

Delitos en la Ciudad de Buenos Aires

Periodo 2016-2019

Grupo 04: Nicolás Sidicaro, Melina Rivero, Julia Martinuzzi
Universidad tecnológica Nacional (UTN)
Ciudad Autónoma de Buenos Aires

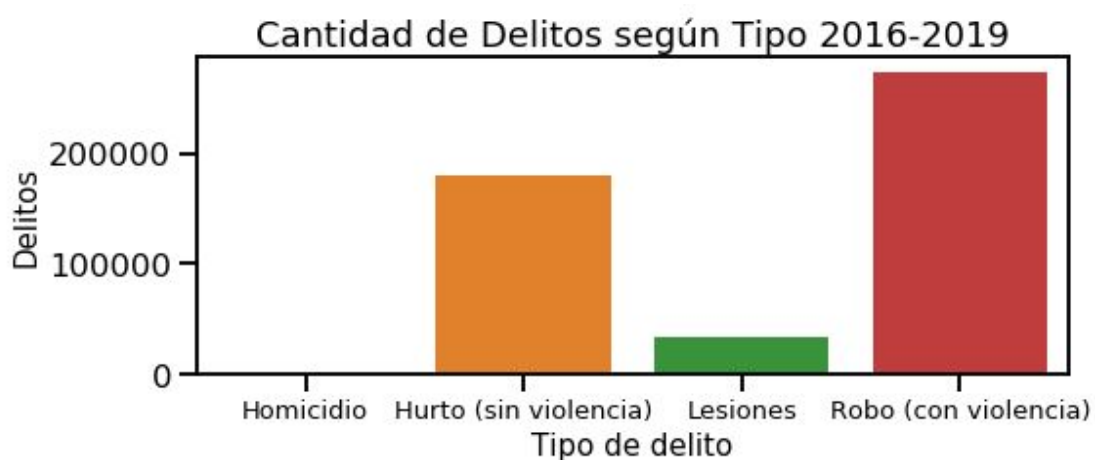
1. Introducción

El presente trabajo pretende analizar los delitos en la Ciudad de Buenos Aires (C.A.B.A) en el periodo 2016-2019 publicados en el portal de BA DATA. Se pretende analizar, en caso de que exista, la relación entre la población y la cantidad de delitos, por barrio y fracción censal.

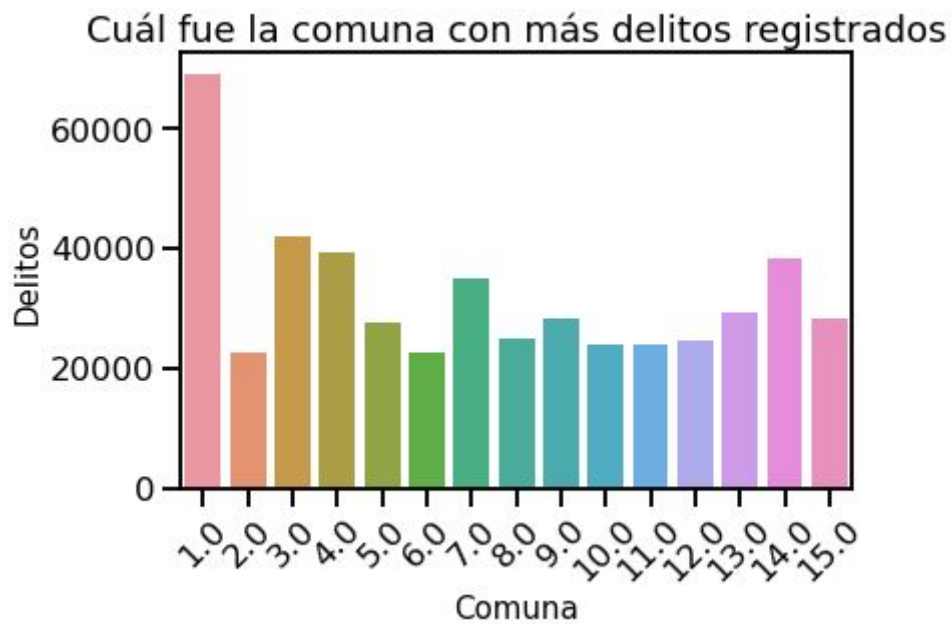
A su vez, teniendo en cuenta que la Ciudad tiene una dinámica particular por zona, día de la semana y horario, se realiza un k-means para poder identificar grupos. Para esto consideraremos tipo de delitos, franja horaria y población.

