Capstone Project - The Battle of Neighborhoods (Week 2)

Ostroukhova Maria

July, 2020

Part 1

Introduction

Saint Petersburg is considered the Northern Capital of Russia. Its Historic Centre and Related Groups of Monuments constitute a UNESCO World Heritage Site, so it's also referred to as Russia's cultural capital. Saint Petersburg is home to the Hermitage, one of the largest art museums in the world, and the Lakhta Center, the tallest skyscraper in Europe.
Many foreign consulates, international corporations, banks and businesses have offices in Saint Petersburg.
Saint Petersburg is one of the most popular touristic city in Russia, famous in other countries.
I live there and would like to understand, what type of places are the most visited for both citizens and tourists. As we know, the city is situated in the north, it can be really important.

Problem Background

Saint Petersburg is a touristic city. A lot of people come to Saint Petersburg for different reasons. Structure data and analytics help them to understand districts and popular places there. A person can choose the most comfortable district for their spending time for a short or long period.
And knowing special places (districts) they can plan something easier than before.

Problem Description

We will find the answers to questions:

- What is the most popular kind of places in the city? In which districts they are?
- How there are grouped by districts?
- Analyzing clusters of districts.

Part 2

Data Section

Saint Petersburg districts from Wikipedia page:

https://ru.wikipedia.org/https://ru.wikipedia.org/wiki/Административно-территориальное_деление_Санкт-Петербурга

I interpreted column names from Russian to English for the understanding of what kind of information it is. And cleaned it from not useful info.

I created a table with districts, latitude and longitude, checking the coordinates for all districts on site:

https://www.openstreetmap.org/#map=10/59.9182/30.0655

Latitude and longitude table:

https://docs.google.com/spreadsheets/d/1LSKSKs-Nq8-VFAijPjW5eT2IwenfFSdird03SyoXVG4/edit?usp=sharing

The Foursquare location data to solve the problems:

- explore all venues of each category;
- explore the most common categories of venues;
- find TOP-3 categories and which the districts are contain its
- split all categories to group and analyze

Part 3

Methodology

I used Foursquare API to get the most common venues of given Districts of Saint Petersburg.
Using k-means clustering algorithm to complete splitting all categories to group and analyzing.
Finally, I used the Folium library to visualize the districts in Saint Petersburg and their emerging clusters.
The result of exploring the data will be structured, grouped and visualized in plots.
The audience is tourists, who is going to visit Saint Petersburg. I would like to help them to find more comfortable district for staying.

Part 4

Discusson:

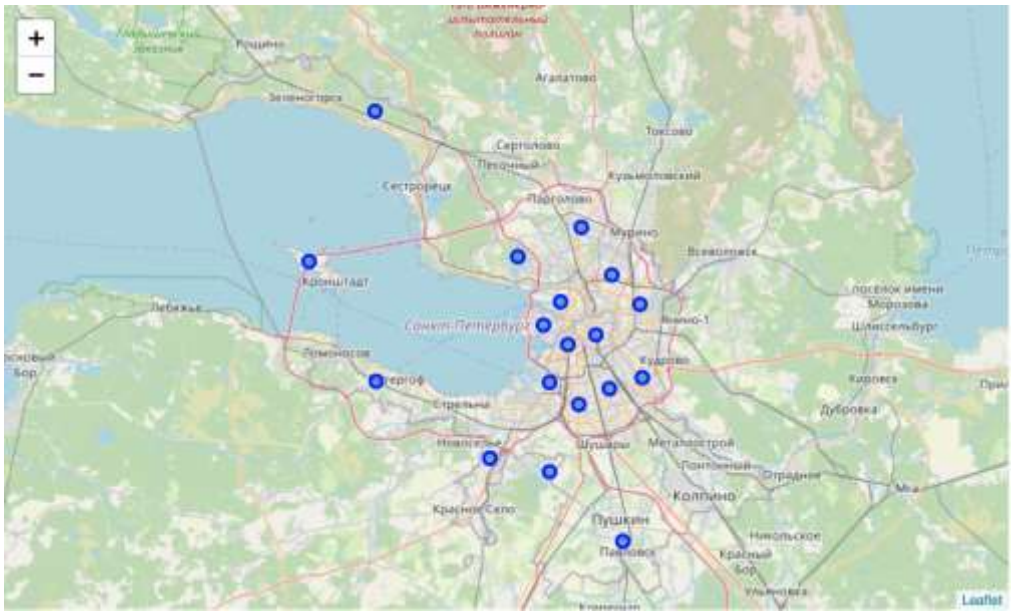
Begin from districts, from which the city is (17 districts):

