European Data Project Summary.

This project aims to answer questions about Europe by using statistical methods and showing the data through visualisations. This project is aimed to reflect my ability to analyse data, it isn’t intended to be entirely factual to this present day.

Here are the sources to the data that I have collected:

* EU member states: (<https://europa.eu/european-union/about-eu/countries_en>)
* Members of the monetary union: (<https://europa.eu/european-union/about-eu/countries_en#countries-using-the-euro>)
* Population: (<https://www.worldometers.info/world-population/population-by-country/>)
* GDP per country: (<https://www.worldometers.info/gdp/gdp-by-country/>)
* Defence budget: (<https://www.globalfirepower.com/defense-spending-budget.asp>)
* External debt: (<https://www.globalfirepower.com/external-debt-by-country.asp>)
* Net exports: (<https://en.wikipedia.org/wiki/List_of_countries_by_net_exports>)
* Quality of life index: (<https://www.numbeo.com/quality-of-life/rankings_by_country.jsp>)
* External Debt for Cyprus: (<https://www.ceicdata.com/en/indicator/cyprus/external-debt>)
* External Debt for Iceland: (<https://tradingeconomics.com/iceland/external-debt#:~:text=Looking%20forward%2C%20we%20estimate%20External,according%20to%20our%20econometric%20models>)
* External Debt for Malta: (<https://en.wikipedia.org/wiki/Armed_Forces_of_Malta>)
* Defence budget for Iceland: (<https://www.icelandreview.com/news/iceland-ups-defence-budget-by-37/>)
* Defence budget for Malta: (<https://en.wikipedia.org/wiki/Armed_Forces_of_Malta>)

DISCLAIMER: This data is collected from the sources above, therefore the credibility of the results within this project is based on my findings using this data.

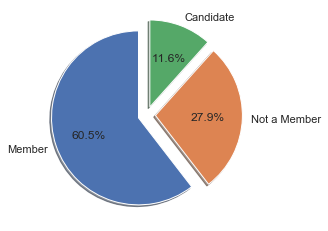
The work produced has been made on Jupyter Notebook and the images seen here are from the Seaborn package based on the matplotlib.pyplot package in Python.

I haven’t included all countries in Europe once I had finished cleaning the data as data wasn’t found for their country, these countries include:

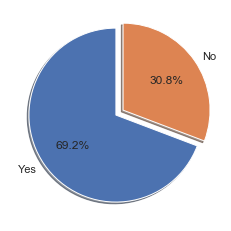
* Andorra
* Kosovo
* Liechtenstein
* Luxembourg
* Monaco
* San Marino
* Vatican City

I filled data in for countries which also had null values which I thought would be more impactful on the data should it have been missing, they are:

* Cyprus
* Iceland
* Malta

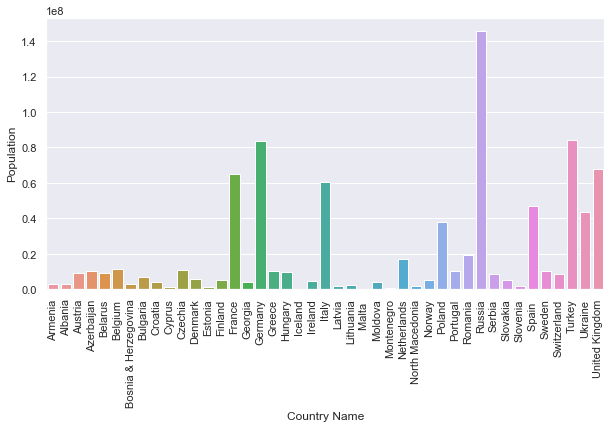
What percentage of European nations are in the European Union?

Here we can see that 60.5% are members, whilst only 28% of Europe are not at all.



What percentage of members are in the monetary union?

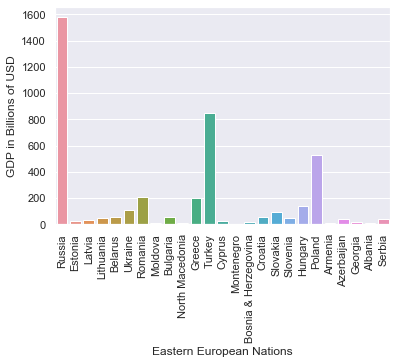
69% of members in the union are using the Euro as their official currency.



