



75.06/95.58 Organización de Datos - 2C 2019

Trabajo Práctico 1

Análisis Exploratorio

Grupo 34: "DataTravellers"

Integrantes:

- Andrés Pablo Silvestri: 85881 (silvestri.andres@gmail.com)
- Juan Manuel González: 79979 (juanmg0511@gmail.com)
- Patricio Pizzini: 97524 (pizzinipatricio@yahoo.com.ar)

Link a repositorio de GitHub:

https://github.com/patopizzini/Organizacion_Datos_2C2019/tree/master/TP1

Fecha de entrega: 30/09/2019

Contenido

1 - Introducción	4
1.1 - ¿A qué se dedica Navent?	4
2 - Estructura y manejo de los datos	5
3 - Análisis sobre la ubicación geográfica de las propiedades	7
3.1 - ¿Cuál es la distribución geográfica de las propiedades?	7
3.2 - ¿Cuáles son los estados con mayor cantidad de propiedades?	8
3.3 - ¿Cómo se distribuyen las propiedades en México DF?	9
3.4 - ¿Cómo se distribuyen las propiedades Edo. de México?	10
3.5 - ¿Cómo se distribuyen las propiedades en Jalisco?	11
3.6 - ¿Cómo se distribuyen las propiedades en Querétaro?	12
3.7 - ¿Cómo se distribuyen las propiedades en Nuevo León?	13
3.8 - ¿Cómo se distribuyen las propiedades en Puebla?	14
3.9 - ¿Cuáles son, para todo el territorio, las ciudades con mayor cantidad de propiedades?	15
4 - Análisis sobre los precios de las propiedades	16
4.1 - ¿Cuál fue la evolución del precio promedio de las propiedades en el tiempo?	16
4.2 - ¿Cuál fue la evolución del precio promedio de las propiedades por estado?	17
4.3 - ¿Cuál fue la evolución del precio promedio de las propiedades por ciudad?	18
4.4 - ¿Cuál fue la evolución del precio promedio por zona de México DF?	19
4.5 - ¿Cuál fue la evolución del precio promedio por tipo de propiedad?	20
4.6 - ¿Varía el precio de las propiedades de acuerdo al mes de publicación?	21
4.7 - ¿Cuál es el precio promedio por metro cuadrado?	22
4.8 - ¿Cuál es el precio promedio por metro cuadrado para las dos ciudades con mayor cantidad de propiedades?	23
4.9 - ¿Cómo varía el precio de acuerdo a la distancia del centro de México DF?	24
4.10 - ¿Cuál es el precio promedio por metro cuadrado para las zonas de México DF?	25
4.11 - ¿Cómo varía el precio del metro cuadrado según la cantidad de habitaciones?	26
4.12 - ¿Cómo varían los precios de las propiedades con y sin amenities?	27
4.13 - ¿Cómo varían los precios de las propiedades con y sin gimnasio?	28
4.14 - ¿Cómo varían los precios de las propiedades con y sin piscina?	29
5 - Análisis sobre las características de las propiedades	30
5.1 - ¿Cómo se distribuyen las propiedades por tipo?	30
5.2 - ¿Cuál es la proporción de propiedades que cuentan con amenities?	31
5.3 - ¿Cuál es la proporción de propiedades que cuentan con gimnasio?	32
5.4 - ¿Cuál es la proporción de propiedades que cuentan con piscina?	33
5.5 - ¿Cómo se distribuyen las propiedades de acuerdo a la cantidad de garajes que poseen?	34
5.6 - ¿Cómo se distribuyen las propiedades de acuerdo a la cantidad de habitaciones que poseen?	35

5.7 - ¿Cuál es la relación entre la cantidad de metros cuadrados y la cantidad de habitaciones?	36
5.8 - ¿Cuál es la relación entre la cantidad de habitaciones de una propiedad y su antigüedad?	37
5.9 - ¿Cuál es la relación entre la cantidad de habitaciones y el tipo de propiedad?	38
6 - Análisis sobre la actividad del sitio	39
6.1 - ¿Cuál es la cantidad de publicaciones por año?	39
6.2 - ¿Cuál es el promedio de publicaciones por mes?	40
6.3 - ¿Cuál es el promedio de publicaciones por día de la semana?	41
6.4 - ¿Cómo fue la evolución en la cantidad de publicaciones para las ciudades TOP?	42
6.5 - ¿Cómo fue la evolución en la cantidad de publicaciones de acuerdo al número de habitaciones?	43
7 - Conclusiones	44

1 - Introducción

Este trabajo está enfocado en hacer un primer análisis de los datos ofrecidos por Navent, de manera que encontremos particularidades que puedan ser de interés para dicha empresa.

Siendo este el caso, lo primero que vamos a hacer es interiorizarnos de uno de los pilares que tiene la ciencia de datos, el negocio.

Se hará un análisis general para entender un poco más el negocio y luego un análisis más fino sobre los datos para contestar algunas preguntas que podrían resultar curiosas.

1.1 - ¿A qué se dedica Navent?

Navent es una empresa argentina con presencia en la región, se trata de un conglomerado de unas 20 marcas enfocado en el segmento de los clasificados online, con dos grandes verticales de negocio: empleo y bienes raíces.

La empresa fue fundada en 2007 y cuenta en el país con algunas marcas muy reconocidas, como ZonaJobs y ZonaProp, ambas líderes en sus respectivas verticales. Hoy cuenta con presencia en 8 países y 25 ciudades, con unos 800 empleados.

En su sitio web, se publica que en la actualidad sus diversas plataformas son utilizadas por unos 95.000 particulares y empresas, contando con unos 4.2 millones de anuncios, conectando 21 millones de personas.

En este caso, Navent nos facilitó un set de datos perteneciente a su portal ZonaProp, del segmento de bienes raíces. El set contiene información sobre las propiedades publicadas en el portal para México, entre los años 2012 y 2016.

2 - Estructura y manejo de los datos

A fin de mejorar la performance de ejecución de las operaciones realizadas, y por otro lado facilitar las tareas de análisis, se ha hecho un tratamiento previo sobre el set de datos proporcionado por la cátedra.

Se puede consultar el tratamiento completo, así como también el código utilizado para generar el análisis y los gráficos, en el notebook disponible en la carpeta "Entrega" del repositorio de GIT presentado en la carátula de este informe. Asimismo, una copia del mismo notebook ha sido agregado como kernel en Kaggle. Puede consultarse en la siguiente URL, de acceso público: <https://www.kaggle.com/juanmg0511/tp1-2c2019-g34>.

Como primer medida, se consideró tipar las columnas de los archivos en forma manual, a fin de mejorar el tiempo de carga y de uso de memoria. Esta estrategia fue recomendada en clase para datasets grandes, pero luego de cargar y analizar las características básicas del set del trabajo práctico se decidió que no era necesario ya que el tamaño era fácilmente manejable por nuestros equipos.

Contamos con un solo archivo para esta entrega, de unos 150MB de tamaño y 240.000 registros, que cubre un período de actividad de 5 años, más precisamente entre el 2012 y el 2016:

- **TRAIN.CSV** - información sobre las publicaciones del portal ZonaProp para México.

Como primer medida, armamos algunas columnas auxiliares que nos permitan separar la fecha de publicación en año, mes y día de la semana, para luego poder aplicar esta información como parámetros de búsqueda y organización en los diferentes gráficos que planteamos. Asimismo, también copiamos el año como string, a fin de evitar que aparezca con decimales en los gráficos.

Un tema que surgió durante el análisis fue el tratamiento de los registros con valores NaN en las columnas siendo investigadas. En estos casos fue necesario adoptar un criterio para solucionar esto: por lo general descartamos estos registros para la consulta que estábamos realizando, pero para algunos otros casos se rellenaron los datos con el promedio de los demás.

Al analizar los datos desde el punto de vista de la ubicación geográfica de las publicaciones, también fue necesario hacer cierta manipulación: se acotaron las coordenadas a las válidas para México, descartando valores fuera de dicho rango antes de realizar el procesamiento.

Otro aspecto importante del manejo de los datos se dió con el parseo del campo de descripción de las publicaciones, ya que en este caso se procedió a buscar algunas palabras clave como "esquina" o "seguridad" para luego guardar este dato en columnas auxiliares. Esta metodología nos permitió generar información "adicional" a la presente en las columnas estándar del dataset, para ser usada como parte del análisis.

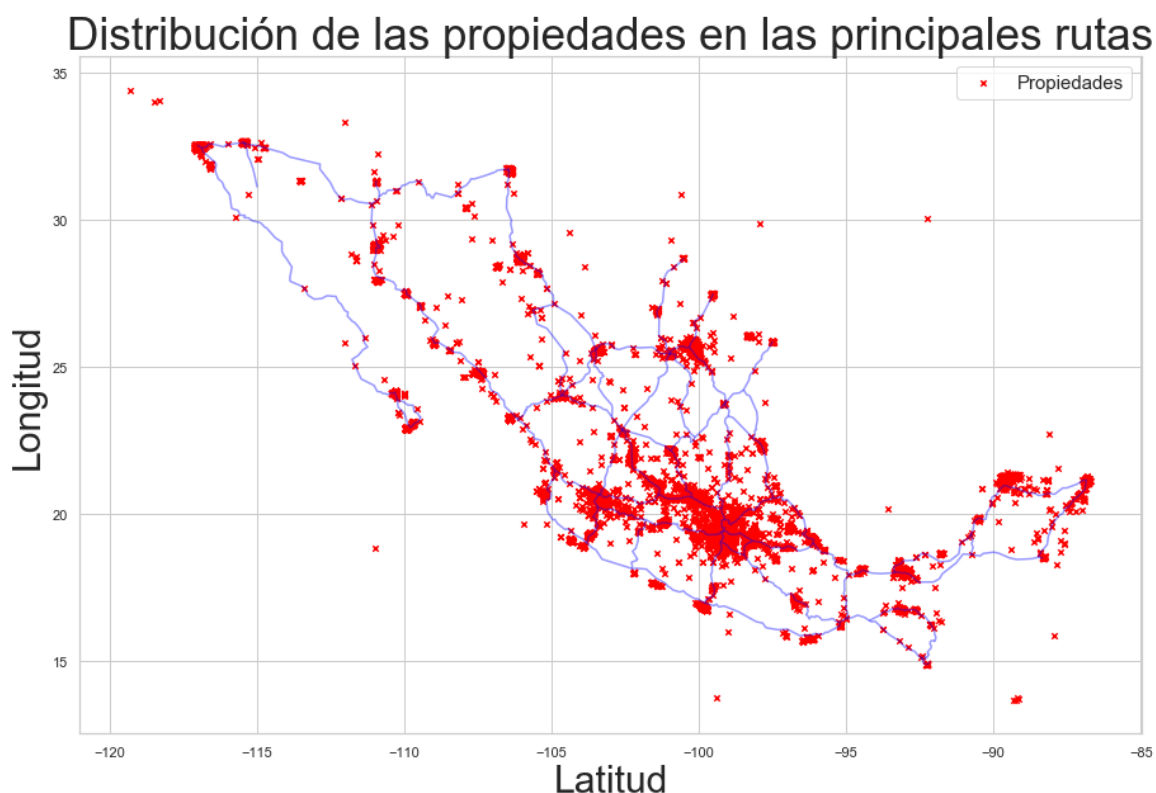
Por último, vale agregar que en algunos casos nos hemos tomado la libertad de considerar algunos valores como despreciables en cuanto a relación de cantidad sobre el total, como así también hacer un análisis sobre los casos que más cantidad de registros nos brindan, lo que nos permite entender la situación con un volumen de datos mayor lo cual nos acerca un poco más a la realidad.

3 - Análisis sobre la ubicación geográfica de las propiedades

Nos pareció interesante comenzar el análisis viendo cómo las distintas propiedades publicadas en el portal se distribuían geográficamente en México, a fin de entender donde se ubican la mayor cantidad de usuarios y extraer conclusiones.

3.1 - ¿Cuál es la distribución geográfica de las propiedades?

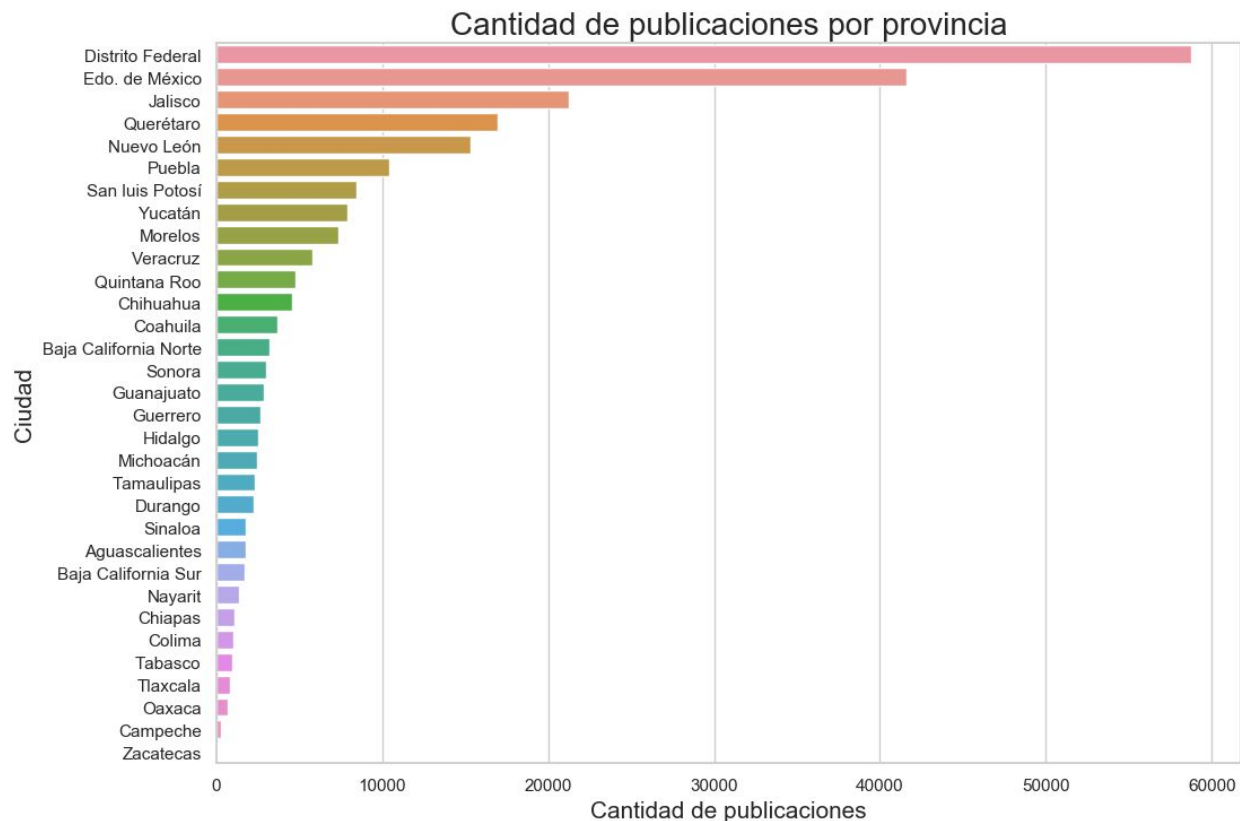
Hemos tomado un mapa de las carreteras de México para evaluar cerca de qué zonas están hechas las publicaciones, obteniendo el resultado que se presenta a continuación.



Comparándolo con un mapa político pudimos notar que, como presentimos, hay un marcado aumento en la concentración de publicaciones cerca de la zona central del país, donde se encuentra una de las urbes más pobladas del mundo: México DF, su área metropolitana y también las ciudades cercanas.

3.2 - ¿Cuáles son los estados con mayor cantidad de propiedades?

Haciendo un análisis similar, se puede apreciar, de otra forma, en que zonas o estados del país se encuentran la mayor cantidad de publicaciones.



Tal como nos mostró el mapa anterior, la zona más poblada del país también es donde más publicaciones hay, lo que suena lógico debido a la ocupación poblacional. De hecho, entre el DF y el Estado de México se reparten aproximadamente el 40% de las publicaciones, una cifra muy superior a la de cualquier otro estado.

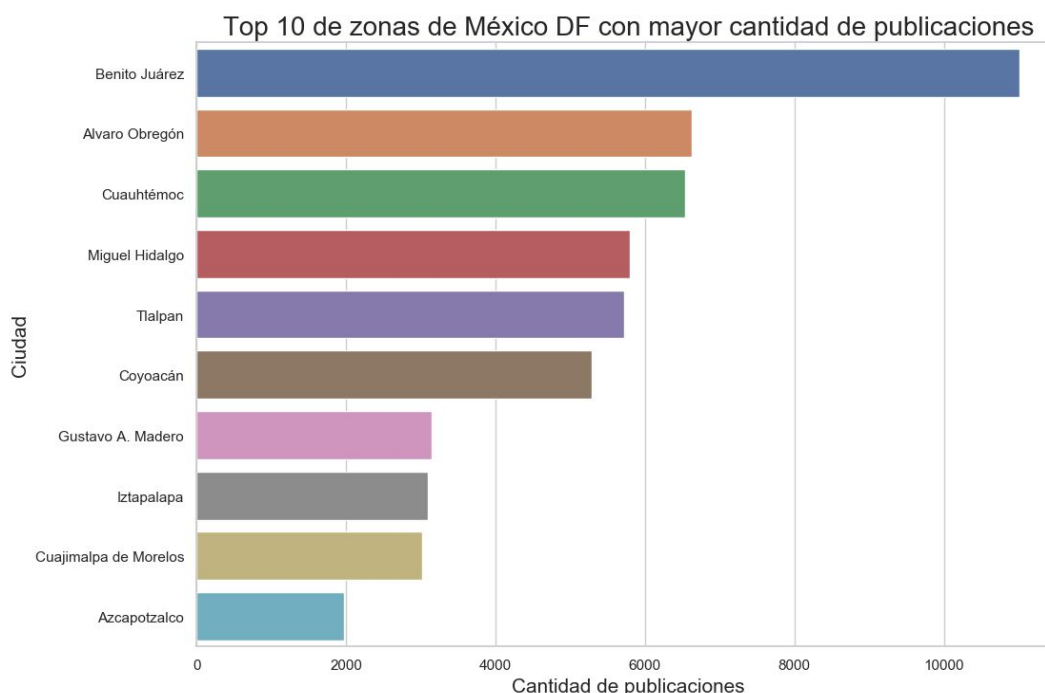
También tenemos estados con ciudades importantes encabezando la lista, como por ejemplo Jalisco (Zapopan / Guadalajara), Querétaro (Querétaro) y Nuevo León (Monterrey).

