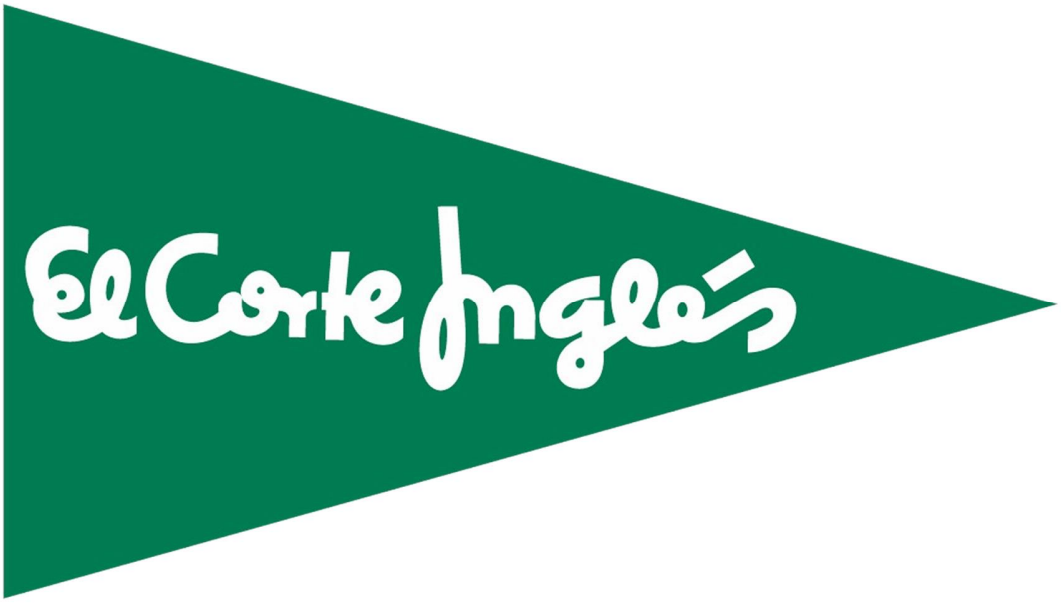
2019



Attracting New Investments with the Establishment of a New Center of El Corte Inglés in Huelva

CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS

ISABEL DELGADO ARENAS

INTRODUCTION

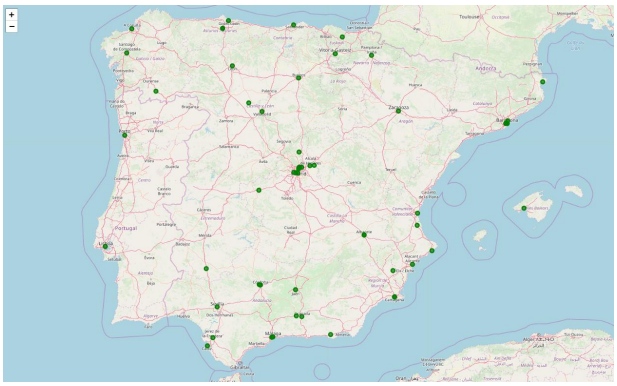
The objective of this document present the final project of the course "Applied Data Science Capstone" by Coursera.

The project use Data Science Methodology to give answer to a question. In this case:

Which kind of business would get successful with the establishment of a center of El Corte Inglés in Huelva?

We must clarify that the new location that we use in this project not a real location for El Corte Ingles Center. We have chosen it at random in order to develop this investigation.

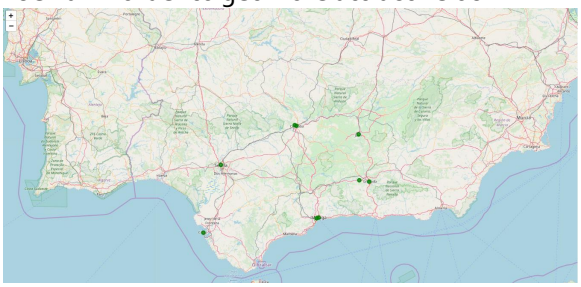
Introduction/Business Problem



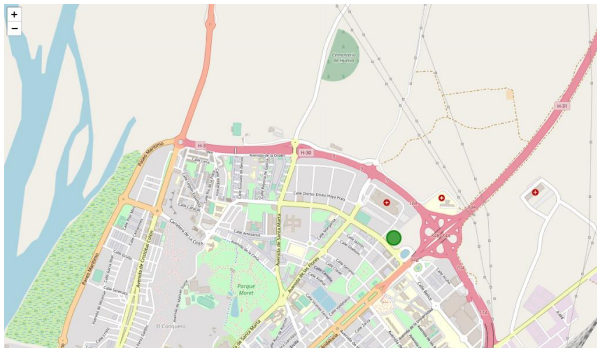
El Corte Inglés is ones of the largest group in Spain that operates in sectors like commerce and distribution, insurance, travel, information technology and it has consolidated the leadership of its department stores in an increasingly demanding and competitive market.

