

A new restaurant in Rome

Introduction

Rome is the largest city in Italy and it is one of the favourite touristic destinations for people all over the world. Apart from hosting an outstanding number of museums and other touristic attractions, Rome is home to a large number of restaurants.

Background

Restaurants' revenues are highly dependent on the number of clients, that tend to prefer certain neighbourhoods over others due to many reasons:

- public transport
- number of restaurants to chose from
- nearness to attractions
- others

Problem

An investor from Brazil would like to invest in Rome by opening several restaurants in the city.

The main idea is to open his Brazilian restaurants in areas where there is already a high number of people going out for lunch or dinner and offering them something different from Italian restaurants, that represent the majority in Rome.

Data (I)

The data used for the problem will be extracted from the Foursquare database and it will regard the venues in 35 neighbourhoods of Rome. In particular, the data imported from the database for each venue will be:

- name
- latitude
- longitude
- category

Data (II)

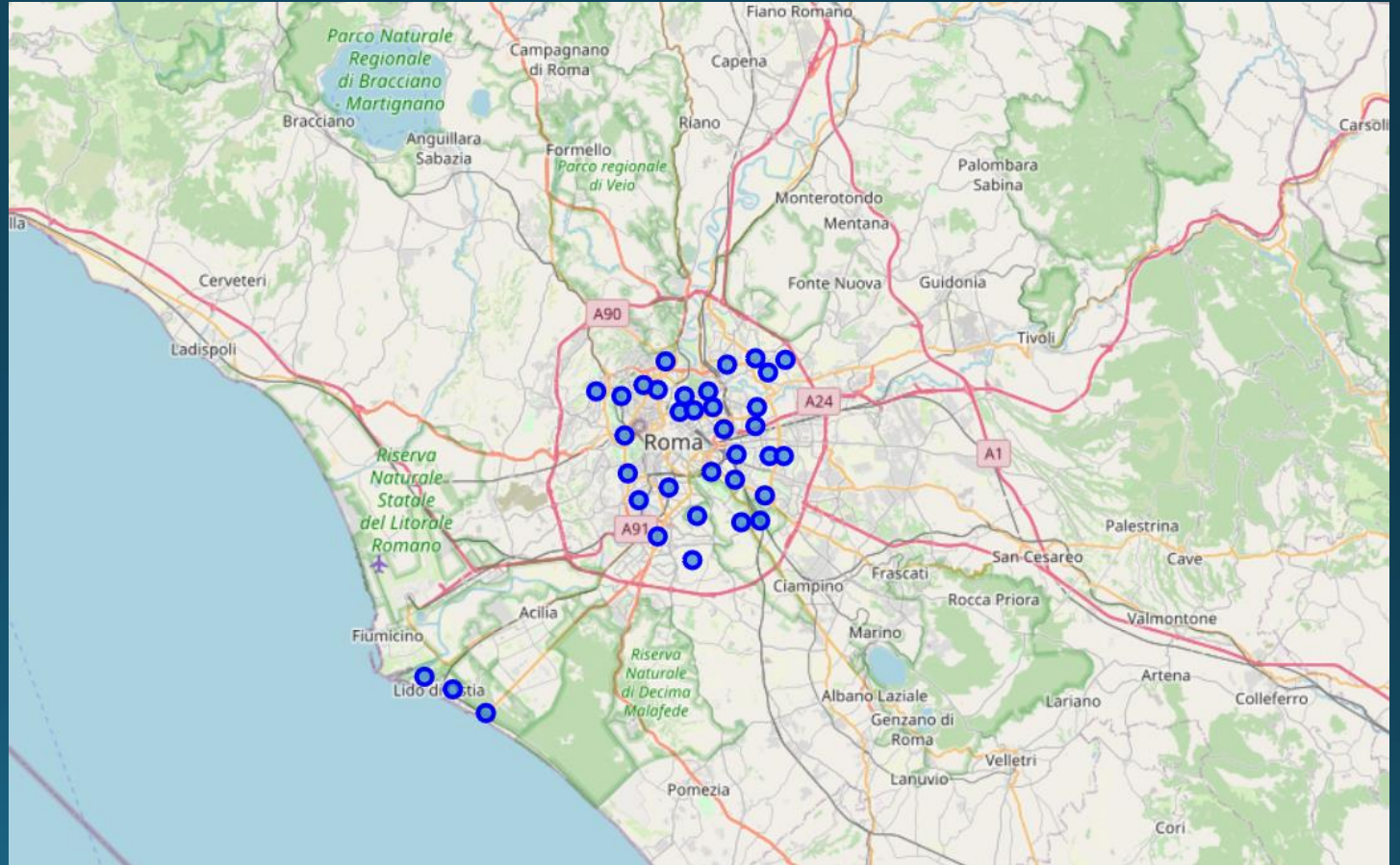
The neighbourhoods' data has been obtained from Wikipedia and includes the following information for each neighbourhood:

- name
- latitude
- longitude

Methodology - exploratory analysis

For each neighbourhood, a maximum of 200 venues in the radius of 1 km have been imported from the Foursquare dataset.

The 10 most common venues have been listed in descending order. From a first glance, restaurants and Italian restaurants, in particular, are very common.



The neighbourhoods analysed

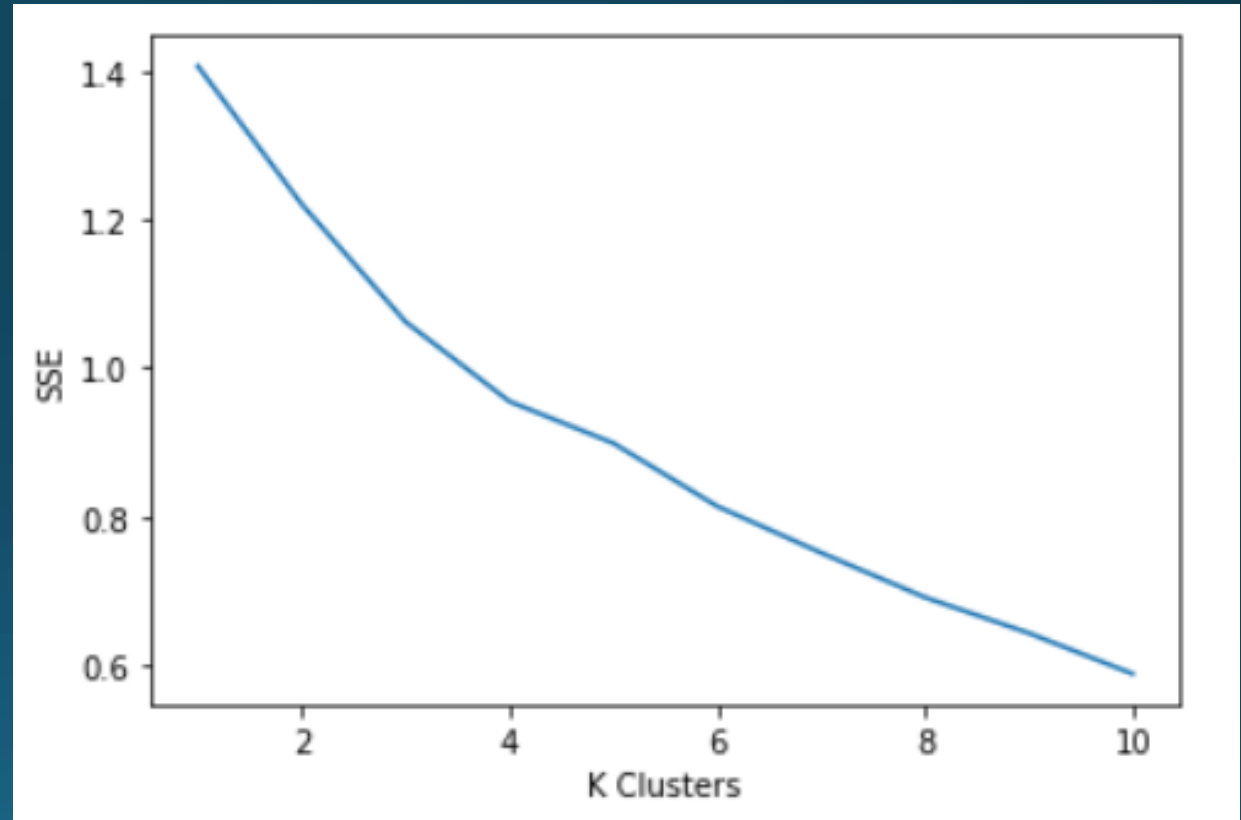
Methodology - clustering

In order to find similar neighbourhoods with the same lively atmosphere and a great number of restaurants a clustering approach will be used. In clustering the most important problem to tackle is the selection of the right number of clusters. Two approaches have been used to make this evaluation: the elbow method and the silhouette method.

