Capstone Project
Ethnic food preferences in the most populated cities in the United States

1. Introduction

Background

The United States of America (USA) is known worldwide as one of the countries with the highest level of immigration, a fact that has allowed this country to have a wide variety of nationalities that greatly enriches its culture. One of the ways this can be reflected is in the great variety of ethnic cuisine that can be found in the USA, which is consumed by many American citizens - especially for the younger ones, who love to try new flavours and experiences.

Problem

I want to answer the following questions: is there a difference in ethnic food preference among the most populated cities in the USA? And if so, what is it, and what could explain it? Also I want to know how the culinary preference is distributed among the cities and which of those have similar ones.

2. Data

This study takes the 30 most populated cities in the United States in order to keep the problem tractable and for it to be representative enough in terms of population. For this I used the list of the 317 most populated cities in the U.S. as estimated by the U.S. Census Bureau, which can be found on https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population.

To define ethnic food preferences, the 5 most populous international cuisines in all of the United States were established according to Technomic, which is cited in the report at https://blogs.voanews.com/all-about-america/2015/05/18/top-10-most-popular-ethnic-cuisines-in-us/. Among these foods are Asian, Mexican, Italian, Greek and French.

Once I have the cities with their latitudes arranged in a dataframe, I will proceed to call Foursquare which will be used to search for each of these ethnic foods through the restaurant in each of the cities under study. With this information I will know what is the favorite ethnic food in each city, how these preferences are distributed among cities and which of them have similar preferences.

Finally, you can see on the map the cities grouped according to their culinary tastes.

3. Methodology

To solve our problem of how ethnic food preferences are distributed in the 30 most populated cities in the United States, the methodology followed was as follows:

1. Install and import all the libraries we need for the analysis.
2. Search for the most populated cities in the U.S. and find the coordinates of each one. For our luck on Wikipedia all this information was found in the same page in a table.
3. Scrape the Wikipedia page, using the "read_html" function of "pandas.io.htm", to get the table with the information of the cities. Then we use this table to make a functional and clear DataFrame with all the information we need.
4. Once the DataFrame was done, we proceeded to make a call to the Foursquare API to find the Asian, Mexican, Italian, Greek and French restaurants in each city. We defined a radius of 1.2 km which is the maximum radius we can choose to get relevant information, this is because the free Foursquare account only returns 50 venues and if we enlarge the radius we exceed this limit and see reflected only 50 for all cities. We also look for the identification of each type of restaurant at https://developer.foursquare.com/docs/build-with-foursquare/categories/, where Foursquare defines a category for each type of restaurant. In this way we made the search more efficient and specific to what we needed to obtain.
5. With the information obtained on Foursquare, we made a new DataFrame to sort the data of the amount of each type of restaurant for each of the cities.
6. After making the new DataFrame, we cluster the cities using the K-mean model. For this analysis we grouped the data by city and defined a k=4.
7. Finally, we made a map using Folium library to visualize how the cities, that were previously grouped, are distributed in the U.S.

4. Results

The results obtained show that the most consumed food in the 30 most populated cities in the U.S. are those included in the Asian category, followed by Mexican food. See table and figure 1. It is important to note that within the Asian category you can find Chinese, Japanese, Korean, and Filipino food among others.

	Asian restaurante	Mexican restaurant	Italian restaurant	Greek restaurant	French restaurant
TOTAL	453	381	111	30	41

