Capstone Project

Pablo de la Fuente Aguilera

The Battle of Neighborhoods

WEEK 2

Pablo de la Fuente Aguilera

Initial Questions

Initial Prologue

Who is carrying out this study?

A friend of us who wants to start a new and successful business in the area of car parking. He´s an architect and he is considering the idea to build an underground parking. He needs to choose a location in order to maximize the incomes of this business.

Where is he thinking about starting his project?

He considers that a good location could be a place surrounded by hotels and car renting services, so he could have some fix clients. He should consider also to build the business in a country or city where there´s a big business and travel life.

Our friend lives in Spain and according to this conditions, Madrid could be a right city due to all its economic activity.

Taking this considerations, Madrid is famous all over the world for having a big density of offices, hotels and main locations of big companies.

Pablo de la Fuente Aguilera

Initial Questions

Initial Prologue

Who are the stakeholders?

The main customers would be the businessmen who visit Madrid. This people usually need to rent a car during their time in the city or to leave their car in a safe place while they´re staying in a Hotel.

Which kind of location?

We need to find something that matches the next prerequisites:

-The price of the ground shouldn´t be to much high

-But must be inside Madrid city center

-It must have a good communication with highways and roads

Pablo de la Fuente Aguilera

Initial Questions

Initial Prologue

Advantages

- It can be made quickly and it can be stored for some time before being sold
- It´s a light food which contains vegetables, fish and rice. The perfect match to meet your feeding requirements in the meal and the dinner
- It doesn´t need to be heated in the microwave and can be eaten while working with almost no risk of having an accident with your papers
- It´s easy to package and transport
- It´s a fashion food and it´s not cheap, so there´s a considerable margin of benefits

Pablo de la Fuente Aguilera

Presentation of the Problem

Initial Presentation

The Battle of Neighborhoods
Week 1

To sum up the Problem

We can suppose that we are thinking about building an underground parking in Madrid. This could be a good business opportunity but we need to carry out a market research in order to establish a long-term success.

To start with, we will analyse the existing hotels and rental services in the different neighborhoods. To do that we will use Foursquare venues and information of the geographical location of every restaurant.

After that we will sort them by neighborhood, in order to identify the best possible location and determine which neighborhood inside Madrid could be the best place.

At the end, we will identify, based on filtering and ponderation of candidates, the best possible location in Madrid. We will support our decision in a map, to give a easy understanding of the problem.

Capstone Project

Pablo de la Fuente Aguilera

Working with Information

Analysis of the Madrid neighborhoods and filter the posible locations

The Battle of Neighborhoods
Week 2

We take data from all Boroughs and prepare it in a json file to be used. To do this, we download the data from the Govern of Madrid Website https://datos.comunidad.madrid and we combine it with the prices of the square meters in every neighborhood, provided and downloaded by the Website https://www.idealista.com/

We load them and build on a Dataframe:

```
# We show the Dataframe
neighborhoods.head()
```

	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
0	Centro	Palacio	-3.713134	40.415325	1.46	15323.287671	4852.0
1	Centro	Embajadores	-3.702543	40.409444	1.03	43345.631068	4479.0
2	Centro	Cortes	-3.696785	40.415439	0.59	17850.847458	5272.0
3	Centro	Justicia	-3.695976	40.423497	0.75	21866.666667	5893.0
4	Centro	Universidad	-3.706963	40.426121	0.93	33051.612903	5282.0

[163]

```
print("We have", neighborhoods.shape[0], " different neighborhoods in total")
```

