New Restaurant in Riga

IBM Data Science project

Andis Orbidāns

20th June 2020

1 Introduction

Riga, capital of Latvia, is metropolis of Baltic states (Latvia, Lithuania, Estonia). It is the largest city with population of 627 thousand and around 1 million in its metro area. It's a vibrant and active city whose citizens are quite outgoing. There are around 8 thousand international students living in the city which adds to vibrant flavor of Riga. Interest from EU and Eastern Europe tourists has been also on the rise. Since the end of previous financial crisis around 2010-2012 Riga has experienced solid economic growth, which, given increasing presence of demographic problems, translates rather strongly to wage growth. The effects of all of this can be easily observed in the menus of local restaurants - prices have visibly inflated. Demand for dining options outside home has surged. This in turn creates wide range of opportunities for new dining ventures.

The restaurant scene is quite saturated in central neighborhoods (especially, Old Town and Center) and is not clear whether there are space for new ventures. Therefore the problem is to determine if there is a market for dinning places outside the center. And what could be the best neighborhood as location for new restaurant. The potential new restaurants might have two

possible focuses - either late evening dining or restaurant focused on cheaper business menu lunches to serve masses.
This projects aims to find best suited neighborhoods for a new restaurant. Main inputs for analysis will be Foursquare data on venues in each neighborhood.
Results of this report would be compelling for food industry entrepreneurs or possibly Riga City Municipality to see which of the neighborhoods could be developed further to increase its appeal for new restaurant related business ventures.

2 Data

2.1 Data acquisition

In order to answer question posed in the Introduction section I firstly will determine which of neighborhoods of Riga are developed enough in terms of commercial, residential infrastructure but don't have sufficient number of dining places. This is be achieved by collecting venues data in each neighborhood through Foursquare API. List of all venues in 750m radius from the center of neighborhood is retrieved. For example, venues will include parks, dog runs, eating places, fitness centers etc. It's going to be full variety of venues which will give a clear picture of what activity is going on in given neighborhood.
Additionally data on size, population in neighborhoods is to be collected from Riga municipality home page.[1] However, that population and people employed data is outdated therefore I look for more recent sources. The more up-to-date data on population (from year 2019) of each neighborhood is downloaded from Central Statistical Bureau of Latvia (the direct download point for data from Central Statistical Bureau of Latvia is The Latvian Open data portal where the data is stored).[2] This will help to determine

inhabitant density of neighborhoods which would be one of the measures of potential demand for dining places.
Specifically to measure demand for eating places offering business menus, data on number of people employed in each neighborhood is be downloaded from Central Statistical Bureau of Latvia.[3]
As a measure of purchasing power data on average wage (EUR/month) in each neighborhood is to be downloaded as well from Central Statistical Bureau of Latvia.[4].
Lastly, location data is gathered trough Google Maps API.

2.2 Data preparation

When searching for Riga neighborhoods location data in Google Maps, some of the neighborhoods are not recognized correctly. For instance, if we search for "Centrs, Riga" (Centrs is one the neighborhoods in Riga), as result we get whole Riga. Additionally, there is a village with same name as one the neighborhoods next to the neighborhood, therefore searching plainly for its name combined with "Riga" yields wrong result. To resolve this issue for problematic neighborhoods (there are only 2 of them) as a second feature I add higher level Riga subdivision - district (I specify it only for problematic neighborhoods as there is no need to go through hassle to collect information for all neighborhoods).
Population data and income is readily available for years 2019, however, number of people employed is only for year 2017. To have them all three on the same level, I artificially calculate number of people employed in 2019 by multiplying number of people employed in 2017 by rate of change of population from year 2017 to 2019. Theoretically, it could be done for each neighborhood separately, but employment structure by neighborhoods and population structure by neighborhoods is visibly different and, for instance, high population decrease in one neighborhood does not necessarily mean high population decrease in its number of people employed since there could easily

be new jobs introduced in the area especially if it is industrial neighborhood.

2.3 Exploratory/Descriptive statistics

In this section I will do a short overview of the main neighborhood data collected. In the first figure descriptive statistics of 4 features is exhibited.

	Size (Ha)	Income 2019	Population 2019	Employed 2019*
count	58.0	58.0	58.0	58.0
mean	515.0	1153.0	10854.0	8134.0
std	356.0	265.0	13409.0	11286.0
min	74.0	654.0	69.0	39.0
25%	244.0	1018.0	1216.0	792.0
50%	443.0	1106.0	4989.0	4085.0
75%	699.0	1243.0	17244.0	13506.0
max	1873.0	2316.0	55579.0	73642.0

Figure 1: Descriptive statistics of 4 data features collected.

As we can see in total there are 58 neighborhoods. All four features have wide dispersion. Area can vary from as small as 58 hectares to as large as 1873 hectares. Most of the largest neighborhoods have large areas of forests, plains, lakes or other uninhabited areas. Income has lower variety - standard deviation is only 23% of mean value. The most affluent is Skanste neighborhood having average wage of 2316 EUR/month is considerably more than the second wealthiest neighborhood Vecpilsēta (Old Town) figure 2, on the opposite side is Spilve with 654 EUR/month (figure 3). Standard deviation for Population and Employed is very high, it's even higher than respective mean values. As we can see bottom 10 neighborhoods in terms of population and number of people employed doesn't go above 500 inhabitants, and Voleri (10th lowest by Employed) barely exceeds 500 employees. The most populated neighborhood is Purvciems. Purvciems is also in the top 10 largest employers (9th place). Most workplaces are located in Centrs with 74 thousand places which is almost three times higher than the second place Maskavas forštate wth 26 thousand places. There are 4 neighborhoods which are both in top 10 neighborhoods by population and top 10 by number of employed. We can consider these neighborhoods very active.

