

UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA

75.06 ORGANIZACIÓN DE DATOS

TRABAJO PRÁCTICO 1

ANÁLISIS EXPLORATORIO

2C - 2019

Grupo 31

Pablo De Giano	95040
Nicolas Nieva	101067
Maximiliano Romero	99118
César Castellanos	81404

<hoja en blanco>

Introducción	4
Introducción a México - División política y ubicación geográfica	4
Gobierno y división política	4
Ubicación geográfica	4
Primera Parte - Análisis inicial	5
Hay publicaciones duplicadas?	5
Wordcount del dataframe	6
Wordcount de las columnas detalle y título	6
Matriz de Correlación	7
Segunda Parte - Análisis de las publicaciones	8
Los 10 tipos de propiedades con más publicaciones	8
¿Cuales son las provincias con más publicaciones?	8
¿Cómo están distribuidas las publicaciones en el mapa?	9
¿Cómo fue la evolución de las publicaciones por día?	9
¿Cual es el día de la semana que se realizan más publicaciones?	10
¿Cómo fue la evolución de las publicaciones por año?	10
¿Cuales son las publicaciones por mes?	11
¿Como es la evolución de las publicaciones por año y mes?	11
Heatmap de las publicaciones por día y mes	12
Mapa de la ubicación de las publicaciones.	13
Boxplot de Precios de propiedades	13
Histogramas por antigüedad y habitaciones	14
Tercera parte - Análisis sobre Estado de México y Distrito Federal	15
Cuales son las ciudades con más publicaciones en el Estado de México	15
Mapa de cantidad de publicaciones en el Estado de México	15
Cuales son las ciudades con más publicaciones en el DF	16
Mapa de cantidad de publicaciones en el DF	16
Cuál es la cantidad de publicaciones por tipo de propiedad	17
Cuarta parte - Análisis sobre los Amenities de las publicaciones	18
Cuáles son los tipos de propiedad según cantidad de Amenities	20
Cantidad de publicaciones según los distintos Amenities	21
Cómo varía el precio promedio de cada propiedad según Amenities	22
Cantidad de publicaciones contabilizando todos los Amenities	24
Cómo varía el precio según la cantidad total de Amenities	25
Quinta parte - Análisis sobre “Otros servicios” en las publicaciones	25
Cuales son las ciudades con mayor algún servicio	26
Cuales son los tipos de propiedad que se encuentran cercanos a estas edificaciones	27
Qué porcentaje de publicaciones cuentan cercanas a estas edificaciones	28
Cómo varía el precio según dichos servicios	29
Cómo varía el precio y la cantidad de publicaciones según la cantidad de servicios prestados	30

Sexta Parte - Análisis de Precios	31
¿Cual es la relación del precio con la antigüedad de todas las propiedades?	31
¿Cual es la relación del precio con la cantidad de habitaciones?	31
¿Cual es el precio promedio por provincia?	32
¿Cuál es la relación entre metros cubiertos y precios dependiendo de la cercanía con escuelas?	33
¿Cuál es la relación entre metros cubiertos y precios dependiendo de la cercanía con centros comerciales?	34
Precio promedio de los 5 tipos de propiedad más publicados	35
Análisis de precios en Apartamentos	36
Precio promedio por cantidad de habitaciones	36
Precio promedio por antigüedad	36
Histograma de precios	37
Precio promedio dependiendo de cantidad de amenities	37
Análisis de precios en Casas	38
Precio promedio por cantidad de habitaciones	38
Precio promedio por antigüedad	38
Histograma de precios	39
Precio promedio dependiendo de cantidad de amenities	39
 Séptima Parte - Análisis de propiedades “con todo” y “sin nada”	 40
¿Cómo varía la cantidad de publicaciones de propiedades con nada y sin todo con respecto a la antigüedad de las mismas?	43
¿Cuales son las ciudades con mayor cantidad de propiedades con todo?	44
¿Cuales son las ciudades con mayor cantidad de propiedades sin nada?	47
¿Cual es la distribución de metros totales y cubiertos en propiedades con todo, sin nada y en general?	50
¿Cual es el precio promedio de propiedades con todo y sin nada?	51
¿Cual es la cantidad de publicaciones de propiedades con todo y sin nada?	51
 Conclusiones	 52

Introducción

El presente trabajo de análisis exploratorio se basa en el sitio zonaprop.com perteneciente a la empresa Navent. ZonaProp es un sitio de compra-venta y alquiler inmobiliario donde desarrolladores, inmobiliarias y anunciantes particulares pueden publicar su oferta, mientras que, los interesados, pueden encontrar la propiedad que están buscando según las características deseadas por cada uno. Durante dicho trabajo se utilizarán las publicaciones referidas a las publicaciones de ofertas inmobiliarias de México.

El repositorio se encuentra en https://github.com/cessar5/7506_20192C

Introducción a México - División política y ubicación geográfica

Debido a que el análisis se basa sobre datos de México haremos una breve mención política y geográfica que nos permitirá tener mayor conocimiento del set de datos. La misma es una referencia a ***El Programa de las Naciones Unidas para el Desarrollo***¹.

Gobierno y división política

México está organizado como una República representativa, democrática y federal. Está dividido políticamente en 32 entidades federativas, de las cuales 31 son estados libres y soberanos en su régimen interior y un Distrito Federal donde residen los Poderes Federales.

Cabe aclarar que desde 2016 el Distrito Federal es Ciudad de México.²

Ubicación geográfica

La República Mexicana está situada en el continente americano en el hemisferio norte; parte de su territorio se encuentra en América del Norte y el resto en América Central. Su superficie territorial es de 1,967,183 Km².

Actualmente una línea fronteriza de 3,152 Km. marca el límite de la frontera norte con los Estados Unidos de América. Al sur, México limita con las Repúblicas de Guatemala y Belice mediante una línea fronteriza sinuosa, que alcanza 1,149 Km de extensión total.

México se extiende entre los paralelos 14° 32 ´ 27" en la desembocadura del río Suchiate y el paralelo 32° 43 ´ 06" que pasa por la confluencia del río Gila con el Colorado; así mismo está comprendido entre las longitudes oeste de Greenwich de 118° 22 ´ 00" y 86° 42 ´ 36" respectivamente.

¹ <http://www.mx.undp.org/content/mexico/es/home/countryinfo.html>

² <https://www.lanacion.com.ar/el-mundo/mexico-df-ciudad-enrique-pena-nieto-estado-nid1867015>

Primera Parte - Análisis inicial

Se observa un buen set de datos con información clara sobre las publicaciones. El periodo comprendido del set de datos corresponde al periodo enero de 2012 a diciembre de 2016

En cuanto a datos nos encontramos con datos consistentes . Los datos que tienen alta tasa de nulos son latitud y longitud con un porcentaje del 50%, es decir, que la mitad de los datos que hacen referencia a la latitud y longitud de cada propiedad no está disponible en el set de datos, y las direcciones y metros totales que tienen una tasa del 20% de datos nulos.

	Cantidad Nulls	Porcentaje(%)
lng	123488	51.45
lat	123488	51.45
direccion	53072	22.11
metrostotales	51467	21.44
antiguedad	43555	18.15
garages	37765	15.74
idzona	28621	11.93
banos	26221	10.93
habitaciones	22471	9.36
metroscubiertos	17400	7.25
titulo	5387	2.24
descripcion	1619	0.67
ciudad	372	0.16
provincia	155	0.06
tipodepropiedad	46	0.02
precio	0	0.00
centrocomercialescercanos	0	0.00
fecha	0	0.00
gimnasio	0	0.00
usosmultiples	0	0.00
piscina	0	0.00
escuelascercanas	0	0.00
id	0	0.00

Hay publicaciones duplicadas?

Luego buscamos en el set de datos si existían posibilidad de que existen para una misma propiedad más de una publicación. Nos encontramos que solo 4 de ellas se encuentran repetidas lo que generaba un cambio determinante en el análisis.

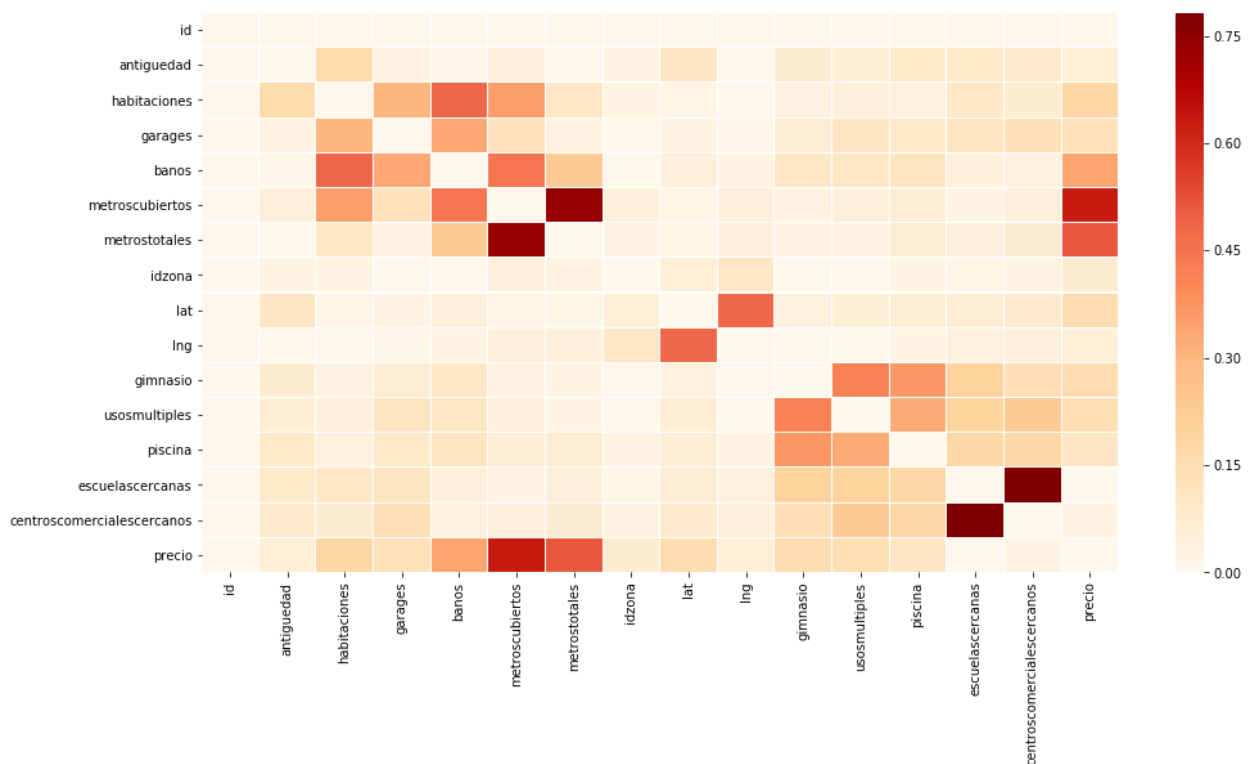
Wordcount del dataframe



Wordcount de las columnas detalle y título



Matriz de Correlación



Se observan algunas fuertes relaciones entre las diferentes características de las propiedades. En primer lugar vemos una gran relación entre las escuelas cercanas y los centros comerciales cercanos a cada propiedad. Lo que podemos empezar a pensar que de que éstas dos edificaciones se encuentran por la misma zona, en otras palabras, que si el inmueble tiene un centro comercial cercano, es muy probable que también tenga escuelas aledañas.

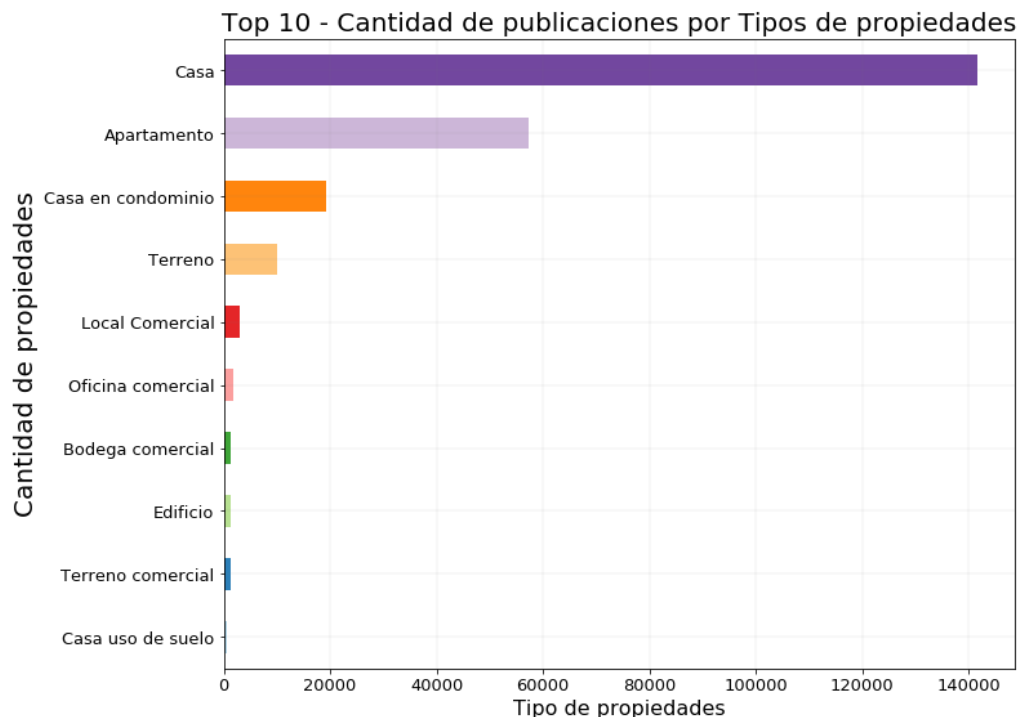
Por otro lado, se puede notar que el precio está más fuertemente relacionado con los metros cubiertos y los metros totales de cada propiedad. Y en menor medida la cantidad de baños que cuenta la misma. A su vez, a primera vista, se ve que dicho valor del inmueble no se encuentra relacionado con la presencia de instituciones cercanas (ya sea escuelas como centros comerciales).

Otra gran relación presente es la cantidad de metros cubiertos con la cantidad de metros totales, esta es una correspondencia bastante intuitiva ya que a mayor cantidad de metros cubiertos, mayor será la cantidad de metros totales (salvo que los datos no sean correctos).

Por último se puede observar otra relación y es entre la cantidad de habitaciones con la cantidad de baños que posee la vivienda. Éstas también se correlacionan, en menor medida, con la cantidad de metros cubiertas.

Segunda Parte - Análisis de las publicaciones

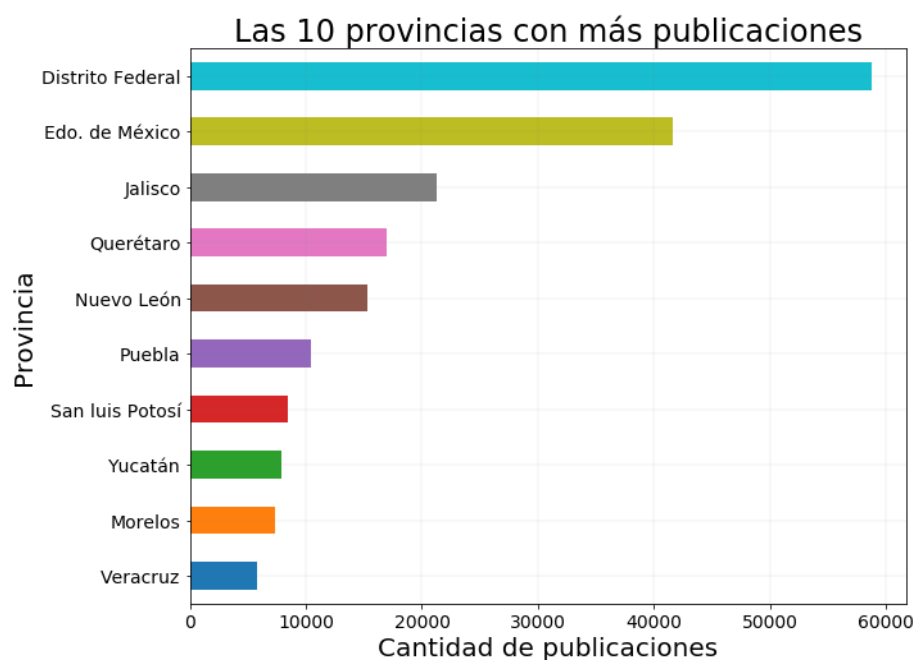
Los 10 tipos de propiedades con más publicaciones



En cuanto a los tipos de propiedades observamos que el set de datos está formado principalmente por un gran tipo número de publicaciones de casas y seguido por Apartamentos, y en menor medida se observan Casa en condominios.

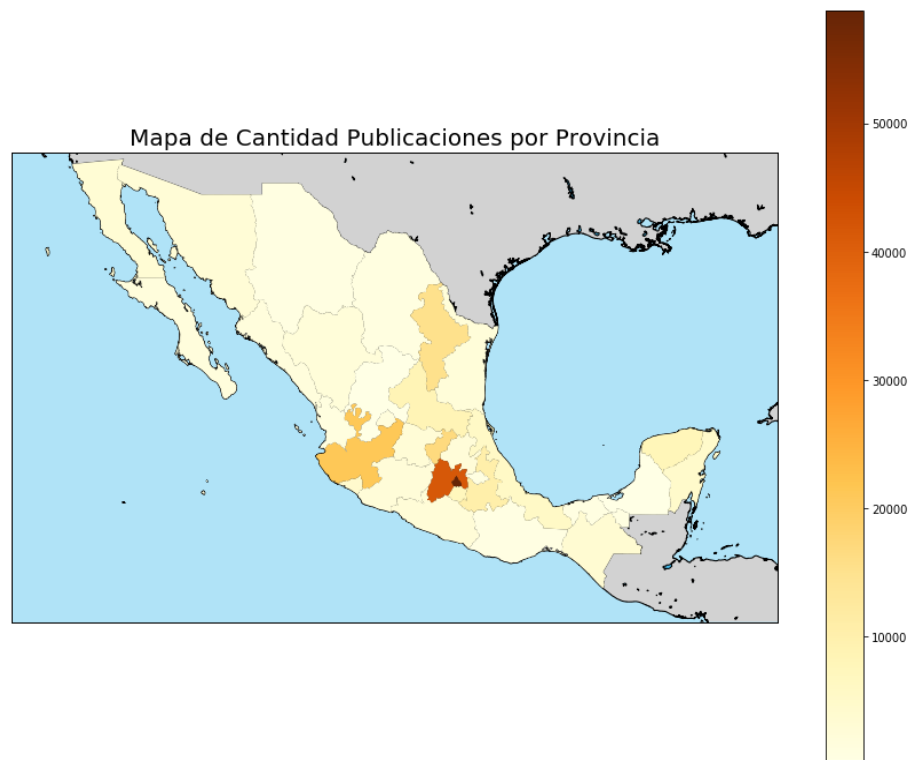
Los tipos de inmuebles que menos aparecen son Edificios, Terreno Comercial y Casa uso de suelo.

¿Cuales son las provincias con más publicaciones?



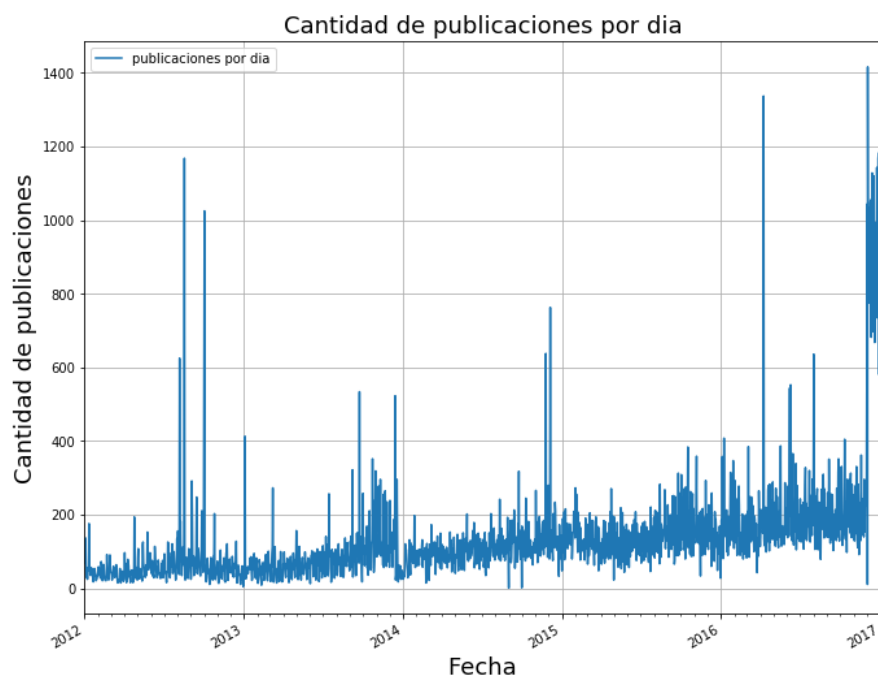
Aquí hacemos una distinción, ya que el Distrito Federal no es una provincia por sí misma pero se evalúa como una de ellas debido a su importancia, obteniendo la mayor cantidad de publicaciones. Luego le sigue el Estado de México, y en menor medida Jalisco.

¿Cómo están distribuidas las publicaciones en el mapa?



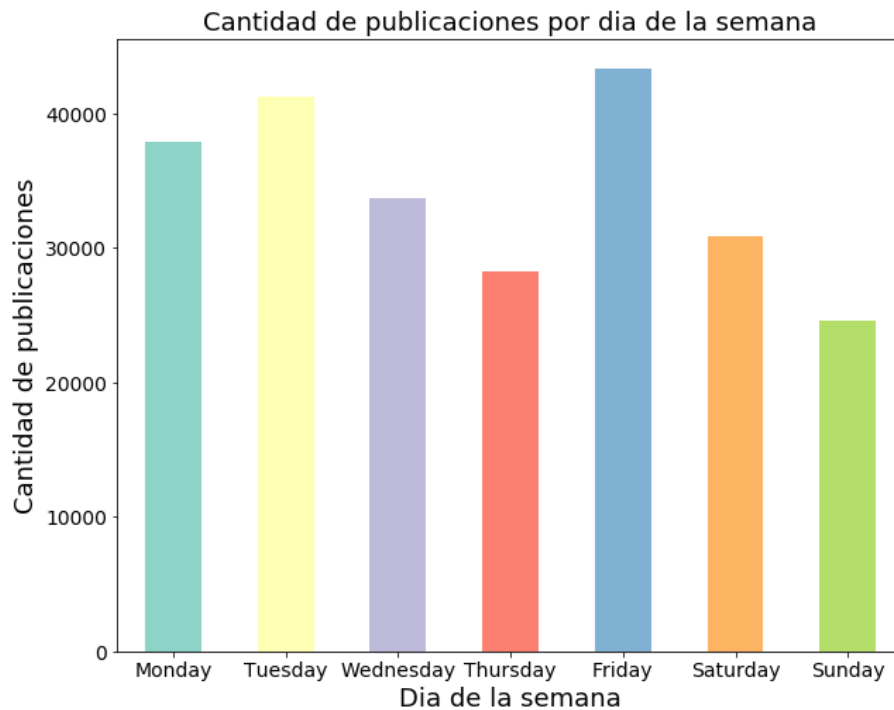
El mapa muestra la distribución de las provincias, donde se observa claramente la importancia del DF y el Estado de México.

¿Cómo fue la evolución de las publicaciones por día?



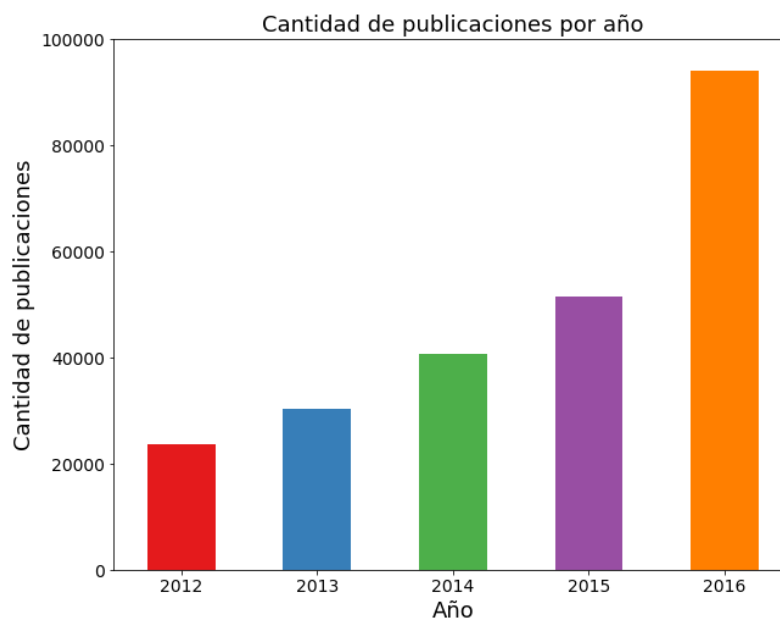
Observamos su evolución, con algunos picos determinados año por año, pero en el 2016 ya aumenta la cantidad de publicaciones regularmente. Ésto puede indicar que la plataforma online ganó su popularidad y confianza a través del transcurso de los años

¿Cual es el día de la semana que se realizan más publicaciones?



