



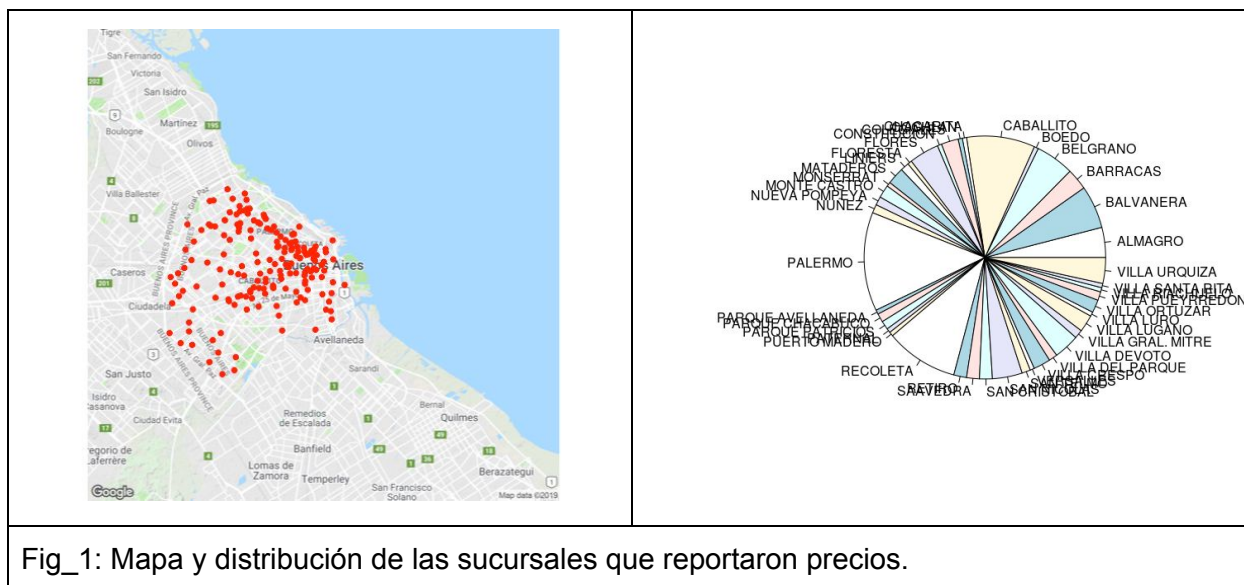
## **Precios CABA**

*Informe TP1 - Datamining*

Ignacio Chiapella, Juan Knebel

*Maestría en Data Mining, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*  
Entregado el 3 de Junio de 2019

<b>1. Abstract</b>	<b>3</b>
<b>2. Introducción</b>	<b>4</b>
<b>3. Datos</b>	<b>5</b>
3.1 Análisis exploratorio	5
<b>4. Resultados</b>	<b>10</b>
4.1 Análisis de la tendencia de los precios	10
4.2 Análisis de datos por barrio	13
4.3 Canasta Básica	17
4.4 Productos navideños	18
4.5 Estudio de variaciones	19
<b>5. Conclusiones</b>	<b>21</b>
5.1 Trabajos futuros	22
<b>6. Referencias bibliográficas</b>	<b>22</b>



En los próximos capítulos presentaremos un análisis previo de los datos obtenidos, los resultados obtenidos de los diferente análisis para terminar con las conclusiones finales.

**Keywords:** *Datos, boxPlot, evolución, inflación, precios, dolar, pesos.*

## 2. Introducción

En el presente trabajo nos disponemos a analizar un grupo de datos con los datos de cierta cantidad de producto en establecimientos de venta de la Ciudad De Buenos Aires. El objetivo de este trabajo será poder extraer información de estos datos para mostrar resultados que confirmen si el precio de los productos acompañó el proceso inflacionario que atraviesa el país en el periodo donde se recolectaron los datos. Es para nosotros de interés, relevar los siguientes ítems:

- Barrios Porteños: Investigar si existe una variación de precios según el barrio para todos los productos informados y también para subconjunto elegidos como representativos.
- Sobre los establecimientos de venta nos interesa conocer si según el tamaño del establecimiento los precios varían y también conocer la evolución de los precios entre las distintas razones sociales.
- Estudiar la evolución general de los precios para ver si en el periodo en que se capturaron los datos, se producen incrementos o decrementos de precios o si hay algún intervalo de tiempo donde haya estabilidad.
- Estudiar la evolución de los productos de mayor, menor y precio promedio tanto en tiempo como también con una granularidad de los barrios donde se comercializan.
- Tomar un subconjunto representativo de la canasta básica y estudiar su variación de precios en el tiempo.
- Elegir algunos subconjunto de datos que nos parezcan interesantes y estudiar su variación de precios en el tiempo.
- Estudiar si la variación de los precios, el valor del Dólar y la inflación guardan alguna relación.

Ya que muchos de los datos que vamos a querer analizar no están en el conjunto de datos iniciales que se nos presentó, es que vamos a tener que consultar otras fuentes de datos y realizar asociaciones entre los mismos para poder ir atacando estos puntos.

La resolución del presente trabajo se realizó utilizando como lenguaje de programación R y como reservorio de datos MongoDB. El código y los script de la base están a disposición de los docentes por si quieren revisarlo.

### 3. Datos

Para comenzar con el análisis de los precios en la Ciudad Autónoma de Buenos Aires se comenzó con un conjunto de datos provisto de crawling realizado entre los meses de Septiembre del 2018 hasta Abril de 2019. Los datos incluyen la información básica de las sucursales que entregan los precios de algunos productos seleccionados, la identificación de los productos informados y los precios de los mismos discriminando la fecha y sucursal que lo informó. Adicionalmente sumamos tres conjuntos de datos. El primero contiene los polígonos de referencia para cada uno de los barrios de la Ciudad de Buenos Aires para poder identificar a qué barrio pertenece cada una de las sucursales. Los otros dos conjuntos de datos contienen información de índole económica para ampliar el análisis realizado. Contamos con el porcentaje de inflación informado por el Indec entre los meses de Septiembre del 2018 hasta Abril de 2019, por último obtuvimos de una casa de cambio los valores día a día de la cotización del dólar.

Los datos provistos de las sucursales no eran suficientes para poder presentar de manera clara un análisis sobre su ubicación, por tal motivo decidimos incluir el barrio al que pertenecen cada una de ellas. Como si contamos con ubicación geográfica, latitud y longitud, y el gobierno de la Ciudad de Buenos Aires ofrece una colección con los barrios y las coordenadas que forman el polígono, pudimos ubicar exactamente los barrios de todas las sucursales. La transformación fue realizada gracias a las funciones de ubicación espacial que provee el motor de base de datos MongoDB.

#### 3.1 Análisis exploratorio

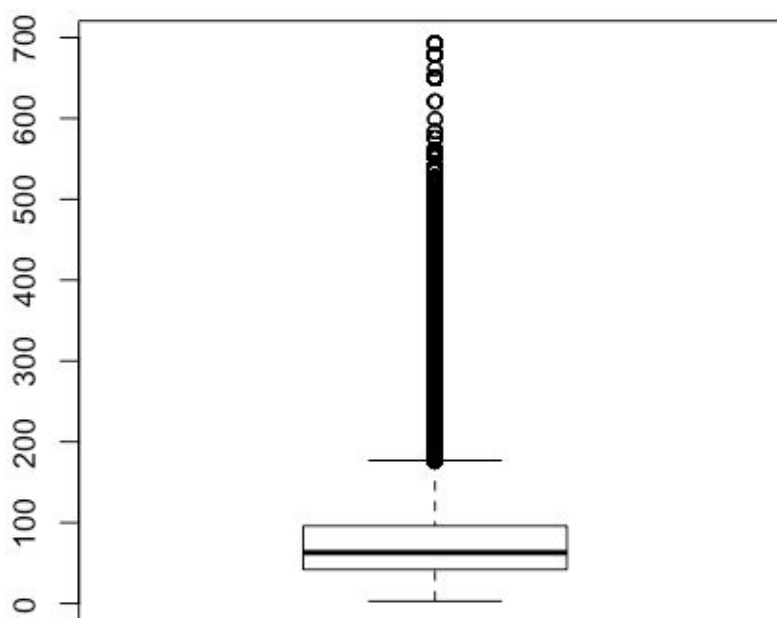
Comenzando con un rápido y breve análisis de los datos pudimos identificar que de las más de 800 sucursales que se encuentran registradas, solo 175, distribuidas en 42 barrios de la Ciudad de Buenos Aires, informaron precios de los 1000 mil productos seleccionados entre el periodo Septiembre 2018 - Abril 2019.

