Selecting neighbourhood for opening new pizza restaurant in Helsinki

Roman Lübitch

January 2, 2021

1. Introduction

1.1. Background

Helsinki is capital of Finland with large selection of restaurants and tough competition between them. Due to Covid19 outbreak demand for eating out has decreased significantly, also there are indications that demand for eating out will not return to the same level as remote working will become will become more common norm.

Pizza restaurants are one of the most popular restaurant types in Finland and Helsinki (quite often they serve also other types of dishes) which also commonly provide home delivery service in areas they operate and take away service, thus taking into account potential trends they might be in more advantageous position compared to other restaurant types. At the same time competition is expected to remain tough and it is essential to get as much information as possible about market potential in which neighbourhood new restaurants should be opened.

1.2. Problem

Finland's national statistical data on neighbourhood level will provide information on market size. To take into account existing level of competition we will use foursquare data. This project aims to predict in which neighbourhood in Helsinki there might be largest potential to pizza restaurant to succeed taking account existing competition

1.3. Interest

This project results will be interesting for who are wanting to open pizza restaurant in Helsinki and wants to know in which neighbourhood there is greater chance to succeed taking into account existing market and competition.

Notice also that model used for making assessment for Helsinki area can be also used with small modification also for other municipalities in Finland and venue type can be changed as well (i.e. Japanese restaurants).

2. Data

Data for the project will acquired from freely available sources – the following datasets will be used:

1. Finland's national statistical data on neighbourhood level. This data will include following datasets on post index (neighbourhood) level:
    a. Area: how large post index area is.
    b. Number of people working in selected area.
    c. Total population.
    d. Adult population (over 18 years old population).
    e. Average income per inhabitant
    f. Median income per inhabitant
    g. Number of households

h. Average size of households
i. Average income per household
j. Median income per household

2. Finnish neighbourhoods' geolocation data.
3. Pizza restaurants' location data (foursquare data).

3. Exploratory data analysis

3.1. Market overview

There are 84 neighbourhoods with total of 72 Pizza restaurants in Helsinki (based on foursquare data). Most of pizza restaurants are located logically in the city centre or in its vicinity – where most of people live and work. There are also a lot of neighbourhoods (35 out of 84) with no pizza restaurants which might give some initial indication for market potential.

Figure 1: Pizza restaurants in Helsinki on neighbourhood level.

