DATA SCIENCE CAPSTONE PROJECT

Luis Cespedes

March 19, 2019

1. Introduction

1.1 Background

When someone travel to work, do tourist, visit another person, etc. One of the most important decision is to choose the place where he/she will stay. If is a business travel maybe the best choice is a place near to work and some specific venues, if is to do tourist maybe the best choice is a place near to tour agencies or restaurants. In this case I will focus in hotels in Rome.

1.2 Problem

Let assume that you work in a travel agency and one day a client come to your services, because he want travel to Rome, Italy to do tourist and need information about hotels near to restaurants, tourist agency and more places to enjoy his journey. Which are the best groups of hotels in Rome, Italy that you could recommend to your client according his specifications?

2. Data acquisition

2.1 Based on the business problem, factors that will influence visitor to choose are:

1) Number of existing facilities around hotel.
2) Type of existing facilities around each hotel.

Following data sources will be needed to generate the proper decision:

1) Number of existing facilities and their type and location in every areas around each hotel will be obtained using Foursquare API

2.2 Explore Data

As soon as the business problem is defined, we need to create our Dataframe of hotels in Rome, to do that we will use Foursquare API to search any hotel in an area with 1.3 km of radius after that we will normalize the json file an create the pre-Dataframe which have 50 rows and 19 columns.

```
Index(['categories', 'hasPerk', 'id', 'location.address', 'location.cc',
       'location.city', 'location.country', 'location.crossStreet',
       'location.distance', 'location.formattedAddress',
       'location.labeledLatLngs', 'location.lat', 'location.lng',
       'location.neighborhood', 'location.postalCode', 'location.state',
       'name', 'referralId', 'venuePage.id'],
      dtype='object')
```

Fig.01 Columns of the pre-Dataframe

Then we will filtered and finally we will use only the columns that could help us to achieve our goal, which are "name","categories","lat" and "lng".

	name	categories	lat	lng
0	Hotel Forum	Hotel	41.893840	12.487730
1	Hotel Pace Helvezia	Hotel	41.896887	12.484984
2	Hotel Apollo	Hotel	41.896701	12.490097
3	Hotel Esposizione	Hotel	41.899328	12.491482
4	Hotel Hassler Roma	Hotel	41.905930	12.483732

Fig.02 Dataframe of hotels in Rome (40 rows and 4 columns)

Using Folium library, we can also be used to visualize geographic information of all these hotels. And I created a map of Rome with the hotels base on with the latitude and longitude values get from our data to the visual as below:
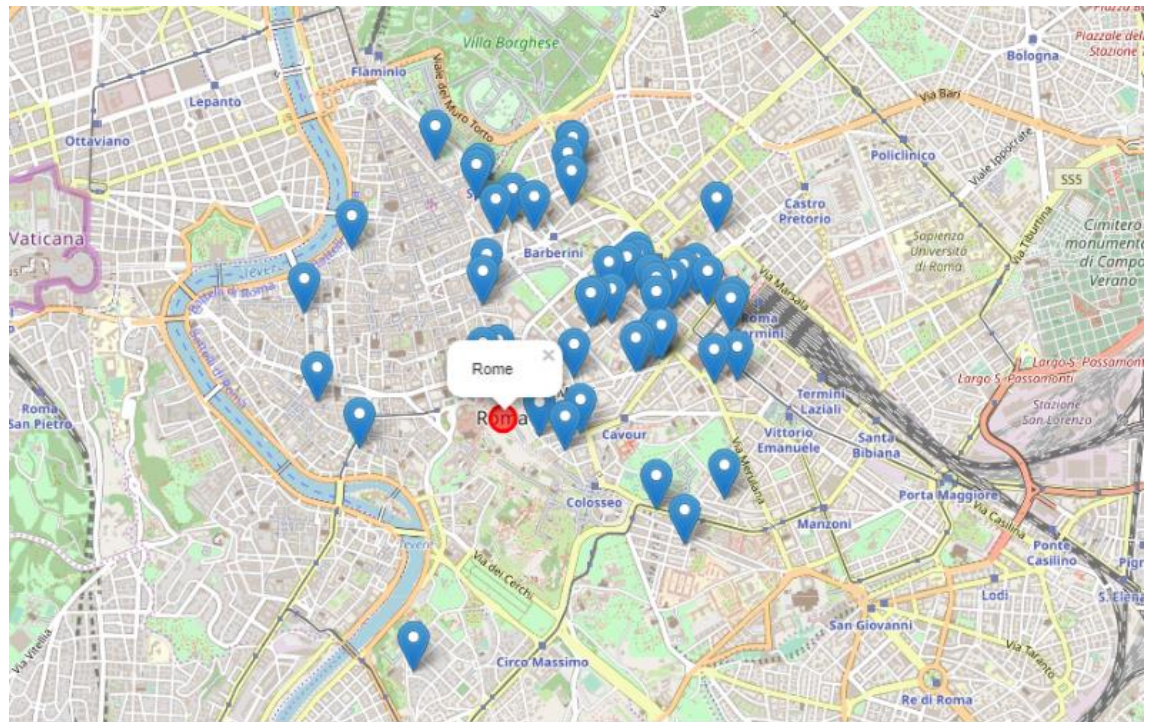

