Research of Minsk bakeries

Volha Halynskaya

08 July 2020

Introduction

Background:

When I was looking topic of research I really want to use my native town at this task and definitely want to explore topic which I interested, but the main purpose was to create research using foursquare application with some restrictions and collect the data about companies.

"What district of Minsk really needs more bakeries? "

This topic I decided to research also because I really like morning's croissant with fresh coffee and would like to know where's places I should visit to find it and the same time I'm planning to move and I'll know which district have enough it now.

Target audience of this research:

It's definitely people who interested in opening bakery and looking for district to open place. In my mind this person should consider how many bakeries Minsk has now and how peoples rated it.

Opening hours in my opinion also important, because I and my colleagues and friends really like to buy fresh croissant and morning coffee when we're going to work.

Problems:

- Information about Minsk district and it coordinates doesn't exist in open source and I scan it manually from yandex map and create a list myself.

- API Foursquare doesn't have a lot of users in Belarus and data is not full.

Plan of my research:

- Download the data – this part is concluding creating of data frame with information about districts of Minsk and it coordinates.

- Using API Foursquare I should to know id of bakeries category for searching all company this type and full their information. *

- Join all data into one data frame including information about each company and its district.

- Deleting duplicate rows, transforming columns.

- Analyze it and appending on map.

- Using K-Means create model and fit it on data frame.

- Give a conclusion.

* Because the free part of API service is limited by calls.

Data

Data description of research:

- Open source data about Minsk districts from yandex map.

- API foursquare to find all bakeries of Minsk.

I should show that in Belarus Foursquare application is not in high demand and data which I got not include all Minsk bakeries and places. The same I can consider about companies information, for example about prices.

Methodology section

Using libraries: pandas, numpy, requests, json, geodesic, translit etc. Machine learning techniques: K-means for clustering data.

Data frame Minsk districts you can see below:

```
The geograpical coordinate of Minsk are 53.902334, 27.5618791.
```

	district	dist_lat	dist_lng
0	Центральный район	53.935766	27.521649
1	Советский район	53.940933	27.585951
2	Первомайский район	53.940165	27.646695
3	Партизанский район	53.902873	27.634289
4	Заводской район	53.869370	27.647405
5	Ленинский район	53.859626	27.585654
6	Фрунзенский район	53.906236	27.453907
7	Октябрьский район	53.858930	27.538223
8	Московский район	53.871021	27.492966

Data frame bakeries of Minsk you can see below:

	district	summary	id	name	lat	lng	categories
0	Центральный район	This spot is popular	5473097b498e2d4d05afda1b	Пан Круассан	53.929749	27.541323	Bakery
1	Центральный район	This spot is popular	59ee1ae8d0a1496688e63150	Brioche Bistro	53.939274	27.478426	Bakery
2	Центральный район	This spot is popular	58ba97964bc2f1632a353e04	Stories	53.900877	27.552514	Bakery
3	Центральный район	This spot is popular	522dc5b311d2920f2460d44c	Cakes.by	53.900149	27.522044	Bakery
4	Центральный район	This spot is popular	4f7afb6ae4b02579039ddc3f	Зерно	53.912388	27.580818	Bakery

According to plan of research:

- I downloaded all names of company with category = "Bakery" from API.

- I downloaded all inside information about each company, for example rating, price etc.

- I joined 2 data frames in one using district as a key.

- I calculated the distance between central coordinate of the district and coordinate of company and delete duplicate rows.

The shape of final data frame contain (87, 21). It's mean that my table is including 87 unique rows with 21 columns.
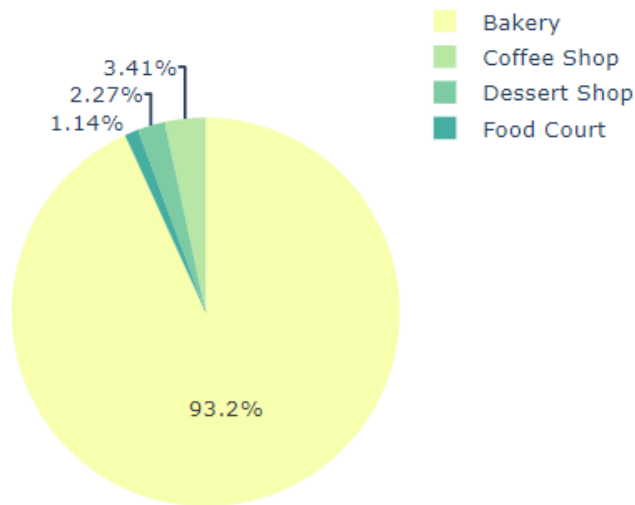
Example:

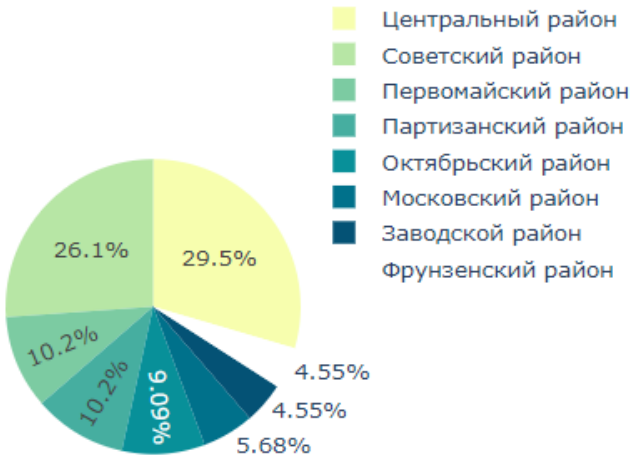|	name	district	categories	rating	price	tips
0	Бейкери дю Солей	Октябрьский район	Bakery	7.6	2	142
1	Каравай	Партизанский район	Bakery	6.1	0	74
2	Хлебны куток	Центральный район	Bakery	6.1	0	6
3	Горячий хлеб	Московский район	Bakery	5.6	0	4
4	Штолле / Stolle	Советский район	Bakery	5.9	2	112
...	...	...	...	...	...	...
82	Файны Ліс	Советский район	Bakery	0.0	0	1
83	PAUL	Советский район	Bakery	6.1	0	4
84	Территория Кофе	Советский район	Coffee Shop	8.4	0	5
85	Птифур	Партизанский район	Bakery	0.0	0	0
86	Brø Bakery	Центральный район	Bakery	0.0	0	0

Below you can see different plots reflective count of bakeries by each district and count of bakeries by categories from Foursquare API.

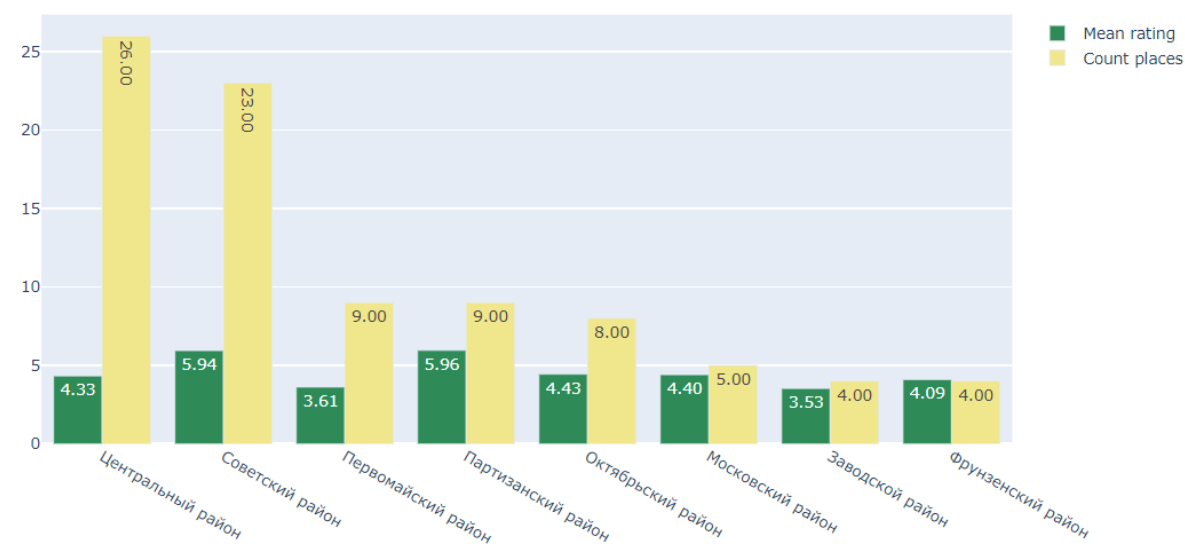**Below you can see Mean rating and count likes of places by district:**

Count of places by categories

- Bakery
- Coffee Shop
- Dessert Shop
- Food Court

3.41%
2.27%
1.14%

93.2%

Count of places by district

- Центральный район
- Советский район
- Первомайский район
- Партизанский район
- Октябрьский район
- Московский район
- Заводской район
- Фрунзенский район

