Cities' Secrets for Success

Mingus Knox White

I. Introduction

International travel had once been an exciting getaway from everyday life, but due to the ongoing coronavirus pandemic, it is little more than a fond memory and a hopeful look into the vaccinated future. Tourism had been a prominent industry, providing jobs and boosting the economy of many countries around the world, but has now all but disappeared. Thus, it is more important than ever that, in the wake of the pandemic, the industry thrives, especially for less visited destinations.

There are countless beautiful locations across the globe, but it may be difficult for them to become known when compared with cities such as Paris or Tokyo, that have had recognition for many years. Comparing the attractions of popular cities could provide insight on what makes them stand apart. The venues in the top 20 most visited cities in the world should show different reasons the cities are so popular.

A city not found in the top 20 could find which popular destinations are most similar to their own, modifying how they market tourism in order to associate (or disassociate) from their similar cities. This could mean opening more resorts, providing infrastructure for nightlife, or prioritizing upkeep of local parks, depending on how the city wants to be seen by new visitors.

This information could also be used by airlines, since it would be likely that if someone enjoyed visiting one city, they would also enjoy visiting a similar city, and could be recommended a flight to a destination they may not have already considered. Similarly, travelers themselves could find cities similar to ones they have enjoyed in the past; to visit, or even move permanently.

II. Data

The top 20 most visited cities were found in a report by Mastercard, titled, "Global Destination Cities Index 2019.[i]" This data is based on the number of overnight international visitors, with Bangkok being the most popular (22.78M visitors), and Hong Kong being the 20th most visited (8.23M visitors). The report showed data from the previous two years, in which the number of visitors for each city had remained relatively the same or grown, so it is safe to assume these cities will remain popular after the worldwide pandemic. The data itself needed to be manually entered, only taking the city names, countries, and the positions in the top 20. The position is to used to scale the size of the points when the cities are plotted on a world map; the higher the visitation of the city, the larger the point.

The ability for cities not found on the list to use the data requires an input of cities not found in the Mastercard report. For the sake of this report, the following cities have been inputted: Cape Town, South Africa; Melbourne, Australia; Los Angeles, USA; Toronto, Canada; and Rio de Janeiro, Brazil. These cities were chosen as they are also major tourist destinations, and are found in areas of the globe that have few cities in the top 20, though other cities could be selected instead.

The GPS coordinates for each city are used in a call to the Foursquare API, returning (up to) the top 100 venues in the city. The API call does not specify a time of day, day of the week, or search radius, so that the top most suggested venues in all of the city are recommended. The categories of each venue are saved to their respective city, in order to understand the types of popular venues found in the city.

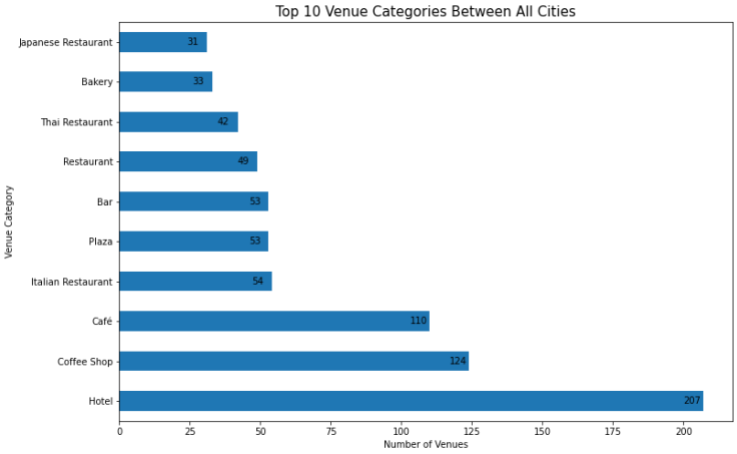
In addition to available venues, weather has a large impact on what sets a city apart from others. For this project, the average high and low temperatures of each month are scraped for each city from "weatherspark.com." This site was chosen due to having a consistent way to access data for cities around the world, as well as containing other forms of weather data, such as average rainfall and humidity. This additional data may be useful to expand the scope of the project in the future.

