Business Problem

Background: Barcelona, an expensive dream

Finding a place to live can be a hard task, specially if you are not living in the city and you are a foreign student. As a student in a different country I learned a lot of tips to find the perfect place to live, unfortunately, everything I learned was at the final period of my studies.

Barcelona is a modern cosmopolitan city endowed with great dynamism. A great place with a rich culture, an identity of its own and an inexhaustible cultural and recreational offer. Therefore, it is not surprising that it is the 3rd most visited city in Europe and the 10th worldwide, with more than 8.4 million tourists annually. But all of this has a cost: it is the city with the most expensive rental in Spain. Also, in Barcelona there are more than 200,000 students where the 40% are foreign and the other 30% comes from other cities of Spain. We have the combination of a city with a lot of attractions and culture and a high flow of students desperate to find the most convenient place to live.

My business idea is to create an application that recommends the best neighborhood to live in Barcelona according to each type of student. Let's take a look.

The quest for the perfect Neighbourhood

Imagine you are a foreign student in Barcelona. You must prepare all the things you need for your trip, your papers, where will you live, which are the most popular neighbourhoods? which means of transport will you to take? So you enter the internet to look for useful information but bad news: your city is big and there's so much information it confuses you. By this point you are already very stressed, so, what could you do?

Enter to the new and fresh app called Ciro! With only a few questions about your personality, college, budget and interests, Ciro recommends you the three best neighbourhoods!

CIRO

THE RIGHT PLACE

Target audience

The target audience of this project are foreign students coming to live to Barcelona, from outside of Spain as well as from other cities in Spain. This project would be likewise interesting to property owners, real-state agencies and investors seeking to invest in an innovative app and other kind of investors related to a variety of sectors such as restaurants, historical sites and monuments etc.

Data

We will need data from:

- Cartographic division of Barcelona by neighborhood
- The most attractive places for each neighborhood
- Nearby libraries
- Cost of the average rental of apartments by neighborhood
- Transport information

The sources used for this project will be:

- Foursquare: to get the venue data from the neighbourhoods
- Open Data BCN: to get information about:
  - Coordinates of the neighbourhoods
  - Apartment rental by neighborhood
  - Public libraries
  - Public transportation

Methodology

Analytic approach

To face the problem, we create a Machine Learning model using a k-means clustering algorithm to divide Barcelona into 10 groups to recommend a zone to each student type.

Knowing Barcelona

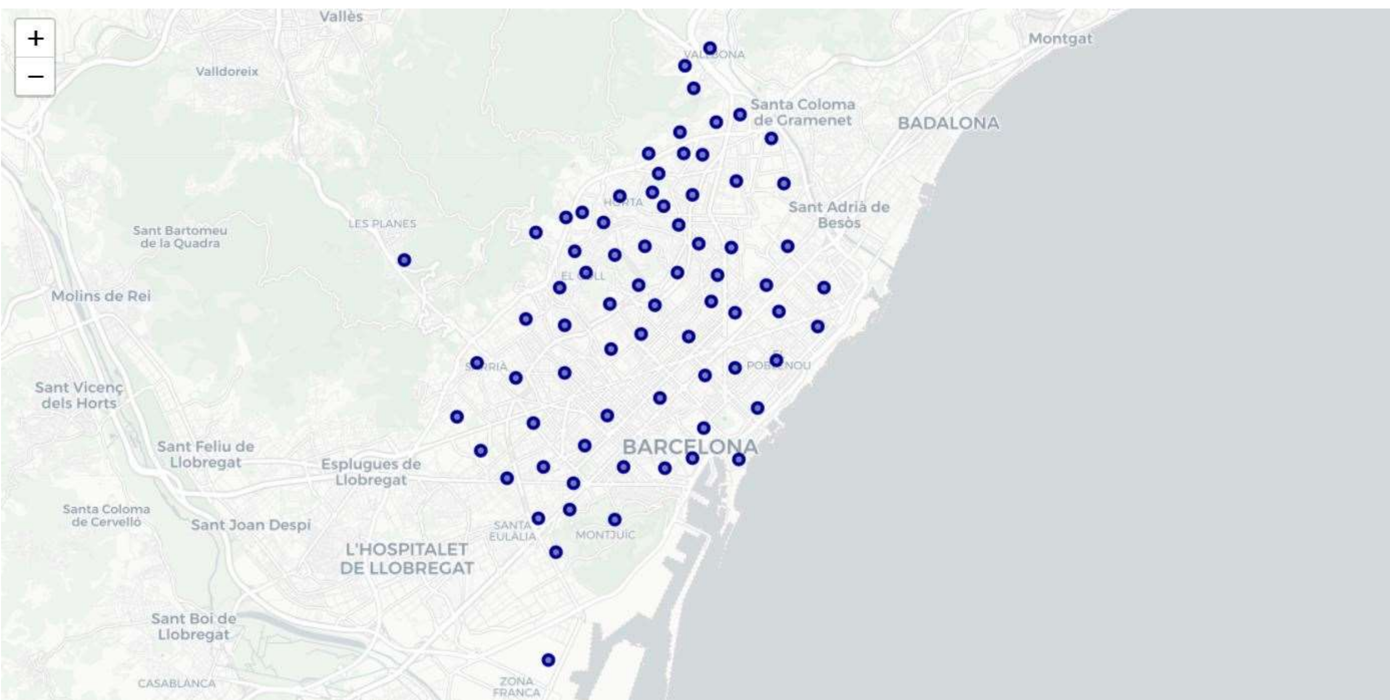
With a population of 1.6 million distributed in 10 districts and **73 neighbourhoods**

