# Exploration of the European Restaurant Market

*Part of the Capstone Project - The Battle of Neighborhoods*


**Simon Gross**

June 7, 2020

# INTRODUCTION

## Background

Investing in a new restaurant is often not an easy decision. The decision made is a long-term project and involves a great deal of resources to find the right location, to renovate it, but also to find out what kind of cuisine to offer. Whether or not an investment will be successful is often not visible for a long time, as a restaurant must first build up a good reputation and a constant flow of customers, which can take up to several years. It is therefore always very important to reduce the risk of investing in a restaurant of a particular category in a geographical area. But how can you find out if the dishes offered in a restaurant are appreciated by the locals or if the location of the restaurant is good?

## Problem

The first problem, whether people want the provided dishes, is often related to cultural preferences. In southern Italy, for example, people like fish and seafood, while in southern Germany they prefer rustic cuisine. Therefore, an investor must find the right cuisine for the region in which he wants to invest. The second problem of whether a location is a good one is often related to how many competitors there are in that particular neighbourhood or town. Therefore, an investor needs to understand the city's surroundings, where there might be room for another Asian restaurant and where there isn't.

## Interest

This market analysis is particularly interesting for large investment companies that want to explore the restaurant market and are not sure where and in what kind of restaurants to invest. This could be a Chinese investor who is not familiar with the European market and does not know the regional preferences. They will obtain detailed information about the European restaurant market. In which region are some restaurants and dishes preferred and some not and would therefore be associated with a high risk.

## Goal

The aim of this project is to get a visual impression of what kind of restaurants are preferred in each region of Europe and therefore would have a better chance of success.

# DATA

## Data Sources

In order to cover the most relevant areas in Europe, the 500 largest cities in Europe will be investigated. A list of the cities together with their country and population can be found [here](). Based on this list of city and country we collect the coordinates through Geopy. These coordinates are then used to find 100 locations within a 1000 meter range through Foursquare.

| | City | Country | Latitude | Longitude |
|---|---|---|---|---|
| **0** | MOSKVA (Moscow) | Russia | 55.750446 | 37.617494 |
| **1** | LONDON | UK | 51.507322 | -0.127647 |
| **2** | St Petersburg | Russia | 59.960674 | 30.158655 |
| **3** | BERLIN | Germany | 52.517037 | 13.388860 |
| **4** | MADRID | Spain | 40.416705 | -3.703582 |

Fig. 1: Head of table after receiving cities from webpage and coordinates from geopy.

## Data cleaning and Feature Selection

In order to use only relevant data, venues that do not include "restaurant" in their category name are deleted. If the category name only contains "Restaurant", it will also be deleted, as it does not contain any information about the dish served.

```
df2['Venue Category'].unique()
```

```
array(['Italian Restaurant', 'Spanish Restaurant', 'Thai Restaurant',
       'Japanese Restaurant', 'French Restaurant', 'Greek Restaurant',
       'Pakistani Restaurant', 'Ramen Restaurant', 'English Restaurant',
       'Lebanese Restaurant', 'Chinese Restaurant',
       'Modern European Restaurant', 'African Restaurant',
       'Vegetarian / Vegan Restaurant', 'German Restaurant',
       'Sushi Restaurant', 'Tapas Restaurant', 'Seafood Restaurant',
       'Mexican Restaurant', 'Argentinian Restaurant', 'Asian Restaurant',
       'Peruvian Restaurant', 'Paella Restaurant',
       'Mediterranean Restaurant', 'Falafel Restaurant',
       'Roman Restaurant', 'Caucasian Restaurant', 'Turkish Restaurant',
       'Scandinavian Restaurant', 'Swiss Restaurant',
       'Ukrainian Restaurant', 'Alsatian Restaurant',
       'Auvergne Restaurant', 'Portuguese Restaurant',
       'Burgundian Restaurant', 'Corsican Restaurant',
       'Romanian Restaurant', 'Eastern European Restaurant',
       'Hungarian Restaurant'], dtype=object)
```

Fig. 2: An example of a list of categories for restaurants after cleaning.