El conjunto de productos no presenta ningún tipo de anomalías ni datos extraños ya que solo se trata de la identificación de los mismos.

En cuanto a los precios que contienen las observaciones que se realizaron sobre los productos se obtiene el primer gran resumen del valor de los productos.

Valor mínimo	2,85
Primer cuartil	42,00
Mediana	62,90
Media	80,79
Tercer cuartil	95,99
Valor máximo	693,00
Rango intercuartil	53,99

En principio el valor máximo y mínimo observado podrían indicar la presencia de valores anómalos o outliers en las mediciones, sobre todo teniendo en cuenta que el valor de mediana es del orden de magnitud diferente a los dos extremos.

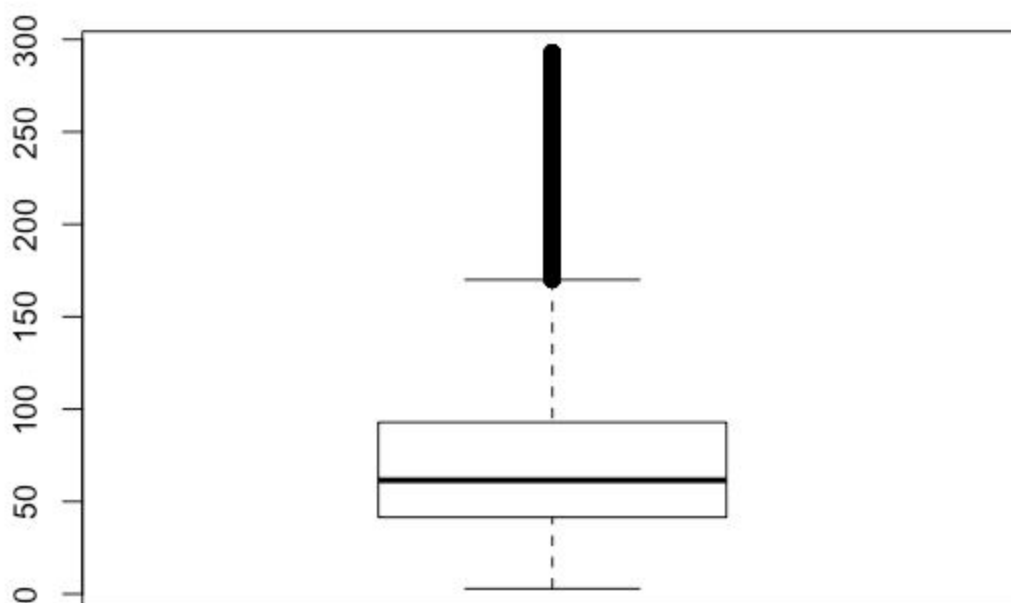


Fig\_2: Boxplot de todas las observaciones de los precios.

Cómo puede observarse en el BoxPlot los datos que podrían considerarse outliers son aquellos precios observados con un valor mayor a \$200,00. Se probó realizar el mismo BoxPlot pero descartando aquellas mediciones que se encuentran 1,5 veces la distancia intercuartil por debajo o por encima. Sólo que en este caso los posibles valores outliers pueden encontrarse en aquellas observaciones de más de \$150,00 y sin ninguna evidencia clara de la ausencia o presencia de datos extraños.

