# Where to put City Longshot Restaurants in the States?

## Introduction:

A chain of Scandinavian restaurants “Longshot City Food” is specialised in opening restaurants in the city centre of big cities in Europe. They are now interested in entering the American marked. They only have experience from the European marked and based on that experience they want to get an idea of which city in the states is most likely to respond well to their kind of food.

## Data:

The “Longshot City Food” cooperation have provided me with a list of the European cities where they have restaurants and a corresponding score describing how well their food was received in those cities with a score from 1-3. They have also provided me with five potential cities (Los Angeles, San Francisco, New York, Minneapolis, Chicago) in the states that they are interesting in opening a restaurant in.

Google maps have been used to get the coordinates of all those European and American cities. The Foursquare API has been used to get a list of 100 categorised restaurants within 1000 meter of the city centre of each of those cities.

## Methodology:

The main objective in this task is to find a way to figure out which American city centre is most similar to the European city centres where “Longshot City Food” have had the most success. I have decided to use the distribution of restaurants (American, asian, sushi, Scandinavian etc.) close to the city centre (1000 m) as an approximation for the preferred kind of food in each city centre. It is our first assumption that similarities in restaurant distribution defines places with similar interest in restaurant. This assumption can of course be tested in various ways and statistics, but that is not within the scope of this project.

The first step in the analysis was to find the coordinates for all the cities from Google Maps and use those coordinates to find the distribution of restaurants within 1000 meter of the city centres using the Foursquare API. The correctness of the city centre coordinates was tested by using the folium map package in Python.

The Foursquare API contains a lot of restaurant categories ranging from bakery to pet cafes. It was decided in the analysis to only include the most common kind of restaurants in the analysis. The decision was done by hand but should in principle be done by consider the food varieties served at the “Longshot City Food” and by additional investigation and analysis. This is again outside the scope of this project.

The categories were first transformed to one-hot vectors for each restaurant. Then the hot vectors for each city was summed up to get a single one-hot vector describing the distribution of restaurants for each city.

Based on the one-hot vectors a K-mean algorithm was first used to categorise the restaurants into clusters. This was done to figure out if similarities could be found within the clusters and also to figure out if some American cities are very similar to some European ones.

The next step was to use a predictor to try to predict the success score by placing a restaurant in one of the American cities. Due to the very limited amount a data a k-nearest neighbour algorithm was used to predict the score based on the k-most similar European cities. The prediction results are useful to the restaurant chain in their decision process.

The data was not split into training and testing sets due to the small amount of cities included in the analysis. The analysis can be seen as a first approach to solving the problem and many improvements can later be built on top.

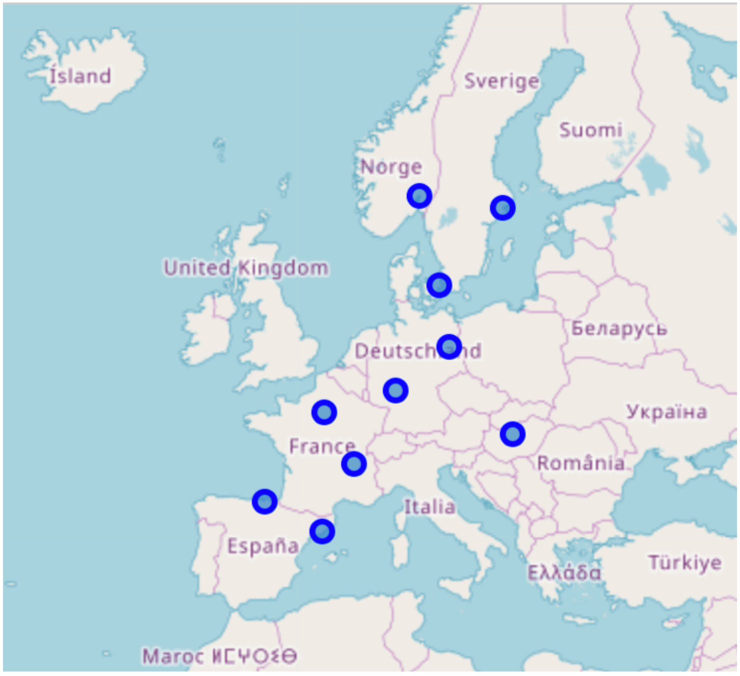
# Results:

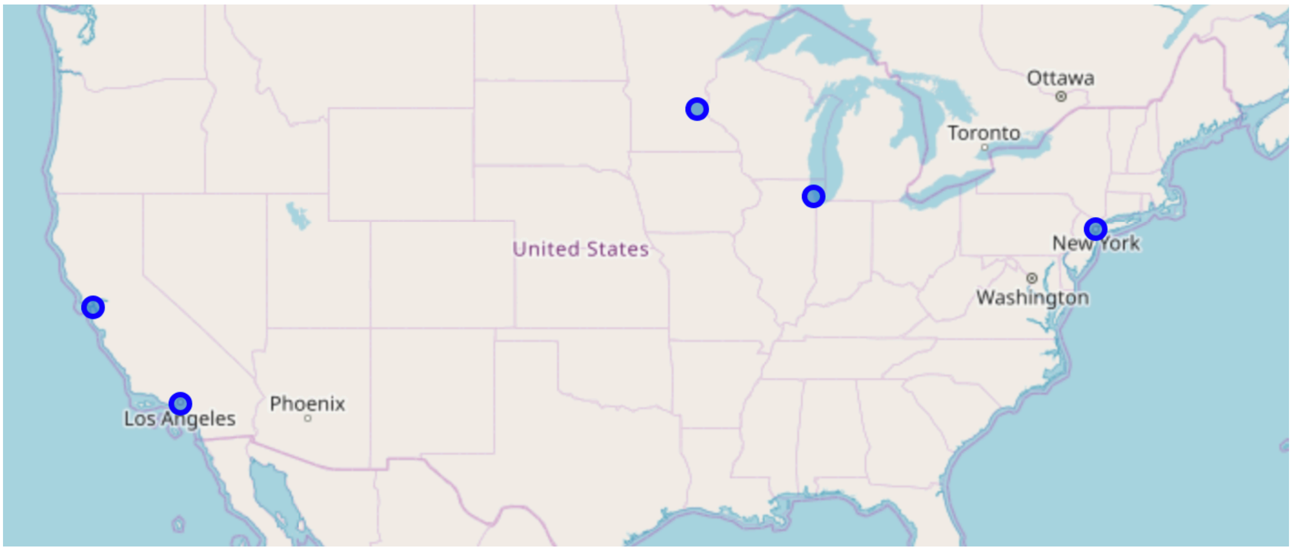
The table shows the European and American cities, their coordinates (from Google Maps) and the success score for the restaurants already placed in European cities.

Cities and their “Success” Score

| City | Latitude | Longitude | Score [1-3] |
|---|---|---|---|
| **Copenhagen** | 55.6760968 | 12.5683371 | 03.00 |
| **Oslo** | 59.913869 | 10.752245 | 03.00 |
| **Stockholm** | 59.329324 | 18.068581 | 03.00 |
| **Frankfurt** | 50.110922 | 8.682127 | 03.00 |
| **Barcelona** | 41.385064 | 2.173404 | 01.00 |
| **Bilbao** | 43.263013 | -2.934985 | 01.00 |
| **Budapest** | 47.49712 | 19.040235 | 02.00 |
| **Paris** | 48.856614 | 2.352222 | 01.00 |
| **Lyon** | 45.764043 | 4.835659 | 01.00 |
| **Berlin** | 52.520007 | 13.404954 | 02.00 |
| **New York** | 40.712784 | -74.005941 | NaN |
| **San Francisco** | 37.774929 | -122.419416 | NaN |
| **Los Angeles** | 34.052234 | -118.243685 | NaN |
| **Chicago** | 41.878114 | -87.629798 | NaN |
| **Minneapolis** | 44.977753 | -93.265011 | NaN |

Two maps were constructed to confirm the location coordinates of the cities.




Then information about all venues and their categorization in a radius for 1000 m around the city centres was extracted (limited to 100 per city) from the Foursquare API. Only restaurants within 25 pre-defined categories was included in the analysis and one-hot vectors were constructed. An example of a subset of such data is shown below:

| Cities | Burger Joint | Scandinavian Restaurant | Steakhouse | French Restaurant | Italian Restaurant | Fast Food Restaurant | American Restaurant | Asian Restaurant | Sushi Restaurant | Indian Restaurant | Mexican Restaurant | Vietnamese Restaurant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Barcelona** | 4 | 0 | 1 | 0 | 6 | 0 | 1 | 2 | 0 | 0 | 2 | 1 |
| **Berlin** | 3 | 0 | 1 | 0 | 8 | 0 | 0 | 4 | 1 | 0 | 0 | 11 |
| **Bilbao** | 2 | 0 | 2 | 0 | 3 | 1 | 0 | 3 | 1 | 0 | 1 | 0 |
| **Budapest** | 2 | 0 | 3 | 3 | 13 | 0 | 1 | 0 | 1 | 2 | 1 | 2 |
| **Chicago** | 4 | 0 | 1 | 2 | 11 | 0 | 4 | 3 | 0 | 1 | 1 | 1 |
| **Copenhagen** | 5 | 18 | 3 | 7 | 3 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| **Frankfurt** | 3 | 0 | 3 | 2 | 10 | 0 | 0 | 2 | 2 | 1 | 1 | 1 |
| **Los Angeles** | 0 | 0 | 1 | 3 | 6 | 0 | 1 | 0 | 10 | 1 | 9 | 0 |
| **Lyon** | 4 | 1 | 2 | 29 | 5 | 4 | 1 | 0 | 2 | 2 | 0 | 0 |

A k-mean (k=6) algorithm from the sklearn module was used to construct six clusters/groups as shown on the left table below. The right table shows the results grouped by "Group" with a description of the group and the average "success score" of the European cities in the Group.

| | City | Group |
|---|---|---|
| **0** | Barcelona | 4 |
| **1** | Berlin | 6 |
| **2** | Bilbao | 4 |
| **3** | Budapest | 6 |
| **4** | Chicago | 6 |
| **5** | Copenhagen | 3 |
| **6** | Frankfurt | 6 |
| **7** | Los Angeles | 5 |
| **8** | Lyon | 1 |
| **9** | Minneapolis | 2 |
| **10** | New York | 2 |
| **11** | Oslo | 3 |
| **12** | Paris | 1 |
| **13** | San Francisco | 2 |
| **14** | Stockholm | 3 |

| X | Cities | Description | Avg. Score |
|---|---|---|---|
| **Group 1:** | Paris, Lyon | French Cities | 1.0 |
| **Group 2:** | Minneapolis, New York, San Francisco | American Cities Group 1 | NaN |
| **Group 3:** | Stockholm, Copenhagen, Oslo | Scandinavian Cities | 3.0 |
| **Group 4:** | Bilbao, Barcelona | Spanish Cities | 1.0 |
| **Group 5:** | Los Angeles | American Cities Group 2 | NaN |
| **Group 6:** | Budapest, Chicago, Berlin, Frankfurt | Mixed Group with one American City | 2.33 |

A predictor based on k=5 nearest neighbours algorithm from sklearn and the results from the European cities predict the following results for the american cities.

| | Score Pred. | City |
|---|---|---|
| **0** | 2.2 | New York |
| **1** | 1.8 | San Francisco |
| **2** | 2.2 | Los Angeles |
| **3** | 1.8 | Chicago |
| **4** | 2.0 | Minneapolis |

# Discussion:

The clustering of the cities look very promising as it managed to isolate categories for Spanish, France and Scandinavian cities. Two categories had only American cities. Only one category had mixed cities with Chicago, Frankfurt, Berlin and Budapest. This suggest that Chicago could be more similar to some European cities than the other American cities in this analysis. The group also had decent average score of 2.33 which suggest that Chicago might be an interesting place to start putting a restaurant as the challenges in adoption might be easier than for the other American cities.

That Chicago is more similar to some European cities does not necessarily mean it is the best place to put the first restaurant. A prediction model based on nearest neighbours showed another picture of the situation. According to the model, Chicago and San Francisco are the worst places to put a restaurant and New York and Los Angeles are the best places to put a restaurant.

This illustrates that more investigations are needed to resolve the situation. The nearest-neighbour model only works if the parameter space is well sampled. If the american cities to be predicted are far away in the parameter space from the European cities, the nearest neighbour algorithm might not give realistic results. This analysis is outside the scope of this project.

Another factor is that only 100 restaurants per city is included in the analysis, which might be improved with a professional access to FourSquare.

# Conclusion:

The clustering algorithm seemed to work pretty well for the dataset, with successfull clustering into Spanish, French, Scandinavian and American cities. The nearest-neighbour algorithm predicted the best “success” score for a new restaurant in New York and Los Angeles, but a deeper investigation of the parameter space is needed to decide if those cities are sufficiently close to their nearest neighbours for a meaningful result. This also shows the complexity of making predictions from a limited dataset with many features.