

DAT405

INTRODUCTION TO DATA SCIENCE AND AI

---

# Assignment 1

---



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

*Author 1:*

Carl ÖSTLING, IT

960925-8257

ostcarl@student.chalmers.se

Hours spent on assignment:

10

*Author 2:*

Erik GUNNARSSON, IT

980905-3078

erig@student.chalmers.se

Hours spent on assignment:

10

January 31, 2020

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

Below table shows the countries that have a life expectancy of atleast one standard deviation higher than mean in the year of 2017.

Entity	Life expectancy (years)
Andorra	83.443
Anguilla	81.568
Australia	83.122
Austria	81.340
Belgium	81.305
Bermuda	82.316
Canada	82.210
Cayman Islands	83.633
Channel Islands	82.766
Falkland Islands	81.163
Finland	81.561
France	82.432
Germany	81.039
Greece	81.901
Guadeloupe	81.708
Hong Kong	84.493
Iceland	82.725
Ireland	81.872
Isle of Man	81.112
Israel	82.663
Italy	83.184
Japan	84.290
Liechtenstein	82.228
Luxembourg	81.955
Macao	83.989
Malta	82.221
Martinique	82.164
Monaco	86.325
Netherlands	82.004
New Zealand	82.002
Norway	82.146
Portugal	81.655
Saint Barthlemy	81.599
Saint Martin (French part)	81.622
San Marino	84.699
Singapore	83.279
Slovenia	81.016
South Korea	82.628
Spain	83.294
Sweden	82.516
Switzerland	83.473
Tokelau	81.476
United Kingdom	81.168

## 2 Which countries have high life expectancy but have low GDP?

Since "high" and "low" are subjective, it is up to us to define them for this question. If we define them as being one standard deviation above or below the mean we get no results as the GDP per capita is negative in this case. Therefore, we chose to define them as simply being above or below the mean. This way of defining "high" and "low" has its flaws, mainly because the mean of GDP per capita is very low compared to the outliers that have a comparatively very high GDP per capita.

Below table shows all countries with GDP per capita under mean and life expectancy above mean along with the values.

Entity	GDP per capita \$	Life expectancy (years)
Albania	11803.430594	78.333
Algeria	13913.839363	76.499
Argentina	18933.907147	76.372
Armenia	8787.579940	74.797
Azerbaijan	15847.418833	72.693
Barbados	16978.067686	78.981
Belarus	17167.967260	74.340
Belize	7824.362459	74.365
Bosnia and Herzegovina	11713.894784	77.128
Brazil	14103.451531	75.456
Bulgaria	18563.306617	74.815
Cape Verde	6222.554166	72.570
China	15308.712123	76.470
Colombia	13254.949218	76.925
Costa Rica	15524.994563	79.914
Dominica	9673.366962	74.597
Dominican Republic	14600.860584	73.689
Ecuador	10581.936429	76.584
El Salvador	7292.457804	72.872
Georgia	9745.078904	73.414
Grenada	13593.876918	72.388
Guatemala	7423.807592	73.810

Honduras	4541.795257	74.898
Jamaica	8193.570650	74.267
Jordan	8337.489701	74.292
Lebanon	13367.565022	78.833
Libya	17881.509351	72.520
Macedonia	13111.213574	75.589
Maldives	15183.616398	78.325
Marshall Islands	3819.202120	73.278
Mexico	17336.469072	74.947
Montenegro	16409.287696	76.667
Morocco	7485.012951	76.218
Nicaragua	5321.443516	74.068
Palau	13240.404810	73.320
Palestine	4449.898325	73.740
Paraguay	8827.010058	73.992
Peru	12236.706152	76.286
Saint Lucia	12951.838877	75.907
Samoa	6021.557424	73.046
Serbia	14048.880637	75.684
Solomon Islands	2205.923205	72.645
Sri Lanka	11669.076679	76.648
Thailand	16277.671151	76.683
Tunisia	10849.297386	76.310
Vietnam	6171.884192	75.241

---

### 3 Does every strong economy have high life expectancy?

If we assume the same definition for "high" as in question 2, and define a strong economy as having a GDP per capita atleast of one standard deviation above the mean, we find that all countries with such a GDP per capita have a "high" life expectancy. This does not come as a surprise as a higher GDP per capita often comes with better access to health care aswell as precautionary means to stay healthy and avoid death.