	Ha	Ha#	Income 2019	Income 2019#	Population 2019	Population 2019#	Employed 2019*	Employed 2019*#
0	Kleisti	1873.0	Skanste	2316	Purvciems	55579	Centrs	73642.0
1	Jugla	1409.9	Vecpilsēta	1805	Ķengarags	46541	Maskavas forštate	25599.0
2	Mežaparks	1182.1	Buļļi	1665	Imanta	44189	Teika	25159.0
3	Trīsciems	1131.9	Ķīpsala	1552	Pļavnieki	42048	Vecpilsēta	22810.0
4	Daugavgrīva	1014.7	Centrs	1521	Ziepniekkalns	32108	Sarkandaugava	18803.0
5	Šķirotava	1005.7	Suži	1472	Centrs	30557	Āgenskalns	17439.0
6	Spilve	957.6	Mežaparks	1468	Teika	28720	Brasa	17190.0
7	Jaunciems	913.2	Vecāķi	1402	Āgenskalns	25047	Torņakalns	16792.0
8	Imanta	900.3	Brasa	1351	Maskavas forštate	24659	Purvciems	16527.0
9	Bolderāja	832.9	Bieriņi	1343	Jugla	24011	Čiekurkalns	16236.0

Figure 2: Top 10 neighborhoods and their respective values for each of the features.

	Ha	Ha#	Income 2019	Income 2019#	Population 2019	Population 2019#	Employed 2019*	Employed 2019*#
0	Atgāzene	74.5	Spilve	654	Spilve	69	Buļļi	39.0
1	Vecpilsēta	94.4	Voleri	732	Salas	70	Suži	62.0
2	Zasulauks	119.0	Katlakalns	829	Katlakalns	185	Mūkupurvs	142.0
3	Beberbeķi	120.4	Rumbula	878	Mūkupurvs	237	Vecdaugava	150.0
4	Šampēteris	136.6	Bolderāja	879	Voleri	255	Beberbeķi	206.0
5	Grīziņkalns	151.7	Kundziņsala	880	Buļļi	295	Vecāķi	316.0
6	Katlakalns	155.4	Daugavgrīva	916	Kundziņsala	345	Brekši	353.0
7	Brasa	174.1	Vecmīlgrāvis	921	Kleisti	437	Trīsciems	398.0
8	Avoti	181.5	Brekši	930	Suži	460	Jaunciems	410.0
9	Ķīpsala	197.5	Maskavas forštate	960	Beberbeķi	462	Voleri	502.0

Figure 3: Bottom 10 neighborhoods and their respective values for each of the features.

3 Methodology

The first step is to collect venues data for each neighborhood in Riga. The max limit of venues to be collected is set to 100. Radius of an area around the center of a neighborhood is set 750 m, it is slightly larger than usually used as to be able to cover larger neighborhoods, which in case of Riga can reach up to 18 km2 in size (in comparison area with radius 750 m covers around 1.8 km2). Nevertheless, larger neighborhoods are mostly industrial or green territories where there are not much venues in general and there is no need try to cover them fully. On the other hand radius can't be to larger since there are also very small neighborhoods (0.7 km2) and with large radius many venues for bordering neighborhoods will be collected.
Based on the venues data, neighborhood profiling will be done. This will be attained by, firstly, calculating what proportion each venue in each neighborhood holds from total number of venues in a given neighborhood. These proportions will be used to cluster neighborhoods. Clustering will be executed by K-means algorithm [5]. The optimal number of clusters to be used

in algorithm, will be determined by calculating sum of squared errors for each cluster (i.e., calculating distance for each neighborhood from the closest cluster center and squaring it) and then taking sum of all these single cluster sums for each possible number of clusters:

$$\sum_{i=1}^{C}\sum_{k=1}^{N_i}({}_in_k - c_i)^2,$$

where C is number of clusters, N_i denotes number of neighborhoods in cluster i, ${}_in_k$ neighborhood value belonging to cluster i and c_i is the center of cluster i.

Then the number of clusters for which decrease in sum of squares versus previous number of clusters will start to slow down, i.e. , for each next number of clusters it will become marginally small, will be considered the optimal. This will be determined graphically by so called ”elbow method”.

Then qualitatively cluster most suitable for new restaurant will be determined. Main criteria will relative number of restaurants in the neighborhood, relative number of other socially appealing venues in the neighborhood and relative number of the most common venues (fitness center, shopping mall etc.) for active public life in the neighborhood. No exact number will be determined, judgment will be based on presence of aforementioned venues in top 10 most frequent venues in a given neighborhood.

For more quantitative analysis Restaurant Attractiveness index will be composed. First component of it will be density of population and number of employed summed together per hectare, second component average wage, and lastly density of restaurants in the neighborhood - number of restaurants per 1000 people living and being employed in a given neighborhood. Number of restaurants will be calculated by counting the following list of venue categories in a neighborhood:

American Restaurant	Japanese Restaurant	Turkish Restaurant
Asian Restaurant	Kebab Restaurant	Vegetarian / Vegan Restaurant
Belgian Restaurant	Mexican Restaurant	Vietnamese Restaurant
Chinese Restaurant	Middle Eastern Restaurant	Breakfast Spot
Comfort Food Restaurant	Modern European Restaurant	Burger Joint
Dumpling Restaurant	Restaurant	Diner
Eastern European Restaurant	Scandinavian Restaurant	Food
Fast Food Restaurant	Seafood Restaurant	Food Court
French Restaurant	Spanish Restaurant	Noodle House
Indian Restaurant	Sushi Restaurant	Salad Place
Italian Restaurant	Theme Restaurant	Soup Place