Al tratarse de precios de distintos productos de supermercado en los que efectivamente hay diferencias grandes de valores entre alguno de ellos, se decidió utilizar otra técnica de detección de outliers basada en el desvío estándar.

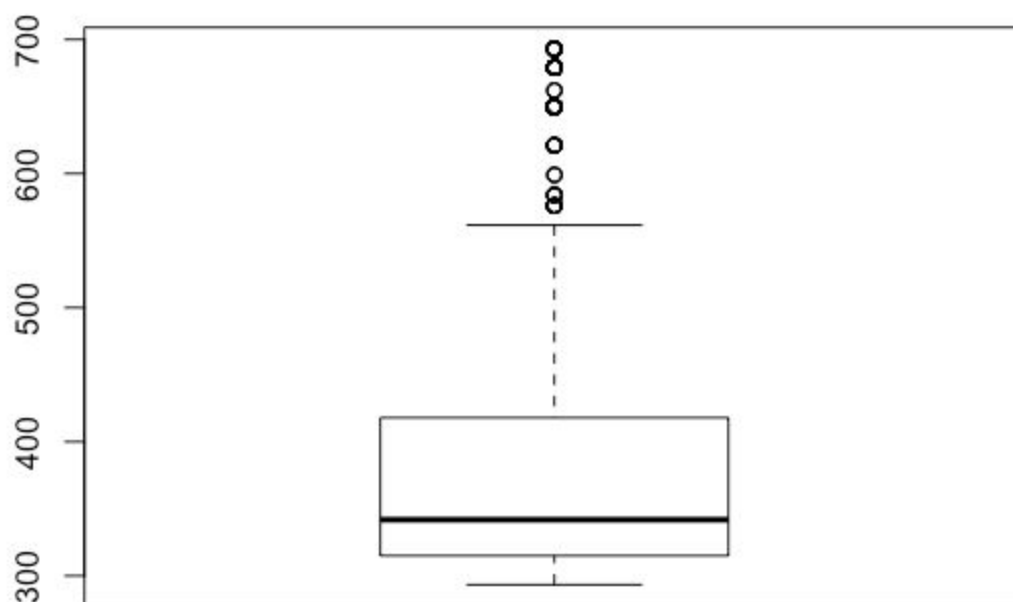
A continuación mostramos el mismo BoxPlot pero solo teniendo en cuenta los valores observados que se encuentran 3 veces alejados del desvío estándar, que si fueran observaciones normales e independientes deberían ser el 99% de los datos.



Fig\_3: Boxplot con las observaciones dentro de los 3 desvíos estándar.

Al igual que antes, se observan muchos datos por encima del bigote superior. Basado en esta información y en que la cantidad de observaciones con valores 3 veces superior a la media más el desvío estándar es superior al 2% del total, podría indicar la existencia de otro grupo de observaciones con valores altos. Se decidió realizar un BoxPlot solo con las observaciones altas y se obtuvo lo siguiente.



