



Organización de Datos

Segundo Cuatrimestre 2017

Trabajo Práctico 1

Análisis Exploratorio de Properati

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 de la relación de las características de las propiedades con el precio/-superficie	1
1. Preparación y Recolección de Datos vía URLs	1
2. Análisis de Cuartos por Propiedad	2
2.1. Relación entre cantidad de habitaciones y el precio	2
2.2. Relación entre cantidad de habitaciones y la superficie	2
2.3. Conclusiones	3
3. Análisis de Ambientes por Propiedad	3
3.1. Relación entre los ambientes y el precio	3
3.2. Relación entre los ambientes y la superficie	4
3.3. Conclusiones	5
4. Conclusiones generales	5
II Análisis con Google Places	5
5. Análisis Instituciones Educativas	6
5.1. Análisis del Precio por Metro Cuadrado	6
5.2. Análisis de la Superficie Total de la Propiedad en Metros Cuadrados	15
6. Análisis de Locales Gastronómicos	19
6.1. Análisis del precio por metro cuadrado	19
7. Análisis de Puntos de Interés Cultural	23
7.1. Análisis del precio por metro cuadrado	23
8. Análisis de Transporte Público	29
8.1. Análisis del precio por metro cuadrado	29
9. Conclusiones Generales	41

Parte I

Análisis de la relación de las características de las propiedades con el precio/superficie

El capítulo a presentarse a continuación, tiene el objetivo de analizar las relaciones posibles entre las distintas características de una propiedad, como por ejemplo habitaciones o ambientes, con el precio o superficie de dicha propiedad.

1. Preparación y Recolección de Datos vía URLs

El set de datos utilizado en primera instancia fue filtrado por las propiedades que no se pudo conseguir:

- *Precio Aproximado en usd*
- *Precio por m² y superficie cubierta*
- *La latitud y longitud*
- *Geonames*

Luego de realizar este filtrado, lo primero que se hizo fue parsear las descripciones y títulos que aparecían en el set de datos. Aquí apareció el primer problema, ya que dichas descripciones/títulos estaban escritos por las personas que realizaban la publicación, por lo tanto no se poseía una entrada estandarizada, sino que una palabra a buscar podía estar escrita de muchas formas distintas. Por ejemplo, si se quería obtener la cantidad de ambientes de una propiedad, podía encontrarse como:

- *Ambientes*
- *Ambiente*
- *Ambientes:*
- *Amb*
- *Amb.*

Cabe aclarar, que solamente se mencionan algunas posibilidades con fines ilustrativos. Por otra parte, además de la falta de estandarización que se tenía en las descripciones/títulos, también se tenía el problema del orden en el que se encontraba la información, y su formato. Es decir, la información que se quería encontrar podía aparecer en distintos formatos y distinto orden. Por ejemplo, para la cantidad de ambientes se podía encontrar:

- *Amb: 2*
- *Amb: dos*
- *2 Ambientes*
- *dos amb*
- *dos ambientes*

Nuevamente solamente se mencionan algunas combinaciones con fines ilustrativos. En definitiva, si se tiene en cuenta los dos problemas anteriores, se vuelve muy tedioso tener que considerar las distintas combinaciones. Por último y no por eso menos importante, aparecía los errores de teclado de la persona que había subido la publicación. De todas formas, se parsearon las descripciones y títulos de una manera no tan exhaustiva en búsqueda de cantidad de ambientes y cuartos, con el fin de completar lo más posible el set de datos.

Por otro lado, con el objetivo de obtener una entrada más estandarizada, se notó que en el set de datos aparecían los links URL de las publicaciones. En dichas publicaciones la entrada era totalmente estandarizada. Siguiendo con el ejemplo de ambientes, se podía obtener como *Ambientes: n*, siendo n la cantidad de ambientes. Entonces se optó por intentar obtener la información de las páginas web. De todas formas, aquí se presentaron 3 problemas: el primero es que obtener la información desde la página web era muy lento, el segundo fue que había publicaciones que habían sido quitadas porque ya se había realizado la venta y por último es que la página web podía no contener la información que se buscaba. Igualmente se prefirió lidiar con estos dos problemas, antes que la falta de estandarización de las descripciones.

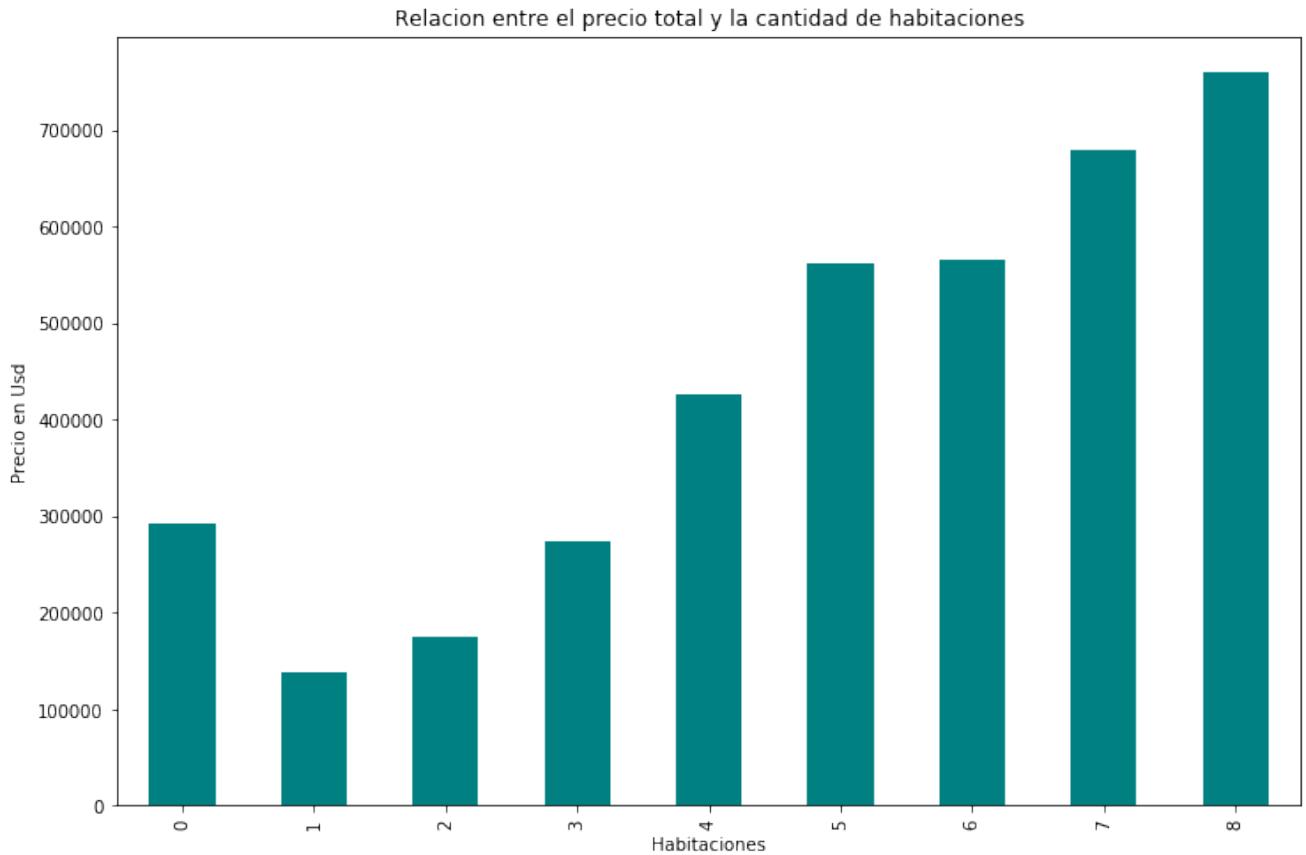
Teniendo en cuenta los problemas mencionados anteriormente, para llevar acabo el proceso de obtener la información desde las páginas web, se realizaron distintas pasadas. La primera pasada, bajaba la información de la página web de los datos que no poseían cantidad de ambientes, ni cantidad de habitaciones. Cabe aclarar, que se lo que se bajó fue todo el código de la página web, es decir que dicha información debía ser parseada. Luego de parsear esta información, se realizó la segunda pasada que bajaba la información de los datos que no poseían cantidad de ambientes o cantidad de habitaciones. Aquí notamos que tanto la cantidad de cuartos como la cantidad de ambientes estaban completas en más del 60% y por falta de tiempo se decidió trabajar con estos datos.

2. Análisis de Cuartos por Propiedad

Esta sección se presenta con el objetivo de analizar la posible relación que puede existir entre la cantidad de Habitaciones de una propiedad, con el precio y/o la superficie.

2.1. Relación entre cantidad de habitaciones y el precio

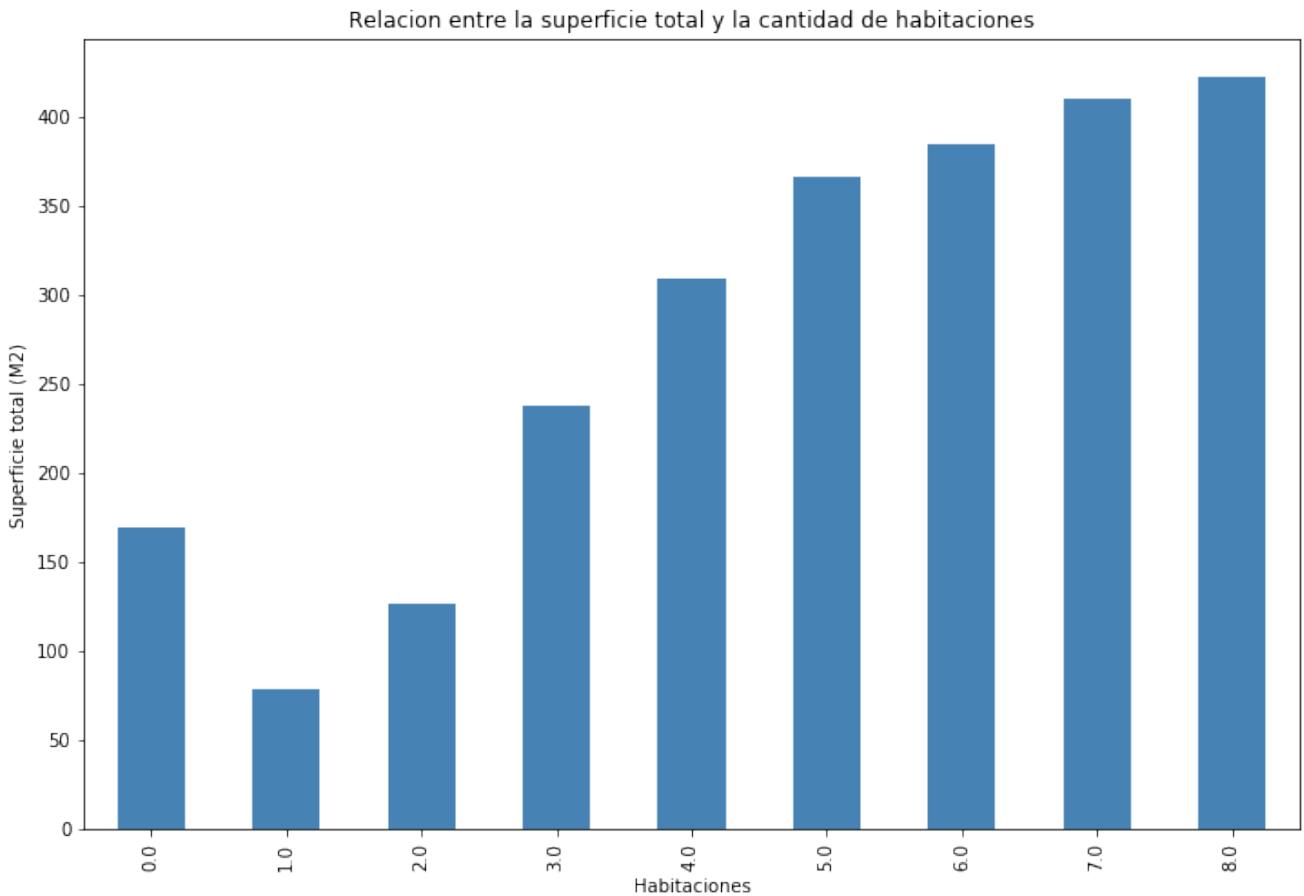
Aquí se estará analizando la relación entre la cantidad de habitaciones y el precio aproximado de una propiedad. Lo que se hizo para esta sección fue agrupar por cantidad de habitaciones, dejando el precio promedio como valor. Con dichos datos se obtuvo el siguiente gráfico:



2.2. Relación entre cantidad de habitaciones y la superficie

En esta sección se analizara la relación que se puede establecer entre la cantidad de habitaciones y el precio aproximado de una propiedad.

Para esto, se siguió un proceso similar a la sección anterior. Se agrupo por cantidad de habitaciones y se dejó el promedio de la superficies como valor. Se presenta el siguiente gráfico que representa dichas relaciones:



2.3. Conclusiones

Como se puede observar en el gráfico de habitaciones vs precio, este sigue una tendencia creciente, es decir que a medida que aumenta la cantidad de habitaciones aumenta el precio de dicha propiedad.

Se puede realizar el mismo análisis para la cantidad de habitaciones vs superficie. Se puede observar claramente que a medida que aumenta la cantidad de habitaciones, también lo hace la superficie de la propiedad.

Estos resultados son totalmente esperables. Además se puede ver que ambos gráficos siguen la misma tendencia de crecimiento, lo cual nos permite inferir que hay una estrecha relación entre las 3 variables analizadas: cantidad de habitaciones, precios y superficie.

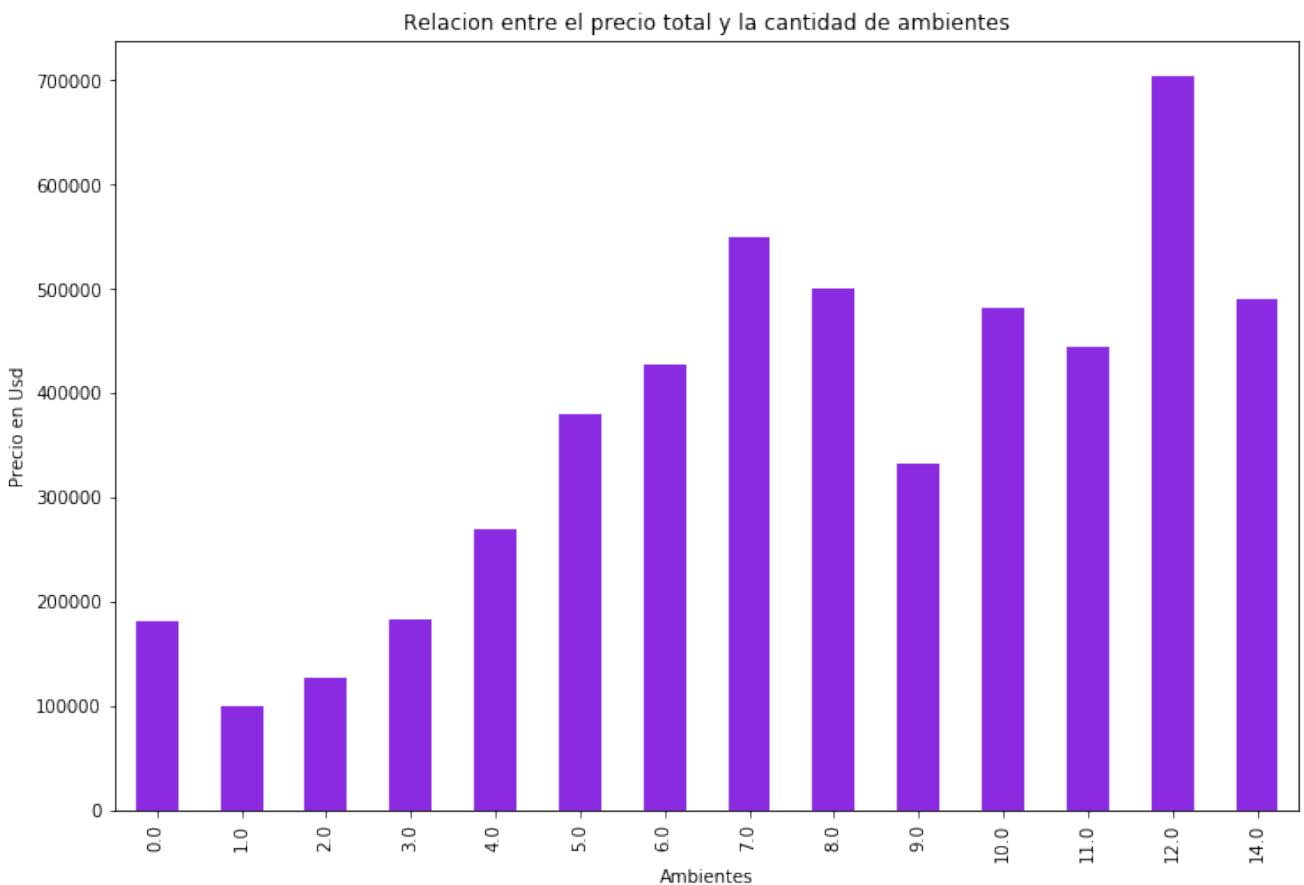
3. Análisis de Ambientes por Propiedad

En esta sección se estará analizando la posible relación entre la cantidad de ambientes de una propiedad con el precio y/o la superficie.

