Analysis of the incidents managed by the local police in Barcelona's neighborhoods in 2019

Amparo González Castaño

Barcelona, May 2020

1. Business Problem

1.1. Background

Barcelona is well known as being one of the most dangerous cities in Europe because the many incidents happening every day. How local police manage these incidents and how effective they worked can improve the security of the city.

Incidents' types, locations and occurrence vary during the year. However, neighborhoods can be group depending on its incident's types occurrence.

Therefore, it is advantageous for the local police to have the neighborhood grouped based on the most common incidents. It could be used to be more efficient and distribute better the officers.

1.2. Problem

This project aims to cluster Barcelona's neighborhoods in groups depending on the occurrence of the top 10 most common incidents in each neighborhood manage by the local police. There are many incidents occurring in the different neighborhoods, if the neighborhoods are clustered in groups with similar incidents it would be useful for the local police to determined where they should send more teams and what kind of incidents they may encounter.

1.3. Interest

Local police would be interested in knowing in advance what incidents might they encounter in each neighborhood, helping them to be prepare and knowing at a first glance what is happening in each neighborhood.

2. Data Acquisition and Cleaning

2.1. Data Sources

The data source chosen was 'Incidents managed for Guàrdia Urbana in the city of Barcelona' from the year 2019 from Open Data Ajuntament Barcelona, https://opendata-ajuntament.barcelona.cat/data/en/dataset/incidents-gestionats-gub, it contains the different categories of incidents managed by the local police of Barcelona in the different neighborhoods of the city as well as the number of each incident in the year 2019. This data source contains the neighborhood code, the neighborhood name, the district code, the district name and the postal code. It is very useful to locate all the different neighborhoods.

Afterwards, to the incidents data, the geospatial coordinates has been added. The obtaining geospatial coordinates process can be found in 'Obtain neighborhoods with coordinates cvs' notebook. The geospatial data for the neighborhoods has been obtain in combination of using the geocoder library https://geocoder.readthedocs.io/ and the neighborhood information in 'Incidents managed for Guàrdia Urbana in the city of Barcelona'. Also, it has been checked that the number of neighborhoods corresponds with the real ones and that none is missing.

2.2. Data Cleaning

It can be observed there are two data cleaning process, one in 'Data visualization of incidents in Barcelona's neighborhoods in 2019' and another one in 'Cluster Barcelona neighborhoods' notebook.

Data used in the notebooks is the same, is the combination of the coordinates found in 'Obtain neighborhoods with coordinates cvs' notebook and the neighborhood information in 'Incidents managed for Guàrdia Urbana in the city of Barcelona'.

They main actions to clean the data were to change the columns name, removed 'NaN' and 'Desconegut' values from 'Neighborhood' Column, change columns types, add the coordinates to the existing data frame and removed 'NaN' values after adding neighborhoods coordinates.

2.3. Feature Selection

After data cleaning, different approaches were done depending on the needs of each notebook.

	Incident	District Code	Disctrict	Neighborhood Code	Neighborhood	Month	Number of Incidents	Neighborhood Latitude	Neighborhood Longitude
0	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS ...	6	Gràcia	31	la Vila de Gràcia	1	85	41.403178	2.157166
1	ACCIDENTS DE TRÀNSIT SENSE FERITS ...	10	Sant Martí	73	la Verneda i la Pau	4	4	41.423220	2.202940
2	INCIDÈNCIES AMB AFECTACIÓ DE VIA ...	2	Eixample	9	la Nova Esquerra de l'Eixample	4	12	41.382816	2.149966
3	CONVIVÈNCIA VEINAL ...	6	Gràcia	31	la Vila de Gràcia	4	105	41.403178	2.157166
4	AGRESSIONS ...	1	Ciutat Vella	3	la Barceloneta	1	12	41.380653	2.189927

Figure 1 Incidents Barcelona 2019

In 'Data visualization of incidents in Barcelona's neighborhoods in 2019' notebook, the 'Month' column was transformed from categorical to quantitative variable first, resulting in 12 different columns, one for each month of the year, and, also, the original 'Month' column was dropped.

	District Code	Disctrict	Neighborhood Code	Neighborhood	Incident	Number of Incidents	Neighborhood Latitude	Neighborhood Longitude	1	2	3	4	5	6	7	8	9	10	11	12
0	6	Gràcia	31	la Vila de Gràcia	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS ...	85	41.403178	2.157166	1	0	0	0	0	0	0	0	0	0	0	0
1	10	Sant Martí	73	la Verneda i la Pau	ACCIDENTS DE TRÀNSIT SENSE FERITS ...	4	41.423220	2.202940	0	0	0	1	0	0	0	0	0	0	0	0
2	2	Eixample	9	la Nova Esquerra de l'Eixample	INCIDÈNCIES AMB AFECTACIÓ DE VIA ...	12	41.382816	2.149966	0	0	0	1	0	0	0	0	0	0	0	0
3	6	Gràcia	31	la Vila de Gràcia	CONVIVÈNCIA VEINAL ...	105	41.403178	2.157166	0	0	0	1	0	0	0	0	0	0	0	0
4	1	Ciutat Vella	3	la Barceloneta	AGRESSIONS ...	12	41.380653	2.189927	1	0	0	0	0	0	0	0	0	0	0	0

Figure 2 Months transformed into categorical variables

Afterwards the 'Number of incidents' were added to the month they occurred.

	District Code	Disctrict	Neighborhood Code	Neighborhood	Incident	Neighborhood Latitude	Neighborhood Longitude	1	2	3	4	5	6	7	8	9	10	11	12
0	6	Gràcia	31	la Vila de Gràcia	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS ...	41.403178	2.157166	85	0	0	0	0	0	0	0	0	0	0	0
1	10	Sant Martí	73	la Verneda i la Pau	ACCIDENTS DE TRÀNSIT SENSE FERITS ...	41.423220	2.202940	0	0	0	4	0	0	0	0	0	0	0	0
2	2	Eixample	9	la Nova Esquerra de l'Eixample	INCIDÈNCIES AMB AFECTACIÓ DE VIA ...	41.382816	2.149966	0	0	0	12	0	0	0	0	0	0	0	0
3	6	Gràcia	31	la Vila de Gràcia	CONVIVÈNCIA VEINAL ...	41.403178	2.157166	0	0	0	105	0	0	0	0	0	0	0	0
4	1	Ciutat Vella	3	la Barceloneta	AGRESSIONS ...	41.380653	2.189927	12	0	0	0	0	0	0	0	0	0	0	0

Figure 3 Number of incidents added in each month

In 'Cluster Barcelona Neighborhoods' notebook, also the 'Incident' column was transformed from categorical to quantitative variable in different columns, one for each incident, and the original 'Incident' column was dropped.

