

# **Departamentos en Venta (2015):**

## **Aproximación analítica al mercado inmobiliario**

### **1. Datos generales**

Este set de datos recabó su información a partir del relevamiento muestral de avisos publicados para la venta de Departamentos del año 2015, realizado por el Gobierno de la Ciudad de Buenos Aires, que incluye las publicaciones comerciales realizadas en distintos sitios Web para la exposición de este tipo de propiedades. Mediante su abordaje, se pretenderá describir y caracterizar desde distintas aristas la composición y desenvolvimiento del mercado inmobiliario, particularmente en lo que respecta a los factores causales de los precios de mercado vigentes (aspecto que se cotejará en el valor total de las propiedades y en su valor por metro cuadrado).

### **2.Contexto comercial**

La Ciudad Autónoma de Buenos Aires es una de las urbes más prominentes y vanguardistas de Latinoamérica, en la cual la proliferación de desarrollos inmobiliarios y las transacciones de propiedades (en este caso particular, departamentos) son moneda corriente.

A efectos de realizar una indagación retroactiva del mercado inmobiliario (para luego poder contrastarlo con su estadio actual), se busca explorar las principales variables pertinentes a este mercado y, en particular, los factores influyentes en la composición de sus valores. A tal efecto, se han recopilado datos de los departamentos publicados a lo largo del año 2015 en los principales sitios web del rubro (desde enero hasta diciembre del aludido año).

Se propenderá a reconocer patrones que delineen las condiciones que valorizan un departamento, en aras de taxonomizar las condiciones más propicias para impulsar desarrollos inmobiliarios, así como reconocer características que el gobierno porteño debería tomar en consideración para promover un desarrollo parejo y sustentable en la ciudad.

Para ello, se recopilaron cualidades consideradas pertinentes entre los departamentos escrutados.

### **3. Contexto Analítico**

En principio se descargó el DataSet pertinente, cotejable en la web:

<https://data.buenosaires.gob.ar/dataset/departamentos-venta/resource/juqdkmgo-7029-resource>

Tras seleccionar las variables de interés, se procedió a leer, transformar y preparar los datos para su visualización:

\_ Reasignar el formato en las columnas, definiéndolas como de tipo Texto, General, Numérica o Monetaria, según el caso.

\_ Reemplazar nombres de títulos y categorías de atributos poco representativas por otras más claras.

\_ Corregir errores en caracteres tergiversados (fundamentalmente Ñ y algunas palabras acentuadas).

Ya con la base de datos debidamente procesada, se formularon algunas preguntas iniciales de investigación, a efectos de realizar un abordaje inicial del DataSet y sacar conclusiones preliminares. A partir de la respuesta a esas preguntas, mediante visualizaciones y sus respectivas conclusiones, se presentan los principales hallazgos de la investigación.

### **3.1 Variables integrantes del dataset**

La base de datos utilizada referencia múltiples características de interés sobre los departamentos en exposición. A continuación se especifica el contenido de cada columna considerada:

\_ ID: variable de número, identificador único de cada propiedad.

\_ CALLE: variable de texto, estipula la dirección del departamento.

\_ NÚMERO: variable de número (si bien su carácter cuantitativo no reviste interés analítico), estipula la numeración en la dirección del departamento.

\_ CANTIDAD DE M2: variable de número que refiere al tamaño de la superficie del departamento.

\_ PRECIO (USD): variable monetaria que expresa la cotización total de la propiedad.

\_ PRECIO M2 (USD): variable monetaria que expresa la cotización por metro cuadrado de la propiedad.

\_ AMBIENTES: variable de número que define la cantidad de ambientes de la propiedad.

\_ A ESTRENAR: variable de texto, dicotómica, que expresa si el departamento es a estrenar o reviste alguna antigüedad.

\_ ORIENTACIÓN: variable de texto, versa sobre la disposición del departamento con respecto al edificio en que se sitúa.

\_ COCHERA: variable de texto, dicotómica, que expresa si el edificio en que se encuentra el departamento cuenta o no con cochera.

\_ WEB PUBLICACIÓN: variable de texto, sitio especializado de internet del cuál se recabó la publicación.

\_ BARRIO: variable de texto, barrio de ubicación del departamento.

\_ COMUNA: variable de número (como en el caso de número de calle, su carácter cuantitativo no es de interés analítico) que expresa la comuna de ubicación del departamento.

\_ COTIZACIÓN: clasificación de la propiedad en Económica (hasta 1800 dólares el M2), Mediana (entre 1801 y 3000 dólares) y Cara (más de 3000 dólares).

\_ DIMENSIÓN: clasificación de la propiedad en Chica (hasta 50 M2), Mediana (entre 51 y 100 M2), y Grande (mayor 100 M2).

\_ ZONA: clasificación que divide a la Ciudad de Buenos Aires en Norte (Comunas 2, 12, 13, 14, 15), Centro (1, 3, 4, 5, 6, 7) y Sur (Comunas 8, 9, 10, 11).

#### **4. Objetivo**

Describir y caracterizar desde distintas aristas la composición y desenvolvimiento del mercado inmobiliario para el período 2015 en el ámbito de la Ciudad de Buenos Aires, particularmente en lo que respecta a los factores causales de los precios de mercado vigentes. Se parte de la premisa de que aspectos como la ubicación geográfica, el tamaño, la presencia de cochera y la orientación son algunos de los factores que ejercen una marcada influencia en la cotización de las propiedades.

#### **5. Hipótesis**

La principal hipótesis de esta presentación versará en que la ubicación en ciertas comunas (1, 2, 13, y 14), el mayor tamaño, la presencia de cochera y la orientación frontal son factores que elevan el valor de las propiedades publicadas.

#### **6. Problema Comercial**

El rol de este análisis consiste en, mediante la gestión de los datos disponibles, la generación de visualizaciones y la utilización de machine learning, generar insights que permitan responder a ciertas preguntas de investigación, a saber:

\_ Precio y Precio por M2 serán variables objetivo recurrentes durante la investigación. ¿Cómo se distribuyen?

\_ ¿Cómo se despliegan las propiedades cotejadas al considerarlas según su tamaño en M2?

\_ ¿Cuál es el comportamiento en el Precio de las propiedades en función del tamaño en M2? Este comportamiento, ¿se refrenda al considerar la cantidad de ambientes?

\_ ¿Cómo se distribuyen los departamentos en función a su ubicación? ¿Y por Zona?

\_ ¿Se observan comportamientos diferenciales en el Precio de los departamentos al segmentarlos según ubicación?

\_ Con respecto a otras características edilicias, como ser Orientación de la propiedad y presencia de Cochera, ¿Cómo se distribuyen las propiedades? ¿Cómo influyen en su valor?