3.1. Relación entre los ambientes y el precio

El objetivo de esta sección es analizar la relación entre la cantidad de ambientes y el precio de una propiedad. Al igual que en la sección anterior, lo que se hizo fue agrupar con cantidad de ambientes, dejando como valor el precio promedio.

A continuación se presenta un gráfico que representa dichas relaciones:

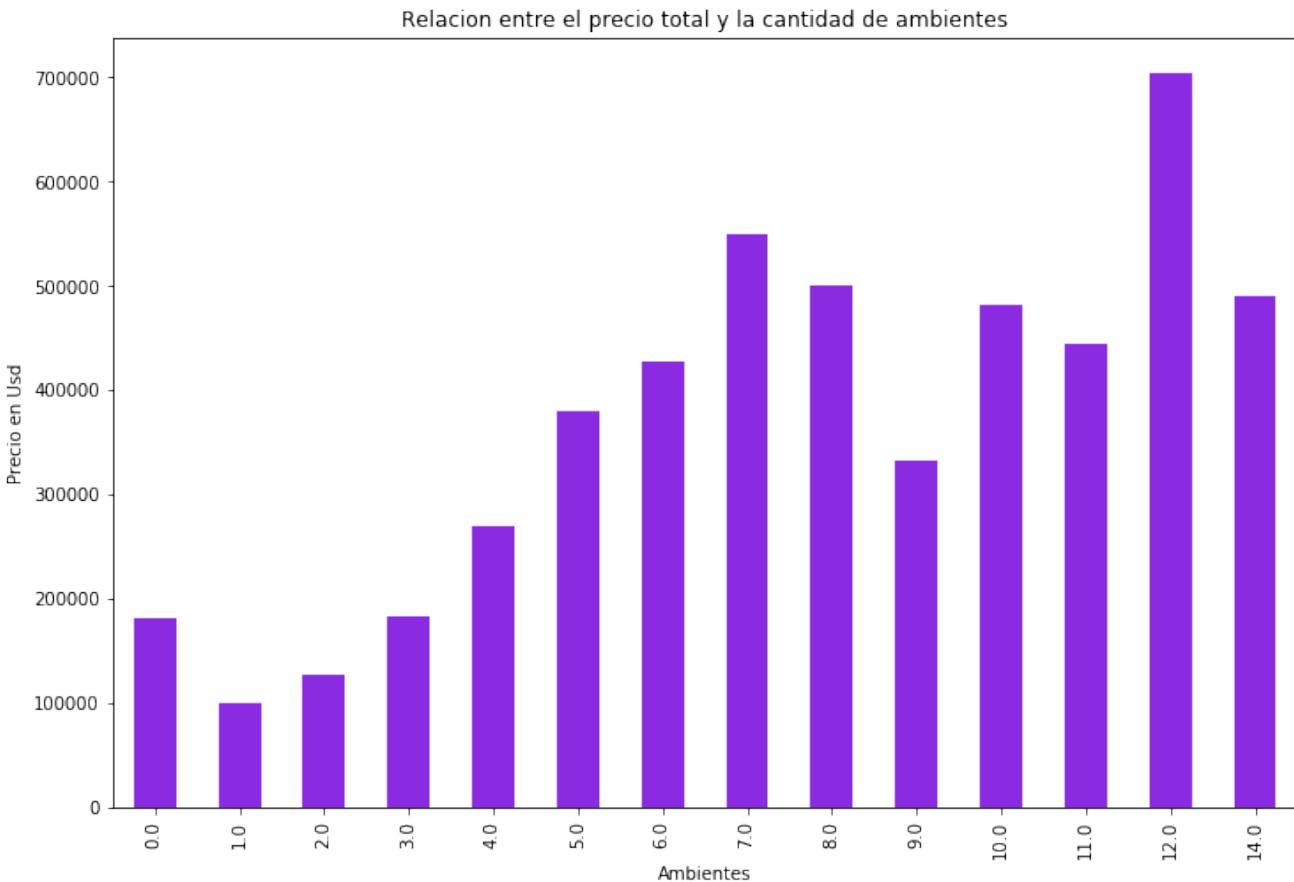


3.2. Relación entre los ambientes y la superficie

Esta sección se presenta con el objetivo de analizar la relación entre los ambientes y la superficie de una propiedad.

Para presentar los datos, lo que se hizo fue agrupar por cantidad de ambientes, y se dejó el precio promedio como valor.

Se pasa a presentar un gráfico con los resultados:



3.3. Conclusiones

Como se puede ver en el gráfico que representa ambientes vs precio, se presenta una tendencia creciente hasta 8 ambientes. Luego, los precios son muy variables dependiendo de la cantidad de ambientes.

Lo mismo ocurre con el gráfico de ambientes vs superficie, en el cual se puede observar una tendencia creciente hasta los 8 ambientes, y luego la superficie varía de acuerdo a la cantidad de ambientes.

Lo que se puede decir es que el precio depende fuertemente de la superficie de una propiedad. Se puede ver que en propiedades con mayor cantidad de ambientes, como por ejemplo las propiedades que tienen mas de 8 ambientes, el precio es acorde a la superficie de dichas propiedades.

4. Conclusiones generales

Se puede observar que en el análisis anterior los gráficos siguen la misma tendencia creciente hasta las 8 habitaciones y ambientes. Entonces, se establece cierta correlación entre los ambientes y las habitaciones.

Pero de todas formas, como se vio anteriormente, se puede notar que una vez que se superan los 8 ambientes, no se obtiene una tendencia creciente. Por lo tanto, se nota que se establece una relación muy fuerte entre el precio y la superficie de las propiedades.

Parte II

Análisis con Google Places

En este segmento se utilizó la API de Google Places¹ para obtener información adicional sobre las propiedades provistas por Properati. Esta API permite entre otras cosas buscar sitios (definidos en esta API como establecimientos, ubicaciones geográficas o puntos de interés destacados) dentro de un área definida, como los límites de un mapa o alrededor de un punto fijo. Para poder utilizarla sin embargo, se necesitan de coordenadas del tipo latitud y longitud.

El set de datos provistos contiene algunas ubicaciones en este formato pero no todas. Por suerte dentro del set de datos está la información denominada como “geoname_id”. Este tipo de identificación pertenece a una base de

¹<https://developers.google.com/places/?hl=es-419>

datos geográfica con mas de 10 millones de entradas únicas. Mediante la base de datos completa descargada de Kaggle ² se completaron las latitudes y longitudes faltantes para así tener una muestra completa mayor.

Para realizar pedidos a Google Places es necesario tener una API Key que esencialmente controla el tráfico diario de pedidos a la API. Esta limitación, utilizando una cuenta prioritaria, es de 150000 pedidos por día. A la vez el tiempo entre que se realiza un pedido y se obtiene su respuesta es muy alto. Es por esto que se limitaron las búsquedas a las siguientes categorías(cantidades por propiedad en un radio de 400 metros):

- *Locales de tipo gastronómico*

Categorías de Google Places: FOOD, BAKERY, BAR, MEAL_TAKEAWAY, MEAL_DELIVERY, RESTAURANT

- *Instituciones educativas*

Categorías de Google Places: SCHOOL, UNIVERSITY

- *Puntos de interés cultural*

Categorías de Google Places: ART_GALLERY, MUSEUM, PLACE_OF_WORSHIP

- *Espacios verdes*

Categorías de Google Places: PARK, NATURAL_FEATURE

- *Estaciones/Paradas de transporte público*

Categorías de Google Places: BUS_STATION, SUBWAY_STATION

5. Análisis Instituciones Educativas

A continuación se encuentra el análisis de las instituciones educativas para un set de datos completo de 72474 entradas.

5.1. Análisis del Precio por Metro Cuadrado

Para los siguientes gráficos se agruparon las propiedades por cantidad de instituciones cercanas y sacando el promedio del precio por metro cuadrado para ellas. Para que los resultados sean significativos, se filtro a todas los conjuntos de propiedades con menos de 300 entradas para las cantidades de instituciones cercanas. Por ejemplo: Si solo hay 15 propiedades con 60 instituciones cercanas, estas no se tuvieron en cuenta para el análisis.