We have 128 different neighborhoods in total

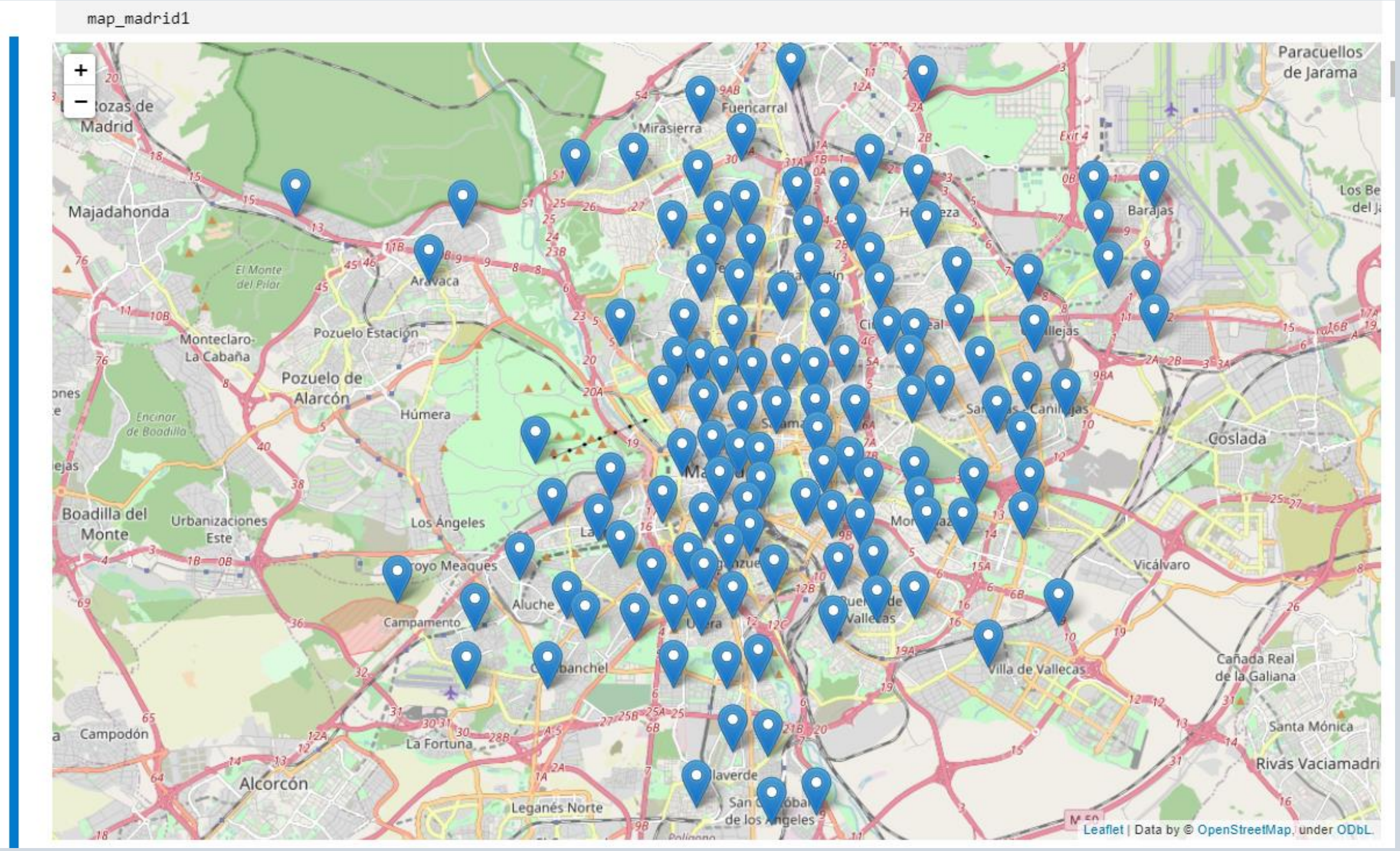
Capstone Project

Pablo de la Fuente Aguilera

Working with Information

Analysis of the Madrid neighborhoods and filter the posible locations

We take the coordinates of Sol (the most central and touristic point in Madrid) as the reference to represent our different neighborhoods from now on in a map. We use Folium to show them:



The Battle of Neighborhoods
Week 2

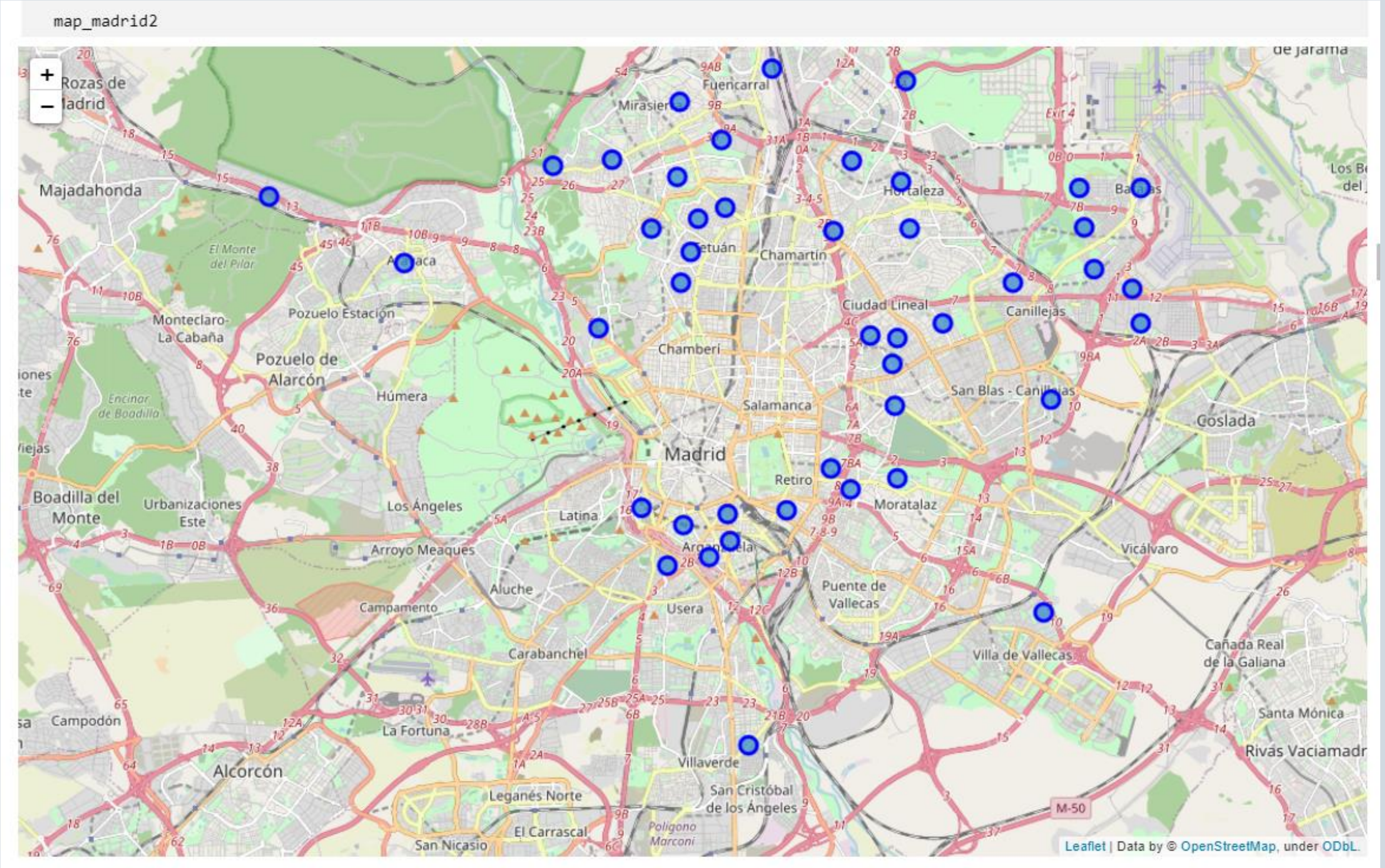
Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Analysis of the Madrid neighborhoods and filter the posible locations

Due to budget restrictions, we apply a filter between 2600 and 4200 €/m². With this change we have now only 44 neighborhoods from the initial 128.