This area (department stores) also generate benefit in the place where this is settled. Establish a center of El Corte Ingles use to attracts other kinds of business and to revitalize this area of the city

In our project, we have a location for a new center in Huelva. In order to get more attractive our establishment, to new investors or town government, we are going to find out **which kind of business would get successful with the establishment of a center of El Corte Inglés.**



We can use the result of our research to contact with future investor interested in the result categories. The town Government would be interested on promoting the area or some of the activities. This analysis would help then to support the establishment of the new center. Also the own company would see convenience develop a new activity as a result of the investigation.



Data

In order to give an answer to this question, we only need two kind data: data of the centers and data of the activities around then

Data of the centers

We will extract from El Corte Inglés (since now ECI) web pages, all the centers of Spain, and especially those that are located in Andalucía. We transform the list below in a data frame which contains

- Name of the center
- Address
- Telephone
- Region
- Longitude
- Latitude

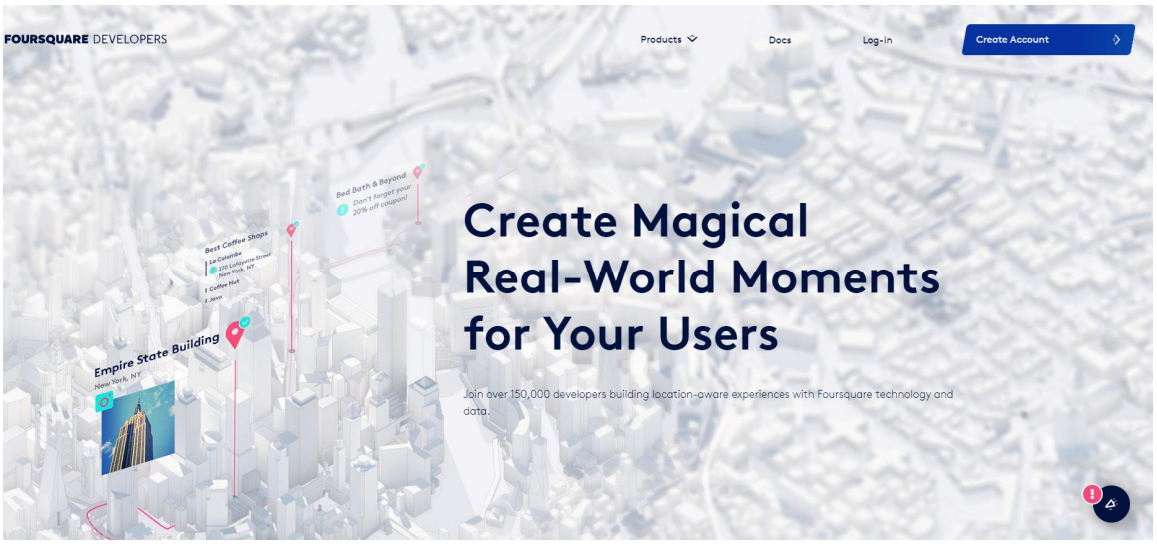
We obtain name, address, telephone and region from the web page. To get geolocation of the centers we use address and Geopy. If we don't find the location we delete that center.

9]:

	Centro	Direccion	Telefono	Latitude	Longitude	Dir_sp	Region
0	El Ejido	Paseo de Pedro Ponce, s/n, 04700 Almería, And...	950 542 000	36.7733	-2.80375	[Paseo de Pedro Ponce, s/n, 04700 Almería, ...	Andalucía
1	Bahía de Cádiz	Avda. de las Cortes de Cádiz,1,11012 Cádiz, An...	956 297 100	36.521	-6.2771	[Avda. de las Cortes de Cádiz, 1, 11012 Cádiz,...	Andalucía
2	Ronda de los Tejares	Av. Ronda de los Tejares, 30, 14008 Córdoba, E...	957 222 881	37.887	-4.78359	[Av. Ronda de los Tejares, 30, 14008 Córdoba...	Andalucía
3	Ronda de Córdoba	Ctra. Sta. María de Trassierra, S/N, Córdoba, ...	957 224 949	37.8959	-4.81079	[Ctra. Sta. María de Trassierra, S/N, Córdob...	Andalucía
4	Arabial	Arabial, 97, 18003 Granada, Andalucía	958 217 600	37.1777	-3.61265	[Arabial, 97, 18003 Granada, Andalucía]	Andalucía
5	Plaça de Catalunya	Plaça de Catalunya, 14 Barcelona	933 063 800.	41.3877	2.17103	[Plaça de Catalunya, 14 Barcelona]	14 Barcelona
6	Diagonal	Avenida Diagonal, 617 08028 Barcelona	933 667 100	41.3891	2.13028	[Avenida Diagonal, 617 08028 Barcelona]	617 08028 Barcelona
7	Francesc Macià	Avinguda Diagonal, 471 08036 Barcelona	934 934 800	41.391	2.13911	[Avinguda Diagonal, 471 08036 Barcelona]	471 08036 Barcelona