2. Análisis exploratorio de los delitos

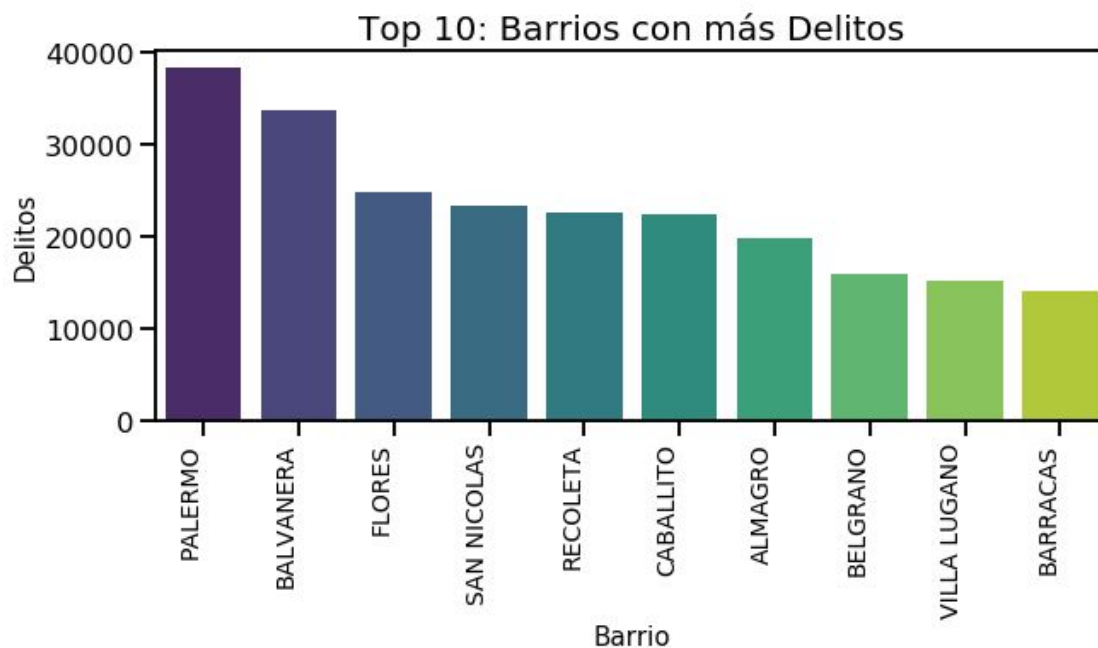
Se observa que los tipos más frecuentes de delitos son: Robo -con violencia- y Hurto -sin violencia-.



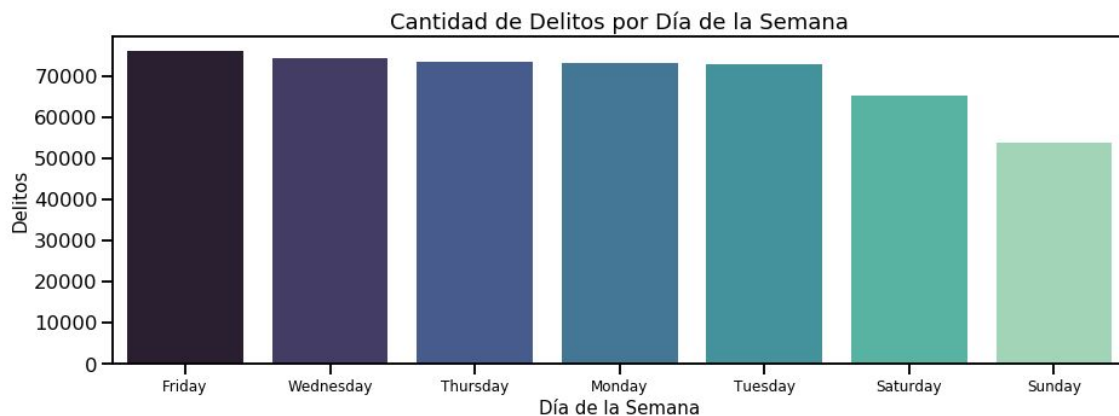
La Comuna 1 tiene la mayor cantidad de delitos en el periodo, registrando 69.208 (más del doble que la media y la mediana de C.A.B.A. por comuna), seguida por las Comunas 3 y 4.



Los barrios con más delitos son Palermo, que registra 38.338, y Balvanera, con 33.749 (más del triple que la media registrada en C.A.B.A. por barrio que es de 10.004).

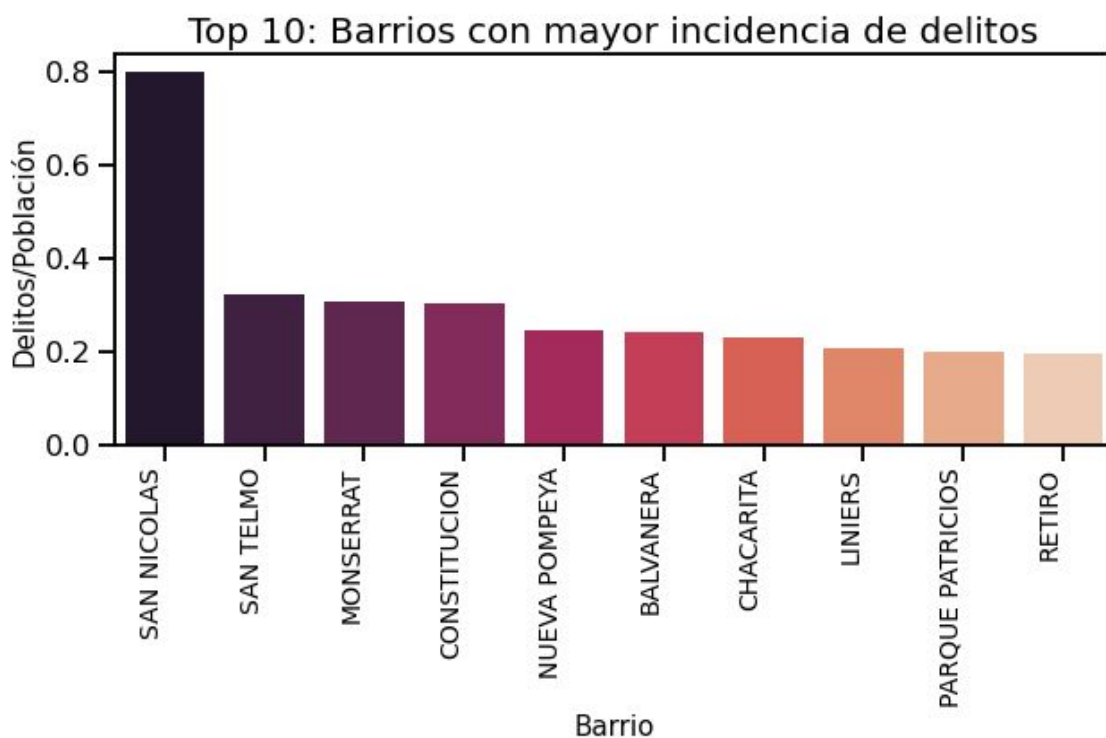


A su vez, en los fines de semana la cantidad de delitos es menor que durante los días de semana. Esto puede deberse a que la transitabilidad en la Ciudad durante la semana es mayor.



Es importante mencionar que no se observa diferencias significativas por mes ni por año.

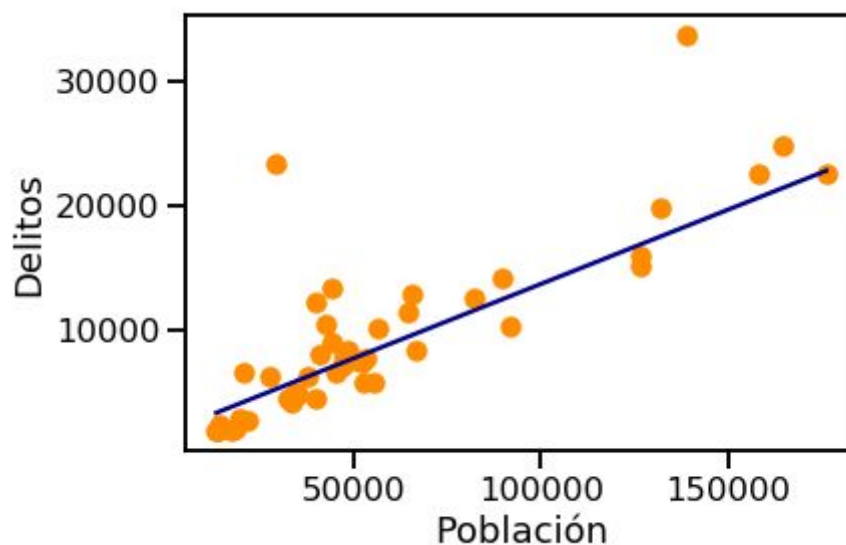
El barrio con mayor incidencia de delitos/población es San Nicolás. Esto puede deberse a la elevada circulación de personas que hay en ese barrio durante los días de semana, personas que trabajan allí y que provienen de otros barrios y de otras ciudades del AMBA.



3. Modelos

Regresión lineal

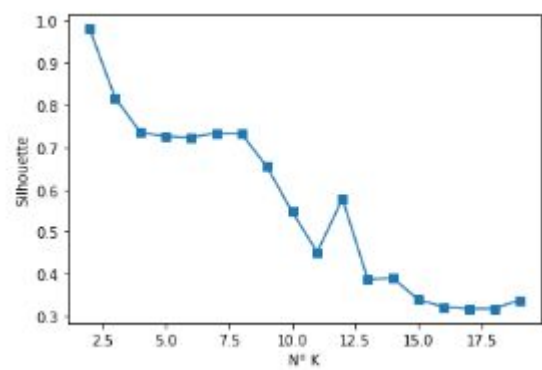
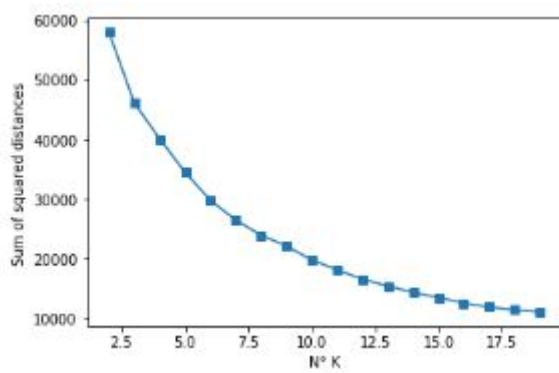
Se realiza un scatterplot para ver si hay una relación entre la población de cada barrio y los delitos. A simple vista pareciera que existe una relación lineal entre ambas variables. Realizamos una regresión lineal y obtenemos un R cuadrado de 0.62.



