IBM / COURSERA
Applied Data Science Capstone Project

The Battle of Neighborhoods: Clustering Milan Real Estate market for students decisions support

BUSINESS PROBLEM

Background
According to the main financial news firms, the Housing Market in Milan (Italy) has been facing for the last five years a growing trend in the households selling prices as well in the rental ones.
A further prices acceleration is expected in 2020 due to the "Brexit", which will probably result in a shift of real estate investments from London to the other UE cities.

Business Problem
In this scenario, many students who intend to pursue their university studies in Milan need to know more about the real estate market, in order to correctly choose the neighborhood where to live according to the available public services and venues and to the households buying or rental prices.

SOLUTION

For this purpose it is possible to run machine learning models (clustering) in order to assist students to make an informed decision about potential neighborhoods where to live.
Hence, we clustered Milan neighborhoods in order to recommend amenities, facilities and buying/rental households prices.
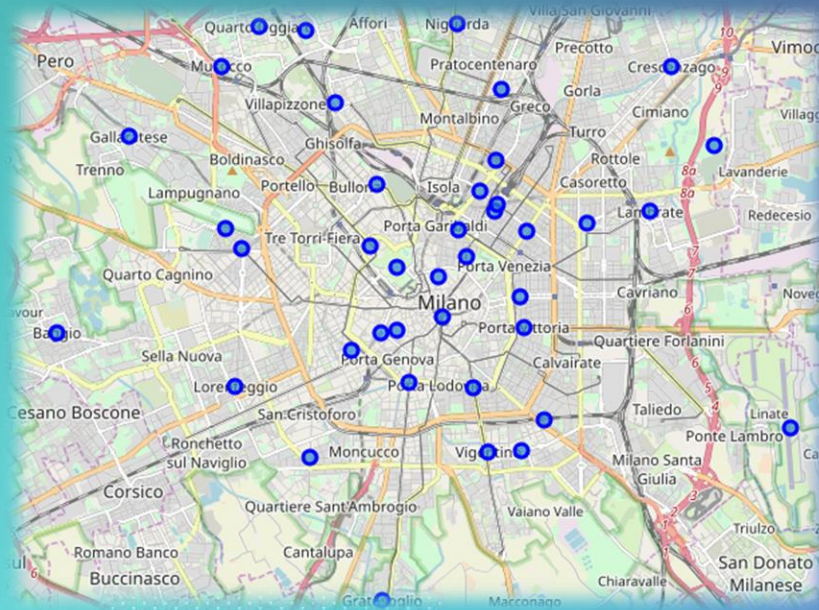
DATA AND METODOLOGY

Data section
Data on Milan neighborhoods and current real estate market are available at:

https://www.agenziaentrate.gov.it/portale/web/guest/schede/fabbricatiterreni/omi/banche-dati/quotazioni-immobiliari/download-gratuito

To explore and target recommended locations we extracted venues using FourSquare API interface.

Merging data on Milan real estates market and venues, we can support the students decision process.

The Methodology section describes the main components of our analysis system and is structured in four steps, as follows:

- Data collection and inspection
- Data exploration and understanding
- Data preparation and preprocessing
- Data modeling

Milan kMeans Clustering
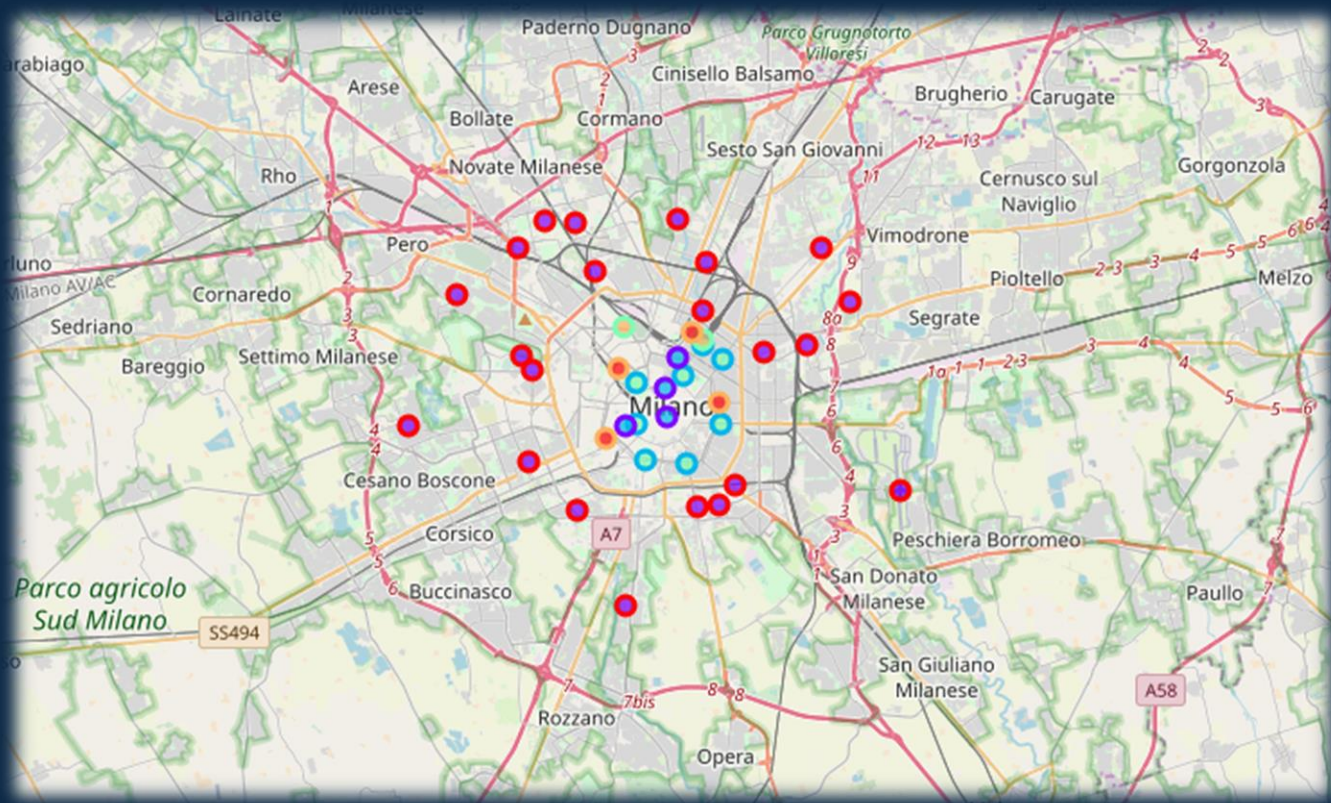
OUTCOMES

Analyzing the Milan neighborhoods and clusters we have identified, we can recognize three main patterns:

- the first pattern refers to Milan "Old town" and includes cluster 1 and 2; characterized by monuments, restaurants, cafes and boutiques, as well as high purchase and rental prices, these clusters might target students who loves amenities and a cultural life, provided they can afford the high cost of living

- the second pattern refers to Milan residential neighborhoods and includes cluster 3 and 4; characterized by several facilities as well as medium purchase and rental prices, this cluster might target students who want to live in stately residential neighborhoods well connected to the city centre

- finally, the third pattern refers to Milan suburban districts and includes cluster 0; characterized by ethnic amenities and lower purchase and rental prices, this cluster might target students who loves multicultural neighborhoods and want to pursue a more affordable lifestyle