



Organización de Datos

Segundo Cuatrimestre 2017

Trabajo Práctico 1

Integrante	Padrón	Correo electrónico
Rodrigo De Rosa	97799	rodrigoderosa@outlook.com
Marcos Schapira	97934	schapiramarcos@gmail.com
Facundo Guerrero	97981	facundoiguerrero@gmail.com

Índice

I	Análisis del precio por m^2	1
1.	Adaptación del DataFrame	1
1.1.	Filtrado de columnas	1
1.2.	Completando el DataFrame	1
2.	Estudio estadístico de los datos	1
2.1.	Análisis de la distribución de precios	1
2.2.	Agrupando por barrios	3
3.	Analizando grupos característicos	5
3.1.	Los diez barrios con mayor precio por m^2	5
3.1.1.	Unificación de Palermo	5
3.1.2.	División de Palermo	6
3.1.3.	Comentario sobre el Top 10	7
3.2.	Los diez barrios con menor precio por m^2	7
3.3.	Dividiendo en secciones	8
3.3.1.	Grupo 1 - $[2500 : \infty)$ USD	9
3.3.2.	Grupo 2 - $[2000 : 2500)$ USD	10
3.3.3.	Grupo 3 - $[1500 : 2000)$ USD	11
3.3.4.	Grupo 4 - $[1200 : 1500)$ USD	13
3.3.5.	Grupo 5 - $[950 : 1200)$ USD	14
3.3.6.	Grupo 6 - $[450 : 950)$ USD	14
4.	Distribución geográfica	15
4.1.	Grupos característicos y su ubicación	16
4.1.1.	Grupo 1	16
4.1.2.	Grupo 2	16
4.1.3.	Grupo 3	17
4.1.4.	Grupo 4	17
4.1.5.	Grupo 5	18
4.1.6.	Grupo 6	18
4.1.7.	Comparación de grupos	19
5.	Progresión del precio por m^2 a través de los años	19
5.1.	Fluctuación del precio del m^2 a través de los años	19
5.2.	Año 2013	20
5.3.	Año 2014	20
5.4.	Año 2015	21
5.5.	Año 2016	22
5.6.	Año 2017	22
5.7.	Comparación de la progresión mensual en cada año	23
6.	Análisis de precio aproximado.	24
6.1.	Objetivo	24
6.2.	Preparación y procesamiento de los datos	24
6.3.	Presentación de los gráficos de promedios	24
6.4.	Ubicación de las propiedades con mayor precio	26
6.5.	Conclusiones de la Fluctuación del precio aproximado.	29
7.	Conclusiones generales	30

Parte I

Análisis del precio por m^2

1. Adaptación del DataFrame

Para el análisis particular de cada característica de la información que se posee, se adaptó el DataFrame original para poder analizar dicha información mas fácil y comodamente.

1.1. Filtrado de columnas

Para el análisis de esta cierta característica de las propiedades, consideramos *importantes* sólo a algunas celdas. Estas son:

- `place name` \leftarrow `location`
- `price aprox usd` \leftarrow `price`
- `surface total in m2` \leftarrow `totalSurface`
- `surface covered in m2` \leftarrow `coveredSurface`
- `price usd per m2` \leftarrow `pricem2`

1.2. Completando el DataFrame

Lo primero que se hizo para realizar este analisis fue completar las columnas faltantes de la mayor cantidad de entradas posibles. Esto es, `location`, `price`, `totalSurface`, `pricem2`. De esta forma, nos permitimos analizar una mayor cantidad de propiedades para realizar un analisis un poco mas correcto.

Para completar el campo de precio por m^2 se necesita que la entrada sobre la que se trabaja cumpla la siguiente condición lógica: $pricem^2 \vee (price \wedge surface)$. Es decir, necesita tener o el precio por metro cuadrado o tanto el precio total como la superficie total.

Si el campo `pricem2` tiene valor, entonces ese será el utilizado. En caso contrario, si tanto el campo `price` como el campo `totalSurface` tienen valor, definimos como nuestro nuevo `pricem2` a la división $\frac{price}{surface}$.

Para esto, necesitamos unificar `coveredSurface` y `totalSurface`, para maximizar nuevamente la cantidad de entradas disponibles. Esto se hace, simplemente, poniendo como `totalSurface` el valor de `coveredSurface` en aquellas entradas donde la primera no tenga valor (consideramos que $total - covered = uncovered$).

Una vez completados todos los `pricem2` posibles, eliminamos todas aquellas entradas que tengan `NaN` como valor (en cualquiera de las celdas que definimos como *importantes*), pues ya no podemos obtener el valor de esa celda de ningun otro lugar.

2. Estudio estadístico de los datos

2.1. Analisis de la distribución de precios

Una vez completado el DataFrame lo mas posible, se realizó un análisis de la distribución de precios. Con esto nos referimos a analizar la variación del precio por metro cuadrado entre todas las propiedades. Es decir, *limpiar los datos que no tienen sentido*.

Para esto le pedimos el `.describe()` a nuestro DataFrame con los percentiles 0,01 y 0,99. Esto nos permite analizar que tan desviados estan los valores máximos y mínimos.

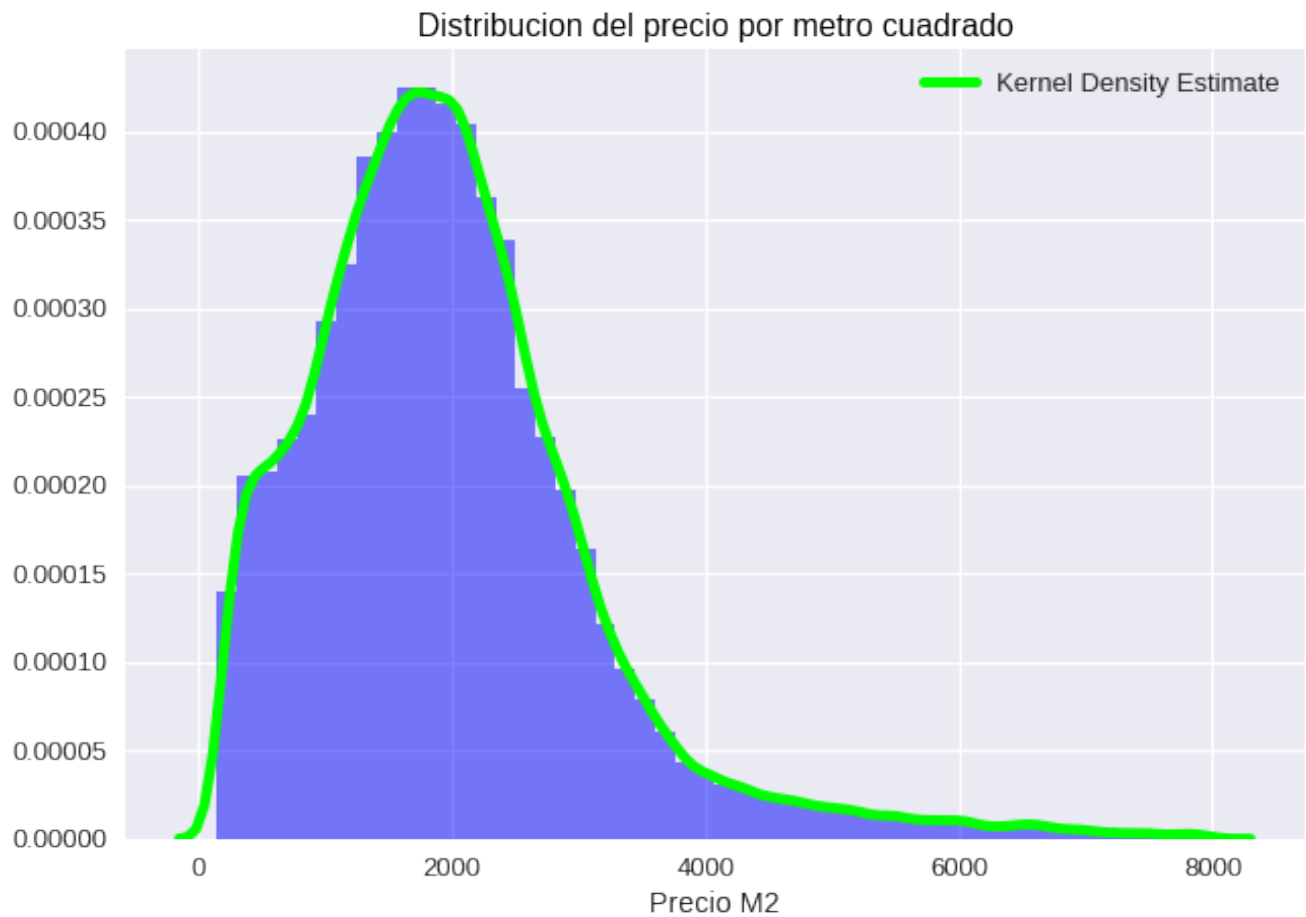
Con los percentiles recién mencionados hacemos un recorte de los datos para lograr una distribución que se asemeje a una Normal lo mas posible. El primer recorte es tanto inferior ($> 150USD$) como superior ($< 18000USD$). Como en este nuevo DataFrame la diferencia entre el percentil 0,99 y el máximo es de más del doble, se vuelve a recortar superiormente ($< 8000USD$).

Luego de esto, la distribución de precios es un poco mejor que antes (las diferencias entre los percentiles 0,25, 0,5, 0,75 son similares).

A continuación se muestra un gráfico de distribución de precios por metro cuadrado que se obtiene del DataFrame original sin realizar el filtrado recién mencionado. Nos hubiese gustado poder mostrar tanto el KDE como el histograma pero al haber tanta diferencia entre el maximo y los valores principales de la distribución, el histograma era solo una linea. El objetivo de este gráfico es hacer incapié en lo mencionado en el previo párrafo: es necesario filtrar los datos para tener un conjunto de datos con sentido.



En el siguiente gráfico de distribución de precios por metro cuadrado se puede ver que la mayor parte de las propiedades están concentradas en el rango de precios [150;4000]USD y luego hay un drástico decaimiento de cantidad de propiedades para el resto de los precios. Si bien se podría considerar que un recorte sería correcto, a partir de fuentes externas se sabe que ciertos barrios (*i.e.* Puerto Madero) tienen, aproximadamente, un valor medio de 6000USD por metro cuadrado.

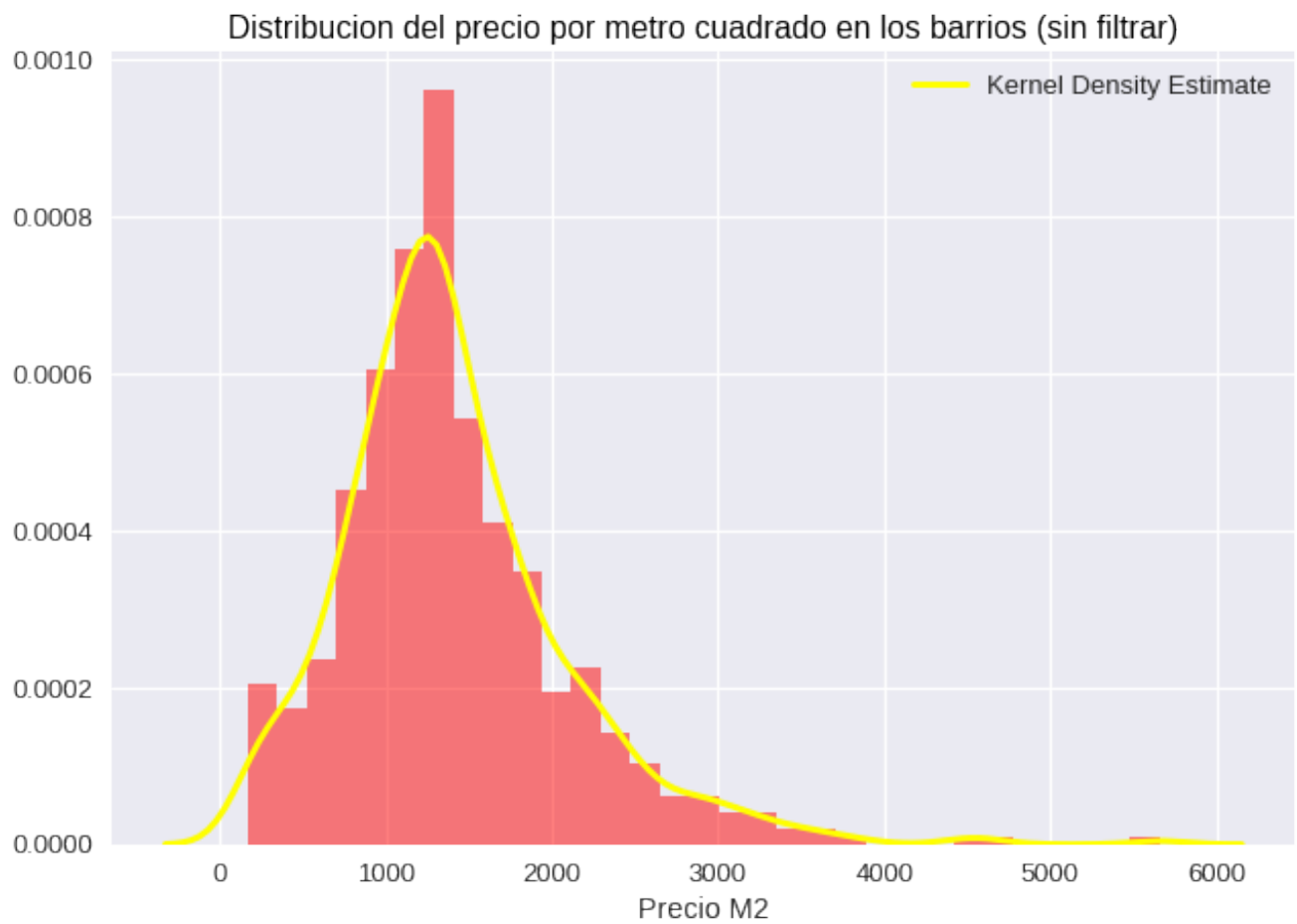


2.2. Agrupando por barrios

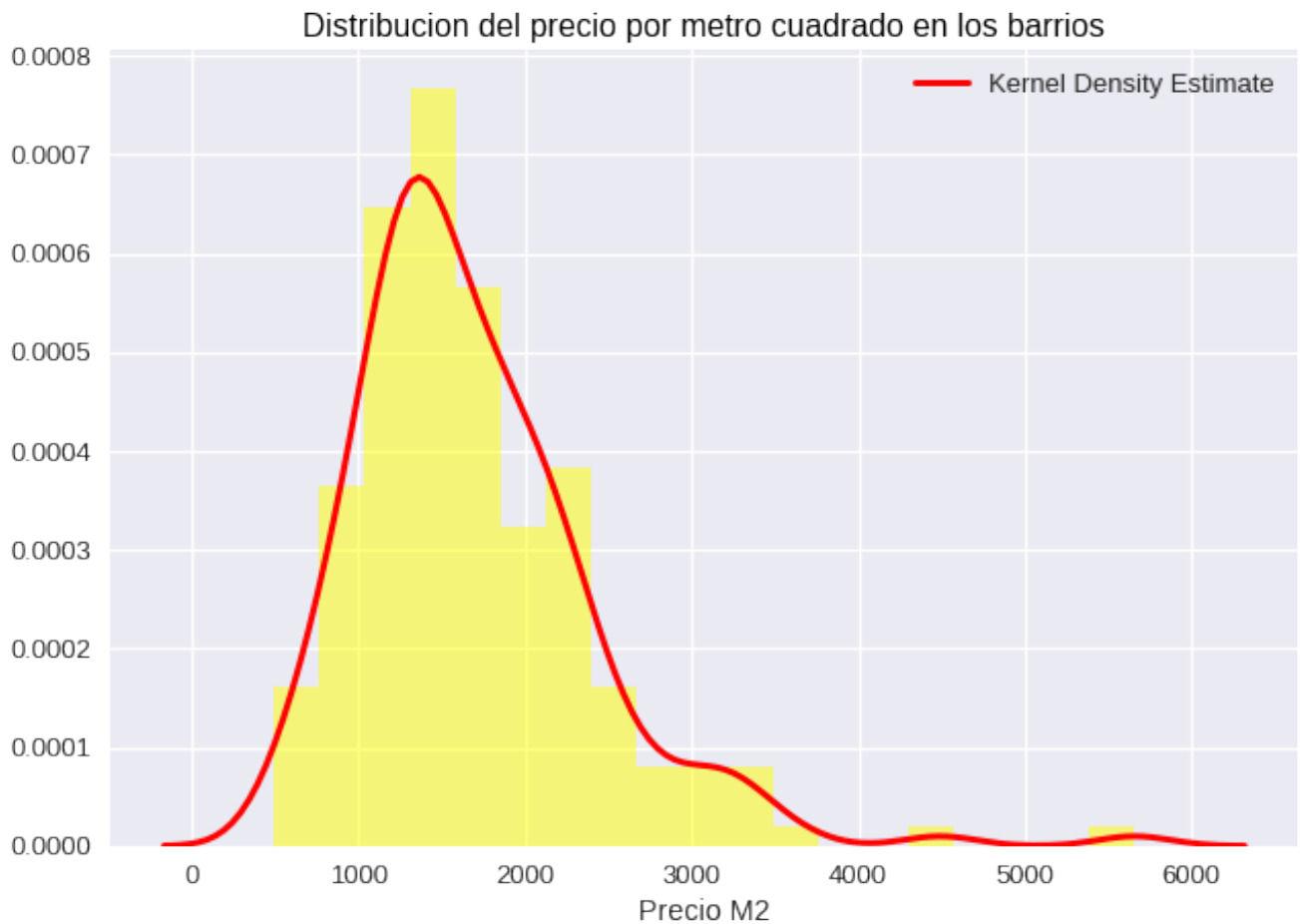
Ahora que nuestros datos están tan completos y retocados como queríamos, procedemos a agrupar todas las propiedades de acuerdo al barrio al que pertenecen. Una vez que los tenemos agrupados, debemos establecer un *minimo de propiedades* por barrio. Pues un barrio que tiene una o dos propiedades podría alterar el estudio de la informacion.

Nuevamente, para esta tarea utilizamos `.describe()` y resolvemos que utilizaremos como cota inferior 50 propiedades (dos mas que el equivalente a una publicación por mes en los últimos cuatro años).

Aquí, al igual que hicimos antes, mostraremos la distribución antes y después del filtro aplicado. Si bien en escencia no son tan diferentes, podemos observar que desaparecen algunos barrios de la zona de precios altos.



Una vez que eliminamos los barrios problemáticos, si analizamos la distribución de precio por barrio podemos ver que la mayor parte está concentrada en el intervalo [500; 3500]USD, mientras que muy pocos (solo tres) superan ese valor.



Podemos ver, además, que la distribución es bastante similar a la anterior (sin agrupar por barrios) aunque, obviamente, con valores menores (pues son promedios).

3. Analizando grupos característicos

En esta sección analizaremos ciertos grupos característicos a partir de la información con la que estamos trabajando.

3.1. Los diez barrios con mayor precio por m^2

Dado que ya estamos felices con la forma en que tenemos dispuestos los datos, comenzaremos por hacer un *Top 10* de los barrios más caros de CABA y GBA.

