

New Vegan Restaurant in Madrid



Location recommendation report.



Problem:



- Vegan restaurants are on the rise in Europe. It's still a good opportunity, but it takes timely market research.
- Today more than ever, the profession of chef is especially valued.
- From an economic point of view, in the most successful cases a return of 25% has been achieved.
- Is vital to define our business concept, its structure, positioning, location and elements of the marketing mix.

Description of the data



- For this work we will first use the data provided by the government of the City of Madrid. <http://www.madrid.org/iestadis/fijas/clasificaciones/descarga/cobar18.xls>
- A second file is obtained which can be configured for downloading at the town hall site with all the coordinates of the neighborhood of the capital of Spain. <https://www.madrid.es/portal/site/munimadrid>
- From the local government statistics website, a new downloadable one can be configured with the information on the evolution of the population in the last ten years. <http://www-2.munimadrid.es/TSE6/control/seleccionDatosBarrio>

Data overview



Government of the City of Madrid
(All neighborhoods by districts).

	A	B	C	D	E	F
1	munic	distr	ldistr	barrio	descrip	secci
2	0796	01	Centro	1	Palacio	001
3	0796	01	Centro	1	Palacio	002
4	0796	01	Centro	1	Palacio	003
5	0796	01	Centro	1	Palacio	004
6	0796	01	Centro	1	Palacio	006
7	0796	01	Centro	1	Palacio	007
8	0796	01	Centro	1	Palacio	008
9	0796	01	Centro	1	Palacio	009
10	0796	01	Centro	1	Palacio	011
11	0796	01	Centro	1	Palacio	012
12	0796	01	Centro	1	Palacio	013
13	0796	01	Centro	1	Palacio	014
14	0796	01	Centro	1	Palacio	015
15	0796	01	Centro	1	Palacio	016
16	0796	01	Centro	1	Palacio	018
17	0796	01	Centro	1	Palacio	019
18	0796	01	Centro	1	Palacio	020
19	0796	01	Centro	1	Palacio	021
20	0796	01	Centro	2	Embajadores	022
21	0796	01	Centro	2	Embajadores	023
22	0796	01	Centro	2	Embajadores	024
23	0796	01	Centro	2	Embajadores	025
24	0796	01	Centro	2	Embajadores	026
25	0796	01	Centro	2	Embajadores	027
26	0796	01	Centro	2	Embajadores	028
27	0796	01	Centro	2	Embajadores	029
28	0796	01	Centro	2	Embajadores	030
29	0796	01	Centro	2	Embajadores	031
30	0796	01	Centro	2	Embajadores	032
31	0796	01	Centro	2	Embajadores	033

Town Hall
(Coordinates).

	A	B	C
1	BARRIOS	LATITUD	LONGITUD
2	Abrantes	40.380556	-3.723889
3	Acacias	40.401422	-3.704936
4	Adelfas	40.400278	-3.670833
5	Aeropuerto	40.494167	-3.566944
6	Alameda de Osuna	40.457222	-3.587778
7	Almagro	40.431667	-3.694167
8	Almenara	40.47129	-3.695505
9	Almendrales	40.383611	-3.698889
10	Aluche	40.3875	-3.754167
11	Amposta	40.430278	-3.619444
12	Apostol Santiago	40.474992	-3.662278
13	Arapiles	40.434167	-3.707778
14	Aravaca	40.45	-3.783333
15	Arcos	40.423889	-3.613056
16	Argüelles	40.429161	-3.718847

Local government statistics website
(evolution of the population).

BARRIO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Amposta	6439	6984	7122	7223	7184	7098	6872	6954	7023	7273	7301	7298	7435	7689	7851	7997	8284	8342	8471
Atalaya	768	801	826	899	961	981	1076	1193	1268	1275	1325	1376	1498	1503	1501	1537	1559	1654	1699
Chopera	14237	15432	16065	16707	17451	17948	18237	18399	18358	18139	17645	17091	17345	17110	16851	16759	16500	16333	16420

Methodology section

(data processing)



- The first thing is to process the datasets to get a clean list of neighborhoods with coordinates.

```
df_all_madrid.head()
```

	distr	Idistr	barrio	descrip	secci
munic					
796	1	Centro	1	Palacio	1
796	1	Centro	1	Palacio	2
796	1	Centro	1	Palacio	3
796	1	Centro	1	Palacio	4
796	1	Centro	1	Palacio	6



```
df = df_coord.rename(columns={"BARRIOS":  
df.head()
```