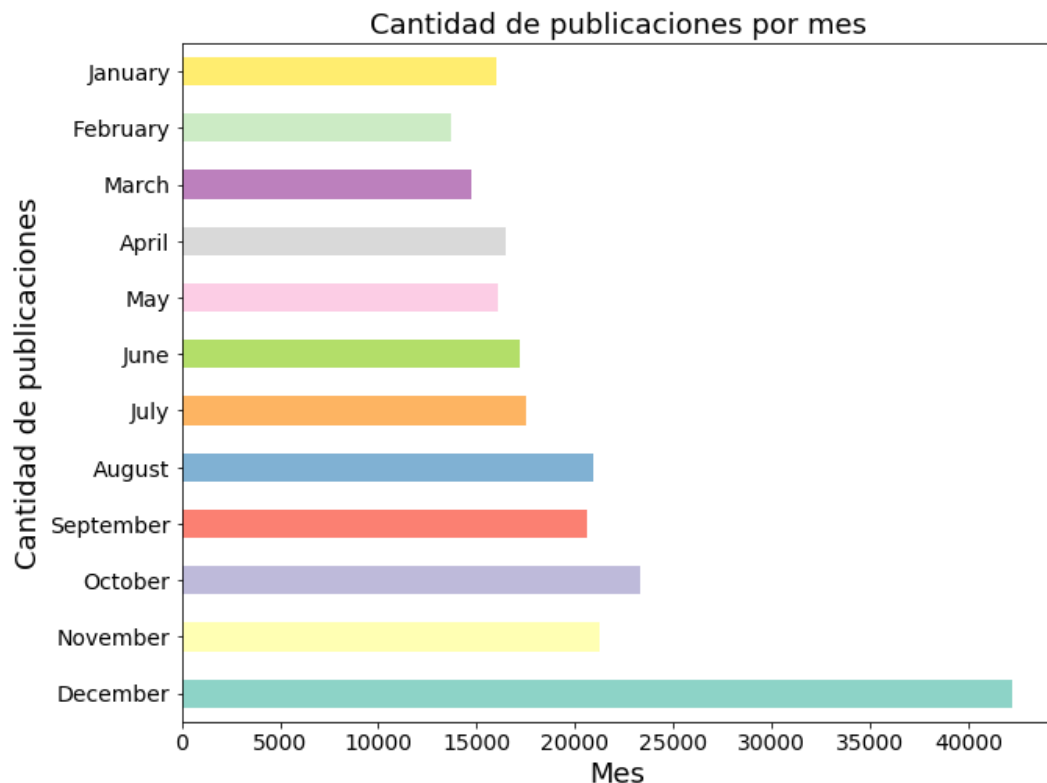
Aquí observamos que el día viernes es del día con mayor publicaciones tal vez el usuario que publica evalúa que el fin de semana una persona puede tener mayor disponibilidad de tiempo para buscar propiedades. En contrario el día domingo es el día con menor cantidad de publicaciones.

¿Cómo fue la evolución de las publicaciones por año?



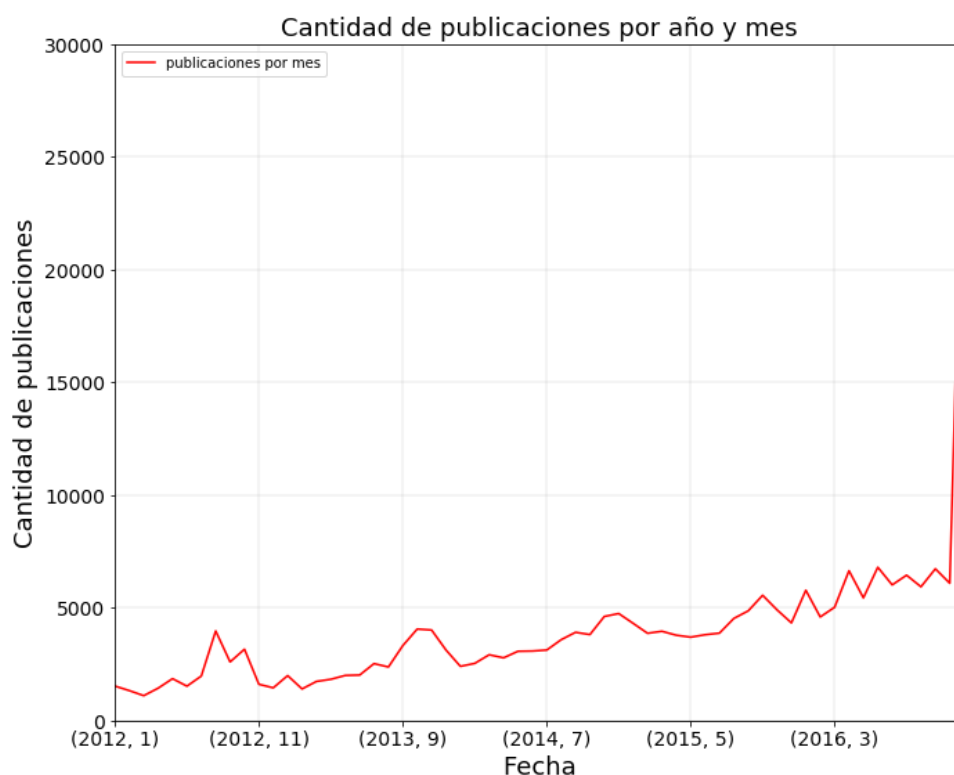
Aquí se observa un crecimiento creciente a medida que pasan los años, donde se observa un gran salto es en el año 2016. Como se dijo anteriormente puede tratarse por una mayor exposición del sitio, ya sea por los años de confianza o por una gran inversión en publicidad.

¿Cuales son las publicaciones por mes?



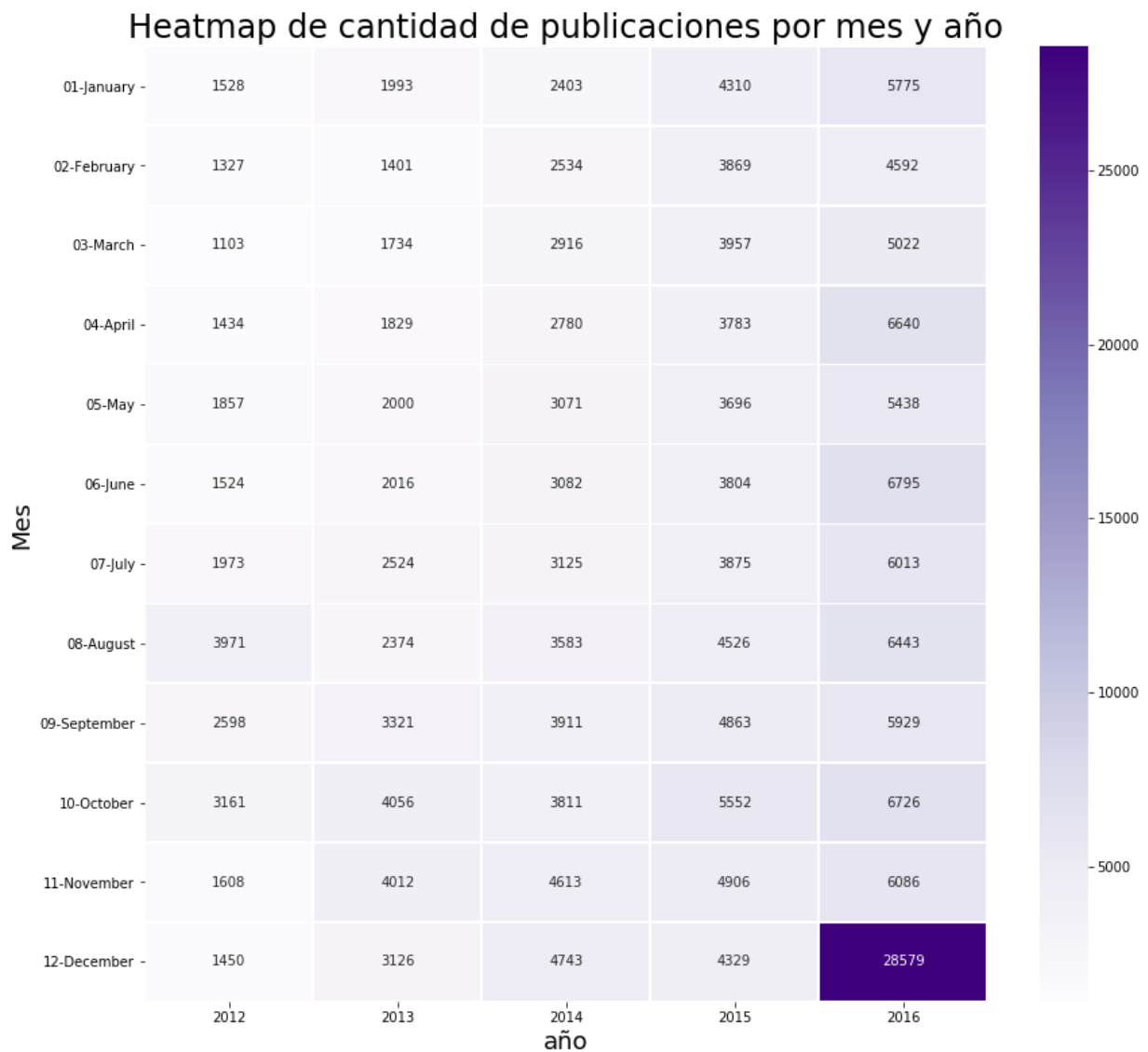
Observamos un gran número de publicaciones en Diciembre siendo parejo su comportamiento en los meses anteriores. Veremos en posteriores gráfica el análisis por año y mes para entender un poco más lo observado aquí.

¿Como es la evolución de las publicaciones por año y mes?



Aquí es donde entendemos lo visto en las gráficas anteriores. En diciembre de 2016 se observa un salto en las cantidad de publicaciones. La evolución hasta esa fecha tiene una evolución razonable pero salta.

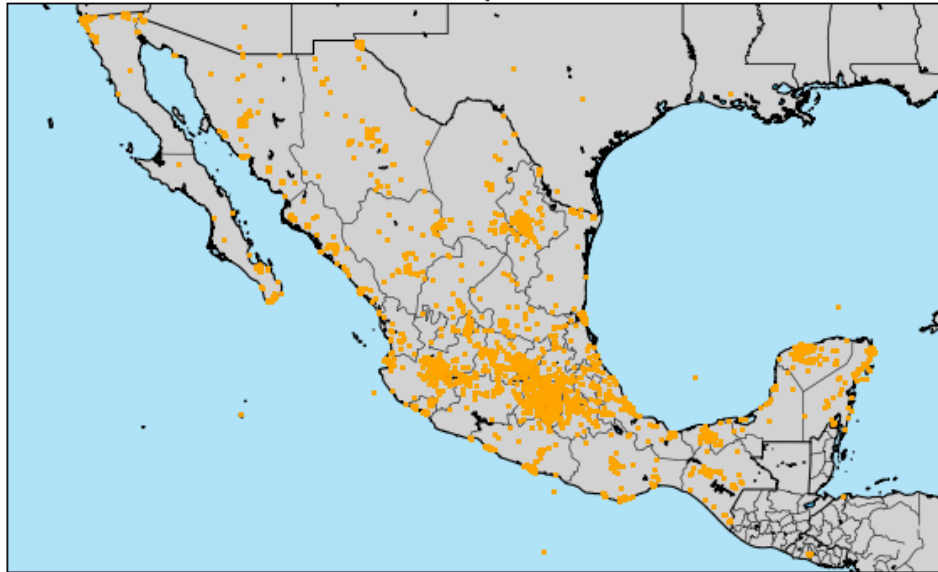
Heatmap de las publicaciones por día y mes



Gracias a dicho HeatMap podemos ver que la mayor ocurrencia de publicaciones se da en Diciembre de 2016, por la que podemos entender por qué es el mes con mayor cantidad de ofertas.

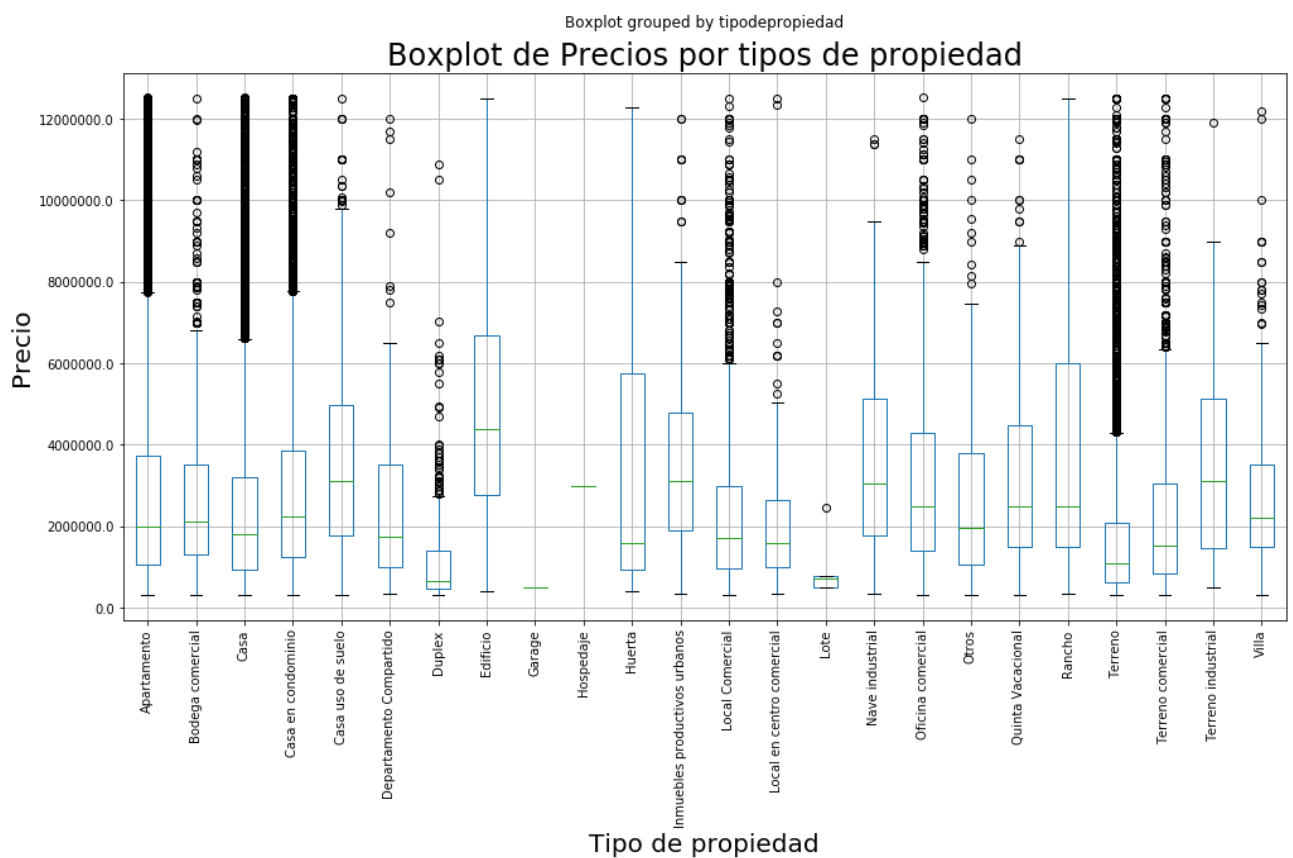
Mapa de la ubicación de las publicaciones.

Publicaciones por ubicación

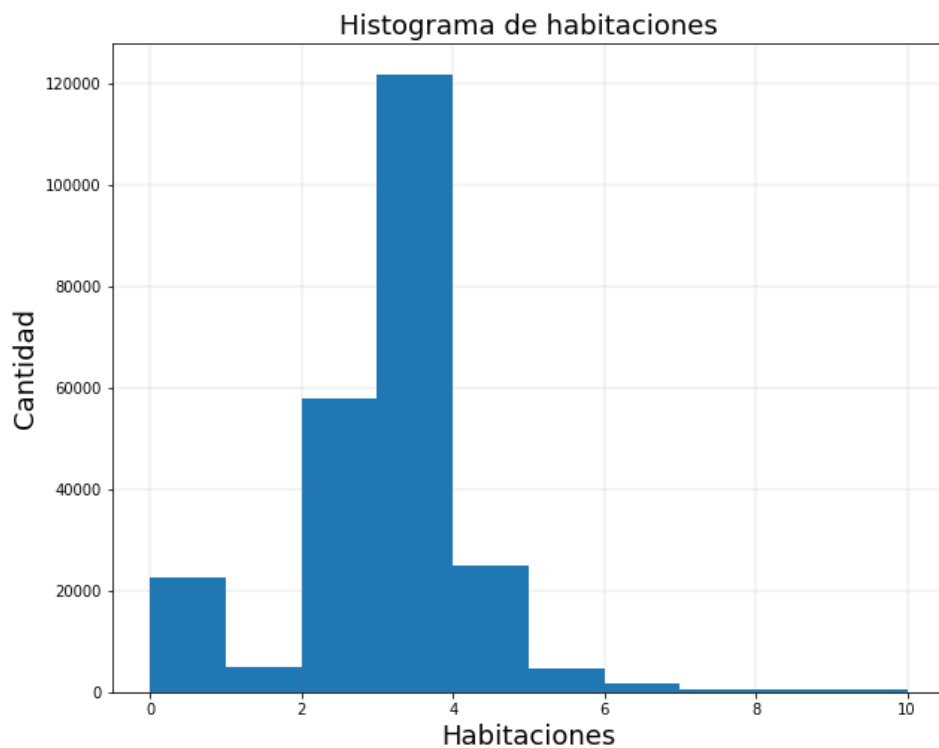
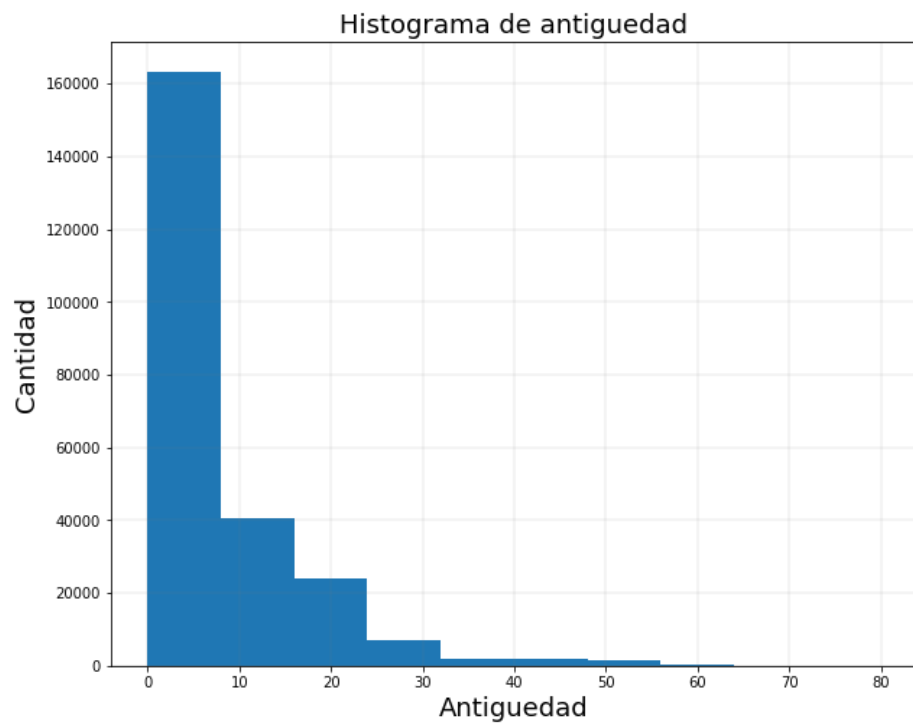


El mapa muestra cómo están distribuidas las publicaciones en el mapa. Si bien se tiene el 50% de los datos queremos mostrar la información disponible. Con los datos utilizados, también nos da la idea de la mayor concentración se da en la zona del Distrito Federal y el Estado de México.

Boxplot de Precios de propiedades



Histogramas por antigüedad y habitaciones



A partir de éstas últimas dos gráficas podemos ver que la mayoría de las propiedades publicadas en el sitio analizado poseen 3 habitaciones, y por otro lado, los inmuebles publicados suelen tener una antigüedad menor a 10 años.

