MACPEC

Menu Augmentation Campaign with Popular Ethnic Cuisine

M. Neukirch

June 10, 2020

1 Introduction

1.1 Outset

Let us assume that we have been contacted by a major restaurant chain operating in Canada and the US. As an effort to attract more customers the enterprise would like to augment their menus with a range of dishes of locally popular foreign cuisine. While they are predisposed to include different menus in different towns, for logistic and administrative reasons they would like to keep the number of different cuisines to a minimum, i.e. 5.

1.2 Problem

Our task is to find out which cuisines are most popular in the major (more than 100’000 inhabitants) US and Canadian cities/towns and provide them in maximally 5 groups with cuisine recommendations.

1.3 Interest

Knowing the local taste of customers and adapt local menus can have a major impact on large restaurant chains and their business value, because serving popular food will attract more customers. Furthermore local customers may appreciate to find all their most favourite food in a single establishment.

1.4 Approach

Our initial assumption is that we can approximate the popularity of a cuisine with its presence in a town relative to other present cuisines. In other words, a cuisine that has many restaurants in a town is likely to be a popular cuisine in that town. With this assumption, we generate a cuisine data set for US/CAN cities/towns and cluster the data into 5 groups, which forms the basis of our final recommendation.

2 Data

2.1 Sources and Cleaning

First, we need a data base of US and Canadian cities and towns with more than 100'000 inhabitants. We can find lists ordered by population on Wikipedia here and here, which we can scrape for city names, regions and coordinates for cities with more than 100'000 inhabitants. This results in a database of the largest US and Canadian cities.

Second, we query Foursquare for restaurants at most 5km around the city centres with the Foursquare API "https://api.foursquare.com" and extract the venue category for each result. This forms a data base of restaurants with their category, the city and region where they are located and the city centre location for restaurants in the largest US and Canadian cities.

2.2 Cleaning

The list of Canadian municipalities does not strictly contain only towns and cities but municipalities. For simplification, we treated all entries equally as cities and towns, assuming that each municipality would have something similar to a city centre. Furthermore, the list of Canadian municipalities does not contain coordinates, therefore we filled missing longitude and latitude information with *geopy* which is a Python module for extracting coordinates.

The restaurant category data retrieved from Foursquare is very heterogenous and difficult to work it as is. Our principal interest are not exactly the categories but the cuisines of each restaurant and several categories can relate to the same cuisine. Fortunately, the Foursquare categories represent largely the restaurant cuisines though, so that only a minimum effort was required to clean the data set. As an example, a typical Foursquare venue category is "Italian Restaurant", which includes the "Italian" cuisine as descriptive adjective with the overall category "Restaurant". Most categories follow this example and only require iterative text scraping to transform into cuisines automatically. Some categories needed manual assignment, however. After identifying categories which do not directly specify the cuisine, e.g. "Pizza Place", "Taco Place" or "Udon Restaurant", we defined a dictionary to translate these categories into their respective cuisines, that is "Italian", "Mexican" and "Japanese", respectively. Any category that is not clearly identifiable as an ethnic cuisine, such as "American Restaurant", "Sandwich Place" or "Salad Bar" are not relevant to our current analysis and thus, were dropped. The dictionary is attached to the Appendix. Finally, we obtain 38 different cuisines for 14402 unique restaurants in 347 North American cities.

Having obtained the cuisine for each restaurant, we transform the cuisine with One-Hot-Encoding. The resulting cuisine data set is grouped by city and each cuisine type averaged in each city into a data set of its relative presence of the cuisine types in each city with their city names and region as well as city centre locations. The value of the relative presence can assume values between 0 and 1, where 1 means it is the only present cuisine type and 0 denotes that no restaurant with this cuisine is within the search parameters of this city. Cities for which we did not find restaurants were dropped. This data set forms the base data for our analysis as we use the relative occurrence of a cuisine in a city as proxy for its popularity.

