**Finding best location for our new shopping mall**

Mindaugas Rudzevicius

April 23, 2019

**Introduction**

A **shopping mall** is a modern, chiefly North American, term for a form of **shopping precinct** or **shopping center,** in which one or more buildings form a complex of shops representing merchandisers with interconnecting walkways that enable customers to walk from unit to unit. A **shopping arcade** is a specific type of shopping precinct which is usually distinguished in English for mall shopping by the fact that connecting walkways are not owned by a single proprietor and are in open air. Shopping malls in 2017 accounted for 8% of retailing space in the United States.

**Lithuania** has around 20 shopping malls with more than 30,000 m2 (320,000 square ft) square meters of space. Another two are under construction. Around half of Lithuania's shopping malls are in **Vilnius**. From 2010 to 2016, Vilnius took 80% of all Lithuanian retail real estate investment. Lithuanian shopping malls produce 7–9% annual ROI. According to real estate analysts, Lithuanian shopping malls are very well developed and competitive, but the market itself is still under-supplied. Vilnius is estimated to have the lowest stock in Northern Europe and the most growth potential in entire European Union.

**Business problem**

As we already understood, the Vilnius’ big shopping mall domain is under-supplied so finding the best place/ location within specified requirements for our stakeholders is one of the most important tasks to do. Using available data, we can find optimal location at a current moment.

All big real estate companies are interested in a new good location, so our project can bring big values for those companies and people.

**Data acquisition and cleaning**

Data was gathered from many different sources. Mostly, it was acquired using web scraping technique.

Data for our main shopping list was scraped from Wikipedia: (<https://en.wikipedia.org/wiki/List_of_shopping_malls_in_Lithuania>) . The main information we got is the name, year built, size and number of shops in the mall.

Data for our smaller shops/chains and other type of shops we craped from : <https://www.1551.lt/prekybos-centrai/vilnius> and from Foursquare API.

The Borough list of Vilnius we acquired from Wikipedia using web scraping: (

https://lt.wikipedia.org/wiki/Vilniaus\_miesto\_savivaldyb%C4%97)

Data was cleaned by multiple steps with every data set.

**Exploratory Data Analysis**

The best location can be described in many different ways but at this project we stick to the idea that place should be in the well populated area, with as less as possible competition.

In our data there were missing some features: like addresses, boroughs.

First, we created some methods to get right addresses using popular Python libraries like Gaocoding. After that, using geojson data provided by Vilnius open data platform (<http://gis-vplanas.opendata.arcgis.com/datasets/6d5088b44dba4643a6611455d5352268_1.geojson>) we go the necessary data about Vilnius’ boroughs and finally we have required information about our main shopping malls in Vilnius.