Si lo comparamos con los datos censales del año 2010¹, encontramos algunas discrepancias, por ejemplo Veracruz, con una gran cantidad de población pero no tan arriba en nuestro gráfico, si bien la distribución obtenida sigue en líneas generales la cantidad de población por estado.

Si en cambio vemos el PBI por estado mexicano para el año 2016², encontraremos que esto ajusta mejor a la lista obtenida: es decir que vemos una mayor cantidad de publicaciones en las zonas que generan la mayor riqueza. Este resultado es lógico, pero nos sirve para entender dónde reforzar actividades como promociones o publicidad, a fin de acaparar más clientes, o entender que extracto de la población utiliza el portal para publicar propiedades.

3.3 - ¿Cómo se distribuyen las propiedades en México DF?

En sintonía con esto último, tomamos las publicaciones de México DF, y analizamos su distribución:



¹ Disponible en:

https://es.wikipedia.org/wiki/Anexo:Indicadores_económicos_de_las_entidades_federativas_de_México

² Disponible en:

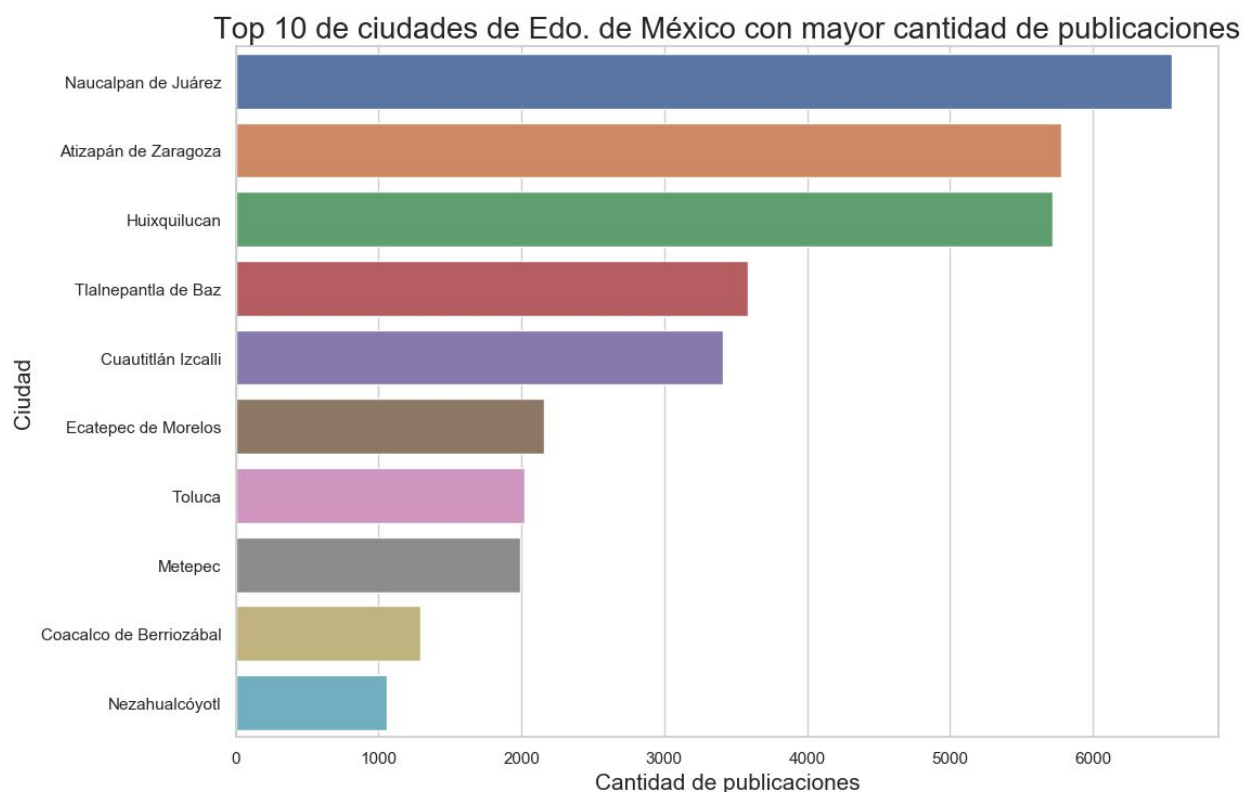
<http://cuentame.inegi.org.mx/monografias/informacion/mex/economia/pib.aspx?tema=me&e=15>

El resultado obtenido es que en Benito Juárez es donde más propiedades se encuentran publicadas, llevándole una amplia ventaja a las demás áreas de la capital.

Investigando un poco podemos ver que varias de las zonas más exclusivas de la ciudad se encuentran entre los resultados obtenidos³, lo que da una idea del target en cuanto al poder adquisitivo de los compradores de las publicaciones, al menos para la capital. Conocer esta información puede ser útil nuevamente para entender si los esfuerzos de captación de nuevas publicaciones están bien enfocados y cumplen con lo esperado.

3.4 - ¿Cómo se distribuyen las propiedades Edo. de México?

Para el resto de los estados no contamos con información tan detallada como para la capital, pero si contamos con la información de a qué ciudades de un estado pertenecen las publicaciones. Por ello, analizaremos la distribución en los 5 estados restantes con mayor cantidad actividad, comenzando por el Estado de México.



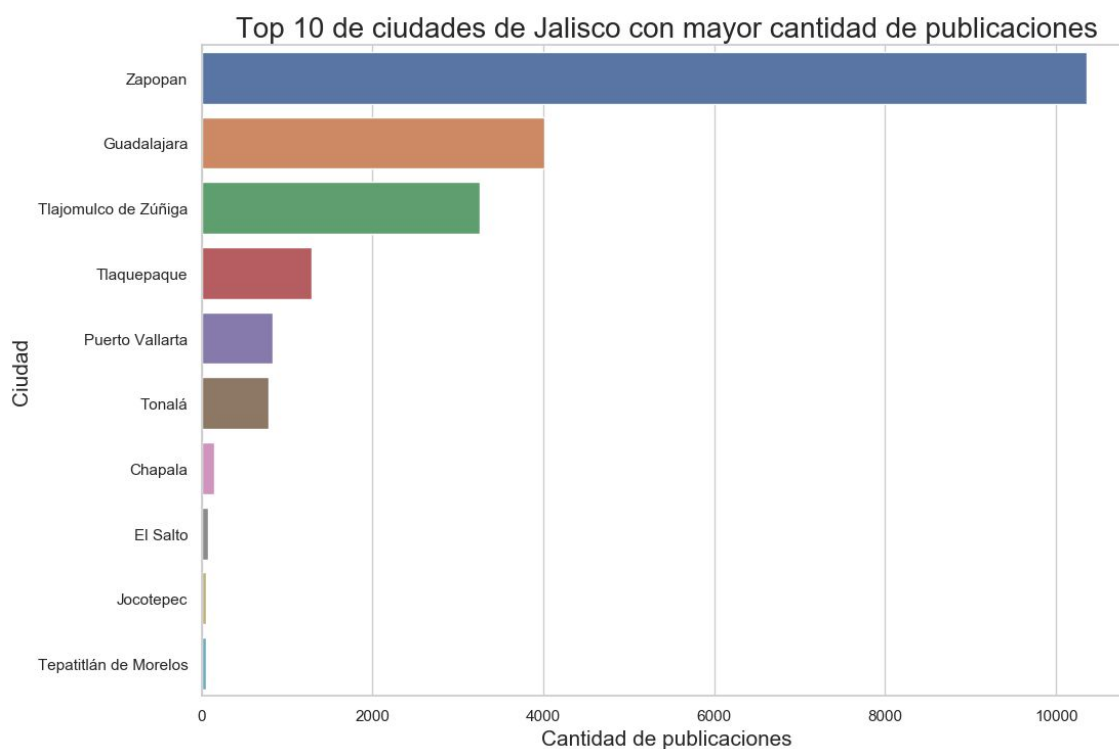
³ Disponible en:

<http://www.metroscubicos.com/articulo/guia-de-precios/2012/11/28/las-10-zonas-mas-exclusivas-para-vivir-en-el-df>

Haciendo el mismo análisis sobre el segundo estado con más propiedades, el Estado de México, notamos que en este caso no hay tanta diferencia entre la ciudad con más publicaciones, Naucalpan de Juárez, y las dos ciudades que le siguen.

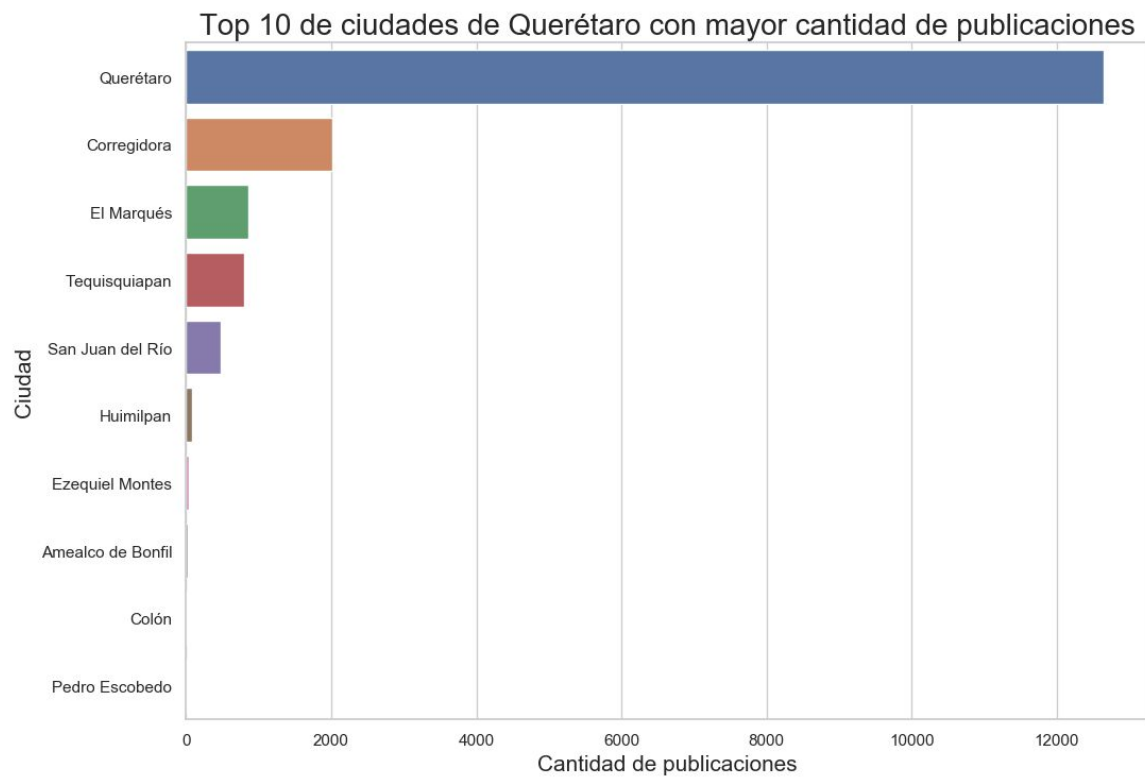
Vale aclarar que este análisis puede potencialmente repetirse para todos los estados del país, o lo que es más importante, para los cuales el cliente haya expresado interés en conocer. Al no contar con esta información, y para no extender el informe innecesariamente, nos concentramos solamente en el Distrito Federal y los 5 estados restantes con mayor cantidad de publicaciones.

3.5 - ¿Cómo se distribuyen las propiedades en Jalisco?



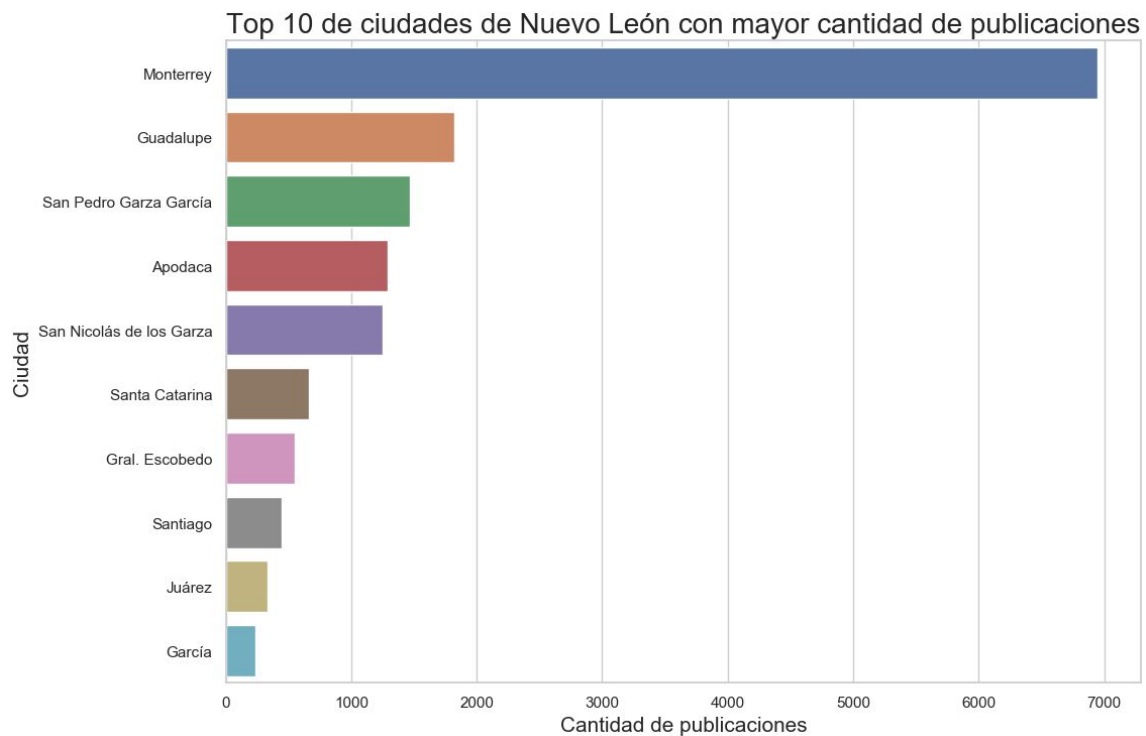
A diferencia de la visualización anterior, aquí uno puede ver que en Jalisco las publicaciones si están en su gran mayoría en la ciudad de Zapopan (un 50% del total del estado), obteniendo una gran ventaja sobre las otras ciudades como Guadalajara o Tlajomulco de Zúñiga.

3.6 - ¿Cómo se distribuyen las propiedades en Querétaro?



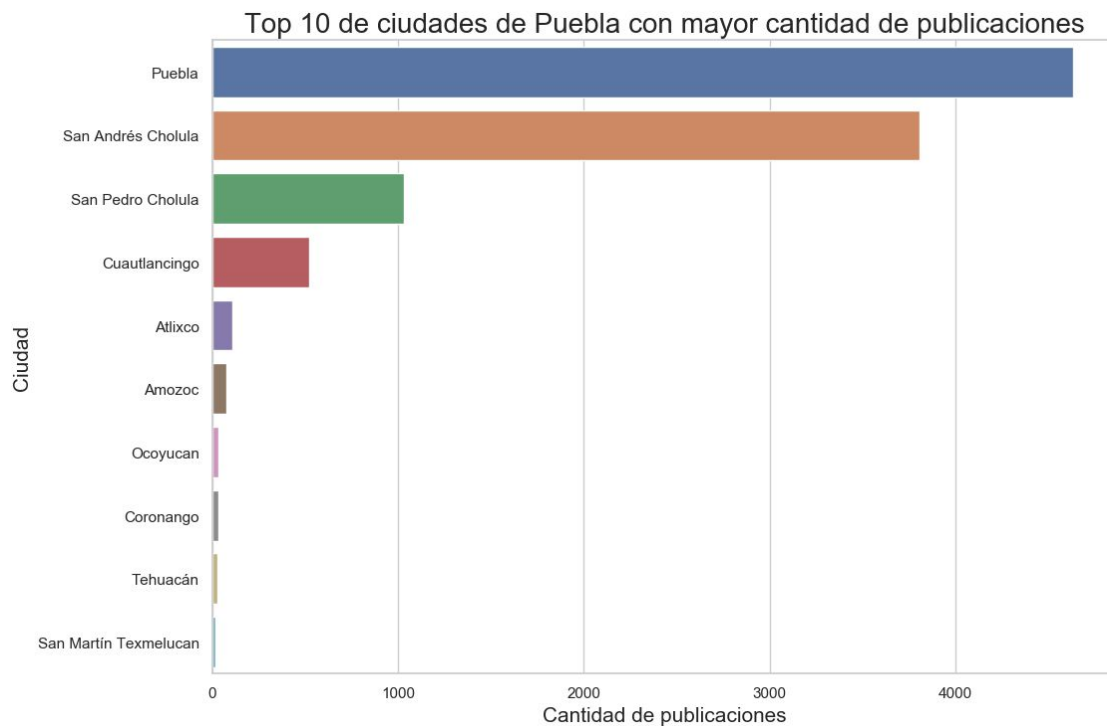
Siguiendo con el mismo análisis, en el estado de Querétaro podemos notar que casi todas las publicaciones están hechas en la ciudad con el mismo nombre siendo seguido por la ciudad de Corregidora.

3.7 - ¿Cómo se distribuyen las propiedades en Nuevo León?