26.1%
29.5%
10.2%
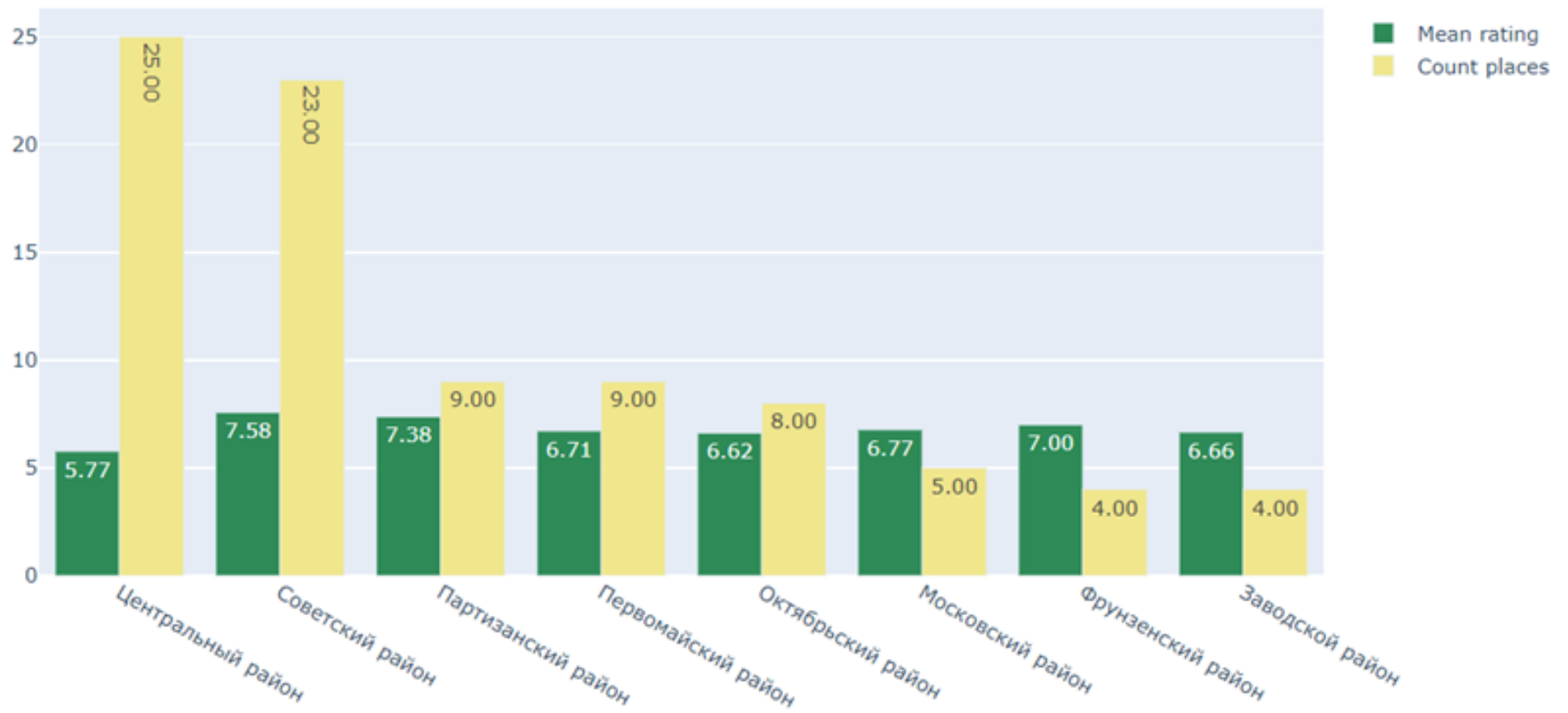10.2%
9.09%
4.55%
4.55%
5.68%

Mean rating and count likes of places by district

As we've seen the plot higher:

1. 'Центральный район' has most of bakeries in Minsk.
2. 'Партизанский район' has highest rating bakeries in Minsk.

1. 'Центральный район' has most of bakeries in Minsk.
2. 'Советский район' has highest rating bakeries in Minsk.

