

Introducción

Precios claros es un programa implementado por el gobierno Nacional, en la que, mediante un sistema electrónico de publicidad de precios, te brinda información sobre los **valores** de productos de consumo masivos en los comercios.

En el siguiente informe se analiza los distintos precios reportados por los establecimientos ubicados en diferentes provincias de Argentina con el objetivo de indagar si los valores de los distintos productos han aumentado o disminuido en términos de otro bien. La forma de medir el aumento de los precios de los bienes en términos de otros es a través del cálculo de precios relativos.

A continuación, se comienza con la presentación del dataset y su limpieza correspondiente.

1. Presentación del Dataset

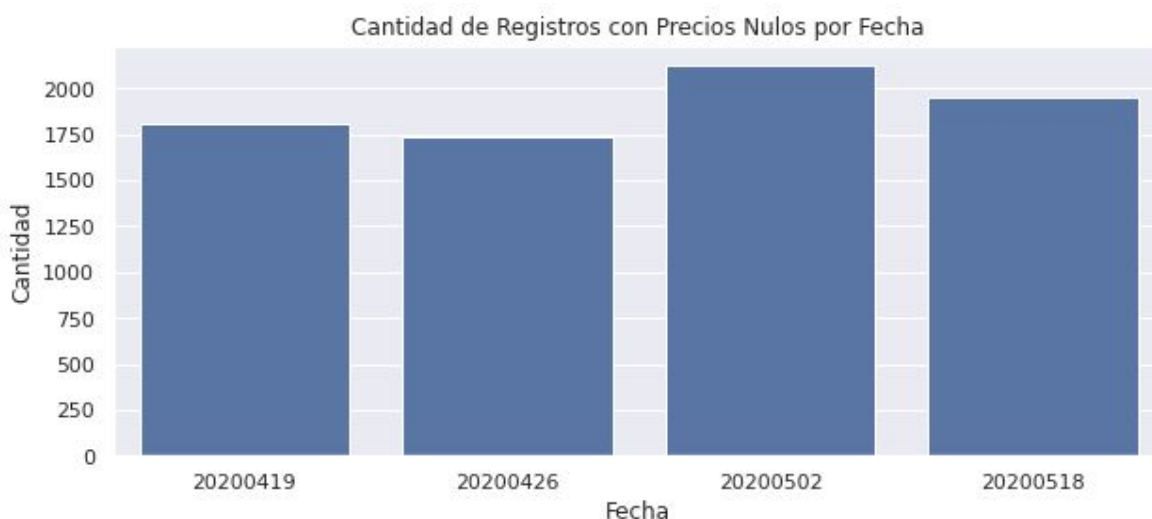
El dataset utilizado contiene 2 millones 200 mil productos reportados y 26 variables que brindan información relacionada. Algunas de estas variables son:

- ❖ Fecha de reporte
- ❖ Información de la sucursal que reporta el producto (nombre, región, provincia)
- ❖ Información sobre el producto (marca, unidad de medida, cantidad, precio)

Limpieza del datasets

Para realizar un análisis detallado sobre los precios de los distintos productos reportados se observaron los valores nulos encontrados en la base de datos. En la principal variable de análisis, el precio, se han reportado sólo el 0.34 % de registros nulos, lo cual una vez analizados, se excluyeron.

Valores nulos



La mayor cantidad de registros con precios nulos se encuentra en las fechas de mayo del año 2020. Indagando más sobre estos *missings*, el 75% de los datos fueron reportados por el Supermercado “Diarco”, principalmente en las provincias de Buenos Aires y en CABA. El resto de los valores removidos, no tienen asociado un supermercado ni información geográfica.

	Sucursal	cantidad
0	Diarco	5621
1	NaN	2002

	Provincia	cantidad
0	Provincia de Buenos Aires	5475
1	NaN	2002
2	Ciudad Autónoma de Buenos Aires	146

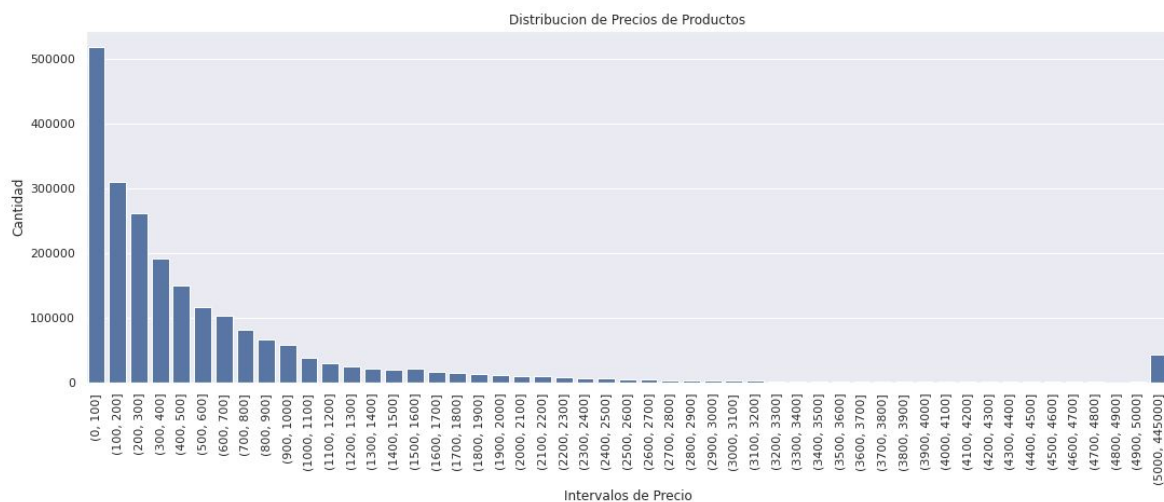
Algunos de estos registros que se excluyeron del análisis fueron principalmente de la marca Loreal, Arcor, Pantene, Sin Marca, entre otras.

	index	cantidad
0	SIN MARCA	64702
1	LOREAL	26580
2	ARCOR	26535
3	PANTENE	24952
4	DOVE	24665
...