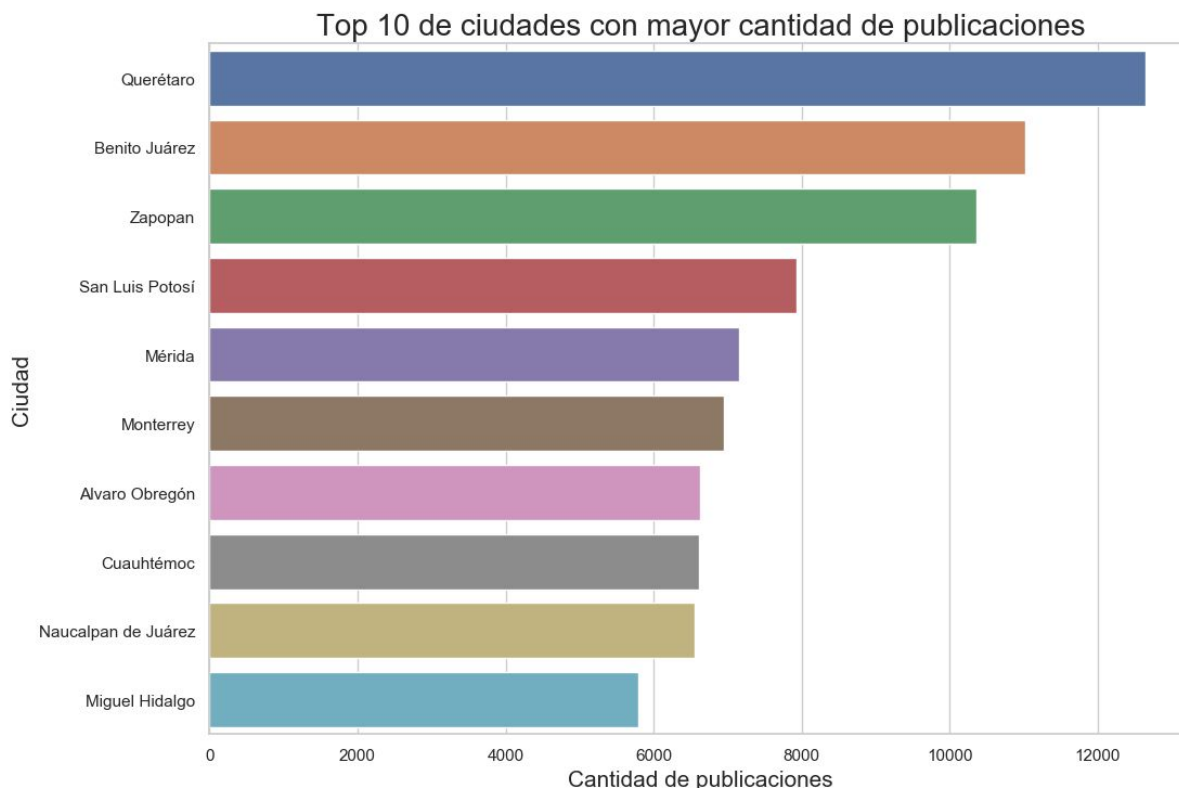
En el estado de Nuevo León podemos decir que gran parte de las publicaciones se encuentran en la ciudad de Monterrey con casi 7000. Lo sigue la ciudad de Guadalupe con un aproximado de 1800 publicaciones menos. Podemos decir también que, sacando a la primer ciudad, el resto de las publicaciones se encuentran bastante distribuidas en el resto de las ciudades del estado.

3.8 - ¿Cómo se distribuyen las propiedades en Puebla?



Finalizando el análisis de estados, en Puebla, se diferencia del resto ya que las dos ciudades con más publicaciones, Puebla y San Andrés Cholula, le llevan una amplia diferencia al resto, donde se destacan dos ciudades pero con 1000 o menos publicaciones.

3.9 - ¿Cuáles son, para todo el territorio, las ciudades con mayor cantidad de propiedades?



Por último, tenemos un gráfico comparativo de las 10 ciudades con más publicaciones en todo el territorio de México. Debe notarse que en realidad también se incluyen mezcladas las zonas de México DF como ciudades individuales, lo que muestra también la gran cantidad de publicaciones que hay en la capital del país.

En esta visualización se destaca Querétaro, del estado que lleva el mismo nombre como la ciudad con más propiedades publicadas, teniendo más de 12000. Las dos ciudades que le siguen son Benito Juárez y Zapopan, de los estados de México DF y Jalisco respectivamente, con más de 10000.

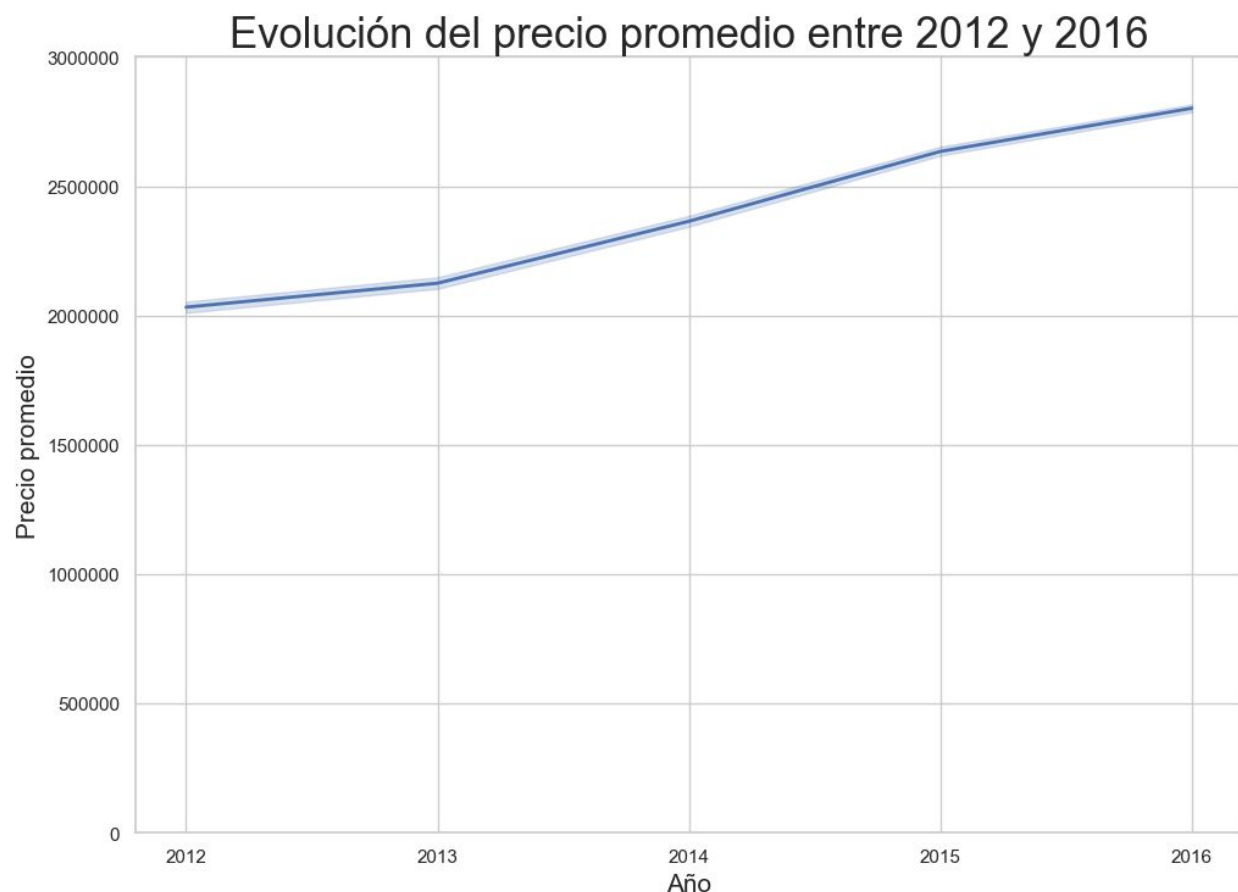
El resto de las ciudades ya cuentan con menos de 8000 publicaciones, este número va descendiendo hasta menos de 6000, pero en ese rango se encuentran 7 de las 10 ciudades con más publicaciones.

4 - Análisis sobre los precios de las propiedades

Continuamos el análisis investigando los precios de las propiedades. Consideramos que el trabajo de esta sección es particularmente importante, en vistas a la segunda etapa del trabajo práctico, donde ya nos fue adelantado deberemos predecir el precio de una propiedad.

4.1 - ¿Cuál fue la evolución del precio promedio de las propiedades en el tiempo?

Con este apartado vamos a tratar de entender cómo fueron evolucionando los precios de las propiedades en México realizando un análisis económico.



Como base planteamos el precio general de las propiedades en todo el país y vemos como creció año a año, obteniendo como resultado que fue un ascenso casi lineal,

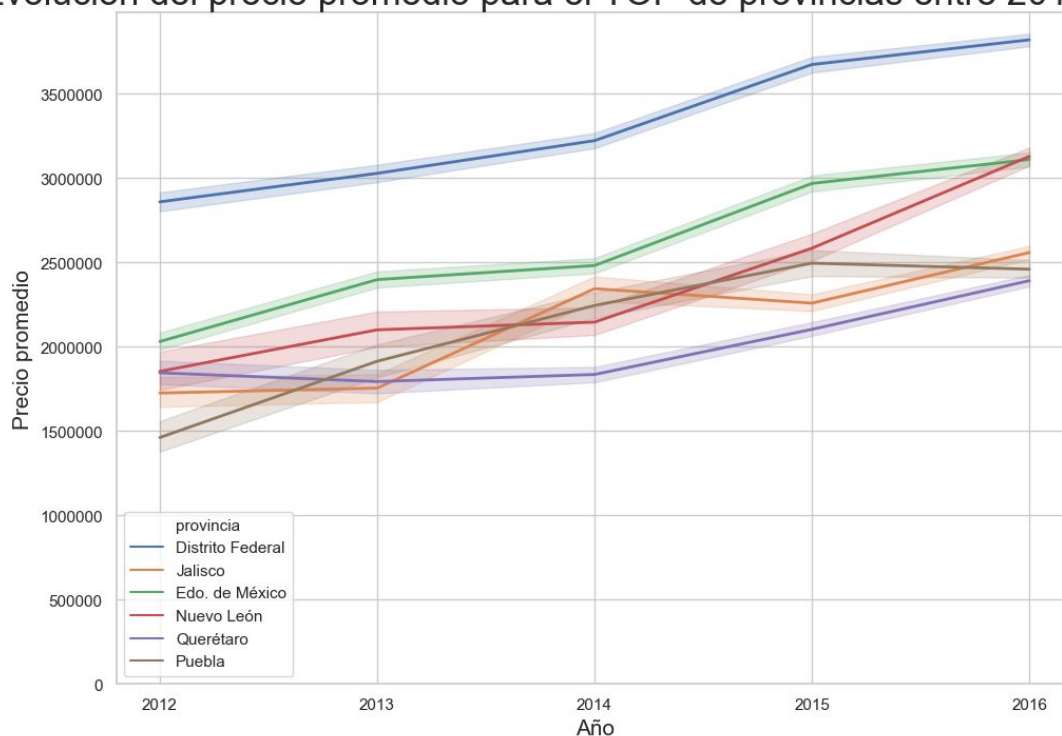
pero no tan ascendente. Si se compara con datos de la inflación de precios⁴ durante el mismo periodo nos da una similitud con lo obtenido aquí.

Si bien entre 2012 y 2013 no hubo un gran aumento de precios, si se nota más durante los años sucesivos.

4.2 - ¿Cuál fue la evolución del precio promedio de las propiedades por estado?

Análogo al análisis anterior, lo que buscamos con el siguiente gráfico es ver como creció el precio de las propiedades en los 6 estados con más propiedades publicadas, los mismos para los que hicimos el análisis de ciudades.

Evolución del precio promedio para el TOP de provincias entre 2012 y 2016



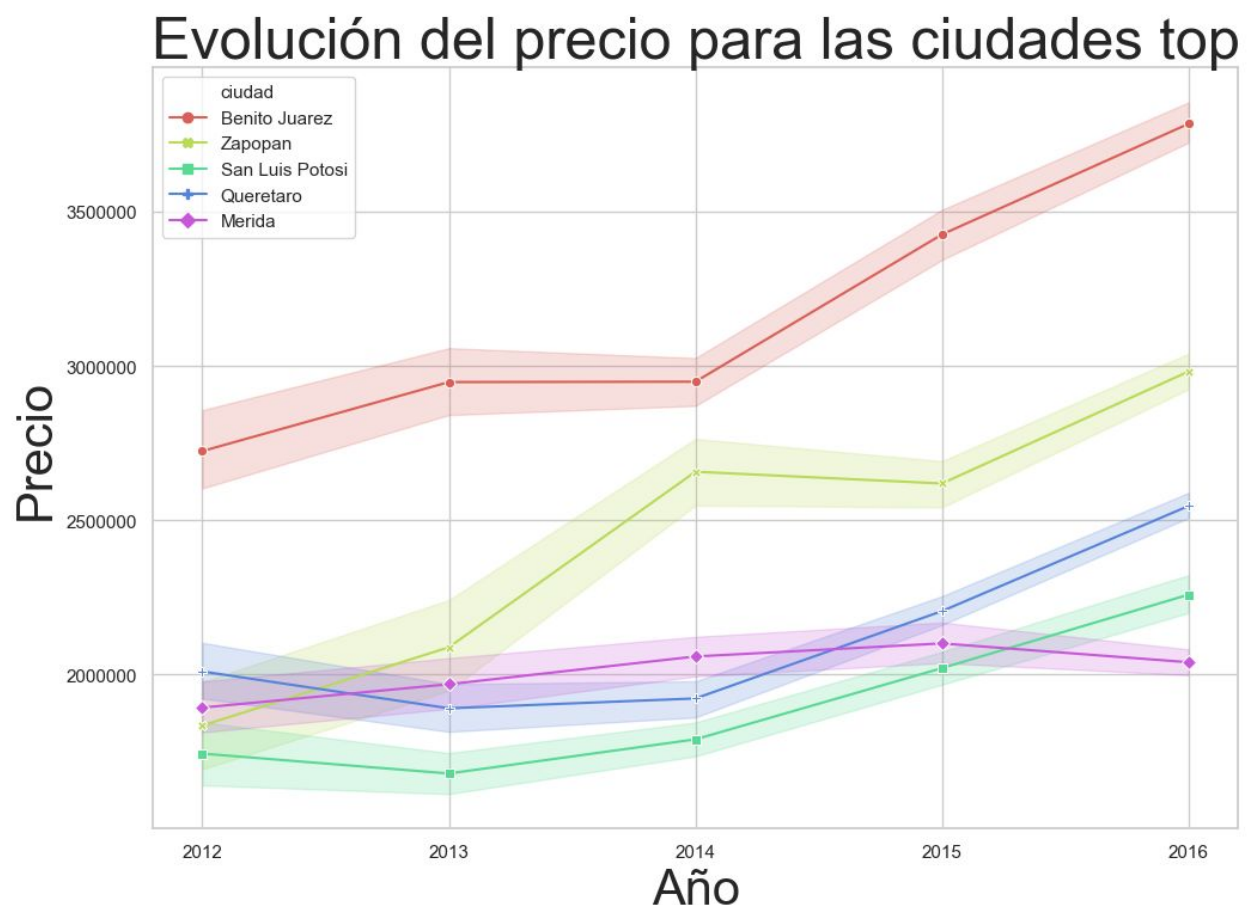
Obtenemos como resultado que en el Distrito Federal se encuentran las propiedades más caras, por otro lado vemos que la distribución del estado de México es

⁴ Disponible en: <https://www.banxico.org.mx/tipcamb/main.do?page=inf&idioma=sp>

similar pero con precios menores y en Puebla sucede lo mismo, similar distribución pero con menor precio.

Jalisco es el caso más particular donde vemos que en varias ocasiones el precio ha disminuido, mientras que para el resto se percibe una tendencia ascendente en los últimos años, cuestión entendible por la inflación conocida de dicho País.

4.3 - ¿Cuál fue la evolución del precio promedio de las propiedades por ciudad?

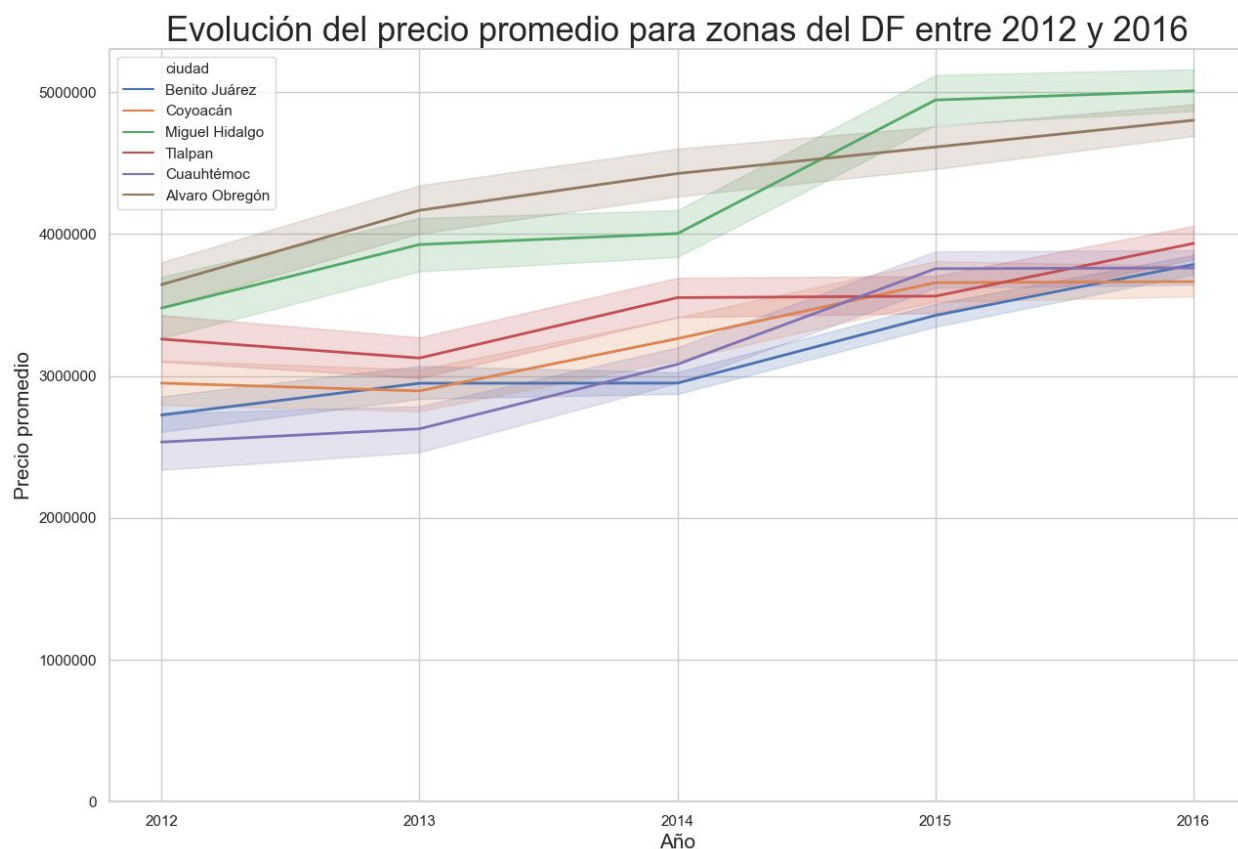


Como se diferenció anteriormente, aquí tenemos la comparación de precios de las ciudades con más propiedades, y podemos notar que Benito Juárez es la ciudad con los precios más elevados y también se puede decir que después del año 2014 fue la que más aumentó sufrió, junto a Queretaro.

