What kind of restaurant in Riga?

February 2020

Cornelis Jan Roele
cees.roele@gmail.com

Table of Contents

Introduction

Opportunity for starting new restaurants

A record number of tourists visited Riga last year - 3.5 million, which was 1.4 times more than in 2017, the Riga Tourism Development Bureau informed LETA.

The largest number of tourists visiting the Latvian capital in 2018 were from Germany - 176,000 or 11.9% of the total number of foreign tourists visiting the city. This was followed by tourists from Russia - 173,800, or 11.6%, Estonia - 109,800, or 7.4%, Lithuania - 100,400 or 6.7%, and Great Britain - 98,700 or 6.6%.

As the image above illustratives, year over year the city of Riga receives more visitors. They come on business, to see family and friends, to see the sights, and generally to enjoy spending time in the city. They need a place to stay. And they need a place to eat.

The rise in the number of visitors means more potential restaurant customers and hence opportunity for starting new restaurants in the city.

Obviously, entrepreneurs starting a restaurant want it to be successful. Introducing his *“Top 20 factors for success in the restaurant business“*[1] (2016), Geoff Wilson writes:

> *"Success means different things to different people. Some operators want to just make a living. Others have loftier goals – maximization of market share, achievement of targeted return on investment and so on. Regardless of one’s definition of success, the basics in the restaurant industry never change."*

Some factors for success are:
- Location
- Quality food
- Scheduling to balance labour to demand
- Empowered staff
- Menus engineered to yield optimum gross margin
- Participative management

---

[1] “*Top 20 factors for success in the restaurant business”* (2016) Geoff Wilson, https://www.restobiz.ca/top-20-factors-success-restaurant-business/

- Winning attitude

They are out of scope of the current investigation.

We will focus on the ***first*** of Wilson's success factors:

> *Validated concept definition. Can you clearly state what experience your restaurant offers, what products it serves and what service-style it employs? If not, you’re not sure what your restaurant is all about and neither are your customers.*

Or rather, ignoring service-style, on *products served*.

Visitors coming from abroad have built up expectations and tastes, possibly developed by visiting other cities. At the same time local customers develop new tastes, possibly also by travelling abroad.

Business problem

Above we have seen that there is opportunity for opening new restaurants in Riga, that the first thing to be clear about is to know what products to serve, but that it is uncertain what product has a relatively good chance of success.

The present investigation seeks to answer the following question of prospective entrepreneurs wanting to open a restaurant in Riga:

What type of restaurant, serving what type of food, is relatively likely to be successful in Riga?

Data

Source

FourSquare is a location data company. Its website can be found at https://foursquare.com/

The company writes on itself:

> Our human-sourced database of 105 million places provides the richest index of venue information—from ratings and reviews to tips and tastes, our data paints an intricate story of place.

The FourSquare service allows developers to obtain data about locations world wide. This can include pictures, number of people who have checked into that location, reviews, and ratings.

Currently, we are interested in restaurants. The data made available by FourSquare, includes the following fields which we are interested in:

Field	Description
id	unique identifier for venues, e.g. "5284cb6511d2e6ccc1a464f3"
name	name of the venue, e.g. "Rijksmuseum Café"
category::name	name of the FourSquare category of the venue, e.g. "Italian Restaurant"
location::latitude	float identifying the venue's latitude
location::longitude	float identifying the venue's longitude

In the table above the "::" denotes that the right hand side is an attribute of the attribute on the left hand side.

Selection of data

Methodology, to be discussed later, in the full report, is based on comparing data from different cities. As we have taken our starting point from the growth of the number of tourists, we will first identify a number of touristic venues spread over Europe. We take as our beginning museums as they are a kind of venue that is present in many cities.

An overview of museums was found by looking at a number of travel blogs and manually adding latitude/longitude coordinates using the Nominatim geolocation service on the address of each museum.

We select data from the FourSquare service as follows:

- Query FourSquare for venues within a certain *radius* of the museum which we have taken as a basis for search.
- Query FourSquare for venues *falling under the top level category "Food"*, which is the category under which restaurants fall.

Filtering of data

After receiving the data from the FourSquare Service, we filter the result to only include those venues with the word "Restaurant" in the category, so as to filter out e.g. coffeeshops, wine bars, and foodstands.

How will this data serve to answer our question?

Remember that we try to answer the question: ***What kind of food served is relatively likely to make a restaurant successful in Riga?***

The selected data will allow us to compare the composition of restaurant categories among those venues near museums in different cities. Where the composition of restaurant types is relatively similar, we look at what restaurant categories in Riga are underrepresented. Based on the experience of other cities where restaurant customers turn out to have similar tastes, those underrepresented categories of restaurants are relatively likely to have a market in Riga and hence become successful.

Methodology

We seek to answer our question through comparison: what works in multiple similar cities is relatively probable to work in Riga.

Selecting areas for comparison

Our first step is to:

- Identify cities to compare Riga with
- Identify similar areas in cities

As we start with the assumption that tourists form a significant segment of restaurant customers - and growth of the number of tourists provides market opportunity for opening more restaurants - we take museums as one type of tourist attraction as the criterion for including cities and as the central point for comparing areas with restaurants.