	District	Population	Square kilometers / km²	Population density, people / km²
0	Адмиралтейский	159 795	1382	11 562,6
1	Василеостровский	207 482	2147	96638
2	Выборгский	522 746	11552	45252
3	Калининский	529 187	4018	13 170,4
4	Кировский	336 157	4746	7083
5	Колпинский	193 839	10225	18957
6	Красногвардейский	356 628	5635	63288
7	Красносельский	408 026	9049	45091
8	Кронштадтский	44 461	1953	22765
9	Курортный	79 067	26819	2948
10	Московский	352 172	7307	48197
11	Невский	536 137	6066	88384
12	Петроградский	128 075	1954	65545
13	Петродворцовый	143 823	10708	13431
14	Приморский	573 024	10990	5214
15	Пушкинский	226 336	24009	9427
16	Фрунзенский	386 537	3752	10 302,2
17	Центральный	214 672	1777	12 075

I cleaned the data, read information about latitude and longitude of districts

and I got preparing data for analyzing districts in Saint Petersburg.

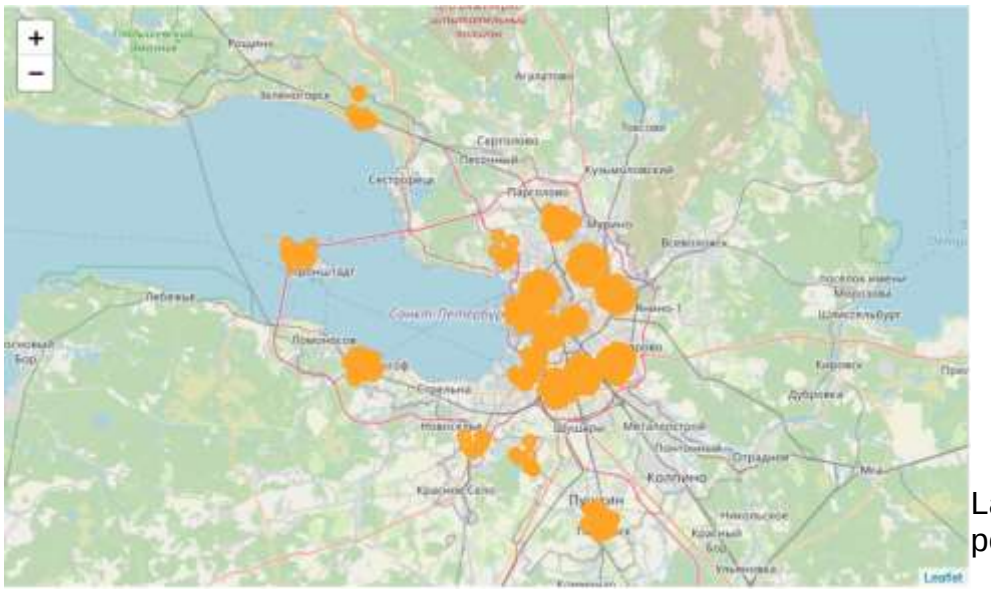
I setup the cridentials for Forsquare API.
Using Forsquare API I got a map with all 17th districts of the city:



In the next step I create a table with venues, their coordinates and categories.
For one of the central districts it looks like:

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1301	Центральный	59.930908	30.361817	come mode medical	59.941132	30.368190	Health & Beauty Service
1302	Центральный	59.930908	30.361817	St. Martin	59.938213	30.344483	Cocktail Bar
1303	Центральный	59.930908	30.361817	Vinissimo	59.935336	30.342868	Wine Shop
1304	Центральный	59.930908	30.361817	Subzero	59.927148	30.342841	Sushi Restaurant
1305	Центральный	59.930908	30.361817	Кофе на кухне	59.935999	30.343889	Coffee Shop

The number of unique categories is 243.
All venues returned by Foursquare on the next map:



Later I found the most popular categories of

venues as shown below. For example, there are Bakeries, Parks and Coffee Shops.

