Coursera Capstone Project Report

INTRODUCTION

Rome is one of the most visited cities across the World and our intent, as a popular Italian Data Science company, is to develop a tool able to provide some information about the main Rome municipalities, in term of their tourist attractions.

In particular we want to develop a model that exploits the Foresquare location data to measure how the different district are similar or dissimilar in terms of touristic venues. By doing so, the tourist agency will be able to offer a customized and client-based tour across different roman neighborhoods just on the basis of their preferences (for instance, there are persons more interested in museum and cultural activities while other persons are more curious to taste local dishes or even go shopping).

DATA

One of the main steps related of our project is to accurately select and collect the data starting from which the model will be developed.

Fist, the location data of all Rome municipalities must be retrieved. Rome is composed by 19 municipalities (https://it.wikipedia.org/wiki/Municipi_di_Roma) and by knowing their name we can retrieve their geospatial coordinates with an ad hoc Python library.

Later these coordinates will be passed to the Foursquare API (https://it.foursquare.com/) and will be used to get information about venues in each neighborhood. For instance, the Foresquare API will return a bunch of information about each venue but we are interested only on the venue category. To limit the number of returned categories only some macro-categories will be investigated such as “food”, “art” and “shopping” and so on.

ORIGINAL FOURSQUARE API CATEGORY	MACRO CATEGORY CONSIDERED
BAKERY	***Food***
ART GALLERY	***Arts & Entertainment***
JAZZ CLUB	***Arts & Entertainment***
CAFÉ	***Food***
SHOPPING MALL	***Shop & Service***