III. Methodology

As the goal of this project is to group similar cities, but there is no proven measure of how closely related different cities are, the use of a clustering machine learning model made the most sense. Knowing a group of similar cities is very useful to determine the overall similarities within each group, but finding the single most similar city and having a measurement of how similar it is could also be helpful. This is why a hierarchical clustering method was determined to be optimal. The use of hierarchical clustering also allows the use of a dendrogram to visualize the differences between each city.

Each city will likely have their own distribution of popular venues, but similar cities should have

similar available venues. It would make sense that popular destinations would have many hotels to accommodate the large number of visitors intending to stay the night, and the venue data supports this hypothesis. In addition, coffee shops and cafés are distinctly the second and third most popular venues (Fig. 1), which is also reasonable, as coffee is a part of daily routines of many people around the world. There are more than twice as many hotels, coffee shops, and cafés as the next most common venues, so any cities that have few or none of them would likely be in their own cluster. On the other hand, cities with a disproportionately large number of any of those venues would likely be in another cluster. Among the remaining clusters of cities with an average number of hotels, coffee shops, and cafés, differences will likely be determined by the next most common venue categories in those cities.



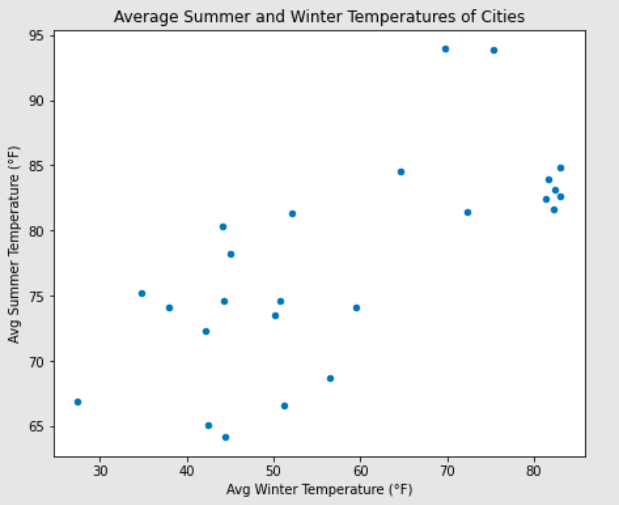
(Fig. 1)

Another factor that would make two cities similar is experiencing similar weather year-round. Using average temperatures of each season would give a concise view of year-round temperatures for each city. However, while summer and winter are generally the hottest and coldest times of the year, spring and fall temperatures tend to lean more towards summer or winter temperatures. Checking the correlation between the average temperatures of all four seasons confirms this. It is clear that spring and fall are highly correlated with each other season, while the correlation between summer and winter is not as strong. Thus, the feature set will be more useful when containing only average winter and summer temperatures than every season.

Correlation of Average Temperatures				
	Avg Winter	Avg Spring	Avg Summer	Avg Fall
Avg Winter	1.000	0.921	0.702	0.985
Avg Spring	0.921	1.000	0.905	0.961
Avg Summer	0.702	0.905	1.000	0.803
Avg Fall	0.985	0.961	0.803	1.000

(Fig. 2)

Geographic location of a city plays a large role in the weather that city will experience throughout the year. Tropical climates often see similar temperatures in the summer and winter due to being so close to the equator, while deserts will have very high or very low temperatures. Clusters would be formed by cities in similar ranges of temperatures, likely formed for cities with different permutations of hot and cold, winters and summers. More detailed clusters would be formed based on ranges of temperatures within those permutations.



(Fig. 3)

Any results produced by this hierarchical model could not be applied globally, due to a discrete number of cities being specified. The benefit of allowing an input of cities, though, is that this model could be applied to a different selection of cities than ones explicitly involved in this project. Any number of cities could be included to allow a widespread use of this model.

IV. Results

Performing hierarchical clustering on the data resulted in 5 helpful clusters:

Cluster 0 had Toronto and Melbourne clustered together, along with Paris, London, New York City, Seoul, and Milan. The cities in this cluster from the original global top 20 destinations seem to be big in the fashion industry. The top venue categories of this cluster were coffee shops, hotels, cafés, Italian restaurants, and plazas. The average winter temperatures of cities in this cluster are around 40°F and summer temperatures around 69°F. These cities typically have the lowest average temperatures compared to the cities in other clusters, though they are most similar to cities in Cluster 1. Individually, Melbourne was most similar to London and Paris, and Toronto was most similar to all three.