Las bases de datos con la que se trabaja, al reportar distintos productos, muestra cantidades, unidades de medida y precios distintos, en consecuencia, para poder compararlos unos con otros se realizó una normalización de unidades de medidas dejando todos los productos en kilogramos, metros y litros de acuerdo a tipo de producto analizado. Una vez obtenidos los precios normalizados, se analizaron las variables más importantes que se relacionan con el precio.

2. Variables más relevantes o de interés

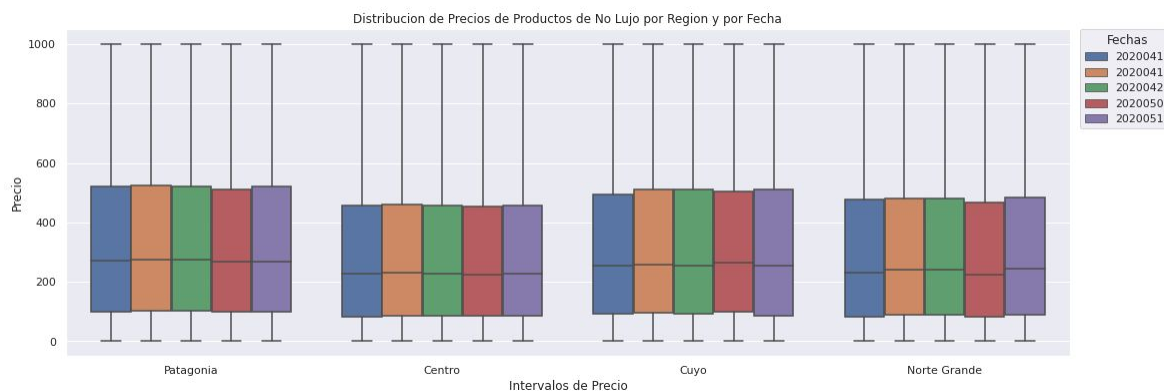
Precio



La distribución de la variable precio, es altamente concentrada a la izquierda, es decir que, la mayoría de los productos reportados muestran precios menores a \$ 1000, mientras que solo algunos pocos se encuentran en los demás precios.

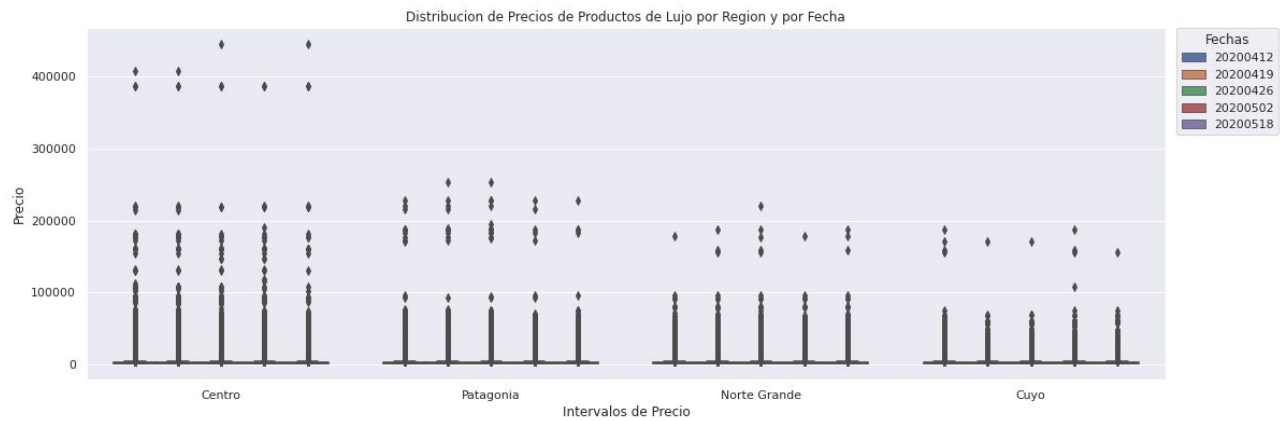
En consecuencia, se realizó una su clasificación de los productos, para analizar mejor su comportamiento. Se llaman “bienes de lujo” a todos aquellos bienes que tienen valores mayores a \$ 1000 pesos, mientras que se llaman “bienes normales o no de lujo” aquellos que tienen precios inferiores a los \$1000.

Cuando se analiza, la distribución de los precios de este último grupo, se observa que el comportamiento es similar. La distribución es asimétrica derecha, y no muestra grandes variaciones para cada una de las regiones y fechas analizadas.



En la mayoría de las regiones el 75% de los registros se encuentran en valores menores de \$ 500 pesos, mientras que solo el 25% se encuentra en la otra mitad del intervalo preestablecido.

Si se analiza la distribución del grupo “bienes de lujo”



Existe una concentración de precios en los valores inferiores de la distribución, pero se puede observar una importante cantidad de registros con precios excesivamente altos.

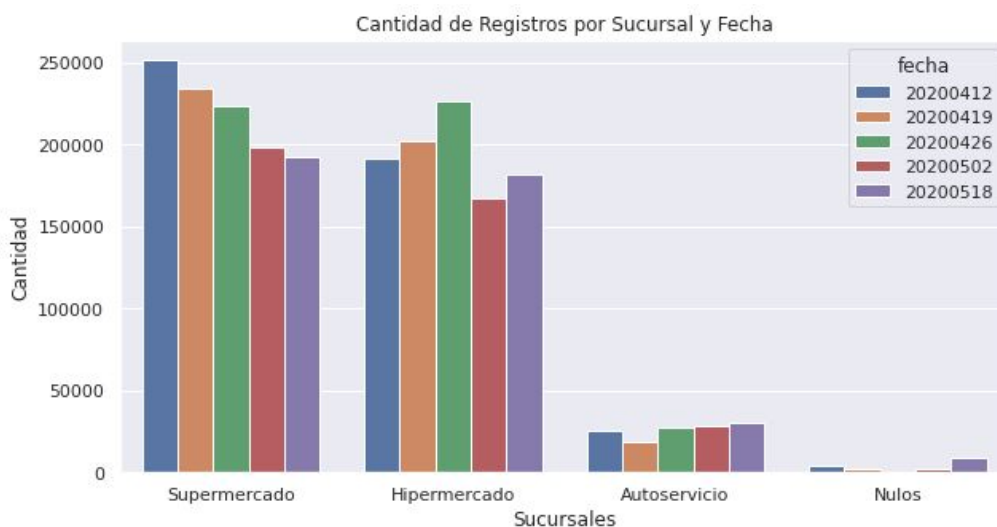
En la región CENTRO, para las 5 fechas analizadas, existen productos valorados a más de \$ 35.000 mientras que en las demás regiones los valores atípicos se encuentran dentro de los \$20.000 -\$25.000. Algunos de estos productos atípicos se encuentran a continuación.

