Coursera Capstone Project

IBM Applied Data Science Capstone

Opening a truly Italian restaurant in Milan

Giuseppe De Nittis

1 Introduction

Food is one of the most important resources for mankind. For Italian people, it becomes even more important, bringing emotions and feelings with it. In fact, in Italy, having a meal is not just feeding ourselves with what we need, but it becomes a fundamental moment to speak, discuss and share ideas (often, about food different w.r.t. the one we are eating).

Thus, deciding to open a restaurant in Italy can result in a very complex decision. Location, type of food and style affect significantly the future of the restaurant, making these decisions crucial. The objective of this capstone project is to decide the best location to open a new restaurant in the city of Milan (Italy). Using data science methodology and machine learning techniques learned during the Specialization, the question I want to answer is the following: where would you open a new truly Italian restaurant in the city of Milan? Which district would you choose?

1.1 Audience

The audience is constituted by investors willing to open a new activity in one of the most important Italian cities, namely Milan. Despite the high number of restaurants already present in the city, unfortunately, a lot of them is just driven by the latest fashion, and recipes are very simple and immediate, just trying to attract non-experienced tourists. An investor would probably decide to open a new and classy place, a unique spot in the city, with high-level chefs and sommeliers. This project can provide her the answer with the best location for her new truly Italian restaurant.

2 Data

To solve the problem, I will need (and use) the following data:

- list of districts in Milan. This defines the boundaries of the project, which is restricted to one of the most important cities in Italy;
- latitude and longitude coordinates of those neighborhoods;
- venue data, specifically the ones related to Italian restaurants, needed to perform the neighborhood clustering.

2.1 Sources of data and methods to extract them

This Wikipedia page (`https://en.wikipedia.org/wiki/Category:Districts_of_Milan`) contains a list of districts in Milan, with a total of 76 districts. I will extract the data from the Wikipedia page by exploiting Python requests and the BeautifulSoup. Then, I will get latitude and longitude coordinates of

the districts using the Geocoder package. Then, I will use Foursquare APIs to retrieve information on the venues present in those districts. Such APIs provide many categories of the venue data: I am interested in the Italian restaurant category.