

Restaurant analysis for Cities and Towns in Slovakia

Introduction

In this project I am interested in restaurant market segmentation in all cities/towns in Slovakia. I am a fan of foreign cuisine. But unfortunately in my small home town (around 30 000 people) there is not much to offer. In this analysis i would like to find whether simmilar situation is the case for other towns as well.

This analysis could be helpfull for a starting restaurateur who does not know what cuisine should offer. Or for immigrant who want to share his/her native cousine, but don't know in which town to be most profitable. To be profitable, the owner should start a bussines in place with the highest possible average income of local population and also lowest possible crime rate, unemployment rate and poverty rate.

Used data

For the current state of restaurant segmentation Foursquare API will be used. Instead of location data (in API LL) about latitude and longitude, parameter near will be used to get venues from whole town.

The source of demografic information as

- Town population
- District population
- Region population
- Average wage
- Crime rate
- Unemployment rate

was found by Statistical Office of the Slovak Republic API. <https://slovak.statistics.sk/>.

For some other data like town affiliations to regions and district is scraped from wikipedia https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Slovakia

Metodology

At first all mentioned demografic data are downloaded by tatistical Office of the Slovak Republic API. Next step is to use Foursquare API to get information about restaurant category for respective cities / towns. Data cleaning is necessary because after recieving data from Squarespace many venues are in category like "Caffe", "Bakery" or "Creperie" but there are not interesting for this analysis since the main focus is on restaurants. So, they were removed Categories like "Restaurant", "Diner", "Bistro" or "Buffet"

are not specific enough. That is why they were put to single category "Others" and were removed from main analysis, however the data was used to find information about restaurant per capita.

Main algorithm of town selection for specific restaurant category

The main idea how to decide where the specific category of restaurants should be opened is based on following points. Restaurant should be opened in town where there is:

- 1) Lowest ratio of restaurant per capita
- 2) Lowest ratio of restaurant specific category per capita
- 3) Lowest district unemployment
- 4) Highest district wage
- 5) Lowest region crime rate

For every restaurant category a score for every city / town is calculated. The most profitable place of specific restaurant category is in the town where there is highest score. The score is calculated based on following equation:

$$\text{score} = (1 - \text{District Unemployment}) * A * \text{District Average Wage} * B * (1 - \text{Crimes per 100k for Region}) * C * (1 - \text{Restaurants per capita}) * D * (1 - \text{Specific Restaurants per capita}) * E$$

where A, B, C, D and E are penalty factors which can be set according to severity of each feature. Penalties were selected as follows

A = 0.2 (unemployment)

B = 0.2 (wage)

C = 0.1 (crime)

E = 1 (restaurants per capita)

F = 1 (specific restaurant category per capita)

