

TP N°1 — Analisis Exploratorio

[75.06/95.58] Organización de Datos
Curso 1
Segundo cuatrimestre de 2019

Grupo 39

Padron:	Alumno:
102740	CONTI, Gianfranco
102445	FARETTA, Yanina Belen
100560	GARCIA, Ailen
101807	GIMENEZ, Lorenzo

Link del Repositorio

<http://github.com/lzogimenez/orgadatostp1>

Link del Kernel

<https://www.kaggle.com/lzogimenez/orgadatostp1>

Índice

1. Introducción	2
2. Evolución del sitio	3
2.1. Alcance geográfico	3
2.2. Volumen de anuncios	4
2.3. Arribo a Ciudades	4
2.4. Campos Nulos	5
3. Uso del sitio	6
3.1. Actividad por día	6
3.2. Actividad en el año	6
4. Características de las propiedades y sus anuncios	7
4.1. Tipos de propiedades publicadas	7
4.2. Distribución de las categorías de los servicios	8
4.3. Presencia de Baños por cantidad de Habitaciones	9
4.4. Antigüedad de las propiedades	10
4.4.1. Años en los que se construyeron las propiedades	12
5. Precio de las propiedades	13
5.1. Distribución general de precios en el sitio	13
5.2. Precio por ubicación	14
5.2.1. Precio por zona geográfica	14
5.2.2. Precio general por provincia de publicación del anuncio	15
5.2.3. Precio en tipos principales sobre las 5 provincias más costosas	16
5.2.4. Precio en tipos principales sobre las 5 provincias menos costosas	16
5.2.5. Distribución precios de casas por provincia	17
5.2.6. Distribución precios de apartamentos por provincia	18
5.3. Precio por año de publicación del anuncio	19
5.4. Precio por antigüedad según propiedad	20
5.5. Precio por metros cuadrados	21
5.5.1. Precio general por metros cuadrados	21
5.5.2. Precio por metros cubiertos en tipos principales	23
5.6. Precio de las propiedades según la categoría de los servicios	24
6. Conclusiones	25

1. Introducción

A lo largo del siguiente trabajo práctico se analizarán los datos correspondientes a los anuncios de ZonaProp, una plataforma de anuncios para la compra/venta de inmuebles.

El set de datos contiene a todos los anuncios publicados entre 2012 y 2016 en México con los campos completados por el anunciante al momento de crear su aviso:

- Título, descripción y fecha de publicación
- Dirección, ciudad, provincia, idzona, latitud y longitud
- Tipo de propiedad y cantidad de habitaciones, garages y baños
- Antigüedad, metros cubiertos, metros totales y precio
- Indicadores de gimnasio, salón de usos múltiples, piscina y escuelas y centros comerciales cercanos

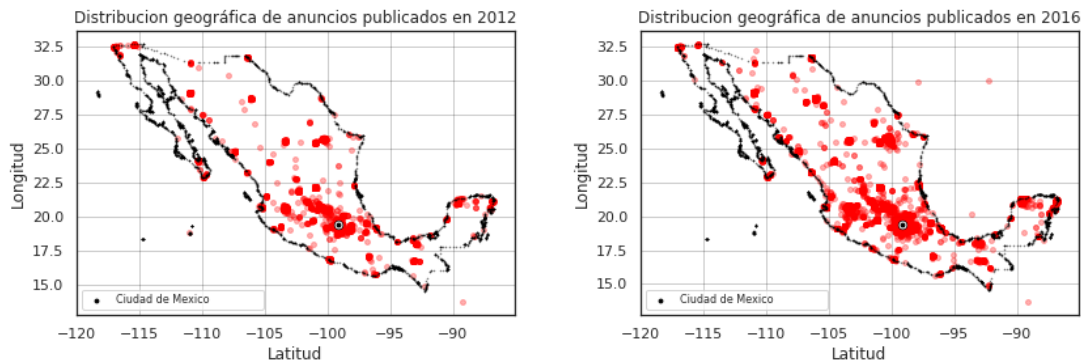
El objetivo de esta práctica es encontrar relaciones y comportamientos relevantes dentro de éstos campos, para confeccionar un reporte que pueda ser útil para:

- Navent (Zonaprop)
- Personas que buscan comprar/vender sus propiedades
- Agentes que trabajan en el mercado inmobiliario

2. Evolución del sitio

2.1. Alcance geográfico

En primer medida, y para determinar el alcance de ZonaProp en México, se graficaron los pares de coordenadas de los anuncios publicados para el primer y el último año registrado en el set de datos, donde se puede observar un gran crecimiento de la extensión de los mismos a lo largo del país.



Se puede observar comparando ambos gráficos que en 2016 existe una mayor dispersión geográfica de los anuncios publicados provocando que ya no se concentren los anuncios exclusivamente en la capital. Esto último se ve respaldado por la evolución en el alcance de ciudades descripto en la sección [2.3].

NOTA: Gran parte de los campos latitud y longitud contenían valores nulos <ver sección [2.4]>y valores no nulos que no correspondían a las coordenadas de México, de manera que se empleó un filtro sobre ellas

2.2. Volumen de anuncios

Para cuantificar ese crecimiento, se estudió la cantidad de anuncios publicados año a año en la plataforma, donde se evidencia el aumento logrado.

Elegimos destacar como el numero de publicaciones en 2016 fue más del cuádruple que en 2012 y más del doble que en 2015.



Este crecimiento en la cantidad de anuncios publicados es un indicador directo de un mayor uso del sitio por parte de los anunciantes y de un aumento de la oferta, lo cual puede atraer a un mayor volumen de compradores y por ende resultar en un sitio más popular.

2.3. Arribo a Ciudades

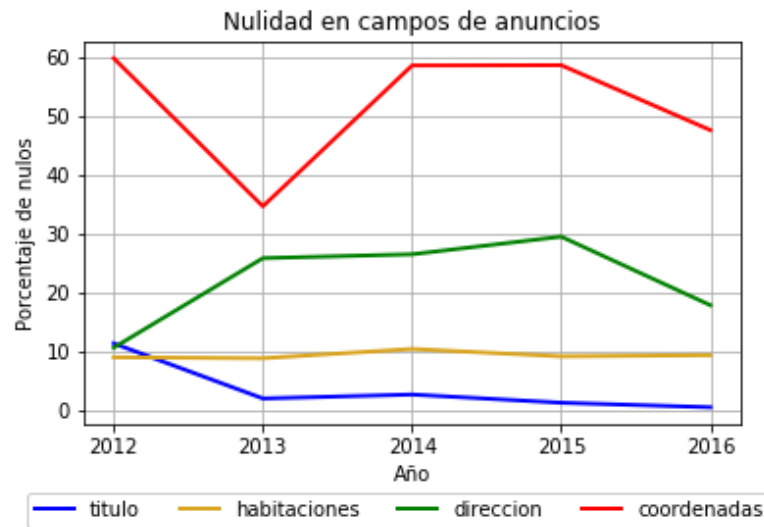
Acompañando también al estudio del crecimiento, se contabilizaron las distintas ciudades sobre las que se publicaron anuncios año a año y nuevamente se observa un gran avance, partiendo desde 454 en 2012, y alcanzando 641 en 2016. Ésta expansión es la que se observa en el mapa geográfico en la sección [2.1] con el aumento en la dispersión de los puntos a lo largo de todo el territorio.



2.4. Campos Nulos

Desde los primeros acercamientos a éste análisis, se fue notando que en el set de datos existían para ciertas categorías una gran cantidad de campos con valor nulo, es por eso que nos pareció importante cuantificar e interpretar por qué los anunciantes no completan éstas categorías.

Para ello se confeccionó un gráfico que muestra el porcentaje de anuncios que tienen valor 'NULL' en las categorías elegidas, para los publicados por cada año, con la finalidad de observar la evolución.



Idealmente, el porcentaje de nulidad en los campos tendría un comportamiento como el de 'título', con algún porcentaje inicial, pero que luego con el correr del tiempo los anunciantes se familiaricen con la plataforma y completan en su totalidad.

Distinto es el caso para 'Dirección' y los campos 'Latitud' y 'Longitud' (éstos últimos 2 expresados como 'coordenadas') y es que en la evolución temporal no solo fluctúan, sino que para 'dirección' se tiene en 2016 un porcentaje de nulos incluso mayor al de 2012.

Por lo tanto cabe preguntarse ¿El 20 % de los anunciantes no conoce la dirección de su inmueble? Damos por supuesto que la conocen, pero eligen no publicarla en el sitio.

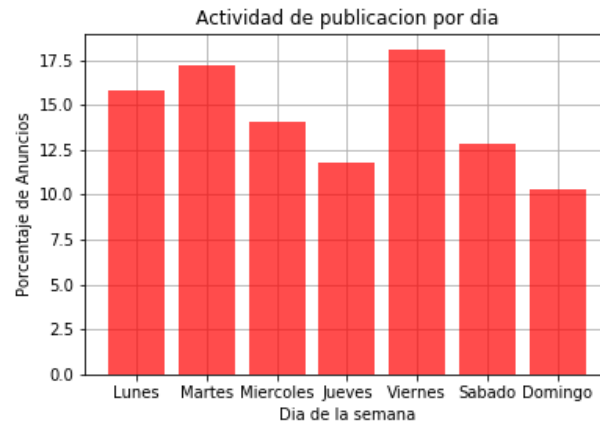
Ésta es entonces una excelente oportunidad para Navent de despejar a los anunciantes de cualquier inquietud que los prive de completar este campo, como una sugerencia, proponemos garantizarle la seguridad de que la dirección solo será provista en caso de que el interesado posea una cuenta en el sitio y quiera hacer una reserva o conocer el solar (como en el caso de Airbnb)

Una propuesta más detallada puede hallarse en las **Conclusiones**.

