

# A NEW RESTAURANT IN RIGA

---

IBM DATA SCIENCE PROJECT



# Riga<sup>1</sup> offers opportunities

---

- Around 1 million population in metro area
- Vibrant, socially active and outgoing city
- Expanding international community
- Increasing interest from EU and Eastern Europe tourists
- Growing economics with strongly increasing household income
- Inflating business menu and dinner prices & increasing demand for eating outside

A new restaurant is an attractive option

# Saturated restaurant scene in central neighborhoods

---

- There are already many restaurants in Centrs and Vecpilsēta (Old Town) neighborhoods
- Not clear if there is space for new restaurant ventures
- **Project aims to find suitable neighborhood in outskirts of Riga**
- Employ Foursquare venues data to analyse neighborhoods
- Food industry entrepreneurs and possibly Riga City Municipality would benefit

# Data

---

- List of neighborhoods from [apkaimes.lv](http://apkaimes.lv)
- Geographical coordinates from Google Maps API
- Retrieve venues data in 750 m radius from neighborhood center from by Foursquare API
- General information on neighborhoods from Central Statistical Bureau of Latvia (through Latvian Open Data portal) and [apkaimes.lv](http://apkaimes.lv):
  - Population
  - Size
  - Number of people employed
  - Average wage

# Descriptive statistics

---

- In total 58 neighborhoods
- High variety of neighborhoods with respect to Size, Population and Employed
- Lots of small neighborhoods (population <500 or employed <500)
- The most populous is Purvciems, employing the largest workforce is Centrs
- Both Purvciems and Centrs are large also in terms of people employed and population, respectively

	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

# Methodology

---

- Collect Foursquare venues data in 750 m radius from neighborhood center
- Based on frequency of venue categories in each neighborhood cluster the neighborhoods
- Apply k-means clustering algorithm
- Determine optimal number of clusters by graphing relationship between sum of squared errors and number of clusters (choose by applying «elbow method»)
- Develop Restaurant Attractiveness index (each components is standardized):

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

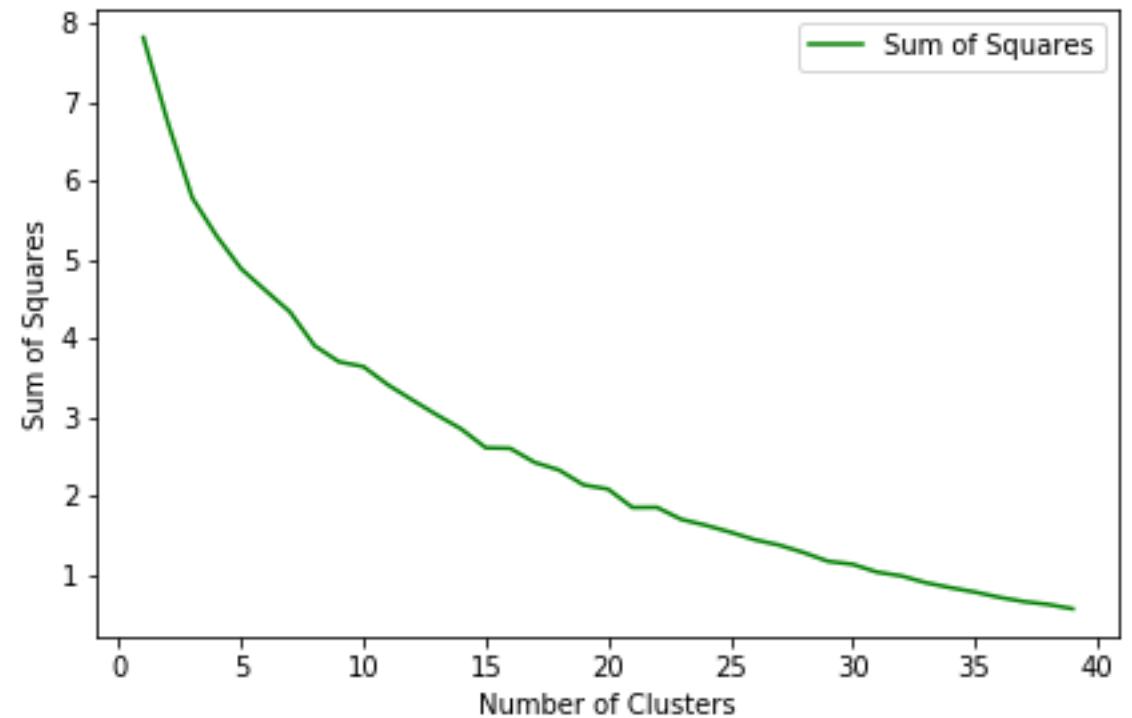
- The higher the index the more attractive is the neighborhood for a new restaurant

# Clustering based on venue frequency

---

- Very slow convergence of sum of squared errors
- Thus high number of clusters is chosen: 10
- Most of the produced clusters have only one member (similar result even with smaller cluster count)
- Neighborhoods are very diverse in terms of venue types, not many are repeated over various neighborhoods

**We drop this method as it's not yielding any value**



# RA index results

---

	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	Ļaivnieki	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



# Most attractive neighborhoods

---

**Skanste:** The most affluent neighborhood, strong purchasing power. Low people turnaround. Focus should be on employed people.