Feature scaling

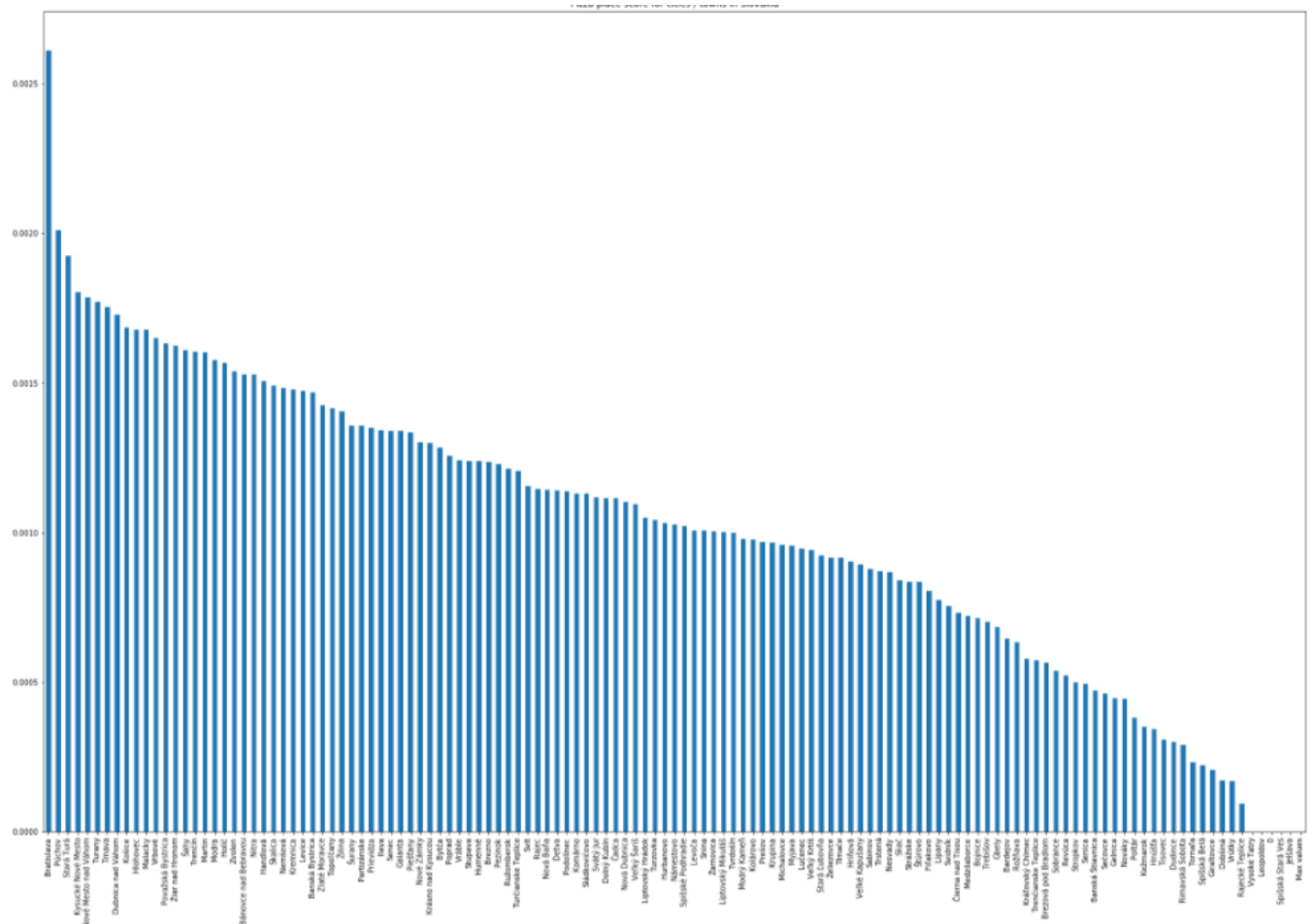
It is necessary to scale the features before putting data to main algorithm. MinMax scaler was used for this purpose. All features are scaled between 0 and 1. However, 0 and 1 are not representing minimum and maximum of particular feature, because score equation would not work for instances with 0 or 1. 0 is representing by real zero number and 1 is for demographic data represented by selected maximum values:

- Wage of 1975 € as 0.9 percentile
- Unemployment of 19.2 % as historical maximum for Slovakia (year 2001)
- Crime rate of 1295.6 cases per 100 000 people

For restaurant per capita and specific category restaurant per capita is 1 defined as just maximum of features.

Results

The Main algorithm was used for 52 restaurant categories (according to Squarespace venue categories). In following chart can be seen the score for Pizza places for all cities / towns



Five most frequent restaurant categories with 5 towns with highest score are shown below.

Pizza Places

Bratislava	0.002611
Púchov	0.002010
Stará Turá	0.001925
Kysucké Nové Mesto	0.001804
Nové Mesto nad Váhom	0.001785

Slovak Restaurants

Bratislava	0.002619
Púchov	0.002010
Stará Turá	0.001925
Trnava	0.001916
Nové Mesto nad Váhom	0.001916

Fast Food Restaurants	
Bratislava	0.002636
Púchov	0.002010
Stará Turá	0.001925
Nemšová	0.001886
Vráble	0.001823

Italian Restaurants	
Bratislava	0.002502
Púchov	0.002010
Stará Turá	0.001925
Hlohovec	0.001924
Nové Mesto nad Váhom	0.001916

Gastropubs	
Bratislava	0.002636
Púchov	0.002010
Stará Turá	0.001925
Hlohovec	0.001924
Nové Mesto nad Váhom	0.001916

Result discussion

It can be observed that for all five restaurant categories is highest score for Bratislava and second highest for Púchov. In case of Bratislava it is quite obvious, because strongest feature in solved equation is restaurants per capita and Foursquare API can send only the results of 100 venues. Because Bratislava is by far most populated city in Slovakia it has always by far the best score for restaurant per capita. In case of Púchov for some reason there are not much restaurant in Squarespace database. Therefore, the restaurant per capita feature is always high.

Many different result were obtained by different setting of penalties for individual features. Which make this analysis quite variable.

Conclusion

An algorithm for city / town selection for specific restaurant type was proposed. It is quite important to set penalties according to future restaurant owner preferences. It is necessary to understand this analysis works as an example. To provide more robust results much more demographic features would be needed as well as other factors like lease price or energy cost.