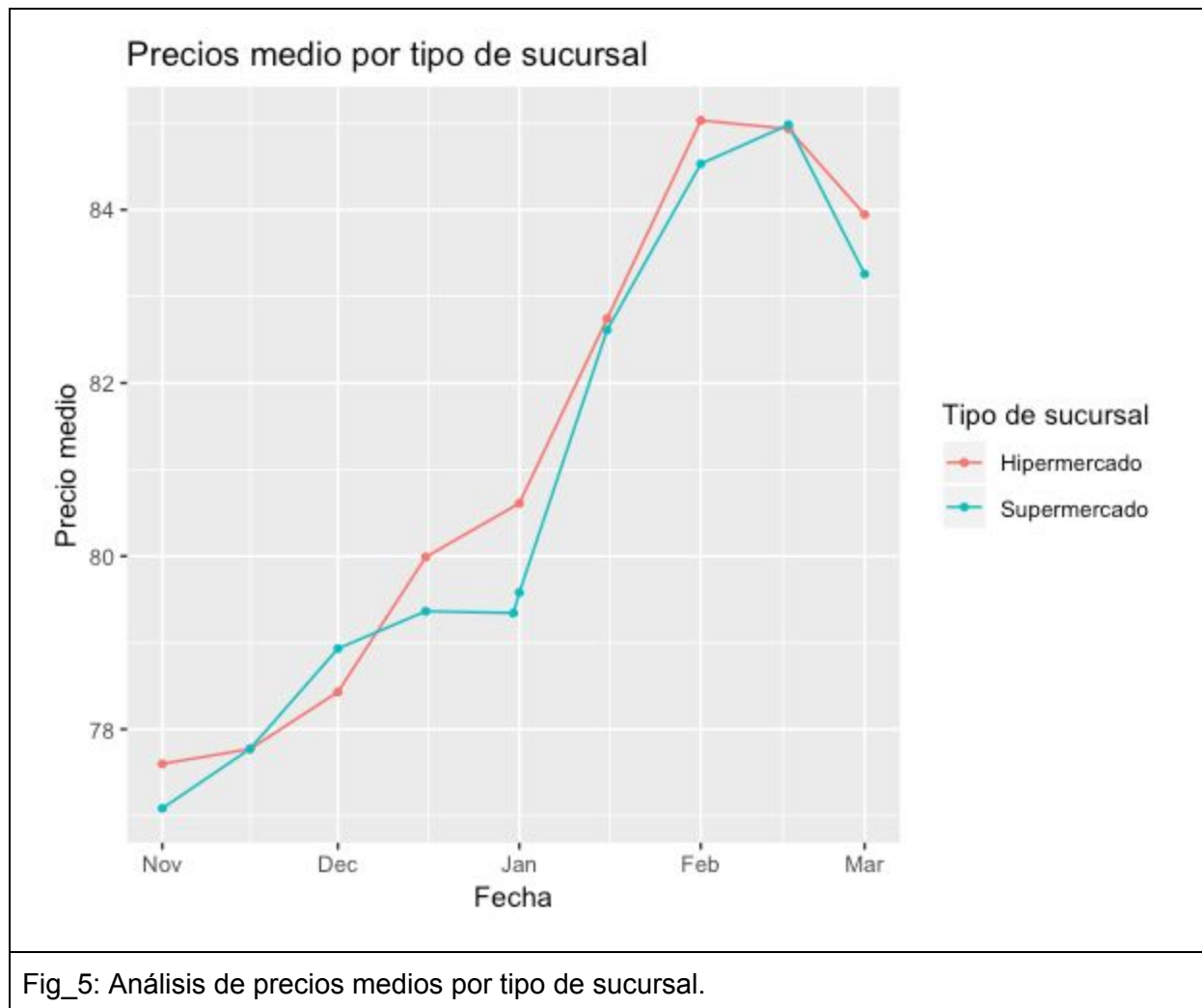


Fig\_4: Boxplot de las observaciones con precios altos.

Que se puede apreciar muy pocas observaciones por encima del bigote superior, que fueron vistas una por una y se trata efectivamente de un producto Whiskey de alto valor que en todas sus apariciones se comportó de la misma manera.

