

Trabajo Práctico 1

Análisis Exploratorio de Datos

[75.06/95.58] Organización de Datos
Segundo cuatrimestre 2019
Grupo 25

Nombre	Padrón	E-mail
Diaz, Nicolas Andres	101369	diaz.nicolasandres@gmail.com
Mastricchio, Facundo Rodrigo	100874	facundomastriichio@gmail.com
Bustamante, Jorge Tomás	89603	jbustamante@fi.uba.ar

Repositorio: <https://github.com/tomasBustamante/zonaprop-mexico-12-16>

Índice

1. Introducción	2
2. Información general de los datos	2
3. Análisis de cantidad de publicaciones a nivel geográfico	3
3.1. Cantidad de publicaciones según provincia	3
3.2. Cantidad de publicaciones de cada tipo de propiedad por provincia	4
4. Análisis a nivel de precio	5
4.1. Precio promedio de propiedades por provincia	6
4.2. Precio promedio por tipo de propiedad	7
4.3. Precio promedio por tipo de propiedad por provincia	8
5. Análisis a nivel temporal	8
5.1. Cantidad de publicaciones por año	9
5.2. Cantidad de publicaciones por año por tipo de propiedad	9
5.3. Cantidad de publicaciones según año y mes	10
5.4. Precio promedio de propiedades según año y mes	11
5.5. Precio promedio de tipo de propiedades según año	12
5.6. Provincias con mas publicaciones por año	13
5.7. ¿En qué mes se realizan más publicaciones?	13
5.8. ¿Qué ocurre en diciembre?	14
6. Análisis de propiedades según atributos	17
6.1. Propiedades según cantidad de baños	18
6.2. Propiedades según cantidad de habitaciones	18
6.3. Relación entre precio y metros cubiertos de algunos tipos de propiedad	19
6.4. Precio promedio de propiedades según cantidad de habitaciones y baños	21
6.5. Precio promedio de propiedades según gimnasio, piscina y SUM	21
7. Otras exploraciones	23
7.1. Histograma de precios (Casas y departamentos)	23
7.2. Relaciones entre cantidad de propiedades por ciudad	23
7.3. Propiedades con más publicaciones	24
7.4. Cantidad de publicaciones de casas y departamentos a lo largo de los años	25
7.5. Predominancia de casas sobre departamentos	25
7.6. Análisis precio por mt2	26
8. Análisis de Locales (Vía publica y Centro comercial)	28
8.1. Histograma precio de un local (Calle) según cercanía a Shopping	28
8.2. Cantidad de locales publicados por año	29

1. Introducción

Se propone analizar en el presente informe los datos obtenidos del registro histórico de publicaciones en www.zonaprop.com.mx, un sitio de compra y venta de inmuebles. Para ello, **Navent** nos proporcionó acceso a los datos a través del archivo **train.csv**.

El objetivo principal de este informe es poder realizar un análisis exploratorio donde a medida que se exploren los datos se vayan encontrando tanto las preguntas como las respuestas a las incógnitas que fueron surgiendo.

Finalmente, entre lo descubierto en el análisis exploratorio y los items marcados, se busca obtener un listado de *insights* aprendidos sobre los mismos y con ellos realizar un aporte a la empresa Navent de manera que el análisis realizado proporcione datos útiles para mejorar su servicio o generar algún tipo de estadística.

2. Información general de los datos

Lo primero a analizar es la estructura general de los datos proporcionados, para comenzar a tener una idea de qué es lo que se tiene y qué se puede hacer con ello. Se observa que:

- Los datos corresponden al período comprendido entre los años 2012 y 2016 en México.
- Se tratan de 220000 registros con 22 atributos.

Los atributos son:

- **id**: id numérico para identificar la propiedad.
- **título**: título de la propiedad publicada.
- **descripcion**: descripción de la propiedad publicada.
- **direccion**: dirección de la propiedad.
- **ciudad**: ciudad de la propiedad.
- **provincia**: provincia donde esté localizada la propiedad.
- **lat**: latitud.
- **lng**: longitud.
- **tipodepropiedad**: tipo de propiedad (Casa, departamento, etc.)
- **metrostotales**: metros totales de la propiedad.
- **metroscubiertos**: metros cubiertos de la propiedad
- **antigüedad**: antigüedad de la propiedad.
- **habitaciones**: cantidad de habitaciones.
- **garages**: cantidad de garages.
- **banos**: cantidad de baños.
- **fecha**: fecha de publicación.
- **gimnasio**: si el edificio o la propiedad tiene gimnasio.
- **usosmultiples**: si el edificio o la propiedad tiene SUM.
- **piscina**: si el edificio o la propiedad tiene piscina.

- **escuelascercanas**: si la propiedad tiene centros comerciales cerca.
- **centroscommercialescercanos**
- **precio**: valor de la propiedad en pesos mexicanos.

3. Análisis de cantidad de publicaciones a nivel geográfico

La idea de este apartado radica en analizar como están distribuidas las publicaciones según la provincia en que se encuentran y así identificar algún tipo de patrón de publicación, como por ejemplo cuál es la provincia con mayor cantidad de publicaciones, o cuál es el tipo de propiedad más publicado dentro de cada provincia.

3.1. Cantidad de publicaciones según provincia

Se realiza un gráfico para visualizar de manera rápida y a grandes rasgos la cantidad de publicaciones por provincia. No se encontraron mayor inconvenientes puesto que los atributos de los registros utilizados estaban en 'buen estado', es decir, el porcentaje de nulos era muy bajo de tal manera que no generaban discrepancias durante el procesamiento.

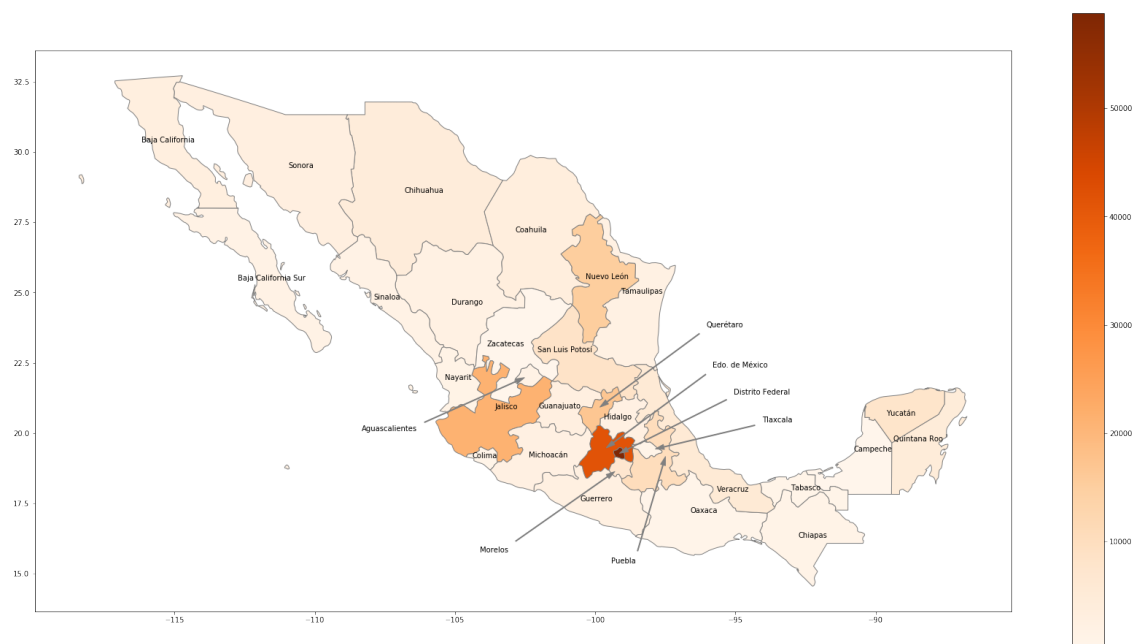


Figura 1: Mapa coroplético de la cantidad de propiedades por provincia

Se puede observar tanto en la figura 1 como en la 2 que la mayor concentración de publicaciones de propiedades provienen del Distrito Federal, Estado de México, Jalisco, Querétaro, Nueva León y Puebla. Por otro lado, el Distrito Federal y el Estado de México son las provincias con mayor cantidad de publicaciones, lo cual guarda relación con sus poblaciones.

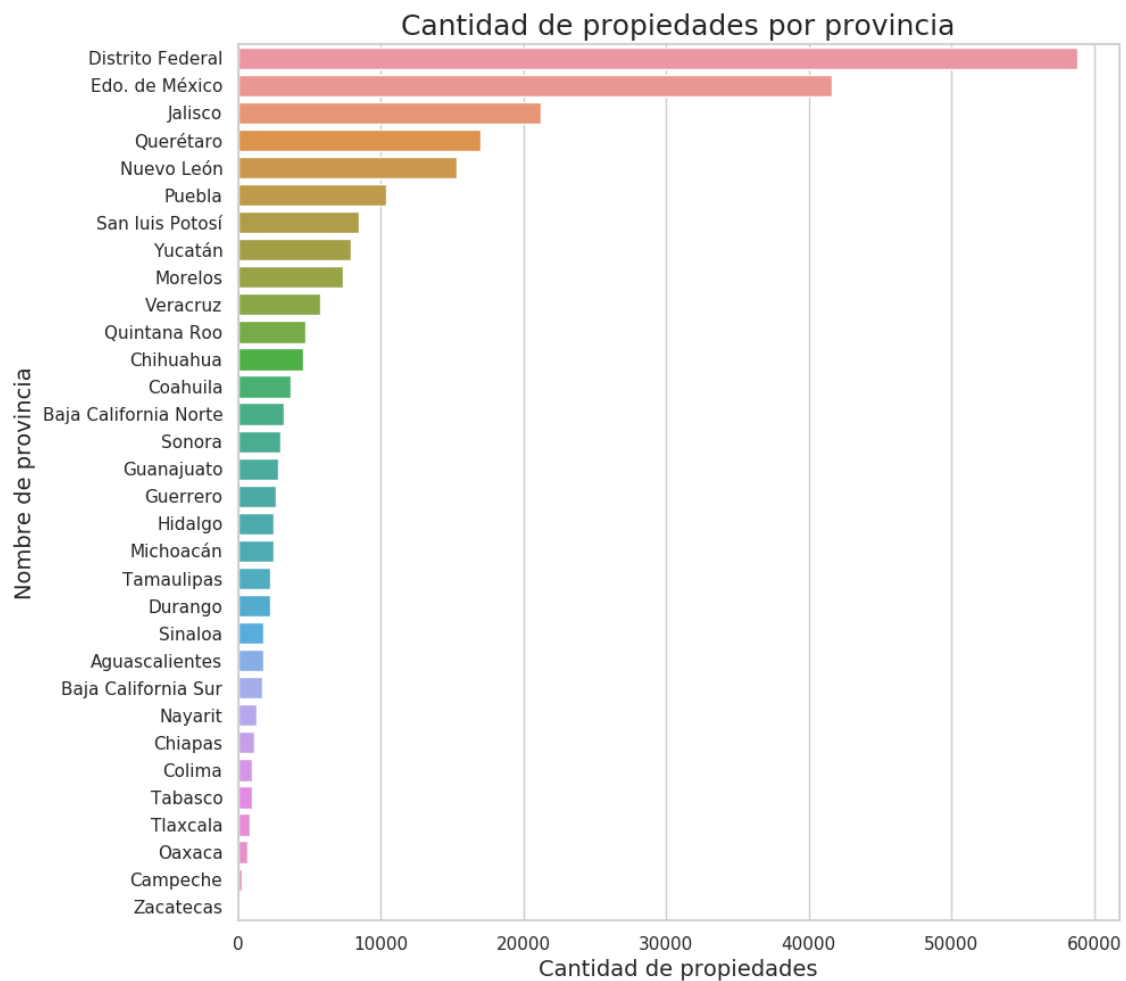


Figura 2: Cantidad de propiedades por provincia

3.2. Cantidad de publicaciones de cada tipo de propiedad por provincia

Se analiza un poco mas en profundidad cual es el tipo de propiedad predominante dentro de cada provincia. Para ello se realiza un gráfico similar a la figura 1, pero en donde cada provincia tendrá una división proporcional de acuerdo a la cantidad de publicaciones de cada tipo de propiedad.