The four dataframes used were:

- Coordinates of the neighbourhoods
- Apartment rental by neighborhood
- Public libraries
- Public transportation

First, we obtained the coordinates -latitude and longitude- of the **73 neighbourhoods** of Barcelona:

	Neighbourhood	Latitud	Longitud
0	Baró de Viver	41.446030	2.199706
1	Can Baró	41.415980	2.163248
2	Can Peguera	41.435088	2.166949
3	Canyelles	41.443092	2.166052
4	Ciutat Meridiana	41.461113	2.175853



Data wrangling

Each of the dataframes underwent a process of data wrangling in which the data was cleaned and the relevant information was extracted to analyze the problem. After this, we show an adequate data visualization, in the case of the apartment rental by neighbourhood a bar chart is presented and for the other 3 dataframes the python folium library is used. Its location is shown on a map.

	Any	Indicador	Equipament	Valor	Notes_Dades	Notes_Equipament	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Titularitat	TipusGeneral	TipusEquipament	Ambit	Lati
0	2017	Assistents_Activitats	Biblioteca Barceloneta - La Fraternitat	1007	NaN	NaN	1	Ciutat Vella	3	la Barceloneta	Consorci o fundació amb presència municipal	Biblioteques de Barcelona	Biblioteques	Lletres	41.379
1	2017	Assistents_Activitats	Biblioteca Francesca Bonnemaison	2732	NaN	NaN	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	Consorci o fundació amb presència municipal	Biblioteques de Barcelona	Biblioteques	Lletres	41.386
2	2017	Assistents_Activitats	Biblioteca Gòtic - Andreu Nin	1430	NaN	NaN	1	Ciutat Vella	2	el Barri Gòtic	Consorci o fundació amb presència municipal	Biblioteques de Barcelona	Biblioteques	Lletres	41.378
3	2017	Assistents_Activitats	Biblioteca Sant Pau i Santa Creu	1009	NaN	NaN	1	Ciutat Vella	1	el Raval	Consorci o fundació amb presència municipal	Biblioteques de Barcelona	Biblioteques	Lletres	41.381
4	2017	Assistents_Activitats	Biblioteca Esquerra de l'Eixample-Agustí Cente...	4754	NaN	NaN	2	Eixample	9	la Nova Esquerra de l'Eixample	Consorci o fundació amb presència municipal	Biblioteques de Barcelona	Biblioteques	Lletres	41.386