**Brasa:** Dense neighborhood. Also quite affluent. Quite many inhabitants and employees. No distinct advantages.

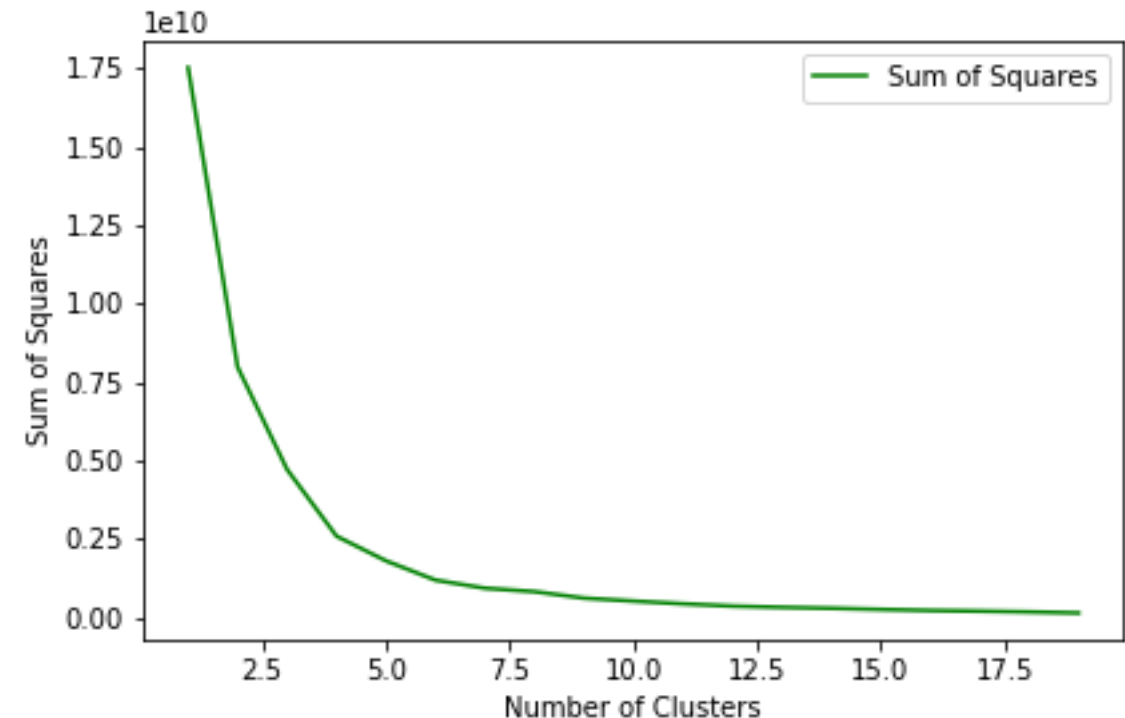
**Purvciems, Pļavnieki:** Large populations as well as number of employed. Not many restaurants, which could be very advantageous

**Teika:** Emerging neighborhood. Employs high skilled workforce. Good potential for restaurant oriented in lunch offerings

# Clustering based on general characteristics

---

- Size, income, population, number of employed and number of restaurants are selected as features
- 6 is the optimal number of clusters
- Cluster sizes: 7, 4, 1, 12, 30, 4