\_ Una característica usualmente debatida es si los departamentos "A Estrenar" son mejor valorados. Suele aducirse que, si bien comprar un departamento puede suponer beneficios, en muchos casos adquirir unidades que no han sido debidamente probadas puede traer problemas edilicios no esperados y necesidad de arreglos adicionales. ¿Cómo valora el mercado los departamentos "A Estrenar"?

#### **7. Alcance. Motivación y Audiencia**

Este trabajo pretende brindar un paneo descriptivo sobre las principales aristas respecto al comportamiento del mercado inmobiliario porteño, esbozándose un perfilamiento general de los factores que podrían incidir en su valorización. En ese

sentido, su utilidad analítica es encuadrable fundamentalmente en el nivel estratégico, como soporte para la toma de decisiones que involucren políticas en un sentido amplio, tanto para el sector público como en el rubro inmobiliario.

Como ya fuera mencionado, definir y clasificar las características que redundan en un mayor valor de los departamentos permitiría brindar información pertinente a empresas desarrolladoras del rubro inmobiliario, para orientar las cualidades de tales desarrollos a la búsqueda de un mayor rédito económico. Asimismo, se buscará generar un análisis de utilidad para asesorar a las autoridades municipales en sentido de tener una intervención virtuosa en el mercado, que permita el desarrollo de una ciudad más sustentable y equitativa.

## **8. Data Acquisition**

Procedemos a ejecutar el código pertinente para importar la base de datos de interés, que se obtendrá desde un archivo Excel ubicado en Google Drive.

Mediante este código, se referencia la librería de "drive" para luego montar 'Google Drive', plataforma en la que se encuentra ubicado el archivo con el dataset. Asimismo, se importa la librería "pandas" para transformar el dataset desde excel a un formato más ameno para trabajar desde el notebook.

Finalmente, mediante el método head, se solicita una visual de las primeras filas del dataframe ya transformado.

ID	Calle	Número	Cantidad de M2	Precio	Precio M2	Ambientes	A Estrenar	Orientación	Cochera	Web Publicación	Barrio	Comuna	Cotización	Dimens
0	11 DE SEPTIEMBRE DE 1888	900	350	880000	2514.300	4	NO	FRENTE	SI	ZONAPROP	PALERMO	14	MEDIANA	GRAN
1	11 DE SEPTIEMBRE DE 1888	1100	200	545000	2726.200	5	NO	FRENTE	SI	ZONAPROP	PALERMO	14	MEDIANA	GRAN
2	11 DE SEPTIEMBRE DE 1888	1300	350	1490000	4257.100	5	NO	FRENTE	SI	ZONAPROP	PALERMO	14	CARA	GRAN
3	11 DE SEPTIEMBRE DE 1888	1400	255	780000	3058.800	5	NO	FRENTE	SI	ZONAPROP	PALERMO	14	CARA	GRAN

## **9. Data Wrangling**

Conociendo ya las variables integrantes del Dataset, se acometió un preprocesamiento de los datos que permitiera trabajar con una base de datos más acotada (quitando categorías que no sean de interés, así como eventuales filas vacías).

### **9.1 Preselección**

Como primera medida, se descartan variables que se sabe de antemano que no tendrán ninguna incidencia en el modelo: ID, Calle, Número (de la calle) y Web de publicación.

Dicho esto, se procede a generar una copia del dataset sin estas columnas.

	Cantidad de M2	Dimensión	Ambientes	Precio	Precio M2	A Estrenar	Orientación	Cochera	Barrio	Comuna	Cotización	Zona
0	350	GRANDE	4	880000	2514.300	NO	FRENTE	SI	PALERMO	14	MEDIANA	NORTE
1	200	GRANDE	5	545000	2726.200	NO	FRENTE	SI	PALERMO	14	MEDIANA	NORTE
2	350	GRANDE	5	1490000	4257.100	NO	FRENTE	SI	PALERMO	14	CARA	NORTE
3	255	GRANDE	5	780000	3058.800	NO	FRENTE	SI	PALERMO	14	CARA	NORTE
4	200	GRANDE	4	1100000	5500.000	NO	FRENTE	SI	BELGRANO	13	CARA	NORTE

## 9.2 Borrado de duplicados

Utilizando el método `drop_duplicates`, se intenta eliminar las filas duplicadas, corroborando finalmente que no había ninguna de tales filas.

## 9.3 Análisis de nulos

Asimismo, mediante el método `is_null`, cotejamos si existen valores nulos en el dataset, concluyendo que no los había.

## 9.4 Análisis del tipo de dato de las variables

¿Se corresponde el tipo de dato de las columnas del dataset con aquello que se pretende para el subsecuente análisis de las variables? Procedemos a vislumbrarlo.

```
Cantidad de M2      int64
Dimensión          object
Ambientes          int64
Precio             int64
Precio M2          float64
A Estrenar         object
Orientación        object
Cochera            object
Barrio             object
Comuna             object
Cotización         object
Zona               object
dtype: object
```

Se observa que Comuna es procesada como variable cuantitativa, no obstante no debería serlo y calcular sus medidas de tendencia central y dispersión no aporta información útil. Es menester ajustar los tipos de datos antes de proseguir con el EDA. Realizamos las correcciones pertinentes.

```
Cantidad de M2      int64
Dimensión          object
Ambientes          int64
Precio             int64
Precio M2          float64
A Estrenar         object
Orientación        object
Cochera            object
Barrio             object
Comuna             object
Cotización         object
Zona               object
dtype: object
```

## **9.5 Anomalías**

¿Cuántas anomalías tiene cada columna numérica? A efectos de averiguarlo, generemos una función (ver en el notebook “buscar\_anomalías”) que permita recabar outliers, para luego aplicarla al dataframe. Se obtuvieron estos resultados.

```
Total de anomalías por columna:  
Cantidad de M2 : 420  
Ambientes : 70  
Precio : 307  
Precio M2 : 359
```

Se intentó inicialmente abordar la presencia de estos outliers mediante la aplicación de técnicas como el método de rango intercuartílico y otras metodologías recomendadas. Sin embargo, tras un análisis detallado, se observó que estas técnicas no lograban reducir significativamente la cantidad de outliers en el conjunto de datos. A pesar de explorar diversas estrategias, las características particulares de los datos y su distribución desafiaron los intentos de tratamiento de outliers de manera efectiva. Por lo tanto, en aras de preservar la integridad de los datos y evitar una potencial distorsión en el análisis resultante, se ha tomado la decisión de dejar los outliers sin tratar por el momento. Se reconoce la importancia de este aspecto en la etapa de Análisis Exploratorio de Datos (EDA) y se planea visitar este tema durante futuras iteraciones del proceso de análisis, cuando se disponga de un contexto más amplio y detallado para tomar decisiones informadas sobre cómo manejar adecuadamente los outliers.