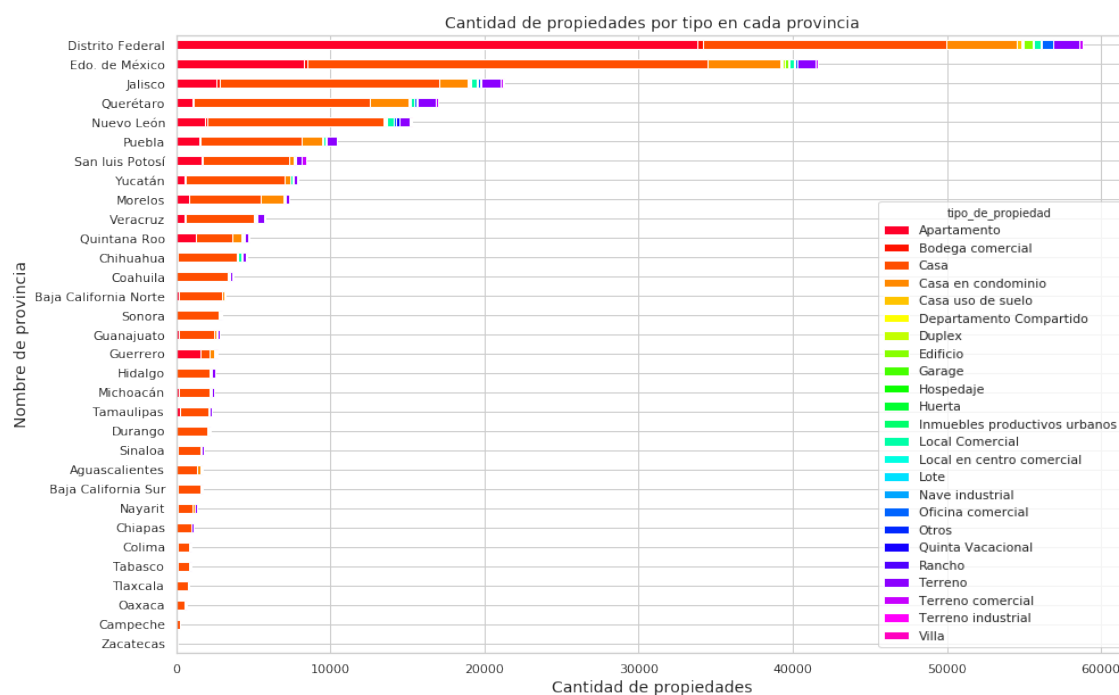


Figura 3: Cantidad de propiedades por tipo por provincia

Como podemos ver en la figura 2, en casi todas las provincias el tipo de propiedad predominante es Casa por una amplia diferencia, seguido por los Apartamentos y las Casas en condominio. La excepción se produce en el Distrito Federal, en donde los Apartamentos poseen la mayor cantidad de publicaciones, seguida por las Casas y luego por las Casas en condominio. Esto es, de nuevo, algo que uno podría llegar a sospechar ya que el Distrito Federal está ubicado dentro de la zona céntrica de México y por lo tanto es sensato pensar que los tipos de propiedades predominantes sean aquellos del tipo habitable.

4. Análisis a nivel de precio

Esta sección tiene como objetivo analizar el precio promedio tanto como a nivel provincial, a nivel del tipo de propiedad y dentro de cada provincia según el tipo de propiedad. La intención es poder averiguar cuál es el tipo de propiedad mas costosa, cuál o cuáles son los tipos de propiedades mas costosas dentro de cada provincia, y cuáles son las provincias con un precio promedio de propiedad elevado.

4.1. Precio promedio de propiedades por provincia

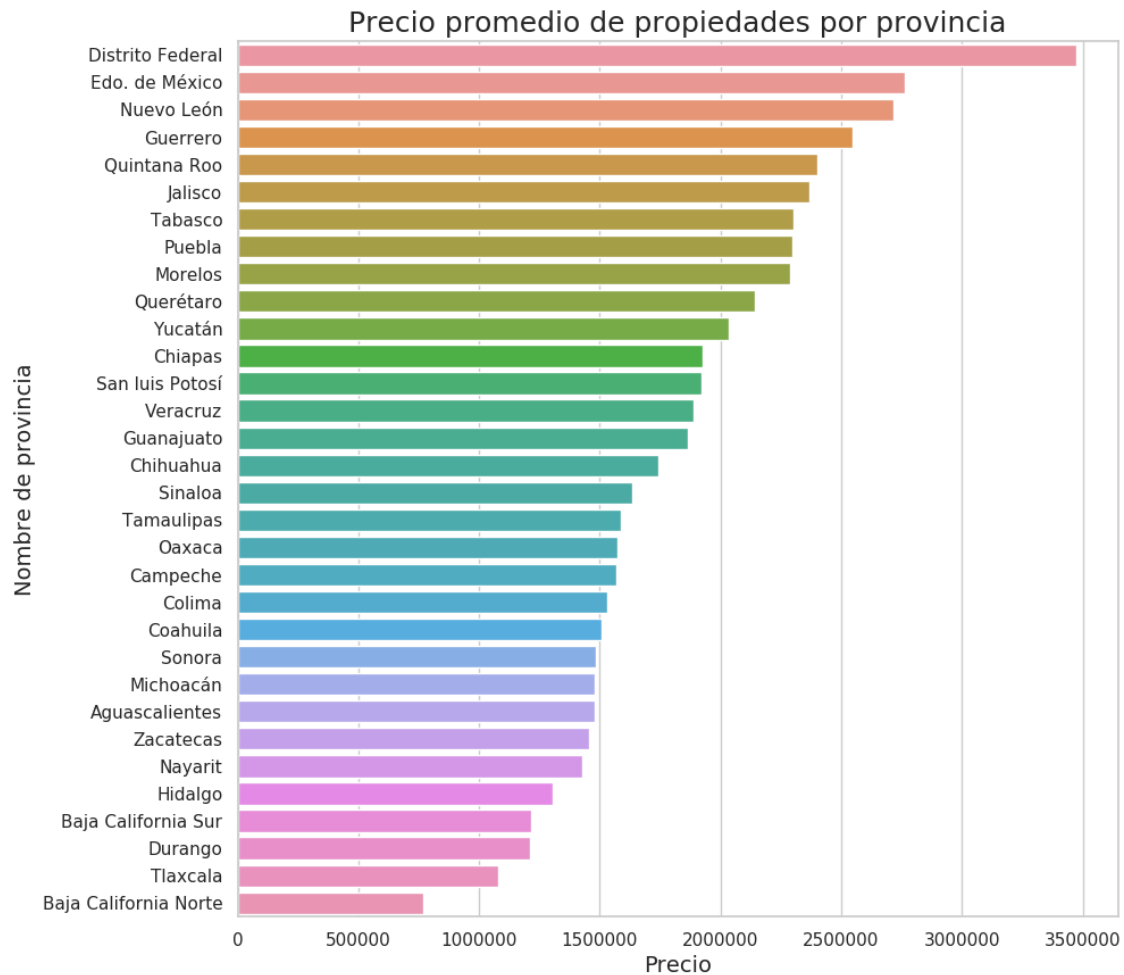


Figura 4: Precio promedio de propiedades por provincia

Comparando con el gráfico de tipo de propiedades por provincia, en general se puede apreciar que aquellas provincias con mayor cantidad de publicaciones también tienen un precio promedio de propiedades elevado. No obstante, Tabasco y Guerrero no poseen tantas propiedades publicadas como, por ejemplo, el Distrito Federal o Estado de México, pero se encuentran dentro de las provincias con un mayor precio promedio por propiedades.