²<https://www.kaggle.com/geonames/geonames-database>

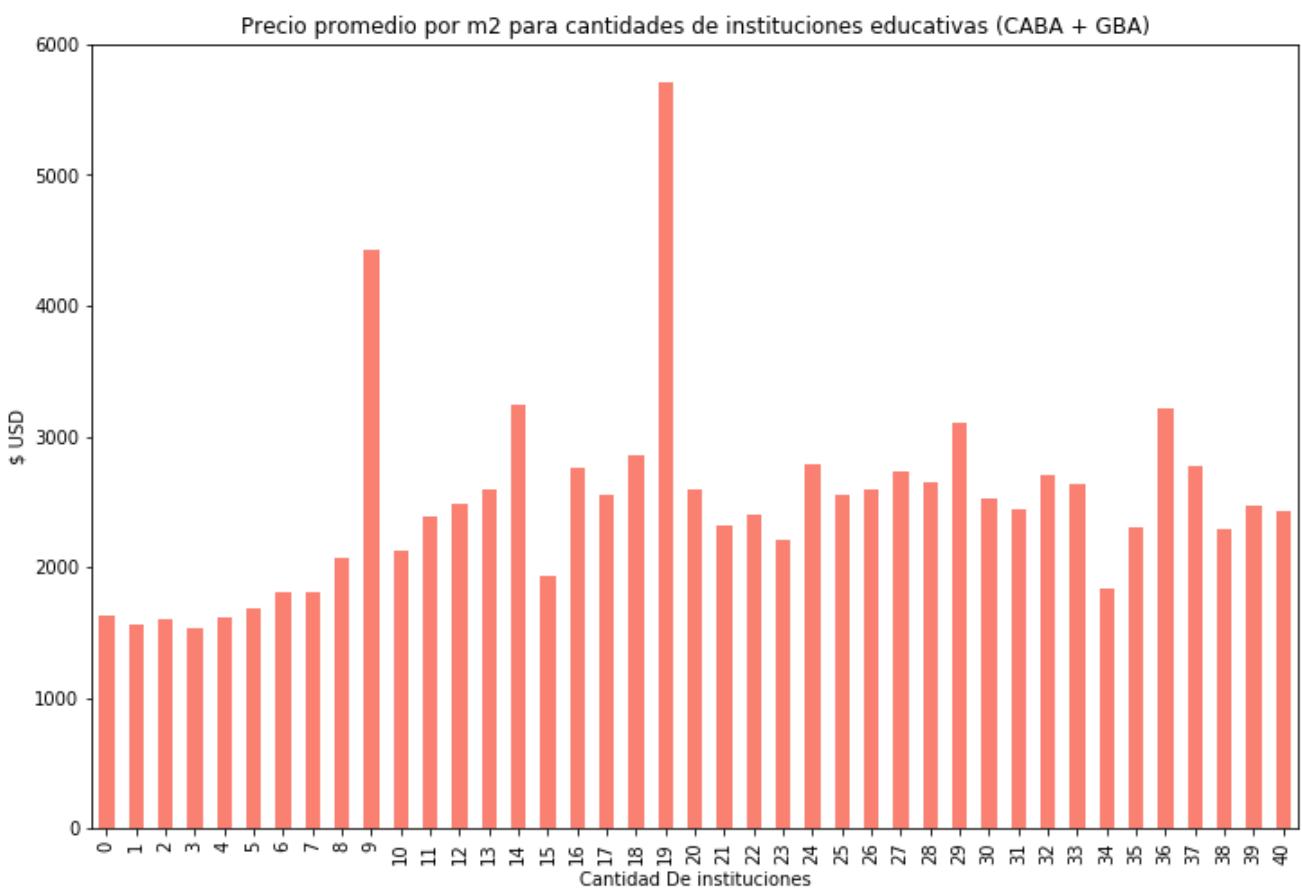


Figura 1: CABA + GBA

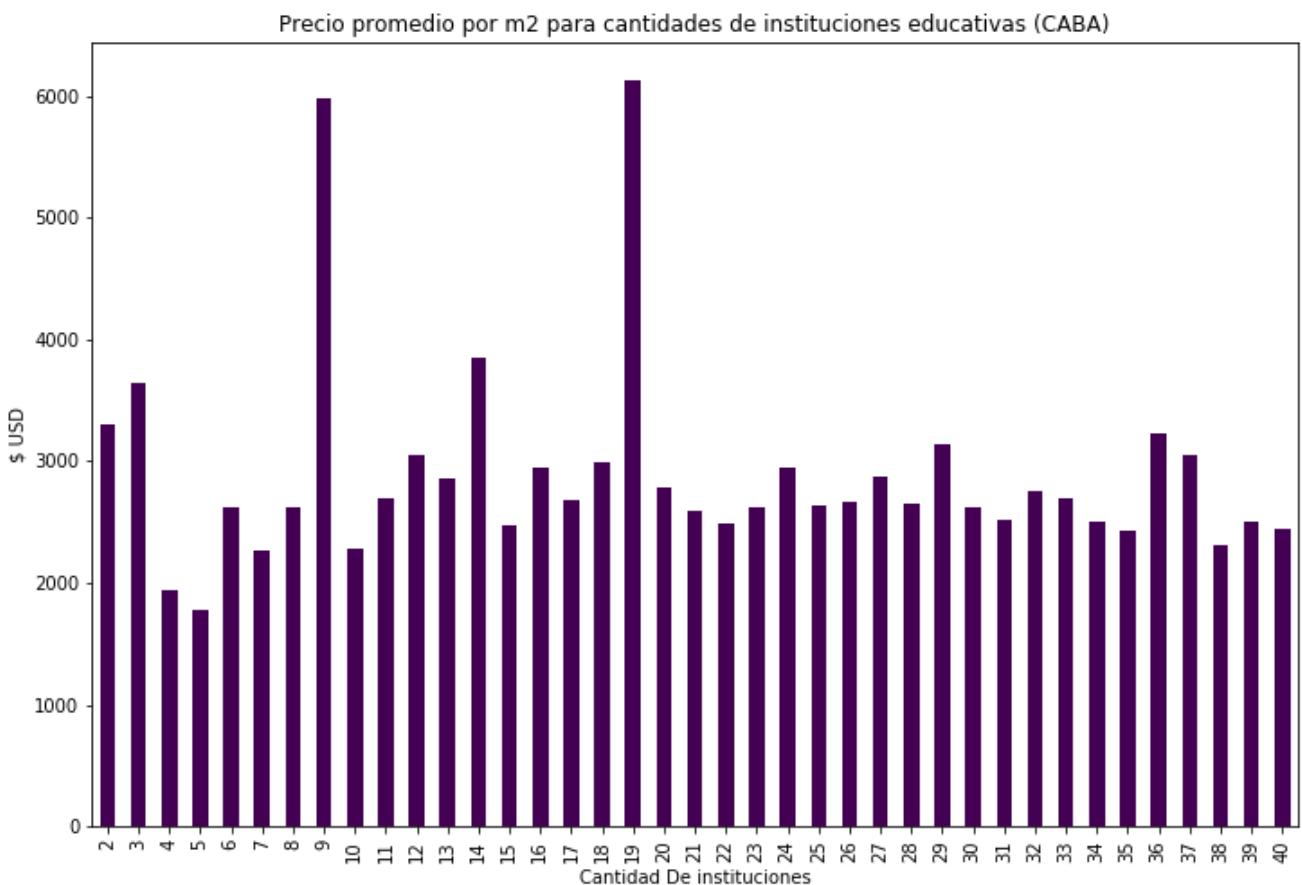


Figura 2: CABA

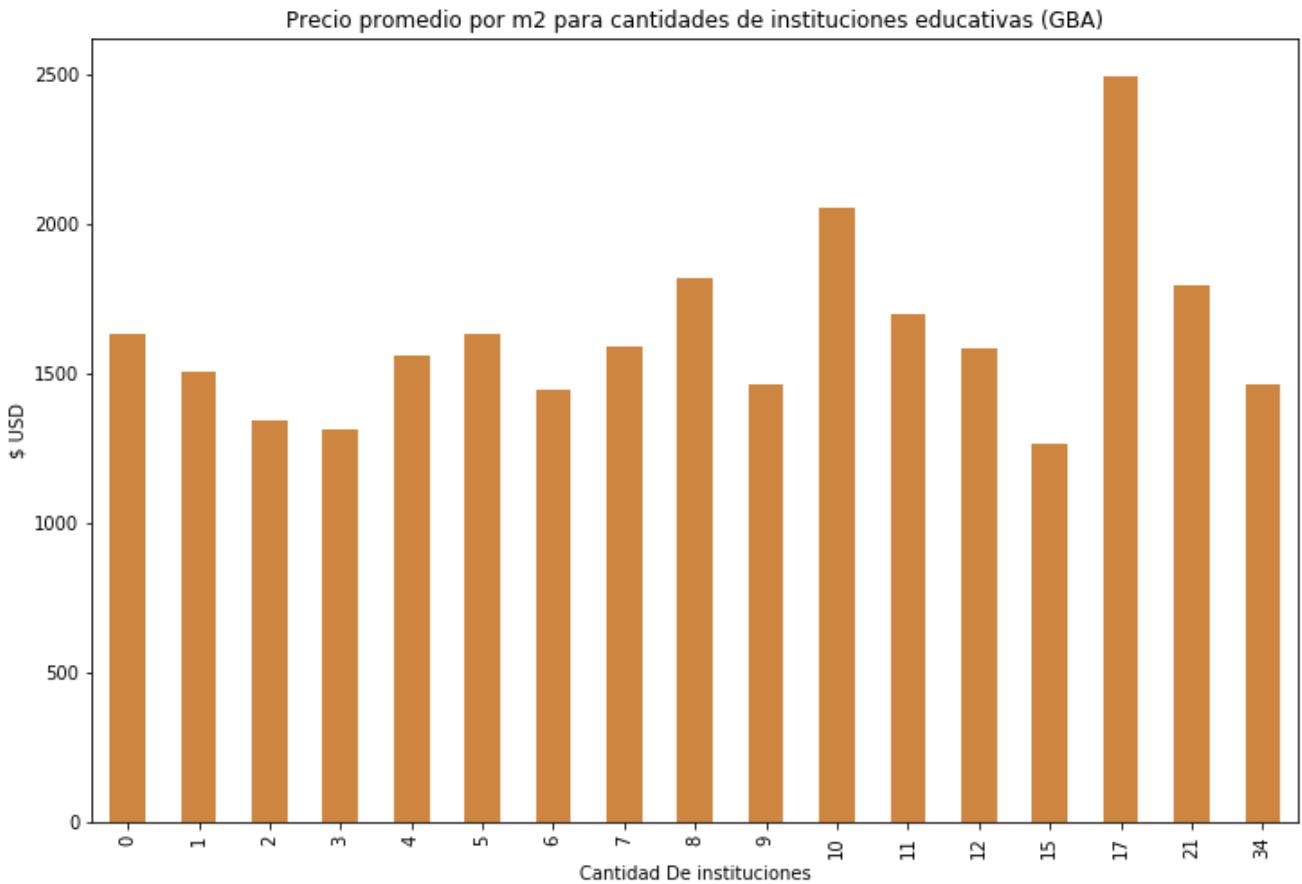


Figura 3: GBA

Como se puede observar en los tres gráficos, hay una tendencia de un pico máximo en todos para una cantidad de instituciones cercana a 20 instituciones. Es importante notar sin embargo, que dichos máximos son muy distintos siendo el de CABA casi el triple que el de GBA (6000 contra 2400 respectivamente).

Se procede a aislarlas y observar en que zonas se concentran.

HEATMAP por ubicaciones



Figura 4: CABA con 19 instituciones

HEATMAP por precio del metro cuadrado



Figura 5: CABA con 19 instituciones

HEATMAP por ubicaciones



Figura 6: CABA con 9 instituciones

HEATMAP por precio del metro cuadrado



Figura 7: CABA con 9 instituciones

HEATMAP por ubicaciones



Figura 8: GBA con 17 instituciones

HEATMAP por precio del metro cuadrado



Figura 9: GBA con 17 instituciones

¿Que conclusiones se obtienen de estos heatmaps?

Se pueden ubicar polos en donde hay una cantidad importante de instituciones educativas (entre 17 y 20 en un rango de 400 metros de radio) y a la vez los precios por metro cuadrado son altos en comparación a la totalidad de las propiedades. En ellos se genera una tendencia que podría marcar una relación directa entre estos dos factores. Estos lugares son:

- Olivos
- San Cristóbal
- Villa Urquiza/General Urquiza
- Parque Patricios
- Colegiales

A menor escala, zonas céntricas de los siguientes barrios:

- Merlo
- San Miguel
- Quilmes
- Recoleta
- Belgrano
- Puerto Madero
- Almagro

Resulta interesante notar una situación particular que ocurre en lugares como Villa Crespo y Boedo en donde el promedio de instituciones cercanas es 9 pero sin embargo hay precios muy altos en las propiedades. Esto genera una problemática en cuanto a la validez del análisis por instituciones educativas y sugiere quizás realizar un filtro más estricto en los datos para futuros análisis. Sin embargo 10 instituciones educativas en un rango de 400 metros sigue siendo un número importante.

¿Cuales son las instituciones educativas en estos polos?

A continuación se muestran los nombres de dichas instituciones por barrio (Solo de los 3 barrios más significativos) para tenerlas en cuenta en futuros análisis cuando se busque predecir precios por propiedades. Cabe aclarar que son más de 20 ya que se toman el conjunto de todas las instituciones del barrio y no solo las de una ubicación en particular.

Olivos:

- Northlands
- Colegio San Andrés secondary
- Centro Cultural Italiano - Colegio Alessandro Manzoni
- St. Andrew's Scots School
- Instituto Jesús en el Huerto de los Olivos
- Escuela Montessori Olivos SRL
- Colegio San Ignacio
- Colegio Nuestra Señora de la Paz
- St. Luke's College
- Action Integral Institute of Performing Arts
- Escuela Municipal Paula Albarracín de Sarmiento
- San Andres Secundario Olivos
- St. Nicholas College

- Instituto Superior De Musica Jose Hernandez
- Escuela EPB N° 2 “Benemérito Teniente Gral. Bartolomé Mitre”
- UCES OLIVOS
- Colegio Tarbut
- Estudio Cambrée Tatiana Flaker
- UCES UNIVERSITY OF BUSINESS AND SOCIAL SCIENCES
- Fundacion Universidad de San Isidro
- Colegio San Nicolas Primario
- COLEGIO SAN NICOLAS JARDIN
- Escuela Superior De Informatica De La Prefectura Naval Argentina
- CENTRO PAMPA / escuela de diseño
- Colegio Santa Magdalena
- Colegio Eidep
- De Los O Colegio Jesus En El Huerto
- Jardin San Ignacio
- E.M.P.A.S
- Ganesha YOGA
- Colegio Feli
- Escuela Hija St Andrews
- Auditorio niño Jesus De Praga
- Escuela Municipal De Musica
- Niño Jesús Del Praga
- Jardín Maternal Niño Jesús de Praga
- Auditorio Northlands School Olivos
- Escuela De Tomas, Francisco Borges Y Rosales
- Jardin Jho
- Jardin de infantes CCI - Centro Cultural Italiano -
- Jardin centro cultural italiano
- Colegio Centro Cultural Italiano
- SCUOLE CCI
- Jardin Jesus en el Huerto de los Olivos
- Jardin Dante
- ArtBA
- ITBA
- Jardin Maternal Osecac
- Instituto San Migue
- St. Nicholas College
- English Boutique

- Scout Huerto De Los Olivos
- Grupo Capoeira Brasil Buenos Aires GCB - La Lucila
- CID vicente lopez
- Colegio Nuestra Sra De La Paz
- Toefl
- Centro Cultural y Político Micaela García
- Centro de Instrucción Aeronáutica C.I.A.
- Escuela N 16 Marcelino Ugarte
- Escuela EST Nro 3

San Cristóbal:

- DE LAS VICTORIAS
- La Aldea del Buen Ayre
- Instit Salesiana - Colegio San Antonio
- Escuela Generación del Futuro
- Danza Árabe Escuela Aldana Arguello
- Sol de America
- Crema y Chocolate
- Crema y Chocolate
- Fundacion tomas eloy martinez
- Curso de cerrajería presencial e intensivo
- Cenedi
- AUDITORIO NAMUNCURA
- Colegio San José de Calasanz
- Colegio Calasanz
- San Antonio
- ILEC - Instituto Laico de Estudios Contemporaneos
- Special Education Institute OUR LADY OF LUJAN
- Curso Sublimacion
- Escuela Domiciliaria N 2
- JIC N 4 DE 6 MARIANO BOEDO
- Ciber Pibes
- Escuela Infantil Cyberpibes
- Instructorado De KIZOMBA
- Esc de Com N° 22 DE 6 "G. M. Zubiria "
- Espacio De Creacion Yapeyu
- ESCUELA N. 6 D.E. 8 SAN JOSÉ DE CALASANZ
- Escuela N 25 Paula A De Sarmiento
- Pasillo al fondo Centro Cultural"

- Instituto Calazans
- Escuela Lucia
- ESCUELA PAULA ALBARRACIN DE SARMIENTO
- Escuela Infantil La Torrecita
- ”Puente Azul” Jardín de Infantes
- Taller De Arte Hilodearbol
- San Antonio Salesian house
- Curso Calidad
- SANTA MARIA INSTITUTE
- Instituto San Antonio-A.226
- Escuela De TANTRACLASICO
- CFP No.30
- Yoga
- Escuela No9 D.E. 8 - Florentino Ameghino
- Supervision D E 8 Primaria

Villa Urquiza/General Urquiza:

- School No. 24 Francisco Morazan
- San Patricio Secondary Institute
- Nuevos Aires SRL
- Sir Thomas Malory
- Mad Escuela
- Sir Thomas Malory School
- Estudio Joya
- Instituto Superior del Profesorado en Educación Especial
- INA - Instituto Nuevos Aires
- St. Patrick's School
- Escuela Infantil Chiquilines
- Instituto Junín
- The Garden of the Fund
- Special Education School 11
- Burdel de maderas
- Clases de Guitarra en Villa Urquiza - Música y creatividad
- St. Patrick's School instituto San Patricio
- Saint Patrick
- Acha Club
- Caebt 56 - Parroquia Jesús Misericordioso
- St patrick's Kinder
- Naranon grupo

- Escuela Nro. 4 D.E. 15
- Escuela Nro 24 D.E.15 - Escuela Nro 8
- Ispee
- Facu Aye
- Facultad Moron
- ESc Infantil N 8 DE 15
- San Pablo
- Island of My Dreams
- escuela republica de costa rica
- Escuela No 24 SIGLO XXI
- Escuela n•15 acevedo
- Universidad -ciclo basico
- Drago Uba
- CBC Drago
- Colegio Franco
- UBA - Drago
- UBA Sede Drago
- Cbc
- CBC UBA - Sede Drago
- Sede Drago

Es importante notar, como se puede ver en las ultimas instituciones de Villa Urquiza, la sede Drago del CBC aparece subida repetidas veces pero escrita de distinta manera. Esto muestra que mas allá del gran poder que tiene Google Places, los datos pueden no ser del todo fehacientes.

5.2. Análisis de la Superficie Total de la Propiedad en Metros Cuadrados

A partir de este análisis se busca encontrar alguna relación entre el tamaño de la propiedad y la cantidad de instituciones en su cercanía. Previo a los resultados se supone que puede llegar a haber una relación teniendo en cuenta que mientras mas grande sea, es mas probable que mas personas vivan allí y por consiguiente necesiten de variadas instituciones educativas. A la vez se podría dar también que pequeñas propiedades estén en zonas donde la demanda de instituciones educativas se muy alta y por esta razón priorizar la cercanía a las instituciones dejando de lado otras comodidades como puede ser un mayor espacio.