Fig.03 Map of Rome and hotels in an area with 1.3km of radius

3. Methodology

In this project we will direct our efforts on detecting venues around the hotels. We will limit our analysis to area 500 meters around each hotel.

In first step we have collected the required data: location and type (category) of every venue within 500 meters around each hotel, we will use maps to visualize the hotels in Rome.

Second step in our analysis will be calculation and exploration of 'venues' across different areas of hotels.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet the requirements established in discussion with stakeholders.

4. Analysis

4.1 Analysis areas around hotels

This project target to explore all venues around each hotels in Rome, to recommend the best hotels for each tourist. As it is widely known not all tourist travel to other cities to see sight some of them go for shopping, adventures, business, etc.

We will obtain number of existing venues and their type and location around each hotel with Foursquare API with a limit as 100 venues and the radius 500 meter for each from their given latitude and longitude information. Here is a head of the list from Foursquare API.

	name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	939 Hotel (Hotel Marcus)	Italian Restaurant	Plaza	Ice Cream Shop	Pizza Place	Trattoria/Osteria	Boutique	Art Museum	Bistro	Restaurant	Sandwich Place
1	Aleph Rome Hotel, Curio Collection by Hilton	Italian Restaurant	Trattoria/Osteria	Bed & Breakfast	Pizza Place	Restaurant	Cocktail Bar	Bar	Middle Eastern Restaurant	Fountain	Plaza
2	B&B Trevi Hotel	Italian Restaurant	Ice Cream Shop	Plaza	Art Museum	Café	Sandwich Place	Restaurant	Bed & Breakfast	Toy / Game Store	Monument / Landmark
3	Baglioni Hotel Regina	Italian Restaurant	Bistro	Trattoria/Osteria	Restaurant	Movie Theater	Cocktail Bar	Pizza Place	Fountain	Nightclub	Gym
4	Downtown Accommodation Hotel	Italian Restaurant	Historic Site	Wine Bar	Pizza Place	Ice Cream Shop	Sandwich Place	Monument / Landmark	Temple	Pub	Cocktail Bar

Fig.04 top 10 common venues around each hotel

Again, with the support of Folium, I visualized geographic details of each cluster, which should be a starting point for individuals to explore and search for optimal hotels.

