

Clustering of Restaurants from Vancouver and Seoul

k- means Clustering

Lea Scharpf, | Business Analytics | 29.11.2022

# Introduction

Opening a restaurant in big capital cities can be quite a challenge. Not only good food needs to be saved in order to not go bankrupt, but also the people living there should value the food. To find the best location to open a restaurant depends on a lot of different factors. People in that area should earn enough money to be able to go to a restaurant. Next, if you want to open a restaurant with a cuisine from another country, some people from that country should live close to the restaurant, to be sure to have some customers. Also, it makes a restaurant more authentic, when people from that country enjoy the food. Third, there should not be too many competitors, since people than have to choose between your restaurant and others.

Therefore, we choose to cluster the restaurants from Vancouver and Seoul in order to find the best location to open a restaurant from a specific type of cuisine. The goal is to reduce business failures due to incorrect market analysis.

# Data import

To create such a clustering method, datasets from both cities including income, citizenship and a restaurant dataset are needed.

Kaggle provided the restaurant dataset for Vancouver:

* <https://www.kaggle.com/datasets/banaveenkumar/vancouver-restaurent-dataset>

The dataset for the income and the citizenship was found at Opendata Vancouver:

* <https://opendata.vancouver.ca/explore/?refine.theme=Demographics&disjunctive.features&disjunctive.theme&disjunctive.keyword&disjunctive.data-owner&disjunctive.data-team&sort=modified>

For the city Seoul the datasets were available on data Seoul:

* Korean dataset of restaurant:
  + <https://data.seoul.go.kr/dataList/OA-16094/S/1/datasetView.do?tab=S>
* Korean dataset of foreigner:
  + <https://data.seoul.go.kr/dataList/803/S/2/datasetView.do?tab=S>

# Preprocessing

After importing these data sets it was necessary to do preprocessing steps to clean datasets.

## Vancouver dataset

# Model and Training

## vancouver dataset