

Tbilisi cafeterias, restaurants and food courts segmentation and clustering

1. Introduction

1.1. Background

Strategic business approaches are based on reasoning and motivations behind actions. Data science gives entrepreneurs opportunities to make decisions while hedging risks. Big data for business offers an evolution of management and marketing tools as a reflection of innovation advancements. Data science techniques can uncover patterns about client profiles, their preferences, the efficiency of advertising methods and even projected ROI for each activity field. Requirements of each customer are in demand, and the only reasonable approach to know what they want is to use lots of data. There are some tools like www.api.foursquare.com that offer rich data according to client interests. By looking carefully at the data, it is possible to identify new niches to target.

1.2. Problem

Data that might contribute to determining food court place in Tbilisi might include geographical location of court, client profiles, paying capacity etc. This project aims to find reasonable segmentation and clusters to define leverage for efficient business.

1.3. Interest

Obviously, food court startup entrepreneur would be very interested in accurate recommendations, for competitive advantage and business values.

2. Data acquisition and cleaning

2.1 Data sources

The cafeterias, restaurants and food courts data can be found by using web scraping techniques. Data segmentation and clustering problems were discussed in the frame of unsupervised learning approaches. Geocoder.google function for data acquisition were used.

2.2 Data cleaning

Data are scraped from multiple sources such as <https://www.gps-coordinates.net/gps-coordinates-converter> and www.api.foursquare.com were combined into one table. Data were enough clean. I

have found only couple of misprinting in Tbilisi post services addresses, which were corrected manually.

2.3 Feature selection

As a features I have taken all companies which are responsible for food services: food courts, restaurants, fast foods etc. By carefully analyzing data, it is possible to validate decisions. Just having data is not enough; we have to transform it into an easy-to-use format, like a dashboard. Easy-to-use format was provided by python pandas library opportunities.

3. Exploratory Data Analysis

3.1 Segmentation & Clustering

Segments divided the cafeterias, restaurants and food courts. Each segment was made around the Tbilisi post service centers with $r=3\text{km}$ radius. Then the clustering problem with higher ratings was discussed. In the frame of K-means algorithm clustering problem has been solved, too.

4. Conclusion

In this study, I analyzed the clustering and segmentation problems for Tbilisi cafeterias, restaurants and food courts.

5. Future directions

Models in this study mainly focused on cafeterias, restaurants and food courts features. However, interactions with client feedback, might also contribute to efficiency improvement. NLP techniques applications will be able to design models that are more efficient.