Coursera Capstone Project

Applied Data Science Capstone

Table of contents

Opening a gym in Vilnius

Introduction: Business Problem

Vilnius is the capital and the largest city of Lithuania. The city itself has a population of ~550 000 people. Vilnius consist of 21 elderships, a statewide administrative division, which are based on neighborhoods. Many people around the world are facing serious health issues associated with

unhealthy eating habits and a lack of exercise. With growing health concerns, more people are joining gyms for exercise and the motivation to change their eating habits. Lately Vilnius has beed growing very fast resulting in growing demand of gyms.

This final project explores the best places to open new gym throughout the city of Vilnius. Questions that will be answered in this project:

1. Is there any location without gym?
2. Which areas have large number of gyms?
3. Which areas have least number of gyms?
4. What are areas with least competition? (at least one gym in that area)

Targert audience

This project is particularly useful to developers and investors looking to open or invest in fitness industy in Vilnius. Why did we decide to focus on fitness industry in our project? Now when the idea of a healthy lifestyle conquered the minds of people all over the world, gyms became extremely popular as minimal amount of daily exercise will help maintain a long and healthy life.

Data

Based on definition of our problem, factor that will influence our decission is number of existing gyms in the neighborhood (or any type of fitness center)

To solve the problem, we will need the following data:

1. Vilnius city data contain neighborhoods.
2. Coordinates of those neighborhoods.
3. Venue data, particular data related to fitness industry.

Data source

Vilnius city data containing the neihborhoods will be obrained from wikipedia. After it, we will get geografical coordinates of the neighborhoods using Python Geocoder package. To get the venue data for the neighborhoods we will use Foursquare API. We are particulartly interested in the gyms/fitness industry data to solve business problem defined above.

Data scraped from multiple sources were combined into one table. There weren't any missing data, but there was a lot of dublicates. Script for cleaning dublicates was written and executed.

Methodology

In first step we will collect the required data: location and type (category) of every gym in each neighborhood.

Second step in our analysis will be calculation and exploration of number of gyms across different neighborhoods.

In third and final step we will focus clusters of locations. We want locations where numbers of gyms is least. We will present map of all clusters and identify most promising neigborhoods.

Analysis

Number of gyms were counted in every area.
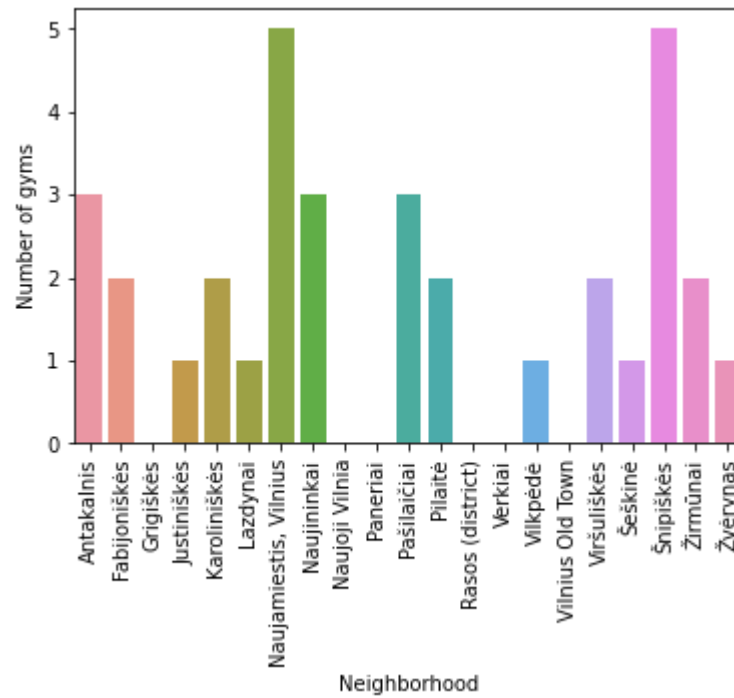
As we can see, there are some areas where there aren't any gyms. Area without gyms:

- Grigiškes
- Naujoji Vilnia
- Paneriai
- Rasos
- Verkiai

Areas with least number of gyms:

- Justiniškes
- Lazdynai
- Vilkpede
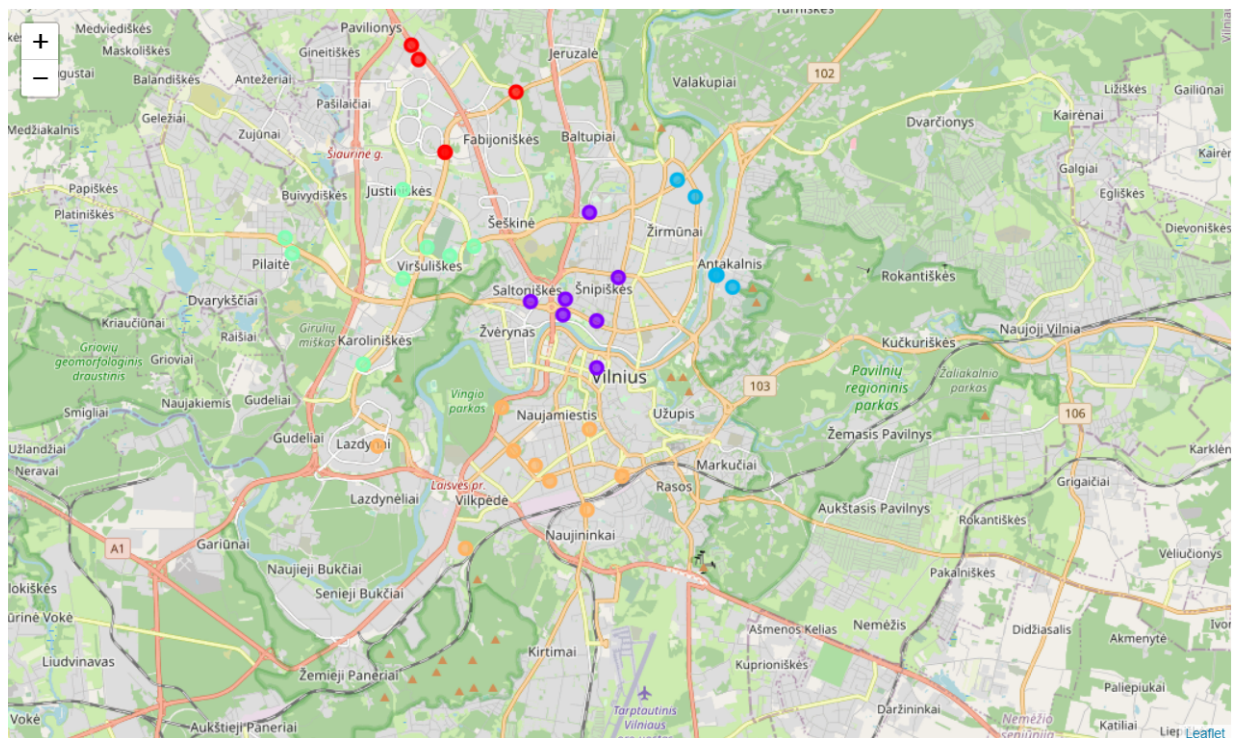- Seskine
- Zverynas

There are two area with most gyms:

- Naujamiestis
- Šnipiškės

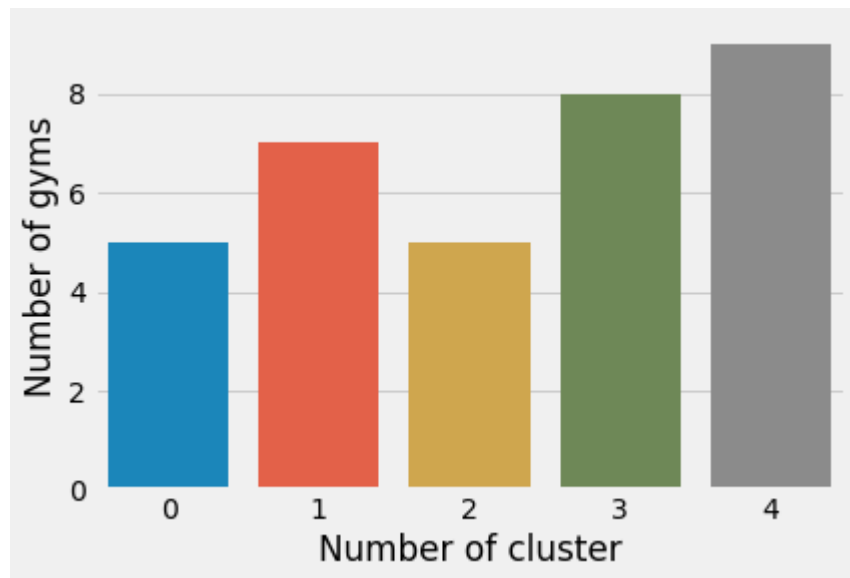For further analysis neighborhoods without gyms were droped out.
Let's test clustering and find out what is proper amount of clusters is needed for analysis. We will use elbow method to determine number of clusters.
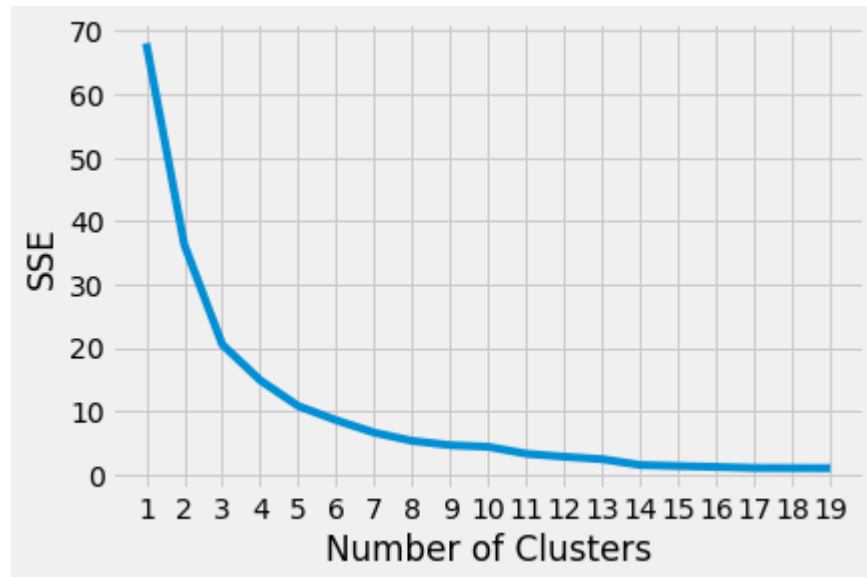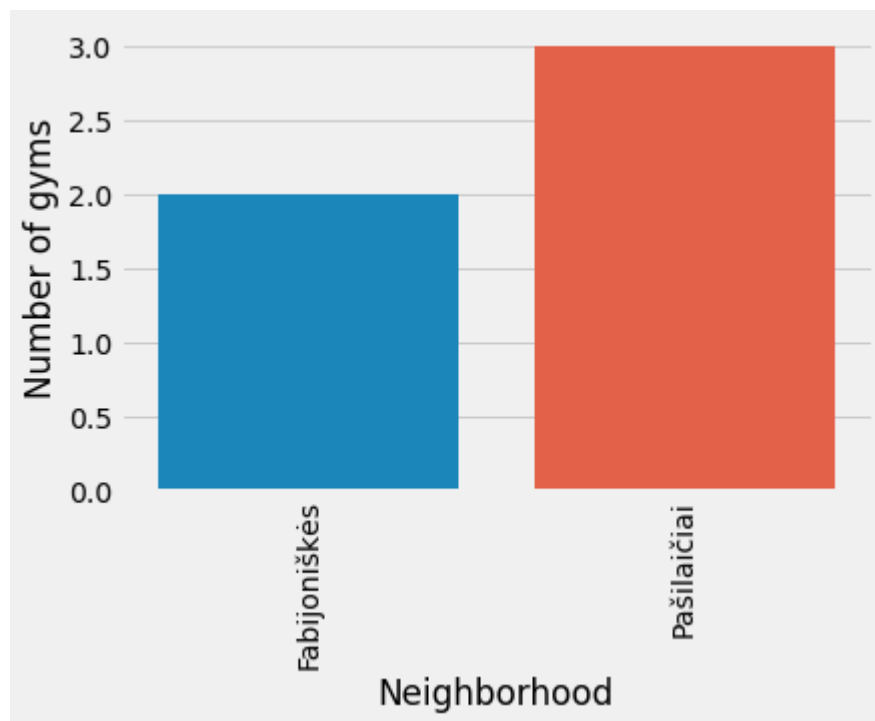

