

COURSERA
IBM CAPSTONE PROJECT

STUDY OF LAUSANNE NEIGHBOURHOODS

Lavinia Saccoccio
Data Scientist
Sapienza University of Rome , EPFL

INTRODUCTION - BUSINESS CASE

For this project, I will introduce a hypothetical business case.

Imagine that you are a business man/student/retired couple who wants to move to the centre of Lausanne for personal or job reasons. But before moving in you want to find the perfect place to live, so that all your needs are fulfilled. Let's say you are a young worker, so you want to find a place near pubs and restaurant, close to the heart of nightlife of the city.

The idea is to study the city of Lausanne in order to find a suitable place to live. The assumption behind the analysis is that we can use unsupervised machine learning (in particular the K-means algorithm) to create clusters of venues that will provide us with a list of areas for consideration.

DATA

The data used to solve this problem is geolocation data collected from the FourSquare API. There is no data for the neighbourhoods in Lausanne, so I decided only to use the ones I got from the API.

The data will be structured in one dataframe, which will contain all the data we will need for the analysis. The dataframe is a simple one, and will contain only the necessary data, such as the name of the venue, the category and the geospatial coordinates (so that we can plot them on a map with the folium library)

	name	categories	lat	lng
0	Sleepy Bear Coffee	Coffee Shop	46.515338	6.631369
1	Les Trois Rois	French Restaurant	46.515515	6.631223
2	Pasta e Sfizi	Italian Restaurant	46.516374	6.633477
3	Café du Simplon	Mediterranean Restaurant	46.515961	6.629948
4	Bar Tabac	Bar	46.518066	6.634775

Figure 1: Example of dataframe present in the study.

Regarding the analysis we will use K-Means, un unsupervised machine learning algorithm to cluster the different kind on venues retrieved by Foursquare. Before that, we will try to find the best number of cluster with the elbow method, and then we will proceed with the study.