The Battle of Neighborhoods

Coursera Applied Data Science Capstone - Final Project

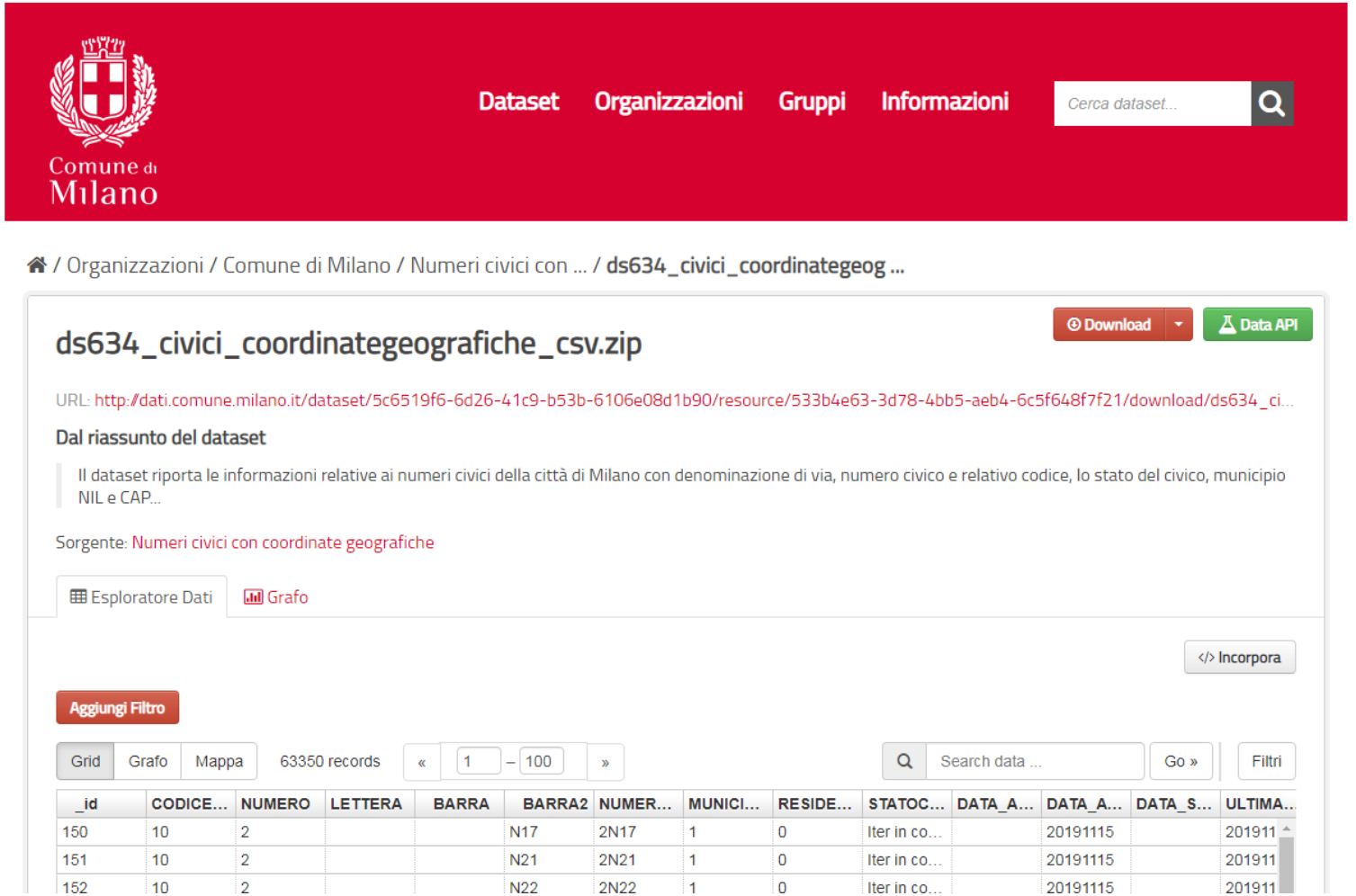
Enrico Palmieri

Introduction and Business Problem

- Opening a new wine bar in Milan
- Location to be determined based on neighborhoods characteristics and existing venues
- Use of K-Means clustering