We obtain a list of museums from overviews in travel blogs and identify their latitudes / longitudes through the Nominatim geolocation service. These coordinates will be the center of the areas we are interested in. Museums play no other role in our investigation than providing these central coordinates.

Exploratory data analysis and cleaning: not all addresses found of museums can be located automatically through the geolocator. Where missing, the latitude/longitude coordinates had to be provided manually by identifying the museum through Google Maps.

Selecting restaurants per area

We find restaurants through using the FourSquare service. We query it to obtain:

- Venues within a certain *radius* from each area center defined above
- Only those venues under the FourSquare “Food” category, which includes restaurants

As radius we take a walking distance of circa 700 meter from the area center (museum).

We clean up the obtained data by including only those venues that have a category including the word “Restaurant”.

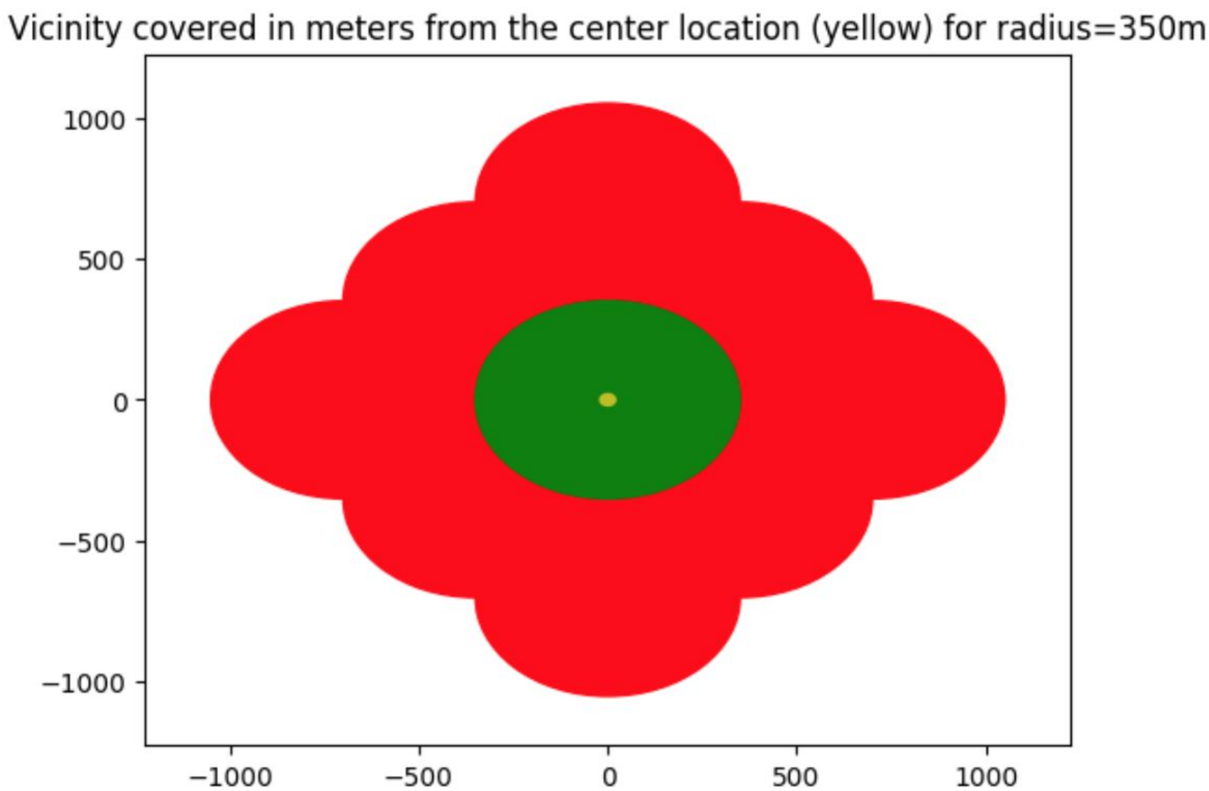
This will effectively give us an overview of how many restaurants of what type there are in each area.

Exploratory data analysis and fix:

The FourSquare REST API limits the result of querying to fifty venues. In other words, even if we set the radius for searching venues around a central point to two kilometers, we will still find no more than fifty venues.

1. The ‘radius’ parameter used to create a query to FourSquare should be small enough to lead to a return of all the actual venues. Some experimentation led to a value of 350 meters.
2. As a significant number of the returned venues are not restaurants, to be filtered out programmatically, even fifty venues leaves us in practice with a maximum result of some twenty restaurants per search. This is less than we want. Therefor, the present investigation is done using a workaround: around the location coordinates central to our search for venues, we automatically create a number of locations surrounding the central location and create a FourSquare query for venues with a relatively small radius parameter for each of them and yet in total cover a larger area.