4.2. Precio promedio por tipo de propiedad

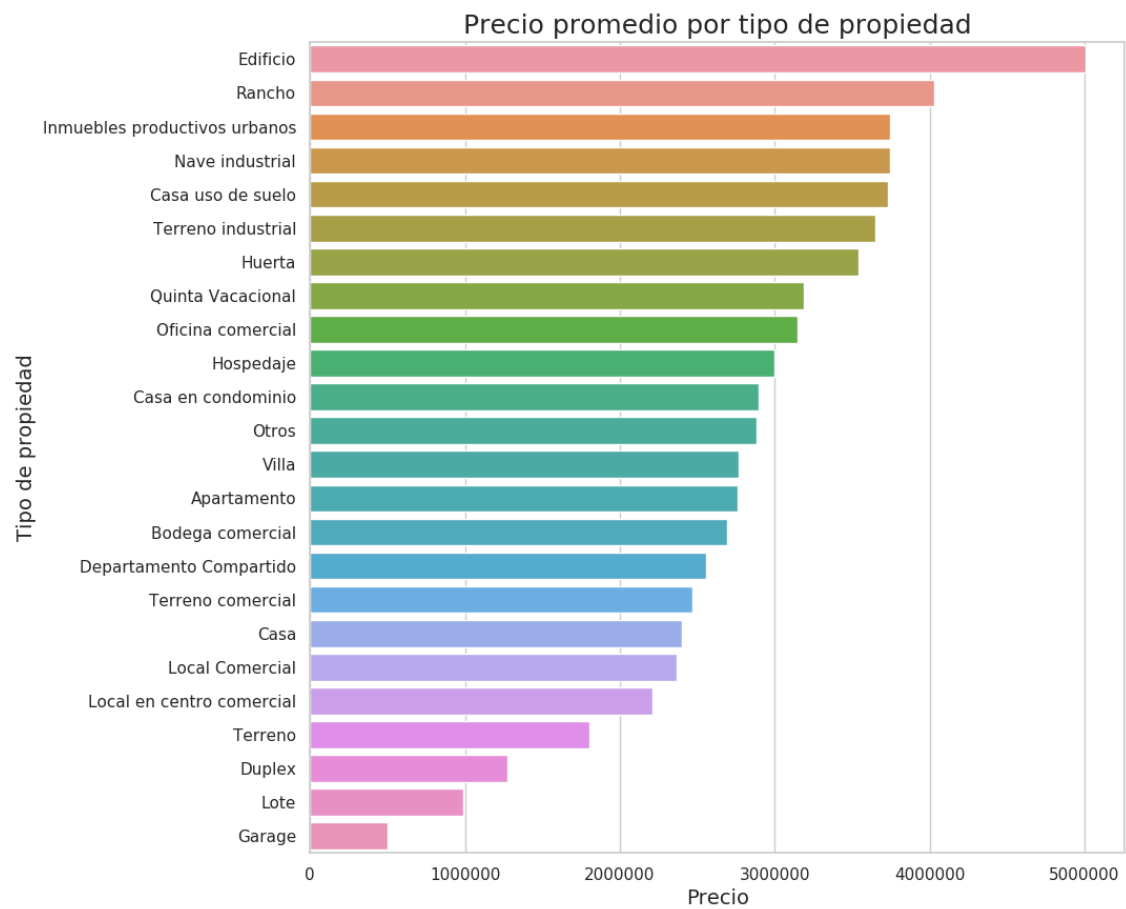


Figura 5: Precio promedio por tipo de propiedad

4.3. Precio promedio por tipo de propiedad por provincia

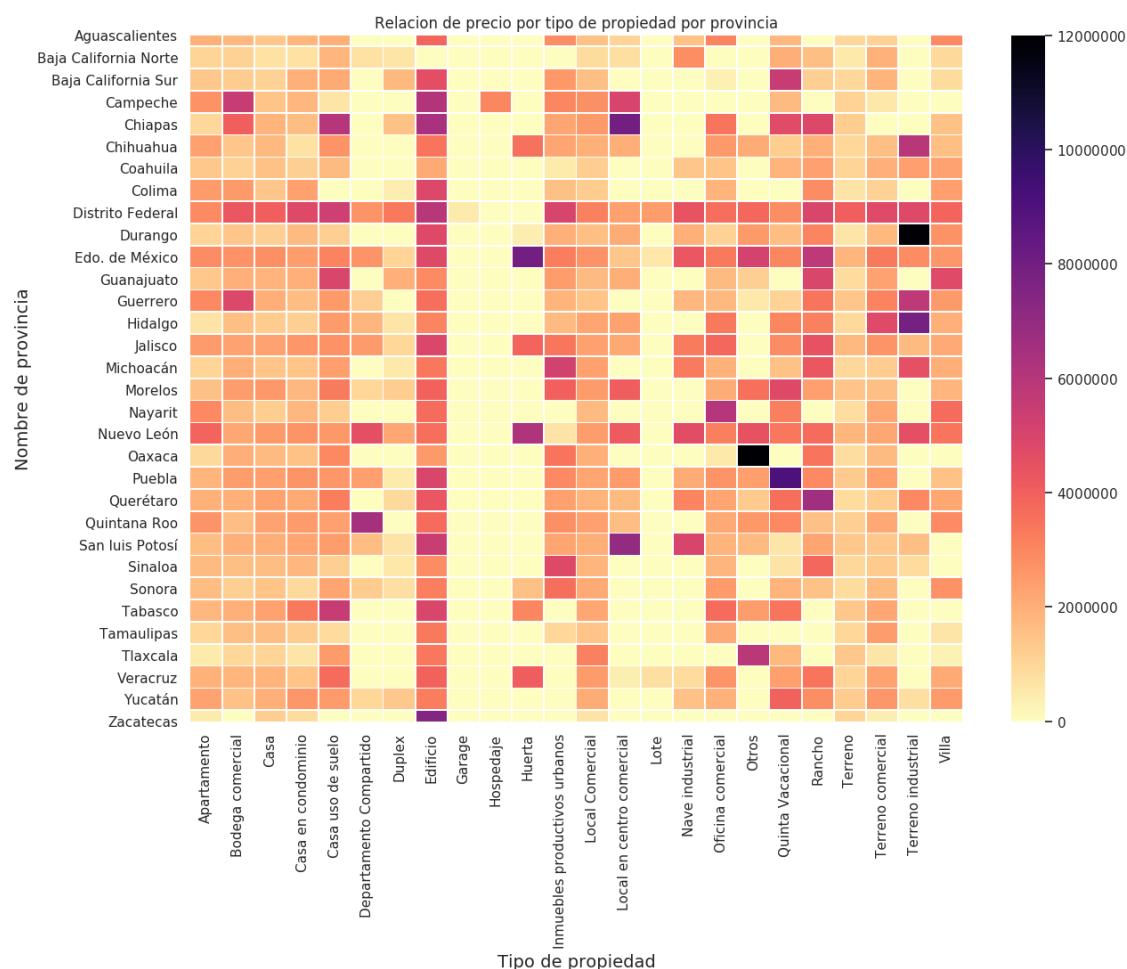


Figura 6: Precio promedio por tipo de propiedad por provincia

En general, se aprecia que los precios promedios de los distintos tipos de propiedades dentro de cada provincia son similares. De todas formas, en algunos casos se puede ver que el precio es mas alto en algunas zonas para un tipo de específico de propiedad. Por ejemplo, un edificio en Zacatecas puede llegar a ser mas costoso que en el Distrito Federal; caso similar sucede con un rancho en la provincia de Querétaro, o un terreno industrial en Durango.

5. Análisis a nivel temporal

Esta sección tiene como objetivo analizar el volumen de la cantidad de publicaciones. La idea es poder ver cuál o cuales fueron los tipos de propiedades con mayor incremento de publicaciones, cómo fueron distribuidas las publicaciones a lo largo de los meses para cada año y si hay meses en donde se presentó una mayor cantidad de publicaciones en general. Además, se desea ver cómo fue variando el precio promedio de propiedades en general tanto de manera interanual, como dentro de un mismo año, y en relación a esto pero solo de manera anual, ver cómo fue variando el precio promedio de cada tipo de propiedad para poder observar si la variación de los precios fue homogénea o si se presentan particularidades o anomalías en ciertas propiedades.

5.1. Cantidad de publicaciones por año

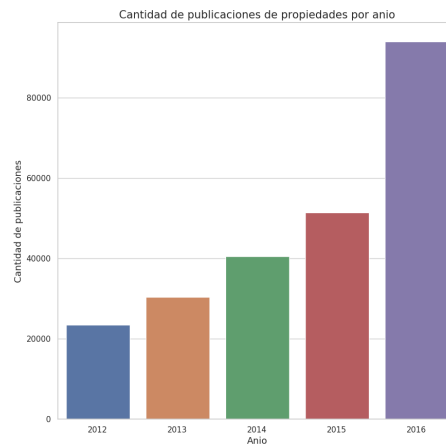


Figura 7: Cantidad de publicaciones por año

En la figura se puede ver un continuo aumento de la cantidad de publicaciones en ZonaProp. Esto puede deberse, entre otras cosas, a que el sitio fue haciéndose más conocido gradualmente.

5.2. Cantidad de publicaciones por año por tipo de propiedad

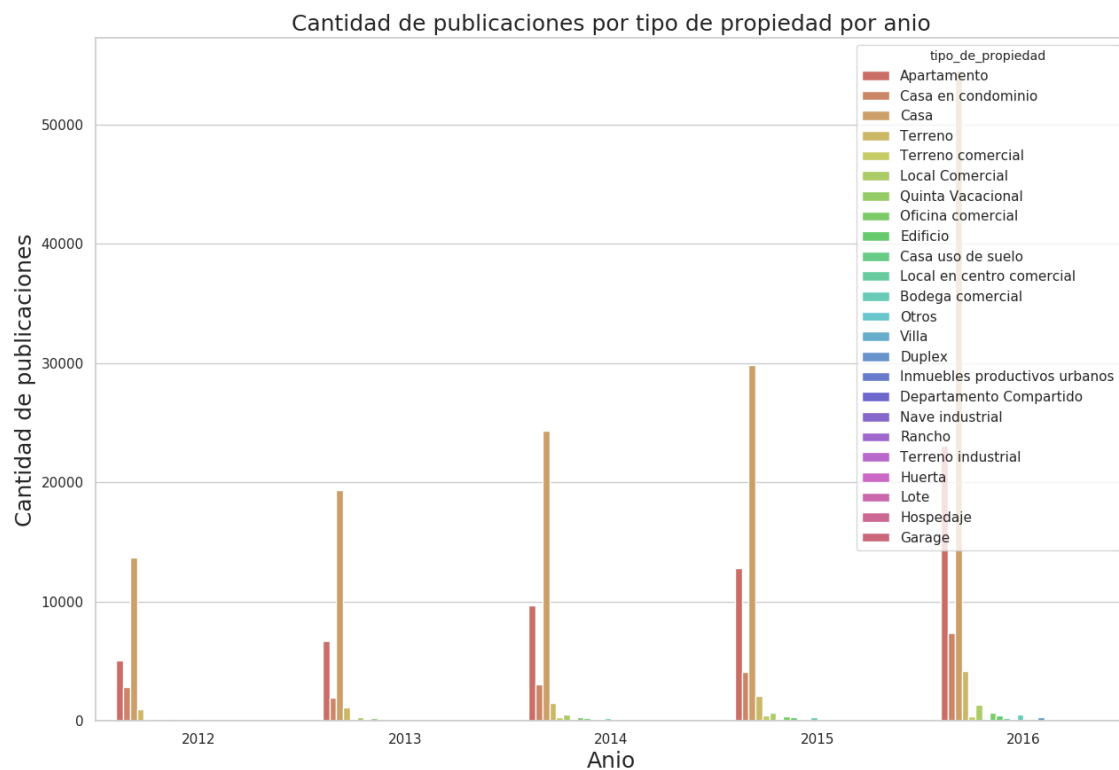


Figura 8: Cantidad de publicaciones por año por tipo de propiedad

La figura nos permite observar a grandes rasgos, que las casas, las casas en condominio y los apartamentos fueron las propiedades que mas se destacaron en cuestión de publicaciones, mostrando un amplio crecimiento dentro del periodo analizado. Debe hacerse una mención a los terrenos, que pese a no ser del mismo tipo 'habitable' que las mencionadas, también presenta un aumento sostenido a lo largo de los años.

5.3. Cantidad de publicaciones según año y mes

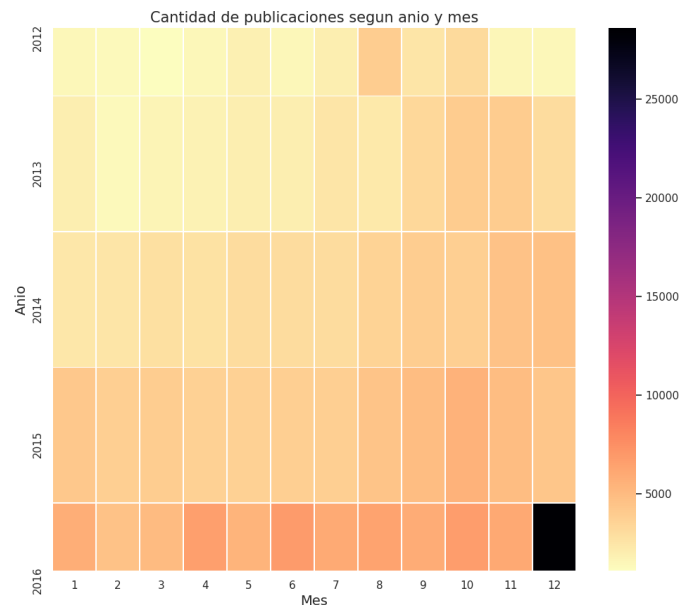


Figura 9: Cantidad de publicaciones por año por tipo de propiedad

Profundizando un poco el análisis realizado en la sección de cantidad de publicaciones por año, durante los meses de agosto, septiembre, octubre de cada año, la cantidad de publicaciones fueron levemente mayor en comparación con los demás meses de un mismo año. Esto puede deberse a un aumento de publicidad por parte de ZonaProp para aumentar su visibilidad. Por otro lado, durante el mes de diciembre de 2016, la cantidad de publicaciones fueron ampliamente superiores a las del resto del año.

5.4. Precio promedio de propiedades según año y mes

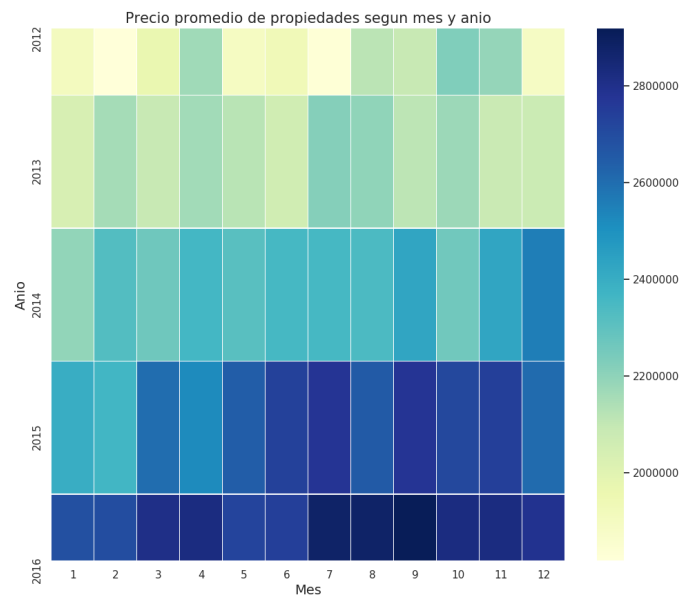


Figura 10: Cantidad de publicaciones por año por tipo de propiedad

En la figura se ve claramente el aumento de precios de las propiedades según mes y año. En general, el precio dentro de un mismo año se mantuvo estable. Por su parte, el precio interanual muestra un crecimiento importante para cada año y cada mes. En los casos donde se ve una disminución representativa del precio, puede atribuirse a que no fueron publicadas tantas propiedades como en otros meses, o que los tipos de propiedades publicadas fueron dispares, afectando al precio general.

