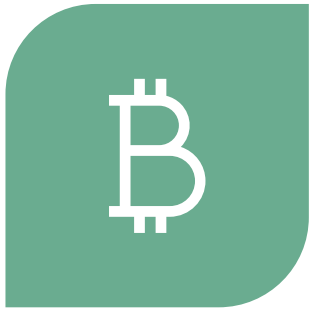
Saint Petersburg restaurant location

BY ARTEM F., 2020

Need to find the best high cuisine venue place with:

FOCUS ON CUSTOMERS WITH HIGH-INCOME

TOTAL POPULATION OF DISTRICTS



LESS POSSIBLE COMPETITIVE VENUES ALONG THE AREA

VENUE'S CUISINE IDEAS FOR DISTRICT

Data

- Foursquare data about venues (via Foursquare API)
- Administrative divisions of Saint Petersburg (parsed directly from Wikipedia with BeautifulSoup4 library)
- Statistics data of average income from Official Statistics Department of Russia in Saint Petersburg from jan to sept 2018
- Price of land from Rusland SP
- Price of land from “Restate” agency

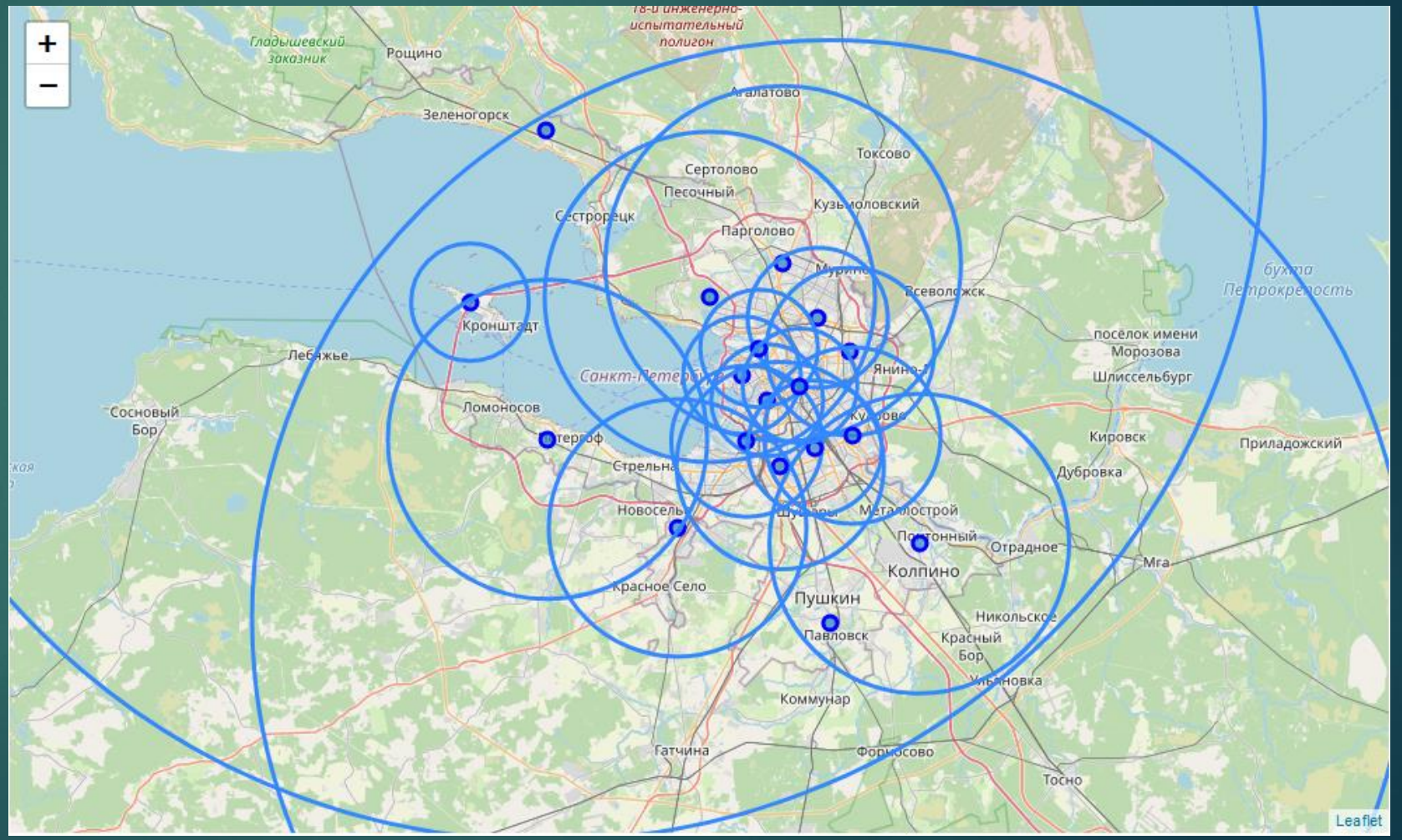
Basic data compilation

	DNumber	District	Population	Square	Density	Latitude	Longitude	Income	RentCost	FlatCostPerM	FCChange
0	1	Адмиралтейский	159795	13.82	11562.599609	59.916824	30.297542	81481	1408.8	136000.0	0.205
1	2	Василеостровский	207482	21.47	9663.799805	59.941425	30.248045	71982	4684.6	110000.0	-0.057
2	3	Выборгский	522746	115.52	4525.200195	60.050449	30.328696	64784	1121.1	117000.0	0.330
3	4	Калининский	529187	40.18	13170.400391	59.997684	30.396824	61920	1065.7	110000.0	0.085
4	5	Кировский	336157	47.46	7083.000000	59.876392	30.257603	60380	1055.7	109500.0	0.232

Collecting venues from Foursquare

- Foursquare API queries are limited
- We need to find best Radius for venues search