	District Code	Disctrict	Neighborhood Code	Neighborhood	Month	Number of Incidents	Neighborhood Latitude	Neighborhood Longitude	ABOCAMENT DE RESIDUS	ACCIDENTS DE TRANSIT AMB MATERIES PERILLOSSES	ACCIDENTS DE TRÀNSIT AMB FERITS	ACCIDENTS DE TRÀNSIT SENSE FERITS	ACTES CONTRA LA PROPIETAT	ACTES CONTRA LA PROPIETAT PRIVADA	ACTES CONTRA LA PROPIETAT PÚBLICA	ACTIVITAT AMB PRESOS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	AGRESSIONS	ALARMES D'INCENDI / ROBATORI	ALTRES ACTUACIONS DE S, C,	ALTRES ACTUACIONS EN MEDI AMBIENT
0	6	Gràcia	31	la Vila de Gràcia	1	85	41.403178	2.157166	0	0	0	0	0	0	0	0	1	0	0	0	0
1	10	Sant Martí	73	la Verneda i la Pau	4	4	41.423220	2.202940	0	0	0	1	0	0	0	0	0	0	0	0	0
2	2	Eixample	9	la Nova Esquerra de l'Eixample	4	12	41.382816	2.149966	0	0	0	0	0	0	0	0	0	0	0	0	0
3	6	Gràcia	31	la Vila de Gràcia	4	105	41.403178	2.157166	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	Ciutat Vella	3	la Barceloneta	1	12	41.380653	2.189927	0	0	0	0	0	0	0	0	1	0	0	0	0

Figure 4 Incidents transformed into categorical variables

Afterwards, the mean of the occurrence of each type incident in each neighborhood was calculated.

	Neighborhood	ABOCAMENT DE RESIDUS	ACCIDENTS DE TRANSIT AMB MATERIES PERILLOSSES	ACCIDENTS DE TRÀNSIT AMB FERITS	ACCIDENTS DE TRÀNSIT SENSE FERITS	ACTES CONTRA LA PROPIETAT	ACTES CONTRA LA PROPIETAT PRIVADA	ACTES CONTRA LA PROPIETAT PÚBLICA	ACTIVITAT AMB PRESOS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	AGRESSIONS	ALARMES D'INCENDI / ROBATORI	ALTRES ACTUACIONS DE S, C,	ALTRES ACTUACIONS EN MEDI AMBIENT	ALTRES INCIDÈNCIES AMB ANIMALS	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	ANIMALS / INSECTES PERILLOSOS	ANIMALS DE COMPANYIA	ASSISTÈNCIA A CONDUCTORS EN V, P,
0	Baró de Viver	0.000000	0.000000	0.028409	0.028409	0.000000	0.039773	0.005682	0.005682	0.056818	0.022727	0.017045	0.056818	0.000000	0.000000	0.017045	0.017045	0.022727	0.000000
1	Can Baró	0.003021	0.000000	0.033233	0.030211	0.000000	0.030211	0.009063	0.000000	0.036254	0.027190	0.012085	0.036254	0.000000	0.003021	0.030211	0.015106	0.033233	0.012085
2	Can Peguera	0.009050	0.000000	0.022624	0.018100	0.000000	0.031674	0.004525	0.000000	0.054299	0.036199	0.022624	0.054299	0.000000	0.000000	0.018100	0.004525	0.036199	0.000000
3	Canyelles	0.007143	0.000000	0.039286	0.042857	0.000000	0.039286	0.003571	0.000000	0.035714	0.028571	0.014286	0.042857	0.000000	0.000000	0.032143	0.014286	0.021429	0.010714
4	Ciutat Meridiana	0.000000	0.000000	0.021978	0.016484	0.000000	0.030220	0.000000	0.000000	0.032967	0.030220	0.021978	0.032967	0.000000	0.013736	0.032967	0.016484	0.032967	0.005495
5	Diagonal Mar i el Front Marítim del Poblenou	0.008889	0.000000	0.026667	0.026667	0.000000	0.026667	0.002222	0.000000	0.026667	0.022222	0.026667	0.026667	0.002222	0.002222	0.026667	0.008889	0.026667	0.015556

Figure 5 Mean occurrence of each incident in each neighborhood

Finally, a new data frame was created with the top 10 incidents in each neighborhood.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
0	Baró de Viver	CONVIVÈNCIA VEINAL	VIGILÀNCIA POLICIAL	ALTRES ACTUACIONS DE S, C,	INFRACCIONS EN GUALS I RESERVES	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INFRACCIONS EN MOVIMENT	INCENDIS	COL·LABORACIÓ AMB ALTRES SERVEIS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	ACTES CONTRA LA PROPIETAT PRIVADA
1	Can Baró	INFRACCIONS EN GUALS I RESERVES	RESTRICCIONS TEMPORALS DE TRÀNSIT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INFRACCIONS PER ESTACIONAMENTS	CONVIVÈNCIA VEINAL	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	ALTRES ACTUACIONS DE S, C,	INCIDÈNCIES AMB AFECTACIÓ DE VIA	ANIMALS DE COMPANYIA	ACCIDENTS DE TRÀNSIT AMB FERITS
2	Can Peguera	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	CONVIVÈNCIA VEINAL	ALTRES ACTUACIONS DE S, C,	VIGILÀNCIA POLICIAL	INFRACCIONS PER ESTACIONAMENTS	ASSISTÈNCIA A PERSONES EN EDIFICI	ANIMALS DE COMPANYIA	INFRACCIONS EN GUALS I RESERVES	AGRESSIONS
3	Canyelles	INFRACCIONS EN MOVIMENT	VIGILÀNCIA POLICIAL	CONVIVÈNCIA VEINAL	ALTRES ACTUACIONS DE S, C,	ACCIDENTS DE TRÀNSIT SENSE FERITS	ACTES CONTRA LA PROPIETAT PRIVADA	INFRACCIONS EN GUALS I RESERVES	INFRACCIONS PER ESTACIONAMENTS	ACCIDENTS DE TRÀNSIT AMB FERITS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS
4	Ciutat Meridiana	ALTRES ACTUACIONS DE S, C,	INFRACCIONS EN GUALS I RESERVES	VIOLÈNCIA DOMÈSTICA	VIGILÀNCIA POLICIAL	INFRACCIONS PER ESTACIONAMENTS	ASSISTÈNCIA A PERSONES EN EDIFICI	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	CONVIVÈNCIA VEINAL	INFRACCIONS EN MOVIMENT	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA

Figure 6 Top 10 most common incidents in each neighborhood

3. Data Analysis

All the data analysis can be found in 'Data visualization of incidents in Barcelona's neighborhoods in 2019' notebook.