5.5. Precio promedio de tipo de propiedades según año

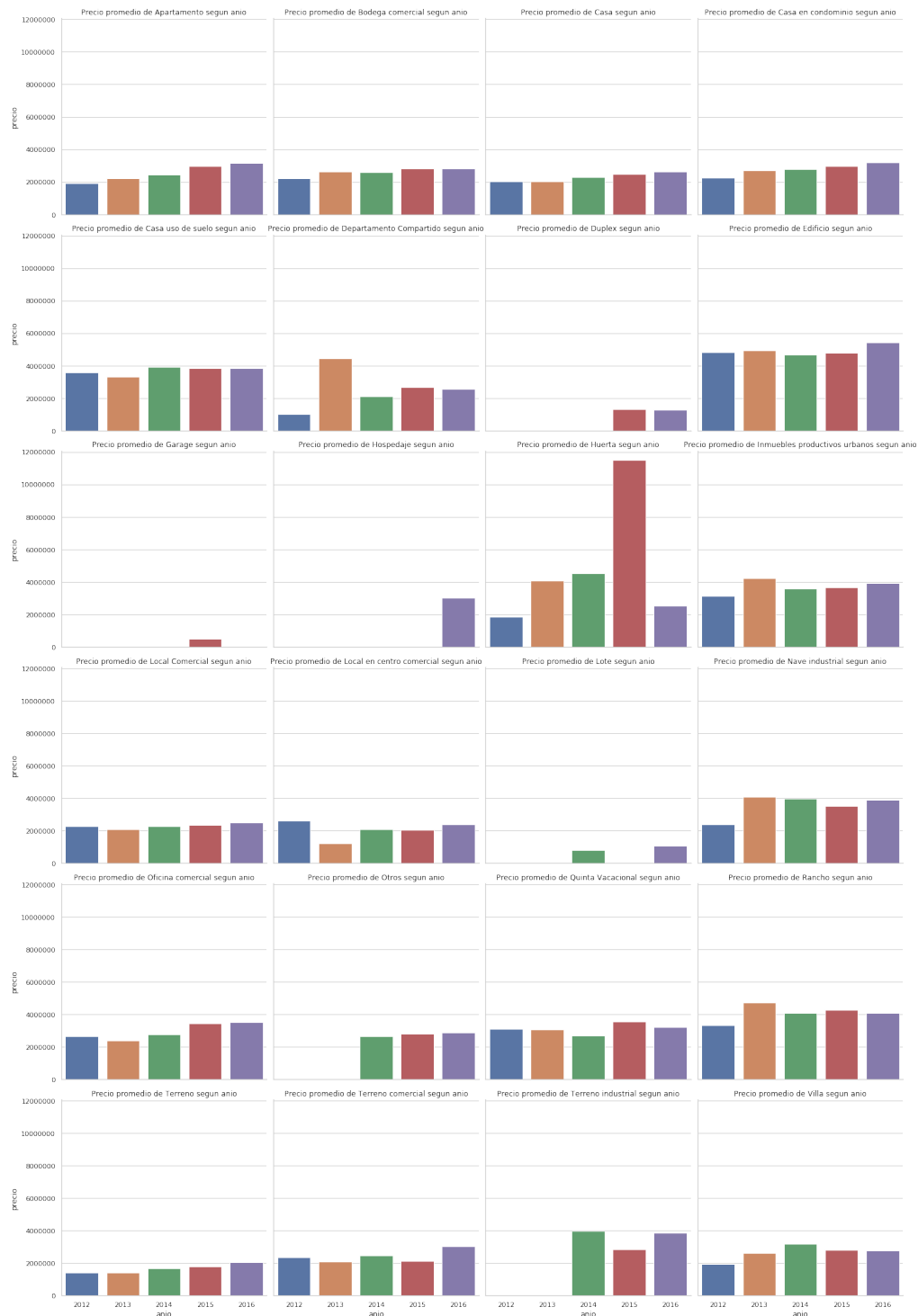


Figura 11: Cantidad de publicaciones por año por tipo de propiedad

Siguiendo la idea de la sección anterior, se muestra la variación de precios para cada tipo de propiedad según los distintos años. Cabe resaltar que en los casos de garage, duplex, lote, hospedaje

no se puede sacar ninguna conclusión significativa puesto que son pocas las publicaciones de esos tipos de propiedades. Por su parte, el aumento de una huerta en el año 2015 pudo deberse a unas pocas publicaciones con un precio muy alto, ya que en general este tipo de propiedad no presenta muchas publicaciones como se vio anteriormente.

5.6. Provincias con mas publicaciones por año

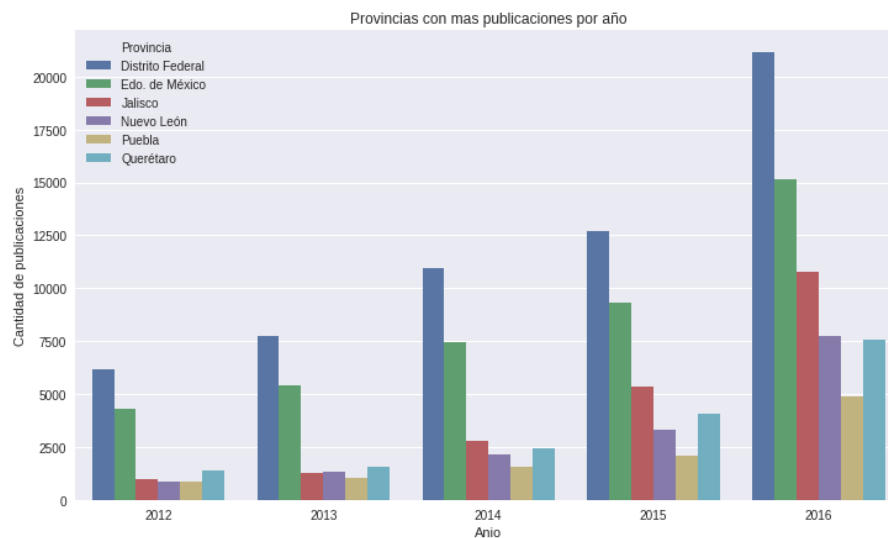


Figura 12: Provincias con más publicaciones por año

Distrito federal siempre tuvo mas publicaciones, además se observa como Jalisco comenzó a crecer año a año.

5.7. ¿En qué mes se realizan más publicaciones?

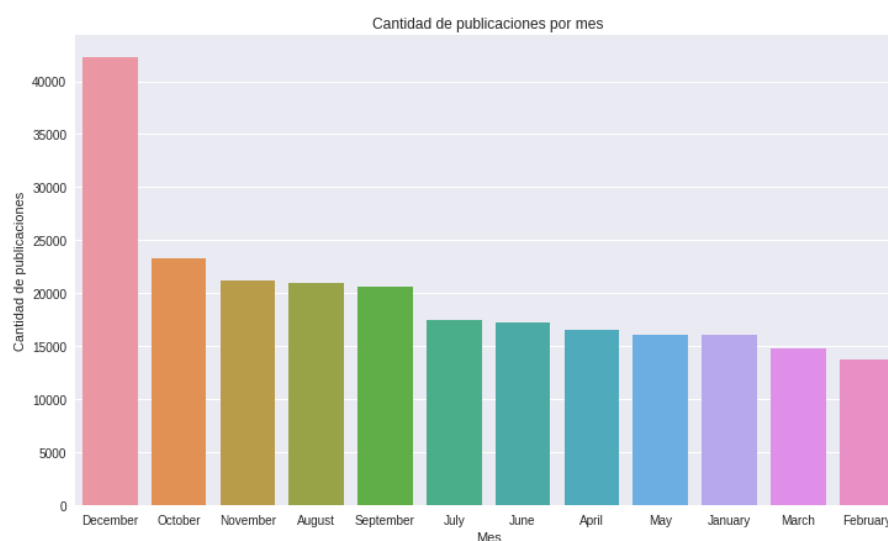


Figura 13: Cantidad de publicaciones por mes

5.8. ¿Qué ocurre en diciembre?

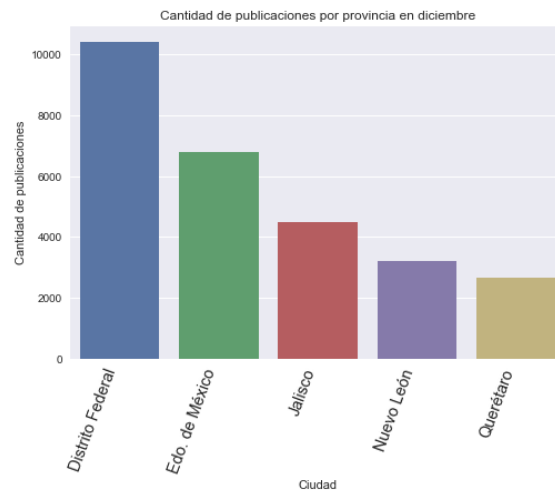


Figura 14: Publicaciones en el mes de diciembre por provincia

Se mantiene el comportamiento de las provincias, solo aumentan la cantidad de publicaciones

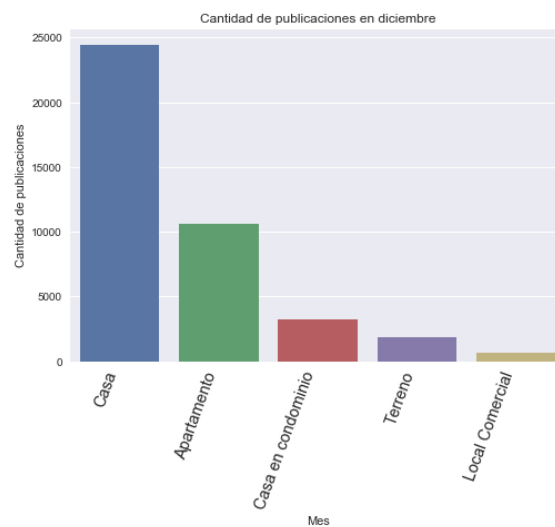


Figura 15: Publicaciones en el mes de diciembre

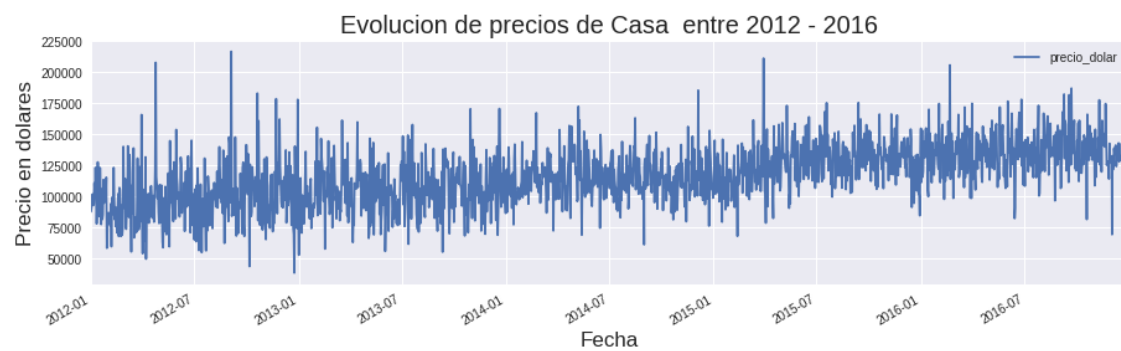


Figura 16: Evolución de precios de casas

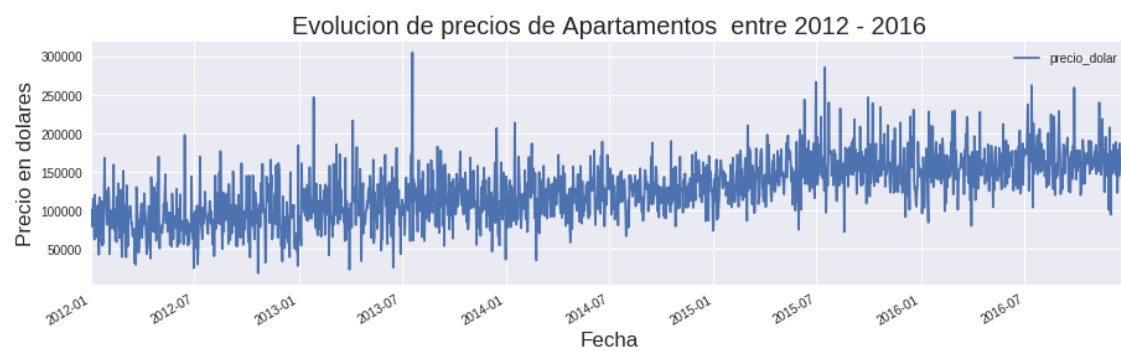


Figura 17: Evolución de precios de apartamentos

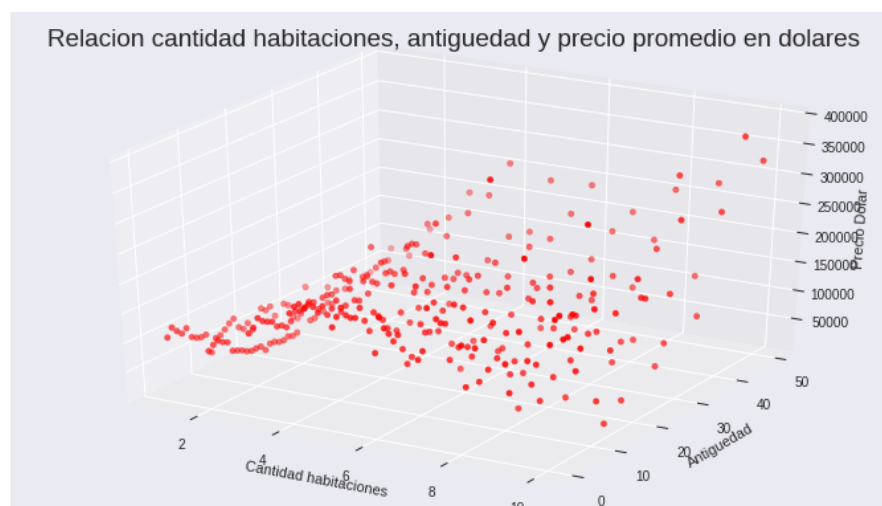


Figura 18: Relación entre cantidad de habitaciones, precio promedio en dólares y antigüedad



Figura 19: Relación entre cantidad de habitaciones, antigüedad y precio promedio en dólares

En los casos de 3 y 4 habitaciones mientras mas antiguas son, mas dinero cuestan en promedio.