Data of Venues

Ones we get geolocation we will use foursquare to explore business around it. We use explore utility to get Venues around the centers



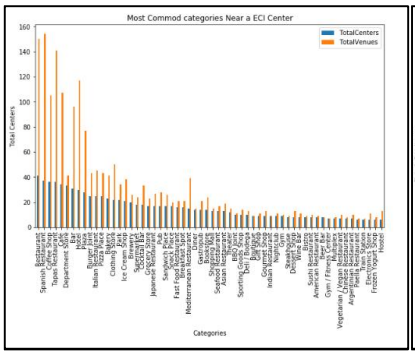
We will extract name, categories and location

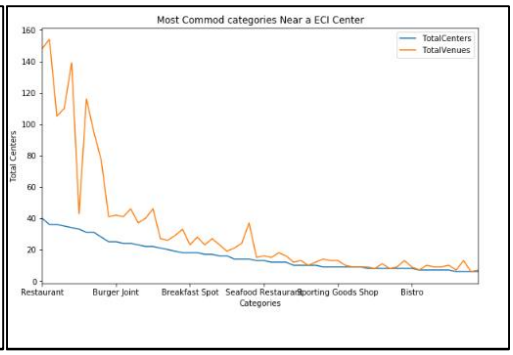
	Center	Center Latitude	Center Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	El Ejido	36.773334	-2.80375	Cañas Y Adobo	36.776378	-2.802690	Spanish Restaurant
1	El Ejido	36.773334	-2.80375	El Corte Inglés	36.772692	-2.806002	Shopping Mall
2	El Ejido	36.773334	-2.80375	La Venia	36.773163	-2.801375	Café
3	El Ejido	36.773334	-2.80375	Burger King	36.774772	-2.804316	Fast Food Restaurant
4	El Ejido	36.773334	-2.80375	Wok slowly	36.774030	-2.803892	Chinese Restaurant

Methodology

The objective of this analysis is discover which categories of venues would be settle around our new center in Huelva. In all this analysis we will use two points of view. Consider Venues categories around centers of ECI across all over Spain or only Andalusia.

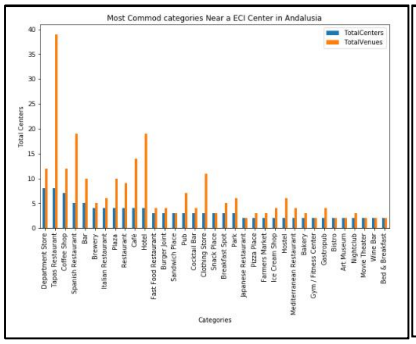
We started representing relation between venues categories, centers of ECI and numbers of establishment of this categories.

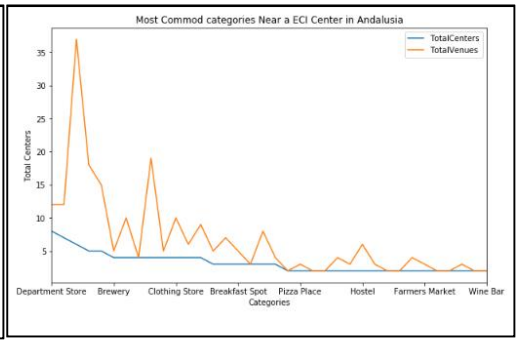




"TotalCenters" represent number of centers of ECI that have, unless one, establishment of the categories fix in **x** axis. "TotalVenues" represent number of enterprise around an ECI's center.

Since the perspective of all Spain the most popular Venues Categories around ECI use to have the majority of the establishment.





From Andalusian perspective this relationship is not as clear as in the Spanish point of view

So there is a relation between centers and categories Venue but this relation change depend of the geographical area. After that we decide do a segmentation of center using venues categories around then.