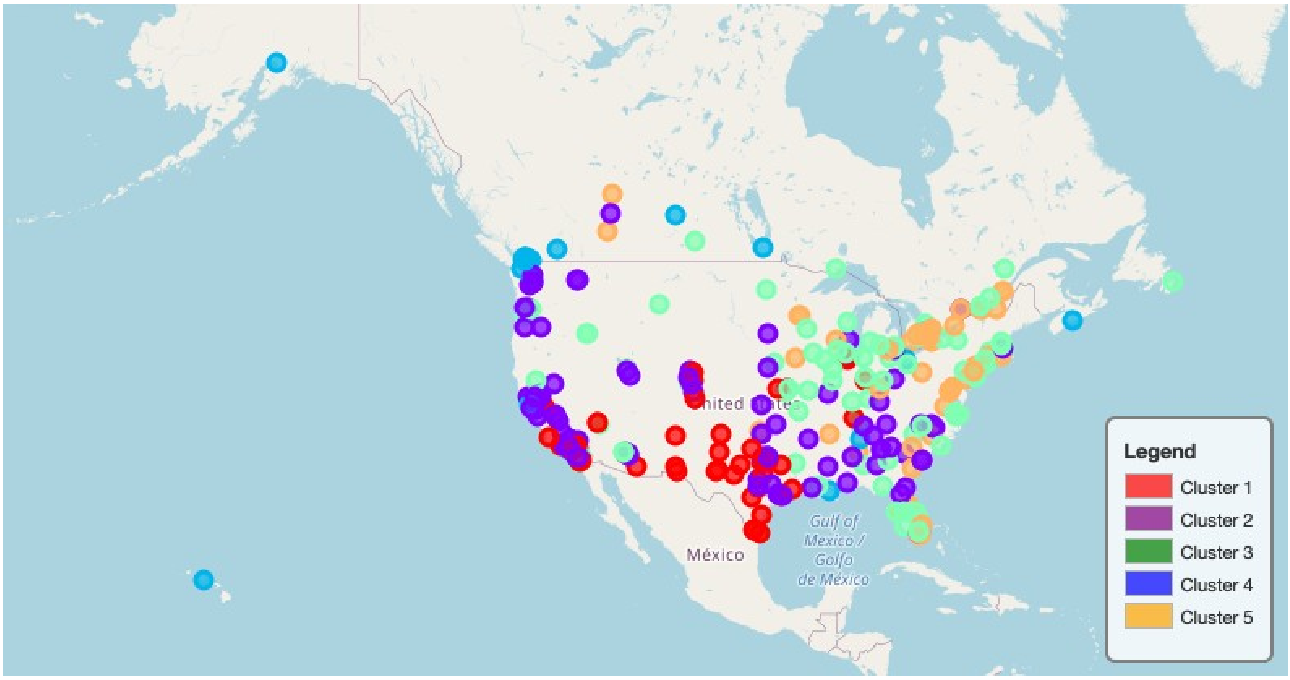

Figure 1: Cuisine category distribution of Canadian and US cities (>100'000 inhabitants)

3 Methodology and Data Analysis

As mentioned above, we use the relative occurrence of a cuisine as a measure for its popularity in the local populace. After all, demand controls supply, such that if there were many restaurants of a certain cuisine and no one went there, many of those restaurants would close until the demand matches the supply in an economical balance. Therefore, we argue that the occurrence rate of a restaurant with a specific cuisine does represent the demand (and therewith its popularity) of it in the local populace.

3.1 K-Means

We use K-Means to group the cities based on the culinary preferences of their inhabitants. As of the project's goal we use k=5 in order to obtain 5 distinct groups of cities so that logistic and administrative strains remain at a minimum for the client. Should similar groups emerge we can reduce k or combine similar groups manually. The five groups are illustrated on a map in Figure 1 and can be summarised into the following groups by location:

(1) Southern USA

(2) Western/Central USA

(3) Eastern/Central USA

(4) Canada, Eastern USA, Alaska and Hawaii

(5) Canada and Eastern USA

It was not expected that the cities would cluster into geographical regions but it is a welcome result for the client as it may reduce logistical strain.

Figure 2: Popular cuisines for cluster 1.

Figure 3: Popular cuisines for cluster 2.

Figure 4: Popular cuisines for cluster 3.

Figure 5: Popular cuisines for cluster 4.

Figure 6: Popular cuisines for cluster 5.

3.2 Word Cloud

In order to recommend cuisines for each cluster, we sorted the cuisines for each city by their popularity and labeled the 6 most popular cuisines. For each cluster, we concatenated the 6 most popular cuisine names of all cities in that cluster in a string variable and drew word clouds representing the cuisine occurrences in the most-popular list. Figures 2 to 6 illustrate the word clouds and provide insight into the preference of the populace in each cluster's cities. Based on the word clouds we can find the following recommendations for the clusters:

(1) Mexican and International

(2) Mexican and Asian

(3) Mediterranean and Asian

(4) predominantly Asian

(5) predominantly Mediterranean and European

In general it appears that there are three cuisines among the absolute most popular, Mexican, Asian and Mediterranean. We can verify this with a word cloud of all five clusters together (Figure 7), which finds Mexican, Japanese and Italian as the three most popular cuisines, which closely represents Mexican, Asian and Mediterranean cuisine.