The Battle of Neighborhoods
Week 2

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Clustering the candidates

The Battle of Neighborhoods
Week 2

To understand and visualize better our candidates, we cluster them into 4 groups

Cluster Nr. 0

[19]

```
barrios_cluster_0=barrios_data1.loc[barrios_data1['Cluster Labels'] == 0]
barrios_cluster_0.head(barrios_cluster_0.shape[0])
```

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
12	0	Fuencarral-El Pardo	PeÃ±a Grande	-3.726833	40.479017	2.86	15476.923077	3283.0
14	0	Fuencarral-El Pardo	La Paz	-3.696045	40.483378	2.17	15578.801843	3580.0
22	0	Moratalaz	Marroquina	-3.647142	40.411364	1.75	15633.714286	2868.0
23	0	Moratalaz	Media Legua	-3.660231	40.408978	1.02	17466.666667	2664.0
24	0	Ciudad Lineal	Ventas	-3.647657	40.426883	3.22	14700.931677	2665.0
27	0	Ciudad Lineal	San Pascual	-3.654677	40.441785	1.06	17079.245283	3406.0
30	0	Hortaleza	Canillas	-3.643754	40.464324	2.56	15820.312500	3241.0
32	0	Hortaleza	Apostol Santiago	-3.659881	40.478759	1.20	12627.500000	2850.0
35	0	Villa de Vallecas	Santa Eugenia	-3.606203	40.382919	2.04	11822.058824	2695.0
39	0	Barajas	Alameda de Osuna	-3.592139	40.455778	1.98	9795.454545	3273.0
41	0	Barajas	Casco HistÃ³rico de Barajas	-3.579222	40.472985	0.64	11415.625000	3104.0

Cluster Nr. 1

[20]

```
barrios_cluster_1=barrios_data1.loc[barrios_data1['Cluster Labels'] == 1]
barrios_cluster_1.head(barrios_cluster_1.shape[0])
```

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
1	1	Arganzuela	Las Acacias	-3.706734	40.401523	1.10	33318.181818	4188.0
2	1	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
4	1	Arganzuela	Palos de Moguer	-3.694600	40.403759	0.65	39372.307692	3943.0
5	1	Retiro	PacÃfico	-3.677976	40.404721	0.76	44271.052632	4149.0
7	1	Tetuan	Bellas Vistas	-3.707527	40.453071	0.73	38923.287671	3365.0
10	1	Tetuan	Berruguete	-3.704891	40.459591	0.60	40533.333333	3263.0
13	1	Fuencarral-El Pardo	El Pilar	-3.708572	40.475544	1.37	33451.824818	3303.0
21	1	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0
25	1	Ciudad Lineal	Quintana	-3.648602	40.435811	0.71	34056.338028	2828.0

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Clustering the candidates

Cluster Nr. 2

[21]

```
barrios_cluster_2=barrios_data1.loc[barrios_data1['Cluster Labels'] == 2]
barrios_cluster_2.head(barrios_cluster_2.shape[0])
```

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
11	2	Fuencarral-El Pardo	Fuentelarreina	-3.743380	40.477849	1.51	2113.907285	3439.0
15	2	Fuencarral-El Pardo	Valverde	-3.682246	40.498379	9.03	6642.857143	3462.0
16	2	Fuencarral-El Pardo	Mirasierra	-3.707996	40.491460	6.91	4465.701881	3645.0
17	2	Moncloa-Aravaca	Ciudad Universitaria	-3.730594	40.443199	14.14	1135.007072	3652.0
19	2	Moncloa-Aravaca	El PlantÃo	-3.822546	40.471137	3.56	764.887640	3742.0
20	2	Moncloa-Aravaca	Aravaca	-3.784802	40.457048	5.86	4384.300341	3742.0
28	2	Ciudad Lineal	Atalaya	-3.665093	40.463866	0.25	6240.000000	3020.0
29	2	Hortaleza	Palomas	-3.614780	40.452846	1.14	5859.649123	3732.0
33	2	Hortaleza	Valdefuentes	-3.644558	40.495833	15.19	3400.526662	3813.0
36	2	San Blas-Canillejas	Rosas	-3.604156	40.428114	9.26	3403.023758	3121.0
37	2	San Blas-Canillejas	Rejas	-3.579308	40.444413	4.98	3165.662651	3022.0
38	2	San Blas-Canillejas	El Salvador	-3.634497	40.444446	1.86	6112.903226	3074.0
40	2	Barajas	Aeropuerto	-3.581604	40.451745	19.87	89.884248	3203.0
42	2	Barajas	TimÃ³n	-3.596474	40.472952	16.40	672.012195	3088.0
43	2	Barajas	Corralejos	-3.594800	40.464725	4.67	1577.087794	3476.0

Cluster Nr. 3

[22]

```
barrios_cluster_3=barrios_data1.loc[barrios_data1['Cluster Labels'] == 3]
barrios_cluster_3.head(barrios_cluster_3.shape[0])
```

	Cluster Labels	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
0	3	Arganzuela	Imperial	-3.718656	40.405161	0.98	23105.102041	3922.0
3	3	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
6	3	Retiro	Estrella	-3.665524	40.413454	1.02	22701.960784	4155.0
8	3	Tetuan	Almenara	-3.695287	40.468889	0.99	22232.323232	3419.0
9	3	Tetuan	Valdeacederas	-3.702735	40.466435	1.17	21574.358974	3099.0
18	3	Moncloa-Aravaca	Valdezarza	-3.715617	40.464359	1.44	20317.361111	2989.0
26	3	Ciudad Lineal	ConcepciÃ³n	-3.647228	40.441070	0.90	22828.888889	3259.0
31	3	Hortaleza	Pinar del Rey	-3.646148	40.474450	2.64	19682.954545	2887.0
34	3	Villaverde	Los Rosales	-3.688816	40.354694	1.51	24163.576159	2869.0

