FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Organización de datos - 75.06/95.58

Trabajo Práctico 1

Análisis exploratorio de datos

2º cuatrimestre 2019

Grupo 28 - Datoy Story

Apellido, Nombre	Nº Padrón
Mac Gaul, Pedro	101503
Macía, Tomás	99248
Perez, Cristian	95536
Riborati, Franco	99411

Link al repositorio: github.com/tomasmacia/orgaDatos

1. Introducción general

Este informe se encargará de analizar los datos de las publicaciones que se han realizado en la pagina https://www.zonaprops.com.ar/. Estas publicaciones fueron realizadas entre 2012 y 2016 en México.

Estas en estas publicaciones se encontraban la siguiente información:

- *id*: Un id numérico para identificar la propiedad
- *titulo*: El título de la propiedad publicada
- *descripcion*: La descripción de la propiedad publicada
- *direccion*: La dirección de la propiedad
- *ciudad*: La ciudad de la propiedad
- *provincia*: La provincia donde está localizada la propiedad
- *lat*: Latitud
- *lng*: Longitud
- *tipodepropiedad*: El tipo de propiedad (Casa, departamento, etc)
- *metrostotales*: Metros totales de la propiedad
- *metroscubiertos*: Metros cubiertos de la propiedad
- *antiguedad*: Antigüedad de la propiedad
- *habitaciones*: Cantidad de habitaciones
- *garages*: Cantidad de garages
- *banos*: Cantidad de baños
- *fecha*: Fecha de publicación
- *gimnasio*: Si el edificio o la propiedad tiene un gimnasio
- *usosmultiples*: Si el edificio o la propiedad tiene un SUM
- *piscina*: Si el edificio o la propiedad tiene un piscina
- *escuelascercanas*: Si la propiedad tiene escuelas cerca
- *centroscomercialescercanos*: Si la propiedad tiene centros comerciales cerca
- *precio*: Valor de publicación de la propiedad en pesos mexicanos

El objetivo de todo este informe es analizar los datos descritos anteriormente para observar si se encuentra alguna tendencia en los datos que fluya a lo largo de todos los registros, encontrar si hay alguna irregularidad en alguno de los mismos y también brindarle a la empresa que nos proporcionó los datos, en este caso Navent, un análisis que puede llegar a proporcionarles datos útiles para mejorar su servicio o generar algún tipo de estadística. Sin ir más lejos, este informe también ayudará al lector a comprender la relación entre datos, y a su vez, puede que le sea útil por si quiere utilizar este servicio en algún futuro.

2. Análisis General

2.1 Datos faltantes

Lo primero que se realizó fue ver qué datos faltan, para poder realizar un análisis que sea representativo del set de datos y no llegar a una generalización si faltasen el 50% de los datos cómo se puede ver en la siguiente imagen, respecto a la latitud y longitud, que faltan más del 50% de los datos como para llegar a decir o generalizar algo a partir de ellos, ya que no estaría siendo representado nuestro set de datos inicial solo representa al 50% de ellos.

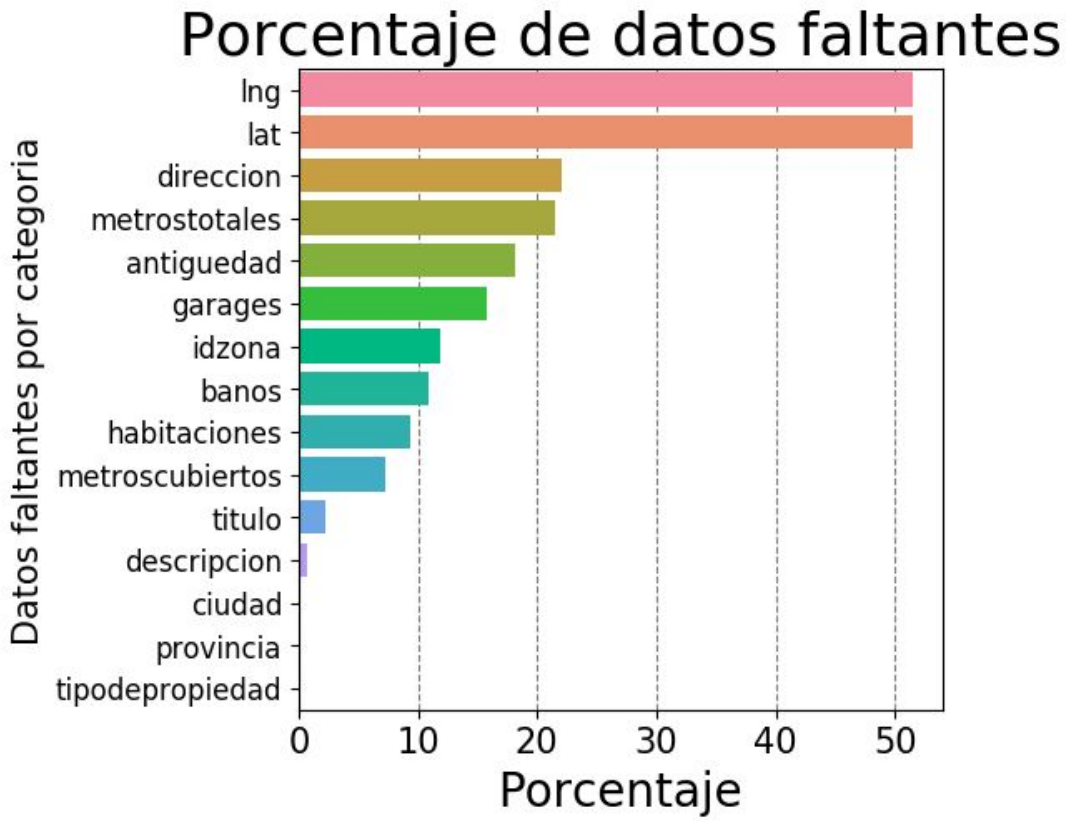

Gráfico 2.1

Para la visualización, se eliminaron todas los campos que estaban completos, es decir que no había datos faltantes. Se puede ver a simple vista que hay muchos datos faltantes en algunas categorías, en especial para las latitudes y longitudes.

Como vamos a ver más adelante, analizando un poco más en profundidad la falta de latitudes y longitudes, se vio que además de los datos faltantes, algunos estaban mal cargados, generando outliers (puntos anómalos), mostrando longitudes y latitudes que no pertenecían a México. Incluso hay puntos cuya latitud y longitud pertenecen a México pero dicen ser de ciudades y provincias que no se encuentran ese lugar. Como conclusión de esto, no se puede definir si estamos representando los datos correctamente. Por esto mismo en los gráficos geográficos, se utilizaron el nombre de la provincia, cómo se puede ver en el gráfico anterior, a la categoría “provincia” sus datos son muy pocos los datos faltantes.

2.2 Correlación entre categorías

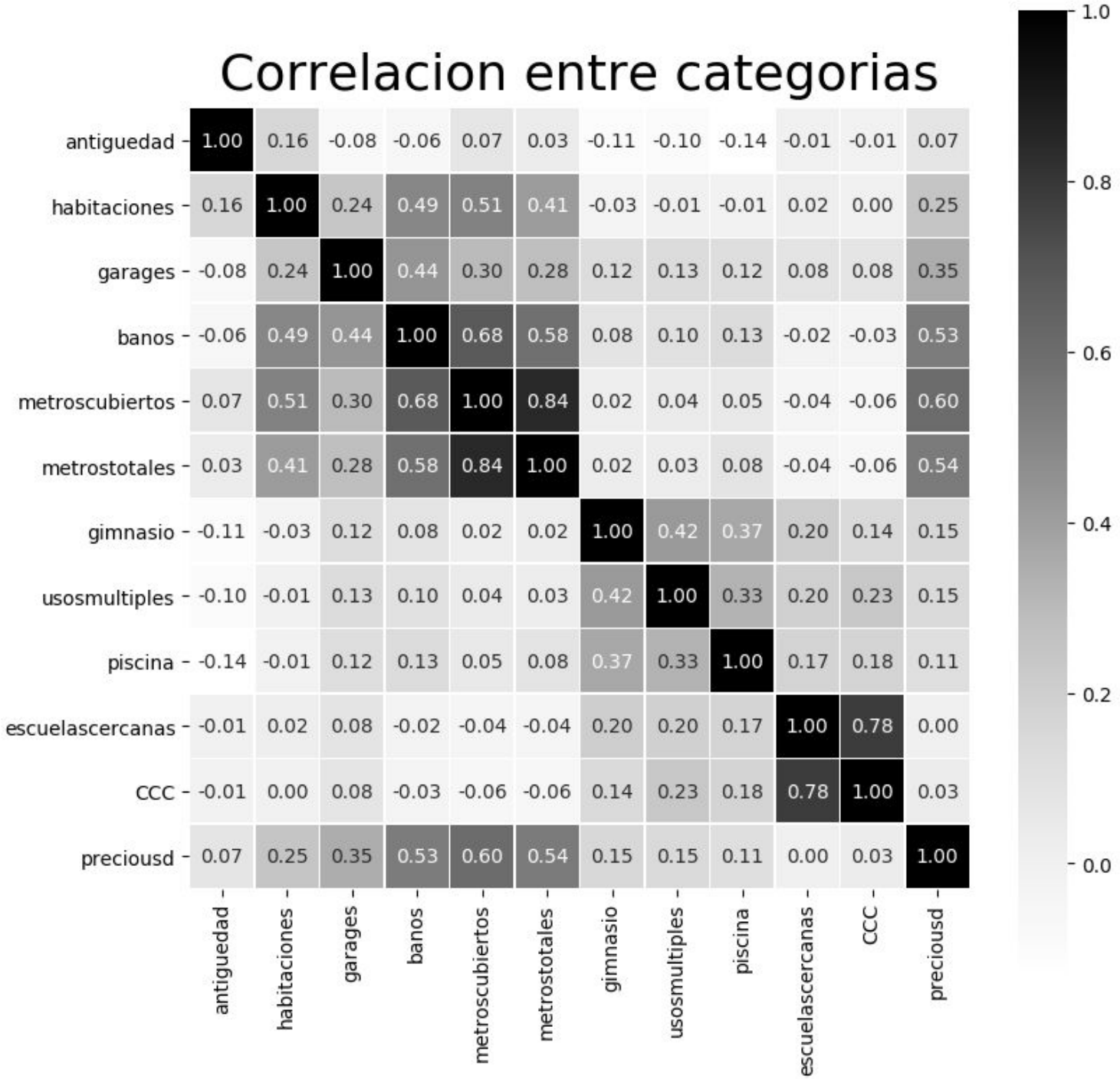

Gráfico 2.2.1

Lo que se observa en este gráfico es la correlación entre las categorías, esto quiere decir que si crece una categoría, la que se está analizando puede crecer cómo decrecer. Lo interesante que se puede destacar, es por ejemplo el caso del precio en dólares, se ve que a medida que aumentan los metros cubiertos, el precio aumenta, es trivial, pero lo que no lo es que a medida que aumentan los baños el precio se eleva aun más todavía y es mayor cómo aumenta en comparación que las habitaciones, que es algo que no es de esperarse.

De las otras categorías podemos decir que, en el caso de los centros comerciales cercanos (CCC) de las propiedades, hay también una escuela cercana, esto probablemente pase por ser lugares más céntricos que tienen centros comerciales y mayor volumen de gente, por ende también más escuelas para esas personas.

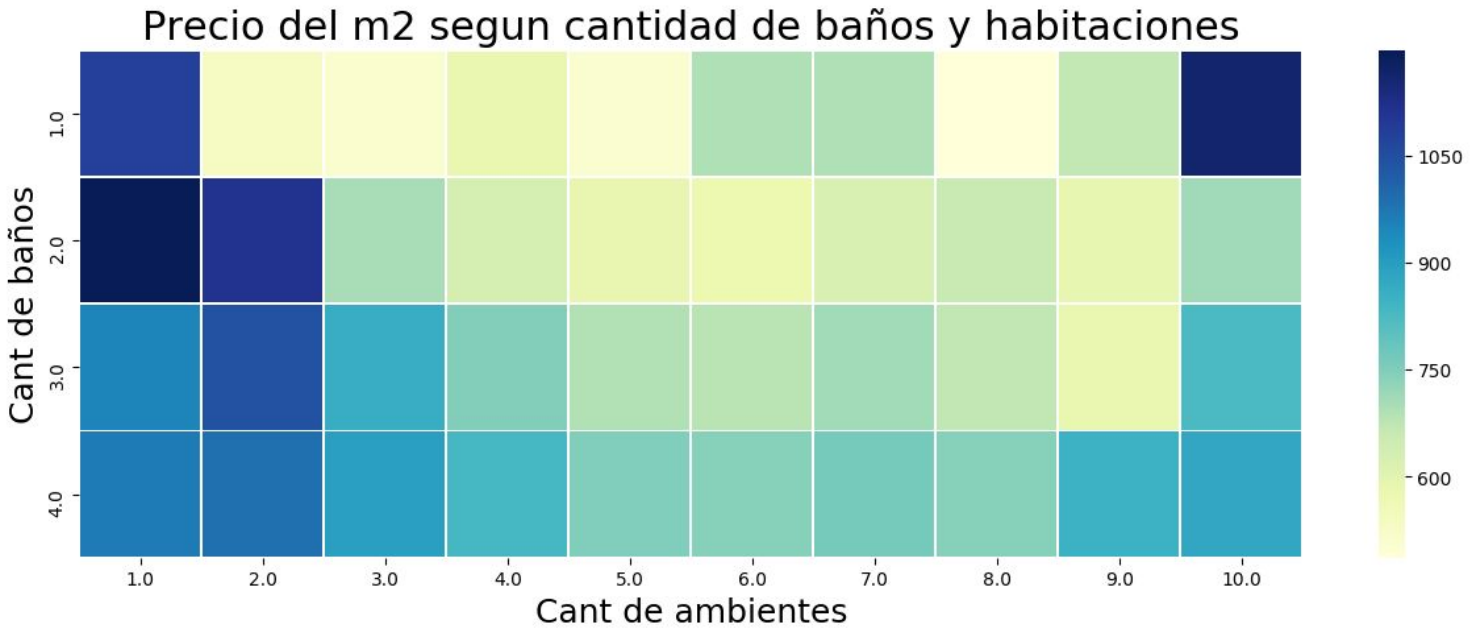

Gráfico 2.2.2

Lo que muestra el gráfico es muy interesante de ver ya que a los monoambientes y de dos ambientes son las publicaciones más costosas, y se ve que a medida que se van aumentando los ambientes el precio por metro cuadrado es más barato, quiere decir que es más barato una vivienda con mayor cantidad habitaciones que una de una sola. Esto puede deberse a que en zonas donde hay un universidades/comercios, son elevados los precios y las viviendas mas chicas porque está centralizado en un único lugar.

Otra conclusión que se puede ver de este gráfico es que las propiedades con 10 ambientes, con un más de un baño son más baratas que las que tienen solo un baño, es un raro comportamiento que vemos de las publicaciones. También podemos llegar a decir que las propiedades de 10 ambientes, tienen el mismo valor por metro cuadrado que las de un ambiente con 2 baños, esto es interesante de destacar ya que las propiedades con 10 ambientes tienen una mayor superficie que las propiedades con un ambiente, por ende deberían ser más caras que los monoambientes. Analizaremos más adelante los monoambientes para ver bien cómo se comportan ya que vimos en este gráfico que son un caso especial dentro de nuestro set de datos.

3. Análisis de publicaciones

3.1 Evolución anual de publicaciones

Gráfico 3.1

Al analizar la cantidad de publicaciones a través de los años, se puede apreciar que del año 2015 al año 2016, la cantidad de anuncios casi se duplicaron en comparación a la evolución anual que sucedía en años anteriores. Del año 2012 al 2015 la evolución era casi lineal (24000/25000 en 2012, casi 30000 en 2013, 40000 en 2014 y 50000 en 2015) y del 2015 al 2016 pegó un salto, de casi 50000 anuncios en 2015 a poco más de 90000 en 2016. Duplicando sus números.

Ahora al analizar el aporte de cada provincia al total anual en la evolución anteriormente dicha, para hacer la vista al gráfico más fácil, se hará el análisis por tramos de 2 años.

3.1.A Tramo 2012/2014

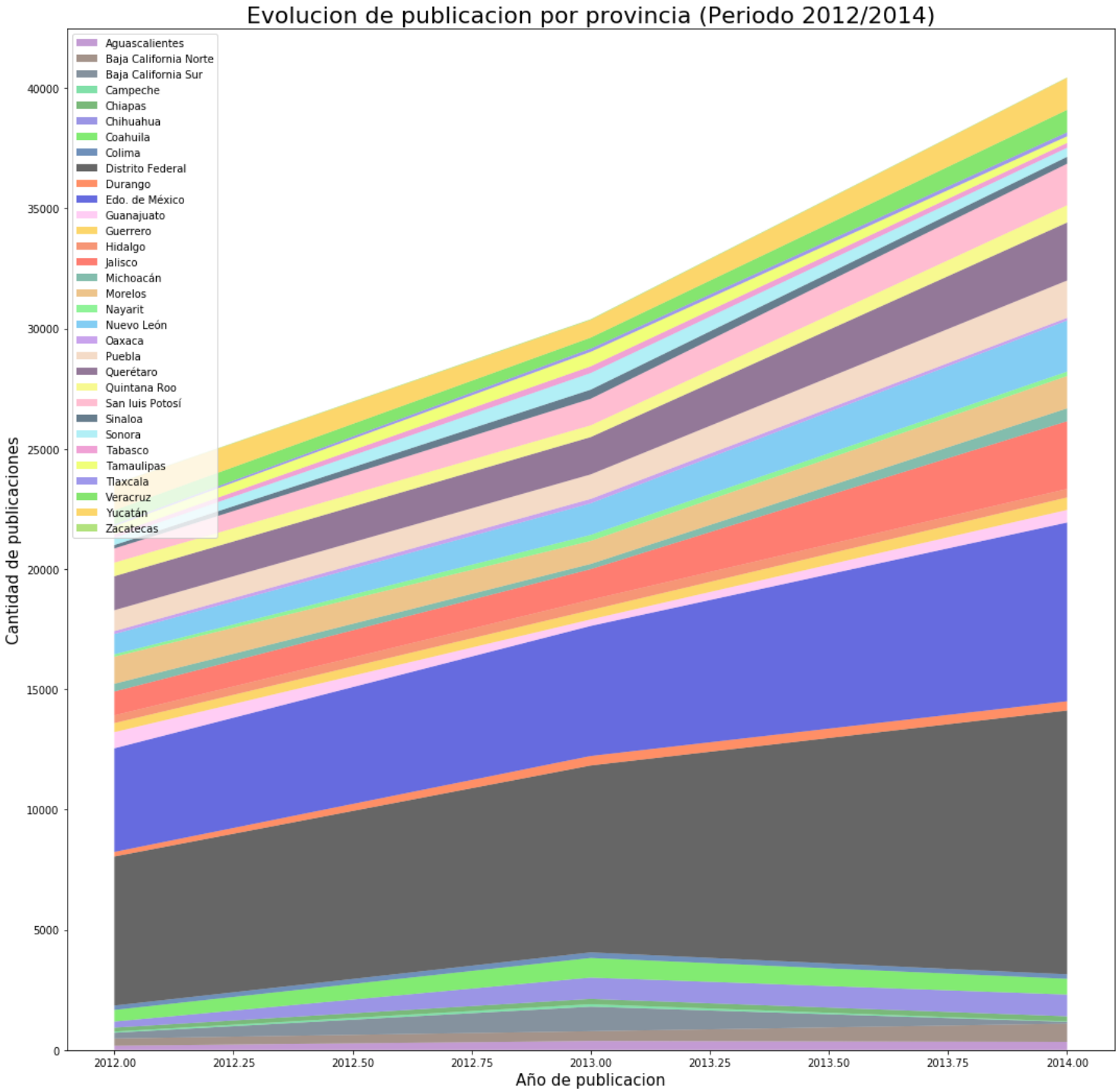

Gráfico 3.1.A

En el periodo 2012/2014, que tiene una evolución lineal, se puede ver que tanto Distrito Federal como Edo. de México están en el top de provincias que más publicaciones aportan y por otro lado provincias como Zacatecas, Oaxaca y Campeche están entre las que menos publicaciones tienen. En tanto Jalisco, Querétaro, Puebla y Nuevo León tienen una cantidad buena de anuncios, pero no se le acercan a Distrito Federal.

3.1.B Tramo 2014/2016

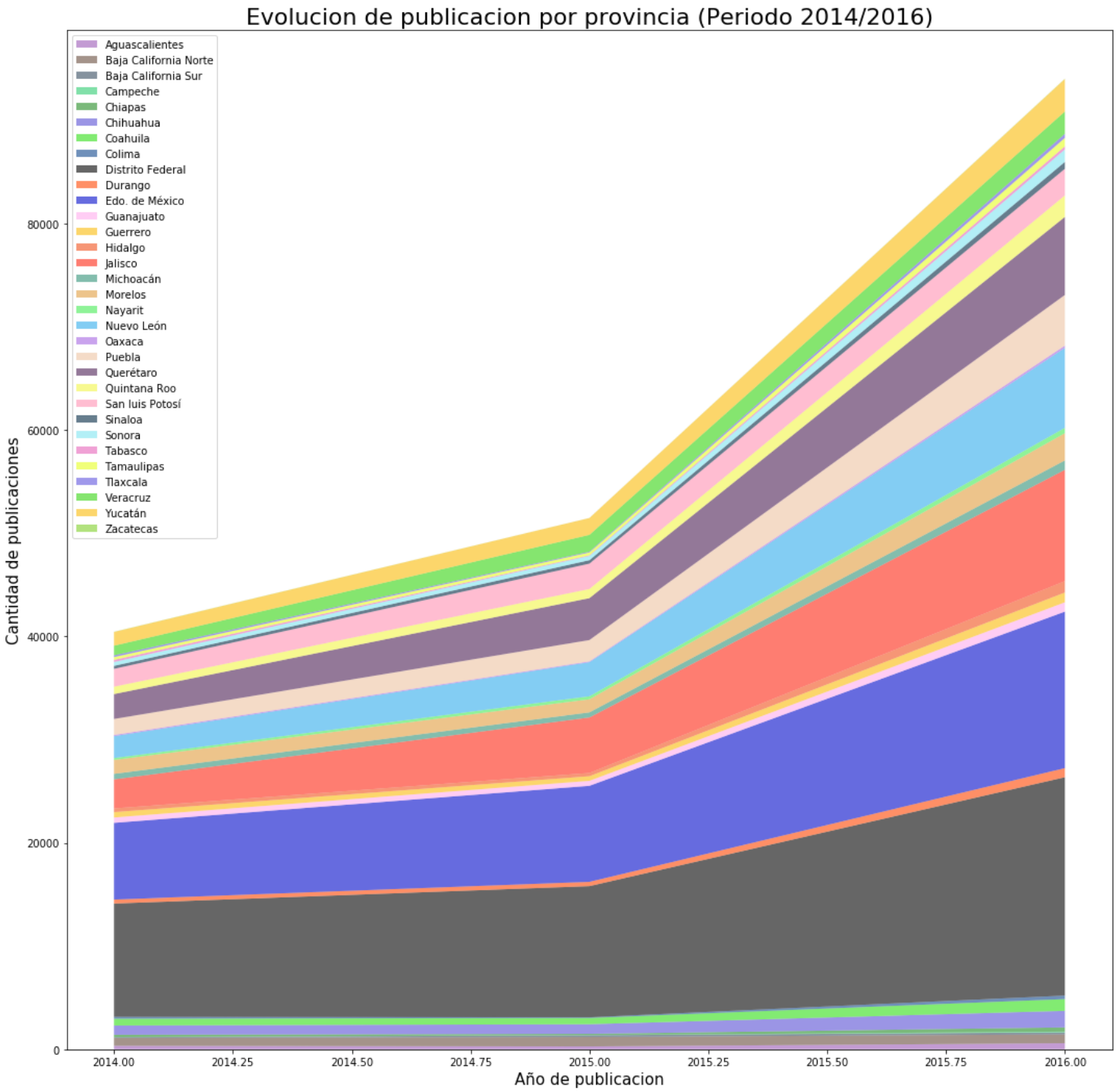

Gráfico 3.1.B

En este último gráfico que corresponden a publicaciones del periodo 2014/2016, las aportaciones en cantidad de anuncios de Distrito Federal y Edo. de México siguen arriba, pero tanto Jalisco, Querétaro, Puebla y Nuevo León tuvieron un aumento en cantidad de anuncios bastante notable.

Aunque en realidad casi todas las provincias, a excepción de las más populares (DF y Edo de Mex), casi duplicaron sus publicaciones. Las más importantes en número de publicaciones son las anteriores mencionadas.

3.2 Publicaciones por provincia

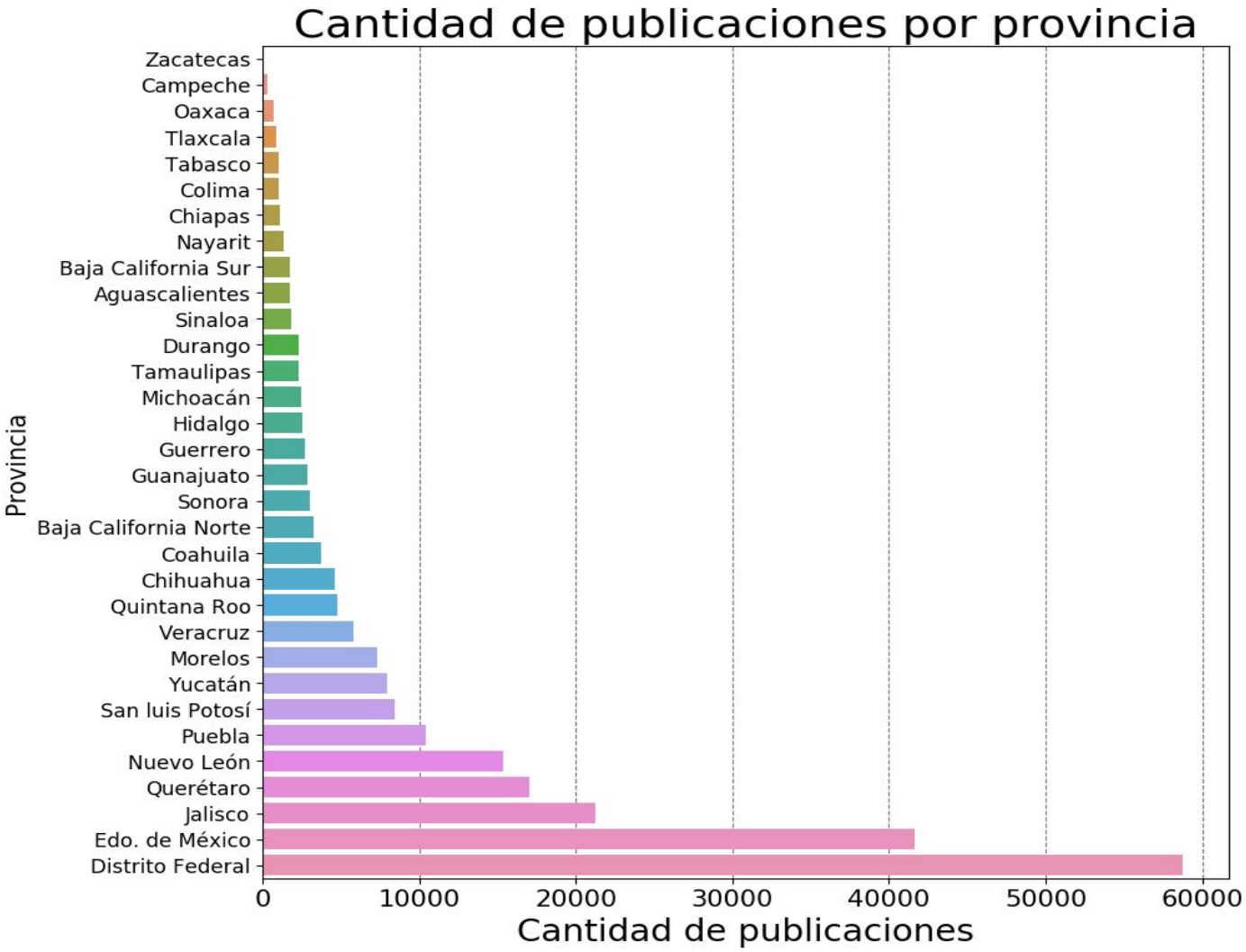

Gráfico 3.2

Ahora que analizamos las correlaciones entre las variables vamos a analizar la cantidad de publicaciones por provincia. Claramente las publicaciones predominan entre las primeras 5 ciudades, siendo Distrito Federal la provincia con más publicaciones, llegando alrededor 60000 publicaciones, y en segundo puesto, la provincia con más publicaciones, es Edo. de México, que tiene un poco más de 40000 publicaciones.