## **10. Exploratory Data Analysis (EDA)**

### **10.1 Exploración preliminar del dataset**

#### **10.1.1 Cantidad de filas y columnas**

Si bien ya se cotejaron las variables inmanentes al dataset, se procede a cotejar sus dimensiones para, además, contemplar su cantidad de filas.

```
Cantidad de filas y columnas: (8979, 16)
```

#### **10.1.2 Nomenclatura de las columnas**

```
Nombre de las columnas: Index(['Cantidad de M2', 'Dimensión', 'Ambientes', 'Precio', 'Precio M2',  
                              'A Estrenar', 'Orientación', 'Cochera', 'Barrio', 'Comuna',  
                              'Cotización', 'Zona'],  
                             dtype='object')
```

Aquí observamos que el dataset, atravesada la instancia de Data Wrangling, cuenta aún con una cantidad considerable de filas y columnas (que, además, representan

aspectos bastante diversos sobre las propiedades muestradas), lo cual debería permitirnos encarar un estudio bastante robusto sobre el tema.

## **10.2 Comportamiento individual de las variables (cuantitativas)**

### **10.2.1 Medidas de tendencia central, de localización y de dispersión**

#### **Media, desvío, percentiles**

	Cantidad de M2	Ambientes	Precio	Precio M2
count	8979.000	8979.000	8979.000	8979.000
mean	92.115	2.723	215530.938	2240.213
std	70.475	1.181	260831.011	840.156
min	10.000	1.000	28100.000	421.000
25%	48.000	2.000	95000.000	1744.200
50%	70.000	3.000	141000.000	2100.000
75%	110.000	3.500	228000.000	2551.500
max	850.000	10.000	5250000.000	12000.000

Algunas prenociones de interés:

La media en M2 es de 92, un 10,8% del metraje del máximo caso registrado. La mediana es de 70 metros, bastante inferior, denotando la influencia de las propiedades más grandes. Esto se asevera al observar una altísima desviación estándar, en torno a los 70 metros.

La media en Precio es \$215000, lo cual no llega a un 5% del valor de la propiedad más cara. Nuevamente, la mediana es ostensiblemente inferior, expresando una tensión hacia arriba por parte de las propiedades más caras. Esto se refleja en una desviación que incluso sobrepasa el valor de la media.

En tanto, los comportamientos en Precio M2 y Cantidad de Ambientes son menos oscilantes: sus desvíos estándar son visiblemente inferiores, y sus diferencias entre Media y Mediana mucho menos considerables.

#### **Moda**

Ahora, cotejemos el comportamiento de la Moda de cada una de estas variables. El resultado por consola es este:

```
La moda de la variable Cantidad de M2 es: 50
La moda de la variable Ambientes es: 3
La moda de la variable Precio es: 120000
La moda de la variable Precio M2 es: 2000.0
```

\_ La moda de 'Cantidad de M2', de 50, es bastante inferior a su media de 92 metros. Esto viene a marcar una asimetría positiva, o asimetría hacia la derecha.

\_ Igualmente, la moda de 'Precio' (de 120.000 dólares) es por mucho inferior a la media de 215.000, denotando nuevamente asimetría positiva. Los altos valores de metraje y precio tensan sus medias.

\_ La moda de 'Precio M2' arroja un valor de 2000 dólares, levemente inferior a la media de 2240 dólares. Esto se asimila a una leve asimetría positiva.

\_ La moda de 'Ambientes' iguala a la mediana en 3, siendo superior a la media de 2,7. Si bien esto marca una ligera asimetría a izquierda, esta variable se comporta similarmente a una simetría.

## **Asimetría**

Ahora, cotejemos estas aseveraciones analizando las asimetrías con su respectiva métrica.

```
La asimetría de la variable Cantidad de M2 es: 2.811114536919886
La asimetría de la variable Ambientes es: 0.3973945891815999
La asimetría de la variable Precio es: 6.279775392534225
La asimetría de la variable Precio M2 es: 2.3380101182408417
```

---

\_ Se refrenda la muy alta asimetría positiva en 'Precio', y las asimetrías igualmente positivas pero menos marcadas en 'Cantidad de M2' y 'Precio M2'.

\_ Sorpresivamente, la variable 'Ambientes' (que exponía una moda superior en valor a la media aritmética) también marca una muy leve asimetría positiva, aunque corrobora una aproximación a la simetría.

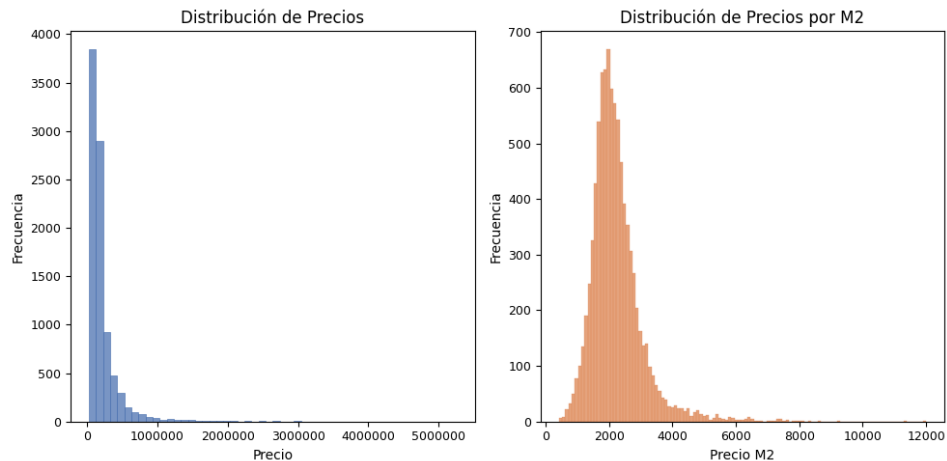
## **10.2.2 Análisis de distribución (variables numéricas)**

Evaluemos ahora la distribución de las variables, primeramente las numéricas por ser las recién evaluadas.

Como se mencionó en las preguntas de investigación, Precio y Precio por M2 serán variables objetivo recurrentes durante la investigación. Por ello, antes de indagar en los factores que influyen en ellas, sería pertinente contemplar su comportamiento individual, en cada caso.

### **Precio y Precio M2**



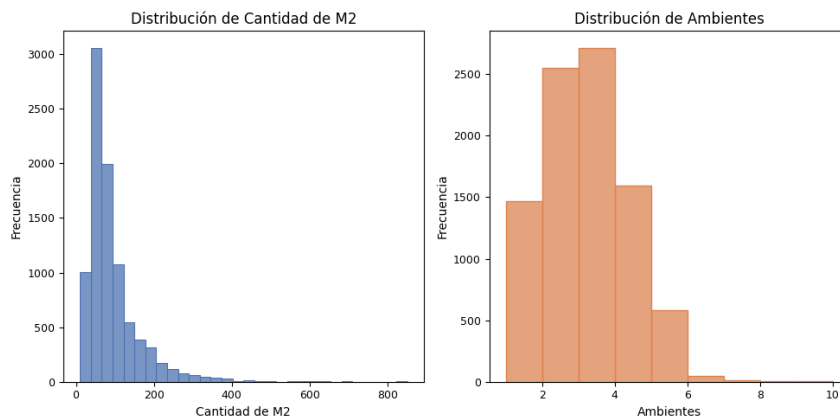


Podemos contemplar que los valores globales de los departamentos denotan una distribución decreciente en función de su valor, con una amplia mayoría concentrada hasta los 300 mil dólares, siendo muy excepcionales los que sobrepasan el millón de dólares.