3.1. Data Transformation

In order to visualize and analyze the data, was needed to continue transforming it. The starting data frame used was the one shown in Figure 3, 'Number of incidents added in each month'. This data frame was transformed three times to achieve the following purposes: knowing the number of incidents each months by neighborhood, knowing the total number of incidents in 2019 by neighborhood, knowing the total number of incidents grouped by neighborhood and type of incident, and finally knowing the total number of each incident.

In the first case, the incidents were grouped by neighborhood, afterwards they were sum in each month. Resulting the following data frame.

Neighborhood	1	2	3	4	5	6	7	8	9	10	11	12
Baró de Viver	60	54	41	34	41	53	30	36	37	38	23	25
Can Baró	74	93	111	96	81	87	113	106	91	93	93	113
Can Peguera	46	39	48	35	33	36	50	42	49	40	32	37
Canyelles	55	57	40	50	60	68	79	64	64	59	54	52
Ciutat Meridiana	183	138	142	137	166	169	185	160	152	176	129	114

Figure 7 Total number of incidents each month by neighborhood

In the second case, using the previous data frame the total number of incidents is added as a new column and the neighborhoods are sorted by descending order of number of incidents.

Neighborhood	1	2	3	4	5	6	7	8	9	10	11	12	Total
el Raval	1887	1825	2161	2158	2338	2302	2428	2242	2317	2223	1940	1877	25698
la Dreta de l'Eixample	1636	1685	1719	1585	1719	1718	2117	1684	1646	2035	1451	1547	20542
el Barri Gòtic	1330	1381	1523	1593	1685	1776	1832	1697	1669	1534	1207	1312	18539
el Poble-sec	814	828	1002	993	1143	1251	1338	993	1199	1039	900	899	12399
la Vila de Gràcia	795	838	916	915	1076	1095	1198	1279	980	920	841	902	11755

Figure 8 Total number of incidents in each neighborhood sorted by descending order

In the third case, starting also form the data frame shown in Figure 3, the incidents were grouped by neighborhood and incident, afterwards all the number of incidents were sum. Finally, the total number of incidents were added in a column and they were sorted by descending order.

Neighborhood	Incident	1	2	3	4	5	6	7	8	9	10	11	12	Total
el Raval	VIGILÀNCIA POLICIAL	399	350	317	315	353	331	283	274	306	341	356	275	3900
	ACTES CONTRA LA PROPIETAT PRIVADA	291	260	283	298	301	288	247	216	250	200	185	169	2988
el Barri Gòtic	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	120	134	203	230	237	317	356	257	253	259	143	136	2645
el Raval	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	147	117	200	218	245	249	307	246	269	240	191	167	2596
el Barri Gòtic	ACTES CONTRA LA PROPIETAT PRIVADA	256	260	236	259	264	195	213	196	196	138	124	148	2485

Figure 9 Total number of incidents grouped by neighborhood and type of incident sorted by descending order

Finally, in the last case, the total number of incidents are shown. Starting from the data frame in Figure 3, the incidents are grouped by type of incidents and they are sum. To end up, the total number of incidents are added in column and the are sorted by descending order.

Incident	1	2	3	4	5	6	7	8	9	10	11	12	Total
CONVIVÈNCIA VEINAL	2445	2369	2869	2700	3079	3437	3680	3320	3266	2874	2618	2699	35356
ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	1576	1567	2259	2130	2569	3552	4203	3383	3463	2688	1814	1843	31047
VIGILÀNCIA POLICIAL	2785	2829	2776	2458	2634	2354	2293	2343	2636	2680	2452	2371	30611
INFRACCIONS EN GUALS I RESERVES	2096	1926	2078	1856	2138	1942	1949	1434	2091	2096	2335	2602	24543
ACTES CONTRA LA PROPIETAT PRIVADA	1762	1706	1639	1709	1715	1664	1777	1618	1516	1255	1132	1152	18645

Figure 10 Total number of incidents grouped by incident type and sorted by descending order

3.2. Data Visualization

Using the data frames obtained in the section 3.1 Data Transformation, three main visualizations can give a big insight about the problem.

Using the data frame of case 1, it can be displayed the total number of incidents in each neighborhood by month.

As there are many neighborhoods, in order to the data be shown clearly only the top 10 first neighborhoods with higher incidents have been displayed.

```python
#As there are many neighborhoods, just the 10th firts with the higher incidents rates has been displayed
df_neigh_sum_sorted=df_neigh_sum.sort_values(by='Total', ascending=False)
top_df=df_neigh_sum_sorted.head(10)
top_df=top_df.loc[top_df.index, months]
top_df=top_df.transpose()
```

Figure 11 Selection of top 10 neighborhoods with more incidents

```python
top_df.plot(kind='line', figsize=(10, 6))

plt.title('Top 10 neighborhoods with highest number of incidents in each month of 2019')
plt.ylabel('Number of incidents')
plt.xlabel('Months')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0.)
plt.show()
```

