

Marcelo Diniz

08/01/2021

Applied Data Science Capstone Project

Choose location to focus a beverage distribution business in Madrid

Introduction

This paper intends to choose the best focus area for a local beverage distribution center located in Madrid Spain using data from internet and machine learning techniques.

As in every big city in Europe, it's quite difficult to transport and deliver products with rigorous time schedule due to heavy traffic, restrictions on trucks mobility and rare parking spots. Taking in account the highly competitive Madrid beverage business, focusing on a specific high profitable area is one of the most useful strategies to reduce logistics cost and ensuring client fidelity.

To decide where to focus the business three factors will be evaluated:

- Amount of bars
- Category of bars
- Average square meter price (which intends to indicate the average income of people living in surrounding area)

Data

The first information needed to initiate this analysis is a list of Madrid's neighborhoods name. This data will be gathered from a Wikipedia website

(https://en.wikipedia.org/wiki/List_of_neighborhoods_of_Madrid)

Afterwards, Geopy API will be used to get the exact coordinates of each neighborhoods.

Geopy API documentation can be found in <https://geopy.readthedocs.io/en/stable/>.

With the location in hands, it's time to look for the venues in each neighborhood.

Foursquared API will be suitable for this job. Official documentation in

<https://developer.foursquare.com/docs/>.

Finally, the average income of the area will be estimated by the square meter prices of places for rent in each neighborhood. This data will be mined from Idealista API, which documentation can be found in <https://github.com/esri-es/idealista-api>.