	Neighborhood	Latitud	Longitud
0	Abrantes	40.380556	-3.723889
1	Acacias	40.401422	-3.704936
2	Adelfas	40.400278	-3.670833
3	Aeropuerto	40.494167	-3.566944
4	Alameda de Osuna	40.457222	-3.587778

Methodology section

(Geolocator and Folium)



- We use **geolocator** to find coordinates of the city.

```
address = 'Madrid'

geolocator = Nominatim(user_agent="madrid_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Madrid City are {}, {}'.format(latitude, longitude))

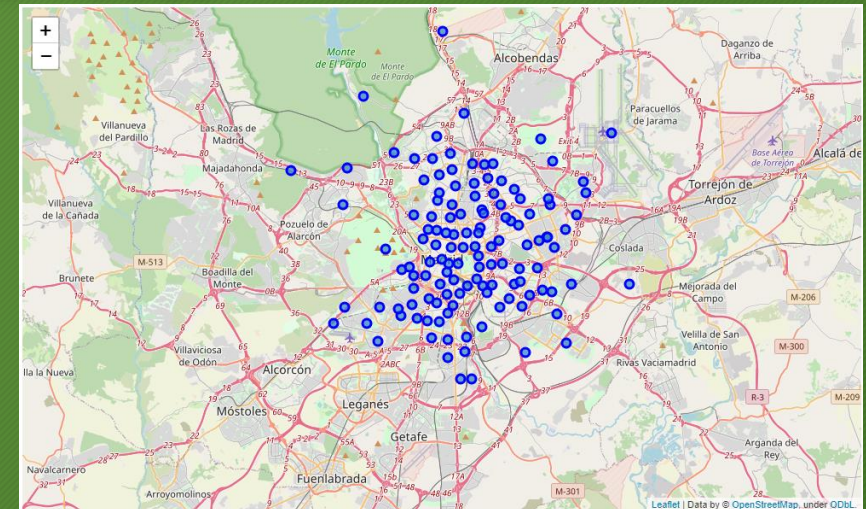
The geograpical coordinate of Madrid City are 40.4167047, -3.7035825.
```

- At this point we are ready for visualize the neighborhoods distribution using **folium**:

```
map_madrid = folium.Map(location=[latitude, longitude], zoom_start=11)

for lat, lng, neighborhood in zip(df['Latitud'], df['Longitud'], df['Neighborhood']):
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_madrid)

map_madrid
```



Methodology section

(Foursquare API)



- To get the total number of places that offer vegan food by neighborhood we use Foursquare.
 - `categoryIds='4bf58dd8d48988d1d3941735'`

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Acacias	40.401422	-3.704936	Distrito Vegano	40.406972	-3.699679	Vegetarian / Vegan Restaurant
1	Acacias	40.401422	-3.704936	La Tia Carlota	40.407516	-3.699776	Vegetarian / Vegan Restaurant
2	Acacias	40.401422	-3.704936	la oveja negra	40.409322	-3.699811	Vegetarian / Vegan Restaurant
3	Acacias	40.401422	-3.704936	El Granero	40.408275	-3.697631	Vegetarian / Vegan Restaurant
4	Acacias	40.401422	-3.704936	Yatiri	40.412690	-3.703373	Vegetarian / Vegan Restaurant

`madrid_venues_vegan.shape`

`(584, 7)`

584 records of vegan restaurants.