3.3 Publicaciones por ciudad

En cuanto a las ciudades nos vamos a quedar con el top 15 de ellas en cuanto a las publicaciones, por el hecho de que son 876 ciudades, y algunas de ellas solo tienen una publicación.

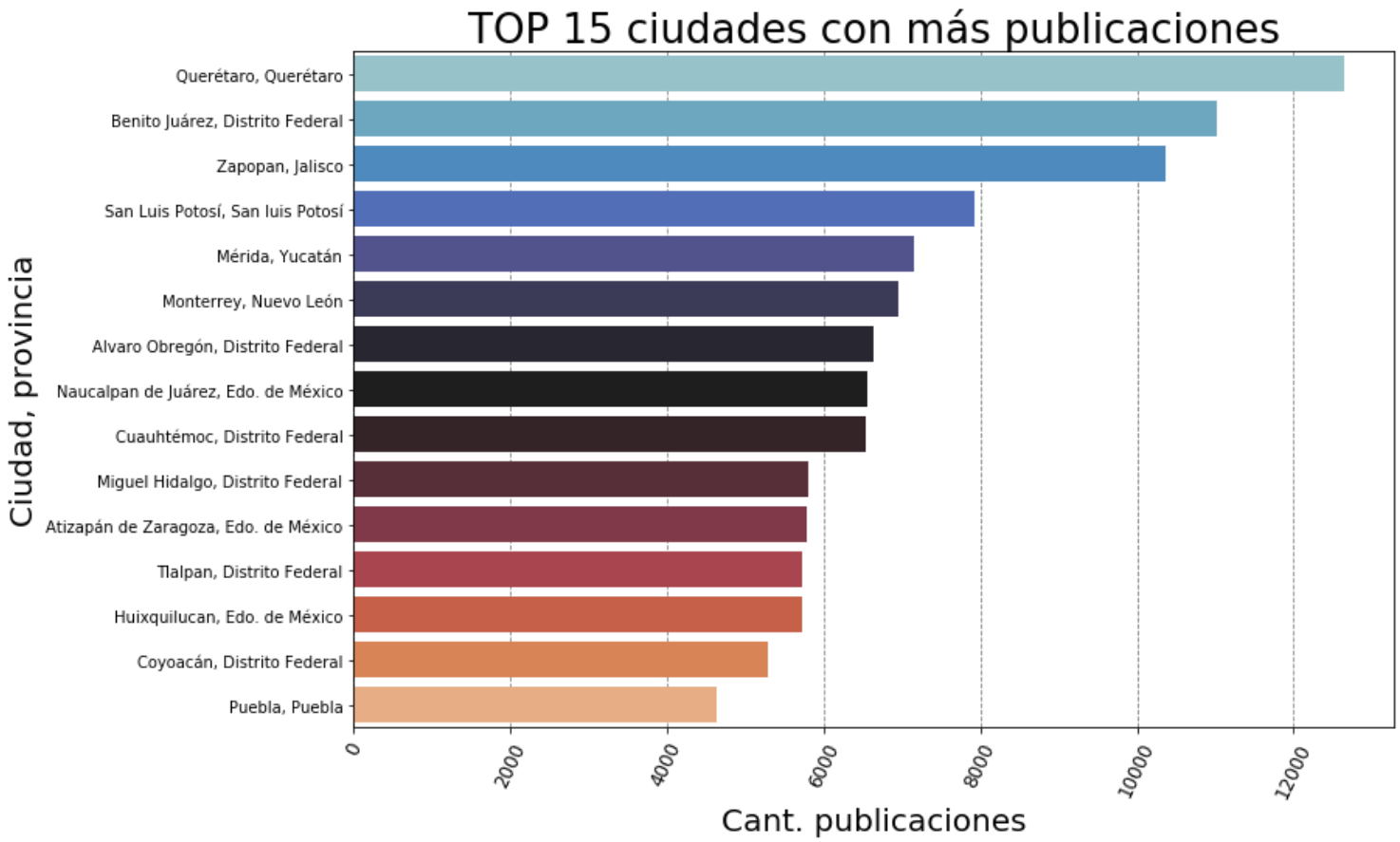

Gráfico 3.3.1

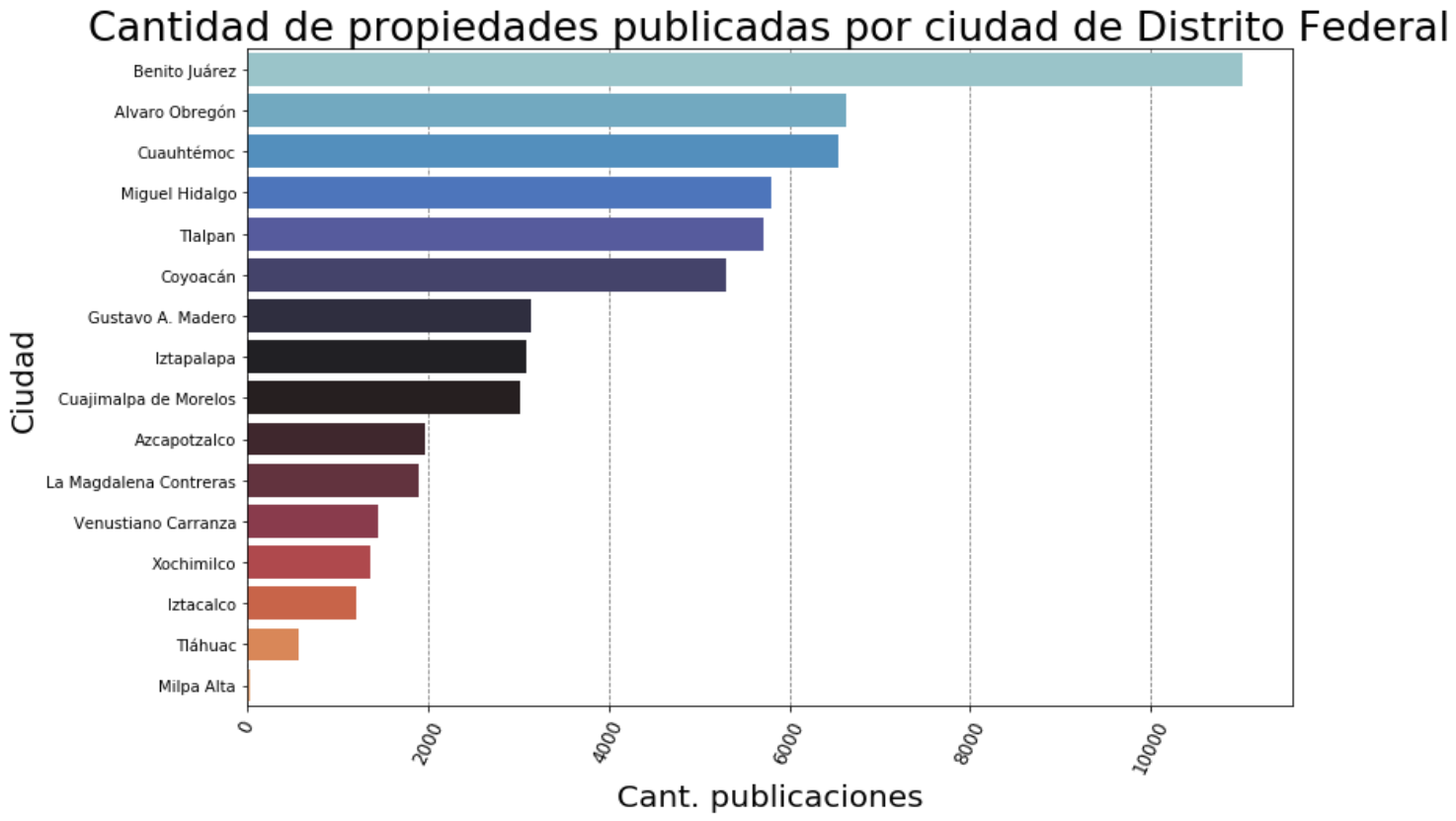

Gráfico 3.3.2

En cuanto a las ciudades se puede ver que las ciudades con más publicaciones son muy variadas en comparación con las publicaciones por provincias. Uno podría llegar a pensar que Distrito Federal tendría más publicaciones por ciudad pero sus publicaciones están distribuidas en todas sus ciudades, siendo Querétaro la ciudad con más publicaciones ya que la mayoría de las publicaciones de Querétaro están agrupadas en su capital, siendo así la primer ciudad con más publicaciones. La ciudad Querétaro tiene más de 11000 publicaciones y cómo se puede ver en el gráfico 3.1, la provincia Querétaro tiene un poco más de 15000 publicaciones nada más, por eso mismo se puede decir que están agrupadas en una única ciudad.

En cuanto a Distrito federal sus publicaciones se dividen en varias ciudades, cómo se puede ver en el segundo gráfico, al ser la provincia con más publicaciones, varias de sus ciudades aparecen en el TOP 15 de ciudades.

3.4 Provincias con menos publicaciones

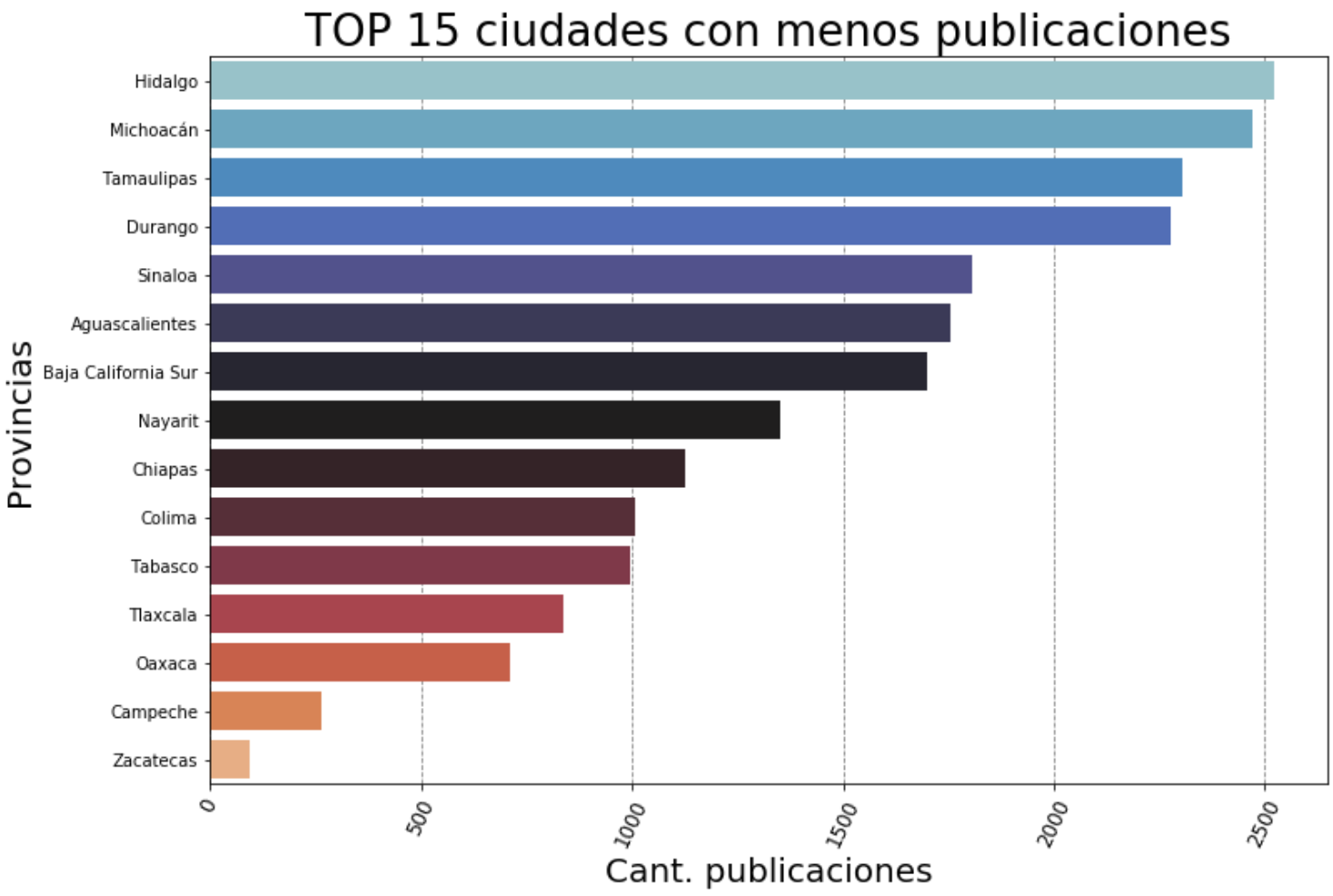

Gráfico 3.4

Agrandando la cola del gráfico 3.1, se puede ver con más detalle las provincias con menor cantidad de publicaciones. Cómo ya habíamos visto cuáles eran, ahora podemos observar la cantidad que tienen, ya que Zacatecas en el gráfico 3.1 parece que no tenía ninguna publicaciones. Siendo la que más tiene de este TOP 15, es Hidalgo con tan solo 2500 publicaciones, tiene menos publicaciones que la última ciudad del TOP 15 del gráfico 3.3.1

3.5 Publicaciones por tipo de propiedad

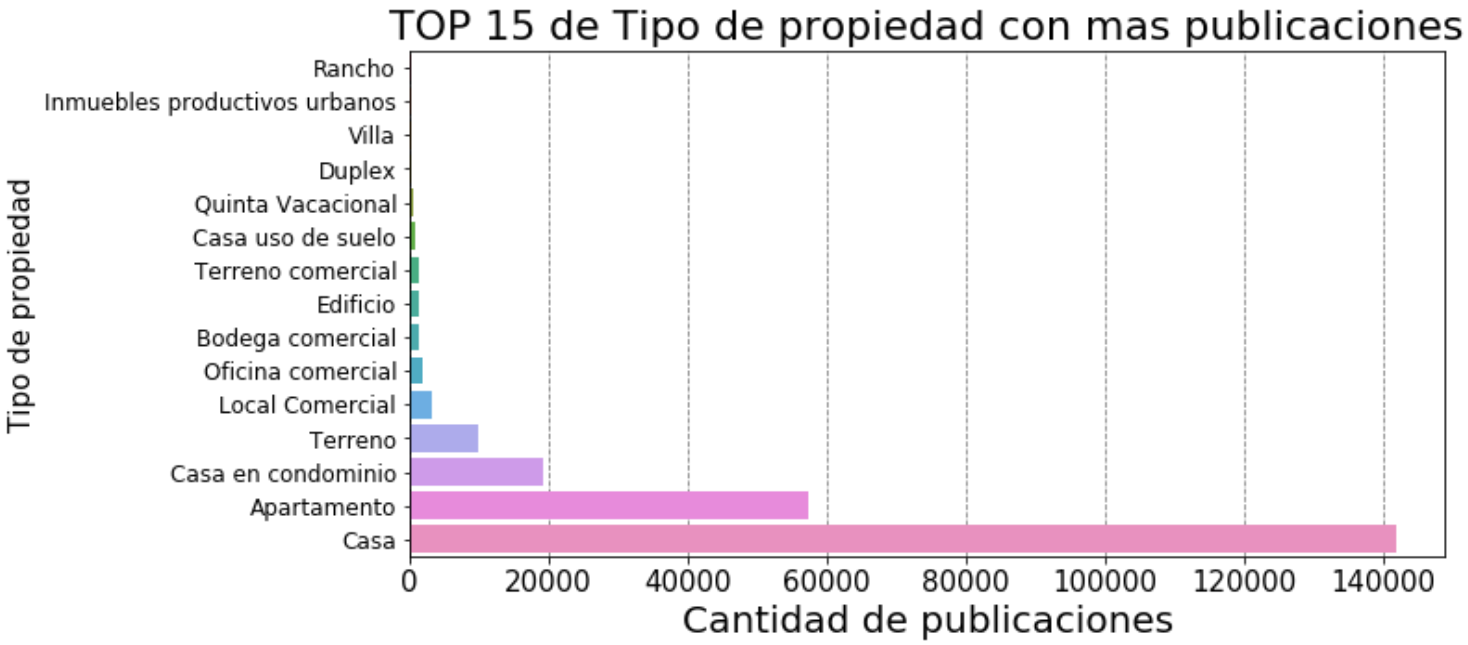

Gráfico 3.5.1

Esquematizando los tipos de propiedades por publicación podemos ver que en el primer puesto del ranking están las Casas con 140000 publicaciones y en segundo lugar los Apartamentos con un poco más de 50000.

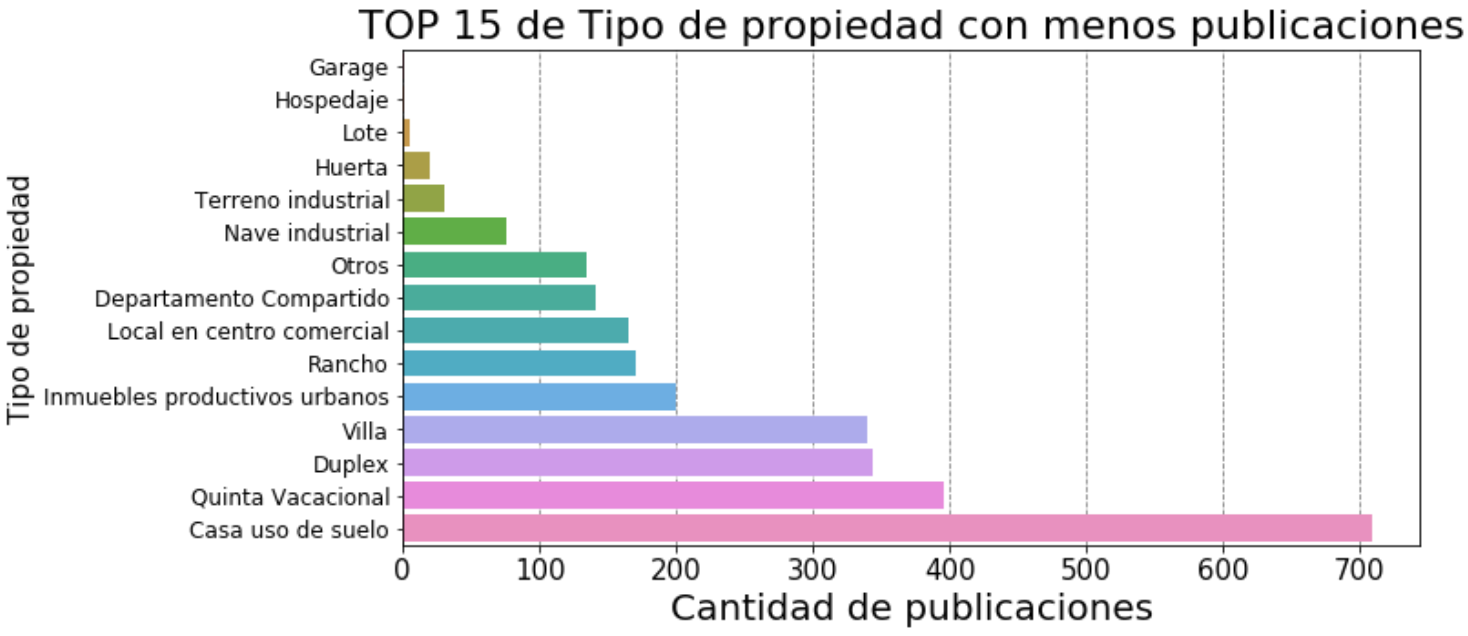

Gráfico 3.5.2

Podemos ver los 15 tipo de propiedad con menor cantidad de publicaciones, el primer puesto con tan solo 700 publicaciones, y los últimos con tan pocas publicaciones que no son visibles en esta escala tan chica.

4. Cantidad de publicaciones por fecha

Analizaremos ahora, la cantidad de publicaciones por su fecha de publicación, para poder ver cual es el momento en el cual hay más publicaciones y llegar a alguna conclusión viendo cuando se publican. Hay que ver también que hay meses que tienen más días que otros, por eso el dia 31, pueden llegar a ser los días con menos publicaciones por el hecho de que solo hay 7 en todo el año, en vez de 12.

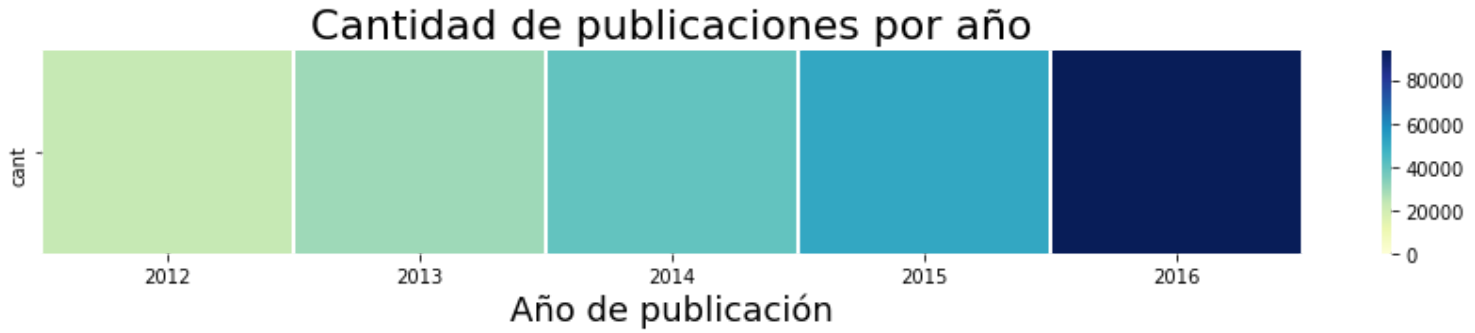

Gráfico 4

En una primera impresión podemos ver que a medida que pasaron los años hay mayor cantidad de publicaciones, siendo el predominante el 2016, frente a los anteriores, aunque todos superan las 20000 publicaciones, el 2016 tiene más de 80000 publicaciones.

4.1 Cantidad de publicaciones por año y por mes

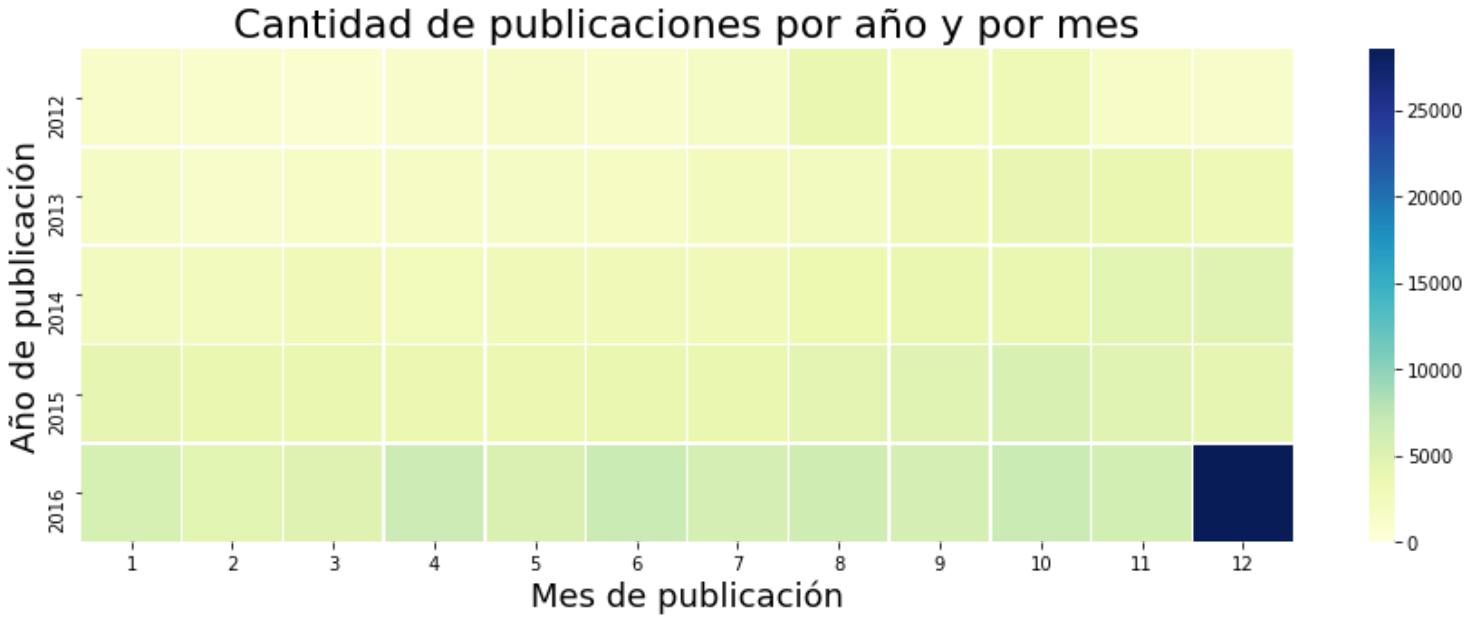

Gráfico 4.1

En una primer impresión podemos ver que en diciembre del año 2016 se realizaron una gran parte de las publicaciones, aunque durante todo el 2016 se realizaron una gran parte de las publicaciones más que en otros años, podemos decir que en el 2016 la página se hizo más conocida y la gente publicó más casas durante ese año.

4.2 Cantidad de publicaciones por día del mes.

Un análisis interesante para ver es que dia se realizaron más publicaciones y analizar especialmente si únicamente diciembre es el mes con más publicaciones o en qué momento del mes se realizan más publicaciones.