# Exploratory Data Analysis

## Data processing

The venue categories are ranked according to how often they occur in each city, and a ranking of the most frequent restaurants in each city is created. This was achieved by grouping the dataset by city and then calculating the average for each category or type of restaurant in that city. The value, which is between 0 and 1, thus represents the percentage of each type of restaurant in a city. The highest value indicates the type of restaurant that is most common in that city.

| | City | Afghan Restaurant | African Restaurant | Alsatian Restaurant | American Restaurant | Arepa Restaurant | Argentinian Restaurant | Asian Restaurant | Australian Restaurant | R |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | AMSTERDAM | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | |
| 1 | ATHINAI (Athens) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.066667 | 0.0 | |
| 2 | Aachen | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | |
| 3 | Abakan | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.142857 | 0.0 | |
| 4 | Aberdeen | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.0 | |

Fig. 3: Percentage of each restaurant in a city

The first analysis of the data leads to very many different restaurants, because the category of the venues is not written uniformly. Therefore, in a second round of data processing, the restaurants are predefined in cuisine clusters. These clusters were divided into North American cuisine, South American cuisine, North European cuisine, South European cuisine, Far Eastern cuisine, and Asian cuisine. Each cluster included geographically selected types of restaurants. For example, Italian restaurants belong to South European cuisine and Argentinean restaurants to South American cuisine.

```
cuisine_dict = {'North American cuisine' : ['New American Restaurant','Grilled Meat Restaurant','American Restaurant', 'Fast Food Restaurant'],
                'South American cuisine' : ['Mexican Restaurant','Caribbean Restaurant','Argentinian Restaurant', 'Latin American Restaurant'],
                'North European cuisine' : ['Scandinavian Restaurant','Polish Restaurant','Hungarian Restaurant','English Restaurant','Dutch Rest
                'South European cuisine' : ['Sicilian Restaurant','Cretan Restaurant','Italian Restaurant', 'Greek Restaurant', 'Mediterranean Re
                'Far Eastern cuisine' : ['Middle Eastern Restaurant','Kebab Restaurant','Halal Restaurant','Eastern European Restaurant', 'Doner
                'Asian cusine' : ['Ramen Restaurant','Asian Restaurant', 'Chinese Restaurant', 'Indian Restaurant', 'Japanese Restaurant', 'Sushi
```

Fig. 4: Shows the dictionary of the cuisines with it restaurant

This results in a table of data that more clearly highlights certain regional preferences and makes it easier for stakeholders to see whether certain groups of restaurants are well accepted in these regions.

| | City | North American cuisine | South American cuisine | North European cuisine | South European cuisine | Far Eastern cuisine | Asian cusine |
|---|---|---|---|---|---|---|---|
| 0 | AMSTERDAM | 0.0 | 0.230769 | 0.307692 | 0.153846 | 0.0 | 0.000000 |
| 1 | ATHINAI (Athens) | 0.0 | 0.000000 | 0.200000 | 0.333333 | 0.0 | 0.133333 |
| 2 | Aachen | 0.0 | 0.000000 | 0.304348 | 0.347826 | 0.0 | 0.173913 |
| 3 | Abakan | 0.0 | 0.000000 | 0.000000 | 0.285714 | 0.0 | 0.571429 |
| 4 | Aberdeen | 0.0 | 0.083333 | 0.166667 | 0.416667 | 0.0 | 0.250000 |

Fig. 5: Table after combining restaurants together to bigger classes of cuisines.

## Analysis of data

The aim of the analysis for this dataset is to group the cities according to the frequency of certain types of restaurants. To achieve this, a k-mean cluster is used, which takes the normalized (mean) value of each restaurant category. Three clusters are selected to be searched for.

## Cluster visualization

Following the analysis, maps are displayed. Two different maps are created. For both, each city has a marker on a map. Each marker is colored based on the cluster category in which it is located. For the first map, it stays that way. Since there are hundreds of markers on the map, a second map is created that combines markers of the same category.