Segmentation

As we say before we use two point of view: all Spain, Andalusian view. To segment we use the k-mean clustering model.

We start with Spanish point of view. We segment activities in 5 clusters

[87]:

Labels	Accessories Store	American Restaurant	Amphitheater	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	...	Turkish Restaurant	Udon Restaurant	University	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Wine Bar	Wine Shop	Winery	Women's Store
0	0.000393	0.003616	0.000363	0.000393	0.000000	0.004120	0.001785	0.001344	0.005392	0.004967	...	0.000925	0.000868	0.000429	0.002934	0.000439	0.000000	0.00603	0.000000	0.00051	0.0014
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.022222	0.000000	...	0.000000	0.000000	0.000000	0.022222	0.000000	0.000000	0.00000	0.000000	0.00000	0.0000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.017857	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.017857	0.00000	0.017857	0.00000	0.0000
3	0.000000	0.031250	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.031250	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.03125	0.0000
4	0.019231	0.000000	0.000000	0.000000	0.019231	0.019231	0.000000	0.019231	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.019231	0.000000	0.000000	0.00000	0.000000	0.00000	0.0000

5 rows × 213 columns

In the case of Andalusian view we choose 3 segments

Labels	American Restaurant	Art Gallery	Art Museum	Asian Restaurant	Bakery	Bar	Bed & Breakfast	Beer Bar	Beer Garden	Beer Store	...	Tapas Restaurant	Tennis Court	Thai Restaurant	Theater	Theme Park Ride / Attraction	Toy / Game Store	Train Station	Wine Bar	Winery	Women's Store
0	0.005495	0.000	0.000000	0.000000	0.000000	0.011988	0.005495	0.000000	0.005495	0.000	...	0.059777	0.020408	0.005495	0.005495	0.047619	0.000000	0.009524	0.005495	0.000000	0.000000
1	0.000000	0.025	0.025000	0.000000	0.025000	0.025000	0.025000	0.000000	0.000000	0.025	...	0.025000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.025000	0.000000	0.000000
2	0.000000	0.000	0.027027	0.027027	0.027027	0.027027	0.000000	0.027027	0.000000	0.000	...	0.000000	0.000000	0.000000	0.000000	0.000000	0.027027	0.000000	0.000000	0.027027	0.027027

3 rows × 84 columns

Now, we have to find out in which segment is our new location.

We search, using Foursquare, all venues categories of the new location

1]:

	Venue Category	0
0	Grocery Store	1
1	Hotel	1
2	Tapas Restaurant	1
3	Women's Store	1

We look in Spanish Segmentation the local categories and extract mean value. We sum values for all activities and the biggest ones show as the Spanish Segment

	0	1	2	3	4
Grocery Store	0,01002	0,02222	0	0,03125	0
Hotel	0,02443	0,02222	0,01785	0	0,01923
Tapas Restaurant	0,03937	0,02222	0,01785	0,03125	0,01923
Women's Store	0,00139	0	0	0	0
	0,07521	**0,06667**	**0,03572**	**0,06253**	**0,0385**

We repeat the same analysis for Andalusia

	0	1	2
Grocery Store	0	0	0,02702
Hotel	0,01198	0,025	0,02702
Tapas Restaurant	0,05977	0,025	0,02702
Women's Store	0	0	0,02702
	0,07175	**0,05001**	**0,1081**

RESULTS

	Category	Value
0	Accessories Store	1
1	American Restaurant	1
2	Amphitheater	1
3	Arcade	1
4	Argentinian Restaurant	1
...	...	...
242	Wine Bar	2
243	Grocery Store	3
244	Hotel	3
245	Tapas Restaurant	3
246	Women's Store	3

247 rows × 3 columns

As a result get a data frame with two columns:

· Category: is the name of the venue category that would be success in our new location.

· Value: we will chose this value to identify origin

1. Spain
2. Andalusian
3. Huelva

With this we get all the activities that are develop around centers of the same segment.

Discussion Section

Category	Value	Region
Hotel	6	3
Tapas Restaurant	6	3
Women's Store	4	2
Grocery Store	4	2
Market	3	2
Japanese Restaurant	3	2
American Restaurant	3	2
Ice Cream Shop	3	2
Hotel Bar	3	2
Gym / Fitness Center	3	2

Will we choose all this activities? Well we can refine the results.

Firsts of all we would choose those activities that are develop on more than one region. We add a column to count regions. The result data frame aggregate categories, sum "Value" and "Region". We select those who "Region">1.

Also we pay attention to value. We give higher value to closer location, because we thought that is more probably get success

Conclusion

These are ten categories that we recommended to promote.