Restaurant analysis for Cities and Towns in Slovakia

APPLIED DATA SCIENCE CAPSTONE PROJECT

Introduction

- If someone is interested in open a new Restaurant where to open it and what category of restaurant should it be?
- What features should be used to find the most suitable city / town for certain restaurant type?
- What is current restaurant distribution in Slovak cities / towns

Used data

- Statistical Office of the Slovak Republic API
  - Town population
  - District population
  - Region population
  - Average wage
  - Crime rate
  - Unemployment rate
- Foursquare API
  - Restaurant categories in cities / towns in Slovakia
- Wikipedia
  - town affiliations to regions and districts

Data cleaning

- Category like "Caffe", "Bakery" or "Creperie" are not interesting for this analysis since the main focus is on restaurants. These categories were deleted.
- Categories like "Restaurant", "Diner","Bistro" or"Buffet" are not specific enough. That is whay they were put to single catagory "Others" and were removed from main analysis, however the data was used to find information about restaurant per capita.
- "Eastern European Restaurant" and "Slovak restaurant" are clearly the same categories. They was be merged as "Slovak restaurant".
- Districts of Bratislava I, II, III, IV, V was merged to only Bratislava. It is not administratively correct, but it is not so important for the purpose of this analyses. The same approach was applied for districts of Košice I, II, III and IV.

Main algorithm of town selection for specific restaurant category

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

Score calculation

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

A = 0.2 (unemployment penalty)

B = 0.2 (wage penalty)

C = 0.1 (crime penalty)

E = 1 (restaurants per capita penalty)

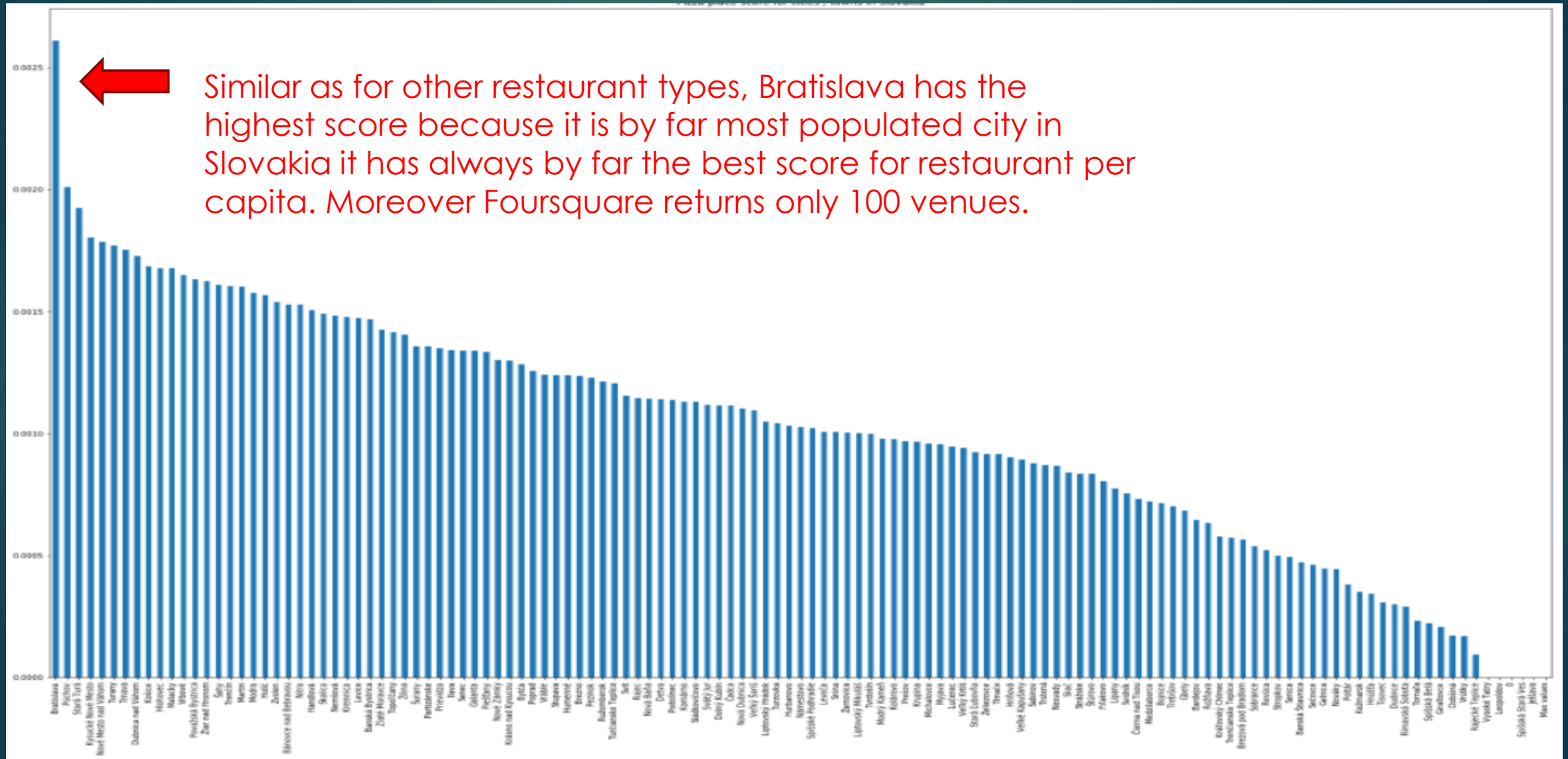
F = 1 (specific restaurant category per capita penalty)

Highest score for town means best place for particular restaurant category.

Feature scaling

- All features are scaled between 0 and 1
- 0 is not representing minimum of the feature but actual 0 value.
- 1 is for demografic 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

Results for Pizza Places



Conclusion and possible improvements of this method

- 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.