En el período 2013-2014 podemos decir que Zapopan tuvo un gran aumento de precios pero después sufrió una disminución.

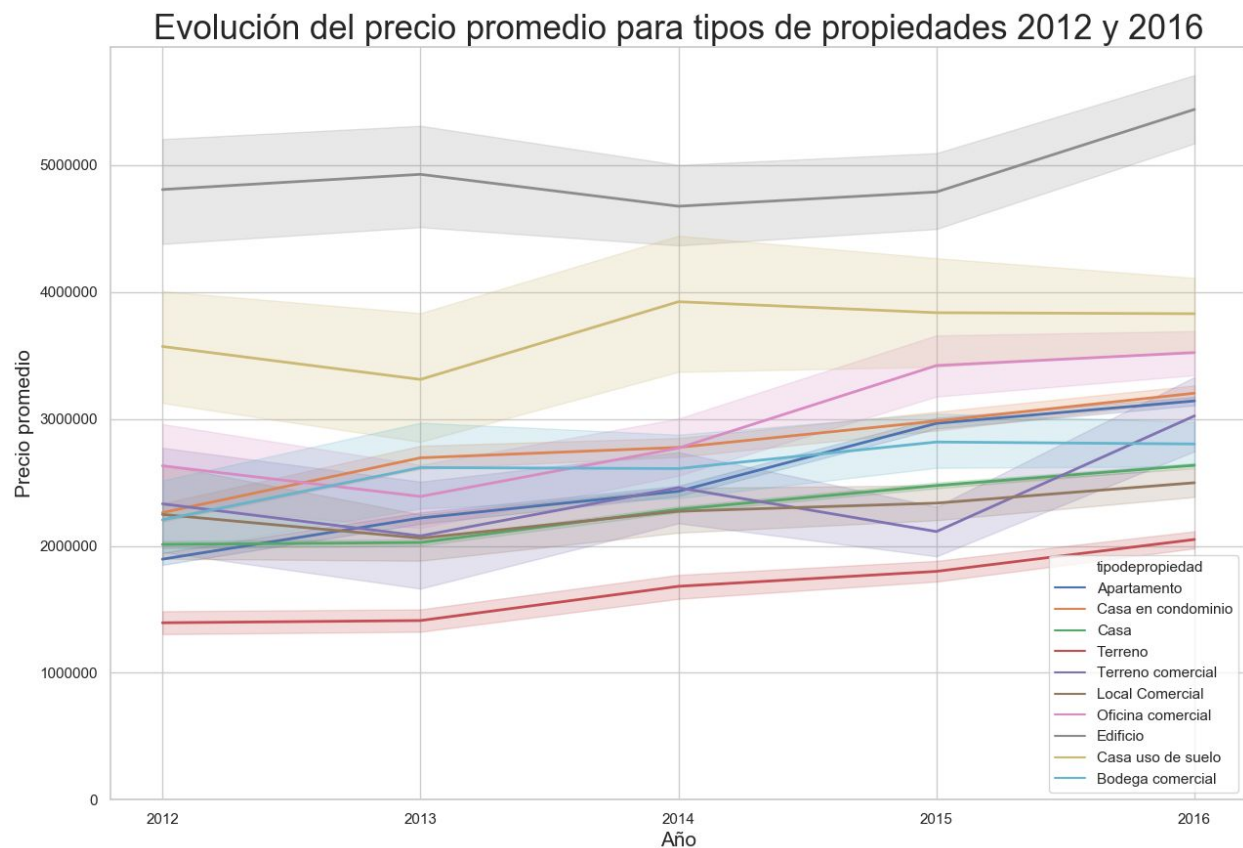
Otra cosa que podemos decir es que San Luis de Potosí es la ciudad con los precios más bajos, aunque luego del año 2015 estos aumentaron y quedó Mérida como la ciudad con precios más bajos, ya que sufrió una disminución en los mismos respecto del año anterior.

4.4 - ¿Cuál fue la evolución del precio promedio por zona de México DF?



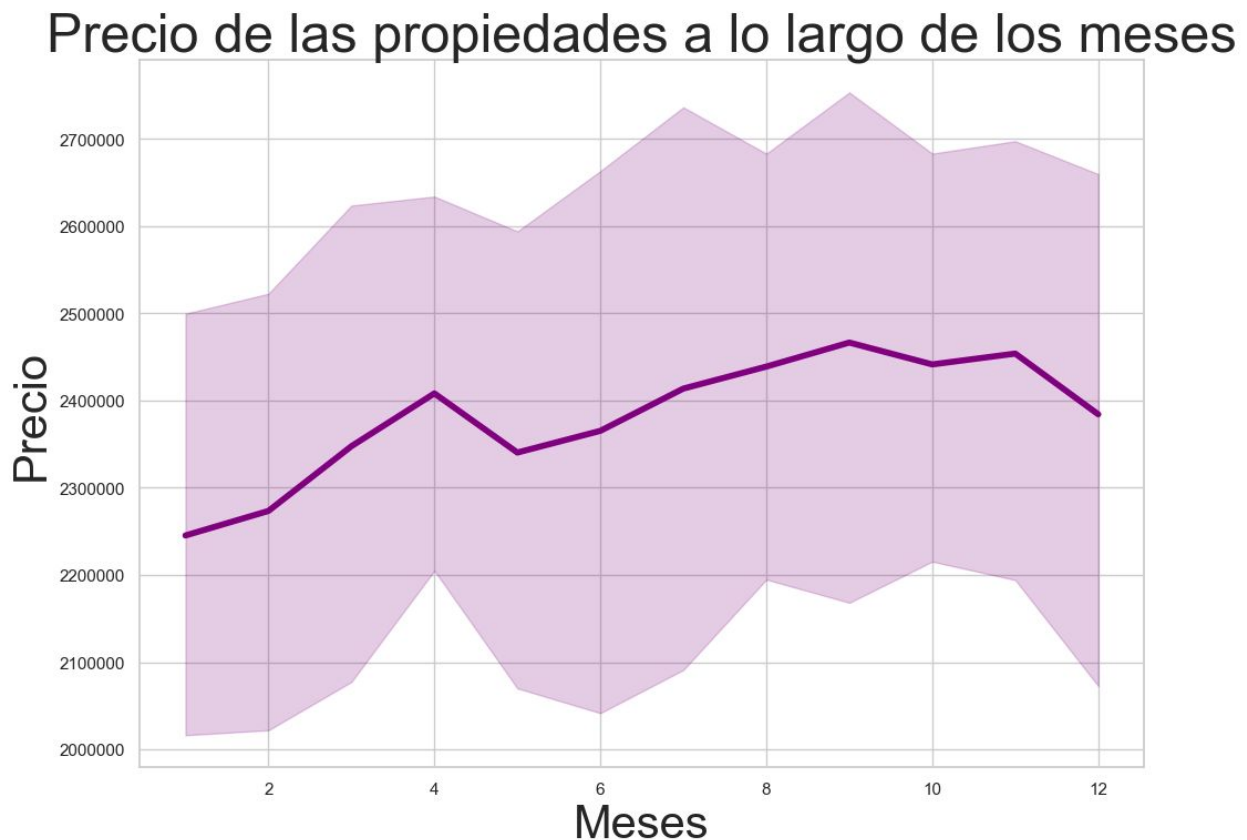
En este caso, si tomamos solo el Distrito Federal, podemos ver que las ciudades Álvaro Obregón y Miguel Hidalgo tienen un precio superior al resto, podría decirse tranquilamente que están un escalón por encima del resto, las demás se mantienen muy similares entre sí, algo a tener en cuenta es el crecimiento notable de Miguel Hidalgo entre los años 2014 y 2015 pudiendo reflejar algún evento o explosión inmobiliaria en la ciudad.

4.5 - ¿Cuál fue la evolución del precio promedio por tipo de propiedad?



Con este gráfico podemos observar como el precio de los terrenos es el más bajo y se mantiene bastante uniforme, algo que a priori uno puede pensar como razonable. Por otro lado y como contrapunto, tenemos a los edificios que tienen el valor más alto y que en particular vemos un incremento destacable en el último año analizado. Un crecimiento similar se puede observar en lo referente a las oficinas comerciales, mientras que las casas con uso de suelo son las que se diferencian estando en un segundo escalón de precios. Luego, como se percibe en el gráfico, los demás son casos similares que terminan oscilando entre los mismos valores.

4.6 - ¿Varía el precio de las propiedades de acuerdo al mes de publicación?

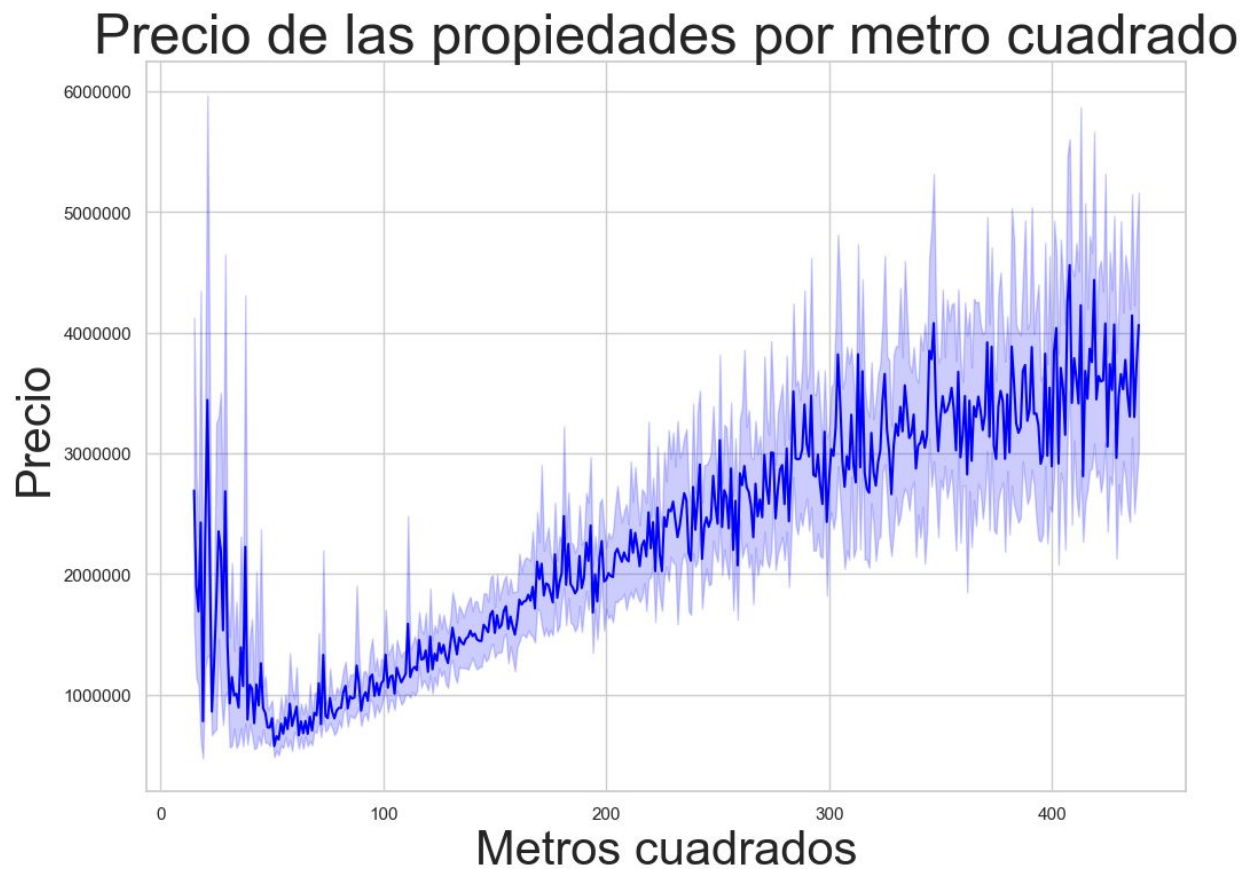


Si ahora hacemos un análisis económico sobre la evolución de precios sobre los meses del año, podemos determinar que los precios van a aumentando a lo largo del año, sufriendo un descenso en abril y en noviembre, teniendo la baja más pronunciada en el mes de enero, momento que empieza a subir el precio.

En el mes de septiembre se puede notar que los precios llegan a su punto máximo, después de una subida constante de varios meses que comienza en mayo.

Podemos determinar que una buena elección para vender propiedades es en el segundo semestre del año y, por el contrario, en cada primer trimestre del año es buen momento para comprar propiedades.

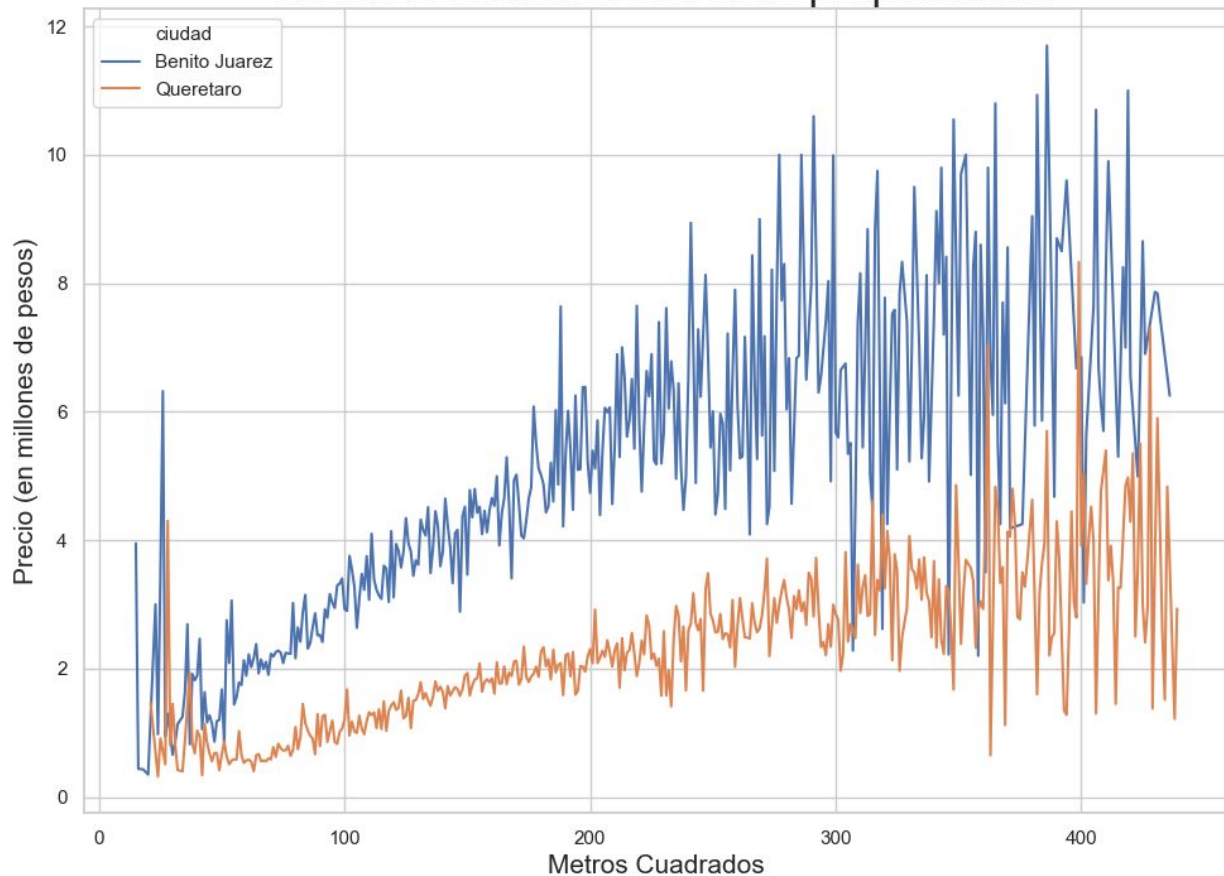
4.7 - ¿Cuál es el precio promedio por metro cuadrado?



En este caso podemos ver como el precio aumenta en una proporción coherente en relación a la cantidad de metros cuadrados, pero hay que entender que por ejemplo tenemos ciertos picos, tanto en los lugares muy chicos como así también los picos se van haciendo más volátiles cuanto mayor es el tamaño, esto se entiende por la diversidad de propiedades en relación a la cantidad de metros cuadrados que ocupan.