## Statistical analysis

No advanced statistical analysis is performed. To identify the cluster categories, the most common type of restaurant or cuisine is examined for each cluster category by examining the number of counts of the restaurant or cuisine.

| 1st Most Common Venue | City |
|---|---|
| Eastern European Restaurant | 3 |
| English Restaurant | 2 |
| French Restaurant | 3 |
| Greek Restaurant | 1 |
| Italian Restaurant | 86 |
| Japanese Restaurant | 2 |
| Romanian Restaurant | 2 |
| Seafood Restaurant | 3 |
| Sushi Restaurant | 2 |
| Vegetarian / Vegan Restaurant | 1 |

Fig. 6: Example of investigating most common restaurant for a cluster category

## Results

The results section is divided into two parts. As already described, the first attempt was to use all restaurants and let the k-mean cluster take effect and find clusters within the more than 100 different types of restaurants. The second attempt was based on a more supervised version, where the number of features was drastically reduced by pre-defining culinary clusters. This should make it easier to understand the clusters found by the k-mean cluster.

### First round: all types of restaurants

For this analysis 450 cities with 38446 event locations were found. After selecting only the venues that contain restaurants in the series, the number was reduced to 5916. Of these restaurant locations, 108 unique types of restaurants were identified. This means that 108 features are used to classify the 3 cluster labels.

### Defining the labels

Three cluster categories are selected, Fig. 7 shows the most frequent restaurant location for all three cluster categories. This was already sufficient for Label 1 and Label 3, but not for Label 2, where no clear preference was indicated.

**Label 1**

| 1st Most Common Venue | City |
|---|---|
| Eastern European Restaurant | 3 |
| English Restaurant | 2 |
| French Restaurant | 3 |
| Greek Restaurant | 1 |
| Italian Restaurant | 86 |
| Japanese Restaurant | 2 |
| Romanian Restaurant | 2 |
| Seafood Restaurant | 3 |
| Sushi Restaurant | 2 |
| Vegetarian / Vegan Restaurant | 1 |

**Label 2**

| 1st Most Common Venue | City |
|---|---|
| Afghan Restaurant | 1 |
| American Restaurant | 4 |
| Asian Restaurant | 6 |
| Austrian Restaurant | 1 |
| Belarusian Restaurant | 1 |
| Bulgarian Restaurant | 2 |
| Caucasian Restaurant | 4 |
| Chinese Restaurant | 6 |
| Czech Restaurant | 1 |
| Doner Restaurant | 3 |
| Dumpling Restaurant | 4 |
| Eastern European Restaurant | 35 |
| English Restaurant | 5 |
| Falafel Restaurant | 1 |
| Fast Food Restaurant | 8 |
| French Restaurant | 15 |
| German Restaurant | 17 |
| Greek Restaurant | 2 |
| Hungarian Restaurant | 1 |
| Indian Restaurant | 12 |
| Italian Restaurant | 24 |

**Label 2**

| | |
|---|---|
| Italian Restaurant | 24 |
| Japanese Restaurant | 8 |
| Kebab Restaurant | 1 |
| Korean Restaurant | 3 |
| Mediterranean Restaurant | 10 |
| Mexican Restaurant | 2 |
| Meze Restaurant | 2 |
| Middle Eastern Restaurant | 1 |
| Modern European Restaurant | 7 |
| Polish Restaurant | 2 |
| Portuguese Restaurant | 5 |
| Romanian Restaurant | 2 |
| Scandinavian Restaurant | 8 |
| Seafood Restaurant | 4 |
| Spanish Restaurant | 16 |
| Sushi Restaurant | 23 |
| Swiss Restaurant | 2 |
| Tapas Restaurant | 14 |
| Thai Restaurant | 2 |
| Turkish Restaurant | 3 |
| Vegetarian / Vegan Restaurant | 5 |
| Vietnamese Restaurant | 5 |
| West-Ukrainian Restaurant | 1 |

**Label 3**

| 1st Most Common Venue | City |
|---|---|
| American Restaurant | 1 |
| Eastern European Restaurant | 2 |
| Falafel Restaurant | 1 |
| Fast Food Restaurant | 55 |
| Indian Restaurant | 3 |
| Italian Restaurant | 2 |
| Russian Restaurant | 1 |
| Sushi Restaurant | 1 |

Fig. 7: The most common kind of restaurants for each cluster category

For label 2, therefore, the second most frequent venues are also examined.

