MACPEC

Menu Augmentation Campaign with Popular Ethnic Cuisines

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 town. Towns for which we did not find restaurants were dropped. This data set forms the base data for our analysis.

3 Methodology and Data Analysis

3.1 K-Means

3.2 Word Cloud

4 Discussion

5 Conclusion

6 Appendix

Code:

sort a dictionary by values, use for category to cuisine dictionary to create a list to add to appendix

```
x = {1: 2, 3: 4, 4: 3, 2: 1, 0: 0}
{k: v for k, v in sorted(x.items(), key=lambda item: item[1])}
{0: 0, 2: 1, 1: 2, 4: 3, 3: 4}
```