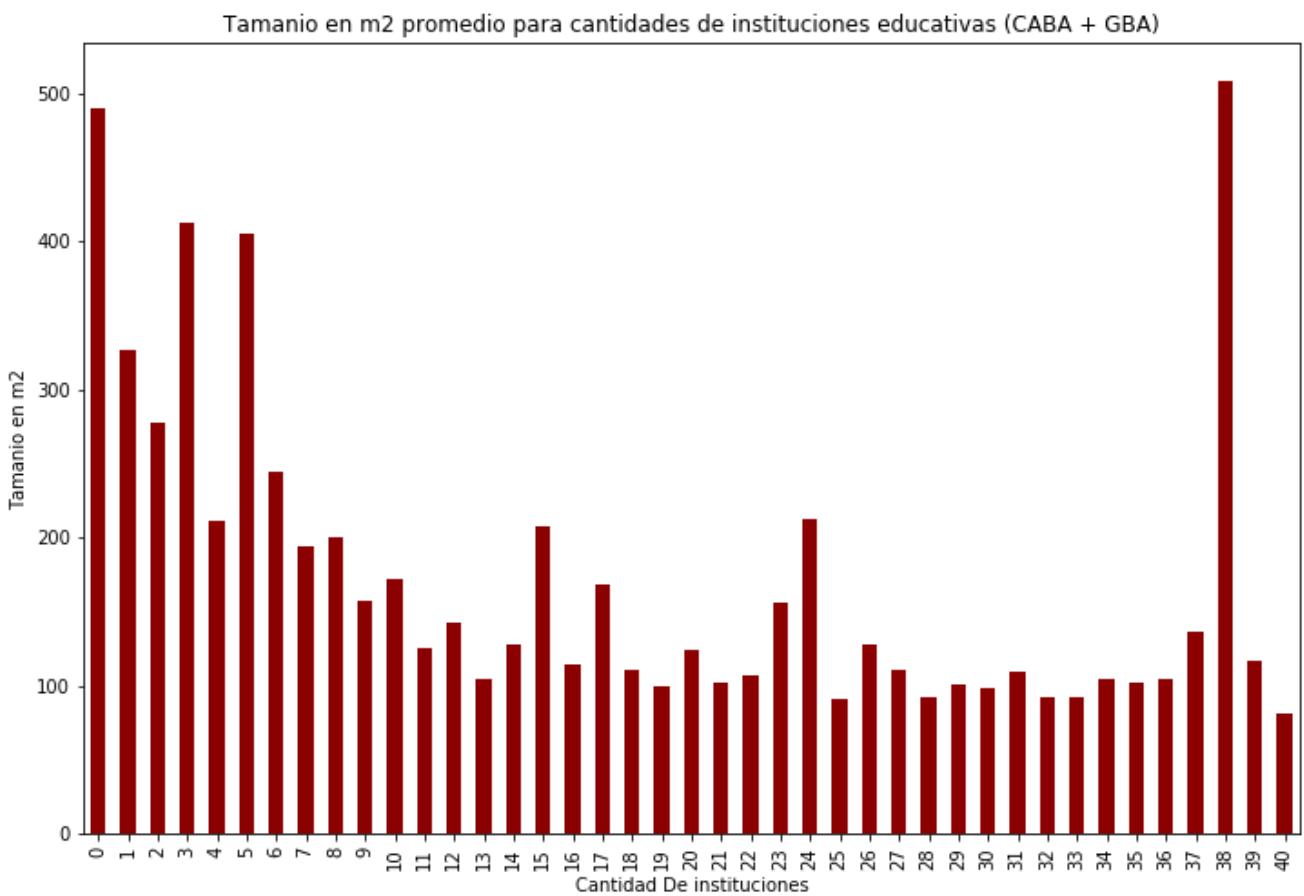


Figura 10: CABA + GBA

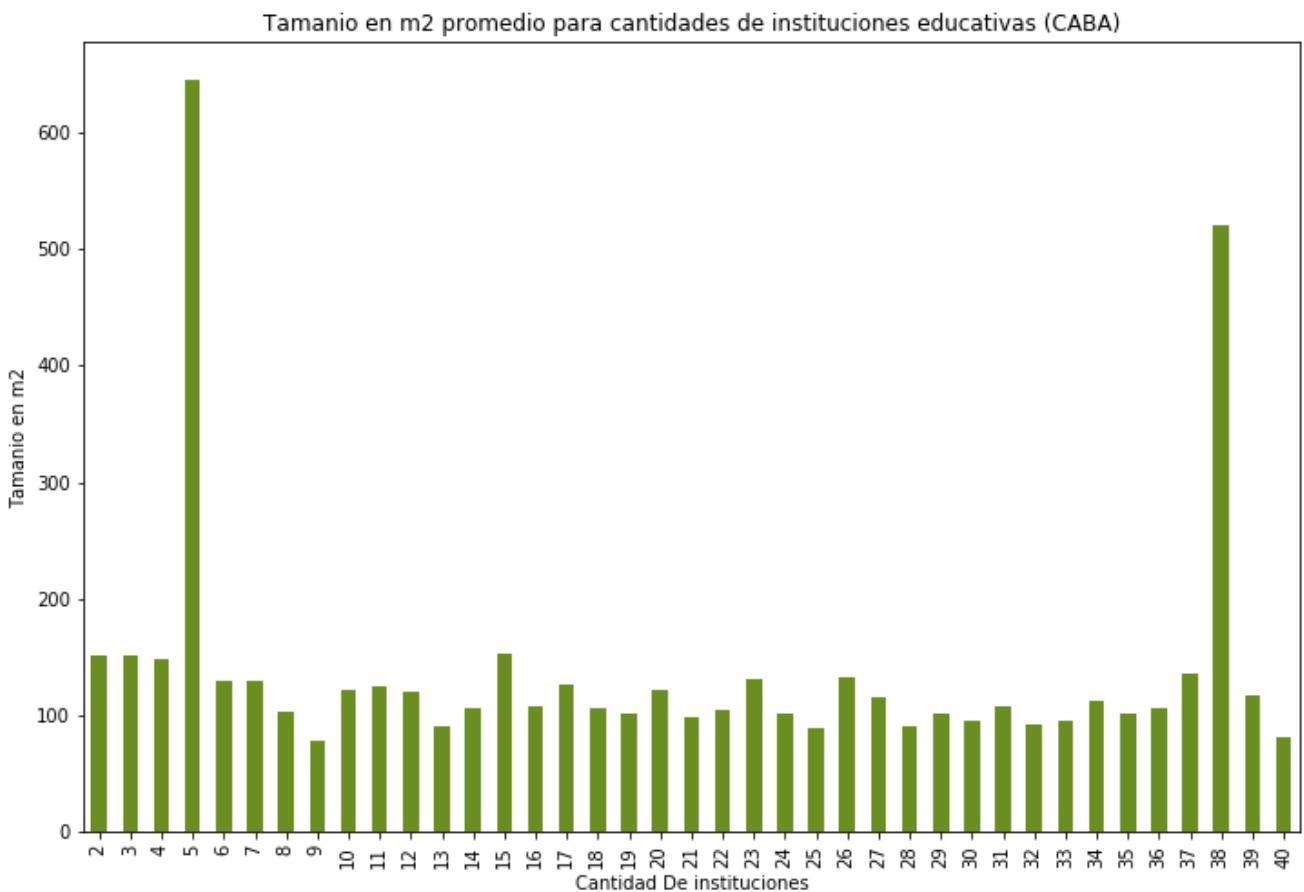


Figura 11: CABA

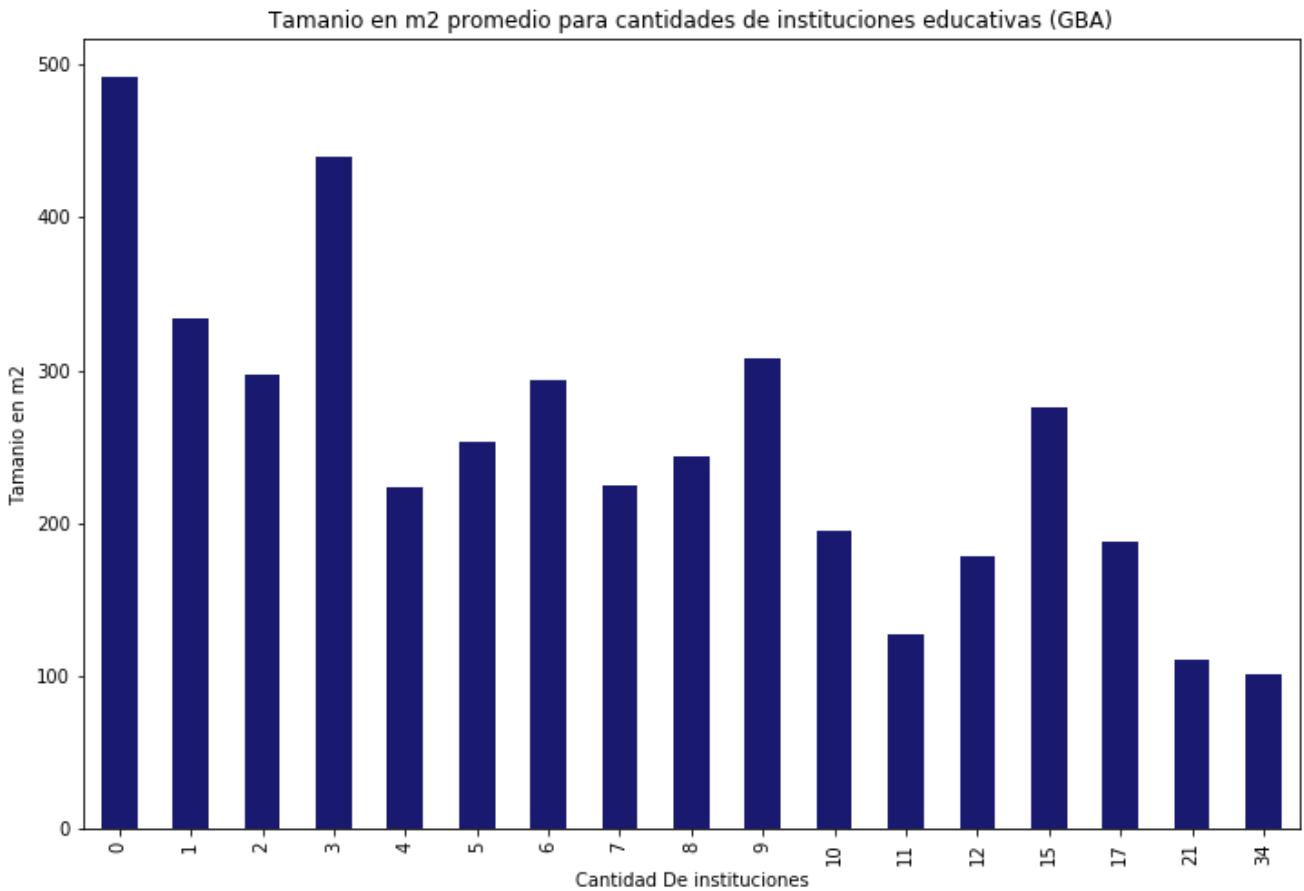


Figura 12: GBA

Al darle una vista rápida a los gráficos se ven que los comportamientos de CABA Y GBA son muy distintos. Pero en este caso los picos máximos tienen valores similares rondando entre 500 y 600 metros cuadrados. Se procede a analizar los datos por separado para obtener conclusiones mas precisas.

¿Que conclusiones se obtienen de GBA?

En GBA se ve una tendencia a la baja de tamaños, siendo que a mayor cantidad de instituciones, las propiedades tienen un tamaño menor. Una posible razón puede adjudicarse a que muchas de las propiedades de GBA pertenecientes al Dataframe son de barrios cerrados en donde los terrenos suelen ser particularmente grandes y a la vez las distancias a, no solo instituciones educativas si no también a locales o zonas con mayor población, son superiores.

Relación entre el tamaño total de la superficie y la cantidad de instituciones educativas para CABA

Se distinguen dos claros picos máximos en 5 instituciones, y en 38. Mas allá de esto no se ven otras relaciones que resulten de interés para el análisis. Para ubicar las propiedades en cuestión se realiza un heatmap con ellas.

HEATMAP por ubicaciones



Figura 13: CABA con 5 instituciones

HEATMAP por tamaño de superficie



Figura 14: CABA con 5 instituciones

HEATMAP por ubicaciones

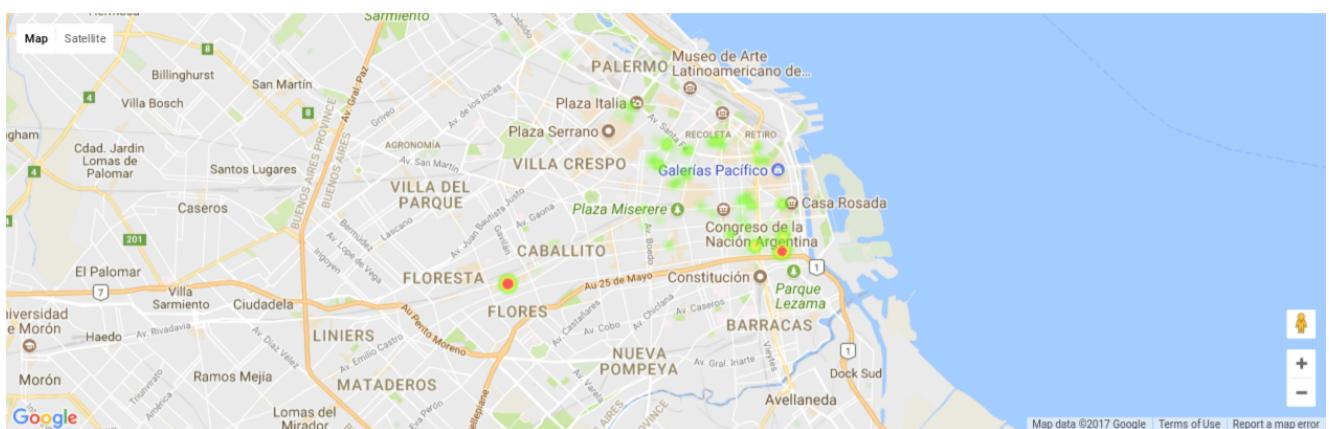


Figura 15: CABA con 38 instituciones

HEATMAP por tamaño de superficie

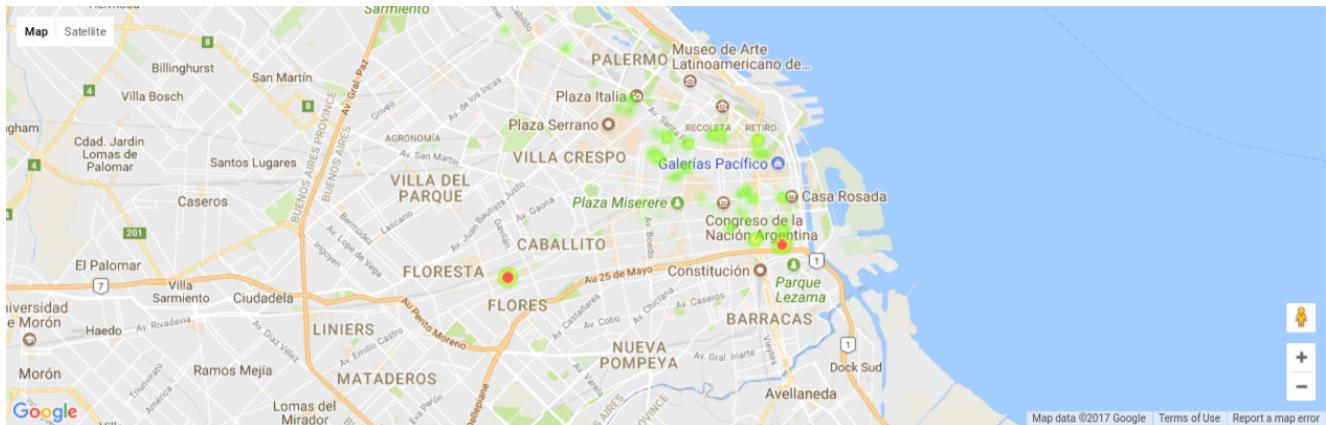


Figura 16: CABA con 38 instituciones

¿Que conclusiones se obtienen de estos heatmaps?

En los heatmaps con 5 propiedades en CABA se marca la tendencia que ubicaba a Villa Crespo como un polo. En donde hay una base sólida pero no muy grande de instituciones educativas y a la vez hay precios altos por metro cuadrado y propiedades de tamaños importantes. A esto se agrega la zona de puerto madero en donde se observan precios muy altos, grandes propiedades, y un buen numero de instituciones educativas, particularmente universidades.

Finalmente se ubican dos nuevos polos con propiedades muy grandes y realmente un numero enorme de instituciones (38 instituciones) siendo estos Plaza Dorrego y en Flores cerca de la Av. Rivadavia entre Av.Nazca y Av.Carabobo.

6. Análisis de Locales Gastronómicos

A continuación se encuentra el análisis de las instituciones de tipo gastronómicas para un set de datos completo de 72474 entradas.

6.1. Análisis del precio por metro cuadrado

Para los siguientes gráficos se agruparon las propiedades por cantidad de instituciones cercanas y sacando el promedio del precio por metro cuadrado para ellas. Para que los resultados sean significativos, se filtro a todas los conjuntos de propiedades con un precio menor a \$USD8000 teniendo en cuenta el análisis inicial.