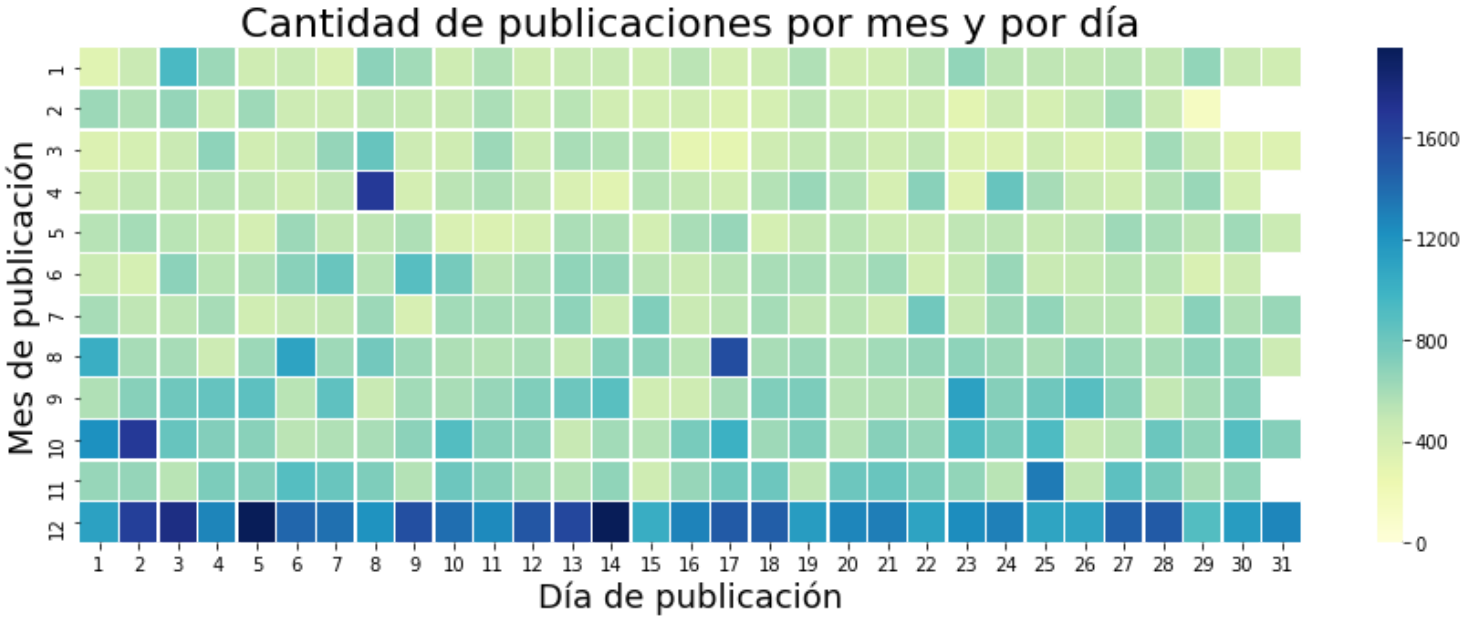

Gráfico 4.2

Podemos ver claramente en el gráfico 4.2 que en el mes de diciembre predominan las publicación de las propiedades, y durante todo ese mes se realizan más cantidad de publicaciones. Hay algunas excepciones que se realizan en algunos meses y fechas exactas, por ejemplo en el mes de abril, el día 8 se realizaron muchas publicaciones. En cuanto a los meses con menos publicaciones se puede observar que los primeros meses hasta el mes 8 (agosto) son los meses con menos publicaciones, en comparación con los últimos 4 meses que se realizan más publicaciones.

4.3 Publicaciones en 2012

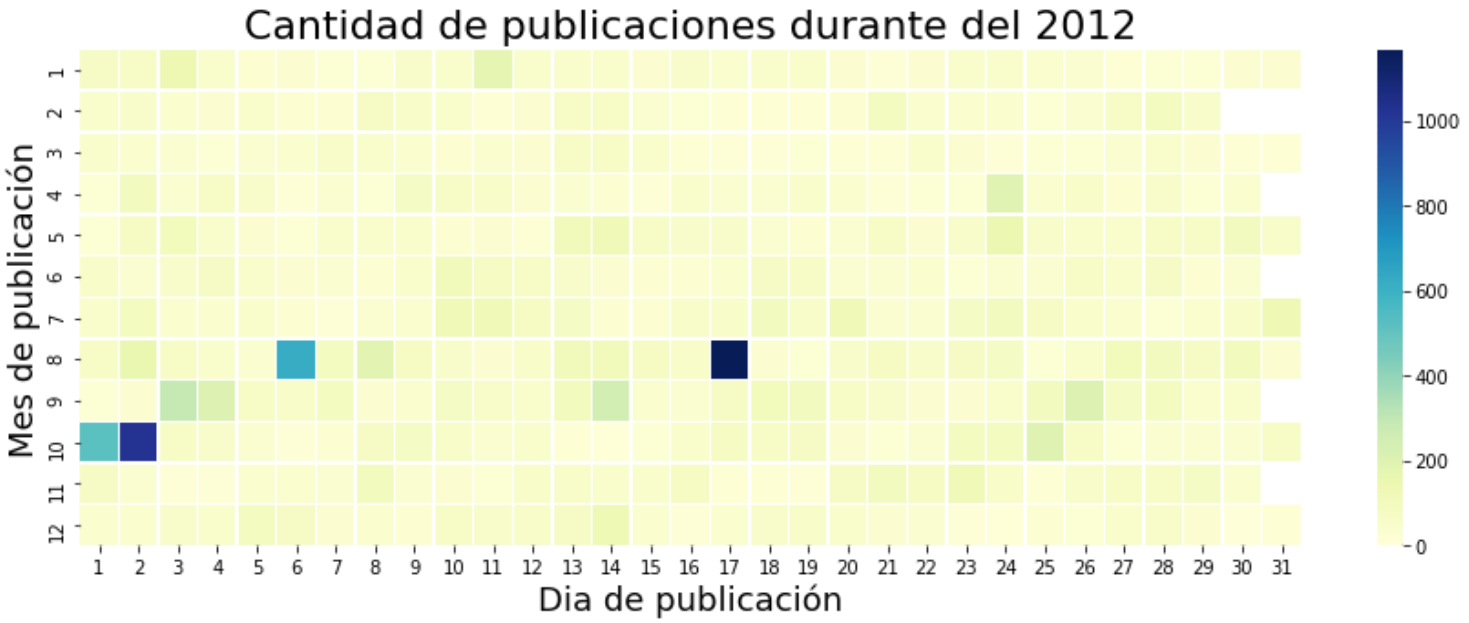

Gráfico 4.3

En todo el 2012 se realizaron muy pocas publicaciones, durante el año no hubo publicaciones a excepción del 6 y 17 de agosto, y 1 y 2 de octubre. Los días que hubo muchas publicaciones son pocos, pero se realizaron más de 600 publicaciones hasta 1200 publicaciones.

4.4 Publicaciones en 2013

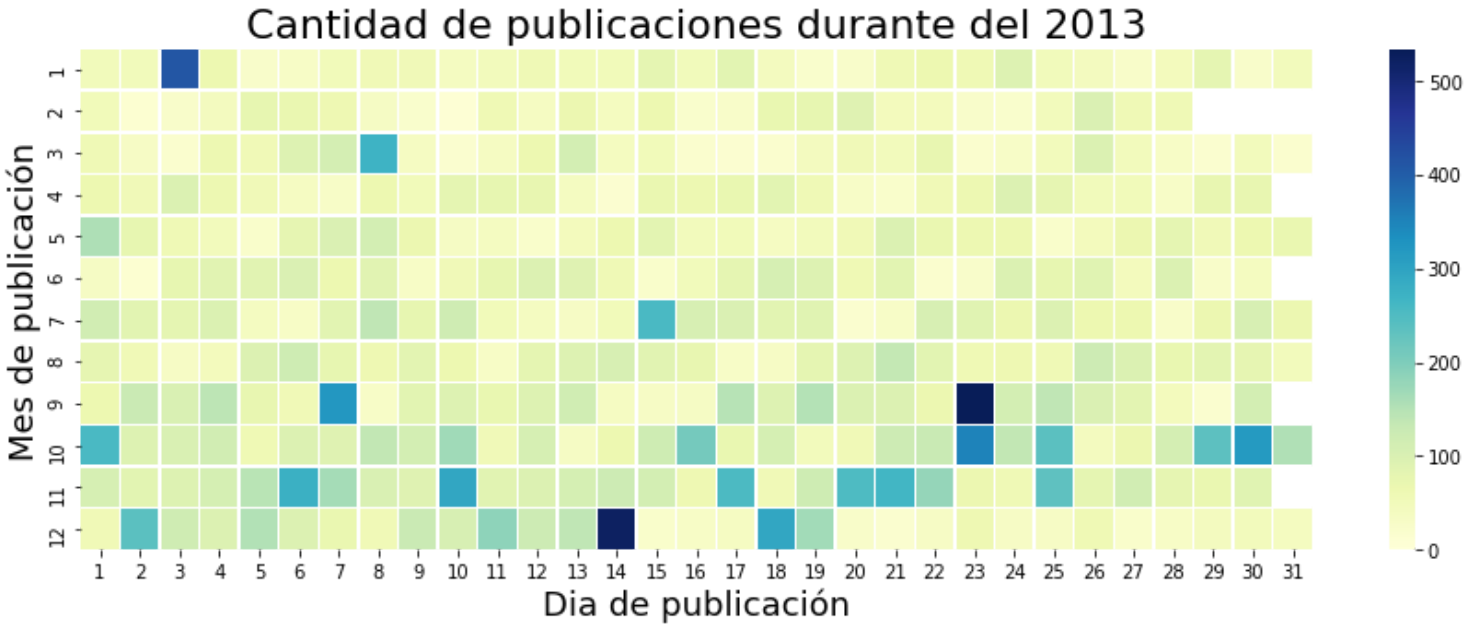

Gráfico 4.4

En este gráfico se puede ver que solo hay un par de días en el 2013 que se realizaron muchas publicaciones. Se ve en especial para los últimos tres meses que hay muchas más publicaciones que a lo largo

de todo el año y un par de dias en especifico como por ejemplo el 3 de enero, y el 23 de septiembre pero con tan solo 500 publicaciones como máximo.

4.5 Publicaciones en 2014

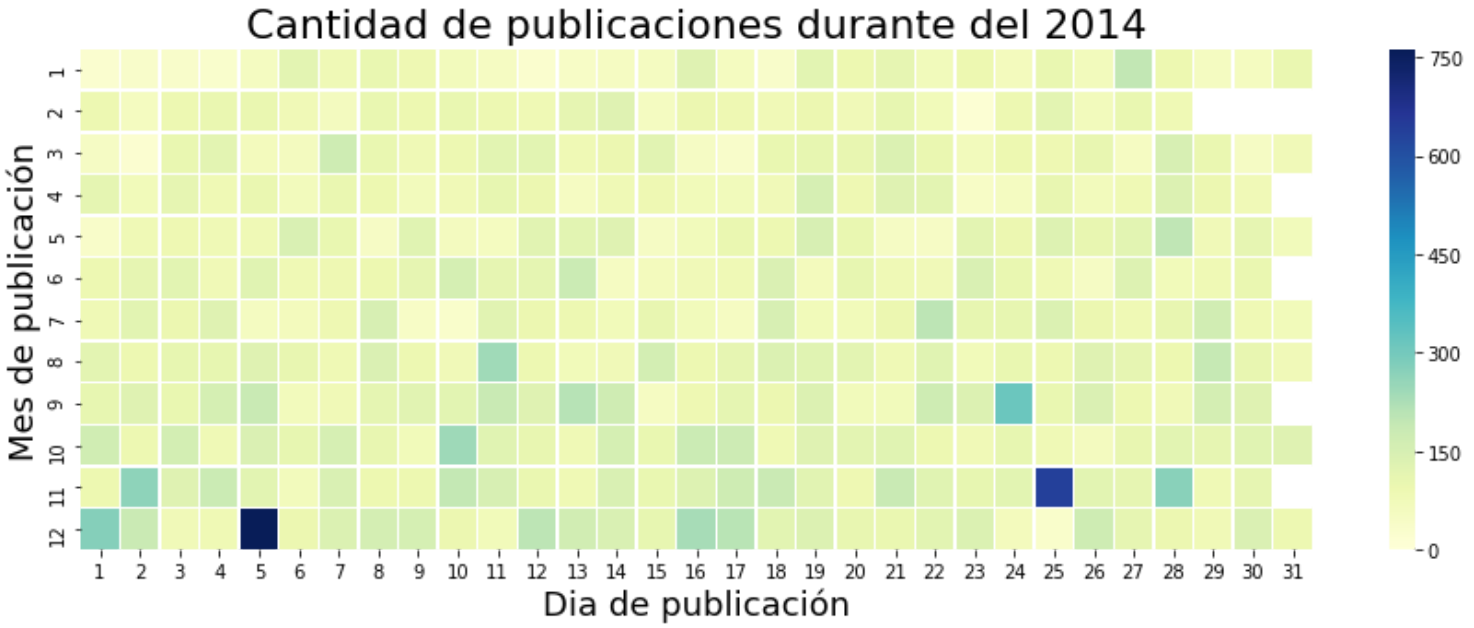

Gráfico 4.5

En el 2014, hubo una caída en cuanto a las publicaciones, se puede ver fácilmente que hubo más publicaciones en el 2012 que durante el 2014. Pocos días hubo muchas publicaciones y con un máximo de 750 publicaciones, menor al del 2012, este año fue un mal año en cuanto a publicaciones. Durante los dos primero cuatrimestres no se realizaron publicaciones.

4.6 Publicaciones en 2015

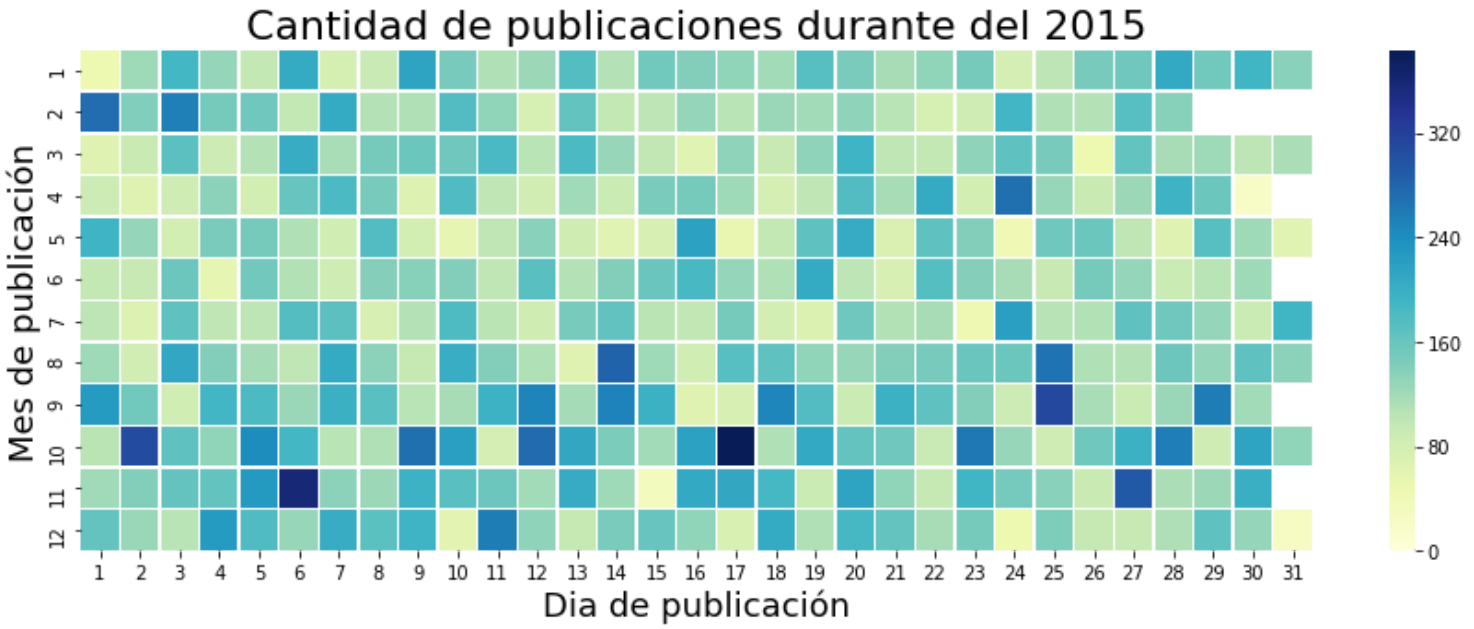

Gráfico 4.6.1

Gráfico 4.6.2

A diferencia de los años anteriores con publicaciones muy concentrada en pocos días, el 2015 fue un año en el cual durante todos los días se publicaban propiedades uniformemente, hubo 10000 más publicaciones que en el 2014 pero estas fueron distribuidas en todos los días del mes, aunque ninguna superando las 400 publicaciones en el dia.

En el gráfico 4.6.2 se puede ver bien la diferencia de publicaciones entre cada mes siendo octubre el mes con más publicaciones. Por más de que los otros meses tienen muchas publicaciones, mínimo 3600, el mes de octubre tuvo más 5500 publicaciones

4.7 Publicaciones en 2016

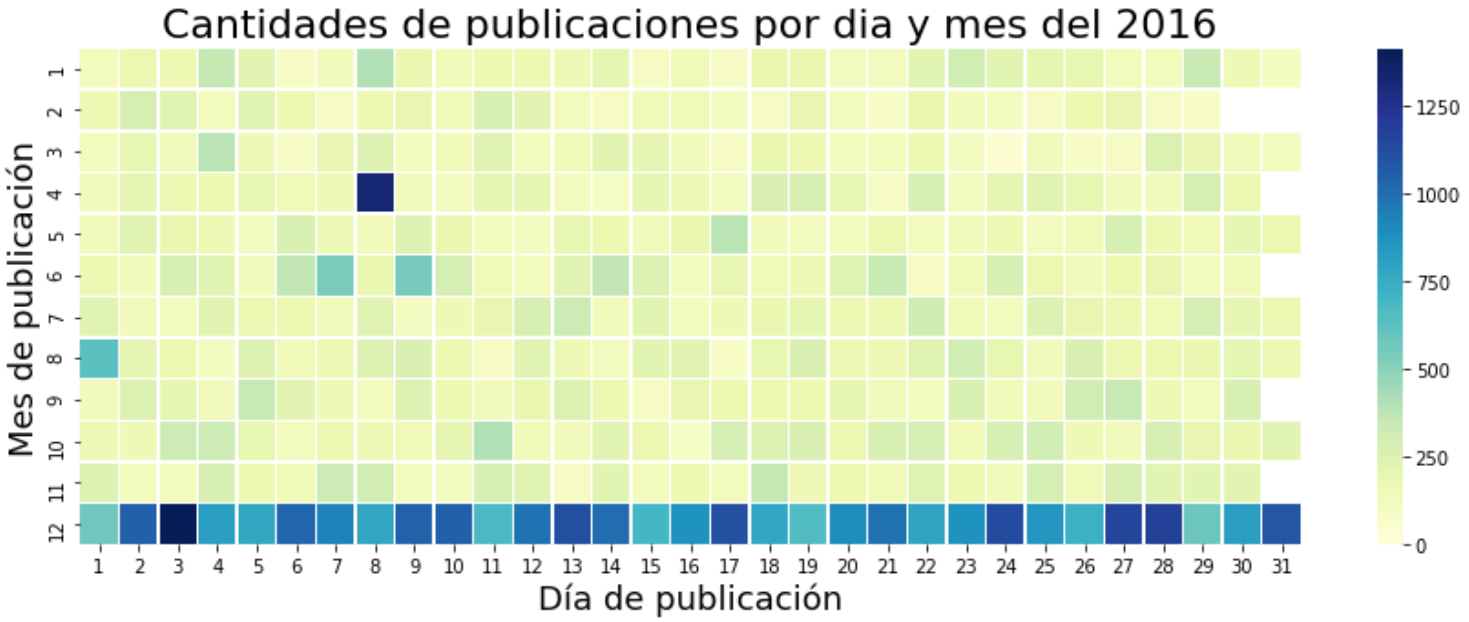

Gráfico 4.7

Para ver un poco más de cerca el gráfico 4.1, queremos ver en especial cómo se comporta el mes de diciembre que es claramente el mes con más publicaciones de todos los años. Durante todo el año no hubo muchas publicaciones excepto el 8 de abril, y el 1 de agosto, durante los demás días del año fueron pocas las publicaciones de propiedades a lo largo del año hasta diciembre.

En el mes de diciembre, la página explotó de publicaciones, en comparación con el resto del año, se realizaron gran parte de las publicaciones analizadas en todo el set de datos, llegando a las 1400 publicaciones y como mínimo 580 publicaciones en un dia. Podemos decir que en el 2016 la página se volvió aún más conocida.

5. Precio del metro cuadrado

El tópico más interesante para hacer es el precio del metro cuadrado (m^2) de cada propiedad. Para este análisis se utilizará el valor en dólares (USD) para saber de cuanto dinero se está manejando por propiedad. La tasación que se utilizó fue que 1 dolar son 20 pesos mexicanos.

Para esta sección, como hacemos un análisis en base a los atributos “metroscubiertos” y “metrostotales”, que como vimos al inicio tenían valores nulos en algunos de los registros, procedimos a completarlos.

El plan fue el siguiente:

- si vienen ambos campos, no hacemos nada,
- si viene solo “metroscubiertos”, completamos “metrostotales” con “metroscubiertos”,
- si viene solo “metrostotales”, completamos “metroscubiertos” con “metrostotales”
- si no viene ninguno de los campos, descartamos dicha publicación. Afortunadamente, no nos fue necesario desechar datos.

Dicho esto, procedimos con el análisis pertinente.

5.1 Comportamiento.

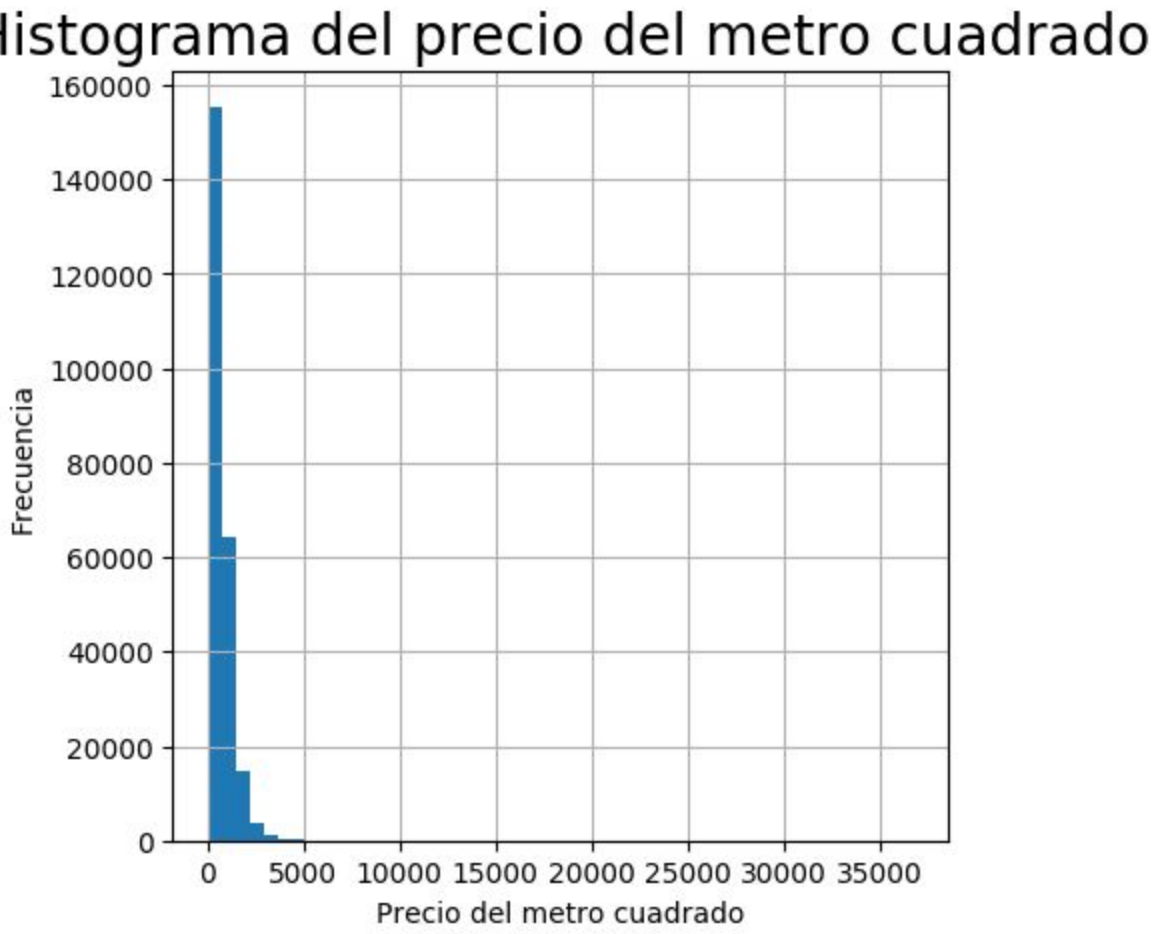

Gráfico 5.1.1

Primero realizamos un histograma, para poder analizar en qué estado se encuentra el dataset. Nos encontramos con una gran cantidad de outliers, representado por una larga cola hacia la derecha. Esto es: propiedades que tienen un precio muy fuera de la media que maneja el set de datos.

Para este análisis en particular, luego de ver los percentiles de nuestro DataFrame, acotamos y desechamos el 1% máximo que era el que generaba la cola larga del histograma anterior. De esta forma, desechamos 2402 propiedades para que no nos contamine el análisis, y volvimos a graficar para observar el cambio:

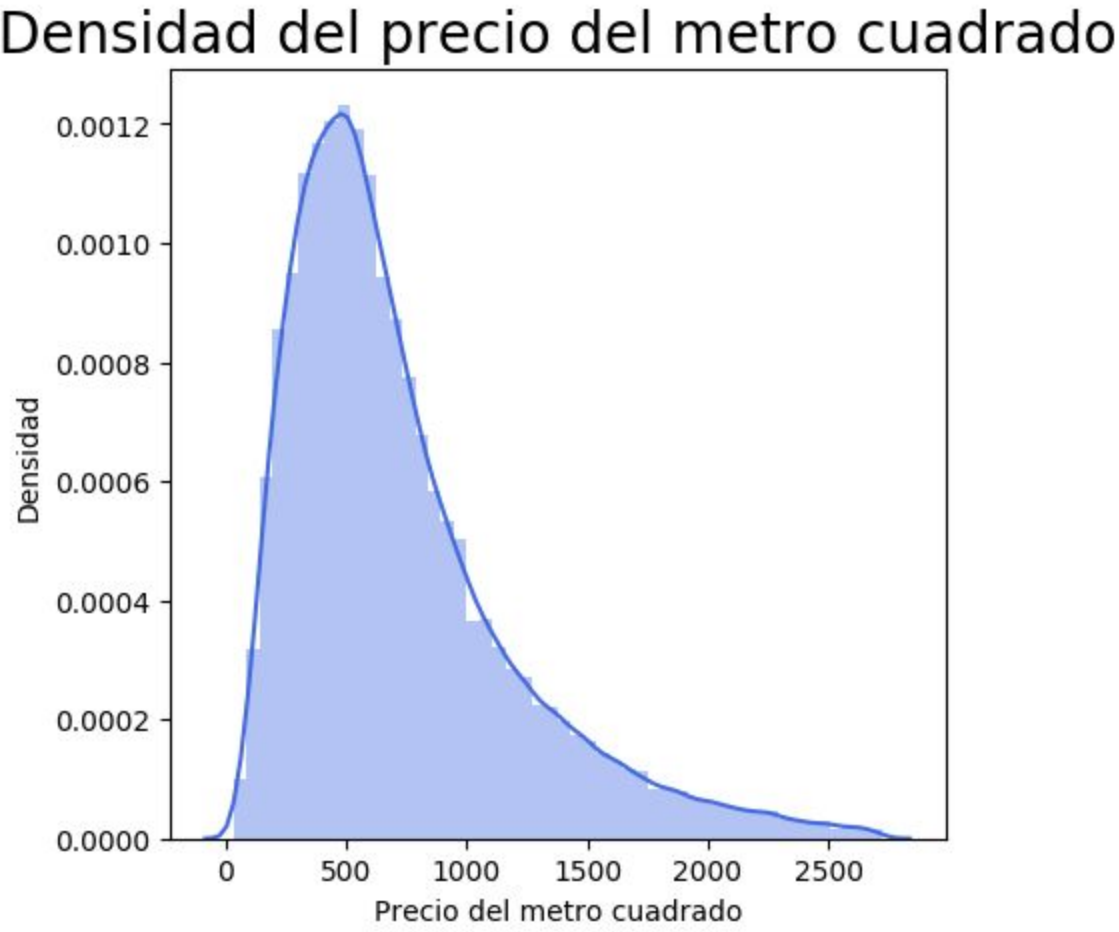

Gráfico 5.1.2

Como se puede ver en el gráfico de densidad, una vez eliminados estos outliers, notamos un pico en 500 dólares, que luego cae drásticamente hasta un poco más de los 2500 dólares por metros cuadrados. Aun habiendo filtrado todos los valores anómalos, se ve que queda una cola larga hacia la derecha, pero representativa del set de datos.