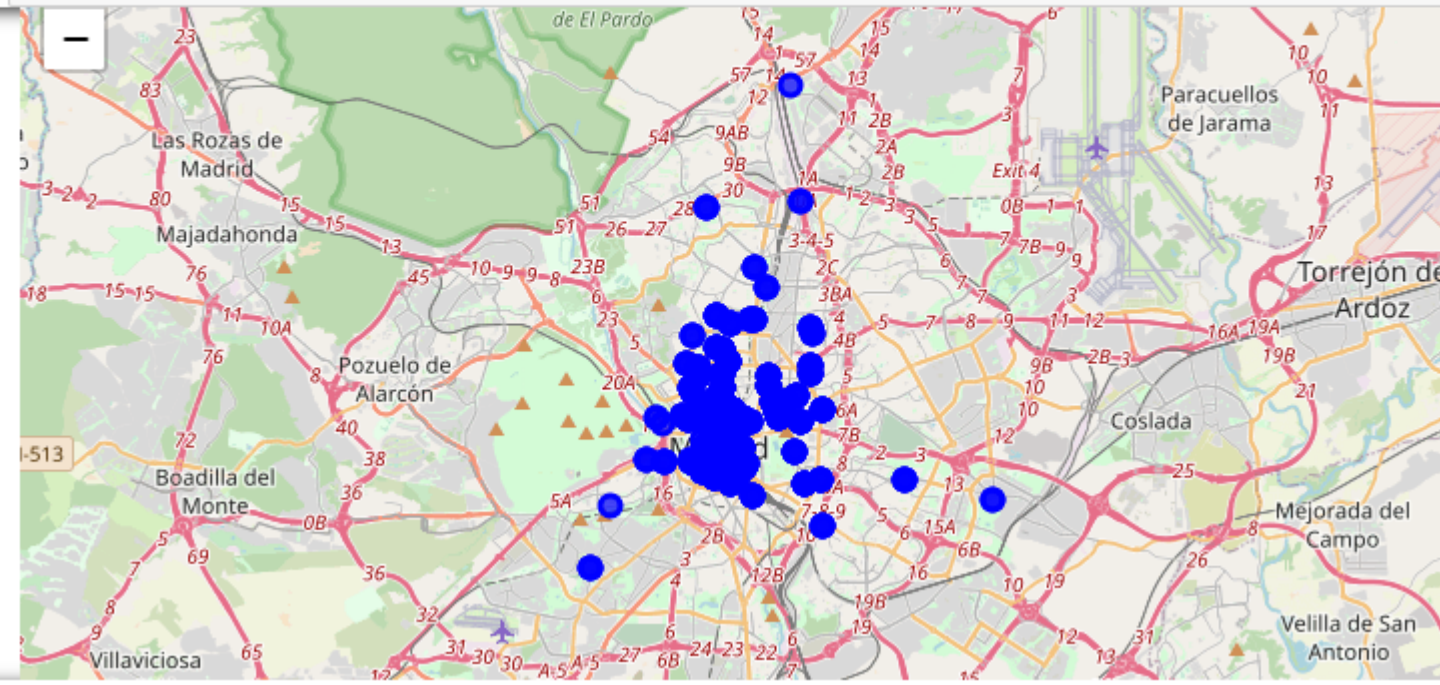
Methodology section

(Folium for Restaurants distribution)



```
map_madrid_vegan = folium.Map(location=[latitude, longitude], zoom_start=12)  
addToMap(madrid_venues_vegan, 'blue', map_madrid_vegan)
```

map_madrid_vegan



Methodology section

(Processing results)



- How many per neighborhood:

```
madrid_grouped = madrid_venues_vegan.groupby('Neighborhood').count()  
madrid_grouped
```

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Acacias	8	8	8	8	8	8
Adelfas	3	3	3	3	3	3
Almagro	28	28	28	28	28	28
Almenara	1	1	1	1	1	1
Aluche	1	1	1	1	1	1
Arapiles	29	29	29	29	29	29
Argüelles	18	18	18	18	18	18
Atocha	2	2	2	2	2	2
Bellas Vistas	3	3	3	3	3	3
Berruguete	2	2	2	2	2	2
Casco Histórico de Vicálvaro	1	1	1	1	1	1
Castellana	10	10	10	10	10	10

Methodology section

(Processing the results)



- The most common categories by neighborhood that include vegan food:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# Create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# Create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = madrid_grouped['Neighborhood']

for ind in np.arange(madrid_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(
        madrid_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Acacias	Vegetarian / Vegan Restaurant	Thai Restaurant	Spanish Restaurant	Sandwich Place	Salad Place
1	Adelfas	Vegetarian / Vegan Restaurant	Thai Restaurant	Spanish Restaurant	Sandwich Place	Salad Place
2	Almagro	Vegetarian / Vegan Restaurant	Spanish Restaurant	Pizza Place	Organic Grocery	Indian Restaurant
3	Almenara	Vegetarian / Vegan Restaurant	Thai Restaurant	Spanish Restaurant	Sandwich Place	Salad Place
4	Aluche	Vegetarian / Vegan Restaurant	Thai Restaurant	Spanish Restaurant	Sandwich Place	Salad Place

Methodology section

(k-means clustering)



- Split the group of restaurants by clusters:

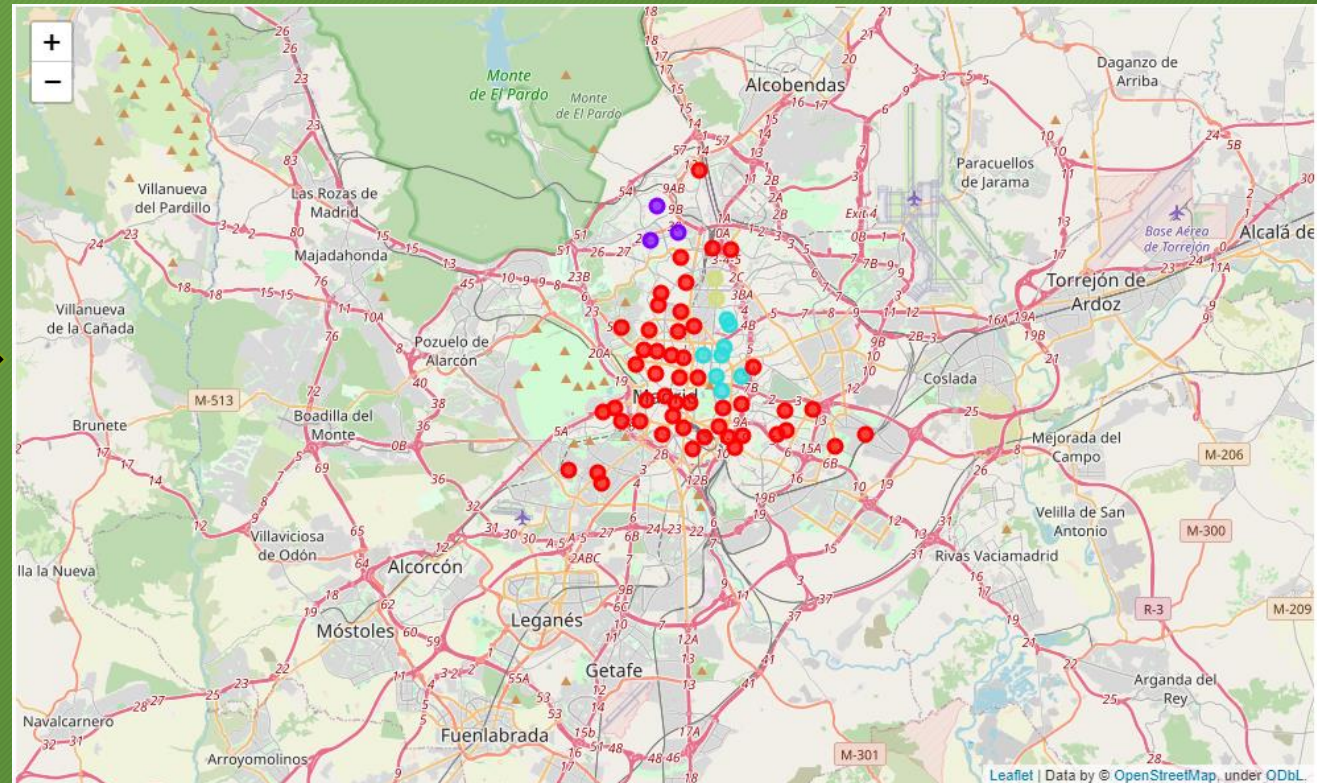
```
# Set number of clusters
kclusters = 5

madrid_grouped_clustering = madrid_grouped.drop('Neighborhood', 1)

# Run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(madrid_grouped_clustering)

# Check cluster labels generated for each row in the dataframe
kmeans.labels_[0:100]

array([0, 0, 3, 0, 0, 3, 0, 0, 0, 0, 0, 2, 0, 3, 2, 0, 0, 0, 0, 1, 3,
       0, 0, 0, 3, 0, 3, 3, 4, 0, 3, 0, 3, 1, 2, 0, 0, 0, 0, 1, 3, 4, 0,
       0, 0, 0, 2, 0, 0, 3, 3, 0, 0, 3, 0, 0, 3, 0, 0, 0, 0])
```