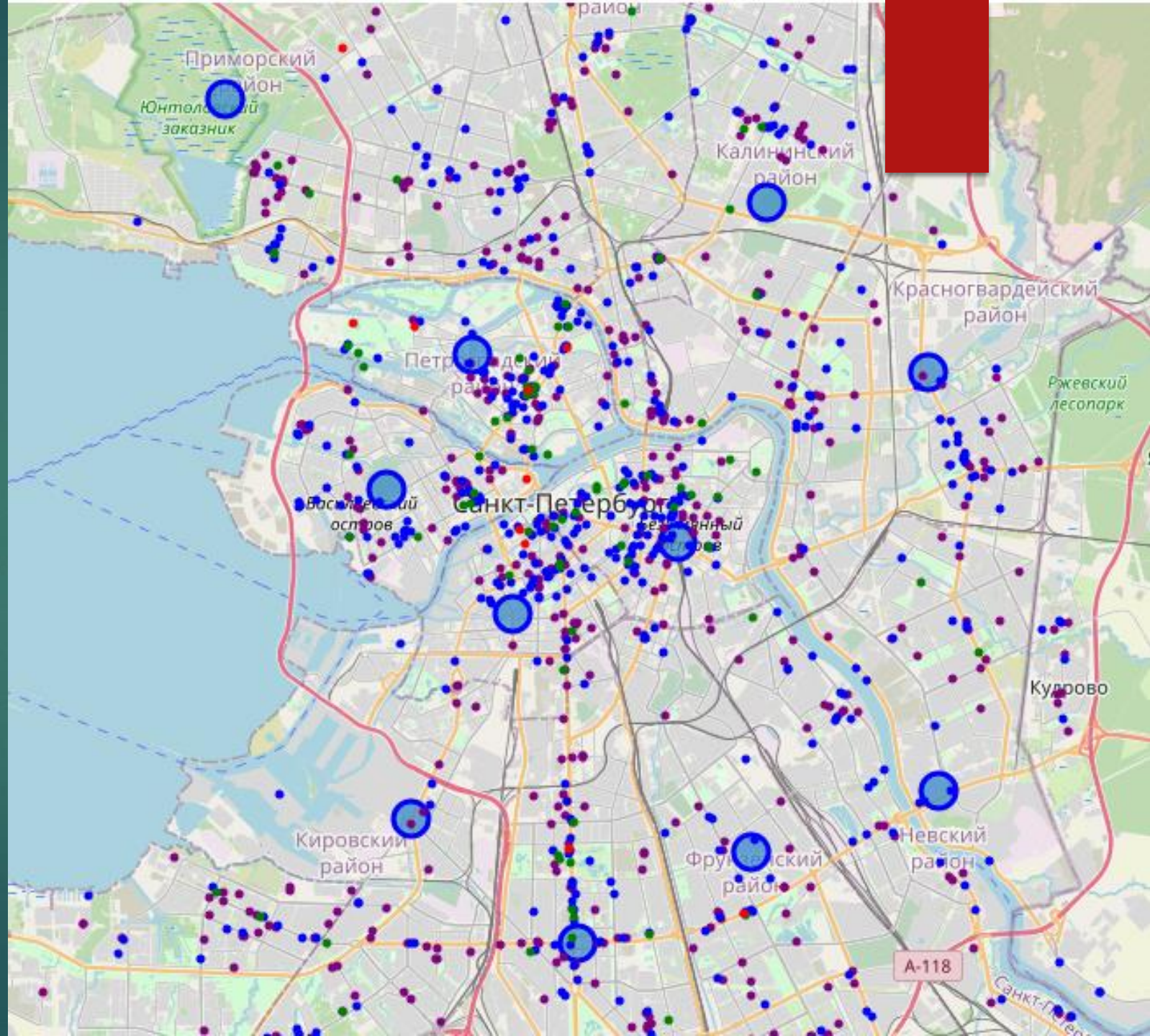
1. Building Voronoi diagram and closest vertice calculation
2. R-function (geometric)
3. R-function (with square)
4. R-mean (limited to 12k)



Collecting venues

Major steps:

- Collecting data for each Districts
- Calculating best Venue/District match according to the distance
- Removing duplicates