Figure 12 Implementation plot top 10 neighborhoods with more incidents in each month
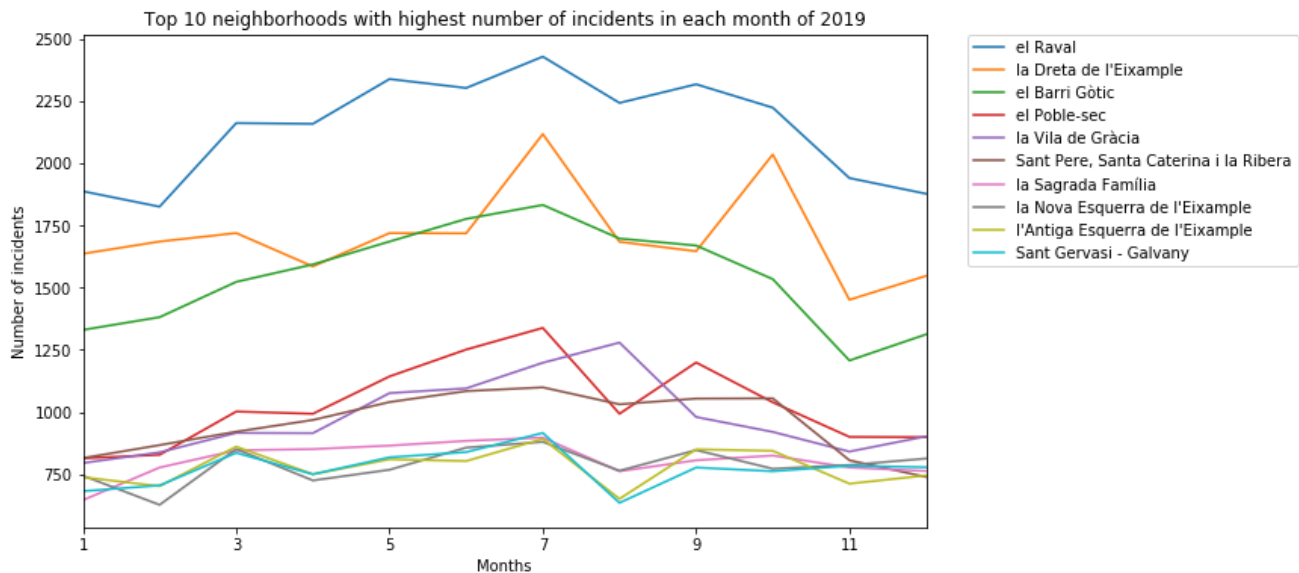
Figure 13 Top 10 neighborhoods with higher incidents occurrence during 2019

Using the data frame of case 2, it can be displayed the total number of incidents in each neighborhood in the year 2019.

```
df_neigh_total.plot(kind='bar', figsize=(10, 6))

plt.xlabel('Neighborhood')
plt.ylabel('Total')
plt.title('Total number of incidents in 2019 by neighborhood')

plt.show()
```

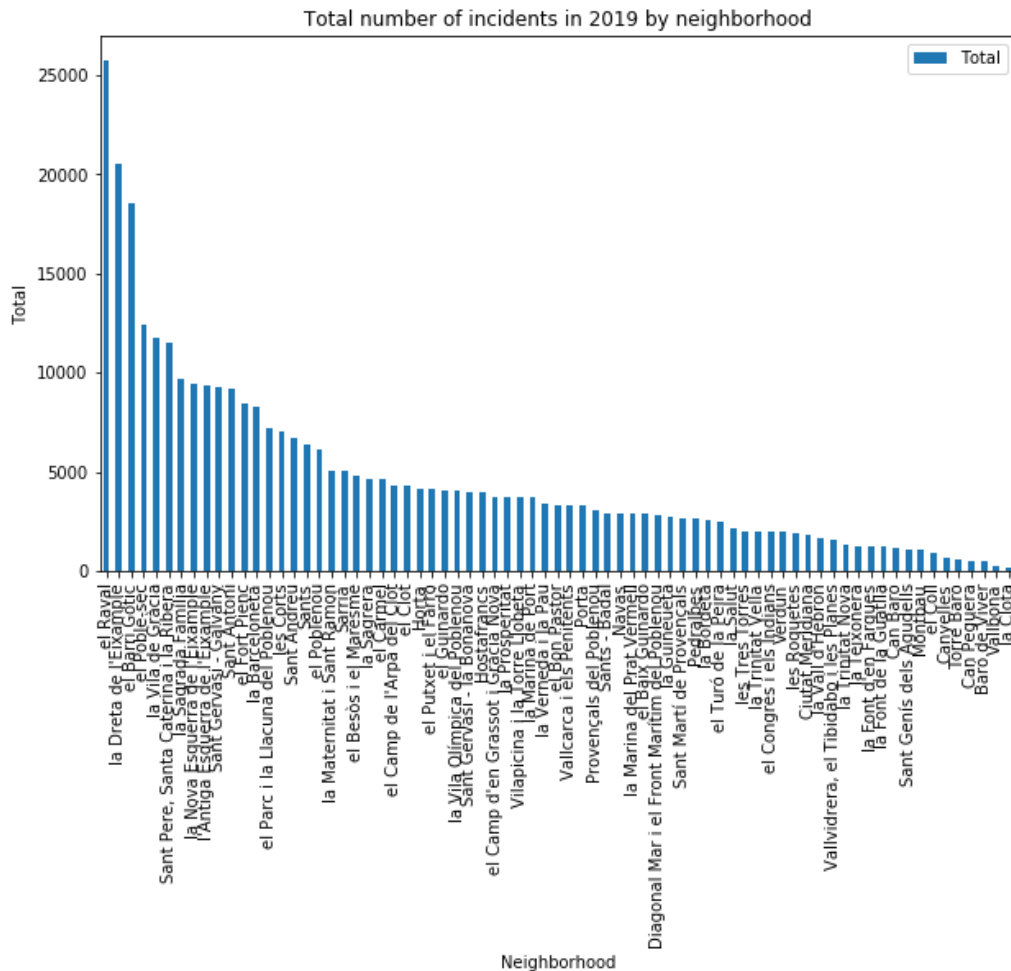*Figure 14 Implementation plot total number of incidents by neighborhood*

Figure 15 Total number of incidents in 2019 by neighborhood

Finally, using the data frame of the last case, it can be displayed the total occurrence of each incident in 2019. As there are many incidents just the top 10 with higher occurrence are displayed.

```
# As there are many types of incidents, just the 10th firts with the higher incidents rates has been displayed
df_incidents_total_sorted=df_incidents_total.sort_values(by='Total', ascending=False)
top_incidents_df=df_incidents_total_sorted.head(10)
```