4.8 - ¿Cuál es el precio promedio por metro cuadrado para las dos ciudades con mayor cantidad de propiedades?

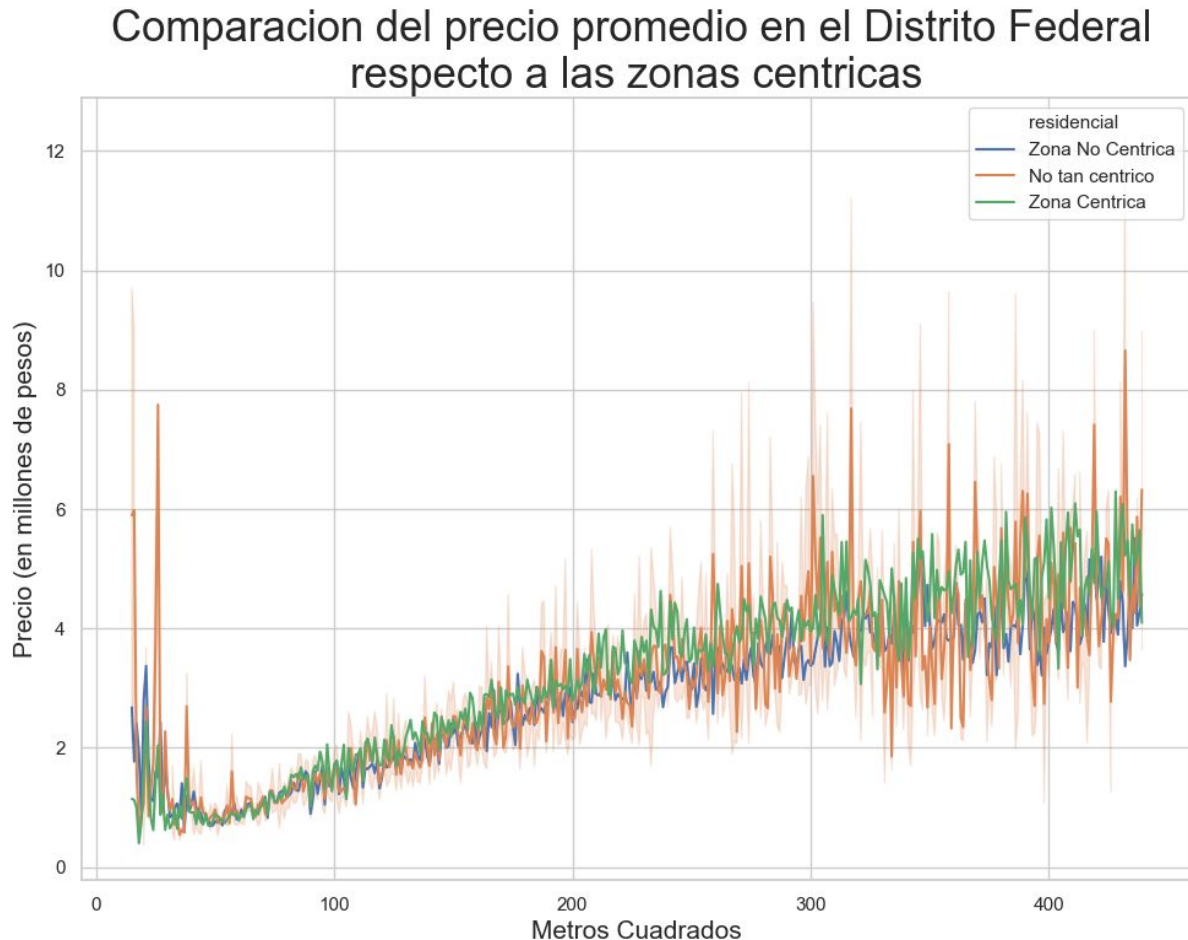
Comparación del precio promedio segun metros cuadrados en las 2 ciudades con mas propiedades



En este caso repetimos el análisis anterior, pero incluimos las dos ciudades con mayor cantidad de propiedades publicadas. Como podemos observar, y de acuerdo a los en relación a los precios, Benito Juárez (México DF) tiene precios por lo general superiores a Querétaro, salvo casos muy puntuales.

Por lo tanto, podemos ver algo que es esperable, y es que los precios dependen, además de la superficie de una propiedad, de su ubicación.

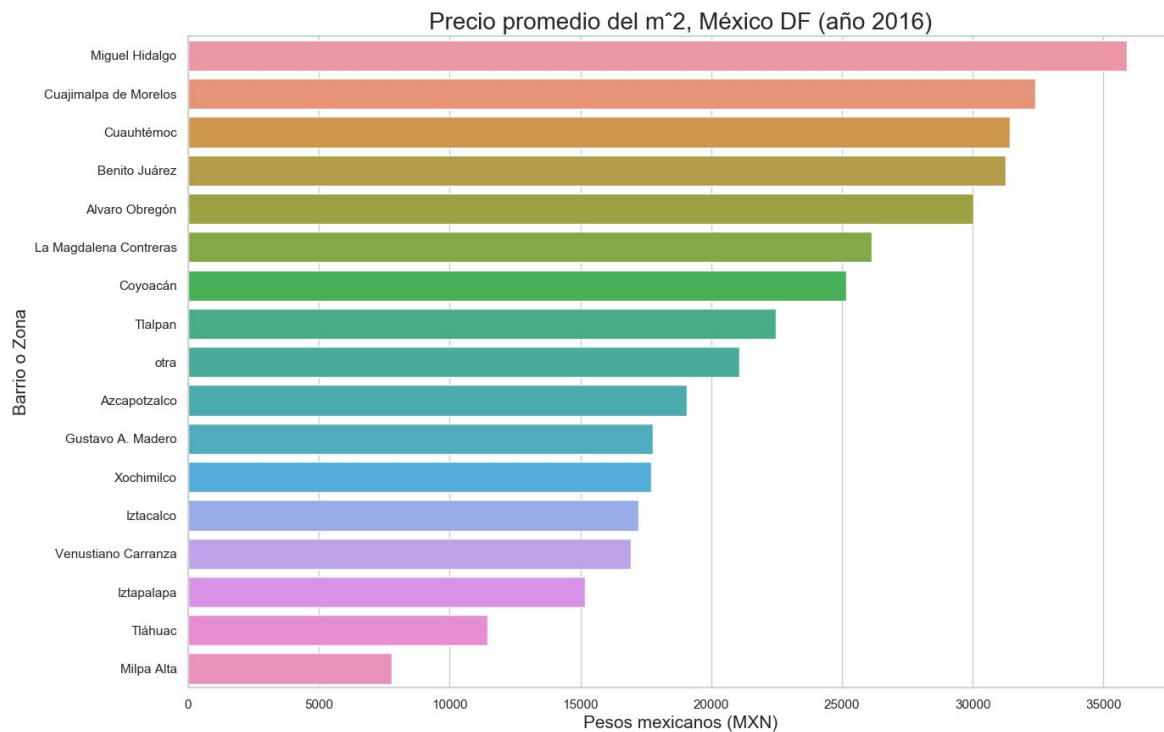
4.9 - ¿Cómo varía el precio de acuerdo a la distancia del centro de México DF?



Haciendo el análisis para responder esta pregunta, descubrimos que el 99% está relacionado al mismo anunciante, lo que nos deja muy poco margen para relacionar o aprovechar este dato, que a priori podría ser sumamente interesante, con lo cual solo decidimos mencionar el análisis básico que hicimos y la magnitud que ese anunciante tiene para este set de datos.

4.10 - ¿Cuál es el precio promedio por metro cuadrado para las zonas de México DF?

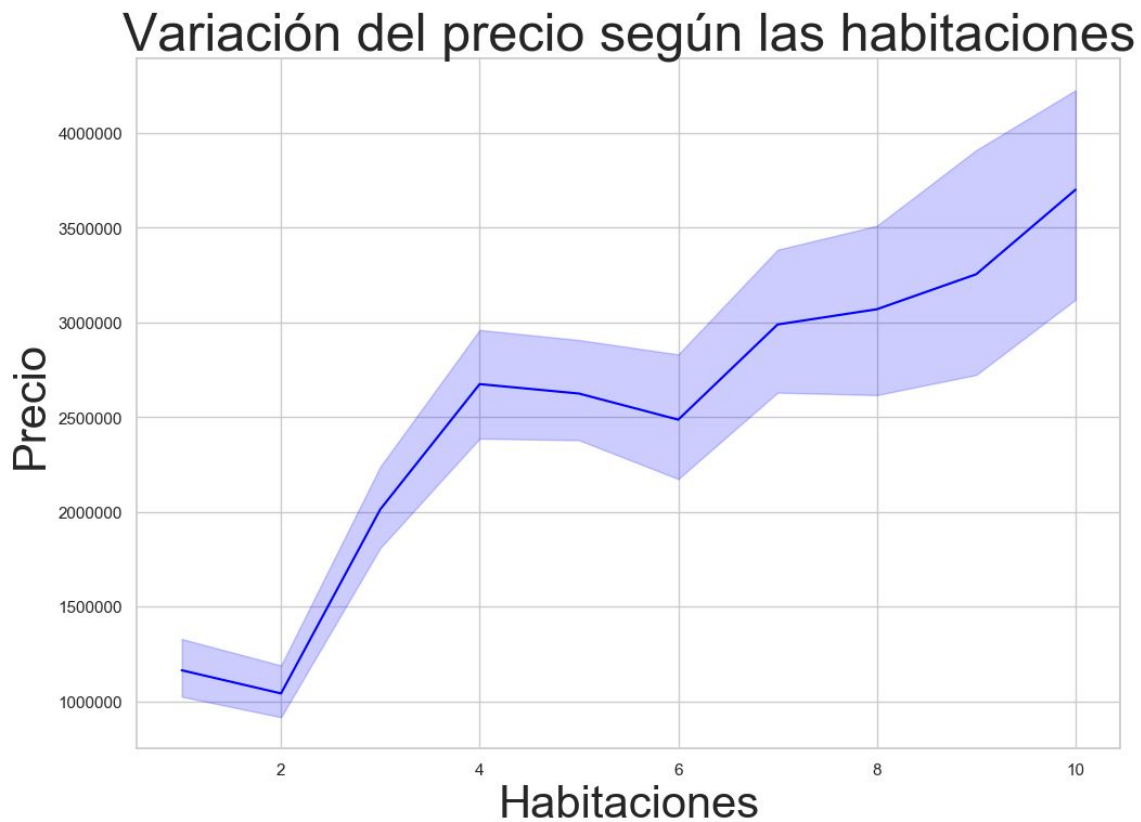
En este caso, nos preguntamos cómo varía el precio por metro cuadrado para México DF. Tomamos solamente el último año dado la variación de precios evidenciada en esta sección.



Como es posible observar, existe una marcada diferencia entre la zona más cara y el otro extremo. Contar con esta información por ciudad es importante para poder determinar el precio de una propiedad con precisión.

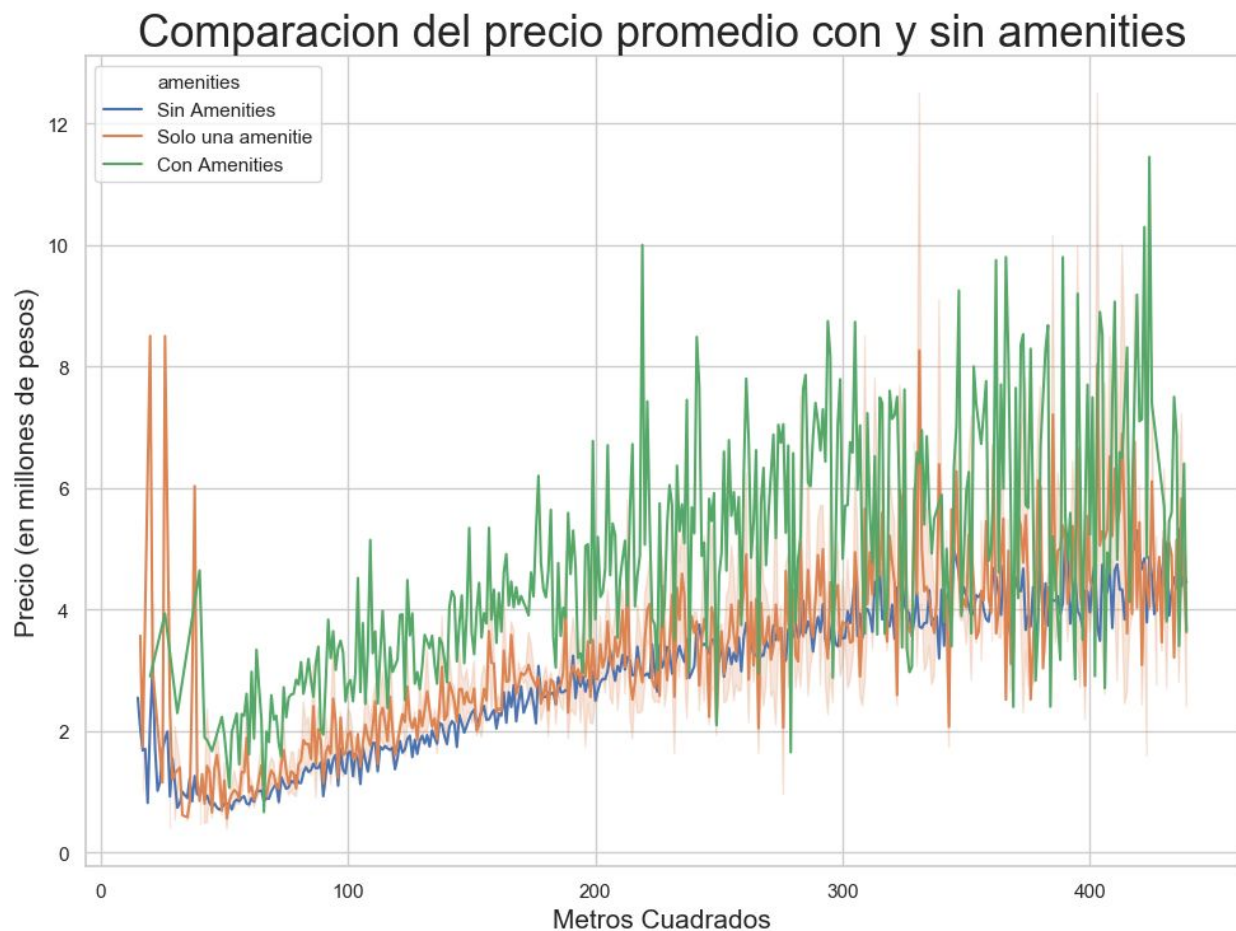
Así como se hizo la aclaración al momento de hacer el análisis geográfico, destacamos que este tipo de análisis se puede hacer con otras más ciudades o estados, pero no los incluimos para no extender innecesariamente el informe.

4.11 - ¿Cómo varía el precio del metro cuadrado según la cantidad de habitaciones?



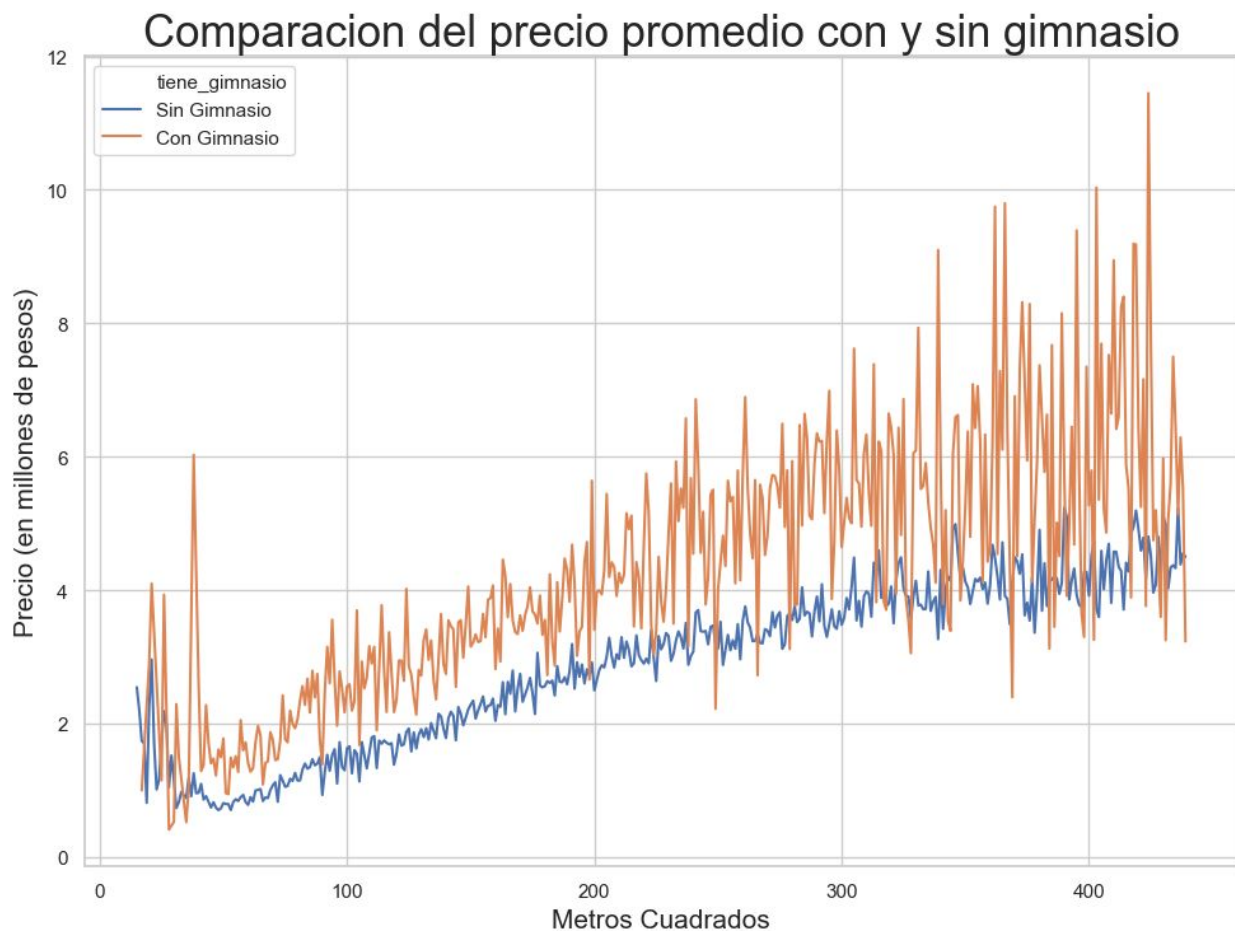
Con este gráfico podemos ver varias cosas, por un lado tenemos que los precios para los monoambientes son mayores en proporción a los apartamentos de dos ambientes, por otro lado vemos cómo a partir de este punto se da un salto grande en el precio hasta las propiedades con cuatro ambientes. Luego puede observarse que a pesar de haber un crecimiento este resulta mucho más discreto.