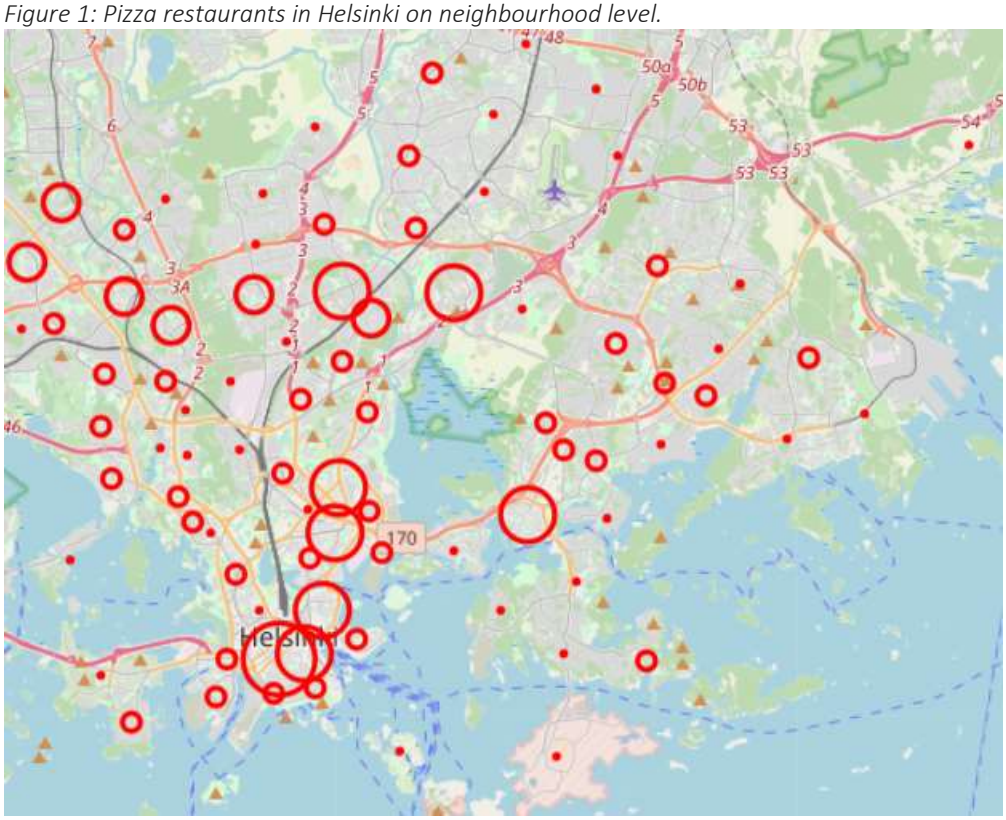
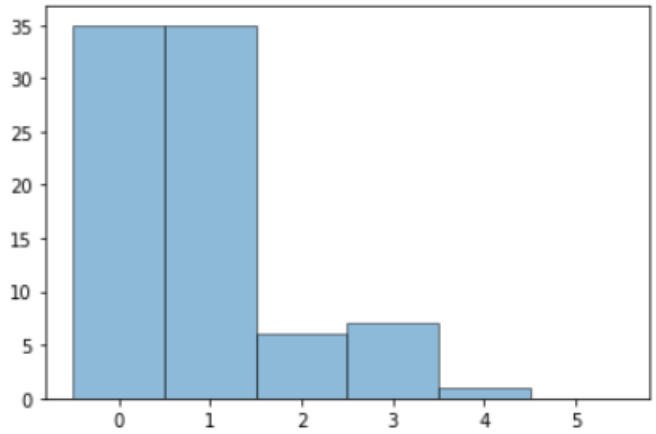


Figure 2: Number of Pizza restaurants on neighbourhood level

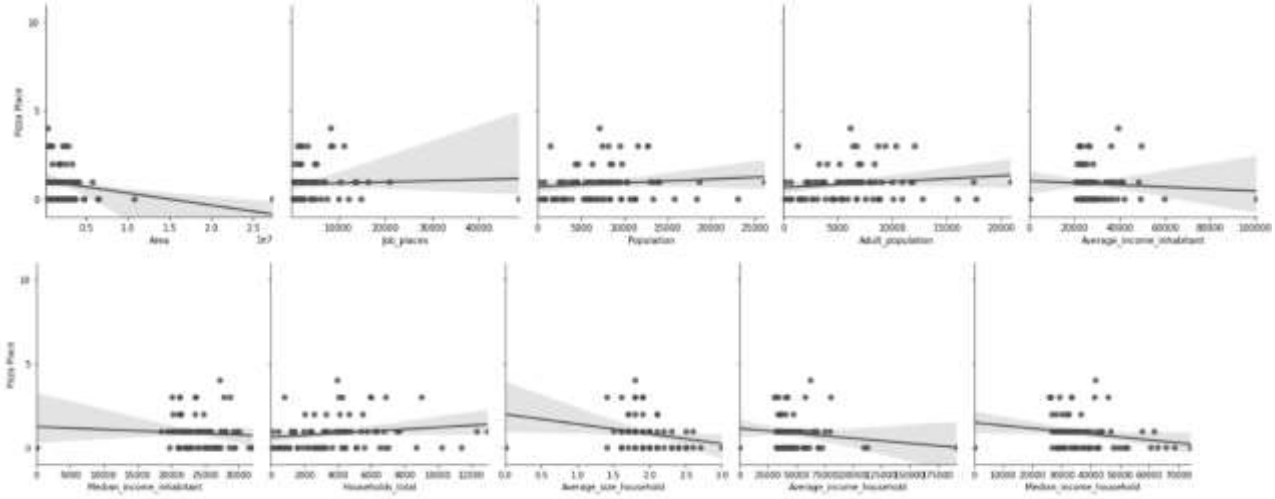


3.2. Relationships between independent and dependent variables.

Next we will have look is there clear relationship between independent variables listed below and dependent variable – number of pizza restaurants:

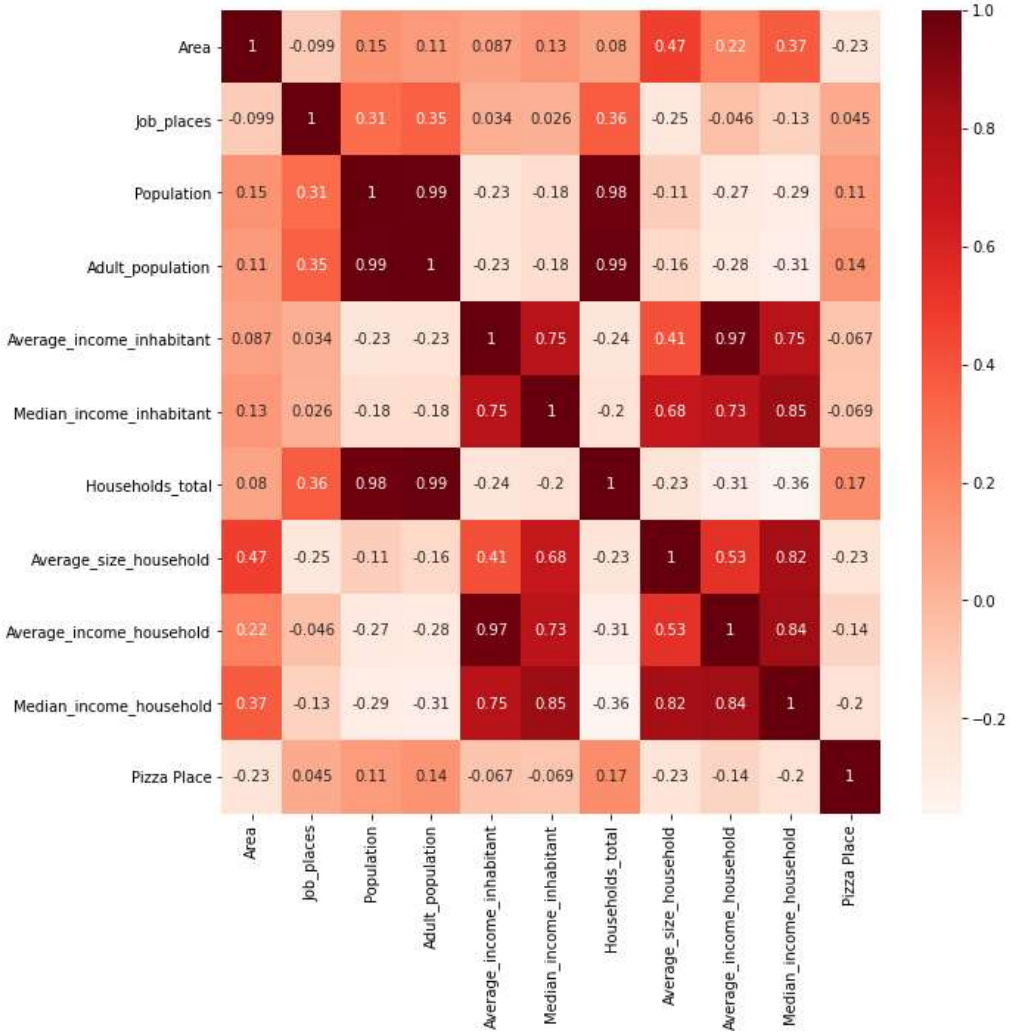
a. Area: how large post index area is.
b. Number of people working in selected area.
c. Total population.
d. Adult population (over 18 years old population).
e. Average income per inhabitant
f. Median income per inhabitant
g. Number of households
h. Average size of households
i. Average income per household
j. Median income per household

Figure 3: Visualizing relationship between independent and dependent variables



Based on data visualizations there are no clear indications about interrelationship between independent and dependent variables. Next we will have a look at correlation matrix.

Figure 4: Correlation matrix



Values show only negligible/low correlation between independent and dependent variables. By using filtering method following features will be selected for predictive modelling (based on their predictive power and removing parameters which correlate strongly with each other → removing multicollinearity problem):

- Area
- Job_places (though having low correlation - feature will be added to improve slightly estimations)
- Households_total (by selecting this feature we will need to exclude Population and Adult_population)
- Average_size_of_household
- Median_income_of_household.

