World Cities

Final Report

1. Table of Contents

1. Introduction **3**

2. Methodology **4**

2.1. Cities 4

2.2. Features 4

2.3. Clustering 4

3. Results **5**

4. Discussion **7**

5. Conclusion **8**

6. References **9**

7. Appendix A **10**

2. Introduction

Travel agencies usually need to recommend travel packages and routes to customers who don't exactly know where they want to go. This recommendation must be given according to the customer preferences, which can be obtained by their opinion about other visited cities.
In this subject, it is useful for a travel agency to know how the most famous cities in the world are related to themselves, and which cities are in the same clusters.
Therefore, collecting information about the cities allow the application of a machine learning technique called clusterization, where similar instances are grouped.

3. Methodology

The project consists in choosing the cities to cluster, and collection information that will be used as features for each city. In this section, the cities used will be described, followed by the selected features and how they were collected.

3.1. Data

The data is formed features of 25 cities around the world. In the next paragraphs, it will be described how the cities and features were chosen and manipulated.

3.1.1. Cities

The cities where choosens based on the website <https://www.worldatlas.com/articles/the-most-popular-cities-in-the-world-to-visit.html>, which present the 25 most popular cities to visit in the world. The information is presented in a html table, and a script was used to transform this data in a pandas DataFrame, so it could be used further.

3.1.2. Features

The Foursquare API allows to request information about venues located in the city. Each venue is described by a series of data, one of them beeing category. Thus, the data selected to be used as a city feature was the number of venues for each venue category found in all the 25 cities. The list of categories can be found at Appendix A.

3.2. Clustering

The KMeans algorithm was used to cluster the cities in similar groups, based on its features. This machine learning technique finds the cities that are more similar and group them together, at the same time that separate different cities in different clusters.

4. Results

The Figure 1 shows a map where each city is represented by a circle, and the circle color represents the city cluster. Cities with the same color where clustered together.

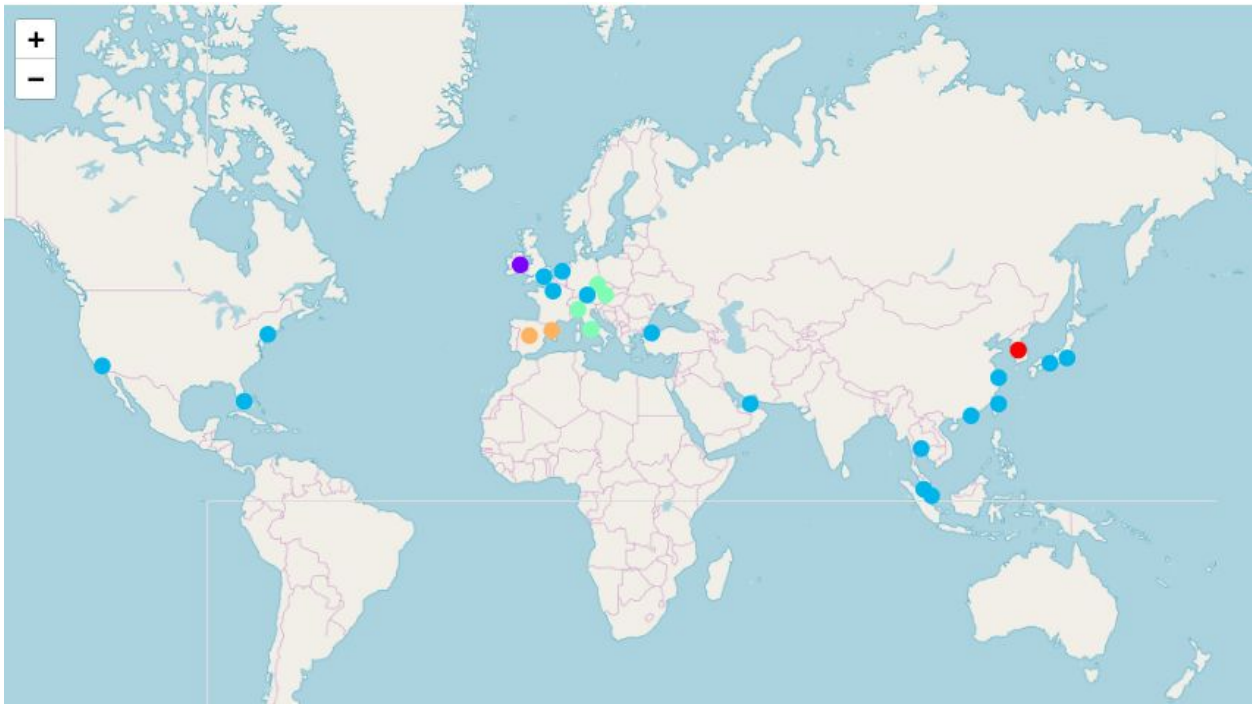

Figure 1. Result Map

The following Figure presents the relation of each city with its cluster.

City	ClusterId
Bangkok	0
London	0
Paris	0
Dubai	0
New York	0
Singapore	0
Kuala Lumpur	0
Istanbul	3
Tokyo	0
Seoul	4
Hong Kong	0
Barcelona	2
Amsterdam	0
Milan	1
Taipei	0
Rome	1
Osaka	0
Vienna	1
Shanghai	0
Prague	1
Los Angeles	0
Madrid	2
Munich	0
Miami	0
Dublin	0

Figure 2. Cities clusters

5. Discussion

As seen, the cities where clustered in similar groups, based on the categories of venues. This information is useful for detecting related and non-related cities.

6. Conclusion

As proposed, the cities where clustered based on their features. This information may be used to solve different problems, among them the travel agency situation described earlier.
As improvement, different clustering techniques may be applied, comparing the results and determining better ways to cluster cities.

7. References

Jessica Dillinger. 2018. *The Most Popular Cities In The World To Visit*. [ONLINE] Available at: https://www.worldatlas.com/articles/the-most-popular-cities-in-the-world-to-visit.html. [Accessed 4 December 2018].

Foursquare Developers. [ONLINE] Available at: https://developer.foursquare.com/. [Accessed 4 December 2018]

8. Appendix A

City Features

Fast Food Restaurant	Pub	Memorial Site	Alsatian Restaurant	Hunan Restaurant
Gukbap Restaurant	Italian Restaurant	Organic Grocery	Hotel Bar	Udon Restaurant
Argentinian Restaurant	Yoga Studio	Travel Agency	Dive Bar	Chocolate Shop
Boutique	Sake Bar	Post Office	Restaurant	Night Market
Wine Bar	Shopping Mall	Bookstore	Diner	Coffee Shop
Unagi Restaurant	Tailor Shop	Middle Eastern Restaurant	Pizza Place	Gourmet Shop
Fish Market	Bistro	Video Game Store	Music Store	Bossam/Jokbal Restaurant
Botanical Garden	French Restaurant	Szechuan Restaurant	Hot Dog Joint	Bank
European Restaurant	Leather Goods Store	Waterfront	Ice Cream Shop	Deli / Bodega
Tour Provider	Korean Restaurant	Japanese Restaurant	Cuban Restaurant	Eastern
Yoshoku Restaurant	Filipino Restaurant	German Restaurant	Bakery	
Cupcake Shop	Gym	Sushi Restaurant	Gift Shop	
Cheese Shop	Indian Restaurant	Building	Dumpling Restaurant	
	Athletics & Sports	Cocktail Bar	Marijuana Dispensary	
		Intersection	River	
			Kebab Restaurant	