4.2 Examine Clusters

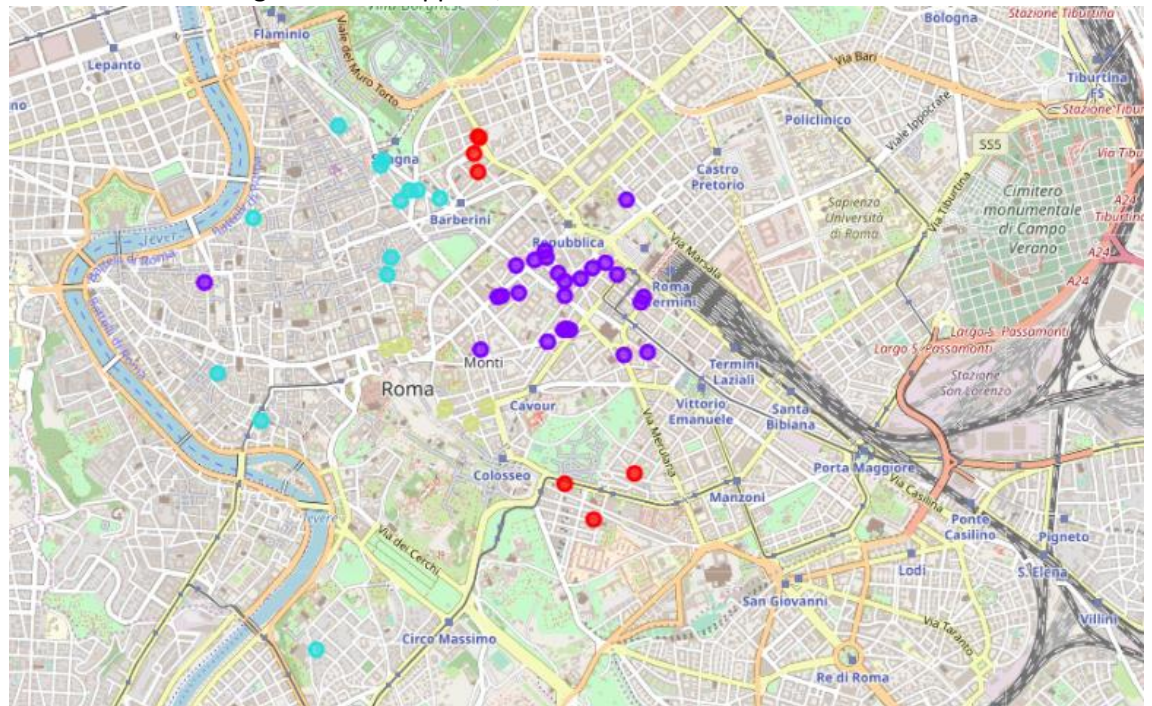After the K-means algorithm was applied, all hotels were divided into 5 clusters:

Fig.05 Visualization of hotels after attached cluster labels using Folium

First Cluster:

name	categories	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Hotel Capo D'Africa Rome	Hotel	Italian Restaurant	Trattoria/Osteria	Café	French Restaurant	Pizza Place	Plaza	Wine Bar	Church	Park	Beer Store
Hotel Ambasciatori Palace	Hotel	Italian Restaurant	Restaurant	Trattoria/Osteria	Cocktail Bar	Pizza Place	Middle Eastern Restaurant	Plaza	Fountain	Bed & Breakfast	Roman Restaurant
Grand Hotel Palace	Hotel	Italian Restaurant	Bistro	Restaurant	Trattoria/Osteria	Pizza Place	Movie Theater	Cocktail Bar	Ice Cream Shop	Café	Plaza
Hotel Gladiatori Palazzo Manfredi	Roof Deck	Italian Restaurant	Café	Historic Site	Trattoria/Osteria	Cocktail Bar	Plaza	French Restaurant	Wine Bar	Pizza Place	Gym / Fitness Center
Baglioni Hotel Regina	Hotel	Italian Restaurant	Bistro	Restaurant	Trattoria/Osteria	Cocktail Bar	Pizza Place	Movie Theater	Café	Gym	Nightclub

Fig.06 head of cluster 1

Second Cluster:

name	categories	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Hotel Apollo	Hotel	Italian Restaurant	Pizza Place	Ice Cream Shop	Wine Bar	Coffee Shop	Bed & Breakfast	Sandwich Place	Historic Site	Indian Restaurant	Roman Restaurant
Hotel Diana	Hotel	Italian Restaurant	Café	Ice Cream Shop	Plaza	Bed & Breakfast	Cocktail Bar	Restaurant	Dessert Shop	Bookstore	Bistro
Hotel Raffaello	Hotel	Italian Restaurant	Ice Cream Shop	Pizza Place	Bed & Breakfast	Wine Bar	Cocktail Bar	Café	Plaza	Sandwich Place	Park
Hotel Galatea	Hotel	Italian Restaurant	Ice Cream Shop	Bed & Breakfast	Pizza Place	Wine Bar	Plaza	Cocktail Bar	Roman Restaurant	Sandwich Place	Bistro
Hotel Esposizione	Hotel	Italian Restaurant	Ice Cream Shop	Pizza Place	Wine Bar	Bed & Breakfast	Cocktail Bar	Plaza	Roman Restaurant	Art Museum	Sandwich Place