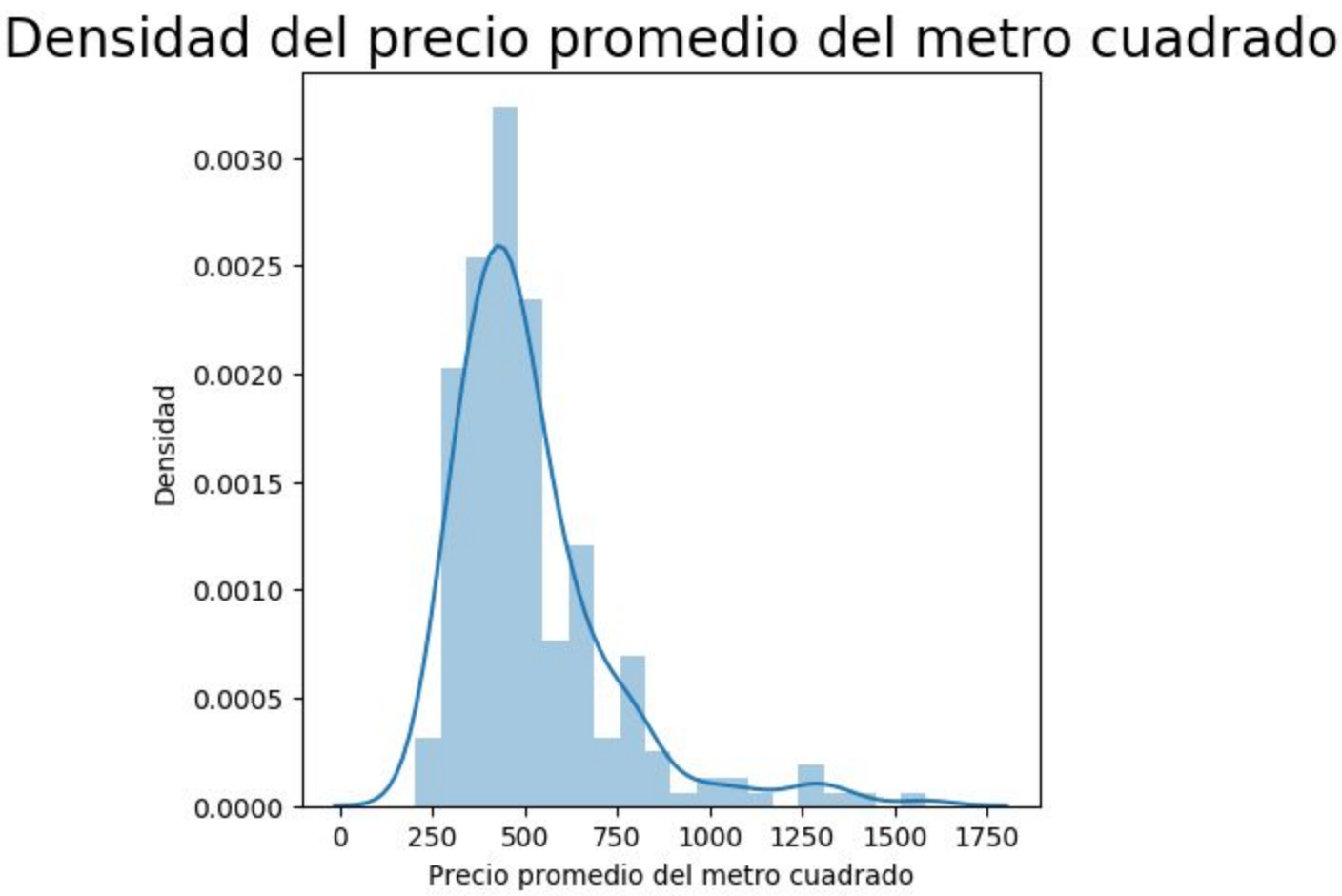

Gráfico 5.1.3

Congruente con el gráfico anterior, se destaca una moda de metro cuadrado en 500 dólares. Nótese lo próxima al valor mínimo que es la moda, dejando a interpretación, que vendedores prefieren cobrar más caro el m^2 que reducirlo en pos de una más rápida venta. También podría culparse al inflar del metro cuadrado, por los distintos “amenities” que dan valor agregado a la propiedad. Éstos dos factores creemos que son en gran medida, los que producen los precios de metro cuadrado dispersos en la cola de promedios mayores.

5.2 Por provincia.

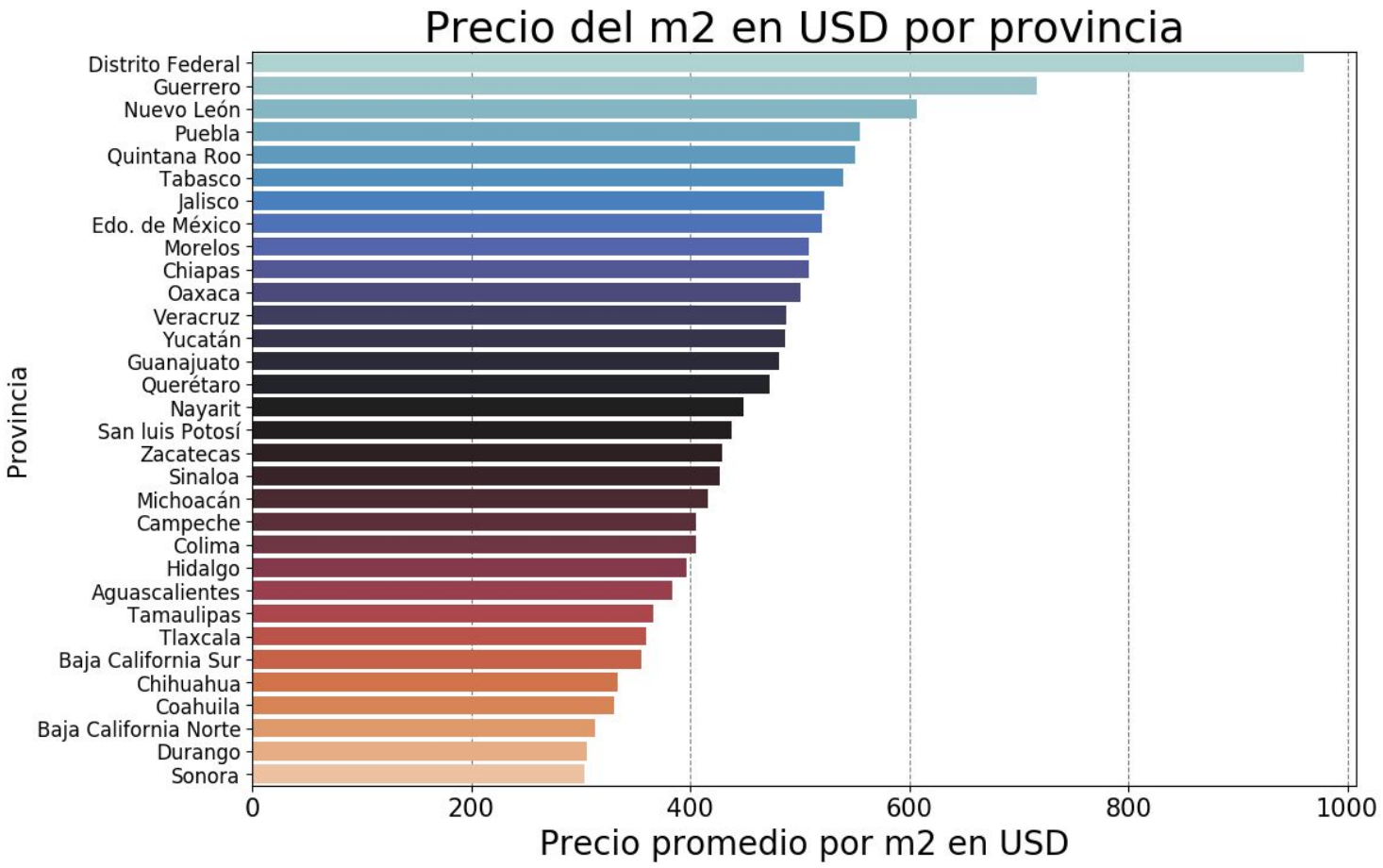

Gráfico 5.2

Para este análisis, para no caer en un análisis poco fidedigno, filtramos todas aquellas ciudades que tenían menos de 25 propiedades, desechando así un 2,5% del dataset. Si no hubiésemos realizado esto, podrían haber aparecido ciudades y provincias cuyo precio promedio no corresponda con la realidad, ya que al tener muy pocos registros para la misma, si la muestra fuera más grande, no veríamos los mismos valores.

En una primera impresión se puede ver cómo el precio promedio del metro cuadrado es muy alto para Distrito Federal en especial y seguido de Guerrero. Cabe destacar que Distrito Federal, como se mencionó anteriormente, es la provincia con mayor cantidad de publicaciones, y también con el mayor valor promedio del metro cuadrado. En cambio Guerrero, tiene un precio promedio de 700 dólares, siendo la 16ava provincia con menor cantidad de publicaciones, pero cada una de esas publicaciones tienen un valor elevado del metro cuadrado dejándola en el segundo puesto.

La provincia con el menor precio por metro cuadrado es Sonora, con un valor de 300 dólares por metro cuadrado aproximadamente y no muy lejos la provincia de Durango, con solo un poco más de 300 dólares.

5.2.A Distrito Federal

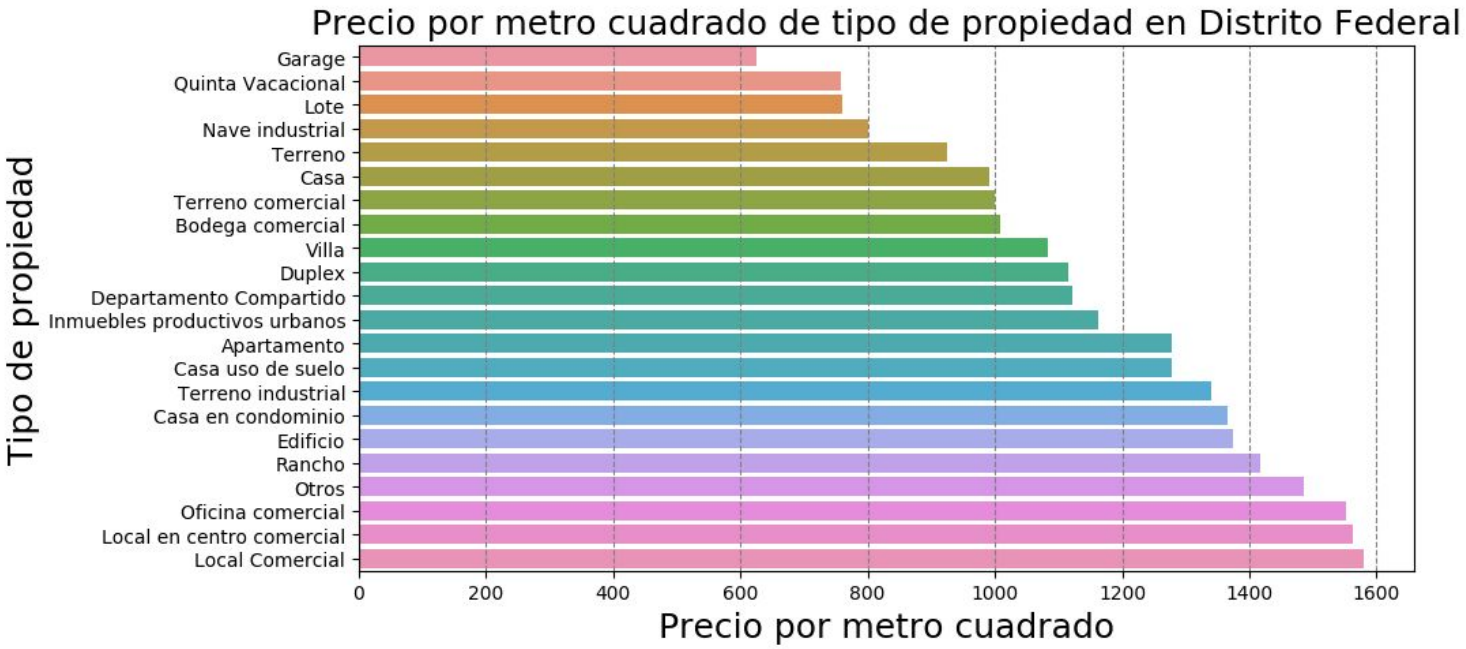

Gráfico 5.2.A

Indagando un poco más de acerca de Distrito Federal, viendo cuáles son los tipos de propiedades de mayor precio por metro cuadrado, encontramos que en Distrito Federal el top 3, lo encaran los comercios, gracias a esto podemos ver qué Distrito Federal por ser la capital, los precios de los comercios son más caros que las propiedades residenciales. Cabe destacar que las Casas tienen un precio menor que el de los Apartamentos.

5.2.B Guerrero

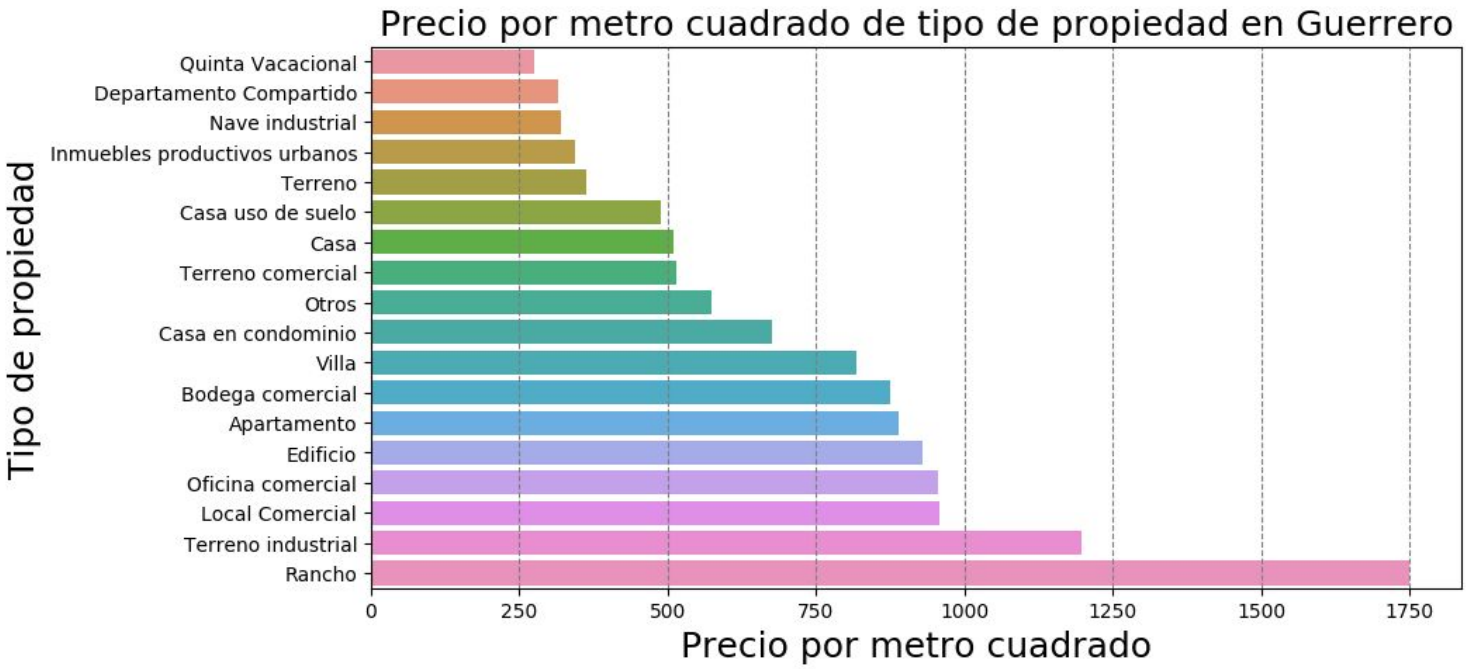

Gráfico 5.2.B

En Guerrero, a diferencia de Distrito Federal, podemos observar que hay un tipo de propiedad de mucho mayor precio que el resto de los tipos de propiedad. Los ranchos, con un valor de precio por metro cuadrado promedio de 1750 dólares, colocan a Guerrero en una posición alta en el gráfico 5.2 siendo de las provincias con menor cantidad de publicaciones, y con un precio por metro cuadrado muy alto.

Dejando de lado los ranchos, se puede ver que los Terrenos industriales tienen un gran valor por metro cuadrado, al igual que las propiedades comerciales con la excepción de terreno comercial.

5.2.C Geográfico

Gráfico 5.2.C

Viendo el mapa de México, y el valor por metro cuadrado se puede ver que las provincias más caras son aquellas que están al sur, cerca de la capital. Se puede ver Distrito Federal, como una pequeña ciudad en el medio de México y cerca de ella Guerrero. En cuanto a Sonora y Durango, ambas se encuentran al norte, estas tienen una considerable superficie en comparación con Distrito Federal.

Nuevo León, una de las provincias al Norte, ocupa el tercer puesto en el ranking por provincia del valor del metro cuadrado (Ver gráfico 5.2), de notoria diferencia por su color entre las provincias aledañas. Esta se destaca por su gran tamaño y por estar rodeado de provincias de mucho menor valor por metro cuadrado.

5.3 Según su latitud y longitud

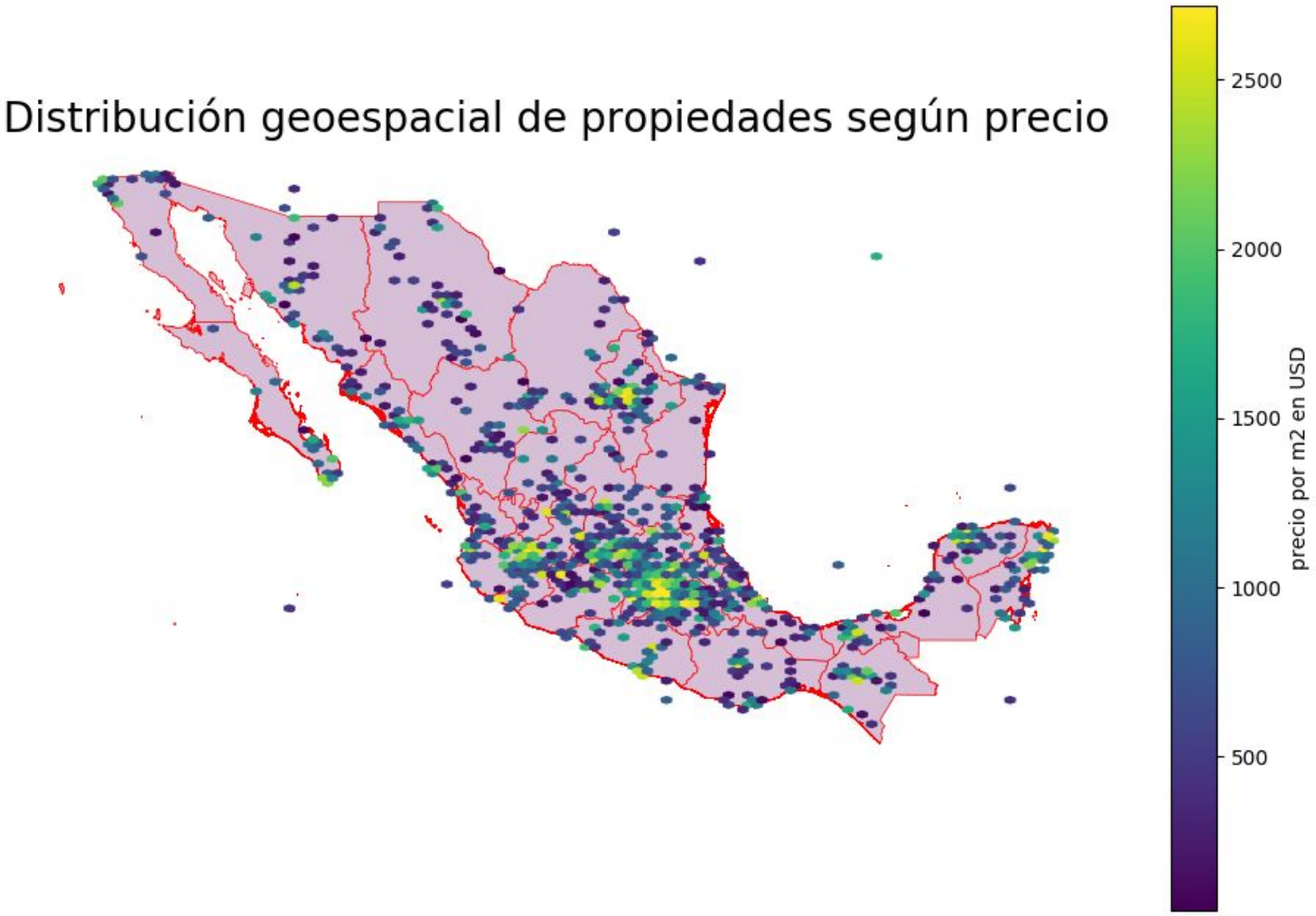

Gráfico 5.3

Como dijimos anteriormente, en este gráfico, no se encuentran todas las propiedades, ya que, no todas tienen bien cargadas su latitud y longitud. Como se puede ver en el gráfico algunas de estas propiedades, según su latitud y longitud, están ubicadas en el mar por más de haber hecho un filtro previamente. Realizamos igualmente el gráfico para ver cómo se distribuía cada propiedad dentro de México, para poder comparar por donde estaban las propiedades.

Vemos que la gran mayoría se encuentra en zonas aledañas a la Ciudad de México, se puede ver a simple vista que Distrito es la que tiene las propiedades agrupadas y de mayor valor. Muchas de las provincias tienen muy focalizadas las propiedades de mayor valor, al igual que Nueva León, pero en cuanto a Guerrero tiene muy distribuidas sus propiedades.

5.4 Según ciudades

Cabe aclarar, como mencionamos anteriormente, que hay ciudades que fueron filtradas ya que no tenían la suficiente cantidad de propiedades (25 propiedades) como para poder considerarlas en el análisis. Dichas ciudades tienen un precio inferior a USD 100 por metro cuadrado, pero no es fidedigno ya que pueden ser casos aislados, como también ciudades con precios altísimos, alcanzando los USD 16000 por metro cuadrado.

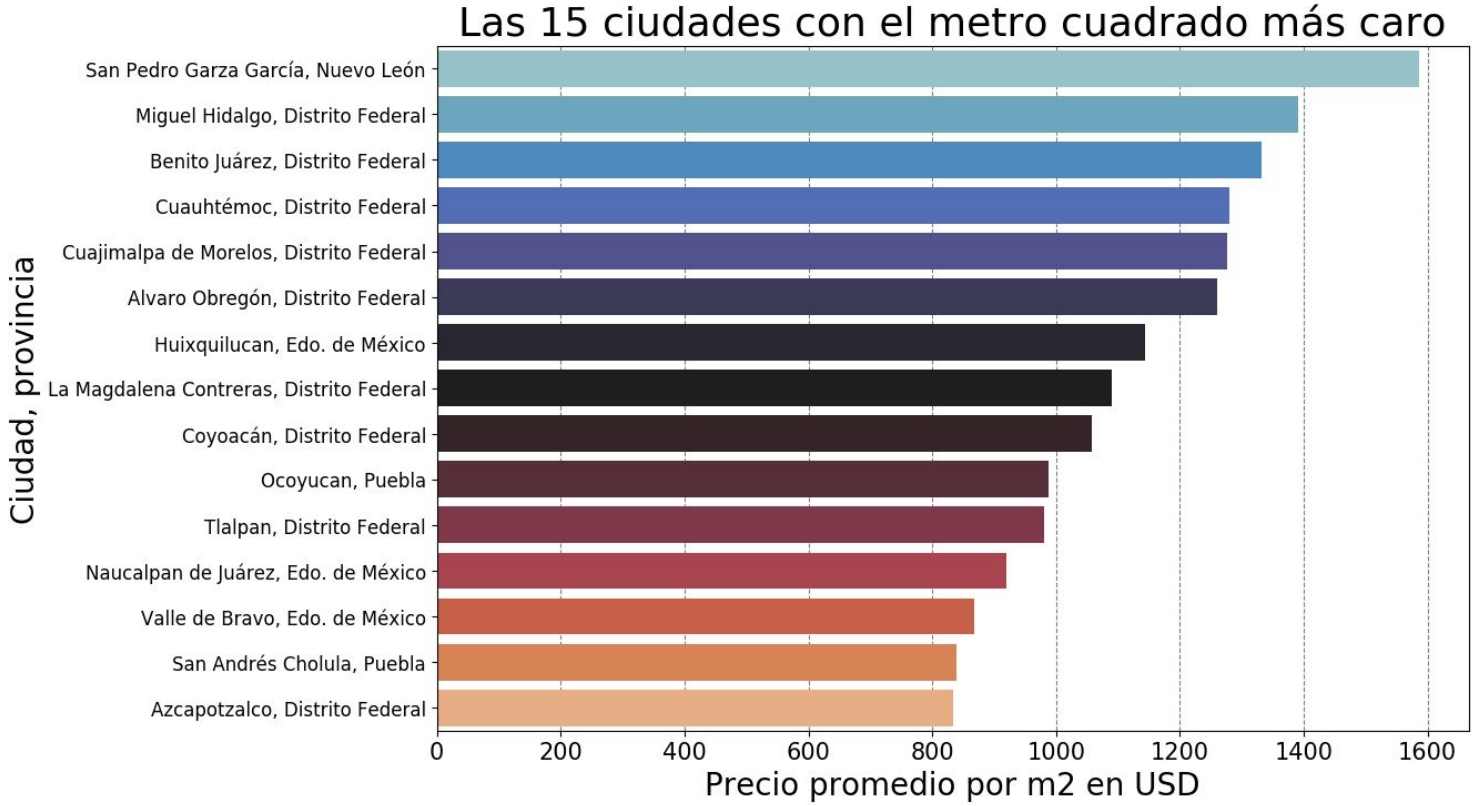

Gráfico 5.4.1

Vemos que las ciudades más caras se encuentran en la provincia más cara, lo cual tiene sentido ya que es la capital del país. Además vemos una gran diferencia en el top, que parece menguar luego de las primeras 5-9 ciudades.

