Closest Venues in the City

Enrique Gámez Flores
Coursera Capstone Presentation
2020-03-04

Introduction

Business Problem

- A tool to show us all the available groups of certain venues around a city or location
  - How many locations, in the city, can be found with:
    - selected types of commercial store, or,
    - a commercial store missing near some special ones

Audience

- Travel agencies
  - Promoting personalized experiences
- Entrepreneurs
  - Explore new locations for expanding their business
- Anyone
  - Planning new experiences

Data

User

Data provided by the user

Used only at the end for final selection

Foursquare Places

For Venue Discovery

Two required features already provided by the API in one single call

- Geographical coordinates, and
- Categorized

for each venue retrieved

Data Cleaning

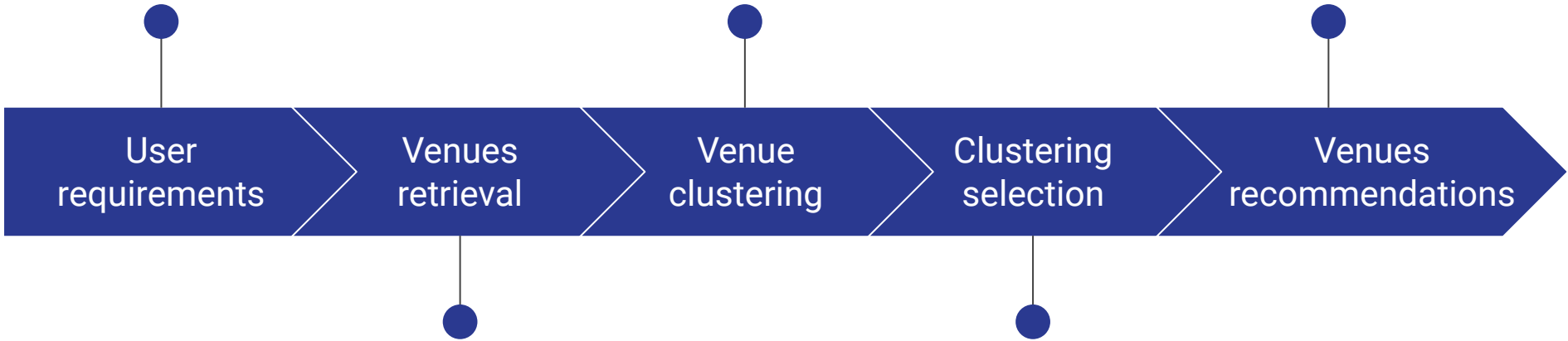
No Data Cleaning

- Geographical coordinates passed directly to the clustering algorithm
- Clustering selection made by user requirements basis