	Library Name	Neighbourhood	Latitud	Longitud
0	Biblioteca Barceloneta - La Fraternitat	la Barceloneta	41.379201	2.188854
1	Biblioteca Francesca Bonnemaison	Sant Pere, Santa Caterina i la Ribera	41.386937	2.175790
2	Biblioteca Gòtic - Andreu Nin	el Barri Gòtic	41.378730	2.175979
3	Biblioteca Sant Pau i Santa Creu	el Raval	41.381057	2.169794
4	Biblioteca Esquerra de l'Eixample-Agustí Cente...	la Nova Esquerra de l'Eixample	41.386449	2.152790

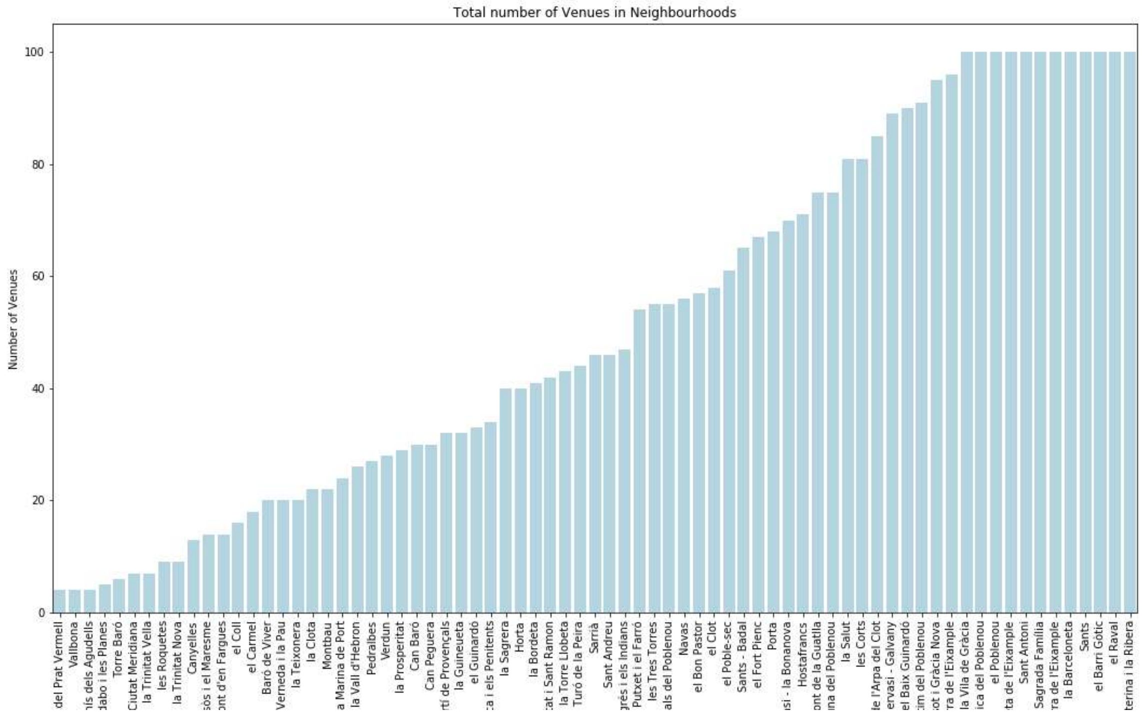
:

Foursquare

Using Foursquare we obtained the different venues for each neighbourhood. We designed the **limit as 100 venue** and the **radius 600 meter** for each neighbourhood from their given latitude and longitude. The dataframe has a size of *3743 rows*.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Baró de Viver	41.44603	2.199706	Ibericus	41.441505	2.197812	Food
1	Baró de Viver	41.44603	2.199706	Pasteleria Buenavista	41.442172	2.199977	Dessert Shop
2	Baró de Viver	41.44603	2.199706	Restaurant Enriqueta	41.445684	2.206801	Spanish Restaurant
3	Baró de Viver	41.44603	2.199706	Lefties	41.441089	2.198103	Clothing Store
4	Baró de Viver	41.44603	2.199706	A Loja do Gato Preto	41.441689	2.197742	Furniture / Home Store

In the following image it can be seen that not all neighbourhoods obtained 100 places. Neighbourhoods such as La Marina del Prat Vermell or Torre Baró have fewer than 10 places whereas more popular neighbourhoods like Sants, Sagrada Família of La dreta de l'Eixample did reach the limit of 100.