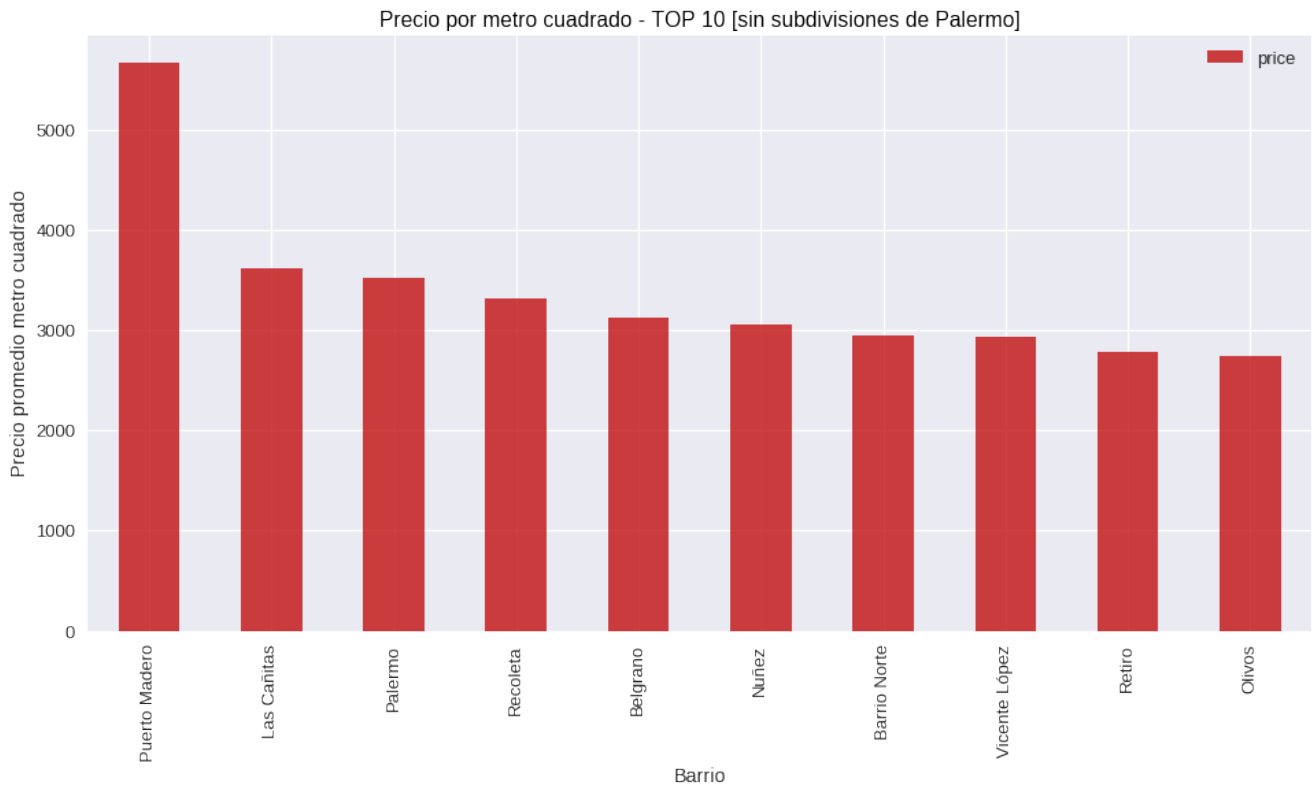
Para esto, como ya tenemos los datos agrupados, simplemente ordenamos el DataFrame y nos quedamos con los primeros diez. Durante el análisis de esta información, notamos que varios de los barrios que aparecían en este *Top 10* eran subdivisiones del barrio de Palermo. Por esta razón, decidimos incluir dos casos: uno en que consideramos que todos los 'Palermos' son uno solo, y otro en que cada uno es considerado un barrio diferente.

3.1.1. Unificación de Palermo

En este caso, consideramos que todas las subdivisiones de Palermo pertenecen a un sólo barrio. El resultado obtenido es el siguiente:

Top 10 [Palermo unificado]		
Puesto	Barrio	Precio m^2 [U\$D]
1	Puerto Madero	5657
2	Las Cañitas	3612
3	Palermo	3518
4	Recoleta	3316
5	Belgrano	3124
6	Nuñez	3056
7	Barrio Norte	2949
8	Vicente López	2925
9	Retiro	2783
10	Olivos	2737

En la tabla se observa que Puerto Madero tiene un valor mucho mas alto que el resto, de hecho, es mayor al doble del precio del décimo. De todos modos, entre el segundo y el último la variación es más suave. Para aportar a este análisis, se realiza un gráfico de barras:



3.1.2. División de Palermo

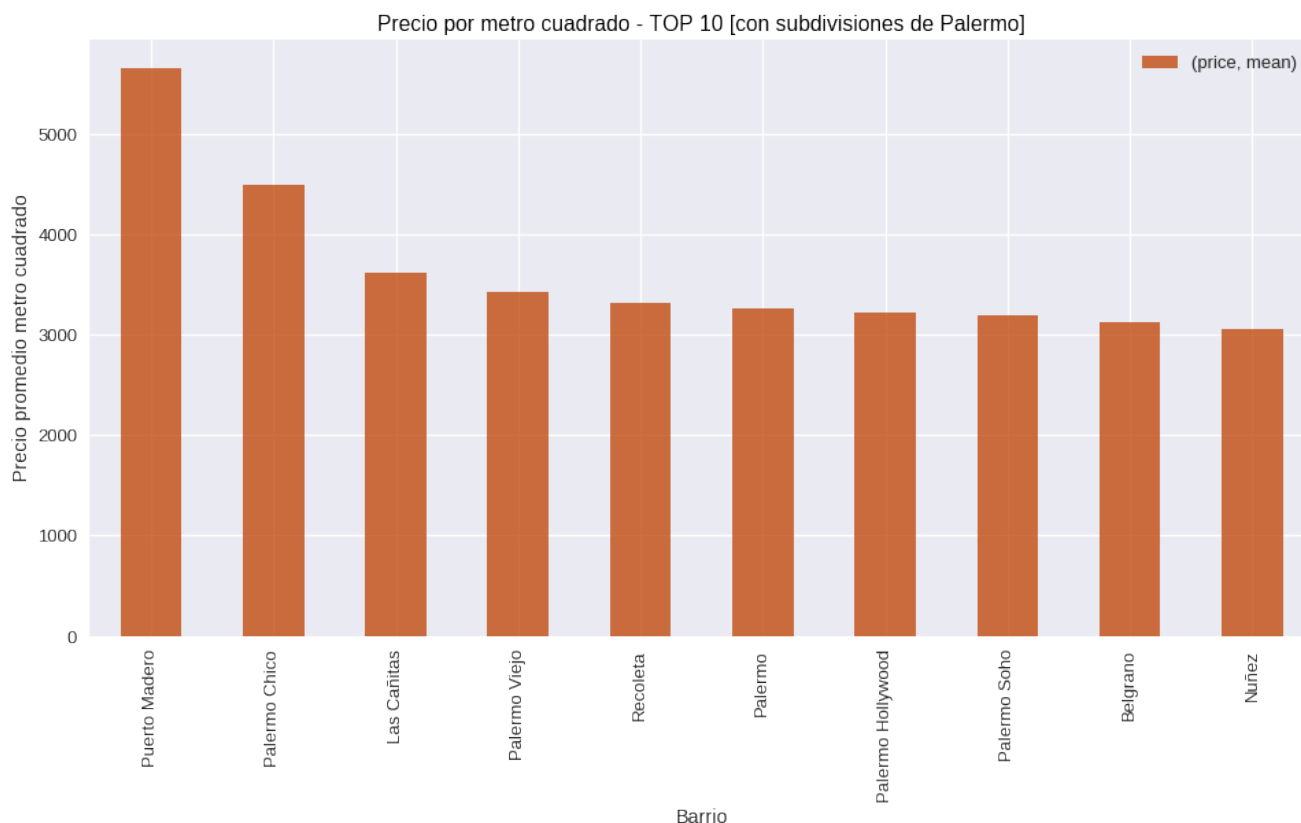
Aquí consideraremos que el barrio al que se nombra Palermo corresponde a todas las secciones de dicho barrio que no son las que ya aparecen en otro grupo.

En este caso, el resultado obtenido es:

Top 10 [Palermo dividido]		
Puesto	Barrio	Precio m^2 [U\$D]
1	Puerto Madero	5657
2	Palermo Chico	4489
3	Las Cañitas	3612
4	Palermo Viejo	3419
5	Recoleta	3316
6	Palermo	3260
7	Palermo Hollywood	3224
8	Palermo Soho	3198
9	Belgrano	3124
10	Nuñez	3056

En la tabla podemos ver que, si bien es correcto y es un *Top 10*, esta plagado de subdivisiones de Palermo y no nos permite tener un plano más general.

Aquí el gráfico de barras es muy similar aunque aparece Palermo Chico, que se acerca un poco mas al valor de Puerto Madero. De todos modos, la diferencia entre el primero y el segundo es muy grande como también lo es entre el segundo y el tercero, dejando la relación entre los valores igual de 'no suave'.



De aquí en más, utilizaremos a Palermo como un barrio unificado.

3.1.3. Comentario sobre el Top 10

Este *Top 10* arroja los resultados que se hubieran esperado, pues los únicos dos valores que no pertenecen a CABA corresponden a los primeros dos barrios de GBA en los que se piensa al pensar en los barrios mas caros de Buenos Aires.

Por otro lado, si nos sorprende el hecho de que el m^2 en Barrio Norte sea más barato que Núñez o en Belgrano.

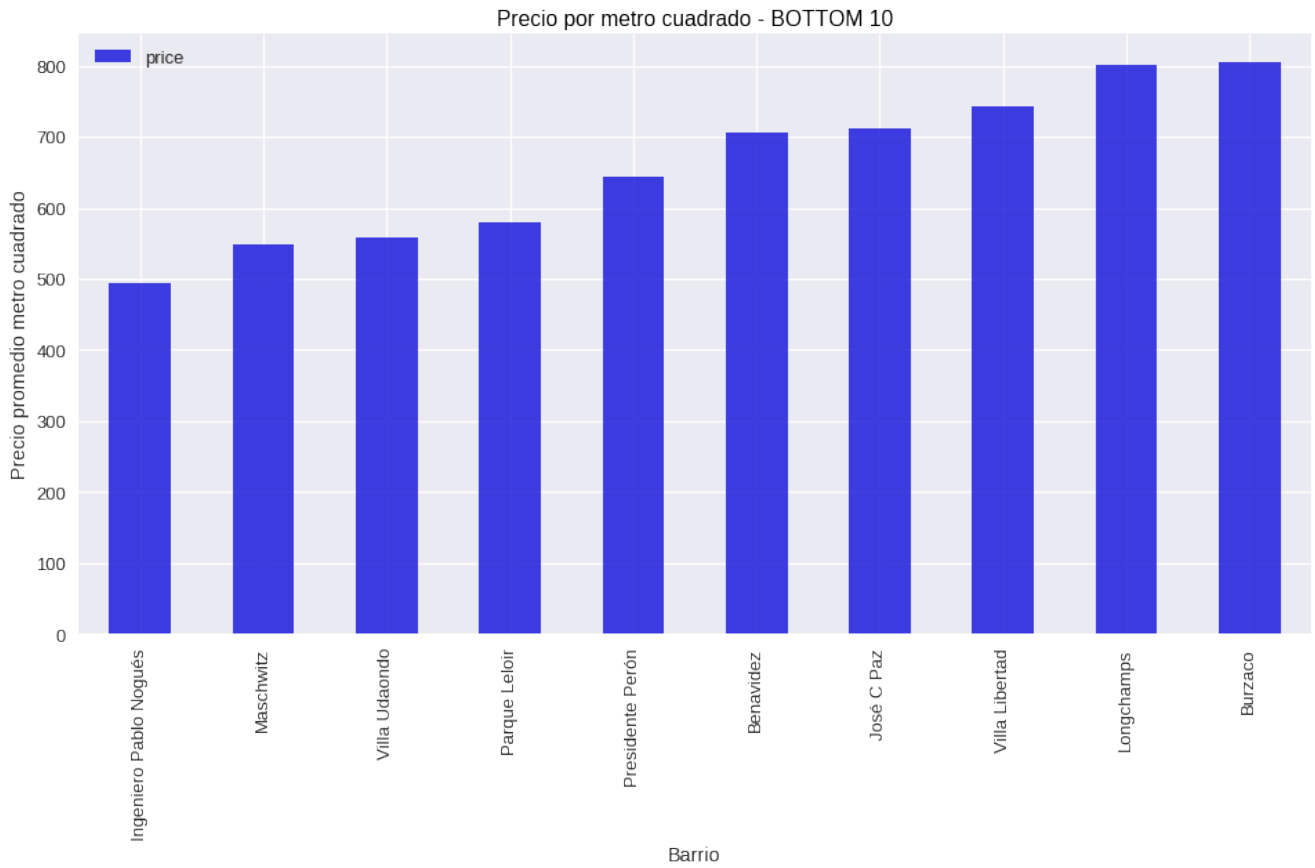
3.2. Los diez barrios con menor precio por m^2

Para esta parte, al igual que antes, ordenamos los datos para analizar cuales son los diez barrios que se encuentran en el *Bottom 10*.

Al realizar este análisis, lo obtenido es:

Bottom 10		
Puesto	Barrio	Precio m^2 [USD]
1	Ingeniero Pablo Nogués	494
2	Maschwitz	548
3	Villa Udaondo	558
4	Parque Leloir	579
5	Presidente Perón	643
6	Benavidez	705
7	José C Paz	711
8	Villa Libertad	743
9	Longchamps	800
10	Burzaco	804

Si graficamos estos valores al igual que antes podremos ver un ascenso (o descenso) más suave que el del *Top 10*. Si bien el primero es casi la mitad de el último, la variación entre puestos es menor.



Aquí, remitiéndonos a la sección 3.1.3, vemos que los barrios del *Bottom 10* son todos barrios alejados de la ciudad, de los cuales es esperable un bajo valor del m^2 .

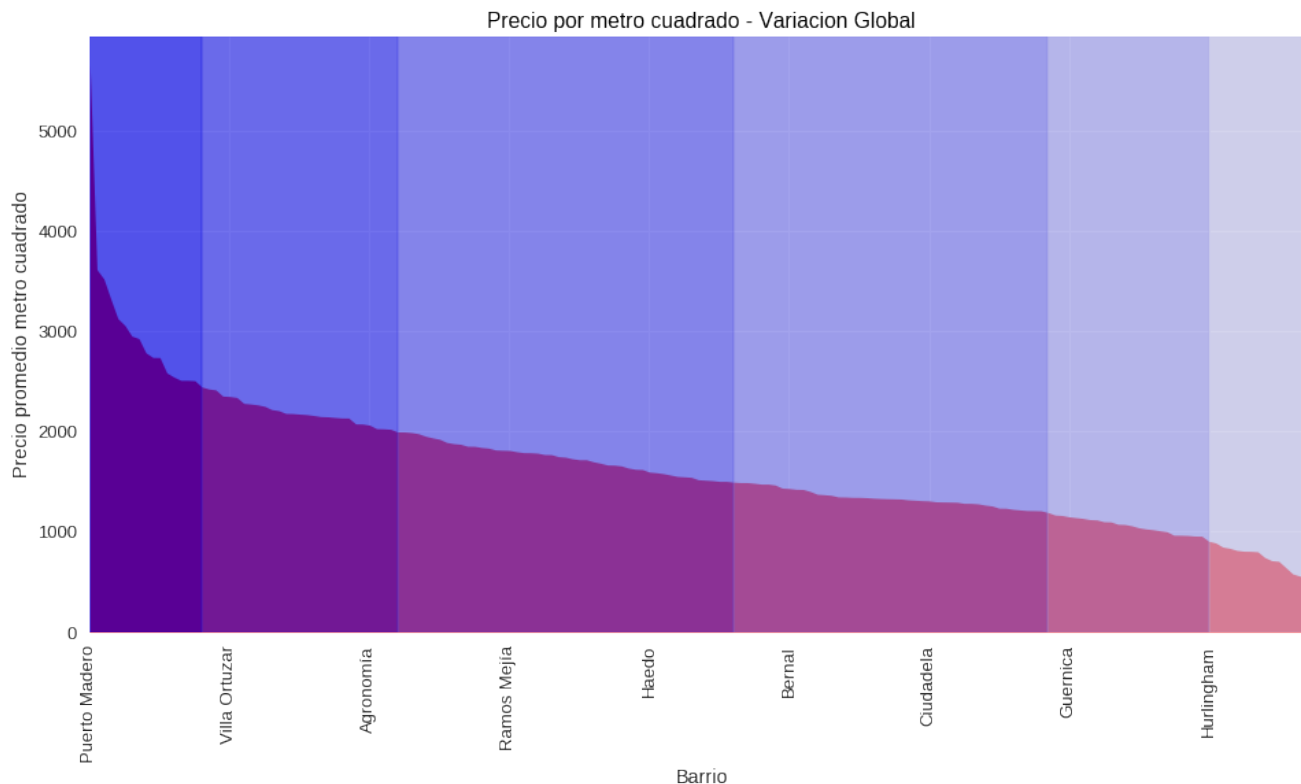
3.3. Dividiendo en secciones

El objetivo de esta parte es determinar diferentes grupos de barrios basados en el precio promedio del m^2 para, de esta manera, analizar que tan suave (o no) es el decrecimiento del valor del m^2 en cada uno de estos grupos.

Los barrios serán divididos en grupos a partir de valores arbitrarios de precio por m^2 , que surgen de un análisis de los datos. Los grupos serán:

Divisiones		
Numero	min(pricem ²) [USD]	max(pricem ²) [USD]
1	2500	∞
2	2000	2499
3	1500	1999
4	1200	1499
5	950	1199
6	450	949

El siguiente gráfico de área muestra el precio del m^2 y las divisiones (diferentes intensidades de azul) indican los diferentes grupos.



Nos interesa mostrar también que porcentaje de los barrios está incluido en cada uno de estos grupos para saber en cuál de ellos se concentra la mayor parte.