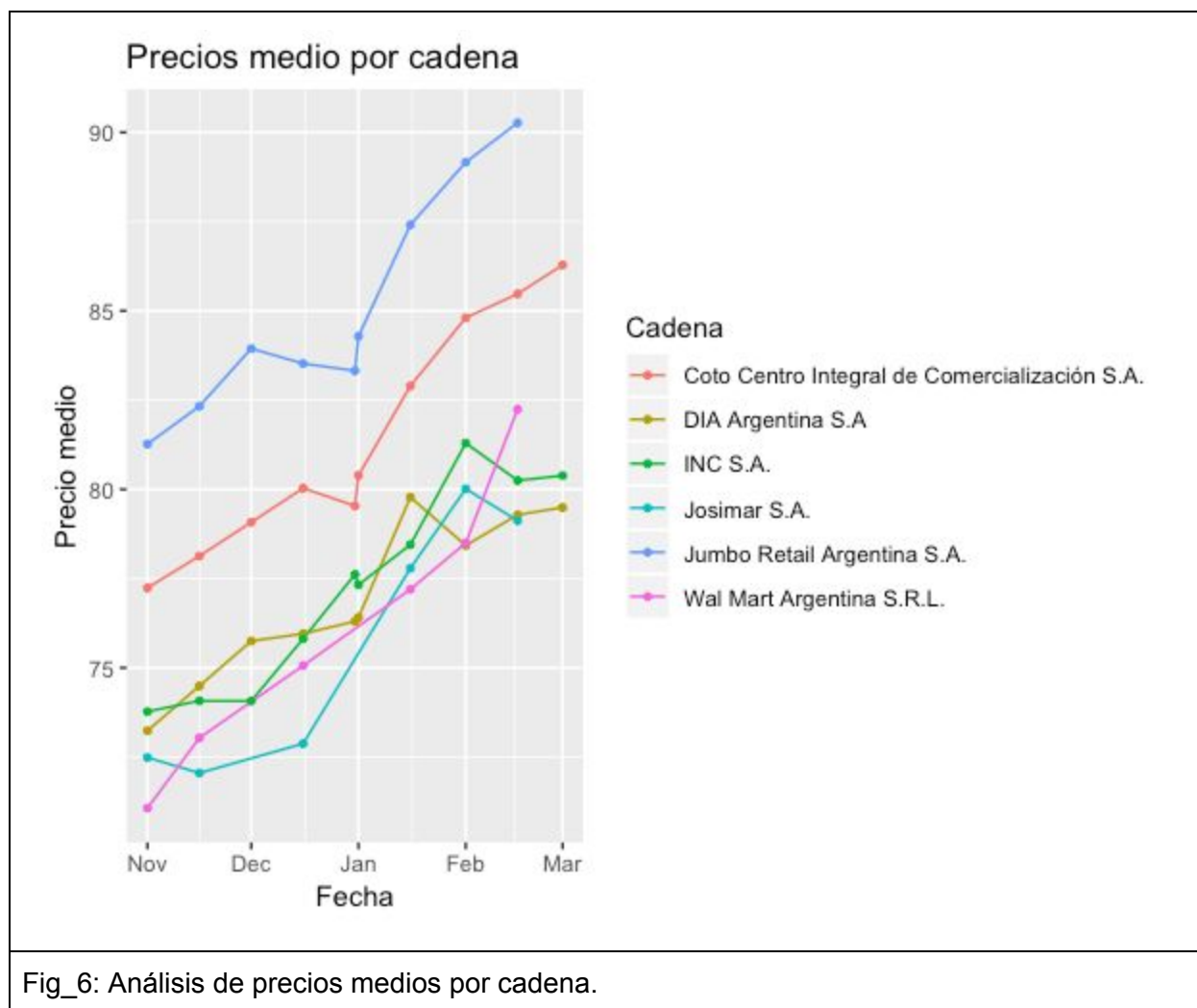
## 4. Resultados

### 4.1 Análisis de la tendencia de los precios



Para identificar si los precios varían entre los diferentes tipos de sucursales de las cadenas de los mercados, agrupamos las mediciones de los productos por tipo de sucursal y por período de 15 días, para finalmente tomar el valor medio de éstos productos. El experimento se repitió con intervalos de 3, 7 y 30 días y la tendencia fue similar en todos, no se observaron diferencias en el comportamiento de la curva.

Al observar que la variación en promedio de los precios no se ve reflejada por el tipo o tamaño de la sucursal, se propuso hacer el mismo análisis pero agrupando por razón social o cadena de supermercados y se obtuvo el siguiente gráfico.