Figure 16 Selecting the top 10 incidents with higher occurrence

```
top_incidents_df.plot(kind='bar', figsize=(10, 6))

plt.xlabel('Type of Incident')
plt.ylabel('Total')
plt.title('Top 10 total number of incidents in 2019 by type of incident')

plt.show()
```
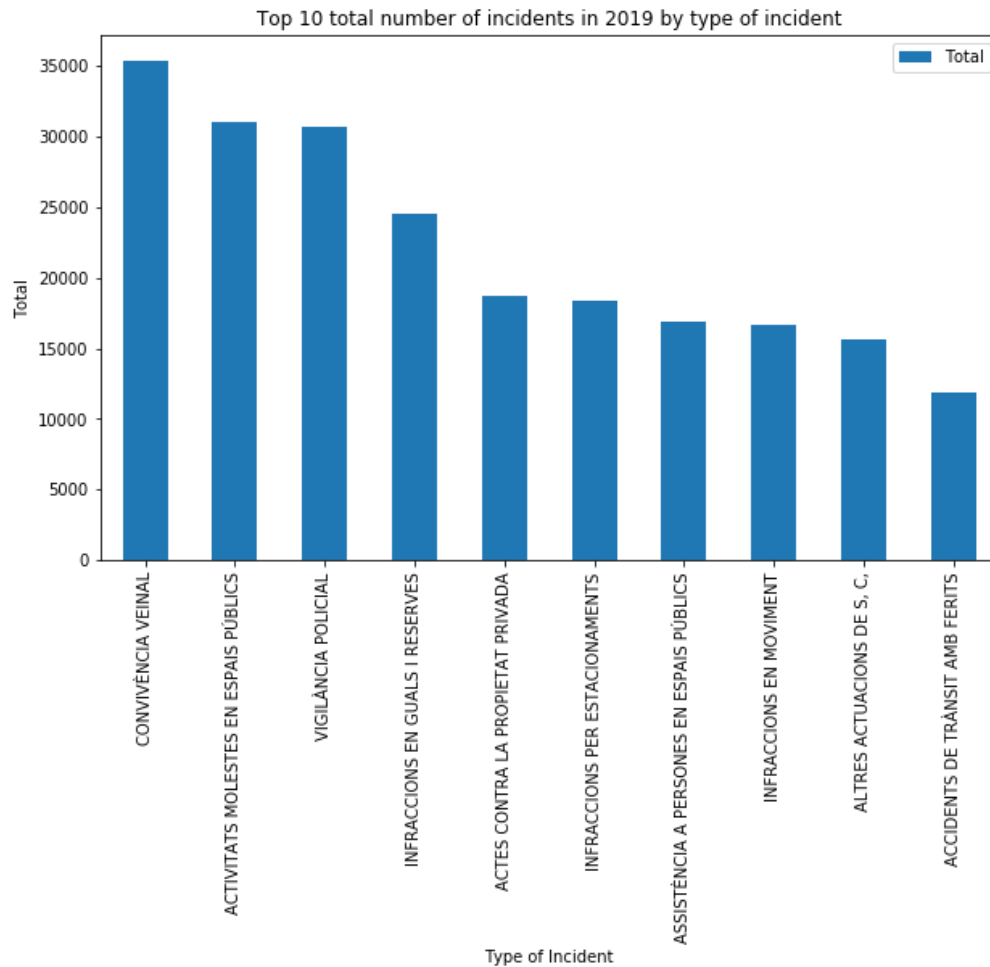
Figure 17 Implementation plot top 10 incidents in 2019

Figure 18 Top 10 incidents with higher occurrence in 2019

4. Model Implementation

All the model implementation can be found in 'Cluster Barcelona Neighborhoods' notebook.

4.1. Model Selection

The idea to resolve the problem was to group the neighborhoods based on their incident's similarity. For example, neighborhoods with the same top incidents occurring during the year will belong to the same group. To achieving that the K-Means model was chosen.

K-Means is a grouping method that groups the data in K groups where each observation belongs to the group with closest mean value.

4.2. Model Testing

The purpose of model testing is to find the best K, or best number of groups, where the observations can be well grouped. Two methods have been implemented to achieve that, the elbow and the silhouette methods.

Also, the data frame used for the testing, as well as for the implementation of the model has been the data frame shown in Figure 6, Top 10 most common incidents in each neighborhood, but dropping the neighborhood column.

The elbow method runs on the dataset K-Means model for a range of K from 1 to 10 and for each K computes the average score for all clusters. The K optimal is the point after which the Euclidean distance start decreasing in a linear way.

The following code computes the elbow method.

```python
from scipy.spatial import distance

# function returns WSS score for k values from 1 to kmax
def calculate_WSS(points, kmax):
    sse = []
    for k in range(1, kmax+1):
        kmeans = KMeans(n_clusters = k).fit(points)
        centroids = kmeans.cluster_centers_
        pred_clusters = kmeans.predict(points)
        curr_sse = 0

        # calculate square of Euclidean distance of each point from its cluster center and add to current WSS
        for i in range(len(points)):
            curr_center = centroids[pred_clusters[i]]
            curr_points = points.loc[i]
            curr_sse = distance.euclidean(curr_points, curr_center)

        sse.append(curr_sse)

    return sse
```

```python
sse = calculate_WSS(incidents_grouped_clustering, 10)
```

```python
plt.plot(range(1, 11), sse,'g')
plt.ylabel('Euclidean Distance')
plt.xlabel('Number of Nabors (K)')
plt.show()
```
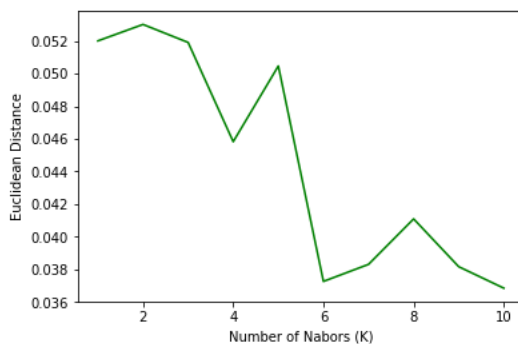

Figure 19 Implementation elbow method

Observing the plot above it is difficult to choose the optimal K. In order to determined which K to choose, another method has been implemented: the silhouette method.

Furthermore, the silhouette method approach measures the quality of a clustering. That is, it determines how well each observation lies within its cluster. A high average silhouette width indicates a good clustering.

The following code computes this method for 2-10 clusters.

```python
from sklearn.metrics import silhouette_score

sil = []
kmax = 10

# dissimilarity would not be defined for a single cluster, thus, minimum number of clusters should be 2
for k in range(2, kmax+1):
    kmeans = KMeans(n_clusters = k).fit(incidents_grouped_clustering)
    labels = kmeans.labels_
    sil.append(silhouette_score(incidents_grouped_clustering, labels, metric = 'euclidean'))
```

```python
plt.plot(range(2, kmax+1), sil,'g')
plt.ylabel('Silhouette score')
plt.xlabel('Number of Nabors (K)')
plt.show()
```
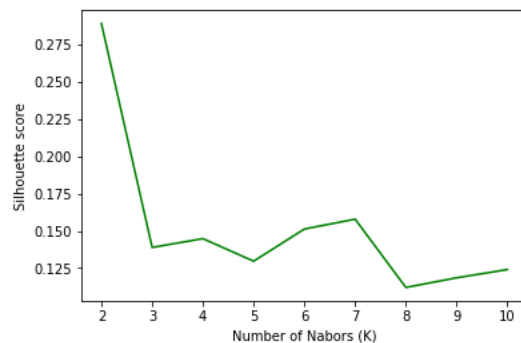

Figure 20 Implementation silhouette method

The results show that 2 clusters maximize the average silhouette values, followed by 7 clusters as second optimal number of clusters.