Label 2

| 2nd Most Common Venue | City |
|---|---|
| American Restaurant | 5 |
| Asian Restaurant | 12 |
| Bavarian Restaurant | 1 |
| Belarusian Restaurant | 1 |
| Caucasian Restaurant | 2 |
| Chinese Restaurant | 7 |
| Eastern European Restaurant | 10 |
| Empanada Restaurant | 1 |
| English Restaurant | 1 |
| Falafel Restaurant | 3 |
| Fast Food Restaurant | 26 |
| French Restaurant | 4 |
| German Restaurant | 7 |
| Greek Restaurant | 6 |
| Greek Restaurant | 6 |
| Halal Restaurant | 3 |
| Indian Restaurant | 7 |
| Italian Restaurant | 49 |
| Japanese Restaurant | 13 |
| Korean Restaurant | 1 |
| Mediterranean Restaurant | 5 |
| Mexican Restaurant | 2 |
| Middle Eastern Restaurant | 5 |
| Modern European Restaurant | 6 |
| Molecular Gastronomy Restaurant | 1 |
| Polish Restaurant | 1 |
| Portuguese Restaurant | 2 |
| Provençal Restaurant | 1 |
| Ramen Restaurant | 1 |
| Romanian Restaurant | 1 |
| Russian Restaurant | 3 |
| Scandinavian Restaurant | 3 |
| Scandinavian Restaurant | 3 |
| Scottish Restaurant | 1 |
| Seafood Restaurant | 9 |
| Spanish Restaurant | 15 |
| Sri Lankan Restaurant | 1 |
| Sushi Restaurant | 17 |
| Tapas Restaurant | 13 |
| Tatar Restaurant | 2 |
| Thai Restaurant | 1 |
| Turkish Restaurant | 3 |
| Ukrainian Restaurant | 3 |
| Vegetarian / Vegan Restaurant | 13 |
| Vietnamese Restaurant | 6 |
| West-Ukrainian Restaurant | 2 |
| Yakitori Restaurant | 3 |

Fig. 8: Second most common restaurants for label 2.

The result of the study of these figures is that Italian restaurants dominate in Label 1 and are the most common in this city in this cluster. At Label 2, Italian restaurants are no longer dominant, but are still frequently found in these cities. Label 3 shows a high number of fast food restaurants in this city.

## Plotting the maps

In a next step, we mark each city with a colored marker regarding its labeling on a European map.

