

Universidad de Buenos Aires
Facultad De Ingeniería
Año 2019 - Segundo Cuatrimestre



75.06 - Organización de Datos
Trabajo Práctico

Análisis Exploratorio de Datos

Integrantes

Nombre	Padrón
Pablo Daniel Garcia	75322
Diego Ripetour	86601
Jhon Valer	93066
Josué Giovanni Valdivia	93075

Repositorio de github: <https://github.com/Giova262/OrgDatTP>

Introducción	3
Descripción de los atributos	3
Carga de datos	5
Análisis exploratorio General¶	6
Tamaño del set de datos	6
Análisis de datos faltantes¶	6
Análisis del atributo título	7
Análisis del atributo precio	8
Análisis del atributo provincia	9
Análisis del atributo dirección	10
Análisis del atributo ciudad	11
Análisis del atributo descripción	12
Análisis del atributo tipo de propiedad	13
Análisis de correlaciones¶	14
Análisis de correlación de todos los datos¶	14
Análisis de correlación entre centrocomercialescercanos y escuelas cercanas	16
Análisis de correlación entre metros cubiertos y metros totales	17
Análisis de correlación entre metros cubiertos y precio	19
Análisis de precios por día de la semana	20
Análisis exploratorio de precios¶	21
Análisis de los precios a lo largo de los años	21
Análisis de precio promedio por cantidad de habitaciones¶	22
Análisis de precio promedio por cantidad de garages¶	23
Análisis de precio promedio por cantidad de baños¶	24
Análisis de precio por tipo de propiedad	25
Análisis de precios por ubicación geográfica	27
Análisis exploratorio de publicaciones	28
Análisis de cantidad de publicaciones por año	28
Análisis de cantidad de publicaciones realizadas por día de la semana	28
Análisis de cantidad de publicaciones realizadas por mes	29
Análisis de cantidad publicaciones por mes y día de la semana	30
Análisis de características de una propiedad	31
Análisis de cantidad de propiedades según cantidad de garages y baños	31
Análisis de las características de cada propiedad individual	32
Análisis de cantidad de habitaciones y la antigüedad de las mismas	34
Análisis de servicios y publicaciones	35
Análisis de servicios y provincias	35
Análisis de la distribución geográfica de las propiedades	37
Análisis de escuelas cercanas	37
Análisis de piscinas	38
Análisis de propiedades con usos múltiples	39
Análisis de propiedades con centrocomerciales cercanos	39
Análisis de propiedades con gimnasios	40
Análisis de tipos de propiedades en el mapa	41
Análisis de precio por metro cuadrado	42
Análisis de precio por metro y provincia para el año 2012 y 2016	44

Análisis de precio por metro y antigüedad de la propiedad	45
Análisis de precio por metro y cantidad de habitaciones de la propiedad	46
Conclusiones	47

Introducción

El presente trabajo práctico se enfoca en el análisis de datos del registro histórico de publicaciones en www.zonaprop.com.ar. realizadas en México.

El set de datos consta de propiedades en venta en México entre los años 2012 y 2016 , valuadas en pesos mexicanos.

Los datos tienen un formato que puede ser leído en forma de filas y columnas, en el cual cada fila representa una propiedad publicada en zonaprop y cada columna representa una característica o atributo distinto de las propiedades.

Descripción de los atributos

Observamos que hay atributos que describen las publicaciones, atributos que describen las localizaciones de las propiedades, atributos que describen las características básicas de las propiedades y atributos que describen características adicionales.

Atributos de publicaciones		
Nombre	Tipo	Descripción
titulo	String	Título de la propiedad publicada
descripcion	String	Descripción de la propiedad publicada
fecha	Date	Fecha de publicación
precio	Float	Valor de la publicación de la propiedad en pesos mexicanos

Atributos de localización		
Nombre	Tipo	Descripción
direccion	String	Dirección de la propiedad
ciudad	String	Ciudad de la propiedad
provincia	String	Provincia de la propiedad
idzona	Integer	La zona es un valor numérico correspondiente a una parte de la ciudad
lat	Float	Latitud geográfica de la propiedad
lng	Float	Longitud geográfica de la propiedad

Atributos de características básicas		
Nombre	Tipo	Descripción
tipopropiedad	String	Tipo de propiedad (casa, apartamento, terreno, etc)
metrostotales	Integer	Metros totales de la propiedad
metroscubiertos	Integer	Metros cubiertos de la propiedad
antigüedad	Integer	Antigüedad de la propiedad
habitaciones	Integer	Cantidad de habitaciones
garages	Integer	Cantidad de garages
banos	Integer	Cantidad de baños

Atributos de características adicionales		
Nombre	Tipo	Descripción
gimnasio	Boolean	Indica si la propiedad tiene un gimnasio
usosmultiples	Boolean	Indica si la propiedad tiene un SUM
piscina	Boolean	Indica si la propiedad tiene una piscina
escuelascercanas	Boolean	Indica si la propiedad tiene escuelas cerca
centroscomercialescercanos	Boolean	Indica si la propiedad tiene centros comerciales cerca

Carga de datos

Si bien el set de datos no era muy pesado, se cambiaron los tipos de datos de algunas columnas para facilitar luego el tratamiento de estas a lo largo del análisis exploratorio , las columnas cambiadas de tipo fueron :

- id : int32
- titulo : category
- fecha : str
- ciudad :category
- provincia : category
- tipodepropiedad : category

Quedando de la siguiente manera :

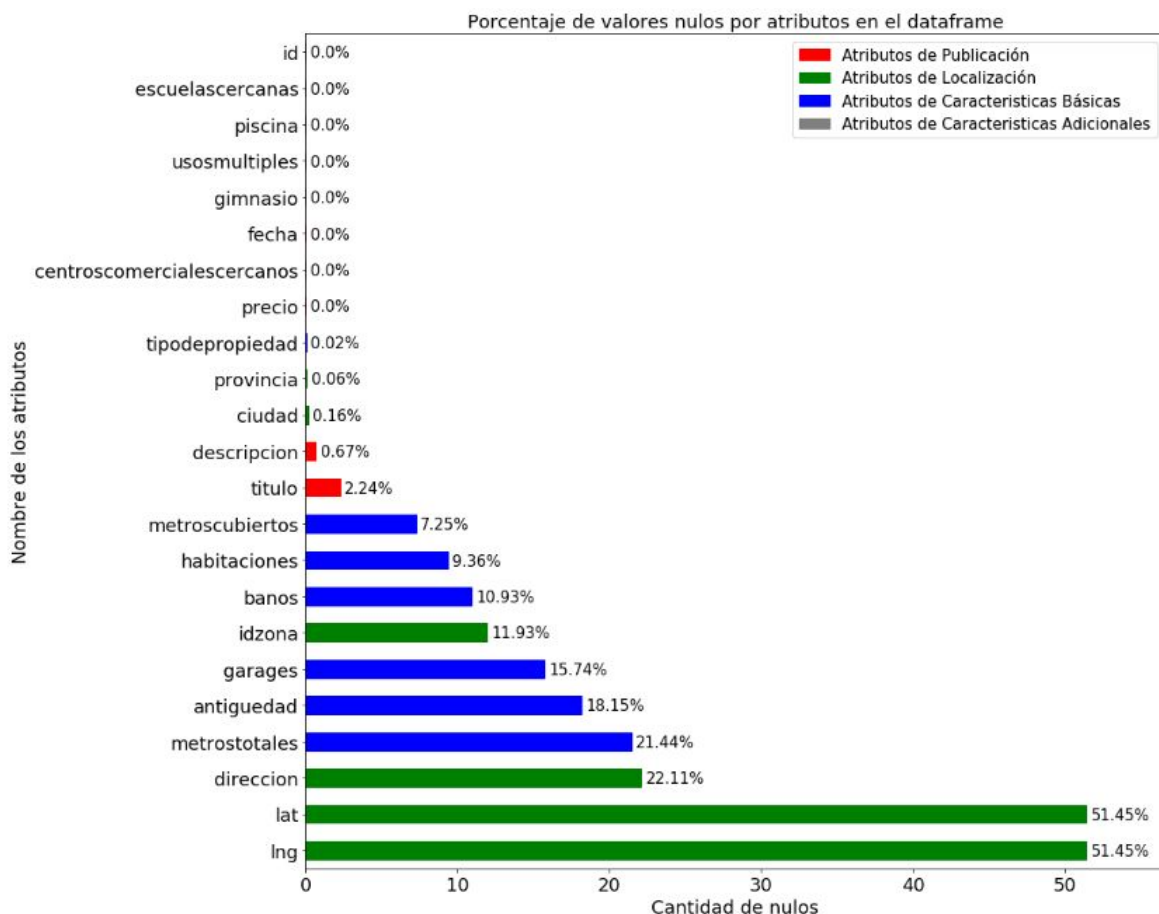
```
Data columns (total 23 columns):
id                240000 non-null int32
titulo            234613 non-null category
descripcion       238381 non-null object
tipodepropiedad  239954 non-null category
direccion         186928 non-null object
ciudad            239628 non-null category
provincia         239845 non-null category
antiguedad        196445 non-null float64
habitaciones      217529 non-null float64
garages           202235 non-null float64
banos             213779 non-null float64
metroscubiertos  222600 non-null float64
metrostotales     188533 non-null float64
idzona           211379 non-null float64
lat              116512 non-null float64
lng              116512 non-null float64
fecha            240000 non-null datetime64[ns]
gimnasio         240000 non-null float64
usosmultiples    240000 non-null float64
piscina          240000 non-null float64
escuelascercanas 240000 non-null float64
centroscomercialescercanos 240000 non-null float64
precio           240000 non-null float64
dtypes: category(4), datetime64[ns](1), float64(15), int32(1), object(2)
memory usage: 42.2+ MB
```

Análisis exploratorio General

Tamaño del set de datos

- Filas : 240000
- Columnas : 23

Análisis de datos faltantes



- Analizando el porcentaje de datos faltantes se puede observar una cantidad demasiado grande de faltantes para los datos de longitud y latitud (faltan más de la mitad de los datos).
- Habrá que tener en cuenta que el peso de estos faltantes es muy grande, y las decisiones que se tomen con estos datos pueden influir mucho en el trabajo de análisis.
- También hay muchos datos faltantes de dirección y metros totales (aunque algo menor, sólo un 20% aproximadamente)
- Se puede observar que hay más datos con faltantes pero el porcentaje es menor al 20%, lo cual es un número más razonable.

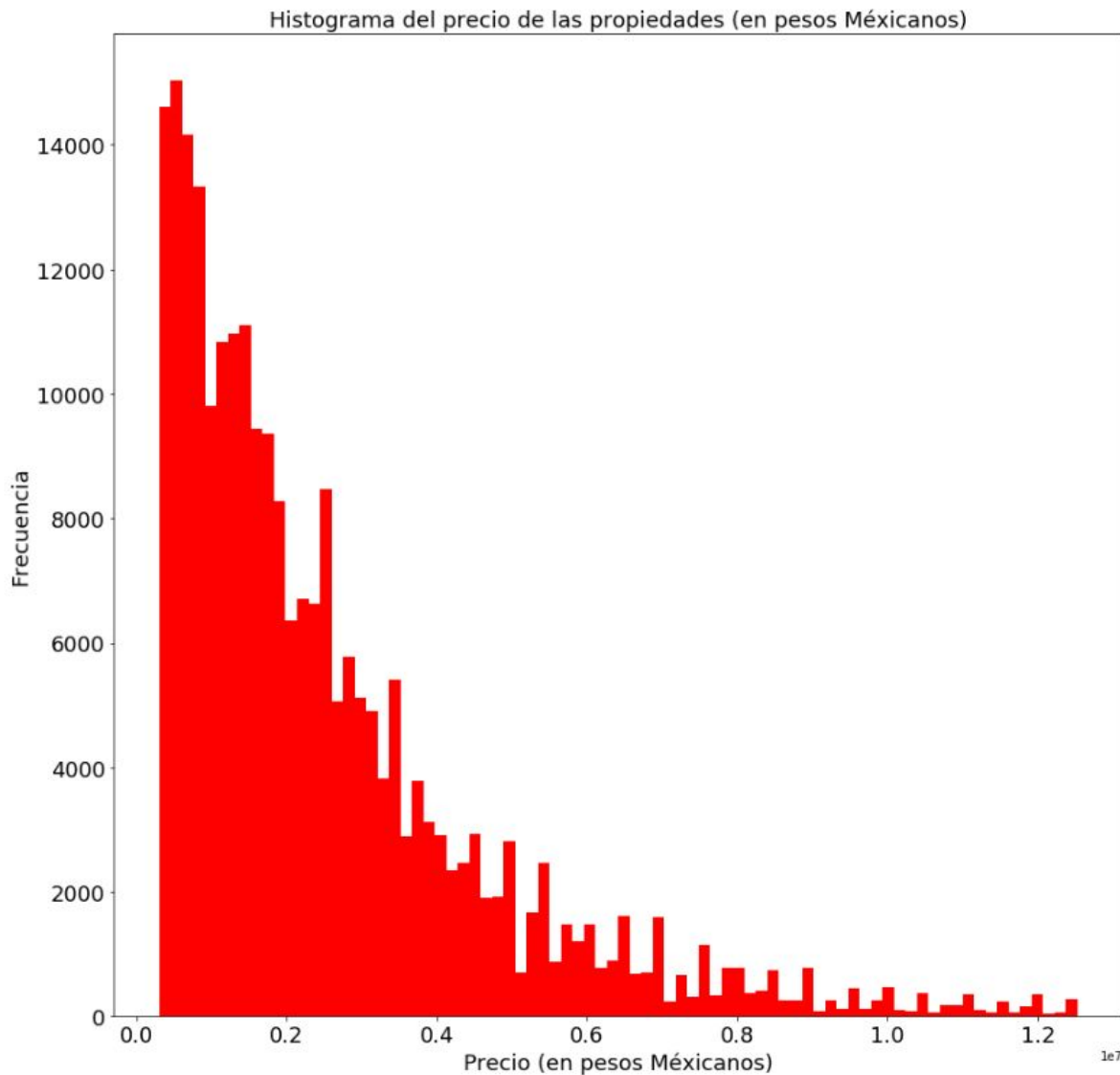
- ## Análisis del atributo título

[illegible]

- Hay una cantidad bastante variada de títulos de publicaciones
- El título más utilizado en las publicaciones es "Casa en venta", seguido por "Departamento en venta"
- Los siguientes títulos más usados en las publicaciones son aquellos que hacen referencia a casa en venta en Chihuahua y en Querétaro

- A continuación los siguientes títulos más usados en las publicaciones son aquellos que hacen referencia a departamento en venta en Benito Juárez y Miguel Hidalgo

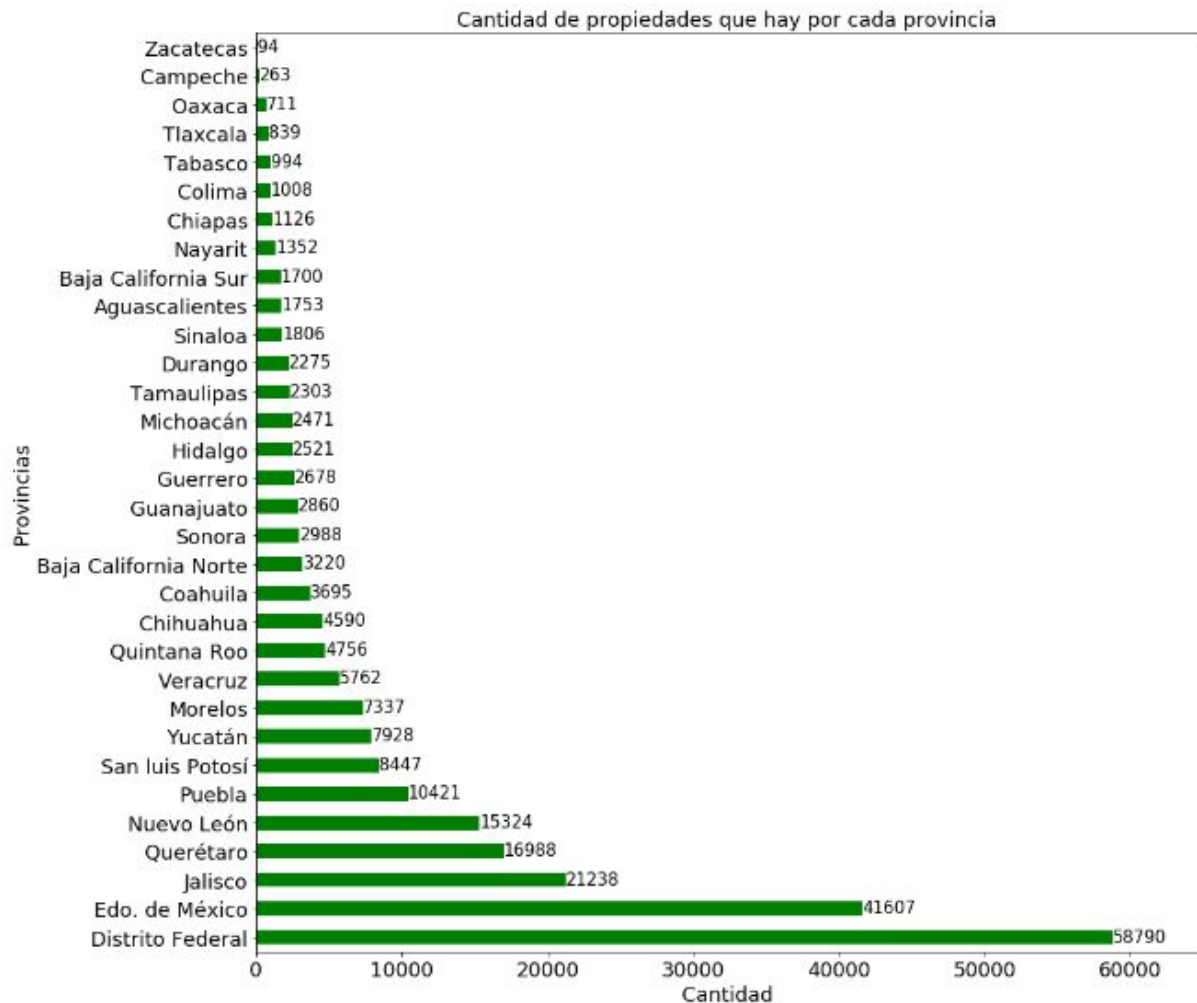
Análisis del atributo precio



Del análisis del gráfico se puede observar que:

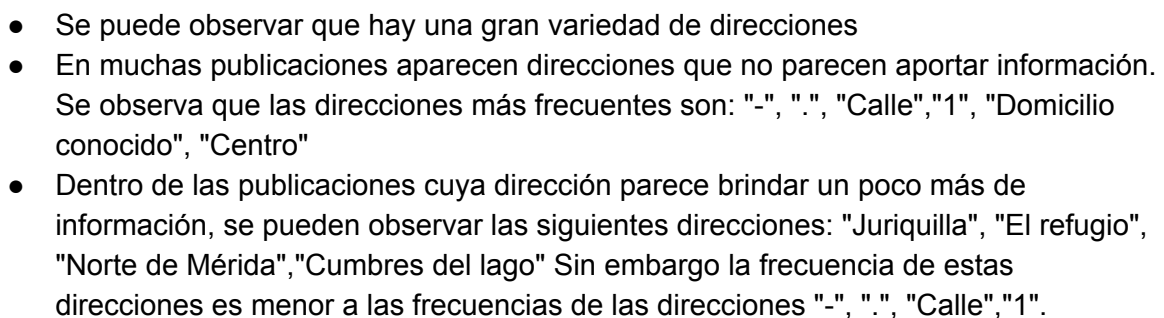
- La mayoría de las propiedades valen menos de 5000000, pero hay valores de algunas propiedades que pueden llegar a valer 12000000
- El gráfico presenta una cola larga hacia la derecha

Análisis del atributo provincia



- Del análisis del gráfico se puede observar que la mayor cantidad de propiedades se encuentran en el Distrito Federal (58790 propiedades), seguida de las que se encuentran en el Estado de México(416007). El Distrito Federal es la capital del país, y el Estado de México limita con el Distrito Federal, siendo estas dos provincias las más pobladas del país. (fuente: https://es.wikipedia.org/wiki/Anexo:Entidades_federativas_de_M%C3%A9xico_por_superficie,_poblaci%C3%B3n_y_densidad)
- Luego la tercer provincia con más propiedades es Jalisco, pero los valores comienzan a caer a cifras mucho más pequeñas (Jalisco posee 21238 propiedades)
- Luego le siguen en orden decreciente: Querétaro ,Nuevo León, Puebla, San Luis Potosí,Yucatán
- La provincia con menos propiedades es Zacatecas, que solo posee 94 propiedades

Direcciones usadas con más frecuencia en publicaciones (cuanto más grande sea más frecuente es)



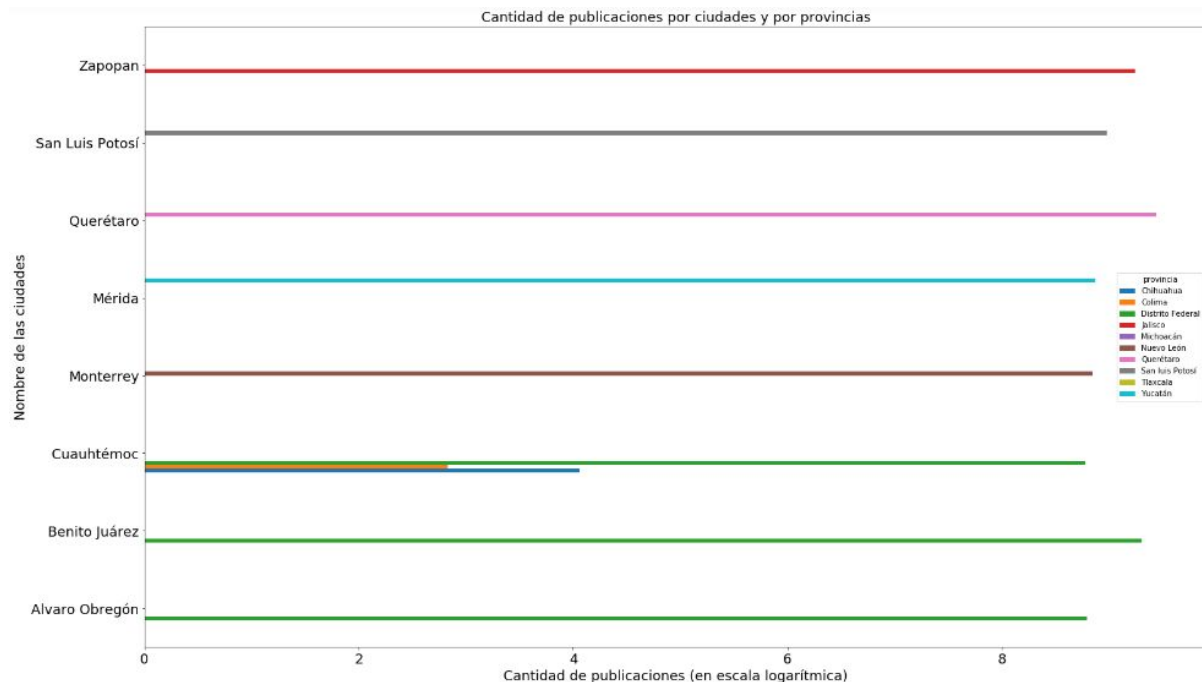
Análisis del atributo ciudad



Se observa que:

- Hay una gran cantidad de ciudades
- Las ciudades con más publicaciones son: Querétaro, Benito Juárez, Zapopan, San Luis Potosí, Mérida, Monterrey, Alvaro Obregón, Cuauhtémoc. Si llegasen a existir muchas ciudades con el mismo nombre dispersas por distintas provincias de México podría parecer que hay muchas publicaciones de una misma ciudad cuando en realidad esto no ocurre.

A continuación analizaremos si hay varias ciudades de las nombradas anteriormente que tienen igual nombre pero están dentro de provincias diferentes, y cuantas publicaciones corresponden a cada una:



- Hay varias ciudades llamadas Cuauhtémoc en distintas provincias de México. Sin embargo la alta frecuencia de publicaciones, correspondiente a estas ciudades, no viene dada porque haya muchas ciudades con el mismo nombre dispersas por todo el país, sino porque la mayoría de publicaciones de estas ciudades son las que corresponden al Distrito Federal (la cual es una provincia con muchas publicaciones)
- Además recordando al análisis de la cantidad de propiedades publicadas por provincia, las demás provincias que aparecen en este gráfico (Jalisco, Querétaro, Nuevo León, San Luis Potosí, Yucatán) también están entre las que tienen más publicaciones

Análisis del atributo descripción

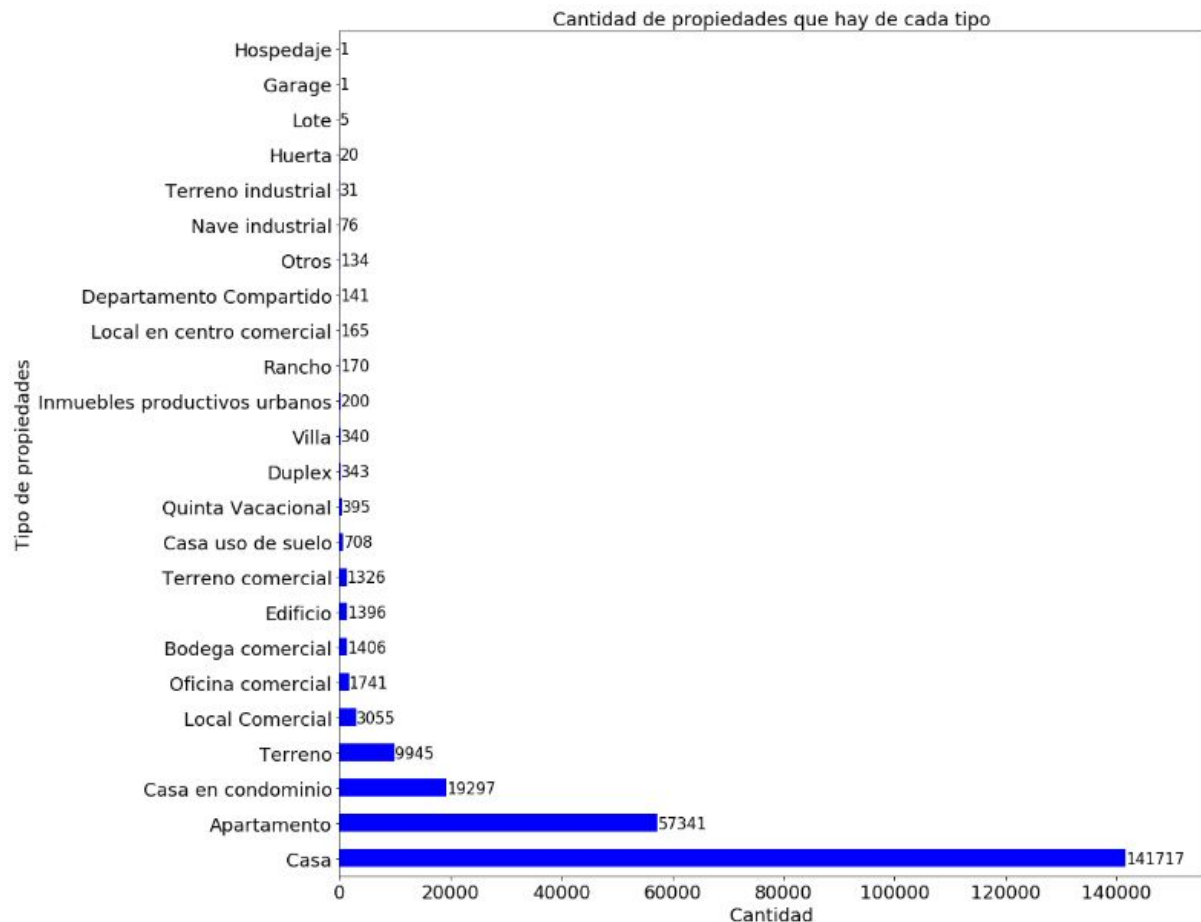
En la descripción de las propiedades publicadas la más frecuente que se presenta en este set de datos es :

empresa dedicada a los ante notario y registro público tiene grandes oportunidades para ti desde un 50-70% debajo de su avalúo, cesiones de derechos con garantía hipotecaria en buenas condiciones en venta únicamente de contado no creditos. si no sabes que es un remate hipotecario no te preocupes nosotros te explicamos. inversión segura, pago de contado.
las imágenes y descripción aquí presentadas pueden no representar con exactitud la realidad por tener diferencias en relación con la numeración, ubicación y superficies. (r) no es inmobiliaria sino media derechos crediticios litigiosos y/o adjudicatarios.

- Se puede ver que la cantidad de propiedades en venta con esta publicación se deben a remates debido a alguna deuda o por otros motivos

Análisis del atributo tipo de propiedad

- Hay 24 tipos de propiedades que se han publicado , la cantidad de publicaciones de cada tipo de propiedad se muestra en el siguiente gráfico :

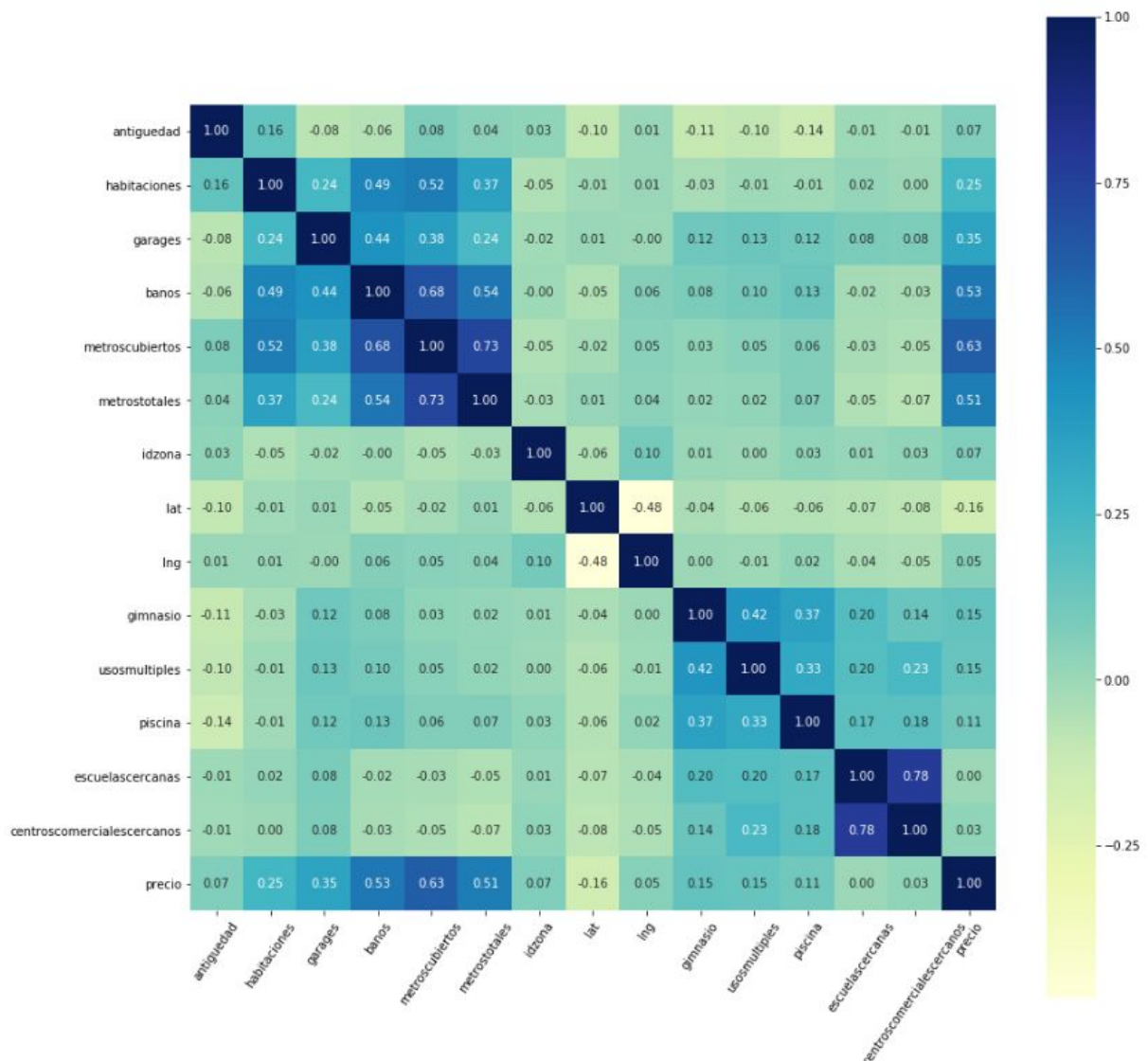


- Del análisis del gráfico se puede observar que la mayor parte de las propiedades son Casas (141717), seguida por los Apartamentos (57341)
- Luego siguen las Casas en condominio (habiendo 19297 publicaciones), pero los valores comienzan a caer a cifras mucho más pequeñas
- Se puede observar que solo hay una propiedad que sea Garage y una sola que es Hospedaje

Análisis de correlaciones

Análisis de correlación de todos los datos

Dos variables se correlacionan positivamente si cuando una variable aumenta, la otra tiende a aumentar