4.12 - ¿Cómo varían los precios de las propiedades con y sin amenities?



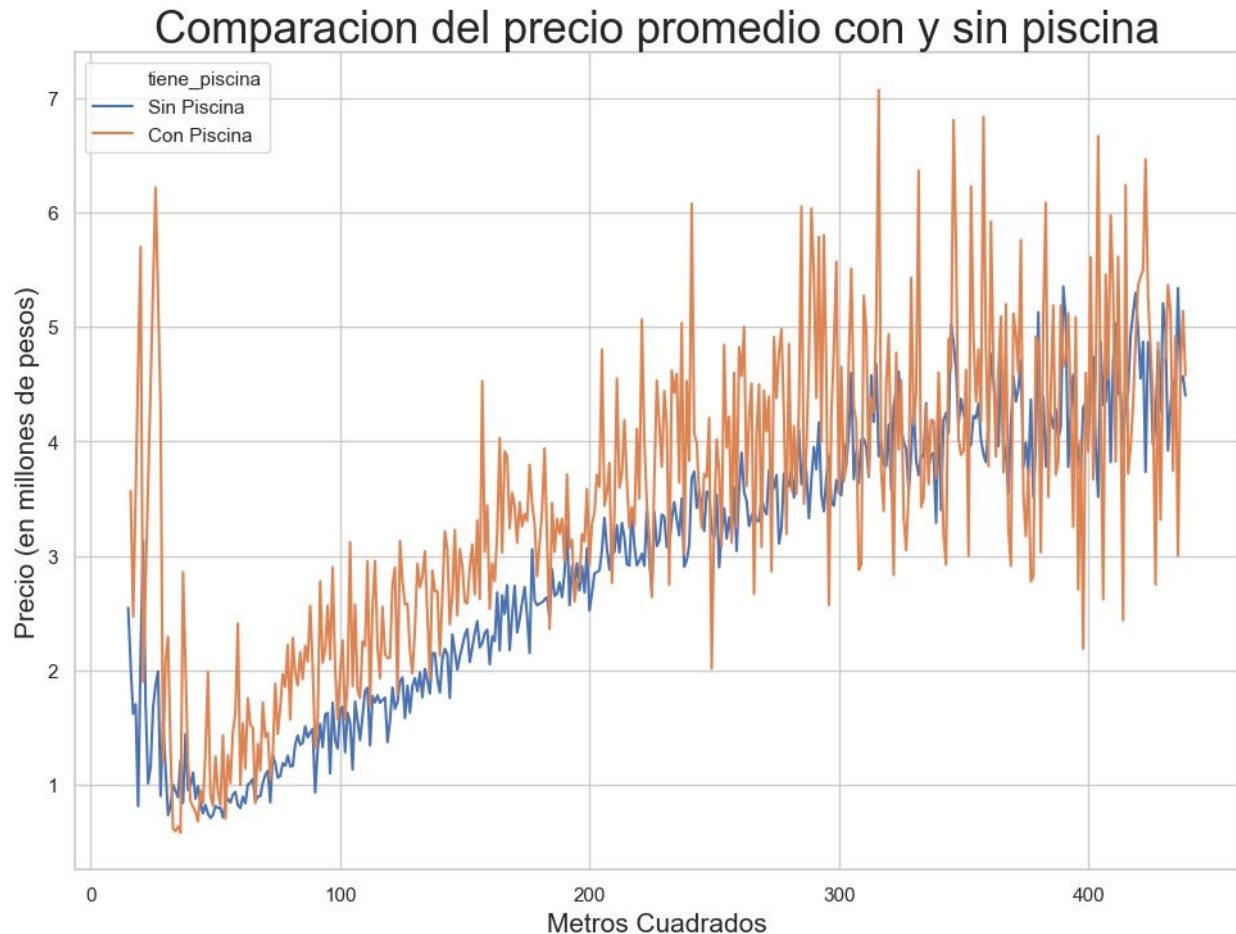
Podemos observar a través de este gráfico que los precios para las propiedades sin amenities no varía tanto como si lo hacen aquellas propiedades que poseen algún tipo de comodidad extra y más aún si la propiedad tiene dos en lugar de una, lo que nos permite inferir que hay una gran volatilidad en el precio de estas propiedades.

4.13 - ¿Cómo varían los precios de las propiedades con y sin gimnasio?



Haciendo un análisis similar pero sobre las propiedades con y sin gimnasio vemos que la diferencia de precio es notable y que además como ya habíamos visto, es mucho más volátil. De todas maneras es bueno apreciar que por más volatilidad que haya, el precio siempre tiene a ser mayor salvo en lo que entendemos son casos muy puntuales.

4.14 - ¿Cómo varían los precios de las propiedades con y sin piscina?

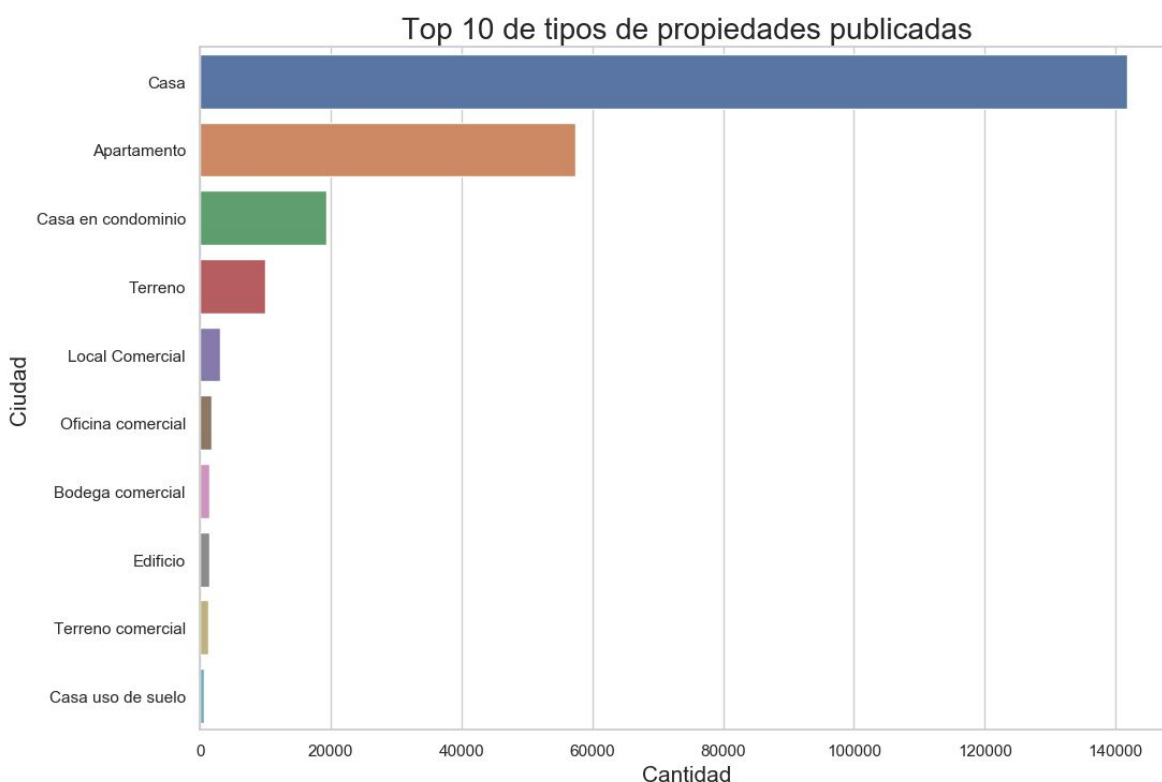


Sin embargo, para el análisis de las propiedades con y sin piscina vemos que la diferencia en el precio en general no está tan separada y que hay muchos más casos donde incluso el precio puede ser menor, esto es entendible si pensamos que los gimnasios solo se encuentran en los edificios, mientras que las piscinas pueden estar tanto en edificios como en casas, lo que a priori puede ser la razón de esta situación.

5 - Análisis sobre las características de las propiedades

A continuación, la idea es realizar un seguimiento sobre las diferentes características que forman parte de las propiedades, y ver cómo estas se distribuyen en los datos que nos han brindado. Asimismo, también buscaremos entender cuáles son las cuestiones que marcan una guía para los siguientes análisis, como para así también entender en parte como se distribuye la vida en la sociedad (siempre limitándose a los datos disponibles).

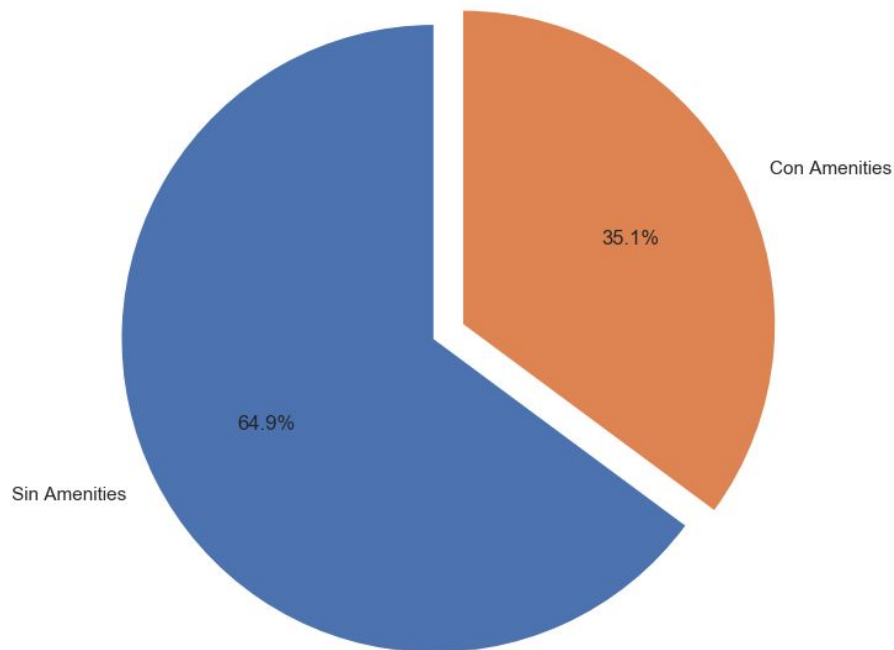
5.1 - ¿Cómo se distribuyen las propiedades por tipo?



Vemos que las publicaciones son mayormente de casas, superando por más del doble a los apartamentos, esto nos permite hacernos una idea de aquellos datos que tendrán mayor preponderancia, considerando que son las casas las que marcarán la tendencia. Del mismo modo, también vale aclarar que a pesar de tener varios apartados comerciales, estos no llegan a ser relevantes, incluso si uno quiere unirlos en una misma categoría.

5.2 - ¿Cuál es la proporción de propiedades que cuentan con amenities?

Porcentaje de propiedades que poseen Amenities (Gimnasio y piscina)

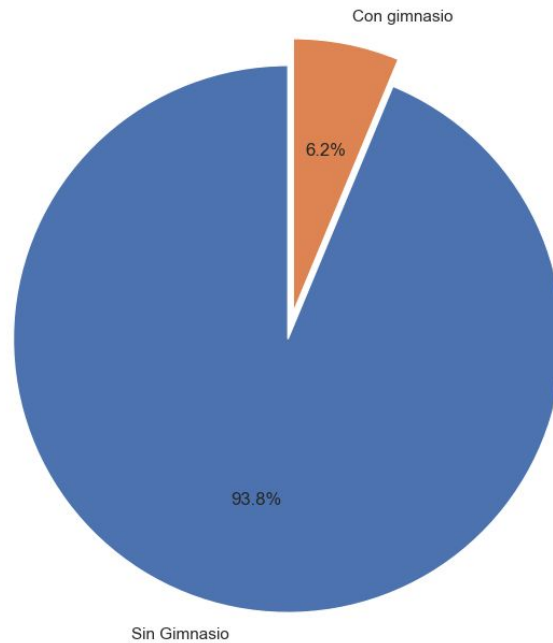


En este gráfico podemos ver cómo es la proporción entre las propiedades que no cuentan con este tipo de comodidades, contra aquellas publicaciones que poseen sea gimnasio, piscina o similares. La proporción es bastante alta, considerando que estas comodidades suelen incrementar mucho el valor tanto de compra como de alquiler. De todos modos, como veremos más adelante, si nos ponemos a ver en particular solo las piscinas y los gimnasios por separado encontramos que su porcentaje por separado es muy pequeño.

Algo que en general sucede si uno se pone a analizar cada comodidad por separado es que podrá notar que son pocas las propiedades que las poseen, pero también se entiende que en caso de unir todo, el porcentaje se incrementa como es de esperar.

5.3 - ¿Cuál es la proporción de propiedades que cuentan con gimnasio?

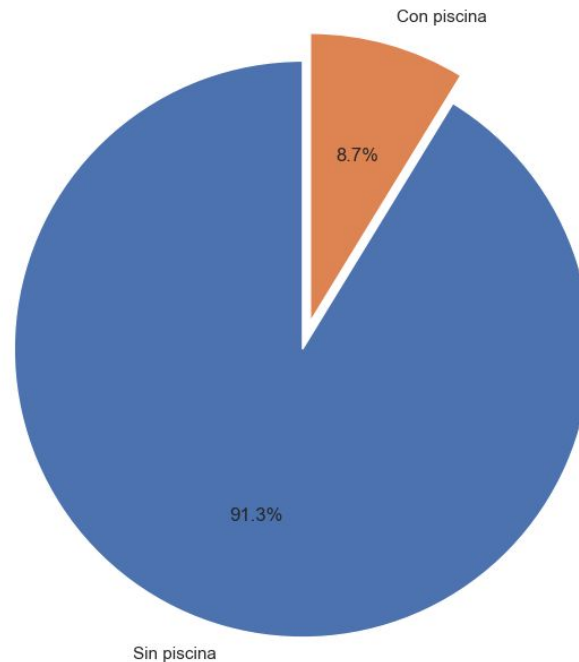
Porcentaje de propiedades que poseen gimnasio



Como podemos ver las propiedades que poseen gimnasio son muy pocas en relación al total de propiedades, esto es bastante claro considerando que los gimnasios solo se encuentran en edificios orientados a un segmento de un gran poder adquisitivo.

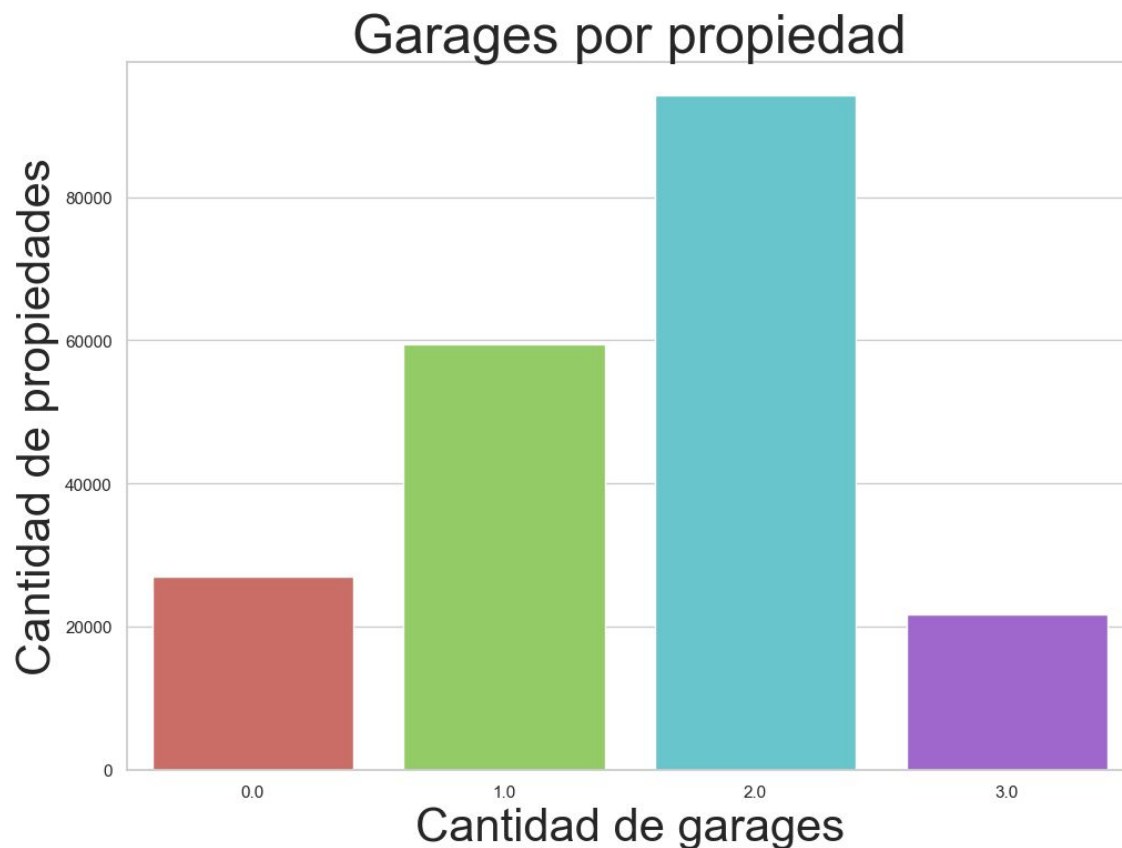
5.4 - ¿Cuál es la proporción de propiedades que cuentan con piscina?