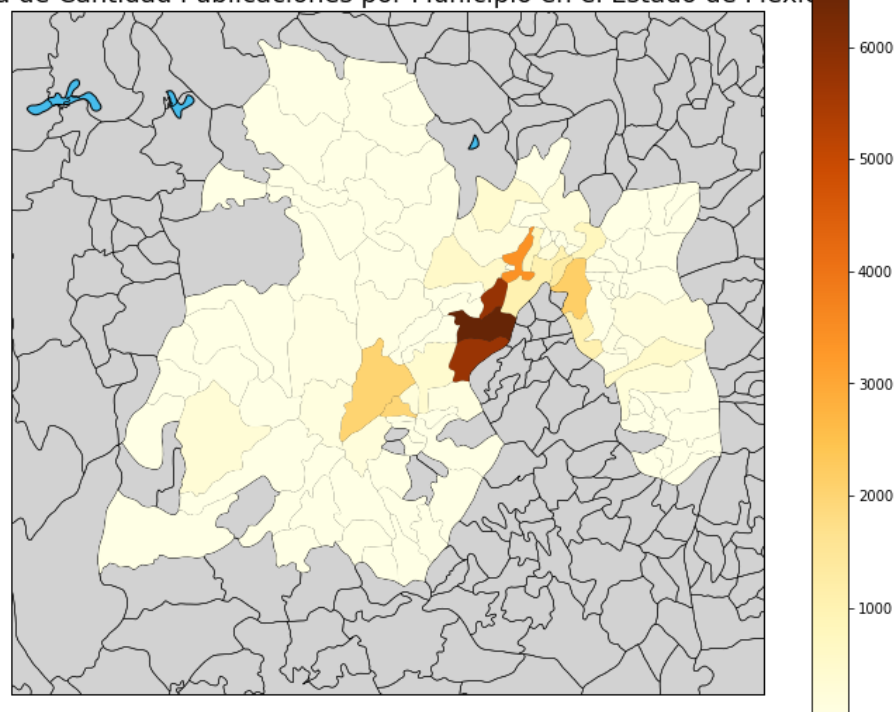
Tercera parte - Análisis sobre Estado de México y Distrito Federal

Cuales son las ciudades con más publicaciones en el Estado de México

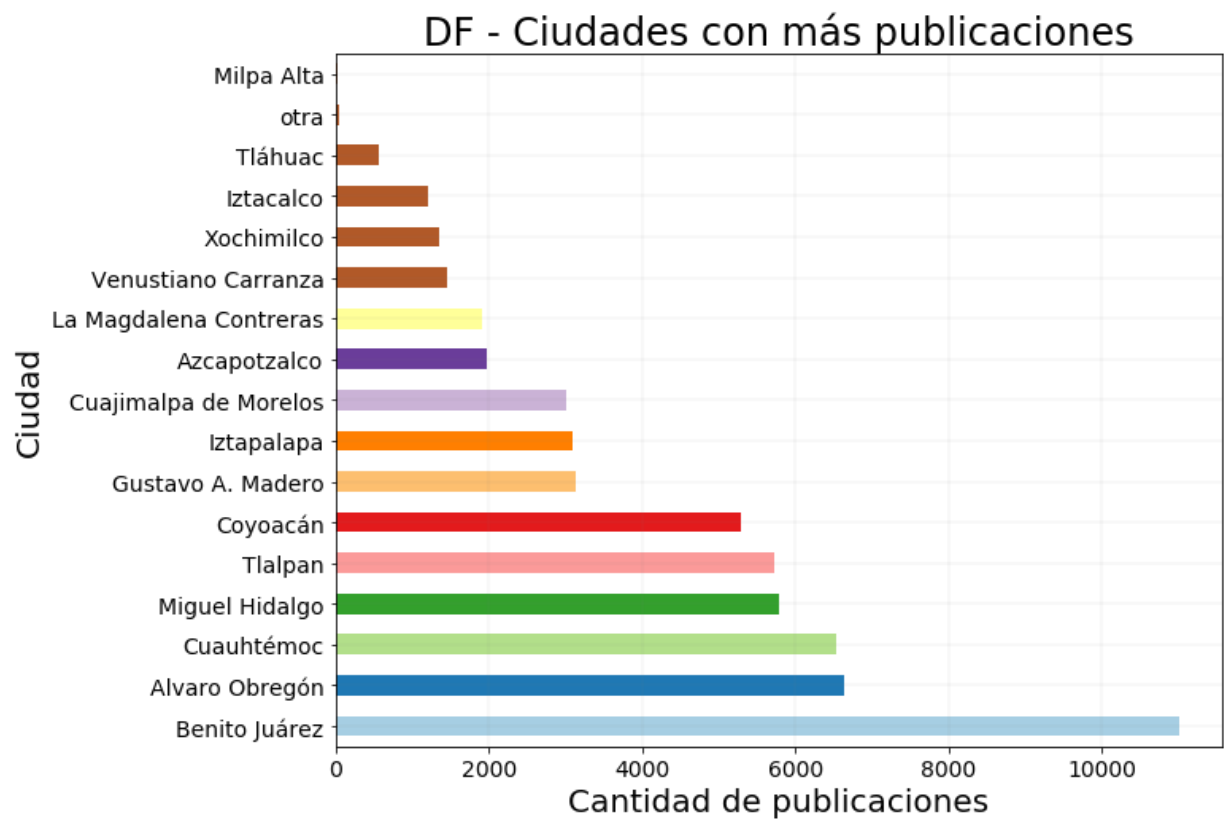


Mapa de cantidad de publicaciones en el Estado de México

Mapa de Cantidad Publicaciones por Municipio en el Estado de Mexico

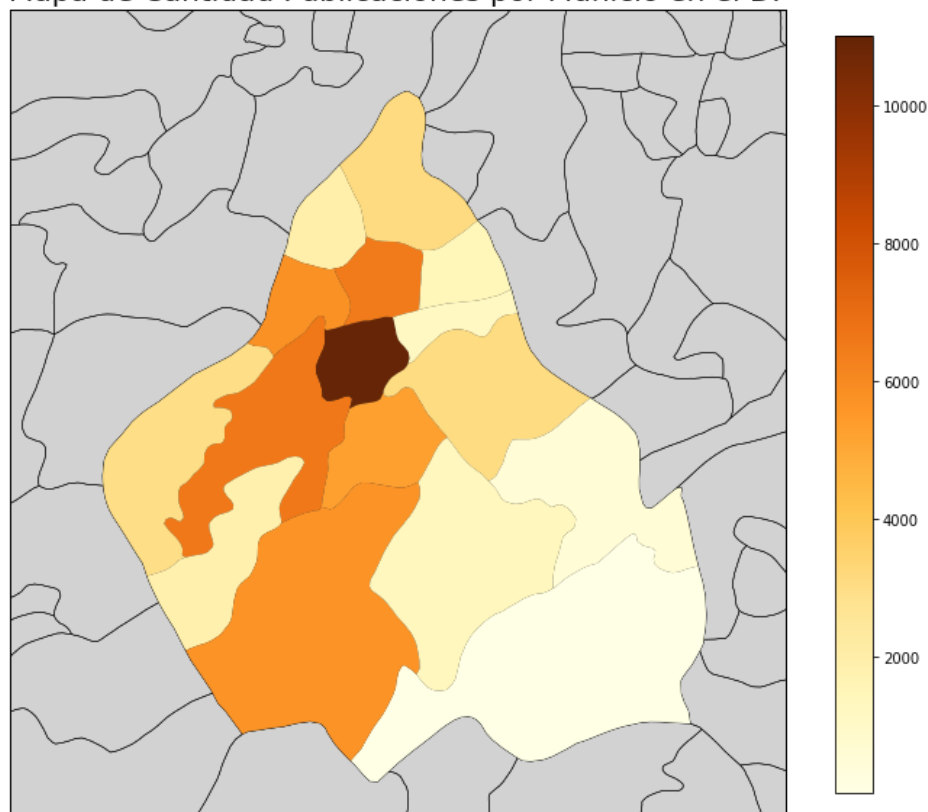


Cuales son las ciudades con más publicaciones en el DF



Mapa de cantidad de publicaciones en el DF

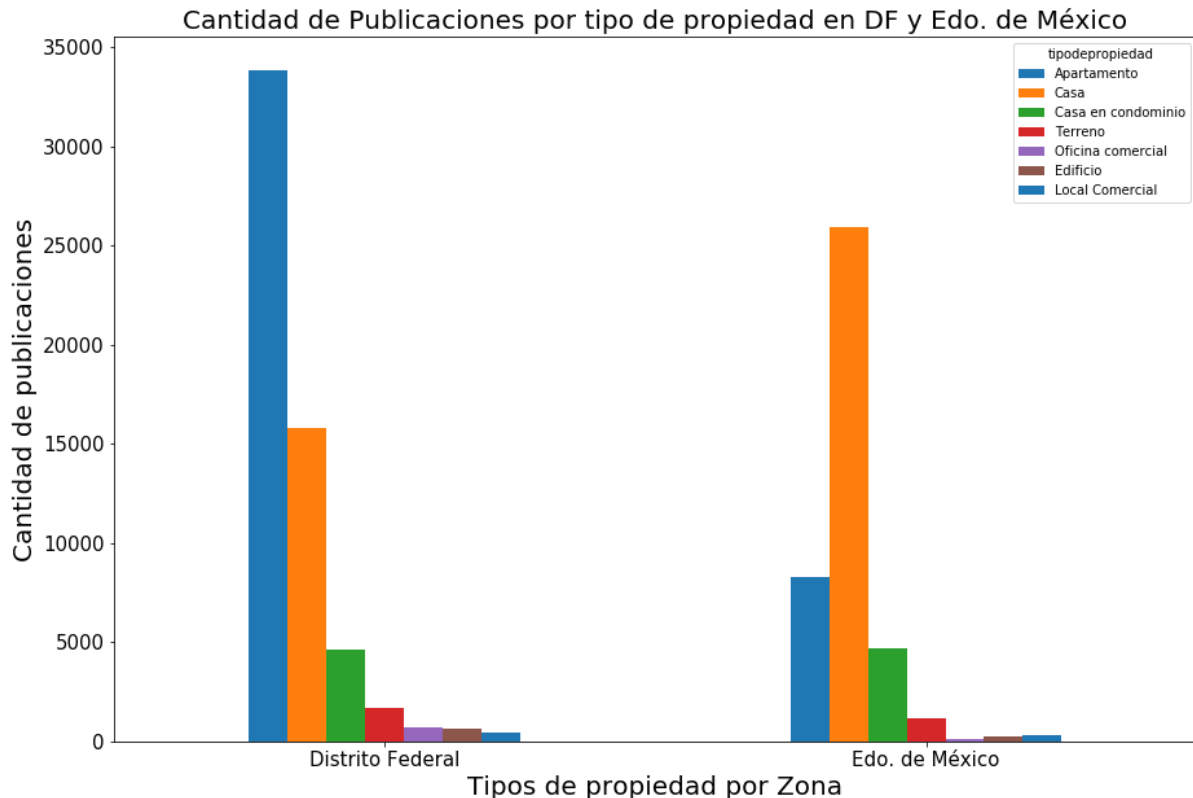
Mapa de Cantidad Publicaciones por Municio en el DF



Analizando las ciudades con más publicaciones tanto en el Estado de México como en el DF, podemos ver, que se trata de Naucalpan de Juárez en el Estado de México y *Benito Juárez en el DF*.

Sí ahora bien, observamos la ubicación de ambas ciudades, podemos darnos cuenta de que se trata de dos ciudades bastante cercanas, ya que se encuentran a una distancia de, aproximadamente, 19.4 km. Por lo que nos infiere que es la zona con mayor tráfico de ofertas dentro del sitio.

Cúal es la cantidad de publicaciones por tipo de propiedad



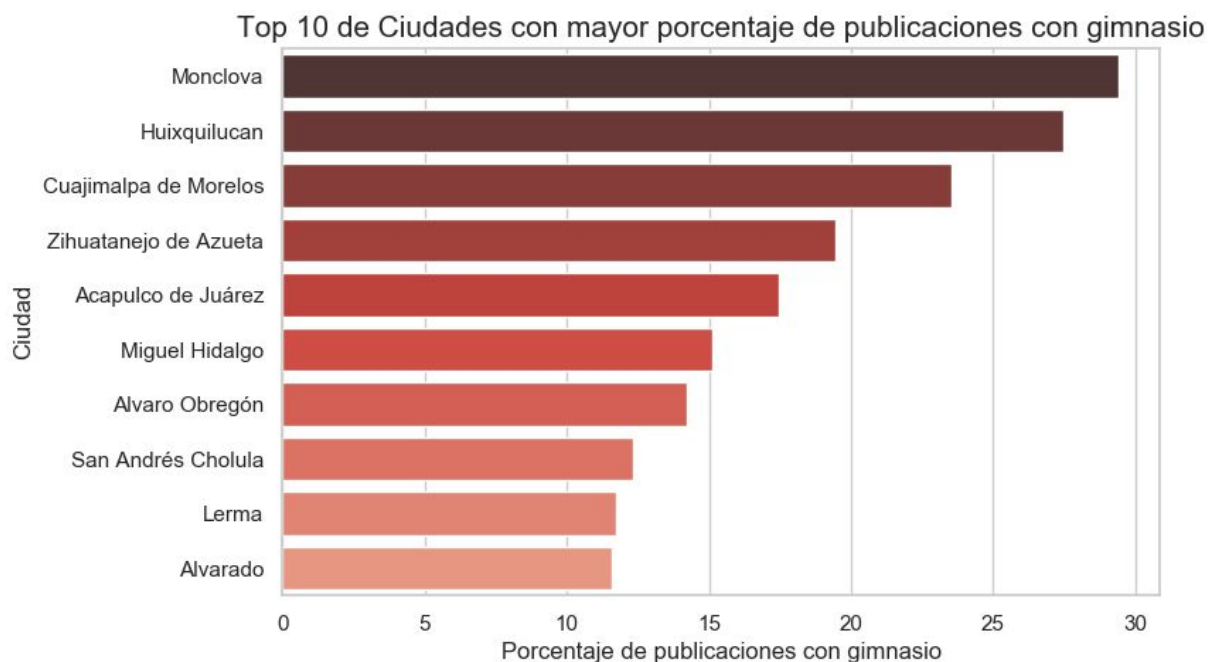
El gráfico anterior es muy interesante ya que hace una diferencia entre las dos provincias con más publicaciones. Podemos ver que en el DF hay una gran cantidad de ofertas de Apartamentos, mientras que en el Estado de México hay una mayor cantidad de publicaciones de casas

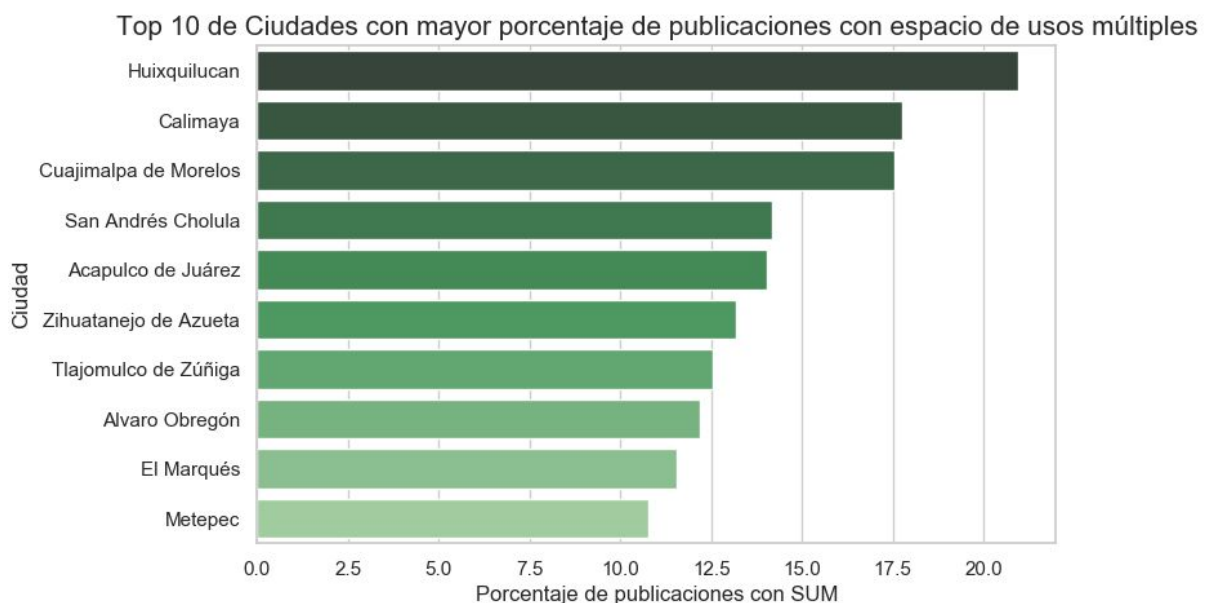
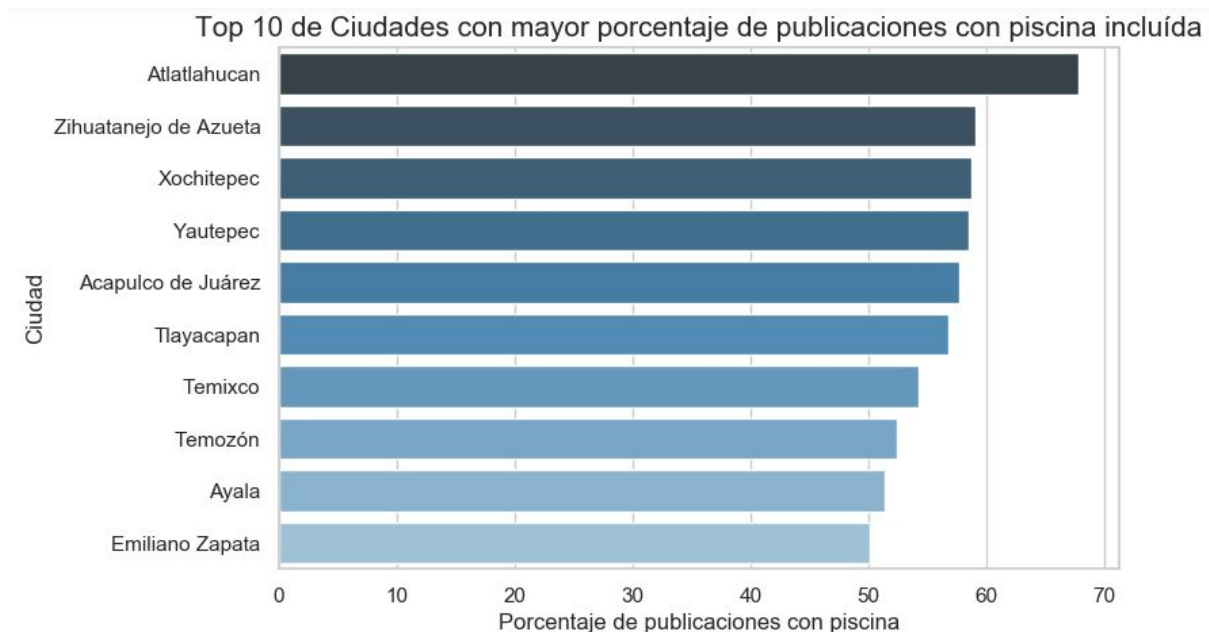
Por otro lado, también podemos observar que hay muchas más oficinas comerciales en el Distrito Federal.

Cuarta parte - Análisis sobre los Amenities de las publicaciones

Amenities se denominan a aquellos espacios de uso común que poseen las propiedades, como pueden ser las piletas de natación, salones de usos múltiples (SUM), gimnasios, entre otros. El fin de estos espacios es mejorar la calidad de vida de los habitantes a través de la comodidad y el confort.

Cuáles son las ciudades con mayor porcentaje de publicaciones según cantidad de Amenities

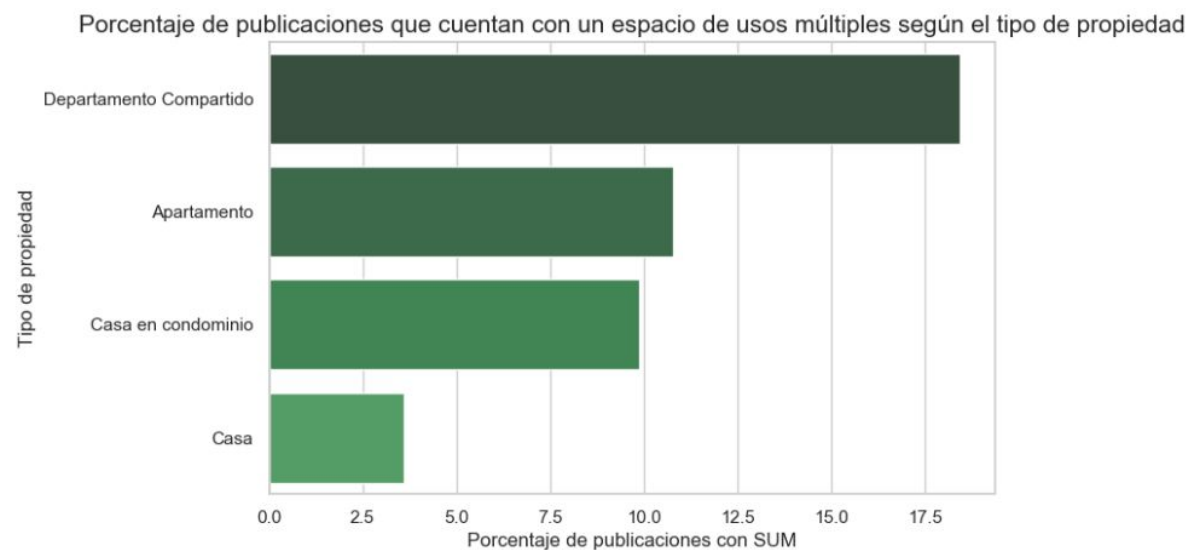
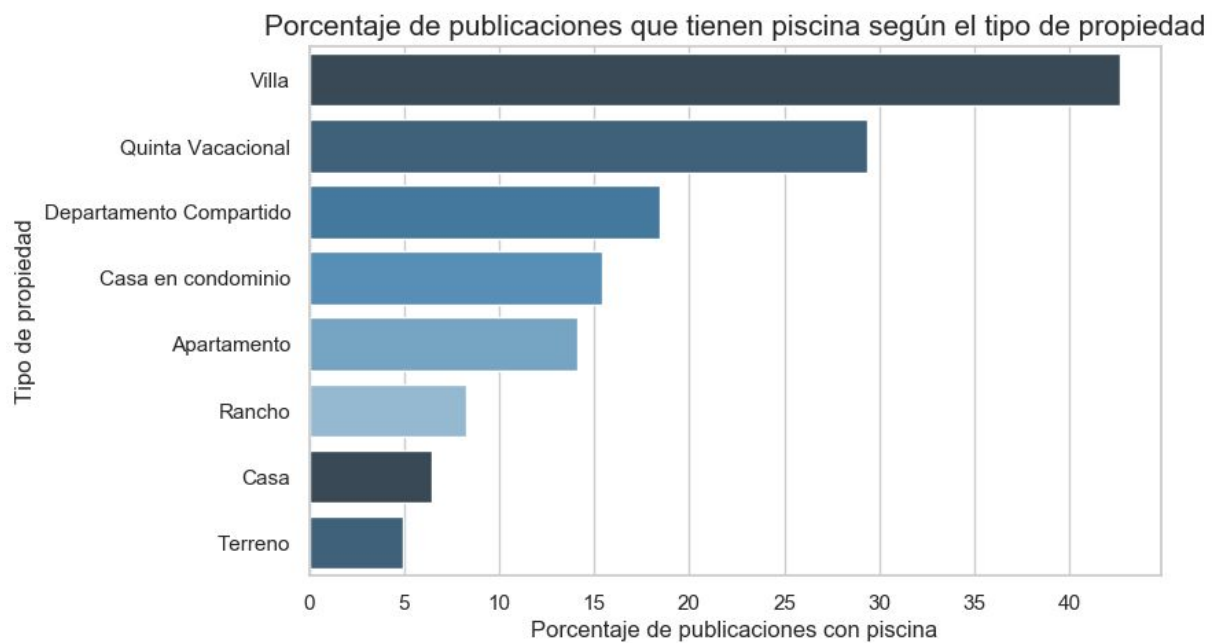
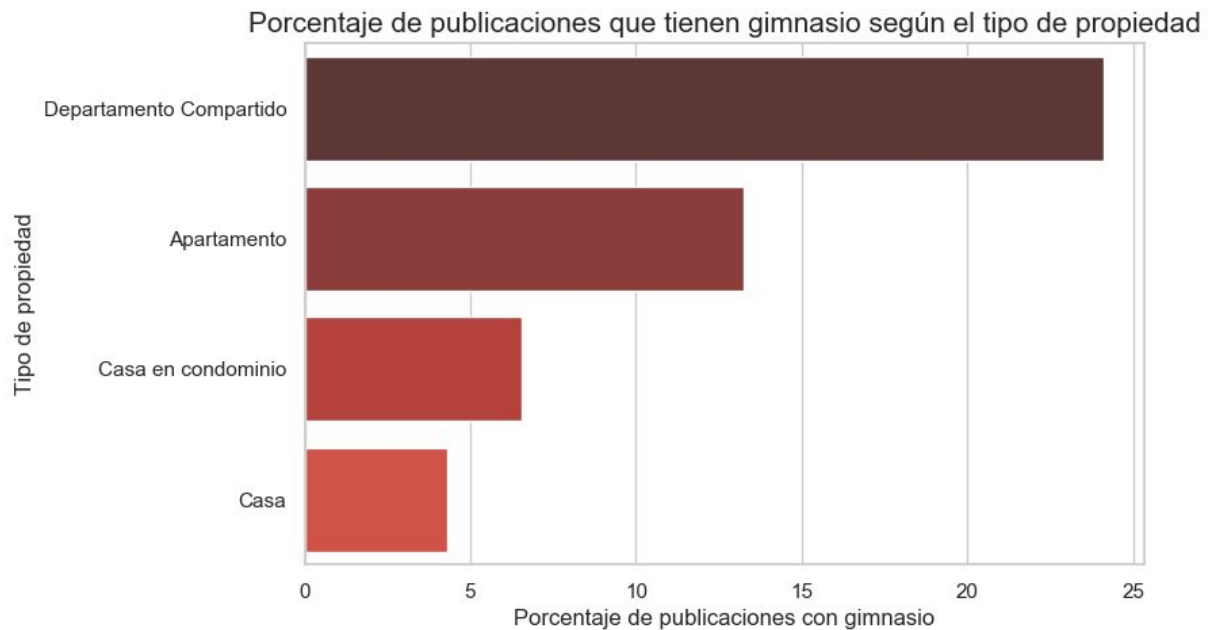




Analizando el **Top 10** de las ciudades con mayor porcentaje de publicaciones con los distintos Amenities podemos observar que hay una diferencia en cuanto al porcentaje de publicaciones con piscina. En él se puede ver que, las 10 ciudades presentes, tienen el 50% ó más de propiedades con pileta; mientras que, en los otros dos casos, no superan ni el 30%.

Por otro lado, la ciudad de **Huixquilucan** se encuentra primera en porcentaje referida a la tenencia de un SUM y segunda referida al gimnasio, mientras que ni figura en el gráfico que hace referencia a la piscina. Dicha ciudad pertenece a la Zona Metropolitana de la Ciudad de México. Gracias a ello, podemos inferir que hay una cierta relación entre la presencia de un espacio de usos múltiples y de un gimnasio. Mirando nuevamente la matriz de correlación podemos observar dicha correspondencia.