## 4 Explanation of data chosen

When choosing data we only want to show data for one specific year in order to better fit the format of a scatterplot. For choosing which year to use, we had two approaches in mind. We first decided to find the year which had the most overlapping countries in both datasets (GDP per capita, Life Expectancy) and remove all the other years. The year that maximized this was 2011 and had 195 overlapping entries. With this approach the data can be considered outdated as we ideally would want to look at more recent data. Thus we decided to go with a second approach and find a more recent year that still had a good amount of data overlap. We found the year of 2017 to have a good amount of overlap with 185 countries. We also did not include the regions of GDP per capita without a country code.

All the cleaned data (except for the data lost in merge due to the data overlap) can be found in the dataframes `gdp_year_not_2017`, `le_year_not_2017`, `gdp_no_country_code`.

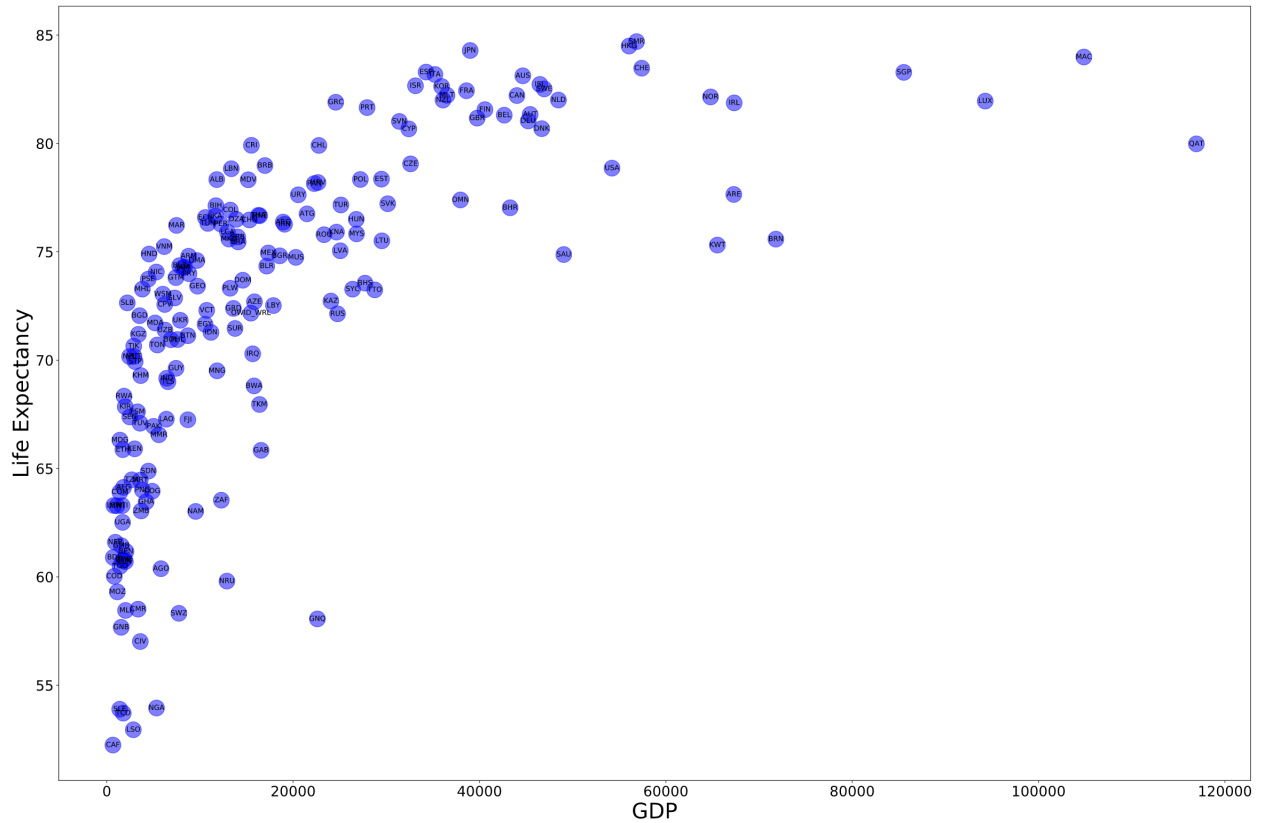


Figure 1: Scatterplot of GDP per capita and Life Expetancy in the year 2017  
(See larger and interactive plot in the code)

## 5 Internet access and science PISA scores

In this section we investigate the correlation between countries' internet usage and their PISA scores in the field of science.

