New Vegan Restaurant in Madrid

Location recommendation report.

Description of the problem

A new vegan restaurant in Madrid:

Food is a frequent topic of conversation, and in recent years there has been an increasing number of initiatives around it: trade shows, fairs, television programs and competitions, thematic channels, specialized magazines, etc.

Undoubtedly, setting up a restaurant has great attractions. Properly managed and with a good public relations policy, it can become a new option in the local leisure scene. Today more than ever, the profession of chef is especially valued and, if our skills are mainly culinary, we can aspire, why not, to the star. From an economic point of view, in the most successful cases a return of 25% has been achieved - it is more normal to expect a return of between 15% and 20%.

It is crucial to be aware of current trends in gastronomy. Healthy and sustainable eating, creating awareness of food waste, using customer-focused technology and offering unique and personalized experiences are some of them. It is also necessary to make a very sincere internal analysis, to know what we can do with the resources we have (economic, technical, human, etc.). A good previous study and the knowledge of our target audience, will allow us to define in the most concrete way possible our business concept, its structure, positioning, location and elements of the marketing mix.

In this study we will make a first approach to the study needed to set up a vegan restaurant, which are a growing trend in all European capitals.

Description of the data

For this work we will first use the data provided by the government of the City of Madrid. Through the following link you can obtain an excel file with all the sections of the city, which we must process to obtain a clean list of the districts of Madrid.

http://www.madrid.org/iestadis/fijas/clasificaciones/descarga/cobar18.xls

This Excel file contains the following columns:

- **munic:** municipal code
- **distr**: district code
- **ldistr:** name of district
- **barrio:** code of neighborhood
- **descript:** name of neighborhood
- **secci:** neighborhood section

Overview:

	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

Once this first set of data has been processed, a second file is obtained which can be configured for downloading at the town hall site with all the coordinates of the districts of the capital of Spain.
https://www.madrid.es/portal/site/munimadrid

- **BARRIOS:** name of neighborhood
- **LATITUD**: neighborhood latitude
- **LONGITUD:** neighborhood longitude
- Overview:

	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

At this point, combining both sets of data we can already obtain an initial distribution on the map of the neighborhoods. Using Foursquare's API we can then see the distribution of all the vegan places in the city.

With these first data we will make several analyses, but as they are not definitive it will be necessary to incorporate new information. 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. This will be the final cross to decide the location of our restaurant.

http://www-2.munimadrid.es/TSE6/control/seleccionDatosBarrio

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

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

```
df_coord.head()
```

Out[8]:

	BARRIOS	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

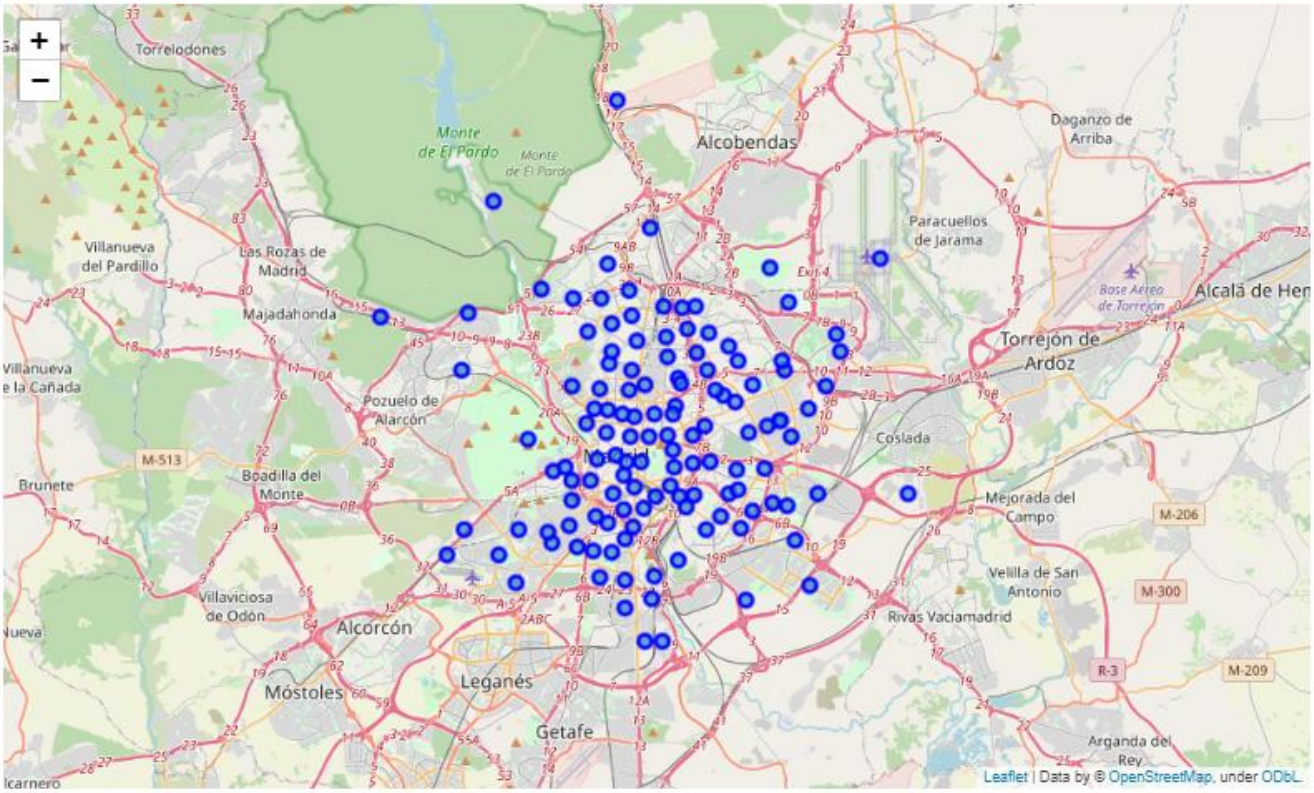
At this point we are already in a position to see the neighborhoods of the community of Madrid.

In [11]:
```
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
```

Out[11]:



By using Foursquare's API we can get the total number of places that serve vegan food in these neighborhoods.

In [18]:
```
madrid_venues_vegan = getNearbyVenues(names=df['Neighborhood'], latitudes=df['Latitud'], longitudes=df['Longitud'], radius=1000,
madrid_venues_vegan.head()
```

Out[18]:

	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	Yatiri	40.412690	-3.703373	Vegetarian / Vegan Restaurant
4	Acacias	40.401422	-3.704936	El Granero	40.408275	-3.697631	Vegetarian / Vegan Restaurant

In [19]:
```
madrid_venues_vegan.shape
```

Out[19]: (587, 7)

We can also see the distribution of these sites on the map.

In [21]:
```
map_madrid_vegan = folium.Map(location=[latitude, longitude], zoom_start=12)
addToMap(madrid_venues_vegan, 'blue', map_madrid_vegan)

map_madrid_vegan
```

Out[21]:

First approximation

We split the group of restaurants by clusters

Cluster Neighborhoods

In [28]:
```
# Set number of clusters
kclusters = 4

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]
```