This is a bar graph showing us the population of each country inside Europe and an insight into which nations are included in this data. 1 on the y-axis represents 100 million people.

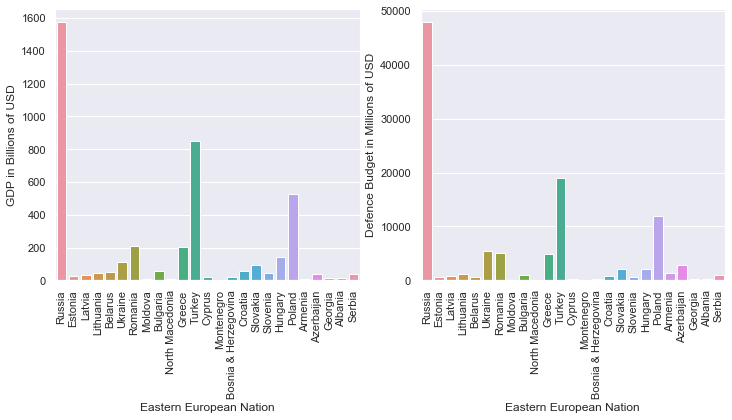
As we can see Russia has a commanding lead at the top of this graph with just over 140 million citizens with Turkey and Germany being a narrow second and third respectively with just over 80 million people each. If we were to take an average of these nations together, it would be heavily outweighed by Russia.

I now begin to split the data into two, creating a dataset for western Europe and one for eastern Europe.



This graph has plot GDP against eastern European nations, we can see that Russia again has the largest economy by far, it’s no surprise given its population.

You’d expect we see a similar pattern when we compare these eastern European countries with military spending.

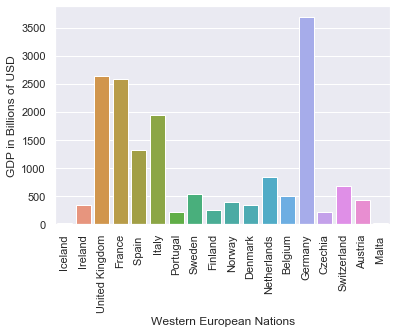


On the right, we have the defence budget plot against eastern European countries. We do have a graph that looks very similar to the last, which means we can expect these nations to have a similar defence budget. More onto that later.

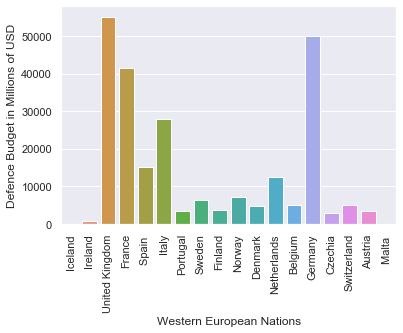
I then calculate the percentage of GDP a nation spends on defence, simply done by dividing military spending by the GDP and multiply it by 100. The top 5 are:

* Armenia : 12.6%
* Azerbaijan: 7%
* Ukraine: 4.8%
* Russia: 3.04%
* Estonia: 2.63%

Other countries follow the trend of spending around 2% then lower than 1% for the small island nations and the neutrals. The data for military spending was taken in June 2020 and having Azerbaijan and Armenia at the top is explained by the current war going on between the countries. And I believe Ukraine can be explained by the threat of Russia on their doorstep. Considering Russia isn’t a large economy, it uses more of its GDP to keep up with the biggest European spenders.

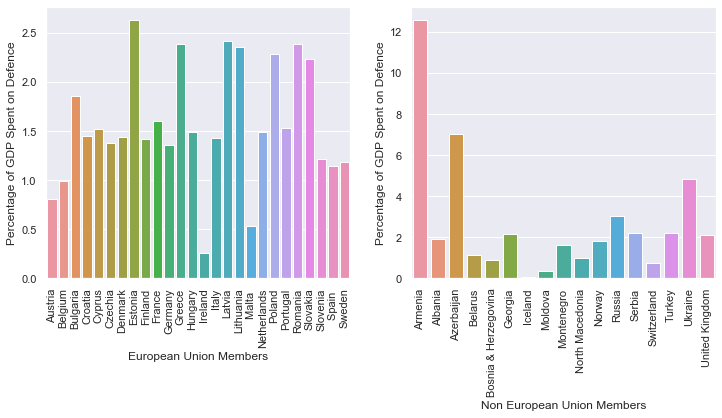
This graph represents GDP up against western European nations. Germany has the highest GDP in western Europe whilst the UK and France are closely followed in second and third place.