4.3. Model Implementation
Using the data frame shown in Figure 6, Top 10 most common incidents in each neighborhood, and with K=7 chosen in the analysis of the elbow and silhouette methods. The K-Means method was implemented.

```python
# set number of clusters
kclusters = 7

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Figure 21 Setting K=7 and running the K-Means model

Afterwards the cluster number of each neighborhoods were added together with the latitude and longitude to the same data frame, with the information of the top 10 incidents happening in each neighborhood.

```
# add clustering labels
neighborhoods_incidents_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

neigborhoods_merged = df
neighborhoods_merged.drop(['District Code', 'Neighborhood Code', 'Incident', 'Month', 'Number of Incidents'], axis=1, inplace=True)

# merge barcelona_grouped with barcelona_data to add latitude/longitude for each neighborhood
neighborhoods_merged = neigborhoods_merged.join(neighborhoods_incidents_sorted.set_index('Neighborhood'), on='Neighborhood')
neighborhoods_merged.dropna(subset=['Cluster Labels'], axis=0, inplace=True)

neighborhoods_merged = neighborhoods_merged.drop_duplicates().reset_index(drop=True) # drop duplicates

neighborhoods_merged.head()# add clustering labels
```

Figure 22 Add a new column with the cluster number where the neighborhood belongs to

	Disctrict	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Cluster Labels	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
0	Gràcia	la Vila de Gràcia	41.403178	2.157166	2	INFRACCIONS EN GUALS I RESERVES	MANIFESTACIONS / CONCENTRACIONS	INFRACCIONS PER ESTACIONAMENTS	INFRACCIONS EN MOVIMENT	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INCIDÈNCIES DE LOCALS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	AGRESSIONS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INCIDÈNCIES AMB AFECTACIÓ DE VIA
1	Sant Martí	la Verneda i la Pau	41.423220	2.202940	1	ALARMES D'INCENDI / ROBATORI	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN EDIFICI	INCIDÈNCIES DE LOCALS	INFRACCIONS EN GUALS I RESERVES	INCENDIS	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ANIMALS DE COMPANYIA	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA
2	Eixample	la Nova Esquerra de l'Eixample	41.382816	2.149966	2	INFRACCIONS EN MOVIMENT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INFRACCIONS EN GUALS I RESERVES	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INCENDIS	INFRACCIONS PER ESTACIONAMENTS	MANIFESTACIONS / CONCENTRACIONS	ANIMALS DE COMPANYIA	ASSISTÈNCIA A CONDUCTORS EN V. P.
3	Ciutat Vella	la Barceloneta	41.380653	2.189927	2	AVARIES DE SERVEIS MUNICIPALS	COL·LABORACIÓ AMB ALTRES SERVEIS	INFRACCIONS EN GUALS I RESERVES	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA	CONFLICTES EN LOCALS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ALTRES ACTUACIONS DE S. C.
4	Eixample	la Dreta de l'Eixample	41.394124	2.166471	2	INFRACCIONS EN GUALS I RESERVES	ALTRES ACTUACIONS DE S. C.	OCUPACIONS D'ESPAIS PÚBLICS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	CONVIVÈNCIA VEÏNAL	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA	MANIFESTACIONS / CONCENTRACIONS	INSPECCIONS / CONTROLS EN LOCALS	INFRACCIONS PER ESTACIONAMENTS	INFRACCIONS EN MOVIMENT (VEHICLES)

Figure 23 Top 10 most common incidents with the cluster number the neighborhood belongs to

Finally, using folium, the different neighborhoods were plot in a map. Each neighborhood is plotted in the map with a specific color depending the clusters they belong to.

```
import folium
import geocoder

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode('Barcelona')
latitude = location.latitude
longitude = location.longitude

# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=12)

# 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(neigborhoods_merged['Neighborhood Latitude'], neigborhoods_merged['Neighborhood Longitude'], n
eigborhoods_merged['Neighborhood'], neigborhoods_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[int(cluster)-1],
        fill=True,
        fill_color=rainbow[int(cluster)-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

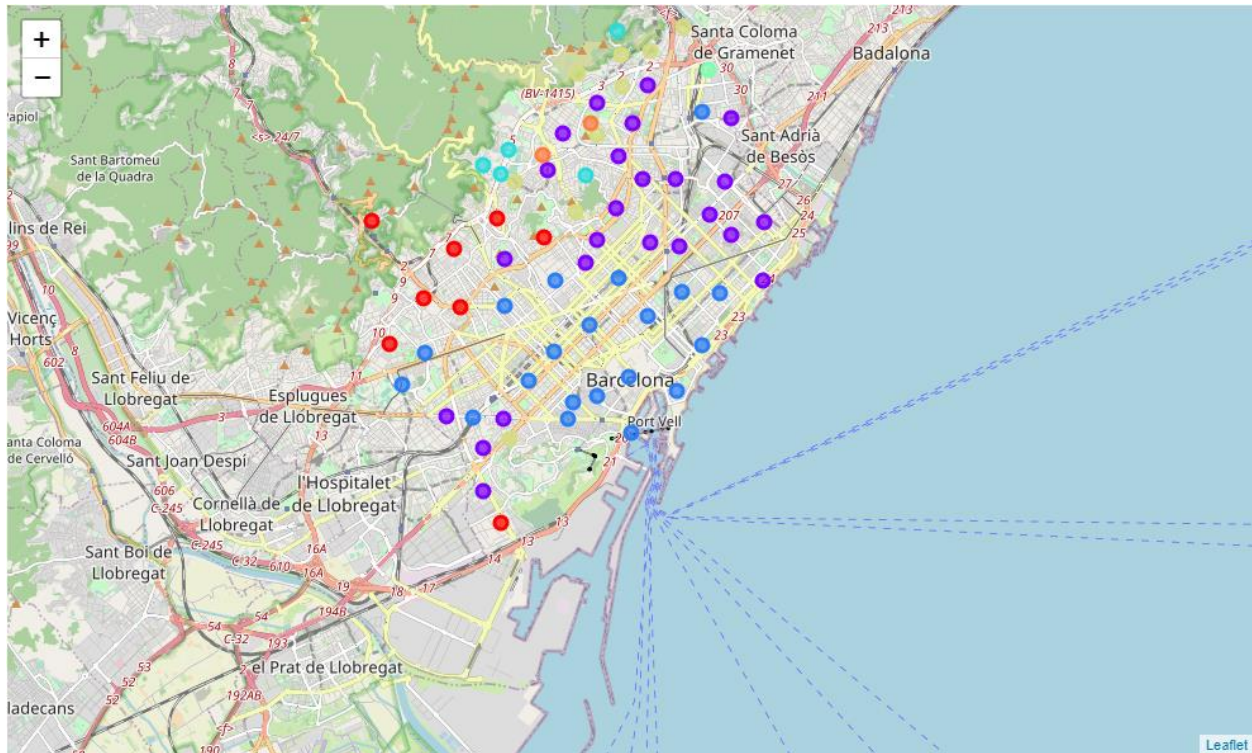*Figure 24 Implementation plot neighborhoods' clusters*