Data

- Data about neighborhoods provided by Milan Municipality website

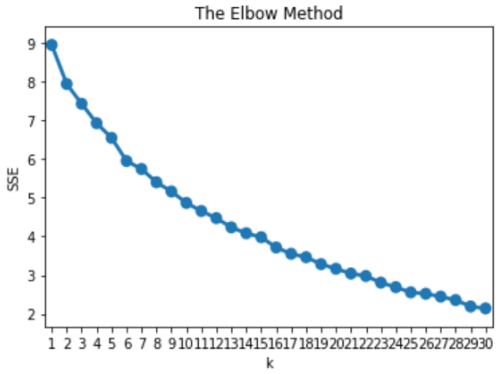


Data

- Data about venues provided by Foursquare

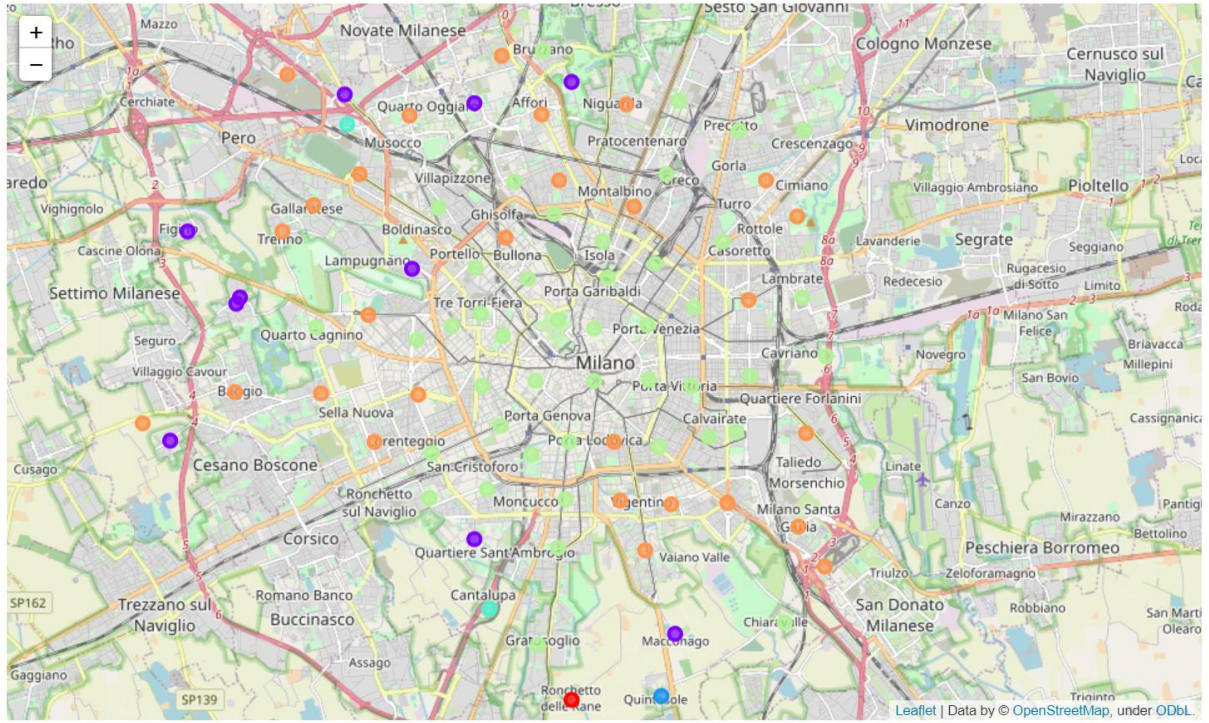
Methodology

- Data scraping, cleaning and filtering in Jupyter Notebook by suitable libraries
- K-Means Clustering



Methodology

- Data visualization by means of Folio library



Results

- Following patterns are identified:

Cluster #	Description
Cluster 1	Pizza and Café
Cluster 2	B&Bs
Cluster 3	Playgrounds
Cluster 4	Shops
Cluster 5	Hotels
Cluster 6	Restaurants

Results

- Most suitable cluster of nighborhoods for our purposes is Cluster 6

Conclusions and further developments

- Further analysis will have to be carried out:
  - market analysis to understand demographic parameters;
  - financial statements of existing wine bars;
  - rental costs analysis.

- Anyway this analysis gives back from an unstructured machine learning algorithm a clear business recommendation about where a new wine bar should be opened