In summary there are **292 uniques categories**.

Preparing Categorical Data: venues

If you try to use categorical data without preprocessing in a machine learning model, you will get an error. One of the approaches to dealing with categorical data is 'One-Hot Encoding' which "creates new columns that indicate the presence (or absence) of each possible value in the original data." (Kaggle, 2019). We can see an example in the next image.

Color
Red
Red
Yellow
Green
Yellow

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0

We combined the 4 dataframes into a master table called 'result'.

Data	Name	Type	Approach
Rent Apartment	bcn_rent	Float	-
Libraries	bcn_lib	Categorical	OneHotEncoding
Transport	bcn_tran	Categorical	OneHotEncoding
Venues	bcn_venues	Categorical	OneHotEncoding

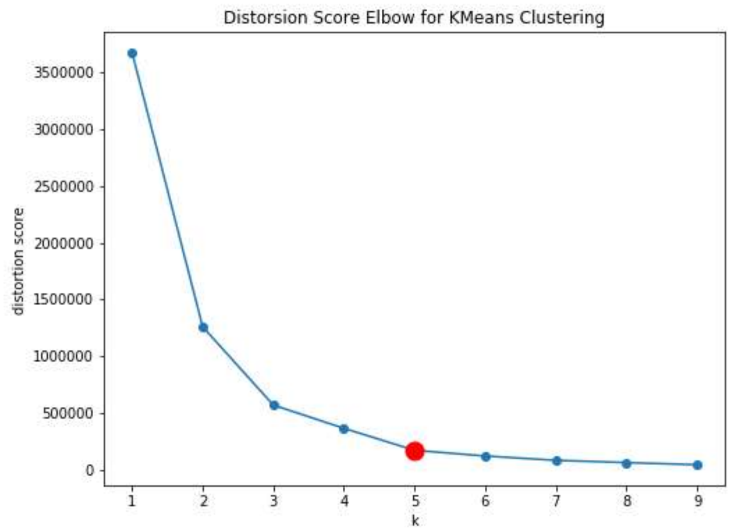
Result dataframe: 73 rows and 758 columns

	Aveg_Rent	Biblioteca Barceloneta - La Fraternitat	Biblioteca Bon Pastor	Biblioteca Camp de l'Arpa - Caterina Albert	Biblioteca Can Rosés	Biblioteca Canyelles	Biblioteca Clarà	Biblioteca Collserola - Josep Miracle	Biblioteca El Carmel - Juan Marsé	Biblioteca El Clot - Josep Benet	...	Video Game Store	Video Store	Vietnamese Restaurant	Water Park	Wine Bar	Wine Shop	Winery	Wings Joint	V
Baró de Viver	716.818900	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Can Baró	738.177500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Can Peguera	648.947561	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Canyelles	716.597500	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Ciutat Meridiana	506.965000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

5 rows × 757 columns

Unsupervised K-means algorithm

To find the appropriate number of clusters in a dataset k, we used the graphic method Elbow. The number of cluster was **k equal 5**.

