Capstone project report

Contents

[Introduction 1](#_Toc21679639)

[Data used 1](#_Toc21679640)

[Data on Toronto neighbourhoods 2](#_Toc21679641)

[Analysis on Milan data 2](#_Toc21679642)

[Data acquisition and cleaning 2](#_Toc21679643)

[Clustering of Milan’s neighbourhoods 2](#_Toc21679644)

[Comparison to Toronto neighbourhoods 2](#_Toc21679645)

[Analysis of the distribution of price tier and rating variables concerning food venues in Milan 2](#_Toc21679646)

# Introduction

I'll be acquiring information regarding Milan's boroughs and will perform an analysis similar to what we've done with Toronto and NY. I'll then proceed to compare it to other cities. The goal is to get insight about how cities from very different countries compare and if their neighbourhoods can be distinguished by a clustering algorithm, depending on the criteria used. The analysis will provide information on how to distinguish between parts of a city, and how different kinds of neighbourhood (residential, touristic, financial...) might compare between cities.

As an example, I am going to use information about the most popular venues per neighbourhood, and use it to cluster them. Each neighbourhood popular venues blend, made of restoration, services and attraction venues in different proportions, can in fact give us insights about its geographical location and position relatively to tourism, business or services centres.

I am then going to carry out an analysis of Milan’s neighbourhood based on food venues rating and price tier. The goal is to correlate those two indicators with the neighbourhood location and nearby presence of tourism, business or services centres, possibly comparing it to the previous analysis.

This project is obviously limited in scope but it provides an interesting starting point for further work on the same data and topics.

# Data used

Milan data are not readily available as for NY or Toronto; basic information about Milan’s boroughs and neighbourhoods can be obtained from Wikipedia (<https://en.wikipedia.org/wiki/Zones_of_Milan>), and the spatial coordinates can be obtained through services as geocoder.

An alternative and more efficient solution is to use an institutional database, which is available online (<http://dati.comune.milano.it/dataset/5c6519f6-6d26-41c9-b53b-6106e08d1b90/resource/533b4e63-3d78-4bb5-aeb4-6c5f648f7f21/download/ds634_civici_coordinategeografiche_20190902_csv.zip>), and contains spatial coordinates of each and every single address in Milan and relative neighbourhood information. By averaging the coordinates for all addresses specific to one neighbourhood it is possible to get a good estimate of each neighbourhood centre location, especially given the size of the database. Location of the coordinates on the map confirms the correctness of the operation.

Information on the neighbourhood popular venues is then obtained through Foursquare API, whereas data on food venues ratings and price tier is available through Yelp API.

Data from Toronto from the previous part of the project course.

## Data on Toronto neighbourhoods

Data on Toronto has been acquired as part of the project course, in the previous module. Information about neighbourhoods and boroughs was scraped from the Wikipedia page <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> , and the coordinates through the geocoder package, or provided with the course material, depending on the correct functioning of geocoder. I have then acquired information about popular venues in central Toronto through the Foursquare API, and clustered the neighbourhoods according to their venues composition. The result is available on <https://github.com/RaffaToSpace/Coursera_Capstone/blob/master/Toronto_clustering_analysis.ipynb> , and shows how neighbourhoods in central Toronto are pretty homogeneous in composition, with some exceptions due to the presence of services or infrastructure (such as schools, parks, etc.) in some neighbourhoods.

# Analysis on Milan data

## Data acquisition and cleaning

## Clustering of Milan’s neighbourhoods

## Comparison to Toronto neighbourhoods

## Analysis of the distribution of price tier and rating variables concerning food venues in Milan