There does seem to be a more even distribution with western European nations over the eastern ones. However, there is still only a few countries here which have a large economy.

When we saw Russia’s GDP in the previous graph when put against its neighbours, it seemingly had a very large economy. However, Russia has an economy of $1.6 trillion, in western Europe though, you have Germany with around $3.7 trillion! If we had Russia here in the graph with the western Europeans, it would lie somewhere between Spain and Italy. This shows although Russia is a large country which boasts military prowess, its economy is average for the biggest European economies.

On the right here we have western Europe’s defence budget, this graph also looks similar to the GDP graph, the difference here is that the UK takes the lead with the highest military budget in Europe.

Below we can see a graphical representation of the percentage of GDP spent on defence between the EU and non-EU members. Interestingly only 9 of the 28 European NATO members pay up to 2% of their GDP on defence, which is what they’re supposed to be contributing.



Facts about military spending between the EU and its neighbours:

What is the difference between how much the EU spends in comparison to the non-EU members?

* The total spent by the EU sums to $208 billion, whilst the others outside of the EU collectively spends $146 billion towards defence. This means the EU spends 42% more on defence than those who aren’t members. Let’s compare this with the population between the union and its neighbours.

What is the combined total of members inside the EU and those outside of it?

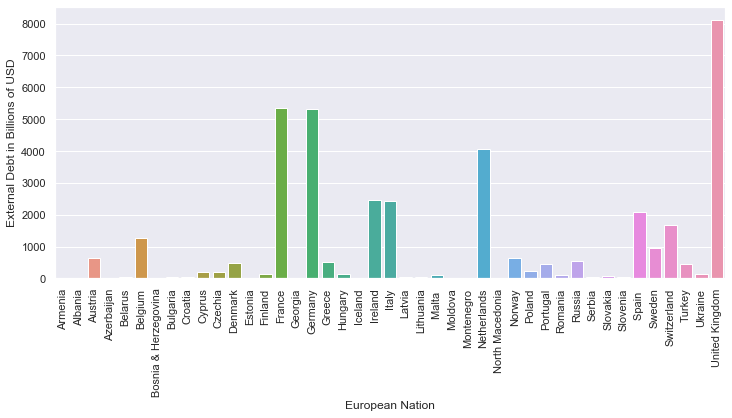
* Inside the EU there are 445 million people whilst the rest of Europe houses 404 million, that’s only a 10% increase, which means that averagely speaking, the EU spends more on defence per capita.

What is the mean spend of each nation between the EU and non-members?

* Calculating mean by taking the sum and dividing it between each nation will calculate the average spend per nation between the union and the non-members. The EU averagely spends $802 million on defence whilst those outside the union spend $861 million, this means the EU averagely spends 6.9% less than its non-members per country.

External debt:

Of all the European countries, we see here a graph of their external debt. External debt is what a country owes towards another nation, we can see here that the UK owes the most amount of money to the world by a long way, and is 2nd in the world in this list owing up to $8 trillion! which is almost 3x their entire economy.



Net Exports:

Net exports are defined as the difference between the spend on your imports to exports, a positive value means it’s the amount of export amount you have over your imports. A negative value means that your spending more on imports than exports and that also tells you how much over you spend on those imports.

Below are the top 5 countries with the most amount earned through exports:

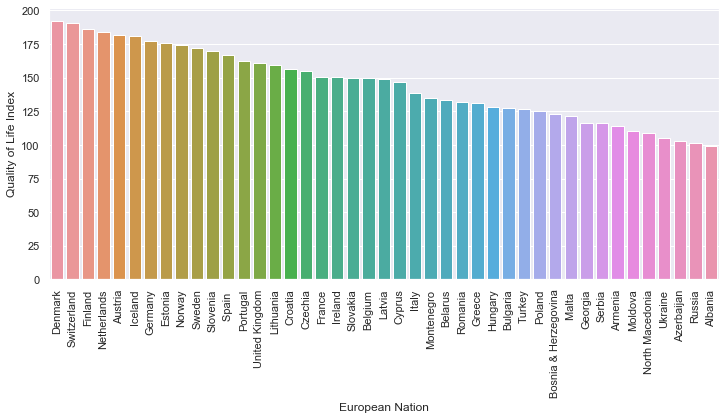
* Germany: $297 billion
* Ireland: $129 billion
* Russia: $124 billion
* Netherlands: $91 billion
* Italy: $72 billion