	nombre	precio	unidad	precio_normalizado
	Champagne Rose Estuche Dom Perignon 750 Cc 1 Un	31000.0	cc	41333.333333
	Whisky 25 Años Chivas Regal 700 Ml	32469.0	ml	46384.285714
	Whisky Chivas Regal con Volca 4.5 Lt	35900.0	lt	7977.777778
	Whisky Chivas Regal con Volca 4.5 Lt	35900.0	lt	7977.777778
	Whisky Royal Salute Chivas Regal 700 Ml	32900.0	ml	47000.000000
	Whisky Royal Salute Chivas Regal 700 Ml	32900.0	ml	47000.000000

Sucursal

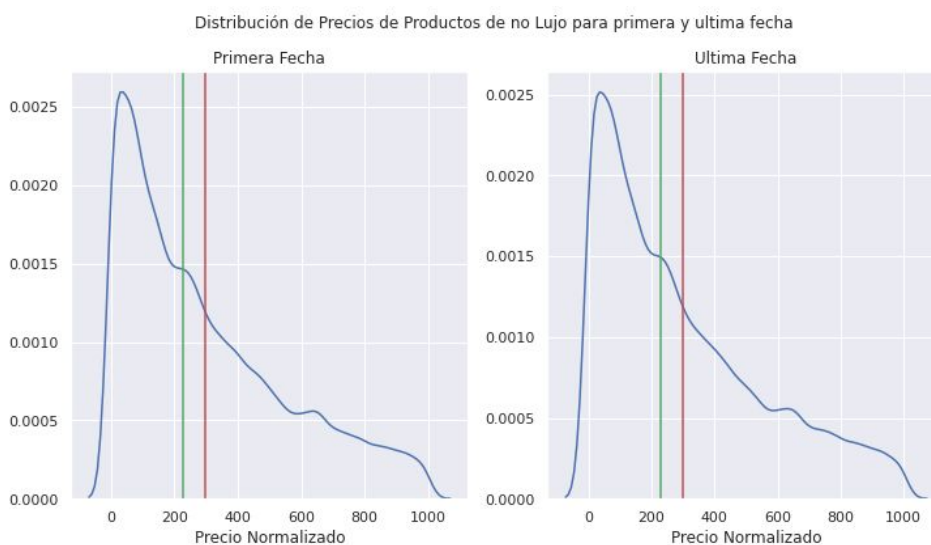
La mayoría de los precios son reportados por supermercados e hipermercados, mientras que existe una mínima cantidad de productos que son reportados por autoservicios.

Es importante aclarar que existe missings en esta categoría, lo cual su explicación está fundada en que son registros que a pesar de tener precios no tienen información en los datos propios de sucursal.



Fecha

Un interrogante planteado mientras se analizan los precios, es si existen variaciones en los mismos, en el periodo analizado, es por ello que si se comparan las distribuciones de precios de los bienes de no lujo para la primer y última fecha, solo para la región centro, y no se encuentran grandes diferencias. El precio promedio (en rojo) para el 12/04/2020 es de \$227.13 mientras que el de \$229.80, y el valor que acumula el 50% de los registros (mediana, en verde) es de \$227 y \$228, respectivamente.