After collecting and cleaning data I used K-Means algorithm of machine learning to separate data frame by 5 clusters. Below you can see explain of clusters and example of bakeries situated on map:
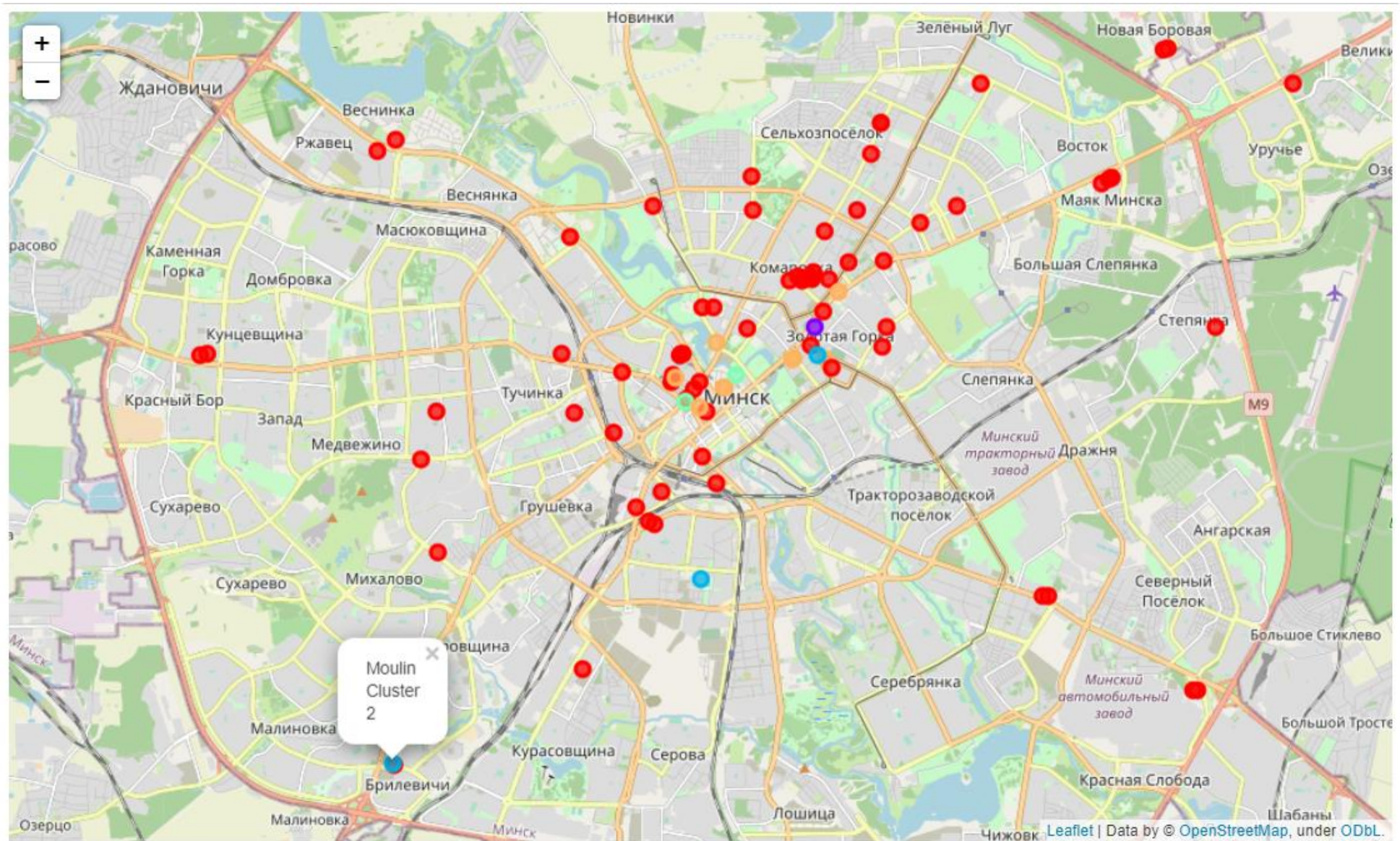
Clusters 1 and 2 have the best rating, prices and likes and its blue and green dots on map.

From 88 bakeries in Minsk just 4 of it has a really good values. Below you can see table of it:

	id	likes	dislike	rating	url	price	ok	ratingSignals	specials	tips	...	summary	name
8	4f7afb6ae4b02579039ddc3f	1011	False	7.9	http://cafezerno.by	2	False	1541	0	357	...	This spot is popular	Зерно

	id	likes	dislike	rating	url	price	ok	ratingSignals	specials	tips	...	summary	name
33	54fee005498e8e467263436b	340	False	7.0	http://www.cafezerno.by	2	False	588	0	103	...	This spot is popular	Зерно
0	4bfcf7b58992a593595fabb0	399	False	7.3	http://cafe-bakery.relax.by	2	False	601	0	142	...	This spot is popular	Бейкери дю Солей
19	50f17208e4b08ef910e45079	326	False	8.7	http://moulin.relax.by	2	False	465	0	131	...	This spot is popular	Moulin

Below you can see example of clusters situated on Minsk map:

Results:

Answering to the question:

Which Minsk district really needs more bakeries?

I should notice that most count of bakeries belong to the center of city and central district.

And it's okay because most of cultural and evening life are using there, all European cities has the same statistic.

But according to my question - bakery for me, it's firstly the morning coffee and croissant and I defiantly recommend opening bakeries not in the center of city, but in the new districts far from the center.

Discussion:

Really great example we can see at 'Moulin' bakery which is according to the high 2 cluster. From all 88 Minsk's bakeries just 3 from it has 2 cluster and it place is situated to new district.

Conclusion:

I invite everyone to Minsk to explore the best bakeries and the freshest croissants. You can visit each of these bakeries and create your own rating.