En tanto, el precio del M2 tiene la mayoría de sus casos entre los 2000 y 3000 dólares por metro cuadrado, siendo excepcionales los casos superiores a los 5000 dólares. Sendas distribuciones presentan una asimetría hacia derecha, siendo esta cola representativa de las propiedades más caras.

En la imagen se corroboran, por cierto, la asimetría positiva muy marcada en 'Precio', y la asimetría positiva más moderada en 'Precio M2'.

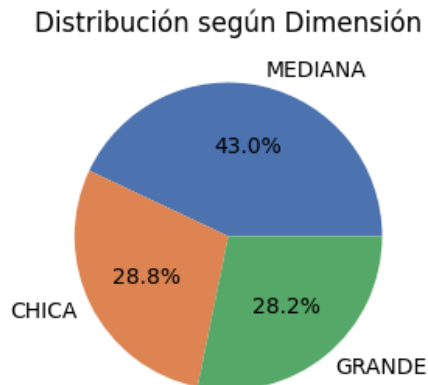
## Cantidad de M2 y Ambientes



El tamaño por cantidad de M2 corrobora la ya vislumbrada moda de 50 M2 y una gran concentración de casos hasta los 100 M2, siendo residuales (y haciendo a una asimetría positiva) los casos más grandes.

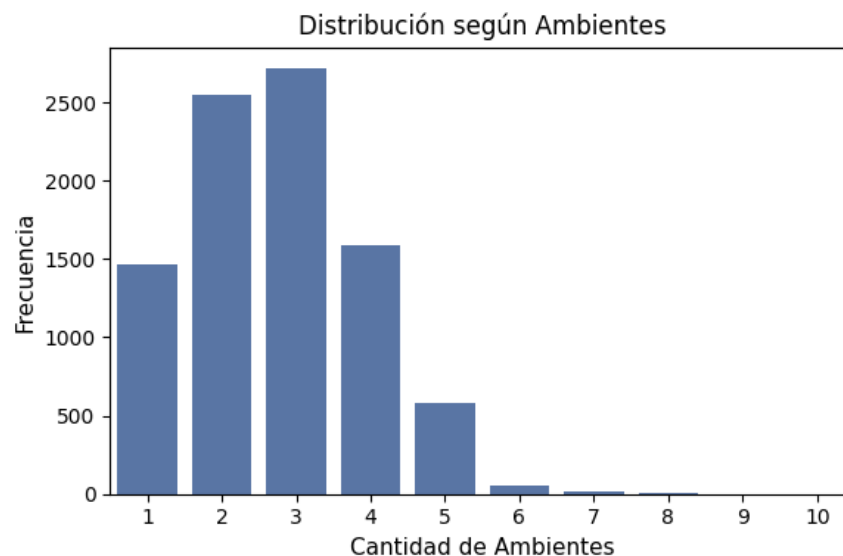
En tanto, la cantidad de Ambientes muestra un comportamiento regular, refrendando la ya vista moda de 3 ambientes, una gran mayoría de casos entre los 2 y 3 ambientes pero un comportamiento más asimilable a la simetría. Aunque presenta una cola a derecha, su cantidad de casos es ínfima.

## Dimensión



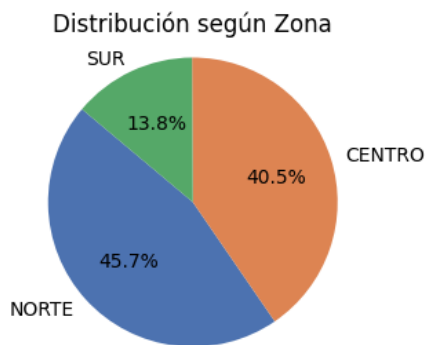
Podemos observar una supremacía de los casos de propiedades Medianas, seguidas por un reparto parejo entre Chicas y Grandes. En todo caso, hay una presencia considerable de todas las categorías. Hay una ligera subrepresentación de propiedades chicas y grandes, pero esto parece desprenderse más del criterio de clasificación que de un sesgo en los datos (recordemos, además, que la muestra en realidad es una muy exhaustiva compilación de propiedades publicadas, realizada por un organismo gubernamental).

## Ambientes



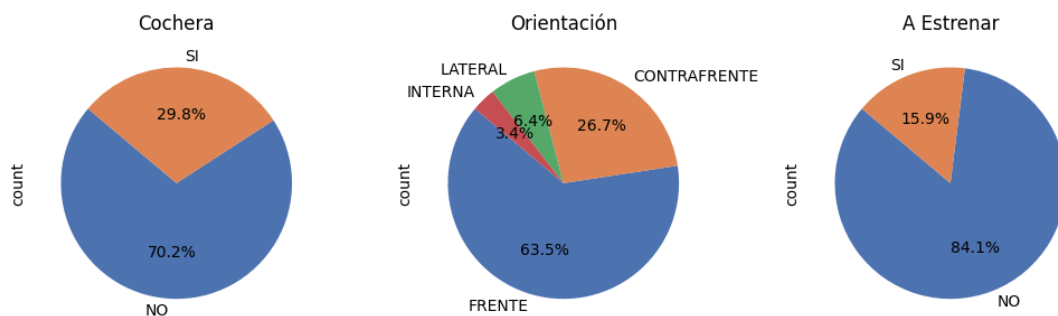
Concluimos que la cantidad de ambientes se concentra entre los 2 y 3 ambientes. Por su parte, los casos a partir de los 6 ambientes es tan poco significativa que no debería ser tenida en cuenta.

## Zona



Hay un reparto parejo de casos entre las zonas Norte y Centro, en tanto la zona Sur queda marcadamente rezagada. Si bien la muestra roza los 9000 casos (lo cual da cierto margen analítico), no deberá perderse de vista en sucesivas conclusiones esta menor representación de la zona sur y sus respectivas comunas.

### Cochera, Orientación, "A Estrenar"



Observamos un 70% de departamentos sin cochera frente a un 30% que la tienen.