Here are the top 5 countries with the most import costs:

* UK: $166 billion
* France: $73 billion
* Turkey: $40 billion
* Spain: $32 billion
* Greece: $21 billion

When we take the mean between the EU members and non-EU members for combined average number of exports to imports, we get -$2.134 billion for non-members ($2.134 billion spent on imports) and $17.559 billion for EU members (money made on exports). This means that the EU makes money from exports whilst as a collective, the non-EU members are losing money to imports.

Quality of life index:

Below is a graph showing each European country in the order of their quality of life index scores from highest to lowest:

These are the factors which go into determining what a country scores on the quality of life index:

* Cost of living
* Purchasing power
* Affordability of housing
* Pollution
* Rates of crime
* The quality of healthcare
* Traffic

From the graph we can see that Denmark is on top with a score of 192.53, this means that Denmark has the best standard of living compared to its European neighbours. We can see here that the quality of life index is greatly dominated by the western nations on the left and eastern ones on the right, with Albania scoring the lowest at 99.20.

What is the average quality of life rating for those inside the European Union and outside of it?

* The EU has an average rating of 155.44 whilst the non-members averagely score 129.41. This tells us that those inside the EU are more likely to have a better quality of living than the non-members.

What about those between the west and the east, how much happier are the west than the east?

* The average quality of life rating for those in western Europe is 166.51 and for those in eastern Europe, the value is 129.78. This concludes that western Europe is 28% happier than their neighbours.

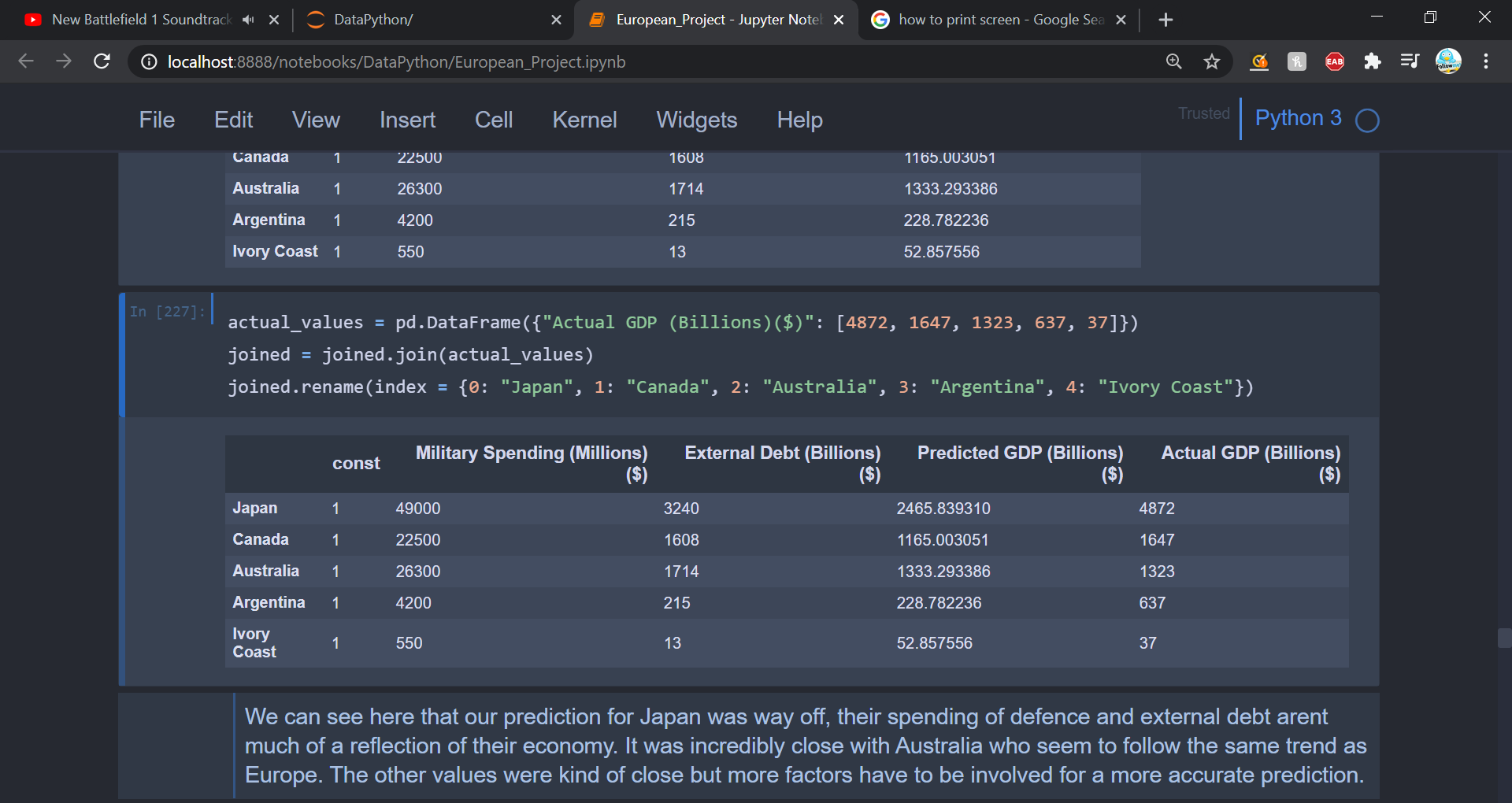
1st Regression:

For the first regression model, I am going to predict some countries GDP based on their military spending and external debt.

When we fit the OLS model, we have an R-Squared rating of 0.923, which means that the data here is highly accurate. We also have a very low Probability (F-Statistic) of 6 x 10-23 which means that our relationship between these variables are strong and hold great explanatory power.

I will attempt to predict the GDP of each of these nations below using the data from Europe to determine how accurate this model is:

* Japan: $49 billion in military spending and $3.2 trillion in external debt
* Australia: $26.3 billion in military spending and $1.7 trillion in external debt
* Canada: $22.5 billion in military spending and $1.6 trillion in external debt
* Argentina: $4.2 billion in military spending and $215 billion in external debt
* Ivory Coast: $550 million in military spending and $13 billion in external debt



Above is an image of the data frame made with the countries the model is trying to predict the GDP of with the predicted results vs the actual results.

The model was a long way off predicting Japans GDP, Japan has the 3rd largest GDP in the world. Their economy may not follow the same pattern as what it does in Europe, as they have a GDP of $4.9 trillion. Japan does have a large population, maybe if we had considered this into our regression, we may have had a different and closer result.

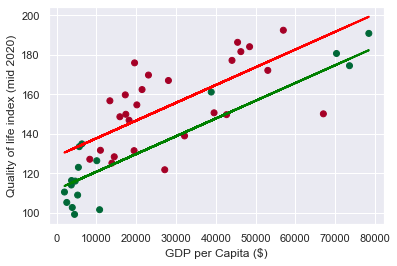
Canada slightly overhit but it wasn’t too far away, predicting a GDP of $1.2 trillion and having an actual GDP of $1.6 trillion.

With a predicted value of $1.333 trillion, Australia hit incredibly close to its true value of $1.323 trillion. The model has very closely predicted the GDP of Australia, this could be because the Australian economy works like that of those in Europe. As there are only two dependent variables in this regression it’s not certain as economies have many other factors in determining their values.

The model overestimated the prediction for Argentina’s GDP and underestimated for the Ivory Coast’s. This goes to show that even though the model was accurate for European countries, other nations economies around the world may have different driving factors in determining their worth.

2nd Regression:

In this next regression model, we’re not going to be predicting any values, but instead viewing the relationship between quality of life and GDP per capita, between members of the EU and non-members.



The formula for multiple linear regression is: ŷ = b0 + b1 x1 + b2 x2 + . . . + bk xk

Ŷ is the independent variable, the value we are aiming to reach.

The x values are our dependent variables, here they are GDP per capita and EU membership status.