Generally, it includes all venue categories with a word ”restaurant” in it with some additional categories, which designate an eating place as well, added (these are listed right after ”Vietnamese Restaurant”).
As next step to ensure same scale for components, standardized values of them are calculated in a following manner:

$$\frac{x_i - \mu}{\sigma}$$

where μ denotes mean value of given component and σ standard deviation of the component.
Then the Restaurant Attractiveness (RA) index will be calculated in a simple way:

$$\begin{aligned} \text{RA index} = {} & (\text{Population} + \text{Employed/Ha})_Z + \text{Average wage}_Z - \\ & (\#\text{ of Restaurants}/(\text{Population}+\text{Employed}))_Z, \end{aligned}$$

where subscript Z denotes standardized value.
The higher the index the more attractive is the neighborhood for a new restaurant. Clearly, the higher the relative density of people going through the neighborhood during a day, the higher likelihood that they will need a

place to eat - either for lunch (gauged by number of employed in the neighborhood) or for dinner (gauged by overall population of the neighborhood). Also the higher average wage the higher propensity to go dinning or visit more exclusive restaurant, In Latvia it is not very popular to have dinners outside home, only the more affluent ones dine in restaurants. Lastly, relative lack of restaurants in the area could be a signal for opportunities to open a new one. Weights for each of the variables of RA index are not added to maintain objectivity of the index.
All neighborhoods will be ranked in accordance with the index, and the ones with highest value will be considered for new restaurant. However, qualitative analysis will have to be applied as well, as the index does not grasp all of the components of an attractive neighborhood and does not model how given three variables interact with each other. In essence is a very simple index.

4 Results

After extracing data on venues in each neighborhood from Foursqaure I obtain 1184 venues. On average it is 20.4 venues per neighborhood, which seems to be sufficient. However if we have a look at the extract of first 10 neighborhoods from table summarizing neighborhoods by number of venues, we see that there is quite some number of neighborhoods with low number of venues.

Neighborhood	Number of Venues
Atgāzene	18
Avoti	70
Beberbeķi	9
Berģi	3
Bieriņi	12
Bišumuiža	9
Bolderāja	7
Brasa	53
Brekši	6
Bukulti	5

Figure 4: First 10 neighborhoods ordered alphabetically with respective number of venues.

There are 23 neighborhoods with venues count below 10 (including Kleisti without any venues). That's around one third of all neighborhoods. Already in the beginning we can see that for large number of neighborhoods a new restaurant won't be an option as they are underdeveloped.

Next let's cluster the neighborhoods based on their relative venue count. Firstly, we have to find optimal number of clusters. It's achieved by calculating sum of squared errors for each cluster count. When running through small number of clusters the sum of squares graph does not distinctively converge to infimum. Therefore I graph sum of squares of cluster count up to 40.

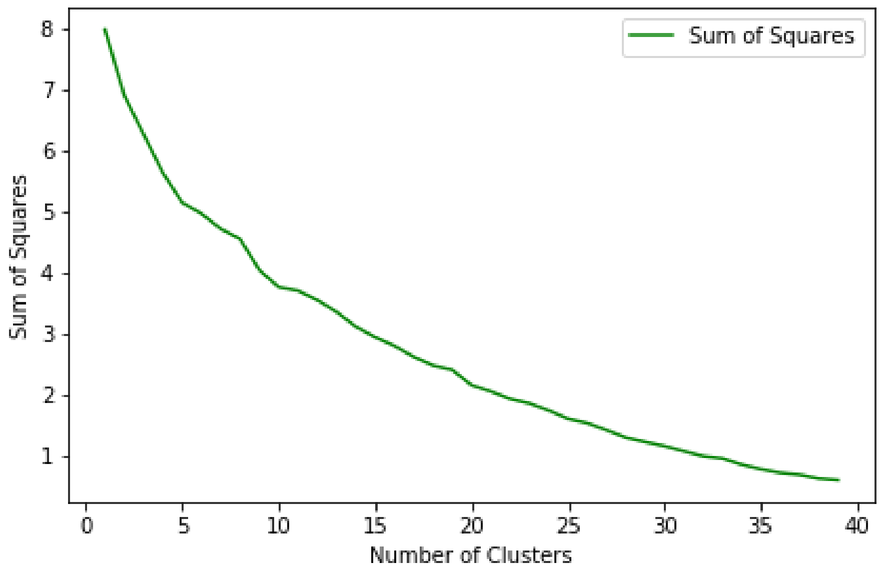

Figure 5: Sum of squares for k-means algorithm with cluster count from 1 to 40.

It still doesn't show a clear convergence. However, no to choose too high number of clusters, 10 clusters are chosen since at around 10 graphs slope to some extent looses it steepness. However, if the algorithm is run with 10 clusters, we see why it converged slowly. Clustering did not manage to actually divide the data groups of similar size. It rather cut off some marginal neighborhoods (with majority of clusters with only just one member) and left majority of neighborhoods in one large cluster of size 39. Algorithm is run also with 6 clusters, but results likewise are not very satisfactory (see figure below).

6 Cluster case

Cluster number	0	1	2	3	4	5
Neighborhood count	4	8	42	1	1	1

10 Cluster case

Cluster number	0	1	2	3	4	5	6	7	8	9
Neighborhood count	1	1	4	3	1	39	1	1	5	1

Figure 6: Neighborhood count by cluster in case of 6 and 10 clusters.

Let's have a look at the most common venues by neighborhood.

