Finding Similar Cities

Mingus Knox White

1. 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.

2. Data

The top 20 most visited cities were found in a report by Mastercard, titled, "Global Destination Cities Index 2019.[1]" 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: Honolulu, USA; Sydney, 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,

[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.

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 venue information, 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.