Distribución en grupos		
Grupo	Cantidad de barrios	Porcentaje
1	16	9 %
2	28	16 %
3	48	27 %
4	45	26 %
5	23	13 %
6	16	9 %

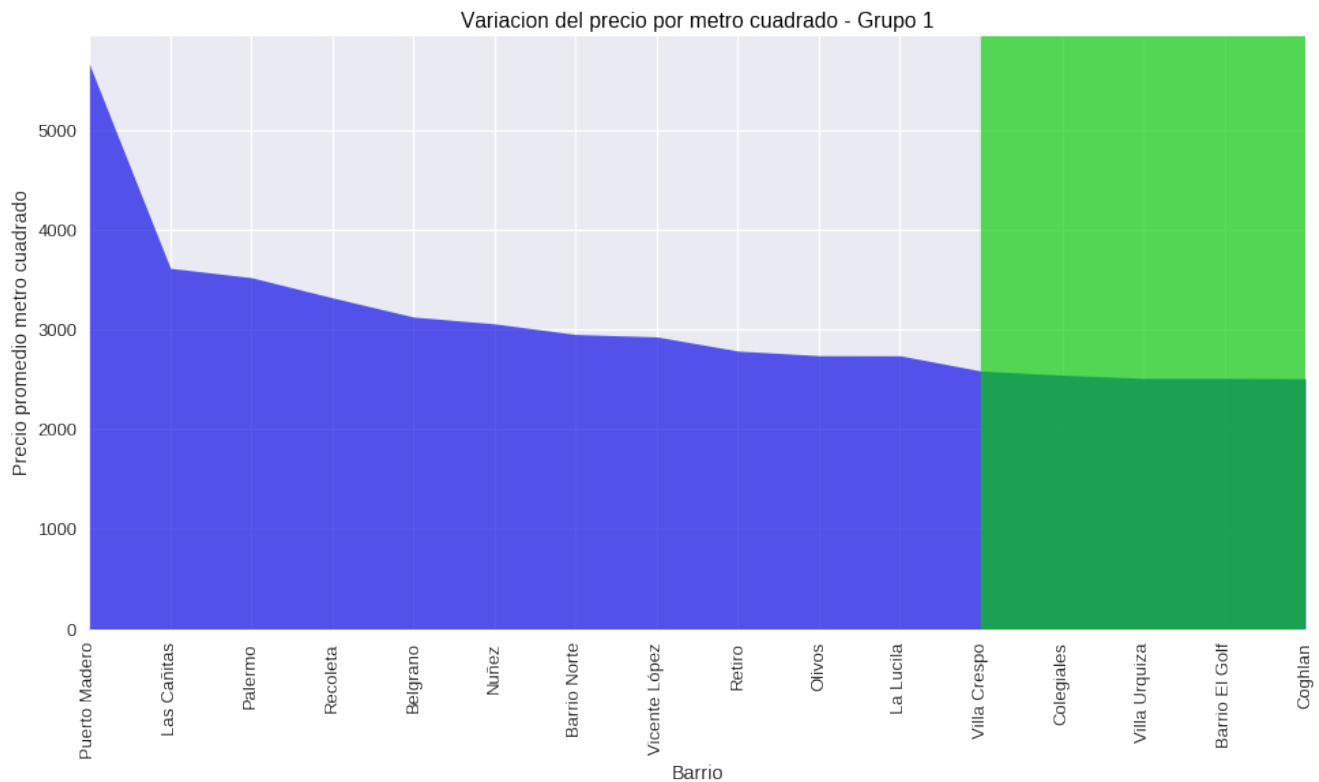
Vemos que, como se puede apreciar en el gráfico, los grupos 3 y 4 son los que concentran a la mayor cantidad de barrios. De hecho, entre ellos dos tienen a más del 50 %.

Ahora procedemos a analizar grupo por grupo.

3.3.1. Grupo 1 - $[2500 : \infty)$ USD

En este grupo se encuentra a los dieciséis barrios más caros de CABA y GBA. Aunque a los primeros diez ya los conocemos de la sección 3.1, lo que nos interesa en esta parte es analizar y comparar cómo varía el precio del metro cuadrado grupo a grupo mas que los nombres propios de cada integrante de cada grupo. En cada uno de los grupos se mostrará un *zoom* del gráfico de área previo, para agregar una visualización al análisis de la variación en cada caso.

En este caso, lo que esperamos encontrar es un inicio con una pendiente muy inclinada y, a medida que nos alejamos del primer barrio (que ya sabemos que es Puerto Madero), un decrecimiento de dicha pendiente, pues la diferencia entre barrio y barrio será mucho menor, aunque seguirá siendo el grupo con mayor variación entre barrios de todos.

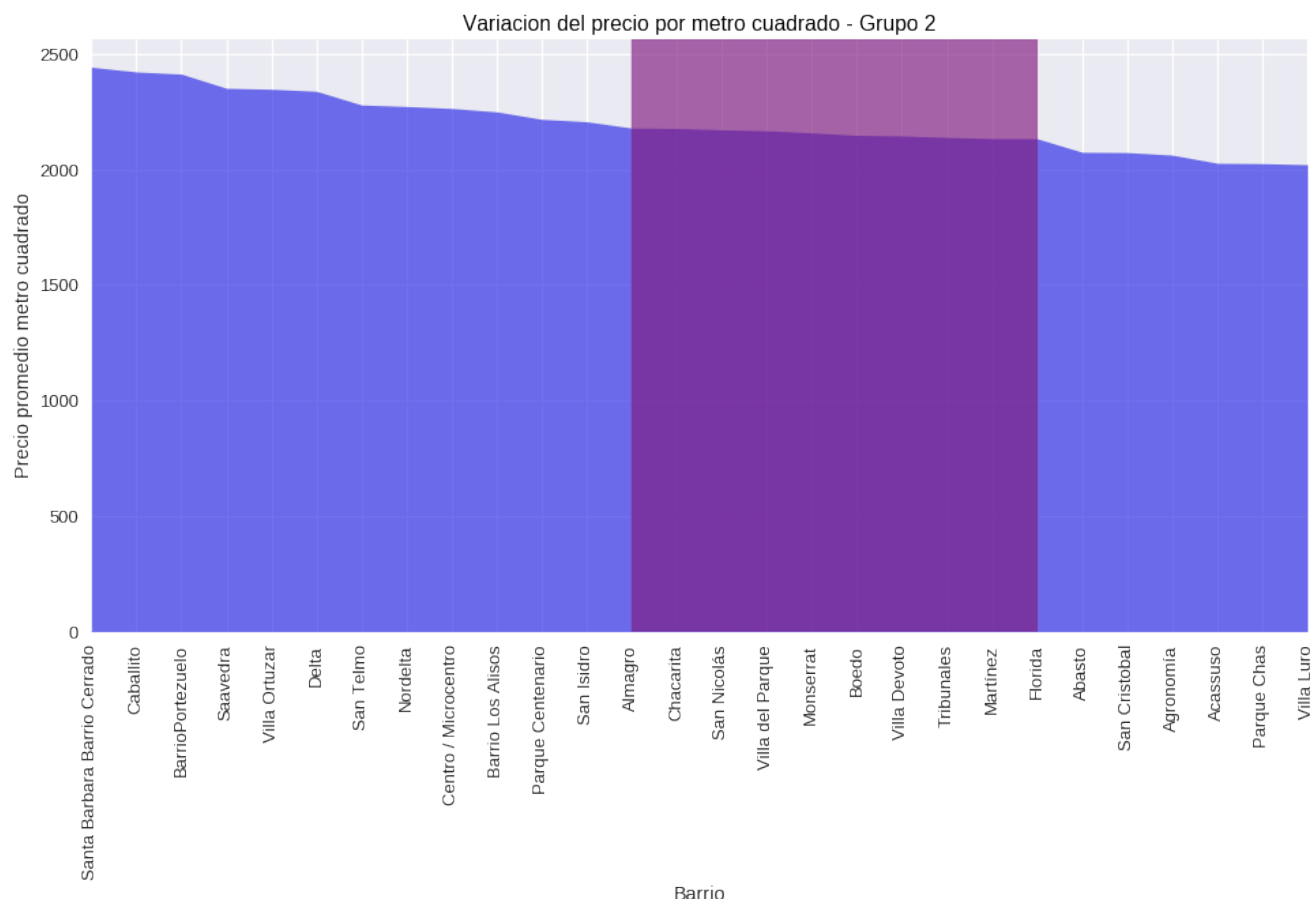


Si bien el gráfico muestra lo que se esperaba, también se ve que a partir del doceavo barrio la curva se vuelve casi constante (en verde), mostrando que hay un cierto subgrupo de barrios con precios muy similares. Esto se puede ver en el gráfico de la sección anterior si se observa el pequeño valle que se forma justo antes de llegar a la división entre el grupo 1 y el 2.

Cabe destacar, además, que si no fuera por el precio extraordinario de Puerto Madero, el grupo tendría una variación mucho menor, pues iría entre 3600USD y 2500USD lo que representaría una variación de 1100USD (30 %) entre el máximo y el mínimo mientras que, actualmente, la variación es de 3100USD (55 %).

3.3.2. Grupo 2 - [2000 : 2500)USD

Observando el gráfico en el que se realizaron las divisiones por grupo esperamos que este segundo grupo tenga un sector medio con pendiente casi nula, donde la diferencia entre el precio en un barrio y el siguiente sea casi nula.



El sector indicado el violeta es aquel que mencionamos previamente con pendiente casi nula. En este subgrupo se encuentran diez barrios y la diferencia entre el primero y el último es de solamente 46USD. Dado los números que se manejan, esa diferencia es casi despreciable, teniendo así un subgrupo de diez barrios con el mismo valor para el m^2 .

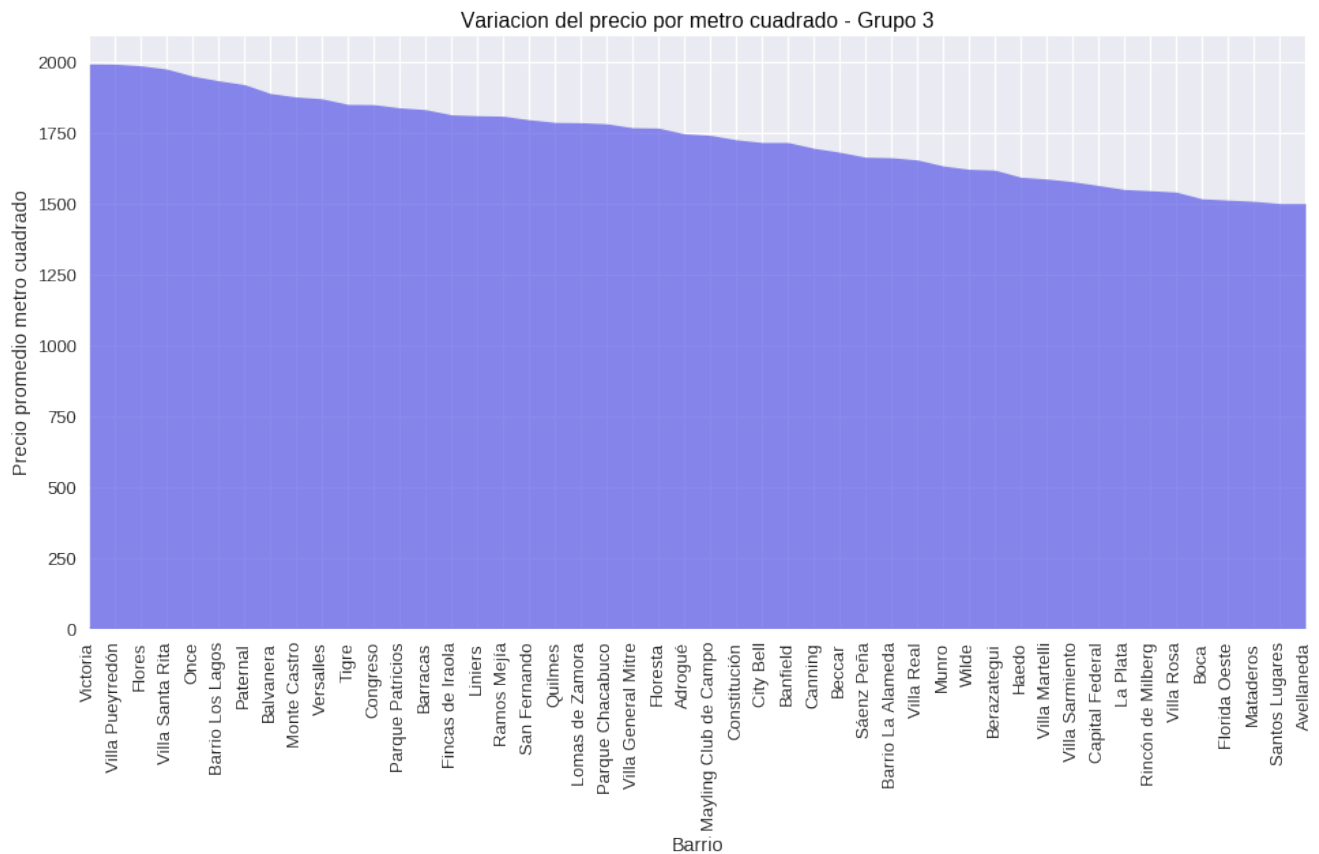
En cuanto al porcentaje de variación, a partir de este segundo grupo lo podemos dividir en dos; uno particular del grupo, en donde el porcentaje se calcula a partir del máximo local, y otro general, en donde el porcentaje se calcula a partir del máximo global (*i.e.* Puerto Madero). La diferencia entre el máximo local y el mínimo es de 422USD, siendo, de esta manera, la variación local del 17% y la general del 7%.

Vemos entonces que la diferencia de variación ya decreció en gran medida pasando del grupo 1 al grupo 2.

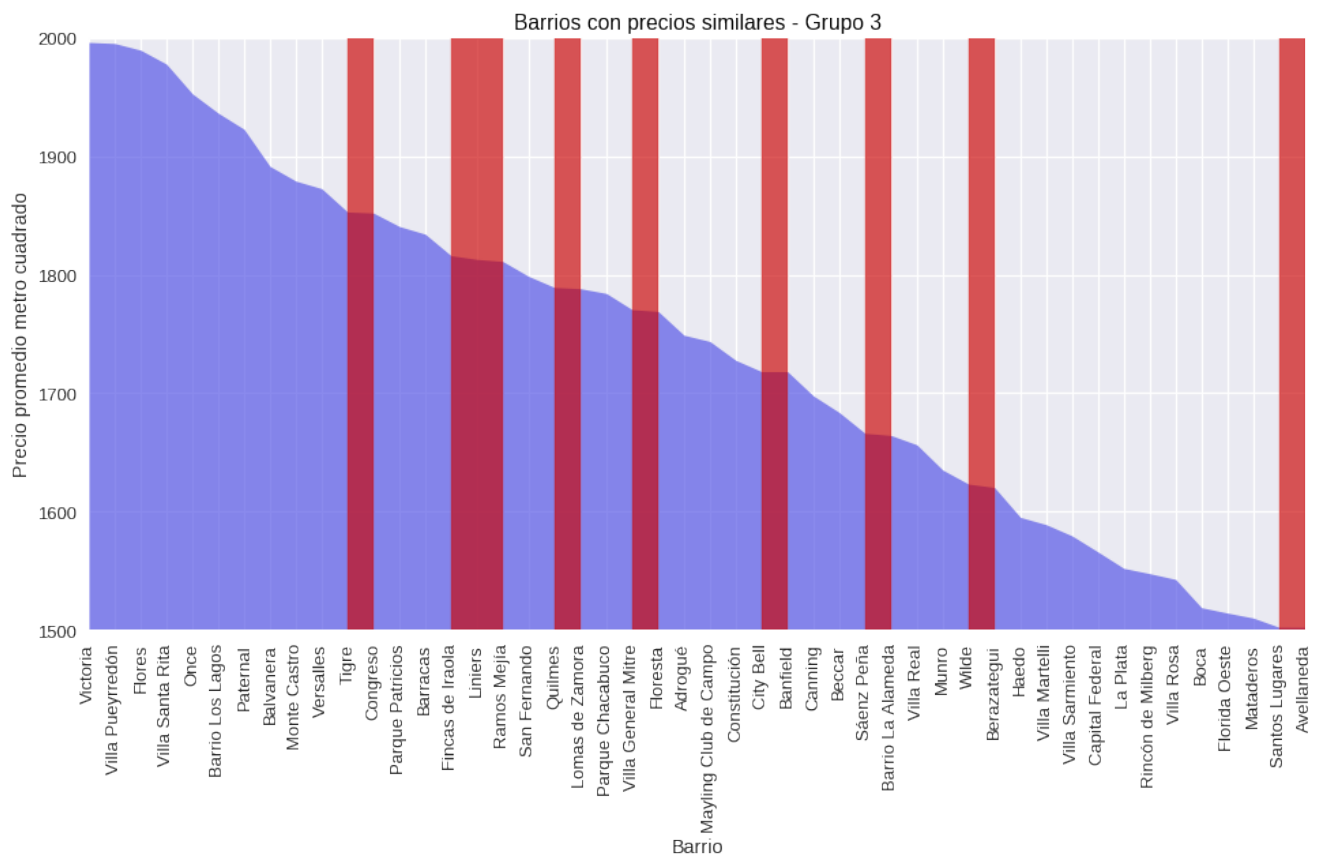
3.3.3. Grupo 3 - [1500 : 2000)USD

Ahora llegamos al primero de los dos grupos principales. Como vimos antes, este grupo concentra al 27% de los barrios de CABA y GBA pero (según nos muestra a grandes rasgos el gráfico general) con mayor variación que el próximo grupo, que también concentra a una gran parte de los barrios.

En este caso, esperamos ver una pendiente casi constante en todo el grupo sin regiones con pendiente casi nula pero con una variación pequeña entre un barrio y el siguiente.



En este gráfico se observa lo previamente indicado, el decrecimiento es bastante lineal y no se detectan zonas con precios constantes. De todas formas, si se amplía un poco, se pueden encontrar pares o tríos de grupos que tienen precios similares, como se ve a continuación.



Aquí podemos ver que si bien hay diecisiete barrios que comparten precio con algún otro, no son contiguos como en el grupo 2. De todas maneras, como en este grupo la variación general de precio es menor y la cantidad de barrios es mayor, podríamos encontrar también un grupo de diez barrios con variación de aproximadamente

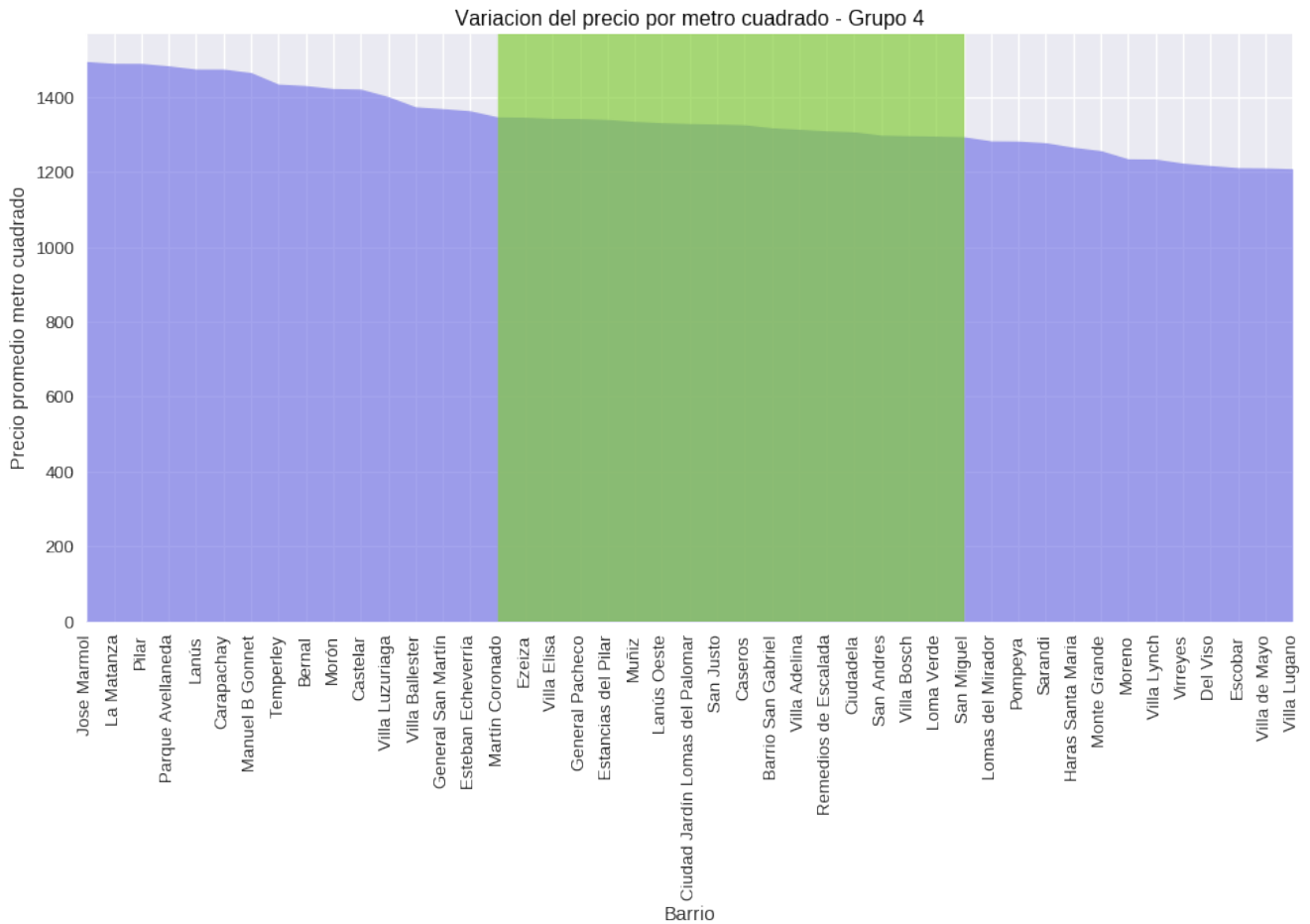
50USD. Por ejemplo, entre Monte Castro y Ramos Mejía hay una diferencia de 60USD y se encierra un subgrupo de nueve barrios.