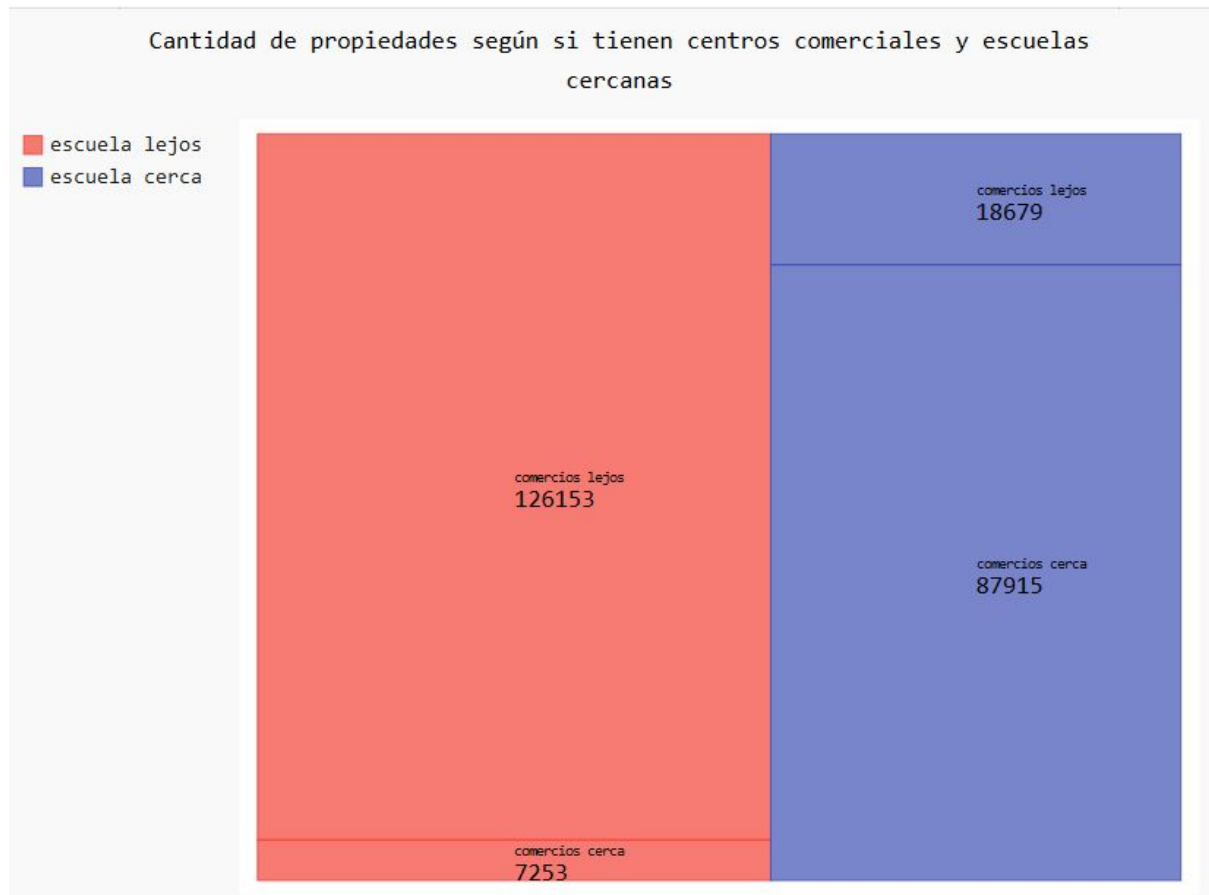
El coeficiente de correlación mide la relación entre variables cuantitativas 1 para la máxima correlación y -1 como el mínimo.

Por ejemplo se ve como metros cubiertos, metros totales, baños, garages, habitaciones están relacionadas con el precio y como el precio no se ve muy relacionado con si hay escuelas cercanas o centros comerciales

A continuación se muestra un cuadro con los cuatro pares de variables con mayor correlación positiva y su análisis:

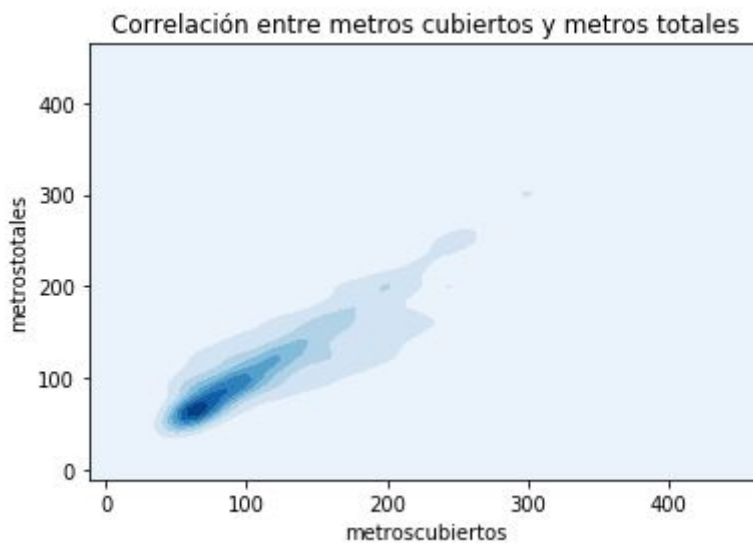
Variable 1	Variable 2	Correlacion	Análisis
centrocomercialescercanos	escuelascercanas	78	Las propiedades que están cerca a centros comerciales tienden a estar cerca de escuelas
metrostotales	metroscubiertos	73	Las propiedades con más metros cubiertos tienden a ser las que tienen más metros totales
banos	metroscubiertos	68	Las propiedades más grandes tienden a tener más baños que las propiedades chicas
precio	metroscubiertos	63	Las propiedades más grandes tienden a ser más caras que las propiedades más chicas

Análisis de correlación entre centros comerciales cercanos y escuelas cercanas



- Hay más propiedades publicadas que se encuentran lejos de escuelas (126153) que propiedades que están cerca de escuelas (7253)
- Como se explicó anteriormente las propiedades que se encuentran cerca de centros comerciales tienden a estar cerca de escuelas
- Hay más propiedades publicadas que se encuentran cerca de escuelas y cerca de centros comerciales (87915), que las que se encuentran cerca de escuelas y lejos de centros comerciales (18679)
- Las propiedades que se encuentran lejos de escuelas tienden a estar lejos de centros comerciales
- Hay más propiedades publicadas que se encuentran lejos de escuelas y lejos de centros comerciales (126153), que las que se encuentran lejos de escuelas y cerca de centros comerciales (7253)

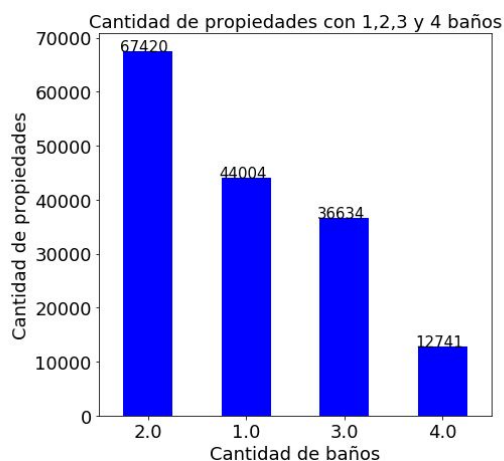
Análisis de correlación entre metros cubiertos y metros totales

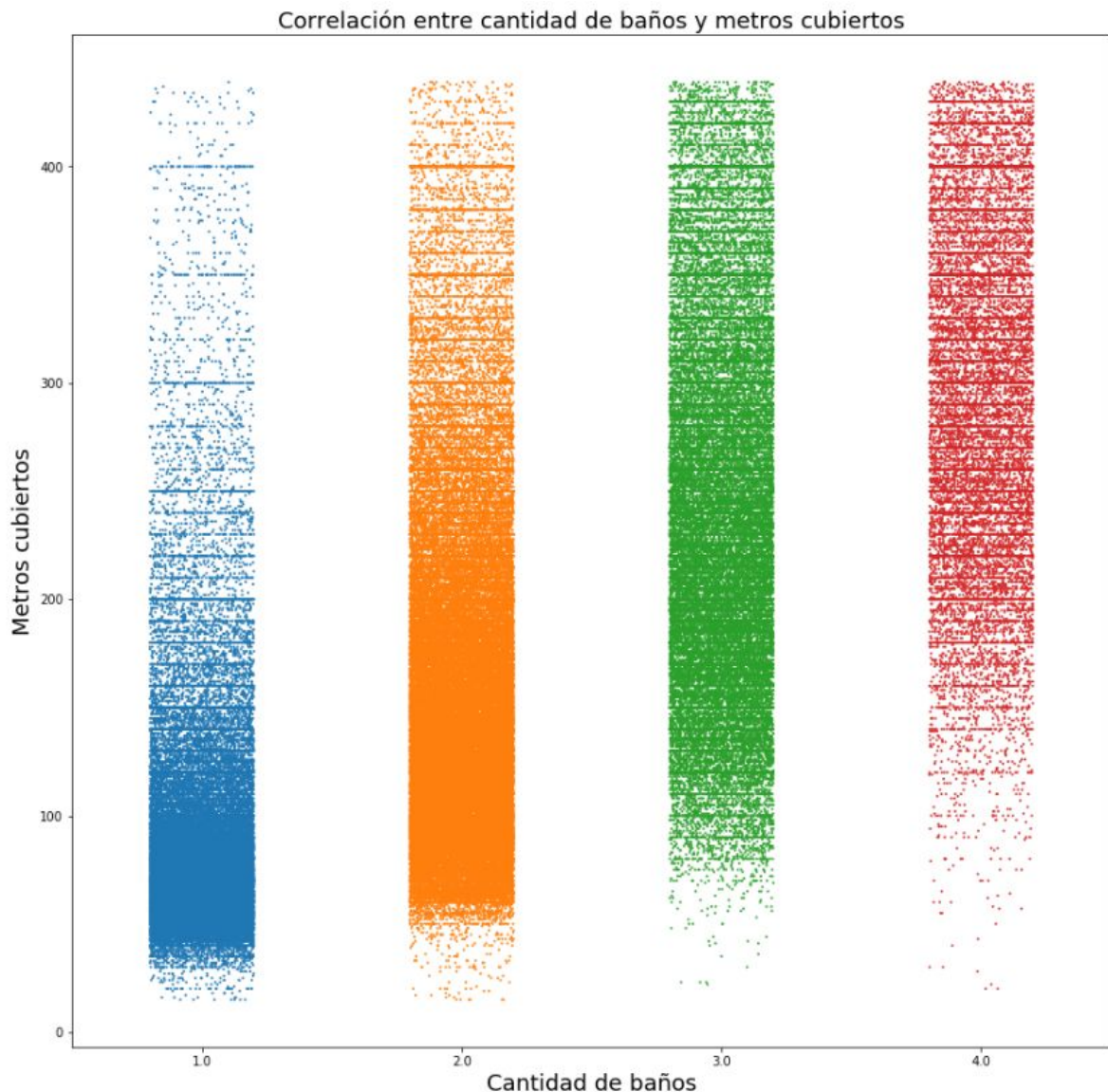


- Se puede observar que, como se explicó anteriormente, hay una tendencia en la cual a medida que las propiedades tienen más metros cubiertos, las mismas tienen más metros totales
- El gradiente de color indica que hay una mayor cantidad de propiedades con pocos metros cubiertos y totales, y que a medida que aumentan los metros cubiertos y totales la cantidad de propiedades disminuye (es decir que tiende a haber más casas chicas que casas grandes publicadas)

Análisis entre cantidad de baños y metros cubiertos

Como la variable cantidad de baños es categórica primero voy a graficar cuantas propiedades hay con 1, 2, 3 y 4 baños para comprender mejor estos datos.





Del análisis del gráfico se puede observar que:

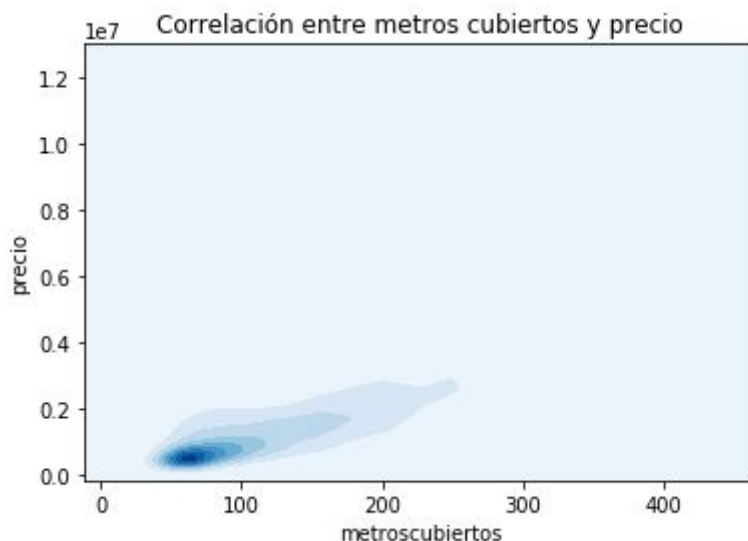
- La cantidad de propiedades con 2 baños es mayor a las que tienen 1 baño y además son las propiedades más publicadas
- A su vez la cantidad de propiedades con 1 baño es mayor a las que tienen 3 baños y hay más propiedades con 3 baños que propiedades con 4 baños
- Como se explicó anteriormente hay una tendencia en la cual a medida que las propiedades tienen más baños, la cantidad de metros cubiertos es mayor
- Es curioso que en el gráfico hay una cierta cantidad de propiedades con muchos metros cubiertos y tan solo 1 o 2 baños, aunque está claro que son las propiedades con muchos metros cubiertos las que tienen más baños (esto podría explicarse en que tal vez esas propiedades no se usan como vivienda sino como garages o galpones por ejemplo)

A continuación voy a analizar cuantas propiedades de cada tipo tienen un baño y más de 400 metros cubiertos para chequear si son viviendas o no:



- Al contrario que lo que indicaría la intuición, la mayor parte de las propiedades con más de 400 metros cubiertos y un solo baño corresponden a casas

Análisis de correlación entre metros cubiertos y precio

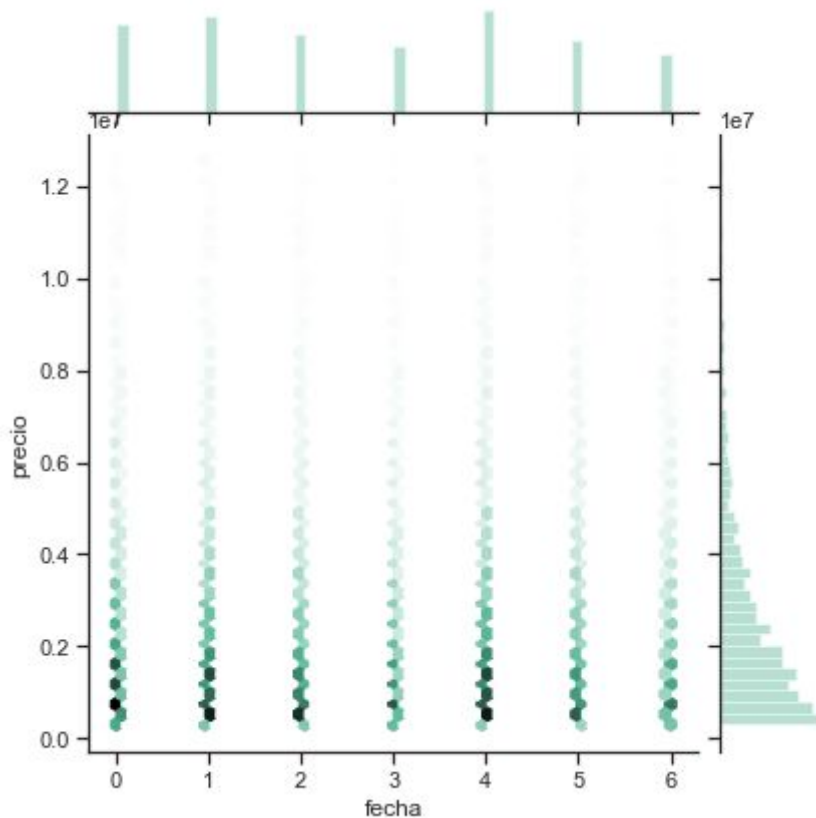


Del análisis del gráfico se puede observar que:

- Como se explicó anteriormente hay una tendencia en la cual a medida que las propiedades tienen más metros cubiertos, las mismas son más caras
- El gradiente de color indica que hay una mayor cantidad de propiedades con pocos metros cubiertos y más baratas, y que a medida que aumentan los metros cubiertos y los precios la cantidad de propiedades disminuye (es decir que tiende a haber más casas chicas y baratas que casas grandes y caras publicadas)

- Adicionalmente se puede ver que la relación creciente entre los metros cubiertos y el precio tiene una pendiente menor a la relación creciente entre los metros cubiertos y los metros totales (mostrados en un gráfico anterior) justamente porque la correlación entre metros cubiertos y metros totales es de 73 y en cambio la correlación entre metros cubiertos y precios es de 63 (o sea la correlación es menor)