Methodology section

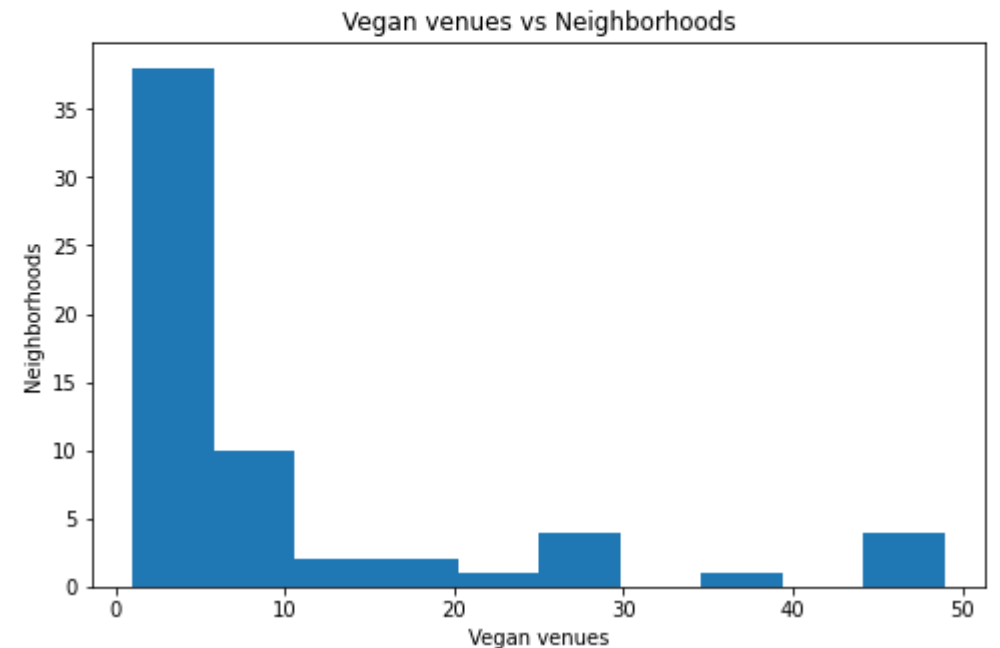
(Vegan venues - Neighborhoods / matplotlib.pyplot)



- Most neighborhoods contain fewer than 10 vegan restaurants.

THIS IS GOOD.

```
madrid_all_vegan['Venue'].plot(kind='hist', figsize=(8, 5))  
plt.title('Vegan venues vs Neighborhoods')  
plt.ylabel('Neighborhoods')  
plt.xlabel('Vegan venues')  
plt.show()
```



Methodology section

(Foursquare API / matplotlib.pyplot)

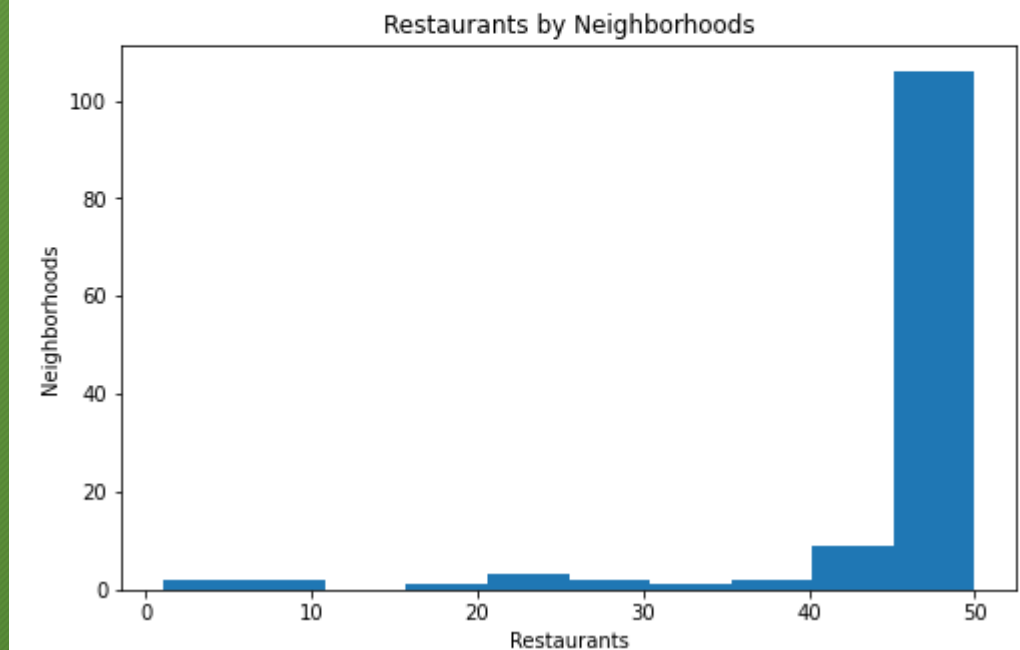


- Distribution of restaurants of any kind in the city (all neighborhoods).

