

How Do Big Cities Differ From Each Other?

Applied Data Science Capstone project by Sophie Tanny

Introduction

We all have pre-conceived ideas regarding the character of old big cities: London is the capital of culture; Milan is the capital of shopping and fashion. But how much of this pre-conceived knowledge is, in fact, true or up-to-date? In a recent trip I made to London I was surprised to find an abundance of international cuisine restaurants, hardly any book stores and only one art supply store. That's not what I remembered from my previous trip there, ten years ago!

Imagine how much I could gain, as a tourist, from consulting with a travel agent who is up-to-date with "what's up" everywhere. In this project, I'll be categorizing big cities around the world according to their distribution of various venues, in order to more clearly define the touristic character of each city.

When selling touristic packages to customers, travel agents can use this categorization to better fit a destination to the specific customer, thus creating a more satisfactory experience for the customer, and a higher chance of their not only enjoying their trip but also becoming returning customers to said travel agent. Therefore, I've gathered a list of the most visited big cities in the world and analyzed them regarding the most abundant touristic venues and clustered them into seven distinct clusters. I believe this information could benefit any travel agent when selling themed touristic packages to tourists, offering a variety to better fit personal tastes, and creating a better experience for the customers.

Data and Tools

Data Sources

The list of cities for this project was gathered from a few online sources: Condé Nast Traveller, World Atlas, Mental Floss and CNN Travel (URLs to follow). Additionally, I

have interviewed the owner of Traveling Star Travel Agency (Tel Aviv, Israel), for their insights regarding popular touristic urban destinations.

For the Geospatial data, I have used a database provided by Simple Maps – Geographic Data Products. Their data is accurate and up-to-date, and was constructed using authoritative sources such as the NGIA, US Geological Survey, US Census Bureau, and NASA .

The following figure shows the distribution of the selected cities around the world:



Tools and Methodology

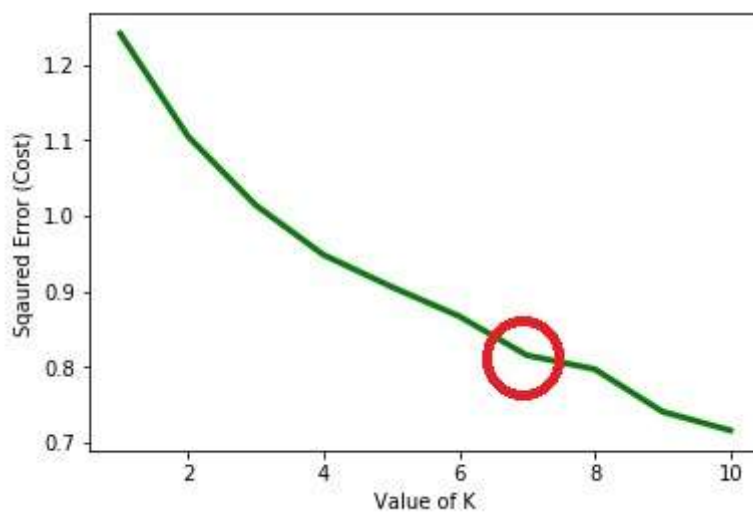
For the exploration of the various venues in each city, I have used the FourSquare API, and harvested my data from it.

For the analysis itself, I've used various python libraries: Numpy and Pandas for handling the data, Matplotlib for graph plots, Folium for map displays and SKlearn for clustering the cities into distinct groups via KMeans analysis. Since the data analyzed is categorical, this type of analysis will give the most comprehensive insights.

Results

Initial analysis of the database has shown that there are 381 unique venue categories. Those were tested along 54 cities and results were given for the ten most common venue category in each city.

The KMeans elbow test result suggested that the cities could be divided into seven distinct clusters, as seen in the following figure:



Their analysis produced the following seven clusters. In the map below you can see their dispersal around the world.