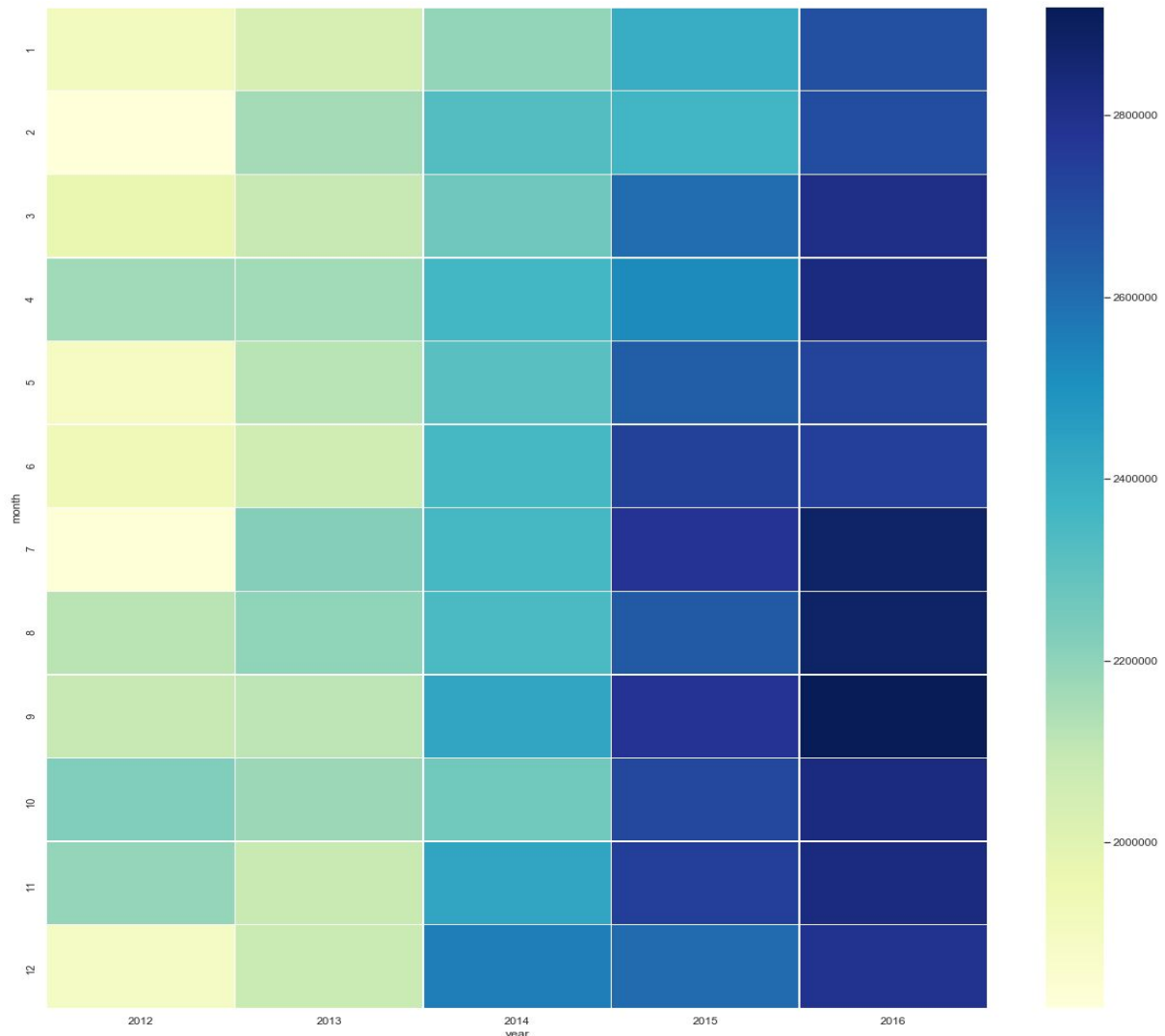
Análisis de precios por día de la semana



- Muestra por cada día de la semana los valores de las propiedades que se publicaron en cada una de ellas , donde 0 = Lunes , 1 = Martes , 2 = Miercoles , 3 = Jueves , 4 = Viernes , 5 = Sabado y 6 = Domingo
- No se ve ninguna fuerte tendencia en que día se publican las propiedades mas caras pero si un leve decrecimiento en los domingos

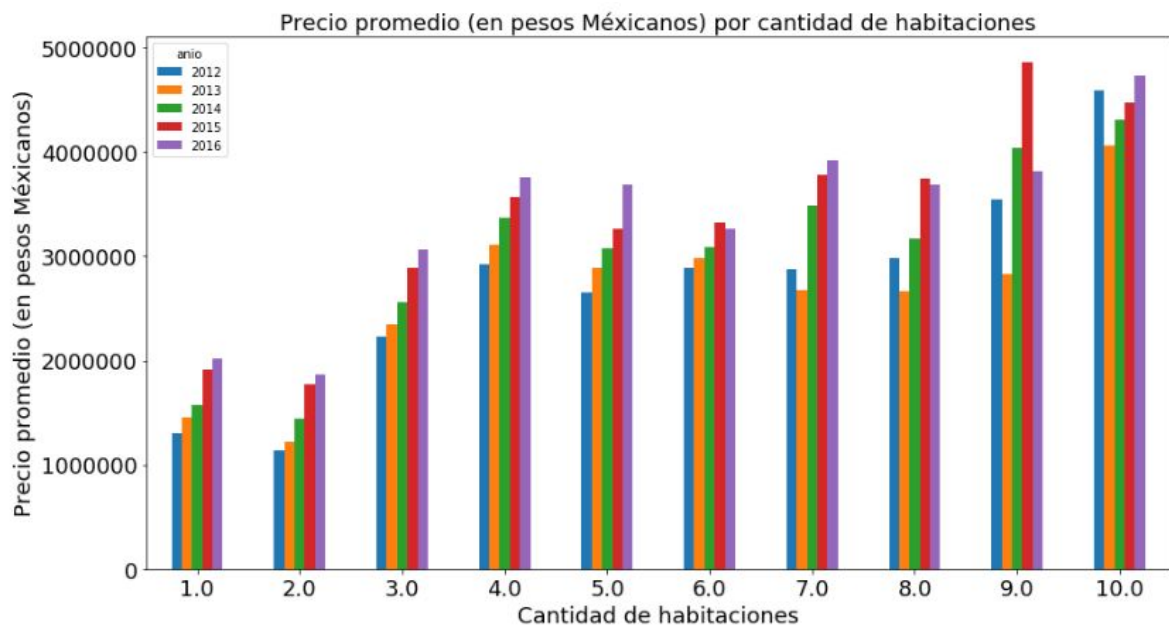
Análisis exploratorio de precios

Análisis de los precios a lo largo de los años



- Durante el transcurso del 2012 al 2016 se pueden ver el aumento de los precios de las propiedades , dando una idea de la inflación de la divisa mexicana que sufrió en ese periodo de tiempo
- El aumento de los precios de año a año se ve claramente en el heat map

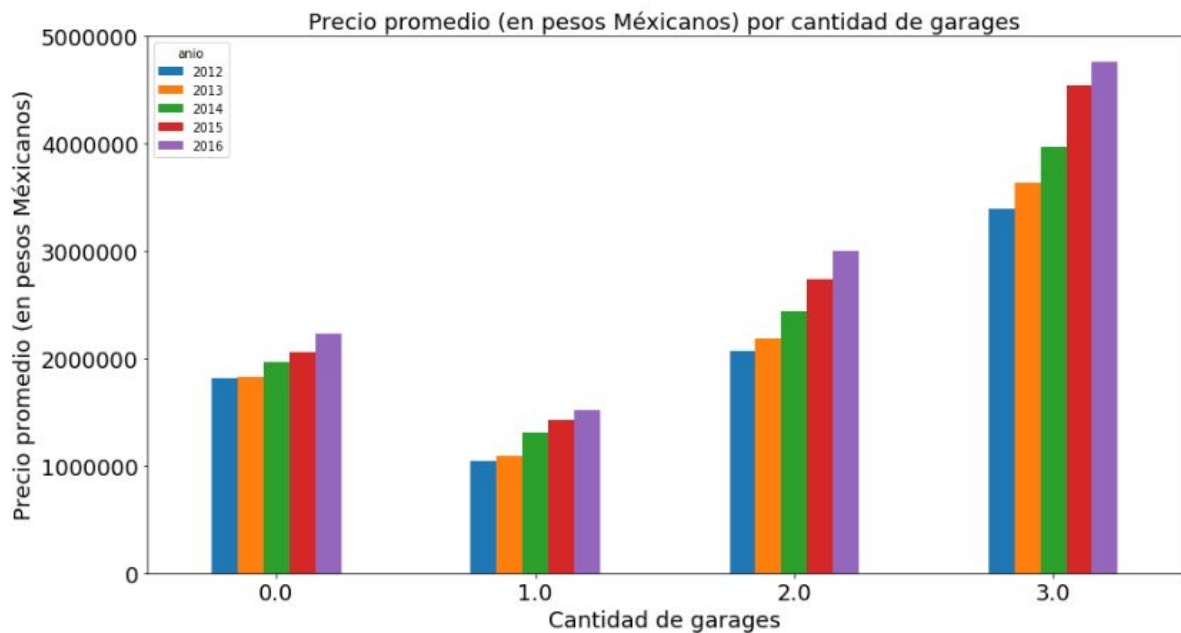
Análisis de precio promedio por cantidad de habitaciones



Del análisis del gráfico se puede observar que:

- A medida que transcurren los años el precio promedio de las propiedades con 1 a 5 habitaciones aumenta
- El precio promedio de las propiedades con 6 habitaciones aumenta año a año, pero en el año 2016 baja respecto del valor del año 2015 pero se mantiene por arriba del año 2014
- El precio promedio de las propiedades con 7 habitaciones baja en el año 2013, pero después aumenta año a año
- El precio promedio de las propiedades con 8 y 9 habitaciones baja en el año 2013, pero después aumenta año a año. Aunque vuelve a bajar en el año 2016. En el caso de propiedades con 9 habitaciones se destaca una suba muy alta en el año 2015 que luego parece corregirse a valores previos en el año siguiente.
- El precio promedio de las propiedades con 10 habitaciones baja significativamente en el año 2013 pero luego aumenta año a año

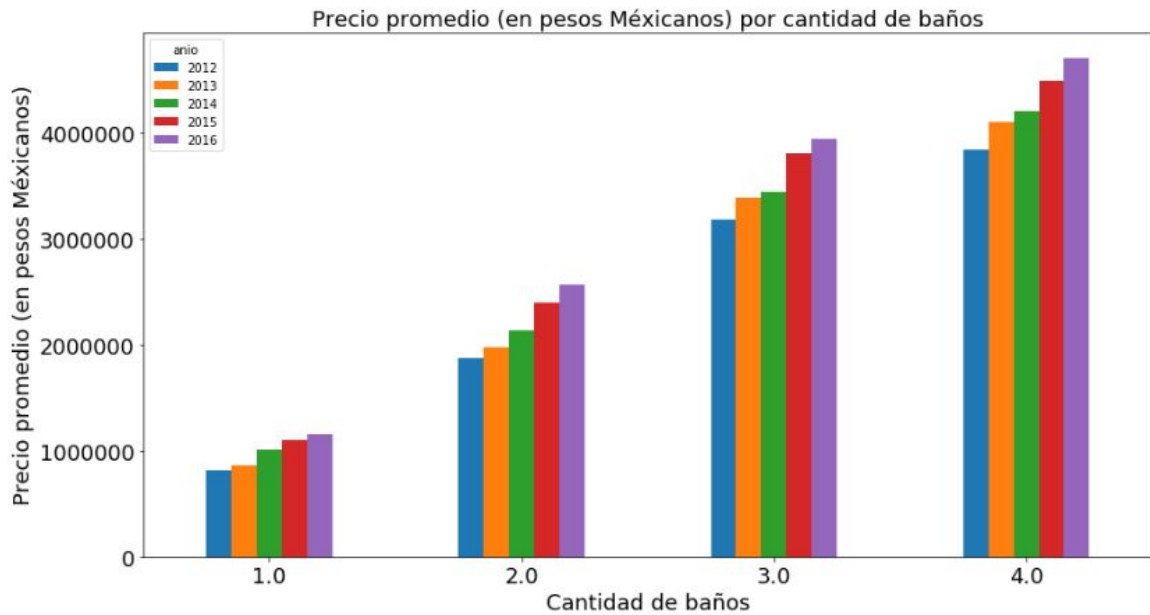
Análisis de precio promedio por cantidad de garages



Del análisis del gráfico se puede observar que:

- Que el precio promedio de las propiedades sin garage es mayor al de las que tienen 1 garage (esto se cumple para todos los años)
- Que el precio promedio de las propiedades sin garage es menor al de las propiedades con 2 garages, que a su vez es menor al precio promedio de las propiedades con 3 garages
- Que a medida que transcurren los años el precio promedio de las propiedades aumenta ya sea que tengan 0,1,2 o 3 garages (una posible explicación de esto podría ser la inflación de la moneda Mexicana)

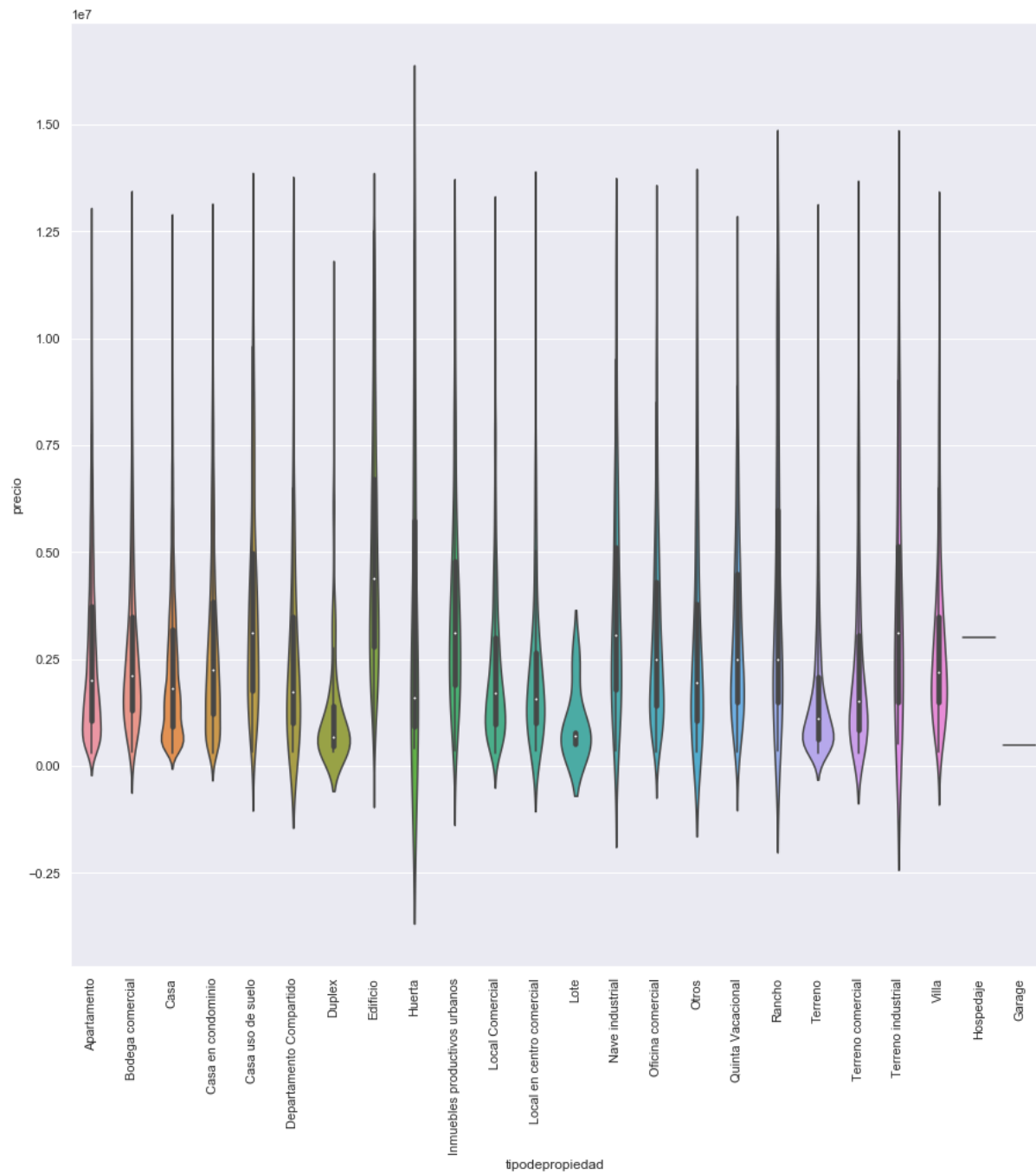
Análisis de precio promedio por cantidad de baños