- Partial conclusion :

Restaurants are very common places in Madrid, practically all the neighborhoods have between 45 and 50 places of this type.

```
madrid_grouped_all['Venue'].plot(kind='hist', figsize=(8, 5))  
plt.title('Restaurants by Neighborhoods')  
plt.ylabel('Neighborhoods')  
plt.xlabel('Restaurants')  
plt.show()
```



Methodology section

(New variable)



- According to the popularity of restaurants by neighborhood we can discard very few of them.
- The vegan movement is very popular among young people, so let's review which neighborhoods have the largest university movement.



Methodology section

(Foursquare API / matplotlib.pyplot)

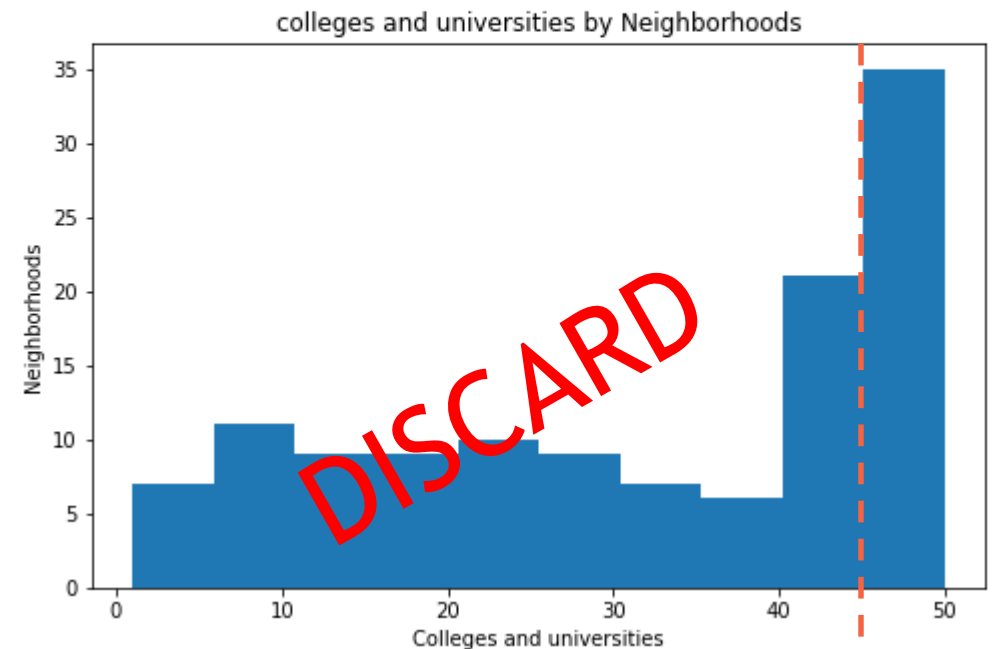


- We focus on neighborhoods with more than 45 campuses that have some relationship with universities.

- Our list is reduced to:

35 candidates.

```
madrid_univ['Venue'].plot(kind='hist', figsize=(8, 5))  
plt.title('colleges and universities by Neighborhoods')  
plt.ylabel('Neighborhoods')  
plt.xlabel('Colleges and universities')  
plt.show()
```



Methodology section

(Processing results / join)



- Joining the data of Vegan Restaurants by neighborhoods, and the neighborhoods with more university movement we can get the neighborhoods that do not have any restaurant of this type yet and that have a great youthful flow.
- Here are the finalists for the poster, “**First Vegan in the neighborhood**”. Areas with high student movement and no vegan options.