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Atgāzene	0	Bus Stop	Bakery	Diner	Fast Food Restaurant	Motorcycle Shop	Bed & Breakfast	Bus Station	Liquor Store	Light Rail Station	Café
1	Avoti	7	Café	Coffee Shop	Restaurant	Bar	Dance Studio	Spa	Burger Joint	Salon / Barbershop	Park	Diner
2	Beberbeķi	0	Diner	Rental Car Location	Bus Stop	Restaurant	Farmers Market	BBQ Joint	Café	Auto Garage	Bus Station	Dog Run
3	Berģi	8	Lake	Beach	Yoga Studio	Discount Store	Farmers Market	Farm	Fabric Shop	Exhibit	Event Space	Electronics Store
4	Bieriņi	0	Bakery	Bus Stop	Miscellaneous Shop	Trail	Sculpture Garden	Eastern European Restaurant	Basketball Court	Track	Bus Station	Dance Studio
5	Bišumuiža	0	Bus Stop	Burger Joint	Auto Garage	Train Station	Forest	Bus Station	Convenience Store	Harbor / Marina	Skate Park	Event Space
6	Bolderāja	9	Food & Drink Shop	Bus Stop	Lake	Deli / Bodega	Park	Grocery Store	Electronics Store	Event Space	Eastern European Restaurant	Disc Golf
7	Brasa	7	Sporting Goods Shop	Arcade	Pizza Place	Diner	Park	Bistro	Gym / Fitness Center	Pool	College Gym	Flower Shop
8	Brekši	9	Grocery Store	Bus Stop	Park	Beach	Supermarket	Dumpling Restaurant	Dive Bar	Doctor's Office	Dog Run	Donut Shop
9	Bukulti	4	Bus Stop	Comfort Food Restaurant	Yoga Studio	Discount Store	Farmers Market	Farm	Fabric Shop	Exhibit	Event Space	Electronics Store

Figure 7: First 10 neighborhoods ordered alphabetically with respective top 10 most common venues.

We can observe that apart from public transit related venues there are not many repeating categories even in neighborhoods from the same cluster. Similarly it also for neighborhoods beyond the ten exhibited here. The neighborhoods in Riga are very diverse and algorithm can't find enough similarities to produce meaningful clusters.

Since clustering apparently does not produce any value, I refrain from further exploring and analysing these results.

As a next step, I calculate RA indexes. As an intermediary step number of restaurants in each neighborhood is computed. The histogram 8 shows that mostly there are only 1 or 2 restaurants in neighborhoods. There are many small, underdeveloped neighborhoods. The leaders with 28 and 25 restaurants are, of course, the central neighborhoods - Center, Old Town, respectively. As asserted in Introduction probably there are not many opportunities for new restaurants in these two neighborhoods.

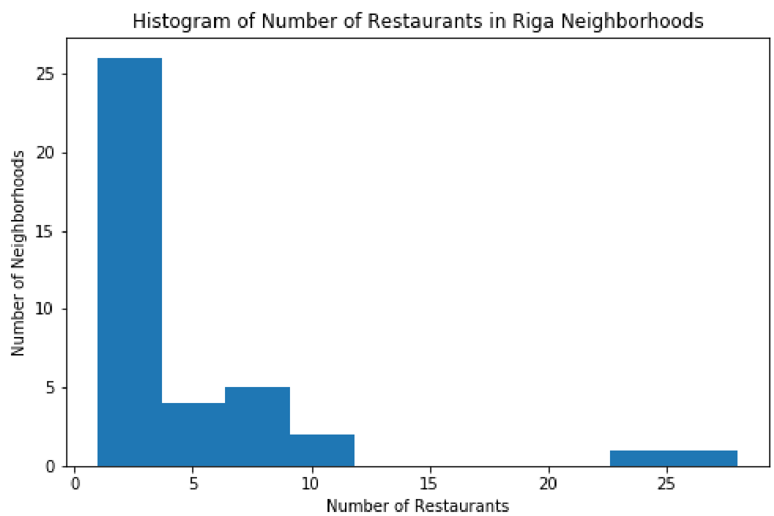

Figure 8: Histogram of number of restaurants in Riga neighborhoods.

Whit respect to RA index, results of top 10 neighborhoods are demonstrated in the table below (calculated standardized values of each component of the index are also shown).

	Code	Neighborhood	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010034	Centrs	373.2	1521	30557	73642.0	28.0	104199.0	279.20	0.27	1.40	3.54	0.03	4.98
1	0010057	Vecpilsēta	94.4	1805	1941	22810.0	23.0	24751.0	262.19	0.93	2.49	3.28	-1.06	4.70
2	0010047	Skanste	214.8	2316	1732	9944.0	6.0	11676.0	54.36	0.51	4.44	0.01	-0.37	4.08
3	0010051	Brasa	174.1	1351	13067	17190.0	9.0	30257.0	173.79	0.30	0.76	1.89	-0.01	2.63
4	0010050	Avoti	181.5	1139	18313	13896.0	11.0	32209.0	177.46	0.34	-0.05	1.95	-0.09	1.81
5	0010025	Purvciems	501.7	1147	55579	16527.0	4.0	72106.0	143.72	0.06	-0.02	1.42	0.39	1.78
6	0010053	Grīziņkalns	151.7	1189	12312	11016.0	6.0	23328.0	153.78	0.26	0.14	1.57	0.05	1.77
7	0010026	Teika	468.2	1278	28720	25159.0	6.0	53879.0	115.08	0.11	0.48	0.97	0.29	1.74
8	0010040	Pļavnieki	298.5	1028	42048	8301.0	3.0	50349.0	168.67	0.06	-0.48	1.81	0.38	1.71
9	0010044	Iļģuciems	244.2	1031	22004	7472.0	0.0	29476.0	120.70	0.00	-0.46	1.06	0.48	1.07

Figure 9: First 10 neighborhoods ordered by RA index. Note, standardized restaurant density $Rest_Z$ is reveresed with respect to restaurant density, i.e., the higher density the lower standardized score.