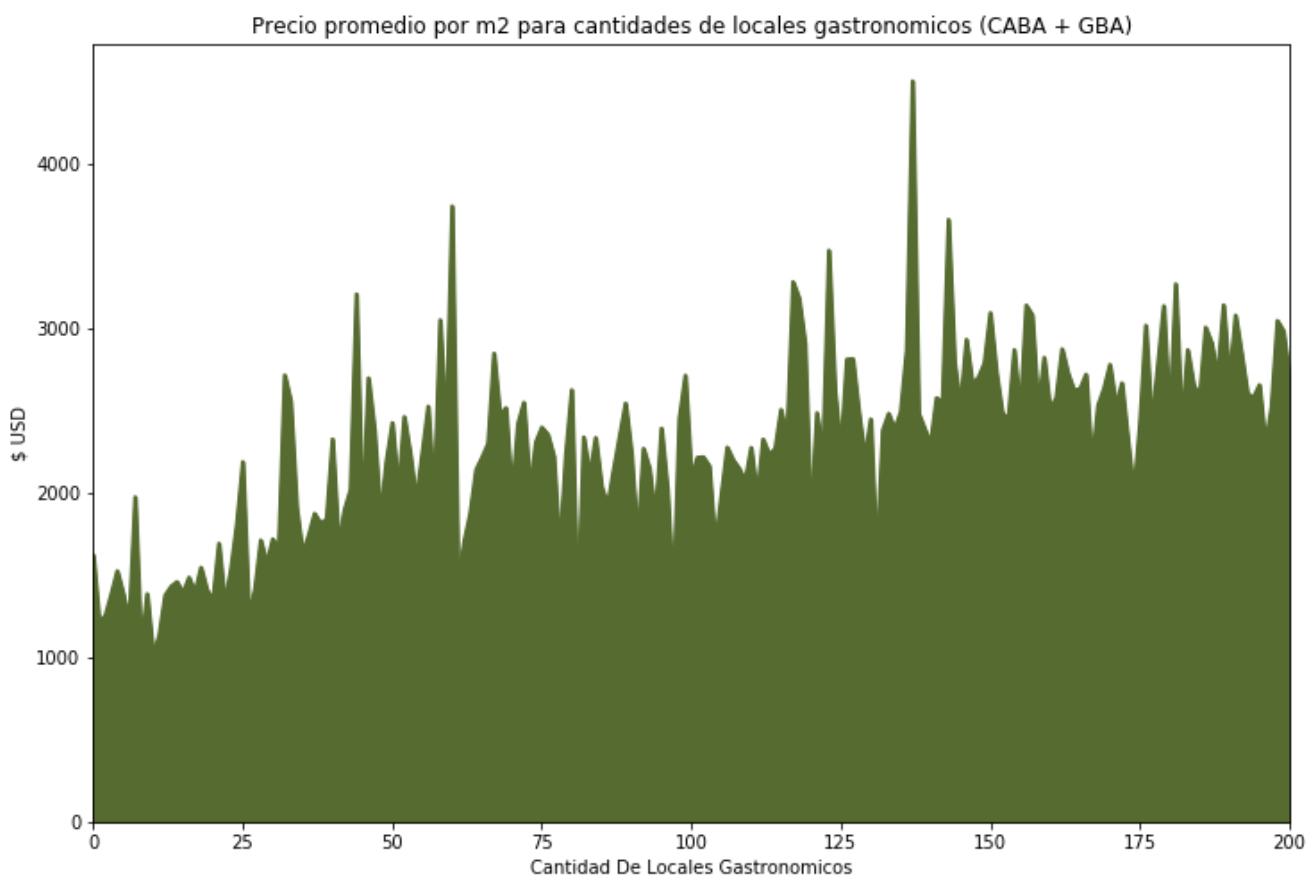


Figura 17: CABA + GBA

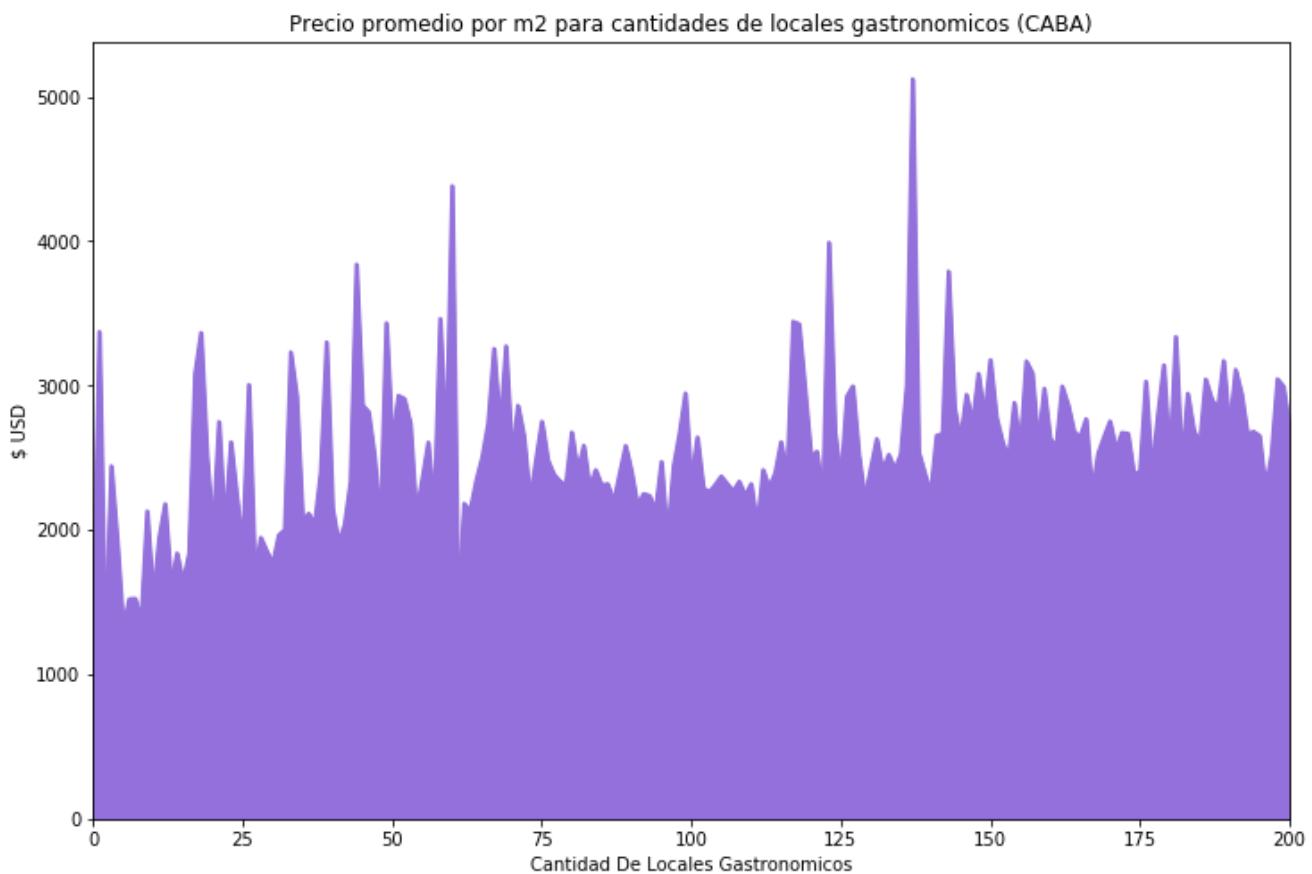


Figura 18: CABA

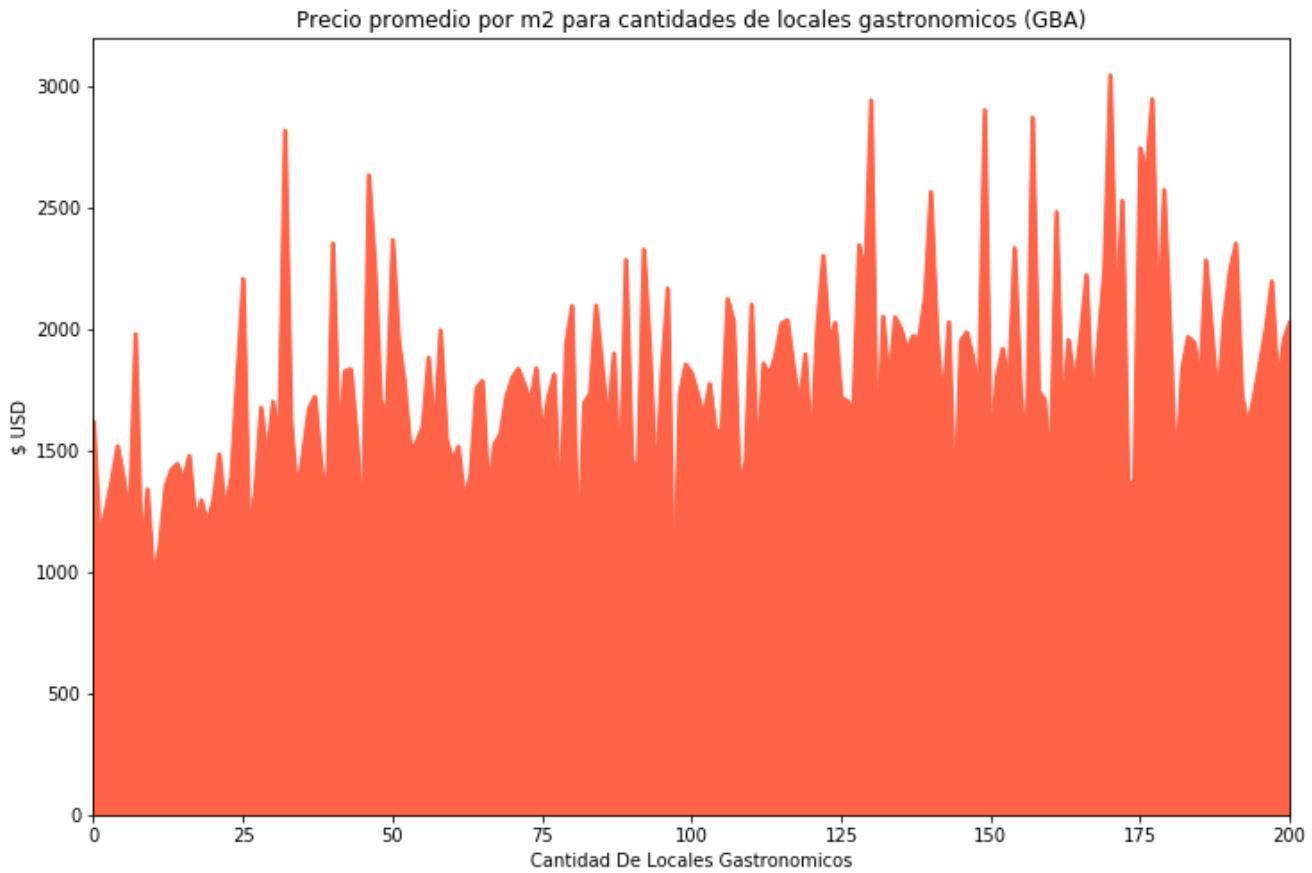


Figura 19: GBA

A simple vista los gráficos no dicen mucho. Es observable una leve tendencia de crecimiento en cuanto a mayor cantidad de locales el precio de las propiedades sería mayor y lo mismo para menor cantidad de locales. Dicha tendencia se puede apreciar mejor en el gráfico de GBA + CABA. En los otros gráficos, la tendencia de punta a punta (mas allá de abruptas variaciones) se asemeja a constante. Se procede a ubicar donde están los lugares con menores locales y donde están los que mas tienen, es decir, los extremos.

Heatmap con el peso en el precio por metro cuadrado de locales gastronómicos



Figura 20: CABA + GBA de 0 a 20 locales

Mismo Heatmap que la figura anterior pero con ZOOM



Figura 21: CABA + GBA de 0 a 20 locales

De estos últimos heatmaps se puede extraer la interesante conclusión de que las propiedades con pocos locales gastronómicos a su alrededor pertenecen a GBA y prácticamente excluyen a CABA.

Heatmap con el peso en el precio por metro cuadrado de locales gastronómicos



Figura 22: CABA + GBA de 180 a 200 locales

Continuando con el análisis previo, las propiedades con entre 180 y 200 locales gastronómicos en sus cercanías están en su totalidad ubicados en CABA, mas particularmente su zona céntrica.

Principales polos gastronómicos

Analizando mas detalladamente las posiciones de las propiedades en el mapa se encuentran los siguientes sectores como principales polos gastronómicos:

- Plaza Serrano
- Recoleta
- Plaza Dorrego
- Belgrano

Cabe aclara que en GBA las mayores concentraciones están ubicadas en la cercanía de centros comerciales (shopping centers).

7. Análisis de Puntos de Interés Cultural

A continuación se encuentra el análisis de los puntos de interés cultural para un set de datos completo de 72474 entradas.

7.1. Análisis del precio por metro cuadrado

Primero se comienza sin agregar ningún recorte extra al set de datos y se analiza una descripción del set de datos. Utilizando la función describe, es posible notar que menos del %50 de los datos tienen una cantidad de entradas menor a 60. Esto quiero decir que puede haber por ejemplo solo una propiedad con un numero muy alto de entradas que distorsiones a todo el resto. Como lo que se esta tratando de buscar son tendencias. Se procede a cortar la cantidad de instituciones por el porcentual del %99 para achicar posibles disruptciones. Finalmente recortando por cantidad de puntos culturales menores a 65 y que estos tengan por lo menos 10 repeticiones se obtienen los siguientes gráficos:

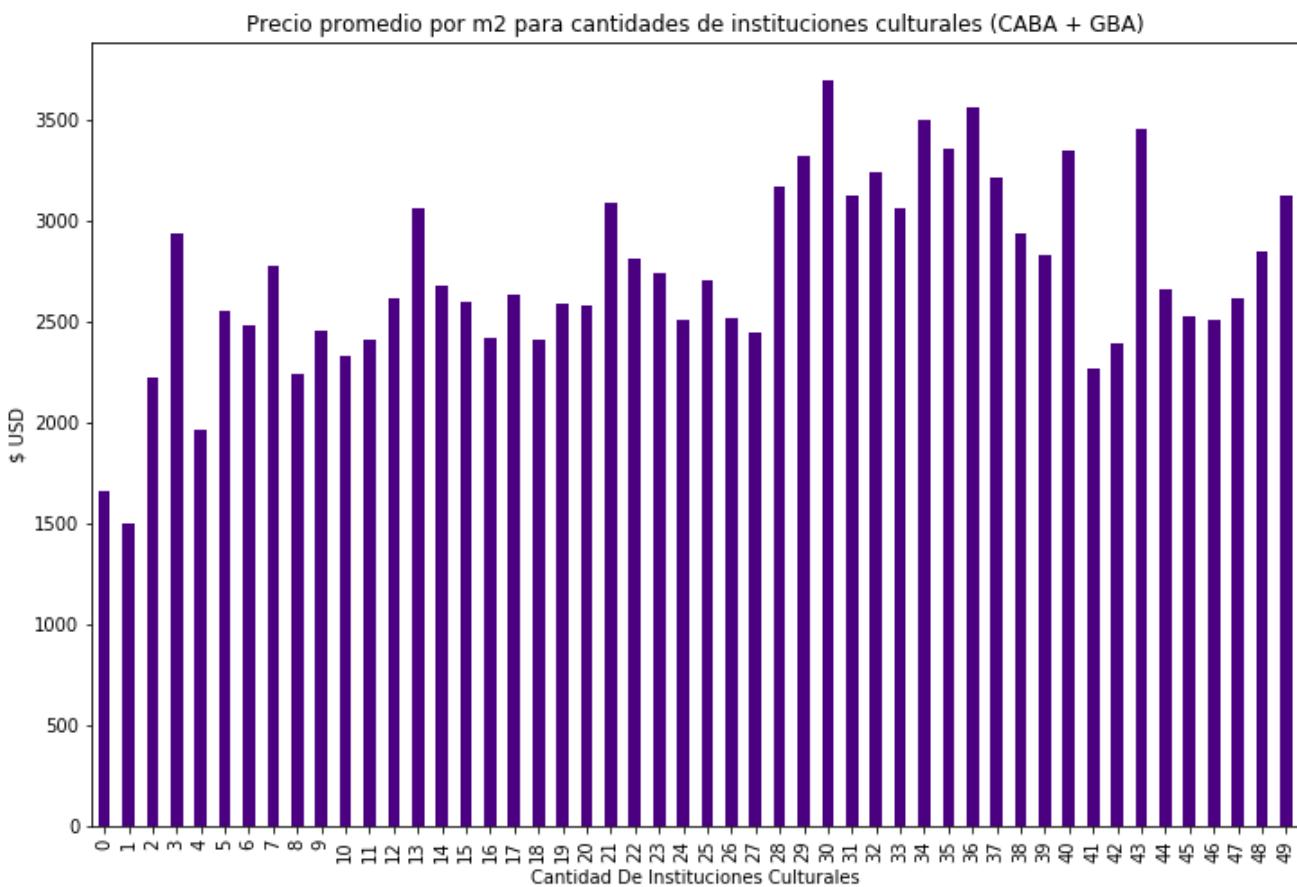


Figura 23: CABA + GBA

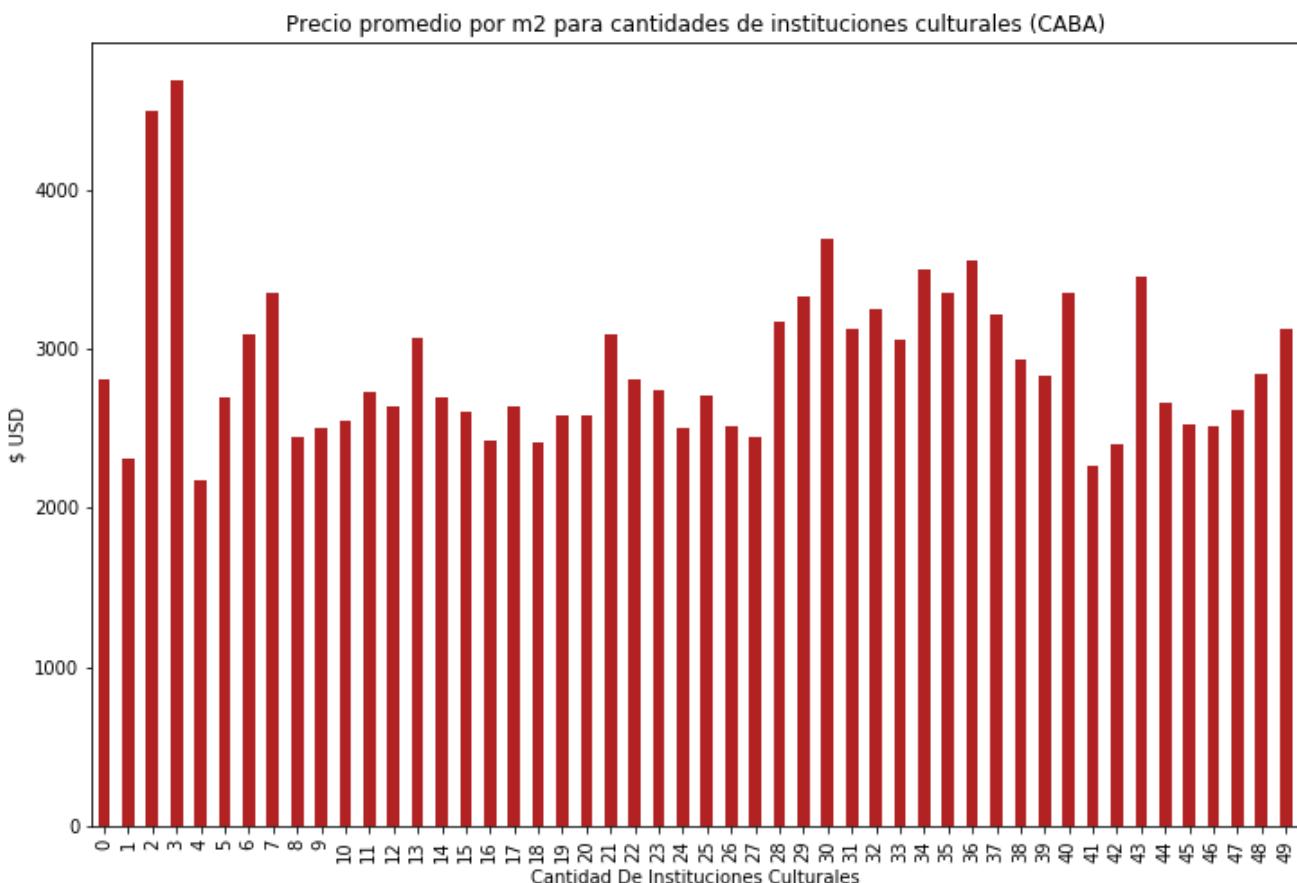


Figura 24: CABA

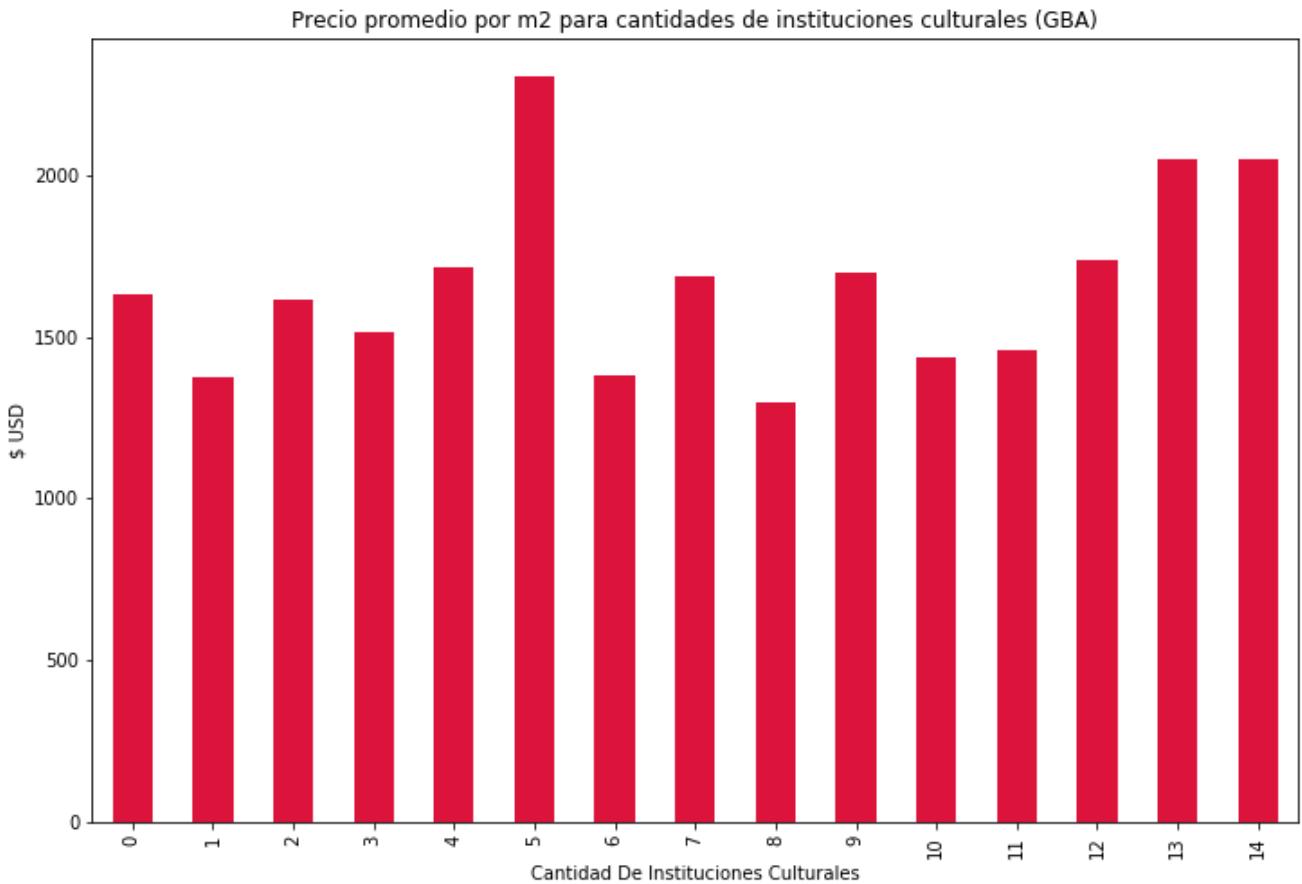


Figura 25: GBA

Instantáneamente se percibe lo que conociendo el territorio seria sospechable. Esto es en primer medida que la cantidad de puntos de interés cultural en CABA contra los de GBA es considerablemente mayor siendo 49 contra 14 las respectivas cantidades máximas. A la vez es notable resaltar que en gráfico de CABA + GBA hay una clara disminución en el precio de la propiedad para lugares con 0 o 1 puntos cercanos. Mas allá de estos conceptos, los tendencias sugieren ser a simple vista constantes.

Se trata de ubicar a las propiedades con pocos puntos cercanos en zonas particulares.

Heatmap con el peso en las ubicaciones



Figura 26: CABA + GBA con 0 o 1 puntos de interés cultural cerca

Como los gráficos anticipaban, casi en su totalidad estas propiedades están ubicadas en GBA. Sin embargo surge un punto interesante y es que se concentran mas que nada en la zona norte de GBA. Esto podría adjudicarse a que un gran porcentaje de las propiedades utilizadas pertenecen efectivamente a esa zona.

Al tratar de analizar los gráficos previos, los diferenciados por CABA muestran indicios de que podría aplicarse una análisis mas particular para lograr obtener sus polos mas importantes como se realizó con el resto de las investigaciones. Es así que se realiza un nuevo gráfico pero esta vez recortando a las cantidades de puntos de interés con bajas frecuencias, para así encontrar concentraciones mas importantes.

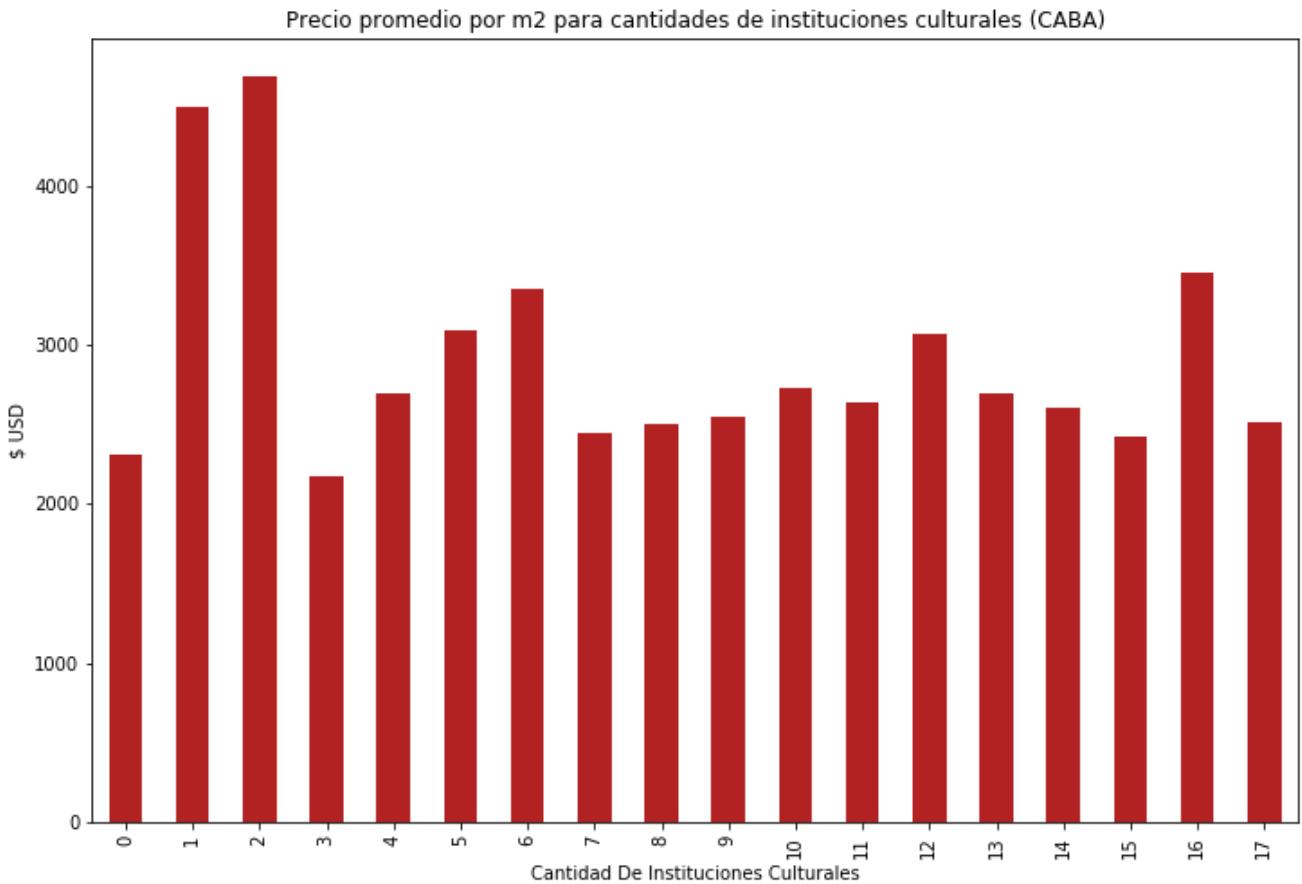


Figura 27: CABA

Los picos en 1 y 2 instituciones se siguen sosteniendo mas allá de la filtración de los datos, esto podría sugerir o que la gran mayoría de las propiedades en CABA tienen esa cantidad de instituciones cerca, o algo mas difícil de demostrar que quizás tener pocas instituciones cerca realmente mejora el precio de la propiedad.

Se asilan los puntos en CABA con una o dos instituciones cerca y se las ubica en un heatmap.

Heatmap con el peso en las ubicaciones

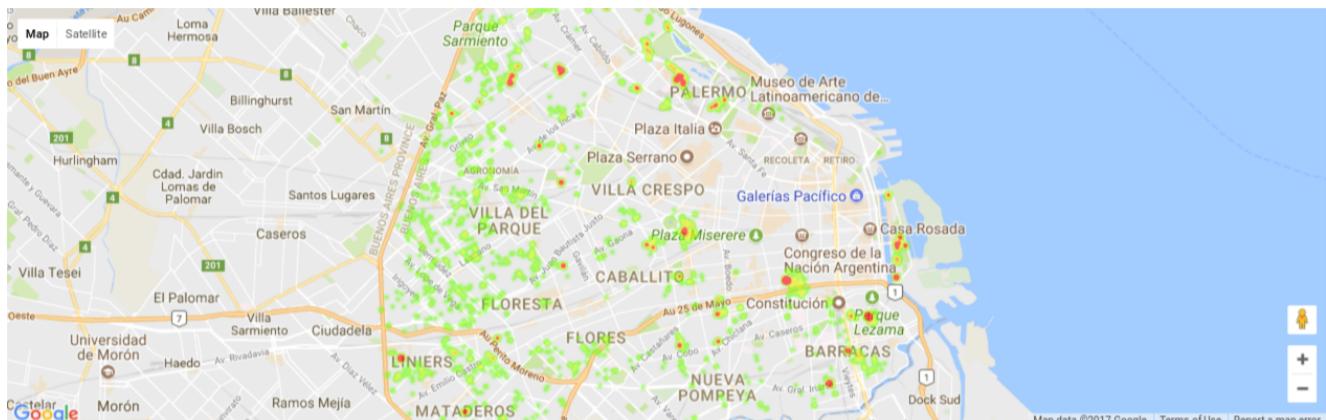


Figura 28: CABA con 2 o 3 puntos de interés cultural cerca

Las propiedades en cuestión están ubicadas en las periferias de la zona céntrica de CABA (salvo excepciones). Esto muestra que hay un numero de propiedades con un valor alto por metro cuadrado que se ubican lejos de las zonas con mas afluencia de personas, o hasta turistas, ya que los puntos de interés culturales suelen ser visitados por ellos. Esto podría sugerir que al estar “aislado” de los lugares mas transitados pero seguir perteneciendo a la capital le da un valor elevado a la propiedad, mas que nada en el sentido de lugares de residencia.

Principales polos de puntos de interés cultural

Heatmap con el peso en la cantidad de puntos de interés cultural cercanos



Figura 29: CABA propiedades con mayor cantidad puntos de interés cultural cerca

Analizando mas detalladamente las posiciones de las propiedades en el mapa se encuentran los siguientes sectores como principales polos de lugares culturales:

- Plaza Serrano
- Plaza Dorrego
- Belgrano

Es interesante comparar estos resultados con los polos gastronómicos. Parecería que al menos a simple vista hay una correlación entre los ambos puntos con mayor concentración de puntos de interés cultural y de locales gastronómicos.

8. Análisis de Transporte Público

A continuación se encuentra el análisis de los servicios de transporte público para un set de datos completo de 72474 entradas.

Con transporte público se refiere a las paradas o estaciones de colectivos o subterráneos en un radio de 400m de la propiedad.

8.1. Análisis del precio por metro cuadrado

Para los siguientes gráficos se agruparon las propiedades por cantidad de paradas de transporte público cercanas y sacando el promedio del precio por metro cuadrado para ellas. Al ser el máximo de 57 paradas relativamente bajo para el radio tomado, se filtro a todas los conjuntos de propiedades con menos de 10 entradas para las cantidades de paradas cercanas y así levemente limpiar de distorsión al set de datos.

