Best neighborhoods for Foodies in Valencia, Spain

Introduction:

The goal of the project is to get insights towards what are the best neighborhoods for food lovers in Valencia, considering for it geographical information of the city and the Foursquare API for recommendations and prices, ratings and categories.

Data Sources:

Two main data sources were used for this project:

- **División administrativa de los barrios municipales** (http://gobiernoabierto.valencia.es/es/dataset/?id=barrios). Geojson describing the different neighborhoods of Valencia. Used to get the geographical center of each neighborhood.
- **Foursquare Venue Recommendations.** Used to get venues (restaurants) around an area. The information about this restaurants is quite scarce, so a second API was used to get more insights.
- **Foursquare Details of a Venue.** Used to get pricing tier, score and category of the restaurants from retrieved from the previous API

Methodology:

Using the geographical centers of the neighborhoods from the administrative city division, we looked for the 100 recommended eateries in a radius of 500 meters using the Foursquare Venue Recommendations API. This radius was selected as being still under walking distance while at the same time covering under it most of the neighborhoods. The fact that there is an overlap between this radius is not only expected but also intended, as it helps to get an idea not only of the venues inside the neighborhoods but also the venues close but in different neighborhoods as there is no prior reason to avoid some venues in favor of another due to neighborhood if they are within a similar distance.

Once we have the list of venues, we retrieve their details using the Foursquare Details of a Venue API. The details used for this project where price tier (Cheap, Moderate, Expensive, Very Expensive or Unknown) and the rating. To help the classification, we converted the rating from a Rational Number to a category as follows:

- 0-2: Very Bad
- 2-4: Bad
- 4-6: Average
- 6-8: Good
- 8-10: Excellent
- Unknown Rating

Once all the venues where classified we use One Hot Encoder to turn all the categories into labels and then aggregated those values by neighborhoods as the mean of all the values. Finally we used Kmeans to get the right distribution of clusters.

Results:
The clusters found are described with the following characteristics:

	Cheap	Expensive	Moderate	Unknown price	Very Expensive	Average	Excelent	Good	Unknown rating
0	0.2	0.2	0.5	0.1	0.0	0.1	0.1	0.4	0.5
1	0.3	0.2	0.5	0.0	0.0	0.1	0.2	0.6	0.1
2	0.6	0.1	0.3	0.0	0.0	0.2	0.0	0.3	0.6
3	0.0	0.1	0.9	0.0	0.0	0.0	0.0	0.1	0.9
4	0.4	0.1	0.4	0.0	0.0	0.2	0.1	0.5	0.2
5	0.4	0.1	0.4	0.0	0.0	0.0	0.0	0.2	0.8

From this results we can conclude the following:

- Cluster 1 seems to be the best option as it has high ratings (80% between Good and Excellent)
- Cluster 4 has slightly lower ratings (60% between Good and Excellent), but is compensated with better prices across the board.
- Clusters 3 and 5 have similar Unknown ratings, although prices in 5 are lower. This could signal that this are neighborhoods not frequented by foodies of tourists, main users of sites like Foursquare.

The complete classification can be found in the file ‘Cluster_output.txt’

Discussion:
This analysis is only a directional indication of where the best venues can be found. There are a few areas that need to be explored in order to have a more comprehensive analysis:

- Quantity of good restaurants VS Ratio of good restaurants: This analysis considers the average quality and price of the restaurants in an area, but that can be misleading as a new ‘bad’ restaurant would have a negative impact on the whole neighborhood. Instead another feature could be engineered such as number of Good or Excellent restaurants in the area, or by inhabitants, in order to focus only on the most relevant locations.
- Homogeneity VS Diversity: For this analysis all restaurants are considered as being interchangeable as long as their scores and prices are similar, ignoring that the diversity of categories can be an attractive factor when choosing an area for dinning.
- Due to the fact that the Venue Details API is considered a Premium API, it was not possible to analyze all the neighborhoods to the same extent. Across several days more information will be retrieved in order to improve the quality of the data and thus the insights.

Conclusion:
After observing the clusters and comparing against the classifications of each neighborhood, it can be concluded that the analysis did a good job grouping together some of the most popular/trendy areas (El Carme, Mestalla, Ruzafa, El Pilar, El Mercat…) and the most residential/less known areas (Favara, Torrefiel...)