4 Results

Based on our analysis we group the major North American cities into the following five groups:

1. Mexican-centred cuisine in Southern USA
2. Mexican and Asian food in whole USA but more concentrated in western part
3. Mediterranean and Asian cuisine in whole USA but slightly favouring eastern part
4. Asian cuisines popular in some cities in Canada, Eastern USA and off-mainland
5. Mediterranean/European cuisine favoured in some cities in Canada and Eastern USA

Detailed lists of cities in each cluster can be found in the attached Jupyter notebook.

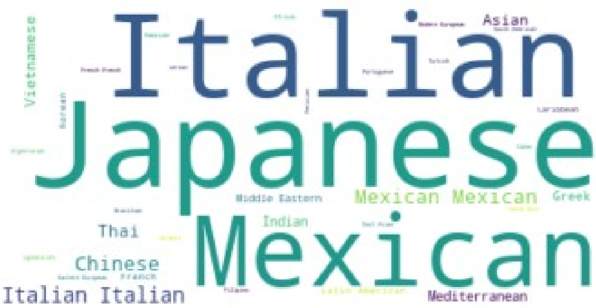

Figure 7: Cuisine popularity among all cities.

5 Conclusion

We have analysed the major cities (with more than 100’000 inhabitants) of Canada and the USA in terms of their populace’s (ethnic) culinary preferences. Then, we identified the dominant cuisines in each city and clustered similar cities based on the occurrence of cuisines into 5 groups. We assumed and argued that the local occurrence of restaurants of a cuisine type is directly related the cuisine’s local popularity and recommended specific cuisine types for each group of cities accordingly.

6 Appendix

6.1 Dictionary for Transforming Venue Category into Cuisine

As mentioned above, for most restaurants the venue category includes the cuisine served in the restaurant, however, in some instances manual intervention was required to assign the correct cuisine to a restaurant. In the following, find the dictionary for manual assignment:

```
manual_assign_cuisine=
{'Moroccan': 'African',
 'Lebanese': 'African',
 'Ethiopian': 'African',
 'Empanada': 'Argentinian',
 'Arepa': 'Argentinian',
 'Salvadoran': 'Caribbean',
 'Hong Kong': 'Chinese',
 'Shanghai': 'Chinese',
 'Cantonese': 'Chinese',
 'Hunan': 'Chinese',
 'Hotpot': 'Chinese',
 'Dim Sum': 'Chinese',
 'Szechuan': 'Chinese',
 'Caucasian': 'East European',
 'Polish': 'East European',
 'Romanian': 'East European',
 'Russian': 'East European',
 'Hungarian': 'East European',
 'Bavarian': 'German',
 'Poke': 'Hawaiian',
 'Indian Chinese': 'Indian',
 'North Indian': 'Indian',
 'Tibetan': 'Indian',
 'Himalayan': 'Indian',
 'South Indian': 'Indian',
 'Chaat': 'Indian',
 'Jewish': 'Israeli',
 'Falafel': 'Israeli',
 'Pizza': 'Italian',
 'Japanese Curry': 'Japanese',
 'Udon': 'Japanese',
 'Cajun / Creole': 'Japanese',
 'Ramen': 'Japanese',
```

```
'Sushi': 'Japanese',
'Soba': 'Japanese',
'Shabu-Shabu': 'Japanese',
'Donburi': 'Japanese',
'Venezuelan': 'Latin American',
'Burrito': 'Mexican',
'Taco': 'Mexican',
'Tex-Mex': 'Mexican',
'Persian': 'Middle Eastern',
'Iraqi': 'Middle Eastern',
'Belgian': 'Modern European',
'English': 'Modern European',
'Fondue': 'Modern European',
'Colombian': 'South American',
'Indonesian': 'South East Asian',
'Satay': 'South East Asian',
'Burmese': 'South East Asian',
'Sri Lankan': 'South East Asian',
'Malay': 'South East Asian',
'Cambodian': 'South East Asian',
'Paella': 'Spanish',
'Tapas': 'Spanish',
'Doner': 'Turkish',
'Kebab': 'Turkish'}
```