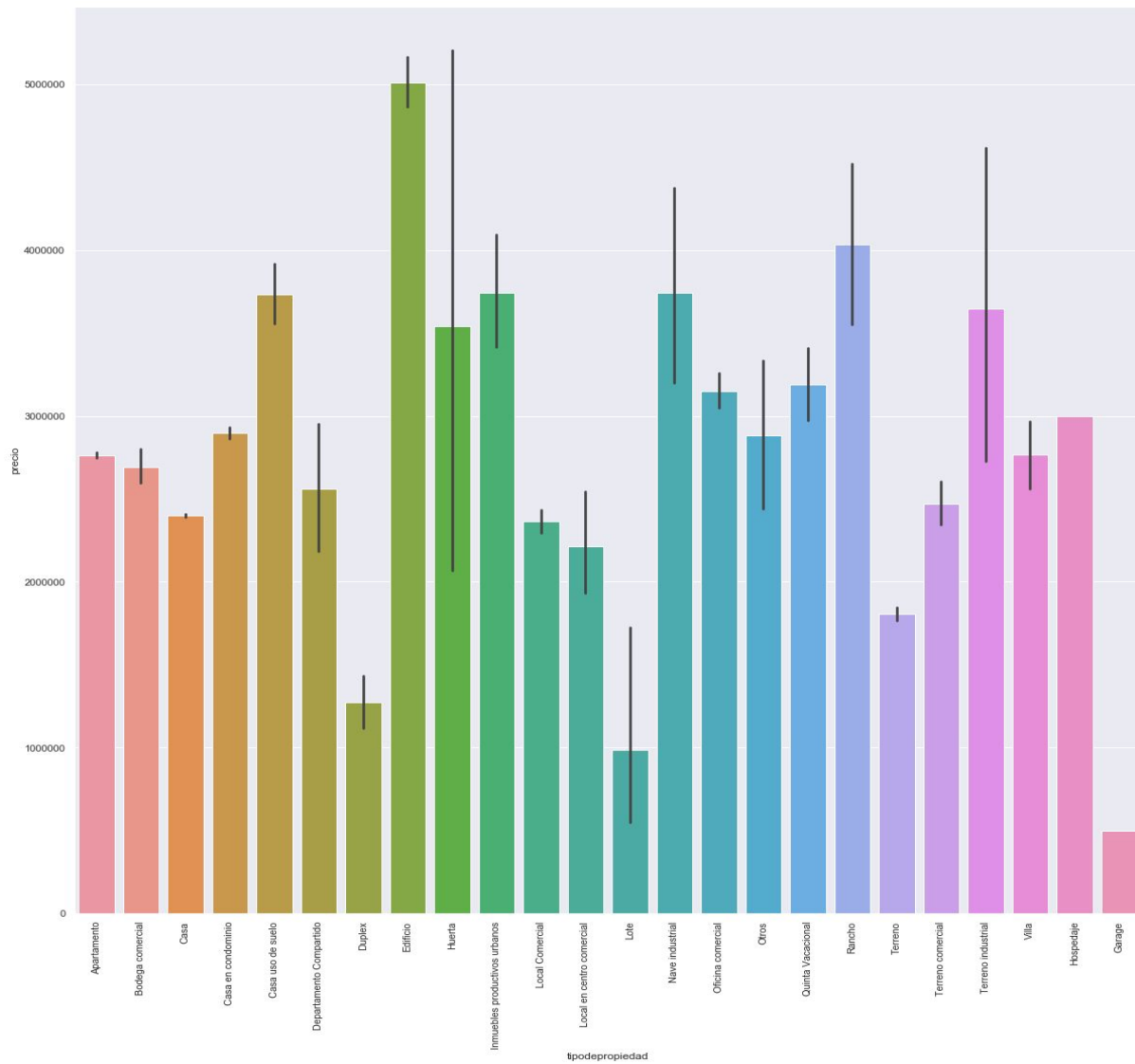


Del análisis del gráfico se puede observar que:

- Que se cumple para todos los años (de los que se tienen datos) cuantos más baños tiene una propiedad más cara resulta en promedio
- Que a medida que transcurren los años el precio promedio de las propiedades aumenta ya sea que tengan 1,2,3 o 4 baños (una posible explicación de esto podría ser la inflación de la moneda Mexicana)

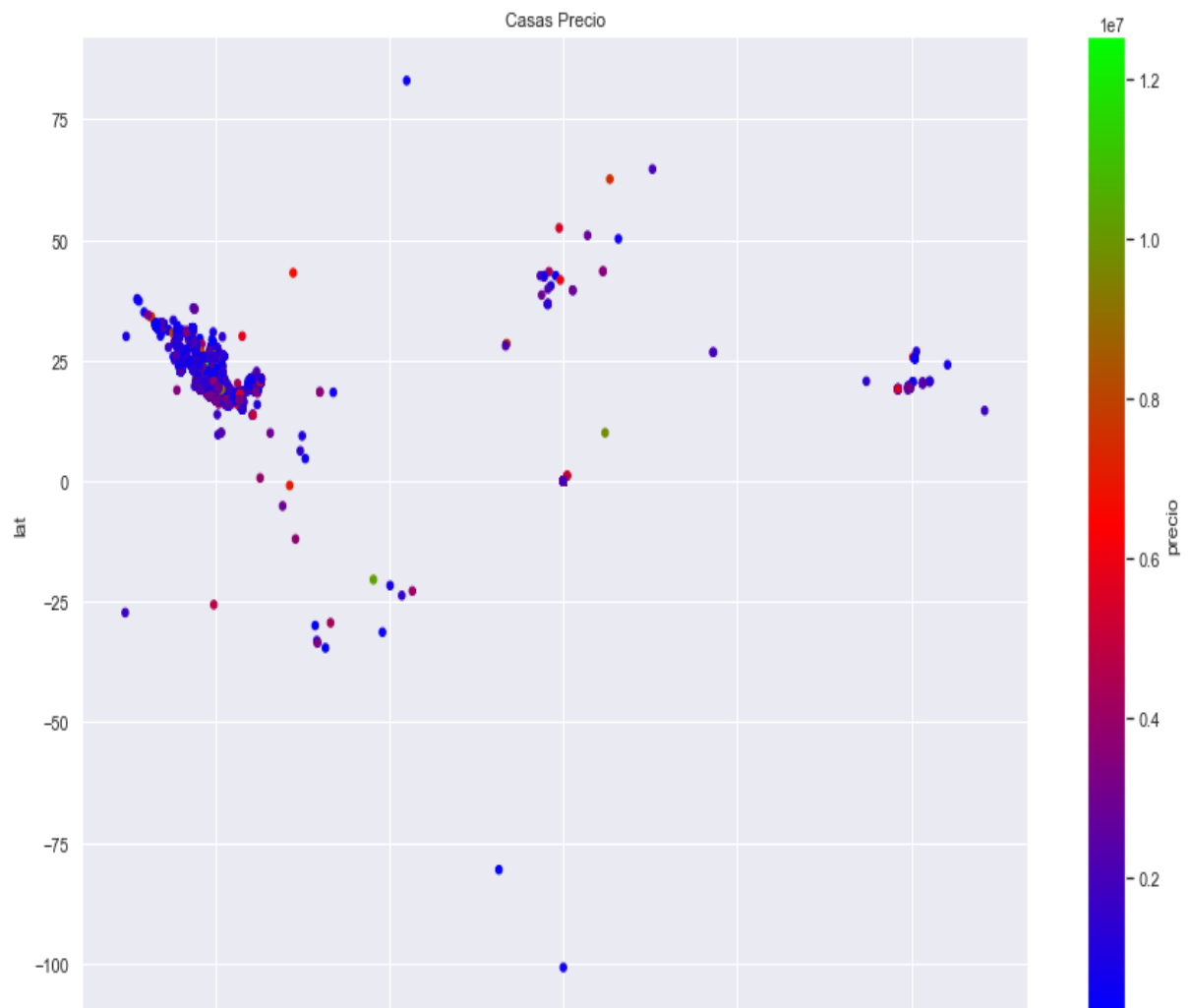
Análisis de precio por tipo de propiedad





- Del análisis del gráfico se puede observar que la tipo de propiedad con más publicaciones caras son las de edificios

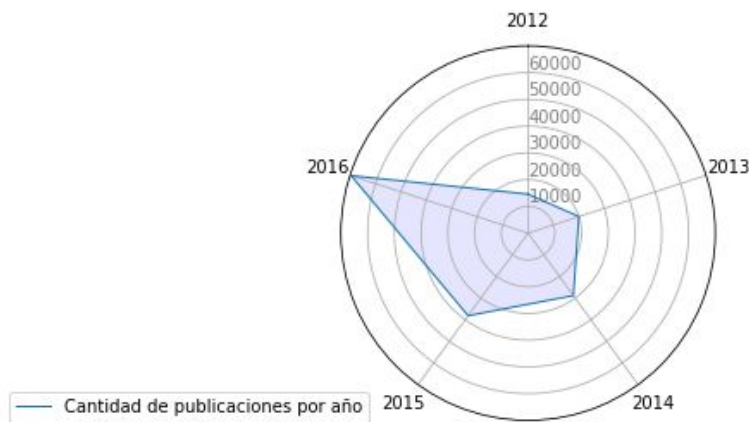
Análisis de precios por ubicación geográfica



- Un mapa donde se ve la ubicación geográfica de las propiedades más baratas

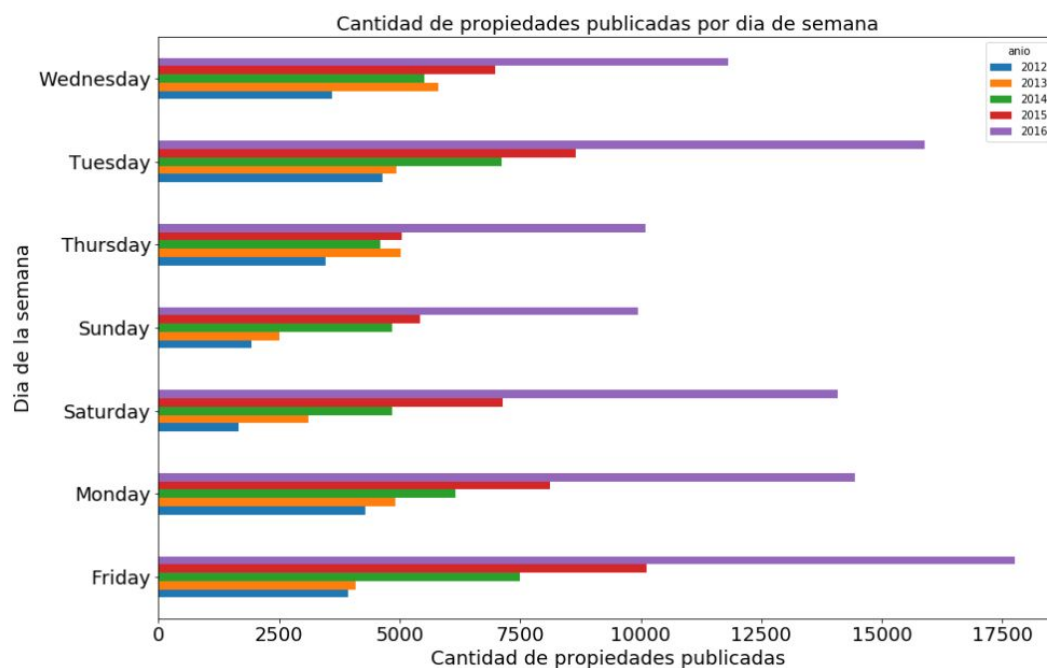
Análisis exploratorio de publicaciones

Análisis de cantidad de publicaciones por año



En este gráfico notamos que la mayoría de publicaciones se dan para el año 2016; esto se puede deber a 2 cosas: la primera es que por algún motivo en especial la gente tuvo mayor intención de venta de sus propiedades o la segunda opción es que la empresa nos brindó mayor información para ese año.

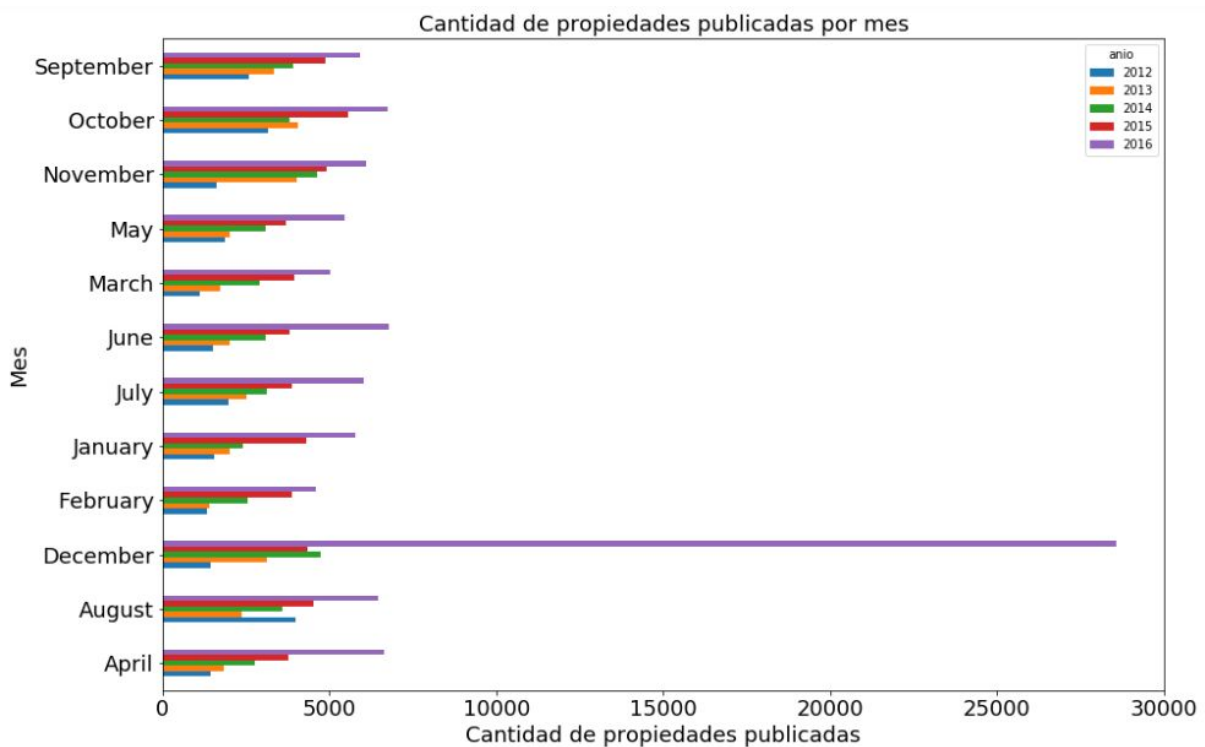
Análisis de cantidad de publicaciones realizadas por día de la semana



- La cantidad de propiedades publicadas aumenta significativamente en el año 2016 respecto a los años anteriores (esto se cumple con propiedades publicadas cualquier día de la semana)

- La cantidad de propiedades publicadas los viernes, sábados, domingos, lunes y martes (friday, saturday, sunday, monday, tuesday) aumenta año a año
- La cantidad de propiedades publicadas los miércoles (wednesday) baja en el año 2014, en el 2015 supera los valores del 2013 y en el 2016 aumenta respecto del año anterior
- La cantidad de propiedades publicadas los jueves (thursday) baja en el año 2014, en el 2015 recupera los valores del 2013 y en el 2016 aumenta respecto del año anterior

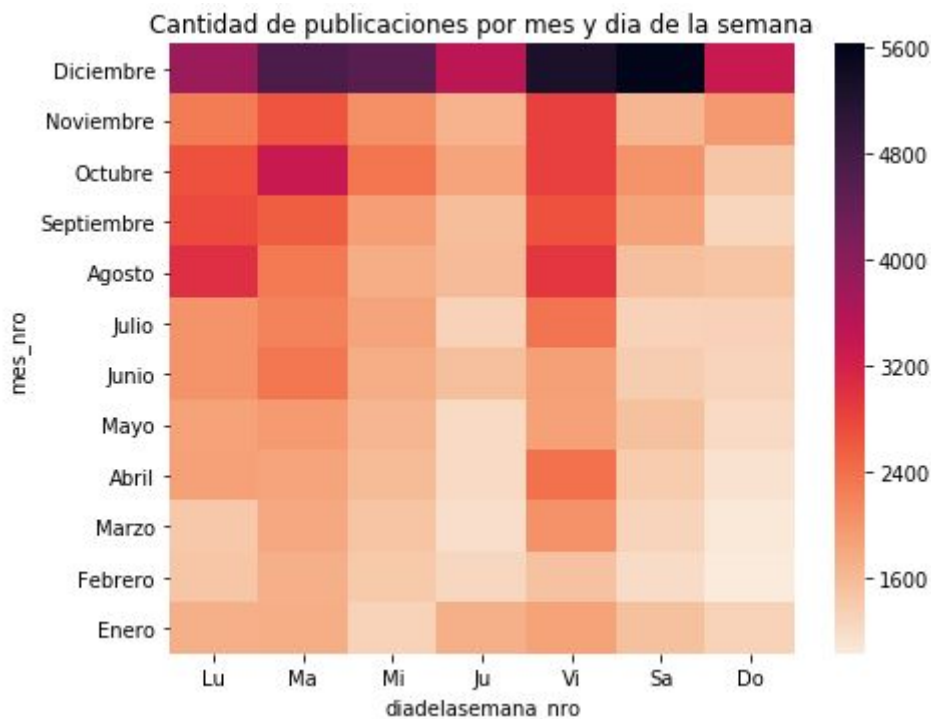
Análisis de cantidad de publicaciones realizadas por mes



- La cantidad de propiedades publicadas aumenta en el año 2016 respecto a los años anteriores (esto se cumple con propiedades publicadas durante el transcurso de cualquier mes). Es llamativo el gran aumento de cantidad de propiedades publicadas que se da en el mes de diciembre del año 2016.
- La cantidad de propiedades publicadas en enero, febrero, marzo, abril, mayo, junio, julio, septiembre y noviembre aumenta año a año
- La cantidad de propiedades publicadas en agosto baja en el año 2013 y luego vuelve a aumentar durante el 2014. Pero es a partir del 2015 que el número de propiedades publicadas supera al número de propiedades publicadas en el 2012
- La cantidad de propiedades publicadas en octubre baja en el año 2014 y luego vuelve a aumentar a partir del 2015