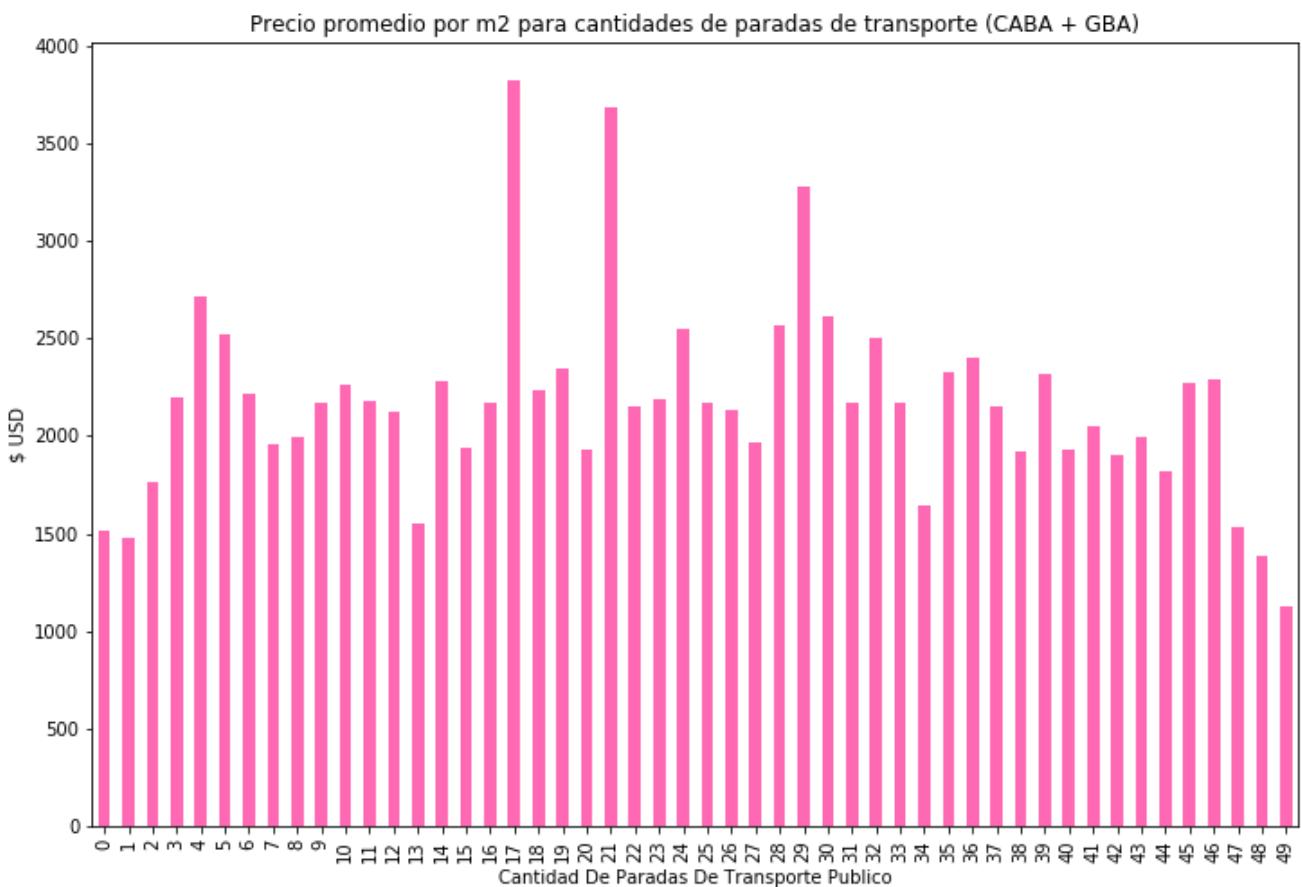


Figura 30: CABA + GBA

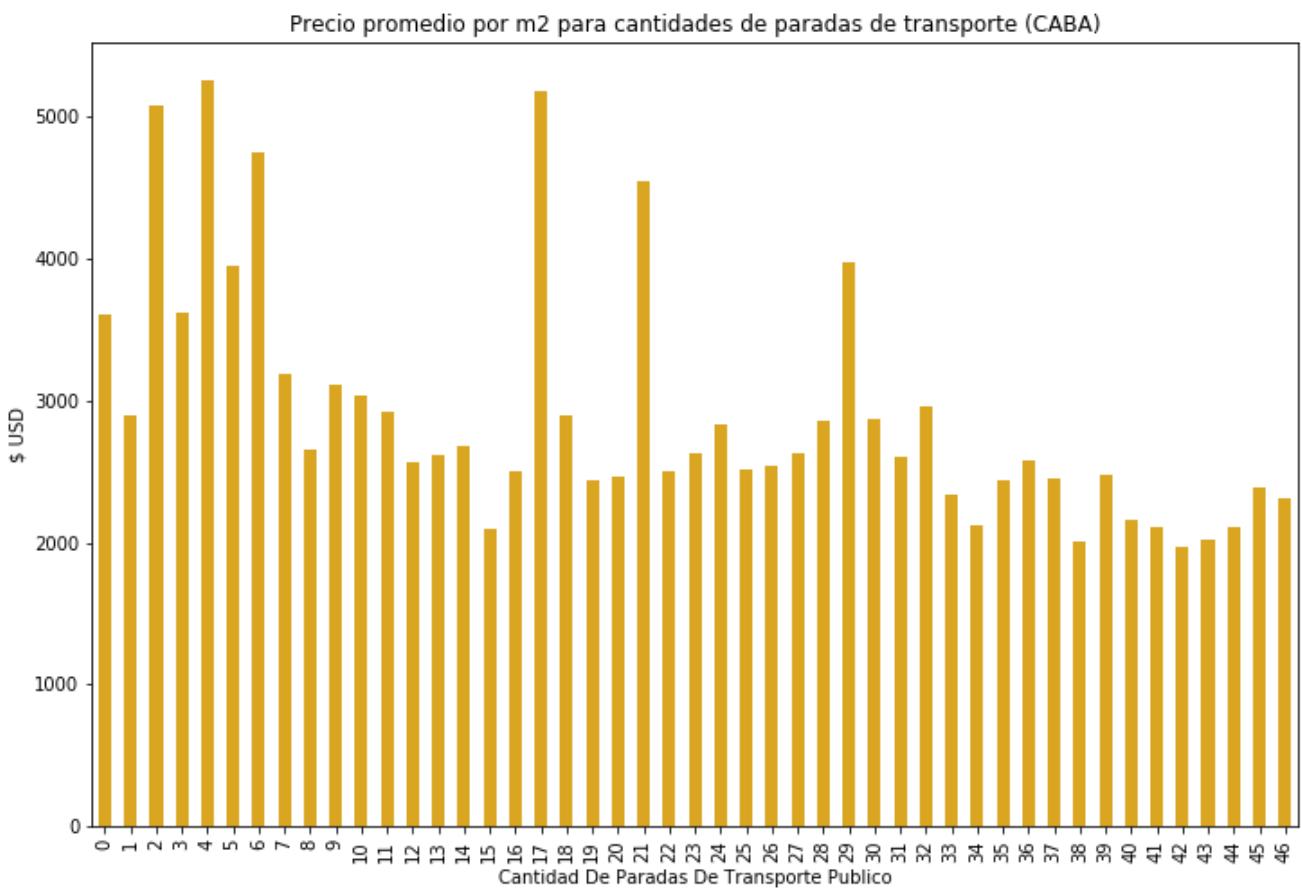


Figura 31: CABA

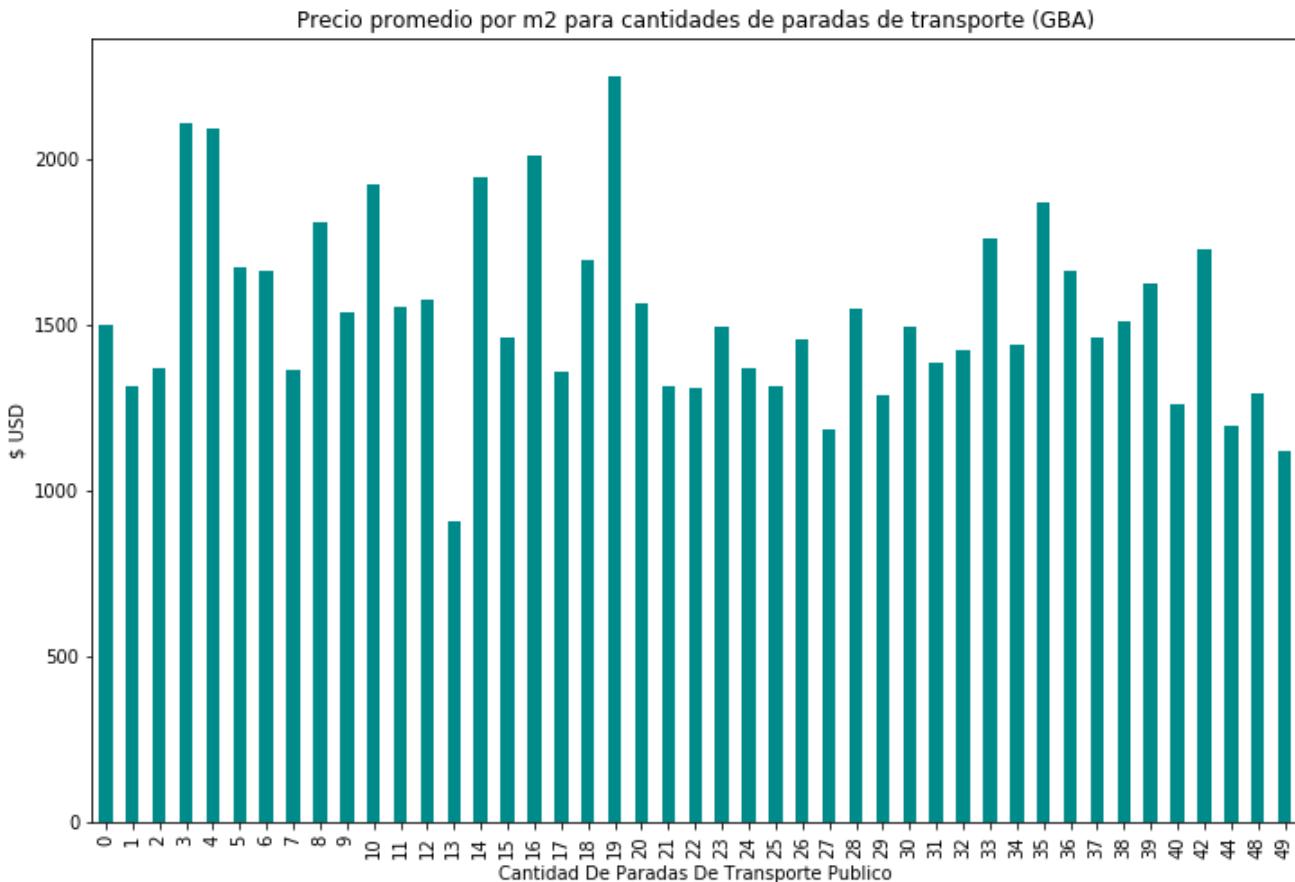


Figura 32: GBA

Analizando el gráfico que contempla GBA + CABA se puede ver una especie de campana en donde los menores precios por metro cuadrado estarían concentrados en los extremos y en el centro los picos mayores. Este tendencia sin embargo no se sostiene en el resto. En CABA parecería ser como hay un leve decaimiento en el precio a medida que aumenta la cantidad de paradas. Nuevamente los máximos de CABA y GBA son significativamente distintos, siendo casi la mitad el de GBA. Se continua analizando las distintas anomalías o tendencias recién mencionadas para encontrar patrones.

Separación de CABA + GBA por precio

\$USD2000 por metro cuadrado parecería ser una linea prudente por la cual dividir las propiedades y analizar la distribución en la cantidad de paradas.

Heatmap con el peso puesto en la cantidad de paradas



Figura 33: CABA + GBA propiedades con precio por metro cuadrado mayor a \$USD2000

Zoom de la figura anterior

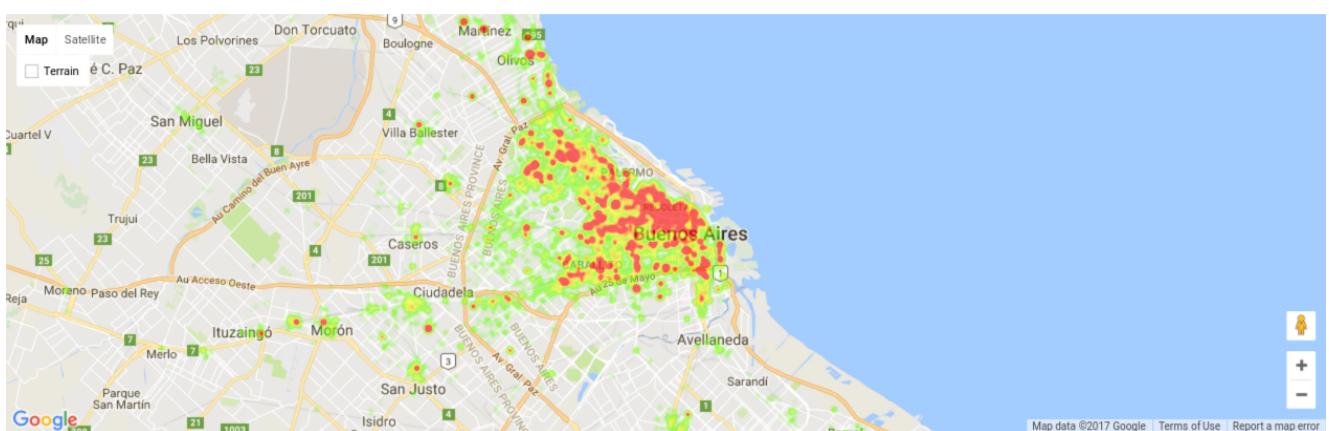


Figura 34: CABA + GBA propiedades con precio por metro cuadrado mayor a \$USD2000

Hay una clara concentración en la capital, con la excepción de centros de los cordones mas inmediatos de GBA. Viendo el siguiente mapa de trenes y subterráneos, este muestra como el tendido de redes se expande a medida que uno se aleja del centro de CABA. El heatmap parecería tener una forma muy similar sugiriendo que las propiedades de mayor precio podrían estar situadas cerca de paradas de subterráneo. Teniendo en cuenta que la distribución de paradas de colectivos por la capital es mas pareja, esto apoyaría a la hipótesis planteada.

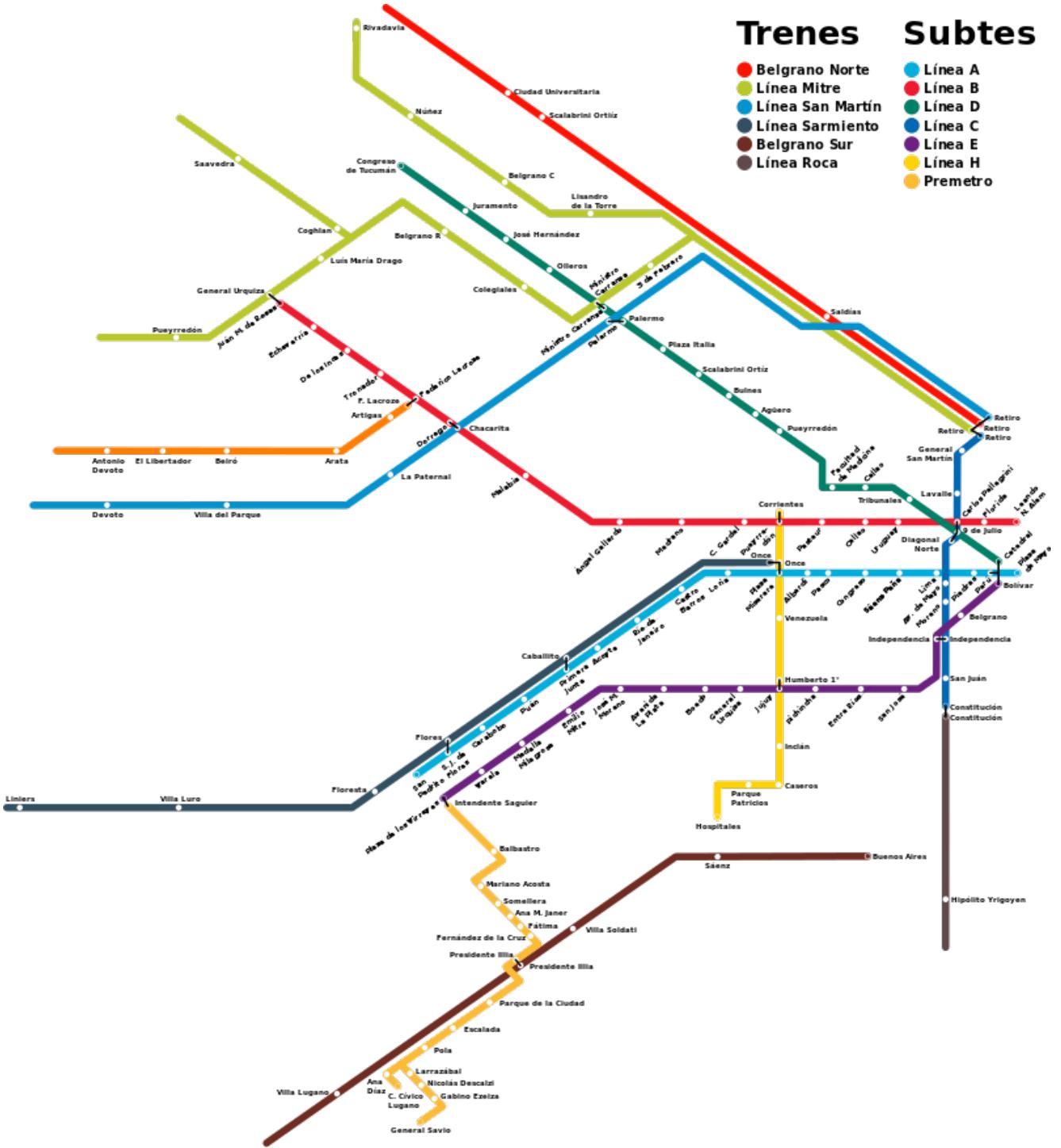


Figura 35: Red de Subterraneos y Trenes de Buenos Aires

Heatmap con el peso puesto en la cantidad de paradas



Figura 36: CABA + GBA propiedades con precio por metro cuadrado menor a \$USD2000

La cantidad de propiedades menores a \$USD2000 parecerían ser las de menor cantidad de paradas de transporte público cercanas. Eso se denota ya que los heatmaps están armados utilizando como peso la cantidad de paradas cercanas. Es decir que mientras mas paradas cercanas tenga, mas roja se vera la ubicación en el mapa. A la vez se ve que las concentraciones de lugares con muchas paradas se ubican en las principales zonas de GBA (a nivel de población) siendo estas Morón, San Miguel, Tigre, Lomas de Zamora, entre otras.