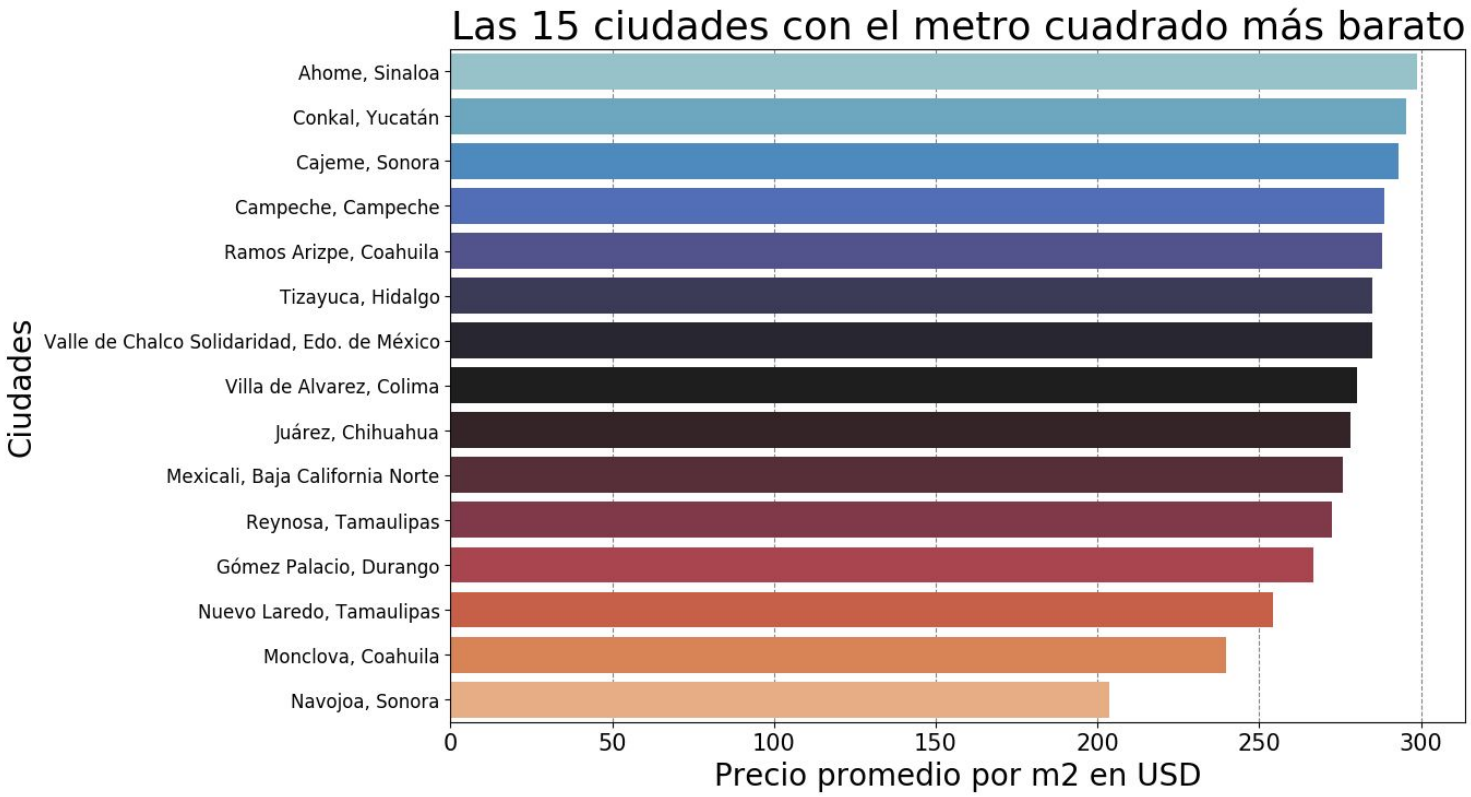

Gráfico 5.4.2

Las ciudades más baratas no parecen seguir el mismo ritmo que las más caras, ya que el precio por metro cuadrado es muy similar (entre los 250 y 300 dólares), salvo en la última, donde la diferencia es un poco más marcada. Como mencionamos anteriormente, hay provincias que tienen un precio promedio más barato, pero que fueron filtradas porque había muy pocas para ser consideradas en el análisis.

5.5 Según fecha

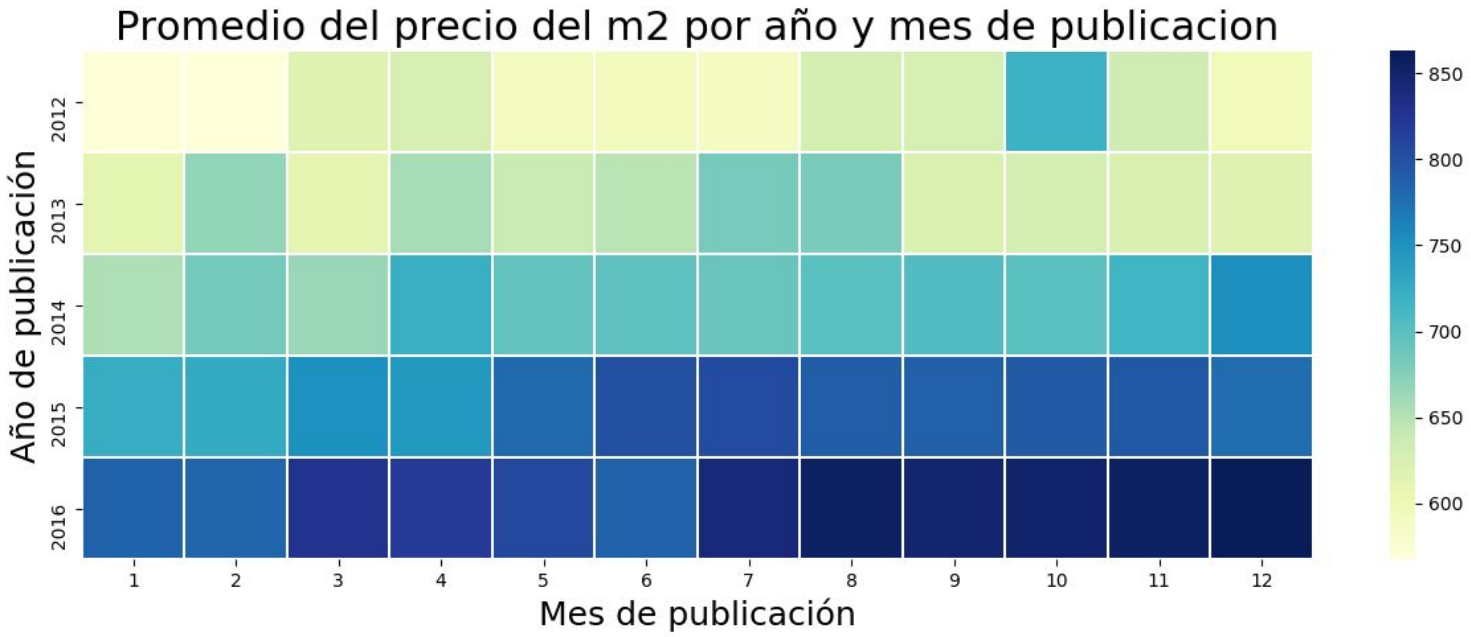

Gráfico 5.5.1

Podemos ver que a medida que pasa el tiempo, el valor de las publicaciones aumentan, esto se debe a la inflación que puede llegar a sufrir el país y además del tipo de publicaciones realizadas.

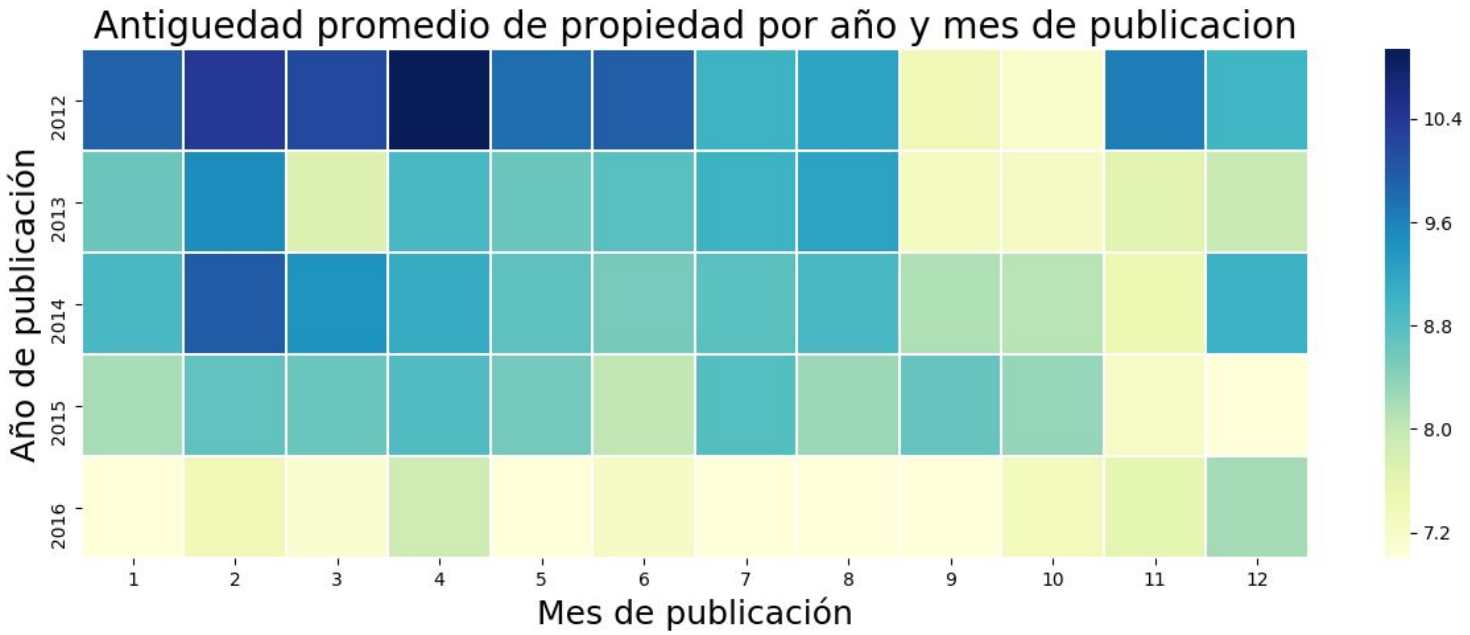

Gráfico 5.5.2

Un gráfico interesante para comparar con el anterior es ver la antigüedad promedio de las propiedades publicadas, para poder analizar si también se debe a esto el valor creciente a medida que pasan los años en las propiedades, ya que una propiedad con menor antigüedad, es más cara.

Comparando ambos gráficos, se puede ver que en el 2012 hay muchas publicaciones realizadas con propiedades antiguas, y por el contrario en el 2016 las publicaciones tienen poca antigüedad promedio. Esto trae un valor más elevado durante el 2016 que se puede ver en el gráfico 5.4.1 . También podemos ver que en octubre del 2012, se realizaron publicaciones de propiedades con poca antigüedad, viéndose reflejado en el precio promedio del metro cuadrado de esa fecha.

5.6 Regresión Linear

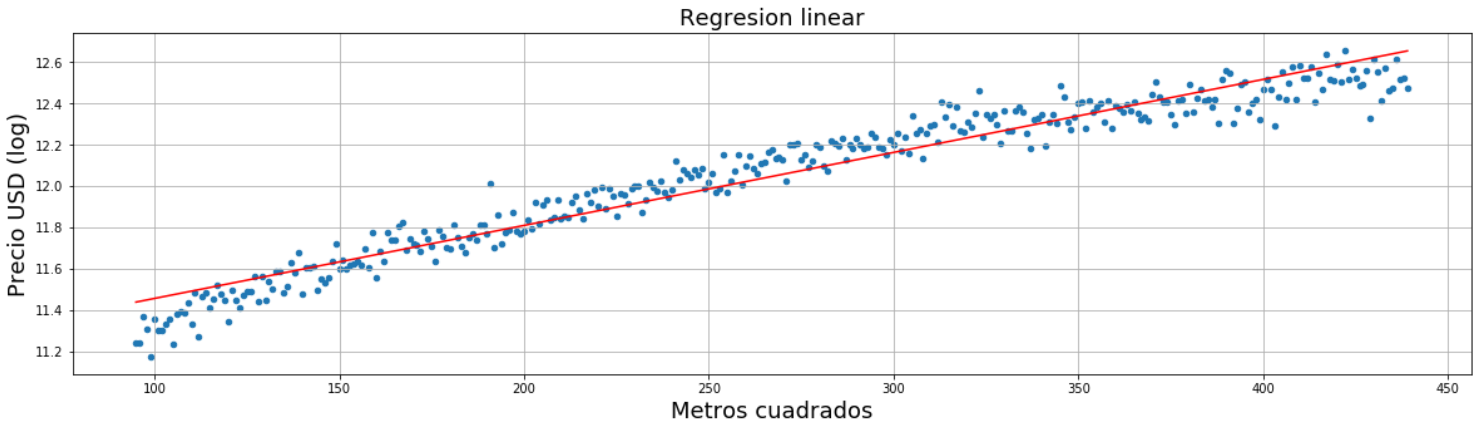

Gráfico 5.6

Para realizar este gráfico filtramos los outliers, que podían llegar a contaminar la muestra, filtrando tanto las propiedades con un precio bajo según sus metros cuadrados y las propiedades muy costosas para una cantidad muy alta de metros cuadrados, para poder ver cómo se comportan los valores que se comportan de la misma manera.

En la visualización se puede observar todas las propiedades con un precio en escala logarítmica, agrupándolos por su valor, viendo cómo eleva su precio a medida de que aumentan los metros cuadrados. La tendencia es uniforme a partir de los 150 metros cuadrados, pero no tanto para las propiedades por debajo de dicho número ni por encima de los 400 metros cuadrados, donde el precio da un ligero revés y tiende a bajar un poco del promedio.

6. Antigüedad de propiedades

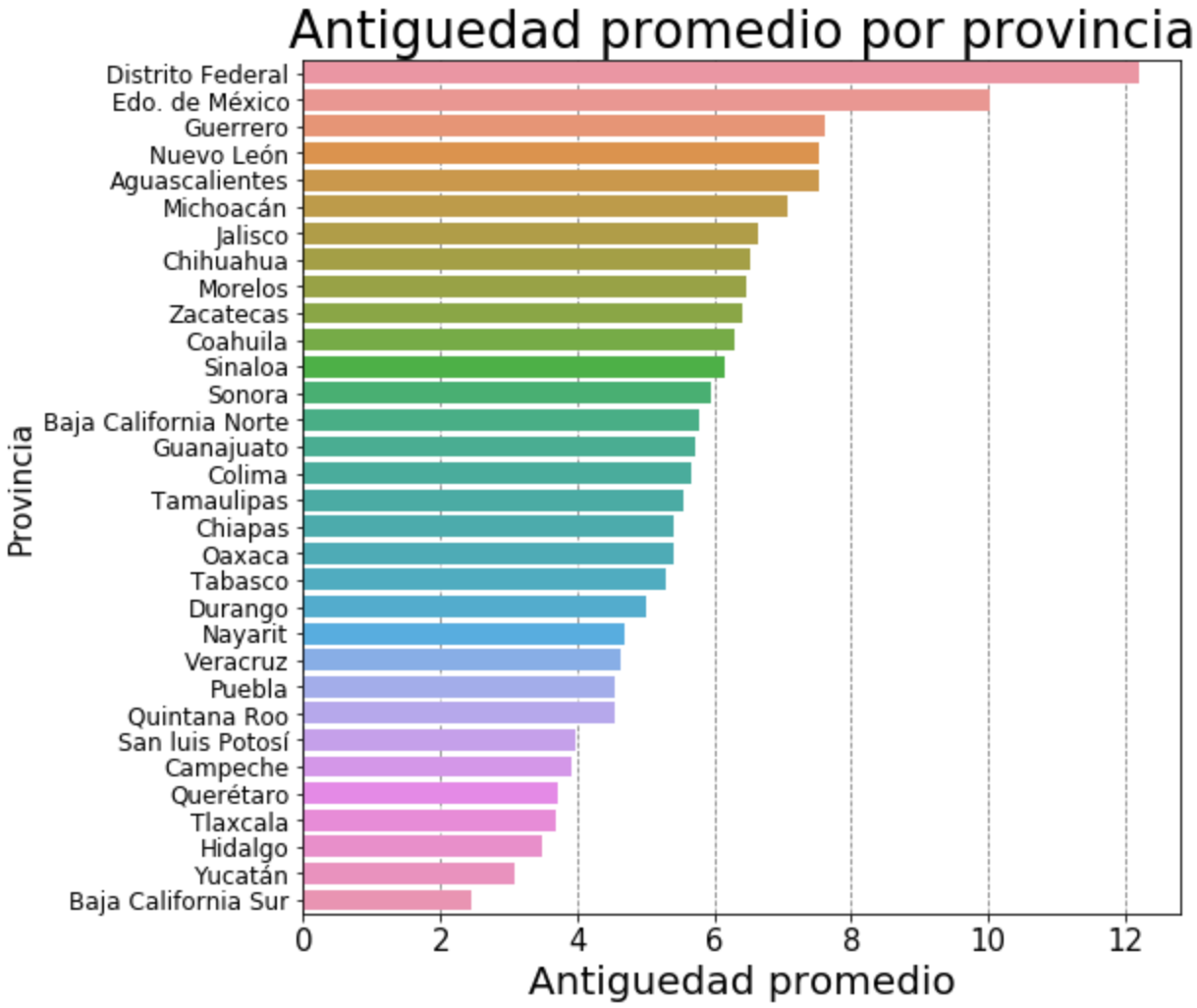

Gráfico 6.1.1

En una relación parecida a la vista anteriormente en el gráfico 5.2, vemos que las provincias que aparecen en el top, son provincias ya de por sí caras. Además de ser caras, hay muchas propiedades con una antigüedad considerable. Dicha correspondencia adquiere sentido cuando vemos el siguiente gráfico (Gráfico 6.1.2), donde las provincias más antiguas son las cercanas al centro del país, donde encontramos el casco histórico y toda la actividad administrativa.

Gráfico 6.1.2

Además , en el gráfico 6.1.2 se puede ver como las propiedades más antiguas se concentran en dos provincias principalmente, Edo de México y Distrito Federal como foco. En tanto, las demás provincias tienden a tener una antigüedad más homogénea .

Los casos opuestos, es decir las provincias con propiedades en venta con menor antigüedad son Baja California Sur a la izquierda y Yucatán a la derecha, por lo que se puede llegar a ver. Resultado ya examinado en el gráfico anterior (6.1.1).

7. Monoambientes

En esta sección se analizará, las propiedades con solo una unica habitacion, las cuales serán precisadas por estudiantes o personas que recién se mudan solos por primera vez. Como vimos anteriormente es un tópico interesante ya que se comporta de manera extraña, en relación a sus habitaciones, baños y precio por metro cuadrado.

7.1 Cantidad de publicaciones

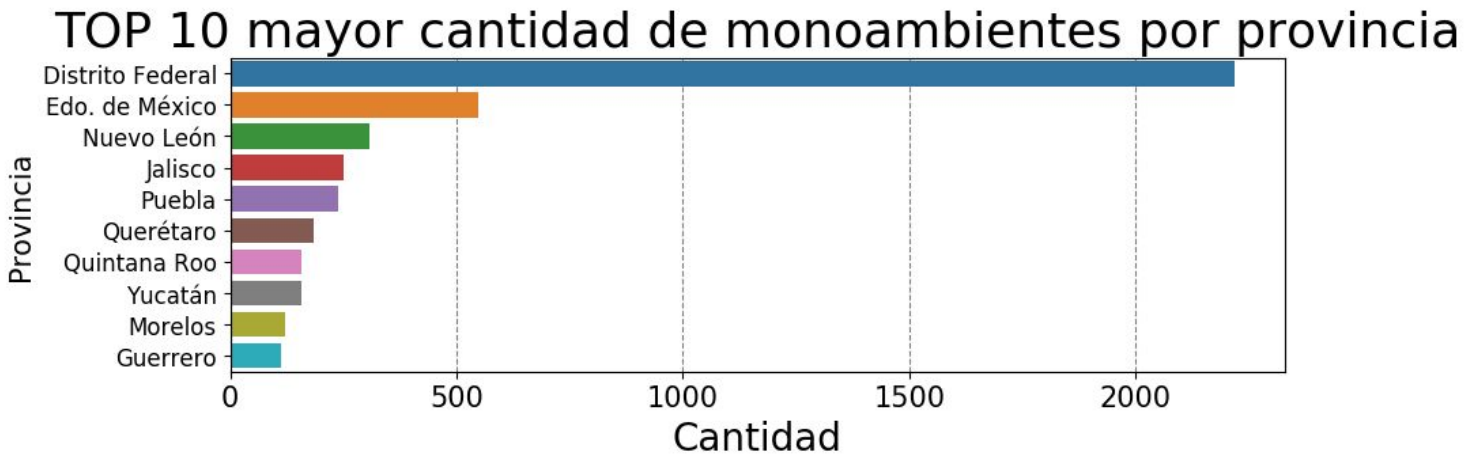

Gráfico 7.1.1

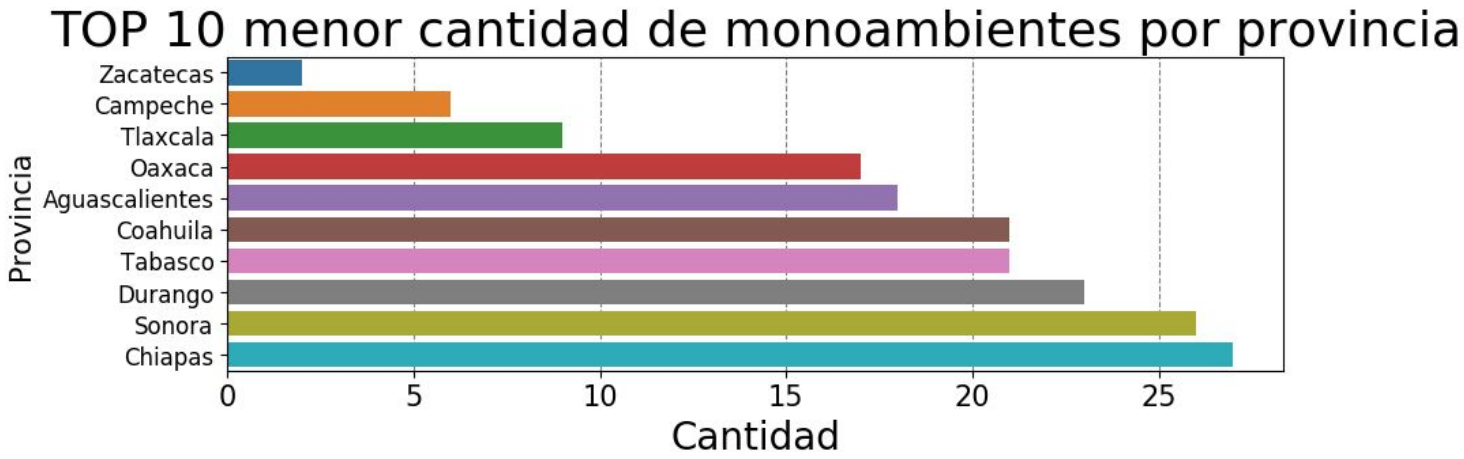

Gráfico 7.1.2

Podemos ver qué Distrito Federal tiene la mayor cantidad de monoambientes publicados, esto se debe por ser la capital del país, por ende la que mayor gente concentra. Es de esperar que nos encontremos con muchos apartamentos de menor tamaño, y con cercanía a las escuelas, universidades, centros comerciales, edificios administrativos, etc. En cuanto a las propiedades con menor cantidad de monoambientes se puede ver que apenas llegan a las 25 publicaciones.

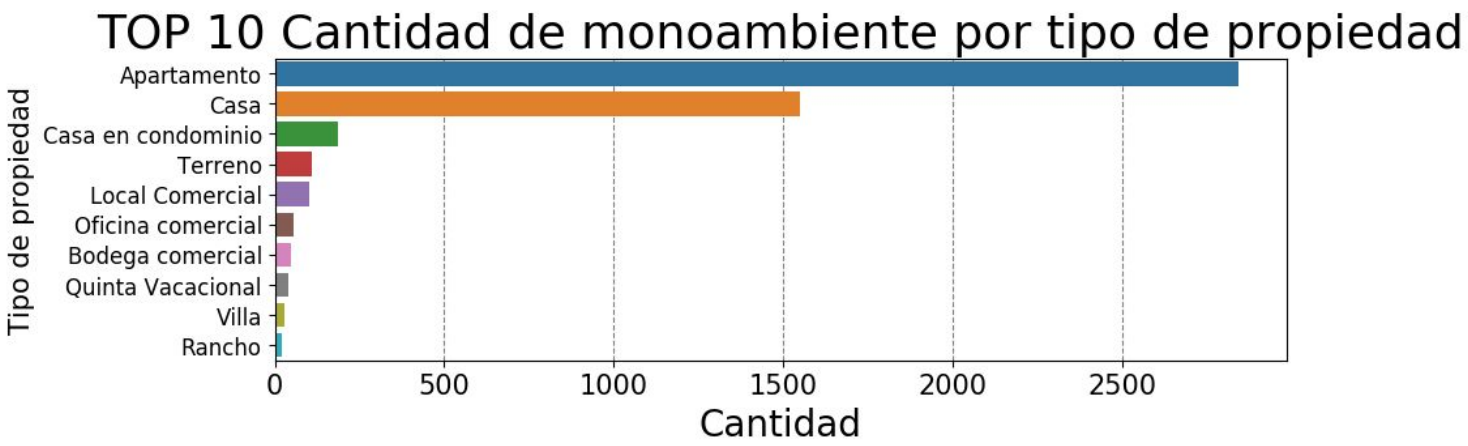

Gráfico 7.1.3

Se puede apreciar que predominan los Apartamentos publicados en comparación a otros tipo de propiedades, y cabe destacar que el top 3 lo abarcan las propiedades residenciales.

7.2 Precio del metro cuadrado

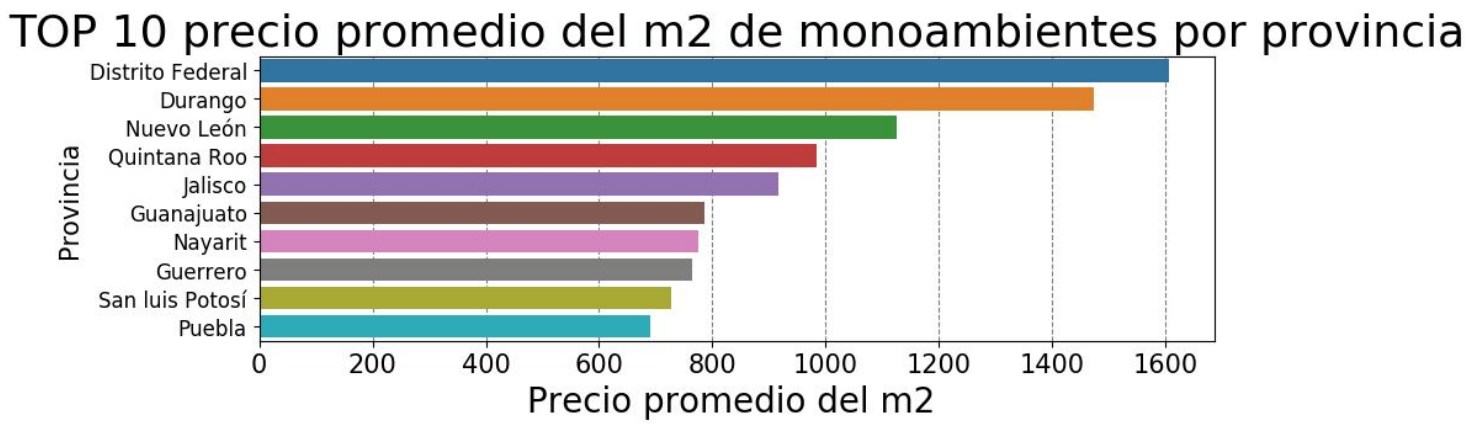

Gráfico 7.2.1

Al representar el precio por metro cuadrado según provincia, nos llevamos una sorpresa, la cual es que Durango, a diferencia de las otras provincias ubicadas en el top 10 de mayor cantidad de monoambientes (Gráfico 7.1.1) esta se encuentra en el gráfico 7.1.2 con un poco más de 20 publicaciones, y se encuentra con el segundo precio más costoso por metro cuadrado, detrás de Distrito Federal.