Cantidad de propiedades que poseen piscina



Similar al caso anterior pero con una cantidad un poco mayor encontramos la situación de las piscinas, las cuáles en relación al total son pocas, si bien uno podría pensar que solo se encuentran en ciertos edificios, también tenemos que incluir casas que poseen terreno para incluir una. De todas maneras, se percibe que son pocas las propiedades que disponen de este tipo de comodidades.

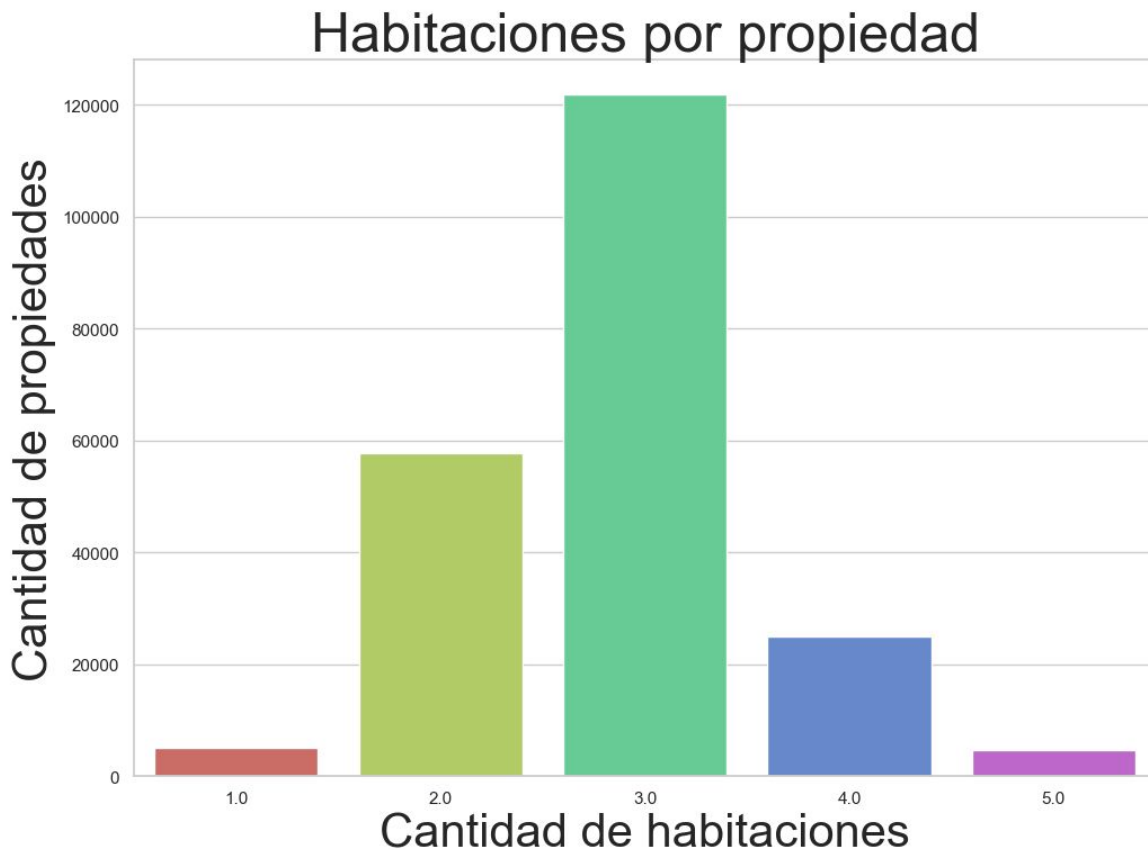
5.5 - ¿Cómo se distribuyen las propiedades de acuerdo a la cantidad de garajes que poseen?



Antes que nada vale aclarar que aportamos un análisis sobre los datos informados, pudiendo haber registros que al no tener dato no es correcto suponer algo en particular, al menos esa es la consideración que tomamos.

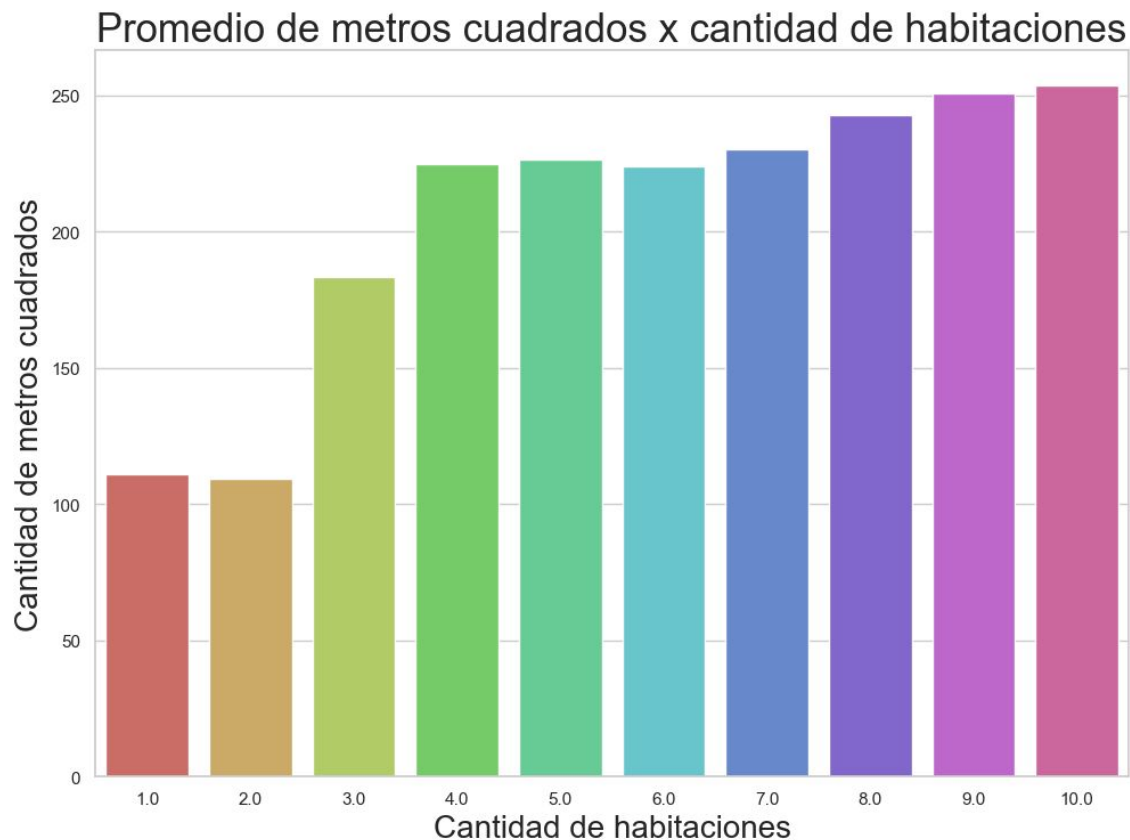
Podemos ver que la mayoría de las propiedades tienen garaje, algo que a priori uno podría suponer que no iba a ser así, pero esto no queda ahí sino que la mayoría de las propiedades tienen no uno sino dos garajes. Esto nos da la pauta de la importancia que marca esta característica dentro de las propiedades que nos brinda este set de datos, y que incide en el precio de las propiedades.

5.6 - ¿Cómo se distribuyen las propiedades de acuerdo a la cantidad de habitaciones que poseen?



Sobre la cantidad de propiedades vemos que hay un importante dominio de las propiedades con dos, tres y hasta cuatro ambientes, quedando en claro que los mono ambientes y las propiedades muy grandes no son lo que predominan en México, o al menos no en este set de datos analizados. De todas maneras, es bastante lógica esta situación por los tipos de familias promedio.

5.7 - ¿Cuál es la relación entre la cantidad de metros cuadrados y la cantidad de habitaciones?

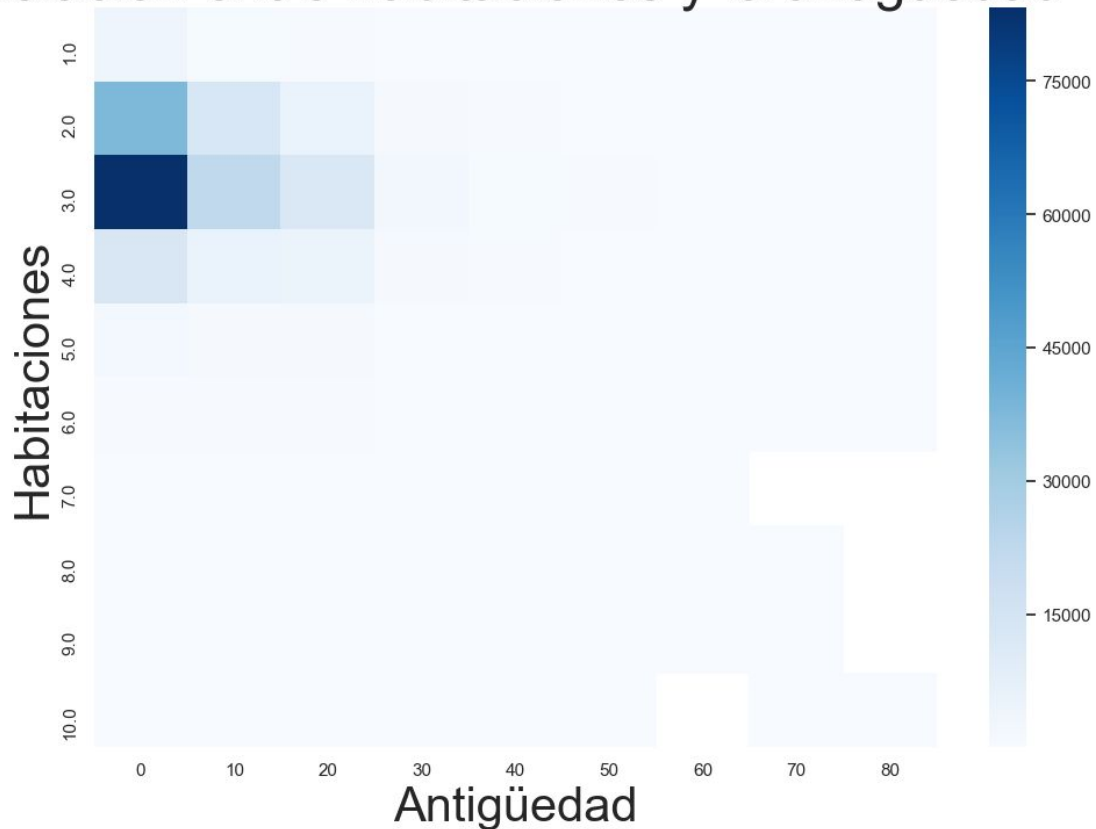


A continuación podemos observar que las propiedades con uno o dos ambientes son muy similares en su tamaño, mientras que a partir de los cuatro ambientes el crecimiento o la diferencia no es tan notable y hasta podría pensarse que se distribuyen las habitaciones de forma diferente, con ambientes más grandes o más chicos pero siempre dentro de un tamaño similar.

Además, y en general, vemos que el tamaño de las propiedades es mucho mayor a lo que uno suele estar acostumbrado en nuestro País, cuestión que se podría profundizar con una investigación interdisciplinaria.

5.8 - ¿Cuál es la relación entre la cantidad de habitaciones de una propiedad y su antigüedad?

Relación entre habitaciones y la antigüedad



Esto nos muestra que las propiedades publicadas en su gran mayoría son con muy poca antigüedad (o en el peor de los casos mal informada o sin informar) y sobre esto además se puede ver que la distribución está muy centrada sobre las propiedades de dos y tres ambientes, cuestión que ya podíamos suponer por los análisis anteriores.

Si bien podrían haber estado distribuidas de otra manera según la antigüedad esto no fue así, además vale mencionar que para el resto se ve bastante uniforme la distribución.

5.9 - ¿Cuál es la relación entre la cantidad de habitaciones y el tipo de propiedad?

Publicaciones según habitación y tipo de propiedad



Tratando de profundizar un poco más, entramos a pensar cómo varían los ambientes según el tipo de propiedad, al menos de aquellos tipos que son más predominantes en el set de datos. Ante esto podemos ver que los apartamentos son mayormente de dos o tres ambientes, con pocos monoambientes. Luego, con respecto a las casas, como era de esperarse la mayoría son de tres ambientes, esto nos remite a los análisis anteriores donde habíamos recopilado que la gran mayoría eran casas y la gran mayoría era de tres ambientes, luego hay un residual de dos y cuatro ambientes, para luego empezar a ser muy parejo en la distribución (y muy menor en la cantidad de datos disponibles).

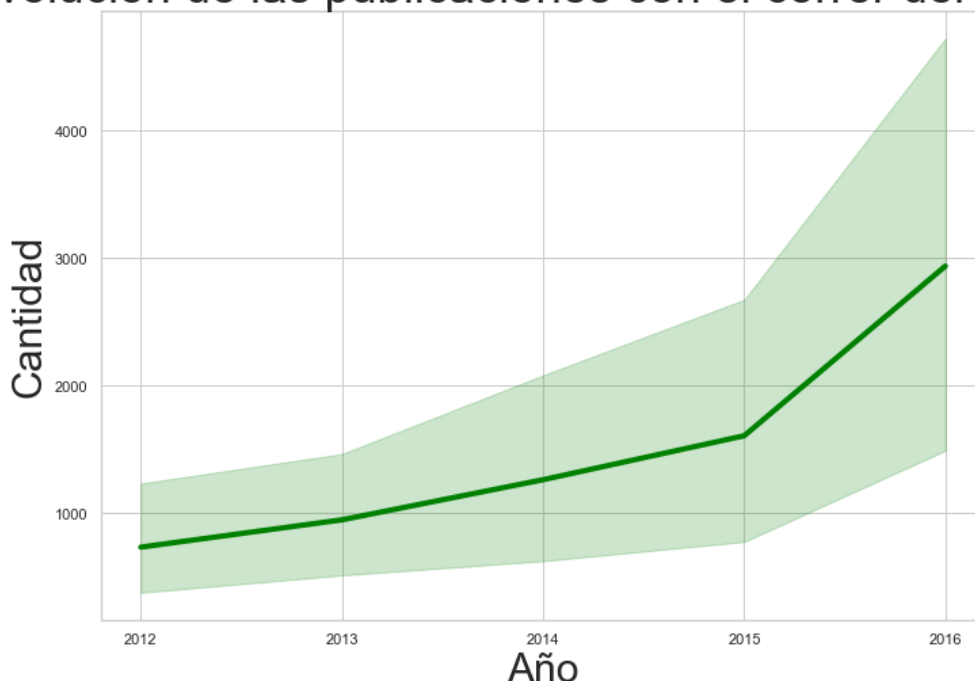
6 - Análisis sobre la actividad del sitio

Finalmente, analizamos las variables relacionadas a la actividad en el portal. Si bien no contamos con mucha información al respecto, sí podemos ver aspectos básicos sobre la cantidad de publicaciones y las fechas de publicación.

6.1 - ¿Cuál es la cantidad de publicaciones por año?

Un aspecto importante que merece estudio es observar cómo evolucionó la cantidad de publicaciones en función del tiempo.

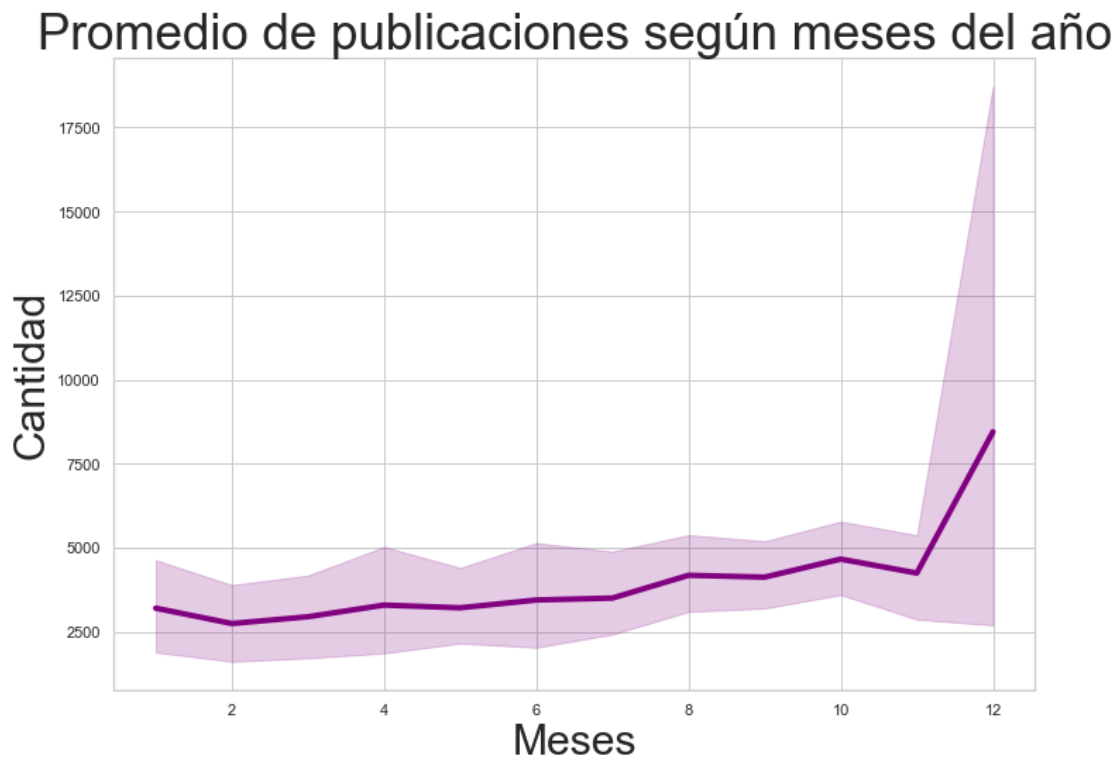
Evolución de las publicaciones con el correr del tiempo



Como se puede apreciar, el portal ha prácticamente logrado triplicar la cantidad de publicaciones en los 5 años que tenemos para analizar, con un crecimiento mucho más marcado en el último año. Sería interesante conocer si la empresa ha incrementado su promoción a través de los medios o de Internet en forma proporcional al crecimiento evidenciado en el gráfico, a fin de explicar esta tendencia.