The Battle of Neighborhoods
Week 2

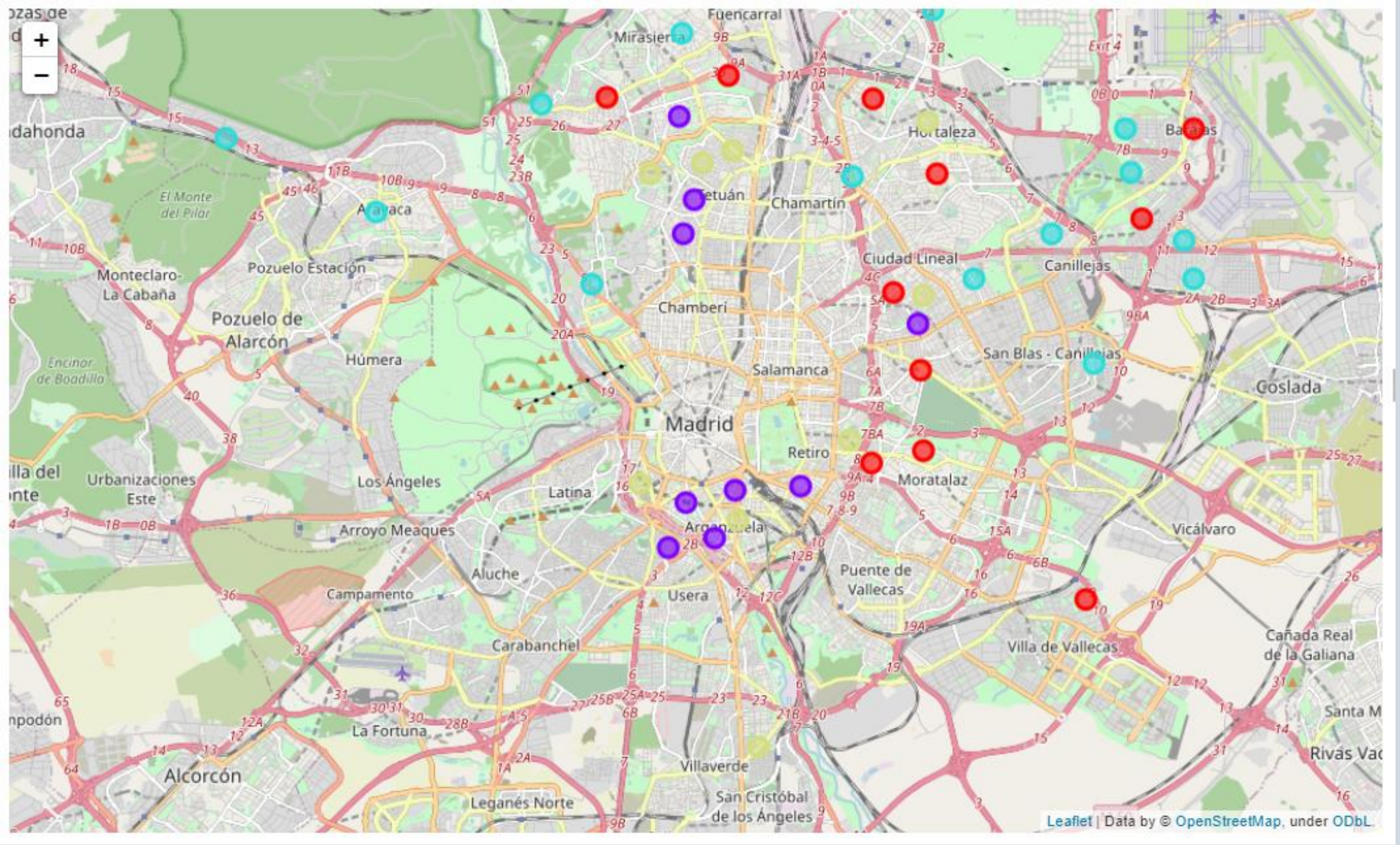
Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Clustering the candidates

And represent them into a new map:



The Battle of Neighborhoods
Week 2

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Using and implementing Foursquare API

The Battle of Neighborhoods
Week 2

We charge the API from Foursquare and download all the categories list and filter them using key words:

	name	id
0	Arts & Entertainment	4d4b7104d754a06370d81259
1	Amphitheater	56aa371be4b08b9a8d5734db
2	Aquarium	4fceea171983d5d06c3e9823
3	Arcade	4bf58dd8d48988d1e1931735
4	Art Gallery	4bf58dd8d48988d1e2931735

[194]

```
print('We have in total ', dfCategoriesRaw.shape[0],' different categories')
```

We have in total 941 different categories

Now, we filter the categories using key words

[195]

```
dfCategoriesFilter = dfCategoriesRaw[dfCategoriesRaw['name'].str.contains('(Hotel|Rental|Car Rent|Taxi|Uber|Airbnb|Driver)',
regex = True, case=False)].sort_values('name').reset_index(drop = True)
dfCategoriesFilter.head(1000)
```

	name	id
0	Bike Rental / Bike Share	4e4c9077bd41f78e849722f9
1	Boat Rental	5744ccdfe4b0c0459246b4c1
2	Hotel	4bf58dd8d48988d1fa931735
3	Hotel Bar	4bf58dd8d48988d1d5941735
4	Hotel Pool	4bf58dd8d48988d132951735
5	Rental Car Location	4bf58dd8d48988d1ef941735
6	Rental Service	56aa371be4b08b9a8d573552
7	Taxi	4bf58dd8d48988d130951735
8	Taxi Stand	53fca564498e1a175f32528b
9	Vacation Rental	56aa371be4b08b9a8d5734e1

