

City against nature

Table of Contents

1.	Introduction	3
2.	Data preparation	4
3.	Analysis	6
4.	Results and Discussion	7
5.	Conclusion	7

1. Introduction.

St. Petersburg is one of the most beautiful cities. There are many park areas here. I bought an apartment in St. Petersburg 10 years ago on the outskirts of the city. There were a lot of green areas nearby. But over the past 10 years, the area has become heavily built up. Fewer and fewer places to stay. This situation is not only in my area. Environmental issues come first. I tried to show the problem centers in this project. Perhaps such mechanisms should be used by the city government for planning construction.

2. Data preparation

This notebook is very inspired by the work I found for the [example](#) I'll keep the idea of clustering the city by area and then build a heatmap to find the best area.

I will change some data:

Country / city: Russia Objective: find areas without access to parks So, I will be crossing data from business days and localizations.

I will be using the following API:

Foursquare API: for finding parks geopy: reverse geolocation

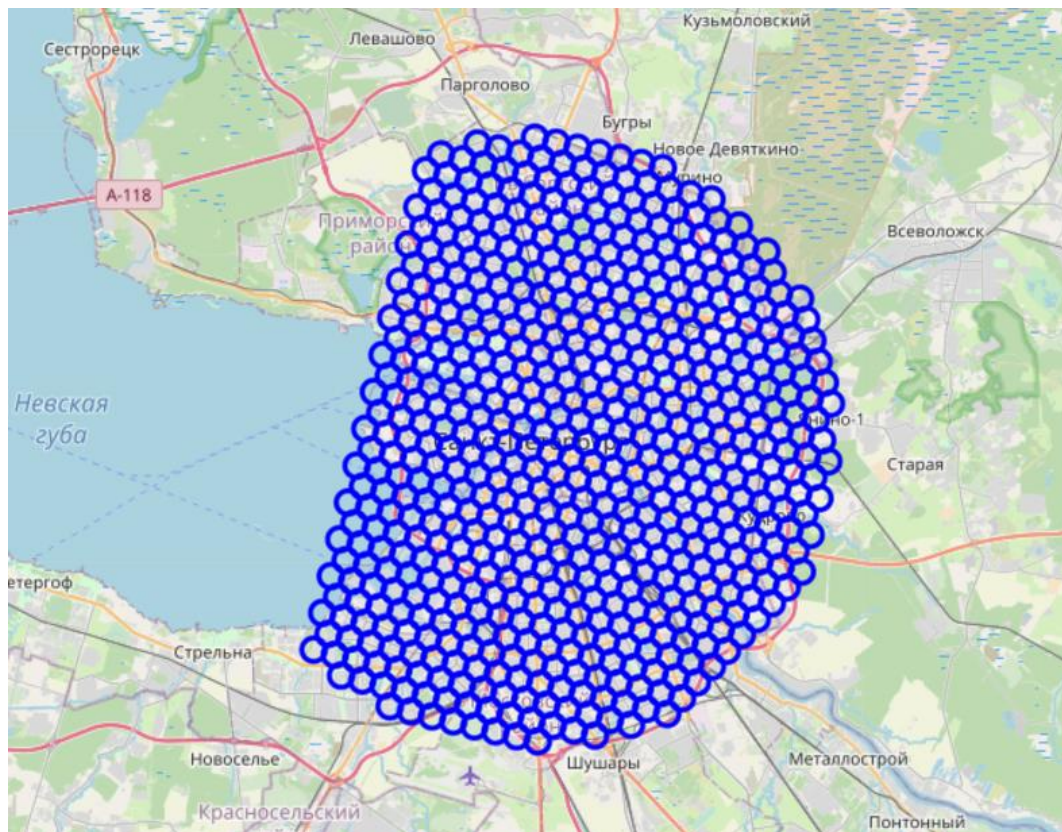
2.1 Dataset 1. Neighborhood Candidates

We create a grid of area candidates, equally spaced, centered around city center and within ~15km. Our neighborhoods will be defined as circular areas with a radius of 500 meters, so our neighborhood centers will be 1000 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

Let's create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all it's neighbors.

Let's visualize the data we have so far: candidate neighborhood centers:



Using the foursquare service, we can get data about nearby places at the specified coordinates to get info on parks in each neighborhood.

We're interested in venues in 'Park' category.

For each neighbor, we request data from the Foursquare API to find nearby parks. We save the coordinates of the point of interest and the found parks to the dataset.

Information we need:

N – number Neighborhood

Lat, lng – geographical coordinates

Id – venue id

Name - venue name

Vlat, Vlng – venue geographical coordinates

Address – venue address

Distance – distance to center of Neighborhood

After transformation we have dataframe with neighborhood and their nearby venues.

N	lat	lng	id	name	vlat	vlng	address	distance
12	59.836981	30.213834	516016ace4b0dcf886fe2a11	Двор	59.840003	30.210297	Россия	390.0
12	59.836981	30.213834	53ceb280498e0eaa7cd0b3bf	Сквер у реки	59.833698	30.211293	просп. Ветеранов, Санкт-Петербург, Россия	392.0
18	59.824717	30.317080	4ff47a14e4b0e8d00c65d6f8	Сквер вдоль Пулковского шоссе	59.827217	30.322655	Пулковское Шоссе, Санкт-Петербург, Россия	418.0
18	59.824717	30.317080	56cb024d498e18ceb1235d6c	сад "Дубовая Роща"	59.824318	30.321292	Россия	239.0
24	59.849572	30.174286	4c5c467294fd0f47ac24c845	Южно-Приморский парк (Парк Ленина)	59.847895	30.172891	Петергофское ш. (ул. Доблести и Петергофское ш...	202.0

From this dataset, we get auxiliary datasets "parks_latlon" with park coordinates and a dataset "dfbad" with places without parks.