3. Uso del sitio

3.1. Actividad por día

Para éste análisis, se estudió el porcentaje de anuncios publicados según el día de la semana, como resultado, se ve que los días Jueves, Sábado y Domingo son los que menor cantidad de anuncios nuevos aportan y los Viernes son los que mayor cantidad de anuncios publicados registran.

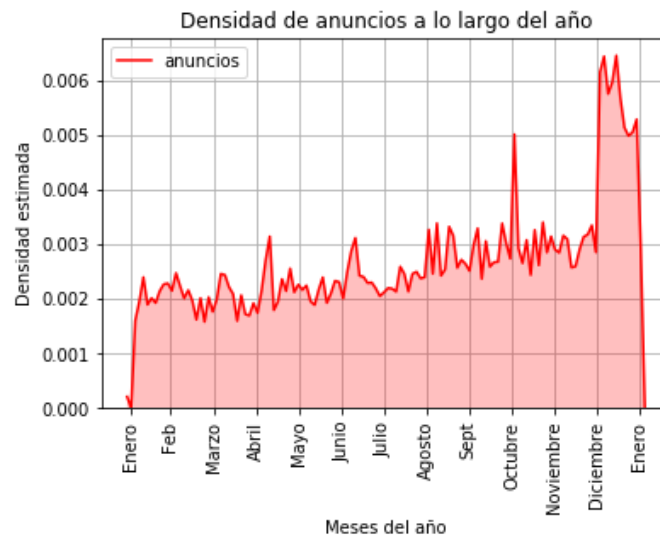


Éste análisis puede ser de gran ayuda para los usuarios que buscan anuncios 'nuevos'. Sabiendo que la mayoría de ellos llegan los días Viernes, una buena idea es realizar sus búsquedas durante el fin de semana.

NOTA: Este análisis incluiría la actividad por hora pero todos los anuncios tienen hora 00:00

3.2. Actividad en el año

A su vez, se graficó la densidad de los anuncios publicados a lo largo del año. Puede observarse que fin de año es el período donde más anuncios son publicados al sitio, conociendo ésta información, un comprador a largo plazo puede estimar cuando es más probable que vea un anuncio de su agrado (considerando de mayor probabilidad los periodos con más anuncios nuevos), además de ser un indicador para Navent de cuándo existe un mayor tráfico en su plataforma.

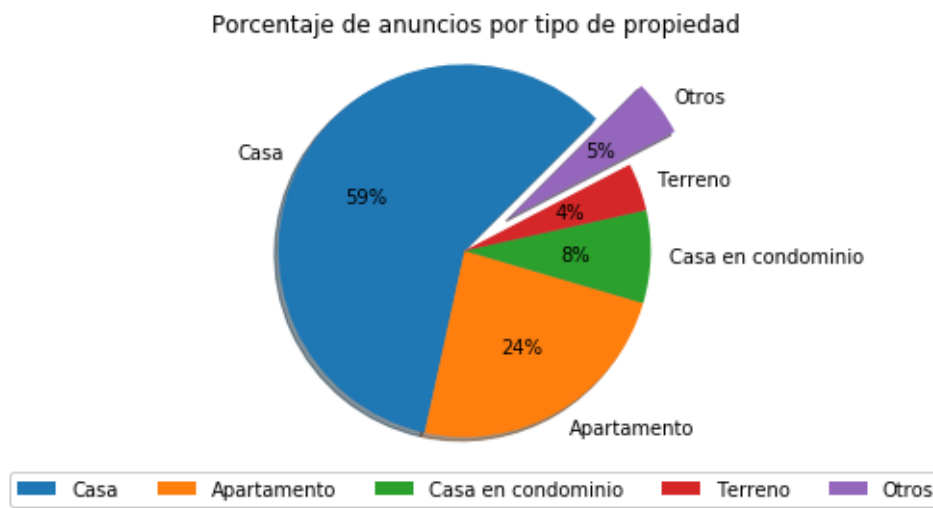


4. Características de las propiedades y sus anuncios

4.1. Tipos de propiedades publicadas

Se estudió qué tipos de propiedades tenían mayor cantidad de anuncios en el sitio: predominan las Casas, Apartamentos y Casas en Condominios, mientras que los Terrenos apenas representan un 4 % y todo el resto de los tipos acumulan un 5 %.

Con esta información, los usuarios que buscan propiedades pueden saber qué encontrarse en el sitio como también Navent conoce más acerca de qué buscan vender la mayoría de sus anunciantes.



Conociendo esta distribución, Navent puede tomar dos caminos: o bien decide abrazar esta tendencia y enfocarse en estos tipos de propiedades, o bien toma medidas dentro del sitio para fomentar nuevas publicaciones de los tipos de propiedades menos comunes, como por ejemplo ofrecer una mayor difusión o algún anuncio 'destacado'.

4.2. Distribución de las categorías de los servicios

En esta subsección, hablaremos del estudio hecho sobre la distribución de los distintos servicios de las propiedades.

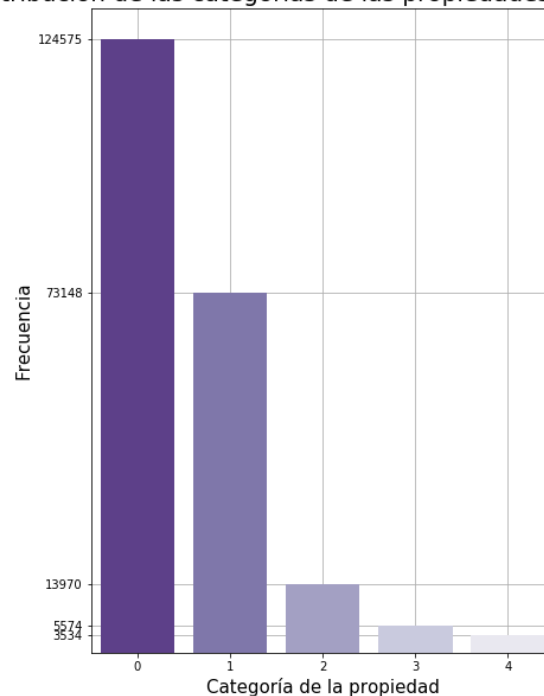
En el set de datos original, existen varias columnas que poseen valores binarios, las llamadas 'gimnasio', 'usosmúltiples', 'piscina', 'escuelascercanas' y 'centrocomercialescercanos'. En este análisis, se consideró a esas columnas como servicios adicionales que poseen las propiedades del tipo "habitables" (como por ejemplo: apartamentos, casas, edificios y demás).

Hemos realizado este filtro de "propiedades habitables", debido a que no tenía sentido hablar de, por ejemplo, un gimnasio en un terreno. Con este panorama, cuantos más de éstos servicios posean las propiedades, mayor será su categoría.

Cada uno de los servicios asigna 1 (un) punto de categoría, que es acumulativo con el resto, los servicios son:

- Gimnasio
- Salón de Usos Múltiples
- Piscina
- Zona Céntrica (Centros comerciales y Escuelas cercanas <ambos>)

Distribución de las categorías de las propiedades habitables

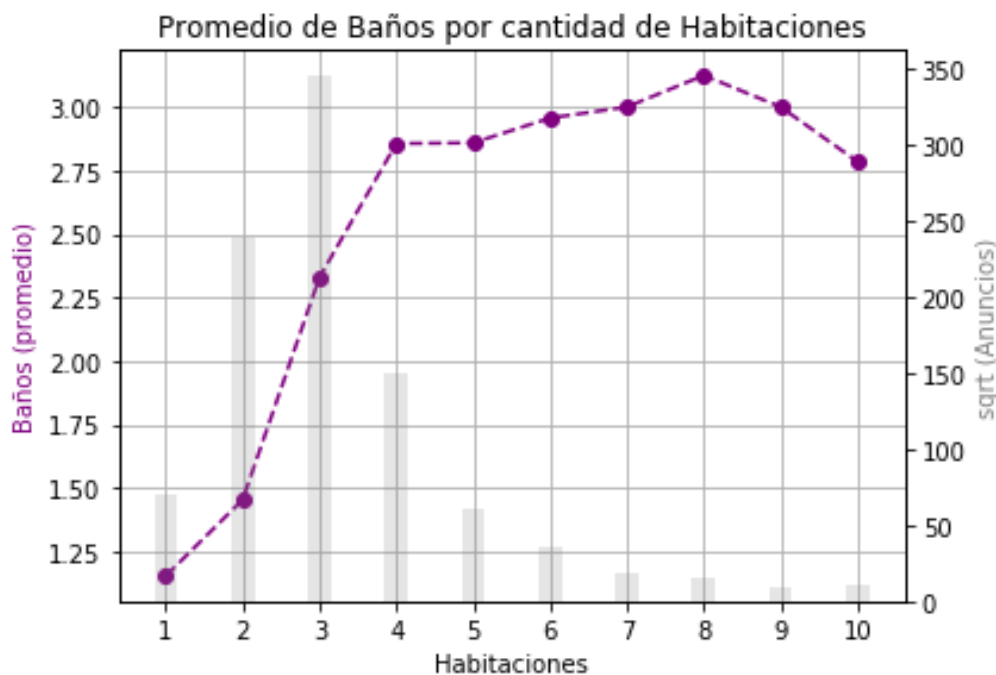


Observando la frecuencia de las categorías entre las propiedades habitables, podemos destacar que predominan las propiedades que casi no poseen servicios adicionales.