Cuáles son los tipos de propiedad según cantidad de Amenities

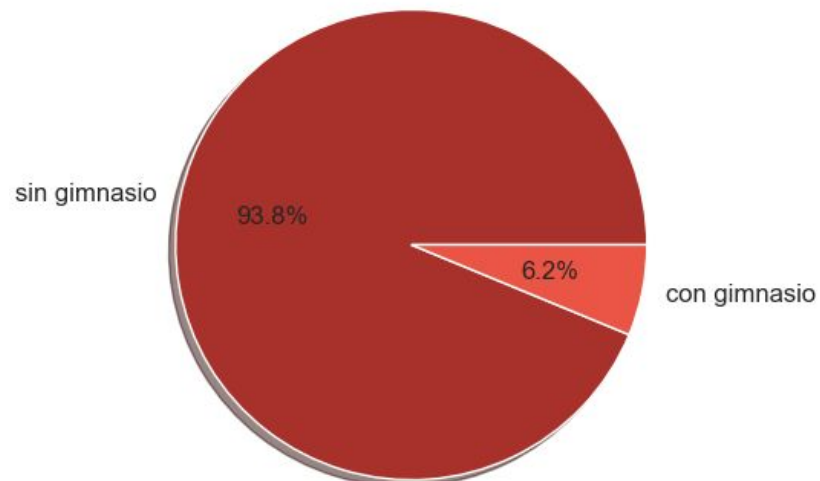


En este punto podemos ver que los tipos de propiedades que aparecen en el gráfico que hace referencia a los espacios de usos múltiples son exactamente iguales (a su vez con el mismo orden) que el que representa la presencia de un gimnasio. Ratifica la correlación entre gimnasio y SUM.

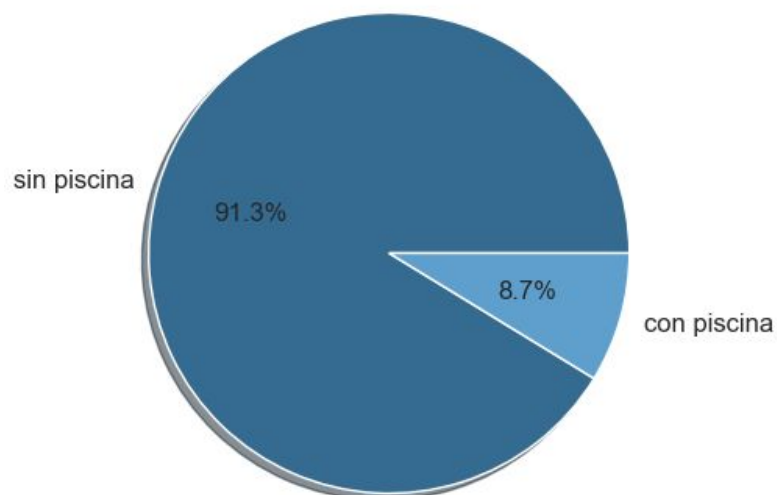
Mientras que para la pileta de natación prevalecen la **Villa** y la **Quinta Vacacional**.

Cantidad de publicaciones según los distintos Amenities

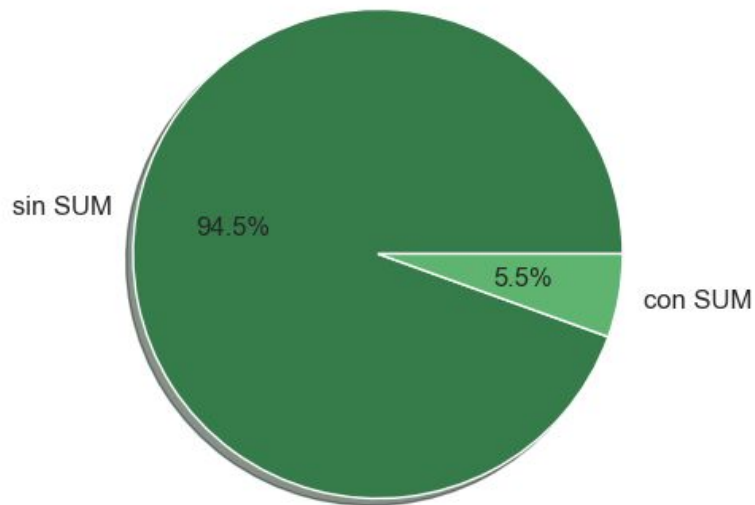
Cantidad de Publicaciones con Gimnasio



Cantidad de Publicaciones con Piscina

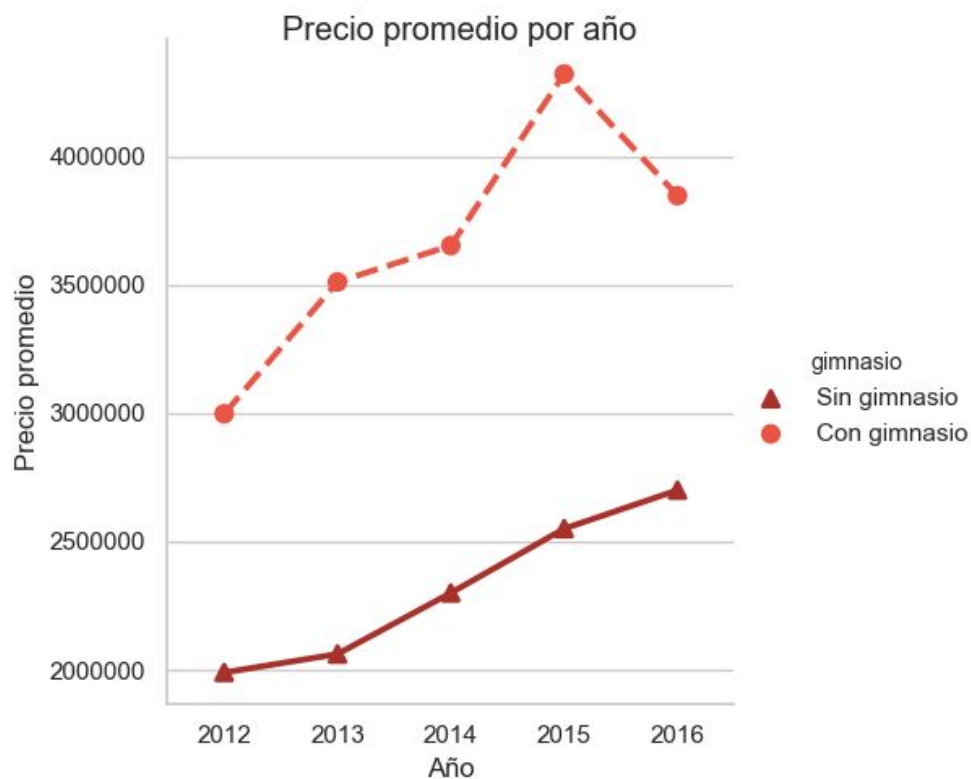


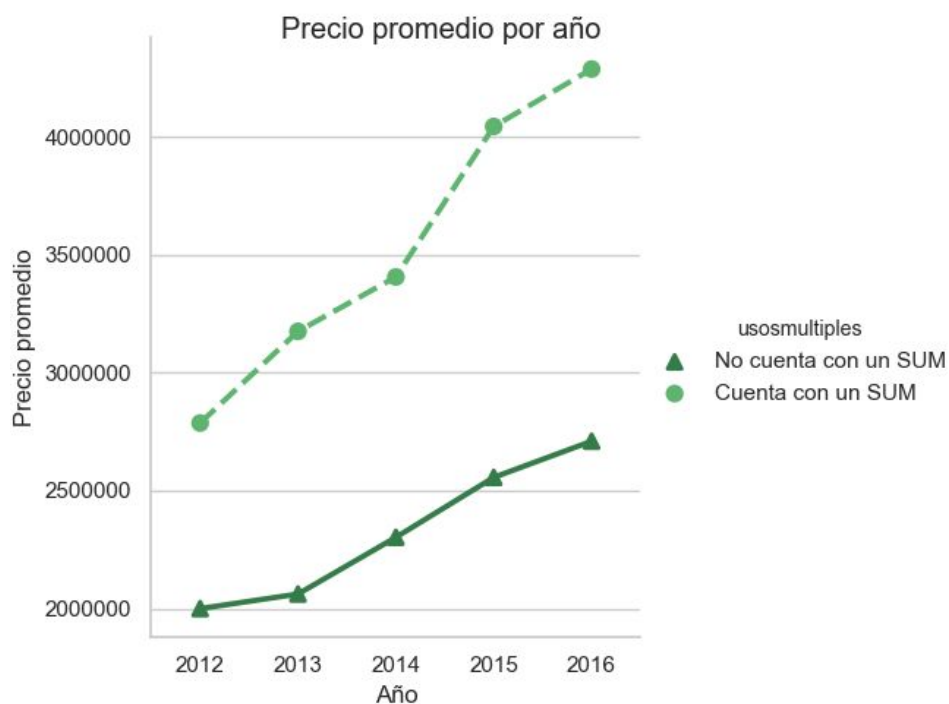
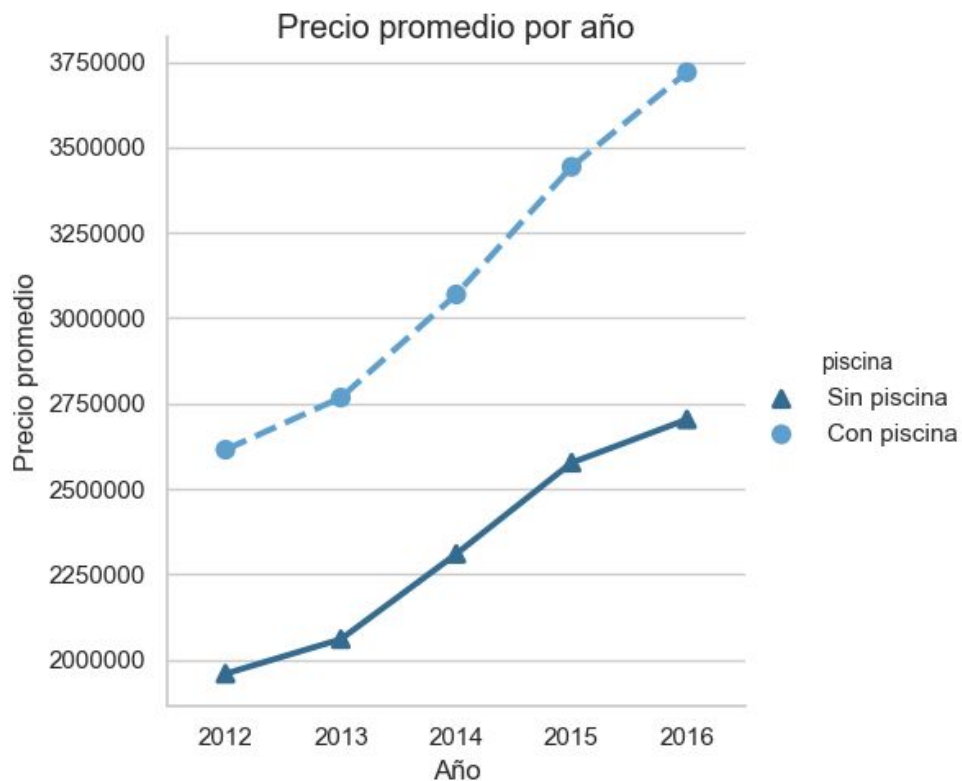
Cantidad de Publicaciones con Salón de Usos Múltiples



Como podemos ver en los tres gráficos anterior, la cantidad de publicaciones en la cual la propiedad cuenta con el amenity analizado no supera ni el 10%. El mayor caso es aquellas propiedades que tienen pileta que llegan a un 8.7% de los publicaciones totales.

Cómo varía el precio promedio de cada propiedad según Amenities



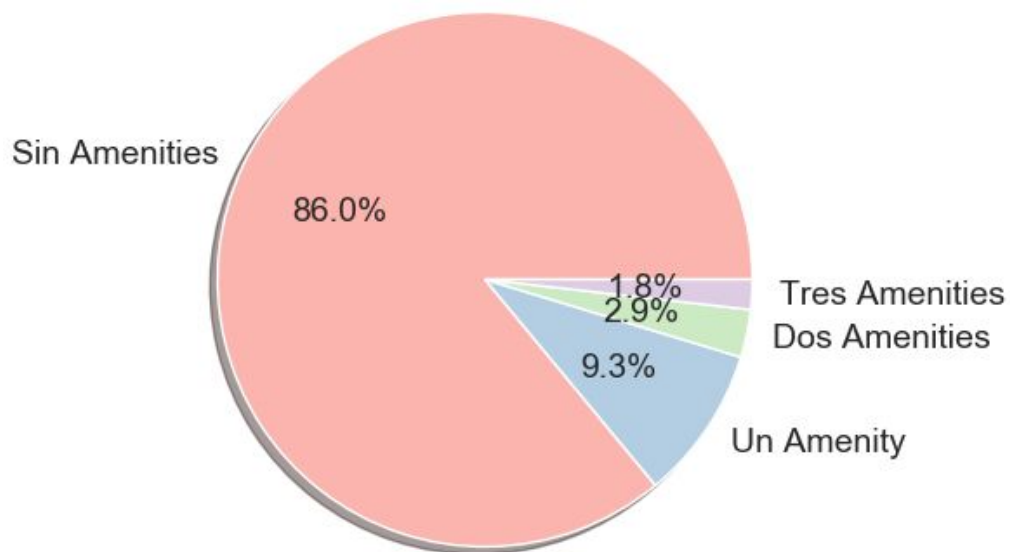


Si ahora lo que nos interesa no es la cantidad de publicaciones sino el precio promedio de los diferentes inmuebles publicados, podemos ver que si ésta cuenta con uno de estos espacios, suele ser más cara.

En todos los casos, podemos ver que el precio promedio, sin importar el año en que se realizó la publicación, es mucho mayor en aquellas propiedades en las cuales se cuenta con por lo menos el amenity en cada caso. Por lo que podemos pensar que no importa el tipo (ya sea piscina, gimnasio o sum), el valor de la propiedad aumenta.

Cantidad de publicaciones contabilizando todos los Amenities

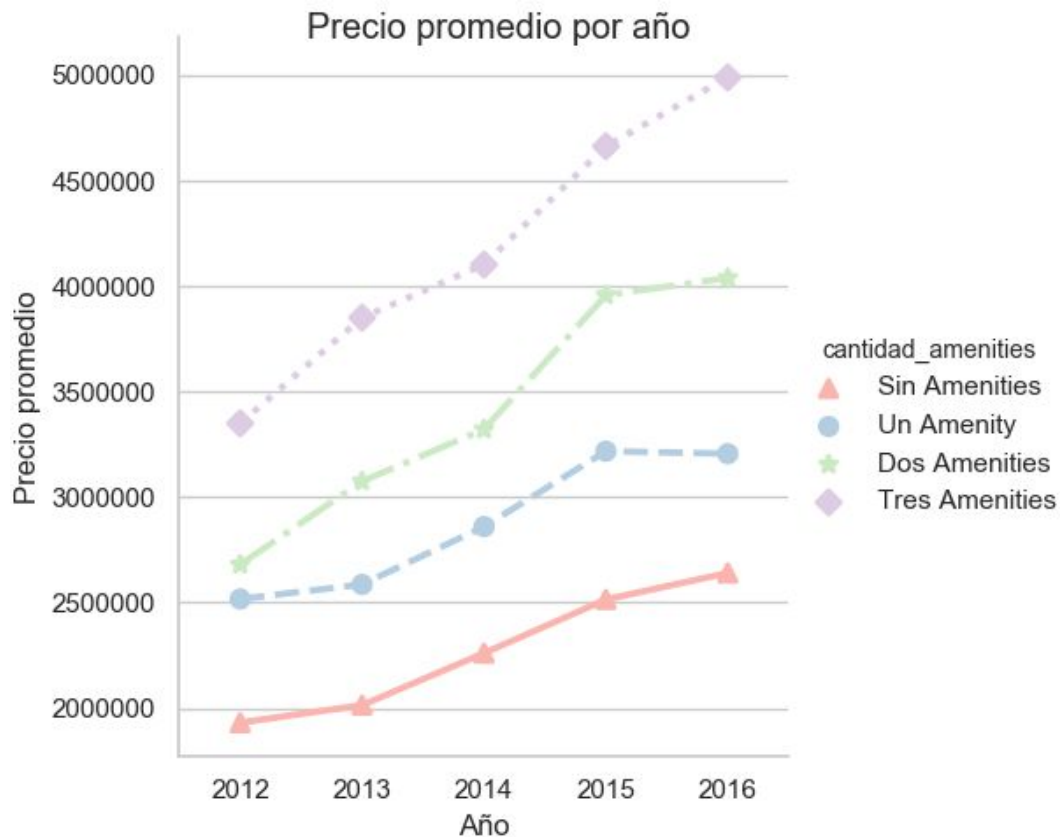
Cantidad de Publicaciones según cantidad de Amenities



A partir de dicha imagen podemos terminar de concluir lo que comenzamos en las secciones anteriores. Se puede ver que la mayoría de las propiedades publicadas no poseen ningún amenity, y el porcentaje de publicaciones disminuye al aumentar la cantidad de amenities.

Siendo muy escasos aquellos inmuebles que poseen todos los servicios presentes en el set de datos.

Cómo varía el precio según la cantidad total de Amenities

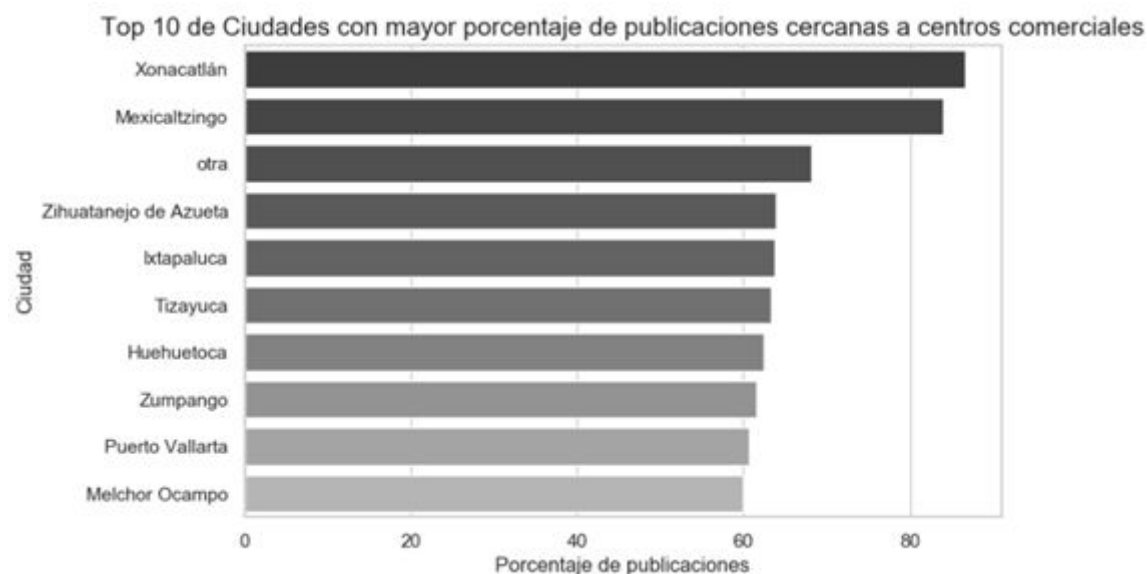
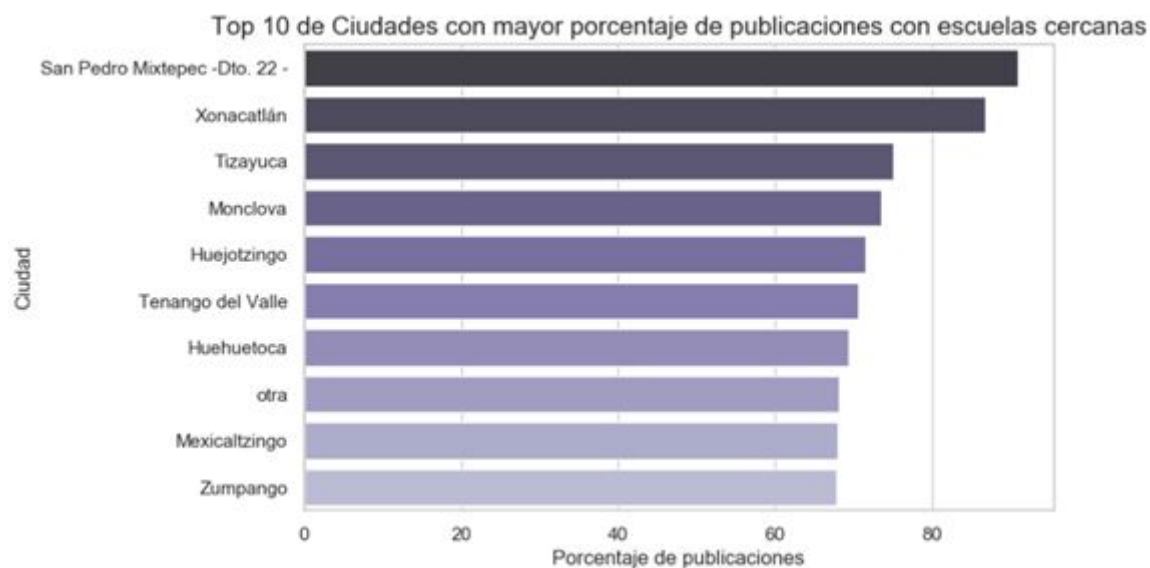


Aquí también podemos ver cómo a medida que aumenta la cantidad total de amenities presentes en la propiedad publicada, a su vez aumenta el precio promedio de la misma de manera considerable, sin importar el año de la publicación.

Quinta parte - Análisis sobre “Otros servicios” en las publicaciones

En esta sección se hará un análisis sobre los “*otros servicios*” que pueden prestar las diferentes propiedades en la publicaciones. En dicho set de datos, con otros servicios, hace referencia a si la propiedad en cuestión tiene escuelas cercanas y/o centros comerciales cercanos.

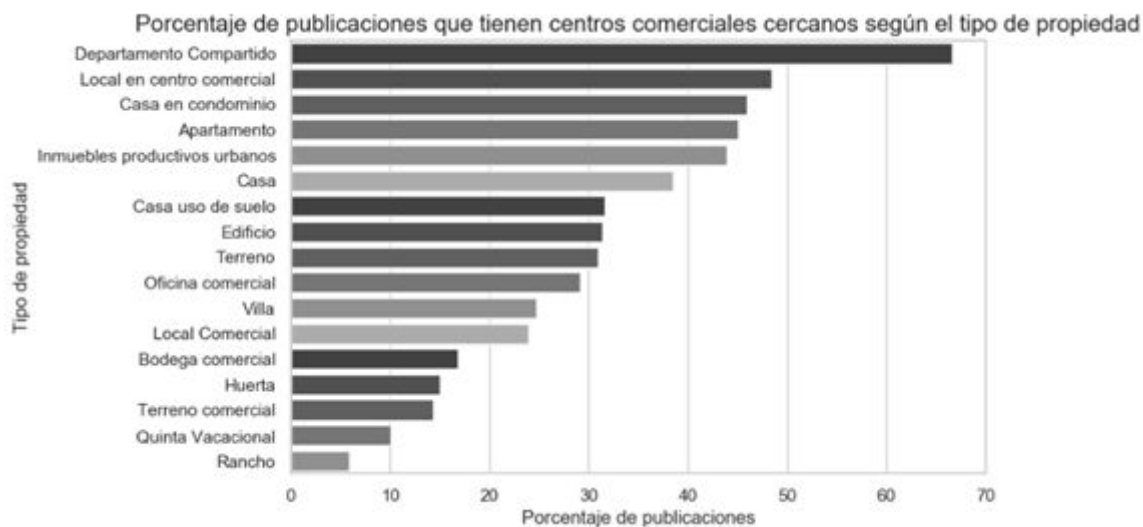
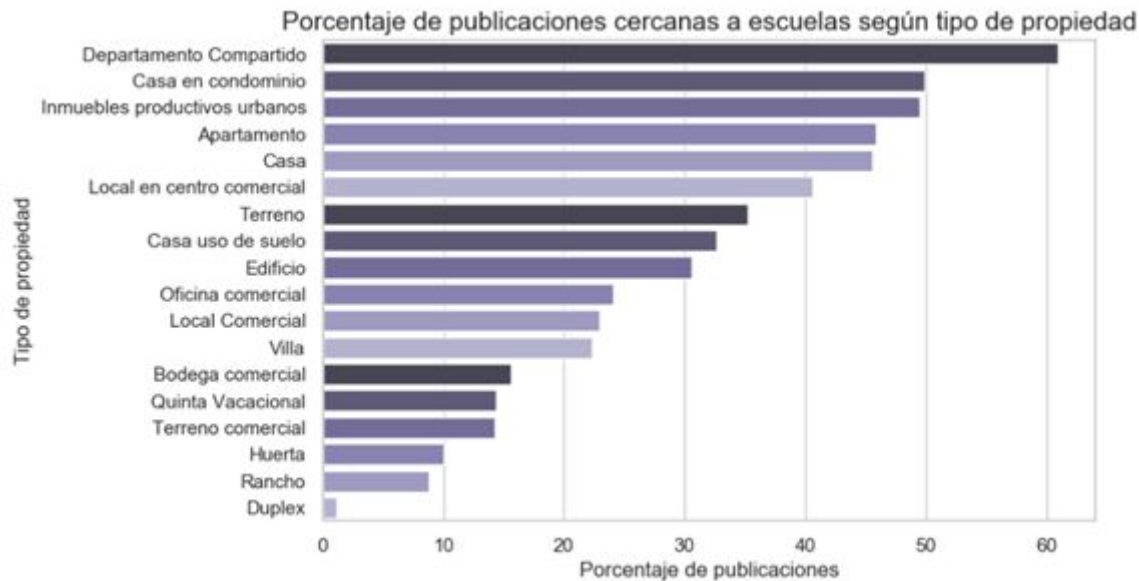
Cuales son las ciudades con mayor algún servicio



La ciudad más destacada, en este caso, se trata de **Xonacatlán**³ con más del 80% de publicaciones tanto para centros comerciales cercanos, como para escuelas. Por otro lado, también en ambas gráficas, figura como ciudad una con nombre *otra* por lo que nos parece que esas publicaciones no tendrían que ser válidas ya que la ciudad es uno de los atributos más importantes de una propiedad.