By launching multiple queries to the FourSquare REST API, a radius parameter set to 350 meters actually covers an area as depicted in the diagram below:



Clustering

Before we compare restaurants in Riga with those in other areas, we want to filter out those areas that are dissimilar with respect to existing restaurant types. Generally, we want to draw conclusions based on different distributions of types of restaurants in different areas, but these are valid only on the basis of an existing similarity. From a similarity of restaurant types we conclude that there is a similarity of taste of customers and that similarity of taste provides the basis for making a prediction on the relative probability of success for starting a new restaurant of a certain type.

We do this by applying K-means clustering of the different areas based on the distribution of their restaurant types.

Comparing

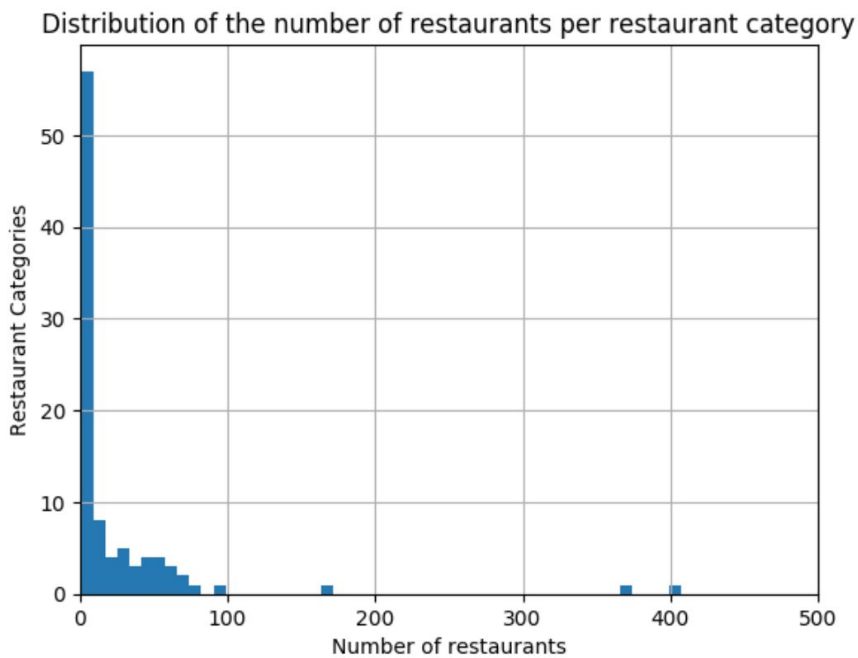
We will compare the distribution of restaurant types in Riga with that in other areas within its cluster. For this we will use visualization through boxplots, which provides a quick insight in how restaurant types are distributed in Riga compared to the other areas.

Results

From a set of *28 museum locations* spread over Europe we found *2468 restaurants* in their vicinity using the FourSquare REST API. As FourSquare limits the search result within a radius to 50 items, we located nearby restaurants by defining a number of locations surrounding the prime location and query nearby venues for all of those, see in the Methodology section above.

Restaurants, restaurant categories, and museum vicinities

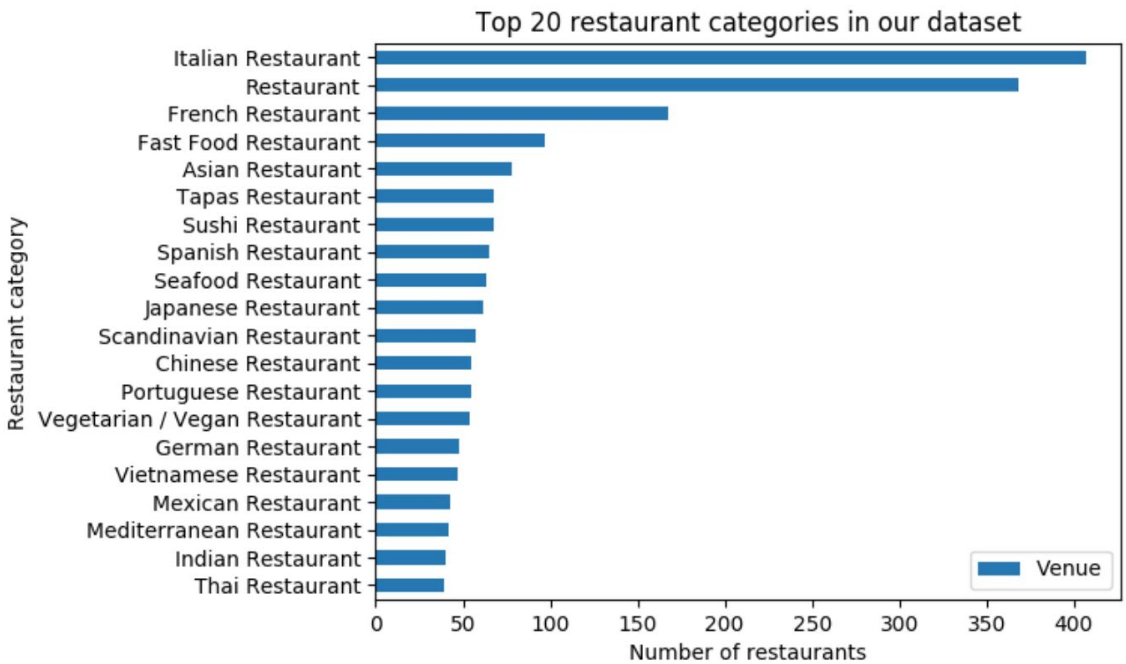
Now, let's see how restaurants are distributed over restaurant categories in our entire dataset:



How should we interpret this histogram? It tells us that:

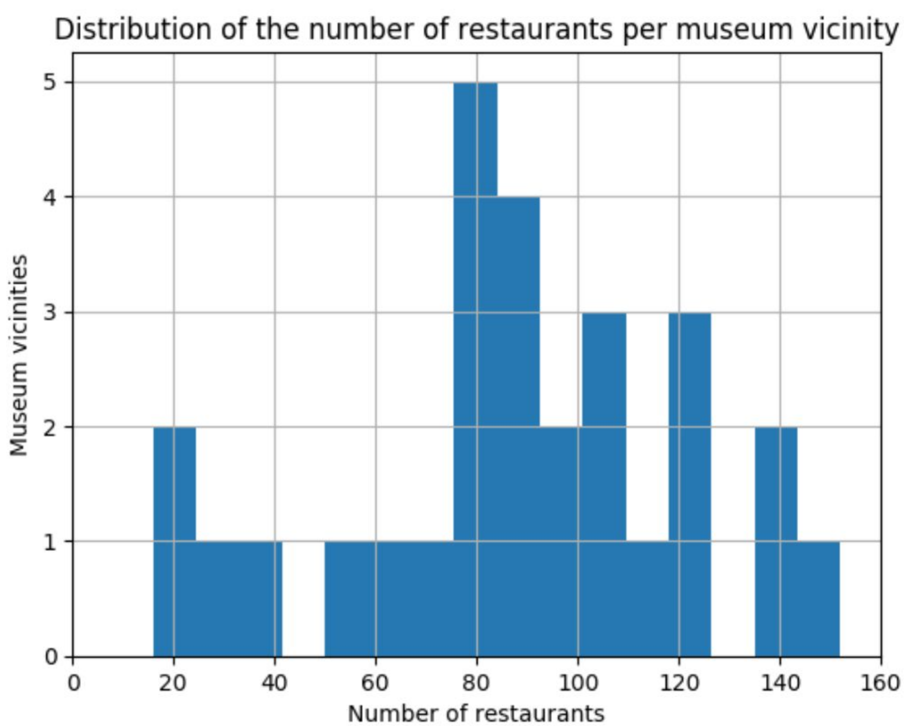
- In the vast majority of restaurant types we find only up to 10 restaurants all over our dataset
- Three restaurant types have more than 100 restaurants and two of them several hundred.

Given the unequal distribution, let's see what the very popular restaurant categories are:



Italian Restaurant and *"Restaurant"* (generic label) are really springing out, with *French Restaurant* also containing significantly more restaurants than the other categories.

Let's see now how the number of restaurants is distributed over the different areas that we have defined, that is, vicinities of museums:

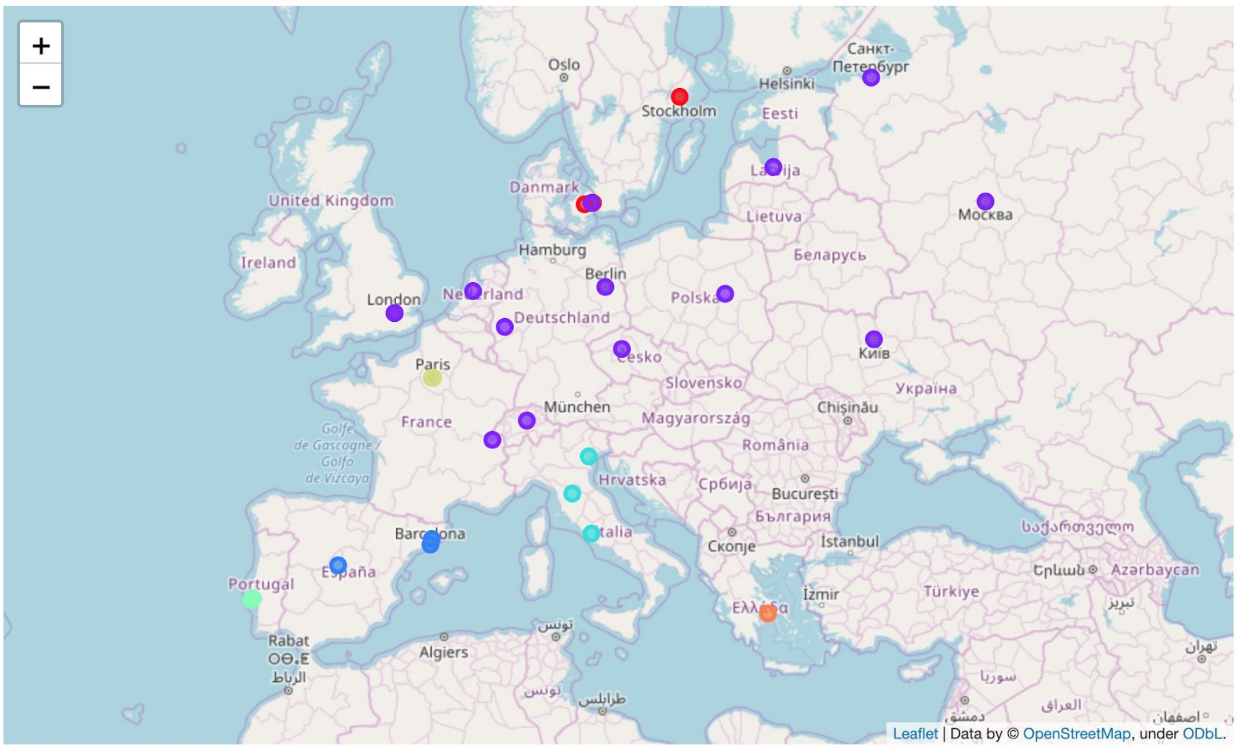


We see in the histogram that the majority of our defined areas contain some 50 to 125 restaurants. In a few we find less than 40 restaurants and those areas might not be sufficiently representative.

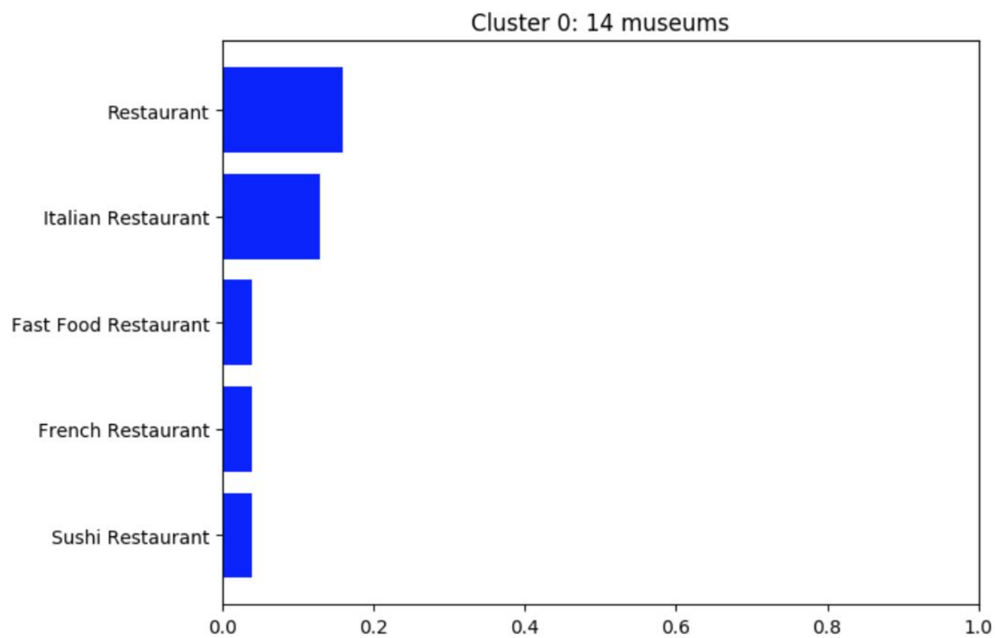
Clustering museum vicinities

After preparing the data with a “one hot” encoding of restaurant types near museums and taking their means, we can using the K-means algorithm to cluster the different restaurant vicinities.

The outcome is visualized below. Different colors designate different clusters.



Half our museum vicinities fall into a single cluster - with the purple dots in the chart above. Let’s look at the most present restaurant categories in that cluster:

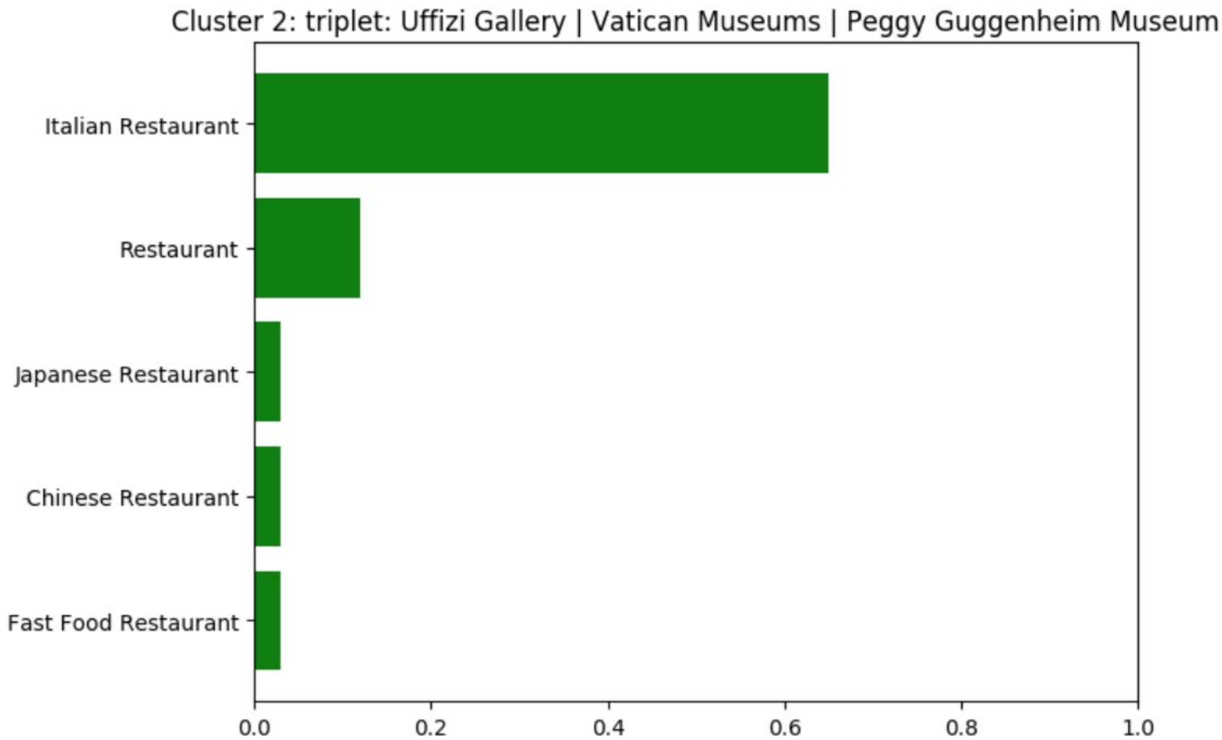


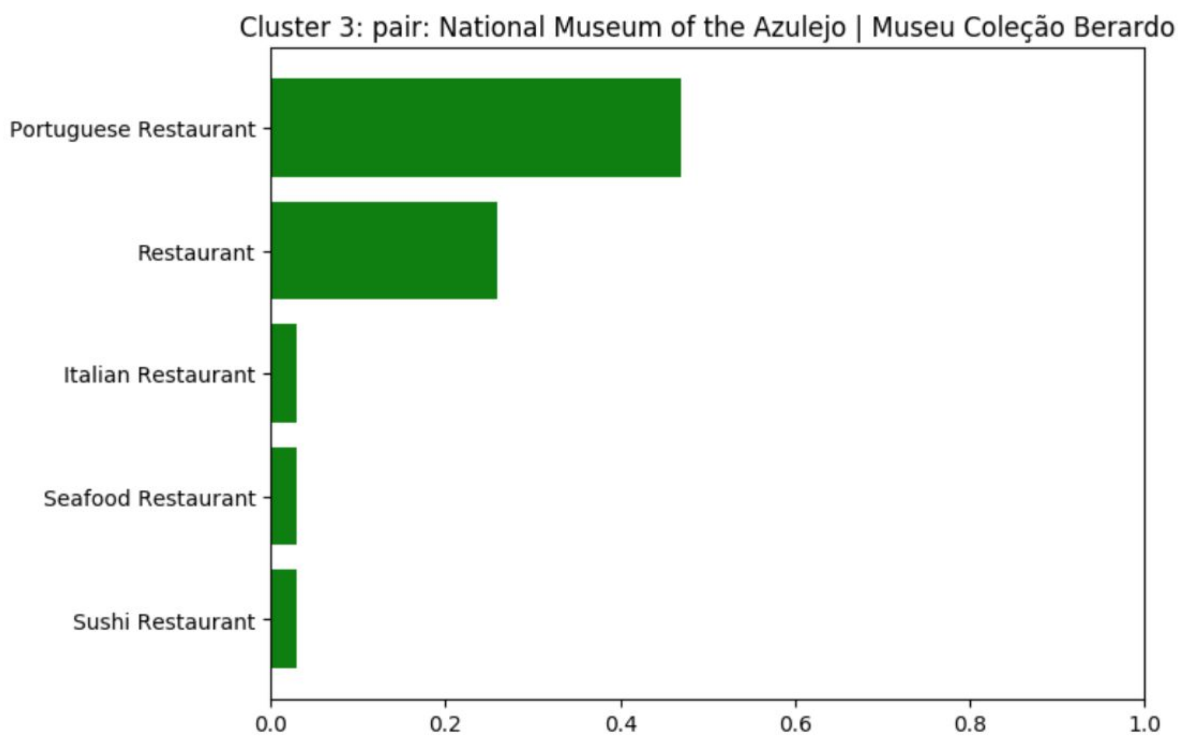
What we see, with “Restaurant”, “Italian Restaurant”, “Fast Food Restaurant” and “French Restaurant” is that the restaurant categories we find most in that cluster are pretty much those that we find most in the entire dataset. We can hence say that this cluster lacks special characteristics.

