Clustering Restaurants in Lisbon Neighbourhoods

Gustavo Bakker

February 13, 2020

1.
Introduction

1.1 Background Lisbon is the capital and the largest city of Portugal, with it's population of about 500 thousand people living in the main area and about 2.8 million people in the metropolitan area, is the 10th most populous urban area of the European union. The city has not only native portuguese people but a lot of emigrant people coming from around the world, being a multicultural city that offers different kinds of events, restaurants, shops, concerts, museums,etc.

1.2 Problem There are tons of restaurants in Lisbon and they are of different kinds and from different cultures. In this project we will see how certain types of restaurant and its food will vary in different neighbourhoods, with the goal of giving understanding of how this can vary with the variance of the location in a certain city. We will use a clustering technique to separate restaurants by categories in each area.

1.3 Interest/Stakeholders This will contribute to the city of Lisbon by providing a valuable information about the kind of restaurants that exist in a certain neighbourhood and also helping entrepreneurs who want to start a new business in town or even future applications that could use this information to recommend different types of food to people.

2. Data acquisition and cleaning

We will use the following resources to get our data:

1. **Lisbon City Neighbourhoods Dataset**

Source**:** https://pt.wikipedia.org/wiki/Lista_de_freguesias_de_Lisboa.

This is a table from Wikipedia (portuguese version), that has all the 24 neighbourhoods of Lisbon city, containing information about the population and area for each one.

2. **Foursquare API**

Source**:** https://developer.foursquare.com/docs

This is a location data provider consisting of a RESTful API that allows to retrieve information of venues of a certain neighbourhood using coordinates, giving us a rich JSON file with details about restaurants and their locations, provided by the neighbourhoods dataset of Lisbon.

3. Exploratory Data Analysis
4. Machine Learning
5. Conclusions
6. Future directions