NEIGHBOURHOOD CULTURAL OPPORTUNITIES INDEX

COURSERA - IBM Applied Data Science Course

Peer-graded Assignment:
Capstone Project - The Battle of Neighborhoods

September 2019

By
Anatolie Poiata

ABSTRACT

This report describe the realization of the Capstone project “The Battle of Neighborhoods” of Coursera course “IBM Applied Data Science Course.

The report consist of two parts, corresponding to 2 weeks.
Part 1:

- A description of the problem and a discussion of the background.
- A description of the data and how it will be used to solve the problem.

Part 2:

- Methodology section which represents the main component of the report where is described exploratory data analysis applied, and used machine learnings methods
- Results section with discussion of the results
- Discussion section with observation during execution and recommendations based on results
- Conclusion section

The final report in Notebook format, with code and data can be found on Github repository at

https://github.com/AnatolPoiata/Capstone-Report

1 INTRODUCTION

Business Understanding

More and more cities trying to identify and evaluate their cultural potential in order to attract more tourists and visitors. Historically, in the same city, different neighborhoods have different number of cultural objects. Some of these attractions are popular, some are not. Thanks to social networks and mobile devices, we know where and when people are active within neighborhoods.
In 2005 by Toronto City Council addopted Toronto Strong Neighbourhoods Strategy (TSNS) 2020. The City uses 140 social planning neighbourhoods for designing programs and services. Each of the City's social planning neighbourhood has at least 7,000 residents and respects important boundaries such as major roads and rivers. More information about these neighbourhoods at http://www.toronto.ca/demographics/neighbourhoods.htm.
The scope of this report is to help to identify neighborhoods with lowest rating of cultural attractions and to recommend a list of these Priority Neighbourhood Areas for Investment (PNIs). The rating of the neighborhoods will be based on number of venues in the neighborhoods per capita, as well as the popularity of venues, according to data on Foursquare. Rated neighborhoods will be clustered and the recommendation for investment will be done for clusters where neighborhoods are close located, as new attractions affect not only neighborhood where is located, but also surrounding neighborhoods.

Analytic Approach

According to the question formulated above we should organize neighborhoods into similar groups based on their ranking scores of cultural potential and popularity. This is a typical K-meaning clustering problem of machine learning. It examines the entire set of interdependent relationships to discover the similarity relationships between the objects in order to identify the clusters.

2 DATA SOURCE

2.1. TORONTO NEIGHBOURHOOD

Neighbourhood refers to the City of Toronto's 140 social planning neighbourhoods. The boundaries of these social planning neighbourhoods are described in GEOJson format and can be uploaded at

https://open.toronto.ca/dataset/neighbourhoods/

Data are located also at
https://github.com/AnatolPoiata/Capstone-Report/blob/master/Neighbourhoods.geojson

2.2. TORONTO NEIGHBOURHOOD PROFILES

In these profiles, "neighbourhood" refers to the City of Toronto's 140 social planning neighbourhoods. The boundaries of these social planning neighbourhoods are consistent over time, allowing for comparison between Census years. We will use the most recent data which refer to 2016, limited to census data . As we will use index calculated as number of objects per capita, only data regarding population will be used.
Data can be uploaded at
https://open.toronto.ca/dataset/neighbourhood-profiles/

Data in .csv format are located also at
https://github.com/AnatolPoiata/Capstone-Report/blob/master/neighbourhood-profiles-2016-csv.csv

Neighbourhoods from mentioned above databases can be linked by common key Neighbourhood Number

2.3. ATTRACTIONS

The data for attractions will be used from Foursquare API. As we are interested in cultural attractions we will use only 3 categories of venues
- Arts & Entertainment
- Event
- Outdoor & Recreation

and 50 subcategories

For each neigbourhood will be identify the list of attractions from Foursquare, located in this neighbourhood, as well as the rating of the venue.