66117877760	81.4
8	Austria	386514223104	81.5
47	Denmark	327737540608	81.5
59	Finland	247606657024	81.9
82	Israel	345459851264	82.4

81	Ireland	392534851584	82.5
76	Iceland	19491448832	82.6
122	New Zealand	204482363392	82.7
154	Singapore	330099785728	82.9
127	Norway	403552927744	83.2
106	Malta	12886269952	83.4
100	Macao	23488188416	85.2
74	Hong Kong	311574462464	85.2

```
In [27]: # Clear the previous plot
plt.clf()

# Defining the axis
xValues = high_le_low_gdp['Life expectancy']
yValues = high_le_low_gdp['GDP']

# Plots the values in a scatter plot
plt.scatter(xValues, yValues)

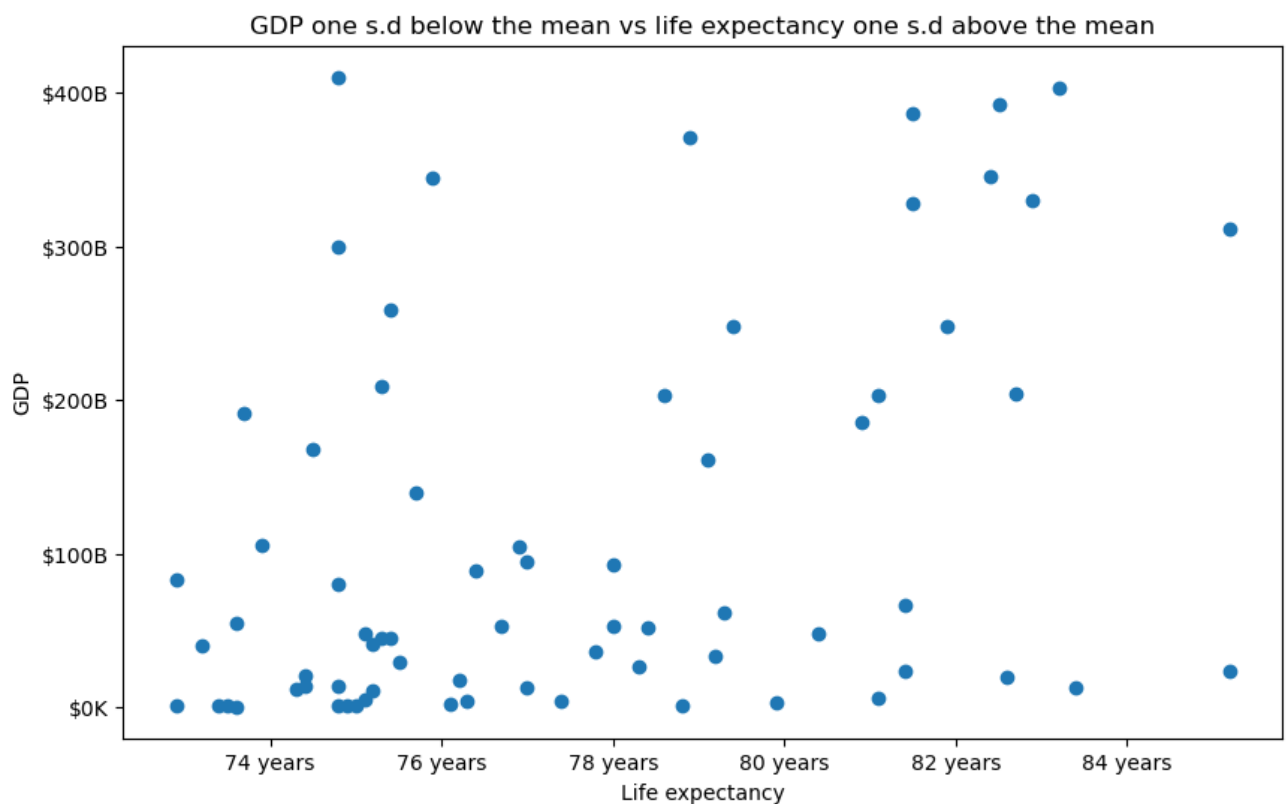
# Formating the figure to make it nicer
plt.title('GDP one s.d below the mean vs life expectancy one s.d above the mean')

plt.xlabel('Life expectancy')
plt.ylabel('GDP')

plt.gca().xaxis.set_major_formatter('{x:1.0f} years')
plt.gca().yaxis.set_major_formatter(num_formatter)

plt.gcf().set_size_inches(10,6)

plt.show()
```



We see that the scatter plot data is fairly spread out. It is however slightly denser in the lower GDP and life expectancy area.