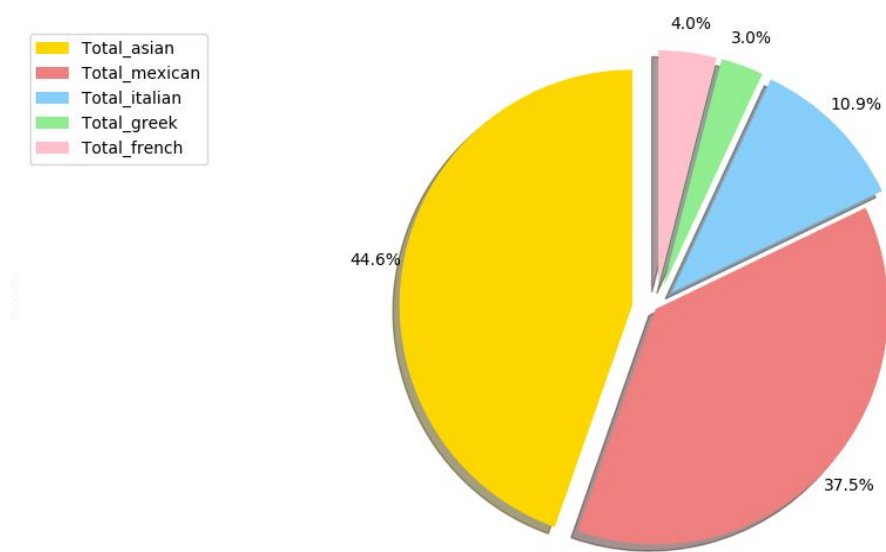*Table 1: Preferences of the 30 most populated cities in the USA*

Figure 1. Pie chart showing the preferences of the 30 most populated cities in the USA

The other result, obtained by grouping the cities with the K-mean model and visualizing them through the map, shows that the cities that are grouped in the same group have very similar culinary tastes and are located in very similar geographical areas. See figure 2.
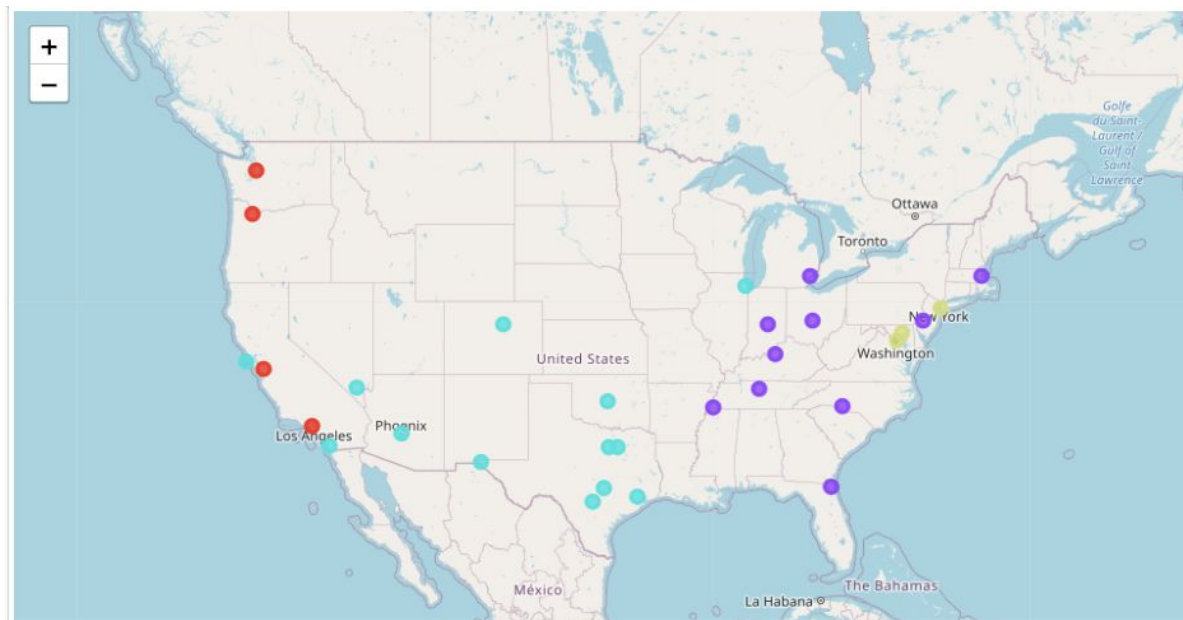

