RESTAURANT-HUNT IN SEVILLE, SPAIN

CAPSTONE PROJECT DESCRIPTION

The aim of this project is to make an analysis of the venues of the city of Seville, Spain. The analysis will be made based on the ratings of the venues and their price.

Seville is a very famous and beautiful city of Spain which attracts a lot of tourism. It is such an historical city where a lot of cultures were mixed and used to live together in peace and harmony. It has lots of magical places to get lost through their streets and their people are actually open-minded and friendly.

However, it is quite complicated for tourists to find good places to have some food since it is not the city with the biggest technological expansion and development.

With this analysis, tourist will be able to select the restaurant that fits them the best in order to have a great experience since Spain is one of the countries with the best gastronomy in the world!!

DATA DESCRIPTION

The data that will be used in order to make this analisys is the data provided by the Foursquare API.

It will be necessary to make a proper data mining and selection among this data.

The venues used for this study will be the ones in a 3 km radius from the center of the city of Seville, in Spain. In this way, tourist that would be visiting the city center could choose the restaurant they like the most..

The raw data provided by the API needs to be treated.The main features that we will be needing to develop this analisys are the following:

- Name of the restaurants
- Location and coordenates
- Ratings
- Categories of the restaurants (type of food they serve)
- Distance from the city center

METHODOLOGY

As it was told before, the data provider was Foursquare. The raw data received by the API needed to be treated correctly, so the first step was the data preprocessing in order to get a filtered dataframe to work with.

During the data collection, there was a problem which needed to be solved. Foursquare doesn´t have any rating attribute for the city of Seville in Spain (I don't know whether this rating stuff is just for Seville or for all the places in Spain, the country I am from).

Due to this, the rating was made with the likes that the restaurants had received in this platform, which I consider it is also a very good indicator of the quality of the different restaurants.

Because of the limitations in the premium calls to the Foursquare API (50 premium calls per day), I needed two days to develop the rating section of every one of the restaurants shown in the analysis.

The second step after data preprocessing was to show the restaurants in the map of Seville, what was made with the Folium tool. This map will be shown later in the results section.

The later step was to group the restaurants according to the rating and the likes received, which was developed using clustering. It is not shown in the Jupyter Notebook due to a problem with the Dataframe, but the criterium followed was this K-means clustering.

To end up, the last step was to include this clustering in the map of Seville in order to get an easy visual recognisation of the goodly and badly rated restaurants. I did this differentiation with colors, which will be explained in the results section.

RESULTS AND DISCUSSION

The Dataframe obtained after the preprocessing part and the adding of the ratings (likes) looks like the figure below:

```
df_final = dataframe_filtered
df_final['rating'] = ratings
df_final.drop('id', axis = 1, inplace=True)
df_final.head()
```

	name	categories	lat	lng	distance	address	rating
0	Restaurant (NH Plaza de Armas)	Buffet	37.392902	-6.002503	792	NaN	1
1	Restaurant Recoveco	Spanish Restaurant	37.391566	-5.990207	559	Calle Ortiz de Zúñiga	0
2	Al Wadi Restaurant	Halal Restaurant	37.386744	-5.994920	213	Jimios nº 32	32
3	Laherre Restaurant	Restaurant	37.387692	-5.991923	319	NaN	5
4	Restaurante Plaza Nueva	Spanish Restaurant	37.387839	-5.995456	88	Calle Barcelona 5	0