6.2 - ¿Cuál es el promedio de publicaciones por mes?

Otro dato que resulta interesante explorar es cómo se comporta la cantidad de publicaciones respecto al mes del año.

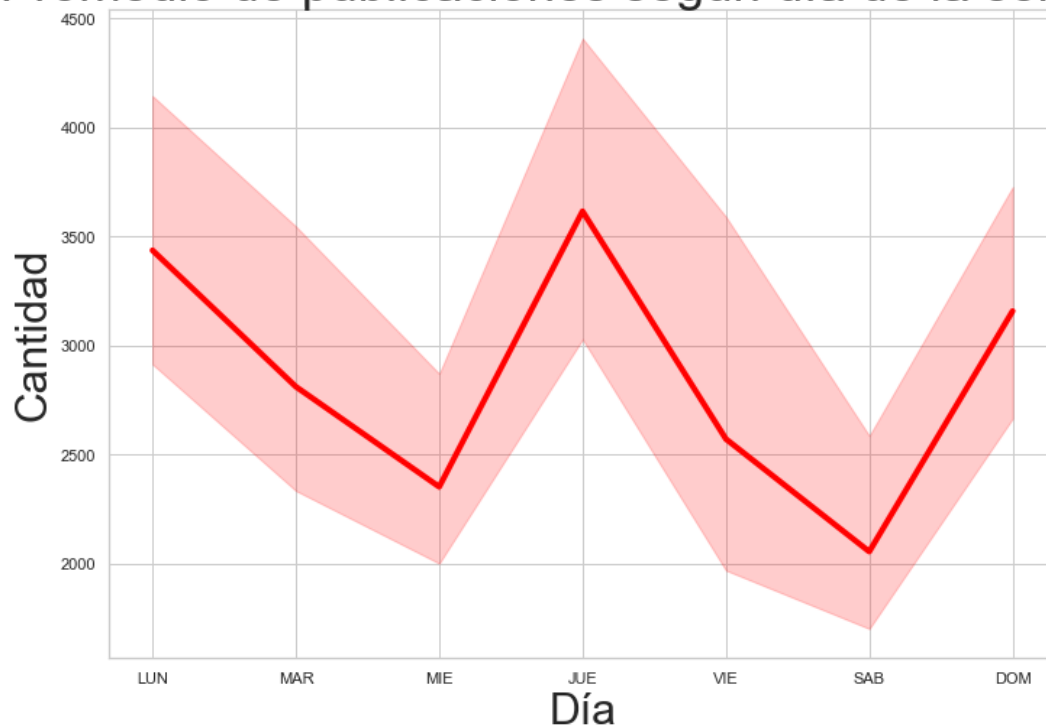


La visualización nos muestra que en febrero se da la menor cantidad de publicaciones, para ir creciendo relativamente parejo hasta octubre. En diciembre se produce un importante salto, siendo un mes que cuenta en promedio con unas 2000 publicaciones más que el resto de los meses.

6.3 - ¿Cuál es el promedio de publicaciones por día de la semana?

Un análisis similar puede realizarse con los días de la semana, viendo la relación entre cantidad de publicaciones y el día de la semana en que fueron realizadas.

Promedio de publicaciones según día de la semana

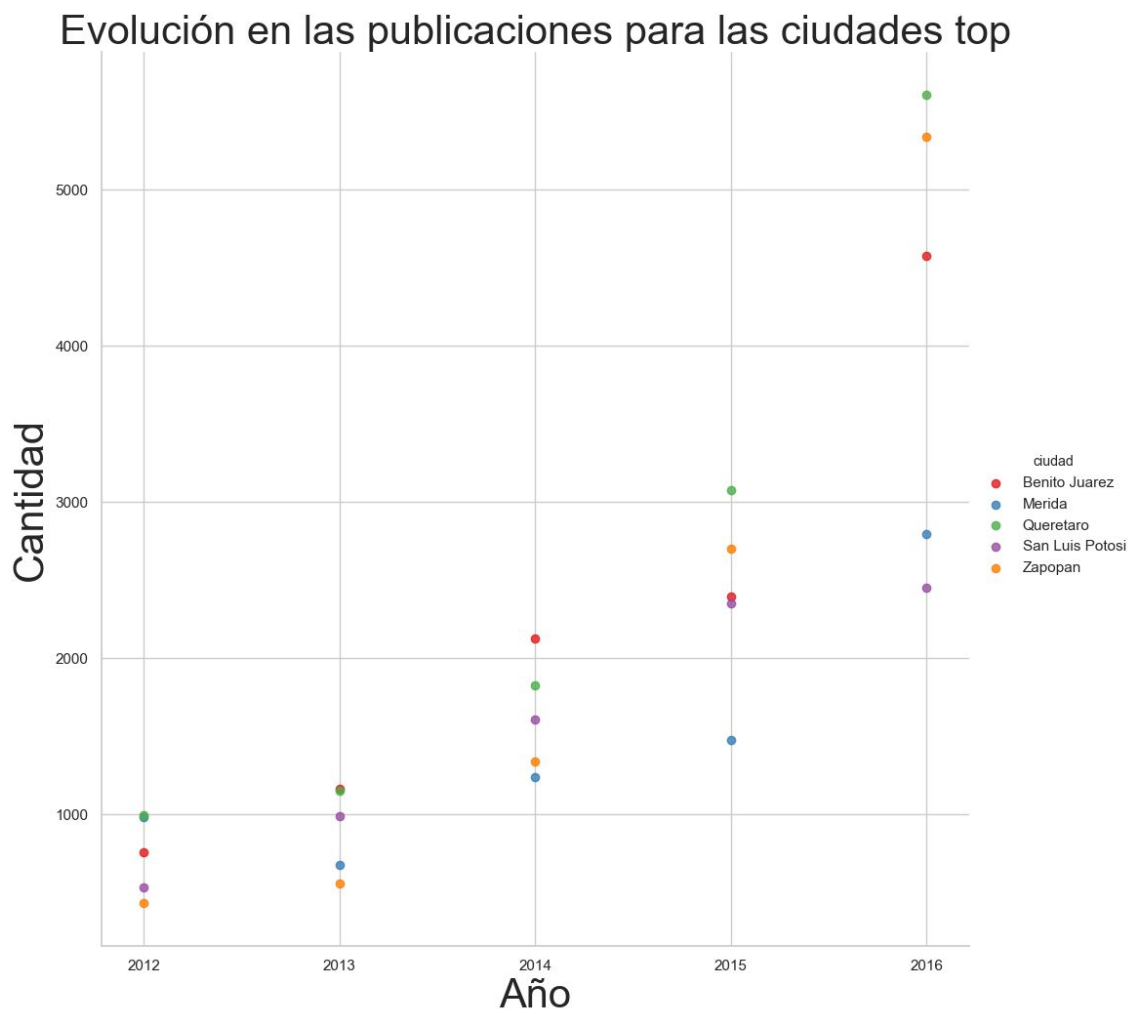


Podemos apreciar cómo hay dos tendencias bien marcadas, los días que menos se publican son los miércoles y los sábados. Los lunes y los jueves tienen la mayor cantidad de publicaciones, seguidos por el día domingo, algo que no esperábamos.

El sábado resulta ser el día con menor actividad, pero no así el domingo. Este dato puede ser aprovechado para realizar tareas de mantenimiento o cualquier otra acción que imponga carga sobre las granjas del sitio, aprovechando la menor cantidad de actividad. Sería interesante sin embargo, contrastar esta información con datos de las visitas al portal (no disponible para este análisis), para ver cómo se relacionan, y poder tomar una decisión de este tipo más informadamente.

6.4 - ¿Cómo fue la evolución en la cantidad de publicaciones para las ciudades TOP?

En esta sección veremos la cantidad de publicaciones a lo largo del tiempo, pero individualmente para las 5 ciudades con mayor cantidad de publicaciones.



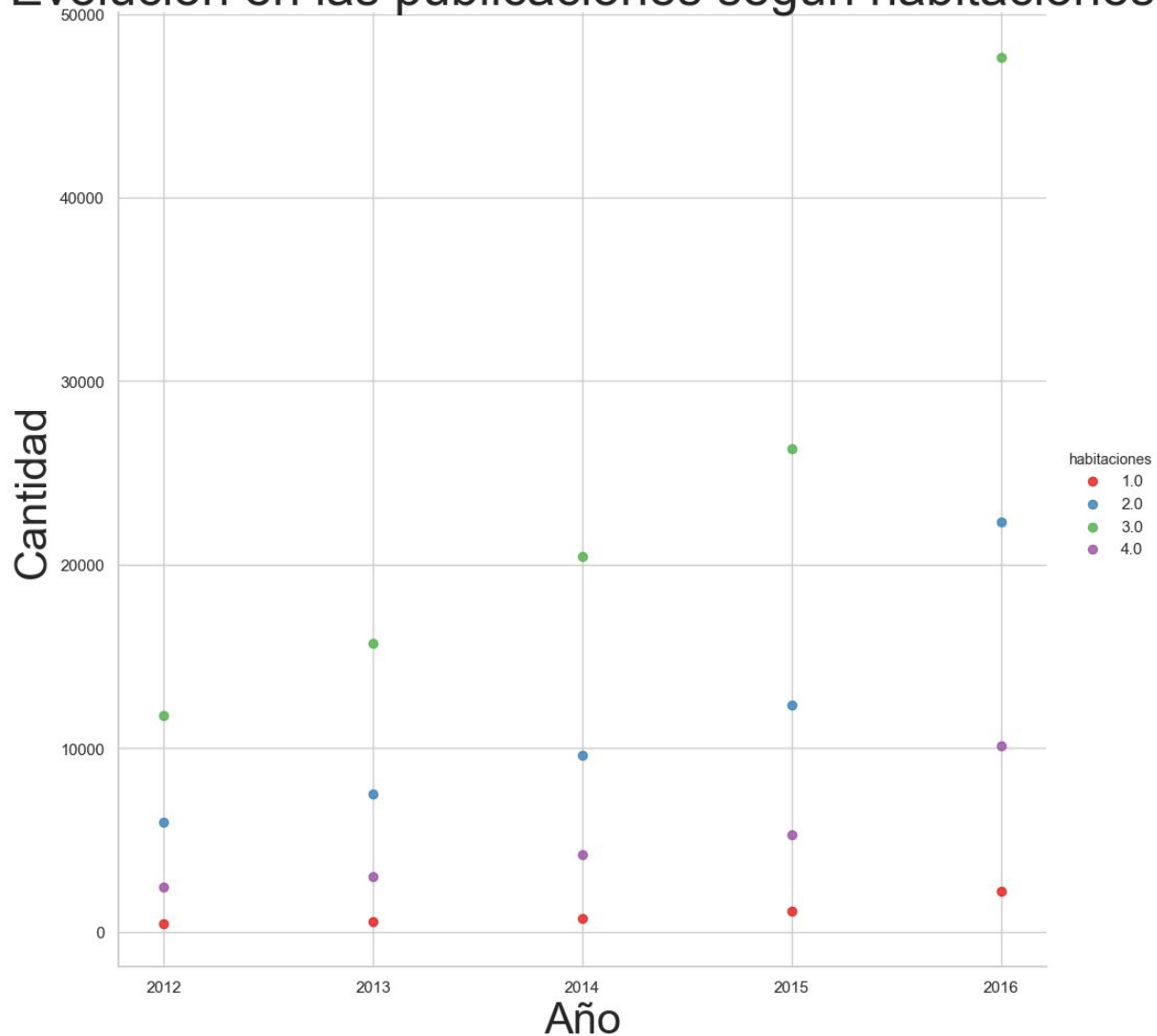
Lo que podemos evidenciar aquí es que la tendencia creciente vista en el gráfico general se repite para todas las ciudades salvo San Luis Potosí, que muestra un crecimiento muy pequeño para el último año (justo el de mayor crecimiento para el resto de las ciudades).

Sería interesante identificar la causa de este comportamiento, y analizar si se debe a algún factor fuera del alcance de Navent, o ver si podría implementarse alguna política promocional para lograr repetir la tendencia del resto del país.

6.5 - ¿Cómo fue la evolución en la cantidad de publicaciones de acuerdo al número de habitaciones?

Para cerrar esta sección, tomamos el análisis del primer punto y lo abrimos por cantidad de habitaciones.

Evolución en las publicaciones según habitaciones



Se puede ver que si bien la tendencia evidenciada anteriormente se repite al abrir el gráfico por esta variable, el mayor crecimiento se ha dado en propiedades de 3 habitaciones, superando ampliamente a al resto.

7 - Conclusiones

A lo largo de este trabajo, hemos estudiado diversos aspectos del set de datos proporcionado por Navent. Se ha analizado la ubicación geográfica de las propiedades en todo el territorio, su distribución por estados y ciudades, y para el caso particular de México DF, por zonas. Siguiendo en la misma línea, se estudiaron los precios de las publicaciones, su variación en el tiempo, tanto en general como por estado o ciudades y por tipo de propiedad. También se trabajó con el precio por metro cuadrado, comparándolo entre distintas ciudades y zonas de México DF, y finalmente la variación de los precios de acuerdo a las características de las propiedades. Otro aspecto interesante resultó ser la relación entre las distintas características, como ser la proporción que cuenta con gimnasio o piscina, así como también otros *amenities* o comodidades. Por último, estudiamos aspectos relativos a la actividad del portal, por ejemplo la evolución en la cantidad de publicaciones a lo largo del tiempo o la distribución de las publicaciones en los días de la semana y meses del año entre otros.

Respecto al análisis realizado sobre la ubicación geográfica, vimos que la gran mayoría de las publicaciones coincide con los grandes centros urbanos del país, en particular toda la zona metropolitana de la capital, México DF. Ahondando un poco, obtuvimos la distribución por estado, confirmando lo observado en el mapa. Encontramos que el orden responde a los estados con mayor participación en el producto del país, a mayor participación, mayor cantidad de publicaciones. Para los estados con mayor cantidad de publicaciones, además, abrimos el análisis en ciudades, encontrando dos aspectos interesantes: en México DF, donde contábamos con información más fina, detectamos que por lo general las publicaciones corresponden a las zonas más acomodadas. Por otro lado, en el resto de los estados la tendencia es a una gran concentración en las ciudades de cabecera (y en algunos casos una segunda ciudad).

Este análisis no se realizó para todos los estados, por cuestiones de espacio y claridad del informe, limitándose solamente a los 6 estados con mayor cantidad de publicaciones, pero en caso de contar con más información sobre las intenciones del cliente podría haberse hecho sobre todos ellos o algunos específicos en particular.

En cuanto a los precios de las propiedades, analizamos su evolución en función del tiempo, encontrando una tendencia al aumento de precios del orden de la inflación registrada en el país para el período analizado. Abrir este gráfico por estado nos permitió observar que en la capital los precios son los más elevados y tuvieron el mayor aumento en los años disponibles para estudio. La visualización por ciudades confirma la tendencia. Ver estos datos por tipo de propiedad, también fue útil para entender cómo variaron los precios, y en qué parte de la escala se encuentra cada una: los terrenos tienen los precios más bajos, mientras que en la otra punta del espectro encontramos a los apartamentos.

Otra visualización interesante fue la que nos permitió observar el precio por mes del año, encontrándose una correlación entre el mes de publicación y el nivel de precios: resulta más bajo en los meses de diciembre y enero, con tendencia al alza el resto de los meses. Para cerrar, consideramos algunos aspectos útiles a la hora de predecir el precio de una propiedad, como ser el precio promedio del metro cuadrado por ubicación, por cantidad de habitaciones, por superficie y por *amenities*/características.

Los datos presentes respecto a las características de cada publicación nos permiten determinar que la gran mayoría de las publicaciones son de casas (60%), seguido por apartamentos (25%) y condominios (8%). De estas propiedades, una cantidad no despreciable del 35% aproximadamente cuenta con algún tipo de comodidad: en el caso de gimnasios, un 6% cuenta con esta facilidad, mientras que un 9% tiene piscina. Un dato interesante que pudimos obtener también fue que la mayoría de las propiedades publicadas cuenta con 2 garajes, algo que no esperábamos. Respecto a la cantidad de habitaciones, pudimos detectar que el 50% de las propiedades publicadas responden a 3 habitaciones, y que por lo general se trata de edificaciones relativamente nuevas, de entre 0 y 10 años de antigüedad.

Finalmente, los datos relacionados a la actividad del portal son útiles para entender cómo fue creciendo la cantidad de publicaciones, cuáles son los meses donde se publica más y qué días de la semana son los que registran mayor uso. Asimismo, abrir la evolución de la cantidad de publicaciones por ciudad y por cantidad de habitaciones nos permite entender dónde está enfocado el grueso de actividad y encontrar casos especiales: la cantidad de publicaciones casi no aumentó en el último año en San Luis Potosí, y lo que más creció fueron los anuncios de propiedades con 3 habitaciones.

A modo de cierre, podemos decir que el análisis realizado nos ha permitido entender la estructura del set de datos y sus principales características, así como también obtener una noción del mercado inmobiliario en México, y de la actividad en el portal de ZonaProp. De esta manera, un análisis de este tipo puede utilizarse para mejorar la oferta de productos, las prestaciones o características de la plataforma, o bien detectar zonas geográficas donde la compañía podría expandir su operación, entre otras aplicaciones.

Asimismo, nos gustaría agregar que el estudio de un set de datos de estas características, que puede obtenerse en forma muy sencilla, por ejemplo en base a logs de actividad, resulta una herramienta poderosísima a la hora de tomar decisiones estratégicas para la compañía, en una amplia variedad de campos, a fin de administrar de la mejor forma posible los recursos disponibles.