Cluster 0 (red circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
8	Cape Town	0	Hotel	Coffee Shop	Café	Indian Restaurant	Gym	Bakery	Cocktail Bar	Italian Restaurant	Restaurant	Burger Joint
21	Dublin	0	Café	Coffee Shop	Pub	Park	Restaurant	Burger Joint	Plaza	Pizza Place	Indian Restaurant	Ice Cream Shop
28	Glasgow	0	Bar	Coffee Shop	Restaurant	Hotel	Café	Seafood Restaurant	Cocktail Bar	Steakhouse	Whisky Bar	Beer Bar
35	Budapest	0	Coffee Shop	Theater	Pizza Place	Ice Cream Shop	Hotel	Restaurant	Park	Bar	Bakery	Burger Joint
37	Bucharest	0	Coffee Shop	Hotel	Tea Room	Dessert Shop	Restaurant	Pizza Place	Pub	Bar	Romanian Restaurant	Theater
38	Larnaka	0	Coffee Shop	Hotel	Greek Restaurant	Bar	Seafood Restaurant	Beach	Café	Supermarket	Fish Taverna	Lounge
39	Johannesburg	0	Coffee Shop	Hotel	Café	African Restaurant	Pizza Place	Restaurant	Steakhouse	Art Gallery	Park	Italian Restaurant

Cluster 0 contains the cities: Cape town, Dublin, Glasgow, Budapest, Bucharest, Larnaka and Johannesburg. The predominant venues in it are hotels and coffee shops, followed by ice cream and pizza places.

Cluster 1 (purple circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	New York	1	Park	Pizza Place	Italian Restaurant	Yoga Studio	Ice Cream Shop	Theater	Indie Movie Theater	Bookstore	Wine Shop	Cocktail Bar
7	Sydney	1	Café	Coffee Shop	Bakery	Bar	Ice Cream Shop	Brewery	Thai Restaurant	Restaurant	Pub	Whisky Bar
10	Berlin	1	Coffee Shop	Bookstore	Concert Hall	Park	Gourmet Shop	Café	Sandwich Place	Vegetarian / Vegan Restaurant	Ice Cream Shop	Bakery
12	San Francisco	1	Park	Ice Cream Shop	Coffee Shop	Bakery	Pizza Place	Yoga Studio	Bookstore	Grocery Store	Scenic Lookout	New American Restaurant
13	Melbourne	1	Café	Cocktail Bar	Coffee Shop	Park	Australian Restaurant	Theater	Italian Restaurant	Wine Bar	Plaza	Music Venue
15	Copenhagen	1	Beer Bar	Café	Scandinavian Restaurant	Cocktail Bar	Coffee Shop	Wine Bar	Sandwich Place	Movie Theater	Park	Art Museum
22	Chicago	1	Coffee Shop	Theater	Brewery	Italian Restaurant	New American Restaurant	Deli / Bodega	Hotel	Bar	Japanese Restaurant	Park
24	Toronto	1	Café	Italian Restaurant	Bar	Park	Grocery Store	Asian Restaurant	Japanese Restaurant	Sandwich Place	BBQ Joint	Vegetarian / Vegan Restaurant
25	Tel Aviv-Yafo	1	Beach	Hotel	Café	Bakery	Ice Cream Shop	Bar	Cocktail Bar	Seafood Restaurant	Park	Plaza
27	Bristol	1	Pub	Bar	Café	Coffee Shop	Italian Restaurant	Theater	Music Venue	Cocktail Bar	Park	Pizza Place
29	Athens	1	Historic Site	Coffee Shop	Bar	Café	Greek Restaurant	Wine Bar	Cocktail Bar	Bookstore	Meze Restaurant	History Museum
30	Istanbul	1	Waterfront	Concert Hall	Hotel	Park	Yoga Studio	Historic Site	Art Gallery	Performing Arts Venue	Restaurant	Soccer Stadium
31	Moscow	1	Yoga Studio	Art Gallery	Hotel	Park	Road	Plaza	Theater	Pizza Place	Bookstore	Jewelry Store
32	Brussels	1	Bar	Italian Restaurant	Plaza	Hotel	Beer Bar	Bakery	Coffee Shop	Sushi Restaurant	Boutique	Bookstore
33	Warsaw	1	Plaza	Café	Coffee Shop	Sushi Restaurant	Bakery	Park	Vegetarian / Vegan Restaurant	Breakfast Spot	Beer Bar	Hotel
34	Kraków	1	Café	Hotel	Bar	Food Truck	Park	Ice Cream Shop	Church	Hostel	Restaurant	Plaza
36	Prague	1	Café	Park	Coffee Shop	Bakery	Bistro	Burger Joint	Hotel	Cocktail Bar	Restaurant	Wine Bar
44	Phuket	1	Thai Restaurant	Coffee Shop	Café	Dim Sum Restaurant	Hotel	Ice Cream Shop	Chinese Restaurant	Bar	Seafood Restaurant	Italian Restaurant
45	Antalya	1	Café	Coffee Shop	Restaurant	Park	Bookstore	Historic Site	Museum	Seafood Restaurant	Bar	Burger Joint
53	Rio de Janeiro	1	Park	Bookstore	Historic Site	Coffee Shop	Bar	Brazilian Restaurant	Italian Restaurant	Middle Eastern Restaurant	Pizza Place	Music Venue