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Using and implementing Foursquare API

The Battle of Neighborhoods
Week 2

We charge the API from Foursquare and download all the categories list and filter them using key words:

[196]

```
dfCategories = dfCategoriesFilter[dfCategoriesFilter['name'].str.contains('^((?!(Boat|Pool|Vacation|Bar)).)*$', regex = True,
case=False)].sort_values('name').reset_index(drop = True)
dfCategories.head(100)
```

	name	id
0	Bike Rental / Bike Share	4e4c9077bd41f78e849722f9
1	Hotel	4bf58dd8d48988d1fa931735
2	Rental Car Location	4bf58dd8d48988d1ef941735
3	Rental Service	56aa371be4b08b9a8d573552
4	Taxi	4bf58dd8d48988d130951735
5	Taxi Stand	53fca564498e1a175f32528b

[197]

```
print('Now we have only ', dfCategories.shape[0],' different categories')
```

```
Now we have only  6  different categories
```

Finally we have a reduced number of categories. And they have these categories id´s:

[198]

```
dfCategories['id'].to_list()
```

```
['4e4c9077bd41f78e849722f9',
 '4bf58dd8d48988d1fa931735',
 '4bf58dd8d48988d1ef941735',
 '56aa371be4b08b9a8d573552',
 '4bf58dd8d48988d130951735',
 '53fca564498e1a175f32528b']
```

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Using and implementing Foursquare API

The Battle of Neighborhoods
Week 2

Using the filtered neighborhood names, we take all the venues according to those categories:

Now we take all the Venues in the filtered Neighborhoods

[199]

```
#dfVenues = getNearbyVenues(barrios_data1['Neighborhood'],barrios_data1['Latitude'], barrios_data1['Longitude'], dfCategories
['id'].to_list())
dfVenues = getNearbyVenues2(names=barrios_data1['Neighborhood'], latitudes=barrios_data1['Latitude'], longitudes=barrios_data1
['Longitude'], radius=1000, categoryIds='')
dfVenues.head(200)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Imperial	-3.718656	40.405161	Europcar	-3.746100	40.327000	Rental Car Location
1	Imperial	-3.718656	40.405161	Europcar	-3.789800	40.464700	Rental Car Location
2	Las Acacias	-3.706734	40.401523	Europcar	-3.746100	40.327000	Rental Car Location
3	Las Acacias	-3.706734	40.401523	Europcar	-3.789800	40.464700	Rental Car Location
4	La Chopera	-3.699556	40.394639	Holiday Inn Express	-3.692243	40.273166	Hotel
5	La Chopera	-3.699556	40.394639	Europcar	-3.789800	40.464700	Rental Car Location
6	La Chopera	-3.699556	40.394639	Europcar	-3.746100	40.327000	Rental Car Location
7	Delicias	-3.693955	40.398093	Europcar	-3.746100	40.327000	Rental Car Location
8	Delicias	-3.693955	40.398093	Holiday Inn Express	-3.692243	40.273166	Hotel
9	Delicias	-3.693955	40.398093	Europcar	-3.789800	40.464700	Rental Car Location
10	Palos de Moguer	-3.694600	40.403759	Europcar	-3.789800	40.464700	Rental Car Location
11	Palos de Moguer	-3.694600	40.403759	Europcar	-3.746100	40.327000	Rental Car Location
12	PacÃfico	-3.677976	40.404721	Europcar	-3.746100	40.327000	Rental Car Location
13	Estrella	-3.665524	40.413454	Europcar	-3.570300	40.462800	Rental Car Location
14	Bellas Vistas	-3.707527	40.453071	Europcar	-3.789800	40.464700	Rental Car Location
15	Almenara	-3.695287	40.468889	Europcar	-3.789800	40.464700	Rental Car Location
16	Valdeacederas	-3.702735	40.466435	Europcar	-3.789800	40.464700	Rental Car Location
17	Berruguete	-3.704891	40.459591	Europcar	-3.789800	40.464700	Rental Car Location
18	Fuentelarreina	-3.743380	40.477849	Europcar	-3.789800	40.464700	Rental Car Location
19	PeÃ±a Grande	-3.726833	40.479017	Europcar	-3.789800	40.464700	Rental Car Location
20	El Pilar	-3.708572	40.475544	Europcar	-3.789800	40.464700	Rental Car Location

Capstone Project

Pablo de la Fuente Aguilera

Processing the Information

Taking a decisión from the Venues

We group the venues by neighborhood and stablish a classification:

[203]

```
#dfVenues_grouped = dfVenues_onehot.groupby('Neighborhood').mean().reset_index()
dfVenues_grouped = dfVenues_onehot.groupby('Neighborhood', axis=0).sum()
dfVenues_grouped
```

Neighborhood	Hotel	Rental Car Location
Aeropuerto	1	1
Alameda de Osuna	1	1
Almenara	0	1
Apostol Santiago	0	1
Aravaca	0	2
Atalaya	0	1
Bellas Vistas	0	1
Berruguete	0	1
Canillas	0	1
Casco HistÃ³rico de Barajas	1	1
Ciudad Universitaria	0	1
Comillas	1	2
ConcepciÃ³n	0	1
Corralejos	1	1
Delicias	1	2
El Pilar	0	1
El PlantÃo	0	2
El Salvador	0	1

Here is our clasification according to our point system:

Neighborhood	Hotel	Rental Car Location	Points
La Chopera	1	2	3
Delicias	1	2	3
Los Rosales	1	2	3
Comillas	1	2	3
Santa Eugenia	1	1	2
Rosas	1	1	2
Rejas	1	1	2
Palos de Moguer	0	2	2
Palomas	1	1	2
Marroquina	1	1	2
Las Acacias	0	2	2
Alameda de Osuna	1	1	2
Imperial	0	2	2

The Battle of Neighborhoods
Week 2

Pablo de la Fuente Aguilera

Processing the Information

Taking a decisión from the Venues

We take the names of the four candidates with the highest score and take all the data from the initial dataframe according to them:

Here we have the Neighborhoods with the highest score:

	Neighborhood
0	La Chopera
1	Delicias
2	Los Rosales
3	Comillas

We take their values in the full table

[208]