The leaders as suspected are Centrs and Old Town. Centrs in all three components is above average having positive Z-score (standardized components) - even it's restaurant density per capita is lower than most of the neighborhoods. Old Town probably would be even higher if tourists would also be accounted for. The next neighborhood almost at same level of RA (above 4) as the first two is Skanste. It scores the highest in average wage component - with 2316 EUR/month. However, currently most of the potential restaurant goers would be employed people as local population is very small. Even number of employed is barely above average. On top that, number of restaurants also seems high. Another attractive neighborhood is Brasa. It is quite small neighborhood with relatively high number of people it is housing. Similarly as in Skanste its advantage is number of employees working in the neighborhood. The number of permanent inhabitants and their wealth also is decently high. Purvciems on other hand has high local population, effect of which is to some extent decreased by size of Purvciems (it is ranked 6th in RA list). Though number of people employed is also one of the highest among all neighborhoods. And there are just a little bit more than couple of restaurants. Very similar is Pļavnieki neighborhood, which maybe has less employed people but is smaller with higher local population density. Also worth mentioning is Teika which has the 3rd highest number of

people employed. It is emerging as one of the most attractive neighborhood both for potential residents and for businesses. As there are located many IT related businesses and other enterprises employing highly skilled workforce, purchasing power for business lunches there would be high (which, though, is not directly observable in the table).
Now I will develop alternative neighborhood clustering method to one based on frequency of venue categories. This method will be based on basic descriptive statistics size, population, number of people employed, average wage and number of restaurants. In this case I also determine optimal number of clusters by graphing relation between number of clusters and its respective sum of sqaured errors.

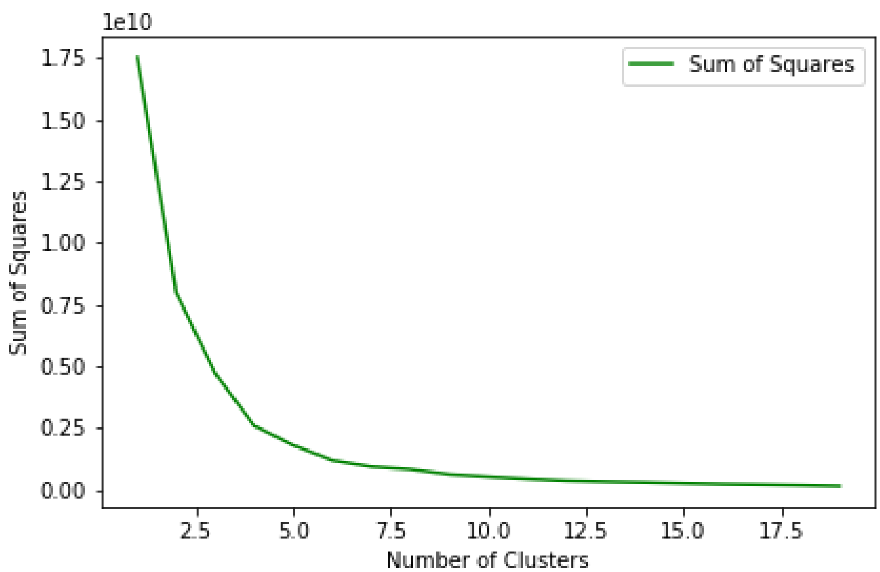

Figure 10: Sum of squares for k-means algorithm with alternative clustering method for cluster count from 1 to 20.

By using the "elbow method" I find that the optimal cluster count is 6. Therefore clustering is executed by dividing neighborhoods in 6 parts. The result is 6 clusters with respective sizes 7, 4, 1, 12, 30, 4 (the list of neighborhoods in each cluster are included in Appendix). The next figure shows profile of each cluster.

Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	424.0	1384.0	4575.0	13884.0	7.0	18458.0	75.0	0.35	0.88	0.33	-0.10	1.12
1	571.0	1134.0	27634.0	20635.0	6.0	48268.0	88.0	0.13	-0.07	0.54	0.27	0.73
2	373.0	1521.0	30557.0	73642.0	28.0	104199.0	279.0	0.27	1.40	3.54	0.03	4.98
3	510.0	1096.0	16835.0	10148.0	3.0	26983.0	85.0	0.11	-0.21	0.49	0.30	0.58
4	530.0	1123.0	2202.0	1476.0	1.0	3677.0	15.0	0.40	-0.11	-0.61	-0.19	-0.91
5	555.0	1064.0	47089.0	13097.0	2.0	60186.0	123.0	0.04	-0.34	1.09	0.41	1.16

Figure 11: Average values of each feature by cluster.

First cluster has neighborhoods small in population, but rather larger number of employed. Also it its residents are the wealthiest. I name this cluster "Business & Industrial centers". Second cluster has neighborhoods large in both population and number of employed. They also are developed in terms of number of restaurants. I call this cluster "Local social centers". Next cluster (3rd) has sole neighborhood - Centrs. It justifies its name - it's heart of Riga. It is large in almost all categories - population, number employed, average wage and number of restaurants. I name it simply "Center". The fourth cluster is no leader in any of the categories. It's neighborhoods have medium sized population, medium number of employed people, not very high count of restaurants and also average in terms of size. I will name it "Average neighborhoods". The next cluster consists of sparsely populated neighborhoods with many of them having 0 or 1 restaurant. There is not much activity going on in them. Furthermore, they are located in the outskirts of Riga. Cluster will be called "Small periphery". The last cluster is the most interesting for new restaurant entrepreneurs as it has the highest average RA index. Though the first cluster is close in the second place, but it has more developed restaurant scene with having on average 7 restaurants per neighborhood. The last cluster is by far largest in population and it has low average income level. It is where workforce spends evenings and nights. I call it "Sleeping neighborhoods".
All neighborhoods are mapped by their cluster affiliation in figure 12.