3. Analysis

After visualizing the data, we can conclude:

Parks are scattered throughout the city, but there are areas without park areas. We will now focus on identifying the worst areas.

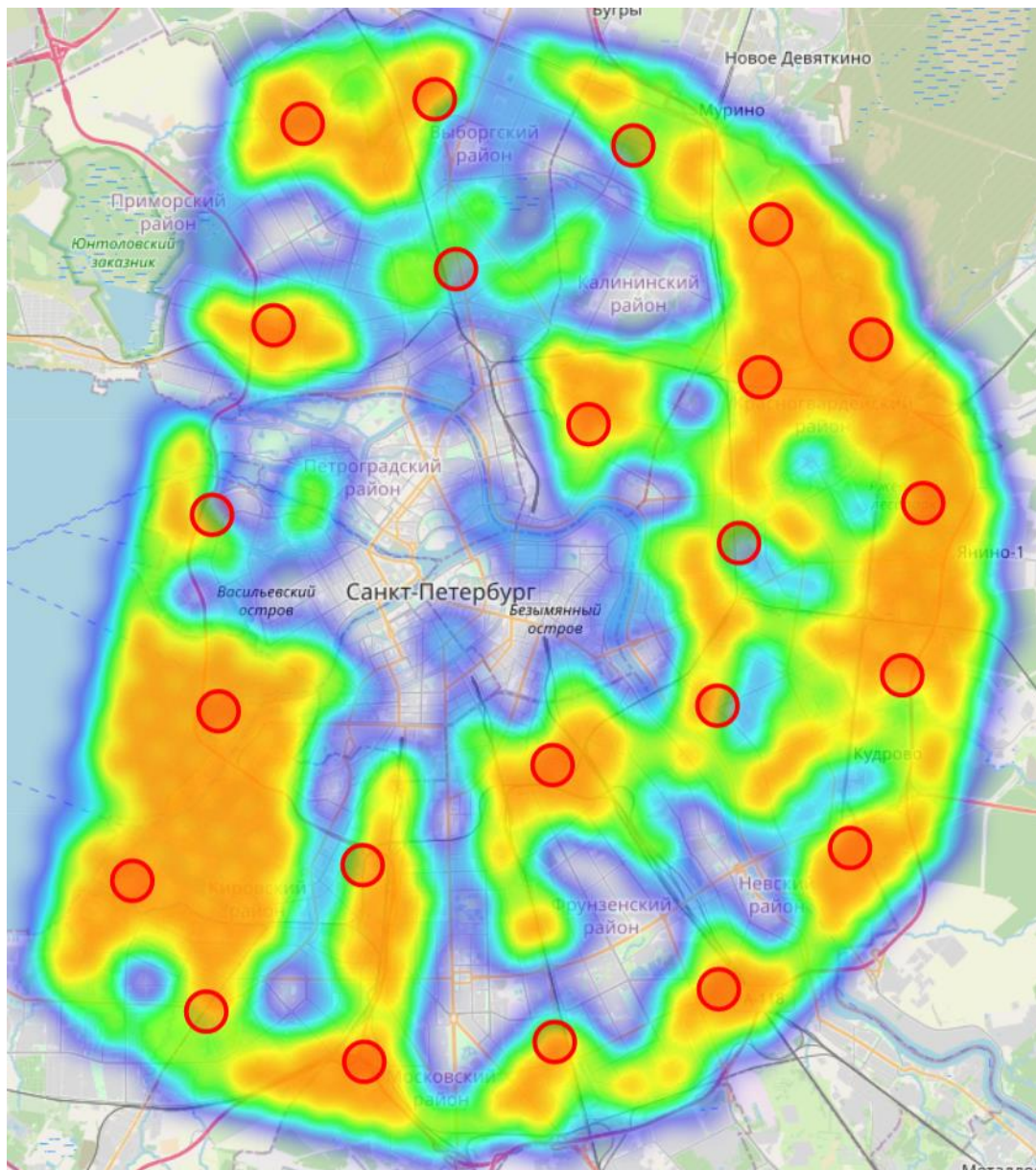
For this we will use the k-means clustering method.

We cluster those bad locations to create centers of zones containing bad locations. Those zones, their centers and addresses will be the final result of our analysis.

Our clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Let's see those zones on a city map with heatmap:



4. Results and discussion

This analysis shows that the situation in the city center is more favorable. Active construction is going on in new districts on the outskirts. Little attention is paid to landscaping. Parks are often sacrificed for lobbying the interests of developers.

The centers of problem areas were identified, in which there are no parks nearby. The data is correct. But for a real project, many other parameters must be taken into account. For example, the population in an area, a park area.

5. Conclusion

This project can be reused for other cities, just think about changing clustering size to adapt to your city.

It's very far from being perfect, a lot of work can be done, other source of data can be found, but in the end the result seems to correlate with the real world, when we know the city, the area predicted seems correct.