4.3. Presencia de Baños por cantidad de Habitaciones

El resultado de éste análisis está dirigido a las personas que trabajan en el mercado inmobiliario, pues propone hallar la relación entre la cantidad de habitaciones y la cantidad de baños de una propiedad. Para al momento de tasar una propiedad, estimar si el número de baños obedece al promedio esperado, si está por encima o si está por debajo.

NOTA: Dado que en el set de datos no existe una distribución uniforme en la cantidad de habitaciones se graficó (detrás de la relación) la cantidad de anuncios sobre la que se calculó el promedio de baños (para tener una noción de las dimensiones de la población y la confianza de su resultado). A éstos valores se les tomo la raíz cuadrada para tener una visualización más sencilla.

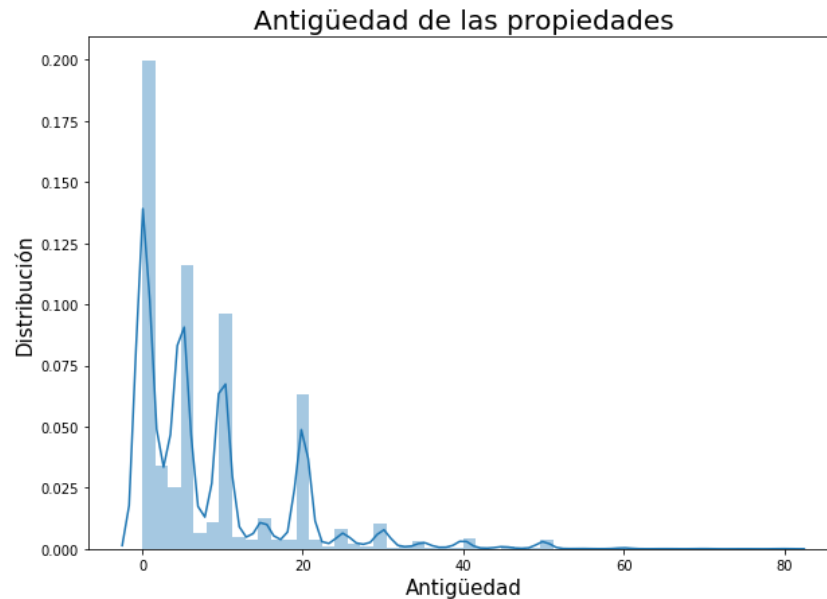


Se observa que hasta las 4 habitaciones, existe un gran crecimiento de la cantidad de baños promedio, pero para números de habitaciones más altos, ya no lo hace con la misma velocidad.

Respecto de la 'caída' del promedio para 9 y 10 habitaciones, concluimos que es debido a que se calculó sobre una población muy pequeña, los consideramos datos poco confiables y dicha caída no se estudiará.

4.4. Antigüedad de las propiedades

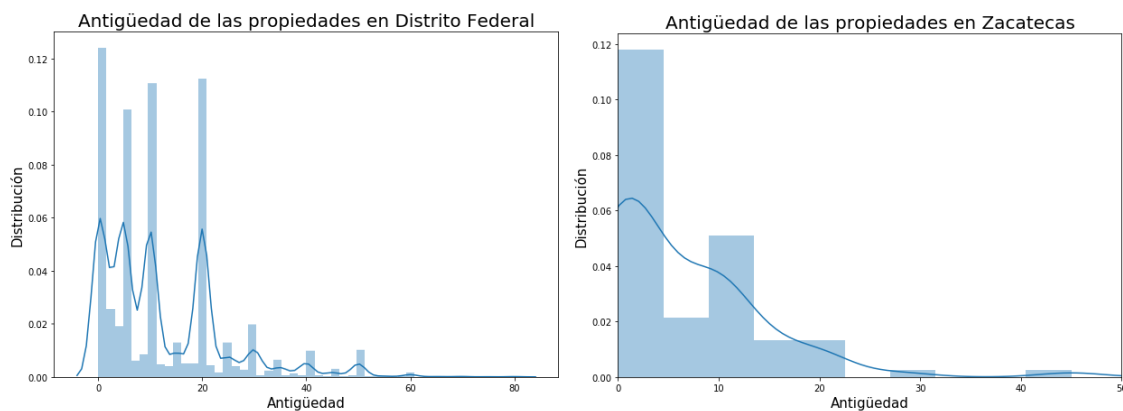
En esta sección, nos centramos en realizar análisis relacionados con la antigüedad de las propiedades. En primer instancia nos pareció conveniente graficar la distribución de esta variable en las propiedades, para ver el rango de años con el que tenemos que trabajar.



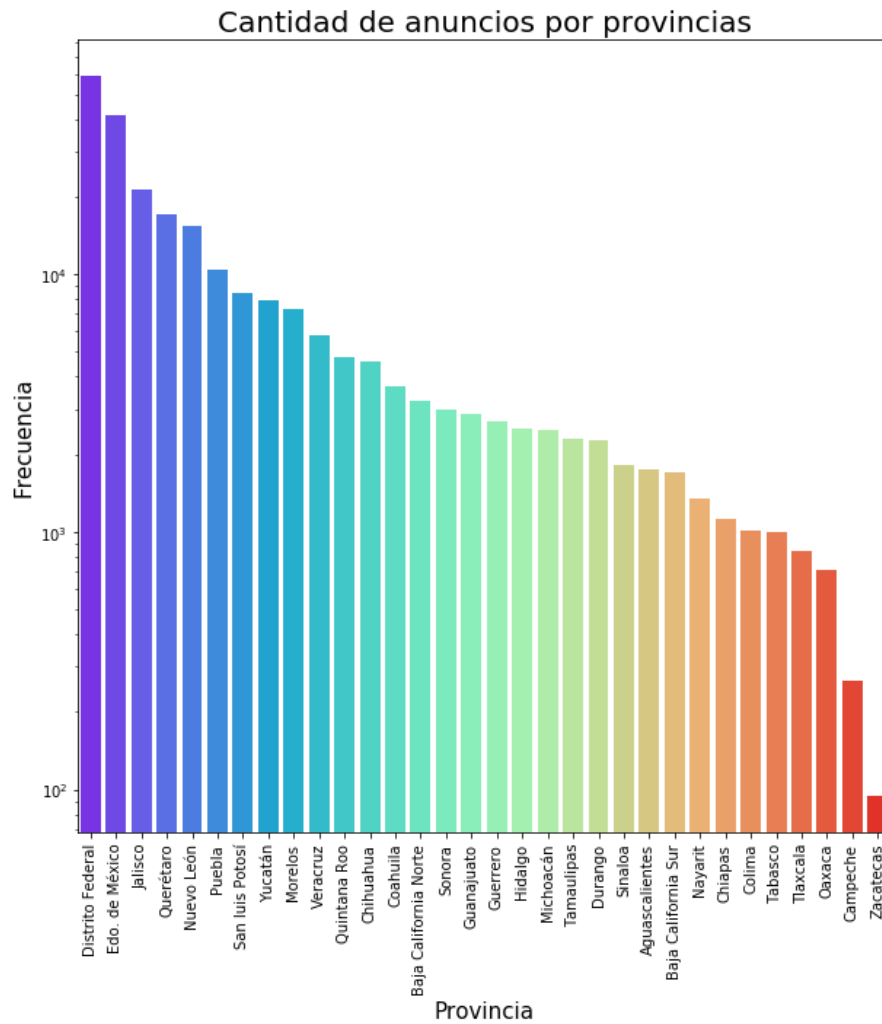
Como podemos observar, estamos frente a un set de datos que posee gran cantidad de propiedades nuevas (con antigüedades cercanas a los 0 años).

Con esta información, nos pareció interesante estudiar este mismo comportamiento en las distintas provincias de México, para ver si existe un determinado grupo de provincias que concentra ese gran número de propiedades nuevas.

En principio, analizamos cómo se distribuía la antigüedad en la provincia más y menos frecuente, es decir, la que mayor cantidad de anuncios posee (DF) y la que menos (Zacatecas) y se observó que en ambas la tendencia es hacia las propiedades más nuevas.



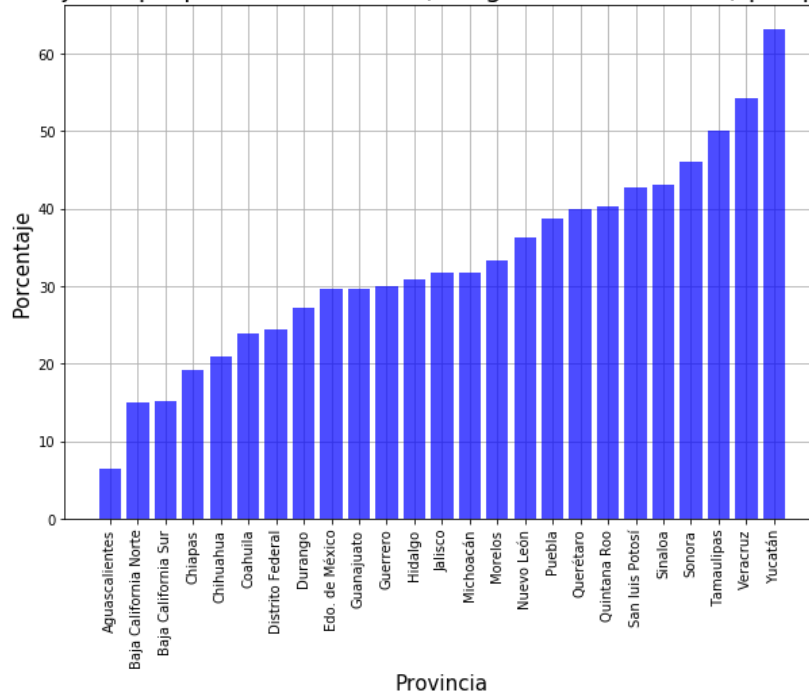
Luego, para evitar calcular promedios sobre poblaciones muy pequeñas de datos, analizamos qué provincias son las que poseen una cantidad admisible de anuncios con un valor de antigüedad no nulo y se obtuvo el siguiente gráfico:



El criterio que se utilizó es que considerar admisibles las provincias que poseen más de 1000 anuncios con antigüedad no nula, para que los promedios sean comparables.