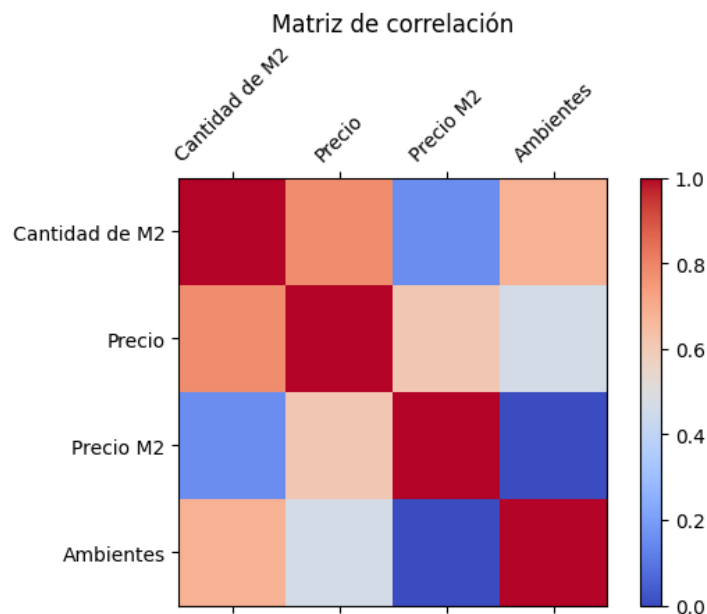
Además, un 63% de unidades tienen orientación frontal, 27% orientación contrafrente, 6,5% orientación lateral y 3,4% interna.

Por último, 84% de departamentos no son "A Estrenar", en tanto 16% lo son.

## 11. Relaciones entre las variables

### 11.1 Correlación entre variables cuantitativas

Sería interesante iniciar este recorrido analítico vislumbrando el grado de correlación entre las principales variables cuantitativas bajo consideración. A tal efecto, se procede a graficar la correspondiente matriz de correlación, para preconcebir hasta que punto unas variables evolucionan en consonancia con otras.



Se observa una notable correlación (en torno al 0.8) entre el tamaño de las propiedades en M2 y el Precio de las propiedades. Sin embargo, al considerar el valor del M2, la correlación con la cantidad de M2 de la propiedad decae hacia el 0.2 aproximadamente.

Asimismo, la cantidad de M2 muestra cierta relación (alrededor de 0.6) con la cantidad de ambientes, lo cual era de esperarse.

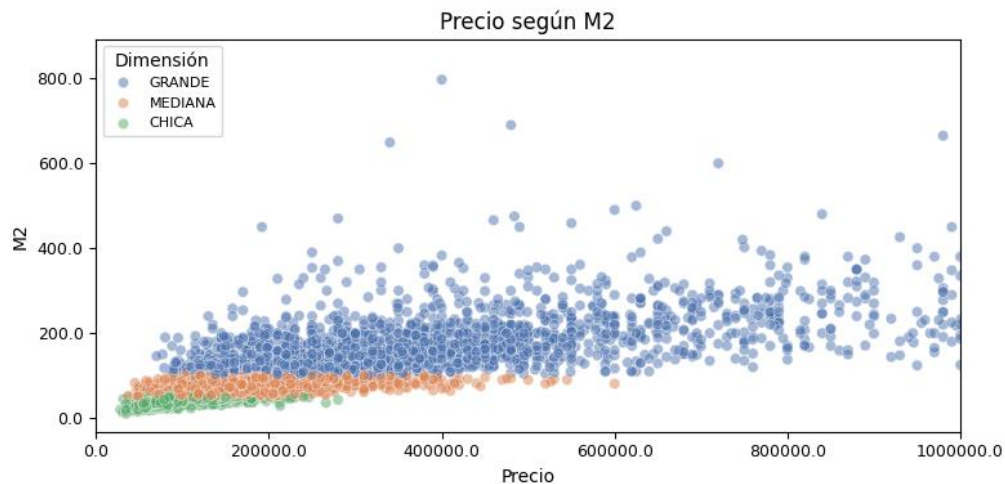
Por otra parte, el Precio de las propiedades expone una considerable correlación con el Precio por M2, en torno al 0.6. Es de esperarse que el Precio total se explique también por otros factores, incluyendo el ya aludido tamaño de la propiedad.

Con respecto a la Cantidad de Ambientes, denota buena correlación con el Precio de la propiedad, de alrededor de 0.5. Sin embargo, esta decae por debajo del 0.2 al cruzarla con Precio por M2, de lo que puede concluirse que su influencia en el valor de la propiedad tiene más que ver con el incremento que supone para su tamaño que por una valoración per se de tener más ambientes.

## **11.2 Influencia del metraje en el precio de las propiedades**

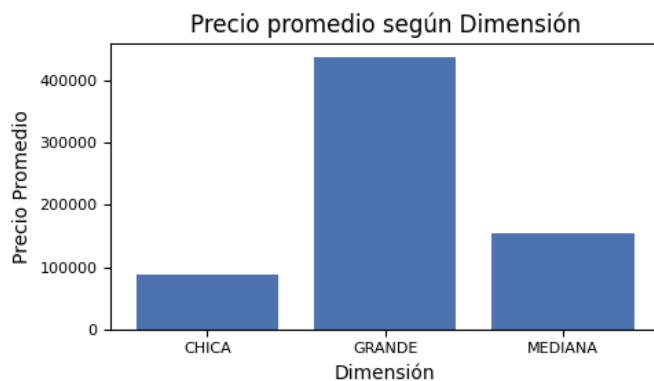
### **Precio según M2**

Ya pasando al ámbito de indagar en las variables candidatas a ser consideradas como factores determinantes en el 'Precio' de los departamentos, sería idóneo iniciar tal recorrido explorando la influencia del tamaño en M2 a tal efecto. Es completamente intuitivo imaginar que una propiedad se valoriza más en la medida en que crece su tamaño, pero... ¿cuán significativo es el peso de este factor? Para responder a esta pregunta, se graficarán 'Precio' y 'Cantidad de M2' segmentando según la categoría "Dimensión" a efectos de una visualización más clara (siendo que la concentración de los casos dificulta considerablemente la misma).



Se observa cierto comportamiento correlativo entre ambas variables. En principio, el tamaño supone un tope máximo para los valores: las chicas (hasta 50 M2) no superan los 300 mil dólares, en tanto las medianas (hasta 100 M2) muy rara vez sobrepasan los 500 mil dólares. Sin embargo, la considerable presencia de valores que no responden linealmente a este comportamiento convergente nos indica que el análisis no capitula acá y deben considerarse otras variables independientes.