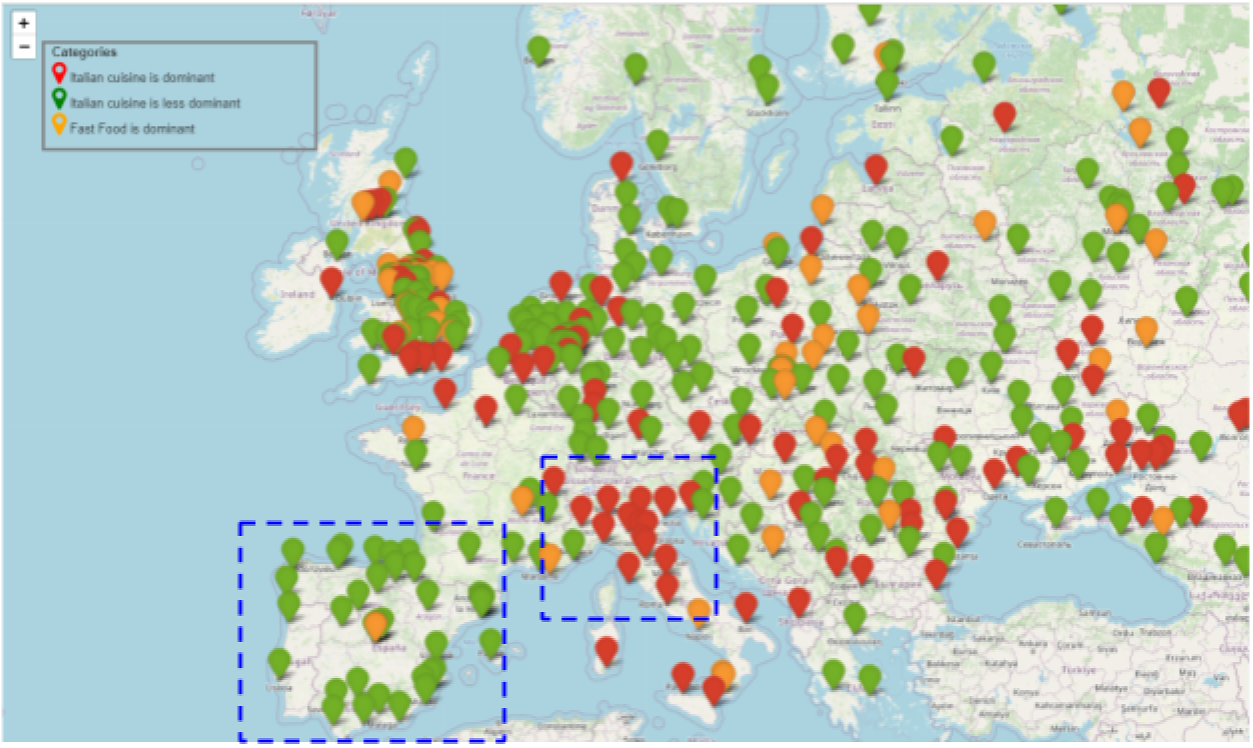


Fig. 9: Map of each city colored by its label

Two regions are highlighted in Figure 9. While Italian restaurants dominate in southern Italy, they seem to play only a minor role in Spain. For the rest of the map the labels are well distributed and no clear preferences can be expressed.