The final user enters the location and the categories to search for

Run KMeans with venue's geographical coordinates as features

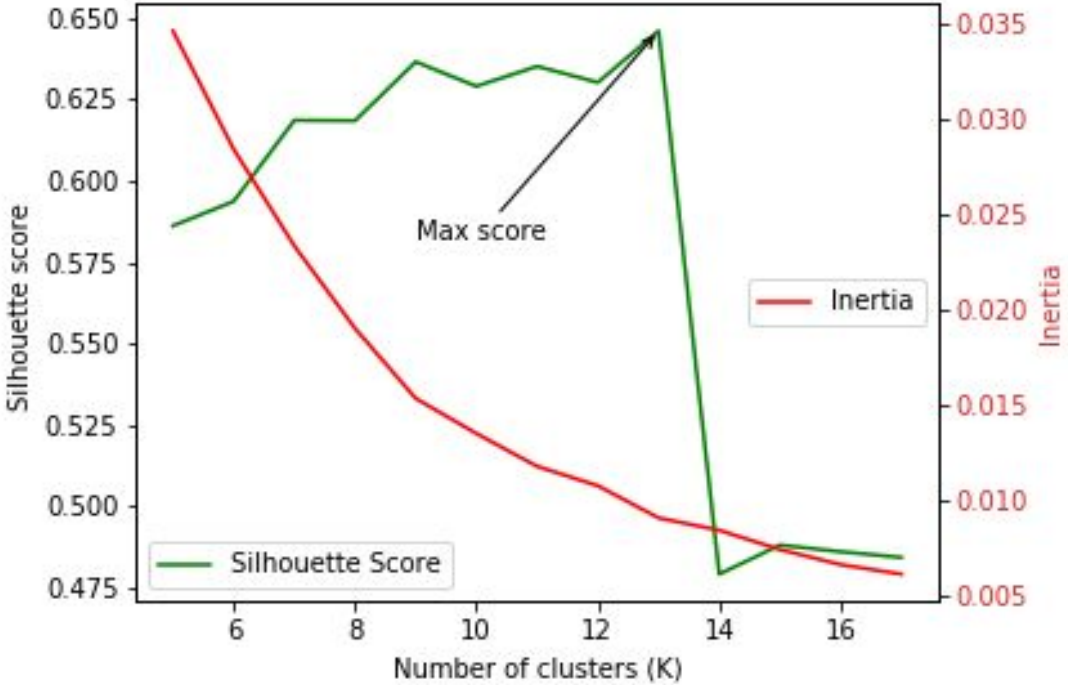
We are able to make recommendations



With the provided information we use the Foursquare Places API for venue discovery