Analizando el porcentaje de variación al igual que en el Grupo 2, sabiendo que la diferencia entre el máximo local y el mínimo es de 493USD, el porcentaje de variación local es de un 25 % y el porcentaje de variación global es del 9 %.

Se observa entonces que, lógicamente, ahora que estamos en una zona de precios menores, una diferencia entre mínimo y máximo similar a la del grupo anterior ahora representa una variación mayor. Por otro lado, la variación global es apenas mayor a la del grupo anterior. De todas formas, estos porcentajes están sujetos a la división arbitraria de los grupos hecha previamente (pues el máximo delta es 500USD).

3.3.4. Grupo 4 - [1200 : 1500)USD

Este cuarto grupo es el segundo grupo principal, concentra el 26 % de todos los barrios y es el de menor rango de valores, con 300USD, salvo por el grupo 5, con 250USD. Esto nos dice que la pendiente en este caso será la menor de todas, como se puede ver en el gráfico general. Además, esperamos ver una zona central con pendiente casi nula, en donde se encontrará un grupo muy grande de barrios con una variación en el precio muy pequeña, a diferencia del grupo 3.



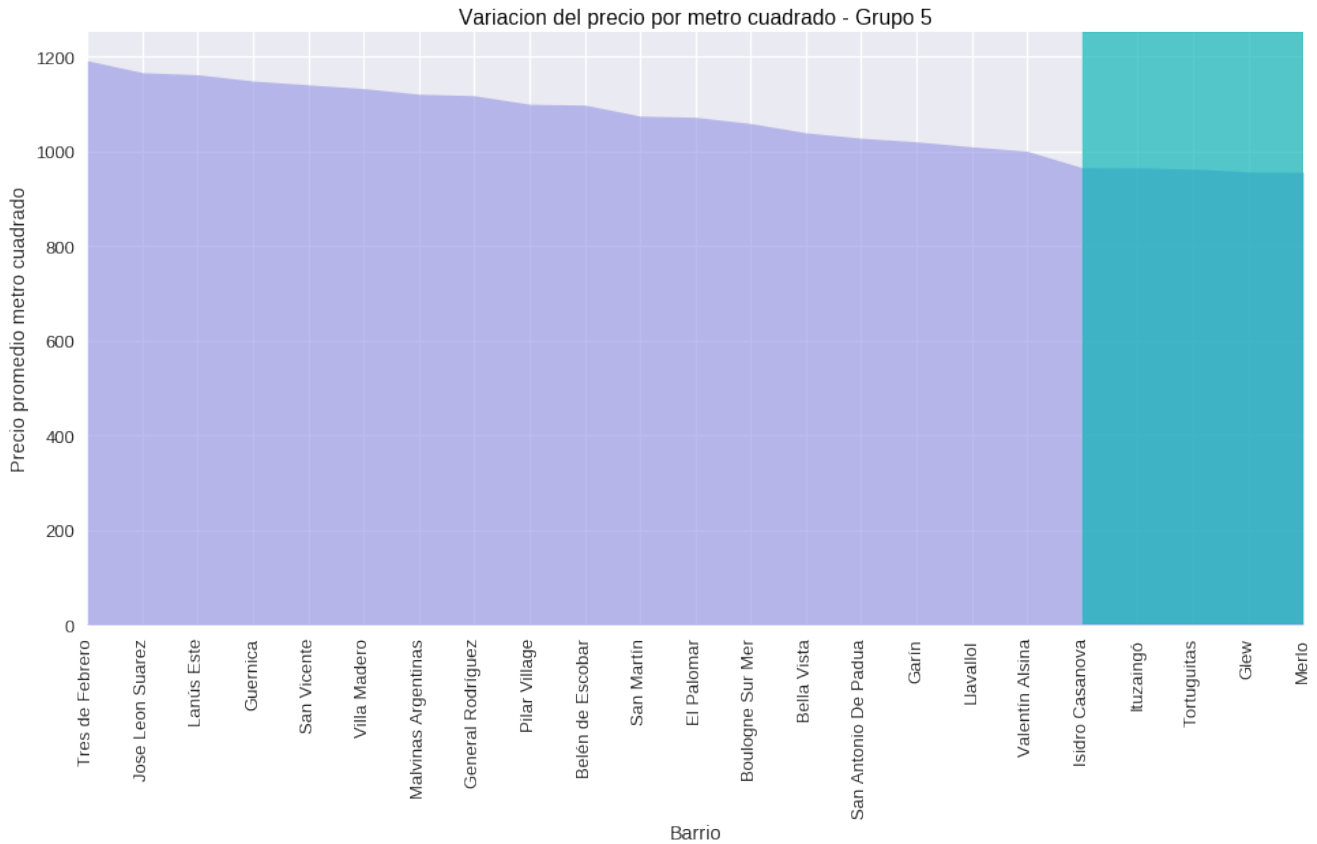
El subgrupo indicado con verde consiste de dieciocho barrios que se diferencian solamente, entre el primero y el último, por 53USD. Como supusimos, en este grupo se encuentra el conjunto de barrios con variación casi nula más grande de todos. Adicionalmente, si se observa el gráfico con atención, en la mayor parte se encuentran subgrupos pequeños con valores casi constantes hasta que hay un pequeño descenso de precio, nuevamente un valor casi constante, etc. dividiendo el grupo en cinco o seis diferentes subgrupos pequeños (o no tanto, como el verde) por lo que no tiene sentido destacarlos como en el grupo 3, pues solo quedarían aquellos pequeños sectores donde la pendiente si toma valor considerable.

En cuanto al porcentaje de variación, sabiendo que la diferencia entre el máximo local y el mínimo es de 286USD, tenemos un porcentaje de variación local del 19 % y uno global del 5 %. Por lo tanto, y como era de esperarse, este grupo tiene los porcentajes mínimos en ambos casos; en el local porque tiene la mayor cantidad de segmentos casi constantes y en el global porque, además de la razón recién mencionada, por tener valores cada vez más bajos. Cabe destacar, además, que el 19 % local está concentrado en ciertos 'saltos' de un barrio a otro; pues, como mencionamos antes, la mayor parte del grupo esta compuesta por subgrupos de precios similares.

3.3.5. Grupo 5 - [950 : 1200)U\$D

Este quinto y penúltimo grupo comienza una pendiente que incrementa progresivamente hacia el último y sexto grupo. Aquí se concentra el 13 % de los barrios, la mitad que en el grupo anterior, y los precios de los barrios ya se acercan a los mínimos que se pueden encontrar.

A partir de la información que brinda el gráfico general, esperamos encontrar una pequeña zona de pendiente casi nula hacia el final del grupo y una variación local relativamente chica.

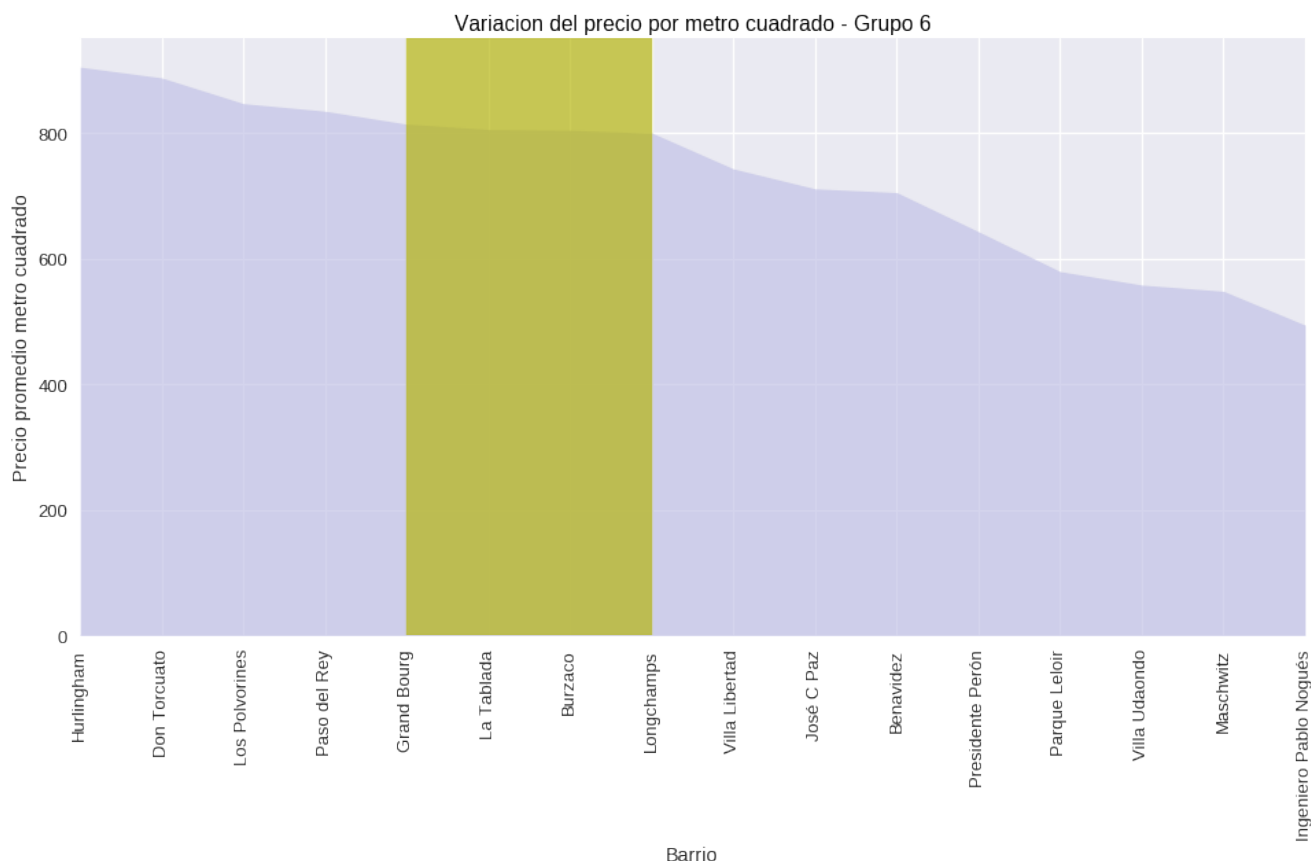


Indicado con color Cyan en el gráfico podemos ver ese pequeño grupo del que hablábamos donde la pendiente es casi nula, mientras que en el resto del grupo hay una pendiente casi constante. El tamaño de este subgrupo es de cinco barrios y la diferencia de precio entre el primero y el último es de 11USD.

Si calculamos el porcentaje de variación, conociendo que la diferencia entre el máximo local y el mínimo es de 235USD, obtenemos un porcentaje de variación local del 20 % y uno global del 4 %. En este caso, el porcentaje local esta distribuido a lo largo de todo el grupo, salvo por el subgrupo final que tiene un valor casi constante.

3.3.6. Grupo 6 - [450 : 950)U\$D

En este último grupo, que contiene sólo al 9 % de los barrios (al igual que el grupo 1), estaremos analizando a aquellos barrios con menor precio por m^2 . A partir del gráfico general sabemos que la pendiente será muy marcada y que la diferencia entre un barrio y el siguiente será importante, salvo por una pequeña parte donde los precios se mantendrán casi constantes.



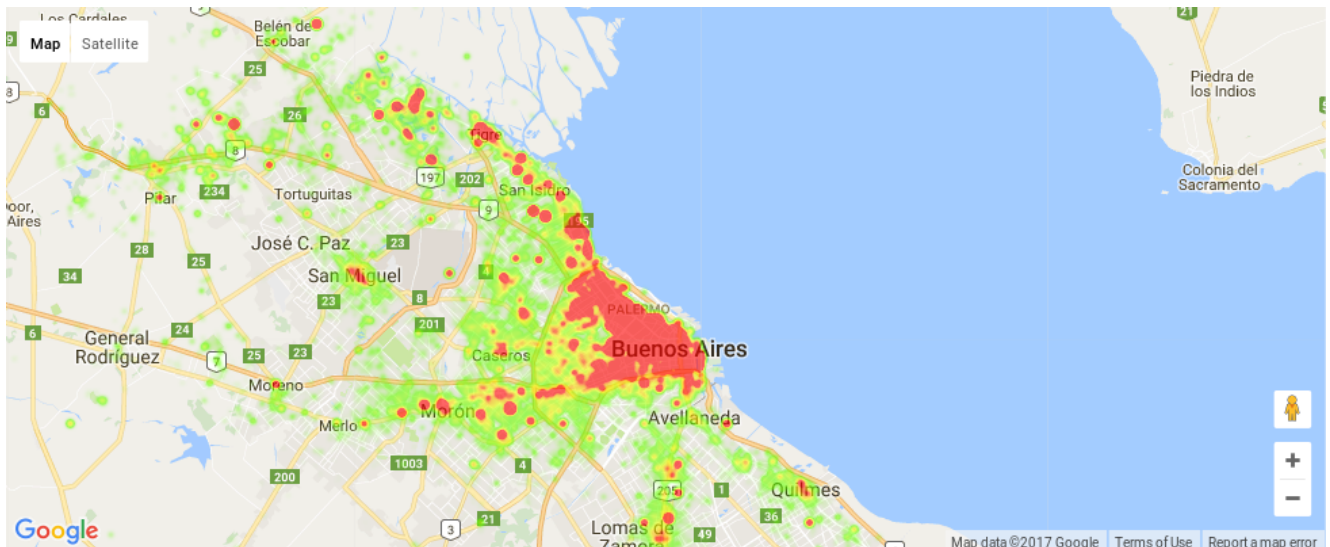
En amarillo se puede observar la única sección casi constante del grupo, que consta sólo de cuatro barrios y que contienen un rango cuya diferencia entre el mínimo y el máximo es de 14USD. Mientras tanto, el resto del grupo tiene una pendiente considerablemente grande y la diferencia entre el valor máximo y el mínimo es casi del doble.

Para remarcar esto último, calcularemos el porcentaje de variación sabiendo que dicha diferencia es de 410USD. El porcentaje de variación local es del 45 % y el porcentaje de variación global es del 7 %. Aquí la variación local está repartida en todo el grupo pero con mayor participación de la segunda mitad, donde la pendiente es mayor, y con menor participación del sector casi constante.

4. Distribución geográfica

En esta sección se mostrará, con la ayuda de *HeatMaps* cómo están distribuidos los precios por m^2 en CABA y GBA.

Para empezar, mostraremos un mapa general de CABA y GBA donde la unidad de medida es el precio por m^2 . Se debe tener en cuenta que el software utilizado determina el color de un pixel a partir de la suma de todas las propiedades que se corresponden con ese pixel, y por dicha razón, podríamos encontrar sectores donde, si bien el precio no es tan elevado, hay muchas propiedades y por ende se las indica con rojo. De todas formas, este gráfico nos sirve para tener un plano general.

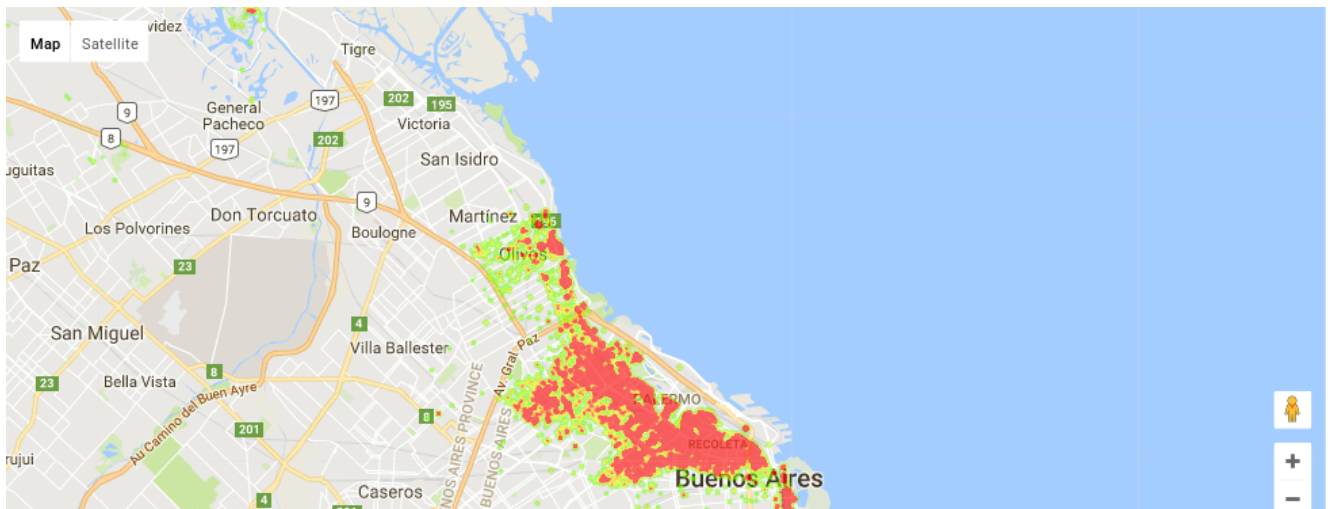


4.1. Grupos característicos y su ubicación

Ahora, lo que haremos será repetir la dinámica pero dividiendo en los mismo grupos característicos de antes, para ver que se puede saber de cada grupo según su ubicación. Cabe destacar que la escala de precios utilizada en los mapas de cada grupo es diferente, pues el objetivo aquí es mostrar la ubicación de estos barrios más que el valor que ellos tienen.

4.1.1. Grupo 1

Para el grupo 1, como se dijo previamente, se espera encontrar a las propiedades en CABA y la zona de GBA que corresponde al inmediato Norte de la Capital Federal.