These parameters will be used in SVM, KNN and logistic regression models as they are all prone to multicollinearity.

4. Predictive modelling and results

4.1. Selecting algorithm

First we will use four different models to predict how many pizza restaurants should be located in each of Helsinki's neighbourhood based on Finland's national statistical data.

Four different models to assess predictive models which will used:

1. Support vector machine
2. KNN
3. Logistic regression
4. Decision tree

For first three models we use parameters selected in previous section – to avoid multicollinearity problem. Decision tree models are not prone to multicollinearity problem and thus all independent variable listed in section 2 (Finland's national statistics) will be used for modelling.

Table 1: Performance of selected algorithms

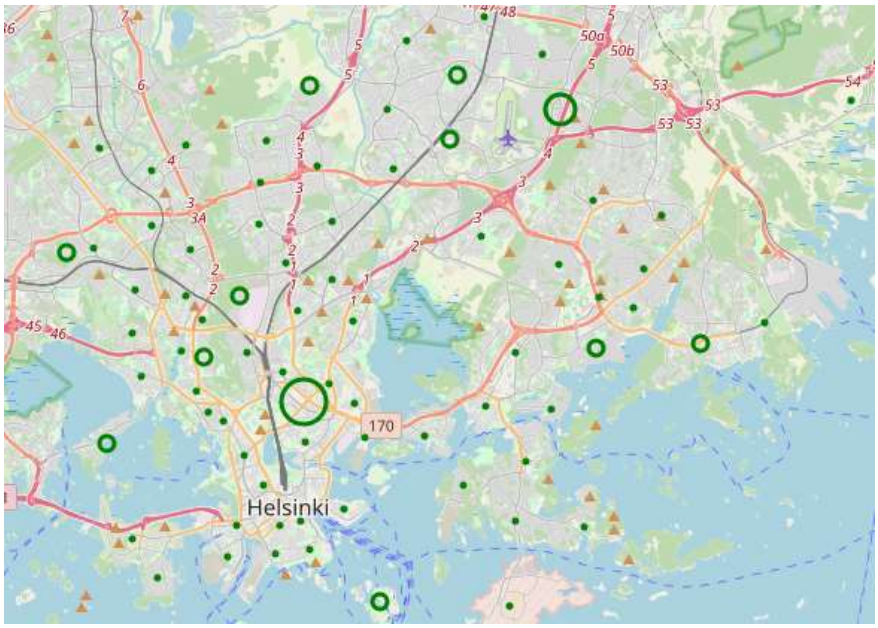
Algorithm	Jaccard	F1-score
KNN	0.312924	0.452886
Decision Tree	0.612223	0.757766
SVM	0.316764	0.456410
Logistic Regression	0.288503	0.427419

Based on results decision tree algorithm gives clearly most accurate results and should be selected. Now we can calculate difference between estimated amount of pizza restaurants in neighbourhood and deduct existing restaurants, by doing this we can plot on map neighbourhoods where new pizza restaurant will most probably be viable.

4.2. Results and recommendations

From map we can clearly see that those neighbourhoods which currently do not have pizza restaurant might have enough market potential for a one to open. In other places – like city centre, competition is already tough and new pizza restaurants might simply not be viable.

Figure 5: Helsinki's neighbourhoods with greatest for market potential for pizza restaurant



5. Conclusions and recommendation

Even though based on algorithm there is market potential for pizza restaurant in neighbourhoods presented on the map, it hard to give clear recommendation before more detailed analysis what kind of restaurants are already available in those neighbourhoods. Nevertheless this approach can clearly add value for decision making process and thus improve one's ability to make better decisions.

Also it is good to notice that created model can be used to analyse other types of venues in similar manner and in different municipalities in Finland.

Data sources:

Finland's national statistics: www.stat.fi, Free Power Query (M) generator for Statistical OpenDATA from PXWeb will be used to collect required data https://stat.qumio.com/

Finland's neighbourhood location data: https://www.avoindata.fi/data/fi/dataset/postcodes

Foursquare venue data