Venue Category	Venue
Bakery	67
Park	42
Coffee Shop	41
Gym / Fitness Center	34
Auto Workshop	34
...	...
Hunting Supply	1
Hot Dog Joint	1
Hockey Field	1
Hobby Shop	1
Zoo Exhibit	1

Analyzing the top 3 venues: Bakery

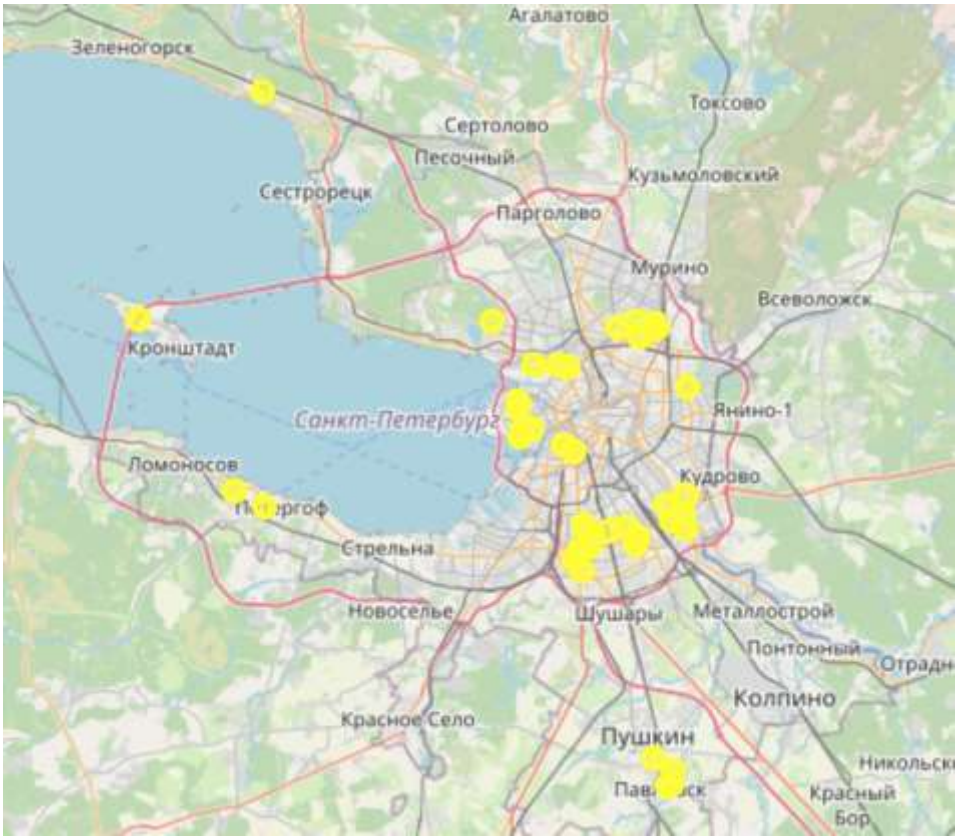
They are concentrated more towards the centre and two districts in the north and south of the city:

Выборгский 9
Центральный 8
Василеостровский 7
Московский 7
Адмиралтейский 6

Analyzing the top 3 venues: Park

They are concentrated not in the centre, in the most in districts in the south of the city and nearby city — in Pushkin town:

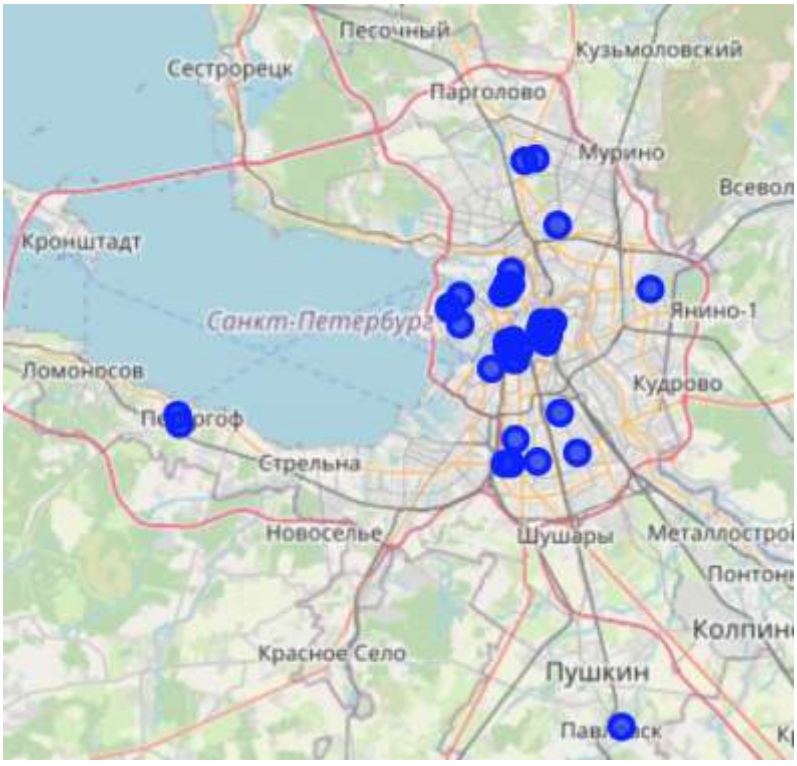
Калининский 7
Московский 6
Василеостровский 5
Невский 5
Пушкинский 4



