Migrating to Spain – Exploring Similar Neighborhood

IBM Data Science Capstone Project Presentation

Problem Statement

- Migrating from Pune, India to Madrid, Spain
- Finding the similar neighbourhood venues in the destination city

Introduction

Migrating from one city to another is many a times a hectic process. New place, new people, new culture, and most importantly, new neighborhood. So exploring the new place is, thus, a new beginning from square one. It would really help one if he/she could find the amenities or restaurants or the venues just like the ones in their current location, in the city where they are migrating.

Here, I am assuming that I am migrating from my current city, Pune, India to city of Madrid, Spain. In this capstone, I will attempt to apply the techniques learned throughout the Data Science courses to explore the neighborhoods in the capital of Spain that is city of Madrid.

I will acquire my places of interest in my current location using the Foursquare API. I will then use the same API and explore the similar kind of venues in the city of Madrid.

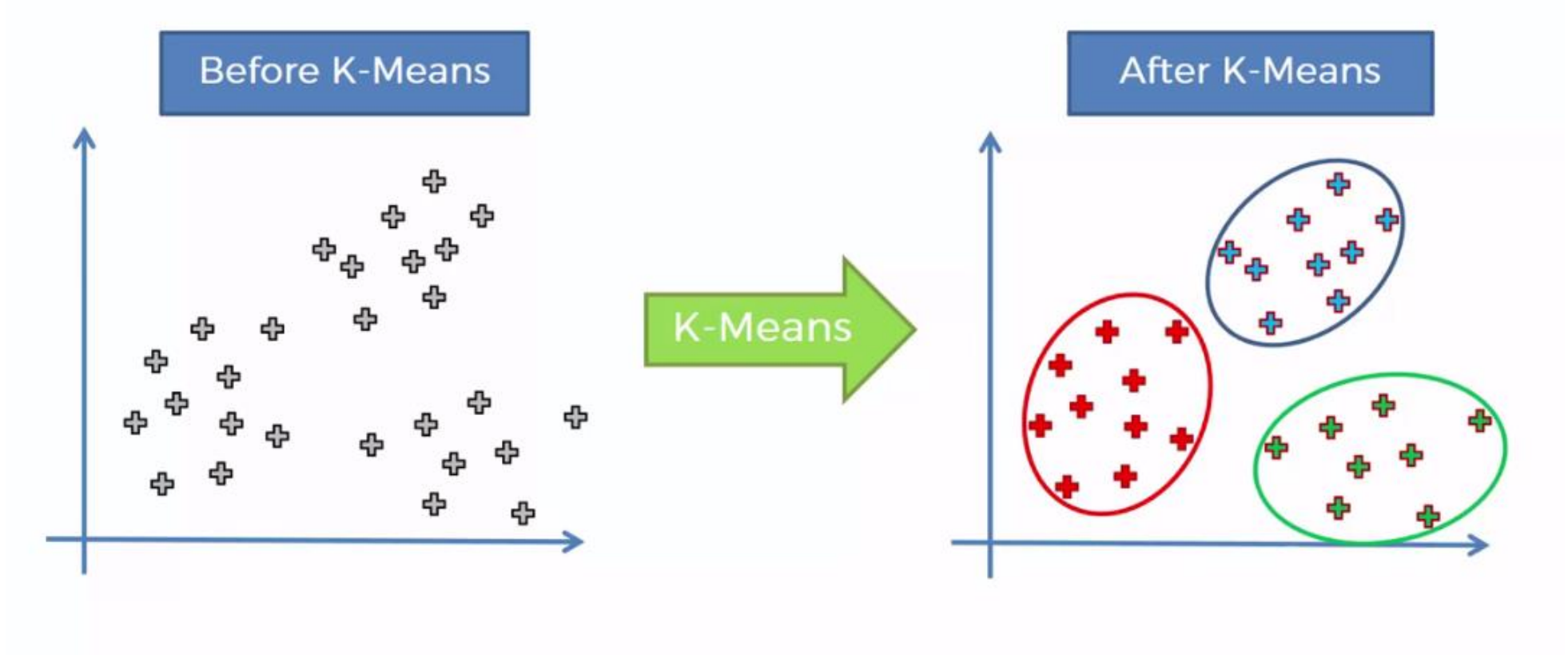
Methodology

Steps included

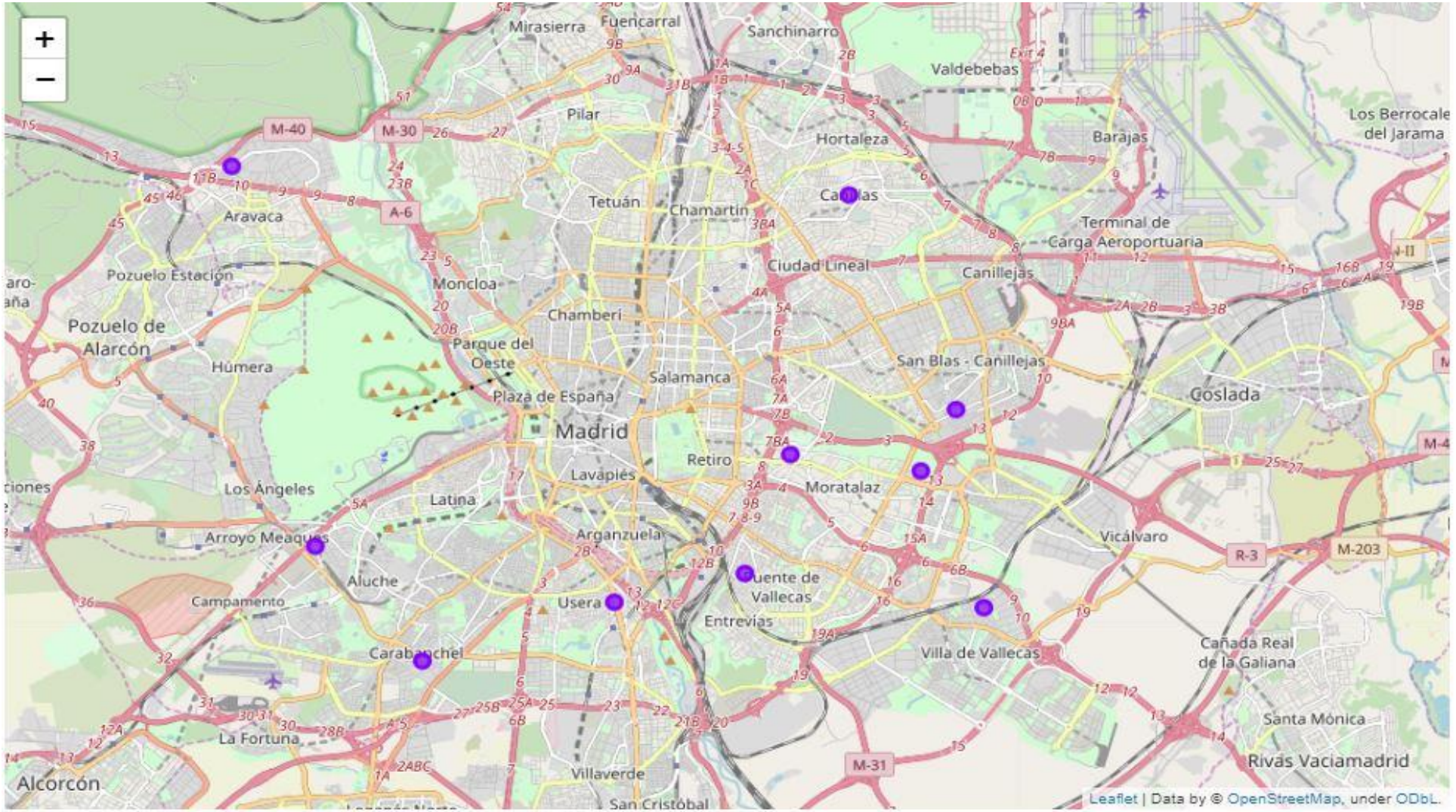
- Using Foursquare API to obtain venues from Pune
    - *Result: 47 rows for 27 categories*
- Use Foursquare API to obtain venues from Madrid
    - *Result: 3,446 rows for 254 categories*
- Group "venue categories" for each neighborhood using the mean.
- Using kmeans algorithm to create clusters of similar neighborhoods
- Finding closest neighborhoods to the current neighborhood using the distance to centroid from kmeans

K-means algorithm

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining



Explored Neighborhood (similar) in Madrid



Result and Conclusion

- It is possible to help people who have to move to another city and want to find conditions similar to those in their current residential location using public data available through the Foursquare API.The final decision cannot be based solely on the results of this analysis.
- One of the difficulties in kmeans algorithms is the choice of the value for K
- The characteristics that distinguish my neighborhood, according to the results of Foursquare, is the diversity of places to eat, shops and places to exercise. These same characteristics are present in almost all the selected neighborhoods