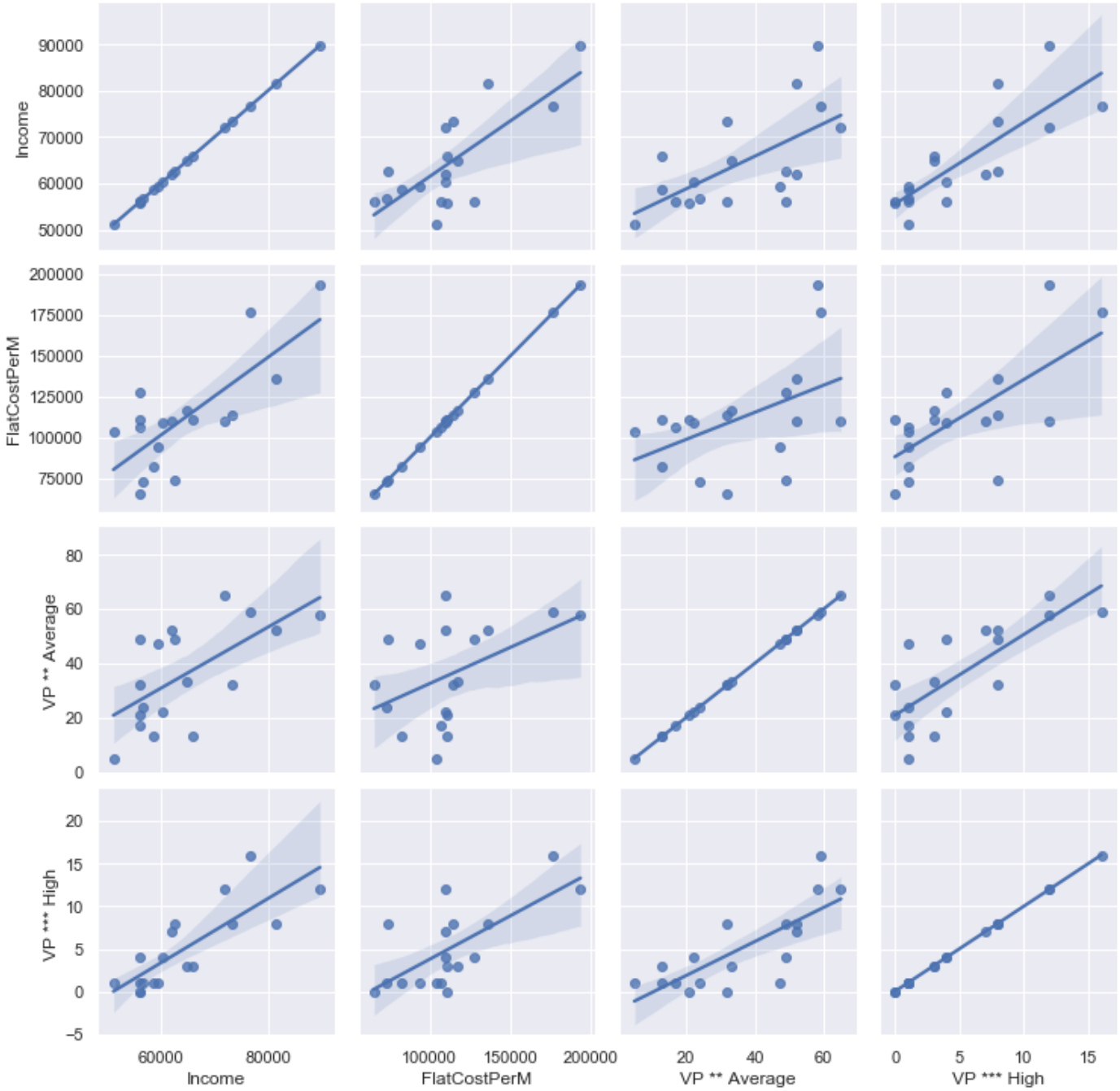
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
0	Адмиралтейский	Bakery	Café	Restaurant	Italian Restaurant	Caucasian Restaurant	Eastern European Restaurant	Asian Restaurant	Japanese Restaurant
1	Василеостровский	Bakery	Café	Restaurant	Italian Restaurant	Gastropub	Burger Joint	Eastern European Restaurant	Chinese Restaurant
2	Выборгский	Bakery	Café	Restaurant	Italian Restaurant	Japanese Restaurant	Pizza Place	Burger Joint	Sushi Restaurant
3	Калининский	Bakery	Café	Restaurant	Fast Food Restaurant	Pizza Place	Middle Eastern Restaurant	Sushi Restaurant	Gastropub
4	Кировский	Bakery	Sushi Restaurant	Pizza Place	Café	Fast Food Restaurant	Restaurant	Italian Restaurant	Eastern European Restaurant

Next step:

Building rating of venues for Districts

	Population	Square	Density	Income	RentCost	FlatCostPerM	VP * Low	VP ** Average	VP *** High	VP **** Top	Venues
Population	1.000000	-0.083035	0.346500	-0.198383	-0.102379	0.004171	0.048830	-0.189282	-0.264271	-0.293274	-0.106918
Square	-0.083035	1.000000	-0.708817	-0.310416	-0.411832	-0.593059	-0.413395	-0.300999	-0.306650	-0.105036	-0.392032
Density	0.346500	-0.708817	1.000000	0.404689	0.438765	0.604991	0.519788	0.577499	0.532066	0.158453	0.606921
Income	-0.198383	-0.310416	0.404689	1.000000	0.395108	0.754910	0.321808	0.629435	0.814706	0.631403	0.580733
RentCost	-0.102379	-0.411832	0.438765	0.395108	1.000000	0.378062	0.129384	0.443224	0.582437	0.177639	0.353782
FlatCostPerM	0.004171	-0.593059	0.604991	0.754910	0.378062	1.000000	0.320912	0.470270	0.693884	0.555862	0.487143
VP * Low	0.048830	-0.413395	0.519788	0.321808	0.129384	0.320912	1.000000	0.721517	0.375490	0.219964	0.898859
VP ** Average	-0.189282	-0.300999	0.577499	0.629435	0.443224	0.470270	0.721517	1.000000	0.771305	0.426360	0.947111
VP *** High	-0.264271	-0.306650	0.532066	0.814706	0.582437	0.693884	0.375490	0.771305	1.000000	0.407620	0.692115
VP **** Top	-0.293274	-0.105036	0.158453	0.631403	0.177639	0.555862	0.219964	0.426360	0.407620	1.000000	0.392552
Venues	-0.106918	-0.392032	0.606921	0.580733	0.353782	0.487143	0.898859	0.947111	0.692115	0.392552	1.000000