```
#neighborhoods_cand = neighborhoods[neighborhoods['Neighborhood']==dfCandBarrios['Neighborhood'].to_list()]
neighborhoods_cand = neighborhoods[neighborhoods['Neighborhood'].isin(dfCandBarrios['Neighborhood'])]
neighborhoods_cand.head()
```

	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
8	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
10	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
65	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0
109	Villaverde	Los Rosales	-3.688816	40.354694	1.51	24163.576159	2869.0

Capstone Project

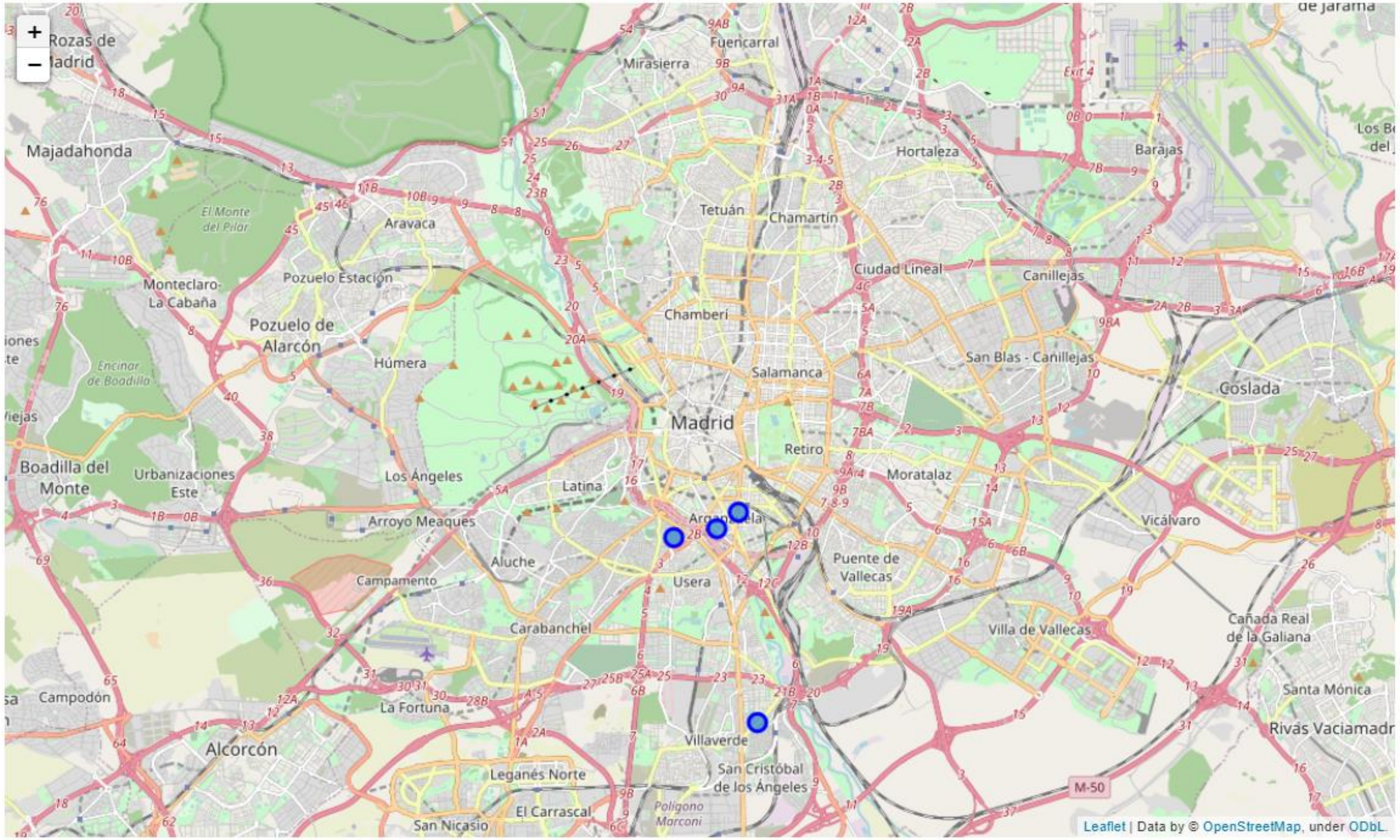
Pablo de la Fuente Aguilera

Solution of the Problem

Final decision