The b values are coefficients, they are given to us in the summary table. Their purpose is for being quantifying factors for reaching the outcomes of ŷ. This now gives us two formulas.

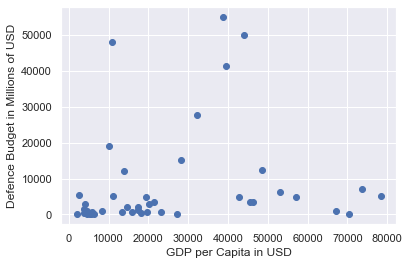
Ŷno = 111.7948 + 0.0009 \* GDP per capita + 0 (coef = 0) (non-members regression line)

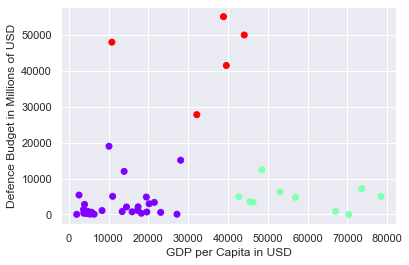
Ŷyes = 111.7948 + 0.0009 \* GDP per capita + 16.9577 \* EU membership (value = 1) (EU members)

The red dots represent European Union members, and the red regression line shows us the line of best fit between the values. The green dots and regression line represent those who aren’t members of the EU.

We can see that the poorer countries who also have a low quality of life all fall outside of the EU, this shows that to become a European Union member, there are requirements. The three richest countries at the far right of this graph are all non-members, yet this doesn’t affect the average wellbeing enough between the union and the non-members. We can see that the EU regression line is higher than that of the non-members, which means averagely the EU member states will be richer and have a better quality of life.

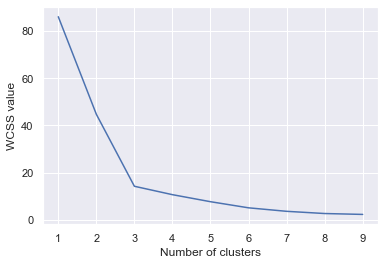
Cluster Analysis:



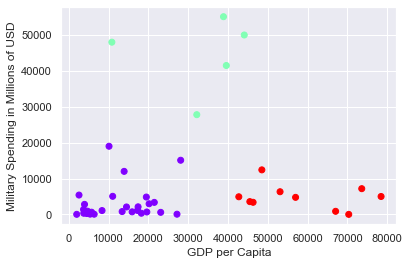
Here is a scatter plot of all the European nations and the relationship between GDP per capita and their defence budget. On this graph above, you most likely see that there are 3 main clusters to group these points into. I have used a K-Means clustering model to group these points into 3 categories and here are the results:

I needed to check that this was the optimal number of clusters for the most accurate model. Therefore, I needed to standardise the values, in this case, it wasn’t required as the two axes were both in increments of 10s of thousands.

To find the optimal number of clusters, we use the elbow method in which we find the Within-Cluster-Sum-of-Squares value (WCSS) which tells you the distance between data points and their centroids, a centroid is essentially the centre of gravity between a group of points and all the points will gather to their closest centroid. We create a plot of the WCSS value against the number of clusters we can use. If there is a sharp bend in the resultant graph, this tells us that it's corresponding x value is the optimal number of clusters to use.



Sometimes there are multiple bends and you need to use your intuition sometimes to determine the categories within the cluster, but in this case, the elbow method is telling us that 3 clusters are by far the most optimal. We can see the same graph again unchanged.



We can identify these clusters into three groups.

Poorer nations who don’t spend big (bottom right), richer nations who don’t spend big (bottom left) and the big economies who do spend big (top centre).

The poorer nations who don’t spend big, we can assume most of them are central and eastern European countries judging by the previous data we have found. Possibly in the area of the ex-Soviet Union members and Balkan nations.

The bottom right countries I believe we could find those in western and central Europe, countries I expect you would find high on the quality of life index, such as Denmark, Finland, Austria, Ireland. These types of countries don’t spend a lot of money on their military but are rich in GDP per capita.

Lastly, we have the big economies who do spend big. If I had to guess I would say Russia is the point at the top left of the graph, only because of their low GDP per Capita and high spending on defence. Other nations in this bloc would be Europe’s big economies such as the UK, France and Germany for example.