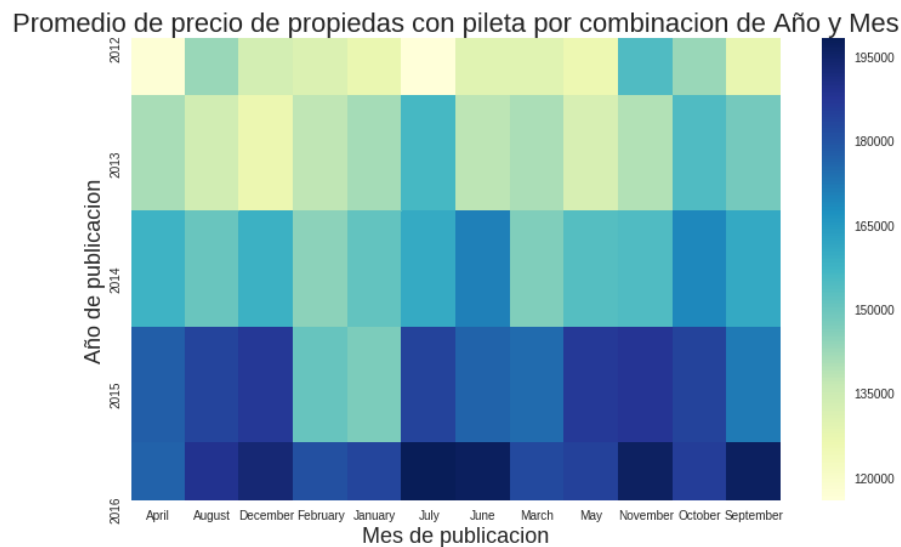


Figura 20: Correlación de promedio de precio de propiedades con pileta y año de publicación

La época de verano en México inicia el 21 de junio y finaliza entre el 21 y 23 de septiembre. Se puede ver como en estos meses el precio de propiedades con piletas es mayor, y como estos fueron aumentando los últimos años

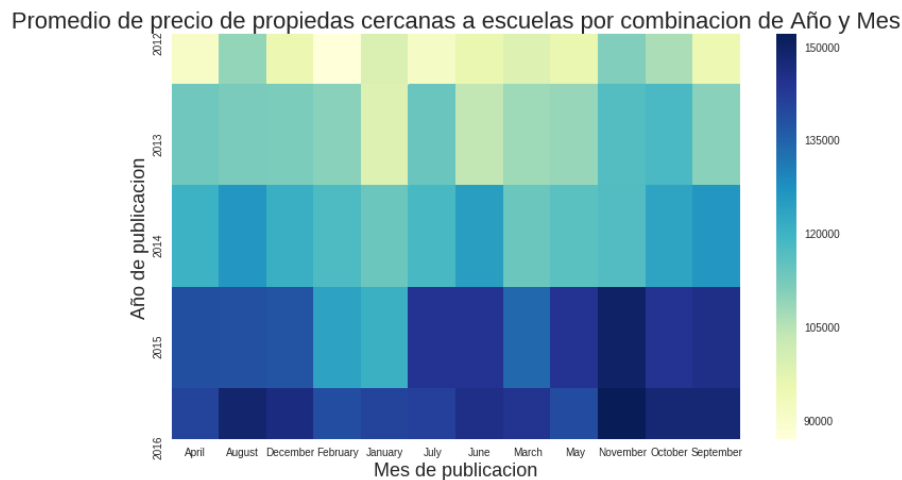


Figura 21: Promedio de precio de propiedades cercanas a escuelas

El ciclo escolar comienza regularmente el segundo lunes de agosto y culmina el tercer o cuarto viernes de abril. Se puede apreciar como en el mes de agosto los precios de las propiedades cercanas a escuelas aumenta. En noviembre también se aprecia un aumento de precios.

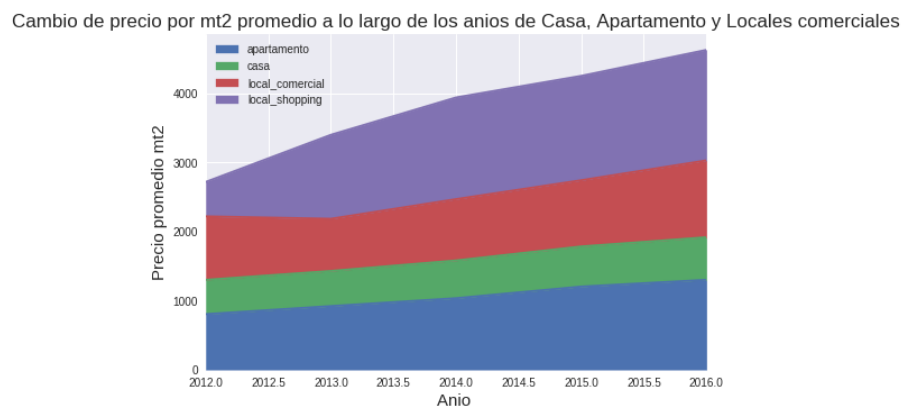


Figura 22: Variación de precio promedio según tipo de propiedad

6. Análisis de propiedades según atributos

La sección que se presenta a continuación tiene como objetivo analizar cómo es la distribución de propiedades en función de la cantidad o porcentaje de baños, habitaciones, precio y metros cubiertos. La idea es ver si se presentan clusters en relación al precio y los metros cubiertos, si existe alguna correlación importante entre la cantidad de habitaciones y la cantidad de baños.

6.1. Propiedades según cantidad de baños

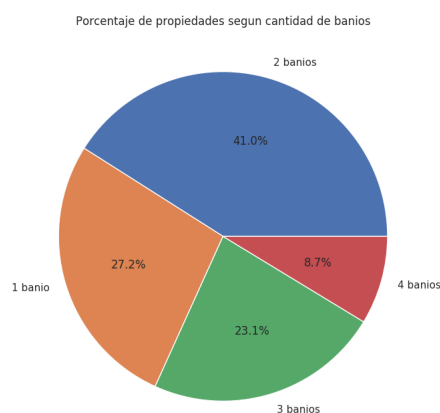


Figura 23: Porcentaje de propiedades según cantidad de baños

La figura muestra algo que tiene relación con lo analizado previamente, como las propiedades mas predominantes son las casas, casas en condominio y apartamento, tiene bastante sentido que un 41 % de las propiedades posean dos baños. Esto puede llevar a pensar que dichas propiedades sean medianamente amplias o tengan varios pisos.

6.2. Propiedades según cantidad de habitaciones

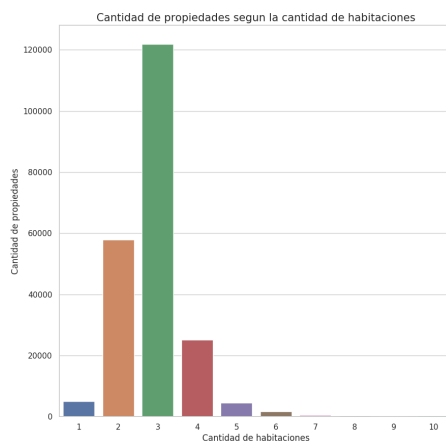


Figura 24: Cantidad de propiedades según cantidad de habitaciones

En la figura se ve que un poco más de la mitad de las publicaciones poseen tres habitaciones. Esto tiene sentido y lleva relación con lo expuesto anteriormente, donde se vio que las casas, los apartamentos y casa en condominio son las principales propiedades publicadas en el sitio.

Similar a la sección anterior, un poco mas de la mitad de las propiedades publicadas tienen tres habitaciones, seguidas por aquellas que poseen dos. Esto tiene sentido y lleva relación con lo expuesto anteriormente donde se vio que las casas, los apartamentos y casa en condominio son las principales propiedades publicadas en el sitio. Además, si se tiene en cuenta el porcentaje de propiedades según cantidad de baños, parecerían ser números razonables para que habite una familia tipo.

6.3. Relación entre precio y metros cubiertos de algunos tipos de propiedad

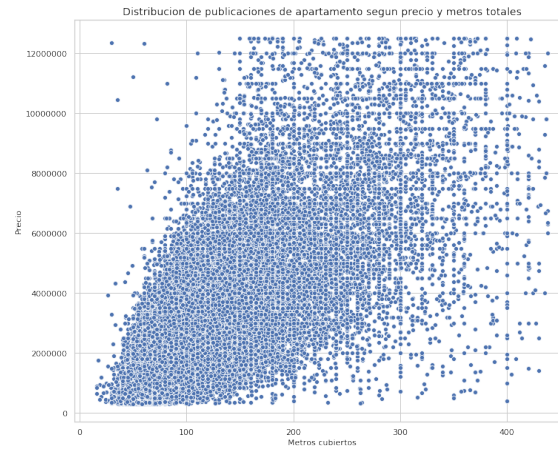


Figura 25: Distribución de publicaciones de apartamento según precio y metros totales

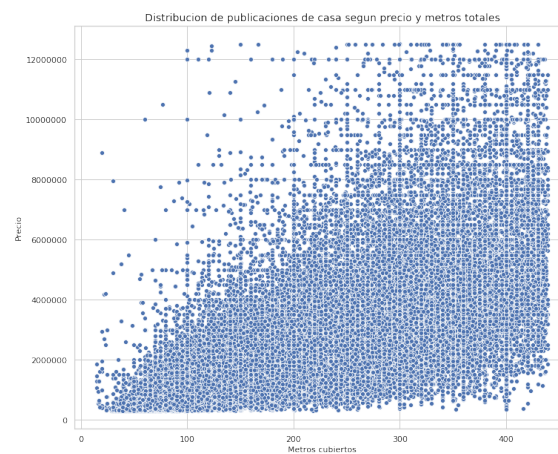


Figura 26: Distribución de publicaciones de casas según precio y metros totales

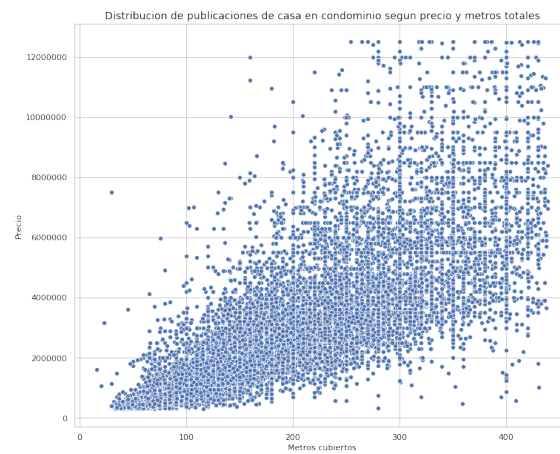


Figura 27: Distribución de publicaciones de casas en condominio según precio y metros totales