The other clusters are more interesting. They are all dominated by national cuisines, that is:

- Italian Restaurant in Italy
- Portuguese Restaurant in Portugal
- Spanish / Tapas Restaurant in Spain
- Scandinavian Restaurant in Denmark/Sweden
- French Restaurant in France
- Greek Restaurant in Greece

Two examples below:





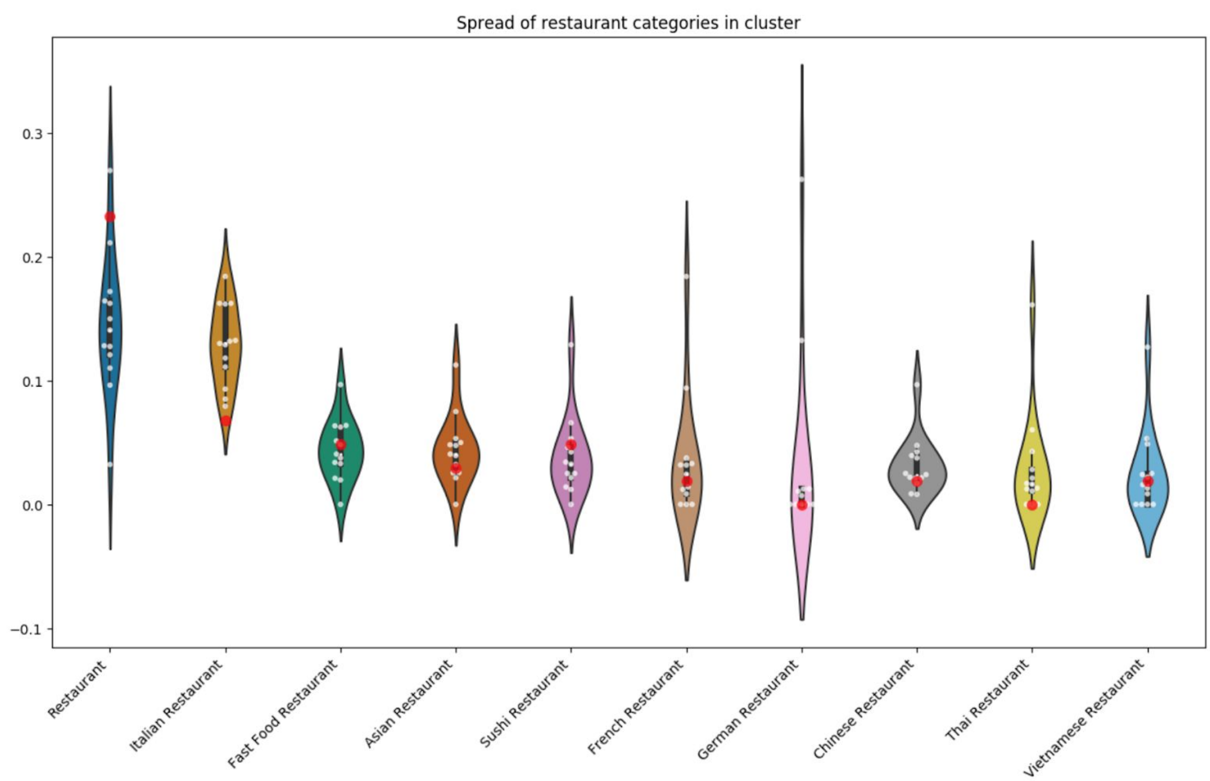
If we apply the K-means algorithm to distinguish more clusters we do find clusters with a presence of national cuisines, e.g. German Restaurant and Polish Restaurant for Germany and Poland respectively, but these are not dominant as in the above examples and clusters become more like different distributions of the dominant restaurant types of Italian, “Restaurant”, French, and Fast Food.

Comparing restaurants in Riga with the cluster it is part of

Our goal was to find what restaurant type would have a relatively good probability of succeeding when opened in Riga. The chosen method was to compare restaurants in Riga with those in “similar” areas. Here “similar” is implemented by comparing Riga with other areas in the same cluster.
We find Riga to be part of the largest clusters and hence not dominated by a national cuisine.

To find where restaurant categories in Riga are underrepresented, our primary concern as that is where new restaurants might be successful, we look at the ten most common restaurant types in the cluster and compare Riga with the rest.

In the diagram below we see a violin plot - a variation on the box plot with variable width - where a swarm plot identifies the mean values for other areas, and a red dot the mean value for Riga. If the red dot for Riga comes out higher than others, that restaurant category is overrepresented. If it comes out lower, it is underrepresented.



Below the comparison of Riga with the other areas in its cluster has been put in a table:

Restaurant type	Comparison, in Riga ...
Italian restaurant	**Very low**
"Restaurant"	**Very high**
Asian restaurant	**Very high**
Fast Food restaurant	**Very low**
German restaurant	**average**
French restaurant	**high**
Sushi restaurant	**average**
Thai restaurant	**average**
Seafood restaurant	**high**
Chinese restaurant	**average**

Following our approach, we were looking for those restaurant categories underrepresented in Riga, as their presence in other similar locations indicates that they have a relatively good probability of success.