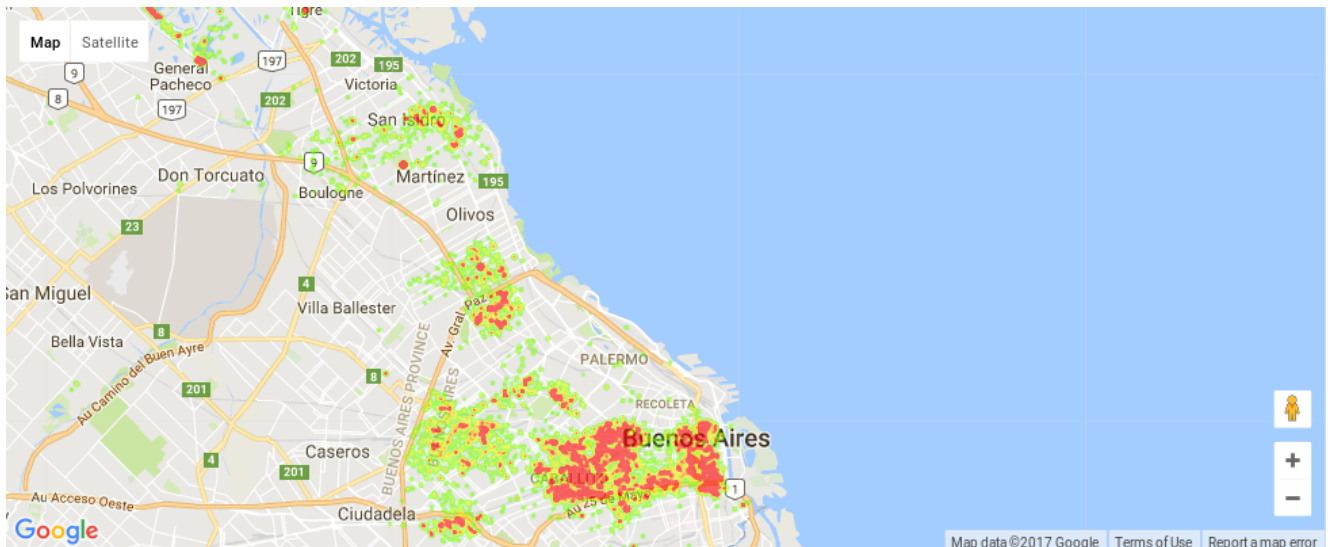


En el gráfico se observa, más específicamente, que las propiedades con valores altos están ubicadas en el Este de la ciudad, esparciéndose desde el Microcentro hacia el Norte de CABA. Se puede ver, también, que las propiedades de la Provincia que se encuentran en este grupo son aquellas cercanas al río. Cabe destacar que Puerto Madero genera una 'discontinuidad' en el mapa, pues si se observa atentamente se ve que no está en una zona contigua al resto de los barrios. De todos modos, esto era de esperarse porque es un barrio construido específicamente como un 'lugar caro'.

Se debe mencionar, además, que el Barrio El Golf está ubicado en la localidad de Tigre y en este mapa no se lo puede ver. Se decidió dejarlo afuera para poder dar una mejor vista de la Capital Federal.

4.1.2. Grupo 2

Este segundo grupo se espera que esté compuesto por los barrios de capital y Zona Norte que no estaban presentes en el grupo previo. Esto es, San Isidro, Caballito, Saavedra, etc.

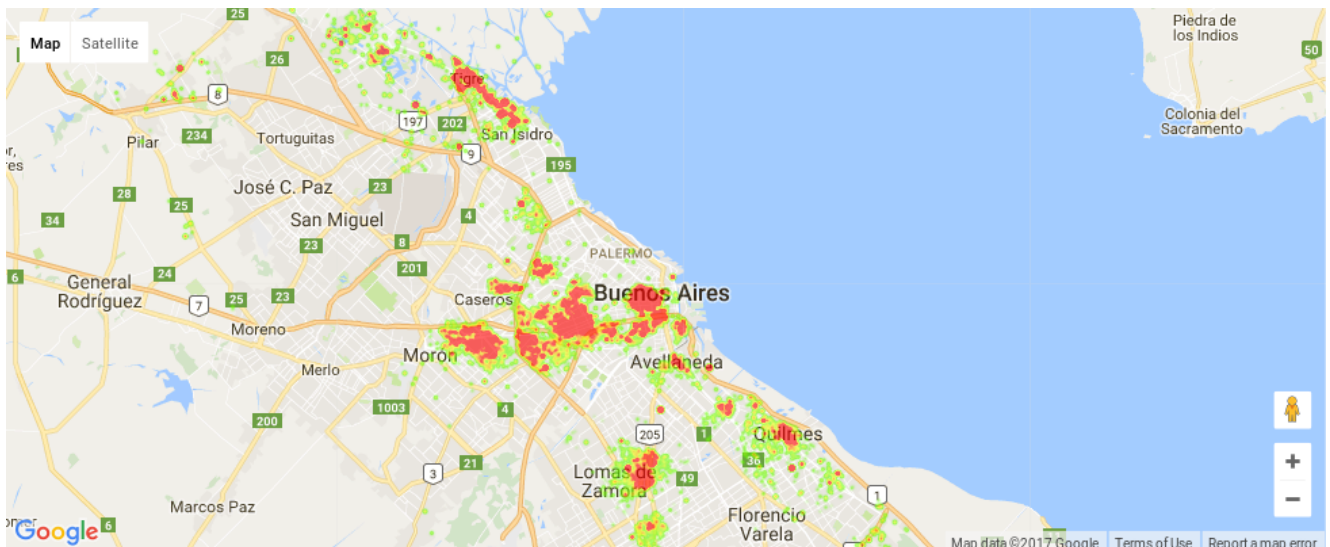


El gráfico nos muestra que este segundo grupo contiene a los barrios del Sur Este de la Capital Federal ('del centro para abajo', al revés que el grupo 1) y aquellos barrios que 'completan' la Zona Norte. Además, se encuentran algunos otros barrios dispersos, como por ejemplo Boedo.

Cabe destacar aquí también que en este segundo grupo se encuentran más barrios cerrados de la localidad de Tigre, como el Barrio Santa Bárbara y el Barrio Los Alisos, que nuevamente no se incluyeron en el mapa para poder dar más detalle a CABA.

4.1.3. Grupo 3

En este tercer grupo se espera que, además de que se cubra la mayor cantidad de superficie, se 'complete' la capital y comiencen a aparecer los barrios del llamado primer cordón del conurbano.

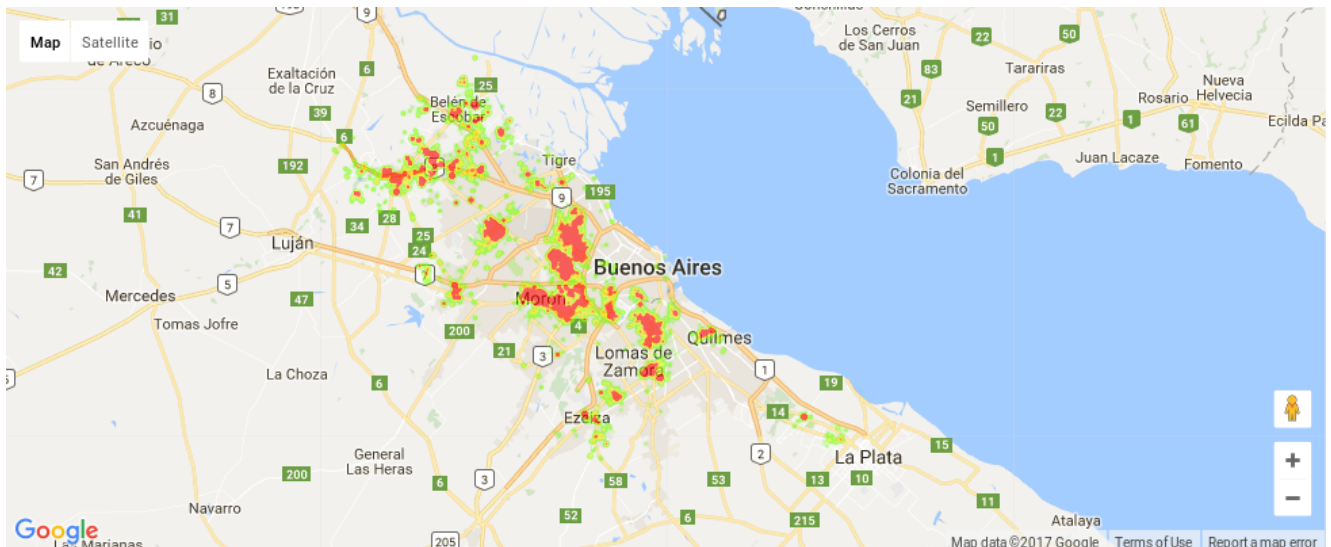


El mapa presentado nos muestra, efectivamente, que está compuesto de los barrios restantes de la Capital Federal, algunos barrios del primer cordón, como Villa Martelli, Munro, Morón y otros más lejanos como Quilmes y Lomas de Zamora. Además, como se puede observar al Norte, se sigue completando la zona de barrios cerrados de la localidad de Tigre y sus alrededores.

También forma parte de este grupo la ciudad de La Plata pero se decidió dejarlo afuera de la visualización para poder enfocarnos más en CABA y GBA.

4.1.4. Grupo 4

Para este cuarto grupo, el segundo grupo principal como fue llamado previamente, se espera que se 'complete' el primer cordón del conurbano y comiencen a aparecer barrios del segundo y tercer cordón. También se espera que ya no aparezcan barrios de CABA.



En este caso se debió utilizar un nivel de zoom menor por la distancia entre los barrios del grupo. De todas maneras, se puede apreciar la aparición de más barrios del primer cordón del conurbano y de Pilar y Escobar, que son, al igual que Tigre, dos centros importantes de barrios cerrados y country clubs.

4.1.5. Grupo 5

Ya acercándonos a los barrios más baratos esperamos, tanto en este grupo como en el siguiente, encontrar barrios alejados de la Capital Federal pertenecientes al segundo y tercer cordón del conurbano.



Se puede observar que se completan algunas zonas faltantes del primer cordón del conurbano y aparecen zonas densas del segundo y tercer cordón, como se esperaba. De todas formas, al poseer cada vez menor cantidad de barrios, es más difícil hacer un mapa claro.

4.1.6. Grupo 6

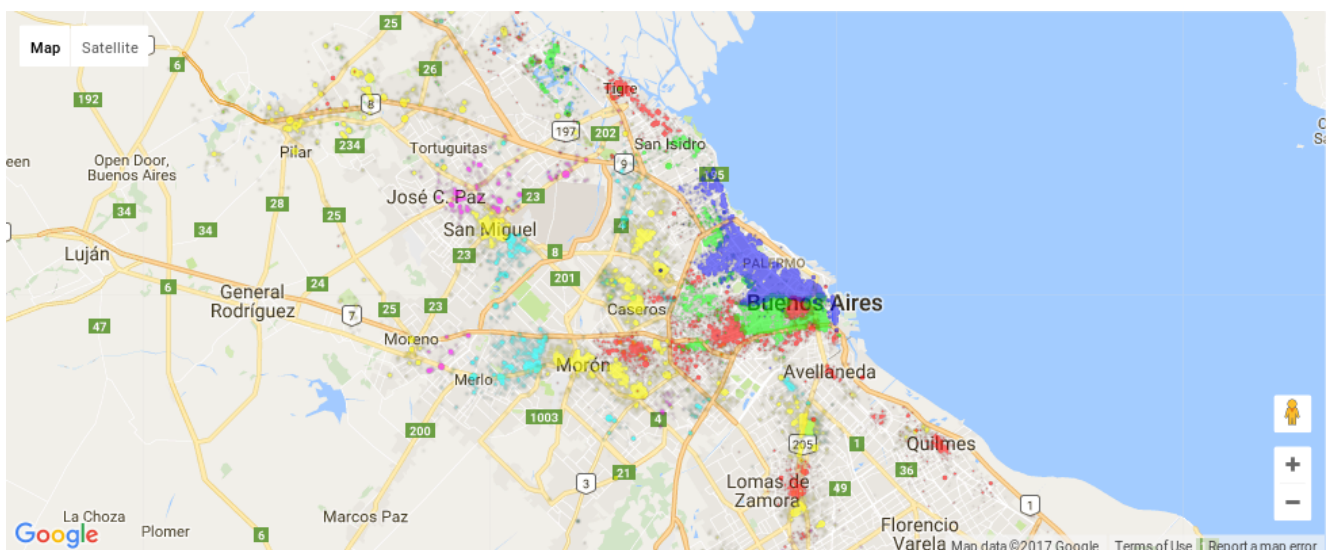
Finalmente, en este sexto grupo esperamos encontrar barrios dispersos y alejados de la Capital Federal, con poca cantidad de propiedades.



Se verifica con este gráfico, entonces, que los integrantes de este último grupo pertenecen al tercer, y más alejado, cordón del conurbano bonaerense.

4.1.7. Comparación de grupos

En este último gráfico mostraremos cómo se distribuyen los grupos en CABA y GBA, unificando todos los gráficos previos para un mejor entendimiento. Se utilizaron diferentes colores para diferenciar los grupos y dichos colores están indicados en la leyenda del gráfico.



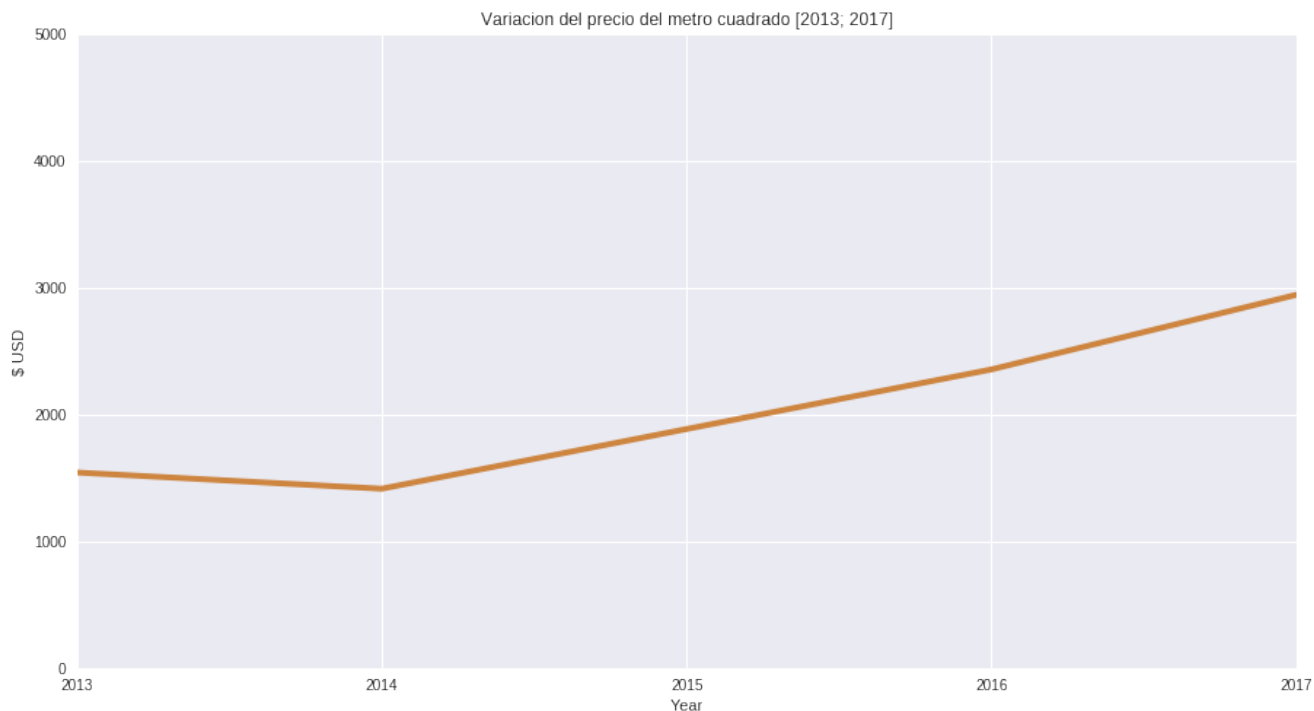
NOTA: Grupo 1 AZUL, Grupo 2 VERDE, Grupo 3 ROJO, Grupo 4 AMARILLO, Grupo 5 CYAN, Grupo 6 MAGENTA. *AGREGARLE LA LEYENDA A LA IMAGEN*

5. Progresión del precio por m^2 a través de los años

En esta sección se analizará como fue progresando el valor del m^2 entre los años 2013 y 2017. Para esto, se realizaron gráficos particulares para cada año como así también dos gráficos que comparan a todos los años entre sí.

5.1. Fluctuación del precio del m^2 a través de los años

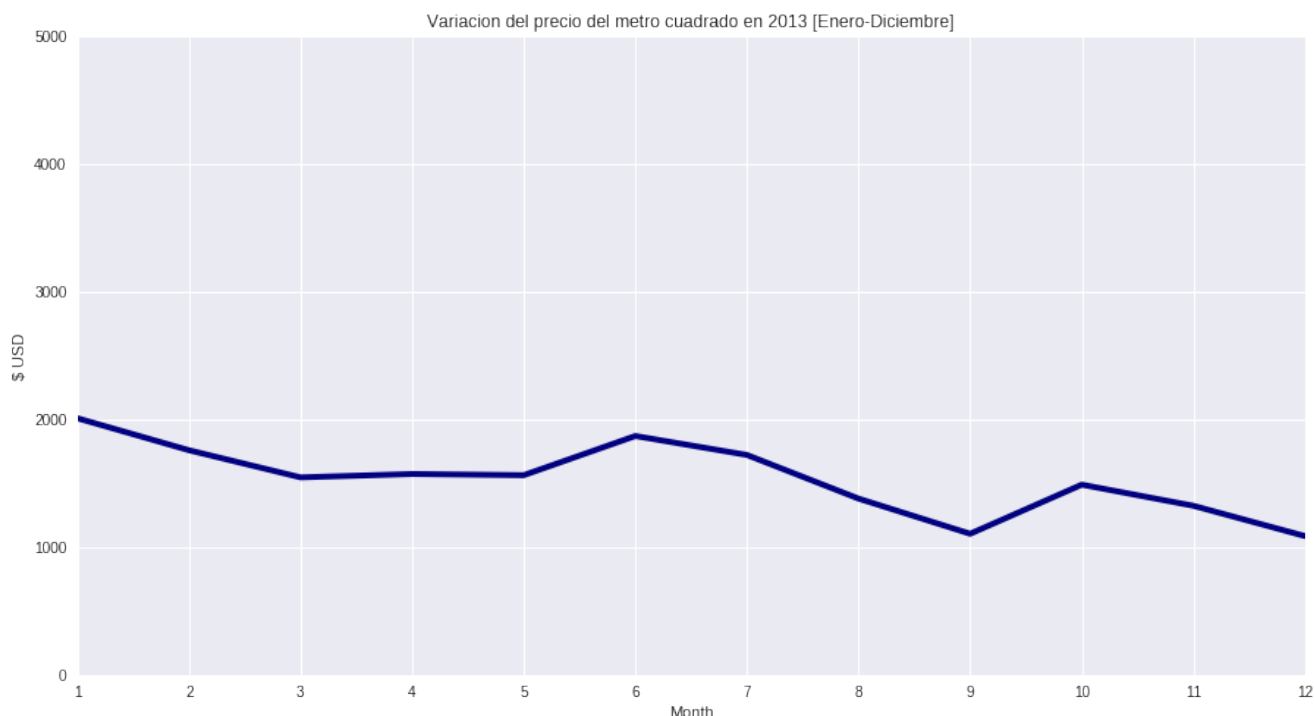
Para comenzar, se mostrará un gráfico en el que se muestra la variación del precio del m^2 entre el año 2013 y el año 2017.



Se observa que entre 2013 y 2014 hubo un descenso del precio promedio pero que a partir de dicho año, el precio de las propiedades estuvo en aumento constante.