- La cantidad de propiedades publicadas en diciembre crece año a año, pero baja en el año 2015 y luego aumenta significativamente en el 2016

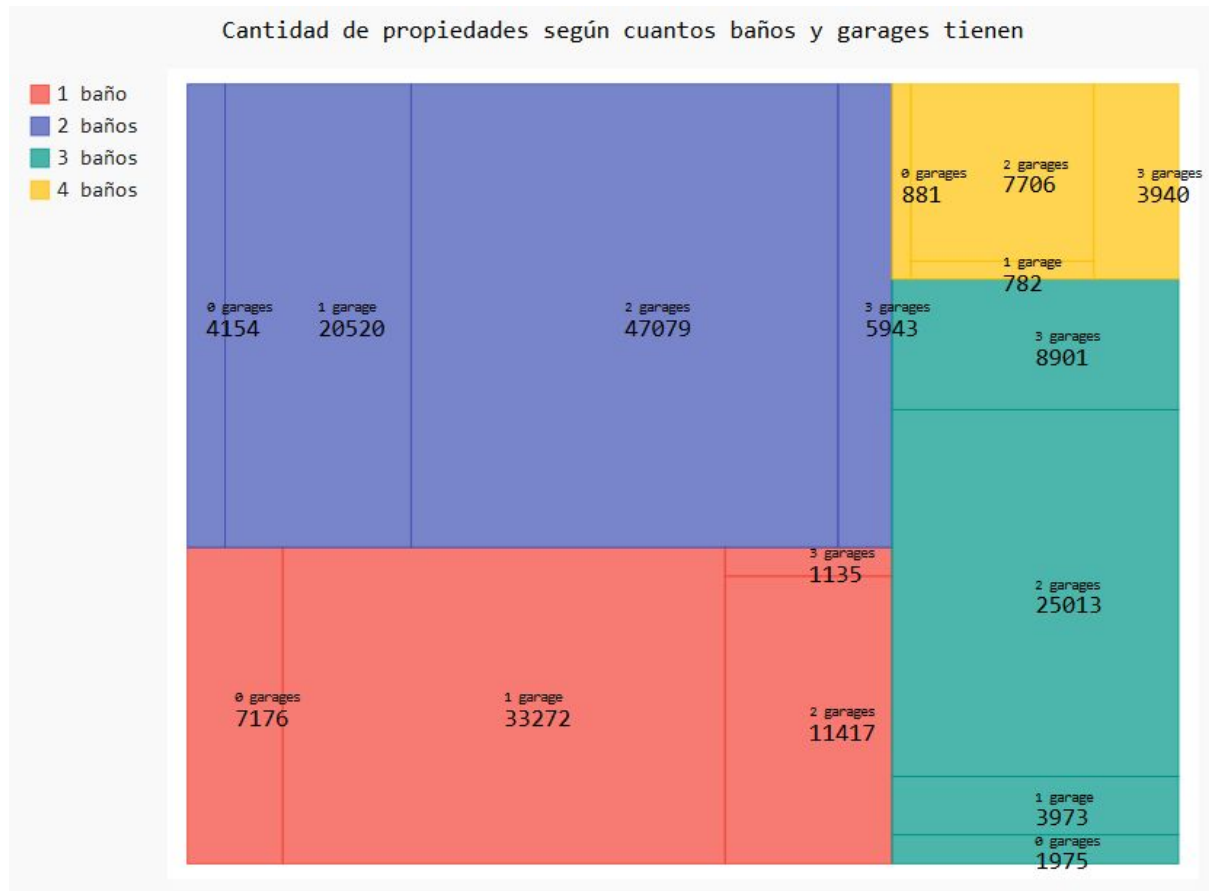
Análisis de cantidad publicaciones por mes y día de la semana



En este gráfico notamos claramente que la mayor cantidad de publicaciones las tenemos para el mes de diciembre, pero este gráfico muestra mas a detalle que para ese mes los días viernes y sábado son los días que más publicaciones tenemos; esto se puede deber a que en general las personas que realizan las publicaciones tienen más tiempo el fin de semana. en general para los demás meses si lo queremos ver de una manera más global, el viernes es en general el día en el que tenemos más publicaciones.

Análisis de características de una propiedad

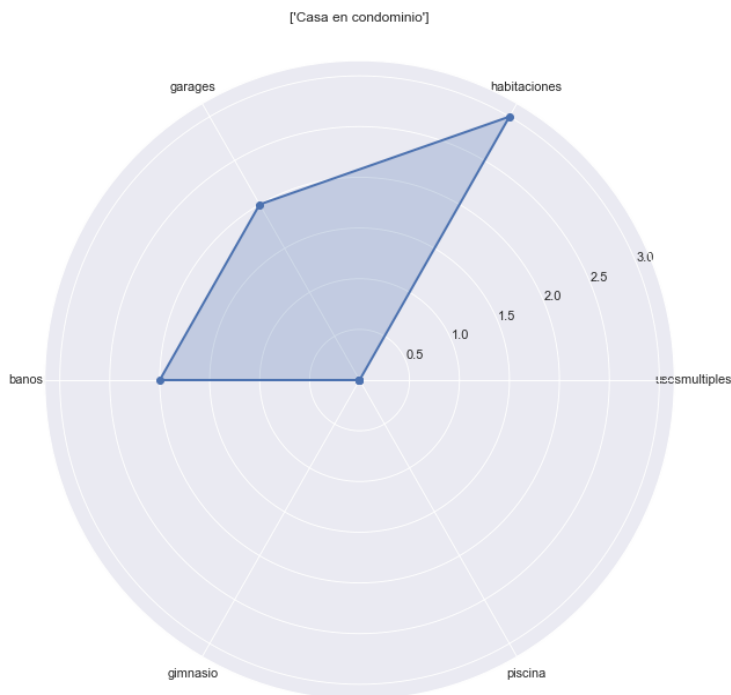
Análisis de cantidad de propiedades según cantidad de garages y baños

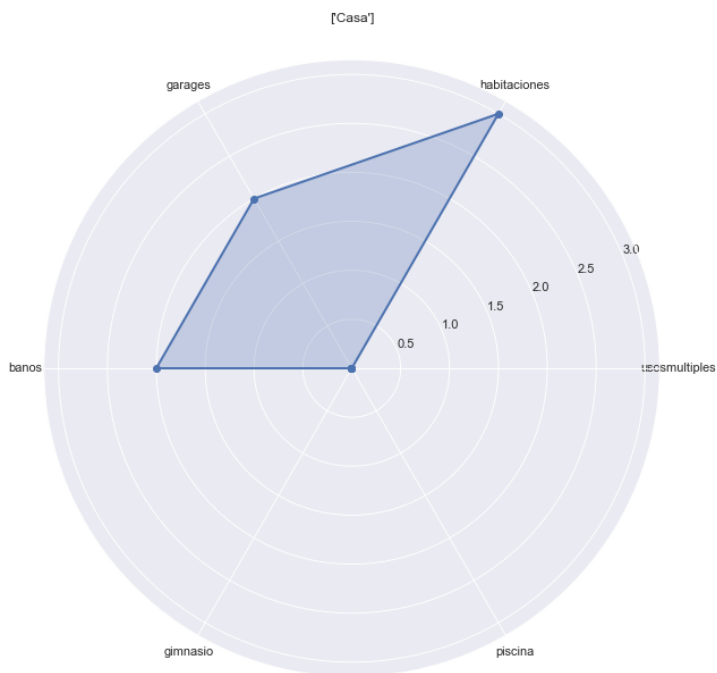
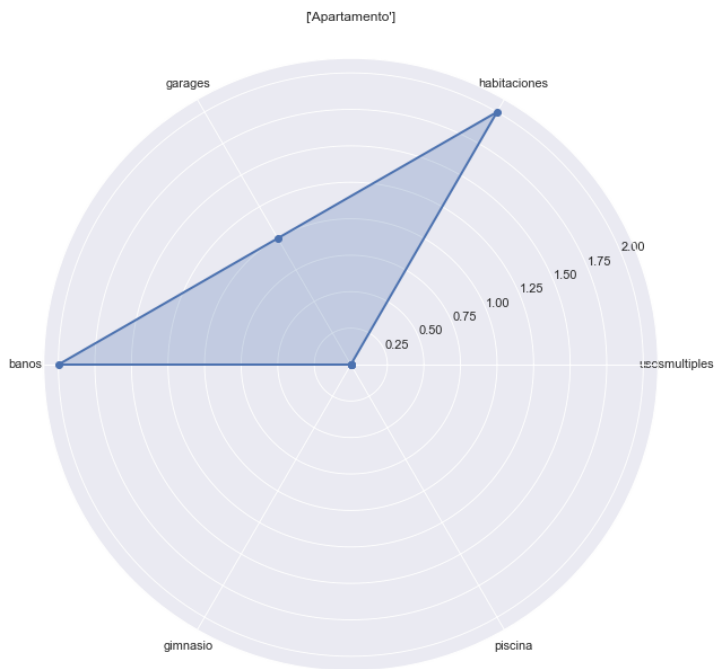


- Se puede observar que las propiedades con más publicaciones son las que tienen 2 baños, seguidas de las que tienen 1 baño, seguida por las que tienen 3 baños y las que menos publicaciones tienen son las que tienen 4 baños
- En el caso de propiedades con 2,3 y 4 baños se observa que la mayor parte de las publicaciones corresponde a aquellas propiedades que tienen además 2 garages
- En cambio en el caso de propiedades con 1 baño, ocurre que la mayor parte de las publicaciones corresponde a las propiedades que tienen 1 solo garage

Análisis de las características de cada propiedad individual

Es posible crear un radar chart para poder visualizar mejor las características de cada propiedad. En este caso realizamos algunos radar chart que nos permiten tener una muestra del tipo de propiedades publicadas y sus características.





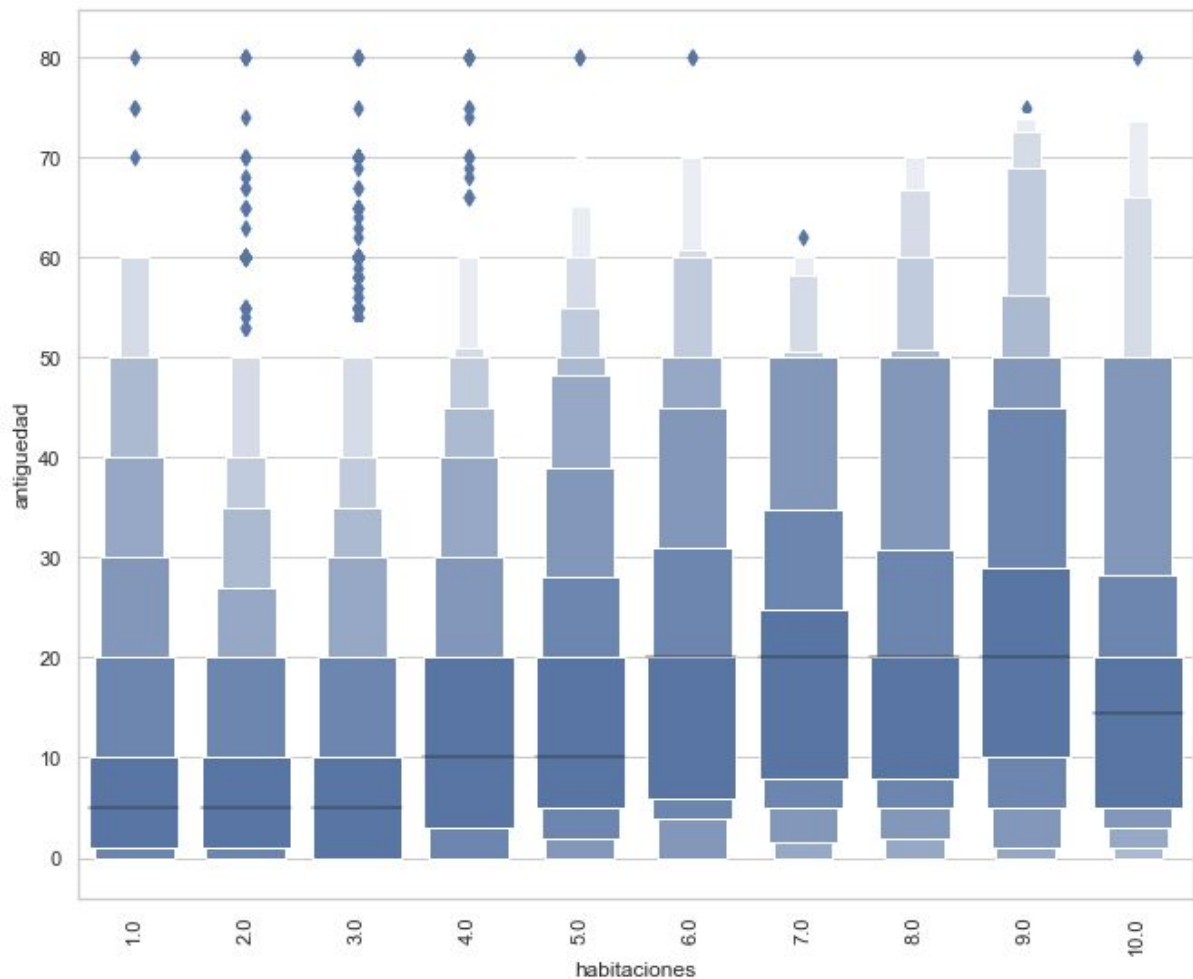
- Como forma de muestra para explorar el tipo de propiedades que se pueden encontrar publicadas tenemos:

Casa en condominio con 3 habitaciones, 2 garages y 2 baños.

Apartamento con 2 habitaciones, 1 garage y 2 baños.

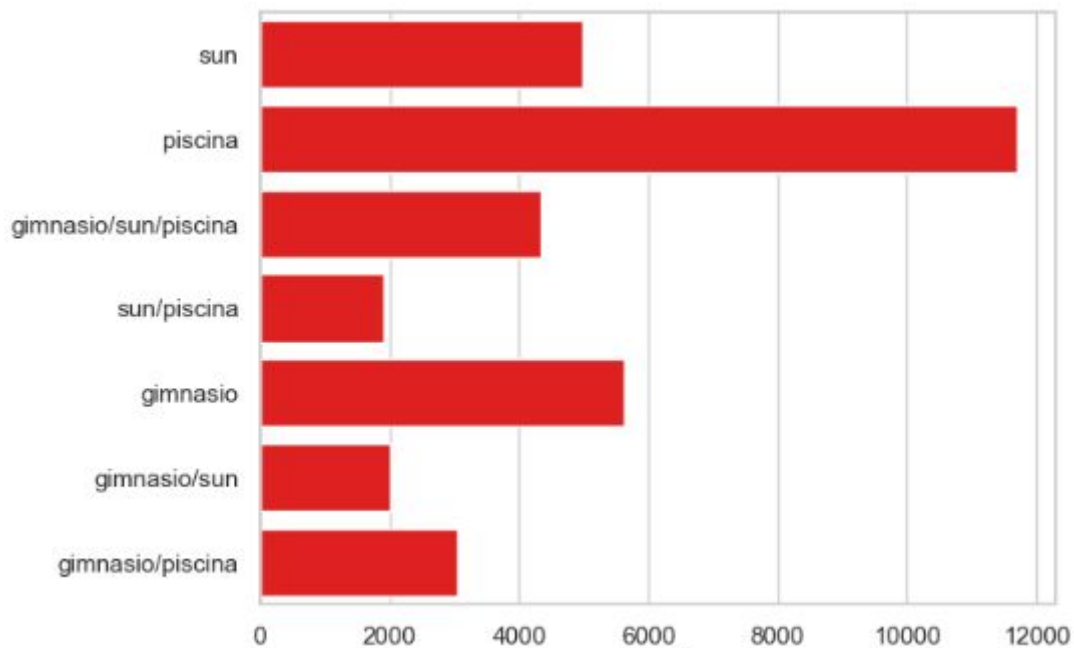
Casa con 3 habitaciones, 2 garages y 2 baños.