Zonas de CABA + GBA por mayor cantidad de paradas

Se ubican las propiedades con mas de 35 paradas de transporte publico cercanas.

Heatmap con el peso puesto en la cantidad de paradas

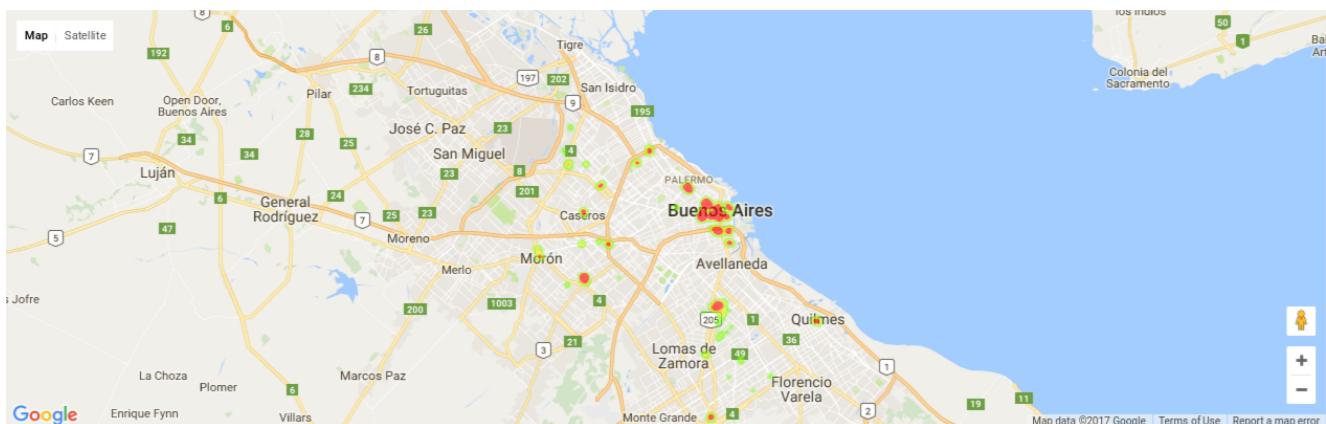


Figura 37: CABA + GBA propiedades con mas de 35 paradas cerca

Gran parte de estos pertenecen al microcentro de CABA, mientras que a la vez se encuentran ciertos puntos con una alto concentracion:

- Lanus
- San Justo
- Constitucion
- Quilmes

Heatmap anterior con ZOOM en CABA



Figura 38: CABA propiedades con mas de 35 paradas cerca

Lógicamente, las propiedades con mayor cantidad de paradas cercanas se encuentran en el micro-centro porteño, de donde salen todas las líneas de subterráneo y una abundante cantidad de colectivos. A la vez Plaza Italia, Plaza Miserere y la zona de la Facultad de Medicina funcionan como puntos principales.

Zonas de CABA + GBA por menor cantidad de paradas

Se ubican las propiedades con 7 o menos paradas de transporte público cercanas. (Heatmap con el peso en la cantidad de paradas)

Heatmap con el peso en la cantidad de paradas

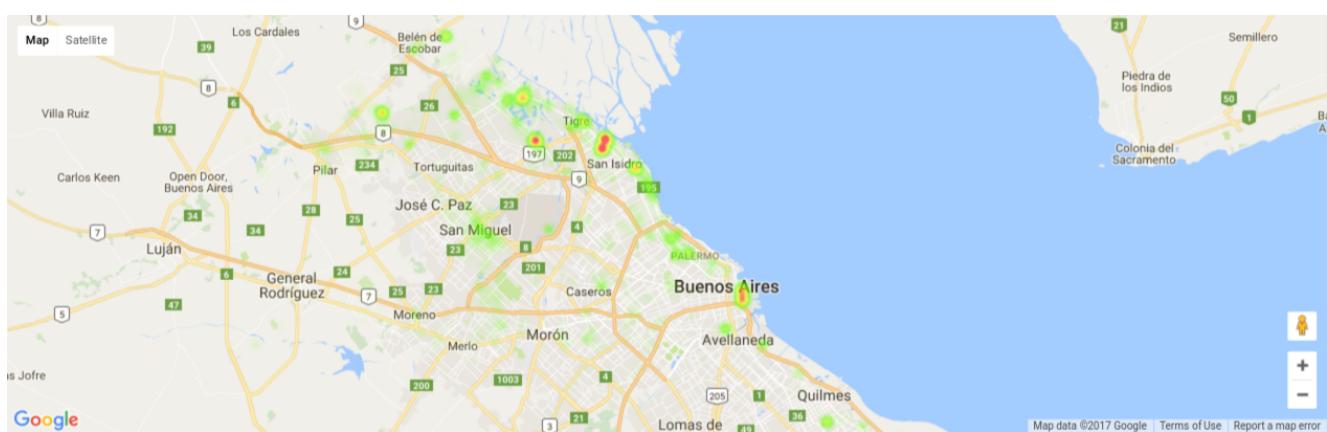


Figura 39: CABA propiedades con 7 o menos paradas de transporte público cercanas

El Heatmap permite rápidamente identificar a Puerto Madero, Tigre, Pacheco y parte de San Isidro como lugares en donde hay una reducida cantidad de paradas de transporte público en comparación al resto del territorio.

Distribución de las cantidad de paradas contra el precio por metro cuadrado en CABA (desarrollo de hipótesis de red de transporte)

El siguiente mapa muestra el tendido de las redes subterráneas por la ciudad de Buenos Aires siendo:

- Línea A (Celeste)
- Línea B (Roja)
- Línea C (Azul)
- Línea D (Verde)
- Línea E (Violeta)
- Línea H (Amarilla)

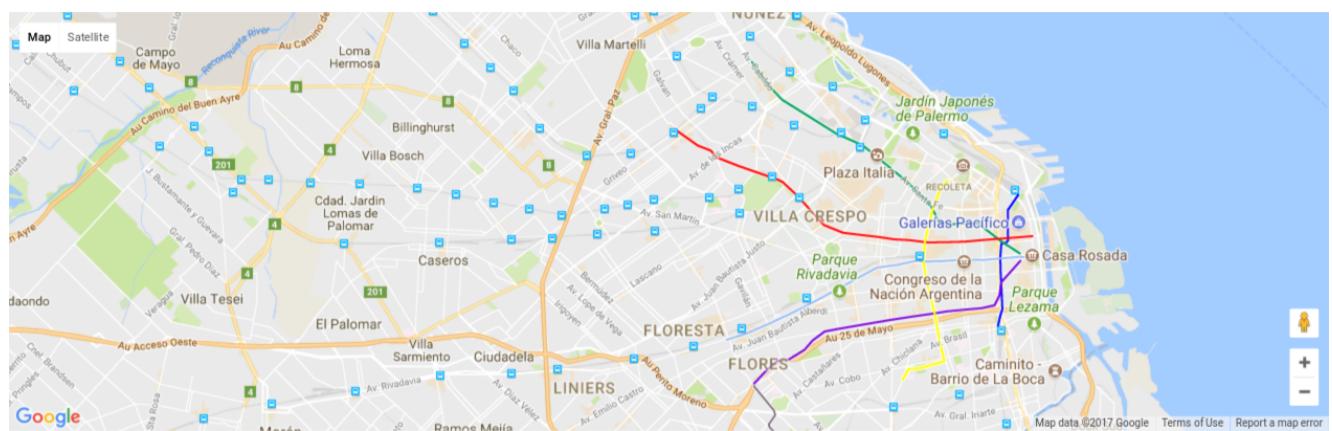


Figura 40: Red de Subterraneos Buenos Aires

Teniendo en cuenta dicho tendido se procede a graficar en un heatmap el conjunto de propiedades (CABA + GBA) teniendo en cuenta como peso del Heatmap la cantidad de paradas cercanas.

Heatmap con el peso en la cantidad de paradas

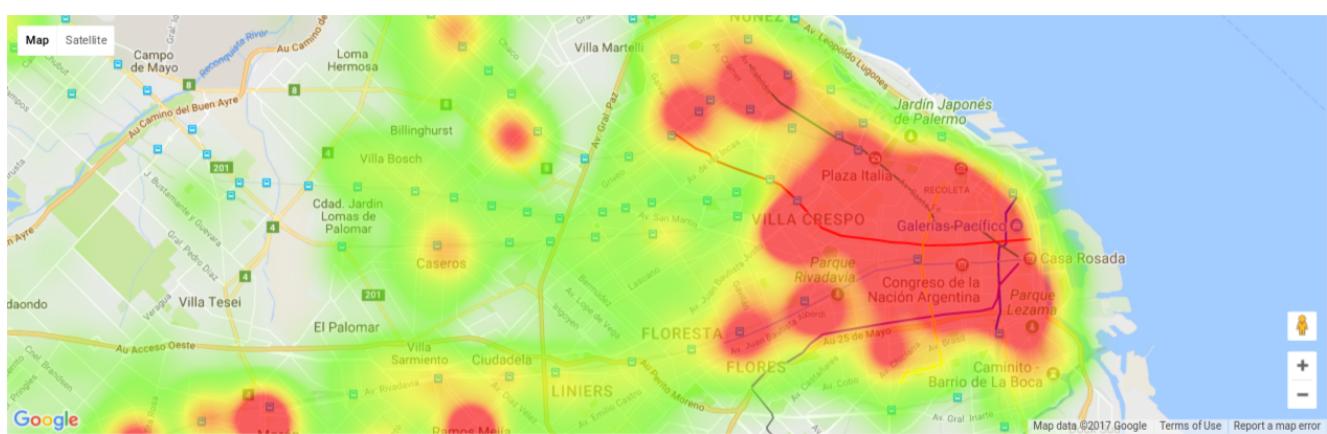


Figura 41: CABA con la Red de Subterráneos de Buenos Aires de fondo

Las zonas rojas se ubican exactamente sobre el camino de los subterráneos. Sin embargo hay huecos en donde esto no ocurre. Este heatmap no es suficiente como para definir un patrón aunque es innegable el hecho de que la hipótesis anteriormente planteada podría corresponderse. Para despejar dudas, se realiza el próximo heatmap utilizando el peso de este como el precio de las propiedades por metro cuadrado.

Heatmap con el peso en el precio por metro cuadrado

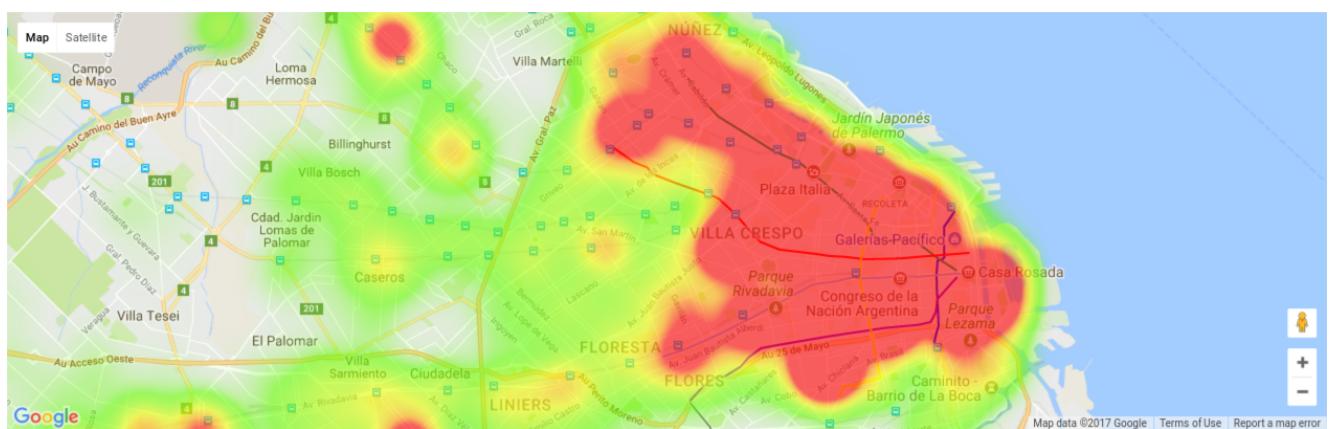


Figura 42: CABA con la Red de Subterráneos de Buenos Aires de fondo

Casi en su totalidad, las propiedades cubren las líneas de subterráneo. Con la salvedad de un pequeño tramo de la línea B. Esto marca una clara tendencia en la relación directa entre el precio de una propiedad con la cercanía a las paradas de subterráneo en la Capital Federal.

9. Conclusiones Generales

Resumiendo, se presentan a continuación las variadas conclusiones obtenidas del análisis de Google Places que serán de gran utilidad para la segunda etapa del trabajo. Se aclara que se encontraron mas tendencias interesantes pero que no son de gran importancia para el próximo trabajo en donde se debe aproximar el precio de una propiedad dada.

- Las relaciones de precios entre CABA y GBA teniendo en cuenta todos los datos calculados por Google Places siempre da una amplia mayoría en el precio de CABA contra el de GBA.
- Principales polos de concentración de instituciones educativas: Olivos, San Cristóbal, Villa Urquiza/General Urquiza, Parque Patricios, Colegiales.
- En GBA se ve una tendencia a la baja de tamaños, siendo que a mayor cantidad de instituciones, las propiedades tienen un tamaño menor.
- Las propiedades con pocos locales gastronómicos a su alrededor pertenecen a GBA y prácticamente excluyen a CABA.
- Las propiedades con la mas grande cantidad de locales gastronómicos cercanos están en su totalidad ubicados en CABA, mas particularmente su zona céntrica.
- Los lugares de Plaza Serrano, Recoleta, Plaza Dorrego y Belgrano comparten la relación de ser centros de concentración tanto gastronómicos como culturales.
- En GBA las mayores concentraciones de locales gastronómicos están ubicadas en la cercanía de centros comerciales (shopping centers).
- La cantidad de puntos de interés cultural en CABA contra los de GBA es considerablemente mayor siendo 49 contra 14 las respectivas cantidades máximas.
- Estar alejado de los lugares "mas culturales" pero seguir perteneciendo a la capital le da un valor elevado a la propiedad.
- En CABA hay un leve decaimiento en el precio a medida que aumenta la cantidad de paradas de transporte cercanas.
- Existe una relación directa entre el precio de una propiedad con la cercanía a las paradas de subterráneo en la Capital Federal.