Figure 25 Neighborhoods' clusters in Barcelona map

5. Results

As it can be seen in the data analysis point, the neighborhood with the higher number of incidents are 'el raval', followed by 'la Dreta de l'Eixample' and 'el barrio Gòtic'.

Furthermore, the months with higher incidents rate are in summer. The incidents start growing in February and reach their peak in July. As during the summer there are more tourists in the city.

Finally, the most common incidents happening are neighborhood coexistence, annoying activities in public spaces, police surveillance, infractions in fords and reservations and acts against private property.

Moreover, in the model implementation point, the resulting optimal number of clusters is seven. Thus, the neighborhoods have been assigned to this clusters by its 10th most common incidents and in Figure 25, Neighborhoods' clusters in Barcelona map, it can be observed how the neighborhoods of high-power acquisition belonging to cluster 0 (red) have among the most common incidents automated signs, administrative interventions and assistance to people in public spaces.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
18	Sarrià	COL·LABORACIÓ AMB ALTRES SERVEIS	SENYALS AUTOMATITZADES	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA	INCIDÈNCIES DE LOCALS	INFRACCIONS EN GUALS I RESERVES	ANIMALS DE COMPANYIA	INFRACCIONS EN MOVIMENT	ASSISTÈNCIA A CONDUCTORS EN V, P,	INFRACCIONS PER ESTACIONAMENTS	ALTRES ACTUACIONS DE S, C,
29	Vallcarca i els Penitents	ALTRES ACTUACIONS DE S, C,	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INFRACCIONS EN MOVIMENT	ASSISTÈNCIA A PERSONES EN EDIFICI	ALARMES D'INCENDI / ROBATORI	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	RESTRICCIONS TEMPORALS DE TRÀNSIT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	COL·LABORACIÓ AMB ALTRES SERVEIS
39	les Tres Torres	INFRACCIONS EN MOVIMENT	ACTES CONTRA LA PROPIETAT PRIVADA	INCENDIS	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INFRACCIONS EN GUALS I RESERVES	CONVIVÈNCIA VEINAL	INFRACCIONS PER ESTACIONAMENTS

Figure 26 Cluster 0

Then in the neighborhoods surrounding the city center (pink) people with lower incomes and immigrants live belong to cluster 1 among the most common incidents are: traffic accidents, fires, annoying activities in public spaces and assistance to people in buildings.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
1	la Verneda i la Pau	ALARMES D'INCENDI / ROBATORI	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN EDIFICI	INCIDÈNCIES DE LOCALS	INFRACCIONS EN GUALS I RESERVES	INCENDIS	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ANIMALS DE COMPANYIA	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA
6	la Prosperitat	INCENDIS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN EDIFICI	CONVIVÈNCIA VEINAL	COL·LABORACIÓ AMB ALTRES SERVEIS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA	ALTRES ACTUACIONS DE S, C,	INFRACCIONS EN GUALS I RESERVES	INCIDÈNCIES DE LOCALS
7	Provençals del Poblenou	INFRACCIONS PER ESTACIONAMENTS	ACCIDENTS DE TRÀNSIT SENSE FERITS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ALTRES ACTUACIONS DE S, C,	ACTES CONTRA LA PROPIETAT PRIVADA	ALTRES INTERVENCIONS DE P. ADMINISTRATIVA	SENYALS AUTOMATITZADES	COL·LABORACIÓ AMB ALTRES SERVEIS	INCENDIS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS

Figure 27 Cluster 1

Neighborhoods in the city center belong to the same cluster (dark blue) belong to cluster 2. This is the cluster with more neighborhoods and the most common incidents are: movement violations, incidents in premises, annoying activities in public spaces, assaults and conflicts in premises among others.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
0	la Vila de Gràcia	INFRACCIONS EN GUALS I RESERVES	MANIFESTACIONS / CONCENTRACIONS	INFRACCIONS PER ESTACIONAMENTS	INFRACCIONS EN MOVIMENT	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INCIDÈNCIES DE LOCALS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	AGRESSIONS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INCIDÈNCIES AMB AFECTACIÓ DE VIA
2	la Nova Esquerra de l'Eixample	INFRACCIONS EN MOVIMENT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INFRACCIONS EN GUALS I RESERVES	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INCENDIS	INFRACCIONS PER ESTACIONAMENTS	MANIFESTACIONS / CONCENTRACIONS	ANIMALS DE COMPANYIA	ASSISTÈNCIA A CONDUCTORS EN V, P,
3	la Barceloneta	AVARIES DE SERVEIS MUNICIPALS	COL·LABORACIÓ AMB ALTRES SERVEIS	INFRACCIONS EN GUALS I RESERVES	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	CONFLICTES EN LOCALS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ALTRES ACTUACIONS DE S, C,
4	la Dreta de l'Eixample	INFRACCIONS EN GUALS I RESERVES	ALTRES ACTUACIONS DE S, C,	OCUPACIONS D'ESPAIS PÚBLICS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	CONVIVÈNCIA VEINAL	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	MANIFESTACIONS / CONCENTRACIONS	INSPECCIONS / CONTROLS EN LOCALS	INFRACCIONS PER ESTACIONAMENTS	INFRACCIONS EN MOVIMENT (VEHICLES)
5	el Poble-sec	INCIDÈNCIES DE LOCALS	ESPECTACLES EN ESPAIS PÚBLICS	INFRACCIONS EN GUALS I RESERVES	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	AGRESSIONS	ALARMES D'INCENDI / ROBATORI	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INCENDIS	ALTRES ACTUACIONS DE S, C,
8	Sants	MANIFESTACIONS / CONCENTRACIONS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INCIDÈNCIES DE LOCALS	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	INCIDÈNCIES AMB AFECTACIÓ DE VIA	INCENDIS	ASSISTÈNCIA A PERSONES EN EDIFICI	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS
10	el Poblenou	COL·LABORACIÓ AMB ALTRES SERVEIS	SENYALS AUTOMATITZADES	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	INCIDÈNCIES DE LOCALS	ANIMALS DE COMPANYIA	INFRACCIONS EN GUALS I RESERVES	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ASSISTÈNCIA A PERSONES EN EDIFICI
11	el Raval	INFRACCIONS PER ESTACIONAMENTS	INCENDIS	MANIFESTACIONS / CONCENTRACIONS	INSPECCIONS / CONTROLS EN LOCALS	COL·LABORACIÓ AMB ALTRES SERVEIS	INFRACCIONS EN MOVIMENT	INFRACCIONS EN GUALS I RESERVES	INCIDÈNCIES DE LOCALS	ANIMALS / INSECTES PERILLOSOS	ANIMALS DE COMPANYIA