5.2. Año 2013

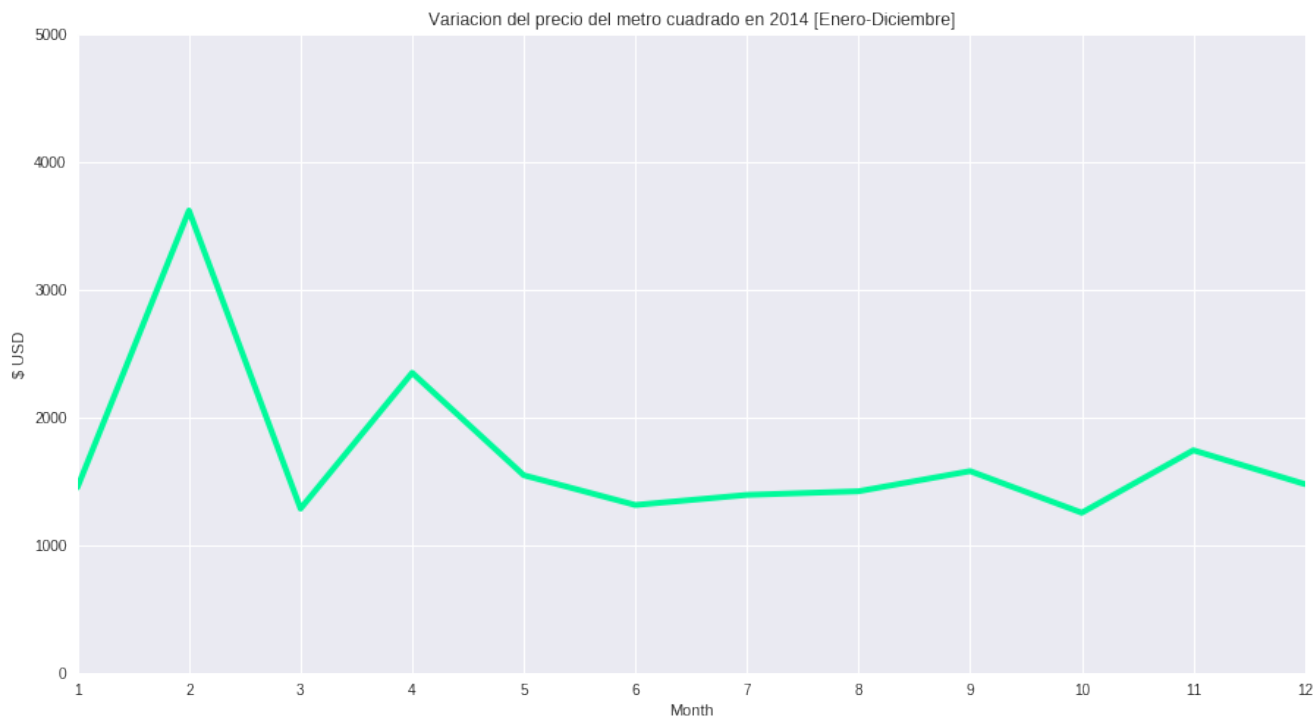
En este caso, se debe tener en cuenta que la cantidad de propiedad es la menor de todas y por lo tanto es posible que los precios disponibles no sean lo suficientemente representativos. De todas maneras, consideramos que es útil para un análisis comparativo con los años más recientes.



Se puede observar que el año 2013 tuvo una tendencia decreciente con aumentos sólo en los meses de Mayo y Septiembre.

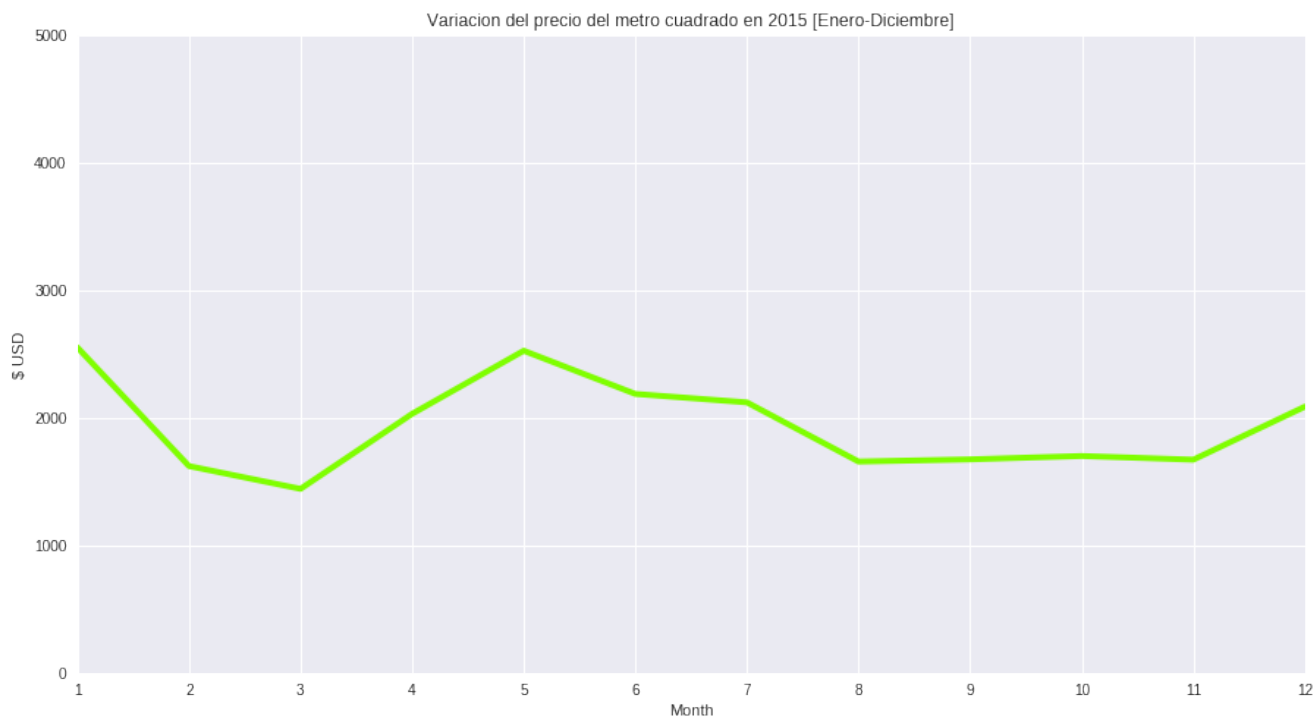
5.3. Año 2014

Aquí la cantidad de propiedades ya es mayor, aunque muy menor a la de los años 2016 y 2017, de todas maneras, al igual que en el año anterior, es lo suficientemente descriptiva como para realizar una comparación más adelante.



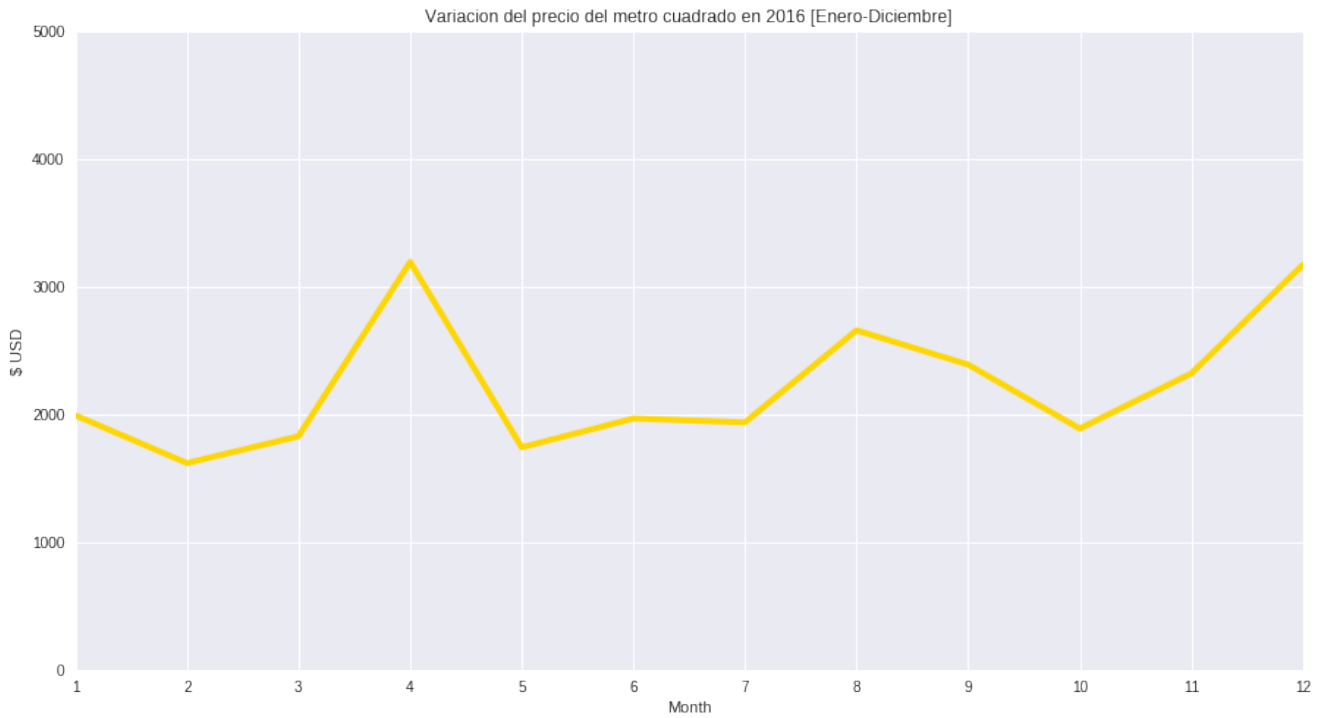
En este caso, vemos que hay una gran fluctuación de precios en los meses iniciales y finales pero se encuentra una región casi constante en los meses entre Mayo y Septiembre. Podemos ver que, al igual que en 2013, el año termina con una baja del valor de m^2 .

5.4. Año 2015



Podemos ver que este año comenzó con un decrecimiento del precio, luego volvió al precio de Enero y, luego de tres meses de precio constante, finalizó creciendo en Diciembre. Cabe destacar que el valor es bastante uniforme para los tres años hasta aquí analizados, aunque con un leve aumento.

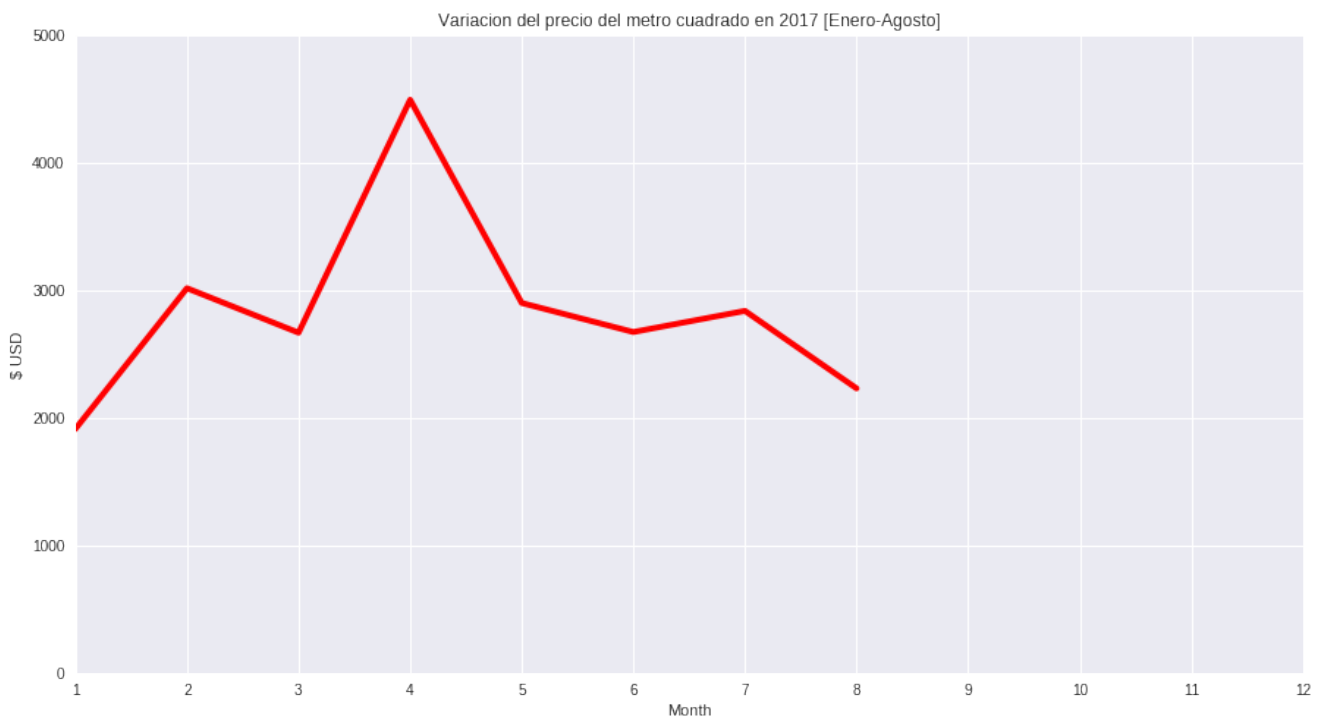
5.5. Año 2016



Este es, hasta aquí, el año con mayor fluctuación de precios. Con los decensos y ascensos más abruptos y con los valores más elevados. Se puede observar que el año finaliza con un aumento de precios muy importante, alcanzando el valor más alto hasta el momento.

5.6. Año 2017

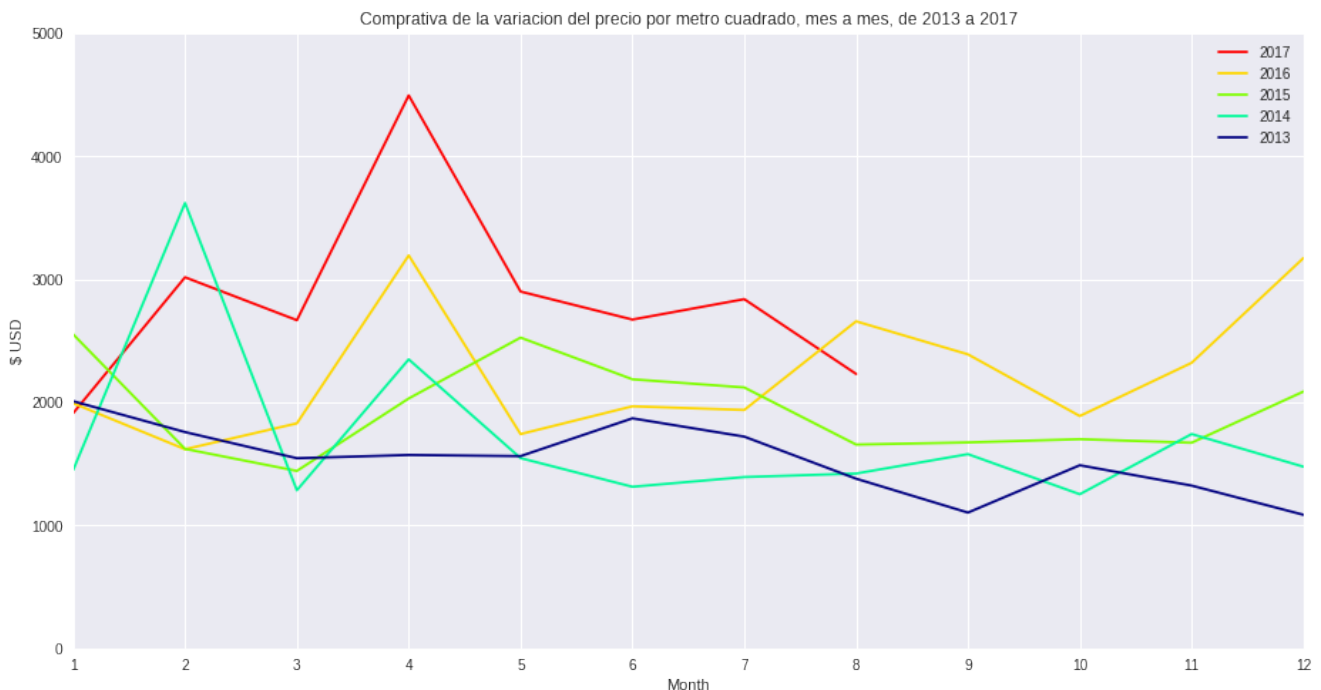
Este es el año del que más información se posee, pues la información disponible es sobre publicaciones activas en la página oficial de *Properati*. Obviamente, la información sólo llega hasta el mes de Agosto.



Aquí podemos ver que el año comenzó con un gran aumento, llegando a un pico de precios de los últimos cuatro años pero que, a partir del mes de Abril, comenzó el descenso y, para el mes de Agosto, ya casi se volvió a los valores de Enero.

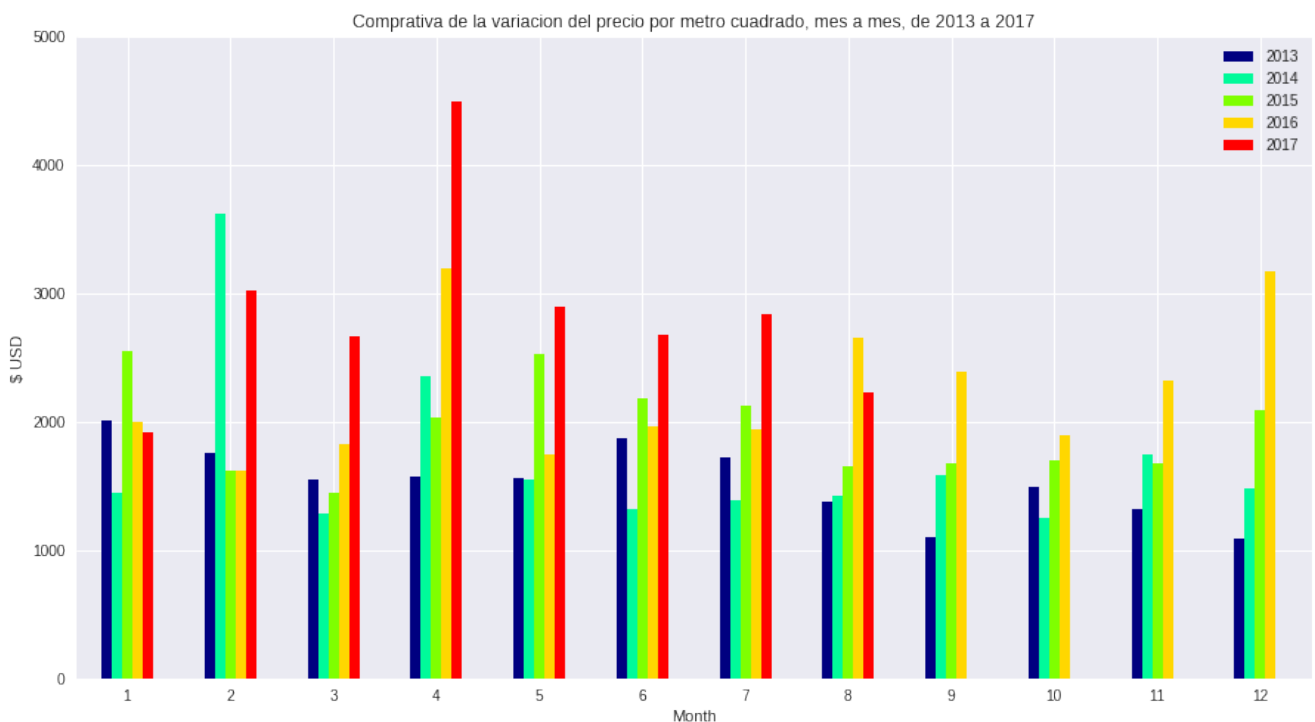
5.7. Comparación de la progresión mensual en cada año

Ahora veremos dos gráficos en los que se unifica todo lo mostrado previamente, el cual nos permite comparar, año a año, cómo fue la variación de los precios. El primero es exactamente la superposición de los cuatro gráficos previos, mientras que el segundo (de barras) se utilizó para analizar mejor, mes por mes, en que año fue mayor el valor del m^2 . Comenzaremos analizando el mencionado gráfico de líneas:



Aquí podemos ver que el año 2017 (o al menos en lo que va del año) es el que mayores precios tiene en cinco de los ocho meses en los que se posee información. Por otro lado, podemos ver que 2015 y 2016 terminan en suba mientras que 2013 y 2014 terminan en baja. También se puede ver que en los meses de Abril, Mayo y Junio hay una tendencia a suba de precios, mientras que no se encuentra una zona particular en la que haya tendencia a la baja de precios.