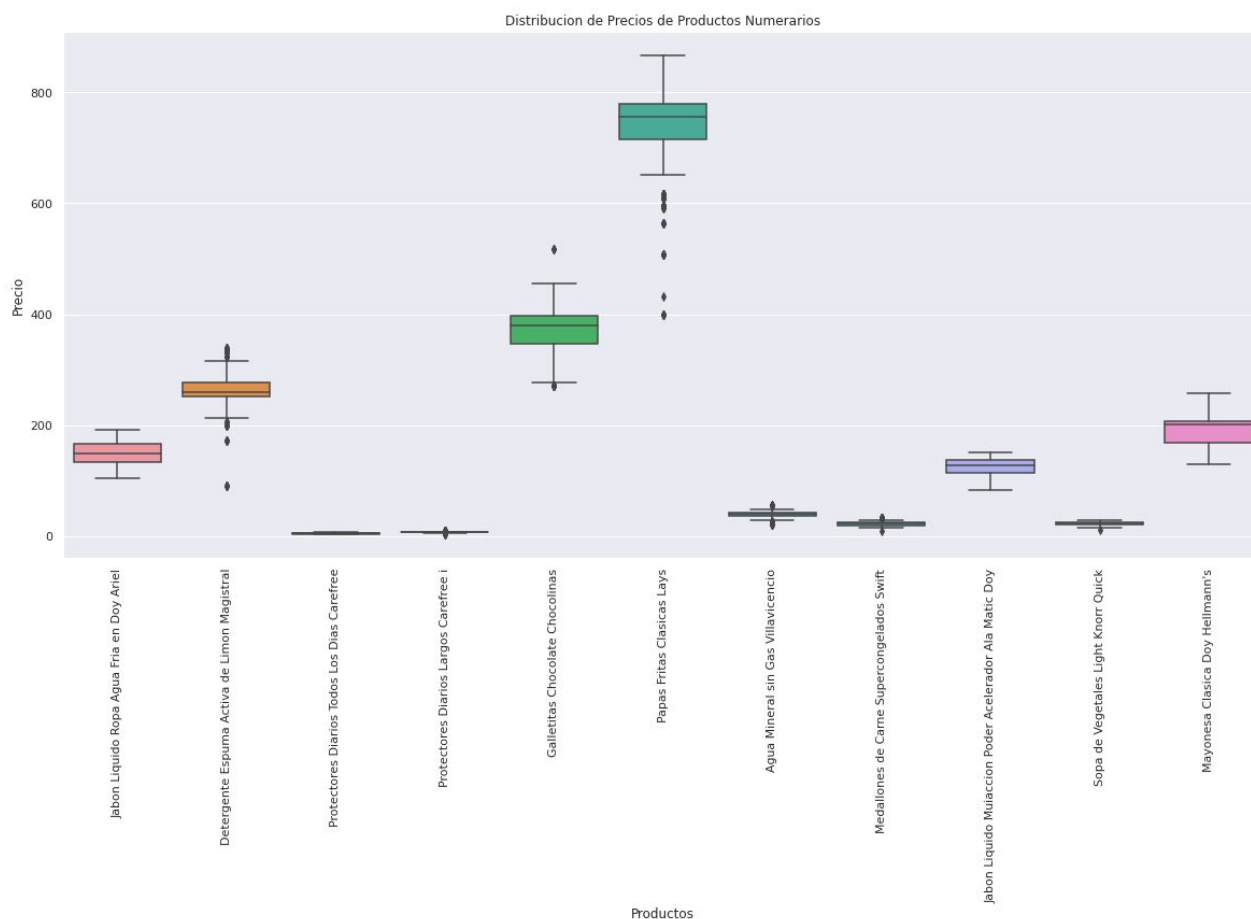


3. Precios relativos

Para realizar el cálculo de los precios relativos, es necesario elegir un producto llamado “bien numerario” que será el bien con el que se va a comparar el resto de los productos para evaluar su precio o valor.

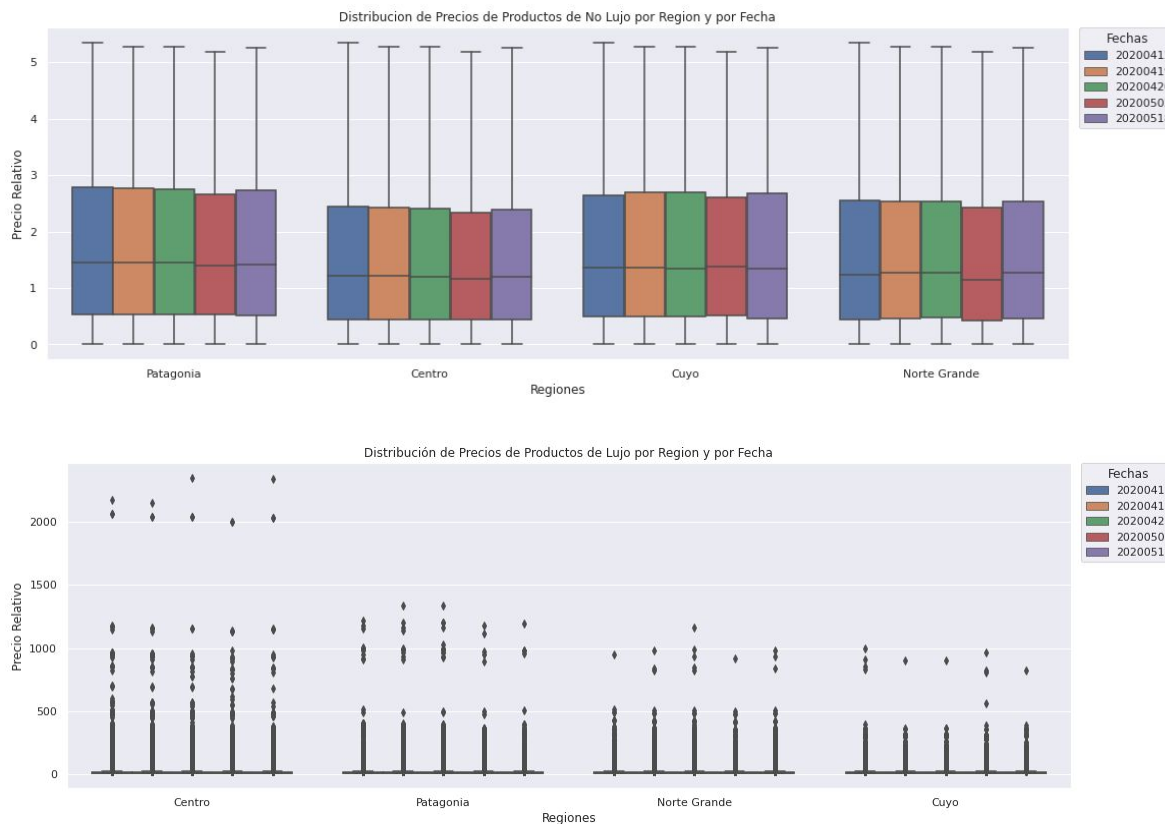
El bien numerario elegido es el precio promedio del **producto_id= 7794000960077**, “**Mayonesa Clasica Doypack Hellmann's 950 Gr**”, cuyo bien se encuentra en todas las fechas y en todas las regiones.

A continuación podemos observar la distribución del bien numerarios y de otros bienes candidatos que fueron analizados:



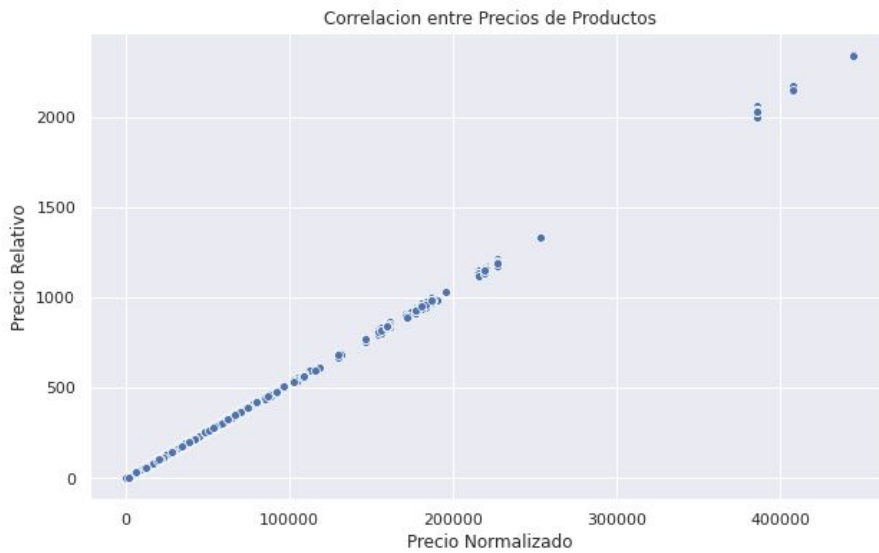
Esta elección, permite realizar el cálculo de los precios relativos de los productos reportados e indagar en su comportamiento.