Fig.07 head of cluster 2

Third Cluster:

name	categories	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Hotel Hassler Roma	Hotel	Italian Restaurant	Boutique	Café	Plaza	Jewelry Store	Japanese Restaurant	Trattoria/Osteria	Pizza Place	Road	Historic Site
939 Hotel (Hotel Marcus)	Hotel	Italian Restaurant	Plaza	Ice Cream Shop	Pizza Place	Boutique	Trattoria/Osteria	Sandwich Place	Restaurant	Art Museum	Toy / Game Store
Hotel Art	Hotel	Italian Restaurant	Boutique	Plaza	Jewelry Store	Ice Cream Shop	Historic Site	Cosmetics Shop	Pizza Place	Café	Road
Hotel Sistina - Roma	Bed & Breakfast	Italian Restaurant	Ice Cream Shop	Boutique	Fountain	Plaza	Bed & Breakfast	Restaurant	Pizza Place	Café	Japanese Restaurant
Hotel San Anselmo	Hotel	Italian Restaurant	Plaza	Park	Café	Roman Restaurant	Scenic Lookout	Ice Cream Shop	Gastropub	Fast Food Restaurant	Creperie

Fig.08 head of cluster 3

Fourth Cluster:

name	categories	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Hotel Forum	Hotel	Italian Restaurant	Historic Site	Temple	History Museum	Wine Bar	Plaza	Pizza Place	Monument / Landmark	Sandwich Place	Ice Cream Shop
Hotel Pace Helvezia	Hotel	Italian Restaurant	Historic Site	Ice Cream Shop	History Museum	Plaza	Restaurant	Art Museum	Monument / Landmark	Pizza Place	Pub
Hotel Fori Imperiali Cavalieri	Hotel	Italian Restaurant	Historic Site	Temple	Ice Cream Shop	Wine Bar	Sandwich Place	Pizza Place	Monument / Landmark	Plaza	Pub
Downtown Accommodation Hotel	Hotel	Italian Restaurant	Historic Site	Wine Bar	Pizza Place	Ice Cream Shop	Sandwich Place	Monument / Landmark	Temple	Café	Roman Restaurant
Hotel Cosmopolita Rome	Hotel	Italian Restaurant	Historic Site	Plaza	Ice Cream Shop	Art Museum	History Museum	Restaurant	Museum	Pub	Monument / Landmark

Fig.09 head of cluster 4

5. Results and Discussion

Our analysis shows that exist 50 hotels registered by Foursquare in Rome in a circular area with a radius of 1.5km, also as we can expect the most common venue in all the Clusters is 'Italian Restaurant', so that venue will not necessary in to the recommendation, also we can see in the first map that the majority of hotels are in the city center, that could be a characteristic to consider if our client prefer a quiet place. Then we can analyze each Cluster and finally get a conclusion.

1) Cluster 1 contains 16 hotels with 10 tops venues mainly contains Café/Ice cream shop/wine bar/pizza place...,it looks like a place where the citizens will enjoy, but maybe not for a tourist who want try and taste different things.

2) Cluster 2 contains 16 hotels with 10 tops venues mainly contains Boutique/Ice cream shop/fountain/shops, it looks like for a tourist who go for shopping and have a good time. 3) Cluster 3 contains 12 hotels with 10 tops venues mainly contains ice cream shop/plaza/historic sites..., it looks balanced for a tourist who want know about the history, but enjoy his free time too.

4) Cluster 4 contains 6 hotels with 10 tops venues mainly contains history site/Temples/Museums/Monuments..., it unlike the first cluster looks good for a tourist who want know more about the culture and history of Rome.

6. Conclusion

Purpose of this project was to identify a group of hotels for a tourists depending of their pleasures. Target to individuals who want to travel to Rome, Italy for sightseeing. The Foursquare location data was leveraged to compare each hotel to provide reliable suggestions for individuals who want to choose a hotel according their reasons to travel (sightseeing, business, shopping, etc.).With unsupervised learning

K-means algorithm, all hotels were clustered in to 4 categories, the advantages of each category was expressed to help individuals choose the ideal group oh hotels to recommend group of hotels to recommend to the client. This Project simply processed the hotels, and cluster them into 4 categories based one the venues around, the results can only help individuals choose the hotel they want. Further analysis can be done base on these 4 clusters, which can help provide more detail information to clarify the advantages of each.