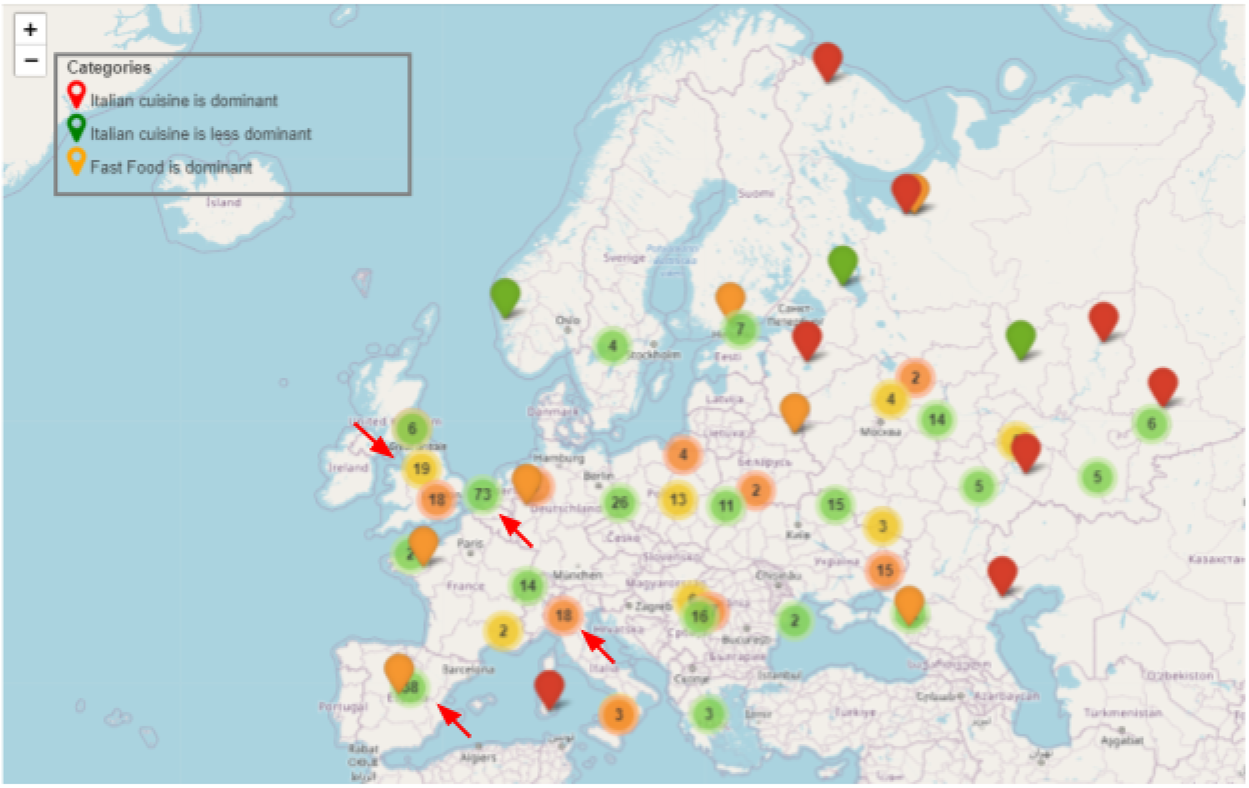


Fig. 10: Map where labels are clustered together to reduce the number of markers on map

The arrows in Figure 10 indicate hot spots of certain cluster labels. The hotspots in Italy or Spain were previously visible, but now there is also a Label 2 hotspot for a lower dominance of Italian restaurants in the Netherlands region and a hotspot of fast food restaurants in England.

## Second round: predefined clusters

In order to better classify and identify the regions of interest to stakeholders, the former 108 characteristics have been reduced to only 6 characteristics. These 6 characteristics group together the most common types of restaurants that refer to different geographical regions, such as North or South European cuisine, North or South American cuisine, Far Eastern cuisine and Asian cuisine, which would be of more interest to Chinese stakeholders, for example.

### Defining the labels

Same procedure as before, checking for each label which cuisine has the highest number.

| Label 1<br>1st Most Common Venue | City | C |
|---|---|---|
| Asian cuisine | 141 | |
| Far Eastern cuisine | 28 | |
| North European cuisine | 50 | |
| South American cuisine | 1 | |
| South European cuisine | 38 | |

| Label 2<br>1st Most Common Venue | City |
|---|---|
| Asian cuisine | 1 |
| Far Eastern cuisine | 1 |
| South European cuisine | 115 |

| Label 3<br>1st Most Common Venue | City |
|---|---|
| Asian cusine | 7 |
| Far Eastern cuisine | 11 |
| North American cuisine | 49 |
| North European cuisine | 3 |
| South European cuisine | 5 |

Fig. 11: Most common cuisine for each label

This time the identification of the label is quite clear. Cities of label 1 show strong preferences for Asian cuisine, while cities of label 2 show strong preferences for southern European cuisine and cities of label 3 for North American cuisine.

### Plotting the maps

The illustration of the cities on the map showed a clear separation of north and south of the labels 1 and 2.