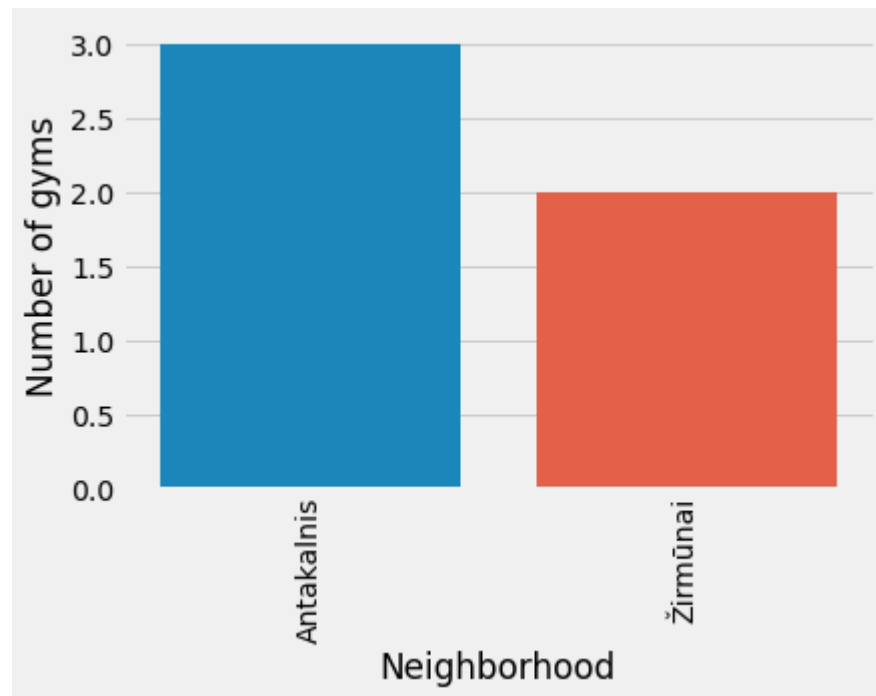As result shows - we need to use 5 clusters in further analysis.

Let's see how those clusters look on barchart:

As we can see cluster 0 and cluster 2 have least gyms. Let's explore them little bit more.

In this cluster neigborhood Fabijoniskes have least gyms, this area might be potencial for new gym.

In cluster 2 neigborhood Zirmunai have least gyms, this area also might be potencial for new gym.

Results and Discussion

Our analysis shows that there is 34 gyms in Vilnius. Some areas don't have any gyms. Those areas might be interesting to open new gym, but there is always a possibility, that area due at the moment unnown reason is not suitable for fitness industy. That why for further analysis we looked at areas where already is competition. Areas with highest and least compeatition were found. In least competition areas we have found two neigbohoods which might be suitable for opening new gym.

Conclusion

Questions that where answered in this project:

1. There is 5 neighborhoods without any gym:

- Grigiškes
- Naujoji Vilnia
- Paneriai
- Rasos
- Verkiai

2. There are two area with most gyms:

- Naujamiestis
- Šnipiškės

3. Areas with least number of gyms:

- Justiniškes
- Lazdynai
- Vilkpede
- Seskine
- Zverynas

4. Areas with least competition:

- Fabijoniskes
- Zirmunai

In [ ]: