




Inmuebles de Buenos Aires  
Trabajo Práctico N 1

GRUPO:

Apellido y Nombre	Padrón	Correo electrónico
Llauró, Manuel Luis	95736	llauromanuel@gmail.com
Lavandeira, Lucas	98042	lucaslavandeira@gmail.com
Funes, Federico	98372	fede.funes96@gmail.com

Github: 

<https://github.com/lucaslavandeira/7506-tp>

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Análisis del Set de Datos</b>	<b>3</b>
2.1. Limpieza previa . . . . .	3
2.2. Pantallazo general . . . . .	3
2.3. Relación de precios con otros atributos . . . . .	4
2.3.1. Expensas . . . . .	4
2.3.2. Superficie . . . . .	4
2.3.3. Ambientes . . . . .	5
<b>3. Barrios más caros</b>	<b>5</b>
<b>4. Relación precio según distancia a un espacio público</b>	<b>7</b>
4.0.1. Según distancia a una estación de subte . . . . .	8
4.0.2. Según distancia a una facultad pública . . . . .	9
4.0.3. Según distancia al metrobus . . . . .	9
4.0.4. Según distancia a un hospital . . . . .	10
<b>5. Precios según atributos de las propiedades</b>	<b>11</b>
5.0.1. Precio según tengan pileta . . . . .	13
5.0.2. Precio según tengan garage . . . . .	13
5.0.3. Precio según tengan jardín . . . . .	14
<b>6. Incremento del precio por metro cuadrado de los Barrios de CABA</b>	<b>14</b>
6.0.1. Metodología . . . . .	14
6.0.2. Periodo 2012 - 2013 . . . . .	16
6.0.3. Periodo 2013 - 2014 . . . . .	16
6.0.4. Periodo 2014 - 2015 . . . . .	17
6.0.5. Periodo 2015 - 2016 . . . . .	18
6.0.6. Periodo 2016 - 2017 . . . . .	19
6.0.7. Conclusiones . . . . .	19
<b>7. Porcentaje de evolución de precios entre los distintos periodos</b>	<b>19</b>
7.0.1. Evolución de precios del 2013 al 2015 . . . . .	20
7.0.2. Evolución de precios del 2014 al 2016 . . . . .	20
7.0.3. Evolución de precios del 2015 al 2017 . . . . .	21
7.0.4. Evolución de precios del 2013 al 2017 . . . . .	22

## 1. Introducción

El objetivo de este informe es mostrar el análisis realizado sobre el set de datos del registro histórico de publicaciones de ventas de propiedades en <http://www.properati.com.ar/> desde el 2012 hasta el 2017, en la Ciudad de Buenos Aires y Gran Buenos Aires. Se estudió:

- El incremento del valor del metro cuadrado promedio por cada Barrio de la Capital por año.
- La relación del precio total de las propiedades con varias de sus atributos, superficie, cantidad de ambientes, ubicación geográfica, expensas, entre otros.
- Precio de las propiedades según el inmueble: negocio, departamento, PH, o casa.
- Precio de las propiedades según atributos de la misma: pileta, jardín, garage, entre otros

Para el estudio apropiado y la visualización de propiedades del set de datos se tuvieron que implementar varias heurísticas que permitan llegar a conclusiones ricas. Este proceso de limpieza de datos consistió en identificar los conocidos casos borde o *outliers*, tales como propiedades con precios muy elevados por encima de la norma, que si fueran tenidos en cuenta afectaría negativamente la visualización de varios gráficos informativos.

Un *notebook* del análisis exploratorio hecho puede encontrarse en el repositorio de GitHub indicado en la carátula.

## 2. Análisis del Set de Datos

### 2.1. Limpieza previa

El set de datos provisto consiste en varios archivos de formato CSV, uno por mes entre los años 2013 a 2017, en los cuales cada fila representa una publicación de una propiedad activa. Varias de las columnas de este CSV no eran relevantes para nuestro estudio de evolución de precios, el ejemplo más notable siendo una columna de URL a imágenes de la propiedad. Además, el set de datos contiene publicaciones en toda Latinoamérica, mientras que el enfoque de nuestro trabajo es sólo en el Gran Buenos Aires. Estos factores llevaron a la decisión de hacer una limpieza previa de los datos, filtrar las propiedades de la región que nos concierne, y borrar columnas que no vamos a usar, y juntar todos los datos en un único archivo .csv fácil de poder procesar. El resultado final de esta operación nos dejó con más de 2 millones de datos listos para usar. En la mayoría de los gráficos se desestimaron las propiedades con precios muy elevados para poder mostrar una mejor visualización de los datos estudiados, y también las que tienen precios considerados como inválidos por ser muy menores.

### 2.2. Pantallazo general

El enfoque principal de nuestro set de datos es el precio. Lo primero que se hizo fue ver cómo se distribuye el mismo. Tomamos siempre el precio de las propiedades en dólares, pues en Argentina los precios en moneda nacional son demasiado volátiles. Los valores en la tabla 1 muestra que la media se ubica alrededor de los \$200.000 dólares, y además que la mayoría de los datos se ubica en ese rango, pues su cuantil 75 % está bastante cercano. Esto quiere decir que casi todos los precios están alrededor de ese rango, pero viendo el máximo, bastante mayor que la media, hay algunas propiedades que tienen un precio bastante más elevado, pero son de una cantidad muy reducida.

count	4.304400e+05
mean	2.068996e+05
std	3.597598e+05
min	2.001125e+04
25 %	8.865248e+04
50 %	1.315810e+05
75 %	2.200000e+05
max	5.000000e+07

Cuadro 1: Valores estadísticos de precios

Lo primero realizado con los datos es poder visualizar la distribución de los precios de las propiedades que tenemos. En la figura 1 se comprueba que la gran mayoría de las propiedades tienen un precio por debajo de los \$200.000 dólares, y que la media es levemente mayor a eso debido a las propiedades con valor muy por encima. Se concluye que hay unas pocas propiedades con precio muy elevado, y luego le sigue el grueso de las mismas con un precio más reducido.

El primer estudio particular es ver cómo se distribuye el precio clasificando las propiedades según el tipo. El set de datos trae cuatro tipos de propiedades

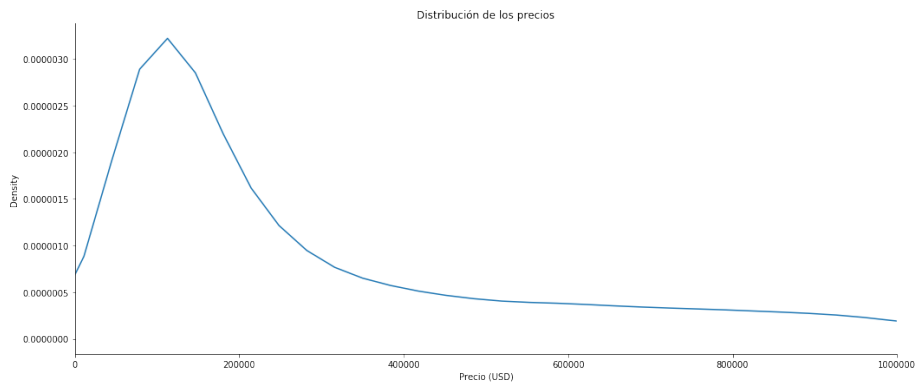


Figura 1: Distribución de los precios

distintos: casa, departamento, negocio y PH. En la figura 2 se puede apreciar que los departamentos son la mayoría de las propiedades estudiadas (luego de hacer un filtro por propiedades que no tuviesen clasificación alguna). Como es de esperar, los precios de las propiedades de este tipo están en norma con la distribución total del precio, mientras que las demás varían un poco: las casas y los negocios tienen precios más volátiles y elevados, lo cual es lógico y concuerda con la realidad.

tipo	cantidad
apartment	289440
house	105353
PH	26864
store	8783

Cuadro 2: Propiedades clasificadas por su tipo

## 2.3. Relación de precios con otros atributos

### 2.3.1. Gastos

Para la cantidad reducida de propiedades que tenían gastos definidos, se hizo un *scatterplot* de los precios contra los gastos a ver si se seguía una relación lineal. Dentro de todo se puede apreciar un aumento de ese tipo sobre el precio. Lamentablemente hay una falta de datos de propiedades de precio elevado con gastos definidos.

### 2.3.2. Superficie

El estudio de superficie es similar al anterior, solo que se puede afirmar con más exactitud que el tamaño de la propiedad está relacionado de manera lineal con el precio de la misma.

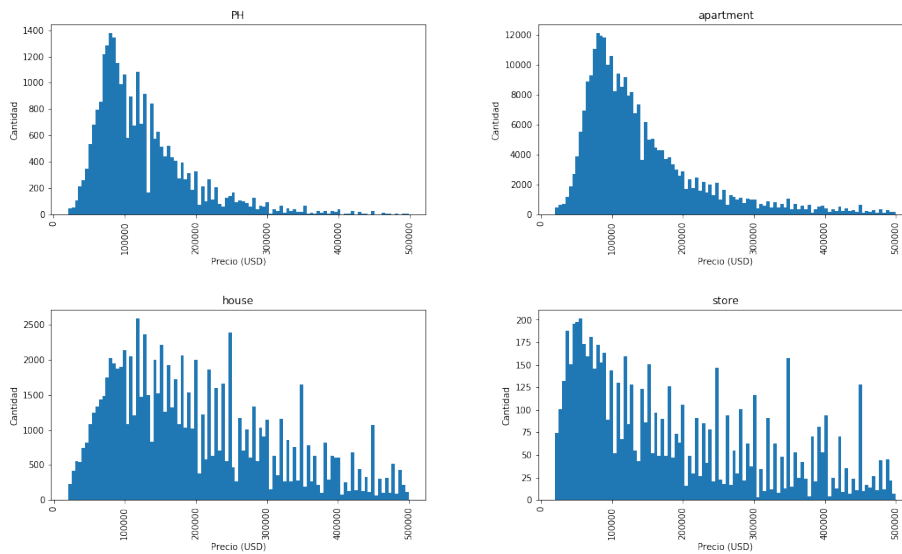


Figura 2: Precios según el tipo de propiedad



Figura 3: Precio vs. Expensas

### 2.3.3. Ambientes

De manera similar se puede apreciar un aumento del precio correspondiente al aumento de los ambientes de una propiedad. La gran mayoría de las propiedades cargadas en el conjunto de datos tiene entre 1 y 3 ambientes, pero se nota que las que tienen más, por lo general son más caras que las demás. En particular, el precio medio de las propiedades agrupadas por cantidad de habitaciones aumenta de manera lineal.

## 3. Barrios más caros

Como una introducción al estudio del precio según la ubicación geográfica de las propiedades, se hallaron los barrios con precio de propiedad promedio más

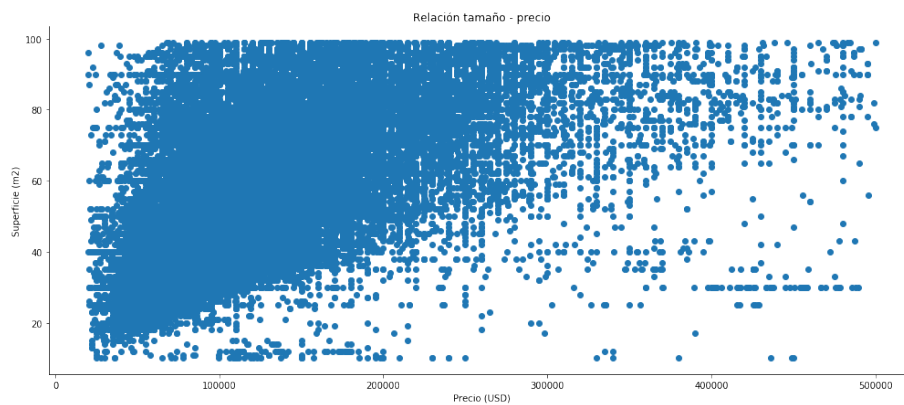


Figura 4: Precio vs. Superficie

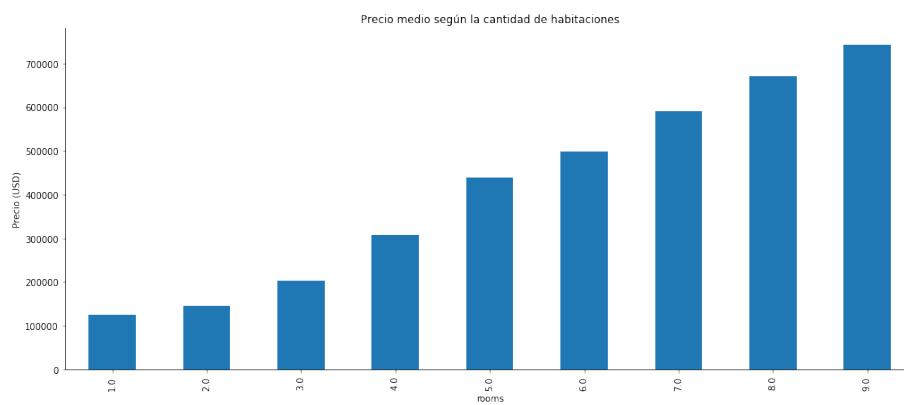


Figura 5: Precio según cantidad de habitaciones

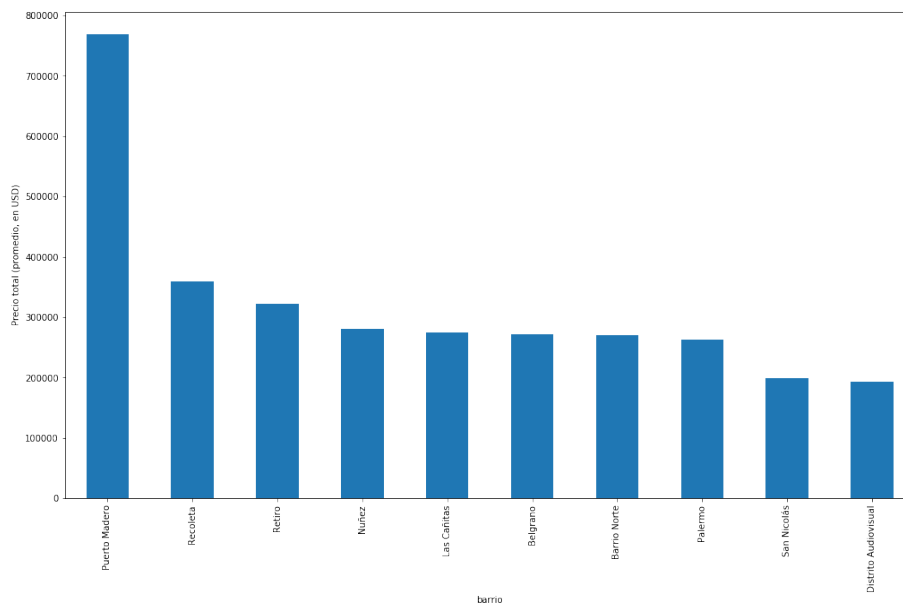


Figura 6: Barrios más caros de Capital Federal

caro. Esta información es sumamente valuable para poder realizar predicciones sobre el precio tasado de una propiedad a futuro, el indicador más fuerte al momento de predecir un precio de manera automática serán los precios de las propiedades más cercanas.

El cálculo fue realizado solo sobre los distritos de Capital Federal debido a la falta de volumen de datos de otros sectores del Gran Buenos Aires.

Los resultados coinciden en mayor parte con publicaciones independientes al estudio, en donde Puerto Madero se ubica como el barrio más caro, seguido de barrios de la zona Norte de la ciudad, como Palermo, Núñez y Belgrano.

#### 4. Relación precio según distancia a un espacio público

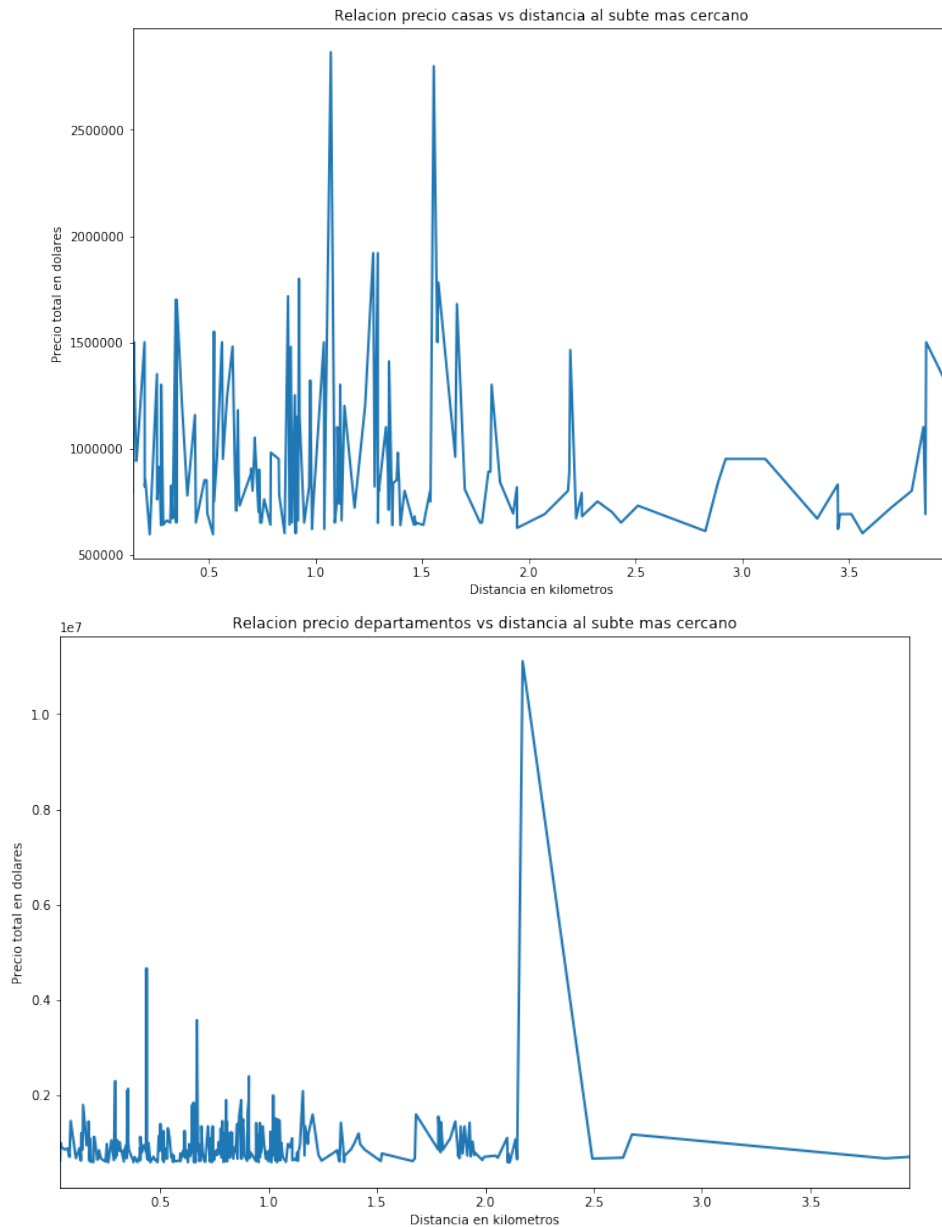
Se propone verificar si existe algún tipo de relación entre los precios de las casas y departamentos de Capital Federal y su cercanía a un espacio público. Se calcularán las distancias entre las propiedades más caras a cada espacio público usando el algoritmo de Haversine, que nos permite calcular la distancia entre dos puntos basándonos en la latitud y longitud de los mismos, y nos quedaremos con la menor distancia según a que espacio corresponda, de esta manera, podremos comparar los valores promedio de las propiedades según las distancias a dichos espacios. Se utilizaran solamente aquellas propiedades que se encuentren en un radio menor a 4km al espacio público más cercano, de esta manera filtramos propiedades que se encuentren con datos erróneos de latitud y/o longitud, también, utilizaremos únicamente las propiedades más caras de Capital Federal, ya que aplicar el algoritmo de Haversine para calcular la distancia más chica entre una propiedad y un espacio público puede tardar



demasiado para muchos datos.

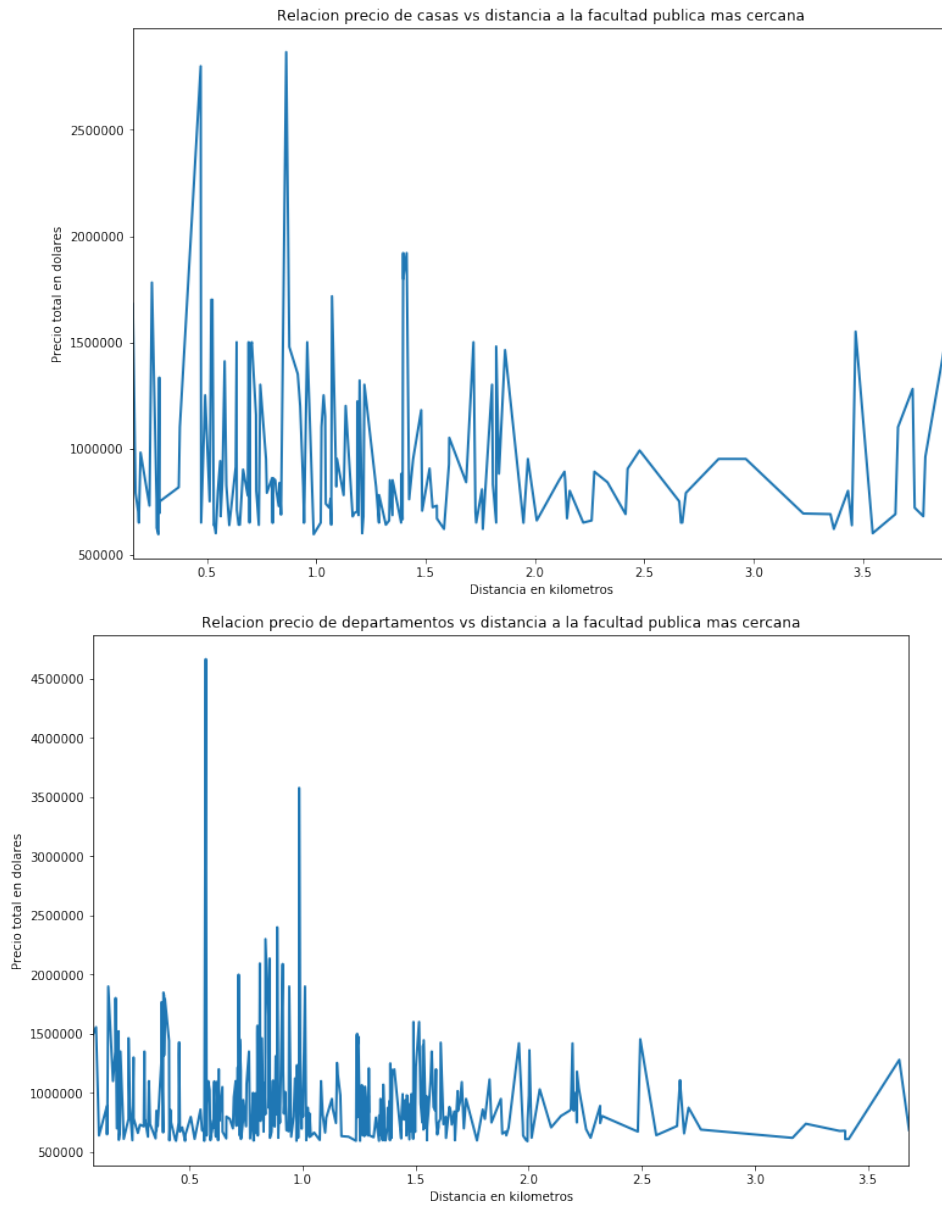
#### 4.0.1. Según distancia a una estación de subte

Los resultados implican que para las casas no existe relación entre la cercanía a una estación de subte y los precios de las mismas, en cambio, para los departamentos, a partir de los 2.2km aproximadamente, puede notarse como los precios de los mismos disminuyen considerablemente.



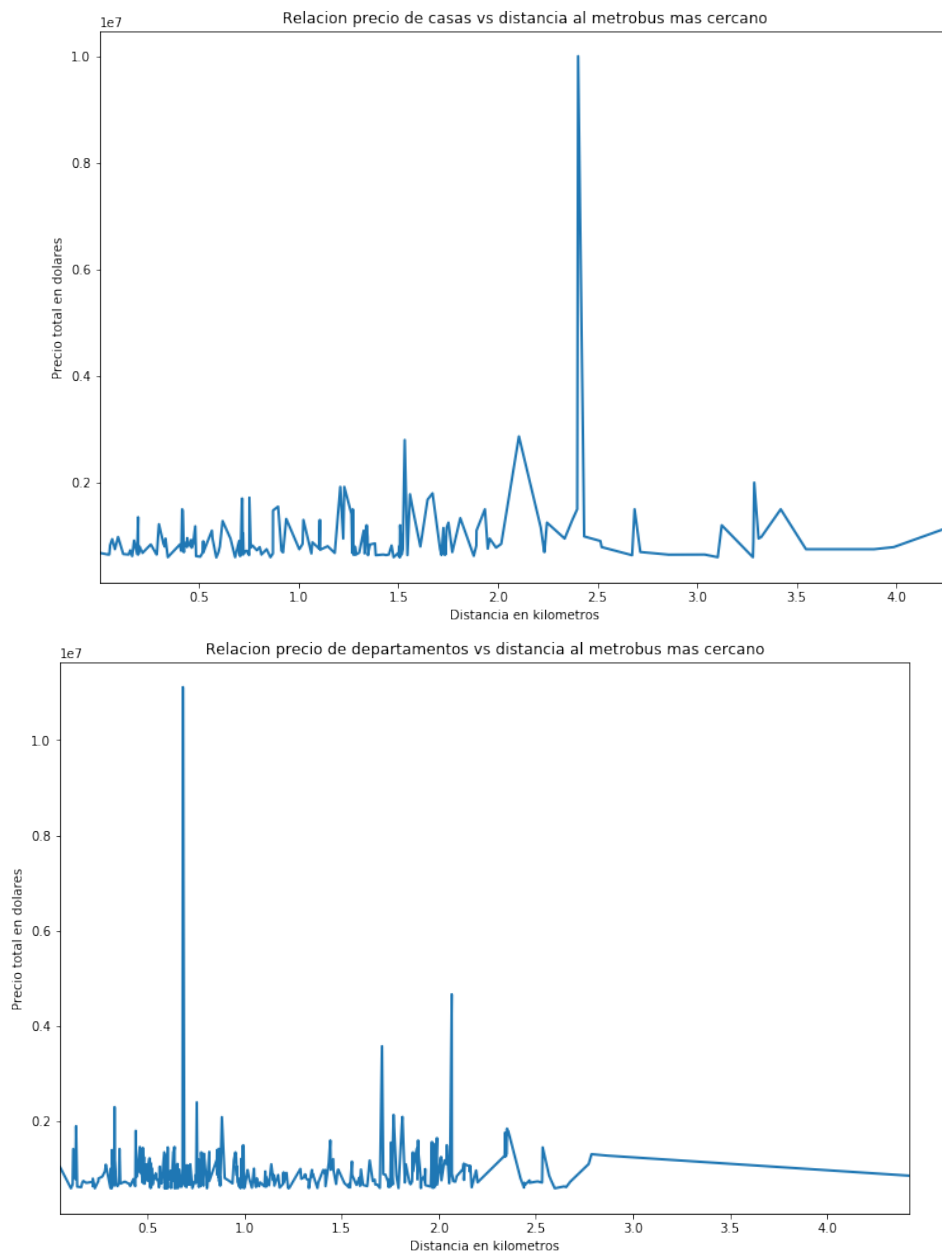
#### 4.0.2. Según distancia a una facultad pública

Igual que el caso anterior de las casas, no existe relación alguna entre la distancia que hay entre las propiedades a la facultad pública más próxima y el precio de las mismas.



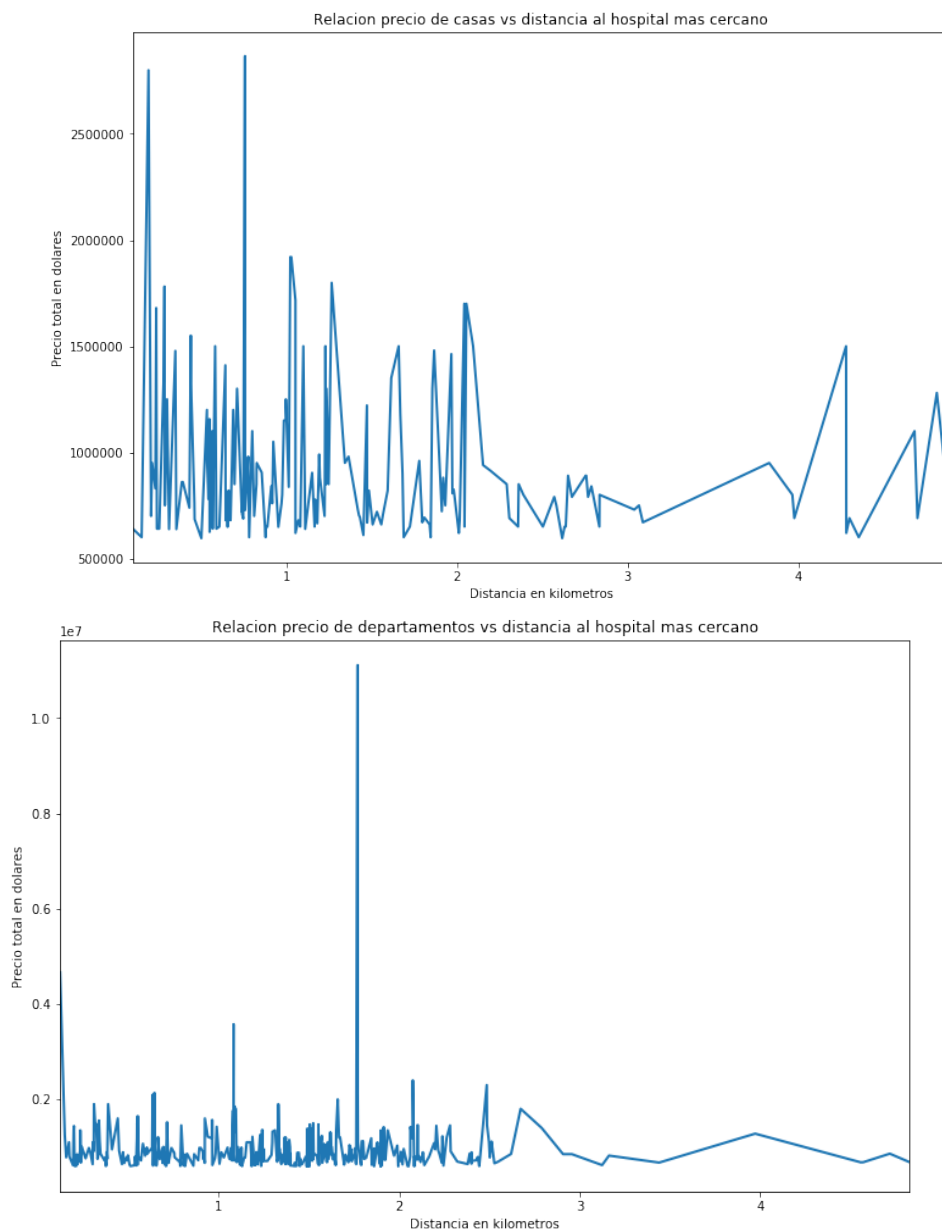
#### 4.0.3. Según distancia al metrobus

Igual que el caso anterior, con la excepción de que se puede observar que las distancias entre 2km a 2.5km de una casa al metrobus, muestran picos elevados, indicando precios muy grandes.



#### 4.0.4. Según distancia a un hospital

Los resultados implican que no existe relación alguna entre la distancia que hay entre las propiedades al hospital más cercano y el precio de las mismas, tanto para departamentos como para casas.



## 5. Precios según atributos de las propiedades

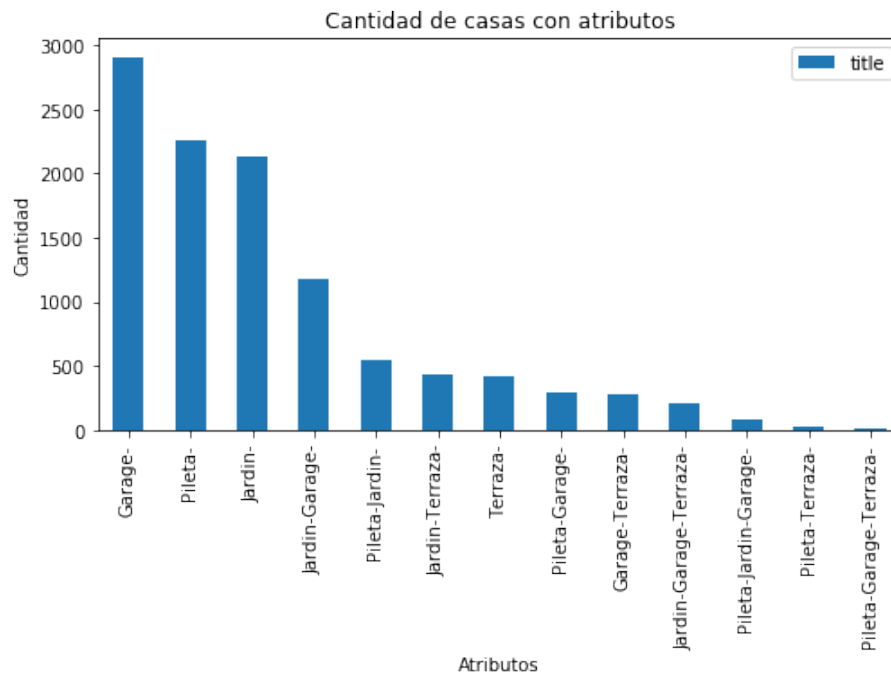
A partir de la descripción de cada propiedad dada por la columna 'title', filtramos a partir de las mismas datos sobre las propiedades, los cuales nos sirven para verificar si dichos datos influyen en los precios de las propiedades.

Como puede verse en la tabla anterior, podemos analizar los precios de las propiedades según tengan pileta, jardín, quincho, etc. Previo al análisis, descartamos todas las propiedades que no tengan una descrip-

Propiedad 1	Excelente 2 Ambientes Luminoso Con Pileta
Propiedad 2	Casa Con Jardín Y Pileta En Ituzaingó 4 Cdas
Propiedad 3	Duplex impecable con quincho y Pileta!

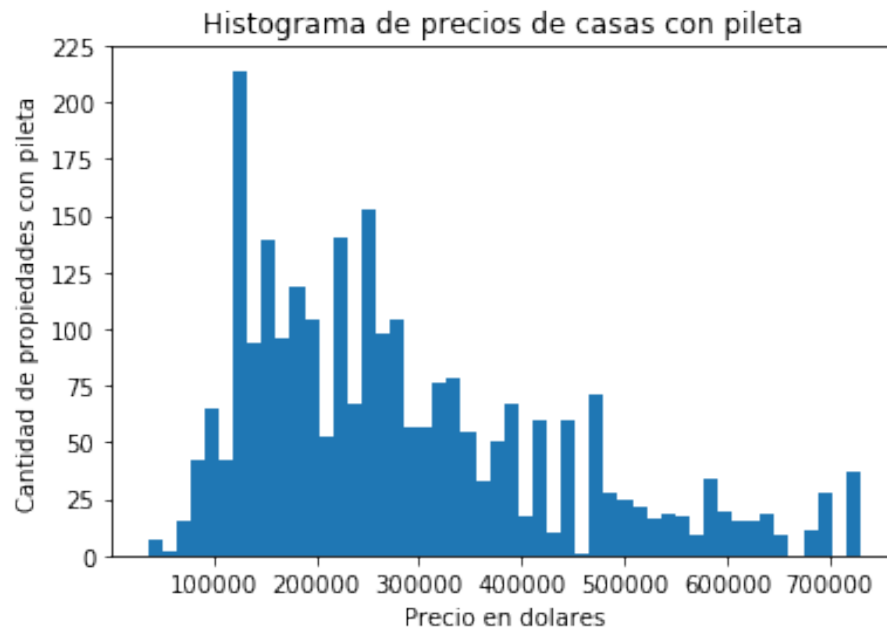
Cuadro 3: Titulos de propiedades

ción y nos quedamos con 543895 propiedades, de las cuales podemos calcular cuantas de estas poseen los datos que buscamos.



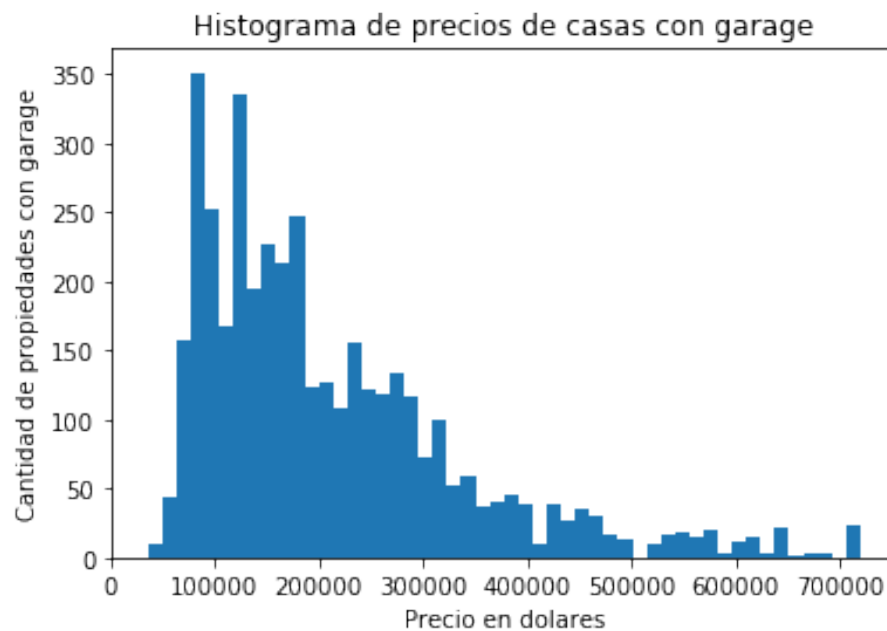
Como puede verse en la imagen anterior, de las 543895, muy pocas propiedades poseen los atributos que buscamos, procedemos a realizar un análisis sobre los tres atributos mayoritarios, los cuales son: pileta, garage y jardín.

### 5.0.1. Precio según tengan pileta



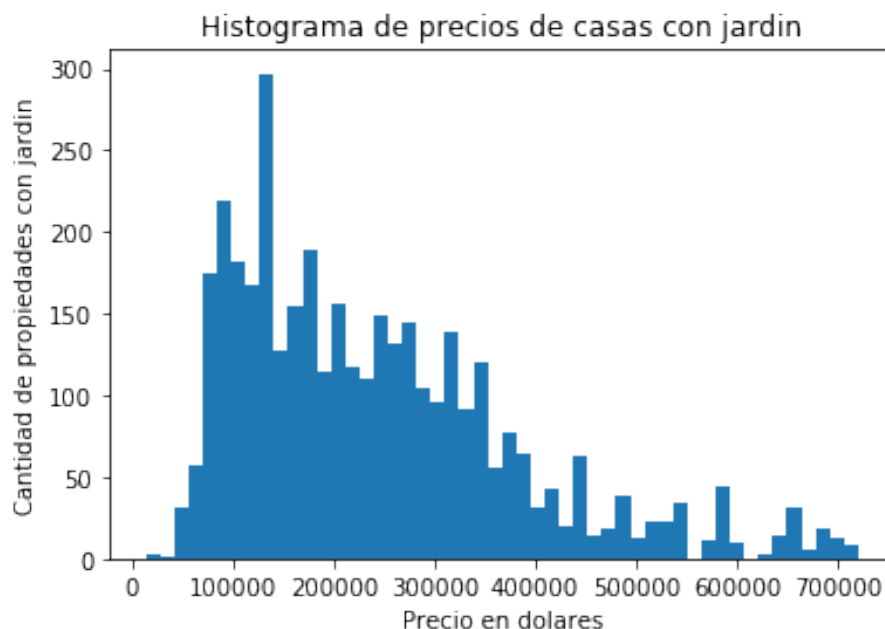
Puede verse que los precios de las casas no dependen en gran medida de si tienen o no pileta

### 5.0.2. Precio según tengan garage



Puede verse que los precios de las casas con garage, en promedio, son bastante bajos.

### 5.0.3. Precio según tengan jardín



Puede verse que los precios de las casas no dependen en gran medida de si tienen o no jardín.

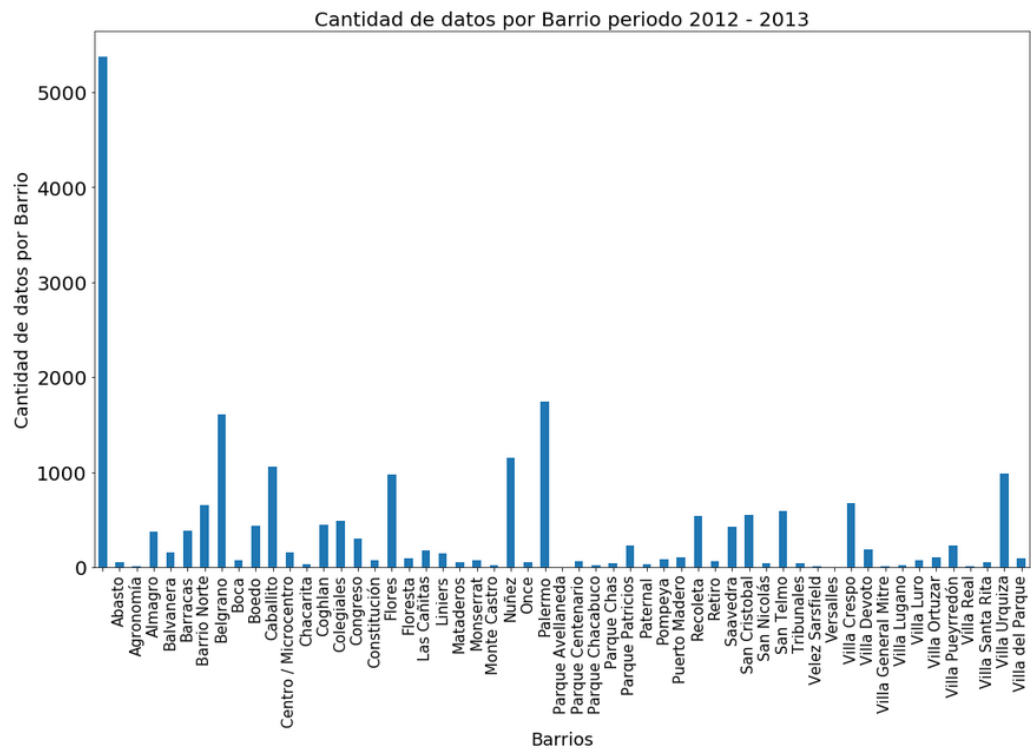
## 6. Incremento del precio por metro cuadrado de los Barrios de CABA

El set de datos abarca desde el 08-08-2012 hasta el 08-08-2017. Para el estudio de como se produjo la evolución de los precios del metro cuadrado promedio por barrio, se subdividieron los datos por periodos que datan desde el 08-08 de un año, hasta la misma fecha del año siguiente, por lo que obtuvimos exactamente 5 subdivisiones. Sobre estos periodos se estudio la evolución del precio.

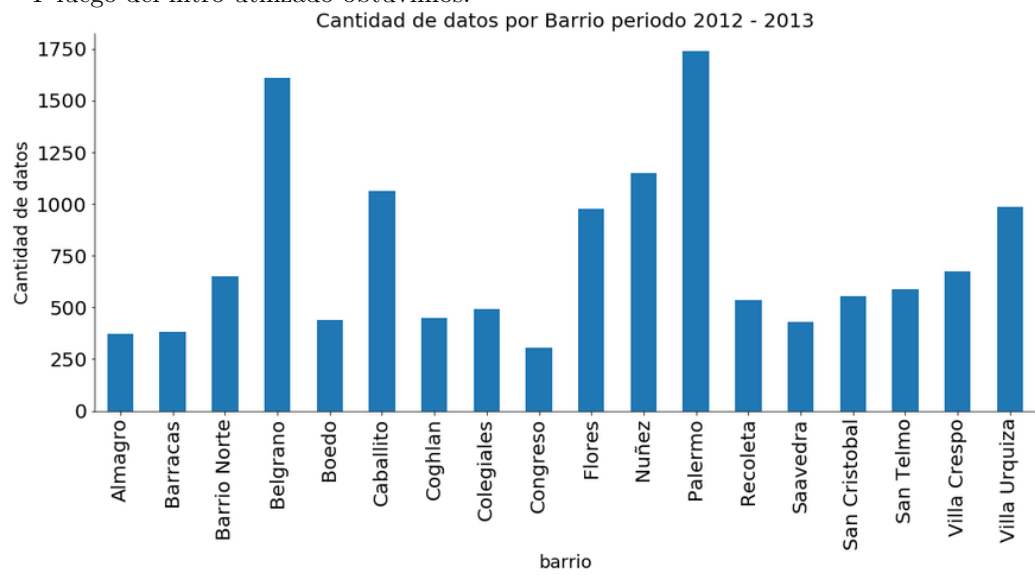
### 6.0.1. Metodología

Para lograr el fin deseado, primero se estudio la cantidad de datos por periodo que se tenia sobre los distintos barrios de la ciudad. Esto se hizo para prevenir aquellos barrios que tengan tan pocos datos que influyeran mucho sobre el estudio sin que sean muy representativos de la realidad.

Por ejemplo, para el periodo 2012 - 2013 se obtuvo lo siguiente:



Y luego del filtro utilizado obtuvimos:



Finalmente se aplica un filtro para descartar datos sin sentido. Por ejemplo, un precio por metro cuadrado igual a 0 o NaN y precios por metro cuadrado mayores a 10000 USD.

Y finalmente obtenemos el resultado de un precio por metro cuadrado promedio de cada barrio que entro en el criterio que se utilizó.



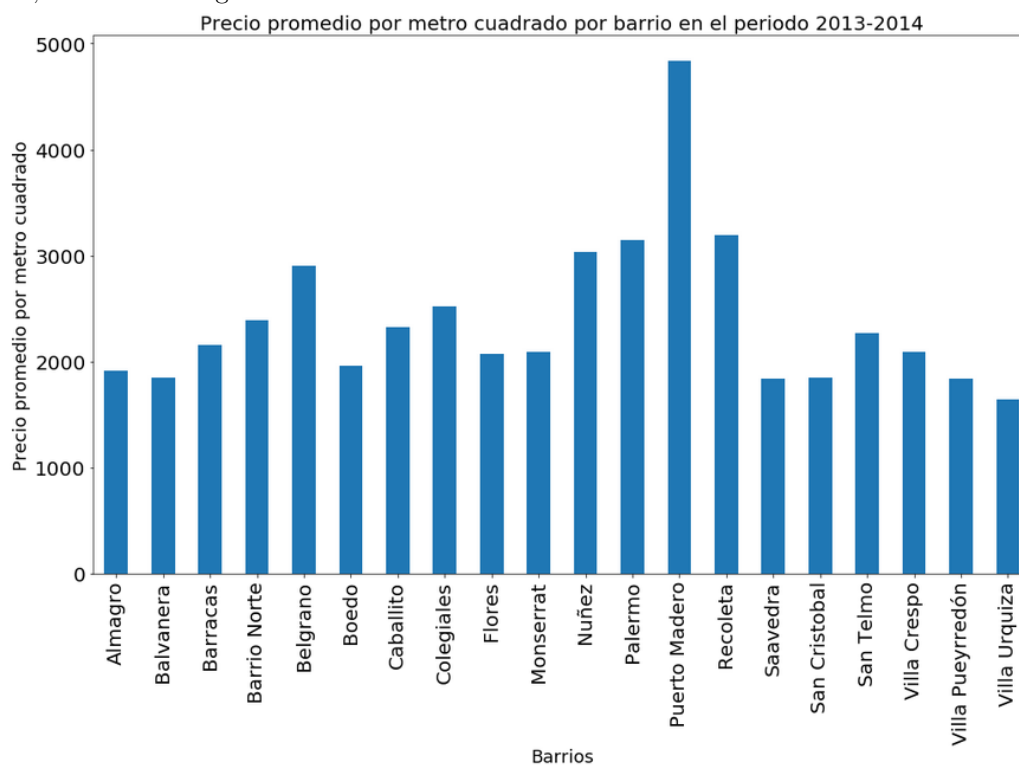
### 6.0.2. Periodo 2012 - 2013

El resultado final de todos los filtros aplicados a este periodo, fue de menos de 100 datos entre todos los barrios, por lo tanto se considero que este periodo no puede ser representativo para el calculo de precio por metro cuadrado promedio de los barrios.

Claramente el set de datos no tiene una buena cantidad de datos significativos en el periodo. Esto puede deberse a un gran error humano a la hora de escribir los datos, o también podría llegar a significar que el mercado de inmuebles estuvo prácticamente congelado y casi sin operaciones en el periodo.

### 6.0.3. Periodo 2013 - 2014

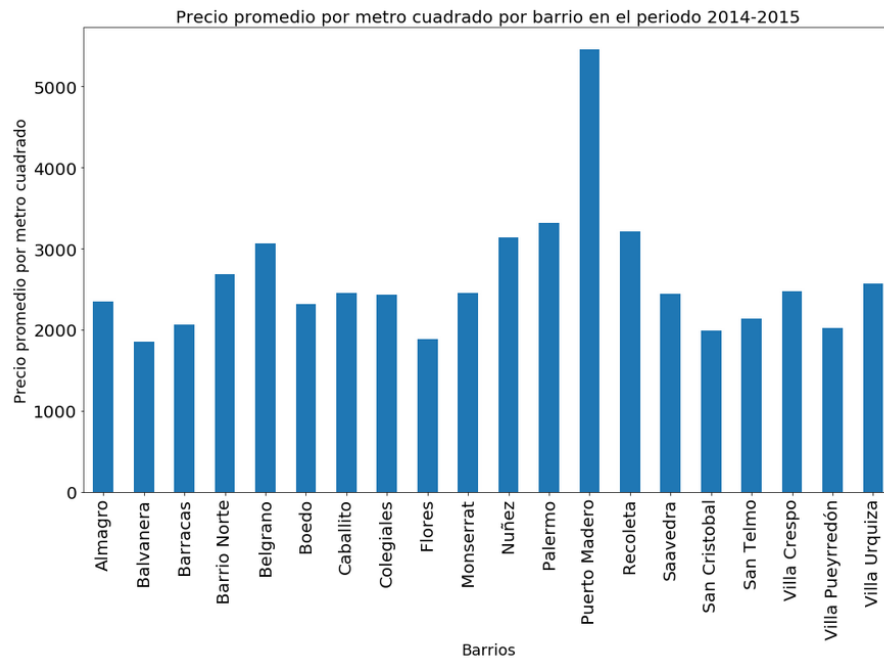
Para este periodo se obtuvieron mas de 381 datos totales luego de los filtros. Si bien no es una gran cantidad y puede influenciar fuertemente en los resultados, se obtuvo lo siguiente:



Como conclusión, los barrios con mayor precio por metro cuadrado son: Puerto Madero, Recoleta, Palermo, Belgrano y Nuñez. Este resultado, por conocimientos previos sobre la Ciudad, si bien pueden tener algunas diferencias con la realidad en el valor exacto en el periodo, por la poca cantidad de datos, se puede decir que tiene sentido el resultado.

#### 6.0.4. Periodo 2014 - 2015

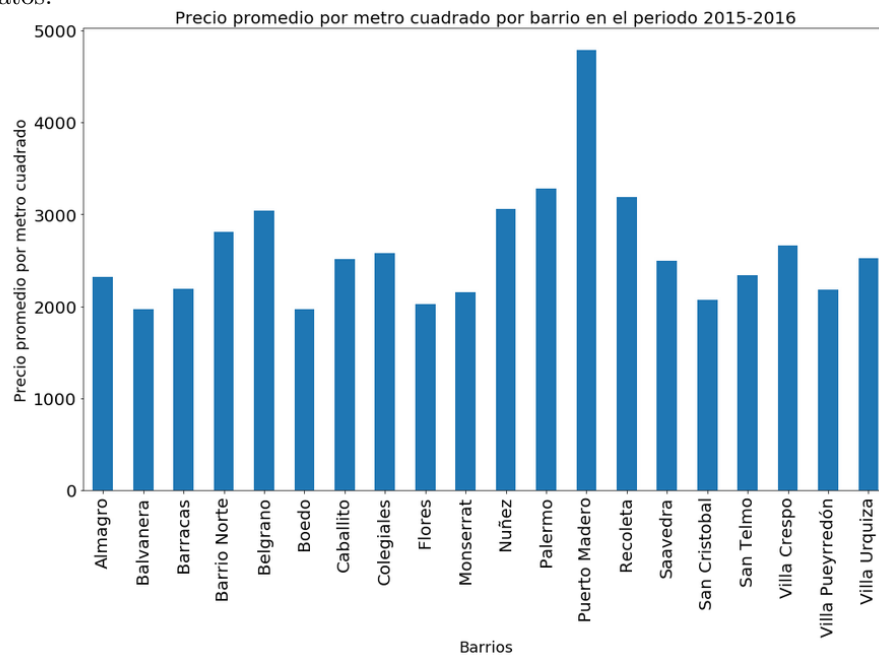
En este periodo se consiguieron un total de 1.190 datos confiables según el criterio tomado. El resultado obtenido sobre el set de datos fue el siguiente:



Se puede observar con estos datos obtenidos una diferencia de precios con el periodo anterior. Además este resultado resulta as confiable que el anterior por la cantidad de datos obtenida.

### 6.0.5. Periodo 2015 - 2016

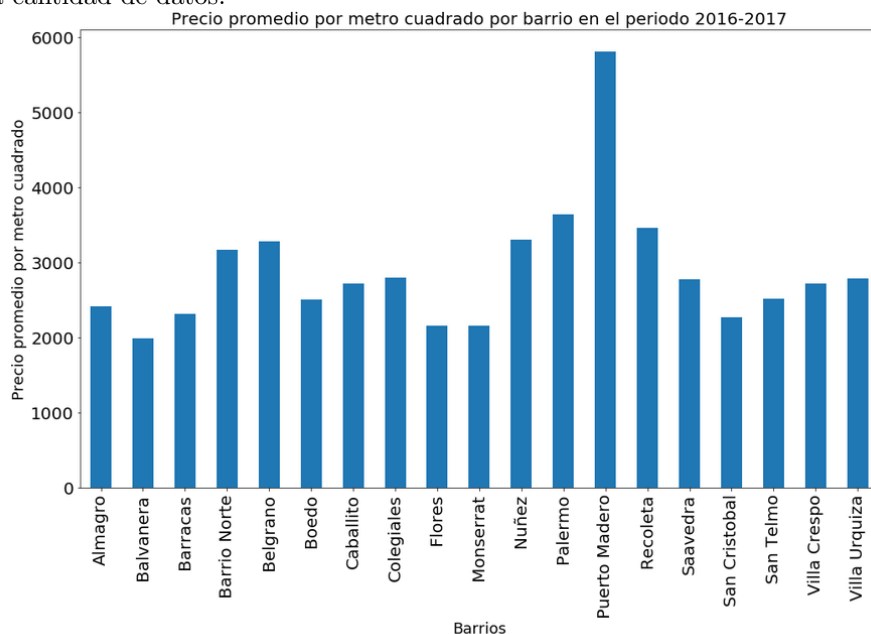
Para este periodo obtuvimos un total de 10.723 datos confiables. Cada vez mas podemos confiar en los datos obtenidos en el resultado por la densidad de datos.



Se obtuvieron diferencias en los valores correspondientes al periodo pasado pero los barrios mas caros se mantuvieron en el tiempo.

### 6.0.6. Periodo 2016 - 2017

Finalmente, para el ultimo periodo se obtuvo 30.573 datos confiables. Este último estudio resulta ser el mas confiable de todos por el gran incremento en la cantidad de datos.



Claramente se observa entre los periodos mas confiables obtenidos, es decir este y el anterior, que hubo un incremento en los precios por metro cuadrado de los inmuebles.

### 6.0.7. Conclusiones

Por los resultados obtenidos, pudimos llegar a las siguientes conclusiones:

- El constante incremento de la densidad de datos entre los distintos periodos nos lleva a la conclusión de que el mercado de inmuebles tuvo un gran crecimiento entre los años 2013 - 2017, por la cantidad de operaciones observada. Y este crecimiento resulto casi exponencial a lo largo de los periodos estudiados, sobre todo desde el 2015 al 2017.
- Los barrios mas caros de la ciudad se mantuvieron constantes en el tiempo a pesar de la diferencia en la cantidad de datos, lo que nos lleva a la conclusión de que, si bien algunos valores pueden no ser tan precisos por la densidad de datos, el ranking de barrios mas caros si coincide con la realidad.

## 7. Porcentaje de evolución de precios entre los distintos periodos

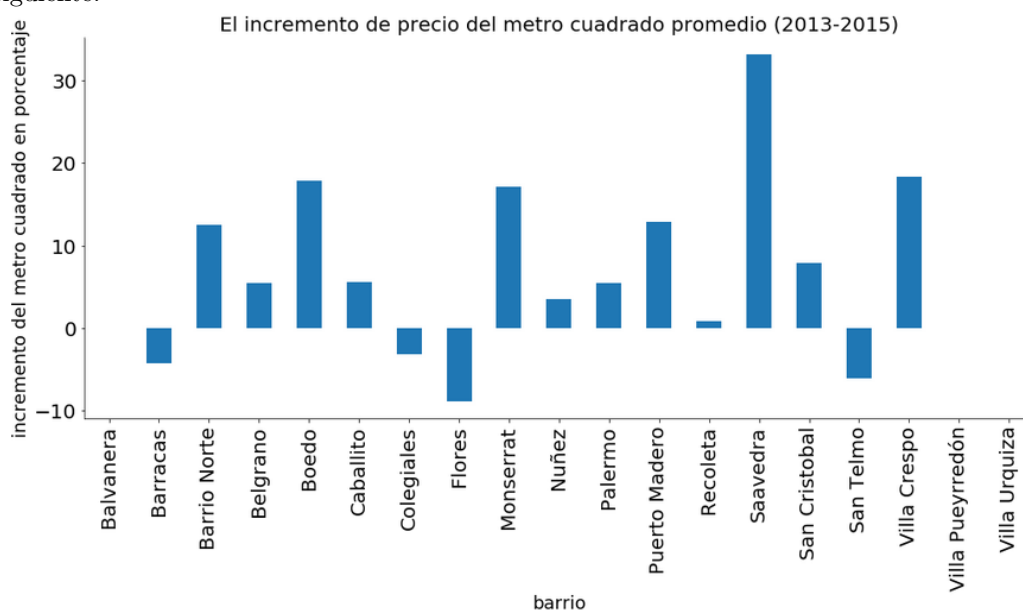
Con los resultados obtenidos en la sección anterior, se procedió a realizar el estudio de como evolucionaron estos precios por metro cuadrado de los barrios

periodo a periodo.

Para este fin se calculo el porcentaje en que incrementaron o decrementaron los precios respecto del periodo anterior.

### 7.0.1. Evolución de precios del 2013 al 2015

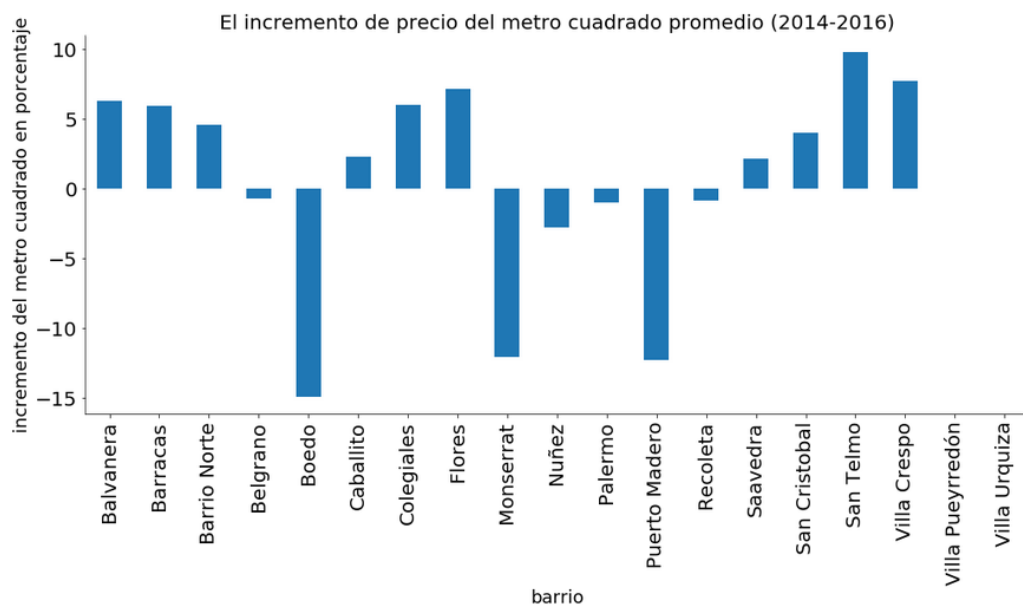
Hay que tener presente que los primeros periodos tienen pocos valores y pueden discrepar con la realidad del momento, pero el resultado obtenido fue el siguiente:



Se puede observar que el resultado dio que el incremento mayor lo percibió el barrio de Saavedra. Esto puede significar un aumento en la demanda de inmuebles en el barrio muy grande por parte de la población, lo que lleva al encarecimiento de los inmuebles. Lo mismo sucedió en otros barrios, pero en otros llegó a disminuir el precio.

### 7.0.2. Evolución de precios del 2014 al 2016

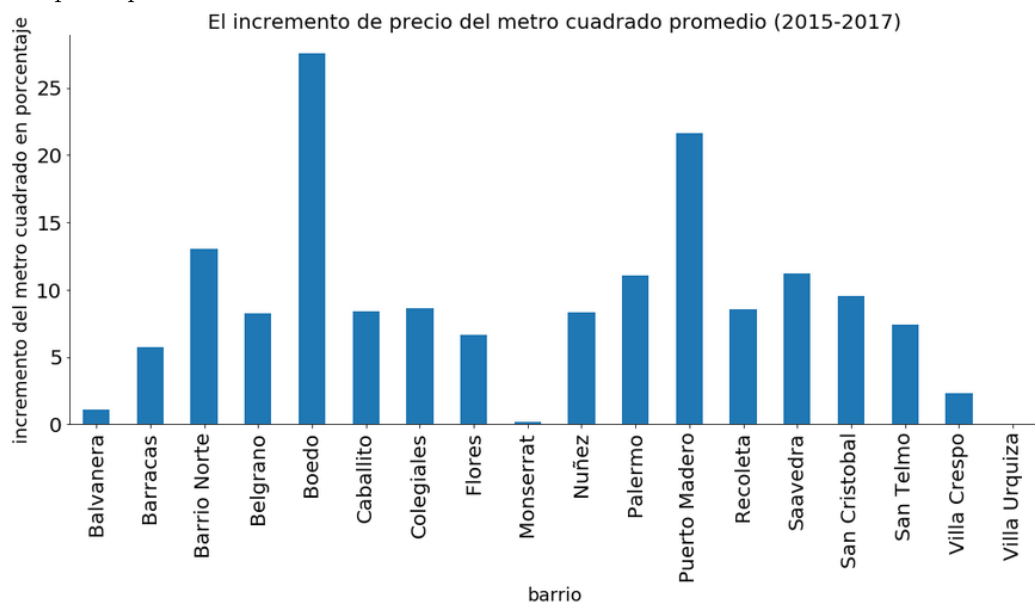
En este periodo ya se obtuvo una mayor cantidad de datos para hacer el análisis, pero no una gran cantidad de datos. Hay que tener presente que el resultado es un porcentaje del crecimiento del periodo 2014-2015 con el 2015-2016, y el resultado obtenido fue el siguiente:



Hubo una gran depresión en los precios en muchos barrios, lo que puede significar que la densidad de datos este influenciando en el resultado de manera significativa.

### 7.0.3. Evolución de precios del 2015 al 2017

Este estudio es el que tiene mayor cantidad de datos totales, y por lo tanto es el que se puede considerar el mas confiable de todos.

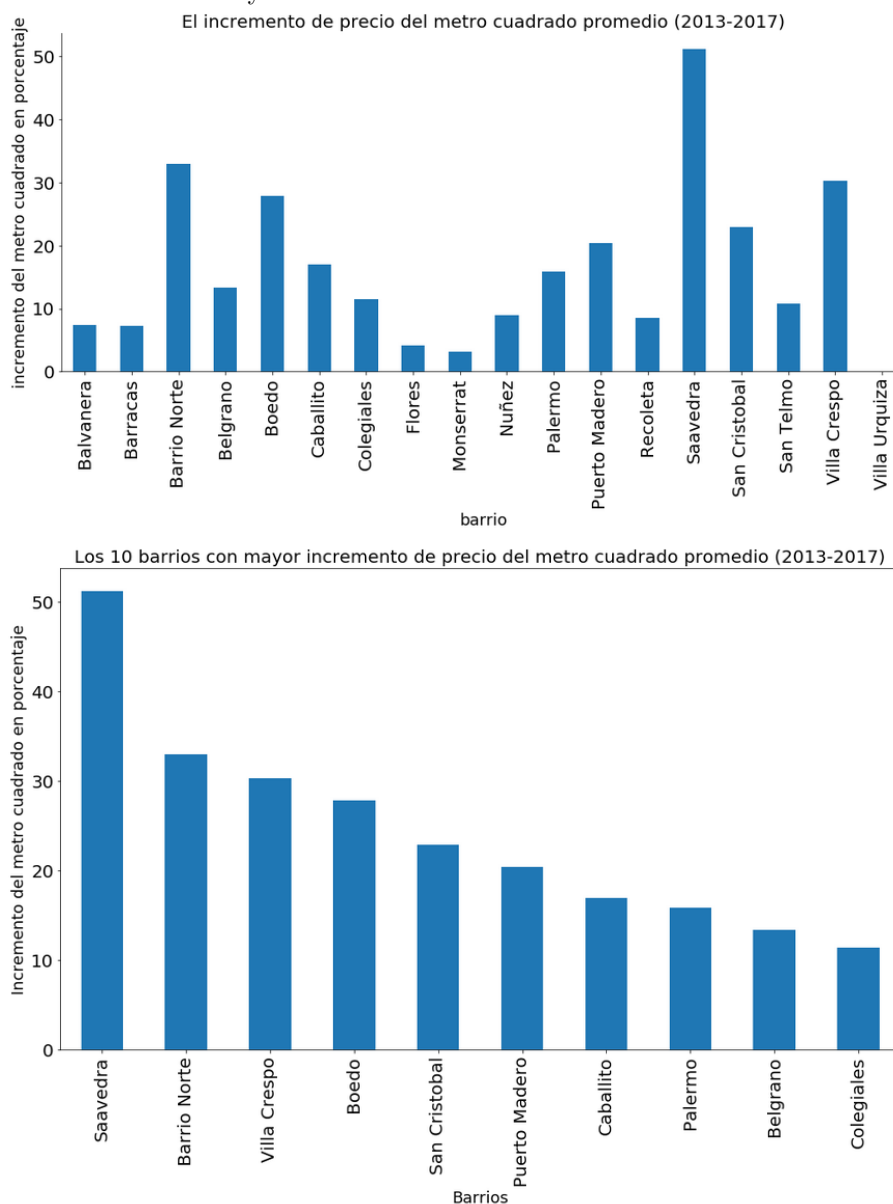


Finalmente, entre los periodos con mayor densidad de datos, observamos que el precio por metro cuadrado promedio aumento significativamente en todos los

barrios estudiados de la ciudad

#### 7.0.4. Evolución de precios del 2013 al 2017

Si bien la diferencia de densidad de datos es muy significativa, este estudio nos puede dar una idea de que tan rentable resultó ser la inversión en un inmueble entre el 2013 y el 2017.



Y los resultados indican que el hubo un incremento de los precios en todos los barrios estudiados, en general. Esto se puede deber al factor del incremento de demanda de inmuebles en distintas zonas de la ciudad. Y esto nos permite ver también donde es que la población esta eligiendo vivir dentro de la Ciudad Autónoma de Buenos Aires