### 5.1 Selection of data

We decided to use a simimilar approach as in the previous problem and chose recent data. We found that the year 2015 was a good candidate that gave us a satisfiable amount of data and being relatively recent.

## 5.2 Conclusions

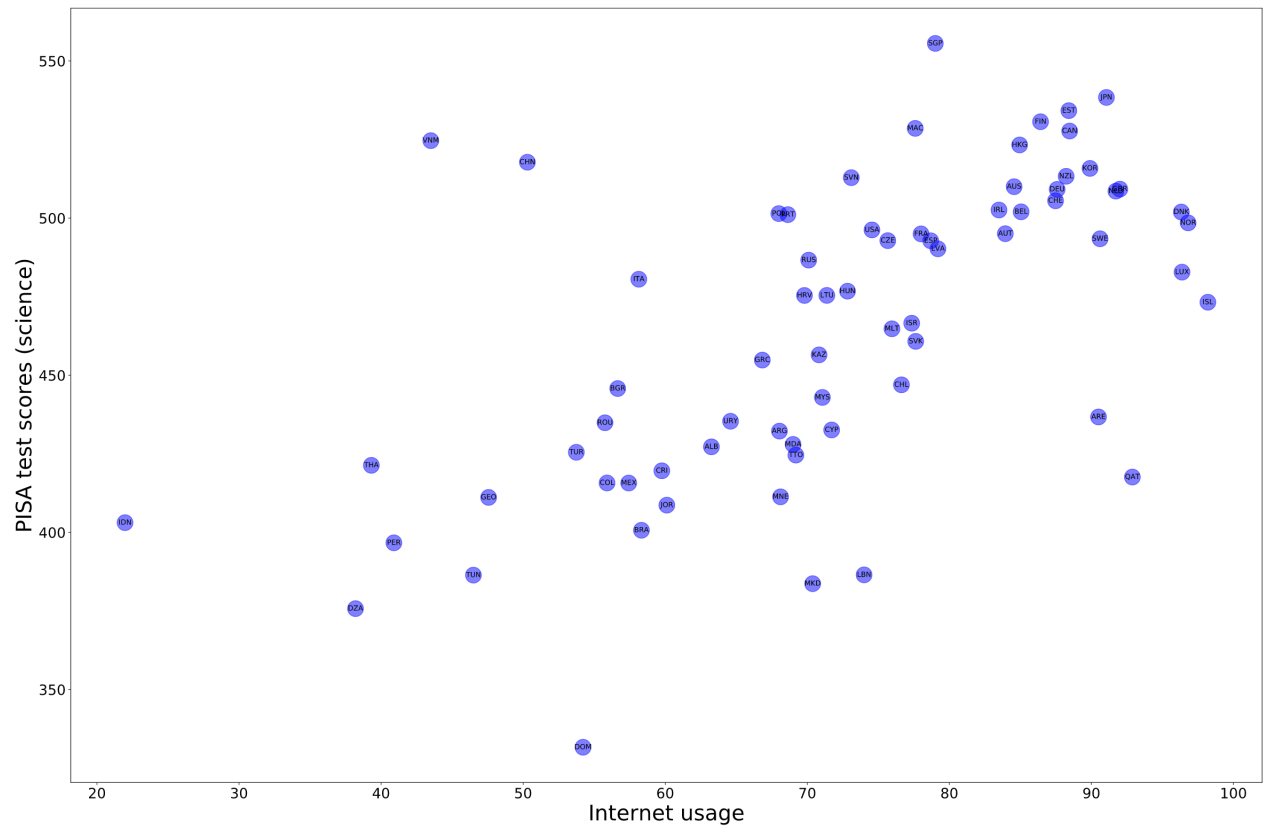


Figure 2: Scatterplot of PISA test scores and internet usage for the year 2016  
(See larger and interactive plot in the code)

Our graph shows that there is a correlation (although small) that countries with a higher internet usage tend to have higher PISA test scores in the science field. However, we cannot say that internet usage has a positive influence on PISA scores from this plot, it could just as well be that high PISA scores promotes internet usage in some way. This the data does not reveal. There are some obvious outliers that have a relatively limited internet usage but high PISA scores.