Analyzing the top 3 venues: Coffee Shops

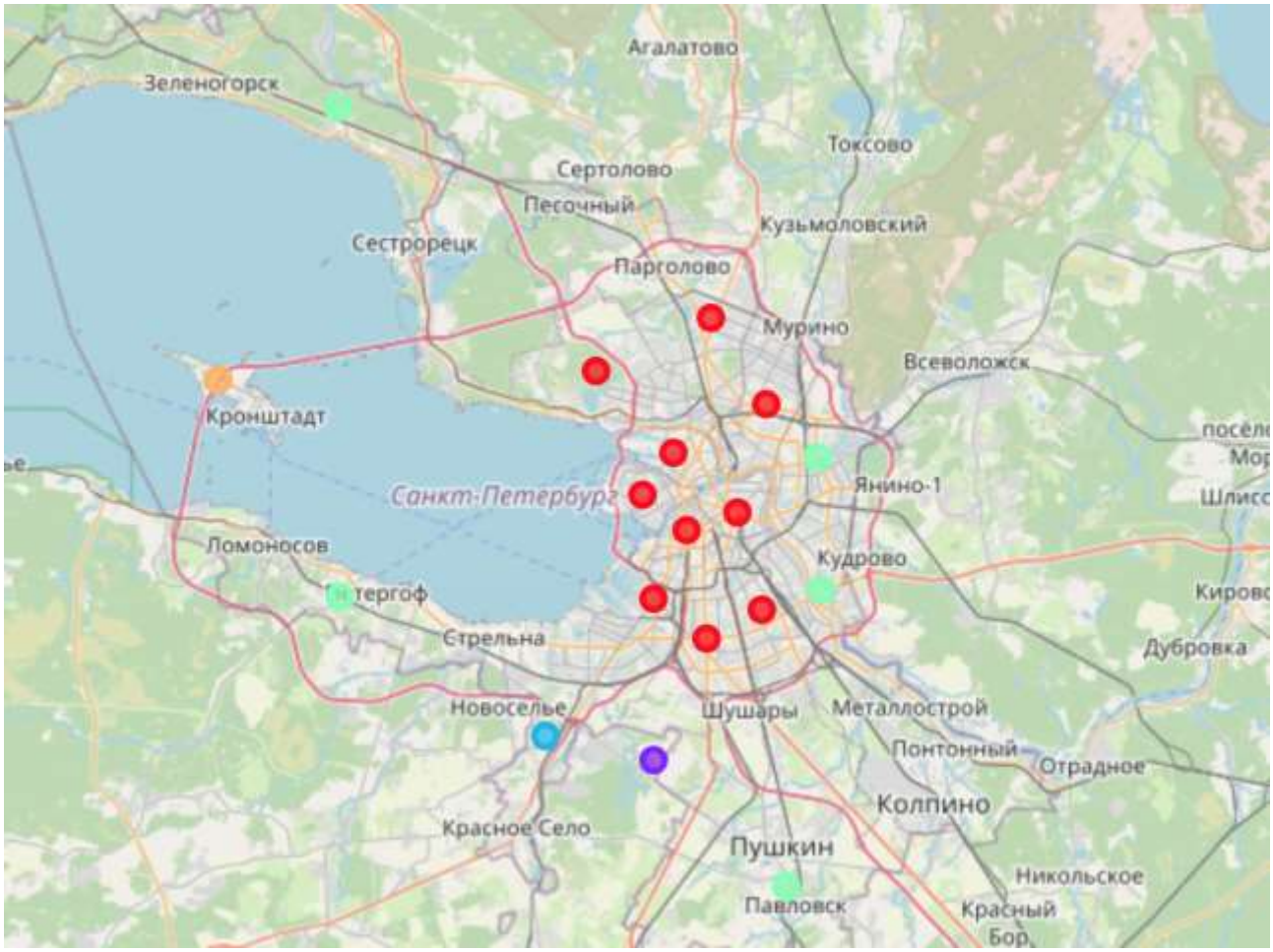
They are concentrated more towards the centre:

Центральный 10
Адмиралтейский 8
Василеостровский 5
Петроградский 5
Московский 4



Then I use one-hot encoding to convert the venue category into categorical variables for each of the venues type.
I proceed with grouping the results per each district. Then I find the top 10 venues of each of districts.