Feature selection from Correlation Matrix



Beast features correlation plots

Finding the best K for KMean clustering algorithm

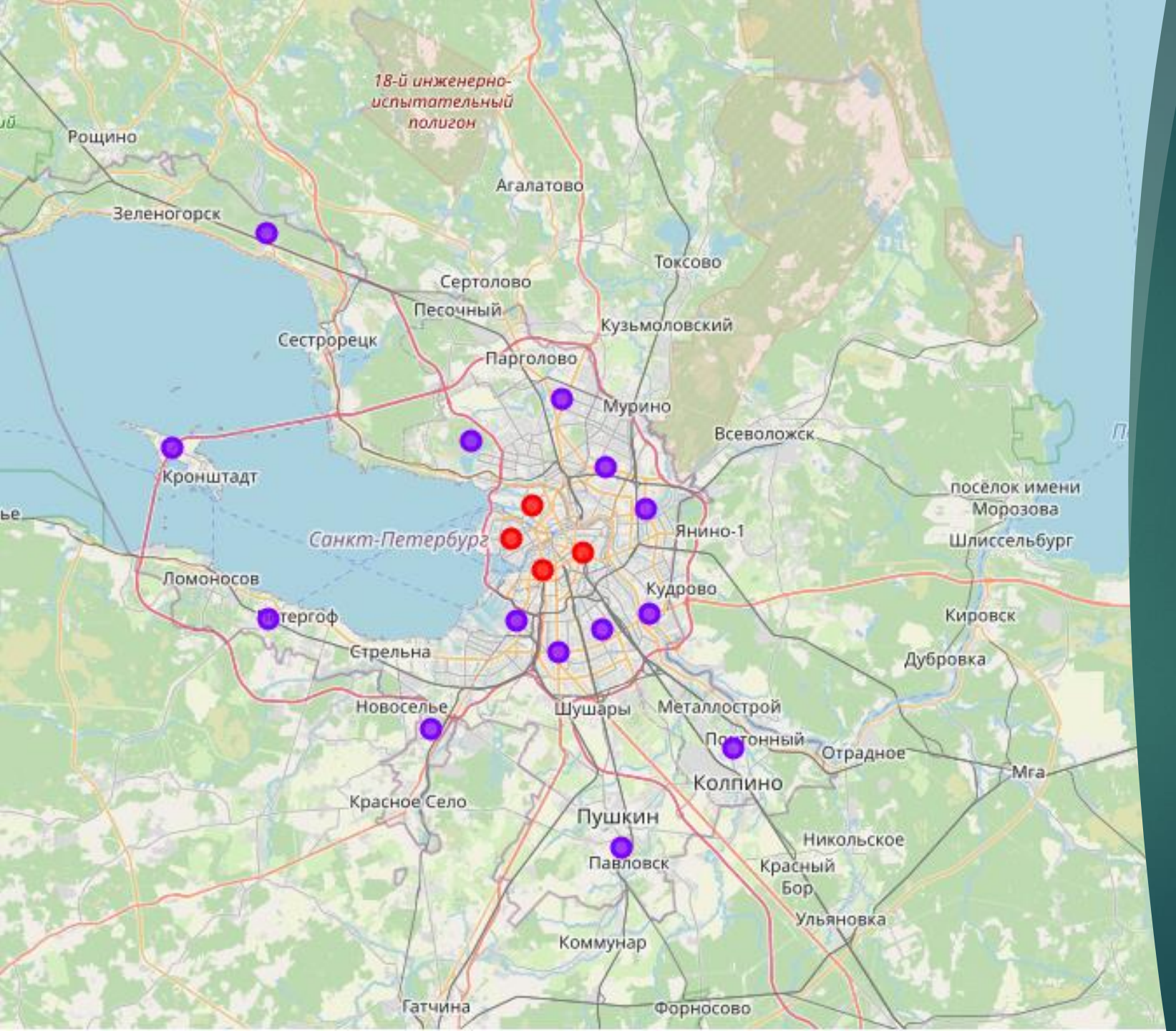
- Preprocess Features data for a model

Scores:

- Davies Bouldin
- Calinski Harabasz
- Silhouette score

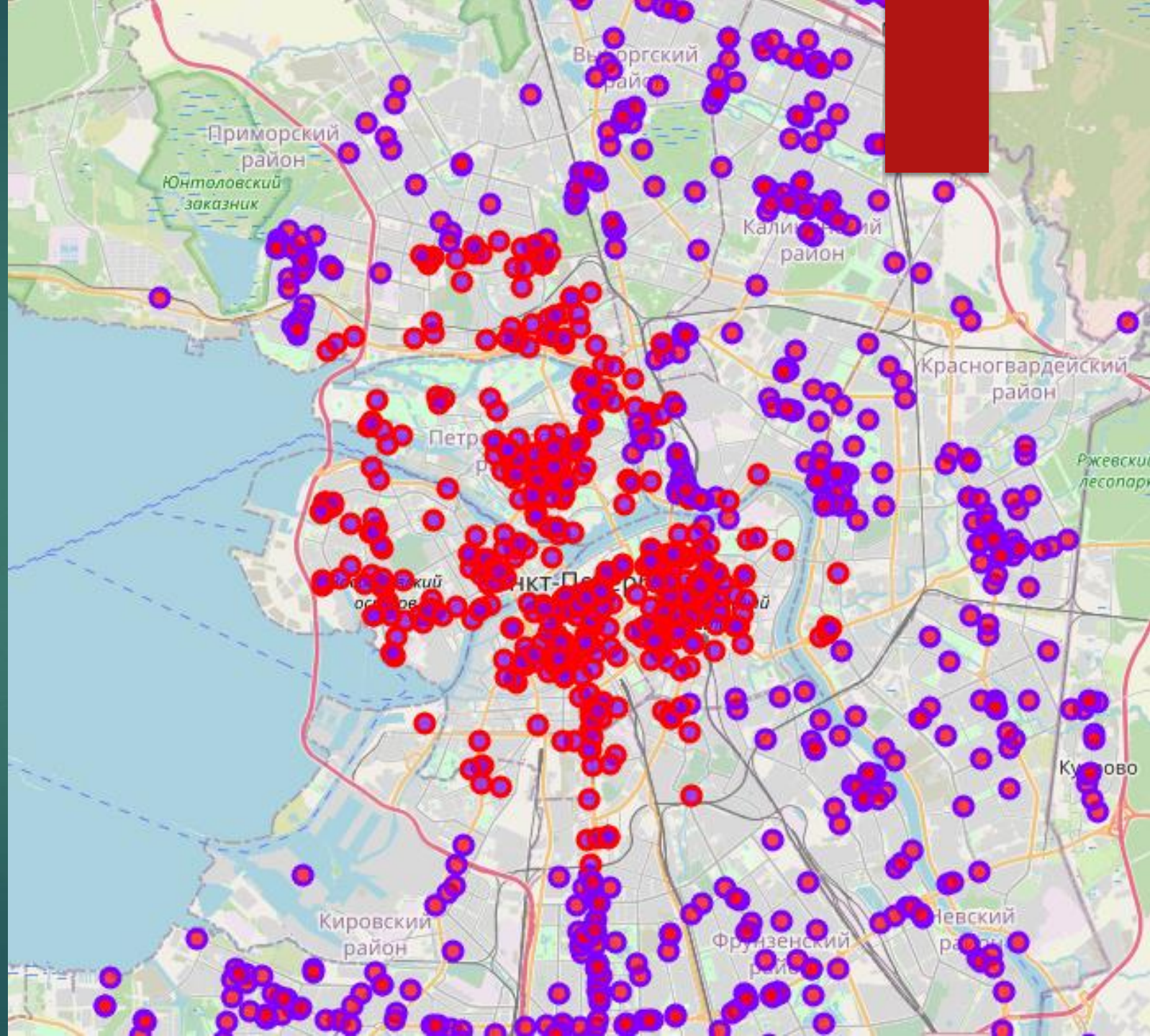
```
[2, 0.7133550239457775, 21.47863670352035, 0.49755437663593344],
[3, 0.9843437910370563, 19.14930250194232, 0.32235763576109444],
[4, 0.9553993509201002, 18.2286951214068, 0.31617697453492477],
[5, 0.8795019028475262, 16.279506854081824, 0.2762016356220343],
[6, 0.9080209563163125, 15.14332937954867, 0.2502131302128696],
[7, 0.7384042041938189, 14.607769995757808, 0.2592375380559034],
[8, 0.6916349103802848, 14.711234279662209, 0.23215197733952875],
[9, 0.6093259826320557, 14.989430273076788, 0.19239956959703897],
[10, 0.5570928700479303, 16.121410546456804, 0.20432169893880017]]
```

