

# Similarity the capital city in the world

## Introduction

There are about 244 countries in the world. The capital city of those countries is all the same as they are the heart of the country. However, each city have their own characteristic. The question is can we do clustering for those capital cities? Can we group those capital city?

If we can group similar cities to each other, a business model which is success in a certain city might also be success if we implement it in the similar city in the other country

Characteristic of the city should be able to be identified by using similarity of the venues located in the center of the city. With this assumption if we use the data containing category of venues nearby the center of the city that we can get it from Foursquare location data, we should be able to do clustering for the cities.

Also, can we use population data to make data visualization for some information display?



## Data section

In order to solve the problem mentioned in the introduction session, we need to have the data set which includes the data of country name, capital name, population, coordinates of the capital cities. Luckily, there are available data in the internet but not all the required data are available in a single data source. Therefore, we need to pull the data from a couple of data sources and process it to be ready for solving our business problem.

Country name, capital name and population from wikipedia.org

[https://en.wikipedia.org/wiki/List\\_of\\_national\\_capitals\\_by\\_population](https://en.wikipedia.org/wiki/List_of_national_capitals_by_population)

List [ edit ]

Rank <sup>▲</sup>	Country/Territory <sup>↕</sup>	Capital <sup>↕</sup>	Population <sup>↕</sup>	Year <sup>↕</sup>	% of country's <sup>↕</sup> population
-	 Hong Kong	Hong Kong	7,482,500 <sup>[14]</sup>	2018	100%
1	 China PR	Beijing	21,542,000 <sup>[1]</sup>	2010	1.5%
2	 Japan	Tokyo	13,929,286 <sup>[2]</sup>	2017	11.03%
3	 Russia	Moscow	12,506,468 <sup>[3]</sup>	2011	8.52%
4	 DR Congo	Kinshasa	11,855,000 <sup>[4]</sup>	2012	12.9%
5	 Indonesia	Jakarta	10,075,310 <sup>[5]</sup>	2011	3.76%
6	 South Korea	Seoul	9,838,892 <sup>[6]</sup>	2015	19.03%
7	 Egypt	Cairo	9,500,000	2012	9.54%
8	 Mexico	Mexico City	8,918,653 <sup>[7]</sup>	2015	7.05%
9	 Bangladesh	Dhaka	8,906,039 <sup>[8]</sup>	2011	5.52%
10	 United Kingdom  England	London	8,908,081 <sup>[9]</sup>	2015	13.19%
11	 Peru	Lima	8,852,000 <sup>[10]</sup>	2012	26.74%
12	 Iran	Tehran	8,693,706	2014	10.53%
13	 Thailand	Bangkok	8,305,218 <sup>[11]</sup>	2010	11.96%
14	 Vietnam	Hanoi	7,781,631 <sup>[12]</sup>	2014	8.14%

Country name, capital name and its coordinates are available in provided webpage.

<https://www.jasom.net/list-of-capital-cities-with-latitude-and-longitude/>

```
Country,Capital,Latitude,Longitude
Abkhazia,Sukhumi,43.001525,41.023415
Afghanistan,Kabul,34.575503,69.240073
Aland Islands,Mariehamn,60.1,19.933333
Albania,Tirana,41.327546,19.818698
Algeria,Algiers,36.752887,3.042048
American Samoa,Pago Pago,-14.275632,-170.702036
Andorra,Andorra la Vella,42.506317,1.521835
Angola,Luanda,-8.839988,13.289437
Anguilla,The Valley,18.214813,-63.057441
Antarctica,South Pole,-90,0
Antigua and Barbuda,St. John's,17.12741,-61.846772
Argentina,Buenos Aires,-34.603684,-58.381559
Armenia,Yerevan,40.179186,44.499103
Aruba,Oranjestad,12.509204,-70.008631
Australia,Canberra,-35.282,149.128684
```

With the accessible data sources, we can process the data and get answer of the business problem. After all data are ready, we can make data visualization to report the project.