The method, I have used next, is k-mean. K-mean clustering is applied to take 5 clusters and proceed to visualize these clusters and the most common venues for each of these. The plot below represents all 5 clusters.

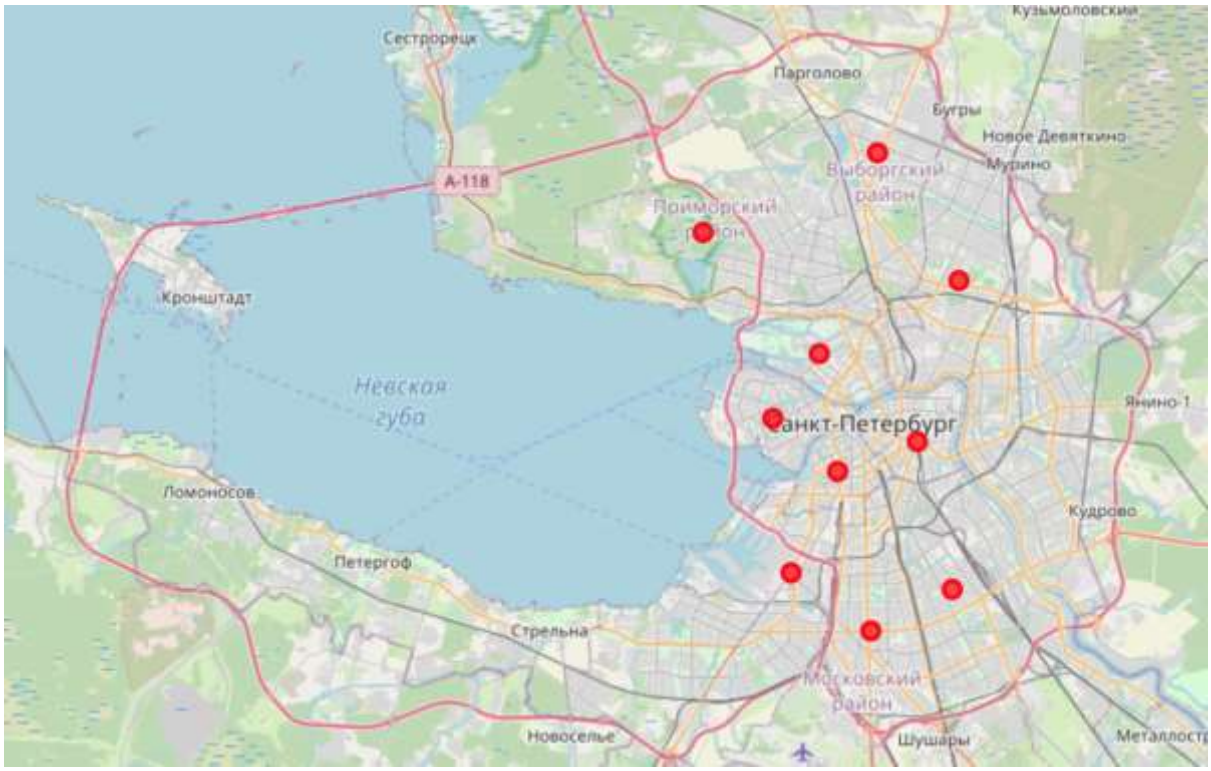


Part 5

Results

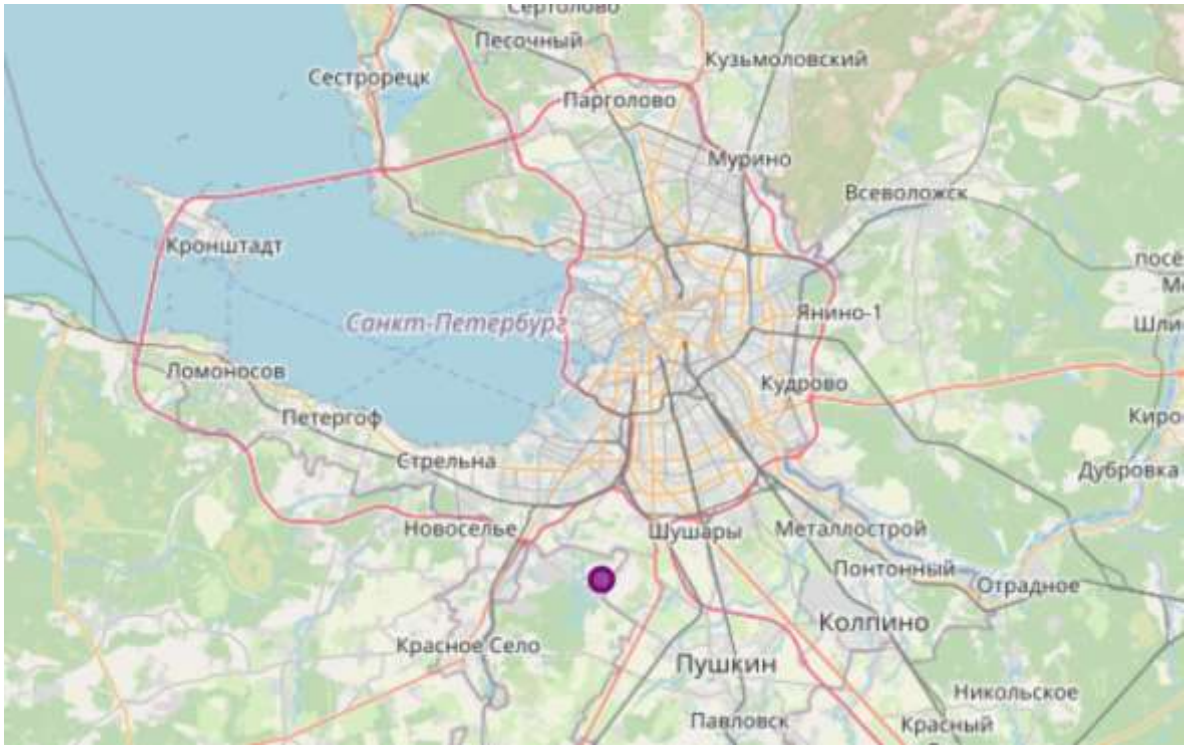
Cluster 1

Cluster 1 combines the majority of districts: Выборгский, Приморский, Калининский, Петроградский, Василеостровский, Центральный, Адмиралтейский, Фрунзенский, Кировский, Московский/
The common venues are Coffee Shops, Bakeries, Parks, Gyms, Dance Studios, Bars and Restaurants.