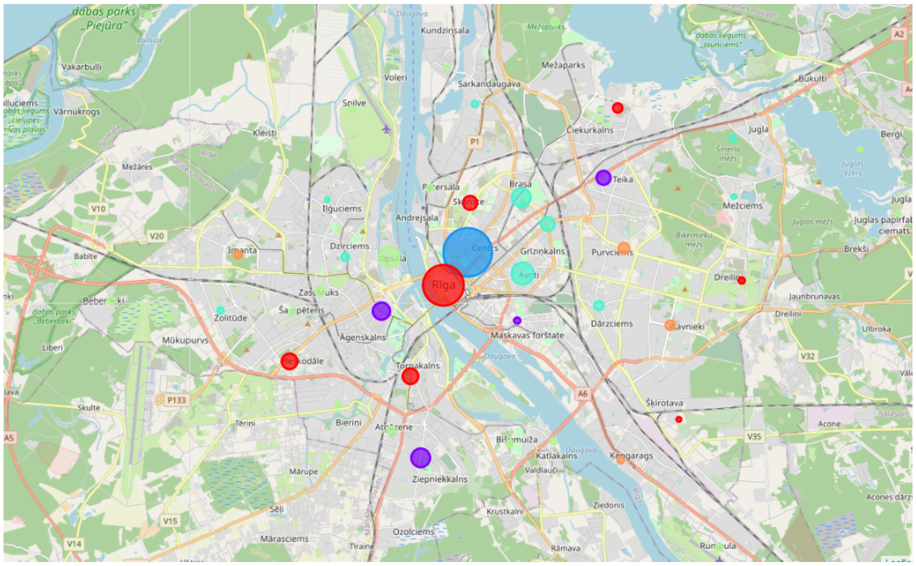

Figure 12: All neighborhoods mapped. Colour of circle determines affiliation to particular cluster. Size of circle determines number of restaurants in neighborhood. Upper size of Riga is cut off as it does not include important neighborhoods and this way map is more readable.

5 Discussion

Although Skanste's district seems to be very appealing for new restaurant given high purchasing power of its inhabitants. However, I don't believe to be suitable since there is not enough turnaround of people and restaurant scene apparently is already developed enough for its size. One the best candidates is Purvciems. It houses the most inhabitants among all neighborhoods. Even though it is large and accordingly it would be difficult to cover its whole population, I believe there could be interest for restaurant in its center. It would be in a walking distance or several minutes ride distance away for most of the folks. Though it should be affordable as wage level of Purvciems' people is not among the highest. Definitely there would be space for some fast food restaurant like a kebab place which currently are popular. Second option similar to Purvciems is Pļavnieki. It is less populous but with higher density. There also would be place for some fast food restaurant (even knowing that it already has two fast food burger places). Additionally, the rest of neighborhoods from "Sleeping nieghborhoods" cluster could considered as their restaurant environment is even less developed with, respectively, 1 and 2 restaurants.

For lunch oriented restaurant better location would be Teika. I favor it over similar Brasa neighborhood because of higher people turnaround and lower restaurant count. It has recently built new business\residential district "Jaunā Teikā" and in territory of old factories lots of new IT businesses are developing. And could have demand also in the evening as Teika is home for increasing number of young professionals with income well above average. Nevertheless, location and restaurant's business plan should be carefully evaluated as there are already restaurants. As demonstrated by map 13, two of them are in "Jaunā Teika" district. Perhaps suitable location would around IT business close to Gaisa Tilts or high quality fast food in "Jaunā Teika" to offer diversity.

Figure 13: Map of restaurants in Teika neighborhood (marked with blue circles).

6 Conclusions

In this project I analysed opportunities to open a new restaurant in one of the neighborhoods of Riga apart from the central ones (Centrs and Old Town). Main input for data was retrieved by Foursqaure API - it was information on what venues are located in each neighborhood. However, since diversity of venues in neighborhoods was high and thus proved to yield no valuable results, the analysis was refocused specifically to number of restaurants in neighborhoods. It was accompanied with data on income level, population, number of employed and size of neighborhoods. Eventually and Restaurant attractiveness (RA) index was developed which was composed of 3 components - density of neighborhood's population and people employed summed, average wage level and number of restaurants per 1000 of people living and people working in neighborhood. Standardized versions of these 3 components was used in the index.

Results of RA index showed that the most viable neighborhoods for restaurant could be Purvciems, Pļavnieki, Teika. For Purvciems and Pļavnieki a more affordable fast food restaurant is suggested. However, in Teika a restaurant oriented on business lunches and active evening life as well. Additionally, clustering of neighborhoods was executed by using k-means algorithm. All neighborhoods were split in 6 clusters. The 6th cluster was the most suitable for new food ventures. In addition to Purvciems and Pļavnieki it included Ķengarags and Imanta. In general it suggested to enter neighborhoods which hosts working class and middle class citizens.
One of the problems I came across was that Foursqaure data in some cases did show full picture of restaurant offerings - some of the information was outdated, clearly it did not include all restaurants (judged by my experience of visiting the neighborhoods), some of the restaurants had wrong locations or did not exist at all. Even more the results retrieved could differ next day when Foursqaure is again accessed. Hence, it casts quite some doubt on the analysis done in this projected. For better research some another mapping services should be used, for instance, Google Maps, although it is also in question if it includes all restaurants.
Another area of improvement is RA index. Current version is too simplified and high RA index might give misleading information. It would be recommended to instead of standardizing values versus mean value standardize them versus some qualitatively determined threshold and reward/penalize neighborhoods further above or further below mean, i.e., for example, in case of average wage introduce level of wealth beyond which there would be clearly higher demand for restaurants' services. As being above mean income level does not necessarily translate to higher propensity for eating outside.
Furthermore, it should be noted - this is a simple analysis which gives very broad guidance in which direction to look for a spot of new restaurant. Proper reals estate availability should also be evaluated, especially this concerns large suburbs in outskirts of Riga (like Purvciems and Pļavnieki). It should also be

suggested to do some proper marketing analysis - surveying neighborhoods' inhabitants to establish their tastes and even if there is need for new eating places. Generally, much more variables should be introduced for proper assessment.