The map of the restaurants located in a 3 km radius area from the city center of Seville is the next one:
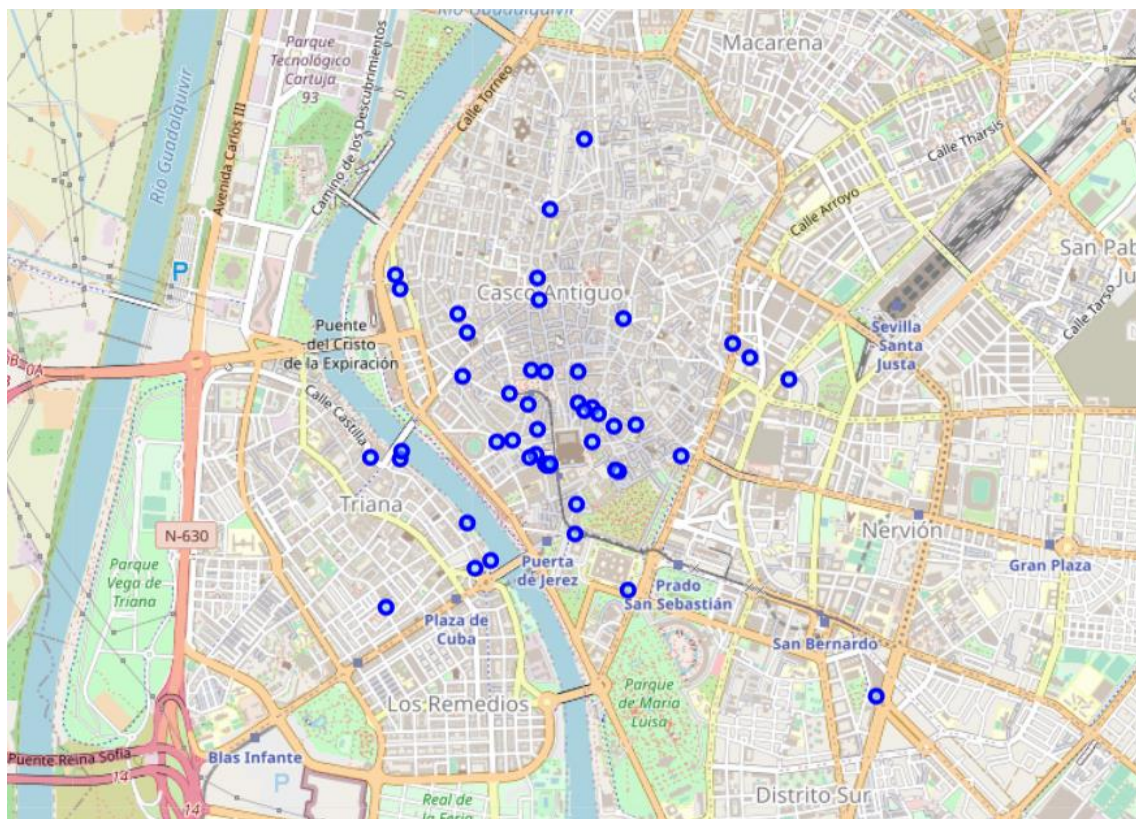
A part of the final Dataframe (the final one contains 50 restaurants) after the clustering process with K-means algorithm, with the restaurants grouped by their rating and opinions is shown in the next figure:

	name	categories	lat	lng	distance	address	rating	group
0	Restaurant (NH Plaza de Armas)	Buffet	37.392902	-6.002503	792	NaN	1	1
1	Restaurant Recoveco	Spanish Restaurant	37.391566	-5.990207	559	Calle Ortiz de Zúñiga	0	1
2	Al Wadi Restaurant	Halal Restaurant	37.386744	-5.994920	213	Jimios nº 32	32	4
3	Laherre Restaurant	Restaurant	37.387692	-5.991923	319	NaN	5	1
4	Restaurante Plaza Nueva	Spanish Restaurant	37.387839	-5.995456	88	Calle Barcelona 5	0	1
5	Restaurante Barbiana	Spanish Restaurant	37.389355	-5.995241	81	C/ Albareda	2	1
6	al cazar arabian restaurant	Diner	37.385172	-5.994265	396	NaN	0	1
7	مطعم القصر AL CAZAR Restaurant	Middle Eastern Restaurant	37.385132	-5.994419	397	NaN	1	1
8	Morrison's Restaurant. Hotel Fontecruz	Restaurant	37.386880	-5.990727	452	Abades 40	0	1
9	Restaurante Don Carlos	Spanish Restaurant	37.389257	-5.994462	104	C. General Polavieja, 18	5	1
10	Restaurant NH	Breakfast Spot	37.375047	-5.976319	2262	NaN	0	1
11	Restaurante Bajo Guía	Restaurant	37.386207	-5.997132	312	Calle Adriano, 5	11	2
12	San Fernando 27 Restaurant	Spanish Restaurant	37.382165	-5.992912	751	NaN	0	1
13	MUMBAI Indian Restaurant	Indian Restaurant	37.396345	-5.994268	863	Calle Amor de Dios 43	1	1
14	Hotel Inglaterra	Hotel	37.388287	-5.996478	107	Plaza Nueva, 7	56	5
15	Restaurante El Cabildo	Spanish Restaurant	37.385617	-5.994999	336	Plaza del Cabildo, s/n	5	1
16	My Way Restaurant	American Restaurant	37.389844	-5.983240	1078	C. Ubeda	12	2
17	Restaurante Taj Mahal	Indian Restaurant	37.389065	-5.999012	328	C. Zaragoza, 6	7	2
18	Restaurante Mediterránea	Restaurant	37.389241	-5.992683	244	NaN	0	1
19	Restaurante La Raza Seises	Restaurant	37.387438	-5.991590	357	Don Remondo, 2	2	1
20	Restaurante Baco	Mediterranean Restaurant	37.387874	-5.992674	250	Calle Francos,42	1	1