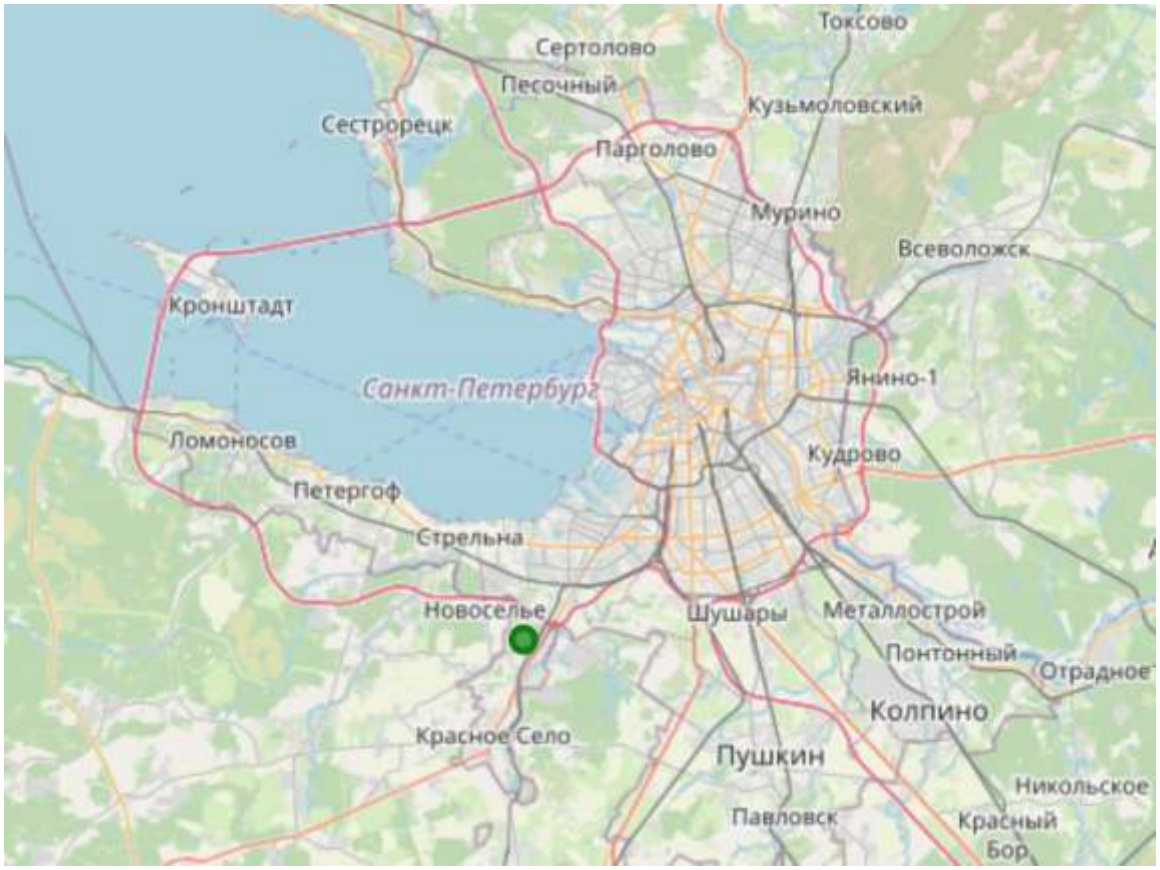
Cluster 2

Cluster 2 is Колпинский district only. It has special venues such as River, Lake, Paintball Field, Scenic Lookout, Airport Service and Fishing Spot. There is everything for high-quality rest on nature.