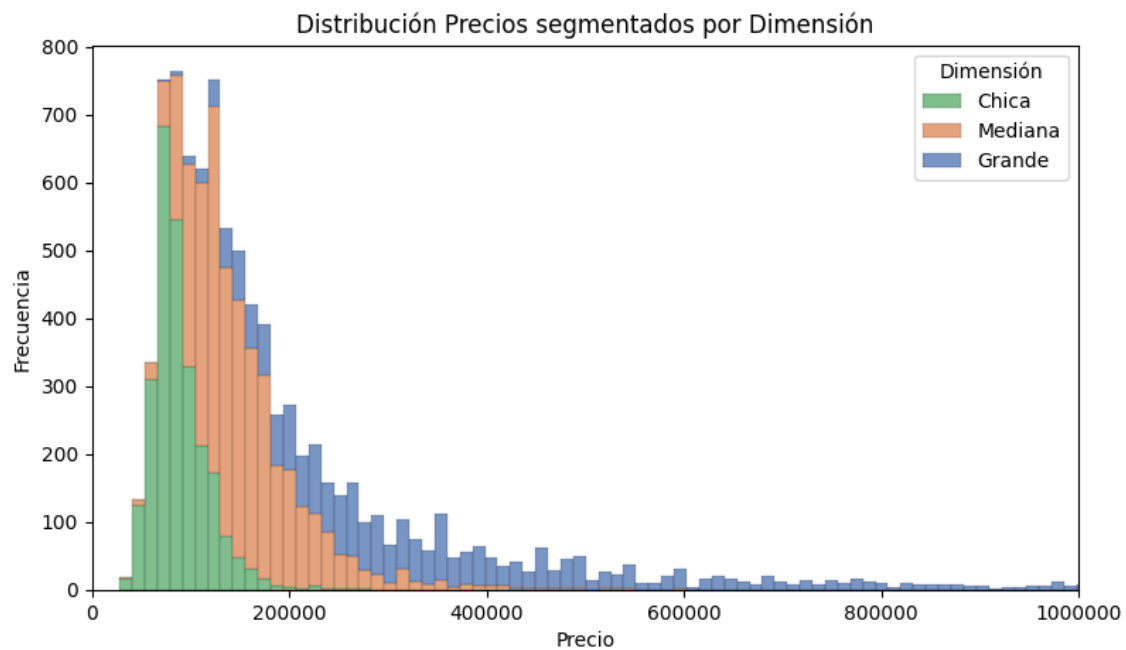
### Precio según Dimensión



Se observa un diferencial notable en las propiedades grandes, que superan los 400.000 dólares. Las medianas rondan los 150.000 dólares y las chicas no alcanzan los 100.000. Este patrón de mayor precio con mayor tamaño es consistente. Aunque las propiedades grandes pueden estar influenciadas por outliers, el análisis de propiedades chicas y medianas también confirma este patrón.

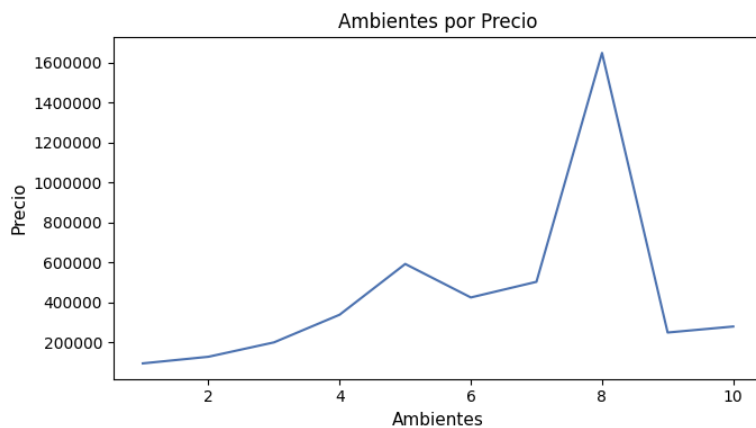
Para salir de dudas, observemos la misma información desde otra perspectiva.

### Distribución de Precios por Dimensión



Se refrenda el mismo patrón visto anteriormente, denotando una influencia sustancial del tamaño en el valor de las propiedades. Las más chicas son las más comunes en torno a los 100 mil dólares. Superado este valor, las Medianas son preponderantes hasta los 200 mil dólares. A partir de allí, ya sin presencia de propiedades Chicas, las Grandes se imponen hasta el final de la distribución.

### Precio por Ambientes

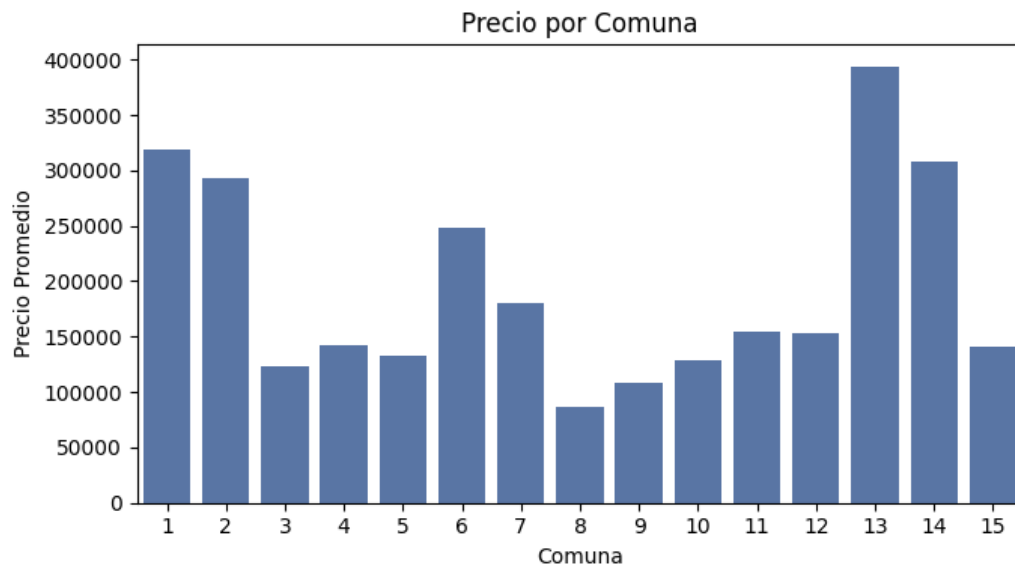


La correlación se plasma entre los monoambientes y los 5 ambientes, y luego entre los 6 y los 8 ambientes, pero hay comportamientos llamativos entre los 5 y 6 ambientes, así como entre los 8 y 10 ambientes. No obstante, como ya observamos en el análisis univariado, esto se debe a la baja cantidad de casos para estas categorías, siendo que tal subrepresentación afecta el normal comportamiento de estas categorías.

## 11.3 Influencia de la Ubicación en el Precio de las propiedades

Remitiéndonos a otra de las preguntas de investigación surgidas en el apartado 3, otro aspecto de interés para imputar las causas del valor de los departamentos bien podría ser la "Ubicación" geográfica de los departamentos. Una primera aproximación sería contemplar cómo se distribuyen las propiedades por Zona. La variable zona toma por categorías Norte (Comunas 2, 12, 13, 14, 15), Centro (1, 3, 4, 5, 6, 7) y Sur (Comunas 8, 9, 10, 11).