The groups considered are 5: 1 for the worst rated restaurants and 5 for the best ones. In the image below we can observe the number of restaurants that belong to every one of the groups clustered:
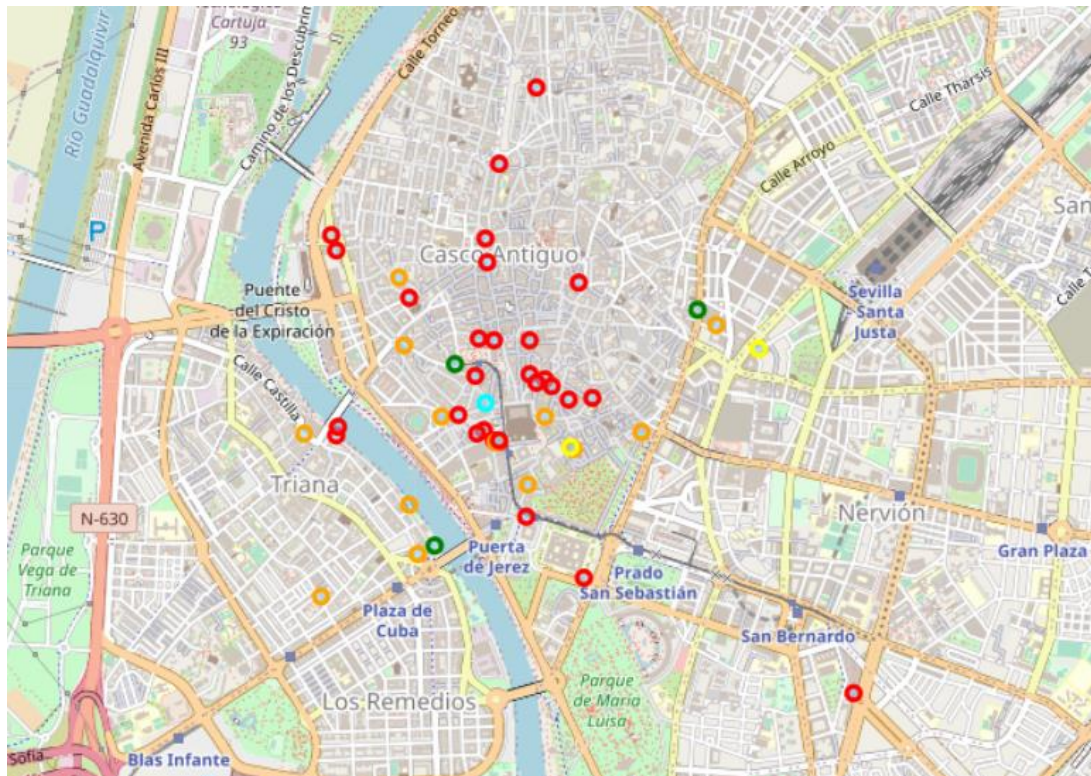
group	
1	30
2	14
3	2
4	1
5	3

Finally, the result of the analysis is shown in the map in the figure below.

The colors show how good are the restaurants rated:

1: Red

2: Orange

3: Yellow

4: Cyan

5: Green

CONCLUSSIONS

The aim of this project is to make easy the restaurant-choosing step for people who would like to visit such an historical city as Seville. Obviously, this is only a first approach to this problem, but it can be a good recommendation for any tourist who want to go to this incredible city and who want to have a delicious meal and experience in Seville.

The map is the graphical representation of the Dataset, but this Dataset contains much more information than the one shown in the map, so that anyone who wants to have a more detailed study of the analysis can do it easily.

This project is an instance of the huge potential of the machine learning algorithms for any kind of problem and the big field and horizon of possibilities that the Internet and the free APIs can provide to anyone who is learning about Data Science and also to the people who want to apply this knowledge to a great amount of problems from the real world.

I hope you can be able to visit the city, it is really worth it and full of historical and magical places within. If you do visit it and want to have a good culinary experience, don't doubt to visit my Github website and take a look at this analysis!

Thank you so much for reading this pall!! Regards!! 😉