³ Es uno de los 125 municipios del Estado de México.

Cuales son los tipos de propiedad que se encuentran cercanos a estas *edificaciones*



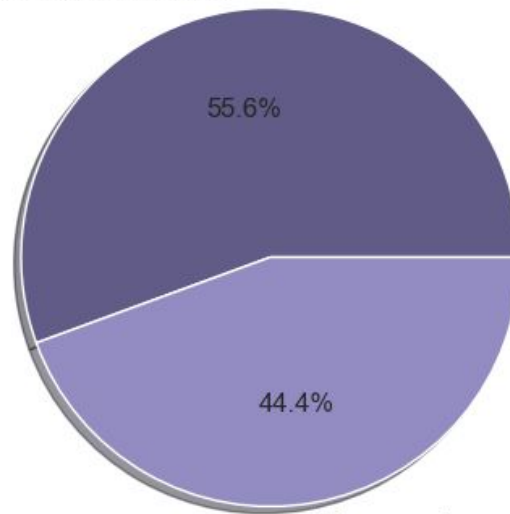
En primer lugar, a partir de la segunda imagen, nos llamó la atención la poca cantidad de publicaciones del tipo **Local en centro comercial** que figuran cerca de centros comerciales. Esto no sabemos si se debe a un error en los datos o por el hecho de que no se cuente el centro comercial al que pertenece el local.

Por otro lado, vemos que tanto el tipo de **Departamento compartido** y **Casa en condominio** figuran en los primeros puestos en tanto a su cercanía con centros comerciales como a escuelas.

Qué porcentaje de publicaciones cuentan cercanas a estas *edificaciones*

Cantidad de Publicaciones según escuelas cercanas

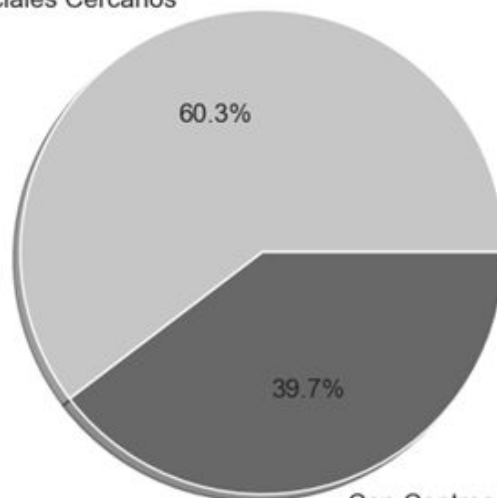
No hay escuelas cercanas



Hay escuelas cercanas

Porcentaje de publicaciones según cercanía con centros comerciales

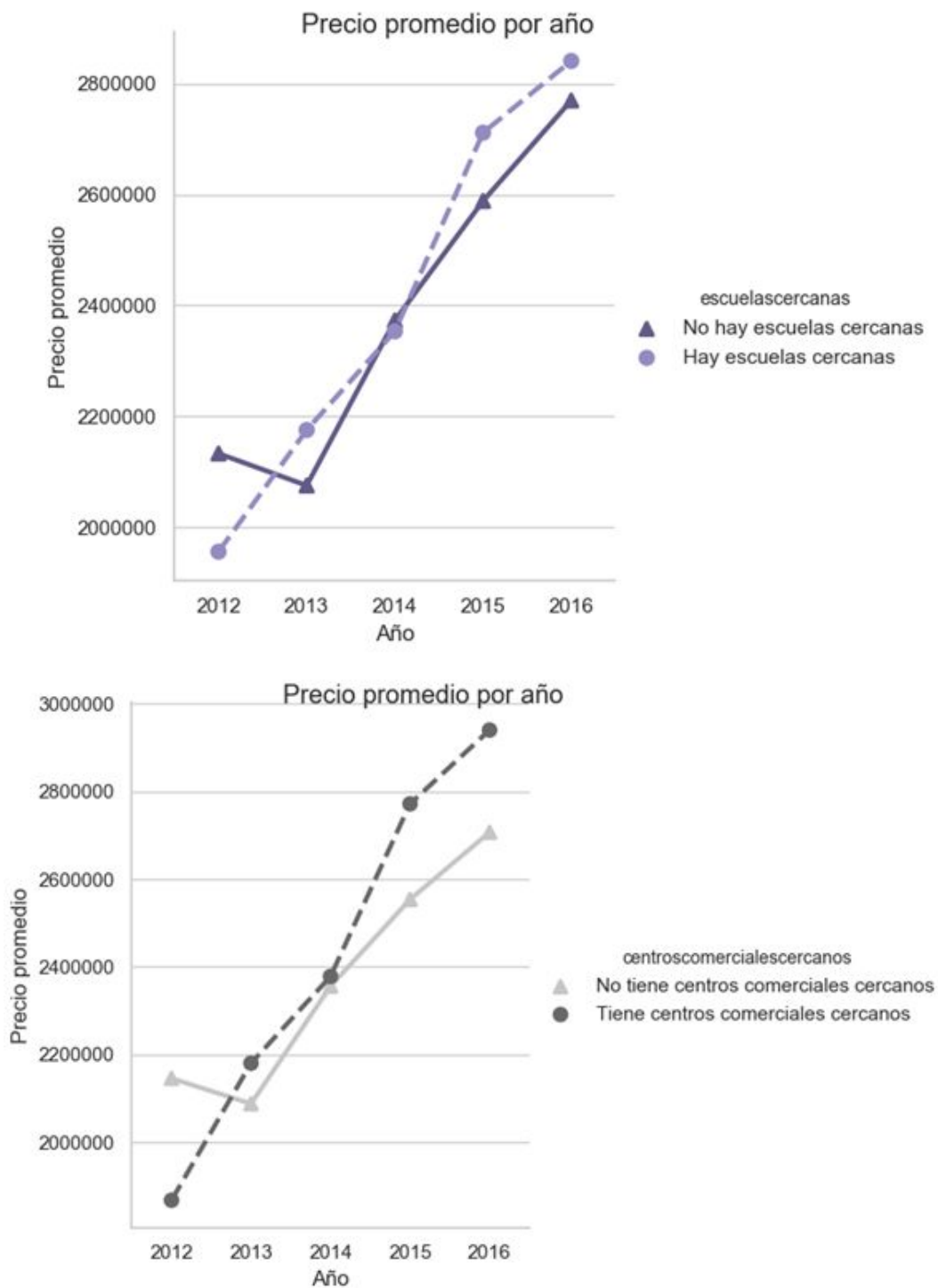
Sin Centros Comerciales Cercanos



Con Centros Comerciales Cercanos

Gracias a estas imágenes podemos ver que bastante parejo la cantidad de publicaciones con escuelas cerca con las cuales no; siendo ya un poco menor la cantidad de propiedades con centros comerciales cercanos.

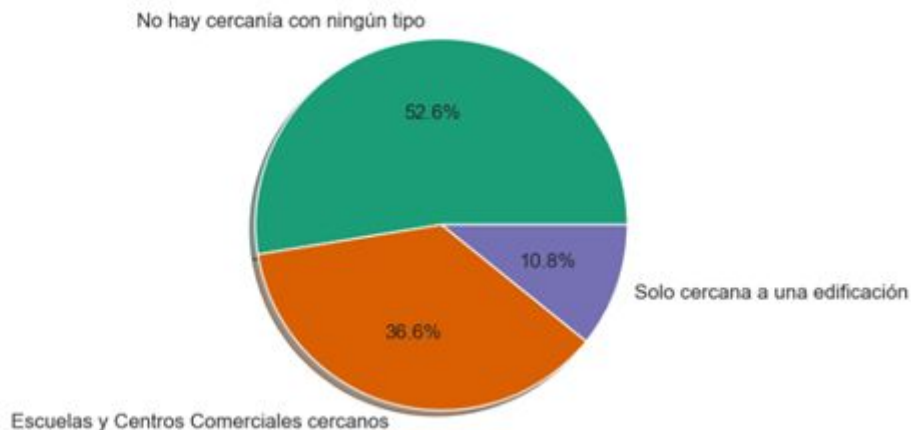
Cómo varía el precio según dichos servicios



Analizando los gráficos llegamos a la conclusión de que no hay una correlación entre el valor de la propiedad respecto a la cercanía a escuelas, la diferencia entre los promedios no es considerable. En cuanto a los centros comerciales podemos ver, a partir del 2015, empezó a tener mayor consideración que se encuentre cerca de éstos para la cotización del inmueble.

Cómo varía el precio y la cantidad de publicaciones según la cantidad de *servicios prestados*

Cantidad de publicaciones según cantidad de edificaciones cerca (escuelas y/o centros comerciales)



A partir del gráfico de torta podemos confirmar lo mostrado en la matriz de correlación presentada en las secciones anteriores. Aquí vemos que la mayoría no cuenta con escuelas o centros comerciales cercanos (52.6%), pero en segundo lugar, con un 36.6%, las propiedades se encuentran cerca de ambas.

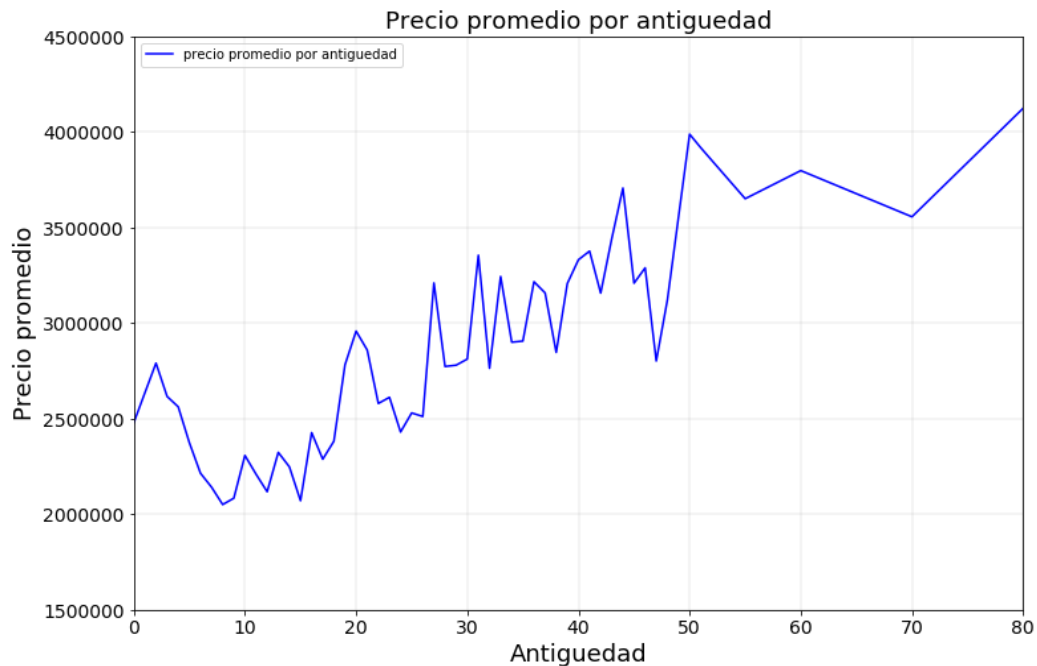
Es decir, es más probable que un inmueble presente en el set de datos se encuentre cerca de tanto de escuelas como de centros comerciales que de uno solo de estos.



El anterior gráfico nos muestra que no hay una correlación entre la cantidad de *servicios* con el valor promedio de la propiedad, por ende, no podemos afirmar que el precio de la publicación dependa de si se encuentra cercana a escuelas y/o a centros comerciales.

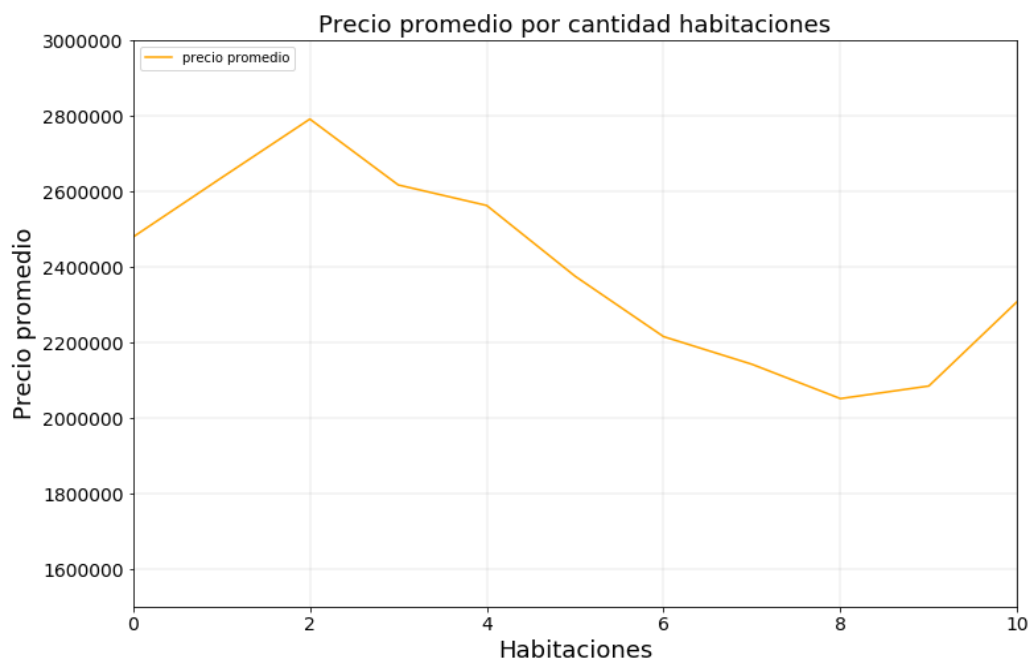
Sexta Parte - Análisis de Precios

¿Cual es la relación del precio con la antigüedad de todas las propiedades?



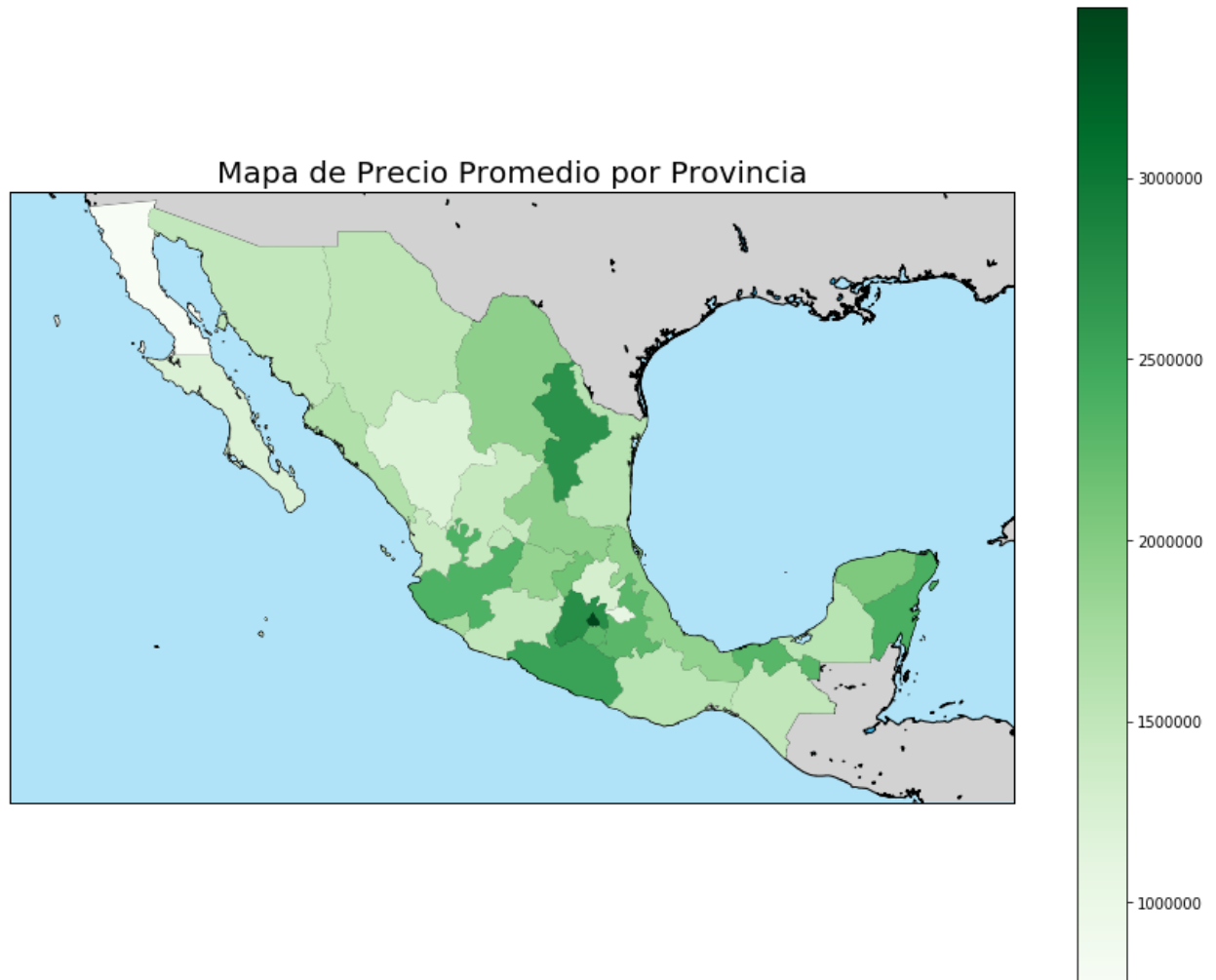
En el gráfico anterior podemos observar una distribución pareja del precio entre las propiedades con antigüedad de 0 a 25 años, luego el promedio tiende a subir para las propiedades con un antigüedad **mayor a 25 años**. Al involucrarse todas las de propiedades, podemos obtener esta estadística general pero no podemos asegurar que se cumpla esta tendencia en todos los tipos de propiedades.

¿Cual es la relación del precio con la cantidad de habitaciones?



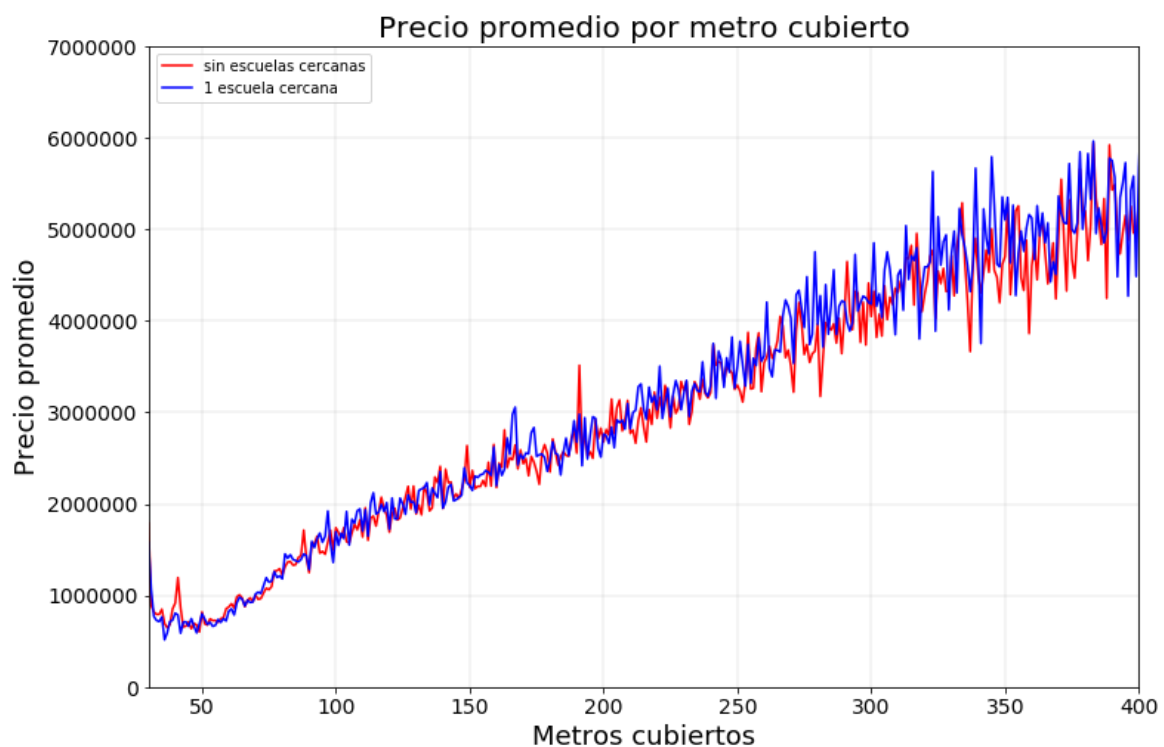
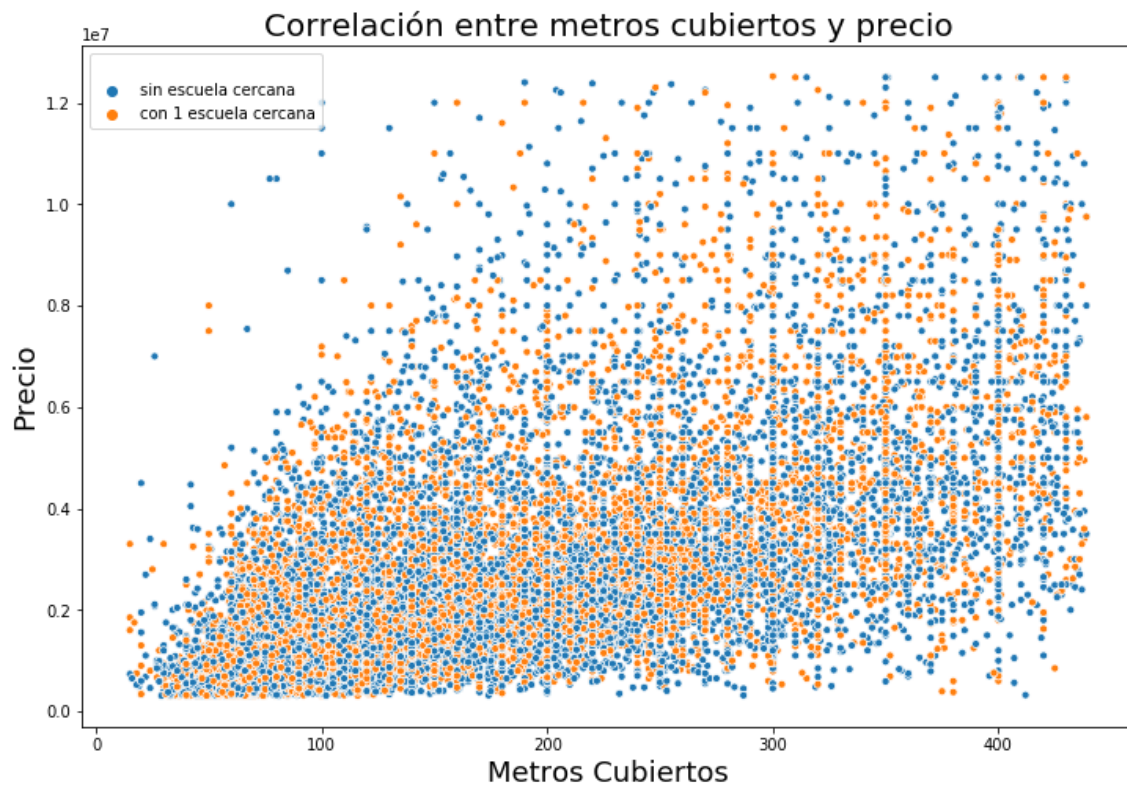
Acorde a lo analizado no se puede determinar alguna relación directa en la cantidad de habitaciones y el precio de las propiedades, pero podemos concluir que las propiedades con 2 habitaciones tienen el promedio más alto con un aproximado de 2.800.000 pesos mexicanos.