Analisis de cantidad de habitaciones y la antigüedad de las mismas



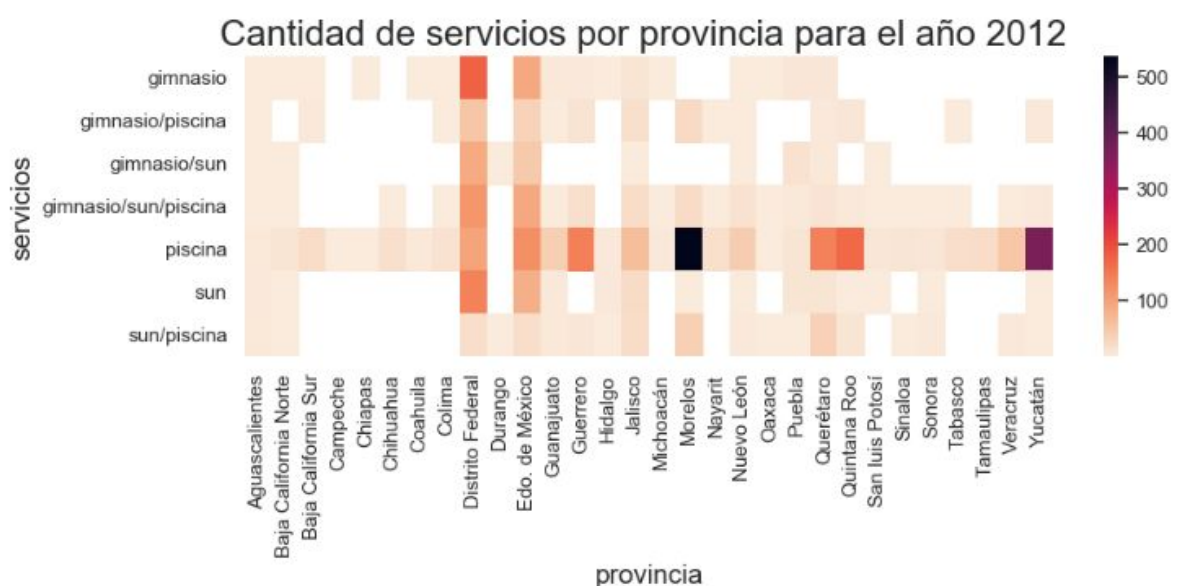
- Las propiedades más antiguas poseen mayor cantidad de habitaciones que las modernas (2016), esto se puede deber a el tipo de estilo de vida de los ciudadanos modernos que buscan propiedades pequeñas pero bien ubicadas, por ejemplo en una ciudad cercana a sus lugares de trabajo o estudio
- Los puntos mostrados sobre cada box indican anomalías , es decir cuando hay muchos valores atípicos

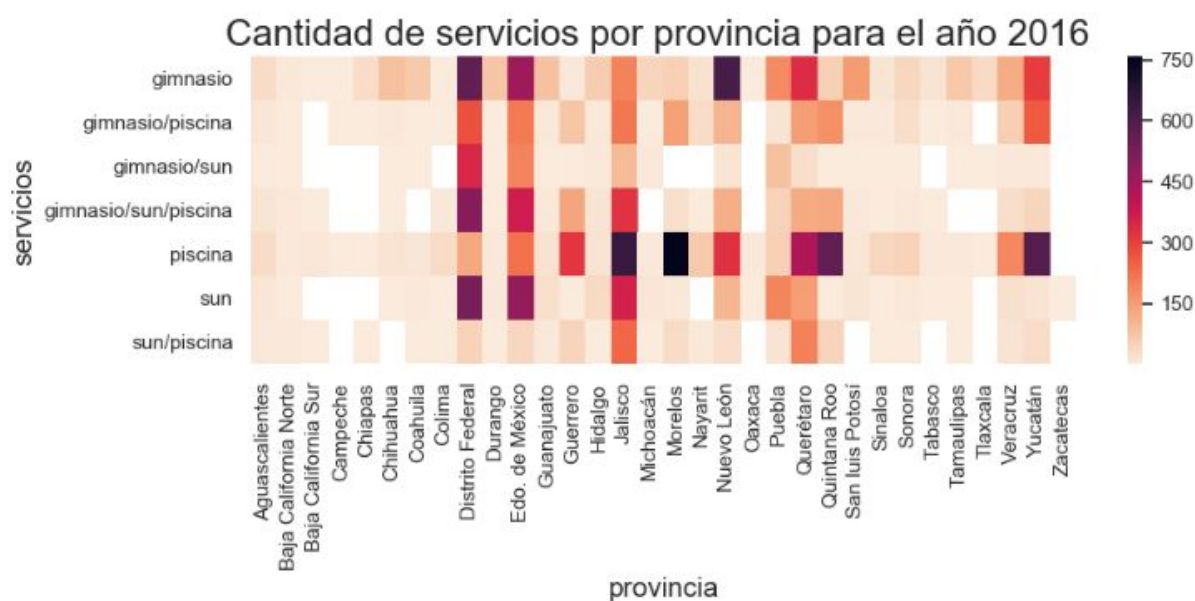
Análisis de servicios y publicaciones



En este gráfico queremos analizar la cantidad de publicaciones que tienen cierto servicio(piscina, sum, gimnasio). en el mismo podemos notar que predominan las publicaciones de propiedades con piscina, en segundo lugar solo con gimnasio y en 3 lugar solo con sum. Cabe destacar que la cantidad de propiedades con los 3 servicios no es la mínima, sino que está casi a la par con propiedades que solo tienen 1 solo servicio.

Análisis de servicios y provincias





El análisis se basa en comparar la cantidad de propiedades que cuentan con tales servicios y en que cantidad, la comparación se hace entre el año 2012 y 2016. Al mirar las 2 imágenes se puede notar una gran evolución entre esos años.

Se observa que en el 2012 la mayoría de propiedades con dichos servicios se centran en Distrito Federal, pero con el paso de los años y en el 2016 para ser más específicos, esto cambia, los servicios se dan en otras provincias.

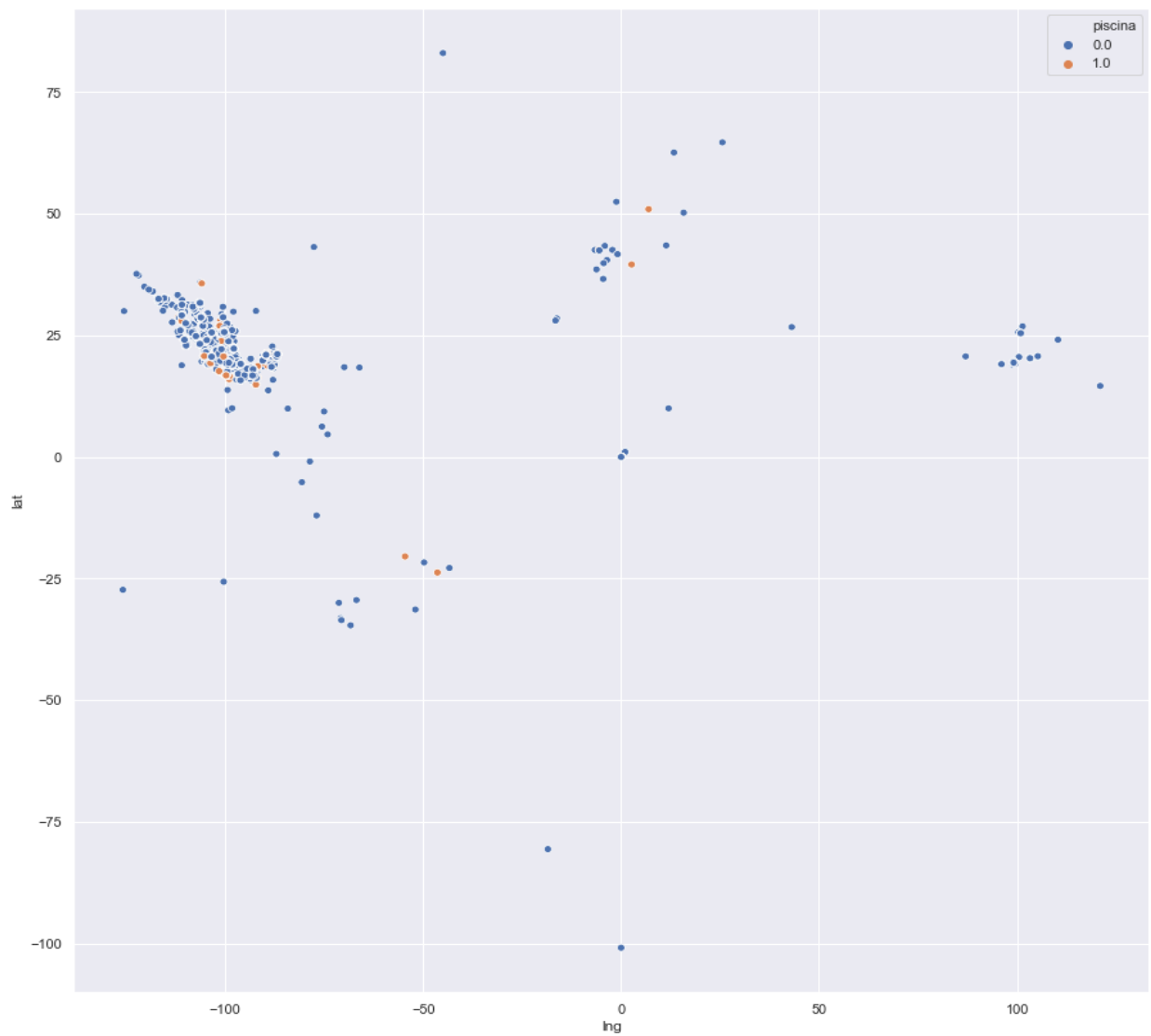
Análisis de la distribución geográfica de las propiedades

Análisis de escuelas cercanas



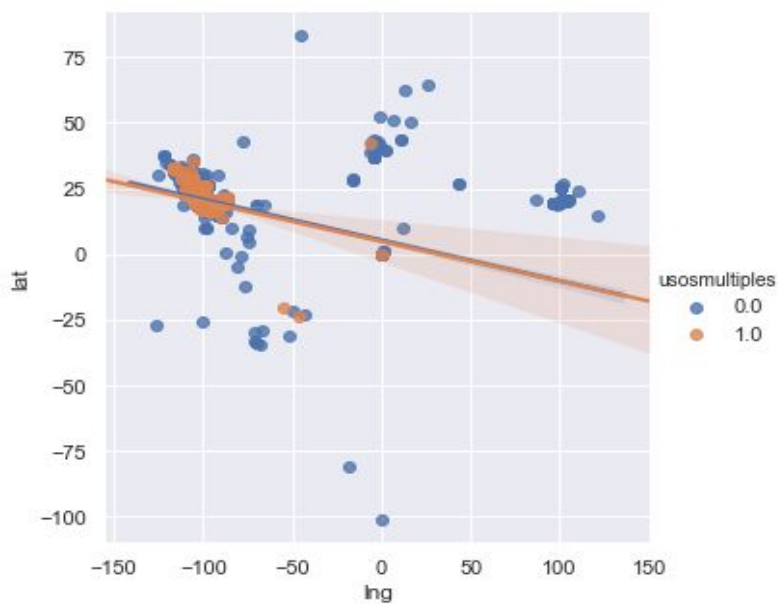
- Las propiedades con escuelas cercanas están distribuidas bastante bien a lo largo del territorio salvo en la zona de la derecha donde parece no haber ninguna o no se especificó bien si lo había

Análisis de piscinas



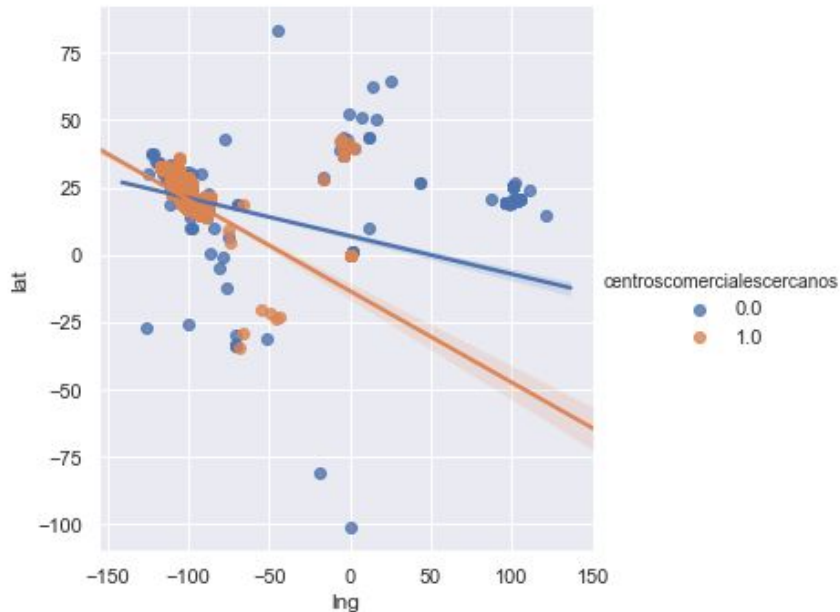
- Las piscinas son muy escasas y apenas se pueden apreciar en el gráfico distribuidas a lo largo del territorio

Análisis de propiedades con usos múltiples



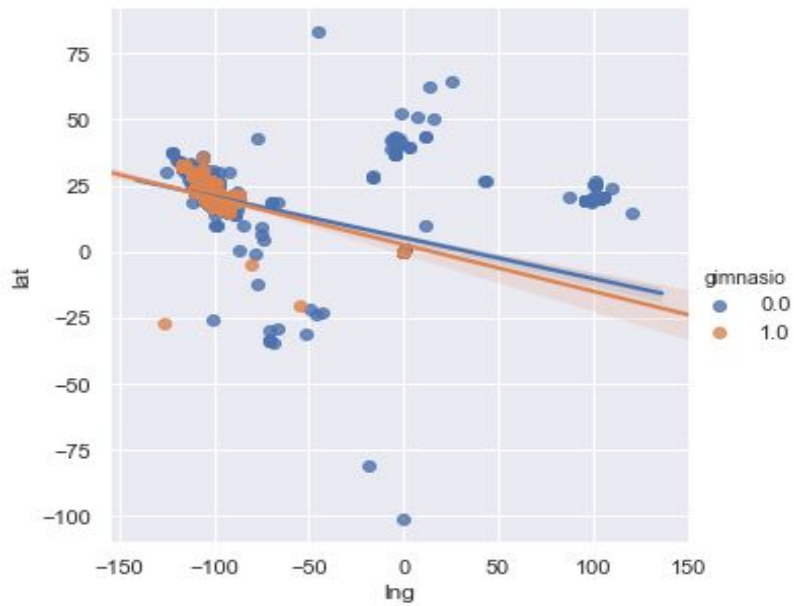
- Normalmente las publicaciones con usos múltiples son edificios por lo que podemos apreciar zonas de una ciudad muy edificada que podría ser el centro de México

Análisis de propiedades con centros comerciales cercanos



- Es posible observar la distribución de propiedades cercanas a centros comerciales , notando que algunas publicaciones que tenían centros comerciales cercanos no fueron puestos correctamente.

Análisis de propiedades con gimnasios



- Los gimnasios se concentran en la zona de la izquierda donde creemos que hay una ciudad

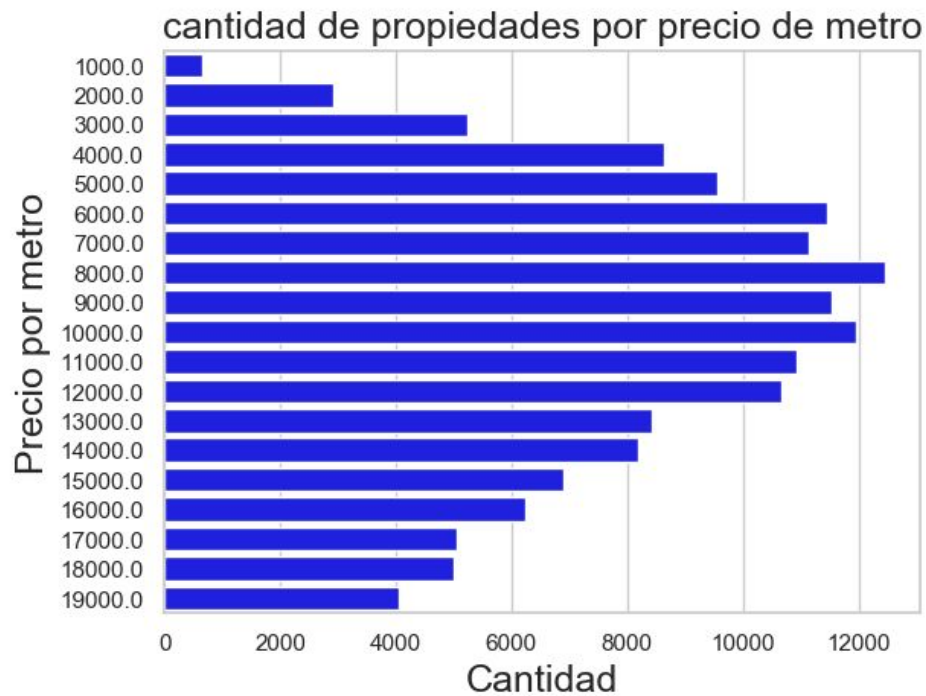
Análisis de tipos de propiedades en el mapa



- Se puede observar que la mayor cantidad de las propiedades publicadas se encuentran en la parte superior izquierda del gráfico.
- Se puede observar que la mayor cantidad de propiedades publicadas corresponden a casas, tanto en los sectores de gran densidad de propiedades publicadas como en los lugares con menor densidad

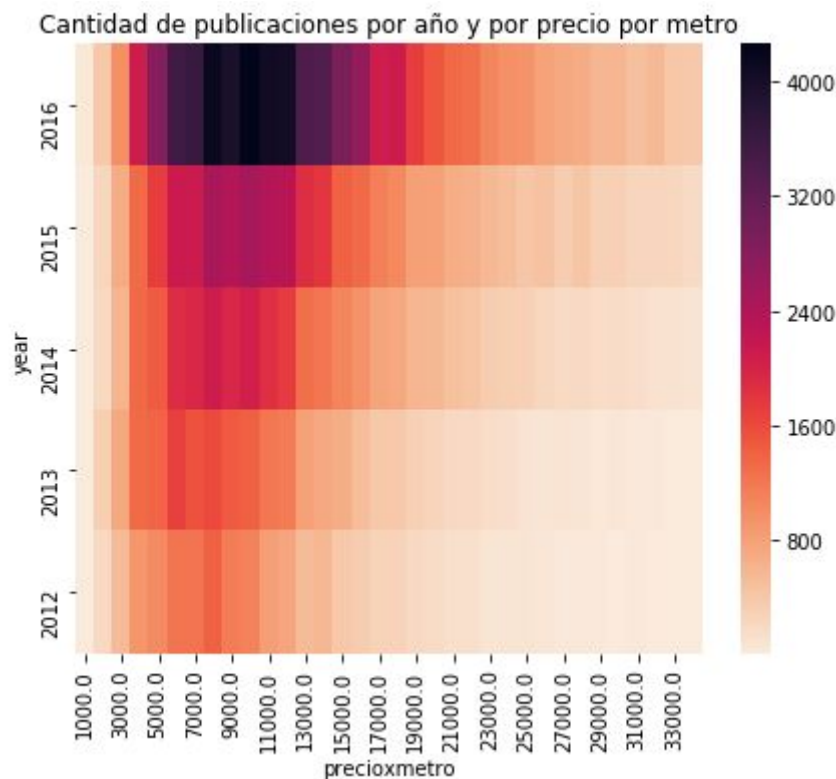
Análisis de precio por metro cuadrado

En este análisis se realizó una división del precio total entre el total de metros, luego de esto se redondeo a miles para poder trabajar con agrupaciones



En la imagen mostramos el rango en el que se encuentra el precio por metro cuadrado. Destacamos este rango por que en el mismo se encuentra la mayor cantidad de publicaciones. Se destaca el rango de entre 5000 y 12000 como el rango en el que se encuentra la mayor cantidad de publicaciones que tienen un precio por metro dentro del mismo.