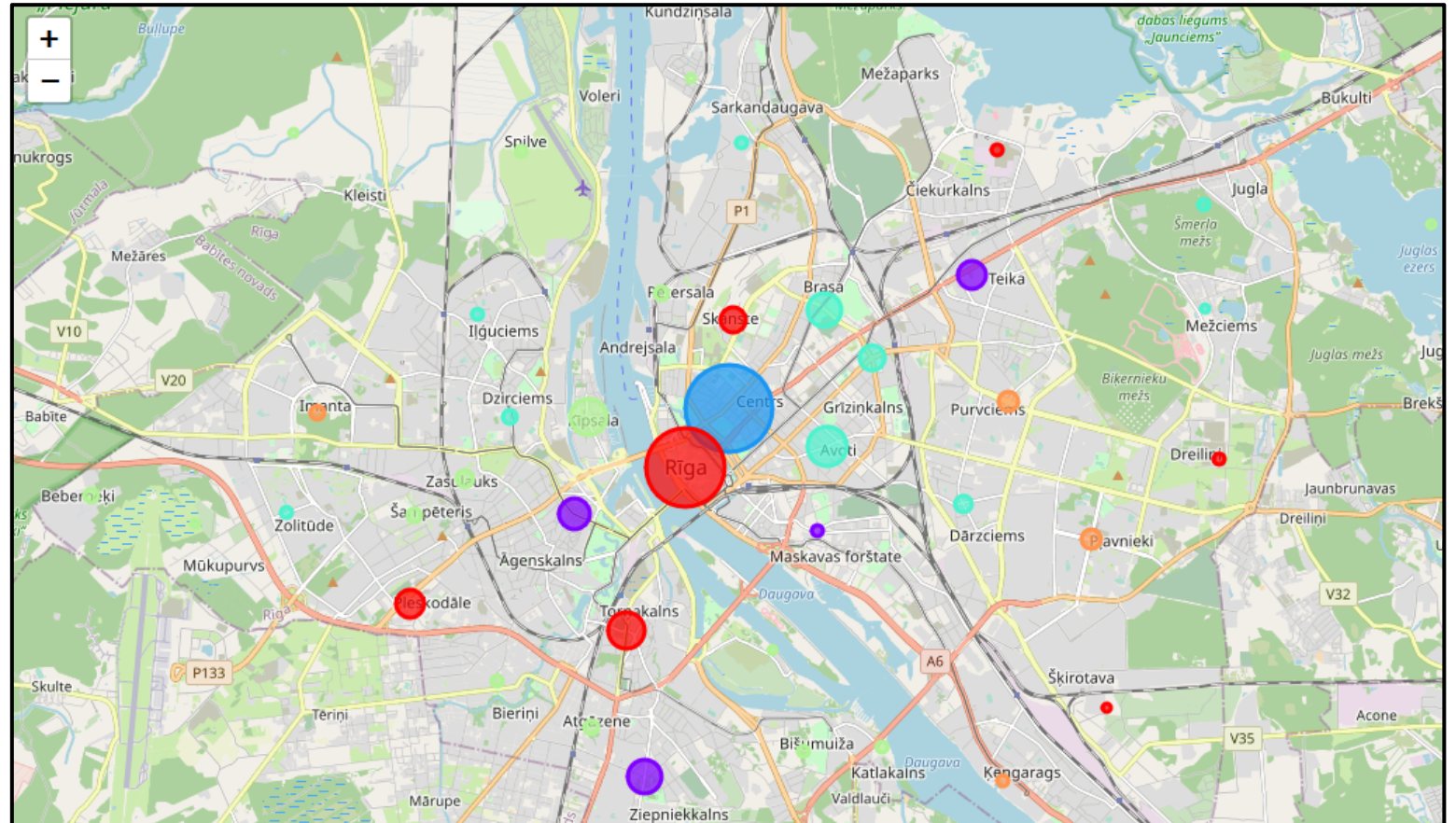
# Cluster profiles

---

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	Cluster Name
0	424	1384	4575	13884	7	18458	75	0.35	0.88	0.33	-0.1	1.12	Business & Industrial centers
1	571	1134	27634	20635	6	48268	88	0.13	-0.07	0.54	0.27	0.73	Local social centers
2	373	1521	30557	73642	28	104199	279	0.27	1.4	3.54	0.03	4.98	Center
3	510	1096	16835	10148	3	26983	85	0.11	-0.21	0.49	0.3	0.58	Average neighborhoods
4	530	1123	2202	1476	1	3677	15	0.4	-0.11	-0.61	-0.19	-0.91	Small periphery
5	555	1064	47089	13097	2	60186	123	0.04	-0.34	1.09	0.41	1.16	Sleeping neighborhoods

# Clusters mapped

- Circle color: affiliation to a particular cluster
- Circle size: restaurant count



# Conclusions

---

- Two candidates for restaurant location:
  - Purvciems (and also Pļavnieki): Affordable restaurant in the very center. Potentially a fast food like kebab restaurant.
  - Teika: Restaurant focusing on business lunches. Should be carefully evaluated as restaurant scene is already developed.
- Future improvements:
  - Foursquare data is not entirely reliable (can change day to day, doesn't include all restaurants, wrong locations or restaurants do not exist). Better map service could be chosen.
  - RA index to be simplified. Thresholds for each component could be introduced.
  - Proper market research and real estate analysis is needed