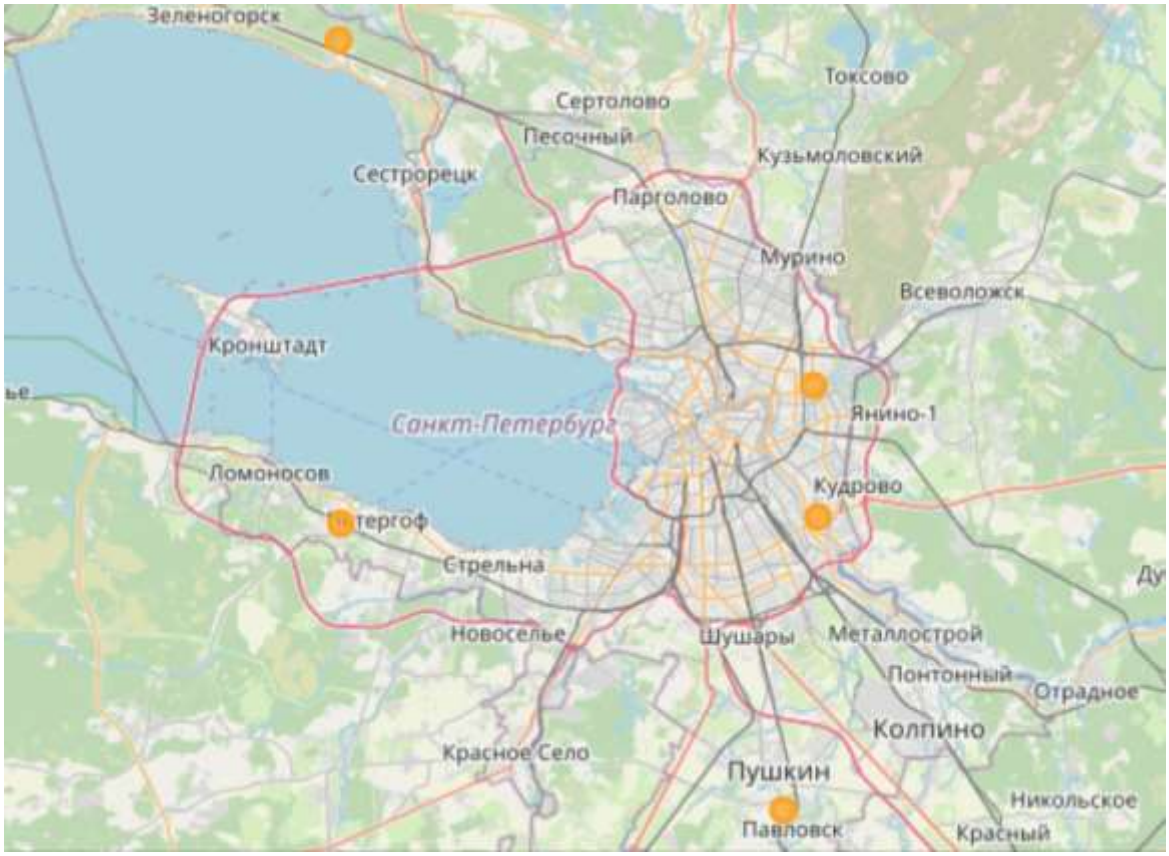
Cluster 3

Cluster 3 is Красносельский district only. It has special venues such as Convenience Store, Baby Store, Gym / Fitness Center, Lake, Grocery Store. It suits for local citizens, not popular for tourists.



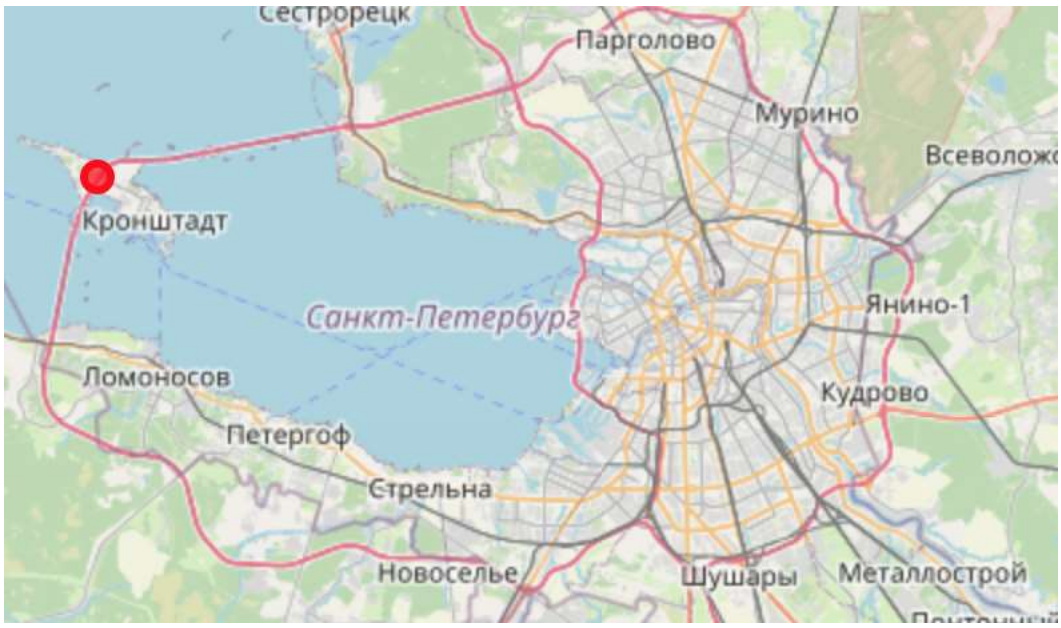
Cluster 4

Cluster 4 combines a part of districts: Красногвардейский, Курортный, Невский, Петродворцовый, Пушкинский.
The common venues are Restaurants, Parks, Café, Auto Workshops, Gym / Fitness Center.
These districts are popular for living of local citizens too, not touristic.



Cluster 5

Cluster 5 is Кронштадтский district only. It has special venues such as Surf Spot, Airport, Historic Site, Soccer Field, Stables. There is everything for high-quality rest on nature.



Part 6

Conclusions

The most popular kind of places in the city are Bakeries, Parks and Coffee Shops. Locations: Bakeries in the centre and some districts in the north and south, Parks in the centre, in the south of the city and nearby city — in Pushkin town. Coffee Shops are more in the centre.

Districts are grouped by the most common venues and I can recommend some of them for city tourism in the centre: Петроградский, Василеостровский, Центральный, Адмиралтейский and the several for beautiful natural rest: Колпинский and Кронштадтский.

After analytics of each district, I can say that although Saint Petersburg is a touristic city, some districts can be not only located far from the centre, but they can be comfortable only for local citizens, not for tourists, for example, Красногвардейский, Курортный, Невский.