Out[28]:
```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 2, 0, 2, 2, 3, 0, 2, 0, 0, 1, 2, 0, 0, 0, 0, 1, 0, 3, 0,
       0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

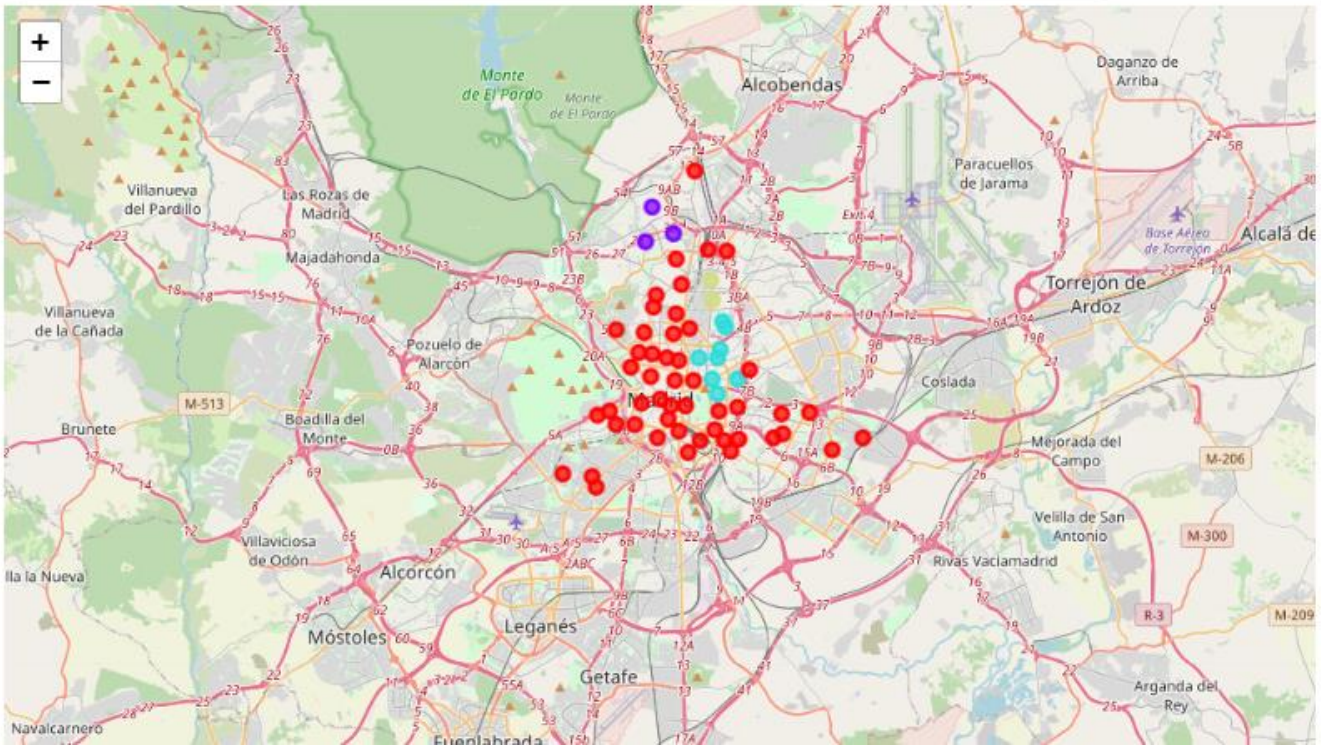
In [34]:
```
# Create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# Set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# Add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(madrid_merged['Latitud'], madrid_merged['Longitud'], madrid_merged['Neighborhood'], madrid_merg
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[34]:



It's good analysis but it doesn't tell us much in this case, we see most of the vegan restaurants concentrated in the center of the city, which is to be expected.

Let's clear the list of the number of vegan sites per neighborhood.

In [92]:
```
madrid_all_vegan.head()
```

Out[92]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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

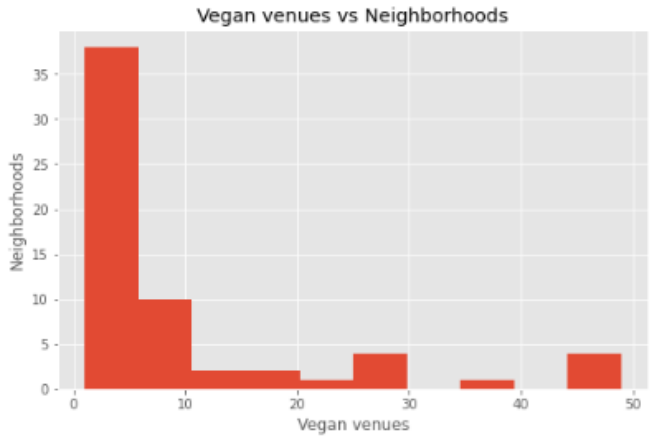
In [93]:
```
madrid_all_vegan.drop(['Neighborhood Latitude', 'Neighborhood Longitude', 'Venue Latitude', 'Venue Longitude','Venue Category'],
```