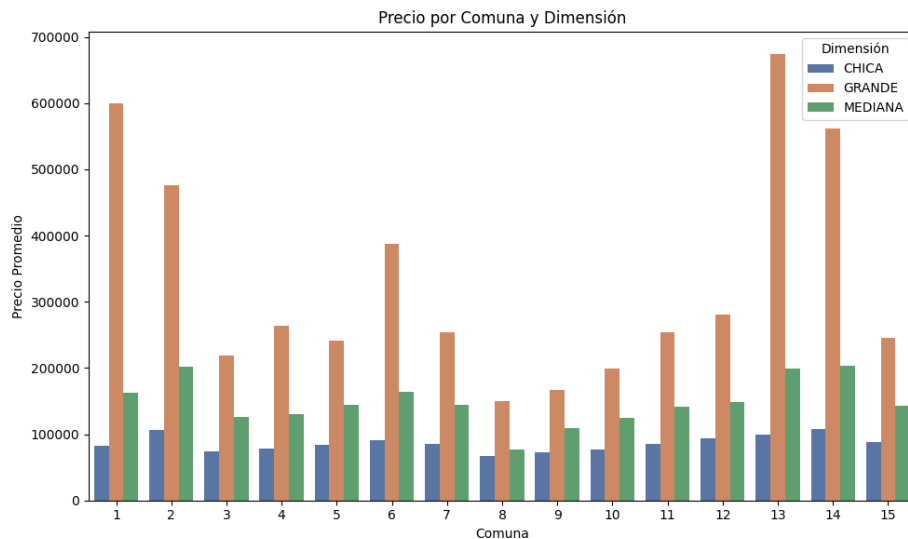
### **Precio por Comuna**



Podemos concluir que este factor es sumamente influyente en el valor de la propiedad. Las comunas 1 (zona centro), 2, 13 y 14 (zona norte) merodean los 300 mil dólares, en tanto otras como la 3, 4, 5, 8, 9, 10, 11, 12 o 15 oscilan entre los 100 mil y los 150 mil dólares. Retomando el análisis por zona, vale mencionar (sin soslayar, nuevamente, que solo representan el 10% de las propiedades publicadas), que 3 de las 4 comunas correspondientes a zona sur (comunas 8, 9 y 10) tienen los menores valores registrados. Las comunas de zona centro tienden a valores más cercanos a los 150 mil dólares, en tanto, por bastante, las comunas de zona norte (excepto comuna 15) registran los valores más altos. Empieza a delinearse un patrón muy marcado alrededor de esta variable.

Ahora, cabe preguntarse: ¿Este patrón discurre parejo entre todo tamaño de propiedades, o se marca más en las más grandes?

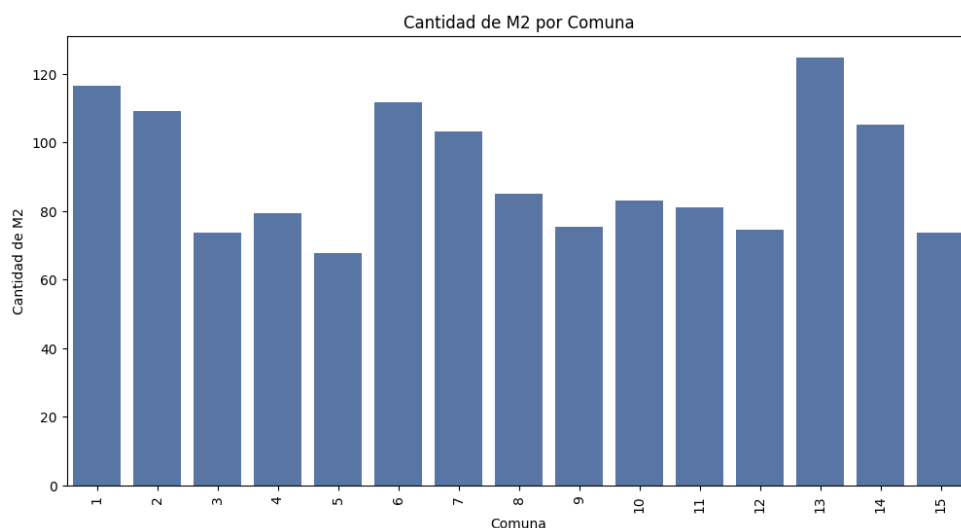
### **Precio por Comuna segmentado por Dimensión**



Las diferencias son leves entre las propiedades chicas en todas las comunas (con oscilaciones entre los 50 mil y 100 mil dólares), se acentúan entre las propiedades medianas (en la comuna 8, por ejemplo, no alcanzan los 100 mil dólares, pero tocan los 200 mil en las comunas 2, 13 y 14), en tanto la varianza se vuelve mucho más significativa en las propiedades grandes.

Vale preguntarse entonces si las comunas con propiedades más caras tienen también propiedades más grandes.

## M2 por Comuna



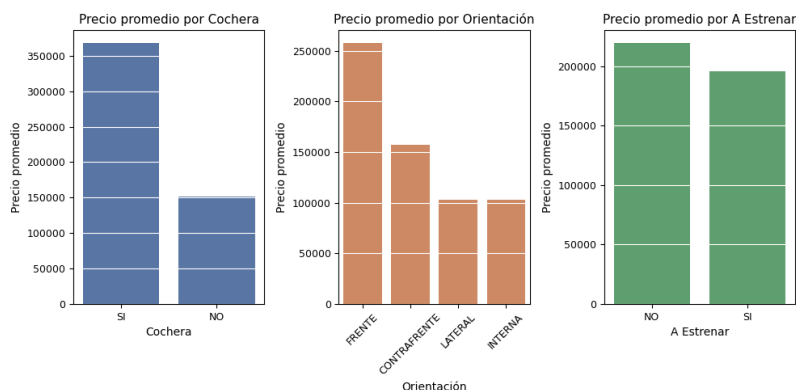
Podemos ver que algunas de las comunas mejor valuadas tienen además una media de metraje más alto, como es el caso de las comunas 1, 13 y 14. Si tomamos en cada caso otras comunas, veremos también cierta proporcionalidad, aunque no concluyente. Tomando conclusiones preliminares: la ubicación es altamente influyente en el precio de las propiedades, pero fundamentalmente en concurrencia con el tamaño de las propiedades (y, potencialmente, de otras variables por cotejar).