Question D

Does every strong economy (normally indicated by GDP) have high life expectancy?

Since we defined high and low as being above or below the mean, we will get the answer for this question by finding countries with GDP above the mean and life expectancy below the mean.

```
In [28]: # Get countries with GDP above the mean and life expectancy below the mean.
high_gdp_low_le = (merged[(merged['Life expectancy'] < le_mean) & (merged['GDP'] > gdp_m)]
                  .sort_values('Life expectancy', ascending = False))
high_gdp_low_le = high_gdp_low_le[['Entity', 'GDP', 'Life expectancy']]

high_gdp_low_le
```

```
Out[28]:
```

	Entity	GDP	Life expectancy
142	Russia	1416124432384	71.3
110	Mexico	1148749217792	70.1
77	India	2500132077568	70.1
78	Indonesia	1027602841600	68.8
125	Nigeria	493917962240	52.9

```
In [29]: # Clear the previous plot
plt.clf()

# Defining the axis
xValues = high_gdp_low_le['Life expectancy']
yValues = high_gdp_low_le['GDP']

# Plots the values in a scatter plot
plt.scatter(xValues, yValues)

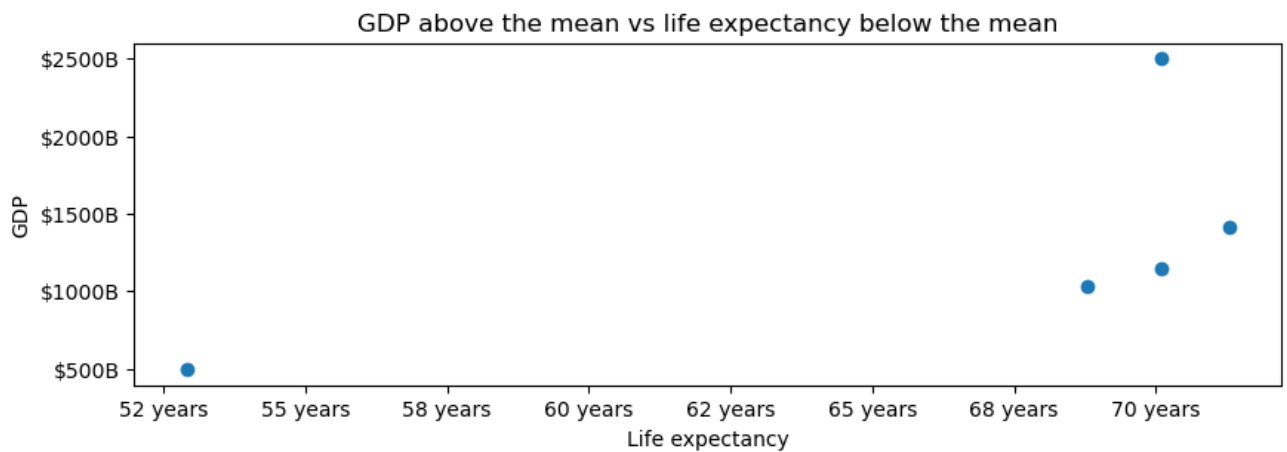
# Formating the figure to make it nicer
plt.title('GDP above the mean vs life expectancy below the mean')

plt.xlabel('Life expectancy')
plt.ylabel('GDP')

plt.gca().xaxis.set_major_formatter('{x:1.0f} years')
plt.gca().yaxis.set_major_formatter(num_formatter)

plt.gcf().set_size_inches(10,3)

plt.show()
```



We see that Russia, Mexico, India, Indonesia and Nigeria have high GDP yet have low life expectancy.

Question E

Related to question d, what would happen if you use GDP per capita as an indicator of a strong economy? Explain the results you obtained, and discuss any insights you get from comparing the results of d and e.

```
In [30]: # Get the GDP pc mean
gdp_pc_mean = gdp_pc_clean['GDP per capita'].mean()

gdp_pc_mean
```

```
Out[30]: 20151.927192486153
```

```
In [31]: # Get countries with GDP pc above the mean and life expectancy below the mean.
high_gdp_pc_low_le = (merged[(merged['Life expectancy'] < le_mean) & (merged['GDP per capita'] > gdp_mean)]
                        .sort_values('Life expectancy', ascending = False))
high_gdp_pc_low_le = high_gdp_pc_low_le[['Entity', 'GDP per capita', 'Life expectancy']]
high_gdp_pc_low_le
```

```
Out[31]:
```