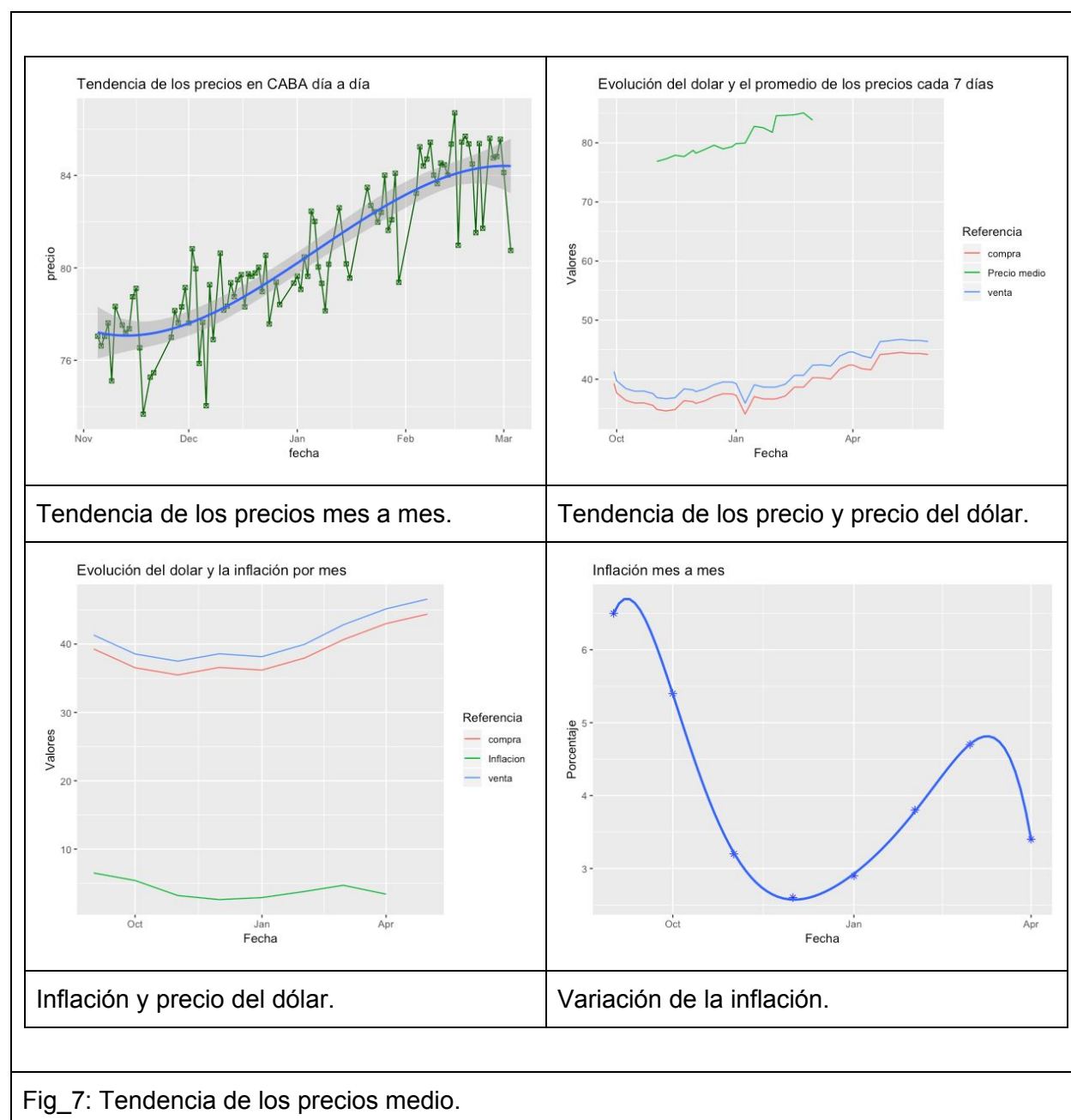


Fig\_6: Análisis de precios medios por cadena.

En este caso se puede ver claramente que las diferentes cadenas de supermercados tienen precios promedios diferentes a lo largo de todas las mediciones. Se destacan Coto y Jumbo como aquellas con mayor valor medio y comportamiento similar. El resto de las cadenas si bien estuvieron por debajo de Coto y Jumbo tuvieron un comportamiento más irregular en la media de los precios, pero igualmente la tendencia de todas es en aumento.

