



André A. G. Barbosa

The Battle of the Cities

February, 2021

A Data Science project

Contents

- The Problem
- The Data
- The Analysis
- The Winner
- Conclusion



The Problem

- Bustopher Jones is a extremely wealthy gentleman who wishes to live in Portugal for a while but has some peculiar requirements about the location:
 - Mr. Jones is afraid of the consequences of global warming, so the city must not be by the sea;
 - He will buy a house by the city centre and will walk everywhere so he can flaunt his designer clothes, so the city must have various restaurants, bars, etc. in a 1km radius of his house;
 - It must be a well-known city, as Mr. Jones doesn't like to explain where he lives.



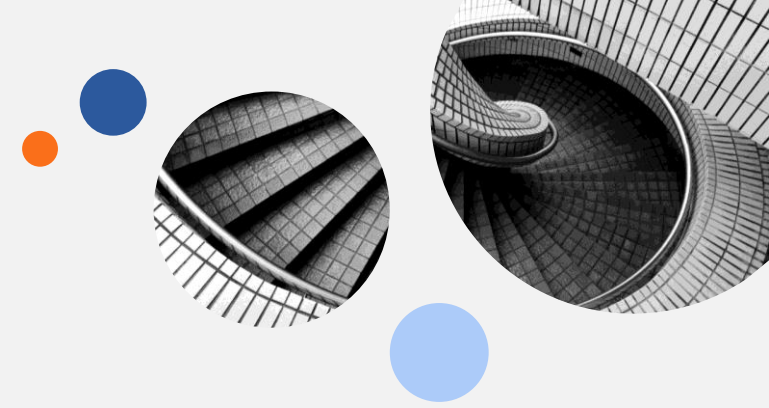


The data

The most valuable asset



The data sources



simplemaps

- The cities data (coordinates) came from a dataset available on <https://simplemaps.com/data/pt-cities>.
- This dataset is the free version of their more complete dataset

Foursquare

- The venues data comes from Foursquare, using their API.

The data

	city	lat	lng	country	iso2	admin_name	capital	population	population_proper
0	Lisbon	38.7452	-9.1604	Portugal	PT	Lisboa	primary	506654.0	506654.0
1	Vila Nova de Gaia	41.1333	-8.6167	Portugal	PT	Porto	minor	302295.0	302295.0
2	Porto	41.1495	-8.6108	Portugal	PT	Porto	admin	237591.0	237591.0
3	Braga	41.5333	-8.4167	Portugal	PT	Braga	admin	181494.0	181494.0
4	Matosinhos	41.2077	-8.6674	Portugal	PT	Porto	minor	175478.0	175478.0

The data frame head

The analysis

Clean the data

- Exclude the minor cities
- Exclude the cities by the sea
- Remove unneeded columns from the data frame.

Cluster the data

- Use the Elbow Curve to determine the best number of clusters
- Use K-means clustering to divide the country in regions

Chose the winner

- Obtain the venues of each city and perform a One Hot Encoding
- Find the most common venues on each cluster
- Find the most common venues of each city on the selected cluster
- Choose the winner



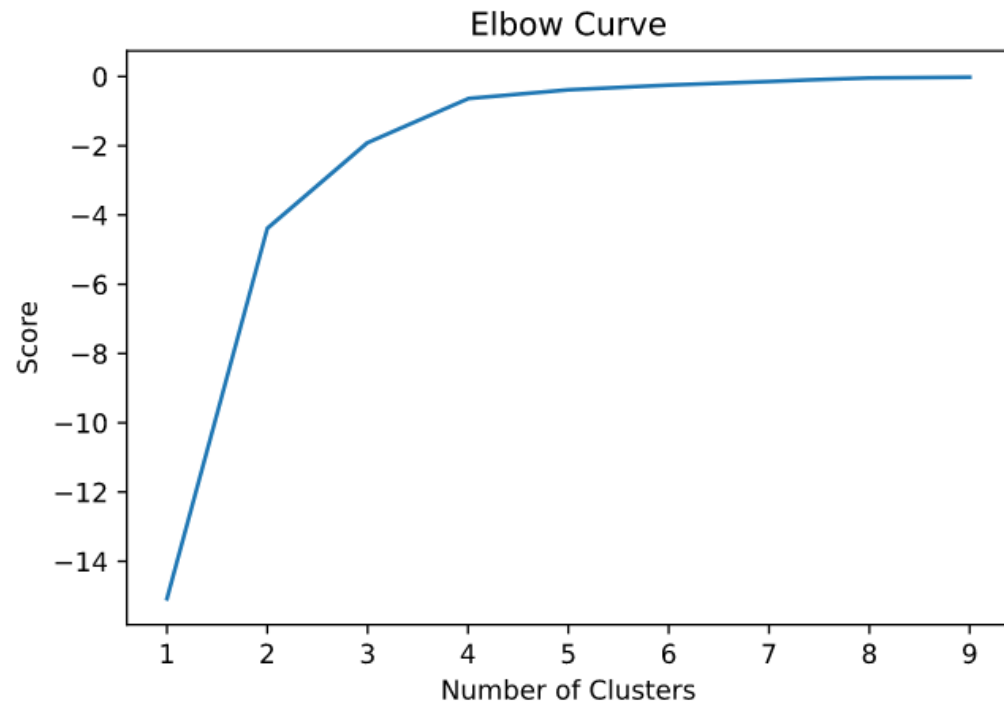
Cleaning the data

	city	lat	lng
0	Braga	41.5333	-8.4167
1	Coimbra	40.2111	-8.4291
2	Leiria	39.7431	-8.8069
3	Viseu	40.6667	-7.9167
4	Santarém	39.2369	-8.6850
5	Castelo Branco	39.8230	-7.4931
6	Évora	38.5667	-7.9000
7	Vila Real	41.3002	-7.7398
8	Guarda	40.5364	-7.2683
9	Beja	38.0333	-7.8833
10	Bragança	41.8000	-6.7500
11	Portalegre	39.3167	-7.4167