In [94]:
```
madrid_all_vegan.head()
```

Out[94]:

Neighborhood	Venue
Acacias	8
Adelfas	3
Almagro	28
Almenara	1
Aluche	1

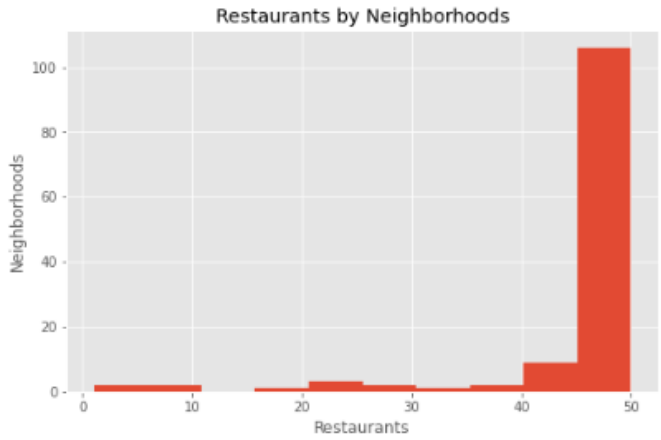
In [130]:
```
madrid_all_vegan['Venue'].plot(kind='hist', figsize=(8, 5))
plt.title('Vegan venues vs Neighborhoods')
plt.ylabel('Neighborhoods')
plt.xlabel('Vegan venues')
plt.show()
```



Most of the neighborhoods have very few vegan sites, this is very good. On the other hand, a minority concentrates the largest number of sites (over 30). The downtown neighborhoods, of course.

In order to try to obtain the general distribution of all types of restaurants in the neighbourhoods, we are once again using the API.

In [131]:
```
madrid_grouped_all['Venue'].plot(kind='hist', figsize=(8, 5))
plt.title('Restaurants by Neighborhoods')
plt.ylabel('Neighborhoods')
plt.xlabel('Restaurants')
plt.show()
```

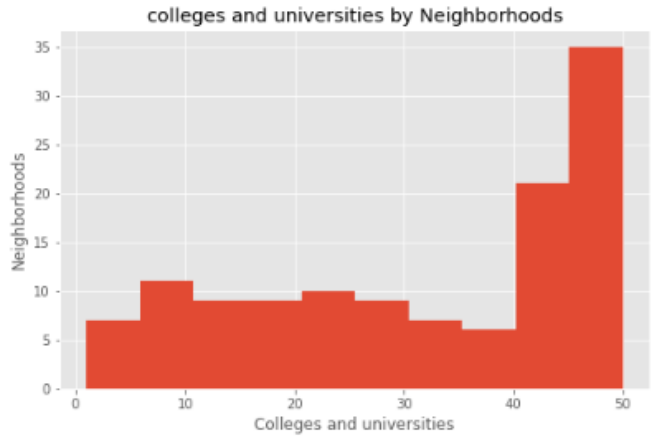


Restaurants are very common places in Madrid, practically all the neighborhoods have between 45 and 50 places of this type.

With the total number of restaurants, we also don't have an approach that will help us much

Another option, we obtain the number of venues that have to do with Universities or similar. This could be a good option since this type of vegan movement is more popular among this sector of the population.

In [134]:
```
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()
```



This is a better distribution, we make a first filter and we keep the neighborhoods that have more than 45 places related to the university world.

The list is reduced to 35 candidate districts.

Out[83]:

Neighborhood	Universities
Acacias	48
Almagro	47
Amposta	48
Arapiles	46
Argüelles	48
Atalaya	49
Bellas Vistas	47
Berruguete	49
Castellana	50
Castillejos	49
Chopera	46
Ciudad Jardín	49
Colina	49
Cortes	48
Cuatro Caminos	48
Ensanche de Vallecas	50
Entrevías	48
Gaztambide	47
Guindalera	49
Hellín	47
Hispanoamérica	50
Horcajo	46
Ibiza	50
Los Angeles	49
Lucero	47
Opañel	50
Palomeras Bajas	46
Peñagrande	47
Pueblo Nuevo	48
Rejas	49
Rosas	50
San Juan Bautista	48
Trafalgar	46
Vallehermoso	48
Ventas	46

If we combine this list with the data set of vegan restaurants by neighborhood that we took out at the beginning, we get 6 neighborhoods with high circulation of young college students and with 0 vegan restaurants, good opportunity.

```
In [102]: madrid_finalists
```

Out[102]:

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

So far we have secured our opening with a big sign.