We present them in a new map using Folium:

map_madrid3



The Battle of Neighborhoods

Week 2

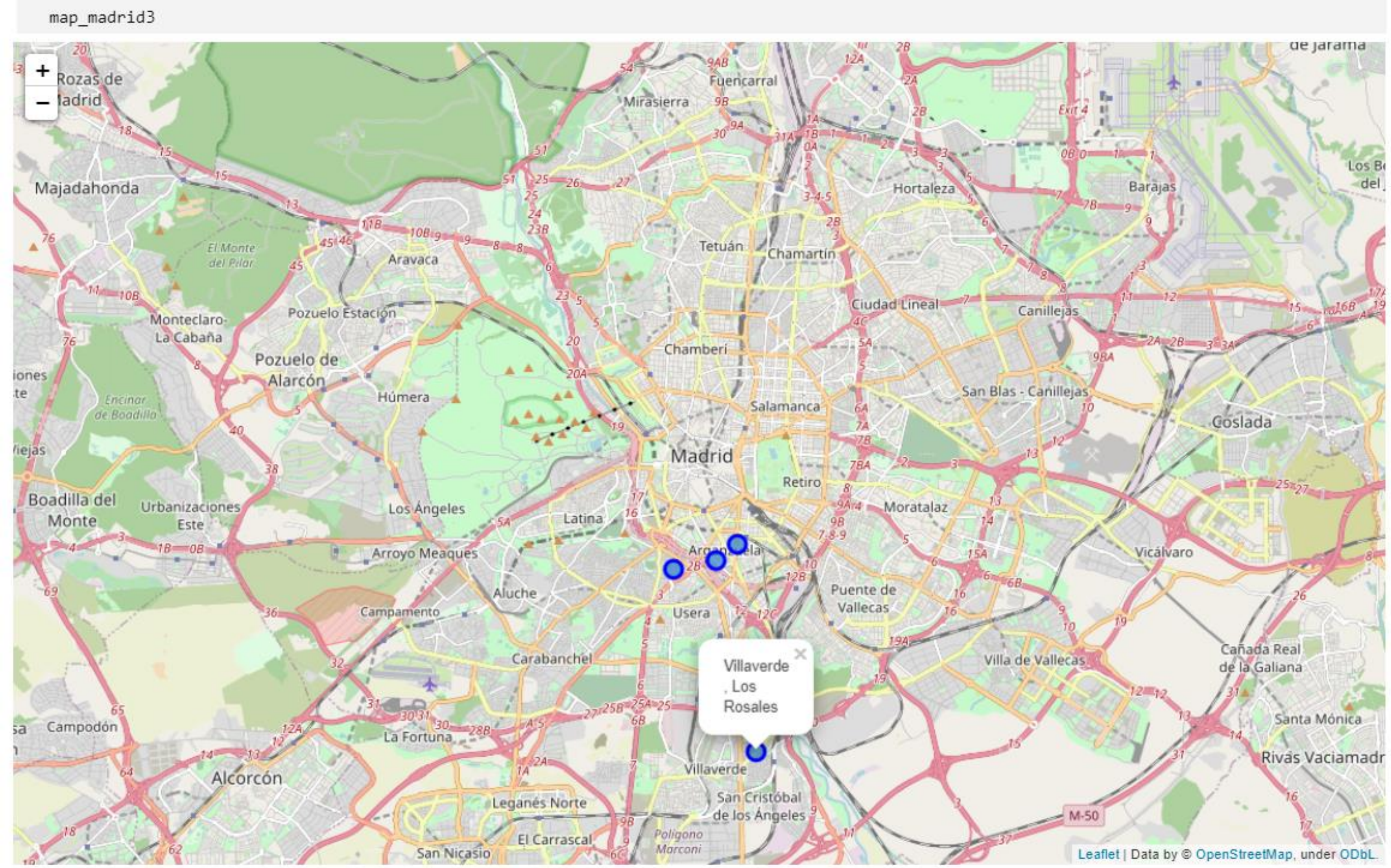
Capstone Project

Pablo de la Fuente Aguilera

Solution of the Problem

Final decision

We can apreciate that Los Rosales (Villaverde) is much far away from the city center than the other 3 candidate locations:



The Battle of Neighborhoods
Week 2

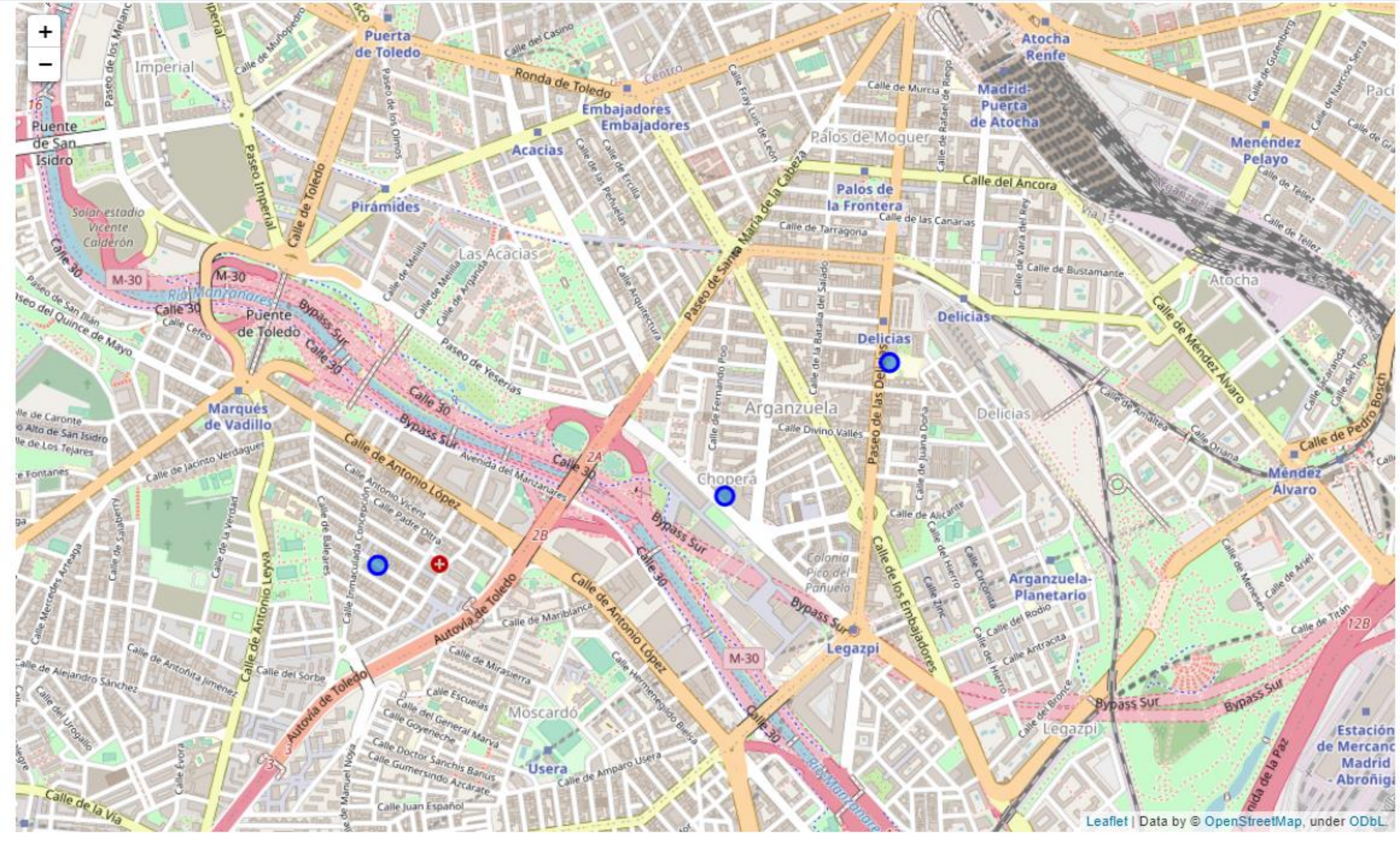
Capstone Project

Pablo de la Fuente Aguilera

Solution of the Problem

Final decision

We concentrate ourselves in the other three locations:



	District	Neighborhood	Latitude	Longitude	Surface (km2)	Density (hab/km2)	Price (€/m2)