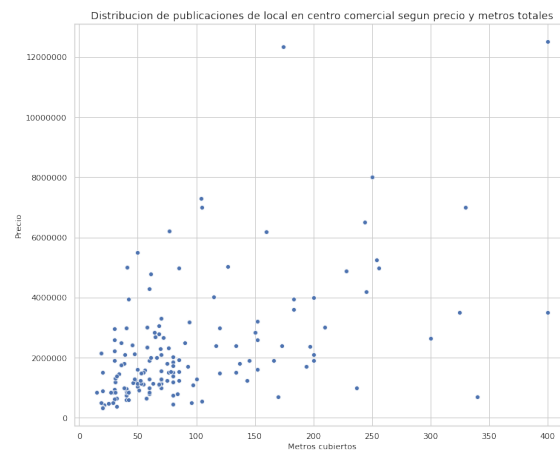


Figura 28: Distribución de publicaciones de locales en centros comerciales según precio y metros totales

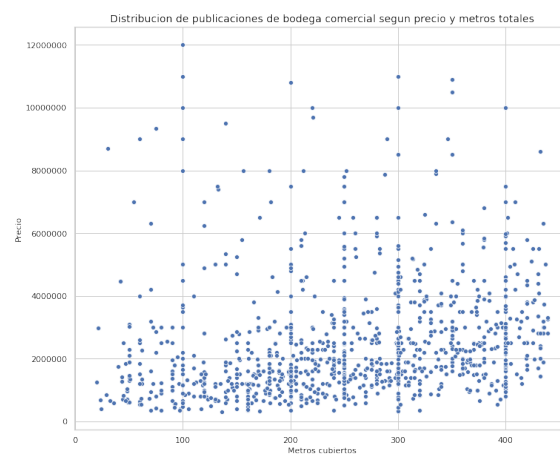


Figura 29: Distribución de publicaciones de bodegas comerciales según precio y metros totales

En cuanto a las casas, no se puede obtener alguna relación o conclusión alguna puesto que es la propiedad con mayor cantidad de publicaciones y existen con diversos precios y diversos metros cubiertos. En cambio, para los apartamentos, casas en condominio, locales en centros comerciales y bodegas si se puede ver algún patrón de clustering.

6.4. Precio promedio de propiedades según cantidad de habitaciones y baños

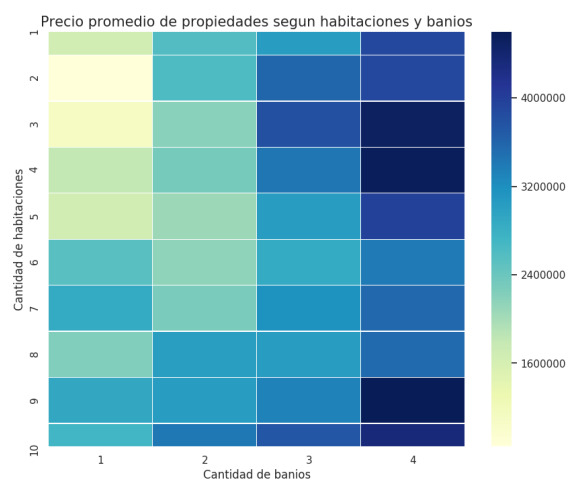


Figura 30: Precio promedio de propiedades según cantidad de habitaciones y baños

La figura nos muestra que influye mas en el precio la cantidad de baños que la cantidad de habitaciones, lo cual es algo poco intuitivo debido a que uno tendería a pensar la cantidad de habitaciones es un factor mucho mas importante.

6.5. Precio promedio de propiedades según gimnasio, piscina y SUM

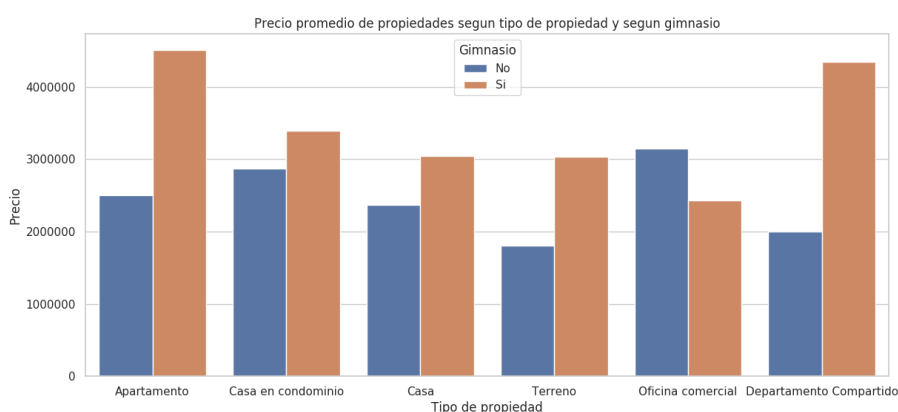


Figura 31: Precio promedio de propiedades según tipo de propiedad y según gimnasio

En la figura se muestra un aumento de precio para aquellas propiedades que tienen gimnasio contra las que no. Sin embargo, para una oficina comercial, esto parece disminuir el precio. Ello puede deberse a que para ese tipo de propiedad no es importante contar con esa comodidad, quitando espacio para otras necesidades u opciones.

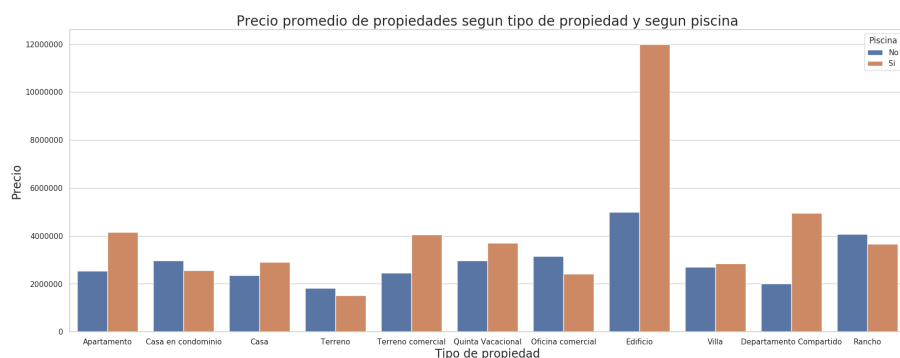


Figura 32: Precio promedio de propiedades según tipo de propiedad y según piscina

En la figura se observa que, para la mayoría de los tipos de propiedad, el poseer una piscina aumenta su valor. Nuevamente, resalta el caso de una oficina comercial donde su precio parece disminuir en el caso de contar con una piscina. Por otro lado, como un podría pensar, un edificio con piscina aumentar notablemente su precio.

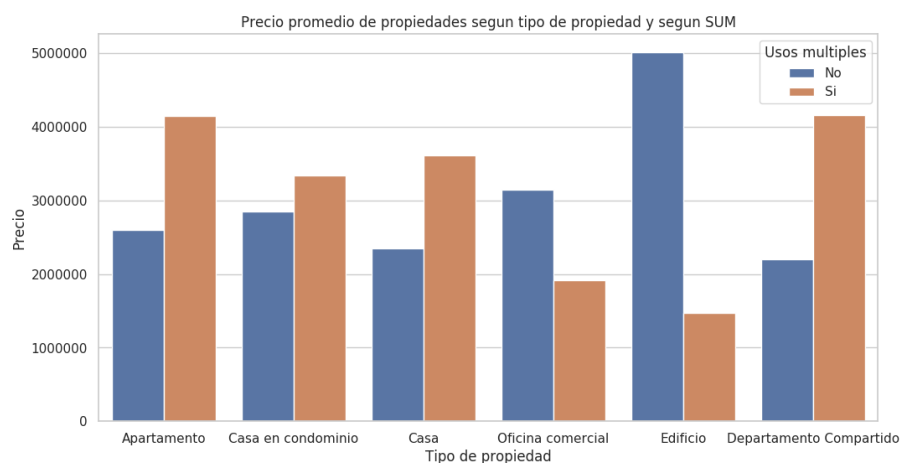


Figura 33: Precio promedio de propiedades según tipo de propiedad y según SUM

En la figura se observa que, para la mayoría de los tipos de propiedad, el poseer un SUM aumenta su valor. Nuevamente, resalta el caso de una oficina comercial donde su precio parece disminuir en el caso de contar con esta comodidad. Sin embargo, en el caso de un edificio, parece ser que su precio es mucho mayor en el caso de no poseer un SUM. Esto puede deberse a que, los edificios con SUM publicados sean pocos y pequeños en comparación con los edificios sin SUM.

7. Otras exploraciones

7.1. Histograma de precios (Casas y departamentos)

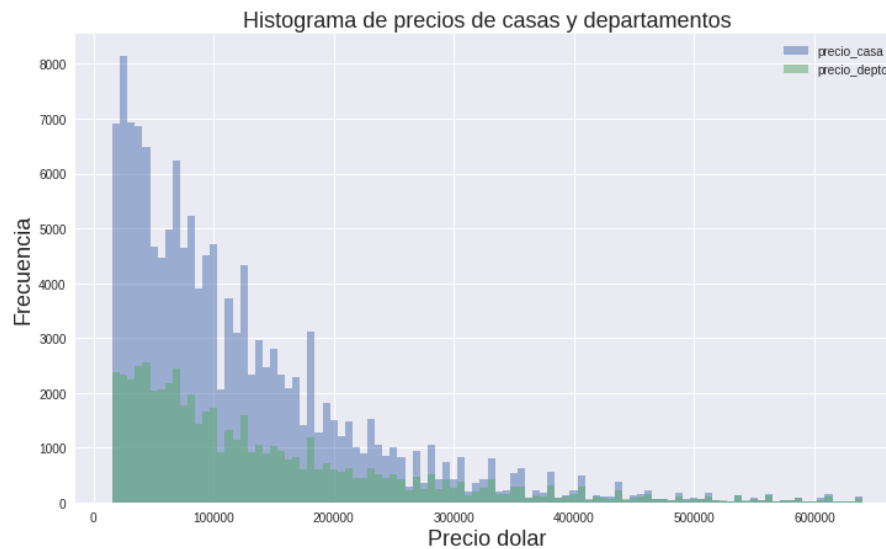


Figura 34: 61

Se puede apreciar que hay mas cantidad de publicaciones de casas que departamentos. Su comportamiento es parecido, hay un pico entre los 25.000 / 50.00 dolares y luego comienzan a decaer. En el caso de las casas existen algunos picos por ej a los 120.000 y 180.000 dolares

7.2. Relaciones entre cantidad de propiedades por ciudad

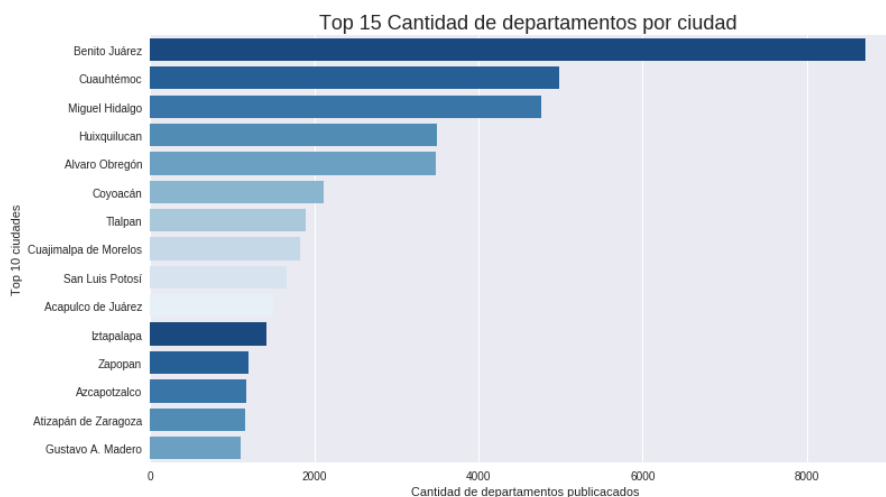


Figura 35: 8

Benito Juárez es una de las ciudades mas céntricas de México, se puede notar que no aparece en el próximo plot de casas. Hay preponderancia de edificios. Y tiene casi el doble de publicaciones que los que se encuentran en 2do y 3er puesto

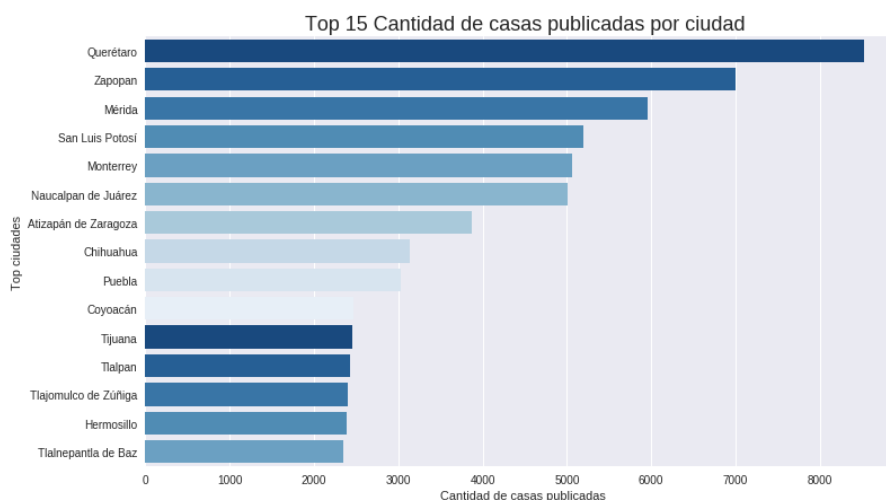


Figura 36: Cantidad de casas por ciudad

En el top 1 encontramos a Querétaro la cual también es una ciudad céntrica pero no figura en el gráfico de departamentos. En esta ciudad predominan las Casas. A que se debe esto?.