Cluster 1 contains the cities: New York, Sydney, Berlin, San Francisco, Melbourne, København (Copenhagen), Chicago, Toronto, Tel Aviv, Bristol, Athens, Istanbul, Moscow, Brussels, Warsaw, Krakow, Prague, Phuket, Antalya and Rio De Janeiro. It is the largest cluster, with mainly parks, beaches and coffee shops. Yoga studios and wine bars also appear frequently in these places, as well as bookstores and theatres.

Cluster 2 (blue circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Bangkok	2	Hotel	Coffee Shop	Shopping Mall	Thai Restaurant	Dessert Shop	Asian Restaurant	Noodle House	Massage Studio	Bookstore	Palace
14	Hong Kong	2	Hotel	Park	Chinese Restaurant	Café	Dessert Shop	Japanese Restaurant	Dim Sum Restaurant	Spa	Shopping Mall	Dumpling Restaurant
19	Dubai	2	Hotel	Coffee Shop	Café	Beach	Jewelry Store	Restaurant	Ice Cream Shop	French Restaurant	Middle Eastern Restaurant	American Restaurant
42	Kuala Lumpur	2	Hotel	Café	Chinese Restaurant	Lounge	Shopping Mall	Park	Indian Restaurant	Spa	Malay Restaurant	Japanese Restaurant
46	Macau	2	Hotel	Café	Portuguese Restaurant	Coffee Shop	Chinese Restaurant	Macanese Restaurant	Italian Restaurant	Plaza	Cantonese Restaurant	Bakery
48	Guangzhou	2	Hotel	Coffee Shop	Turkish Restaurant	Shopping Mall	Chinese Restaurant	Park	Nightclub	Cocktail Bar	Bar	Seafood Restaurant
49	Taipei	2	Café	Hotel	Japanese Restaurant	Dumpling Restaurant	Dessert Shop	Noodle House	Bookstore	Bakery	BBQ Joint	Park
51	Shenzhen	2	Hotel	Coffee Shop	Café	Shopping Mall	Chinese Restaurant	Park	Electronics Store	Hotpot Restaurant	Seafood Restaurant	Nightclub

Cluster 2 contains the cities: Bangkok, Hong Kong, Dubai, Kuala Lumpur, Macau, Guangzhou, Taipei and Shenzhen. All these cities are in the far east, with the exception of Dubai in the middle east. Here you can find mostly hotels, coffee shops (and cafés) and International cuisine.

Cluster 3 (pale blue circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
47	Mumbai	3	Indian Restaurant	Café	Hotel	Scenic Lookout	Bar	Bakery	Dessert Shop	Deli / Bodega	Asian Restaurant	Restaurant
50	Delhi	3	Indian Restaurant	Café	Hotel	Asian Restaurant	Bar	Monument / Landmark	Bakery	Fast Food Restaurant	Restaurant	Market

Cluster 3 contains the cities Mumbai and Delhi, both in India. Both feature Indian restaurants, cafés and hotels in the first places, and are very similar to each other along the rest of the venues list.