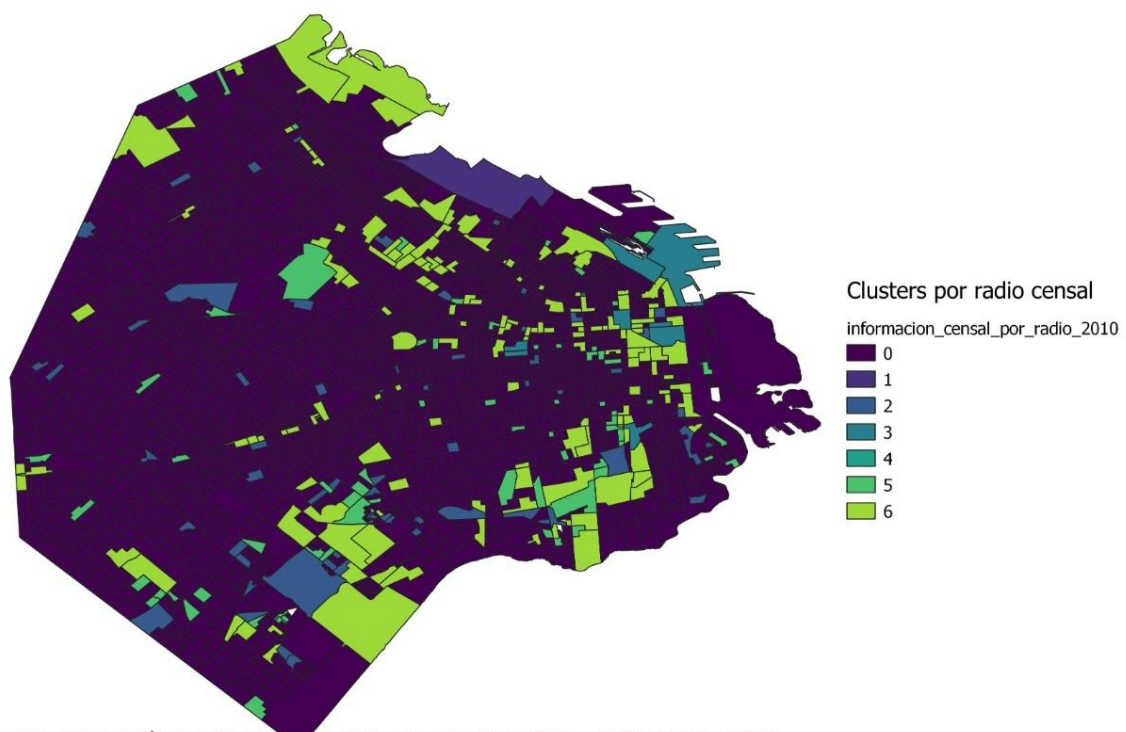
K-Means

Realizamos un algoritmo de k-means para identificar clusters de delitos. Excluimos del análisis los siniestros viales

En una primera instancia lo aplicamos a barrios y establecemos $k=3$. Con este resultado vemos cierta similitud entre barrios con mayor vulnerabilidad social -como Nueva Pompeya, Flores, Balvanera, San Cristobal y Retiro- con Palermo. Sin embargo, esto puede deberse por un lado a la cantidad de personas que viven en Palermo como al movimiento nocturno de ese barrio. Por otro lado encontramos San Nicolás, donde es menor la cantidad de población, pero dada el movimiento diurno que hay por la concentración de oficinas la cantidad de delitos por población habitante es alta y lo lleva a diferenciarse de otros barrios. Luego aplicamos el algoritmo a fracciones censales y establecemos $k=7$ utilizando como features tipo de delitos y franja horaria. El Silhouette index que se obtiene es de 0.73, lo que significa que la similitud entre muestras intra-cluster es alta y que la similitud inter-cluster es baja.



Localización de clusters del delito en base a radio censal



Fuente: elaboración propia en base a datos Buenos Aires Data - Delito 2016 - 2019

4. Conclusiones

La regresión indica que no es suficiente la cantidad de población de un barrio para determinar la cantidad de delitos.

En el cluster realizado por barrio se encontró que no es posible lograr una buena diferenciación entre ellos a nivel agregado, dadas las diferencias hacia dentro de los barrios.

En el cluster a radio censal detectaron mayores diferencias, por lo tanto la mayor granularidad nos permite encontrar mejores relaciones entre los radios censales respecto al delito.