Se obtuvo finalmente, el porcentaje de propiedades nuevas para las provincias admisibles, es decir aquellas donde es más fácil encontrar anuncios de propiedades a estrenar.

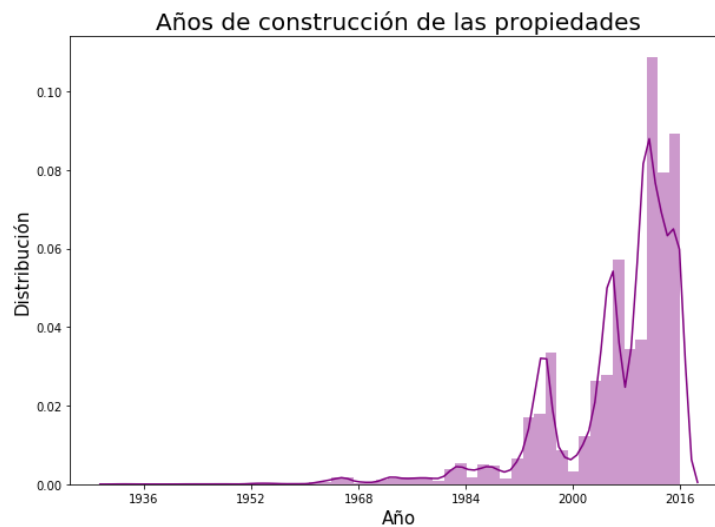
Porcentaje de propiedades nuevas (antigüedad de 0 años) por provincias



Sorprendentemente, el Distrito Federal se encuentra lejos de ser la que alberga un mayor porcentaje de anuncios de propiedades nuevas, siendo la provincia de Yucatán la de mayor porcentaje de oferta 'a estrenar'.

4.4.1. Años en los que se construyeron las propiedades

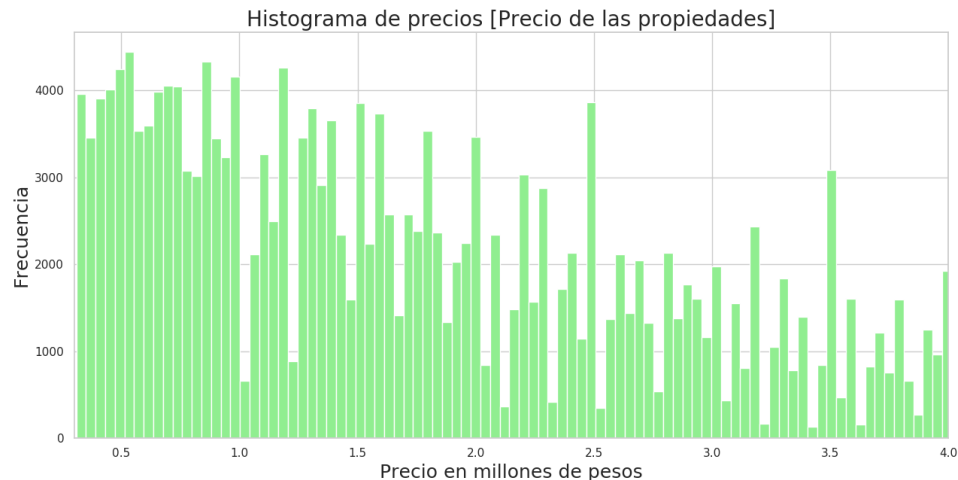
Teniendo la antigüedad de las propiedades, y la fecha de publicación del anuncio, obtuvimos los años en los que se construyeron las mismas, podemos observar, que la mayoría de las propiedades habitables fueron construidas en los años más cercanos a la última fecha de publicación (2016).



5. Precio de las propiedades

5.1. Distribución general de precios en el sitio

Por ser probablemente una de las inversiones de mayor volumen en la vida de una persona, el precio de la propiedad a comprar es un factor fundamental, por lo que consideramos interesante ver como los precios se distribuyen de forma general.

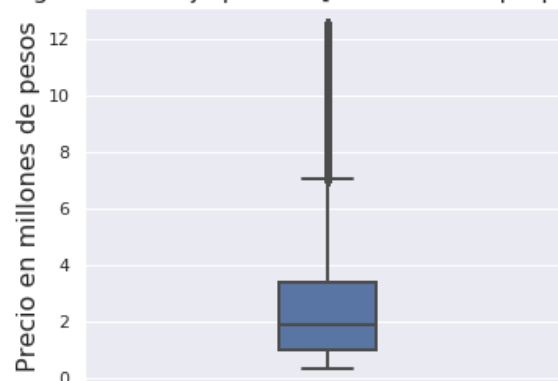


Se reconocen dos particularidades:

- La tendencia es a la baja, la mayoría de los precios son bajos y a medida que los precios aumentan la frecuencia de las publicaciones disminuye (excepto en \$ 2.5 Millones y \$ 3.5 Millones)
- Las publicaciones tienden a juntarse cerca de los valores más “redondos”, ej: \$ 1 Millón, \$ 2 Millón, \$ 3.5 Millones, etc..

La tendencia a la baja de los precios también puede verse reflejada en el siguiente gráfico de caja, puede apreciarse fácilmente que gran parte de los precios tienden a ser bajos.

Diagrama de caja precios [Precio de las propiedades]



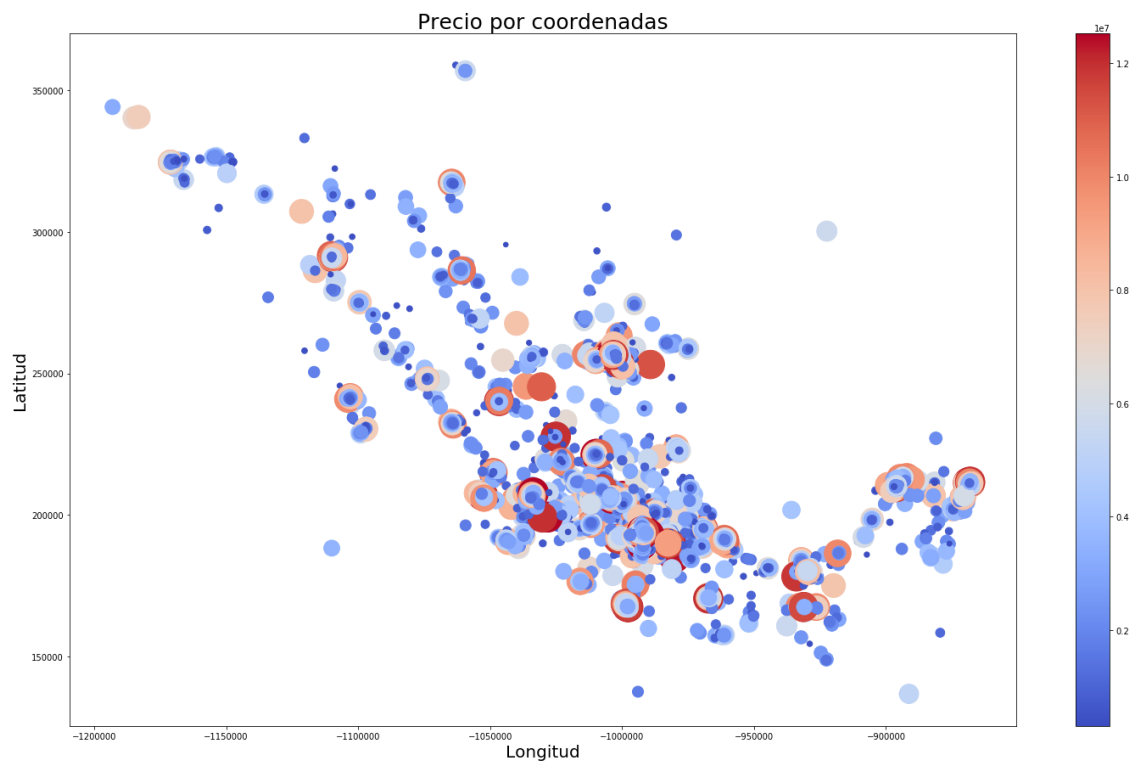
5.2. Precio por ubicación

5.2.1. Precio por zona geográfica

Se estudió, mediante las columnas referentes a la latitud y la longitud (lat y lng), el precio de las propiedades dependiendo de las coordenadas geográficas.

Debido a que estas columnas estaban llenas de NaNs y de valores que no pertenecían al territorio de México, se filtraron para poder observar en los gráficos la forma del país correctamente.