Ahora analizaremos el segundo gráfico, el de barras. El objetivo de este gráfico es, como se mencionó previamente, analizar, mes a mes, qué año tuvo los valores más altos.



Este gráfico permite ver que, si bien en 2017 se obtuvieron casi todos los valores récord en cada mes, mes a mes varía el orden de los años en cuanto a precio. Esto nos dice que los precios fluctuaron mucho en los últimos

años aunque con una tendencia positiva (como ya vimos previamente), pues la mayoría de los máximos pertenecen a años recientes.

6. Análisis de precio aproximado.

6.1. Objetivo

Esta sección tiene el objetivo de analizar como fueron variando los precios de las distintas propiedades en los últimos 4 años. Con fin introductorio, se quiere brindar una mirada global acerca de la fluctuación de los precios de todas las propiedades a lo largo de los últimos 4 años.

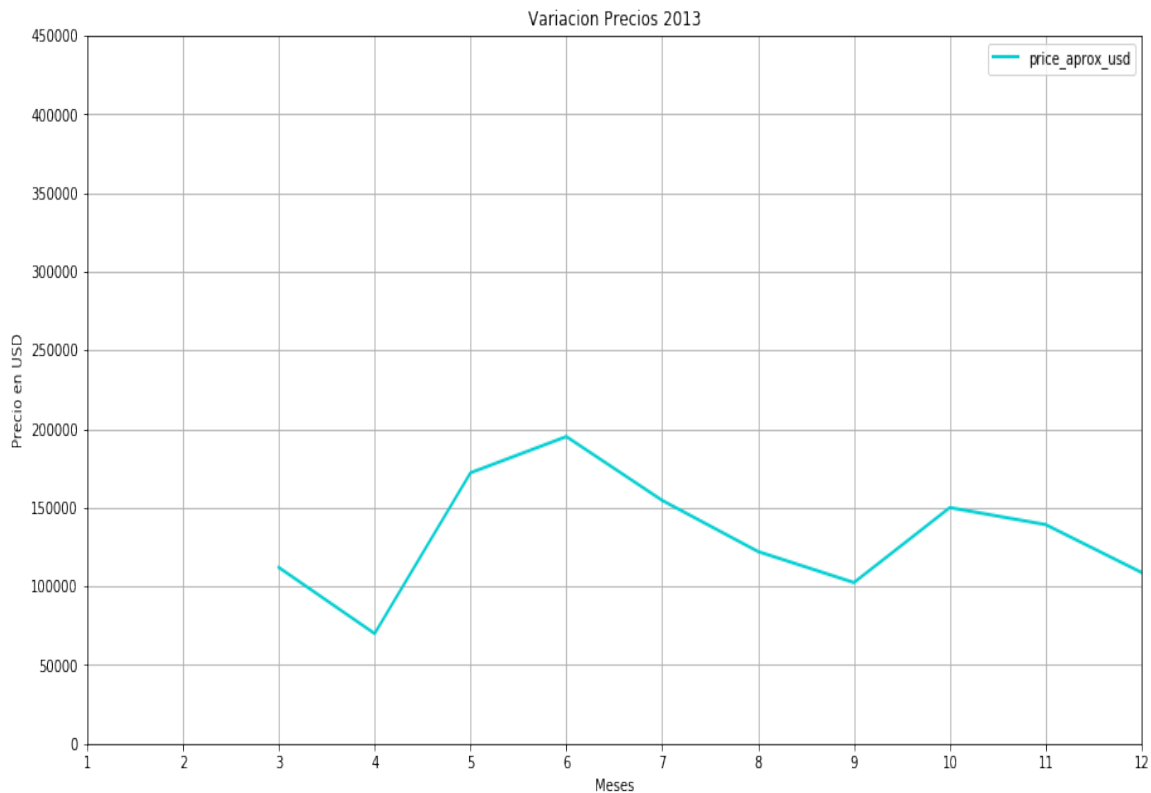
6.2. Preparación y procesamiento de los datos

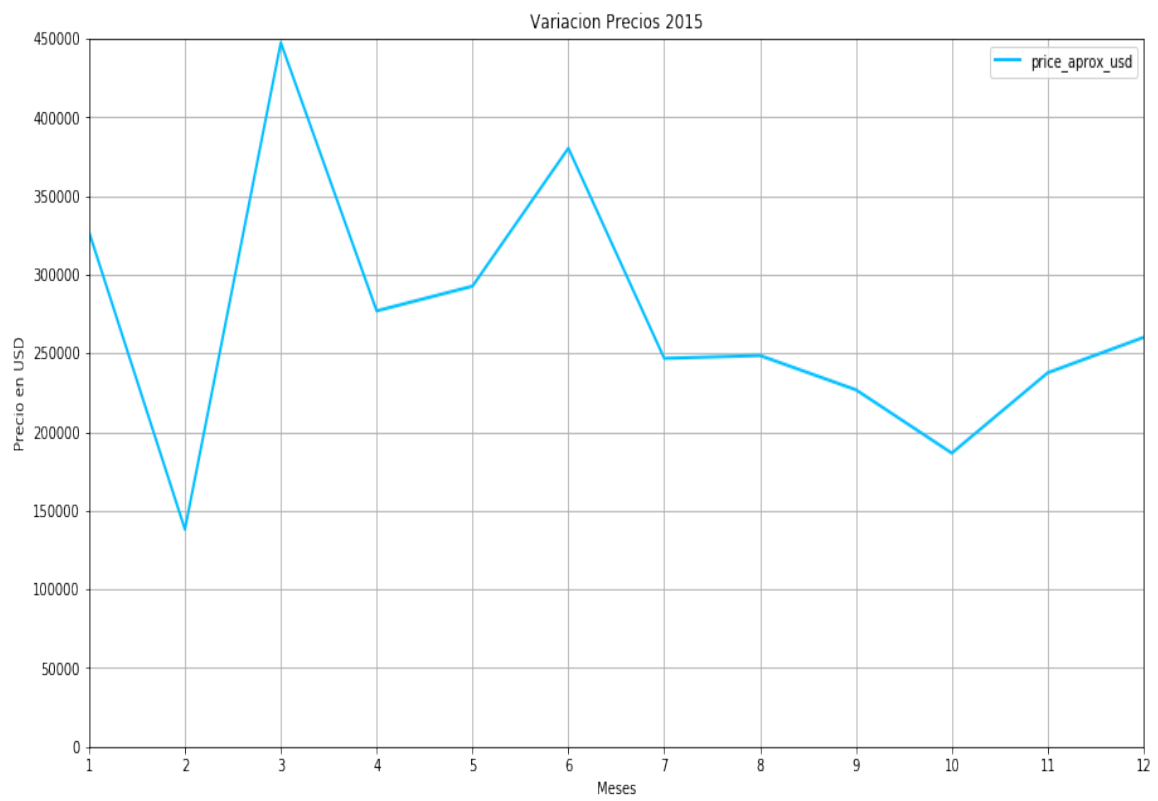
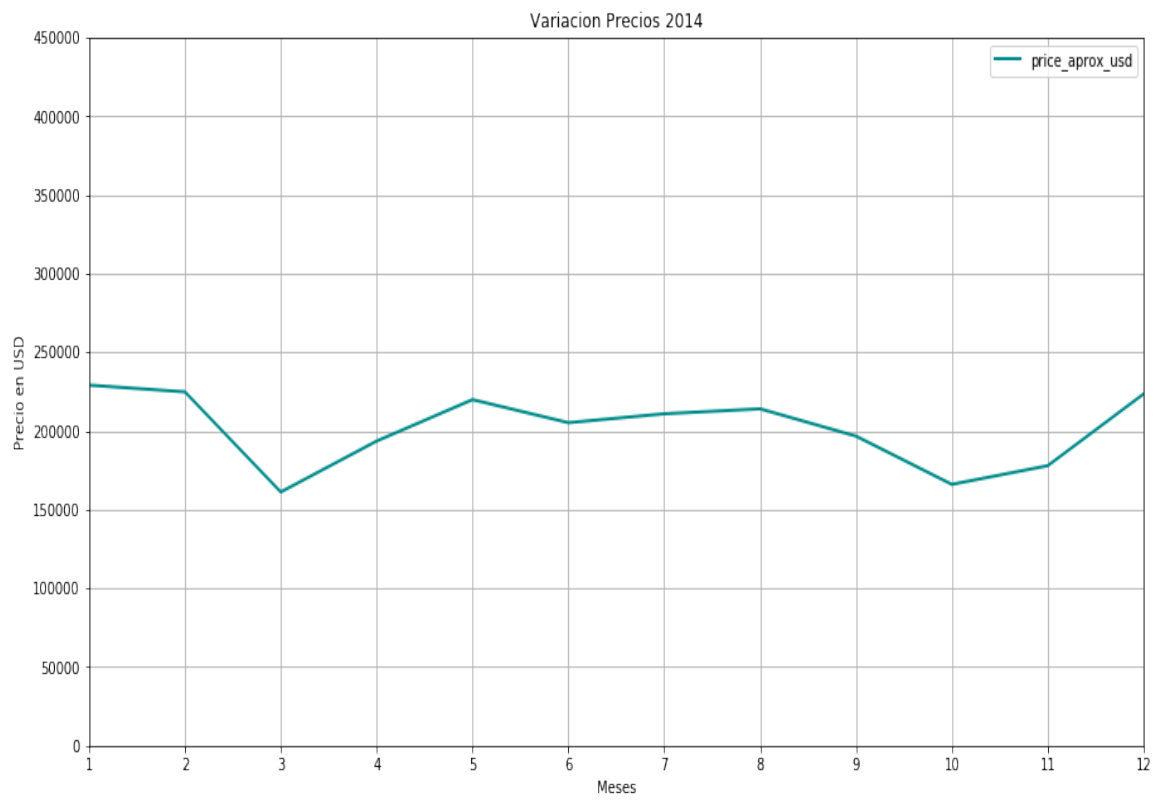
Para procesamiento de los datos, primero se separaron y agruparon los features por años. A continuación, para cada dataframe formado, es decir para las propiedades de un determinado año, lo que se hizo fue separarlas por meses. Entonces hasta el momento tenemos las propiedades separadas por años y por meses. Por último para cada año, se agruparon todas las propiedades de un mismo mes dejando como valor el precio en dolares promedio de las mismas.

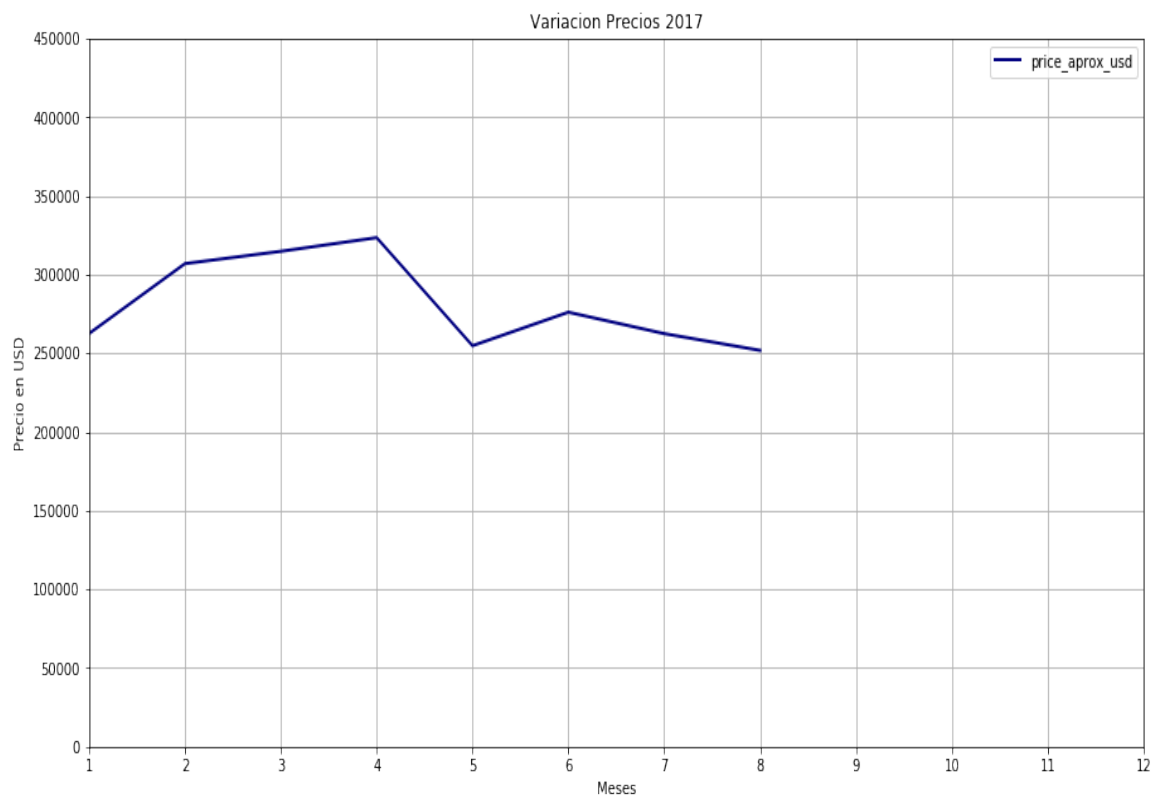
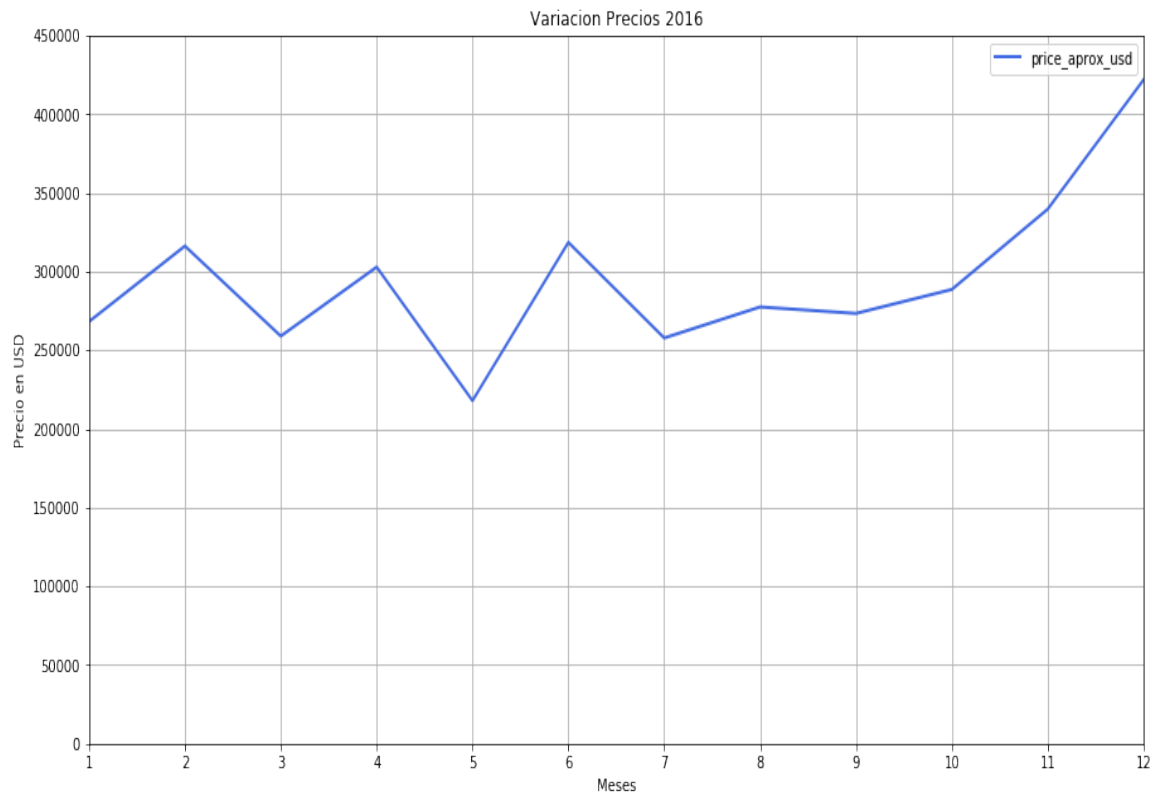
Por último, algo importante a tener en cuenta es que el set de datos esta referido a publicaciones y no a ventas de propiedades en sí. Dichas publicaciones dejan de existir una vez que la propiedad fue vendida, por lo que los datos del 2013-2014 tal vez no sean lo suficientemente representativos por la cantidad de entradas que poseen, como si lo son los datos del 2015 en adelante. De todas maneras, se usaron con fines representativos en el siguiente análisis.

6.3. Presentación de los gráficos de promedios

Aquí se muestran los gráficos que representan los promedios anteriormente mencionados para cada año. Vale aclarar que en los siguientes 5 gráficos el eje x representa los meses de cada año, y el eje y representa el precio de las propiedades en USD. Dicho precio varia entre 0–450000 *usd*, para obtener una mirada objetiva de los gráficos que serán presentados acontinuación.







6.4. Ubicación de las propiedades con mayor precio

Luego de presentados los gráficos que representan la fluctuación del precio a través de los años, se procede a presentar una serie de gráficos con el objetivo de analizar la ubicación de las propiedades con mayor precio.

Para esto, se representa en distintos **Heats Map** la ubicación de las propiedades con mayor precio en cada año, es decir que se filtro el mes que poseía un máximo en cada gráfico anterior y se representaron todas las propiedades que estaban en dicho mes.