Figure 2. Visualization of the 4 cluster on the map

Finally, from this analysis it could also be obtained that cities on the eastern and western coasts have a preference for Asian food, while cities in the central United States prefer Mexican food. Figure 3.
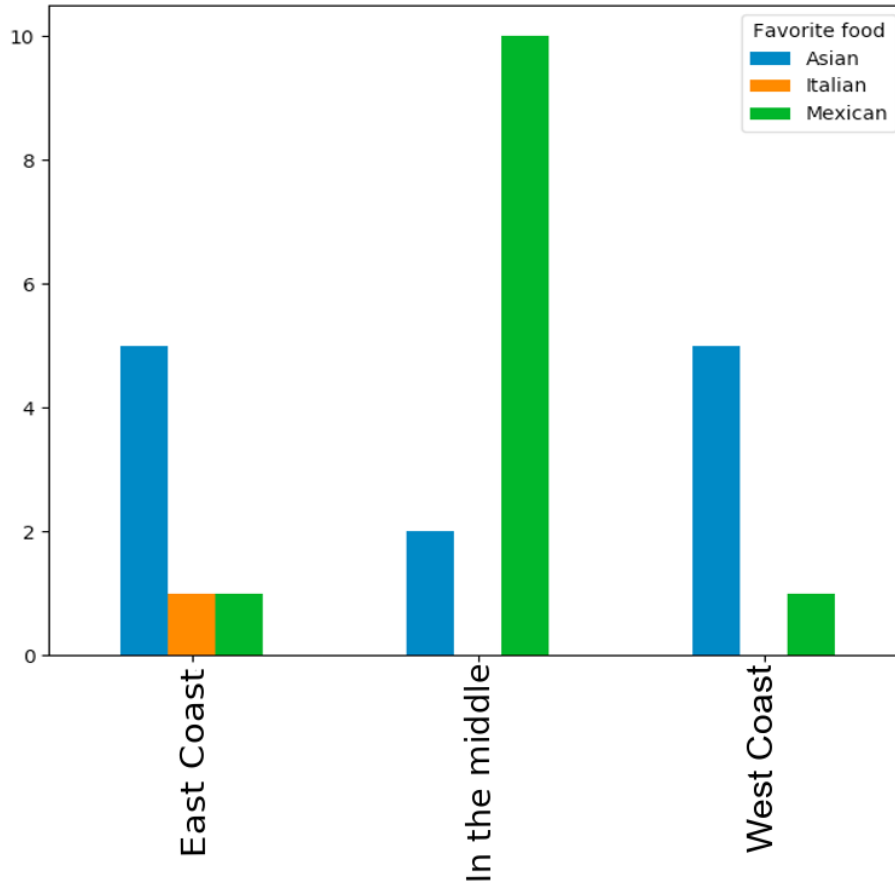

Figura 3. Bar chart showing the location (east, middle, west) of the most consumed foods by the 30 most populated cities in the US.

5. Discussion

An important point to discuss is the coordinates that are assigned to the cities, because these are not located in a place that represents the city. In the case of this study the city of san francisco did not obtain information because the coordinates are in the sea, where the possibility of finding a restaurant is null.

Another point to highlight is the small search radius that had to be defined (1.2 km). We consider this to be unrepresentative of how big all these cities are. Besides, we have to add that this search area is not necessarily in a sector that represents the city, because as we mentioned before, it is around the coordinates that can be located in any part of the city.

With regard to the findings that cities on the east and west coasts have a preference for Asian foods, it may be that these areas are home to the largest Asian population.

Finally, it is important to note that within Asian food there are many subcategories, such as Chinese, Japanese, Korean and others. This makes it a much more varied and larger group compared to the other types of food under study.

6. Conclusion

Despite the reduced number of venues provided by the free "Foursquare" account, it was possible to obtain answers to the problem in question.

However, for future studies it would be good to do the same analysis but with a premium "Foursquare" account. This will allow you to get more than 50 venues in response to each call made. With this access you will be able to define a larger search radius that will allow you to better represent each city in the study.