8	Arganzuela	La Chopera	-3.699556	40.394639	0.56	35276.785714	3625.0
10	Arganzuela	Delicias	-3.693955	40.398093	1.07	25485.046729	3904.0
65	Carabanchel	Comillas	-3.711391	40.392854	0.67	33000.000000	2702.0

The Battle of Neighborhoods
Week 2

Capstone Project

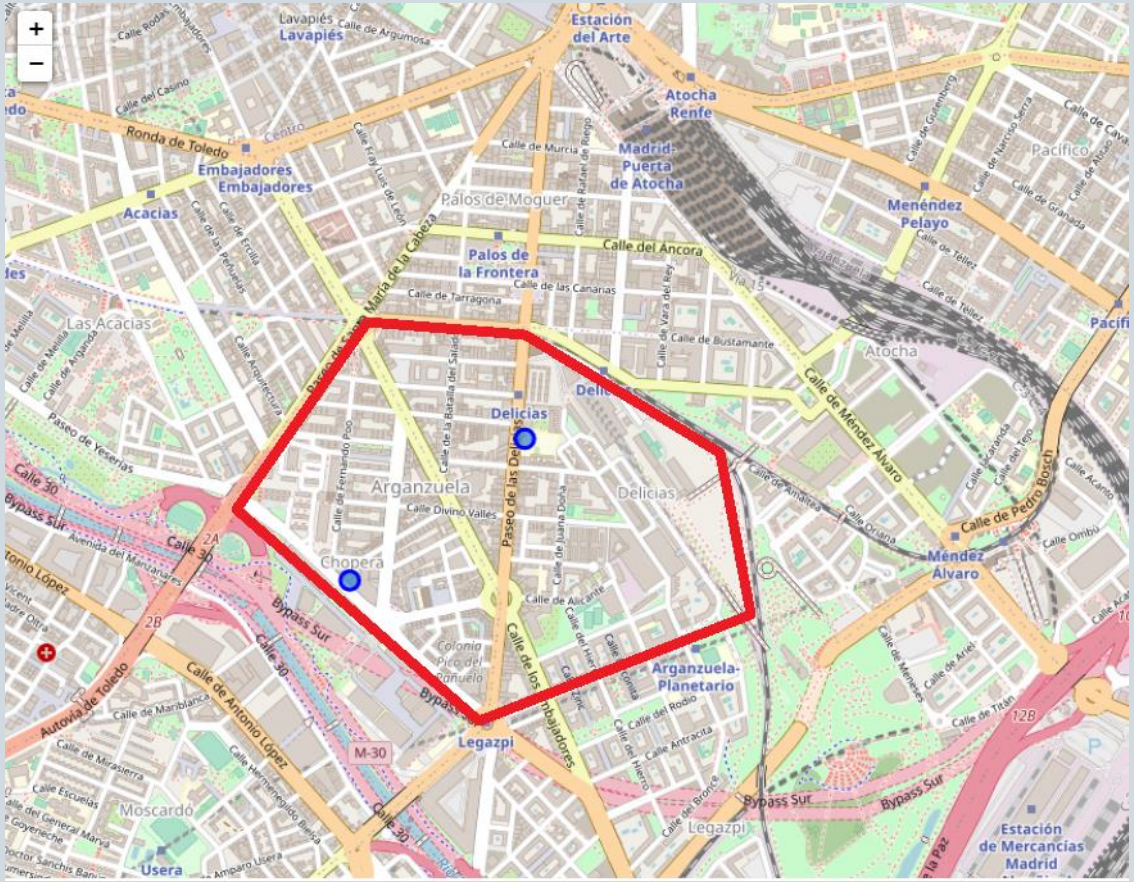
Pablo de la Fuente Aguilera

Solution of the Problem

Final decision

In the map we can see that Delicias and Chopera have a really good location, between the highway Calle 30 (which surrounds with a circular ring all the center of the city) and Atocha Train Station (the biggest Train Station in Spain).

As a result, we will build our parking in the area in red:



The Battle of Neighborhoods
Week 2