1 World’s Largest Companies 2019

1.1 Introduction

1.1.1 Background

The largest companies in the world are getting bigger. The stock valuation of a company can changes quickly. Microsoft briefly dethroned Apple as the most valuable enterprise in the world in 2018, Amazon beat both companies for the top slot in January 2019, and then Microsoft's valuation surged past rivals to cross the $1 trillion threshold for the first time in April 2019. How did that happen? While Apple—which is trying to move beyond hardware to services—suffered due to poor iPhone and MacBook sales, Microsoft's business model is centered around steadily growing streams of recurring revenues. You might not need a new smartphone or laptop every year, but if you purchased a software license, a cloud package or a videogame subscription, you are likely to buy it again in the future.

However, the total dollar value of a company’s outstanding shares, its market capitalization can be affected by a myriad of unpredictable factors. For example, stock markets have repeatedly lost value due to US President Donald Trump's tweets, including companies that have little or nothing to do with China.

Market capitalization indicates the cost of owning a piece of a company today. But does it reflect the company's objective value or inherent worth? Warren Buffett once famously said, “Nothing is further from the truth.” Still, as the founder of Berkshire Hathaway, the sixth most valuable business in the world by market capitalization, Buffett knows that behind such numbers always lies some truth.

This list comprises the world's largest companies by consolidated 2018 revenue, according to the Fortune Global 500 latest tally published on July 22, 2019. American retail corporation Walmart has been the world's largest company by revenue since 2014, with US$514 billion in revenue in 2018. Walmart was also the largest company in the world by revenue from 2002 to 2005, from 2007 to 2008, and from 2010 to 2017.

1.1.2 Problem

Data that might contribute to determining companies improvement might include the revenues, profits, the countries witch government owns 50% or more, and the domain of the industry. This project aims to cluster the he World's Largest Public Companies around the globe to know countries that are really powerful economically.

1.2 Data acquisition and cleaning

1.2.1 Data sources

Our data frame is from Wikipedia, the dataset is limited to the top 50 companies, all of which have annual revenues exceeding US$110 billion. 32 out of the 50 companies are either from the United States or from China. Only companies that publish financial data and report figures to a government agency are included. Therefore, this list is incomplete, as it excludes large companies such as Vitol, Cargill, Koch Industries, Schwarz Group and Kuwait Petroleum Corporation because they do not publish financial data.

1.2.2 Data cleaning

The data frame contains the columns Rank, Name, Industry, Revenue (USD millions), Profit (USD millions), Employees, Country and Ref.

First of all, we are going to delete the column Ref cause it's for the reference information in Wikipedia and we don't need it in our analysis.

Second step is to change some column names to an easiest manipulation.

Old column's name	New column's name
Revenue (USD millions)	Revenue_USDM
Profit (USD millions)	Profit_USDM

Third, we will remove the duplicated countries names in the column "Country".

Forth, we have the value "Netherlands / United Kingdom" separated by "/", let's duplicate the row for each county.

In the last step of data cleaning, the revenue column type is "object", we need to convert it to int.

We need to eliminate the $ sign, then the comma sign "," so we can convert the amounts to int type.

1.2.3 Data completing

In this section we will add some columns to our dataframe. To solve the problem using foursquare we will need the latitude and longitude for each country.

To do so, we are going to use the foursquare API to get the longitude and the latitude of each country.

1.3 Data visualization

In this section, we will create a world map with the localization of each country

Here is a capture of the visualized map.

1.4 Cluster countries based on the revenues of its companies

To know the counties having the largest companies based on the revenue, we have to cluster them. we chose to cluster them in five clusters.

We have grouped the companies by countries. (please check the notebook).

	index	Country	Latitude	Longitude	Revenue_USDM
0	14	United States	39.783730	-100.445882	4169049
1	1	China	35.000074	104.999927	2263520
2	3	Germany	51.083420	10.423447	602635
3	4	Japan	36.574844	139.239418	561157
4	0	United Kingdom	54.702354	-3.276575	396556
6	8	Saudi Arabia	25.624262	42.352833	355905
7	2	France	46.603354	1.888334	309684
9	7	Russia	64.686314	97.745306	250447
10	10	South Korea	36.638392	127.696119	221579
11	11	Switzerland	46.798562	8.231974	219754
12	9	Singapore	1.357107	103.819499	180744
13	12	Taiwan	23.598298	120.835363	175617
14	5	Netherlands	52.500170	5.748082	175009

Then, we clustered the countries based on the revenues from the companies to each country.

We chose the k_means alogorithm and we chose to cluster the countries to 5 clusters.

Here is the result of our kmeans clustering.

Cluster Labels	Country	index	Latitude	Longitude	Revenue_USDM
0	Netherlands	5	52.500170	5.748082	175009
	Russia	7	64.686314	97.745306	250447
	Singapore	9	1.357107	103.819499	180744
	South Korea	10	36.638392	127.696119	221579
	Switzerland	11	46.798562	8.231974	219754
	Taiwan	12	23.598298	120.835363	175617
1	United States	14	39.783730	-100.445882	4169049
2	China	1	35.000074	104.999927	2263520
3	Germany	3	51.083420	10.423447	602635
	Japan	4	36.574844	139.239418	561157
4	United Kingdom	0	54.702354	-3.276575	396556
	France	2	46.603354	1.888334	309684
	Saudi Arabia	8	25.624262	42.352833	355905

Let’s analyze this obvious result:

Our algorithm shows that the first county when it comes to revenues from worldwide largest companies is the united states. Then, we found china in the second cluster.

The third cluster contains Germany and Japan.

The fourth contains: France, Saudi Arabia and the United Kingdom.

The fifth cluster and the last contains Netherlands, Russia, Singapore, South Korea, Switzerland and Taiwan.

1.5 Clustering Visualization with folium

Finally, let's visualize the resulting clusters.

1.6 Conclusion

To conclude, we have to say that we were able in this project to cluster the World's Largest Public Companies contributions around the globe to know countries that are really powerful economically.