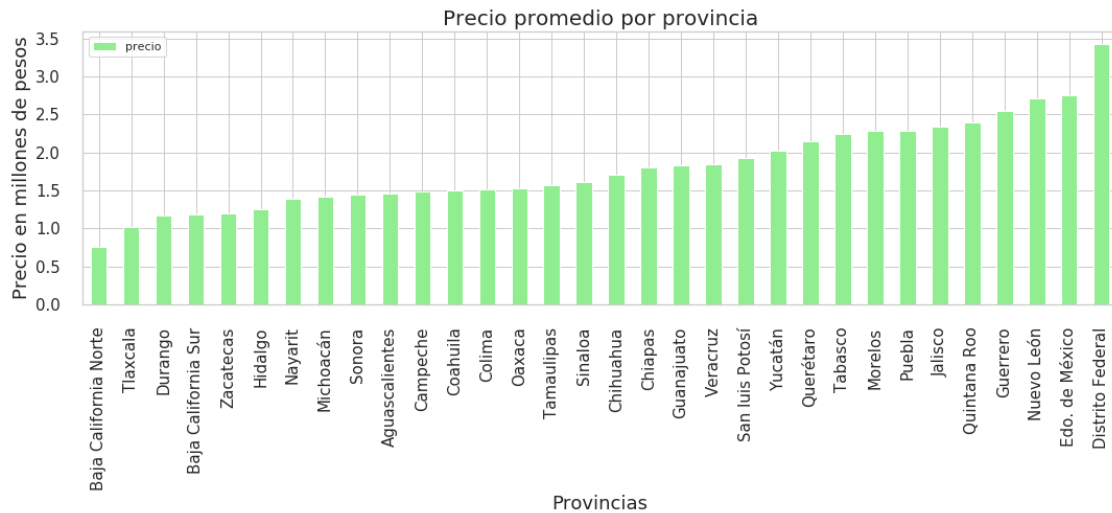
En el siguiente scatterplot cuanto más grande y rojo sea el punto, mayor es el precio de la propiedad representada.



Se observa consistentemente con lo presentado en el análisis anterior que existe una mayor presencia de propiedades con precios pertenecientes al rango más bajo de la escala.

5.2.2. Precio general por provincia de publicación del anuncio

El siguiente gráfico muestra el precio promedio de las publicaciones según la provincia del anuncio teniendo en cuenta todos los tipos de propiedades.

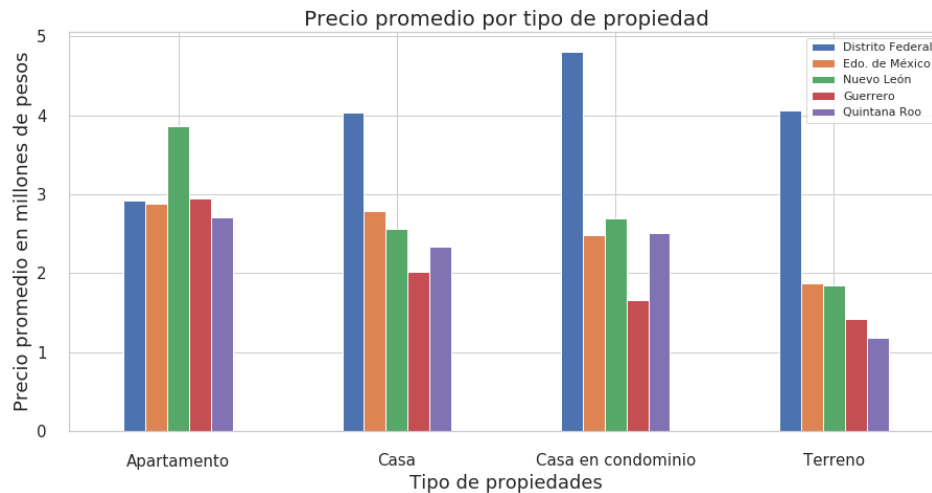


En el gráfico puede observarse que se cubre un gran espectro de precios promedio entre las provincias, siendo DF la provincia con el promedio de precios más alto y Baja California Norte la provincia con el promedio de precios más bajo, a tal punto que su promedio es 5 veces menor que el del promedio de DF.

A partir de este gráfico decidimos analizar el promedio de precios para las 5 provincias más y menos costosas, ésta vez diferenciando entre tipos de propiedad para los más presentes en el set de datos.

5.2.3. Precio en tipos principales sobre las 5 provincias más costosas

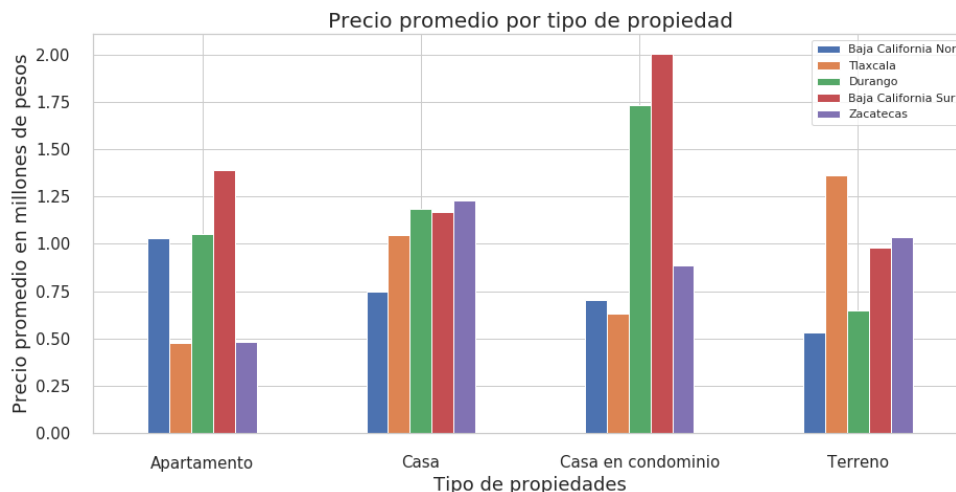
Tal como indica el gráfico en la [sección 4.1], los tipos de propiedad principales son Apartamento, Casa, Casa en condominio y Terreno, para ellos se calculó el promedio de precio en las provincias más costosas:



Para éstas, se observa que el tipo de propiedad más costoso son los Apartamentos, sin embargo, en el Distrito Federal, esa tendencia no se cumple: los Apartamentos poseen el precio promedio más bajo, ésto puede deberse a que, al tratarse de la capital, existe una menor oferta de propiedades que no sean Apartamentos, y esa exclusividad las dota de un precio más alto.

5.2.4. Precio en tipos principales sobre las 5 provincias menos costosas

Entre las provincias menos costosas no hay una tendencia tan marcada mas se puede observar que en líneas generales las casas mantienen un precio promedio similar y los terrenos tienden a ser más baratos.

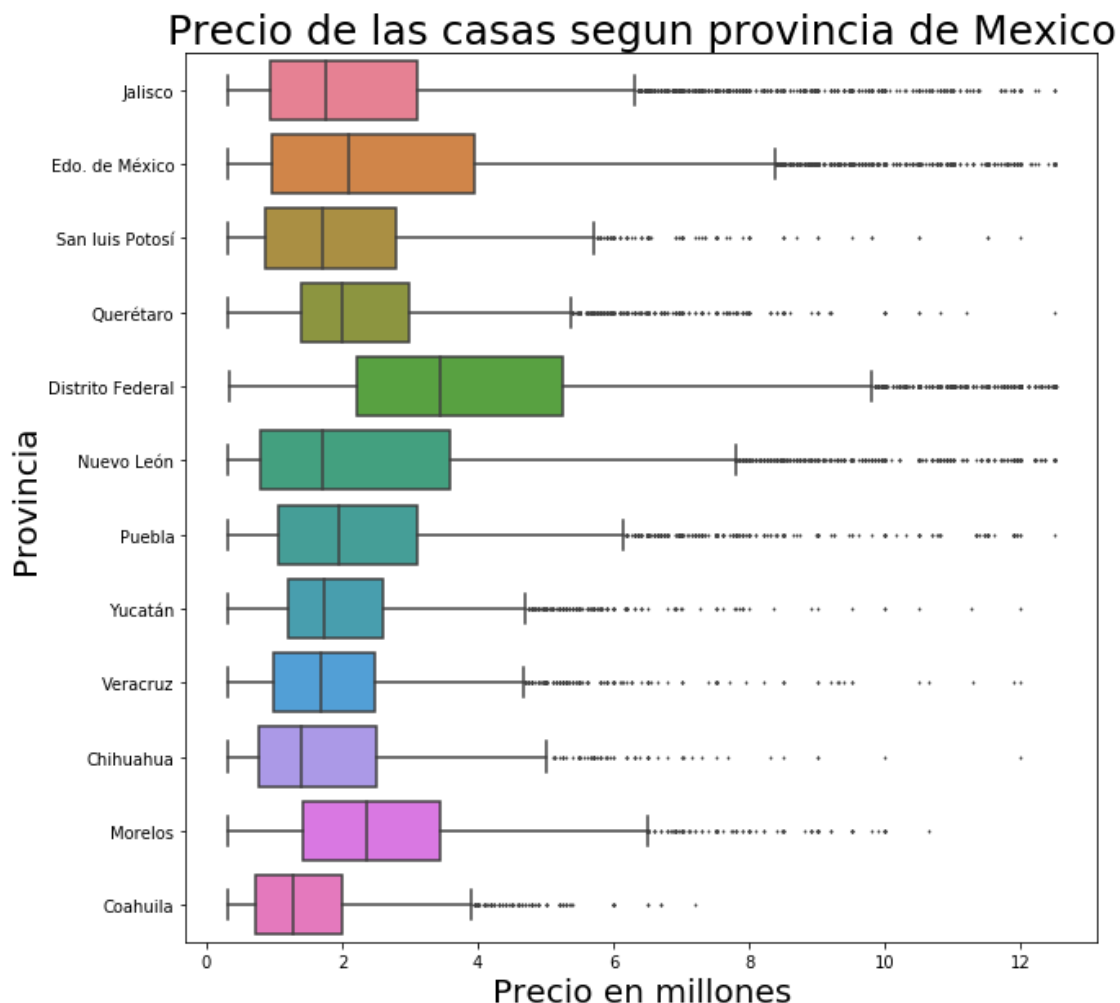


Para construir una visión más completa del comportamiento de precios en las provincias decidimos analizar con más profundidad como están distribuidos los precios mas allá de su valor promedio y concentrarnos en los casos más representativos de la página.