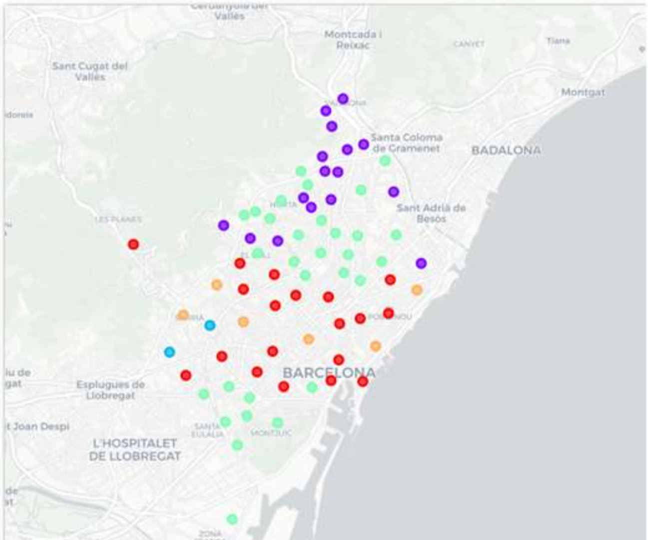


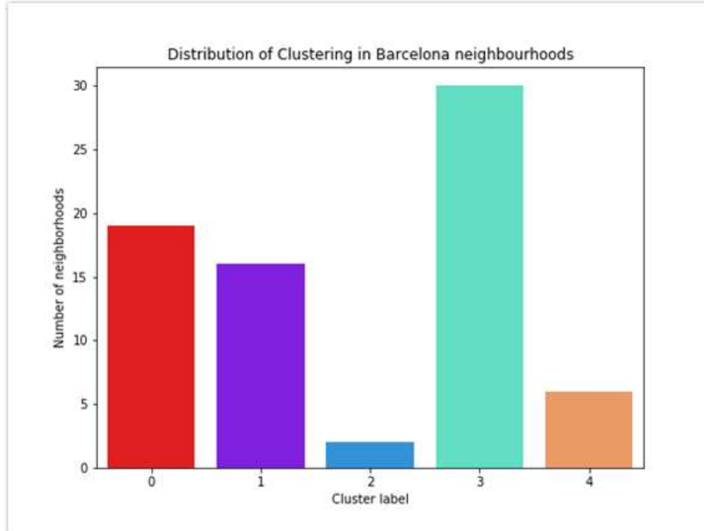
Results and discussion

For each neighbourhood a cluster of label between 0 and 4 was obtained. A new column was created in the dataframe called 'Cluster label'. The 10 most common places for each neighborhood can be appreciated as well.

	Neighbourhood	Cluster label	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	Baró de Viver	3	Spanish Restaurant	Clothing Store	Restaurant	Asian Restaurant	Food	Burger Joint	Bookstore	Steakhouse	Supermarket	Furniture / Home Store
1	Can Baró	3	Spanish Restaurant	Grocery Store	Plaza	Tapas Restaurant	Soccer Stadium	Park	Cambodian Restaurant	Café	Scenic Lookout	Chinese Restaurant
2	Can Peguera	1	Grocery Store	Supermarket	Pizza Place	Café	Tapas Restaurant	Breakfast Spot	Seafood Restaurant	Park	German Restaurant	Burger Joint
3	Canyelles	3	Soccer Field	Market	Plaza	Grocery Store	Food	Skate Park	Café	Seafood Restaurant	Mediterranean Restaurant	Bar
4	Ciutat Meridiana	1	Metro Station	Park	Train Station	Plaza	Supermarket	Mediterranean Restaurant	Fast Food Restaurant	Event Space	Exhibit	Falafel Restaurant
5	Diagonal Mar i el Front Marítim del Poblenou	4	Mediterranean Restaurant	Restaurant	Clothing Store	Hotel	Italian Restaurant	Coffee Shop	Café	Pizza Place	Burger Joint	Tapas Restaurant

The final distribution of the clusters can be seen in the following image:





Discussion

Barcelona is a great city with multiple rates, quite well connected and with an average income of 1000€. Consider these four main features for a specific foreign student: house rent, places around, means of transport and the distribution of public free libraries. The machine learning model shows us five groups, where those with the least neighborhoods are those with the highest rent. It can be appreciated that the neighborhoods of Pedralbes and Les Tres Torres (light blue) are the most expensive and that the Exaimple and Sarrià areas corresponding to group 4 have an average rental price of 1150€.

Conclusions

The main objective of the proposed application Ciro is to recommend to each student according to their personality and interests the best neighbourhoods to live in. In this analysis only the objective of segmenting the neighborhoods of Barcelona according to four features has been achieved, it would be necessary to complement this with an analysis of what type of client (student) each cluster corresponds to.