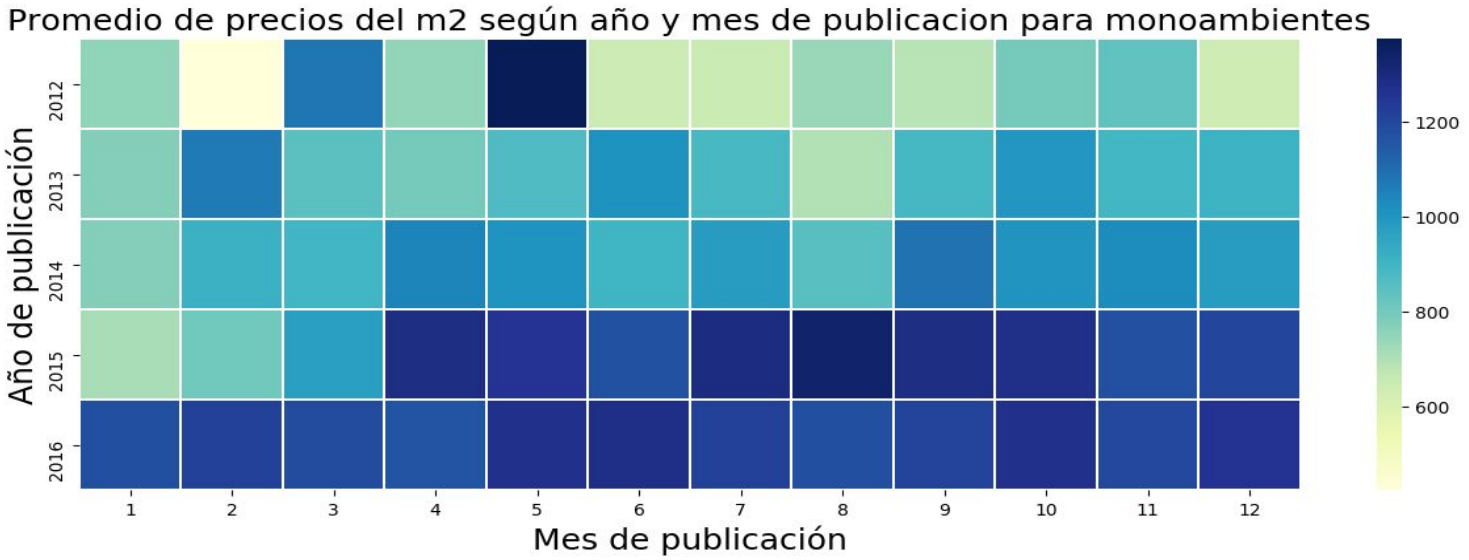

Gráfico 7.2.2

Podemos ver que los monoambientes nunca fueron una propiedad barata, pero a medida que pasan los años siempre fue aumentando el precio del metro cuadrado, mostrando así la inflación en México o cómo al pasar los años los monoambientes fueron tomando importancia en las publicaciones. Cómo vimos anteriormente en el gráfico 2.2.2, son las propiedades con mayor valor por metro cuadrado sorprendentemente, y además de eso, se mantuvieron su precio promedio alrededor de 2 años seguidos.

Gráfico 7.2.3

Otra forma de mostrar los resultados obtenidos en 7.2.1, véase la diferencia de colores entre Durango y DF, con el resto de las provincias. En efecto, en 7.2.1 se ven bien correspondidos éstos destacados.

7.3 Antigüedad

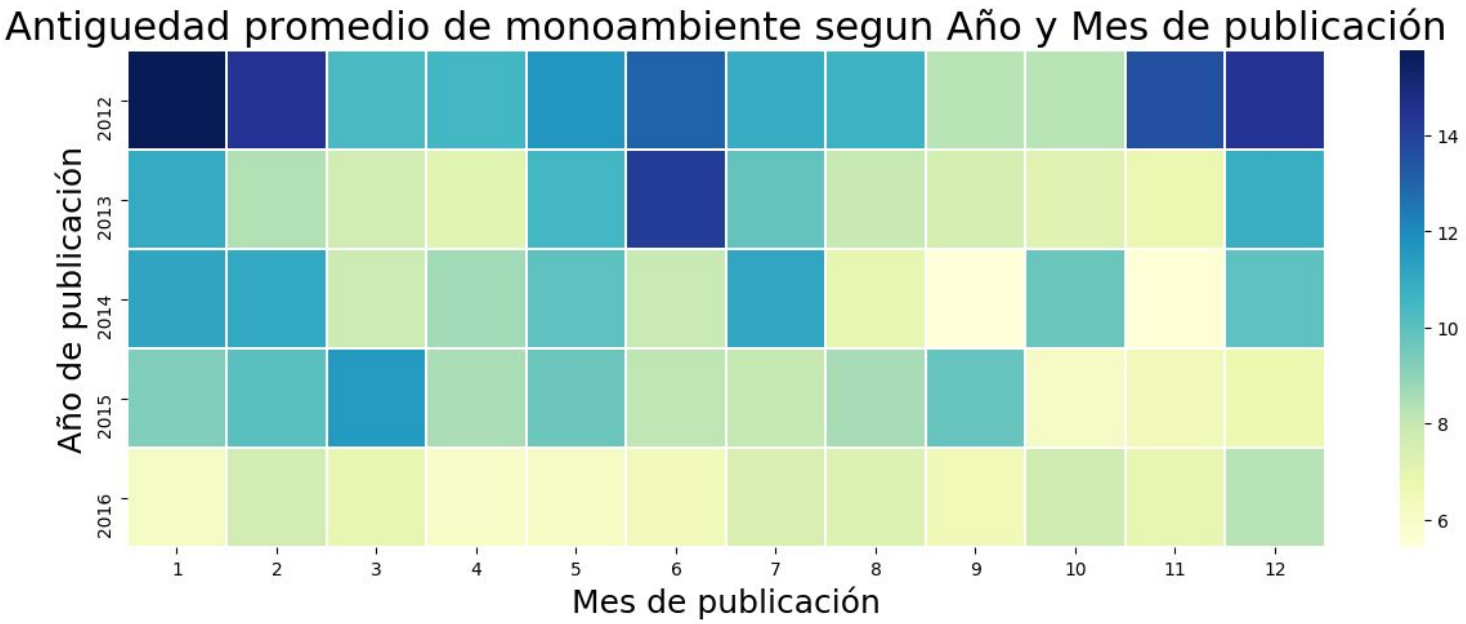

Gráfico 7.3.1

Éste gráfico parecería trivial, claro que las publicaciones más nuevas tienden a ser de monoambientes más modernos, sin embargo, también habla del data frame, la antigüedad de las propiedades tiene un promedio mínimo de 6 años y un máximo de 15.

Comparando con el gráfico 7.2.2, vemos que hay una relación directa entre la antigüedad promedio y el precio de la propiedad, ya que se ve que cuando se publicaron monoambientes más viejos, su precio disminuyó, con la excepción que en mayo del 2012, esto no ocurrió, y se logró el valor pico de los precios por metros cuadrados. En cuanto a los últimos dos años de publicaciones se ve que todas tienen un promedio menor de antigüedad, afectando directamente el valor por metro cuadrado.

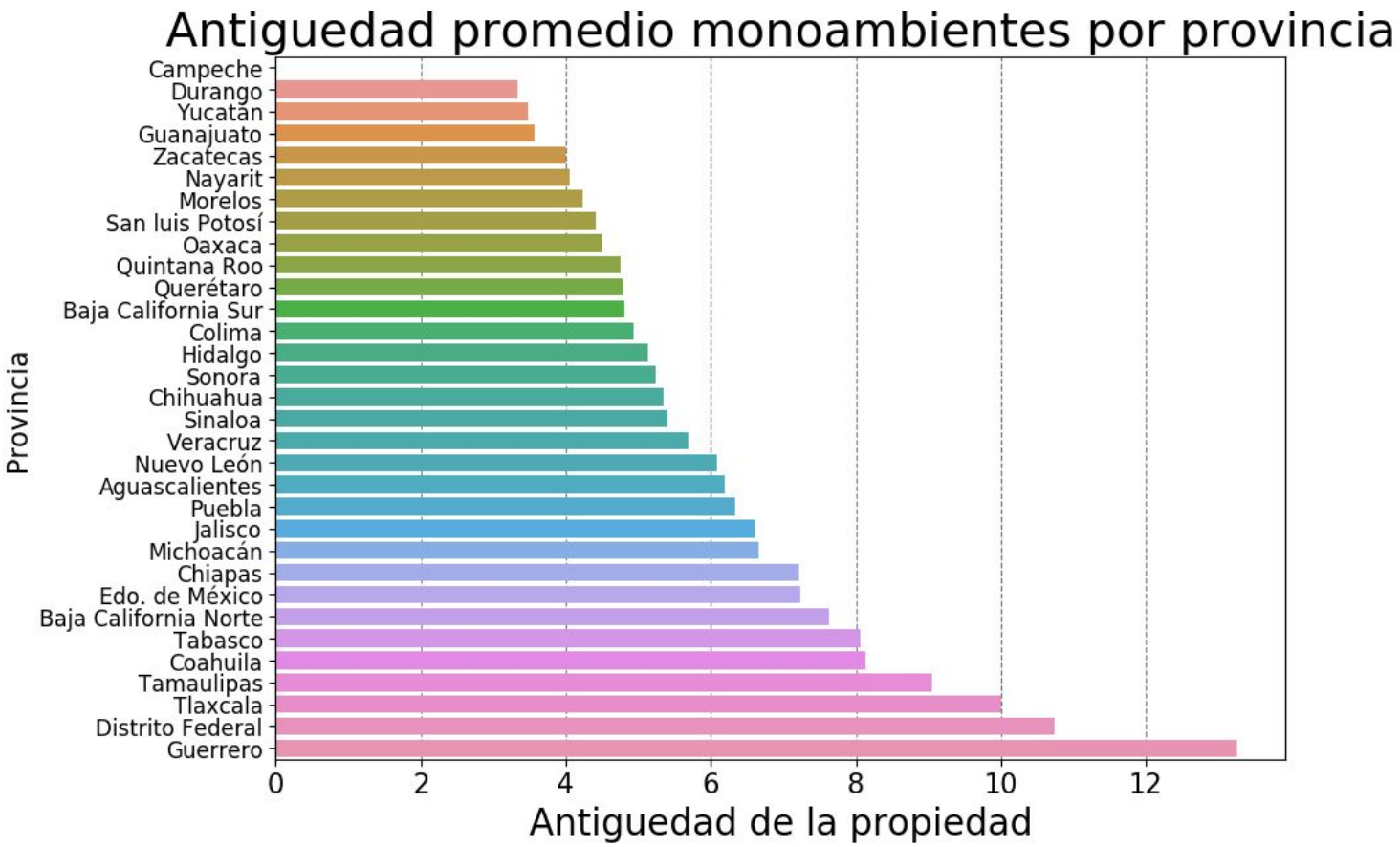

Gráfico 7.3.2

El Distrito Federal, desde su concepción, es un centro comercial y político para México. Por lo tanto es esperable que tenga monoambientes viejos en promedio, puesto a que siempre fue un ambiente donde se promueven este tipo de viviendas. Admitimos esperábamos ganen en cantidad los que se generan constantemente como visto en 7.3.1.

Sin embargo… ¿Guerrero? Una provincia agrocultural esperábamos que se comporte más parecido a las demás. El único análisis que logramos de ésta anomalía es que hay pocos monoambientes en Guerrero y algunos muy antiguos, empujando el promedio hacia los números más altos.

En cuanto a Campeche, al tener pocas publicaciones las cuales su antigüedad promedio es 0 se puede decir que todas sus publicaciones son monoambientes a estrenar, ya que no tienen antigüedad previa.

7.4 Escuelas cercanas

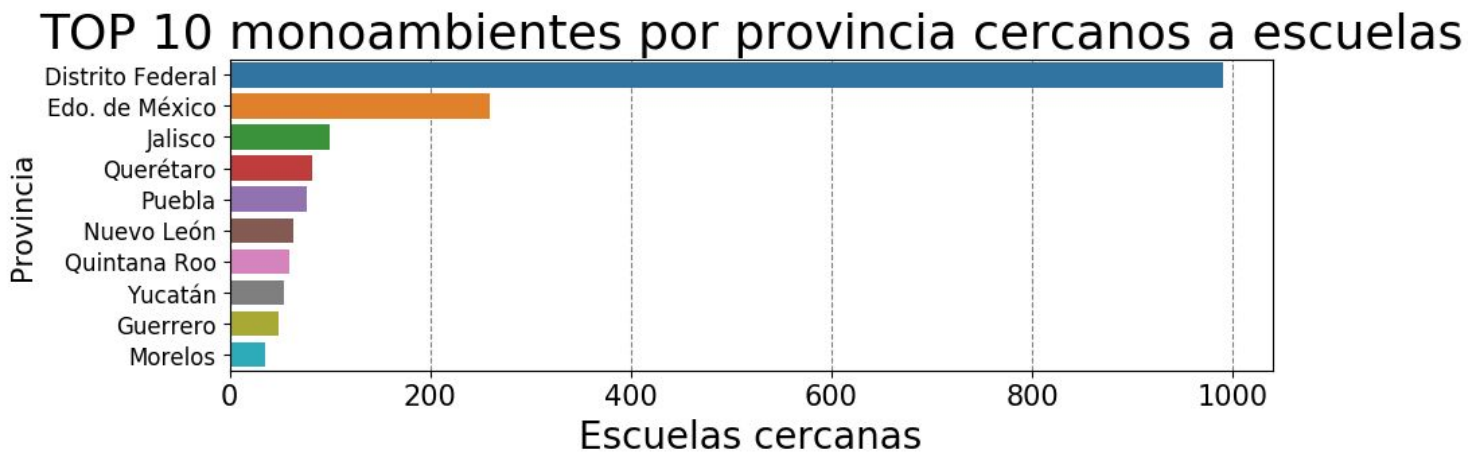

Gráfico 7.4.1

De la capital de México se espera este comportamiento por su alta densidad en construcciones y gran concentración de estudiantes, también el estado de méxico también cumple con esa propiedad pero en menor medida. Las demás provincias, denotan menos centralización en éste gráfico.

Aunque para estos estudiantes/comerciantes tienen un precio de metro cuadrado muy elevado, por ende, es muy costoso la vivienda cuando uno estudia en México, siendo además una propiedad antigua.

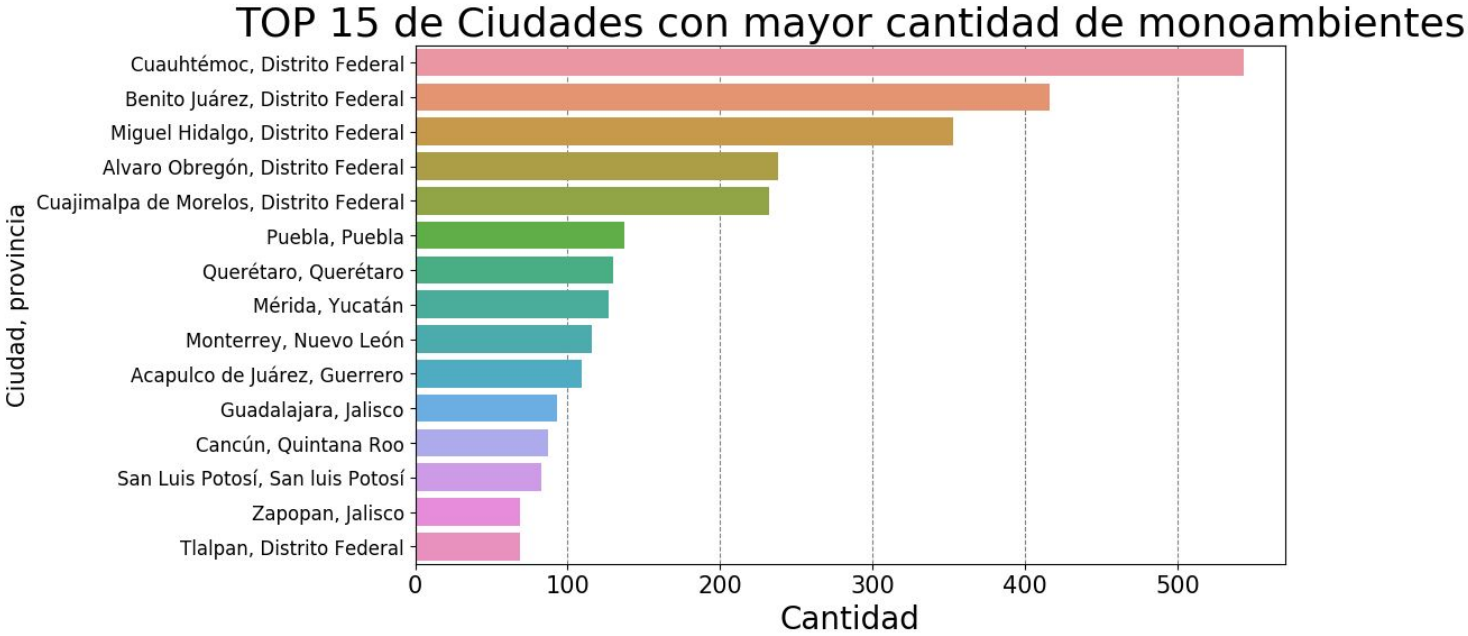

Gráfico 7.4.2

Si bien de todas las ciudades se espera que haya una cantidad sustancial de monoambientes. Nuestro análisis sobre las ciudades del DF, es que tienen gran cantidad de monoambientes no sólo por ser ciudades, creemos que además muchas personas se dirigen con objetivos laborales o estudiantiles al Distrito Federal, promoviendo un mercado de monoambientes, en la capital comercial del país.

Podemos ver como Cuauhtémoc, Benito Juárez y Miguel Hidalgo, acaparan una gran cantidad de monoambientes siendo las ciudades más céntricas de México.

8. Distrito Federal

En esta parte veremos la capital de México (DF), por su cantidad de publicaciones, precio del metro cuadrado.

8.1 Tipos de propiedades

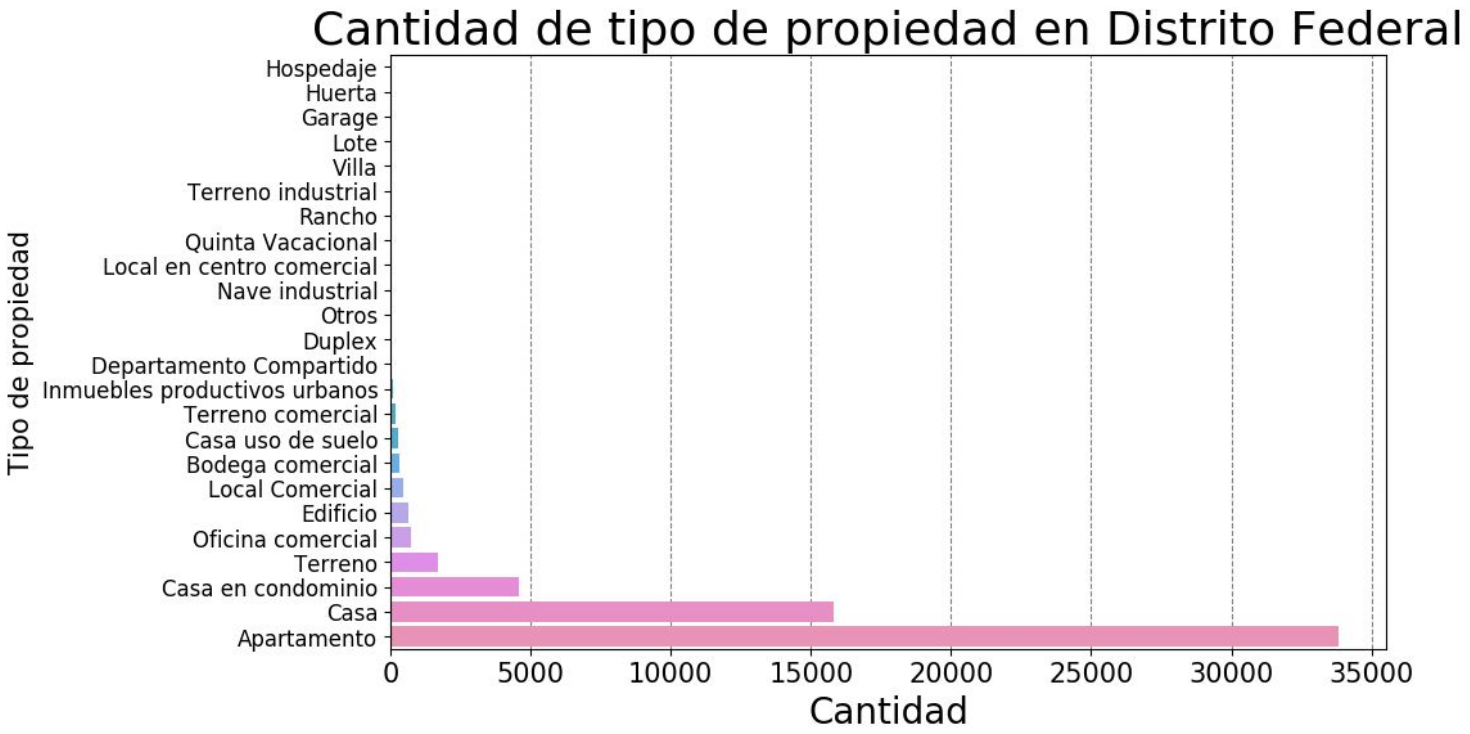

Gráfico 8.1

Viejo: Es sorprendente con la cantidad de publicaciones sobre las casas que se realizaron en la página, en Distrito Federal predominan los Apartamentos, esto debe ser por su pequeña superficie y la cantidad de apartamentos que hay en comparación a las casas.

Sugerido: Es sorprendente la cantidad de apartamentos que se publican en comparación con cualquier otro tipo de vivienda. Creemos que, la alta cantidad de apartamentos está relacionada con la alta cantidad de publicaciones en el D.F. donde predominan los apartamentos.

Como vimos anteriormente, en el gráfico 5.2.A, se vio que Distrito Federal es una provincia comercial, por el precio por metro cuadrado alto en los comercios y bajo en las residencias, cabe destacar que el precio por metro cuadrado es más alto en los Apartamentos que en las Casas, esto se debe a su cercanía con el centro de la ciudad, ya que, en el centro no hay tantas Casas y hay gran cantidad de Apartamentos.

8.2 Ciudades de DF

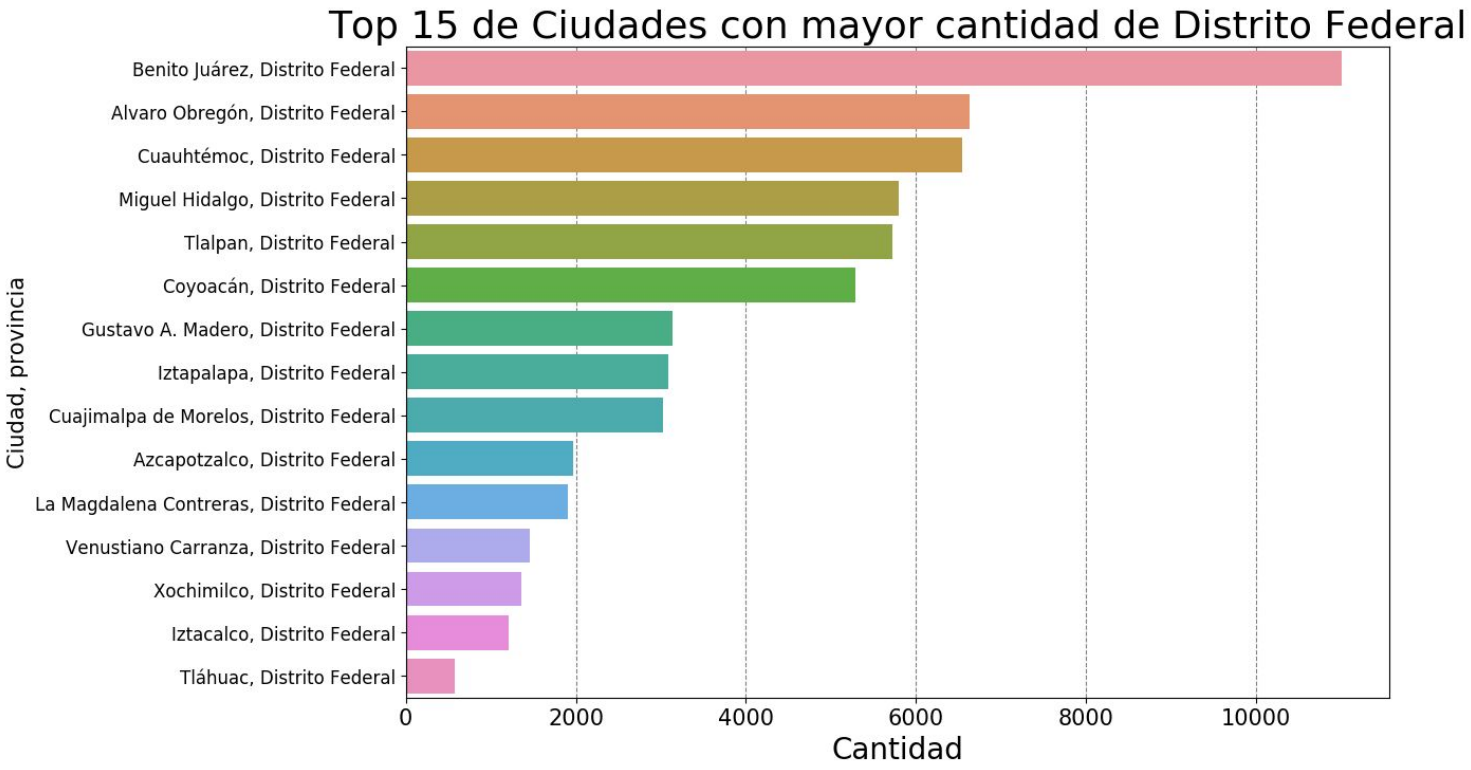

Gráfico 8.2.1

Las ciudades de DF, tienden a ser escalonadas en cuanto a la cantidad de publicaciones que hay en cada una. En Benito Juárez, claramente es la ciudad con más movimiento publicitario, pertenecía a la segunda posición del ranking total de las ciudades con más publicaciones de todo México. En cuanto a las cinco siguientes ciudades apenas tienen la mitad de publicaciones que el primer puesto y el resto de las publicaciones tiene un cuarto que Benito Juárez, haciéndose cada vez más chica la cantidad de publicaciones.