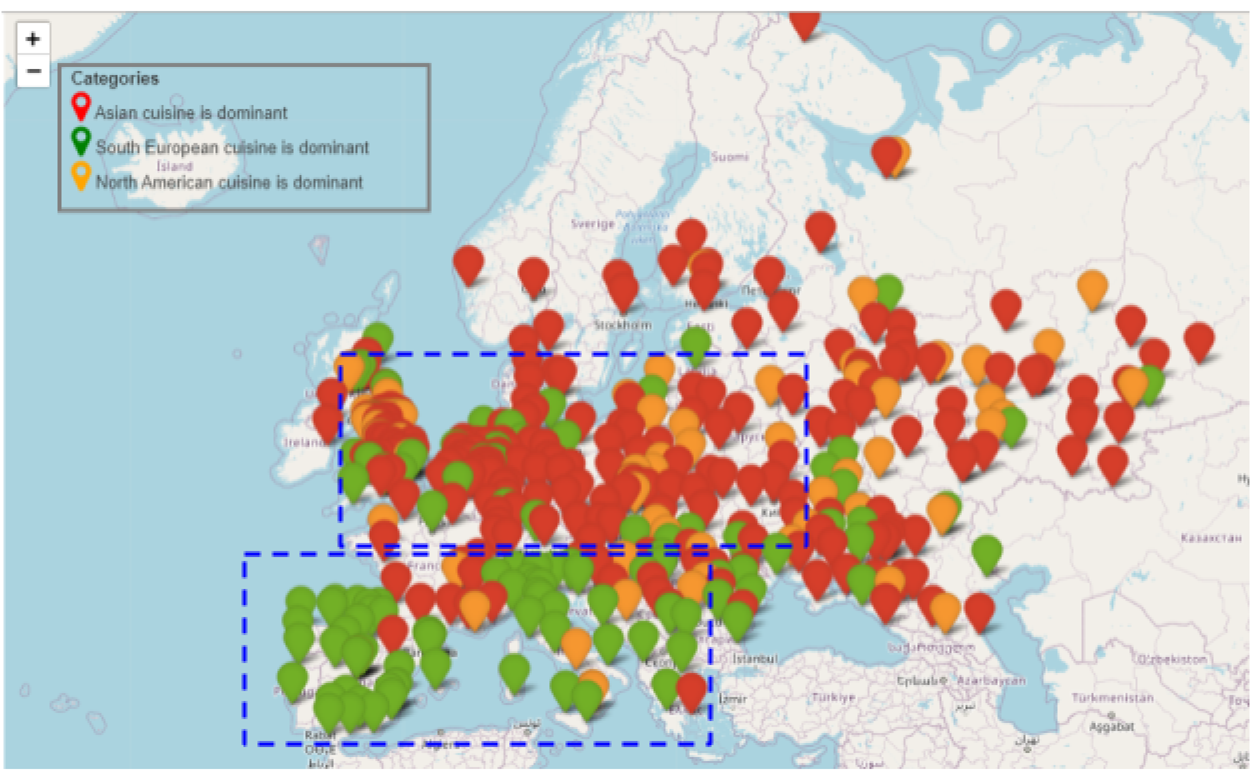


Fig. 12: Map of all cities where marker is colored by correspondant label

The clustering of these labels resulted in clusters for Label 3, where North American cuisine dominates, in England, but also in Eastern Europe.