Figure 28 Cluster 2

Then the neighborhoods close to the mountain can be divided in four: cluster 3 (light blue), cluster 4 (light green), cluster 5 (green) and cluster 6 (orange).

In cluster 3 the most common incidents are assistance to people in public spaces, annoying activities in public spaces and parking violations.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
22	el Coll	ANIMALS DE COMPANYIA	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INFRACCIONS PER ESTACIONAMENTS	CONVIVÈNCIA VEINAL	INFRACCIONS EN GUALS I RESERVES	ACCIDENTS DE TRÀNSIT AMB FERITS	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ALTRES ACTUACIONS DE S, C,	ACCIDENTS DE TRÀNSIT SENSE FERITS	ACTES CONTRA LA PROPIETAT PRIVADA
32	la Font d'en Fargues	INCIDÈNCIES AMB PERILL PEL TRÀNSIT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	INFRACCIONS PER ESTACIONAMENTS	ALTRES ACTUACIONS DE S, C,	CONVIVÈNCIA VEINAL	INFRACCIONS EN MOVIMENT	INFRACCIONS EN GUALS I RESERVES	ACTES CONTRA LA PROPIETAT PRIVADA

Figure 29 Cluster 3

Then in cluster 4 the most common incidents are neighborhood coexistence and police surveillance.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
13	Baró de Viver	CONVIVÈNCIA VEINAL	VIGILÀNCIA POLICIAL	ALTRES ACTUACIONS DE S, C,	INFRACCIONS EN GUALS I RESERVES	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INFRACCIONS EN MOVIMENT	INCENDIS	COL·LABORACIÓ AMB ALTRES SERVEIS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	ACTES CONTRA LA PROPIETAT PRIVADA
62	Vallbona	VIGILÀNCIA POLICIAL	INFRACCIONS EN MOVIMENT	CONVIVÈNCIA VEINAL	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INCENDIS	ACCIDENTS DE TRÀNSIT AMB FERITS	ALTRES ACTUACIONS DE S, C,	ALARMES D'INCENDI / ROBATORI	SUPORTS	ACTES CONTRA LA PROPIETAT PRIVADA

Figure 30 Cluster 4

In cluster 5 the most common incidents are motion violations, parking violations, and infractions in fords and reservations.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
24	el Turó de la Peira	ALTRES ACTUACIONS DE S, C,	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	INCENDIS	INCIDÈNCIES DE LOCALS	INFRACCIONS EN GUALS I RESERVES	INFRACCIONS EN MOVIMENT	INFRACCIONS PER ESTACIONAMENTS	ANIMALS DE COMPANYIA	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN EDIFICI
35	les Roquetes	CONVIVÈNCIA VEINAL	ACCIDENTS DE TRÀNSIT AMB FERITS	INFRACCIONS EN MOVIMENT	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	INFRACCIONS EN GUALS I RESERVES	ALTRES ACTUACIONS DE S, C,	INCIDÈNCIES AMB AFECTACIÓ DE VIA	ACTES CONTRA LA PROPIETAT PRIVADA	ALTRES INTERVENCIONS DE P, ADMINISTRATIVA	INFRACCIONS PER ESTACIONAMENTS

Figure 31 Cluster 5

Finally, in cluster 6, the most common incidents are assistance people in public spaces, parking violations and neighborhood coexistence.

	Neighborhood	1st Most Common Incidents	2nd Most Common Incidents	3rd Most Common Incidents	4th Most Common Incidents	5th Most Common Incidents	6th Most Common Incidents	7th Most Common Incidents	8th Most Common Incidents	9th Most Common Incidents	10th Most Common Incidents
59	Can Peguera	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	CONVIVÈNCIA VEINAL	ALTRES ACTUACIONS DE	VIGILÀNCIA POLICIAL	INFRACCIONS PER ESTACIONAMENTS	ASSISTÈNCIA A PERSONES EN EDIFICI	ANIMALS DE COMPANYIA	INFRACCIONS EN GUALS I RESERVES	AGRESSIONS
70	la Clota	ASSISTÈNCIA A PERSONES EN ESPAIS PÚBLICS	ACCIDENTS DE TRÀNSIT AMB FERITS	INFRACCIONS EN MOVIMENT	ACTES CONTRA LA PROPIETAT PRIVADA	ALTRES ACTUACIONS DE S, C,	INFRACCIONS PER ESTACIONAMENTS	CONVIVÈNCIA VEINAL	ACTIVITATS MOLESTES EN ESPAIS PÚBLICS	ANIMALS DE COMPANYIA	ABOCAMENT DE RESIDUS

Figure 32 Cluster 6

6. Discussion

The neighborhoods belonging to cluster 2, are the more dangerous and where most of the incidents happened there, moreover 9 out of 10 neighborhoods with the higher number of incidents belong to this cluster.

Furthermore, when the neighborhoods are further away from the city center the number of incidents go down, and the neighborhoods in the suburbs have incidents related with neighborhood coexistence and assistance.

7. Conclusion

To conclude the local police should have more resources in the areas where most of the incidents are happening, specially neighborhoods belonging to cluster 1 and 2 should be taken into consideration, and having less in the suburbs where most of the incidents are because of the neighborhood coexistence itself.