

The Battle of Neighborhoods

Content table

Introduction 2

Data 3

Introduction

Relocation has never been an easy decision for anyone. It usually takes time for researching with unstructured information on the Internet. Thanks to advancements in data science, many models and libraries are now available for building such systems that can help people shorten and prioritize the list of neighbourhoods of interest, hence reduce the task load and increase the efficiency.

In order to illustrate the problem, we will take a specific case as an example: Raul graduated recently from computer science on Jaen's University and looked for a data scientist position. He finally got an offer, however, the company bases in Madrid. He decided to relocate to Madrid for something new. The question here is that where exactly he could live in that big city.

One of the concerns of Raul is that he also loves his hometown so much and really wants something similar that could help him to be less homesick.

In order to determine the similarity, we will need to somehow describe each neighbourhood as a numerical vector then apply some machine learning technique, e.g., DBSCAN, to cluster them into different groups.

This problem is not restricted to this example but has wider applications in different situations. For example, if a restaurant or shop decides to open another branch somewhere on the other side of the city, they might also need to find a similar neighbourhood, because there are always interactive effects between stores. Finding similar neighbourhood has a wide application to several situations in the real world.

Data

FourSquare offers free APIs for developers to access their database of venues. Each venue in their dataset is usually categorized into a venue category, which is described in their [Developers Docs](#). There are 10 main categories; each includes subcategories which explicitly describe the venue, e.g., Sushi Restaurant or Fishing Store:

Categories	Number of subcategories
Arts & Entertainment	36
College & University	23
Event	12
Food	91
Nightlife Spot	7
Outdoors & Recreation	62
Professional & Other Places	41
Residence	5
Shop & Service	145
Travel & Transport	34
Total	456