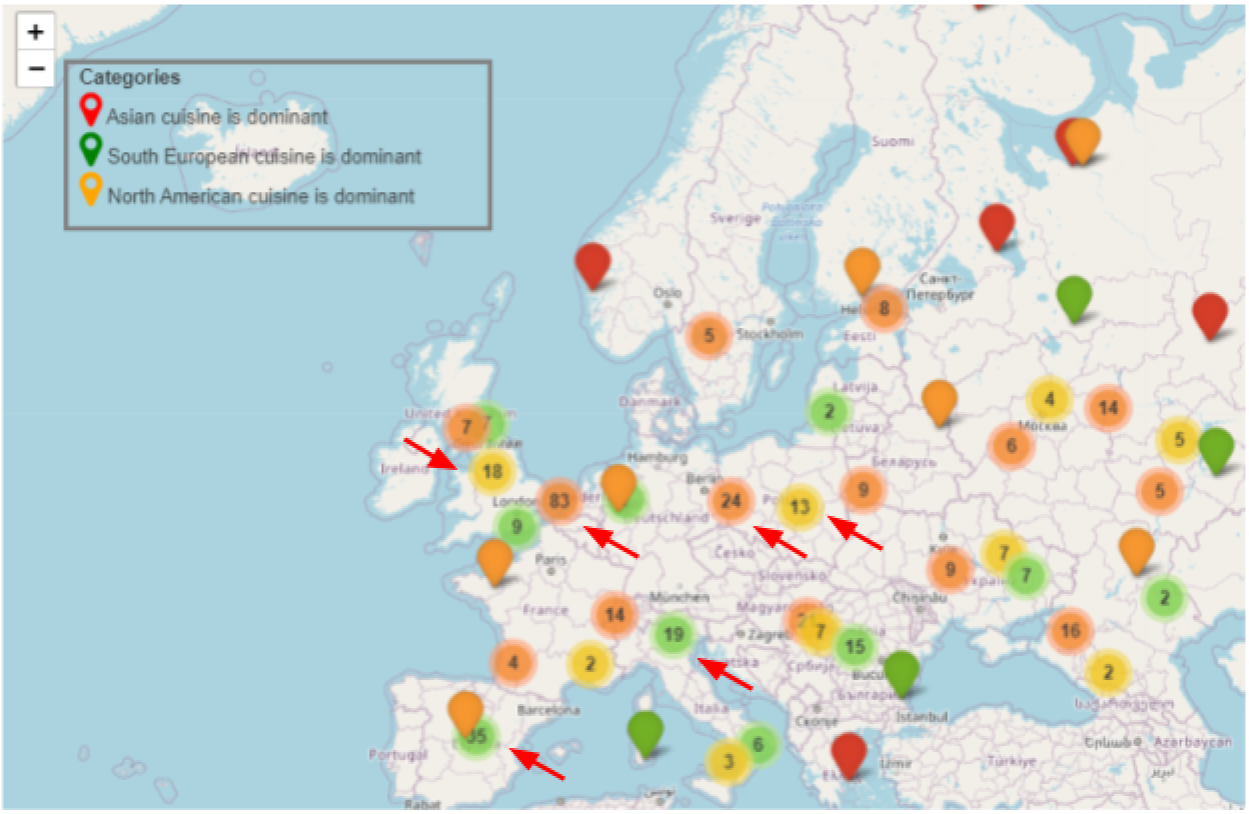


Fig. 13: Showing cluster of labels, were we find hot spots for certain cuisine.

## Discussion

The results for both rounds revealed areas in Europe where certain restaurants or cuisines are more common. This raises the question of what this means. If restaurants are more common in a city, this is an indicator that the citizens of that city appreciate the type of food served in these restaurants, otherwise the type of restaurant would not appear in this number, as it would not survive the competition with the other restaurants. If we think more globally, this means that if we find a group of cities with the same preferences in restaurants or in the cuisine, this will show which food is served most and is most popular in these areas of Europe. Therefore, investing in this type of restaurant in an area where locals appreciate this type of food would find enough potential customers and reduce the risk that the restaurant will not survive.

Looking at the results of this project, it can be seen, for example, that an Italian restaurant is quite common in Europe and finds enough potential customers in most cities. In Italy itself the conditions are even better and in Spain, for example, somewhat more difficult.

During the second round with more specific labels, it was possible to better explore the market for Chinese restaurants. It seems that we have a strong split between Northern and Southern Europe regarding the preference for Asian cuisine. It was obvious that restaurants serving Asian cuisine are more common in the northern part of Europe and therefore have more success there.

Although the third label of North American cuisine or fast food restaurants was the rarest, it was possible in both analysis rounds to identify certain hot spots for this type of restaurant or cuisine. It was found that there is a large market for investments in a fast food restaurant, especially in England, as they are more common there.

## Conclusion

The aim of this project was to identify European areas of interest for stakeholders.

This was possible. Firstly, it can generally be said that investing in an Italian restaurant in Europe carries a much lower risk of not finding enough potential customers than in other restaurants. Secondly, Asian cuisine is more popular in northern Europe than in southern Europe, where the influence of southern cuisine, e.g. mediterranean and Italian restaurants, predominates. This is of particular interest to our Chinese stakeholders, who now know that investing in the southern region of Europe would significantly increase their risk and that it would be advantageous to invest, for example, in Germany or France. Thirdly, we could suggest to our American stakeholders that they invest in England rather than in continental Europe.

This study could be followed up by a more detailed analysis of which city in particular should be invested in. For example, in which city in Germany a Chinese stakeholder would invest. We know that we will generally find enough potential customers for a Chinese restaurant, but the competition would not be too great in relation to the number of Chinese restaurants already present in the city. This could be further investigated by taking a closer look at the identified city and looking at an area of the city where competition with other Asian restaurants would be particularly low.

Overall, it can be concluded that this type of analysis has great potential and contains valuable information about the restaurant industry for any investor.