## 11.4 Influencia de otros factores

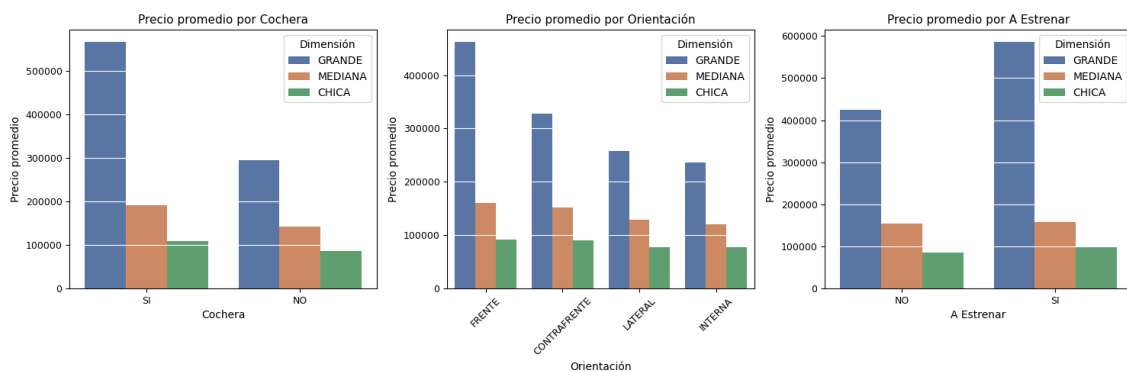
Sería válido, atendiendo al último aspecto abordado en las preguntas de investigación, intuir que la presencia de cochera, la orientación del departamento y el hecho de ser A Estrenar o no pudieran afectar el precio de la propiedad. Indaguemos inicialmente la forma en que las propiedades se distribuyen en relación a estos aspectos.

### Precio por Cochera, Orientación y "A Estrenar"



La variable "Cochera" tiene una influencia notable en la valoración de las propiedades: su presencia incrementa el precio hasta los 350 mil dólares de promedio, reduciéndose a 150 mil dólares en su ausencia. Por su parte, la orientación frontal igualmente denota una importante gravitación, siendo que tales departamentos promedian los 250 mil dólares, reduciéndose a 150 mil dólares para los contrafrente, y a 100 mil dólares en los de orientación lateral e interna. En tanto, la variable A Estrenar no arroja un diferencial significativo: las propiedades que no son a estrenar promedian los 210 mil dólares, y las que son a estrenar se hayan en torno a los 190 mil dólares.

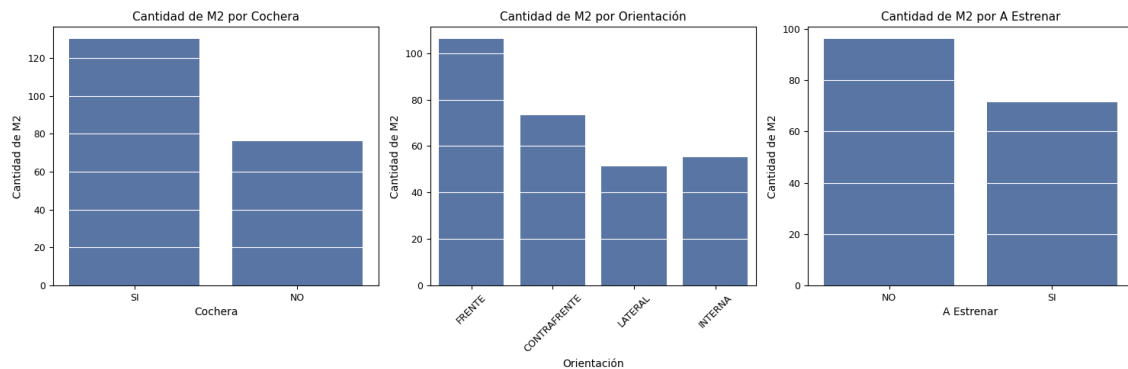
### Precio por Cochera, Orientación y "A Estrenar", segmentados por Dimensión



Tal como en las variables relativas a la Ubicación, el tamaño de la propiedad juega un rol influyente: cuánto más grande la propiedad, más se marca el diferencial en favor de las categorías que más favorecen el valor de las propiedades (tener cochera y orientación frontal). Dicho diferencial va reduciéndose a medida que los departamentos son más chicos. ¿Será que,

nuevamente, las categorías más favorecidas en estas variables son también las más presentes en propiedades de mayor metraje? Cotejémoslo.

## **Metraje por Cochera, Orientación y "A Estrenar"**



Nuevamente, podemos ver que las categorías de estas variables que presentaron mayor precio (presencia de cochera, orientación frontal, propiedades No a Estrenar) se corresponden también con departamentos con (en promedio) mayor metraje.

La línea analítica sigue entorno a una idea común: el tamaño se plasma siempre como una variable preponderante en la valoración de las propiedades. Otros factores marcan también una influencia, pero siempre subsidiaria al tamaño, y en todo caso deberá analizarse en la etapa de aplicación de machine learning (con sus respectivas métricas) cual es el peso específico de cada variable.

## **12. Insights**

Tanto de la correlación entre variables como de las distintas visualizaciones desplegadas, puede desprenderse que el tamaño de las propiedades y, correlativo a esto (pero en menor medida), la cantidad de ambientes, son factores influyentes en el precio de las mismas. Sin embargo, el hecho de que esta sumatoria aritmética de metros y/o ambientes suponga un incremento en el valor de los departamentos solo refrenda algo a priori intuitivo, y esta influencia no se refleja en cambio (tal lo observado en la baja correlación entre Cantidad de M2 y Precio M2) en el valor de cada metro cuadrado. Para esta instancia, deberían considerarse otras variables.

Por otro lado, el valor de las propiedades demostró ser notoriamente superior en ciertas zonas de la capital (Comunas 1, 2, 13 y 14) con diferenciales respecto a las demás comunas de más de 100%. Sin embargo, esas diferencias se plasman mucho más marcadamente entre las propiedades grandes, decreciendo cuanto más chicas. El factor dimensional juega un rol influyente, incluso en el impacto de otras variables.

Pasando a otros factores, la orientación frontal y, especialmente, la presencia de cochera, demostraron ser factores de valorización de los departamentos, en mucha mayor medida que el hecho de ser "A Estrenar". Esto último podría deberse a que los

departamentos nuevos suelen tener algunas dificultades edilicias que se plasman durante la residencia de los primeros inquilinos.

Ahora bien: nuevamente, al segmentar según Dimensión del departamento, el diferencial de estas variables muestra ser mucho mayor en las propiedades grandes y decrecer hacia las medianas, y aún más hacia las chicas.

Como conclusión preliminar, podemos decir que las dimensiones de un departamento son el principal factor en su valorización, en concomitancia con su ubicación, la presencia de cochera y su orientación. Es probable que algunos departamentos grandes no sean (más que proporcionalmente) más valiosos que otros por el mero hecho de su tamaño (lo que pareciera desprenderse de la débil correlación entre Cantidad de M2 y Precio de M2), pero cuando esta variable es asistida por otras condiciones deseables (orientación frontal, presencia de cochera, y probablemente otras variables ausentes en este dataset), parecen definir un perfil de propiedad muy bien valorada en el mercado. No es solo el tamaño, ni solo la ubicación o la presencia de ciertas amenities, sino la concurrencia de varias de esas variables (en sus categorías mejor consideradas) las que potencian el valor inmobiliario en sus unidades mejor consideradas.