Narrowly following that approach, we are to conclude from the table above that Riga could have more:

- Italian restaurants
- Fast Food restaurants

Discussion

Summary of the methodology

Our goal:

Answer the question of potential entrepreneurs in the restaurant business:

> *What type of restaurant has a relatively high probability of success in Riga?*

Methodology:

1. Start with a number of areas in Europe comparable with the area of Riga visited by tourists.
2. Get restaurants and their types for all these areas.
3. Cluster the areas to identify those that are similar to Riga
4. Identify the restaurant types that are underrepresented in Riga compared to the rest of the cluster
5. Conclude that those restaurant types have a relatively high probability of success.

Observations

Clustering

Clustering of different areas based on the types of restaurants served to identify several areas with strong local cuisines.

Thus we find *national/regional clusters* **dominated by a single type of restaurant**:

- *Italian* restaurant
- *French* restaurant
- *Portuguese* restaurant
- *Greek* restaurant
- *Scandinavian* restaurant
- Tapas / *Spanish* restaurant

We wanted to compare Riga with other areas that are similar and now we find Riga to be in the largest cluster.

We find that the top restaurants in that cluster, less dominating than in the previous ones, are:

1. Italian restaurant
2. "Restaurant"
3. Asian restaurant

4. French restaurant
5. Fast food restaurant

This may be a consequence of the clustering process. If instead of 7 clusters we opt for a division into 13 clusters, not visible straight away in the notebook, obtained by re-setting the parameter in the notebook above, we find some new clusters along the same national cuisine pattern, e.g. “German restaurant” as the primary restaurant near the DDR museum in Berlin and "Polish restaurant' and "Eastern European restaurant" in Warsaw. But we also find that rather than dividing up the larger cluster into national outliers, the smaller - already nationally defined - clusters are cut up further.

Also, the large cluster we found with 7 clusters now gets cut into two parts, with the first having "Italian restaurant" and "Restaurant" as the top two restaurant types and the second with "Restaurant" and "Italian restaurant" as the top two restaurant types.

Quite generally, we find that "Italian restaurant" and "Restaurant" are by far the most common types of restaurants in our dataset.

As similarity doesn't seem to increase significantly by having more clusters a size of 7 has been used for further analysis as this removes the strong national cuisines from the set to compare Riga with.

One more observation. Remember the statistics on where tourists to Riga come from? In order: Germany, Russia, the other Baltic countries, and Great Britain. Now look again at the map visualizing clusters: you will see that the visitors of Riga come from areas in the same cluster. That supports the assumption in this investigation that similar clusters designate similar tastes.

Comparison and reconsideration of the data

How does Riga compare to other areas in its cluster? We investigated this by visually comparing the relative presence of restaurant types of the ten most popular restaurant types in the cluster.

We noted at the end of our Result section that the following restaurant categories are underrepresented in Riga and have hence a relatively good probability of success:

- Italian Restaurant
- Fast Food Restaurant

However, the data we have on categories lacks some information:

- The 'Restaurant' category is **over**represented. But what is it? A generic restaurant, serving anything, or can it be just any restaurant that could have been categorized more precisely?
- In the top 10 list with find 'Asian restaurant', but also 'Sushi restaurant', 'Thai restaurant', and 'Chinese restaurant'. The latter three categories are all Asian restaurants and could better be presented as a single category.

When re-checking the FourSquare categories for venues, we find that there is no such thing as “Latvian Restaurant” or “Baltic Restaurant”? The relatively large size of the generic “Restaurant” category may be explained by having all restaurants serving national cuisine being put in this category in the absence of a more precise categorization.

Conclusion

What type of restaurant has a relatively high probability of success in Riga?

We found that two kinds of restaurants are underrepresented in Riga compared with other locations where people have relatively similar tastes:

1. Italian restaurants
2. Fast Food restaurants

Additionally, we found that the generic category of "*Restaurant*" is significantly overrepresented in Riga. What are we to make of this? Is a more precise categorization possible? Can it include restaurants with an emphasis on local cuisine, without labelling it e.g. "*Latvian*" or "*Baltic*"? One reason for the absence of such labels may be that they are not currently available in FourSquare's categorization, so someone would have to introduce them.

In the course of our investigation we found a strong representation of national/regional cuisines in Italy, Spain, Portugal, Greek, and Scandinavia. Given the success of national cuisines elsewhere, the prospective restaurant opener might consider if Riga has more of a market for this, but the present investigation is based on data that fails to identify such a cuisine.

But for the purpose of getting clearer data, and for marketing, restaurant owners in Riga, and in the Baltics generally, might consider branding their cuisine!

References

- FourSquare categories for restaurants (and venues generally): https://developer.foursquare.com/docs/resources/categories
- "Search" in FourSquare REST API: https://developer.foursquare.com/docs/api/venues/search
- Blog article mentioning "must visit" museums in Europe: https://culturetourist.com/museums/15-best-museums-in-europe-you-have-to-visit-this-year/
- *"Top 20 factors for success in the restaurant business"* (2016) Geoff Wilson, https://www.restobiz.ca/top-20-factors-success-restaurant-business/