# Final Capstone Assignment- Identification of Similar Cities for Restaurant Chain Expansion

ANKIT CHAUBAL

May 2020

## Introduction

A famous Restaurant in Los Angeles wants to expand its business globally and is searching for locations across the globe for its expansion. The company has identified major cities across the globe and want to shortlist cities where they can explore expansion opportunities.

The company needs help in shortlisting initial set of cities to begin with. They want to identify cities which are similar to Los Angeles, so that the restaurant can be established without much changes to their core offerings.

The project would be mainly essential for the Strategy team of the Restaurant Chain, who will then, based on the output, do further detailed analysis of the cities and check on feasibility of business expansion in the cities.

## Data Overview

The Restaurant chain has a list of 25 cities which will be the base data, for which we need to perform the cluster analysis. The base data will have only 2 columns- City and Country

For Example, Column A will be 'Toronto' and Column B will be 'Canada'

Data is present in a csv file.

## Methodology

In order to generate latitude and longitude for these cities, we will first generate a new column called 'Address' which will combine the 2 available columns in format- CityName, CountryName

So the address column will have value like 'Toronto, Canada'

| | Cities | Countries | Address |
|---|---|---|---|
| **0** | Mumbai | India | Mumbai, India |
| **1** | Delhi | India | Delhi, India |
| **2** | Madrid | Spain | Madrid, Spain |
| **3** | Los Angeles | USA | Los Angeles, USA |
| **4** | Las Vegas | USA | Las Vegas, USA |

Once the Address is generated, we will use the geocoders package to generate the Lat/Long Details of the cities.

| | Cities | Countries | Address | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | Mumbai | India | Mumbai, India | 18.938771 | 72.835335 |
| 1 | Delhi | India | Delhi, India | 28.651718 | 77.221939 |
| 2 | Madrid | Spain | Madrid, Spain | 40.416705 | -3.703582 |
| 3 | Los Angeles | USA | Los Angeles, USA | 34.053691 | -118.242767 |
| 4 | Las Vegas | USA | Las Vegas, USA | 36.167256 | -115.148516 |

We will also use the Foursquare API to get nearby locations of each of the cities by passing the co-ordinated generated Ultimately, we will have the city name, country name, address, city latitude,city longitude, venue details such as venue name, its location details and category which will be used for further processing

| | City | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Mumbai | 18.938771 | 72.835335 | Town House Cafe | 18.938550 | 72.833464 | Bar |
| 1 | Mumbai | 18.938771 | 72.835335 | Royal China | 18.938715 | 72.832933 | Chinese Restaurant |
| 2 | Mumbai | 18.938771 | 72.835335 | Sher-E-Punjab | 18.937944 | 72.837853 | Indian Restaurant |
| 3 | Mumbai | 18.938771 | 72.835335 | Cafe Excelsior | 18.937701 | 72.833566 | Café |
| 4 | Mumbai | 18.938771 | 72.835335 | Chhatrapati Shivaji Maharaj Terminus | 18.940088 | 72.835257 | Train Station |

After getting the nearby venues for all cities, we will perform on hot encoding and get average scores for each venue type for each city.

| | City | African Restaurant | Alsatian Restaurant | American Restaurant | Amphitheater | Antique Shop | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | ... | Resta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Mumbai | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |
| 1 | Mumbai | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |
| 2 | Mumbai | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |
| 3 | Mumbai | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |
| 4 | Mumbai | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | |

Based on the score, we will identify top 10 most common category types and use only that data for k means clustering We will now have city along with the average score for top 10 most common venues, on which we will be performing the k means.

| | City | African Restaurant | Alsatian Restaurant | American Restaurant | Amphitheater | Antique Shop | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Buenos Aires | 0.000000 | 0.00 | 0.000000 | 0.04 | 0.000000 | 0.200000 | 0.000000 | 0.000000 | 0.000000 |
| 1 | Cairo | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.020833 | 0.000000 | 0.000000 |
| 2 | Dhaka | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.166667 |
| 3 | Dublin | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.250000 | 0.000000 | 0.000000 |
| 4 | Hongkong | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.018519 | 0.000000 | 0.000000 |
| 5 | Istanbul | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.014085 | 0.000000 | 0.000000 | 0.000000 | 0.014085 |
| 6 | Las Vegas | 0.000000 | 0.00 | 0.038462 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 7 | London | 0.012821 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.038462 | 0.012821 | 0.000000 |
| 8 | Los Angeles | 0.000000 | 0.00 | 0.037037 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 9 | Madrid | 0.000000 | 0.00 | 0.000000 | 0.00 | 0.000000 | 0.022727 | 0.011364 | 0.034091 | 0.000000 |

K Means clustering will be performed on this dataset for cluster analysis to identify and group similar cities and results will be visualized and checked for observations. We use K means which is an unsupervised learning method for classification as we do not have output data for performing supervised learning.

Output for K Means-

# Result

After running the cluster analysis for k=6, we get that the city of Los Angeles (the base location for the Client) is classified in the 5$^{th}$ cluster with 13 other cities.

This is also the biggest cluster found for in the output. On analysing the cluster properties, we also observe that the Top 3 most visited venues are Hotels/Cafes/Coffee shops or Restaurants which will help in the final selection as these venues are more relevant to the Restaurant Business.

| | Cities | 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 |
|---|---|---|---|---|---|---|---|---|---|---|
| **2** | Madrid | 4 | Hotel | Hostel | Plaza | Restaurant | Spanish Restaurant | Wine Bar | Gourmet Shop | Art Museum |
| **3** | Los Angeles | 4 | Museum | Shopping Mall | Pop-Up Shop | Sushi Restaurant | Sandwich Place | Candy Store | Bookstore | Noodle House |
| **6** | Hongkong | 4 | Hotel | Café | Lounge | Steakhouse | Cantonese Restaurant | Park | Dim Sum Restaurant | Chinese Restaurant |
| **11** | Perth | 4 | Coffee Shop | Hotel | Café | Korean Restaurant | Wine Bar | Sushi Restaurant | Vietnamese Restaurant | Restaurant |
| **14** | Melbourne | 4 | Coffee Shop | Café | Bar | Cocktail Bar | Dessert Shop | Shopping Mall | Toy / Game Store | Burger Joint |
| **15** | Washington DC | 4 | Garden | Hotel | Outdoor Sculpture | Bar | Government Building | Sandwich Place | Monument / Landmark | American Restaurant |
| **16** | Seol | 4 | Coffee Shop | Korean Restaurant | Chinese Restaurant | Hotel | Historic Site | Lounge | Café | Japanese Restaurant |

# Discussion

As we can observe from the results, the Restaurant Chain can shortlist the 13 cities from the cluster for further analysis on feasibility on expansion.

A City wise detailed analysis is recommended to further shortlist the cities and identify possible locations in cities where the restaurant can be established.

# Conclusion

The study helps the Strategy team to identify possible candidate cities for the business expansion. The shortlisted cities need to be further analysed considering the Cost Analysis/Geo-Political factors and ease of doing business.