



Clustering capitals in Brazil



Tourism is one of the largest and fastest-growing economic sectors

- Over the past six decades, tourism has experienced continued expansion and diversification
- In 2017 a total of 1.326 billion international tourist arrivals were recorded in destinations around the world
- International tourism receipts reached US\$ 1,340 billion in that year
- Considering this context, a market sector that can be very profitable is travel agencies



Problem and Interest

- Something that can be useful for those agencies is clustering the most important cities in a country
- Considering it, this work aims to cluster the capitals of each state in Brazil, 27 in total
- Obviously, travel agencies would be very interested in presenting their clients a nice summary of the cities in a country, because it would improve their service and consequently increase their profits



Data

- Latitudes and longitudes of cities in Brazil can be found [here](#)
- After cleaning, the final dataset had the format below

City	latitude	longitude
Aracaju	-10.90910	-37.0677
Belém	-1.45540	-48.4898
Belo Horizonte	-19.91020	-43.9266
Boa Vista	2.82384	-60.6753
Brasília	-15.77950	-47.9297

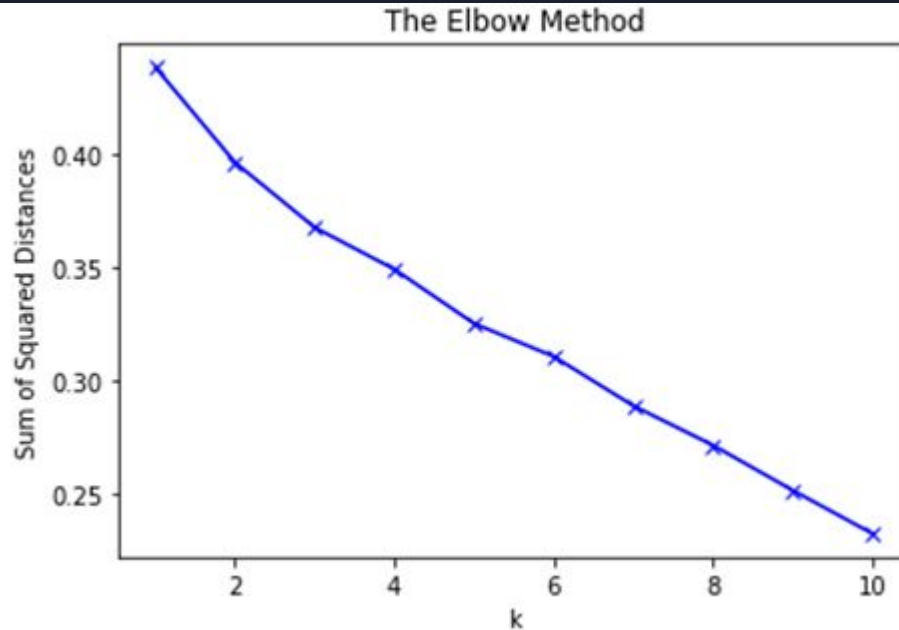


Foursquare data

- Final dataset, ready to be clustered:

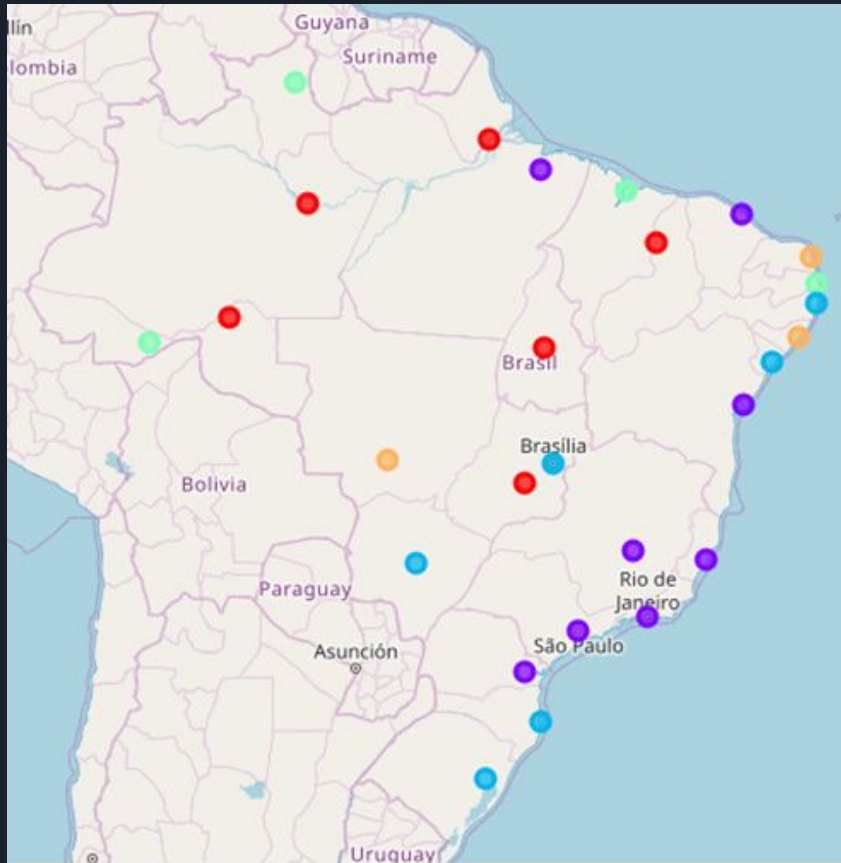
City	Acai House	Accessories Store	Adult Boutique	...
Aracaju	0.016667	0.000000	0.0	...
Belo Horizonte	0.000000	0.000000	0.0	...
Belém	0.029851	0.000000	0.0	...
⋮	⋮	⋮	⋮	⋮

K-means - Choosing K



- The result was not conclusive
- I decided to assign 5 to K, which is the number of regions in Brazil

Map of Brazil showing the clusters





Top 5 most common venues among the cluster

Cluster 1	
Capitals	Goiânia, Macapá, Manaus, Palmas, Porto Velho, Teresina
Top 5 most common venues	Pizza Place, Snack Place, Dessert Shop, Burger Joint, BBQ Joint
Cluster 2	
Capitals	Belém, Belo Horizonte, Curitiba, Fortaleza, Rio de Janeiro, Salvador, São Paulo, Vitória
Top 5 most common venues	Italian Restaurant, Theater, Plaza, Pizza Place, Park
Cluster 3	
Capitals	Aracaju, Brasília, Campo Grande, Florianópolis, Porto Alegre, Recife
Top 5 most common venues	Burger Joint, Pizza Place, Bakery, Italian Restaurant, Northeastern Brazilian Restaurant
Cluster 4	
Capitals	Boa Vista, João Pessoa, Rio Branco, São Luís
Top 5 most common venues	Bakery, Plaza, Seafood Restaurant, Snack Place, Burger Joint
Cluster 5	
Capitals	Cuiabá, Maceió, Natal
Top 5 most common venues	Dessert Shop, Italian Restaurant, Bakery, Burger Joint, Japanese Restaurant



Table of recommendations

	Recommendation
Cluster 1	Where you can try a really good meat
Cluster 2	Italian culture and theaters
Cluster 3	Diversified gastronomy
Cluster 4	It's a must If you like seafood
Cluster 5	Very good dessert shops and Japanese restaurants



Conclusion

- I found very useful insights showing that cities that belong to the same region are not necessarily similar to each other
- As the final result, I could generate a recommendation guide that can be used by clients to choose the cities according to their preferences
- To improve this work, data from other sources besides Foursquare could be used, because there are many other characteristics that a client would like to know about a place that he is intending to visit, such as attractions, prices, security and so on