Santiago de Querétaro es conocida por su arquitectura colonial española bien conservada en donde predominan arquitecturas antiguas.

Por otro lado en el top 2 se encuentra a Zapopan, que a su vez también figura en el top de departamentos. A diferencia de Querétaro, esta también es una ciudad antigua, pero además se modernizó y atrajo muchos turistas. Cuenta con muchos atractivos históricos, religiosos y modernos, con una gran variedad de hoteles, restaurantes y centros comerciales.

7.3. Propiedades con más publicaciones

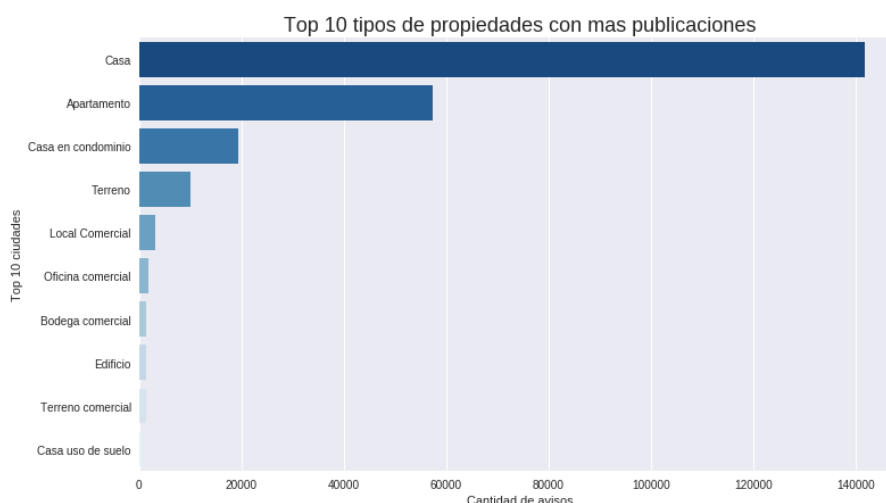


Figura 37: Propiedades con más publicaciones

Pudimos observar que donde mas publicaciones de casas había, se debía a que eran ciudades con edificios antiguos. Podemos analizar que paso en los últimos años, comienza a haber mas departamentos? O se continua con predominio de publicaciones de casas?

7.4. Cantidad de publicaciones de casas y departamentos a lo largo de los años

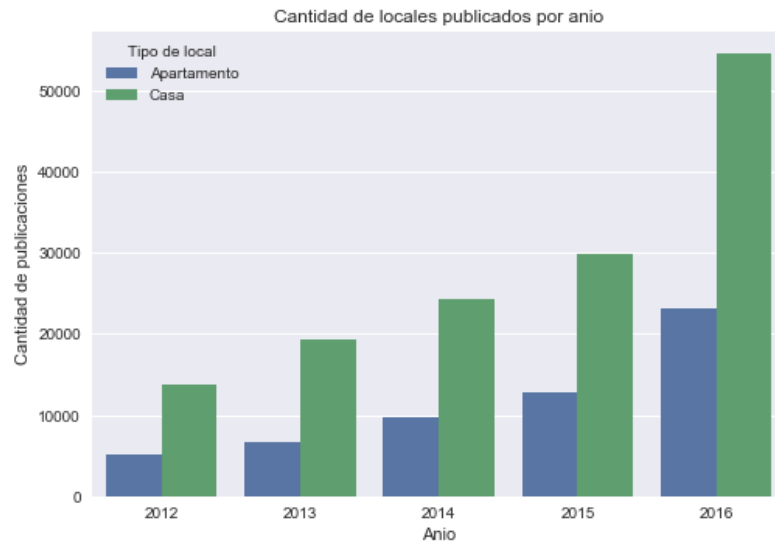


Figura 38: Cantidad de locales publicados por año

Se observa que a pesar del avance del tiempo, sigue habiendo más publicaciones de casas que de departamentos. Podemos filtrar por antigüedad de la propiedad y así poder ver si se están vendiendo muchas casas antiguas y nuevos departamentos, o se siguen construyendo casas.

7.5. Predominancia de casas sobre departamentos

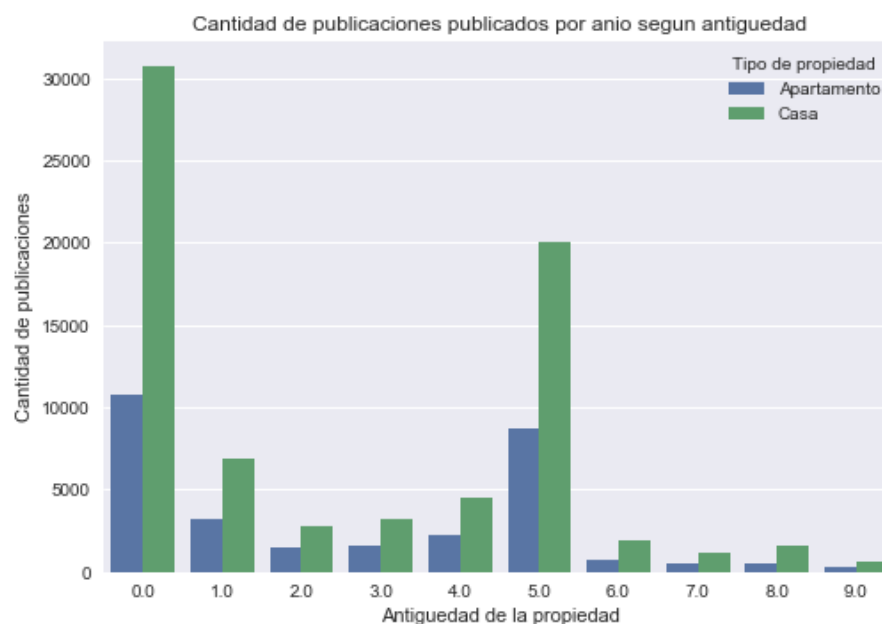


Figura 39: Análisis de publicaciones de apartamentos y casas

Se puede notar que siguen predominando las casas sobre edificios. En el ultimo año se construyeron y publicación 11000 departamentos menos de la mitad de lo que ocurrió con las casas (Al rededor de 30.000)