6 candidates.

```
madrid_finalists = madrid_set[madrid_set['Venue'].isnull()]
```

madrid_finalists

	Universities	Venue
Neighborhood		
Amposta	48	NaN
Atalaya	49	NaN
Chopera	46	NaN
Colina	49	NaN
Horcajo	46	NaN
San Juan Bautista	48	NaN

Methodology section

(Almost done)



- We are going to try to decide for one, in the Web of the Community of Madrid statistical files can be configured to download, we are going to look for the evolution of population of these districts and we will stay with the greater projection of growth.

```
madrid_final_grow = pd.read_excel('C:\\AA-Coursera\\09 - Applied Data Science Capstone\\Madrid\\BARRIOS_FILTRO_1.xlsx')
madrid_final_grow
```

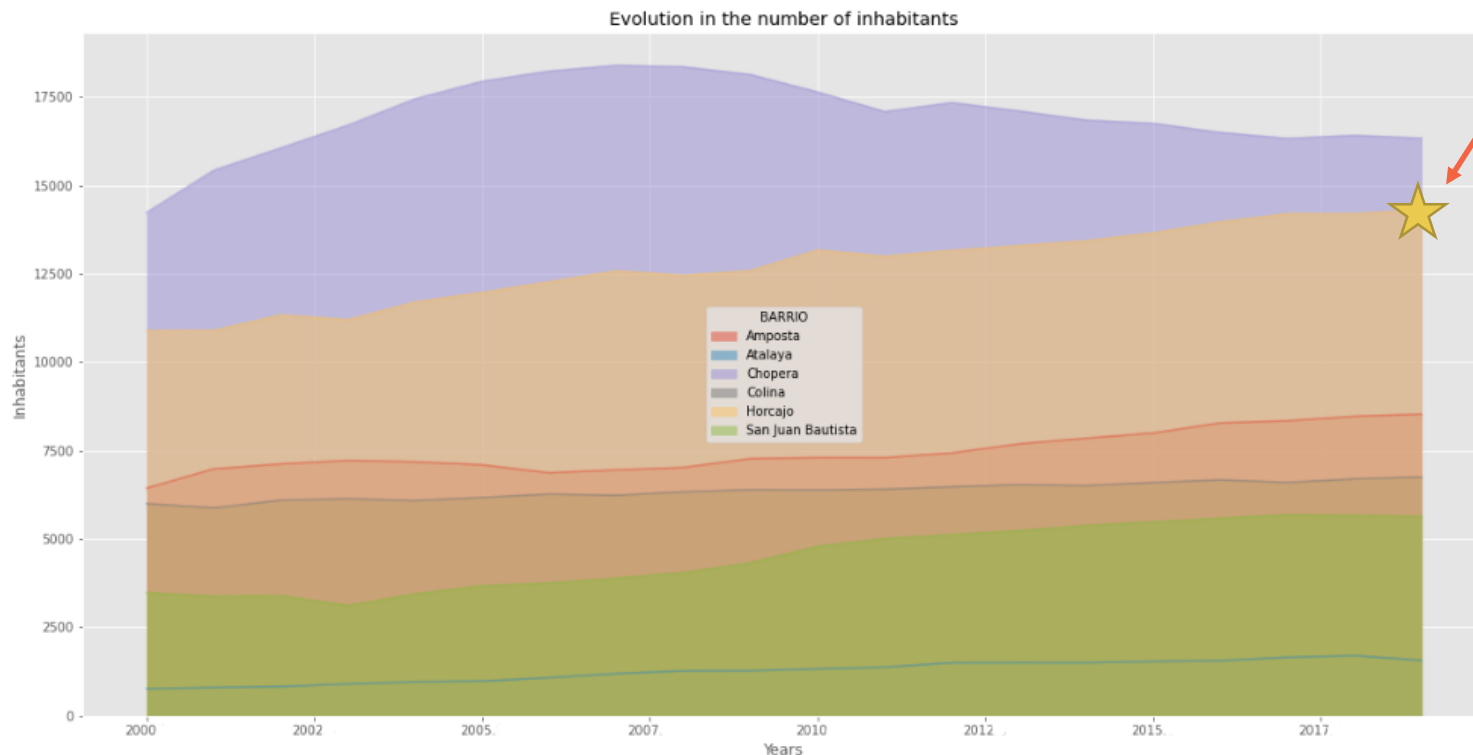
	BARRIO	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0	Amposta	6439	6984	7122	7223	7184	7098	6872	6954	7023	7273	7301	7298	7435	7689	7851	7997	8284	8342	8471	8532
1	Atalaya	768	801	826	899	961	981	1076	1193	1268	1275	1325	1376	1498	1503	1501	1537	1559	1654	1699	1570
2	Chopera	14237	15432	16065	16707	17451	17948	18237	18399	18358	18139	17645	17091	17345	17110	16851	16759	16500	16333	16420	16340
3	Colina	6003	5879	6098	6138	6087	6174	6272	6235	6337	6396	6387	6410	6476	6538	6510	6598	6671	6601	6700	6754
4	Horcajo	10891	10900	11341	11204	11705	11975	12276	12585	12457	12594	13173	13001	13164	13299	13438	13659	13971	14199	14201	14321
5	San Juan Bautista	3487	3386	3399	3122	3451	3684	3755	3899	4052	4332	4798	5016	5129	5238	5387	5487	5591	5688	5677	5643