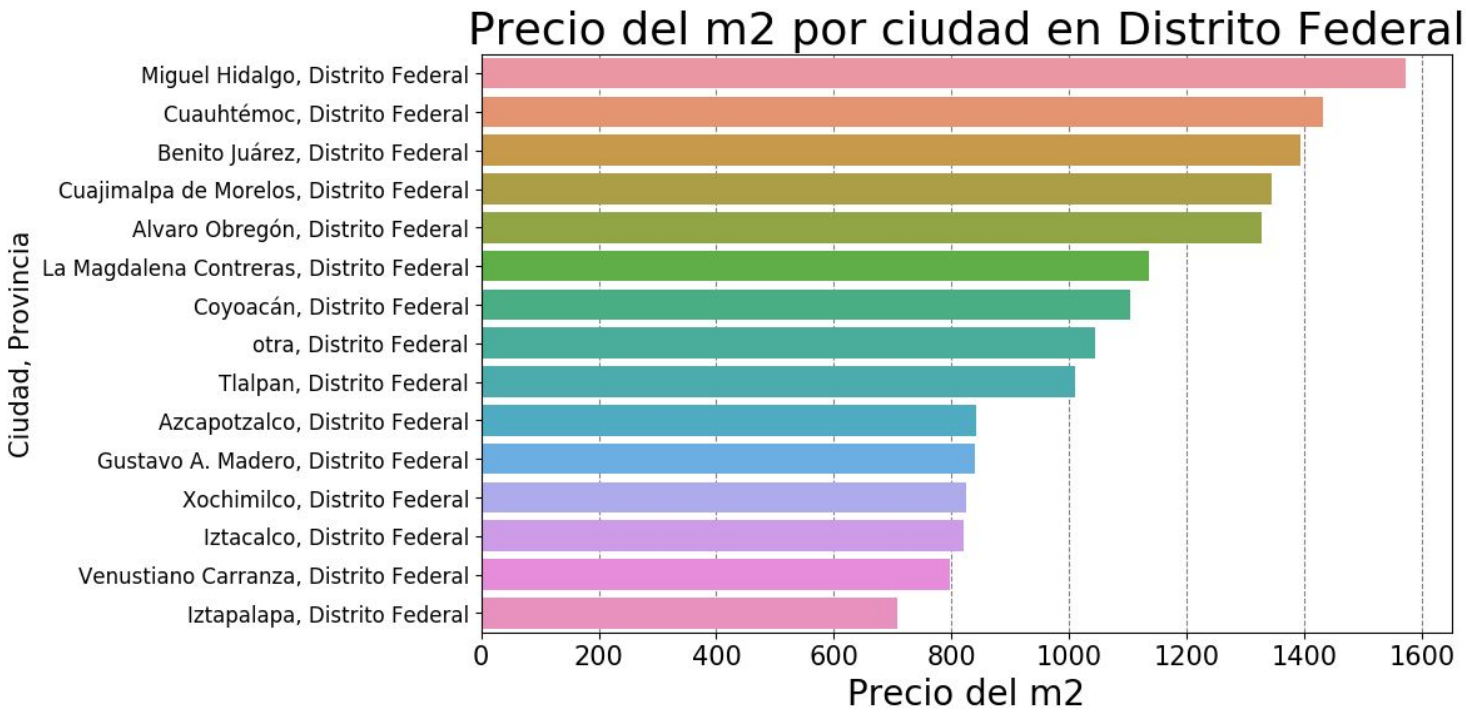

Gráfico 8.2.2

Podemos ver que muchas ciudades tienen un precio muy alto por metro cuadrado, las que mayor interés tienen son aquellas que son pocas publicaciones pero un precio muy elevado, por ejemplo La Magdalena Contreras, que casi no llega a las 2000 publicaciones y el precio promedio esta en 1100 el metro cuadrado.

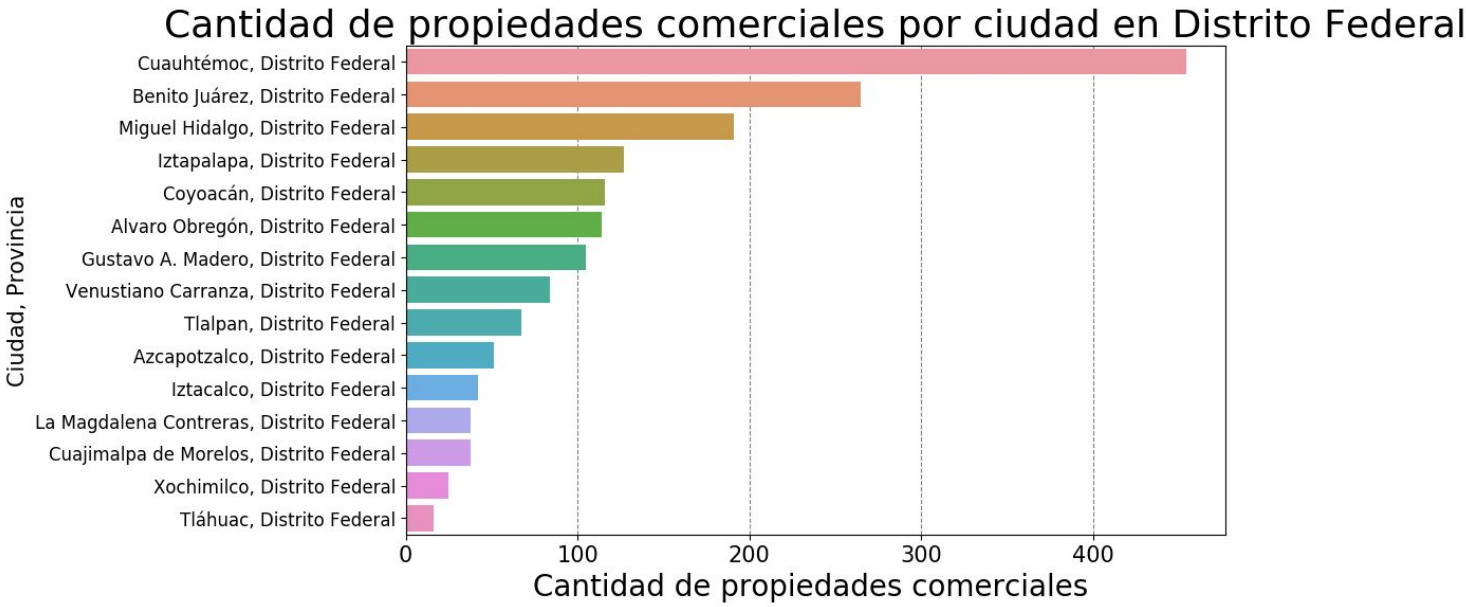

Gráfico 8.2.3

En comparación con el gráfico 8.2.1, filtramos las publicaciones por comercios (Local Comercial, Local en centro comercial, Bodega comercial, Terreno comercial, Oficina comercial). Podemos ver que hay muy pocas publicaciones de las ciudades por comercio que las publicaciones totales. Viendo los órdenes de los ranking podemos notar que Cuauhtémoc, es la ciudad que tienen más publicidades comerciales, pasando a Benito

Juárez podemos ver que hay muy pocas propiedades comerciales, en las cuales el precio por metro cuadrado es muy elevado.

8.3 Densidad de propiedades

Nos propusimos tratar de buscar grupos según el precio por metro cuadrado. Para ello primero generamos un histograma para ver la distribución del precio por metro cuadrado en USD

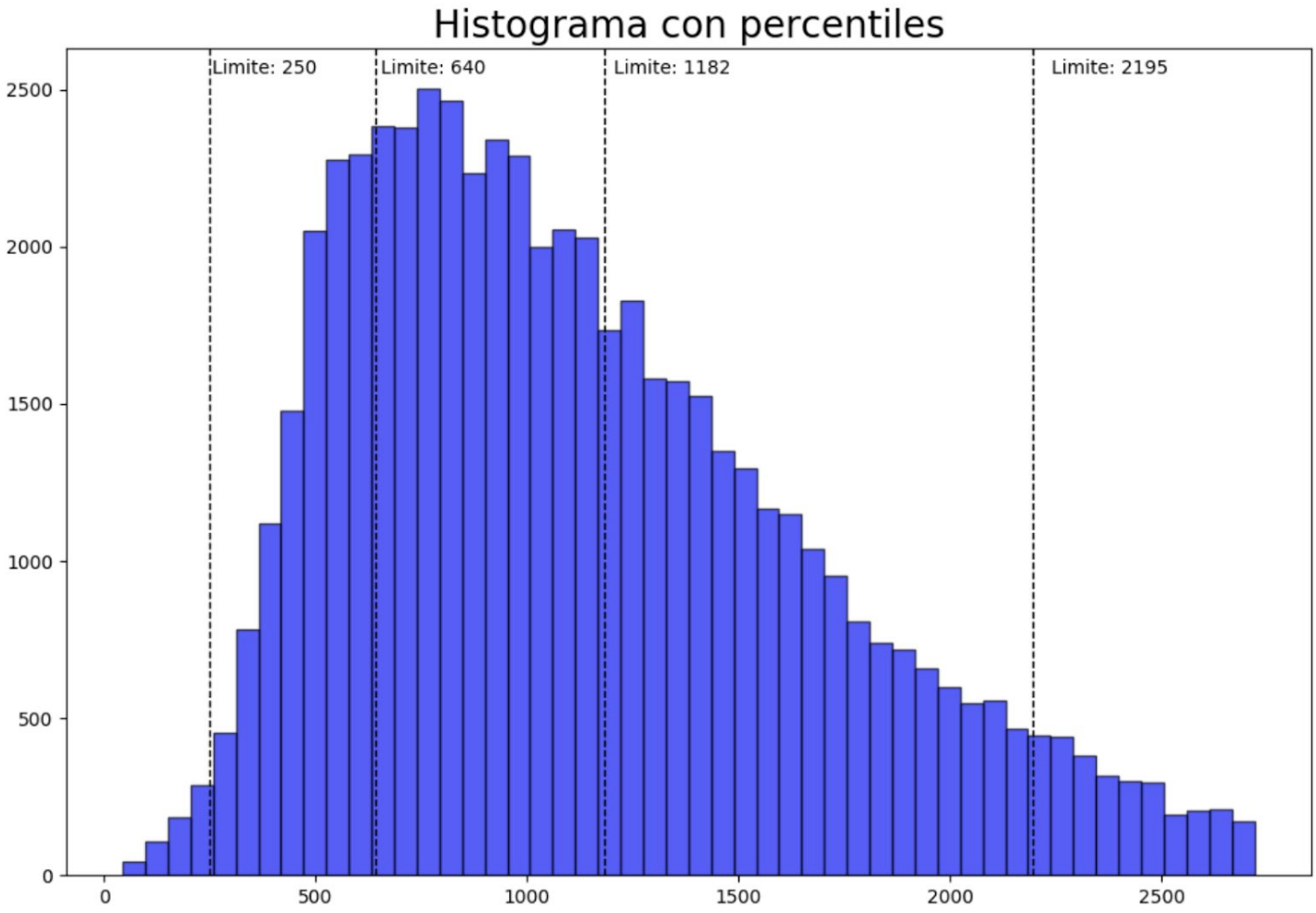

Gráfico 8.3

Vemos que hay muy pocas propiedades con valores bajos, y la distribución tiene forma de cola hacia la derecha. Esto es de esperarse, ya que como vimos antes, Distrito Federal es la provincia con mayor cantidad de propiedades publicadas y la provincia con el precio por metro cuadrado más caro de México para el dataset analizado.

En base a ello, decidimos separar en 5 grupos:

- grupo de propiedades muy baratas: van desde 44 a 250 USD
- grupo de propiedades baratas: van desde 250 hasta 640 USD
- grupo de propiedades “medias” (ni muy caras ni muy baratas): desde 640 hasta 1182 USD
- grupo de propiedades caras: desde 1182 hasta 2195 USD
- grupo de propiedades muy caras: desde 2195 hasta 2719 USD

8.3.1 Grupo: Muy caro

Para el grupo más caro, el top 5 de propiedades se encuentra distribuido de la siguiente forma:

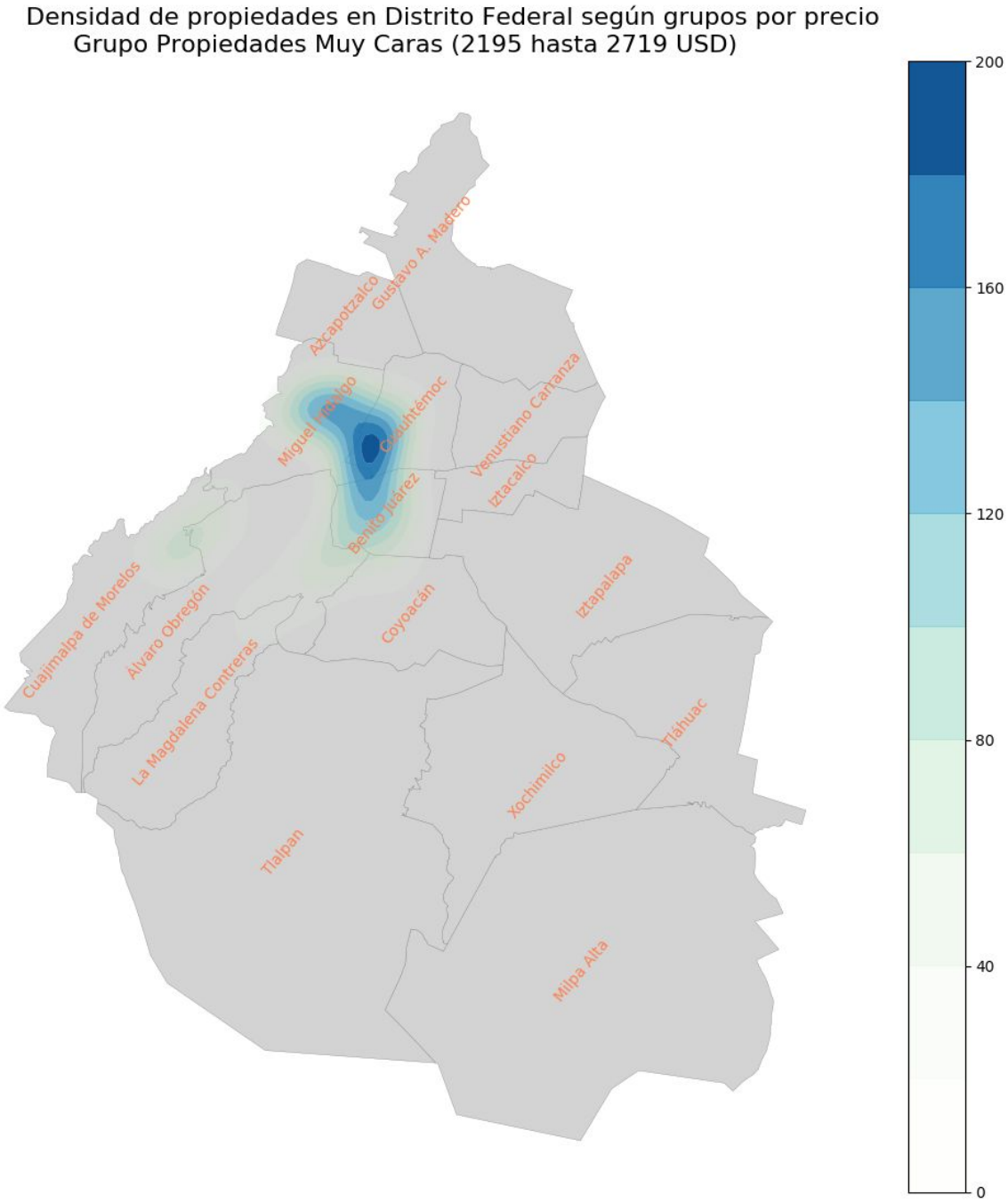

Gráfico 8.3.1

Ciudad	Cantidad
Miguel Hidalgo	648
Cuauhtémoc	644
Benito Juárez	577
Álvaro Obregón	458
Cuajimalpa de Morelos	201

Si bien no tenemos tanta cantidad en este grupo, podemos observar una leve tendencia de concentración en el límite de las tres ciudades del top, Miguel Hidalgo, Cuauhtémoc y Benito Juárez. Esto tiene sentido, ya que en el análisis previo, encontramos que estas ciudades se encuentran en el top de las más caras.

8.3.2 Grupo: Caro

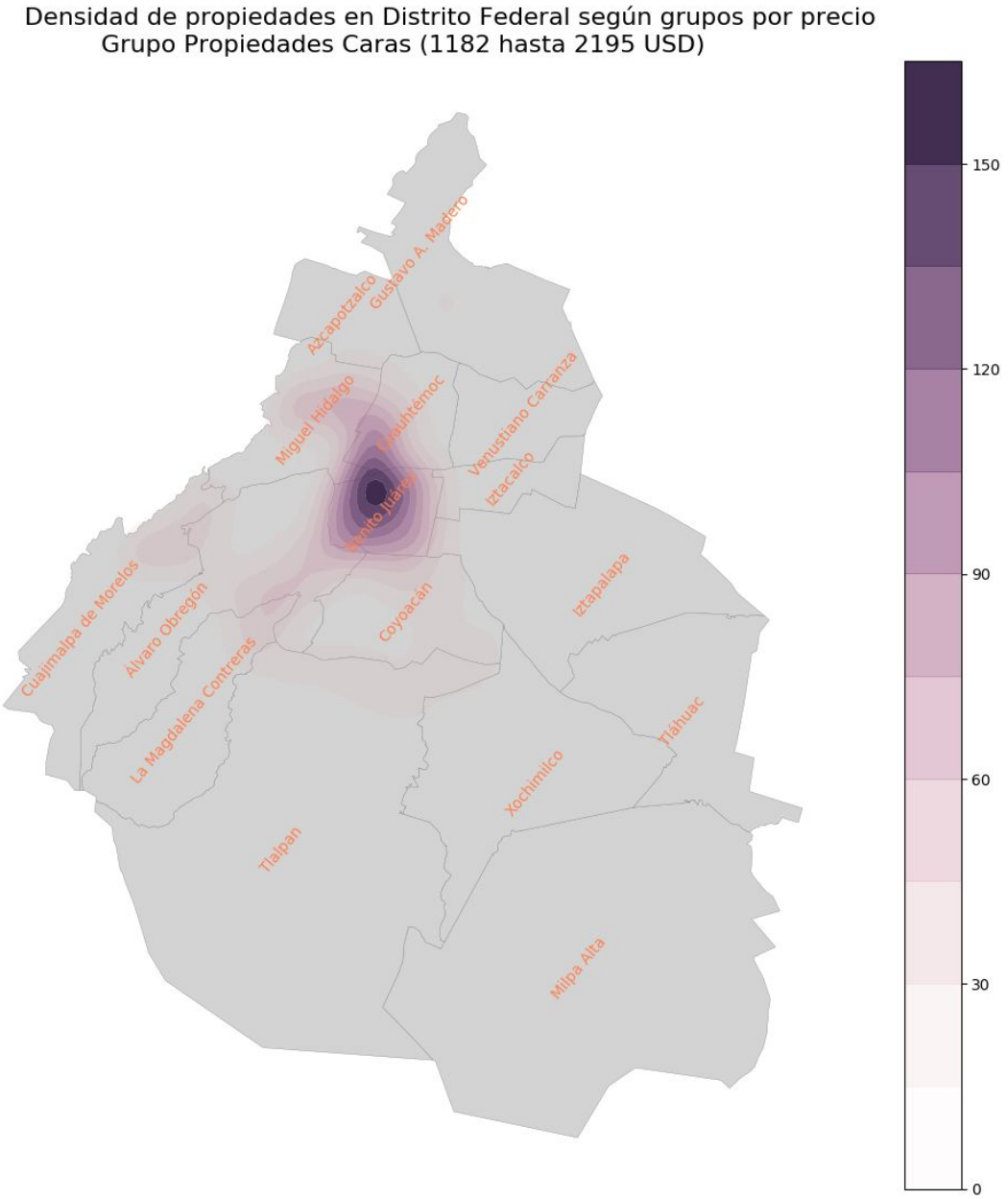

Gráfico 8.3.2

Ciudad	Cantidad
Benito Juárez	5846
Álvaro Obregón	2817
Miguel Hidalgo	2415
Cuauhtémoc	2410
Coyoacán	1595

Ya en este rango de precios empezamos a ver una tendencia hacia el sur, ya que la mayoría de las propiedades de este grupo están concentradas en Benito Juárez.

8.3.3 Grupo: Medio (ni muy caro ni muy barato)

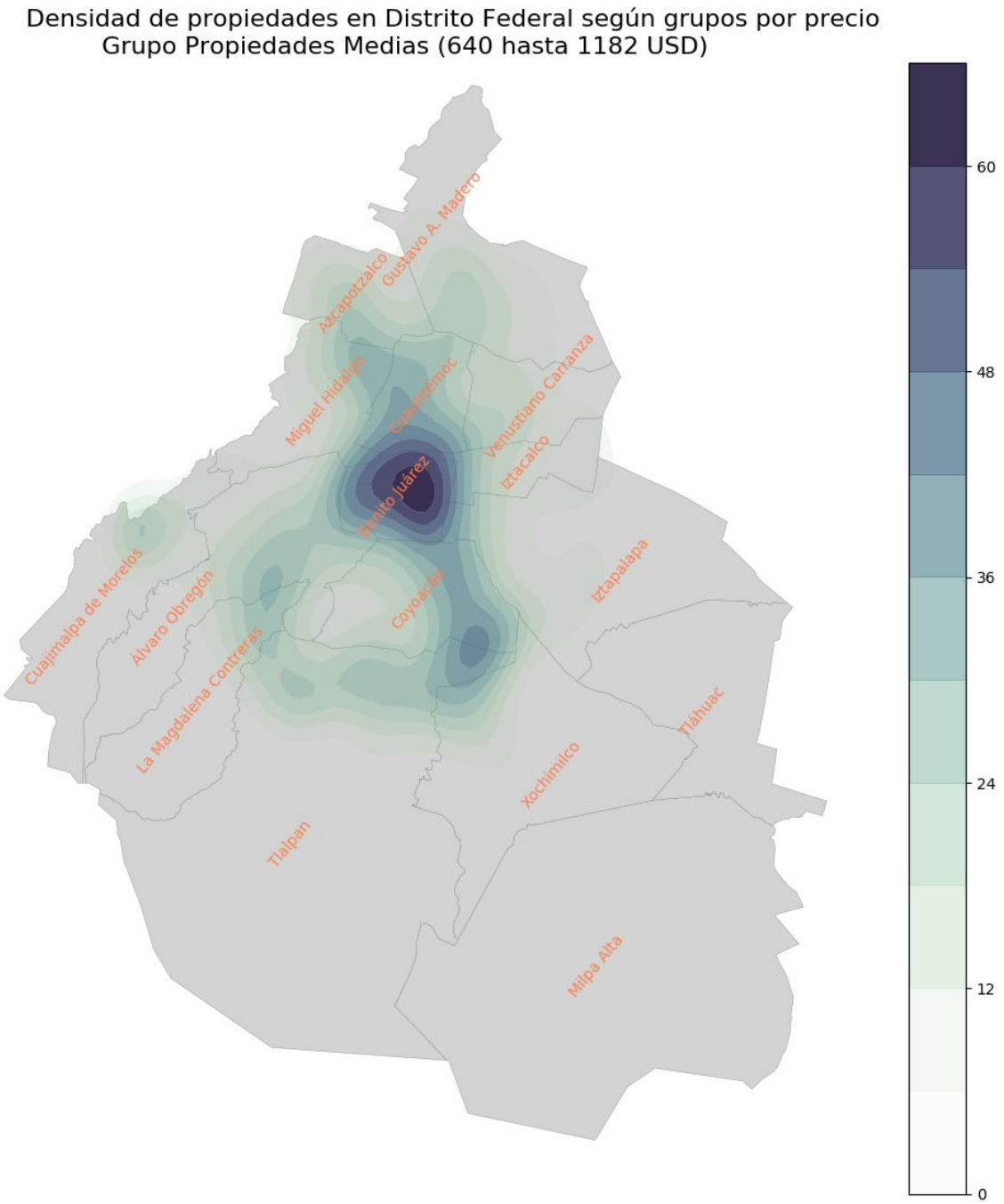

Gráfico 8.3.3

Ciudad	Cantidad
Benito Juárez	3584
Tlalpan	3003
Coyoacán	2640
Álvaro Obregón	2265
Cuauhtémoc	2036

En este grupo, la concentración sigue en Benito Juárez, pero esta vez está más dividido con otras provincias hacia el sur, continuando con la tendencia del grupo anterior.