Methodology – elbow method

In the elbow method we try to find the number of clusters for which a sharp change in the slope of the sum of squared errors (SSE) is observed.

Unfortunately, in this case this method is not very helpful since there is not an abrupt change in the slope of the SSE.

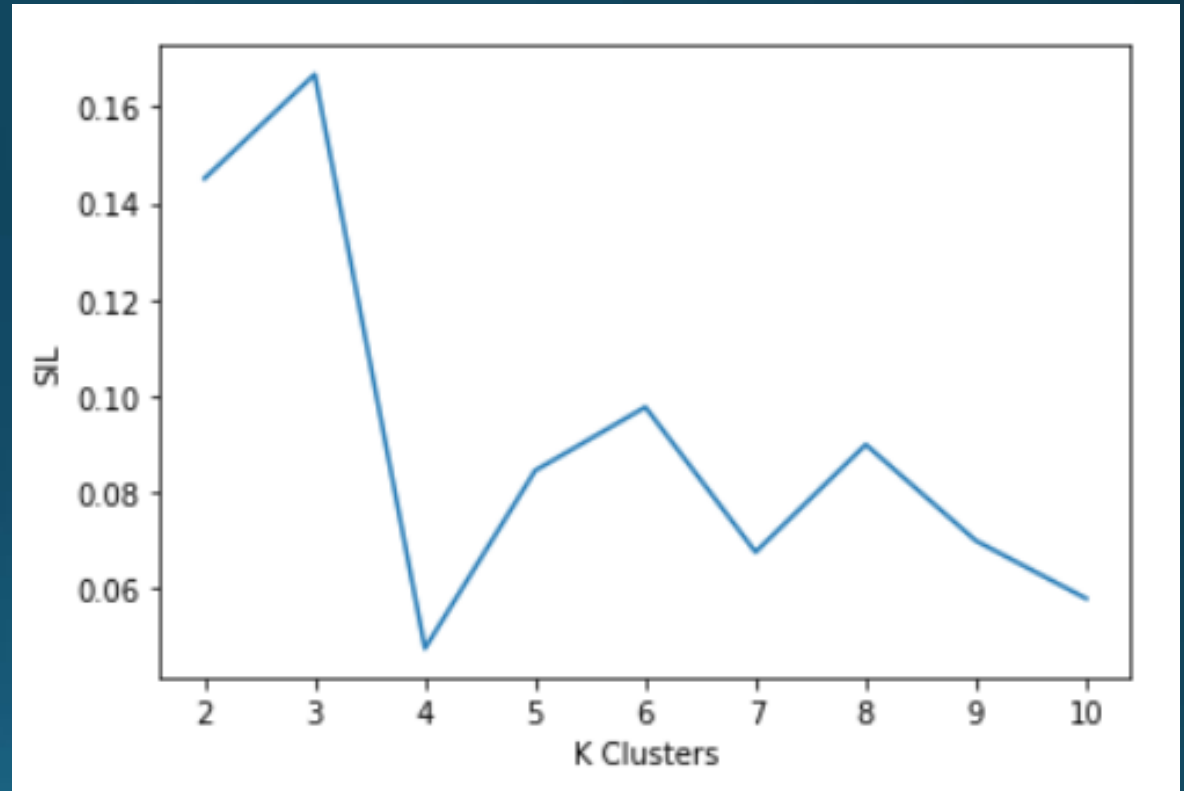


Sum of squared errors as a function of number of clusters K

Methodology – silhouette method

The silhouette method, that measures how similar a point is to its own cluster (cohesion) compared to other clusters (separation) with values ranging from -1 (worst) to +1 (best).