¿Cual es el precio promedio por provincia?

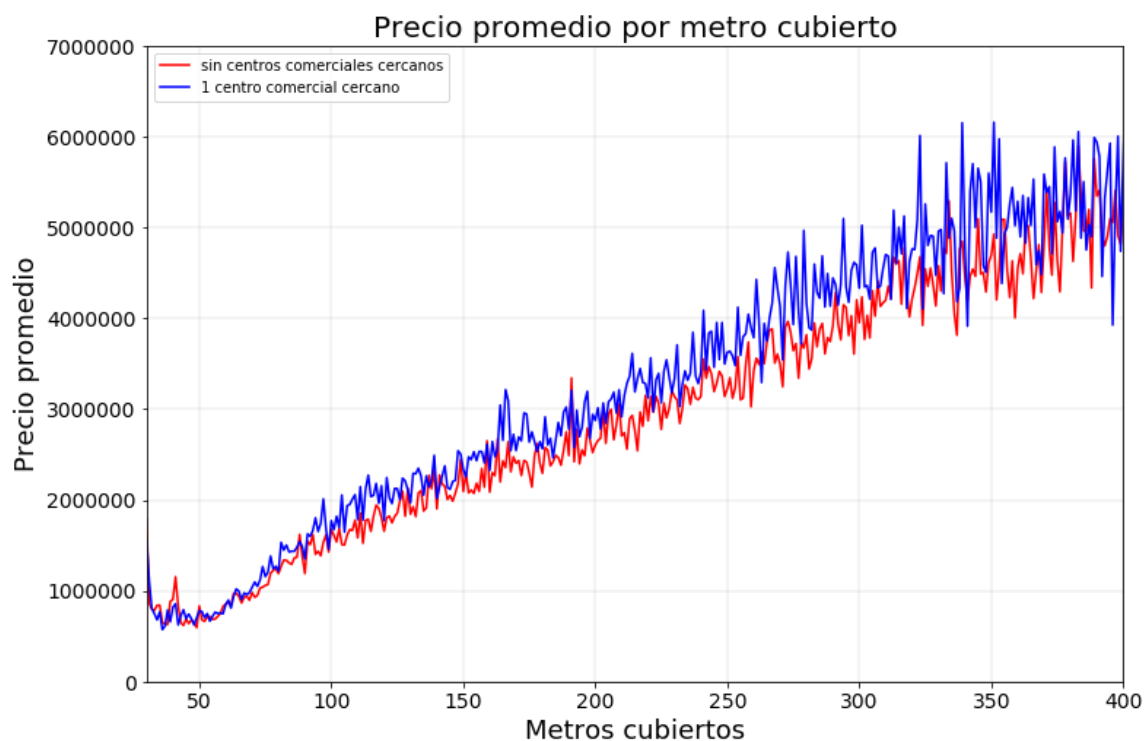
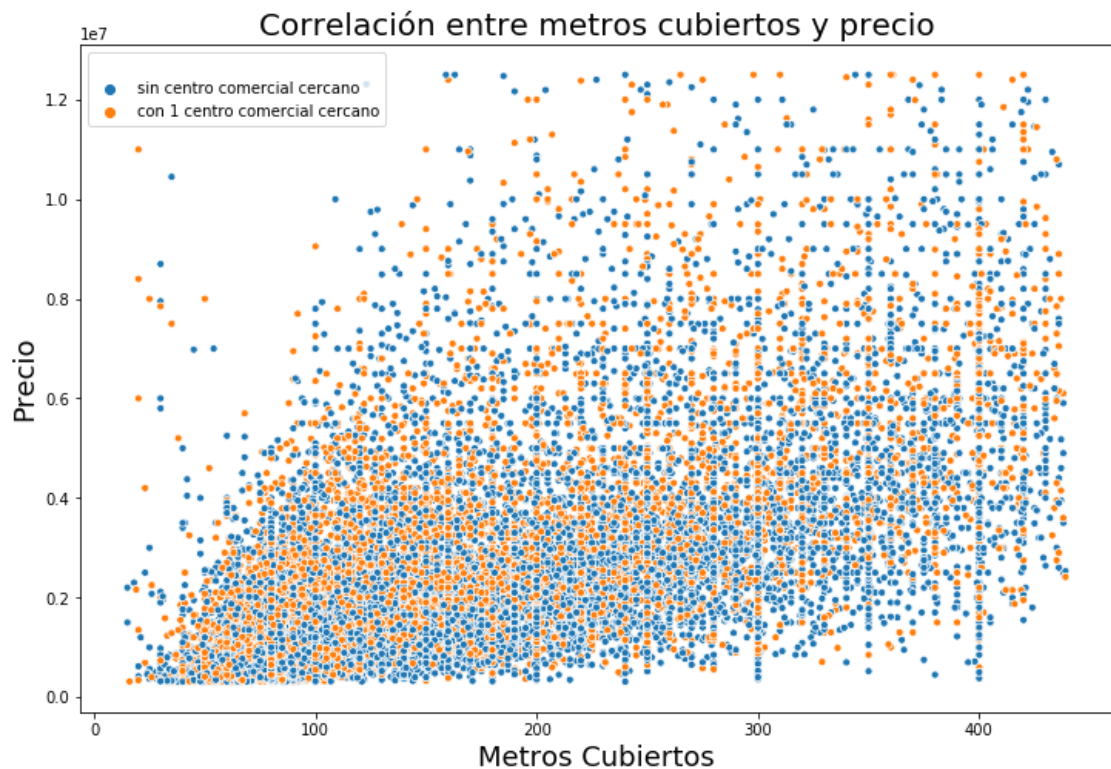


Se obtuvo que el **DF** tienen las propiedades con precio promedio más alto con un valor aproximado de \$3.500.000 seguido por las provincias de Edo. de México, Nuevo León, Guerrero y Quintana Roo. Por otro lado, el promedio de precio más bajo se encuentra en **Baja California Norte** con un promedio de \$ 770.000.

¿Cuál es la relación entre metros cubiertos y precios dependiendo de la cercanía con escuelas?

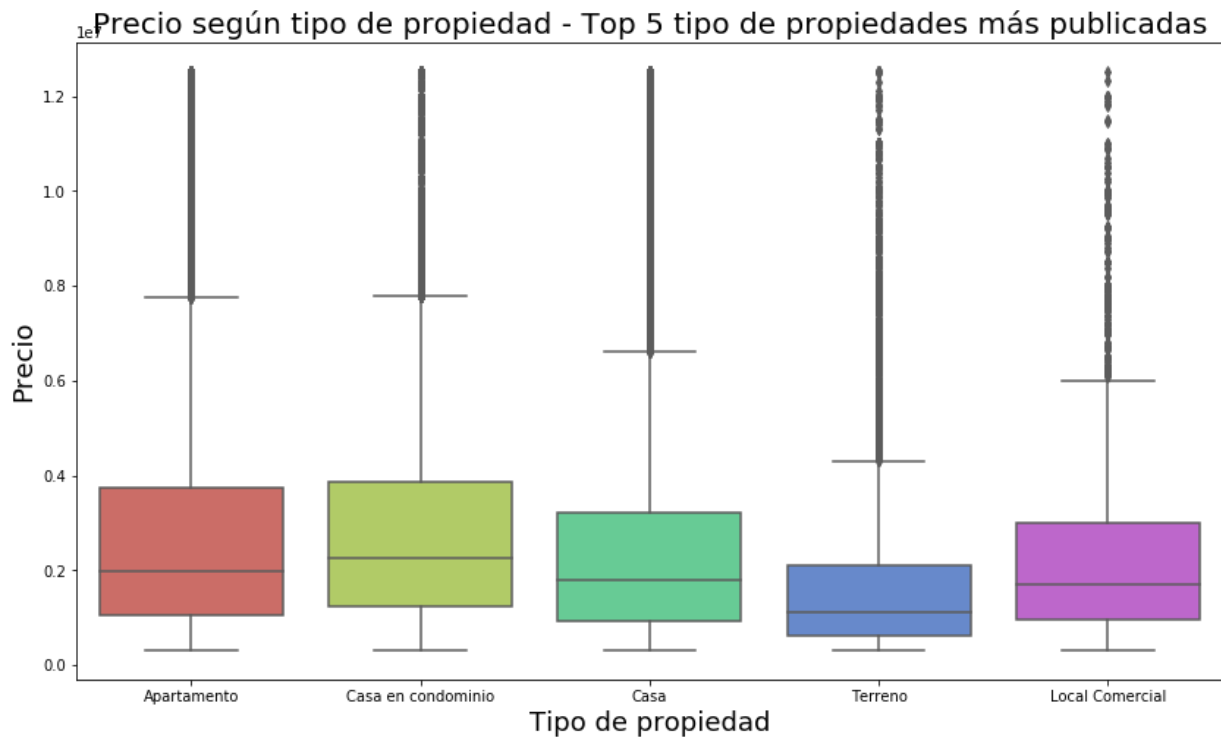


¿Cuál es la relación entre metros cubiertos y precios dependiendo de la cercanía con centros comerciales?



A raíz de estos de las últimas gráficas se ve una mínima diferencia de precio a favor de las propiedad con 1 escuela cercana, pero que no es muy significativa. En el caso de los centros comerciales se puede ver un aumento significativo del precio promedio para las propiedades con 1 centro comercial cercano.

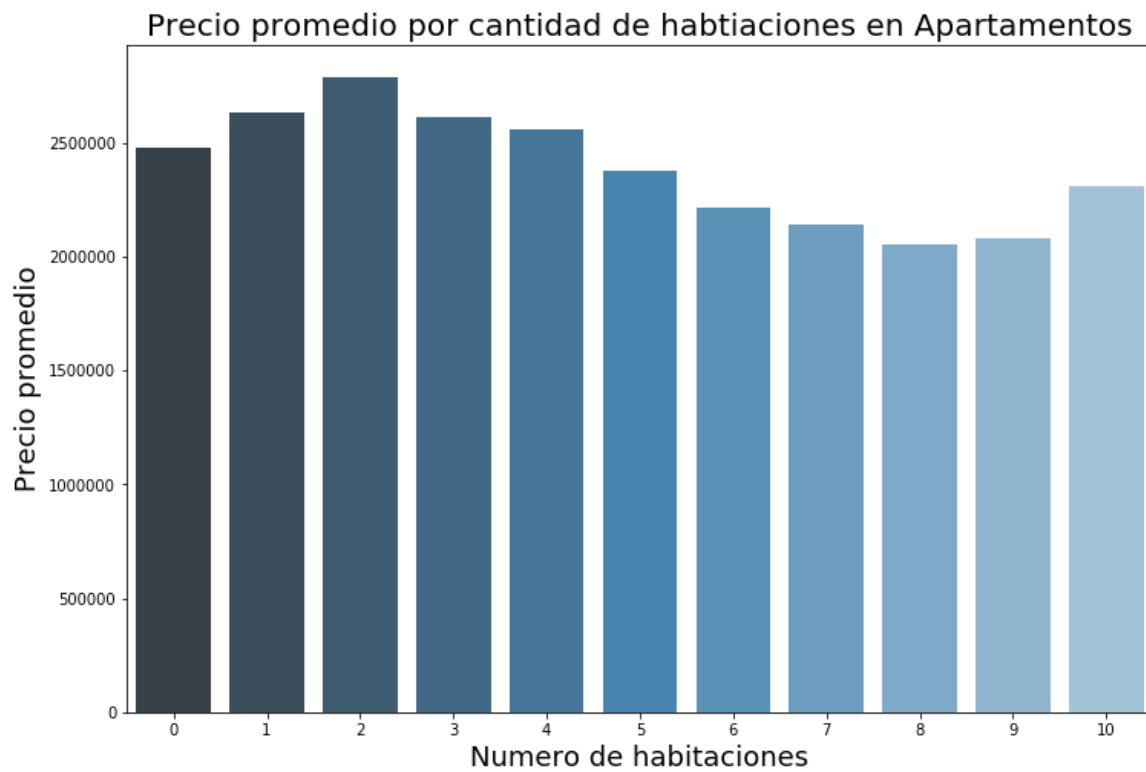
Precio promedio de los 5 tipos de propiedad más publicados



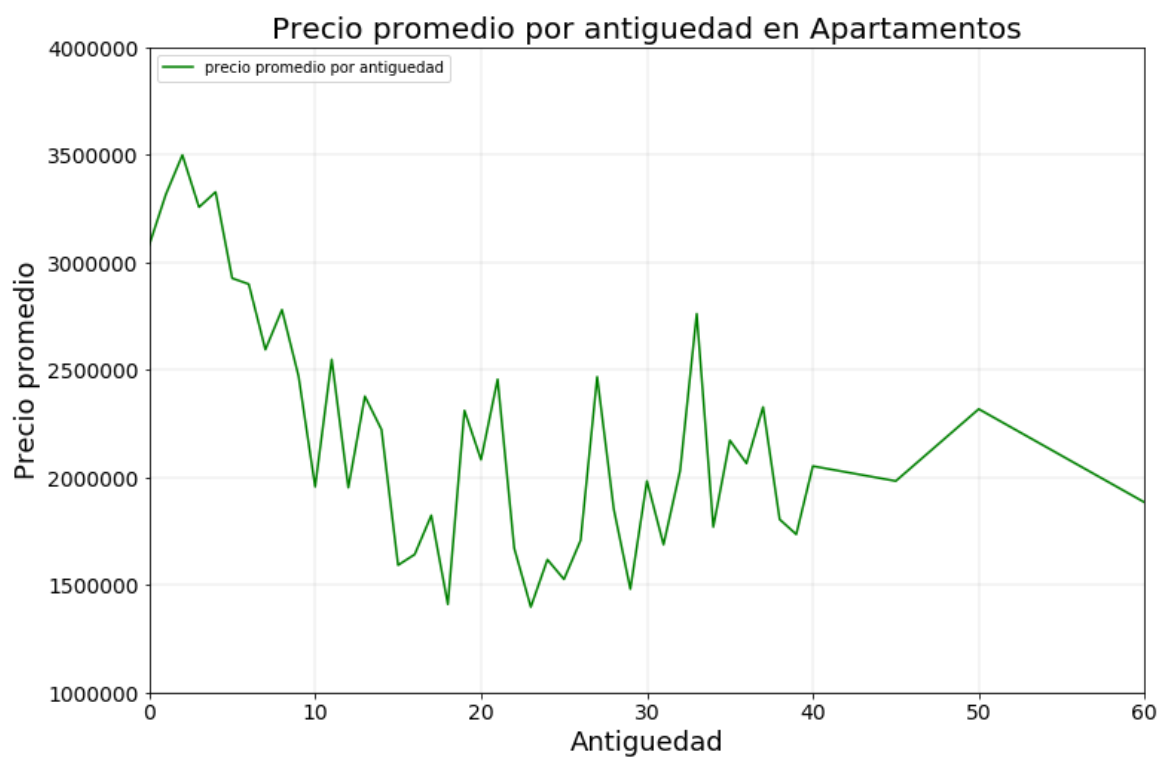
Podemos observar que tanto los apartamentos como las casas en condominio tienen las medias de precio más altas, al igual que las cotas del percentil 75. Los terrenos, en cambio, tienen la media y las cotas del percentil 75 y 25 más bajas. Se puede afirmar que el 50% de las publicaciones de los terrenos se concentra en un rango de precios muy bajo comparado con los otros tipos de propiedad.

Análisis de precios en Apartamentos

Precio promedio por cantidad de habitaciones



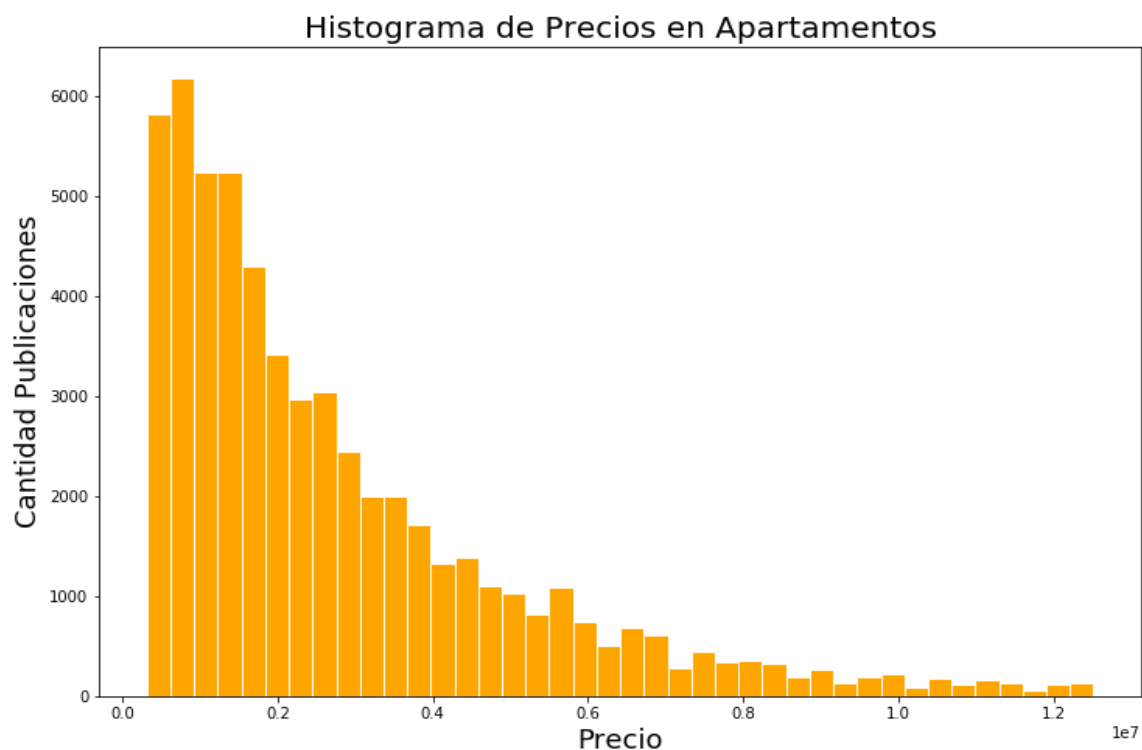
Precio promedio por antigüedad



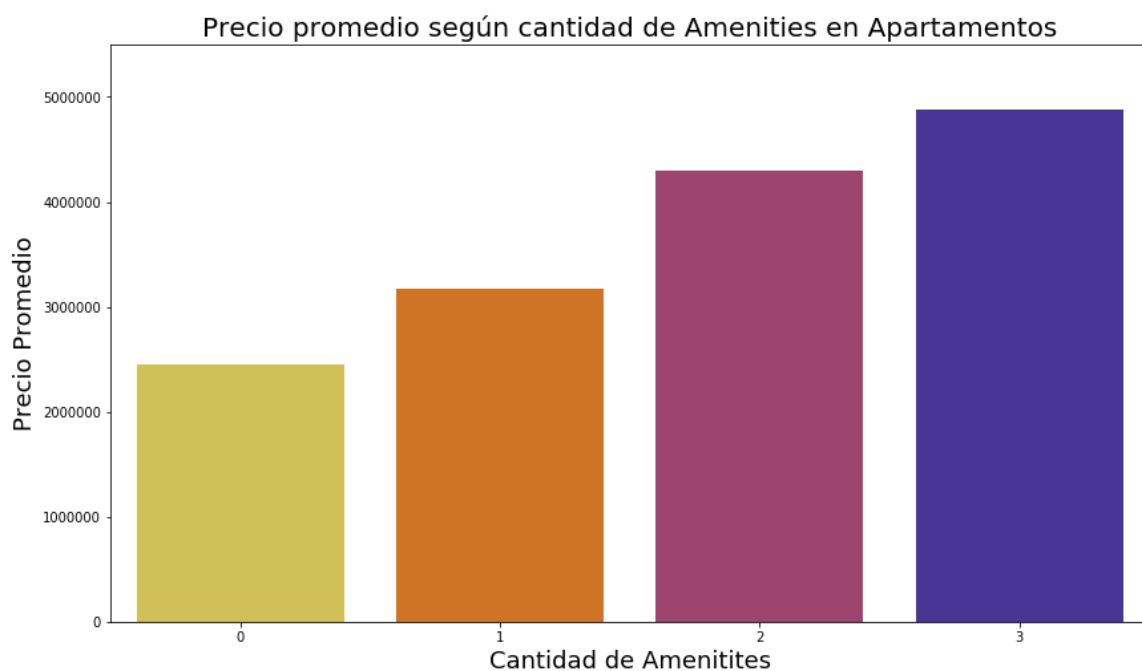
Dada la gráfica anterior, podemos ver que los precios de los apartamentos con poca antigüedad tienen los precios más caros, aunque luego de los 10 años de antigüedad se mantienen una franja de precios constante.

Esto se contrapone con lo visto en gráfico de la página 31 en donde analizamos a todos los tipo de propiedades y se veía una tendencia inversa.

Histograma de precios



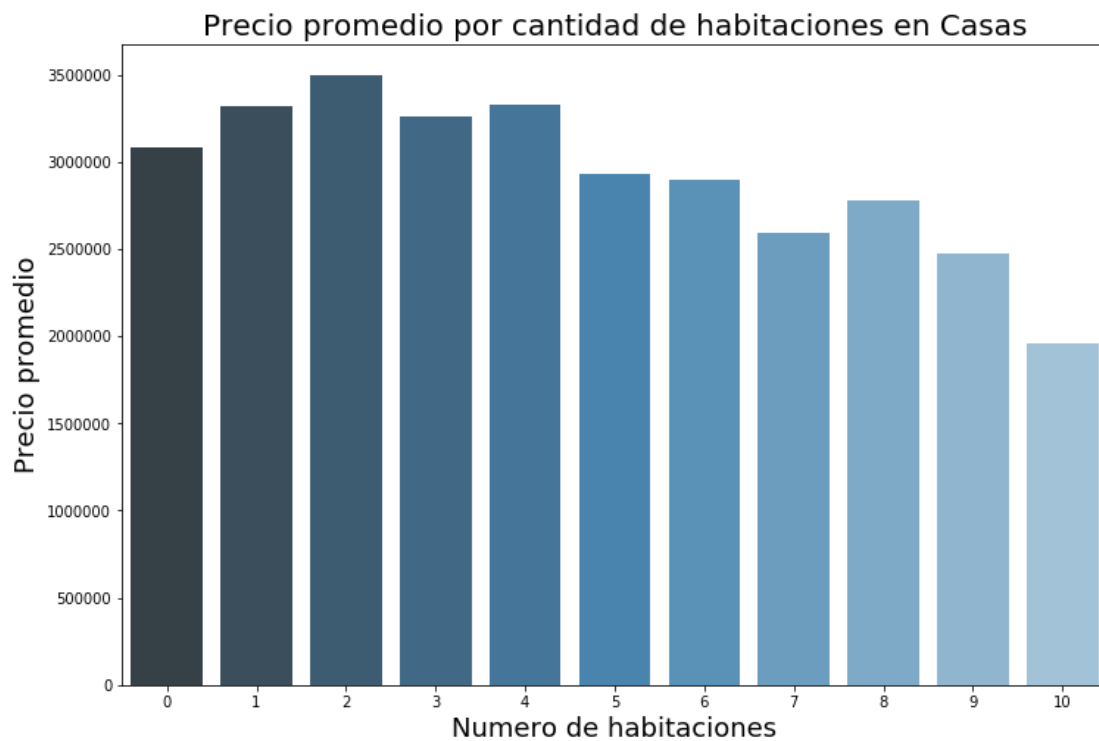
Precio promedio dependiendo de cantidad de amenities



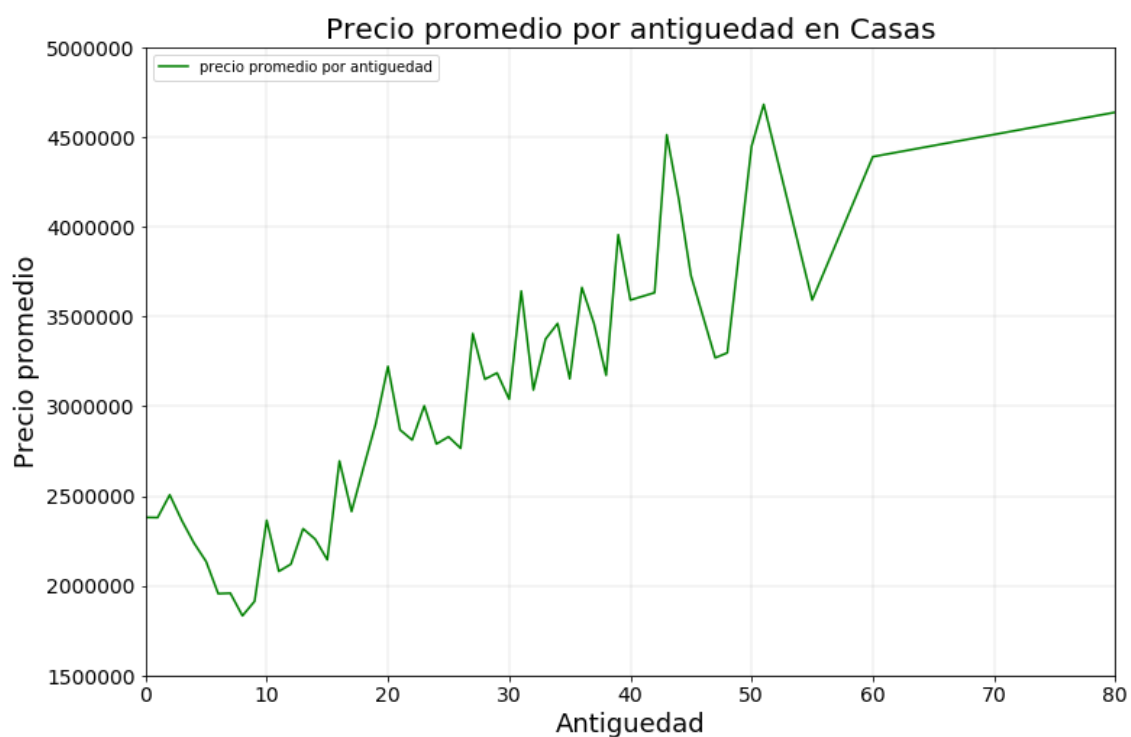
Se puede apreciar un aumento considerable del precio de las propiedades de tipo apartamento de acuerdo a la cantidad de amenities.