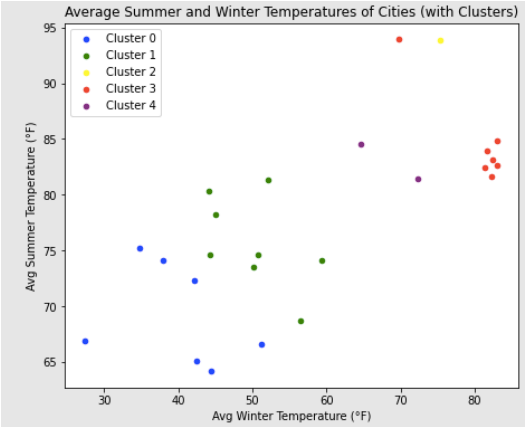
Cluster 1 contained Istanbul, Tokyo, Antalya, Osaka, Barcelona, Palma, Cape Town, and Los Angeles. Hotels, cafés, coffee shops, restaurants, and bars were the top venue categories of this cluster. Winter temperatures average around 50°F, and summer around 76°F. The average temperatures of cities in this cluster are mainly higher than those in Cluster 0, but lower than cities in the other clusters. Cape Town was most similar to Barcelona, and Los Angeles to Antalya.

Cluster 0 and 1 have similar average temperature ranges, though each city in Cluster 1 has a higher average temperature than every city in Cluster 0 during at least one season.

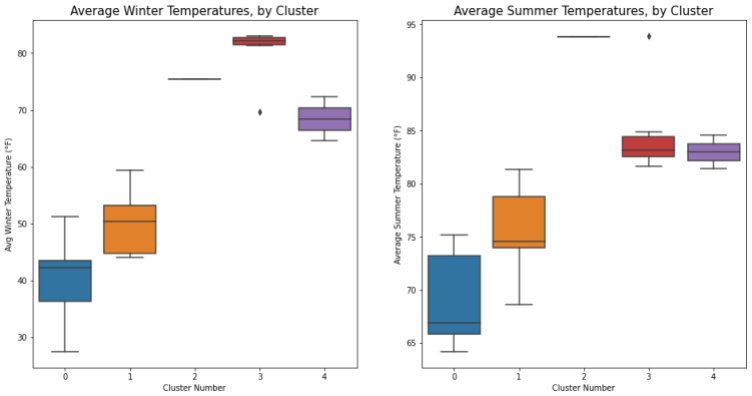
Cluster 2 was the only cluster with only one city: Mecca. Mecca's top venues were hotels, coffee shops, restaurants, middle eastern restaurants, and cafés. In the winter, Mecca's average temperature is around 75°F, and in the summer, around 94°F. Mecca has similar winter and summer temperatures to cities found in Cluster 3 (especially Dubai), but has an extremely large number of hotels compared to every other city in this study. Nearly 50% of all the top venues in Mecca are hotels, while the highest percentage of any other venue category in another city is around 17%. This is expected, since more than two million Muslims pilgrimage to Mecca every year to perform Hajj[2].

Cluster 3 was made up of Bangkok, Dubai, Singapore, Kuala Lumpur, Phuket, Phatthaya, and Bali. The top venues for this cluster were hotels, Thai restaurants, coffee shops, cafés, and Asian restaurants. The cities in this cluster have average winter temperatures of 80°F and summer temperatures of 85°F. These cities are easily identified by having similar temperatures in the winter and the summer, most of them forming a group near the top right in (Fig. 4).

Cluster 4 included only Hong Kong and Rio de Janeiro, whose top venues included Brazilian restaurants, hotels, Italian restaurants, bars, and theaters. These cities have average temperatures of 68°F in the winter, and 83°F in the summer. These cities have similar summer temperatures to the cities of Cluster 3, but having average temperatures about 10°F lower during the winter.