Cluster 4 (green circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Barcelona	4	Hotel	Tapas Restaurant	Spanish Restaurant	Coffee Shop	Plaza	Burger Joint	Wine Bar	Cocktail Bar	Ice Cream Shop	Bookstore
6	Amsterdam	4	Hotel	Park	Breakfast Spot	Ice Cream Shop	French Restaurant	Bar	Coffee Shop	Restaurant	Bakery	Café
9	Lisbon	4	Hotel	Portuguese Restaurant	Plaza	Coffee Shop	Garden	Scenic Lookout	Bar	Ice Cream Shop	Lounge	Cocktail Bar
16	Tokyo	4	Hotel	Tonkatsu Restaurant	BBQ Joint	Coffee Shop	Sake Bar	Garden	Wagashi Place	Chinese Restaurant	Sushi Restaurant	Japanese Curry Restaurant
17	Vancouver	4	Hotel	Coffee Shop	Seafood Restaurant	Bakery	Japanese Restaurant	Dessert Shop	Restaurant	Park	Sandwich Place	Trail
18	Vienna	4	Hotel	Plaza	Austrian Restaurant	Ice Cream Shop	Café	Restaurant	Bar	Palace	Concert Hall	Art Museum
20	Madrid	4	Spanish Restaurant	Restaurant	Hotel	Park	Art Museum	Tapas Restaurant	Art Gallery	Plaza	Café	Theater
23	Los Angeles	4	Hotel	American Restaurant	Ice Cream Shop	Sushi Restaurant	Boutique	Park	Coffee Shop	Theme Park Ride / Attraction	Italian Restaurant	Trail
26	London	4	Hotel	Steakhouse	Theater	Art Museum	Coffee Shop	Grocery Store	Plaza	Pizza Place	Scenic Lookout	Indian Restaurant
40	Cairo	4	Hotel	Historic Site	Middle Eastern Restaurant	Café	Lounge	Hotel Bar	Pastry Shop	Performing Arts Venue	Theater	Italian Restaurant
41	Lagos	4	Hotel	African Restaurant	Lounge	Nightclub	Restaurant	Pizza Place	Movie Theater	Japanese Restaurant	Ice Cream Shop	Bar
43	Singapore	4	Hotel	Event Space	Waterfront	Park	Cocktail Bar	Wine Bar	Plaza	Buffet	Steakhouse	Lounge

Cluster 4 contains the cities: Barcelona, Amsterdam, Lisbon, Tokyo, Vancouver, Vienna, Madrid, Los Angeles, London, Cairo, Lagos and Singapore. Here, again, hotels pick up the first place in common venues, followed by local cuisine. Other venues that are more abundant here are outdoor places like plazas, parks and trails.

Cluster 5 (yellow circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
52	Beijing	5	Historic Site	Hotel	Café	Park	Chinese Restaurant	Bar	Dumpling Restaurant	Yunnan Restaurant	Coffee Shop	Brewery

Cluster 5 is a single city. Beijing features more historic sites than any other venue, as well as local cuisine.

Cluster 6 (orange circle)

	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Paris	6	Plaza	Hotel	French Restaurant	Art Museum	Garden	Historic Site	Gourmet Shop	Bookstore	Cocktail Bar	Concert Hall
1	Rome	6	Plaza	Historic Site	Ice Cream Shop	Sandwich Place	Monument / Landmark	Italian Restaurant	Church	Art Museum	Fountain	Wine Bar
4	Florence	6	Italian Restaurant	Ice Cream Shop	Hotel	Plaza	Sandwich Place	Museum	Art Museum	Trattoria/Osteria	Church	Boutique
5	Venice	6	Italian Restaurant	Hotel	Plaza	Ice Cream Shop	Wine Bar	Winery	Museum	Pizza Place	Church	Historic Site

Cluster 6 contains the cities: Paris, Rome, Florence and Venice. It features mostly plazas and local cuisine, ice cream places and museums.