Methodology section

(Evolution of population / area plot)



```
madrid_final_grow.index = madrid_final_grow.index.map(int)
madrid_final_grow.plot(kind='area', stacked=False, figsize=(20, 10),)
plt.title('Evolution in the number of inhabitants')
plt.ylabel('Inhabitants')
plt.xlabel('Years')
plt.show()
```



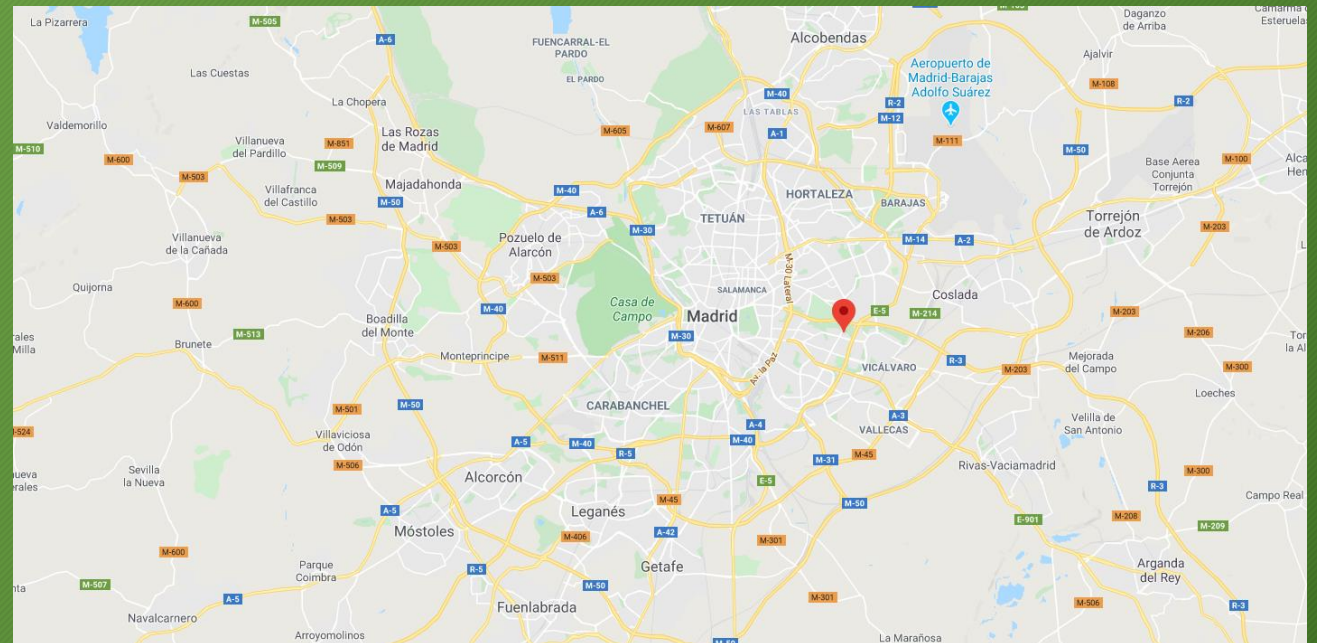
The **Horcajo** district is the second one with the largest fixed population, but it has a more reliable growth trend.

Final decision



- With the flow of students in the area and the projected growth of the fixed population, the market share could be assured.
- Furthermore, there is no similar restaurant in the entire neighborhood.

The **Horcajo** neighborhood, is the area
Selected to open a new restaurant.



Conclusions:



- We have obtained a good approach to the problem and have found a solution that seems to be satisfactory, however, other factors could be studied further, such as the average income of the area, other recreational areas nearby, and the average price per square meter.
- As a conclusion we can say that the objectives of the study have been met.
- The aim was to obtain a recommendation for the location of a vegan restaurant and it has been limited to a considerably small area.