Análisis de precios en Casas

Precio promedio por cantidad de habitaciones

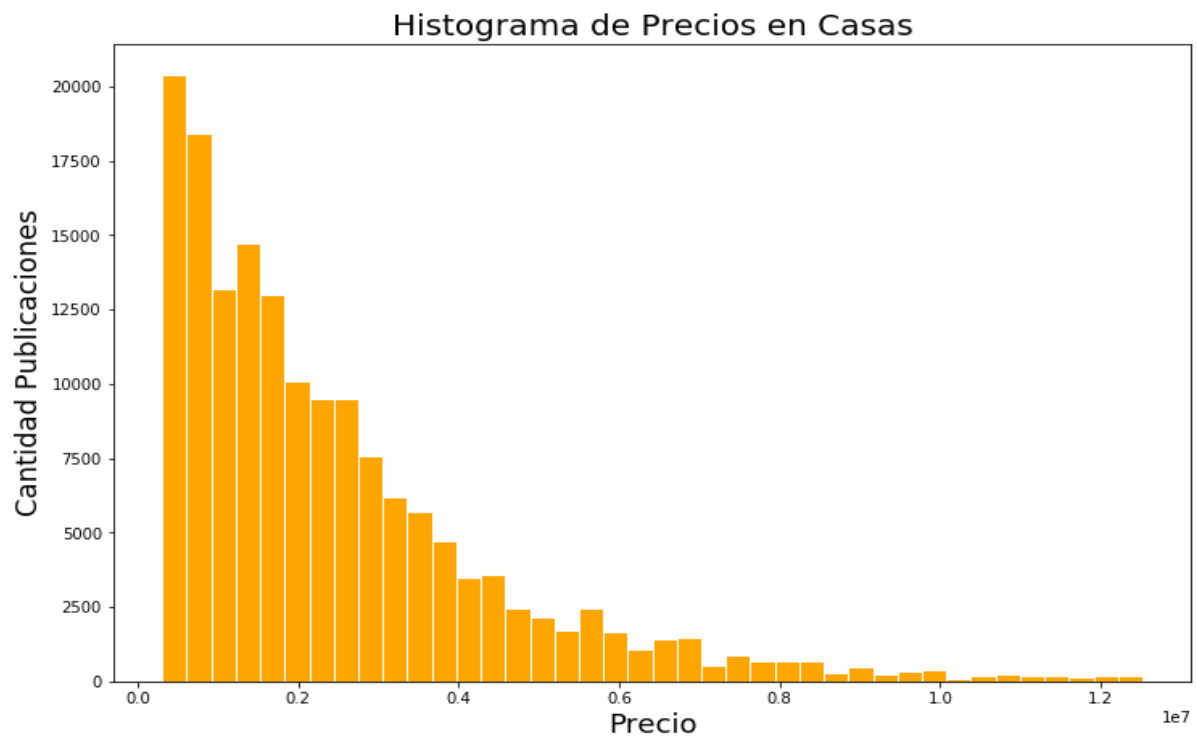


Precio promedio por antigüedad

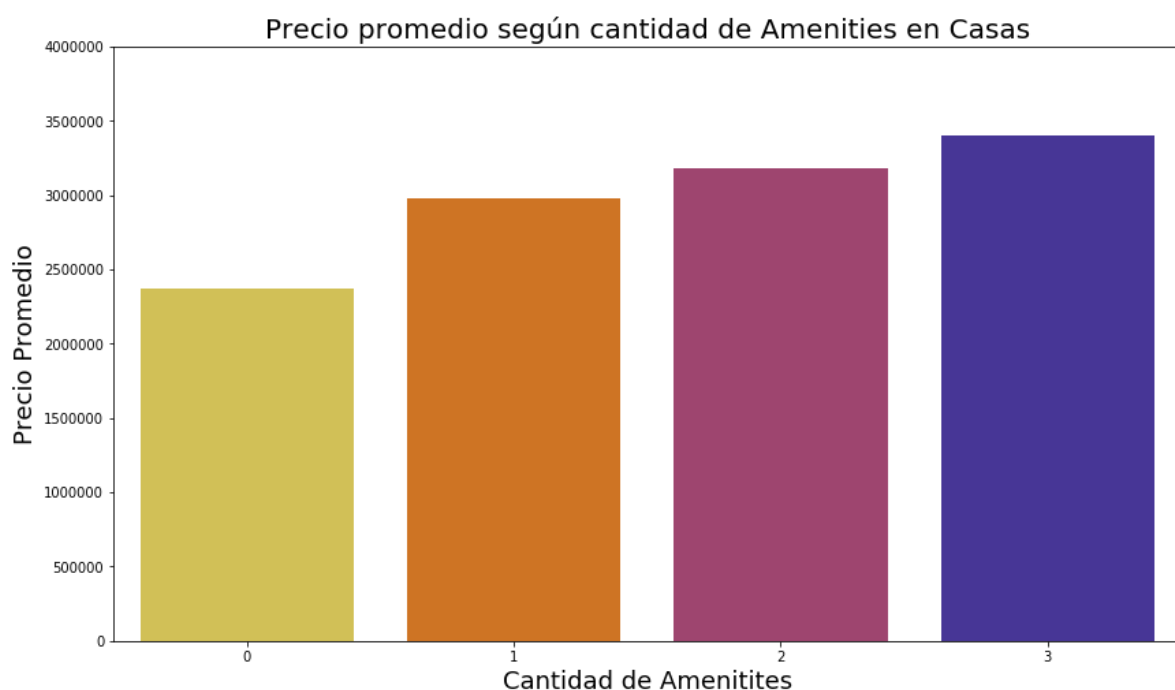


En contrapunto con los apartamentos, el promedio de precio más alto se da en las casas más antiguas y se condice mejor con la gráfica en donde se evaluaban todos los tipos de propiedad.

Histograma de precios



Precio promedio dependiendo de cantidad de amenities



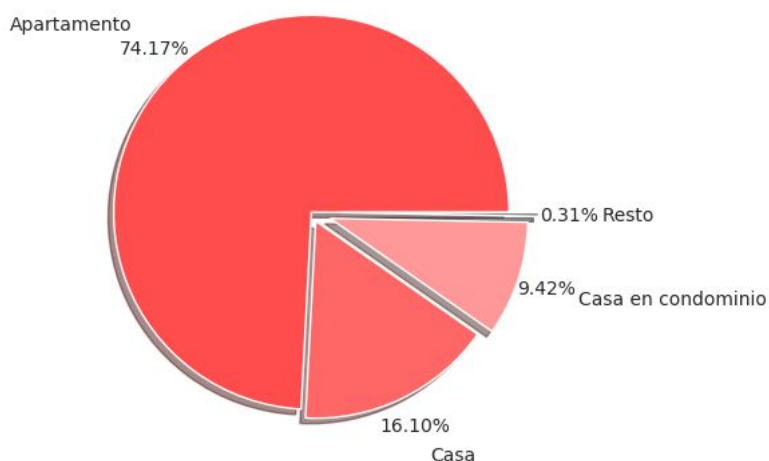
Analizando este último cuadro se puede ver también un aumento del precio de las casas a medida que aumenta la cantidad de amenities, pero este aumento no está tan marcado como en el caso de los apartamentos.

Séptima Parte - Análisis de propiedades “con todo” y “sin nada”

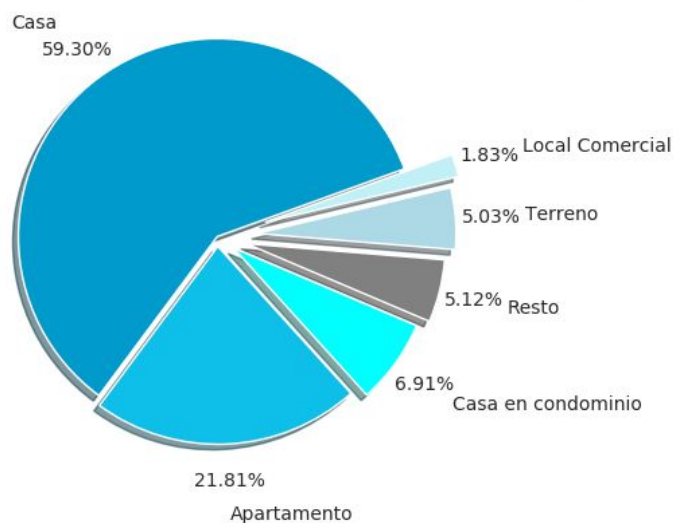
El siguiente análisis se basará en propiedades que tengan gimnasio, piscina, SUM y estén cerca de escuelas y centros comerciales al mismo tiempo (propiedades con todo). Por el contrario una propiedad sin nada sera lo puesto a una que lo tenga todo. Cuando se quiera referir a las propiedades sin condiciones, se hablara de “propiedades en general”.

¿Cual es la distribución en cantidad de publicaciones del tipo de propiedades, en propiedades con todo y nada?

Los tipos de propiedades (con todo) con mas publicaciones



Los tipos de propiedades (sin nada) con mas publicaciones



De la figura roja (propiedades con todo) podemos ver que la mayor parte de las propiedades está formado por apartamentos con una gran diferencia, seguido de casas. Por el contrario la figura azul (propiedades sin nada) muestra una dominancia en cantidad de casas sobre apartamentos, pero se puede ver que hay mas variedad en cuanto a tipo de propiedades. Estas figuras son muy importantes para consecuencias que se verán más adelante.

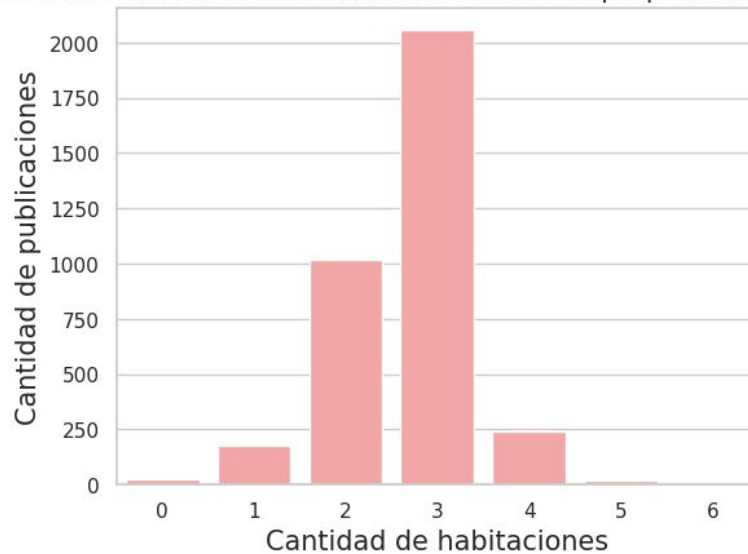
¿Cual es la cantidad promedio de habitaciones por tipo de propiedad en general?



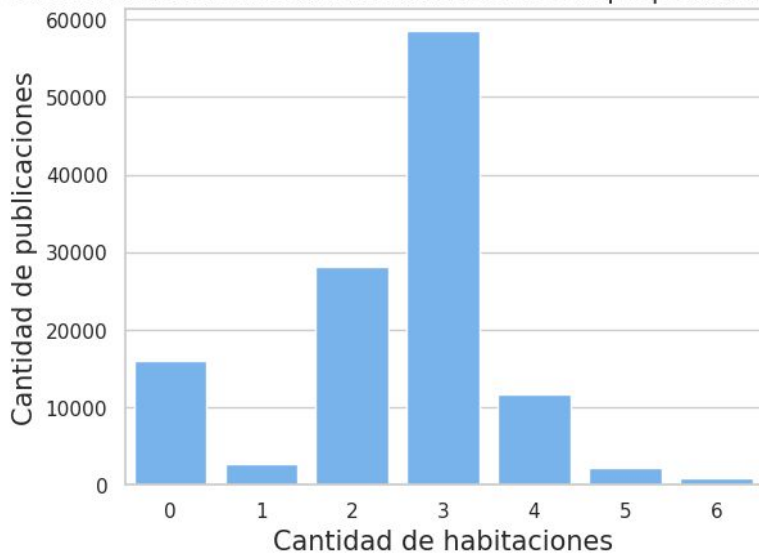
Podemos ver que hay un pico de cantidad de habitaciones en hospedajes, que tiene sentido ya que es esperable que los hospedajes tengan más de 7 habitaciones. Otra observación es que los tipos de propiedades que menos habitaciones tienen suelen ser locales y terrenos.

¿Cual es la frecuencia de la cantidad de habitaciones en propiedades con todo y nada?

Frecuencia de la cantidad de habitaciones en propiedades con todo



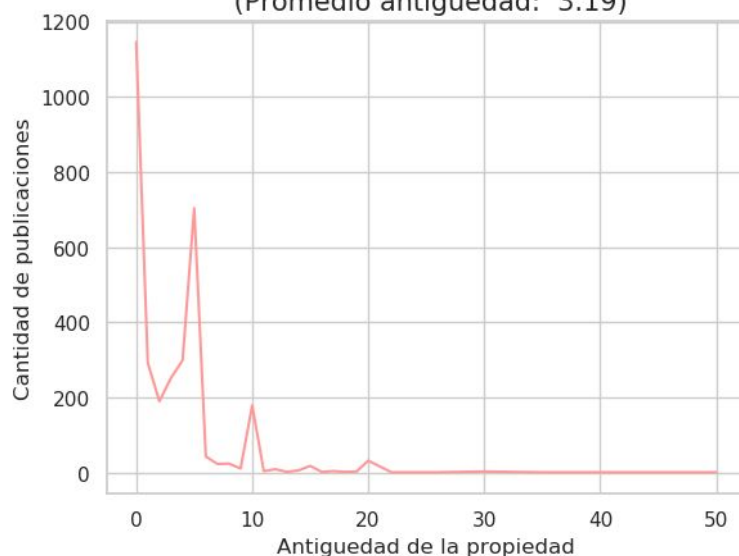
Frecuencia de la cantidad de habitaciones en propiedades sin nada



Se puede ver como ambas figuras son muy similares en los valores que están entre el 1 y el 4. También puede apreciarse como tenemos una gran cantidad de propiedades que tienen 0 habitaciones en el gráfico azul (propiedades sin nada). Esto último podríamos pensar que es consecuencia de que casi un 7% de las propiedades con todo son Terrenos y locales comerciales (como mostramos anteriormente), y que como vimos recientemente los Terrenos y los locales son las propiedades que más tienden a tener menos habitaciones.

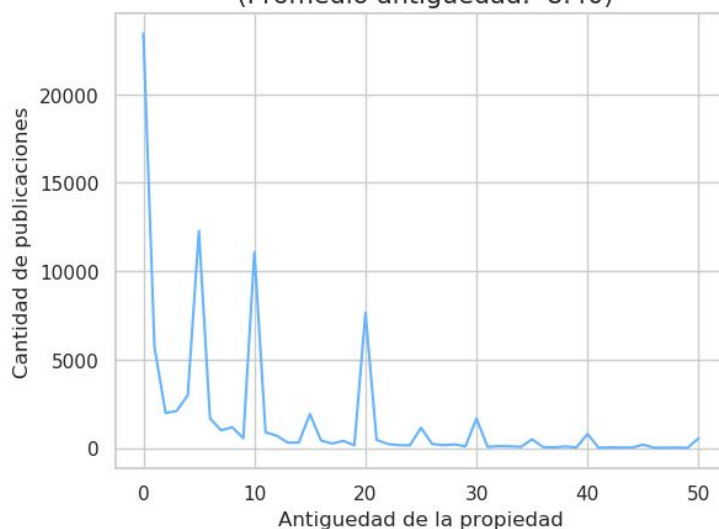
¿Cómo varía la cantidad de publicaciones de propiedades con nada y sin todo con respecto a la antigüedad de las mismas?

Cantidad de publicaciones de propiedades con todo segun antigüedad
(Promedio antigüedad: 3.19)



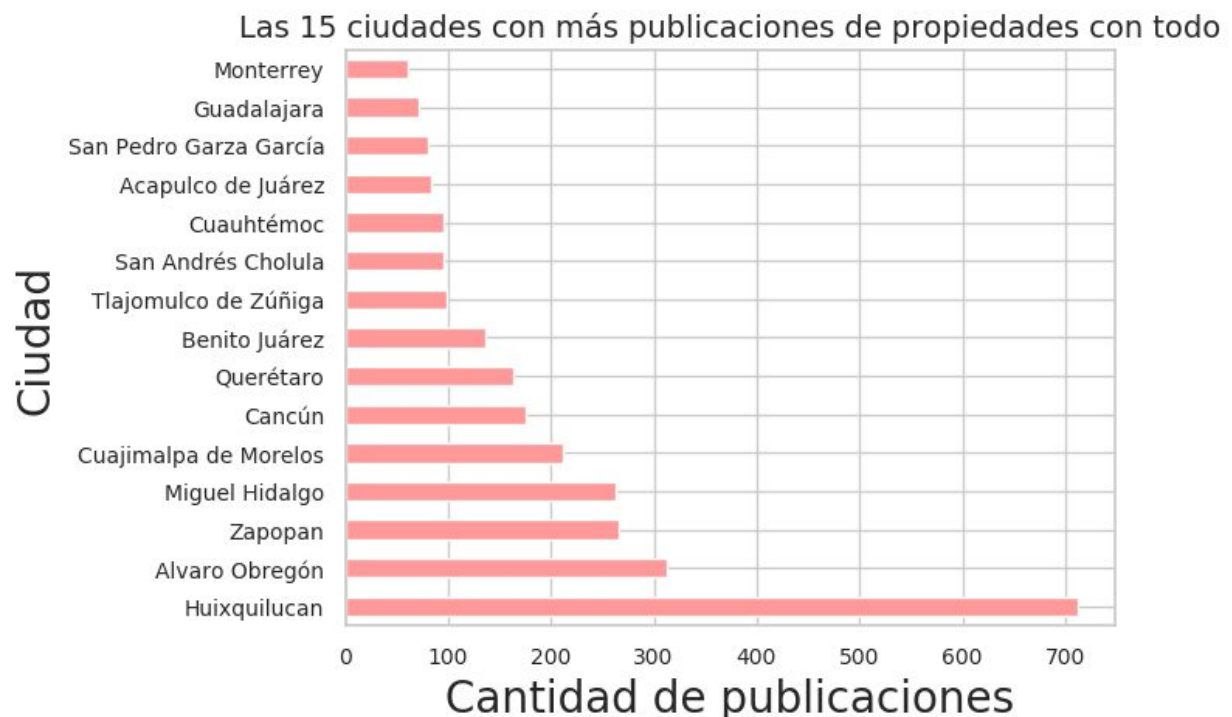
Podemos observar que los valores más altos en cantidad están en valores de antigüedad entre 0 y 5, también se puede apreciar un pequeño pico en valores cercanos a 10. Luego de los 10 años de antigüedad se ve que hay muy pocas cantidades de propiedades. El promedio es de 3.19, que es poco.

Cantidad de publicaciones de propiedades sin nada segun antigüedad
(Promedio antigüedad: 8.40)



Podemos ver que esta figura tiene similitudes y diferencias con la figura anterior de propiedades con todo. Los valores entre los 0 y 5 años de antigüedad son muy parecidos. Una diferencia con la otra figura es que se ven varios picos en años como 10 y 20 seguido de otros pequeños picos. El promedio es 8.4, mayor al de propiedades con todo.

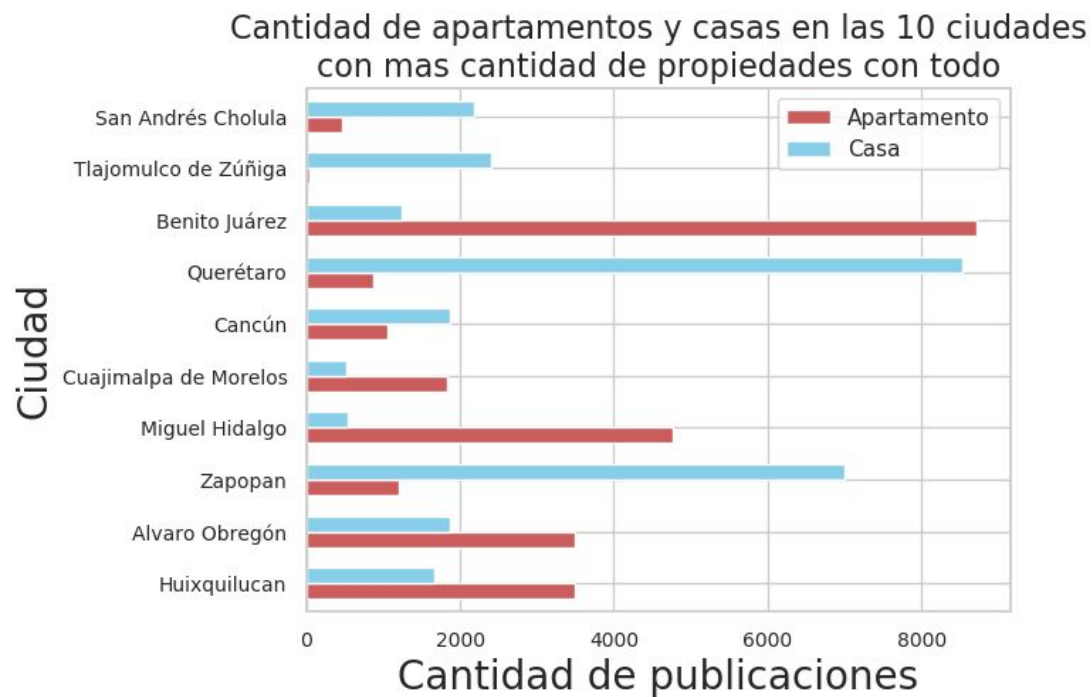
¿Cuales son las ciudades con mayor cantidad de propiedades con todo?



Hay 3 observaciones importantes, la primera es que Huixquilucan tiene una gran diferencia en comparación a las demás ciudades. La segunda es que se puede ver la diferencia de valores entre Alvaro Obregón y Benito Juárez. La tercera es que Tlajomulco de Zúñiga hasta monterrey los valores empiezan a ser muy parecidos y es más difícil apreciar la diferencia. Debido a estas observaciones, analizaremos el caso particular de las ciudades entre Huixquilucan y Benito Juárez (el top 10).