The first vegan restaurant in the neighborhood.

Okay, but which one we decide on.

We take another variable into consideration, according to the student population any of these neighborhoods can be chosen.

But how has the evolution of the fixed population of these neighborhoods been?

The people living in the neighborhood can also be potential customers and that could tip the balance.

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

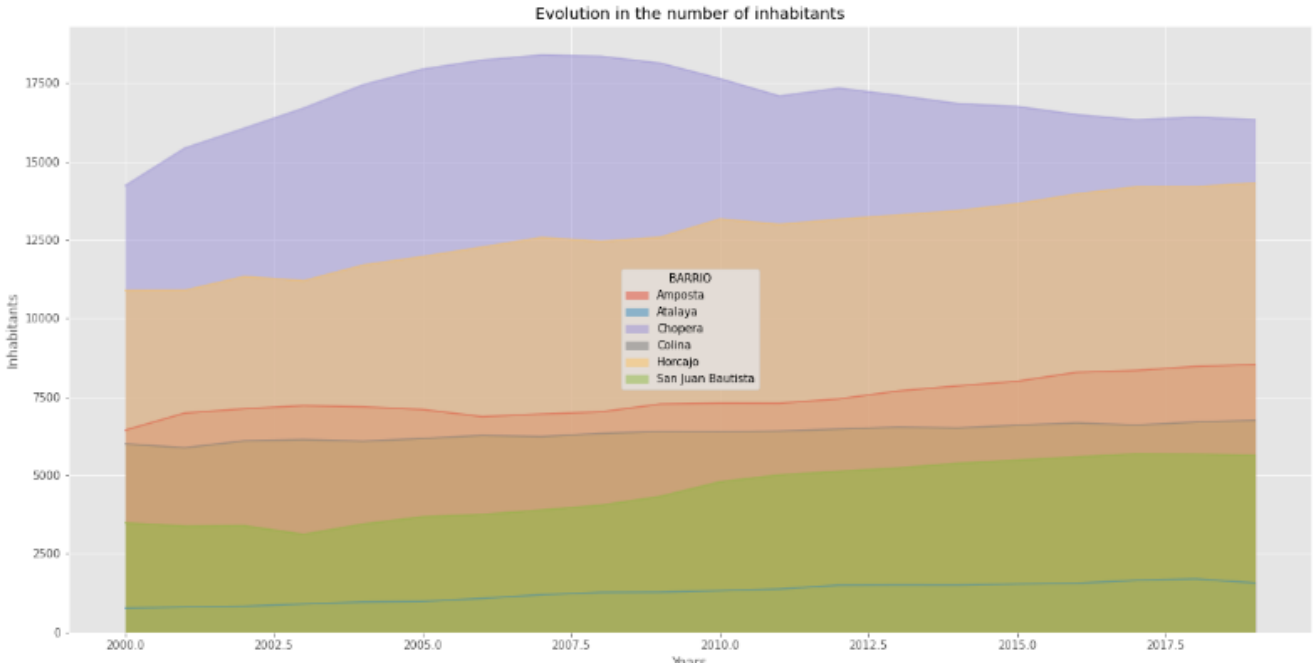
```
In [126]: madrid_final_grow.set_index('BARRIO', inplace=True)
          madrid_final_grow
```

Out[126]:

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

We checked the trend.

```
In [129]: 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()
```