8.3.4 Grupo: Barato

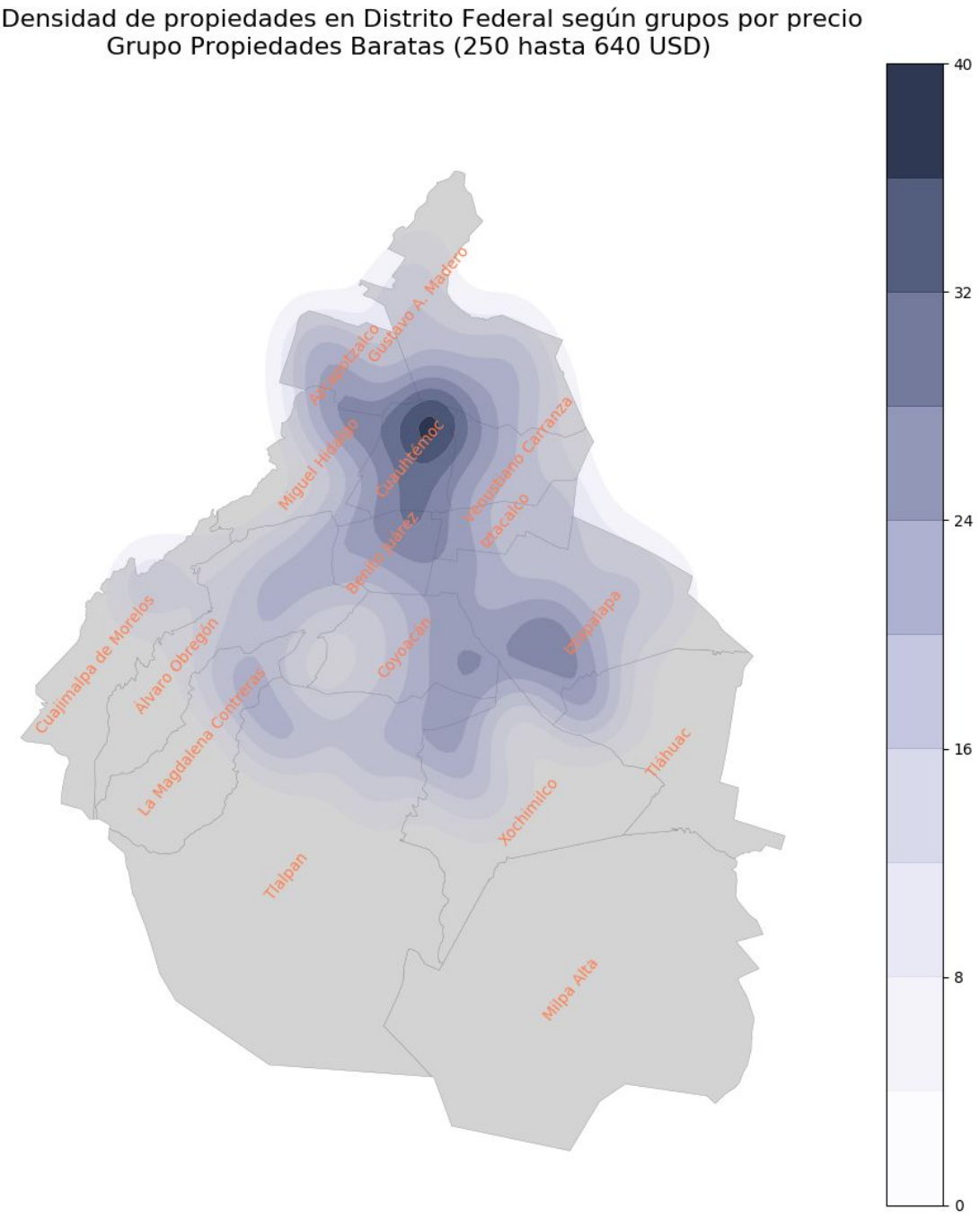

Gráfico 8.3.4

Ciudad	Cantidad
Iztapalapa	1654
Gustavo A. Madero	1113
Tlalpan	1051
Cuauhtémoc	975
Álvaro Obregón	857

Aquí observamos algo distinto a lo que veníamos viendo en el resto, ya que no hay una tendencia hacia el sur, sino que las ciudades baratas están hacia el norte, y en menor medida algunas de las ciudades al sur-este de Benito Juárez

8.3.5 Grupo: Muy barato

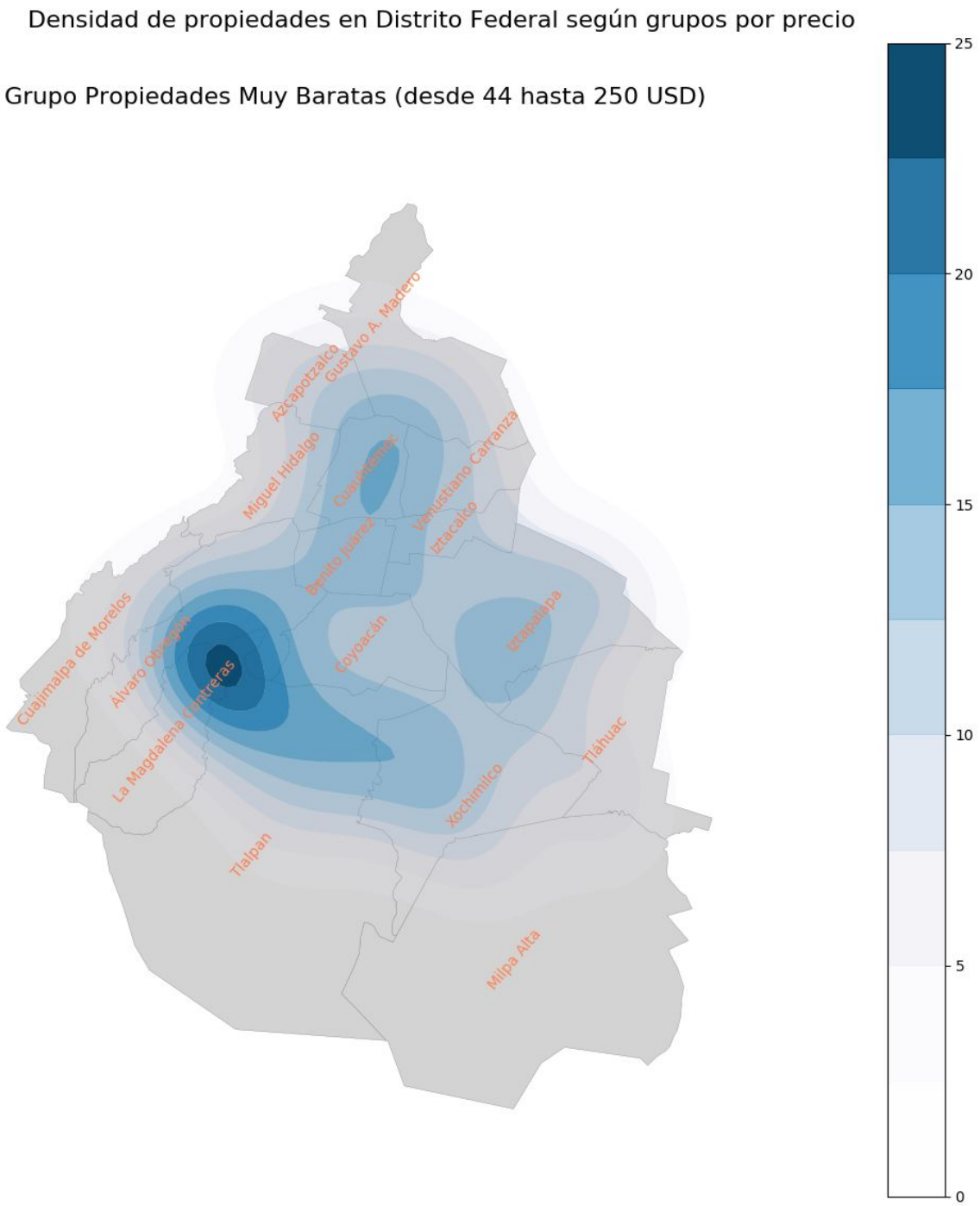

Gráfico 8.3.5

Ciudad	Cantidad
Tlalpan	96
Iztapalapa	87
Xochimilco	52
Álvaro Obregón	46
Gustavo A. Madero	45

Para este grupo aplica lo mismo que el grupo de los más caros: como tomamos el quantile del 1% del dataset para Distrito Federal, la muestra no es del todo representativa debido a la falta de más datos.

9. Otros análisis interesantes

9.1 Influencia de características en el precio por metro cuadrado

En el set de datos contamos con campos categóricos binarios que indican si cierto atributo está presente en una propiedad. Estas características son:

- si tiene gimnasio
- si tiene piscina
- si se encuentra cerca de escuelas
- si se encuentra cerca de centros comerciales
- si es apto para usos múltiples

Si bien ya vimos en el gráfico de correlaciones un número representativo, ¿Cómo influye la presencia de alguna de estas características en el precio por metro cuadrado?

9.1.1 En Apartamentos

Para volver a separar en 5 posibles grupos, analizamos un histograma:

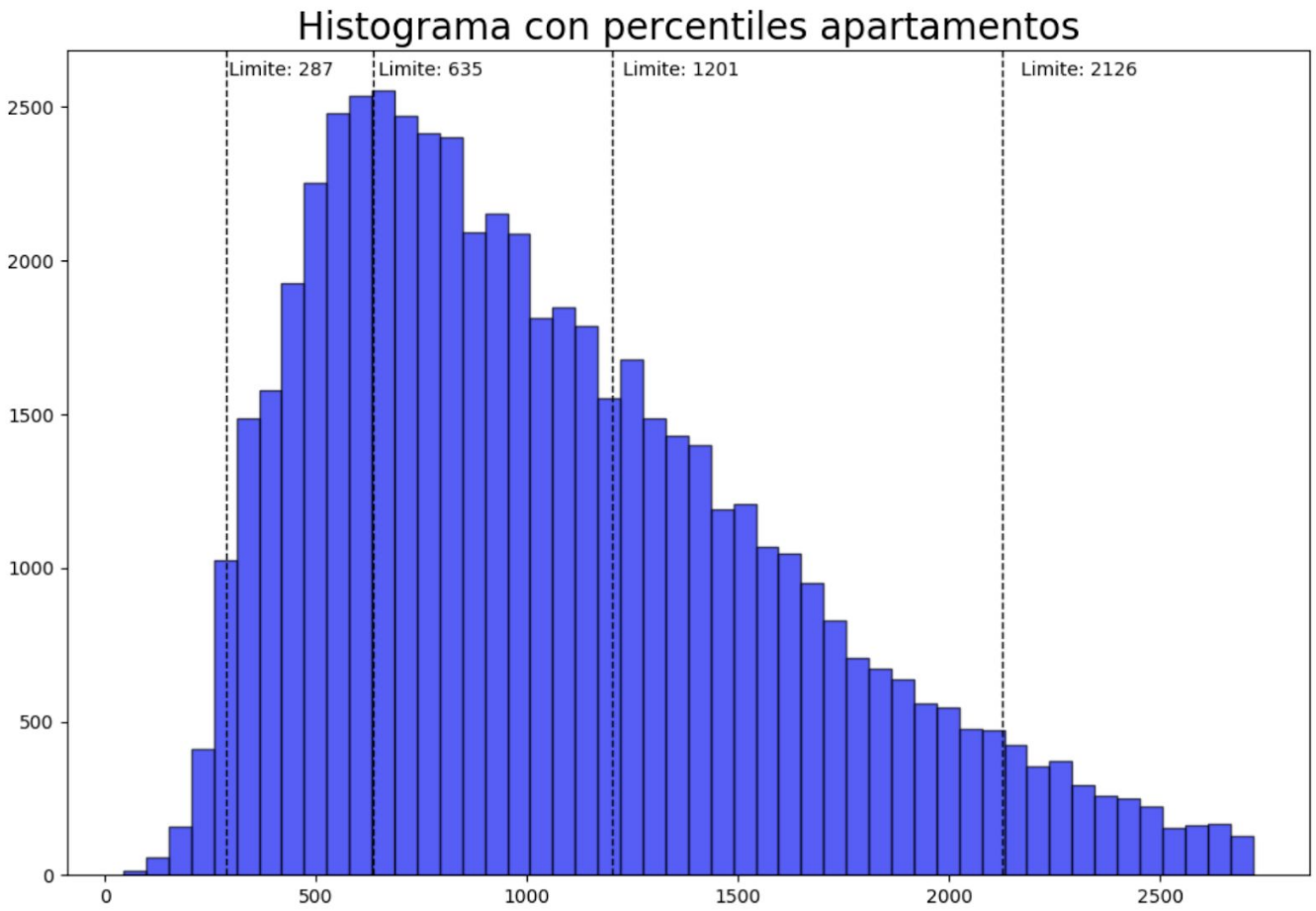

Gráfico 9.1.1.1

A partir del histograma, podemos separar en 5 grupos nuevamente

- grupo de propiedades muy baratas: hasta 287 USD
- grupo de propiedades baratas: van desde 287 hasta 635 USD
- grupo de propiedades “medias” (ni muy caras ni muy baratas): desde 635 hasta 1201 USD
- grupo de propiedades caras: desde 1201 hasta 2126 USD
- grupo de propiedades muy caras: desde 2126 hasta 2719 USD

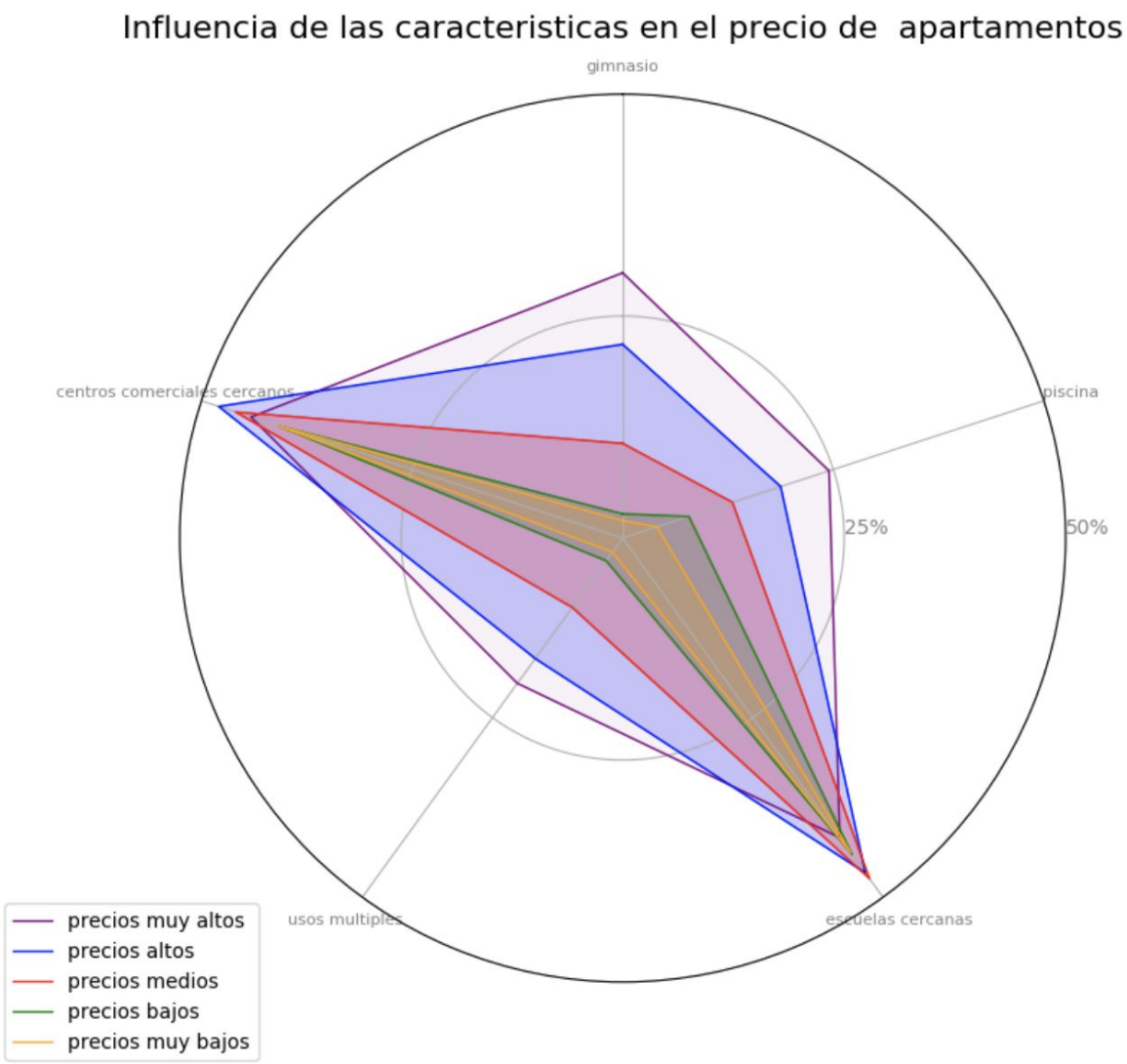

Gráfico 9.1.1.2

En el gráfico observamos que hay ciertas características que son comunes a casi todos los apartamentos, ellas son “cercanía a centros comerciales” y “cercanía a escuelas”. El precio por metro cuadrado no varía demasiado en estas categorías. En cambio, para características como tener gimnasio, piscina, o ser apta para usos múltiples, se nota una diferencia en los rangos de precio. Por ejemplo, el 25% de los apartamentos que pertenecen al grupo de “precios muy altos” poseen una piscina. Incluso un poco más del 25% de los mismos poseen un gimnasio. Vemos cómo estos porcentajes disminuyen a medida que baja el precio por metro cuadrado, lo que indica que son características que empiezan a escasear. Esto tiene sentido, y más en un análisis de apartamentos, donde las piletas y gimnasio son commodities propias del edificio al que pertenecen.

9.1.2 En Casas

Al igual que como hicimos con apartamentos, visualizamos un histograma para poder separar en grupos:

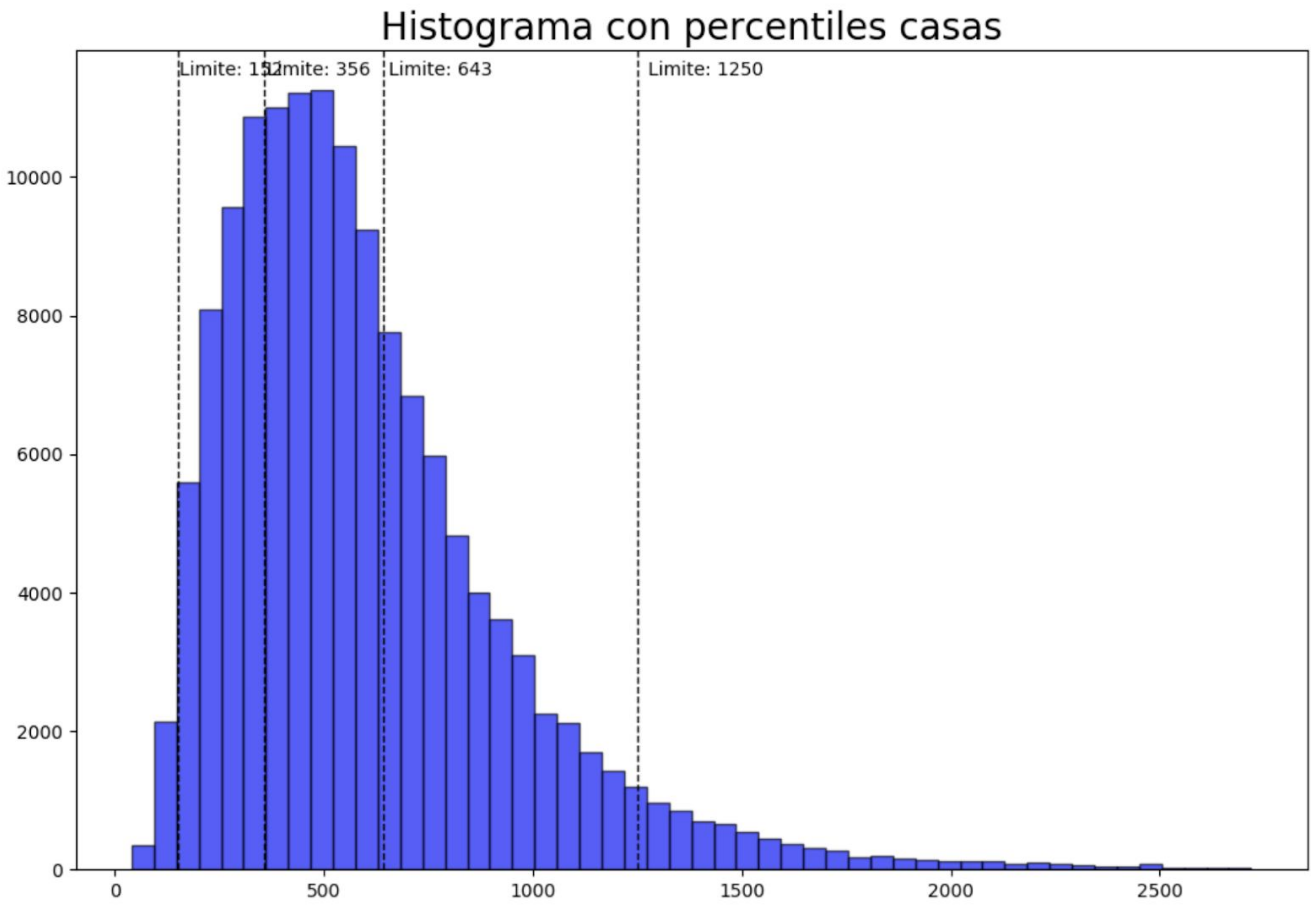

Gráfico 9.1.2.1

- grupo de propiedades muy baratas: hasta 152 USD
- grupo de propiedades baratas: van desde 152 hasta 356 USD
- grupo de propiedades “medias” (ni muy caras ni muy baratas): desde 356 hasta 643 USD
- grupo de propiedades caras: desde 643 hasta 1250 USD
- grupo de propiedades muy caras: desde 1250 hasta 2700 USD

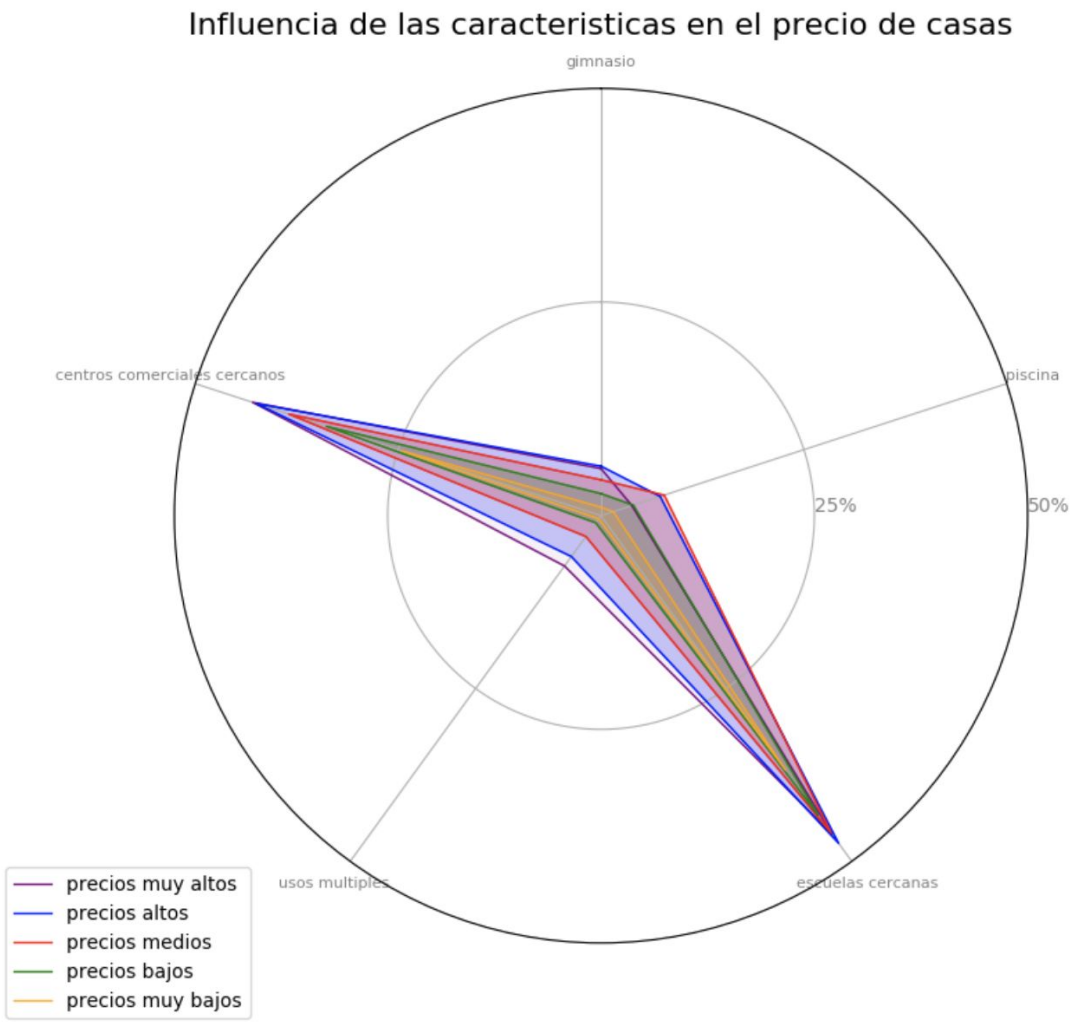

Gráfico 9.1.2.2

Dicho gráfico, a diferencia del Gráfico 9.1.1.2 correspondiente a Apartamentos, no nos aporta nada nuevo. Nos permite ver gráficamente cómo para las casas, el precio por metro cuadrado no está marcado por estas características (porque por lo general carecen de ellas), sino por otras que quedaron fuera de este análisis particular (pueden ser cantidad de ambientes, baños, garage, entre otras). Es interesante también notar cómo se mantiene la relación “cercanía con escuelas” y “cercanía con centros comerciales” respecto a los apartamentos.

9.2 Análisis de frecuencia de palabras

En el Dataset se analizaron las palabras más comunes y frecuentes en los títulos de las publicaciones, ya que es con lo primero que se encuentra uno al realizar búsquedas en la página de zonaprop y algo que debe dar un mini resumen de su contenido.

Se optó por realizar un WordCloud, en donde las palabras más frecuentes tiene mayor tamaño.

9.2.1 Titulos en general

Gráfico 9.2.1

Lo que se puede obtener de aquí corrobora lo se observo del gráfico 3.5.1, las propiedades tipo **casa** son las que predominan entre todas las publicaciones dejando en segundo y tercer lugar lugar a los **departamentos** y **condominios**. Obviamente la palabra “venta” (en venta, venta en) es la más frecuente en los títulos pero no es importante ya que el dataset es de una página de ventas de propiedades.

9.2.2 Títulos de Distrito Federal

Gráfico 9.2.2

En los títulos de las publicaciones de Distrito Federal, como en el anterior predomina “venta” por las mismas razones mencionadas. La palabra más frecuente y de importancia es este bloque es **departamento,** ya se analizó con anterioridad que en Distrito Federal predominan las ventas de departamentos, por lo que este hecho se ve reflejado en los títulos de las publicaciones de dicha provincia.

Se relega a las propiedades tipo **casa** y **condominios** (casa en condominio) a un segundo y tercer puesto respectivamente, tal y como se muestra en su respectivo análisis (Parte 8.1).

10. Conclusiones

A modo de conclusión, observamos que el dataset proporcionado cuenta con muchas propiedades cercanas al centro administrativo del país, y en menor medida ciertas ciudades del resto del país. Esto puede deberse a la popularidad de ZonaProp en México, como así también a la concentración edilicia en zonas administrativas.

En cuanto a la popularidad de ZonaProp en México, al pasar de los años, la página se tornó aún más popular con un auge durante todo el 2016 en especial en diciembre, duplicando la cantidad de publicaciones realizadas en todos los años anteriores. Esta popularidad se volcó sobre la gran cantidad de publicaciones realizadas en las provincias de Distrito Federal y Edo. de México. La gran cantidad de publicaciones son de Casas y Apartamentos, dejando en segundo plano los demás tipo de propiedades.

En cuanto al precio por metro cuadrado podemos concluir que a medida que uno se acerca al centro administrativo y comercial del país, mayor es el precio por metro cuadrado, desde las propiedades residenciales, hasta las comerciales. Al distanciarse del centro administrativo, las propiedades rurales comienzan a tener protagonismo dejando de lado las propiedades residenciales y comerciales cómo es el caso de Guerrero. Correspondiendose con lo anterior las ciudades con las publicaciones más gravosas son las más cercanas a la capital, con algunas excepciones hacia el este del país.

Viendo la antigüedad de las publicaciones y relacionándolo con el precio se puede ver que hay una relación directa, donde a mayor antigüedad, menor es el precio, salvo la irregularidad de Distrito Federal la cual siendo una de las provincias con mayor antigüedad promedio, tiene también el mayor precio por metro cuadrado promedio. Claramente, en esta provincia no importa la propiedad sino la cercanía con las ciudades importantes.

Distrito Federal, teniendo el papel protagónico en la página por ser el centro administrativo, es la provincia con mayor cantidad de publicaciones y precio promedio por metro cuadrado. Viendo que Miguel Hidalgo, Cuauhtémoc y Benito Juárez, son las principales estrellas dentro de esta provincia ya que son las más caras, según su precio por metro cuadrado donde se destacan principalmente las propiedades comerciales.

Cómo recomendación invertiría publicidad en las provincias con menor cantidad de publicaciones pero de mayor valor por metro cuadrado en este caso un ejemplo podría ser Nueva León, ya que, la página recauda plata según la visualización de la propiedad y las propiedades de menor valor, no se arriesgan a pagar un valor de visualización elevado. Siendo así mayor la cantidad de publicaciones de las cuales Navent obtiene un beneficio.

También podríamos repetir muchas más conclusiones, que como mencionamos anteriormente se encuentran dispersadas en los distintos puntos de análisis, pero aquí citamos a aquellas a las cuales se les puede observar una utilidad práctica muy directa.