A continuación, se indaga cómo se comportan los precios relativos respecto a la división entre productos de lujo y no lujo y otras variables de interés.



Al igual que con los precios normalizados, las distribuciones de los precios relativos de los bienes de no lujo como los de lujo son asimétricas derechas, con una gran acumulación de registros en los valores más bajos.

Tanto los precios relativos como los nominales tienen el mismo comportamiento, y esto sucede porque son variables que están correlacionadas de manera perfecta, esto es comprobado por el gráfico de dispersión y el cálculo del coeficiente de pearson, medida descriptiva que mide el grado de asociación que tienen dos variables.



```

precio_normalizado precio_relativo
precio_normalizado    1.000000    0.999947
precio_relativo        0.999947    1.000000

```

Es decir que los precios más baratos o los más caros analizados a través de los precios normalizados serán los mismos que en el análisis de precios relativos.

La diferencia importante cuando se cambia el análisis de precios nominales a precios relativos, es la lectura del precio.

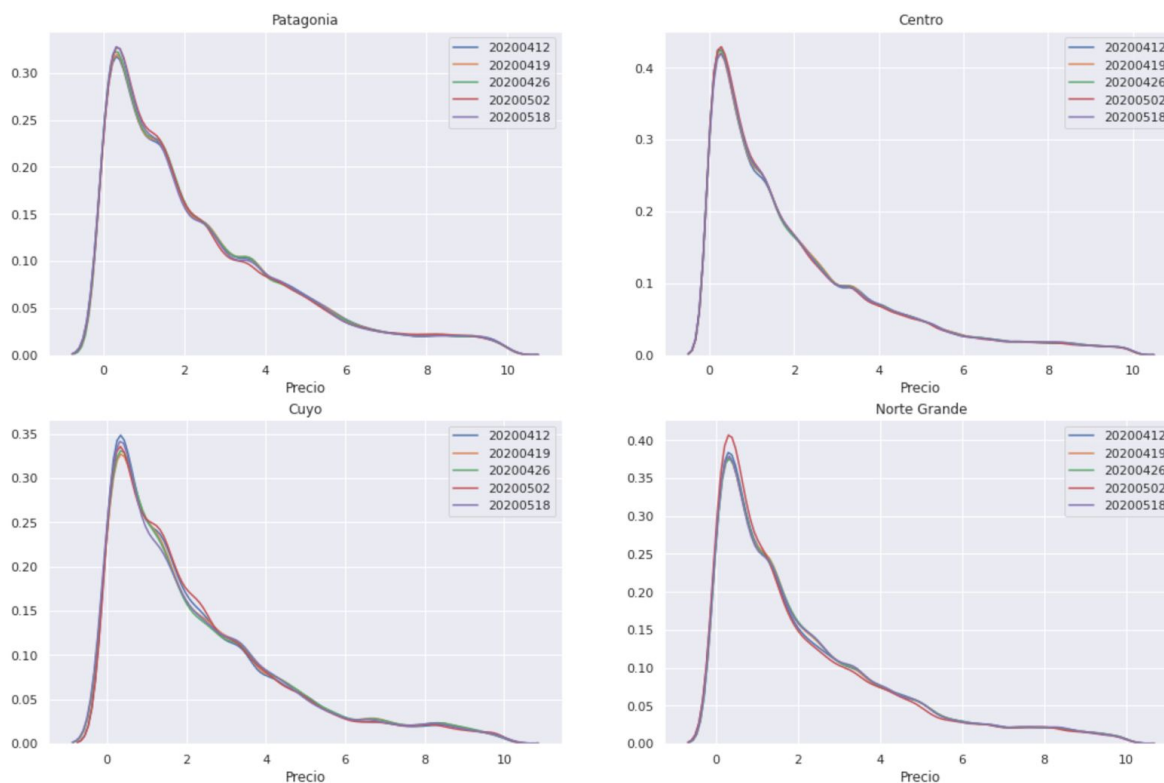
A modo de ejemplo, si se observan los precios relativos de los precios excesivamente altos que se vio en la sección anterior,

	nombre	precio	unidad	precio_normalizado	precio_relativo
953034	Champagne Rose Estuche Dom Perignon 750 Cc 1 Un	31000.0	cc	41333.333333	218.188897
961750	Whisky 25 Años Chivas Regal 700 MI	32469.0	ml	46384.285714	244.851681
1815078	Whisky Chivas Regal con Volca 4.5 Lt	35900.0	lt	7977.777778	41.913686
1815079	Whisky Chivas Regal con Volca 4.5 Lt	35900.0	lt	7977.777778	41.913686
1815206	Whisky Royal Salute Chivas Regal 700 MI	32900.0	ml	47000.000000	246.928817
1815207	Whisky Royal Salute Chivas Regal 700 MI	32900.0	ml	47000.000000	246.928817

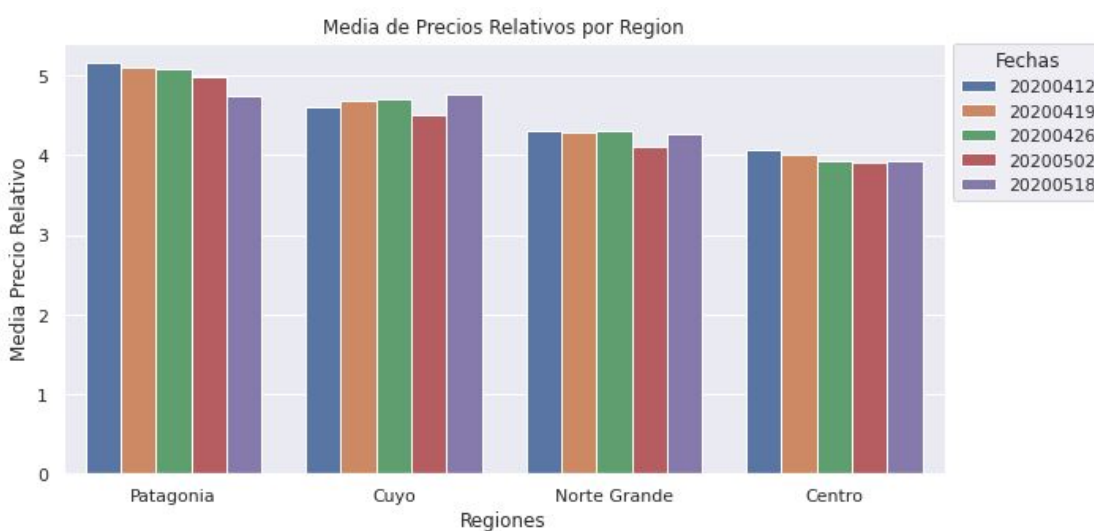
Antes, se necesitaba \$41.333,33 para comprar un litro de Champagne Rosé Estuche Dom Perignon, mientras que ahora se necesitan 218 litros de Mayonesa Clásica Doypack Hellmann's (dicho análisis se extiende para cada uno de los productos analizados).

