

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



Dataset Organizzazioni Gruppi Informazioni

/ Organizzazioni / Comune di Milano / Numeri civici con ... / ds634_civici_coordinategeog ...

ds634_civici_coordinategeografiche_csv.zip

Download Data API

URL: http://dati.comune.milano.it/dataset/5c6519f6-6d26-41c9-b53b-6106e08d1b90/resource/533b4e63-3d78-4bb5-aeb4-6c5f648f7f21/download/ds634_ci...

Dal riassunto del dataset

Il dataset riporta le informazioni relative ai numeri civici della città di Milano con denominazione di via, numero civico e relativo codice, lo stato del civico, municipio NIL e CAP...

Sorgente: Numeri civici con coordinate geografiche

Esploratore Dati Grafo

Incorpora

Aggiungi Filtro

Grid Grafo Mappa 63350 records « 1 - 100 »

Go » Filtri

_id	CODICE...	NUMERO	LETTERA	BARRA	BARRA2	NUMER...	MUNICI...	RESIDE...	STATOC...	DATA_A...	DATA_A...	DATA_S...	ULTIMA...
150	10	2			N17	2N17	1	0	Iter in co...		20191115		201911
151	10	2			N21	2N21	1	0	Iter in co...		20191115		201911
152	10	2			N22	2N22	1	0	Iter in co...		20191115		201911

Data

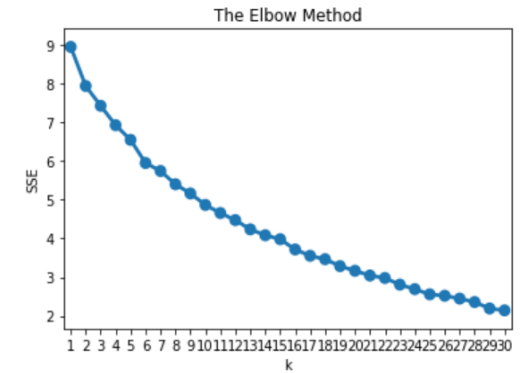
- Data about venues provided by Foursquare



FOURSQUARE

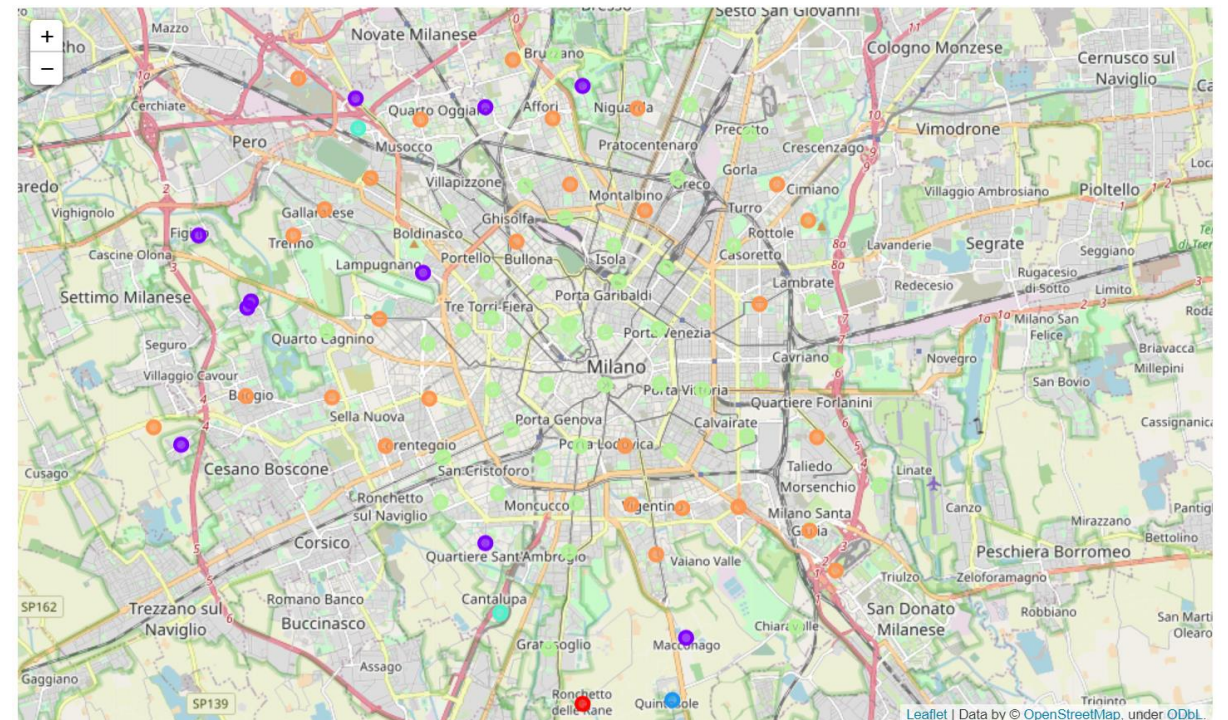
Methodology

- 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 neighborhoods 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