	Entity	GDP per capita	Life expectancy
10	Bahamas	30846.583984	72.7
144	Saint Kitts and Nevis	24319.363281	71.6
142	Russia	26456.388672	71.3
87	Kazakhstan	25363.166016	70.0

```
In [32]: # Clear the previous plot
plt.clf()

# Defining the axis
xValues = high_gdp_pc_low_le['Life expectancy']
yValues = high_gdp_pc_low_le['GDP per capita']

# Plots the values in a scatter plot
plt.scatter(xValues, yValues)

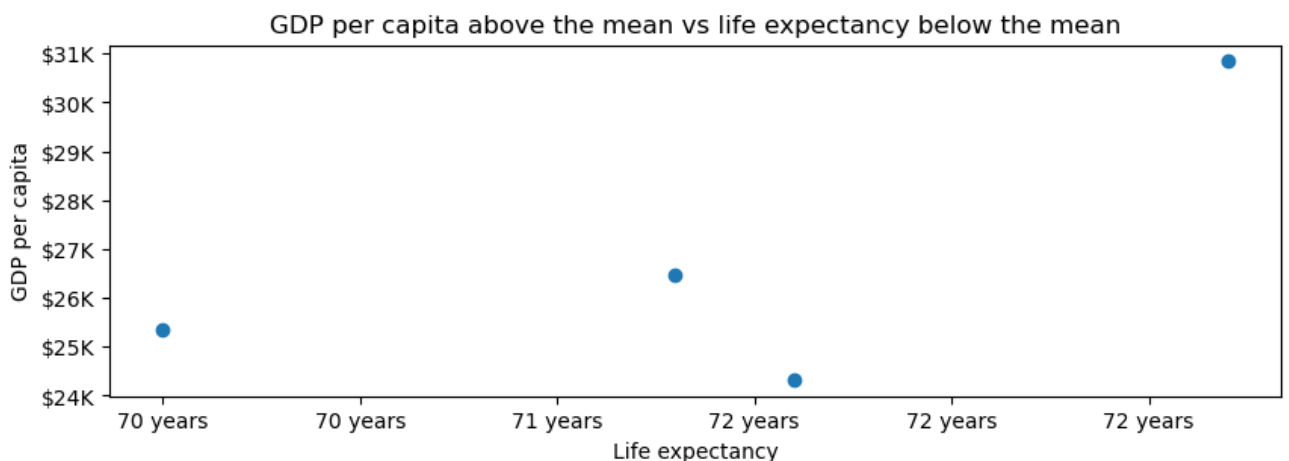
# Formating the figure to make it nicer
plt.title('GDP per capita above the mean vs life expectancy below the mean')

plt.xlabel('Life expectancy')
plt.ylabel('GDP per capita')

plt.gca().xaxis.set_major_formatter('{x:1.0f} years')
plt.gca().yaxis.set_major_formatter(num_formatter)

plt.gcf().set_size_inches(10,3)

plt.show()
```



We see that Bahamas, Saint Kitts and Nevis, Russia and Kazakhstan have high GDP pc yet have low life expectancy.

Since we see Mexico, India, Indonesia and Nigeria in D but not E, we learn that although these countries have a high GDP, they actually have a low GDP pc which makes the low life expectancy more understandable.

On the flip side, since we see Bahamas, Saint Kitts and Nevis and Kazakhstan in E but not D, we learn that although these countries have a high GDP pc, they have a low GDP. This points in the direction that there are other issues these countries have that negatively affect life expectancy such that even though in question A we see a trend with high GDP pc countries having higher life expectancy, these issues are so big that these countries still have a low life expectancy.

And lastly we have Russia which we see in both D and E. This is very interesting since this means that despite having both a high GDP and GDP per capita, they still have a low life expectancy.

However still, looking at the scatter plots we can actually still see a slight trend with wealthier countries having a higher life expectancy.

References

[1] Our World in Data, Economic Growth, GDP per capita 2020, <https://ourworldindata.org/economic-growth#different-data-sets-on-growth-in-the-last-decades> (<https://ourworldindata.org/economic-growth#different-data-sets-on-growth-in-the-last-decades>).

[2] Our World in Data, Economic Growth, GDP 2020, <https://ourworldindata.org/economic-growth#total-output-of-economies> (<https://ourworldindata.org/economic-growth#total-output-of-economies>).

[3] Our World in Data, Life Expectancy, Life Expectancy 2021, <https://ourworldindata.org/life-expectancy#differences-in-life-expectancy-across-the-world> (<https://ourworldindata.org/life-expectancy#differences-in-life-expectancy-across-the-world>).