Región-provincia

El precio también puede ser interesante analizarlo por región para cada una de las fechas.



A simple vista, los precios relativos se comportan de una manera similar, pero si se analizan sus métricas descriptivas, principalmente el precio promedio por fecha y región, se observa que:



- 1) La región patagonia es la que tiene un precio relativo promedio más alto en todas las fechas excepto la última donde la región Cuyo tiene unos centavos más.
- 2) La región centro es la que tiene un promedio menor en los precios relativos en todas las fechas
- 3) La región patagonia ha mostrado un descenso en los precios relativos en el transcurso de los precios, al igual que la zona centro. Las otras regiones, cuyo y norte tienen variaciones más irregulares.

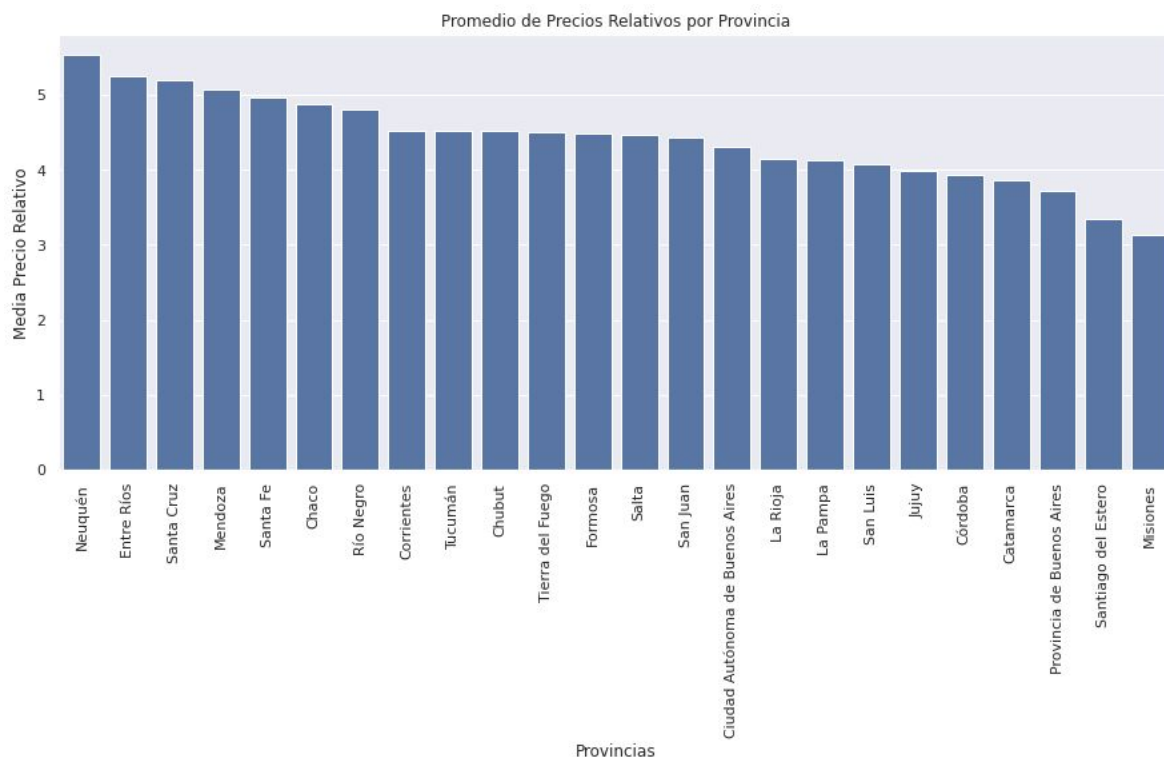
Si se eligen la distribución de precios de la región Cuyo y Patagonia (para la última fecha disponible), las cuales tiene promedios similares en la muestra, para ver si los precios promedios de sus poblaciones son estadísticamente distintos se realiza un test de diferencias de medias para muestras independientes, en la que se plantea como:

H_0) precios promedios de las dos regiones son iguales

H_1) los precios promedios de las regiones son distintos

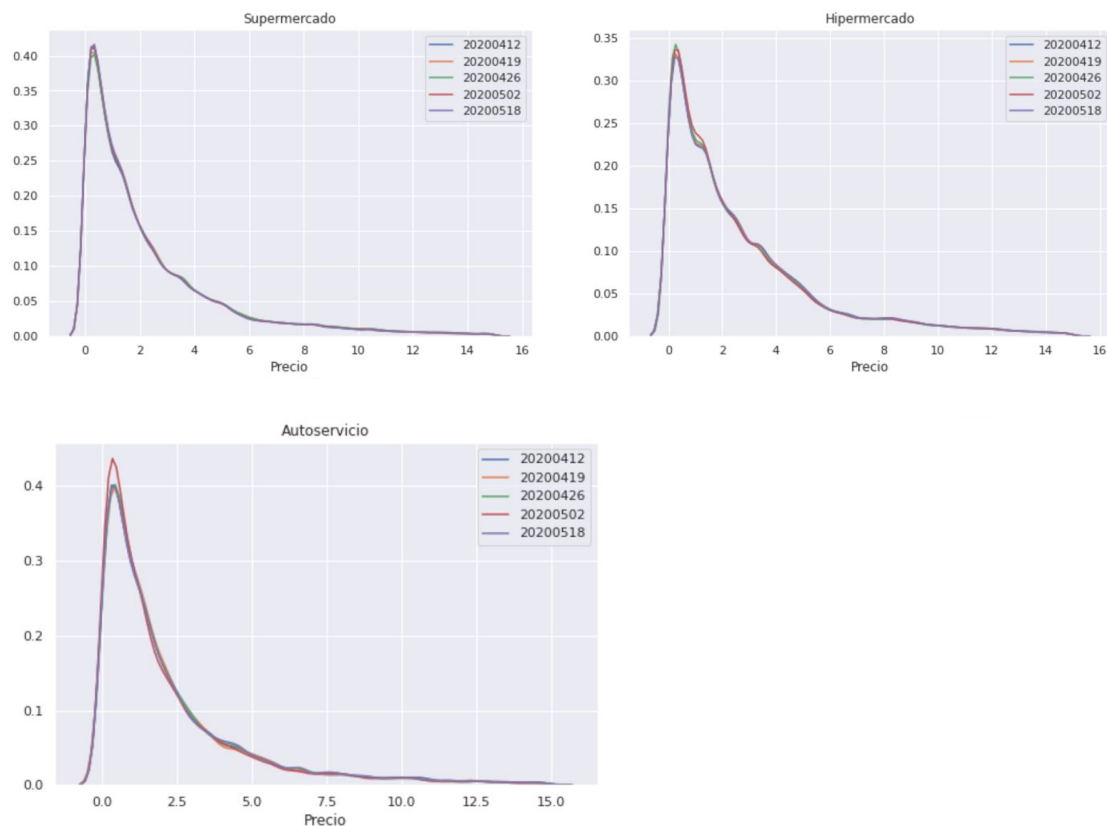
Una vez realizada la prueba, se obtuvo un p valor insignificante y menor a un nivel de significancia del 5%, se rechaza la hipótesis nula, es decir que los precios relativos promedios son significativamente distintas.

Adicionalmente, si se intenta indagar cuales son las provincias que reportan precios mas baratos o mas caros, Neuquén es la provincia que muestra precios promedios relativos más altos mientras que Misiones tiene precios promedios más bajos.

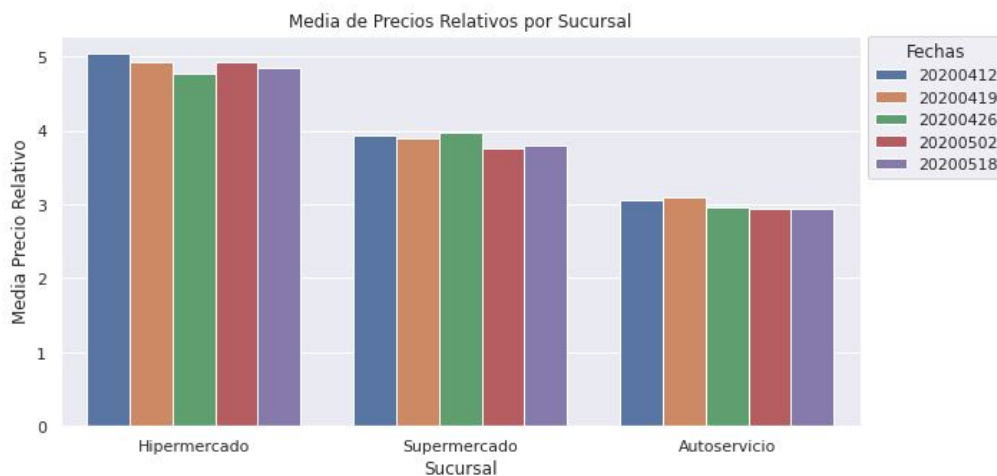


Para el caso particular de la provincia de Córdoba, si se plantea como interrogante, cómo son los precios relativos respecto al promedio nacional, para la última fecha del dataset se observa que hay una probabilidad del 20% de los precios superen el promedio nacional.

Sucursal



A simple vista, las distribuciones de los precios para cada uno de los tipo de sucursales son similares, pero si se analizan los precios promedios, se observan diferencias. El autoservicio, en todas las fechas, tienen precios relativos más baratos mientras que el hipermercado, es el tipo de sucursal que muestra precios más caros.



4. Respuestas a preguntas disparadoras

A lo largo del informe se fueron respondiendo algunas de las preguntas disparadoras. A continuación nos enfocaremos en el resto de las preguntas:

- **¿Qué productos están en todas las provincias?**

La lista de productos que cumplen con la condición de aparecer en toda las provincias es extensa (hemos encontrado más de 641 productos). Por lo tanto listaremos sólo algunos de ellos:

ID	Nombre Producto
7500435004664	Jabon Liquido Ropa Agua Fria en Doypack Ariel 3 Lt
7500435019248	Detergente Espuma Activa de Limon Magistral 750 Cc
7790010616751	Protectores Diarios Todos Los Dias Carefree 100 Un
7790010616812	Protectores Diarios Largos Carefree 60 Uni
7790040929906	Galletitas Chocolate Chocolinas 170 Gr
7790310981191	Papas Fritas Clasicas Lays 250 Gr
7790315000422	Agua Mineral sin Gas Villavicencio 2 Lt
7790360966797	Medallones de Carne Supercongelados Swift 2 Un
7791290011205	Jabon Liquido Multiaccion Poder Acelerador Ala Matic Doypack 3 Lt
7794000598591	Sopa de Vegetales Light Knorr Quick 5 Un
7794000960077	Mayonesa Clasica Doypack Hellmann's 950 Gr