Para ello tomamos las provincias con la mayor cantidad de publicaciones de casas y apartamentos y los analizamos por separado.

5.2.5. Distribución precios de casas por provincia

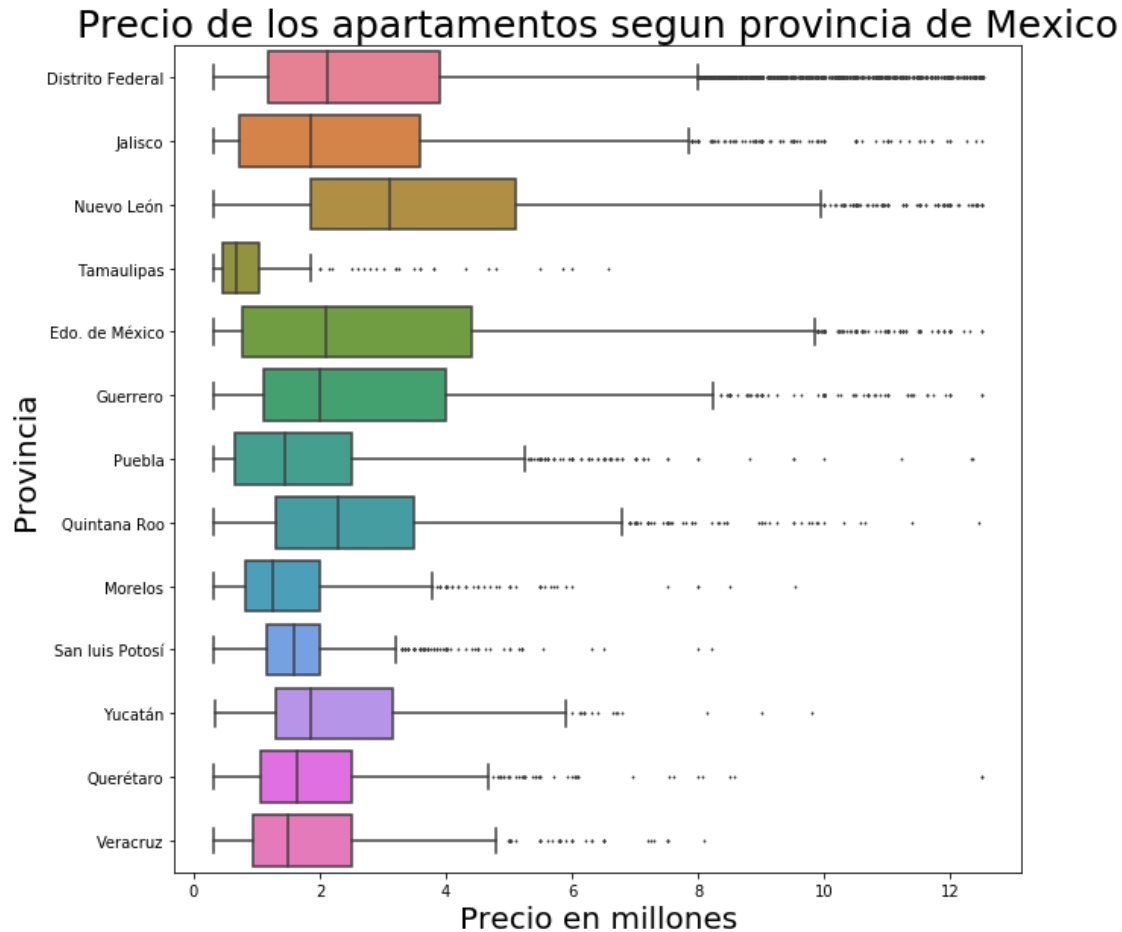
Como podemos apreciar en el siguiente gráfico todas las provincias tienen el mismo precio mínimo, por el contrario existen grandes diferencias en el máximo de precios lo que deja ver que D.F. y Estado de México suelen tener las casas mas costosas respaldando lo visto en los gráficos anteriores. Además, Estado de México, Distrito Federal y Nuevo León, son los que presentan un mayor rango cubierto por el boxplot.



5.2.6. Distribución precios de apartamentos por provincia

Por otro lado, al hacer un análisis sobre los apartamentos vemos que es Nueva León y no Distrito Federal la que alcanza el máximo valor (como se anticipó en la [sección 5.2.3]).

Nuevamente son Estado de México, Distrito Federal y Nuevo León las que presentan los mayores rangos cubiertos.



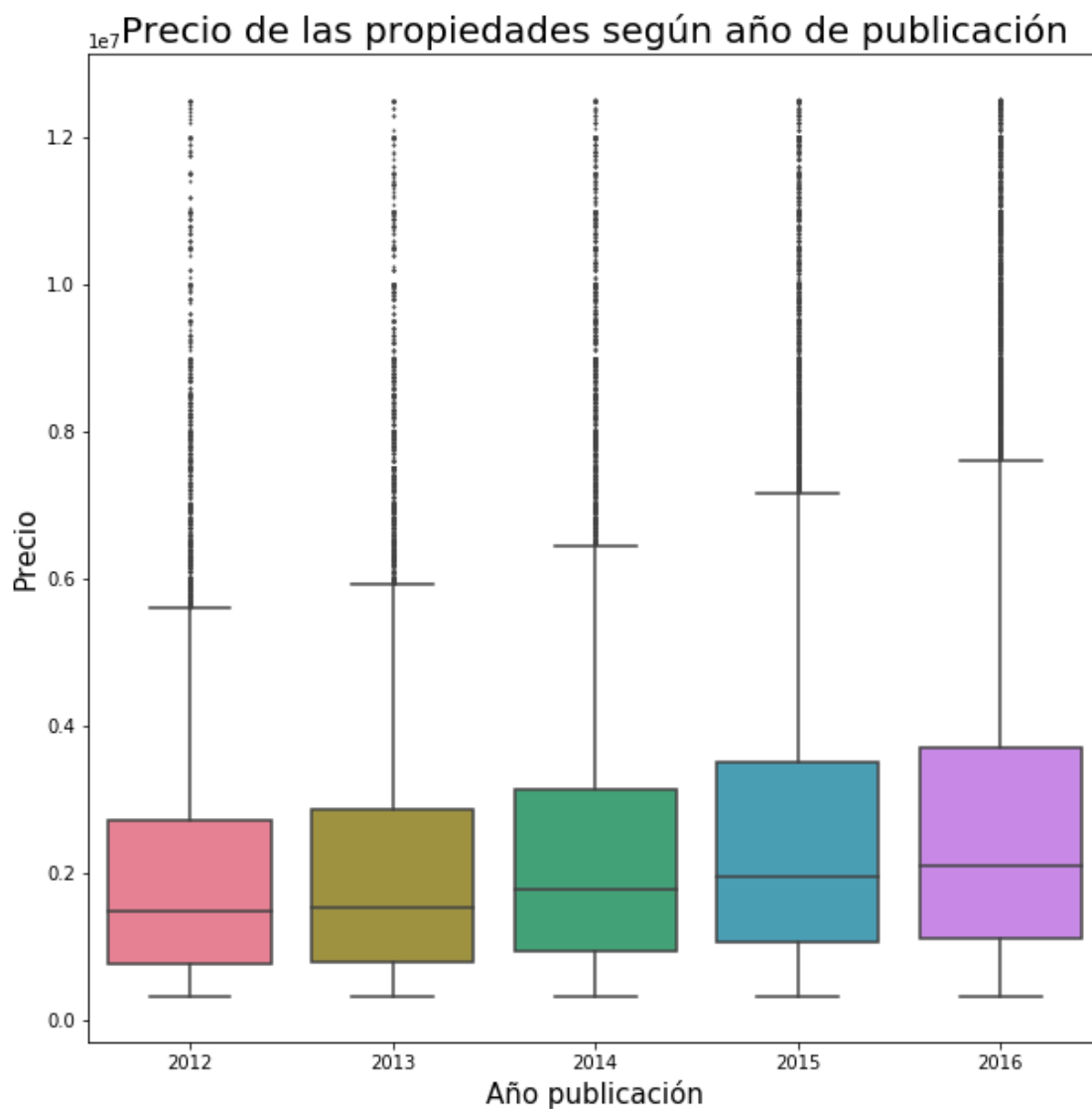
Con éste último gráfico concluye el análisis de precio por provincia del anuncio, que consideramos puede ser útil para un comprador que priorice más el precio que la provincia en la que se encuentra la propiedad, tanto como a agentes en el mercado inmobiliario para contrastar rápidamente la diferencia de propiedades similares entre distintas provincias.

5.3. Precio por año de publicación del anuncio

Para realizar este análisis, nos centramos en la columna "fecha", referida a la fecha de publicación del anuncio de la propiedad. De ésta, obtuvimos el año, y lo comparamos con el precio de las propiedades.

En el siguiente gráfico, podemos observar que, a medida que pasaron los años, aparecieron propiedades con mayor variedad de precios, como así también, propiedades más caras (donde también podría entrar en juego la inflación anual del país).

Esto es una buena señal ya que es una clara muestra de que la pagina tiene gran variedad de ofertas al menos en el rango de precios.