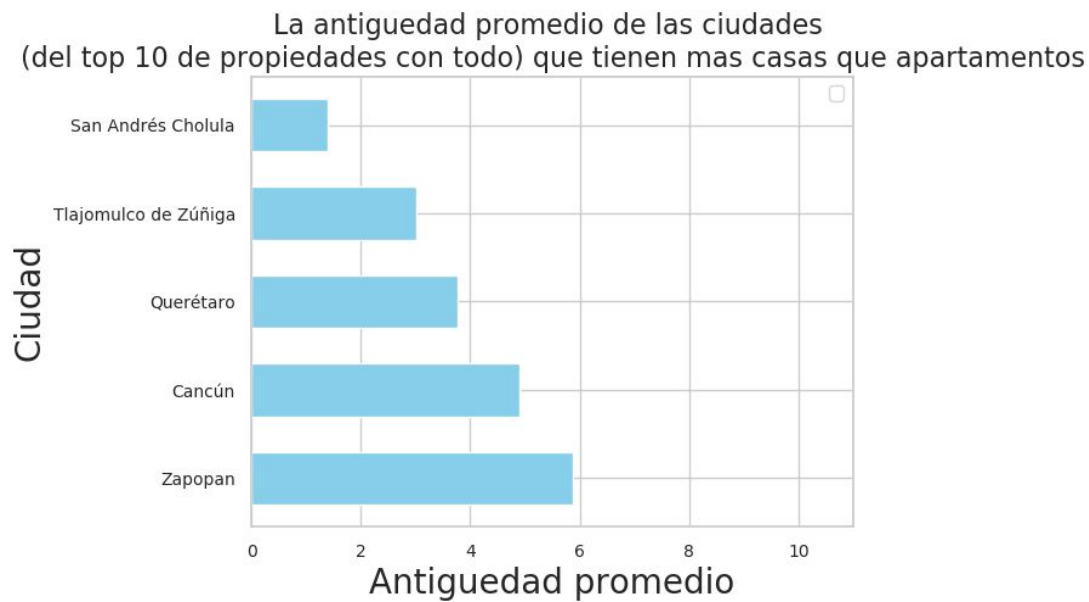
Análisis sobre las 10 ciudades con más cantidad de propiedades con todo

Por información que conseguimos anteriormente, sabemos que las propiedades con todo tienden a ser apartamentos y tener una antigüedad entre 0 y 5 años, con algunas "pocas" de 10 años. Entonces podríamos plantear una hipótesis que sea que las 10 ciudades que observamos antes, estarán formadas mayormente por apartamentos que casas.



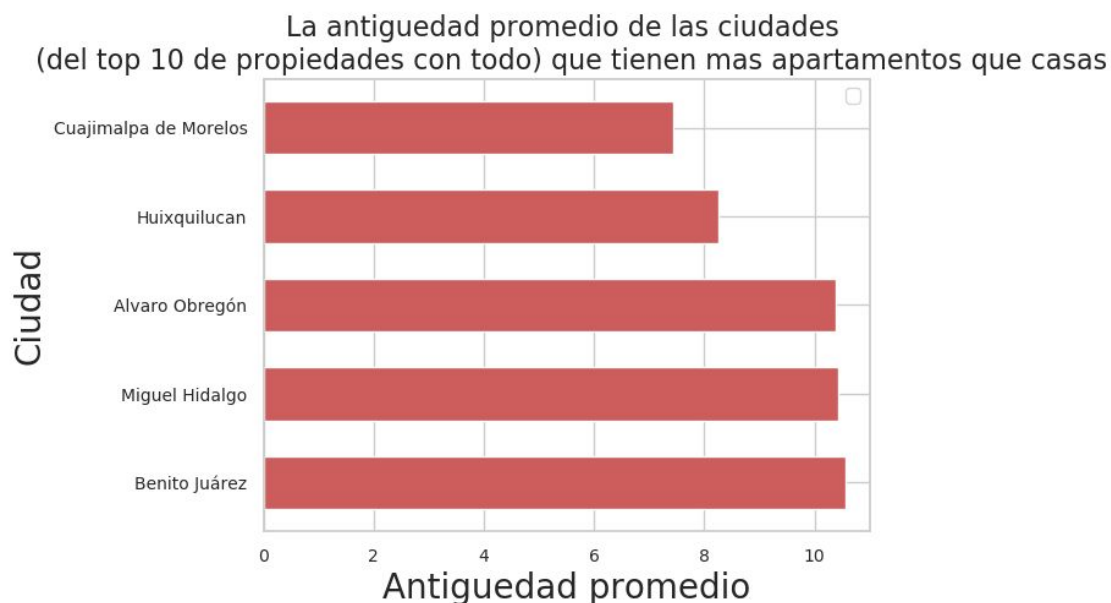
Podemos ver que algunas cumplen con nuestra hipótesis y otras no. Vamos a separar las ciudades en dos grupos, las ciudades que cumplieron nuestra hipótesis (tienen más apartamentos) y las que no (tienen más casas). Géneros otras dos hipótesis basadas en la información que tenemos sobre la antigüedad. Las ciudades con más casas que apartamentos tenderán a tener propiedades menos antiguas, ya que pienso que tienen que “compensar” de algún otro lado. Por el contrario la otra hipótesis sería, como las ciudades que tienen más apartamentos cumplieron con nuestra hipótesis anterior, decimos que tenderán a ser ciudades con propiedades más antiguas.

Ciudades con mayor cantidad de casas (no cumplieron):



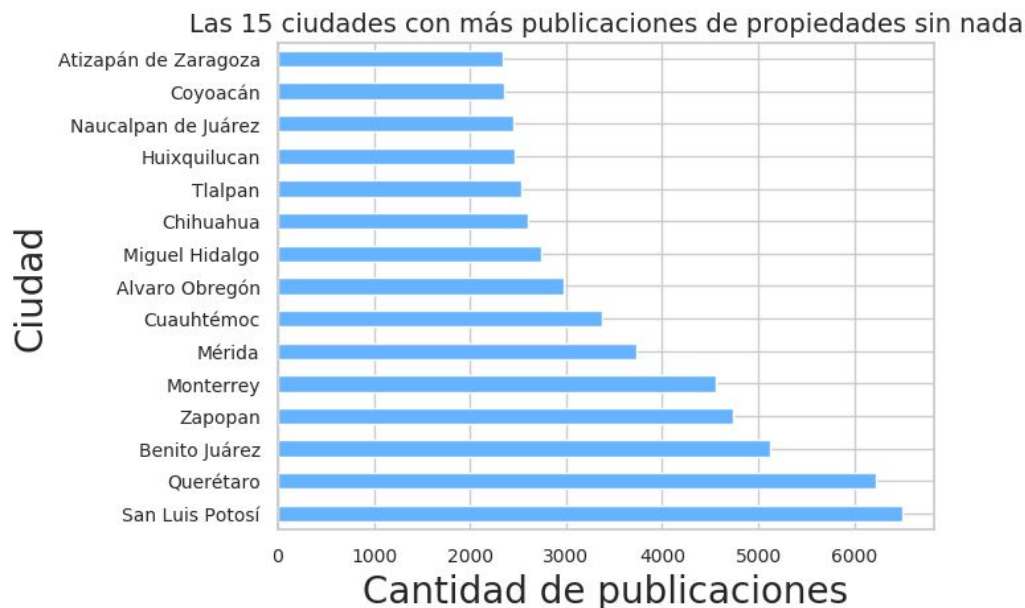
Podemos ver que se cumplio una de las recientes hipótesis, los valores promedio de antigüedad están entre 1.5 y 6. También podemos apreciar que el orden en el que están es exactamente igual al orden en el que estaban en el top de las ciudades con mayor cantidad de propiedades con todo.

Ciudades con mayor cantidad de apartamentos (cumplieron):



Así se cumple nuestra última hipótesis, mostrando que estas ciudades tienen propiedades con mayor antigüedad, con valores entre 6.5 y 10.5 (con una notable concentración en valores cercanos a 10).

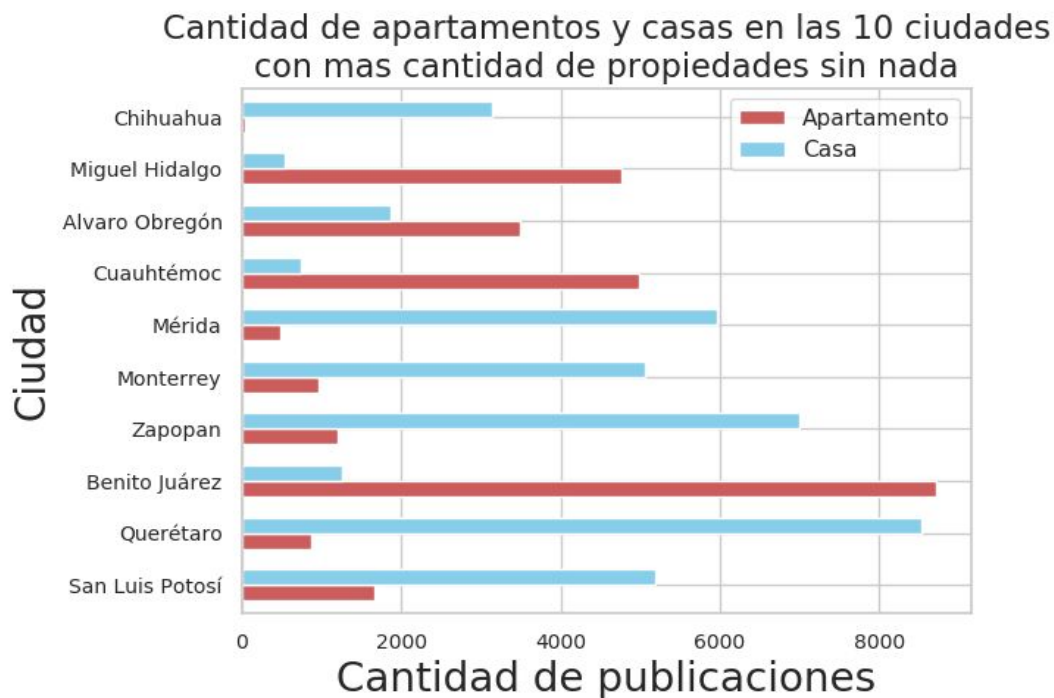
¿Cuales son las ciudades con mayor cantidad de propiedades sin nada?



A diferencia del top de las ciudades con propiedades que tienen todo no hay ningún pico que sobresale. Podemos ver que las ciudades que más se nota la diferencia entre ellos van de San Luis Potosí hasta Chihuahua, ya que desde Chihuahua hasta Atizapán de Zaragoza toman valores muy parecidos. Igual que antes tomamos entonces esas ciudades para hacer un análisis particular, también vuelve a ser un top 10.

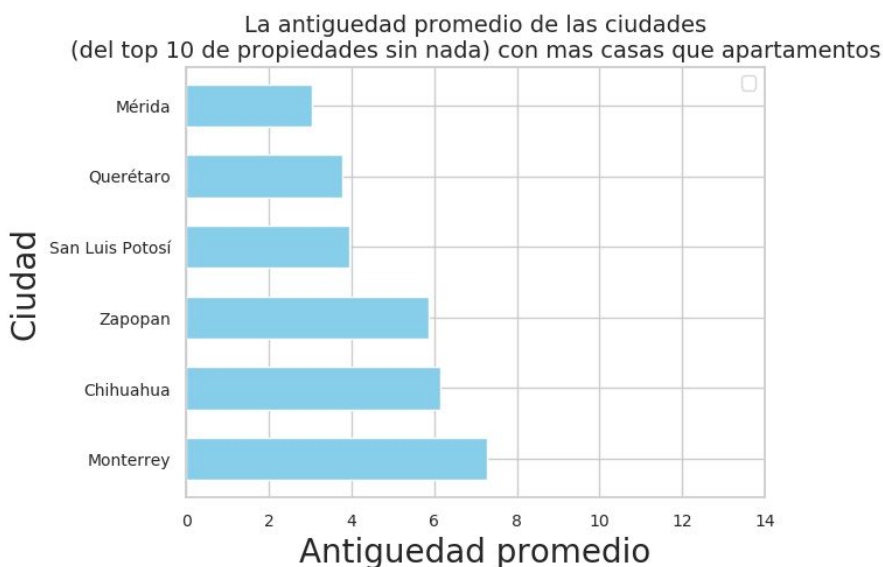
Análisis sobre las 10 ciudades con más cantidad de propiedades sin nada

Por lo que sabemos, las propiedades sin nada suelen ser casas y su antigüedad tiende a ser de 0 años, seguido de 5, 10 y 20. Planteamos la hipótesis de que las 10 ciudades con más cantidad de propiedades sin nada serán mayormente ciudades que tengan más casas que apartamentos.



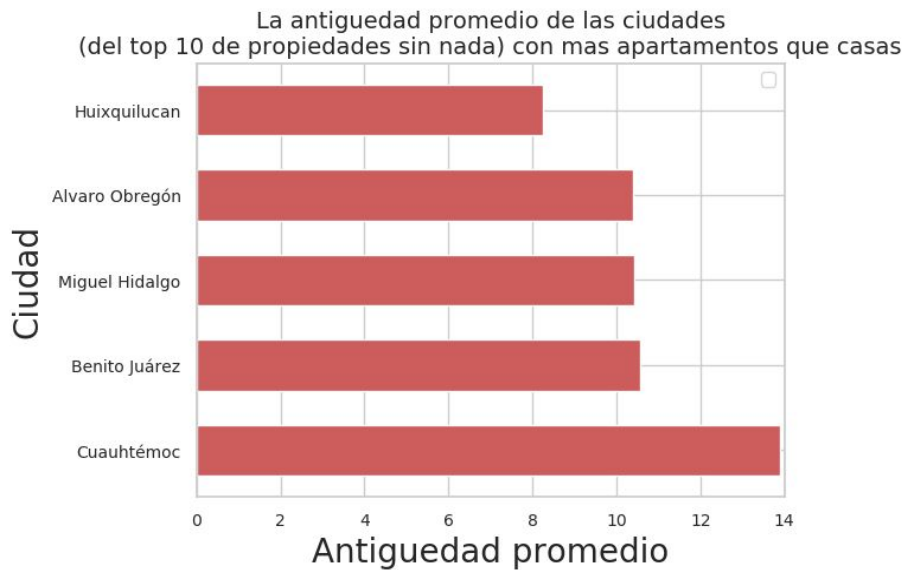
De nuevo vemos que hay algunas no que cumplen con la hipótesis planteada. Analizaremos primero a las ciudades que no cumplieron la hipótesis y luego a las que sí. Planteó que, para las ciudades que no cumplieron tener una mayor cantidad de casas, tenderán a ser más viejas (ya que el promedio de antigüedad de propiedades sin nada es 8.4) y deben compensar. Por otro lado, las ciudades que sí cumplieron tenderán a ser mas nuevas, ya que está más alejado del promedio, pero aun así sigue habiendo una cantidad importante de propiedades que son nuevas.

Ciudades con mayor cantidad de apartamentos (no cumplieron):



Se puede ver como parecería cumplirse una de nuestras hipótesis. Los valores de antigüedad estan entre 3 y 7. Hay 3 ciudades que tienen valores más altos, asique veremos como serán los valores de las ciudades que tienen mayor cantidad de casas para compararlos.

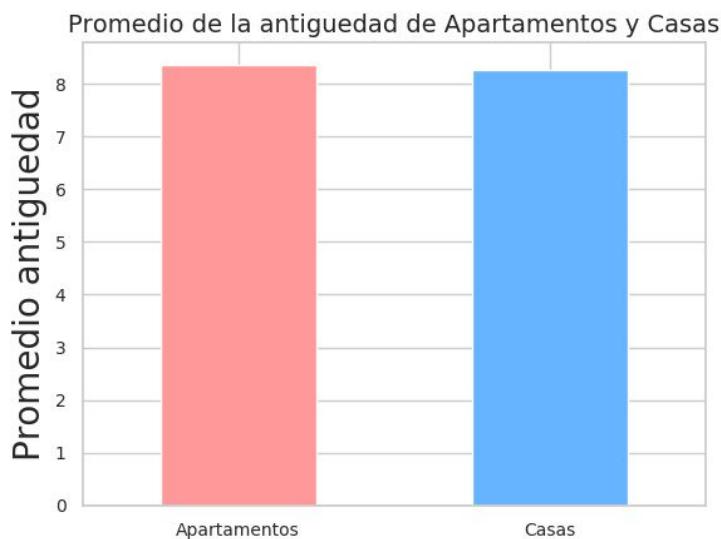
Ciudades con mayor cantidad de casas (cumplieron):



Efectivamente las ciudades con mayor cantidad de casas tienen mayor promedio de antigüedad. Los valores se concentran en 10 con la excepción de Cuauhtémoc que llega casi a 14.

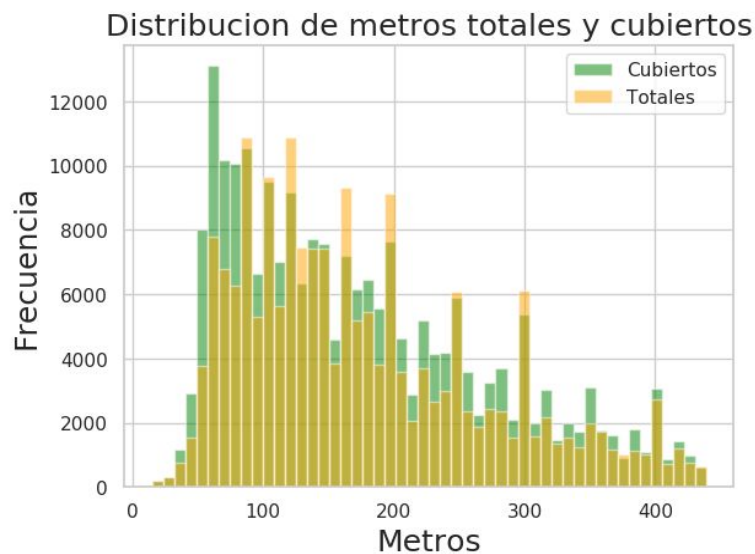
Confirmación de las causas de las consecuencias de los recientes análisis parciales sobre las 10 ciudades con más propiedades con todo y sin nada:

Se podría pensar que nuestras recientes hipótesis son “erróneas”, ya que podría ser que en realidad las ciudades con mayor cantidad de casas tienden a ser más viejas por el hecho de que en general el promedio de las casas es alto y que las ciudades que tienen más apartamentos tienen un promedio de antigüedad más bajo porque el promedio general de apartamentos es bajo. Así que confirmamos que esto no es así mirando el promedio de antigüedad de casas y apartamentos en general.

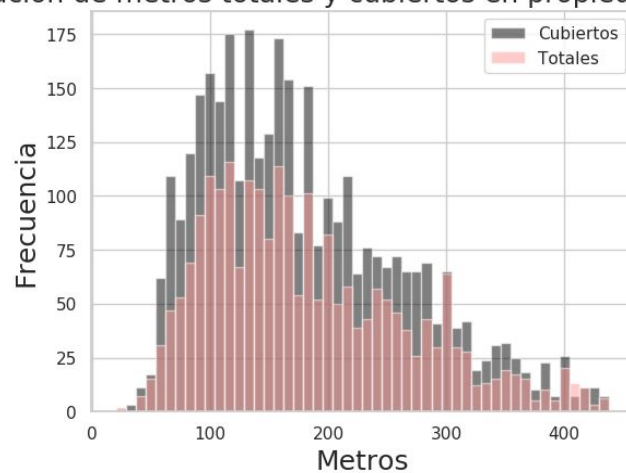


Así vemos como el promedio de ambos tipos de propiedades es de alrededor de 8. Lo que hace invalida la posible lógica recién planteada.

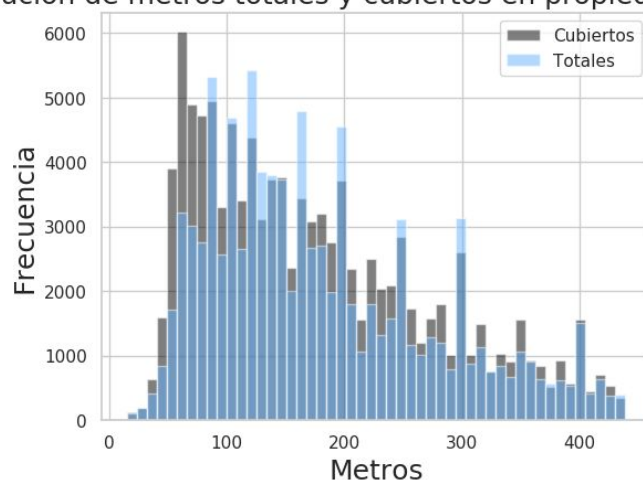
¿Cual es la distribución de metros totales y cubiertos en propiedades con todo, sin nada y en general?



Distribucion de metros totales y cubiertos en propiedades con todo



Distribucion de metros totales y cubiertos en propiedades sin nada



Se ve como las propiedades que tienen todo tienden a “tener” mayor cantidad de metros cubiertos que las propiedades sin nada y en general.

¿Cual es el precio promedio de propiedades con todo y sin nada?



En la figura se muestra como las propiedades con todo tienden a ser más caras que las que no tienen nada

¿Cual es la cantidad de publicaciones de propiedades con todo y sin nada?



Como era de esperarse, la cantidad de propiedades con todo es minima en comparacion a la cantidad de publicaciones que hay de propiedades sin nada.

Conclusiones

Según lo observado en el set de datos de ZonaProp podemos observar que el set de datos está formado con datos mayoritariamente de propiedades de casas y apartamento y de donde contiene mayor cantidad de información. Encontramos que se encuentran la mayor cantidad de publicaciones es en el Distrito Federal (hoy Ciudad de México) y el Estado de México y que en su principales ciudades donde están las publicaciones. En estas ciudades del DF Benito Juárez, Álvaro Obregón, Cuauhtémoc y en estas ciudades pertenecientes al Estado de México, la empresa podría haber localizado campañas para favorecer más publicaciones y por ende aumentar el flujo de navegación en el sitio. En cuanto a publicaciones realizadas vemos un gran incremento en diciembre de 2016 comparado con todos los meses anteriores. No encontramos información adicional en internet que revele el motivo de tal incremento en las publicaciones.

Con respecto al precio de las diferentes propiedades podemos concluir que:

- Aumenta según la cantidad de metros cuadrados que posee, ya sea los cubiertos como los totales.
- Aumenta según la cantidad de amenities que posee el inmueble, sin importar de cuál se trata (salón de usos múltiples, pileta o gimnasio).
- El Distrito Federal cuenta con las propiedades con precio promedio más alto, seguido por las provincias de Edo. de México, Nuevo León, Guerrero y Quintana Roo.
- En Baja California Norte se cuenta con el promedio de precio más bajo de todo el set de datos.
- Tanto los apartamentos como las casas en condominio tienen las medias de precio más altas.
- Se puede afirmar que el 50% de las publicaciones de los terrenos se concentra en un rango de precios muy bajo, comparado con los otros tipos de propiedades.
- Las propiedades con un centro comercial cercano en promedio tienen precio más elevado con respecto a las que no cuentan con esta cercanía.
- En general, tomando todos los datos se pudo apreciar una distribución pareja del precio entre las propiedades con antigüedad de 0 a 25 años, luego el precio tiende a subir para aquellas con una antigüedad mayor a 25 años. Sin embargo, en los apartamentos esta relación se da al inversa teniendo los precios más altos en inmuebles con menos de 10 años de antigüedad.

Con respecto a las propiedades con todo y sin nada se puede concluir:

- Las propiedades con todo tienden a ser apartamentos y las propiedades que no tienen nada suelen ser casas.
- Las propiedades con todo suelen ser mas nuevas y las que no tienen nada tenderán a ser más antiguas.
- Las ciudades que tengan mayor cantidad de propiedades con todo tienden a tener una mayor cantidad de apartamentos que casas. Las propiedades de las ciudades que no cumplan esto último, serían menos antiguas. Las propiedades de las ciudades que sí lo hayan cumplido serán más viejas.
- Las ciudades que tienen una mayor cantidad de propiedades sin nada suelen tener una más casas que apartamentos. Las propiedades de las ciudades que no cumplan esto último, serán menos antiguas mientras que las propiedades de las ciudades que sí lo hayan cumplido serán más viejas.
- Las propiedades con todo suelen ser más caras que las que no tienen nada
- Hay una mínima cantidad de propiedades con todo en comparación a las que no tienen nada
- Si una propiedad tiene 0 habitaciones es mucho más probable que sea una propiedad sin nada que una con todo.
- Mientras una propiedad tenga más de 100 metros totales y tenga más metros cubiertos qué totales, es más probable que sea una propiedad con todo que una sin nada.