References

[1] Apkaimes.lv, Riga Neighborhoods. https://apkaimes.lv/statistika/. Accessed: 2020-06-17.

[2] The Latvian Open data portal, population. https://data.gov.lv/dati/lv/dataset/iedzivotaju-skaits. Accessed: 2020-06-17.

[3] The Latvian Open data portal, economical indicators. https://data.gov.lv/dati/lv/dataset/vdvv-ekonomiskie-raditajii. Accessed: 2020-06-17.

[4] The Latvian Open data portal, monthly gross wage. https://data.gov.lv/dati/lv/dataset/menesa-bruto-darba-ienakumi. Accessed: 2020-06-17.

[5] Wikipedia, K-means clustering. https://en.wikipedia.org/wiki/K-means_clusteringi. Accessed: 2020-06-17.

A Clusters

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010038	Torņakalns	0	321.0	1064	6478	16792.0	7.0	23270.0	72.49	0.30	-0.34	0.30	-0.02	-0.06
1	0010022	Čiekurkalns	0	567.0	1125	7764	16236.0	3.0	24000.0	42.33	0.12	-0.11	-0.18	0.27	-0.01
2	0010036	Pleskodāle	0	348.0	1254	5079	9107.0	7.0	14186.0	40.76	0.49	0.39	-0.20	-0.34	-0.15
3	0010032	Dreiliņi	0	415.5	1152	6944	7573.0	1.0	14517.0	34.94	0.07	-0.00	-0.29	0.37	0.07
4	0010057	Vecpilsēta	0	94.4	1805	1941	22810.0	23.0	24751.0	262.19	0.93	2.49	3.28	-1.06	4.70
5	0010007	Šķirotava	0	1005.7	974	2086	14723.0	0.0	16809.0	16.71	0.00	-0.68	-0.58	0.48	-0.78
6	0010047	Skanste	0	214.8	2316	1732	9944.0	6.0	11676.0	54.36	0.51	4.44	0.01	-0.37	4.08

Figure 14: List of neighborhoods in cluster 1. "Business & Industrial centers".

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010005	Ziepniekkalns	1	594.2	1126	32108	14342.0	9.0	46450.0	78.17	0.19	-0.10	0.39	0.16	0.44
1	0010015	Maskavas forštate	1	759.4	960	24659	25599.0	1.0	50258.0	66.18	0.02	-0.74	0.20	0.45	-0.09
2	0010026	Teika	1	468.2	1278	28720	25159.0	6.0	53879.0	115.08	0.11	0.48	0.97	0.29	1.74
3	0010027	Āgenskalns	1	461.3	1170	25047	17439.0	8.0	42486.0	92.10	0.19	0.07	0.61	0.17	0.84

Figure 15: List of neighborhoods in cluster 2. "Local centers".

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010034	Centrs	2	373.2	1521	30557	73642.0	28.0	104199.0	279.2	0.27	1.4	3.54	0.03	4.98

Figure 16: List of neighborhoods in cluster 3."Center".

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010002	Jugla	3	1409.9	1080	24011	7715.0	1.0	31726.0	22.50	0.03	-0.28	-0.49	0.43	-0.34
1	0010044	Iļģuciems	3	244.2	1031	22004	7472.0	0.0	29476.0	120.70	0.00	-0.46	1.06	0.48	1.07
2	0010018	Vecmīlgrāvis	3	607.3	921	20606	6120.0	2.0	26726.0	44.01	0.07	-0.88	-0.15	0.36	-0.68
3	0010050	Avoti	3	181.5	1139	18313	13896.0	11.0	32209.0	177.46	0.34	-0.05	1.95	-0.09	1.81
4	0010041	Zolitūde	3	288.8	1173	17527	4634.0	1.0	22161.0	76.73	0.05	0.08	0.36	0.40	0.85
5	0010028	Dārzciems	3	457.7	1073	18958	13833.0	3.0	32791.0	71.64	0.09	-0.30	0.28	0.33	0.31
6	0010014	Sarkandaugava	3	759.6	1028	16397	18803.0	1.0	35200.0	46.34	0.03	-0.48	-0.11	0.43	-0.16
7	0010013	Mežciems	3	766.7	1238	15023	9754.0	0.0	24777.0	32.32	0.00	0.33	-0.33	0.48	0.47
8	0010051	Brasa	3	174.1	1351	13067	17190.0	9.0	30257.0	173.79	0.30	0.76	1.89	-0.01	2.63
9	0010012	Bolderāja	3	832.9	879	12437	4109.0	0.0	16546.0	19.87	0.00	-1.04	-0.53	0.48	-1.09
10	0010053	Grīziņkalns	3	151.7	1189	12312	11016.0	6.0	23328.0	153.78	0.26	0.14	1.57	0.05	1.77
11	0010043	Dzirciems	3	244.4	1051	11365	7235.0	2.0	18600.0	76.10	0.11	-0.39	0.35	0.30	0.27

Figure 17: List of neighborhoods in cluster 4."Average neighborhoods".

