Clustering districts in downtown Milan

francesco.bellogi

April 2020

1 Introduction

1.1 Background

Milan is Italy financial capital and it's by far the most economically lively city in Italy. Milan is also particularly famous for its *fashion weeks* and for its *lyfestyle*. With around 1,3 millions inhabitants it's the second most populous city in Italy and also one of the most densely populated. People from all around the country move to Milan to chase a career and in the last year Milan has attracted a lot of people from all around the world.
Milan is a fast pace growing city in which people enjoy going out for dinner, for *aperitivo*, for shopping and that makes it an optimal place to open a restaurant, a shop, a club. On the other hand, not all districts are the same, some of them have been growing over the years and others have fallen behind. Many businesses are opening and many are closing.
Having a more clear picture of the current situation of the various in districts in downtown Milan can be extremely useful for anybody who is looking for business opportunities, for a family wanting to buy a house, for the local government having to distribute finances.

1.2 Description of the data

I built the dataset for this analysis from scratch, collecting pieces from different sources. I obtained the list of names of districts is Milan by scraping the relative Wikipedia page. The resulting list of names needed a lot of cleaning. Once cleaned I was able to obtain the geo-spatial coordinates for each districts using the python geopy package.
After that I collected for each district the first 100 venues in a radius of 300 miles through the Foursquare API. Following some more data wrangling I obtained my dataset ready for clustering.