7.6. Análisis precio por mt2

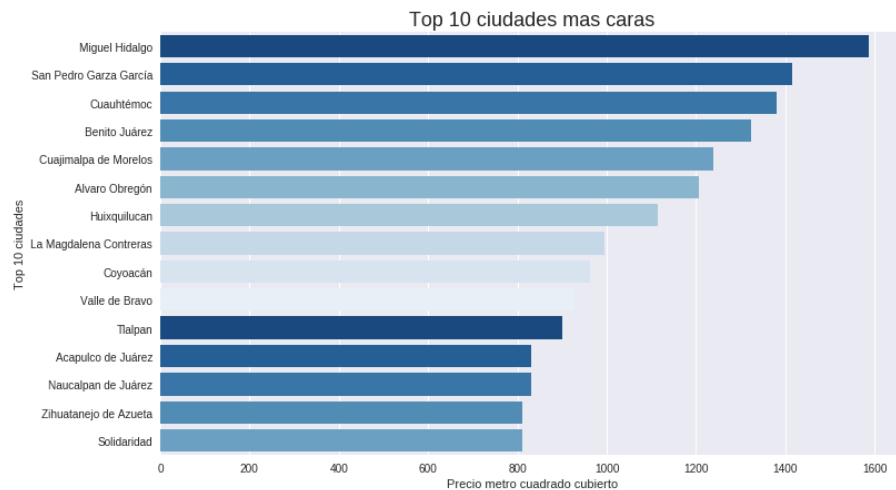


Figura 40: Ciudades con precios más altos

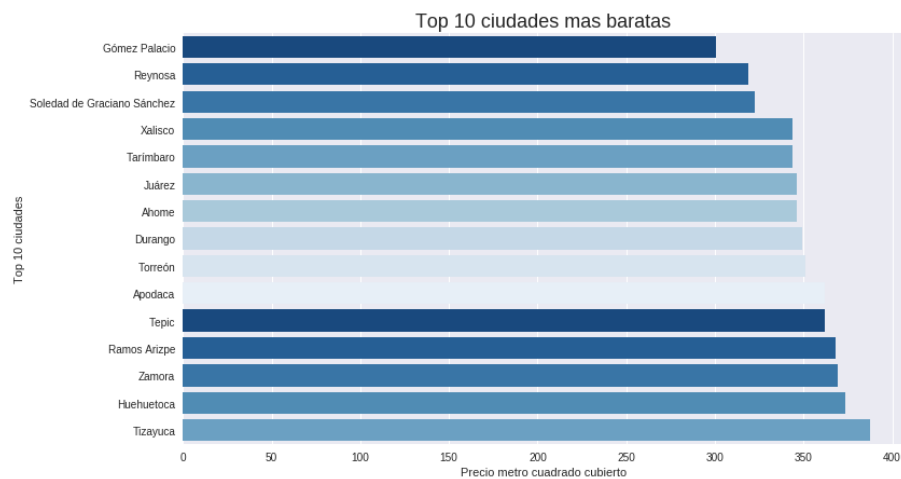


Figura 41: Ciudades con precios más bajos

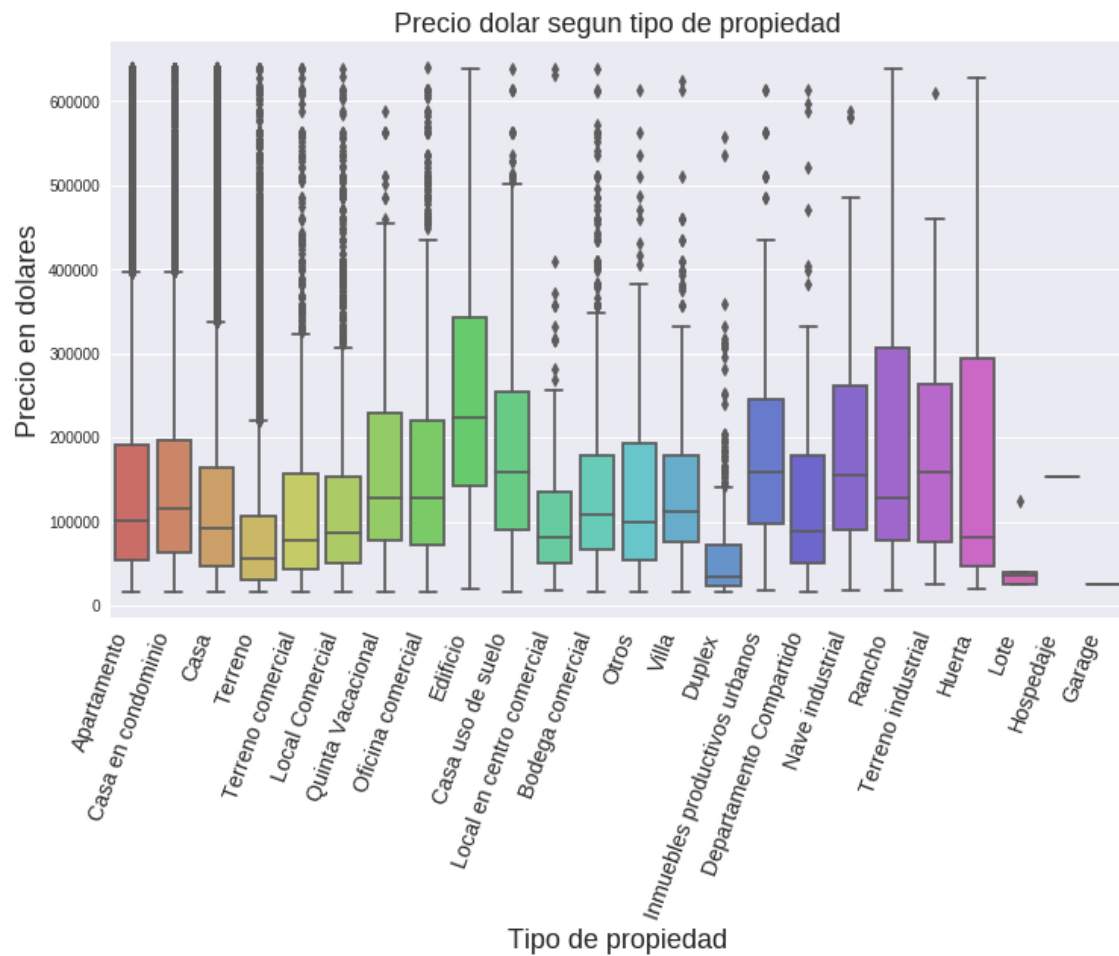


Figura 42: Precio según tipo de propiedad

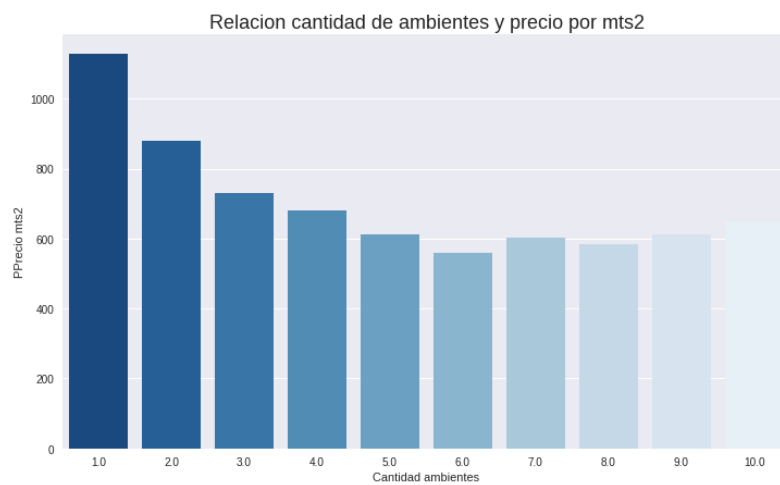


Figura 43: Relación entre cantidad de ambientes y precio por metros cuadrados

8. Análisis de Locales (Vía publica y Centro comercial)

8.1. Histograma precio de un local (Calle) según cercanía a Shopping

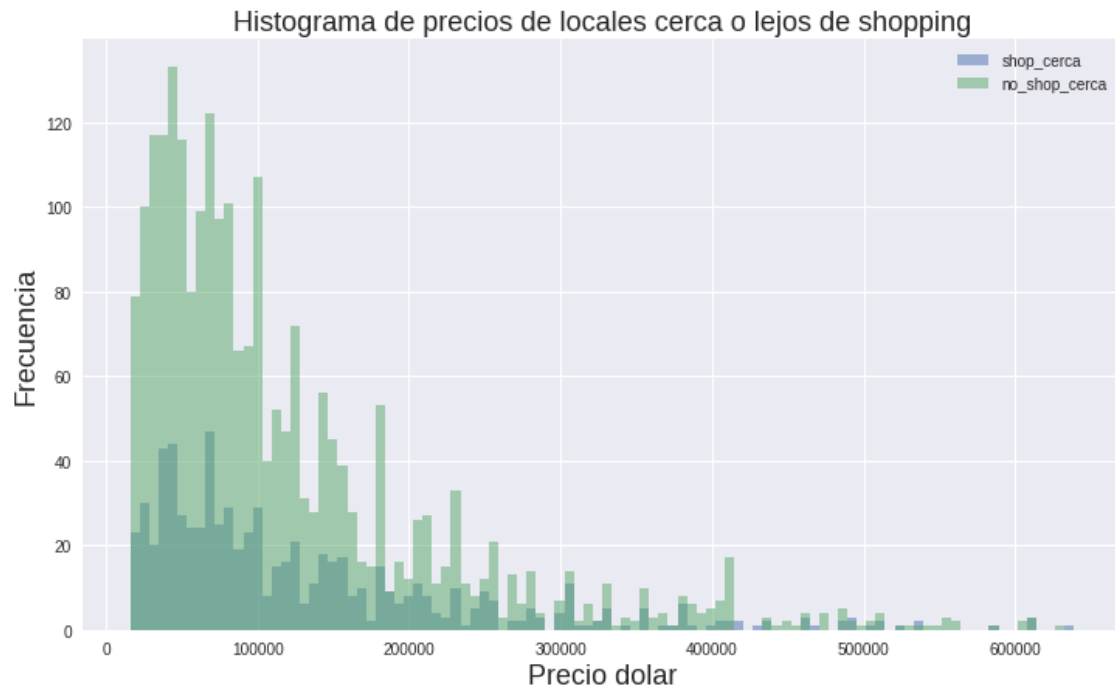


Figura 44: Precios de locales según su cercanía a shoppings

La relación de precios entre una y otra es bastante similar, pero existen muchas mas publicaciones de locales a la calle cuando no existen shoppings cerca.

8.2. Cantidad de locales publicados por año

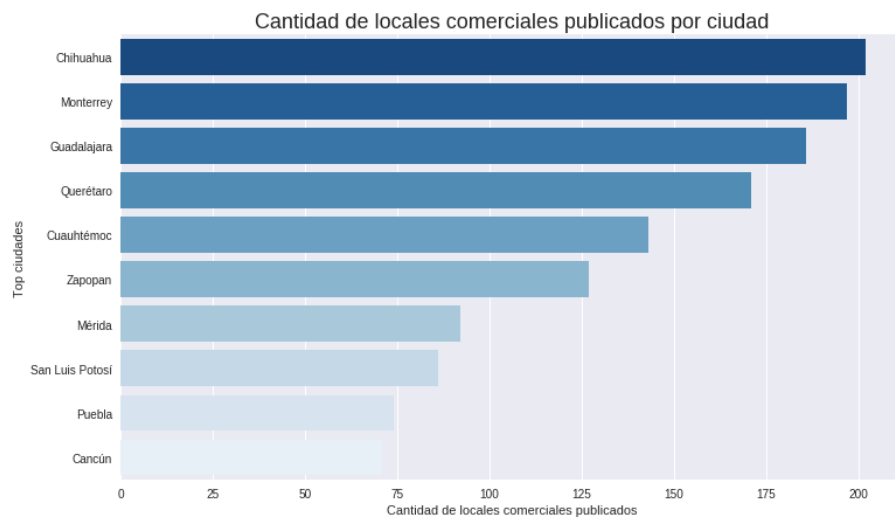


Figura 45: Cantidad de locales comerciales por ciudad

Se puede observar la intersección entre Ciudades con mas locales comerciales (tanto shopping como local a la calle) y los top de ciudades con mas cantidad de publicaciones de casas y departamentos.

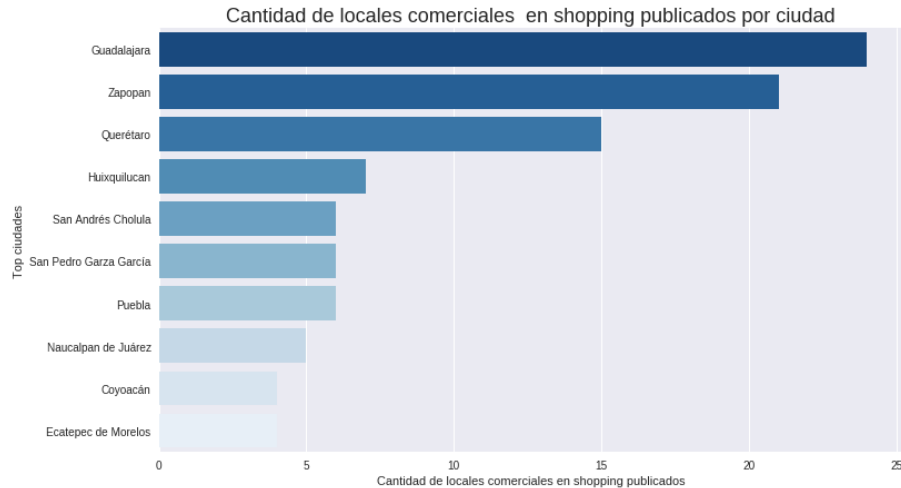


Figura 46: Cantidad de locales comerciales en shoppings por ciudad

Se publican muchos mas locales comerciales en comparación a los que se encuentran en un centro comercial.

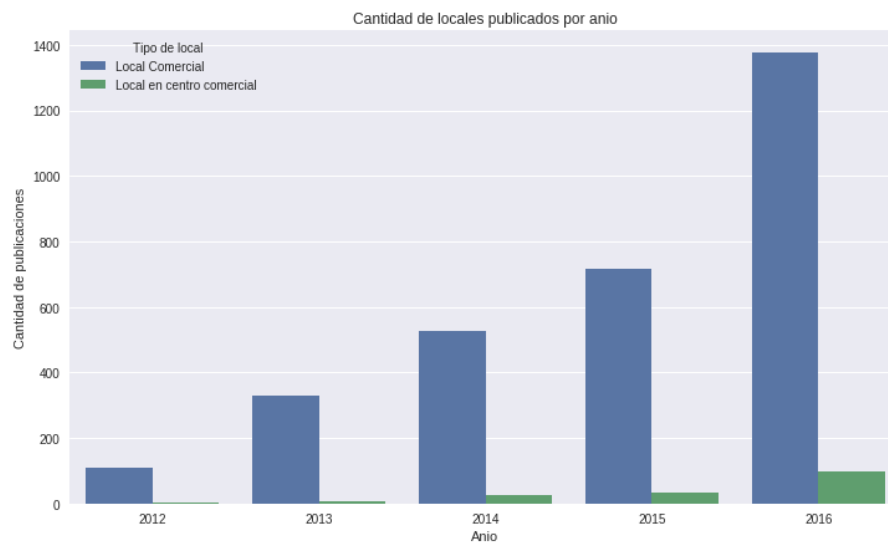


Figura 47: Cantidad de locales por año

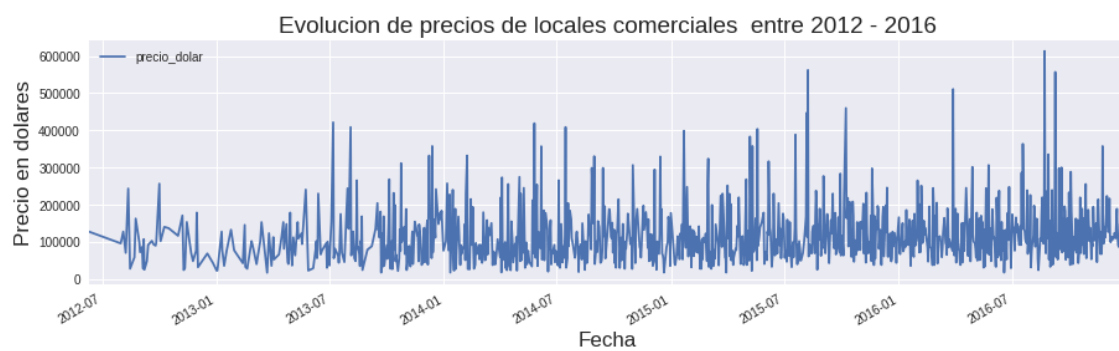


Figura 48: Evolución de precios de locales comerciales