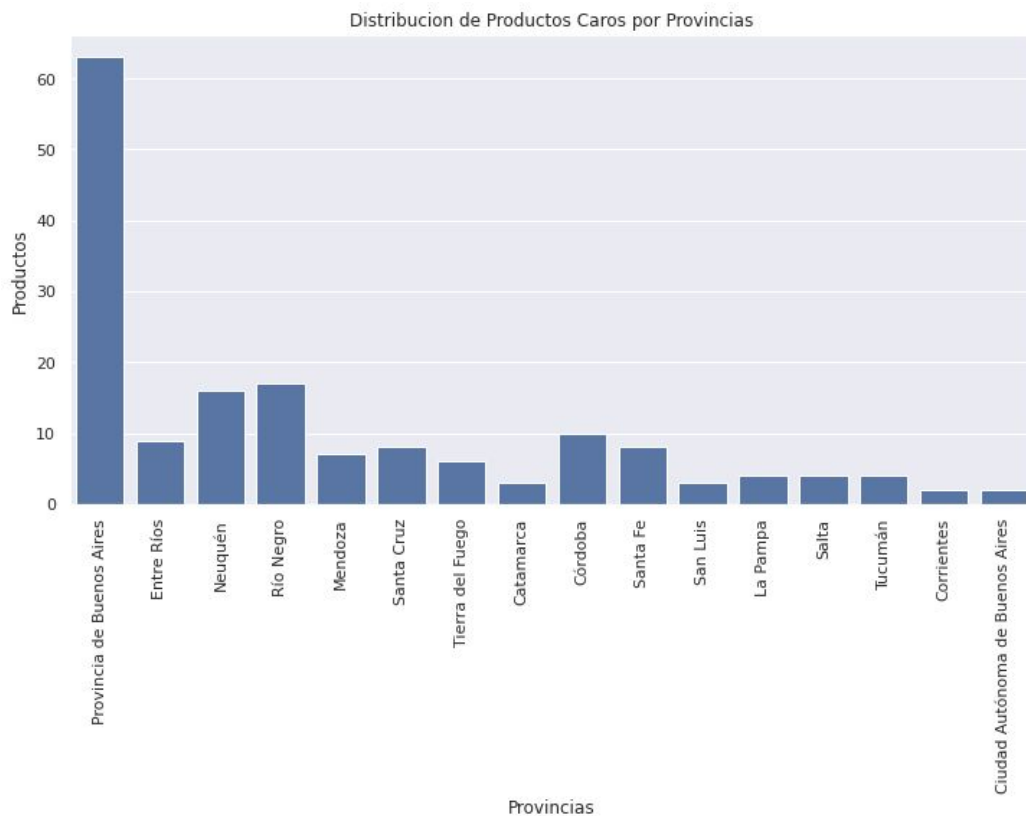
- ¿Qué tipos de productos se encuentran entre los más caros? ¿En qué provincia se encuentran mayoritariamente?

Considerando los precios normalizados, algunos de los productos más caros son:

ID	Producto	Precio Normalizado
41554443493	Mascara para Pestañas Maybelline Falsies 2 Gr	\$407,500.00
41554220186	Delineador de Ojo Maybelline Studio Last 3 Gr	\$215,000.00
7790520022462	Repuesto Aromatizante en Gel Manzana y Canela Glade Sensations 8 Gr	\$174,686.54
3348900417892	Perfume Dior J Adore Mujer 30 MI	\$173,000.00
7790520024688	Aromatizante Acqua Full Glade Electric Car 3.2 MI	\$172,080.34
8411061781197	Perfume Carolina Herrera 212 Vip Rose 30 MI	\$161,666.67

Se observa que los productos, en su mayoría, de cosmética se encuentran entre los más caros considerando su precio relativo.

La mayoría de los productos caros se encuentran en Provincia de Buenos Aires como podemos visualizar en el siguiente gráfico:



- **¿Qué tipos de productos se encuentran entre los más baratos? ¿En qué provincia se encuentran mayoritariamente?**

De igual manera, los productos más baratos de acuerdo a su precio normalizado son:

ID	Producto	Precio Normalizado
7790580108304	Caramelos de Leche Dulcilac 120 Un	0.00258
7798049590233	Papel Higienico Gamuza 240 Mt	0.00912
7798049590189	Papel Higienico Natural Gamuza 360 Mt	0.01386
7798049590417	Papel Higienico Gamuza 300 Mt	0.01597
7791415000053	Palillero Fade Lince 150 Un	0.0166
7790122007164	Escarbadiante JR 60 Un	0.02
7798049590141	Papel Higienico Blanco Gamuza 180 Mt	0.02106
7798063140087	Barrita de Cereal Tunki Crock 13 Gr	0.02692
7791854000089	Papel Higienico Mega Family Pack 2 Un 90 Mt	0.0305
7790430062220	Sopa de Gallina con Cabellos de Angel Maggi Caseritas 68 Un	0.03544

Dichos productos se encuentran en Provincia de Buenos Aires.

5. Reflexiones finales

Los precios relativos permiten comparar si estos varían a través del tiempo. En términos generales, comparando los precios relativos promedios se observa que el comportamiento por provincia, región y dependiendo del tipo de sucursal el comportamiento de las distribuciones son distintos.

A nivel de ejemplo, sabemos que la región Patagonia es una de las regiones más caras, pero muestra disminuciones en sus precios relativos a lo largo de las fechas.

Los próximos interrogantes a resolver están relacionados con el comportamiento de productos en particular que pueden ser analizados de manera completa, es decir saber si han aumentado o disminuido en el tiempo, cuáles de las regiones tienen productos más caros, más baratos y en qué sucursales se encuentran, entre otros.

Consideramos oportuno mencionar que en el análisis de los precios normalizados, hemos encontrado que algunos productos tenían erróneamente cargados los valores de la variable **cantidad**. Por ejemplo:

Alimento para Perros Adulto Carne Pollo y Cereal Pedigree 1.5 Kg

1. Según la descripción del producto la presentación es de '1.5 Kg'
2. Ahora bien, considerando la valor de la variable unidad que acompaña al registro, ésta indica 'gr'

Huevo de Chocolate con Leche Toy Story Arcor 50 Gr

1. Según la descripción del producto la presentación es de '50 Gr'
2. De acuerdo al valor de la variable cantidad que acompaña al registro, ésta indica '1.0'

¡Esto repercute de manera directa en el cálculo del precio normalizado ya que el valor final de la variable toma dimensiones extraordinarias posicionando a estos productos como los más caros. Para el presente análisis no se han extraído/corregido estos datos ya que consideramos que es una tarea para realizar en la próxima entrega, más relacionada al **Análisis y Curación** de los datos.