Best cluster value is 2



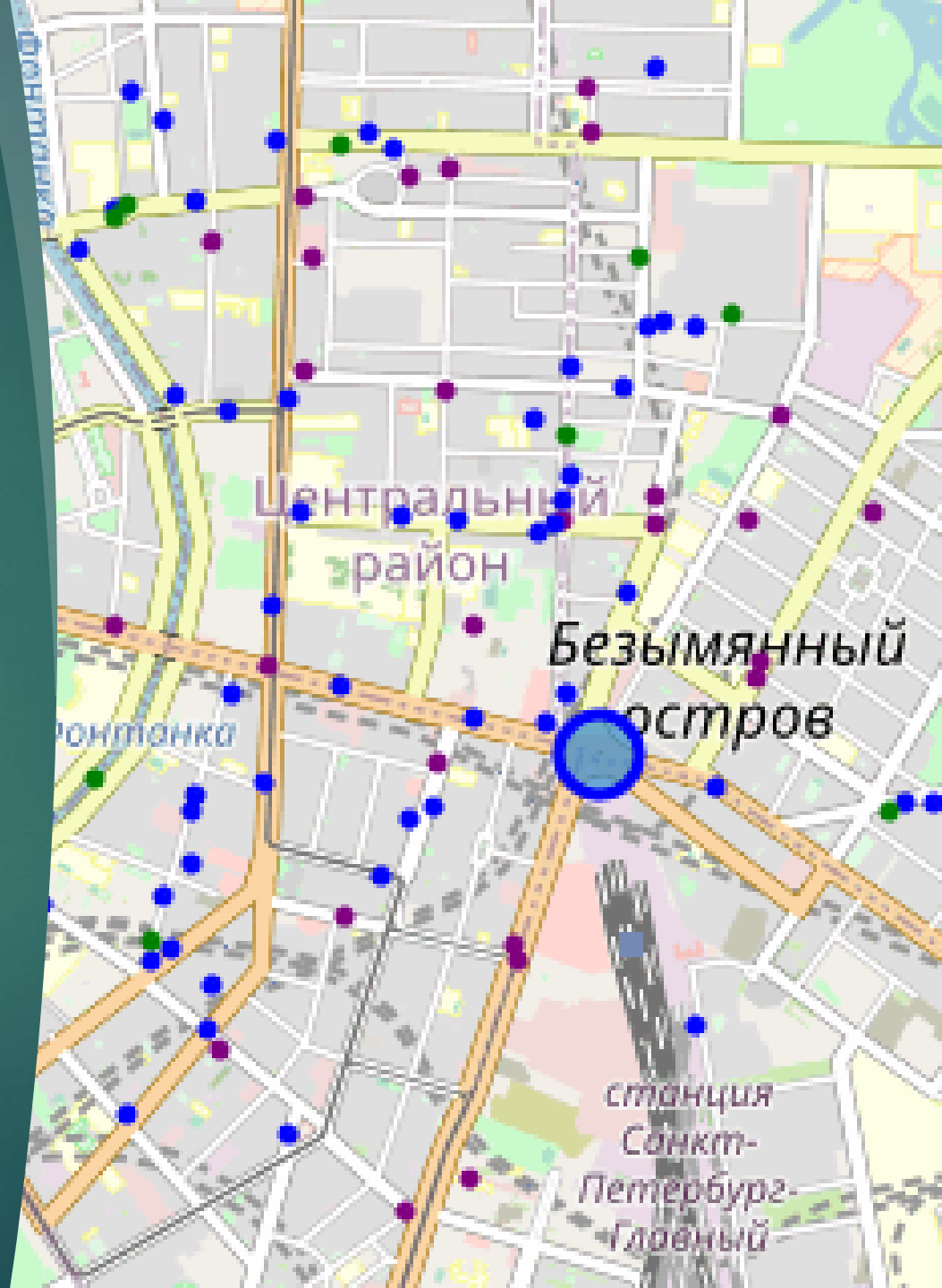
Building a clustering model for Districts (k = 2)

Browsing venues in different clusters



Best match: Centralniy dst.

- Avg income rate is very high
- No competitive venues along the area
- There are over 200k population



Cross-check venues for result District

District top10

```
['Bakery',
 'Restaurant',
 'Italian Restaurant',
 'Café',
 'Gastropub',
 'Burger Joint',
 'Asian Restaurant',
 'Seafood Restaurant',
 'Eastern European Restaurant',
 'Caucasian Restaurant']
```

City top10

```
['Bakery',
 'Café',
 'Restaurant',
 'Eastern European Restaurant',
 'Pizza Place',
 'Sushi Restaurant',
 'Italian Restaurant',
 'Fast Food Restaurant',
 'Middle Eastern Restaurant',
 'Gastropub']
```

The best match for a new venue (according to city top):

- Pizza Place
- Sushi Restaurant
- Middle Eastern Restaurant
- Fast Food Restaurant

Possible drawbacks

- Incomplete or wrong venue data in Foursquare
- Minor Radius inaccuracies
- Venue assignment to the District may be inaccurate due to direct distance calculation and the results may be improved by smart distance matrix
- Venue type selection uses data from all price type venues and may be too noisy as high cuisine restaurants share is low in comparison with all dataset
- We don’t have enough data to evaluate venue type as there is no targeted customers data

Conclusion

This project has approved our main point and made predictions for a central location in Saint-Petersburg. Also it described tastes of Saint-Petersburg citizens which can be used for future investment.