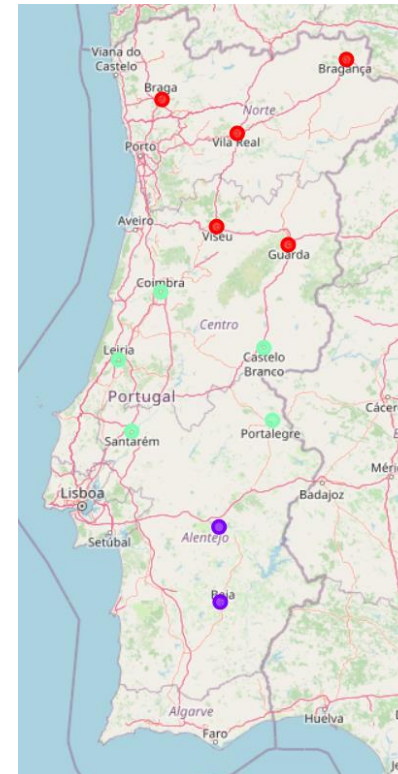
The clean data frame

Clustering the data

The Elbow Curve



The clusters on the map



Choose the winner – One Hot Encoding

	Cluster Labels	city	lat	lng	Art Gallery	Art Museum	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	...	Supermarket	Sushi Restaurant	Tapas Restaurant	Tea Room	Theater	Theme Park Ride / Attraction	Train Station	Vegetarian / Vegan Restaurant	Wine Bar	Winery
0	0	Braga	41.5333	-8.4167	0.000000	0.000000	0.040000	0.04	0.000000	0.080000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
1	2	Coimbra	40.2111	-8.4291	0.010000	0.000000	0.000000	0.00	0.000000	0.060000	...	0.010000	0.000000	0.010000	0.010000	0.000000	0.000000	0.010000	0.01	0.000000	0.000000
2	2	Leiria	39.7431	-8.8069	0.000000	0.000000	0.000000	0.00	0.000000	0.025974	...	0.038961	0.012987	0.064935	0.000000	0.000000	0.012987	0.000000	0.00	0.012987	0.000000
3	0	Viseu	40.6667	-7.9167	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.039216	0.000000	0.000000	0.000000	0.019608	0.000000	0.000000	0.00	0.039216	0.000000
4	2	Santarém	39.2369	-8.6850	0.000000	0.000000	0.033333	0.00	0.000000	0.000000	...	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
5	2	Castelo Branco	39.8230	-7.4931	0.000000	0.000000	0.000000	0.00	0.000000	0.047619	...	0.000000	0.000000	0.023810	0.000000	0.000000	0.000000	0.047619	0.00	0.000000	0.000000
6	1	Évora	38.5667	-7.9000	0.000000	0.026316	0.000000	0.00	0.000000	0.026316	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.026316	0.00	0.000000	0.000000
7	0	Vila Real	41.3002	-7.7398	0.000000	0.000000	0.000000	0.00	0.018868	0.075472	...	0.000000	0.000000	0.018868	0.018868	0.018868	0.000000	0.000000	0.00	0.000000	0.018868
8	0	Guarda	40.5364	-7.2683	0.000000	0.000000	0.000000	0.00	0.000000	0.074074	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
9	1	Beja	38.0333	-7.8833	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
10	0	Bragança	41.8000	-6.7500	0.058824	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
11	2	Portalegre	39.3167	-7.4167	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000

The encoded venues, grouped by city

Choose the winner – The chosen Cluster

	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	0	Café	Portuguese Restaurant	Restaurant	Bar	Bakery
1	1	Restaurant	Food Truck	Portuguese Restaurant	Historic Site	Café
2	2	Portuguese Restaurant	Pool	Bar	Café	Hotel

The most common venues of each cluster

Choose the winner – The top venues per city

	city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Braga	Café	Restaurant	Portuguese Restaurant	Bakery	Italian Restaurant
1	Viseu	Bar	Café	Fast Food Restaurant	Portuguese Restaurant	Hotel
2	Vila Real	Café	Fast Food Restaurant	Bakery	Bar	Park
3	Guarda	Café	Restaurant	Portuguese Restaurant	Bakery	Hotel
4	Bragança	Portuguese Restaurant	Bar	Plaza	Art Gallery	Café

The most common venues of each city on the chosen cluster



The winner is:

BRAGA

as the city where the top 5
most common venues are
“food places”.

Mr. Jones is a happy client.

Conclusion

Based on the analysis, we can conclude that Mr. Bustopher Jones would do good in choosing the city of Braga, being far enough from the sea, well known, and with eateries being the most common venues.

If Braga is not on the taste of Mr. Jones, he can safely choose Viseu or Guarda, also well known, even further from the sea, and with eateries being also common.

André Barbosa



Credits



Author

André A. G. Barbosa



Project link

<https://github.com/Barbosabyte/battle-of-the-cities>



Disclaimer

This work was done during the Covid-19 pandemic; thus, some information may be inaccurate.



Data sources

All data comes from simplemaps and Foursquare.



License

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.