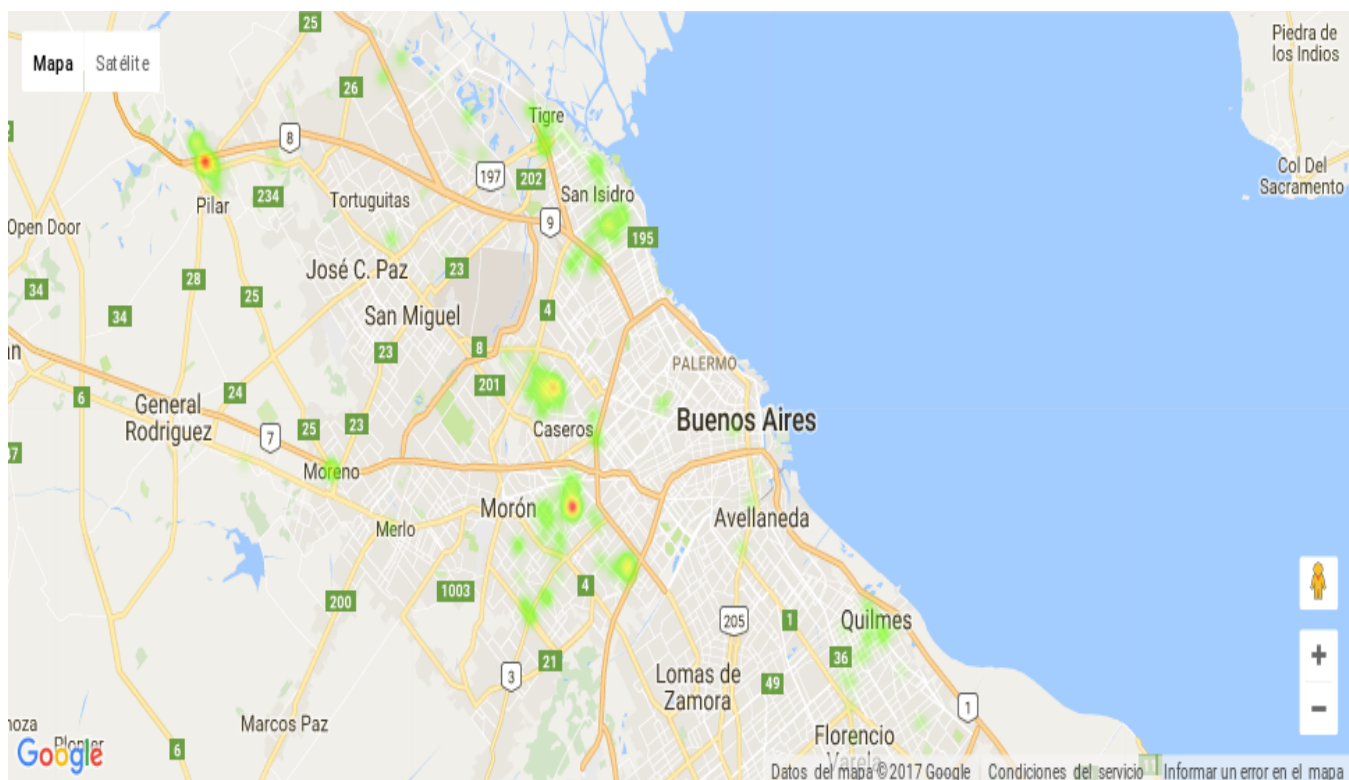


Gráfico de las propiedades con mayor precio en el 2013.

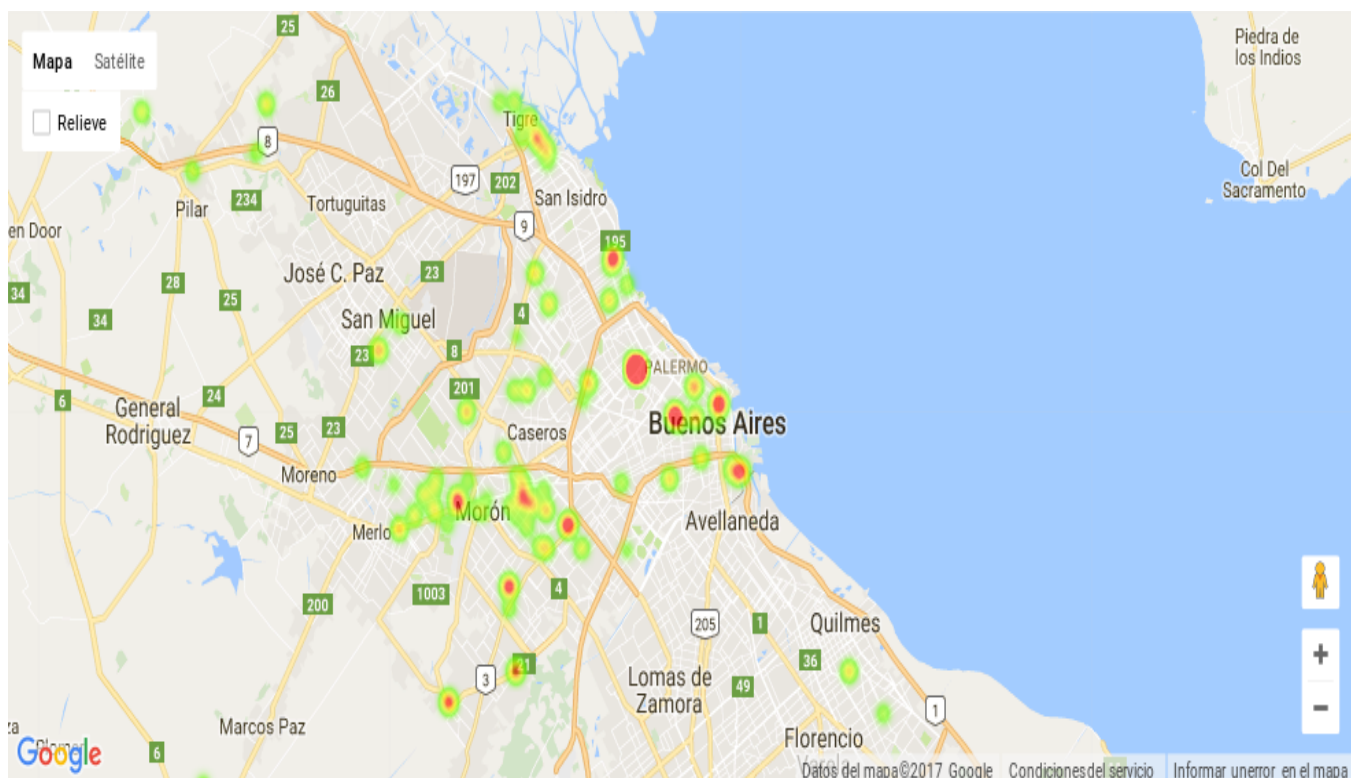


Gráfico de las propiedades con mayor precio en el 2014.



Gráfico de las propiedades con mayor precio en el 2015.

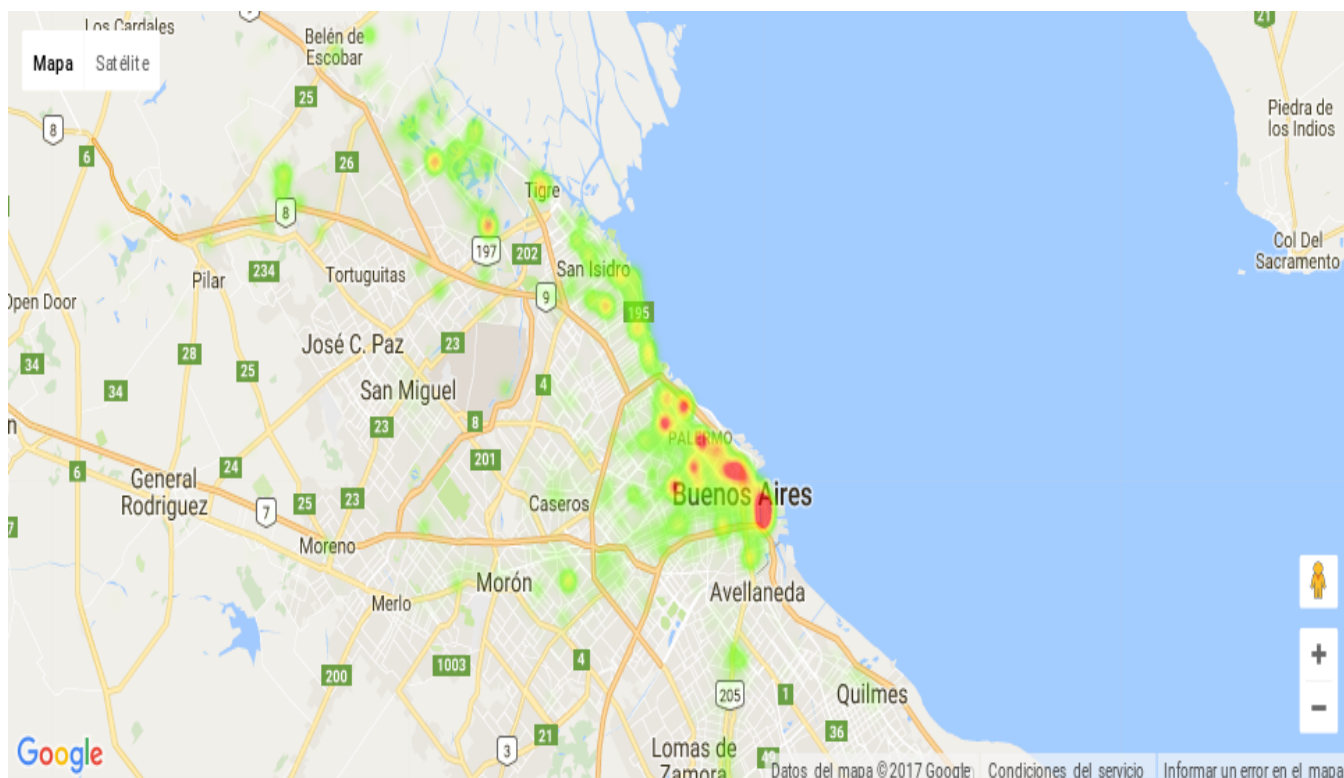


Gráfico de las propiedades con mayor precio en el 2016.

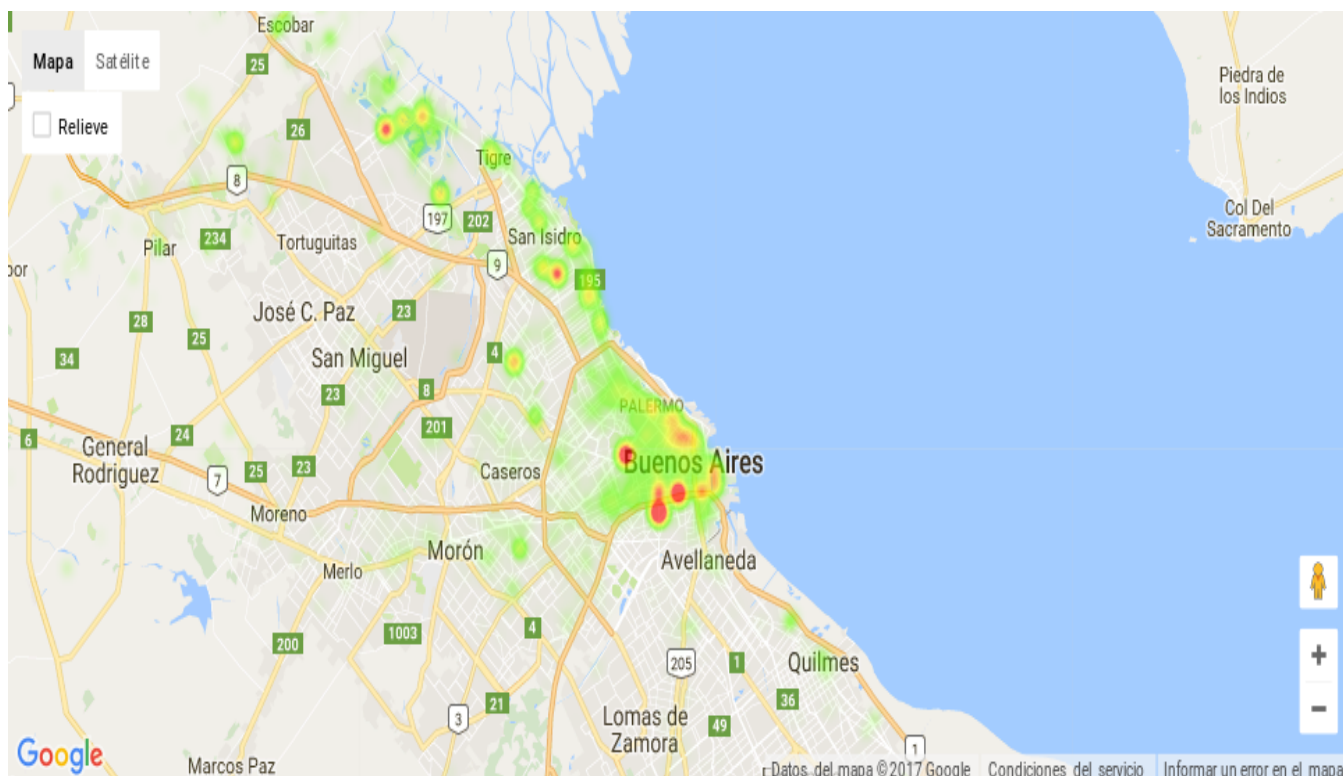


Gráfico de las propiedades con mayor precio en el 2017.

6.5. Conclusiones de la Fluctuación del precio aproximado.

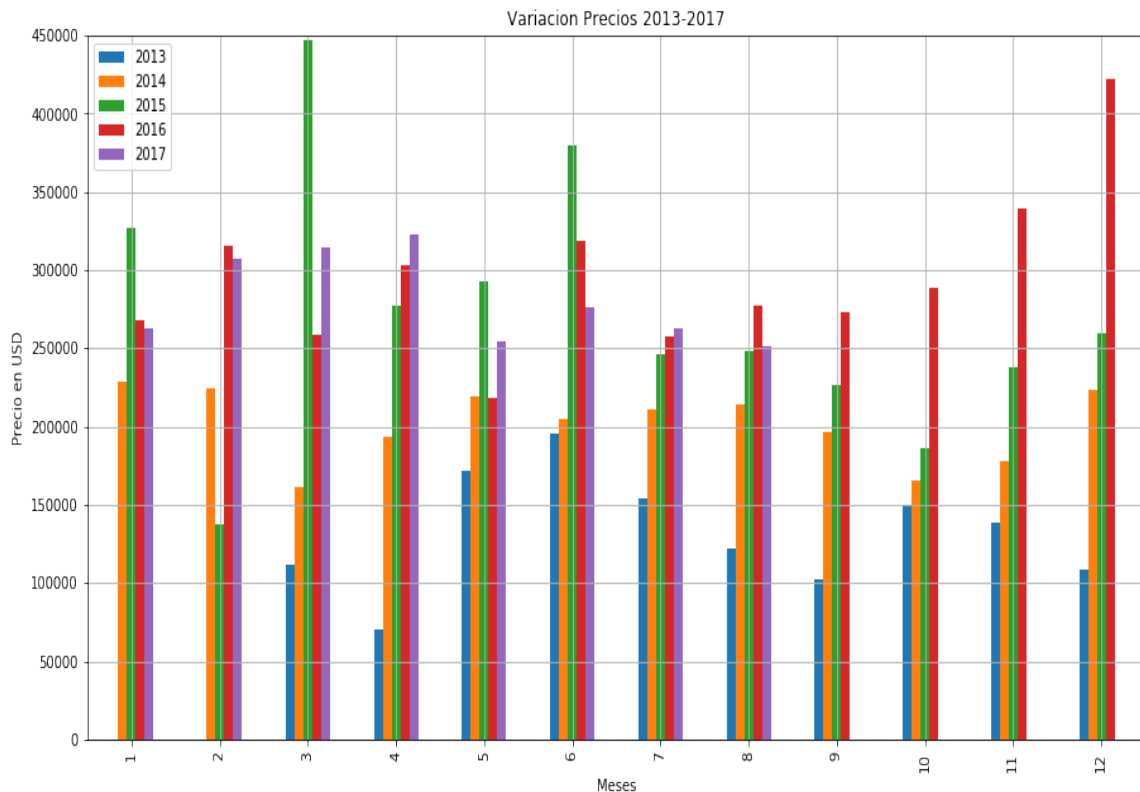
De los gráficos de promedios presentados anteriormente, se pueden visualizar rápidamente los meses con mayor y menor precio de propiedades en cada año. Se presenta una tabla con dichos datos:

Límites de precios por años.		
Año	Mes con Mayor Precio	Mes con menor Precio
2013	Junio	Abril
2014	Enero	Marzo
2015	Marzo	Febrero
2016	Diciembre	Mayo
2017	Abril	Agosto

Cabe mencionar que estos son los datos que fueron filtrados para poder realizar los **Heats Map**.

Por otra parte, se puede ver en la tabla presentada anteriormente que no se puede sugerir una tendencia de meses con mayores-menores precios, ya que los resultados fueron muy variables.

Con el objetivo de brindar otra mirada de la diferencia general entre los precios de distintos años, se construyó otro gráfico que representa los mismo promedios anteriormente mencionados, pero organizados de distinta manera para poder que puedan ser comparados entre distintos años. A continuación presentamos el gráfico:



Como se menciono anteriormente, este gráfico tiene la finalidad de poder brindar una mirada global de los precios a lo largo de los años. Como se puede ver, tenemos en el eje x los meses y para cada mes representamos los 5 años a analizar. Observando detenidamente, se pueden extraer muchas conclusiones.

Lo primero y mas sencillo de ver son los meses con mayor y menor precio de los 5 años. Se puede observar que el mes de Marzo del 2015 fue el que tuvo precios mas elevados de ventas, y por el contrario, el mes de Abril del 2013 fue el que tuvo los precios mas bajos en las mismas. Ademas se puede ver que 2013 tuvo menor precio que el resto de los años en todos los meses. Siguiendo la misma linea, es fácil observar que 2014 es el segundo año con menores precios, salvo en febrero que supero a 2015. De todas formas, se debe considerar que los años recientemente mencionados no son del todo representativos, ya que no se cuenta con la misma cantidad de datos que para los años posteriores. Por otro lado se puede visualizar que los primeros meses de 2015 fueron muy variantes, teniendo su mínimo en Febrero y máximo en Marzo, pero a partir de Junio-Julio se establece una media que supera a los 2 años anteriormente mencionados. Por último se puede ver que el año 2016 en la primera mitad posee precios que se sitúan entre los mas elevados, y en la segunda mitad establece una tendencia creciente la cual supera al resto de los años. A pesar de que los últimos años poseen un precio promedio mas elevado que los años previos, no se puede inferir que haya aumentado el precio en el correr de los años, debido a la abrupta diferencia entre la cantidad de datos. De todas formas, si solo es considerado el periodo 2015-2017, el cual posee la cantidad de entradas necesarias como para ser representativo, obtenemos los siguientes resultados:

Promedios Anuales de Precios.	
Año	Promedio(USD)
2015	292504
2016	315746
2017	275694

De lo cual se puede ver rápidamente que el año 2016 fue el que tuvo precios mas elevados. De todas formas, hay que considerar que el año 2017 solo tiene entradas hasta agosto.

Por otra parte, siguiendo el análisis recientemente hecho, se puede observar una cierta tendencia en los **Heats Maps** presentados anteriormente. En dichos gráficos se puede notar que en los primeros años hay un predominio de propiedades en *Zona Norte* y *Zona Oeste* del *Gran Buenos Aires*. Esto puede deberse, a la baja cantidad de entradas de dichos años, ya solo quedan las propiedades que no fueron vendidas. Siguiendo con este análisis, se puede observar que los años posteriores poseen un mayor porcentaje de viviendas con mayor precio en la zona de *CABA*. Ademas, a esto le podemos sumar el análisis realizado en la sección 3.1, sobre los barrios con mayor **Precio por m²**. En ese análisis se obtuvo que los barrios con mayor valor de metro cuadrado, están ubicados en *CABA*, y mas precisamente, están ubicados en la zona que los últimos 2 **Heats Maps** nos indican mayor precio en las propiedades.

7. Conclusiones generales

Para finalizar esta parte, reflexionaremos sobre la información obtenida respecto del precio del m^2 en la Ciudad Autónoma de Buenos Aires y el Gran Buenos Aires.

Por empezar, es importante mencionar por qué esta es la parte uno; el precio del metro cuadrado será nuestra, por decir de alguna manera, *unidad de medida* a lo largo de todo el trabajo. Por esta razón, parece correcto conocer cómo es que se distribuye en el espacio y cómo varió en los últimos años al inicio.

Conociendo la Ciudad de Buenos Aires y el momento actual del país, los resultados fueron los esperados. Es decir, la distribución espacial de los precios por m^2 es la esperada (disminuye a medida que nos alejamos de las conocidas 'zonas caras' de la ciudad y describe a los llamados cordones del conurbano) y el precio en los últimos años tiene una tendencia creciente, tanto por m^2 como el precio total.