According to user's requirements

Choosing the number of clusters

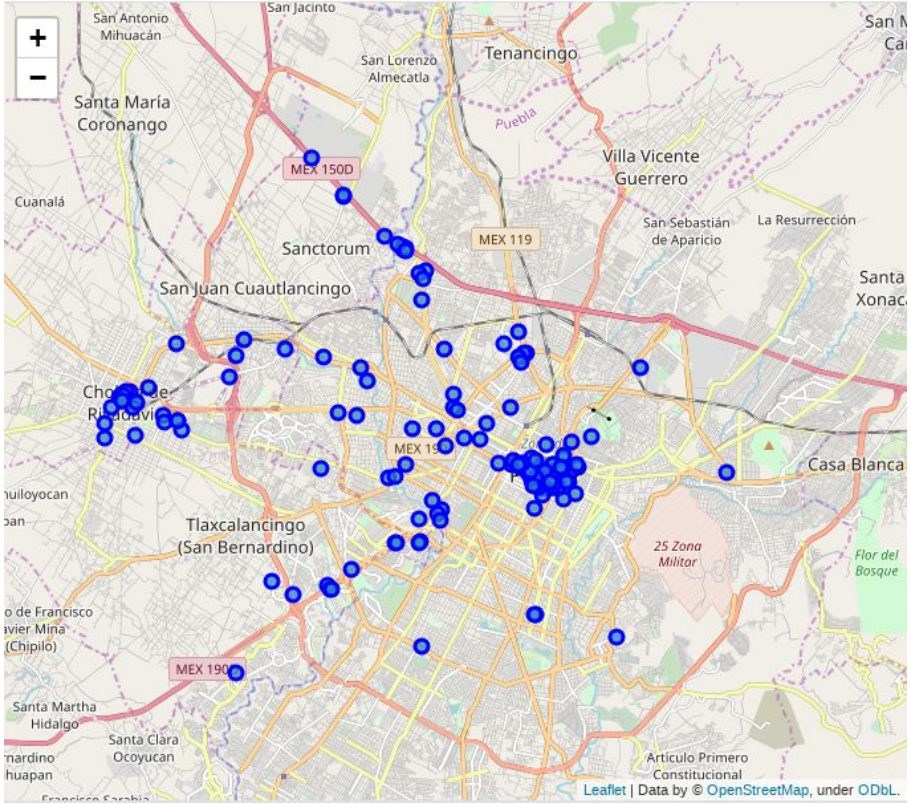


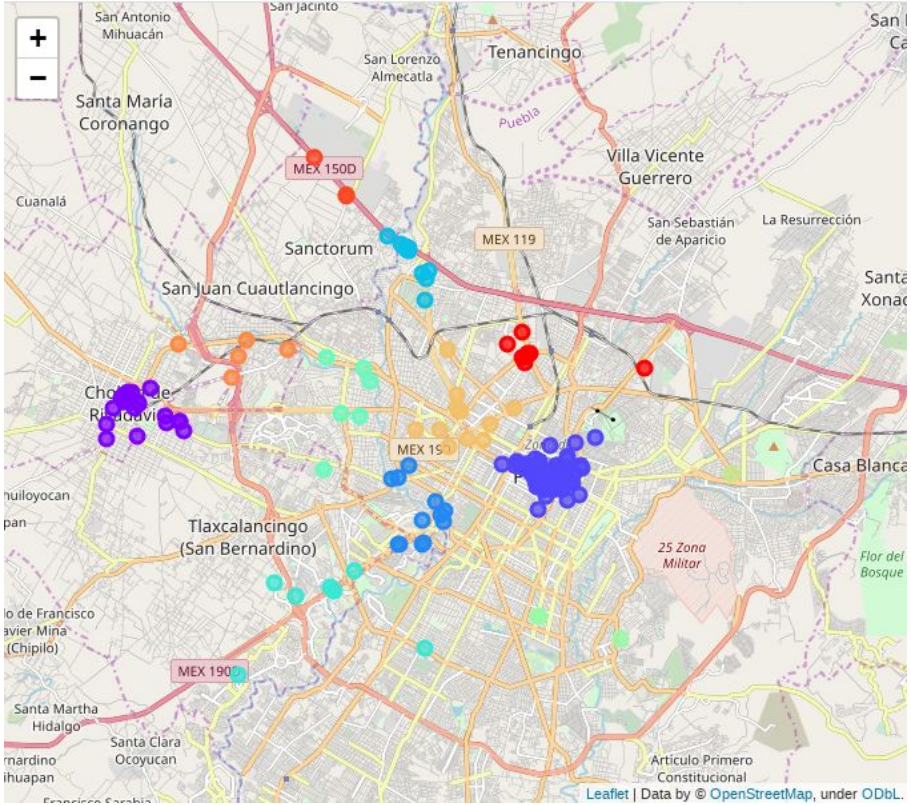
The Silhouette Score, for different number of clusters was selected as a parameter to choose the better number of clusters.

In the image, the Silhouette score is plotted against the number of clusters.

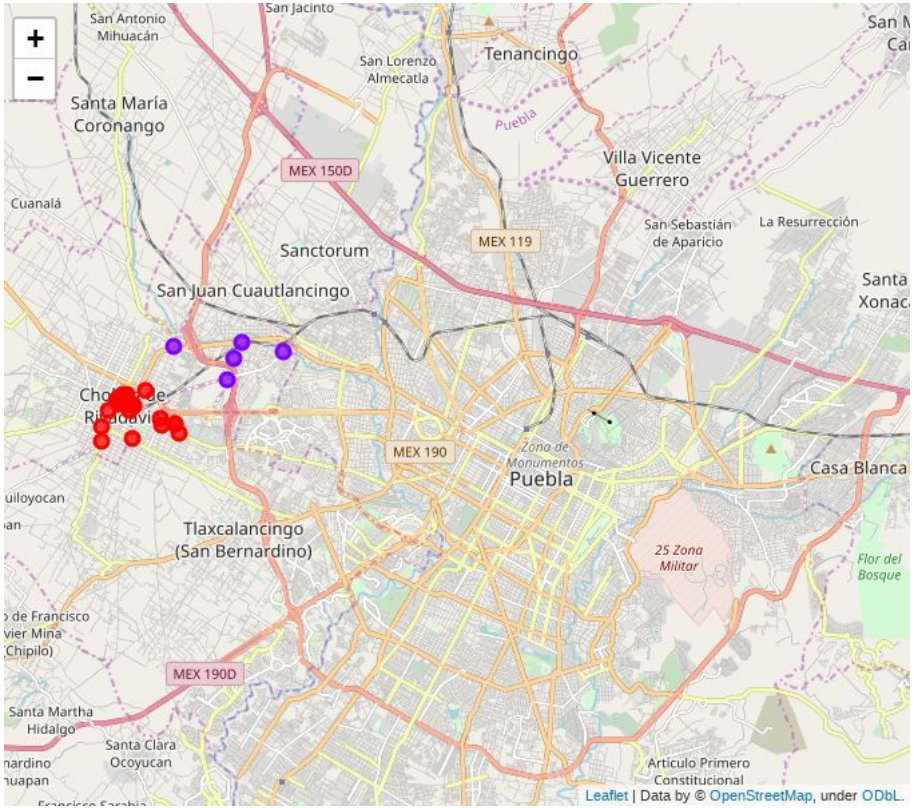
Geographical clusters

(before and after clustering)





Selected clusters



The cluster in **red** has at least one venue of all the categories required. While only two type of venues are cluster in **purple**

Conclusion

- The use of machine learning algorithms, like KMeans in this case, which performs not so trivial tasks effectively, enable us to focus on the specifics of the problem.
- Along with mentioned algorithms, the environment to analyze data (scikit-learn, numpy, pandas, matplotlib) give us the opportunity to deliver recommendations that time ago can only be possible by human inspection.