In this case, we see that the maximum values are reached for 3, 2 and 6 clusters. We exclude $k=2$ and $k=3$ since they appear to low and they will not provide many insights. Therefore, we select $k = 6$.

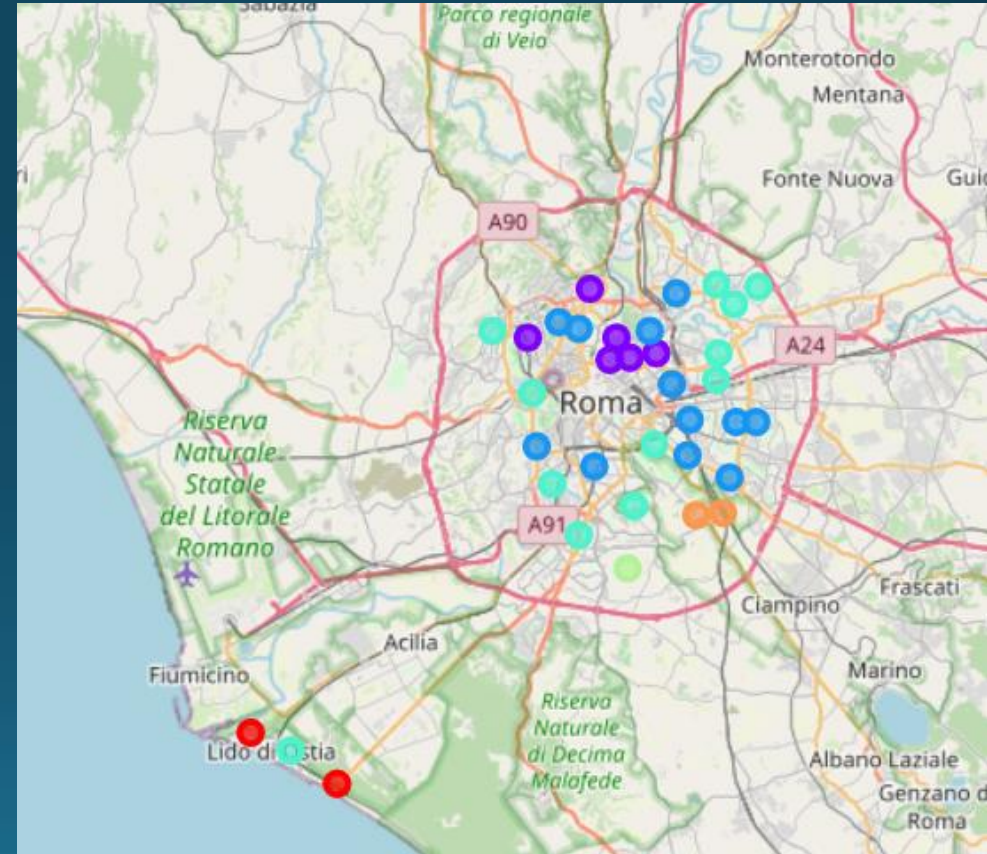


Silhouette score as a function of number of clusters K

Methodology – clusters

The six clusters identified by the algorithm can be categorized as follows:

- Cluster 1 – Beaches
- Cluster 2 - Italian restaurants and hotels (centre-north of Rome)
- Cluster 3 - Italian restaurants and pizza places
- Cluster 4 – Cafés
- Cluster 5 - Hotels and Italian restaurants (suburbs of south Rome)
- Cluster 6 - Hotel and steakhouses



The neighbourhoods analysed grouped in six clusters

Results

The cluster that appears as the most suitable for the Brazilian investor to open his restaurants is the cluster 2, composed of the following neighbourhoods:

- Q. II Parioli
- Q. III Pinciano
- Q. IV Salaria
- Q. V Nomentano
- Q. XIV Trionfale
- Q. XVIII Tor di Quinto

Discussion

The selected neighbourhoods are located in the centre-north of Rome and are considered to be fairly trendy and expensive. It is no surprise that they host a high number of restaurants and hotels.

A further analysis, out of the scope of this study, would be to consider the following aspects:

- rent or cost to acquire a place where to open a restaurant
- nearness to bus stops or metro stations
- preference for places that are more popular during lunch time or dinner time or both.