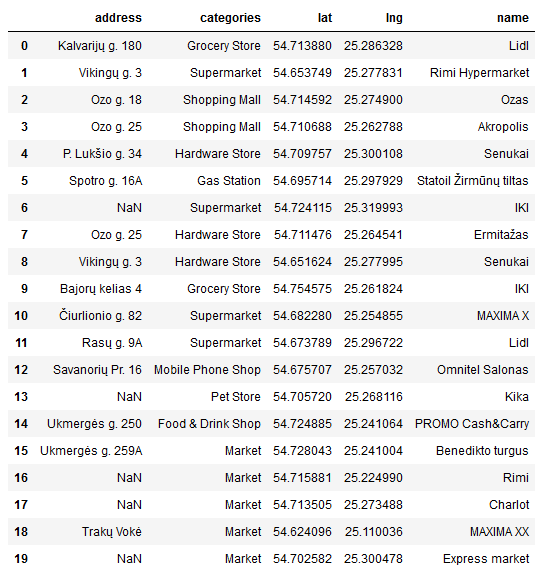


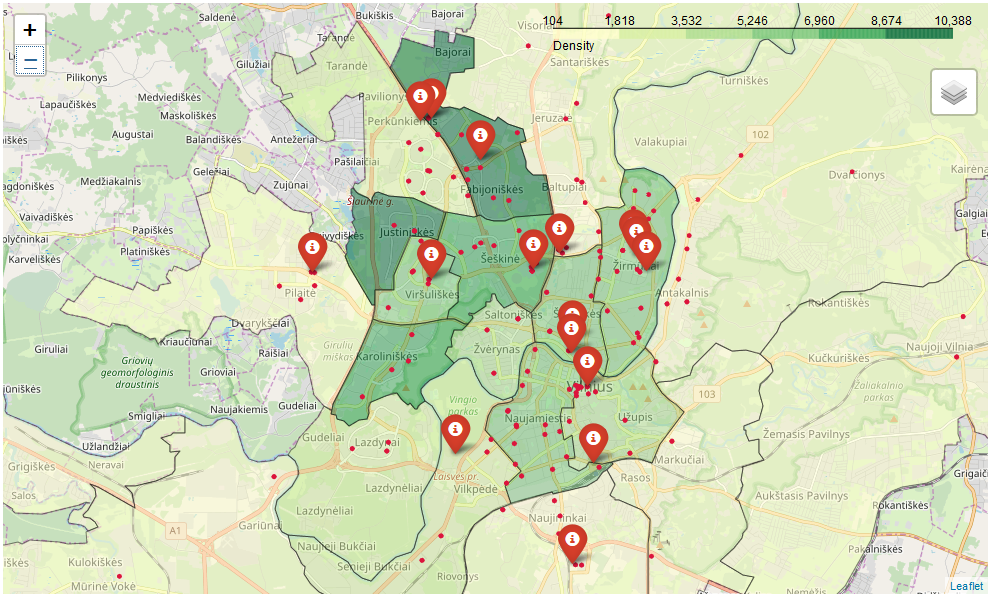
The next step was using Foursquare API and data portal 1551.lt to get information about other shops/chains. The part of the acquired information is below:

1551.lt:

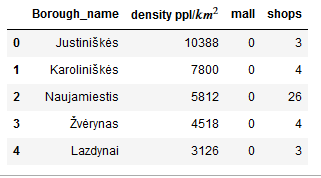


Foursquare:

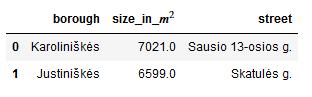


After working out with the tabular data it was easy to show everything on the map:

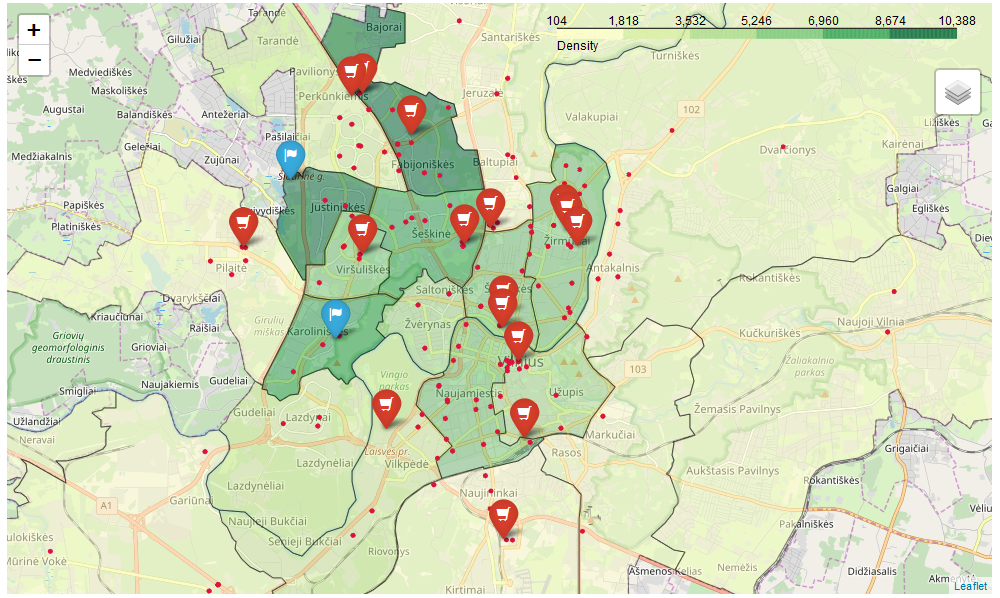
Already from the map we can see that some boroughs like Justiniskes or Karoliniskes don’t have big shopping malls. After querying data, the result shows the same:



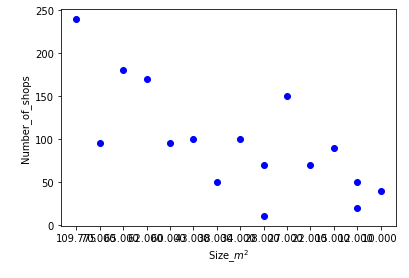
Our top 2 places should be in Justiniskes or Karoliniskes. Now using web scrape method, the information about all available land deals is acquired and extracted only those 2 boroughs we are interested in (both tables are merged):



Our map now looks like this:



We tried to use Machine Learning Regression models to predict how many shops would be optimal to build in a mall with a given parameters. But looking how widely data is populated (see below)



We decided to use averages. Because the machine learning models were overfitted or very inaccurate. The given result is around +/- 100 shops for every mall.

**Conclusion**

In this study I analyzed the locations of shopping centers in Vilnius. From gathered data I was able to determine the best possible location of the given data for new shopping mall building site. Plus the number of the shops was determined using average method.