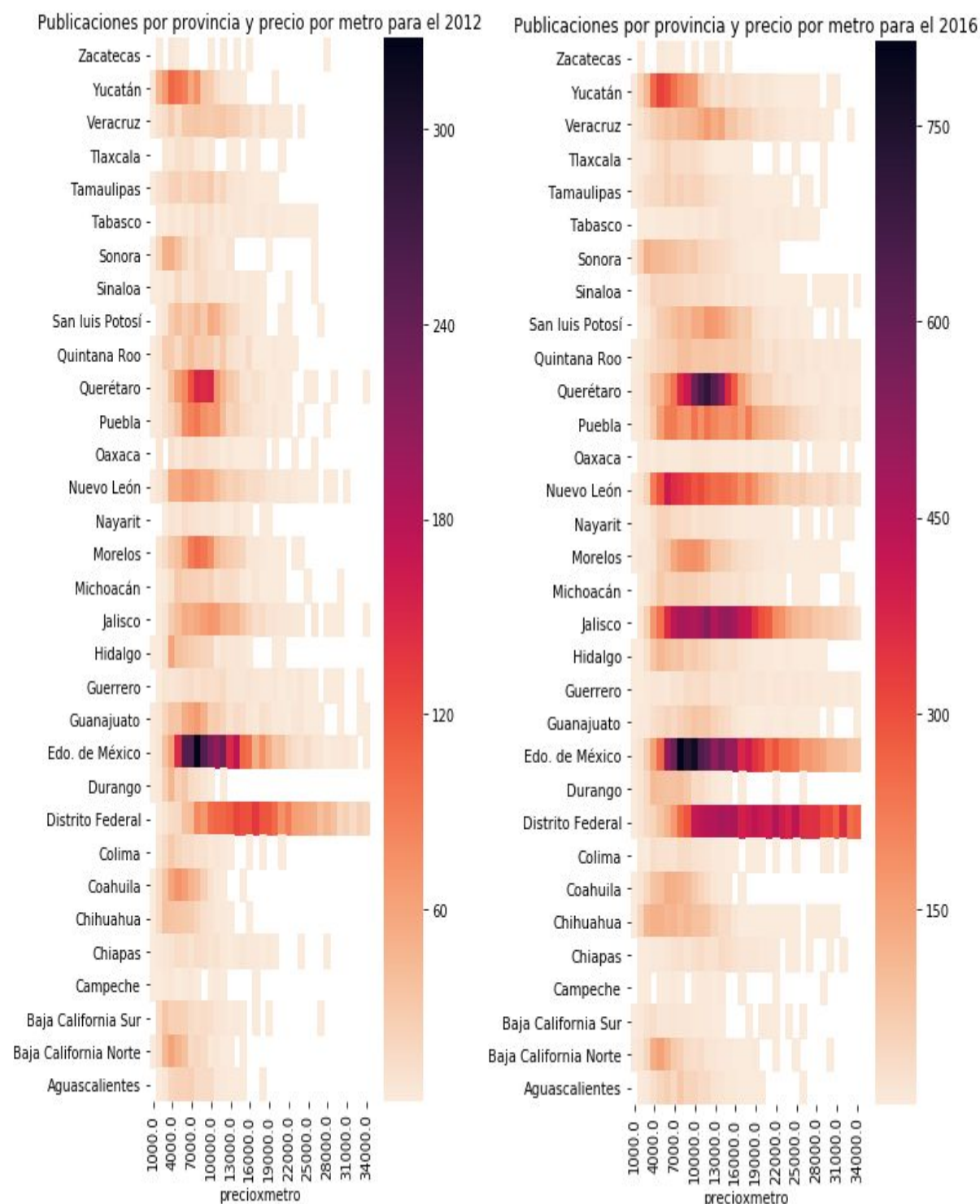
Análisis de precio de metro cuadrado por año



Para este gráfico tomamos el rango de entre 1000 y 35000 por metro cuadrado por que al momento de hacer la agrupación notamos que estas son las cantidades más notables.

Bueno a primera vista notamos que para el 2016 hay una mayor cantidad de resultados, pero esto es obvio por que como vimos antes tenemos una mayor cantidad de publicaciones para dicho año. Lo que podemos notar muy aparte de la cantidad de publicaciones en cada año, es que a medida que aumenta el año de publicación, también aumenta un cierto rango de precio por metro. por ejemplo para el 2012 el rango de precio por metro cuadrado se encuentra entre 5000 y 11000, luego para el año 2013 el rango se amplía entre 4000 y 13000, así hasta el 2016 donde el rango está entre 3000 y 25000. Esto se podría pensar como que a medida de que aumentaron los años aumentó la diversidad de precios por metro cuadrado

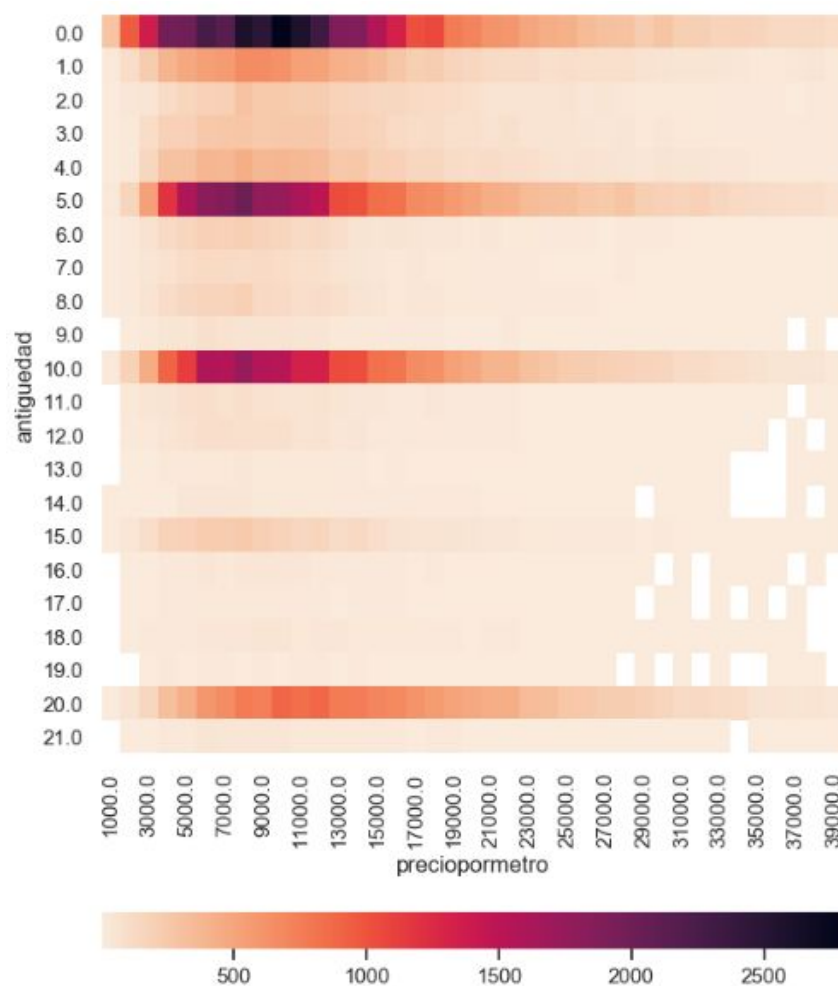
Análisis de precio por metro y provincia para el año 2012 y 2016



En este caso vamos a analizar la diferencia entre los 2 limites de años que tenemos en los datos. El análisis se basa en ver la diferencia entre cantidad de publicaciones por precio de metro y provincia entre el año 2012 y el 2016.

Las 2 imagenes son muy parecidas entre las provincias para los cuales existen un rango de precio por metro; quizás algunas diferencias notables las encontramos en por ejemplo: Distrito Federal, en donde el rango de precios pasó de estar entre 7000 y 22000 en el año 2012 a un rango de entre 7000 y 34000 para el año 2016; otro es Jalisco en el cual de estar en el 2012 con casi ningún rango notable de precios por metro, pasó a tener un gran rango de precios por metro cuadrado; luego esta Nueva León la cual también aumentó en gran medida su rango de precios por metro para el año 2016 con respecto al 2012; luego para las demás provincias el rango de precios por metro cuadrado del 2016 se mantuvo muy similar al rango del 2012.

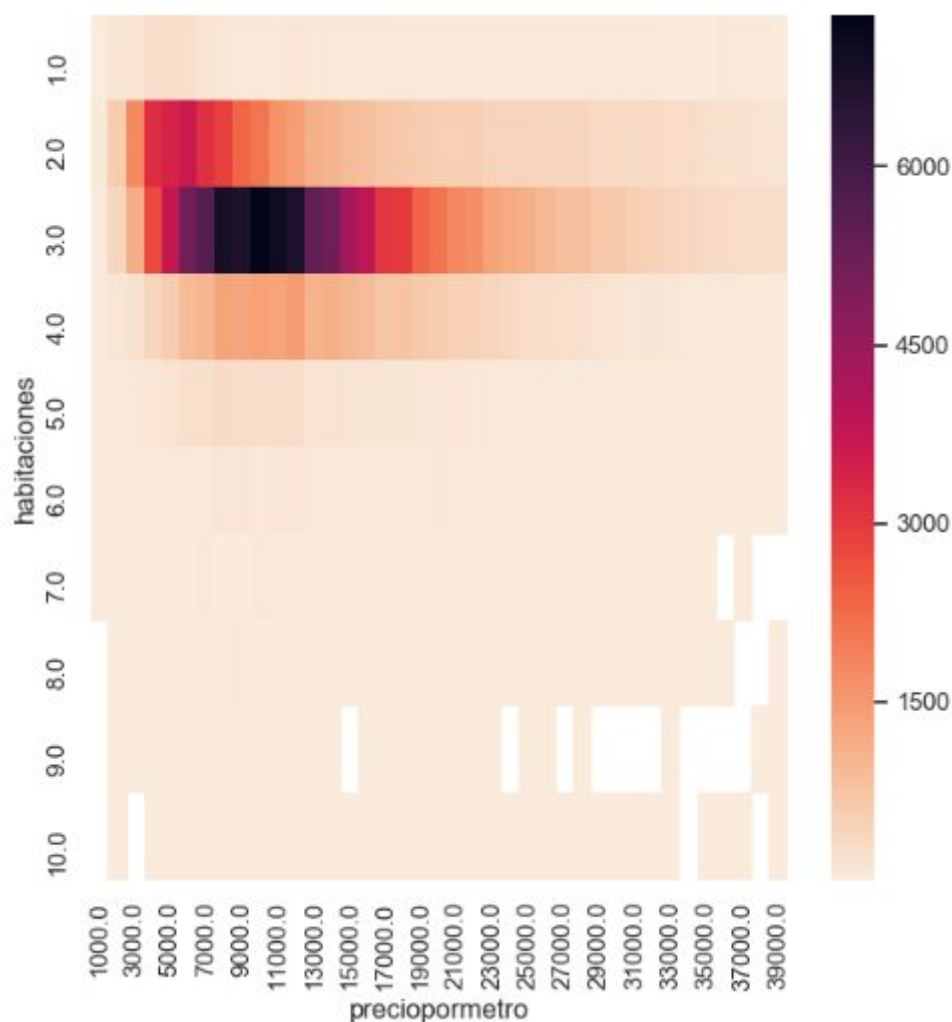
Análisis de precio por metro y antigüedad de la propiedad



En esta imagen se muestra una relación entre el precio por metro y la antigüedad de la propiedad; en la misma se muestra que hay una mayor variedad de precios para propiedades

nuevas, las cuales se encuentran entre un rango de 1000 y 27000 por metro, luego vienen las propiedades de 5 años de antigüedad con un rango más pequeño y le sigue las de 10 años con un rango similar a la de 5 años de antigüedad.

Análisis de precio por metro y cantidad de habitaciones de la propiedad



En este análisis miramos si hay alguna relación entre el número de habitaciones y el precio por metro. Al parecer las propiedades con 3 habitaciones tienen una gran variedad de precios por metro, además de tener muchas publicaciones con esa cantidad de habitaciones. Luego se puede observar a las de 2 habitaciones y en tercer lugar a las de 4. El hecho de la variedad en precios en las propiedades con 3 habitaciones se puede deber al aumento de construcción de departamentos y los mismos con 3 habitaciones.

Conclusiones

Bueno luego de ver toda la variedad de imágenes respecto al análisis que hacemos sobre este set de datos y a pesar de que en cada imagen dimos nuestro punto de vista respecto a lo observado, vamos a dar una vista general de lo más destacado o curioso que notamos en el mismo.

Lo primero que notamos es que hay campos en los que faltan datos que por ahí en lo general se pueden ver como que no son tan necesarios, entre estos puede estar la lon y lat, que son más de la mitad y se podría decir que no son tan relevantes a simple vista. Pero lo que si notamos importante es la falta del dato de la antigüedad y de los metros totales de la propiedad; estos no son datos con poca importancia y al momento de realizar el análisis de los datos faltantes estos era el 18% del total de publicaciones.

Lo segundo que notamos es la gran diferencia de la cantidad de publicaciones entre los años 2012 y 2016. Viéndose una gran diferencia para el 2016 respecto a los otros años. No sabemos si esto se debe a que el set de datos contiene más datos recientes, o si por algún motivo en especial en el año 2016 se produjeron una mayor cantidad de publicaciones respecto a los otros años.

Otra cosa que notamos y que no nos sorprendió tanto , fue el hecho de que la mayoría de publicaciones son de Distrito Federal. Esto es obvio al ser esta la capital de México.

Una cosa respecto a las publicaciones y el tipo de propiedad es el hecho de que la gran mayoría son publicaciones de casas. Uno a primera impresión se piensa que deberían ser departamentos, o por lo menos para nosotros.

Algo respecto al precio por metro, es que hay una diferencia a medida que pasan los años. Al decir diferencia nos referimos al aumento en la variedad del precio por metro. Ya que para publicaciones mas antiguas no existía una gran variedad en el precio por metro, el precio estaba concentrado en un rango pequeño, esto cambió para el 2016 en el cual el espectro de precios por metro aumentó y mucho.

Llama la atención que hay una cantidad de direcciones en donde aparecen por ejemplo caracteres aleatorios

Encontramos que había ciudades con muchas propiedades publicadas y notamos que, en algunos casos, había varias ciudades con el mismo nombre pero en distinta provincia. En particular la ciudad de Cuauhtémoc se encuentra en 3 provincias de México, sin embargo la mayoría de las publicaciones se realizaron desde la que está en la provincia del distrito federal

Viendo las descripciones que se hacen en cada publicación resalta una descripción con un mismo mensaje que corresponde a remates que pudieron ser causados por deudas o medidas judiciales

Hay una correlación entre metros cubiertos y cantidad de baños , analizandola nos resultó curioso cierta cantidad de propiedades con muchos metros cubiertos y con pocos baños (1 o 2) que son la minoría. A partir de esto pensamos que correspondían a propiedades que no eran viviendas. Analizandolas obtuvimos que corresponden a casas.

Es llamativo la cantidad de propiedades publicadas en diciembre del año del 2016 que pasa de 4300 a 28500 , desconocemos la verdadera causa pero corresponde con el periodo de elecciones en estados unidos donde la victoria fue de Donald Trump que pudo aumentar el riesgo de una crisis económica en México (La gente pudo querer vender sus propiedades para que no se devalúen)