(Fig. 4)



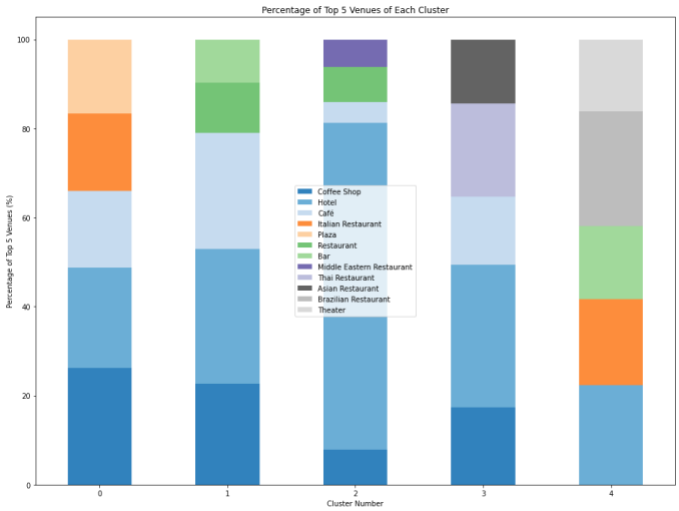
(Fig. 5)

V. Discussion

As noted in the Methodology section, hotels are clearly important venues to have in an ideal travel destination, being the most common venue in a city's top venues by far. Hotels were in the top 5 venue categories of each cluster, so it is clearly important to have accessible accommodation for intended tourists. Clusters 0-3 all had coffee shops and cafés in their top 5 venue categories, so it seems that having easy access to good coffee is also very important in popular destinations. Cities in Cluster 4 have warm average temperatures, so it is possible that warm beverages are less favored, or that tea is preferred over coffee in these places.

Cluster 0 is the only cluster with plazas in the top 5 venue categories, despite having the lowest average temperatures. This could be due to the temperature often not reaching uncomfortable levels, so outdoor areas are more often used. Cluster 3 has more popular

Asian, and more specifically Thai, restaurants than other clusters. This can almost certainly be attributed to most of the cities in the cluster being located in Southeast Asia.



(Fig. 6)

In order to make any implementations of this model, the formed groups should be interpreted. Cluster 0 is made of cities very big in the fashion industry, Cluster 1 has cities very popular in the summer and have access to beaches, Cluster 2 are religious/pilgrimage destinations, Cluster 3 have hot weather and lots of restaurants, and Cluster 4 are warm weathered locations with many options for food and activities.

Belonging to Cluster 0, Melbourne and Toronto could make strides towards becoming fashion capitals themselves, or move toward becoming closer to cities in Cluster 1 by promoting summer tourism. Cape Town and Los Angeles are found in Cluster 1 and could attempt to bring in more visitors during their warm months, or aim to receive more stature in the fashion world to bridge the gap to Cluster 0. Rio de Janeiro is a member of Cluster 4, but certainly has similar average temperatures to cities in Cluster 3, so increasing the number of restaurants could be a potential way to increase its tourist popularity.

VI. Conclusion

In this study, popular global destinations were compared to discover what makes them attractive, as well as finding groupings that gave an overview of what different popular cities may do to succeed. Using local venue and average temperature data, the cities were grouped using a hierarchical clustering method. Hotels, coffee shops, and cafés were found in popular venues between nearly all the destinations included in the study. This model could be used to benefit a city's attempts to increase visitation, due to the ability to input cities, and compare them to the top destinations of the world.

VII. Future Work

Each city has more to it than strictly available venues and average temperatures. Next steps for this project could include adding more weather data, such as humidity or rainfall. Also, including a straightforward visualization of what makes cities similar or dissimilar would make comprehension much easier; these could even be formed from specific cities that the user was interested in. Population of the cities could be added, as well, making higher trafficked, less population-dense areas more clearly popular attractions.

There are some issues with the model that could be improved. Scraping the weather data is extremely inefficient and could likely be implemented using a much quicker approach. Some clusters can end up with much fewer members than other clusters, so adding more base cities could improve reliability of the model. Since the venue data is all collected using the Foursquare API, areas of the world where the data is less often updated impacts the usefulness of the data, so potentially using locally updated sites/apps could be more accurate.

References

[1] Munoz Robino, Diana. "Global Top 20 Destination Cities by International Overnight Visitors (2018)." *Global Destination Cities Index 2019*. Mastercard, 2019, https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf.

[2] "Hajj Pilgrimage Fast Facts." CNN, Cable News Network, 9 Sept. 2020, www.cnn.com/2013/06/21/world/hajj-fast-facts.