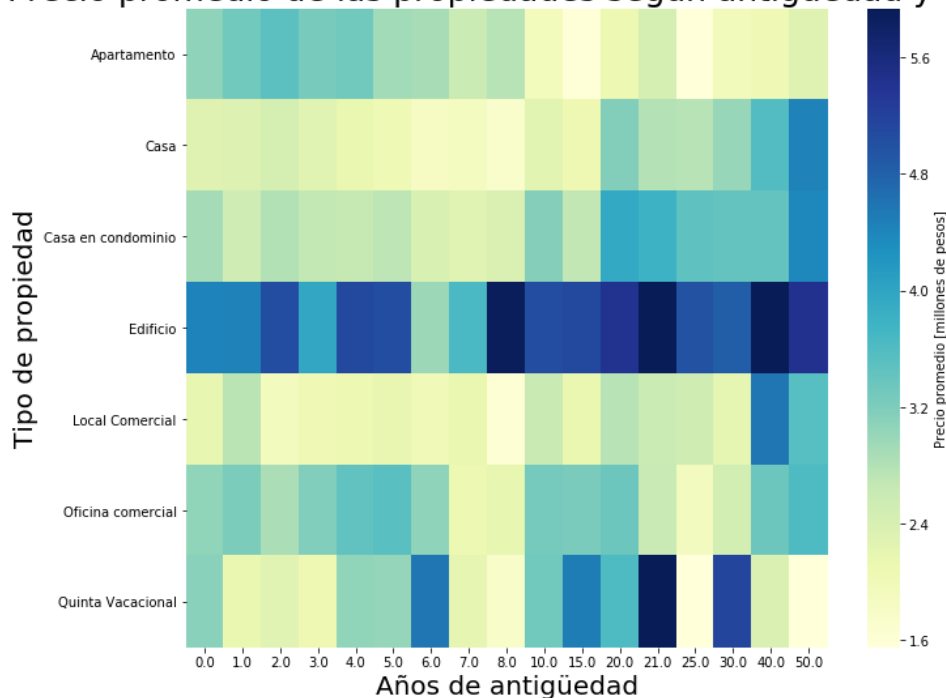
5.4. Precio por antigüedad según propiedad

Para éste análisis se observó qué relación tenía el precio con la antigüedad y tipos de propiedad que estaban publicados confeccionando un gráfico en el que cuanto más intenso sea el azul más costosas son las propiedades que tienen esa antigüedad.

Analizando el gráfico obtenido podemos observar que lo que uno podría suponer sobre la influencia de la antigüedad en el precio de una propiedad (que con el tiempo serian más baratas) no es el caso general, de hecho cada categoría parece tener su propia dinámica de precios para distintos valores de antigüedad.

- Las Casas, Casas en condominio, los Edificios y los Locales de comercio tienden a ser mas caros conforme pasa el tiempo.
- Entre ellos, los edificios suelen mantenerse prácticamente constante pero la tendencia es ligeramente ascendente (además, por tratarse de grandes construcciones poseen un precio mucho más alto)
- Por el contrario en los apartamentos la tendencia de precios es a la baja. Esto puede deberse a que en construcciones de esa envergadura mas antiguas los riesgos y costos de mantenimiento terminan desvalorizando la propiedad.

Precio promedio de las propiedades segun antigüedad y tipo



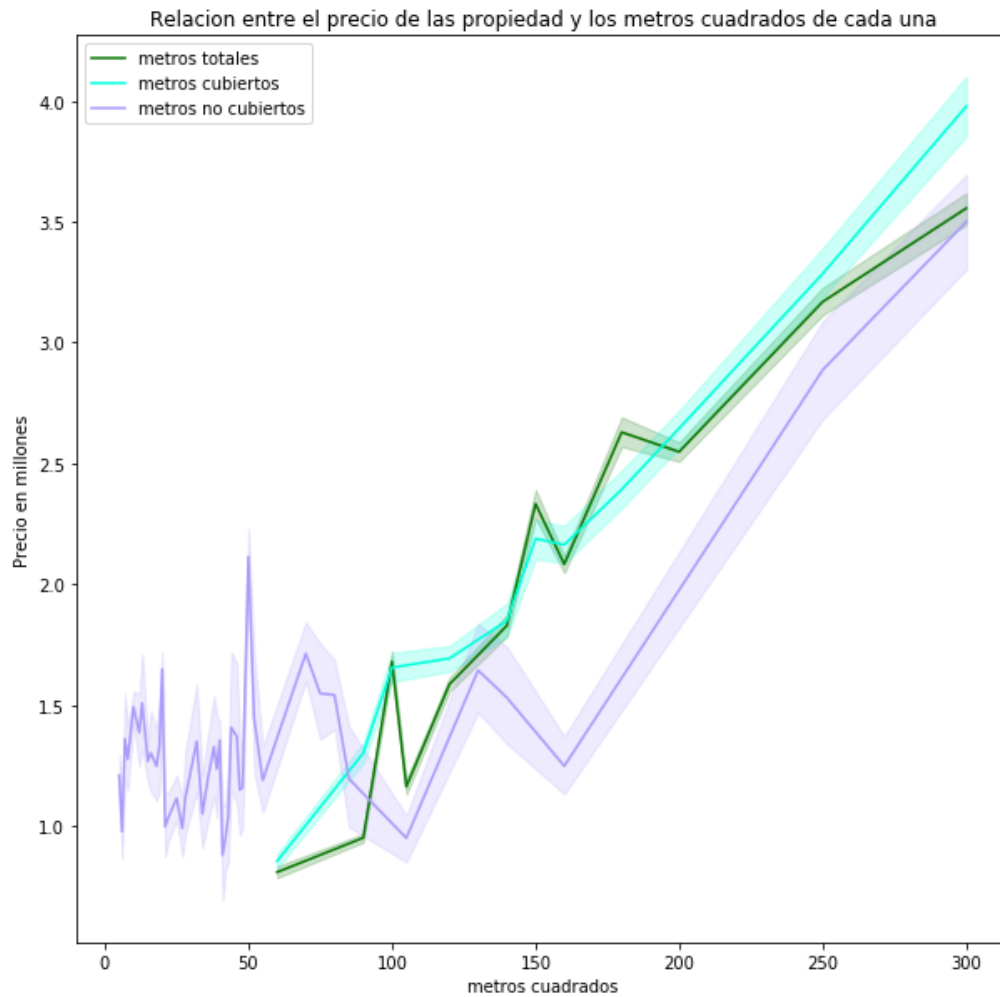
Esto ayuda al comprador en busca de una propiedad ya que la antigüedad es un buen punto a tener en cuenta a la hora de estimar cuánto podría costar una propiedad, por ejemplo, si busca un apartamento a estrenar deberá tener un presupuesto mayor que si no le importase que dicha propiedad tuviera un par de años de antigüedad.

5.5. Precio por metros cuadrados

5.5.1. Precio general por metros cuadrados

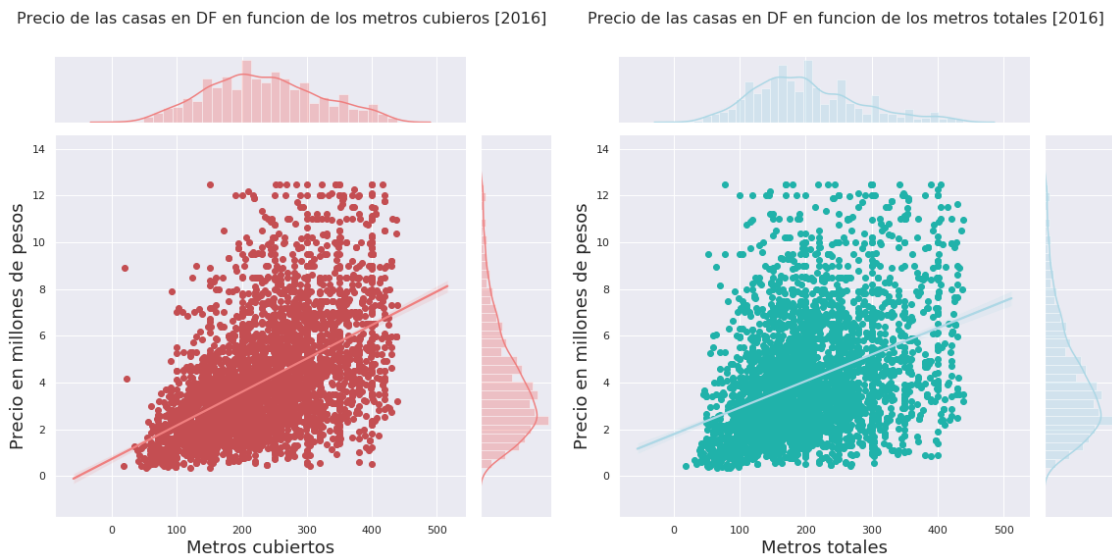
En esta sección analizaremos cómo afecta al precio de una propiedad la cantidad de metros cubiertos, totales y descubiertos que posee.

Esto mismo se puede apreciar en el siguiente gráfico dejando notar que lo mas costoso son los metros cubiertos, como uno podría inicialmente suponer.



Sin embargo, si decidimos representar todos los precios en función de los metros en un gráfico se ve que, pese a que hay una tendencia más o menos lineal, se debe especificar mucho la categoría y otros parámetros para ver mejor dicha linealidad por lo que esta tendencia solo es fácilmente distinguible con esta configuración.

Por ejemplo, en los siguientes dos gráficos se muestra los precios de las propiedades en función de los metros cubiertos para casas en México DF en el año 2016.



En el gráfico de la izquierda se ve que hay una cierta relación entre metros cubierto y el precio de la propiedad sobretodo entre 50m^2 y 250m^2 pero mientras más metros cubiertos posee la propiedad esta relación se hace cada vez menos lineal. Por otro lado, en el gráfico de la derecha se ve que la relación lineal de metros totales y precio se pierde mucho más rápido (entre 100m^2 y 200m^2) siendo los valores observados mucho mas dispersos entre si. Lo que si se puede observar es la presencia de un pequeño cumulo de propiedades con precios entre (\$2 Millones, \$4 Millones) y los (100m^2 , 250m^2).