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010031	Bieriņi	4	427.4	1343	8007	2901.0	1.0	10908.0	25.52	0.09	0.73	-0.44	0.33	0.61
1	0010006	Daugavgrīva	4	1014.7	916	8127	1644.0	0.0	9771.0	9.63	0.00	-0.90	-0.69	0.48	-1.11
2	0010056	Zasulauks	4	119.0	1211	6726	2849.0	3.0	9575.0	80.46	0.31	0.22	0.42	-0.04	0.61
3	0010037	Mīlgrāvis	4	322.5	1004	3865	1957.0	1.0	5822.0	18.05	0.17	-0.57	-0.56	0.19	-0.93
4	0010042	Pētersala-Andrejsala	4	277.3	1307	4979	6523.0	1.0	11502.0	41.48	0.09	0.59	-0.19	0.34	0.73
5	0010054	Šampēteris	4	136.6	1100	4999	3465.0	1.0	8464.0	61.96	0.12	-0.20	0.13	0.28	0.22
6	0010003	Mežaparks	4	1182.1	1468	4165	1602.0	2.0	5767.0	4.88	0.35	1.20	-0.76	-0.09	0.34
7	0010039	Vecdaugava	4	306.6	1048	1242	150.0	0.0	1392.0	4.54	0.00	-0.40	-0.77	0.48	-0.69
8	0010021	Berģi	4	570.6	1166	2956	822.0	0.0	3778.0	6.62	0.00	0.05	-0.74	0.48	-0.21
9	0010030	Dārziņi	4	434.8	1051	3586	694.0	0.0	4280.0	9.84	0.00	-0.39	-0.69	0.48	-0.59
10	0010009	Jaunciems	4	913.2	1030	2339	410.0	0.0	2749.0	3.01	0.00	-0.47	-0.79	0.48	-0.78
11	0010046	Bišumuiža	4	224.3	1313	2510	1033.0	1.0	3543.0	15.80	0.28	0.61	-0.59	0.01	0.03
12	0010048	Brekši	4	200.0	930	1591	353.0	0.0	1944.0	9.72	0.00	-0.85	-0.69	0.48	-1.06
13	0010045	Vecāķi	4	230.3	1402	1698	316.0	0.0	2014.0	8.75	0.00	0.95	-0.70	0.48	0.73
14	0010058	Atgāzene	4	74.5	1015	1649	3896.0	2.0	5545.0	74.43	0.36	-0.53	0.33	-0.12	-0.31
15	0010011	Mangaļsala	4	803.6	1092	1207	784.0	0.0	1991.0	2.48	0.00	-0.23	-0.80	0.48	-0.55
16	0010004	Trīsciems	4	1131.9	1111	1142	398.0	0.0	1540.0	1.36	0.00	-0.16	-0.82	0.48	-0.50
17	0010049	Ķīpsala	4	197.5	1552	928	4061.0	10.0	4989.0	25.26	2.00	1.52	-0.44	-2.84	-1.76

Figure 18: Part 1. List of neighborhoods in cluster 5.”Small periphery”.

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
18	0010017	Rumbula	4	697.8	878	889	3765.0	1.0	4654.0	6.67	0.21	-1.05	-0.73	0.12	-1.66
19	0010024	Bukulti	4	518.3	1116	626	595.0	1.0	1221.0	2.36	0.82	-0.14	-0.80	-0.88	-1.82
20	0010033	Suži	4	402.3	1472	460	62.0	0.0	522.0	1.30	0.00	1.22	-0.82	0.48	0.88
21	0010055	Beberbeķi	4	120.4	1245	462	206.0	2.0	668.0	5.55	2.99	0.35	-0.75	-4.48	-4.88
22	0010020	Kundziņsala	4	555.4	880	345	679.0	0.0	1024.0	1.84	0.00	-1.04	-0.81	0.48	-1.37
23	0010019	Voleri	4	531.5	732	255	502.0	0.0	757.0	1.42	0.00	-1.60	-0.82	0.48	-1.94
24	0010001	Kleisti	4	1873.0	1131	437	744.0	0.0	1181.0	0.63	0.00	-0.08	-0.83	0.48	-0.43
25	0010016	Buļļi	4	700.0	1665	295	39.0	1.0	334.0	0.48	2.99	1.95	-0.83	-4.48	-3.36
26	0010029	Mūkupurvs	4	451.5	1057	237	142.0	0.0	379.0	0.84	0.00	-0.37	-0.83	0.48	-0.71
27	0010052	Katlakalns	4	155.4	829	185	1668.0	1.0	1853.0	11.92	0.54	-1.23	-0.65	-0.41	-2.30
28	0010008	Spilve	4	957.6	654	69	1188.0	1.0	1257.0	1.31	0.80	-1.90	-0.82	-0.84	-3.56
29	0010035	Salas	4	362.8	986	70	818.0	0.0	888.0	2.45	0.00	-0.64	-0.80	0.48	-0.96

Figure 19: Part 2. List of neighborhoods in cluster 5.”Small periphery”.

	Code	Neighborhood	Cluster Labels	Size (Ha)	Income 2019	Population 2019	Employed 2019*	Restaurants	Pop + Employ 2019	Pop + Employ Density	Rest Density	Inc_Z	P+E_Z	-Rest_Z	RA_index
0	0010025	Purvciems	5	501.7	1147	55579	16527.0	4.0	72106.0	143.72	0.06	-0.02	1.42	0.39	1.78
1	0010023	Ķengarags	5	519.0	989	46541	12525.0	1.0	59066.0	113.81	0.02	-0.62	0.95	0.45	0.77
2	0010010	Imanta	5	900.3	1093	44189	15036.0	2.0	59225.0	65.78	0.03	-0.23	0.19	0.42	0.39
3	0010040	Pļavnieki	5	298.5	1028	42048	8301.0	3.0	50349.0	168.67	0.06	-0.48	1.81	0.38	1.71

Figure 20: List of neighborhoods in cluster 6.”Sleeping neighborhoods”.