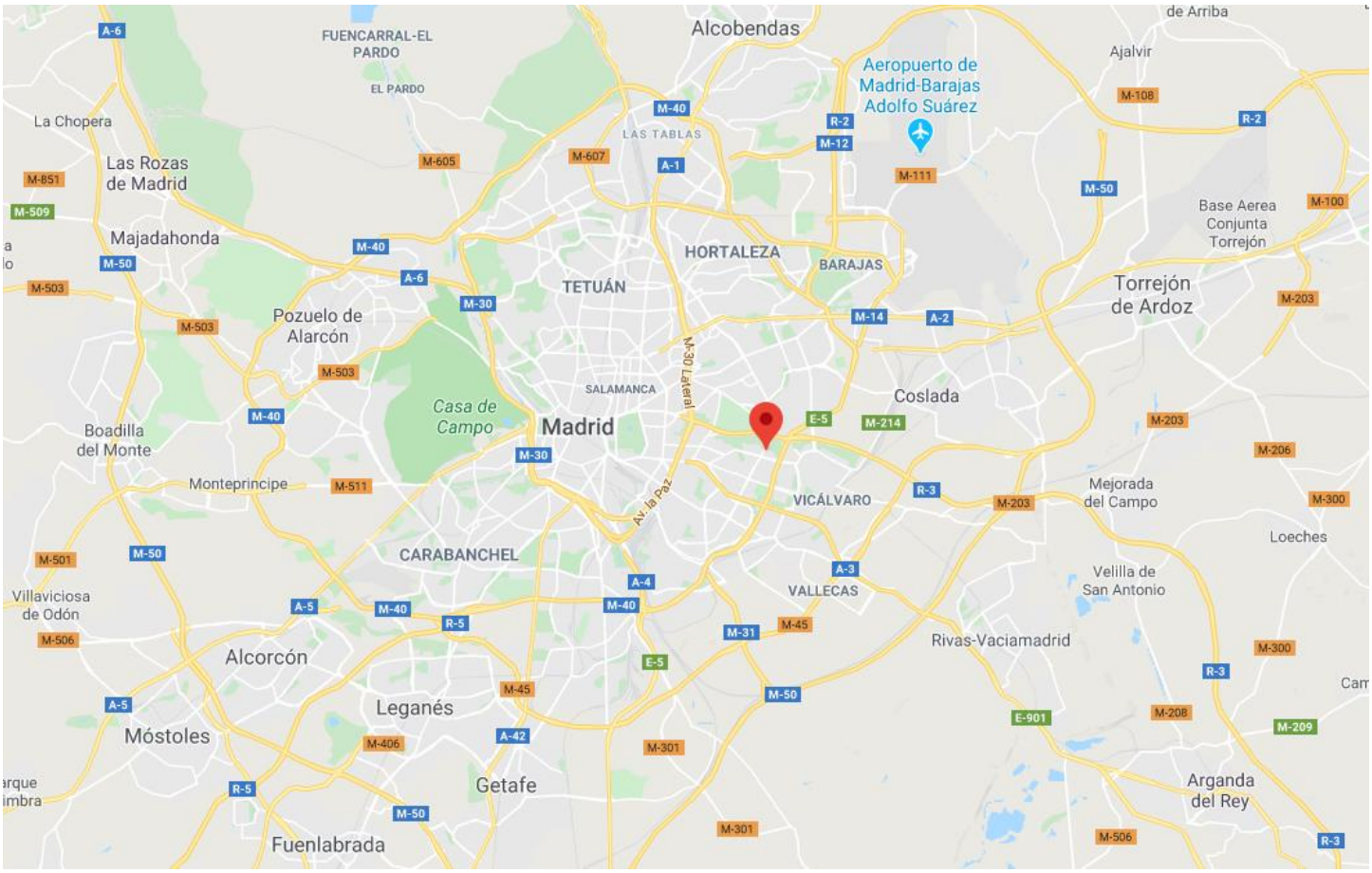


As can be seen in the graph, "Chopera" is the neighbourhood with the largest fixed population, but it seems to be on the decline.

"Horcajo" seems to be a better choice, it is the second in fixed population but it seems to have an upward trend, and that is a good choice.

Results

As a result of this study we can recommend a specific neighborhood for the location of the new restaurant.



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.

Discussion section

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.

Conclusion

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.