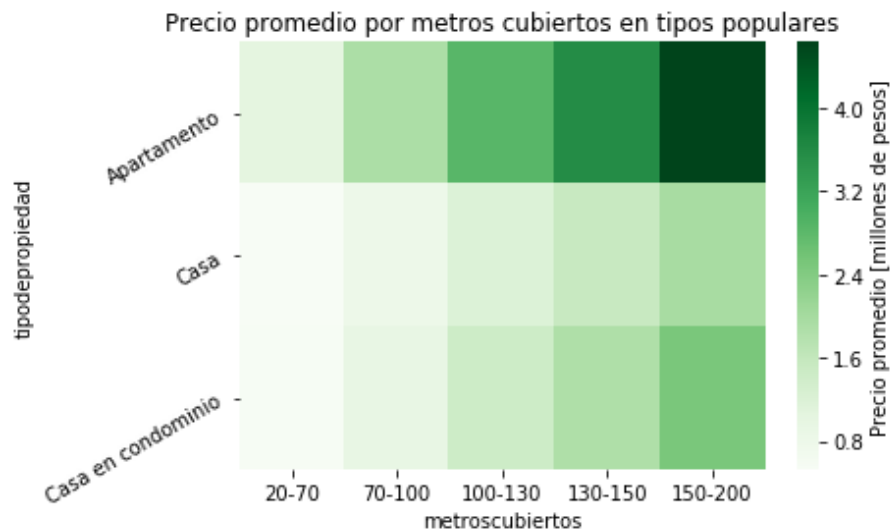
La lectura más importante que se puede sacar de esto es que pese a que se cumple la tendencia que uno esperaría ver (que en general las propiedades con mayor terreno y con mayor terreno cubierto tienden a valer más) esta relación no es tan fácil de medir, por lo que si en algún momento se quisiera aproximar un valor para la propiedad mediante algún algoritmo, se deberían tener en cuenta otros factores para dicha aproximación, por ejemplo: la provincia o la categoría de la propiedad, como ya se discutió anteriormente, o incluso la categoría de los servicios, que se discutirá en la siguiente sección.

5.5.2. Precio por metros cubiertos en tipos principales

Para tener una noción de los precios para viviendas típicas, se tomaron los tres tipos de propiedades con más anuncios (los que indica el gráfico en la sección [4.1]), y se calculó el promedio del precio para rangos de metros cubiertos.

Dado que los metros cubiertos eran valores continuos pero completamente aleatorios, se discretizaron los valores en distintos 'buckets' de interés, siendo éstos viviendas con:

- Entre 20 y 70 metros cubiertos
- Entre 70 y 100 metros cubiertos
- Entre 100 y 130 metros cubiertos
- Entre 130 y 150 metros cubiertos
- Entre 150 y 200 metros cubiertos

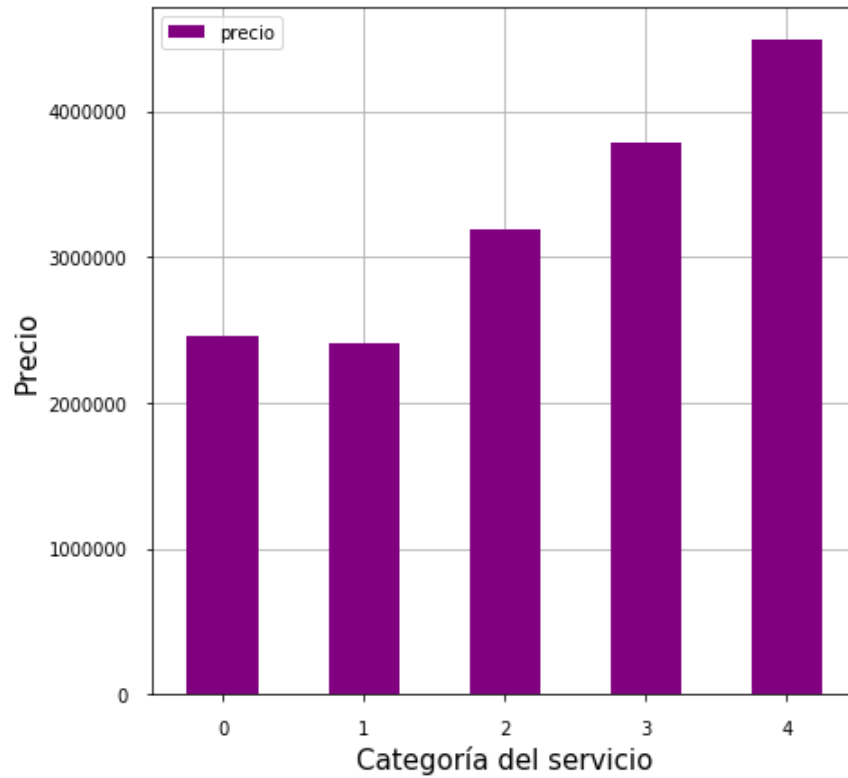


De esta manera, un potencial comprador de 'Apartamento' 'Casa' o 'Casa en Condominio' que definiría que tipo de propiedad comprar según su presupuesto puede observar que los Apartamentos son más caros, para todos los valores de metros cubiertos propuestos, seguidos por las Casas en condominio y finalmente siendo la opción mas económica, las Casas.

5.6. Precio de las propiedades según la categoría de los servicios

Se realizó un análisis que estudie la influencia de la categoría de los servicios adicionales (el criterio presentado en la sección [4.2]) y obtuvimos el siguiente gráfico:

Precio de las propiedades según la categoría de los servicios



Vemos que, como era de esperarse, a mayor categoría, es decir, a mayor cantidad de servicios adicionales provistos, mayor es el precio de las propiedades habitables.

6. Conclusiones

■ Evolución del sitio

Una de las conclusiones que más se evidencia es que la página presento un crecimiento sostenido a lo largo de los años, tanto en la cantidad de anuncios publicados como en la dispersión geográfica y en el alcance a ciudades.

Sin embargo, aún existen zonas de México que prácticamente no tienen anuncios publicados, por lo que una buena opción seria publicitar el sitio en dichos lugares para aumentar la cantidad de anuncios y beneficiar tanto a los compradores al tener más opciones para elegir qué propiedad comprar como al sitio en si al ganar mayor popularidad.

■ Persistencia de campos nulos

Uno de los puntos negativos, fue que en el tiempo no disminuyeron los campos completados con valor **nulo**, como mencionamos en la sección [2.4] nos da la pauta de que los anunciantes no se sienten protegidos por la pagina e intentan brindar solo información indispensable que no los comprometa. Esto podría en un futuro perjudicar al sitio, ya que si la gente no brinda datos como la dirección se podrían perder potenciales compradores que frente a no saber la ubicación de la propiedad pierden el interés en la misma.

Dentro de las variantes que podrían controlar este inconveniente, nuestra propuesta es convertir el campo **dirección** en obligatorio, que también exista una verificación con un servicio como Google Maps que pueda obtener automáticamente las **coordenadas** y garantizar al anunciante que la dirección sólo sera vista por otros usuarios que estén interesados en la propiedad, para que el mismo brinde toda la información requerida sin temor a que la misma pueda perjudicarlo.

Para el caso de los campos como **habitaciones** proponemos que se los haga de tipo *select* con los valores numéricos necesarios, y un valor especial para donde no corresponda hablar de cantidad de habitaciones. i.e.: Terreno, Garage.

■ Uso del Sitio

El sitio presenta más actividad de publicación los días Viernes, con un 18 % de los anuncios y presenta su menor actividad los Domingos con tan solo un 10 %. Además, la cantidad de anuncios publicados por día aumenta a medida que nos acercamos a fin de año, de hecho el período en el que más anuncios se publican es entre Diciembre y Enero.

■ Características de las propiedades y sus anuncios

En primer lugar, observamos que la mayor cantidad de anuncios son de **Casas, Apartamentos y Casas en Condominio** Como ya dijimos en secciones anteriores, se podría fomentar nuevas publicaciones de otro tipo de propiedades.

Además son más frecuentes las publicaciones de propiedades que no poseen tantos servicios, estimamos es debido a que la tendencia en la distribución de precios es hacia el extremo de los mas baratos, y estos son los que poseen menos servicios adicionales.

Por otro lado, la cantidad de baños por número de habitaciones para las propiedades disminuye considerablemente su crecimiento a partir de la habitación número 5.

Otro de los aspectos a destacar es que la mayoría de los anuncios son de propiedades relativamente nuevas (con una antigüedad cercana o igual a 0 años).

Al analizar cuáles eran las provincias que alojan mayor cantidad de propiedades nuevas, se obtuvo que la provincia con más porcentaje de este tipo de propiedades es Yucatán.

Por último, quisimos ver cuáles eran los años de construcción de propiedades habitables más frecuentes, y obtuvimos que la mayoría de las propiedades fueron construidas alrededor del año 2016 (año más actual que posee el set de datos).

■ Precio de las propiedades

Con respecto al precio de las propiedades, hemos visto que responde a la ubicación, a la cantidad de metros cuadrados, a la antigüedad y a la categoría de servicios.

Sería interesante considerar la posibilidad de incluir filtros según el presupuesto con el que cuenta el usuario, que tenga en cuenta todas estas variantes para la zona en que se hace la búsqueda y en consecuencia recomendarle la propiedad que mejor se ajusta a su bolsillo. Por ejemplo: un usuario de México DF con un rango de precio limitado se le podrían recomendar apartamentos, ya que estos son las viviendas más accesibles en dicha provincia, mientras que en otra provincia como Veracruz las recomendaciones en ese caso serían Casas.