Discussion

Most of the cities analyzed in this study are located in Europe and the USA, which lends a tendency of similarity between them. The question is, does that mean "You've seen one – you've seen them all"? The intuitive answer is "no". We all have our favourite city in the world, so much so that we sometimes tend to keep returning to the same place and avoid exploring other cities.

Looking at the clustering results, some observations can be made:

Cluster 1, predominant with parks, beaches and coffee shops, lends a feeling of being attractive to young tourists – people who find joy in being outside, just sitting around and relaxing, maybe people watching.

Cluster 2 seems more business-like. The predominant venues are hotels, coffee shops and international restaurants. This feels aimed at people with more money to spend and a cosmopolitan state of mind. Some of the cities here are considered expensive, such as Dubai and Hong Kong. This research might benefit from another analysis of the data, in relation to the level of expensiveness.

Clusters 3 and 5 are unique in that they contain single cities: Cluster 3 contains the Indian cities of Mumbai and Delhi. Both appear to be characterized by traditional restaurants and other food places, hotels being in third place. Beijing in cluster 4, on the other hand, is abundant with historic sites, hotels and local cuisine. Are these two groups really all that different from other cities in the world? When thinking of a city with historic sites and local food, Rome comes to mind. Looking at the results, it seems the only big difference between Rome and Beijing (re types of venues) is that hotels in Rome don't make it into the 10 most common venues. This might indicate that a refinement of the test is in place.

Clusters 0 and 4 seem at first rather similar to each other. Hotels feature in first place in both clusters, with light-eating food places in cluster 0 and local cuisine in cluster 4. The big difference is that in cluster 4 we can also find outdoor places like plazas and parks. It seems like the cities in cluster 4 are more suited for families traveling together. Cluster 0, on the other hand, seems to have nothing special in it. It may be that with a more specialized analysis, this cluster would merge into another cluster.

The last cluster to discuss is cluster 6, which contains three Italian cities and Paris. The abundance of plazas and local cuisine, as well as ice cream places and museums, lends itself towards tourists who like to walk around the "feel the city".

Conclusion

The analysis I have performed shows that indeed, not all big cities are alike. My conception of London as "city of bookstores" has had to change, seeing that bookstores hadn't even made it into the ten most common venues. Barcelona, on the other hand, does have bookstores in tenth place, and I have only to regret I don't read Spanish. At large, we can see that different types of people would be attracted to different types of cities, and that such a distinction can be made when considering and putting together a touristic package to sell to customers.

However, further tests and analyses are still due, to sharpen the differences. My suggestions for further analysis include:

Refinement of the types of venues, for example: coffee shops and cafés should be a single category, as do museums and galleries. This way, instead of 381 unique categories there might be a lot less, and so the differences between the clusters would be sharper.

Adding more features to test the difference between the cities, such as how expensive the city is, is it a capital city or not and its size, might add more input relevant to the clustering process.

To summarize, big cities are not all alike. We know this intuitively, and the general looks of the city are not measurable in this method. Still, we can better refine our decision when booking a trip or when selling one to our customers.

Sources

- Condé Nast Traveller list of best cities in the world:
<https://www.cntraveller.com/gallery/best-cities-in-the-world>)
- World Atlas – African attractions:
<https://www.worldatlas.com/articles/african-attractions-which-are-the-most-visited-cities-in-africa.html>
- Mental Floss – Most visited cities in the world:
<http://mentalfloss.com/article/558406/worlds-10-most-visited-cities-and-what-it-costs-spend-day-there>
- CNN Travel - Most visited cities in the world:
<https://edition.cnn.com/travel/article/most-visited-cities-euromonitor-2018/index.html>
- Simple Maps: <https://simplemaps.com/data/world-cities>
- FourSquare API: <https://foursquare.com/developers/apps>
- Traveling Star Travel Agency: <https://www.facebook.com/TravelingStar1/>