Continuando con el análisis de la tendencia de los precios, presentamos una serie de comparaciones que ayudan a visualizar el comportamiento de los mismos junto con eventos

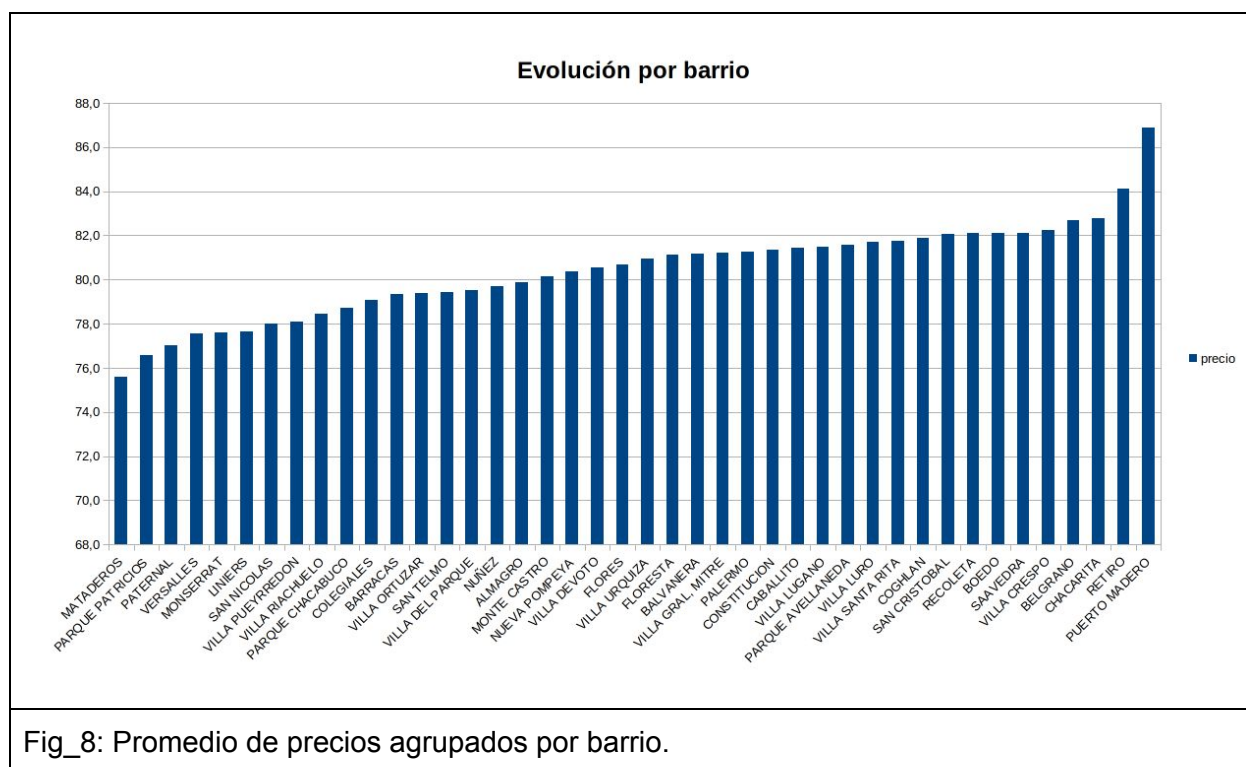
económicos que sucedieron en el mismo periodo de tiempo. Primero se observa una clara tendencia a la alta de los precios, salvo excepciones que se corresponden con que no todos los productos son reportados todos los días. Comparando los precios y la variación del dólar cada 7 días se puede apreciar que ambos tienen la misma tendencia a la alta. En el mes de Enero la divisa baja su valor, y luego de un mes los precios también bajan. Se puede apreciar que la relación entre la cotización del dólar y la inflación no siempre es la misma y en consecuencia no aparenta tener un relación directa en los precios.



Fig\_7: Tendencia de los precios medio.

## 4.2 Análisis de datos por barrio

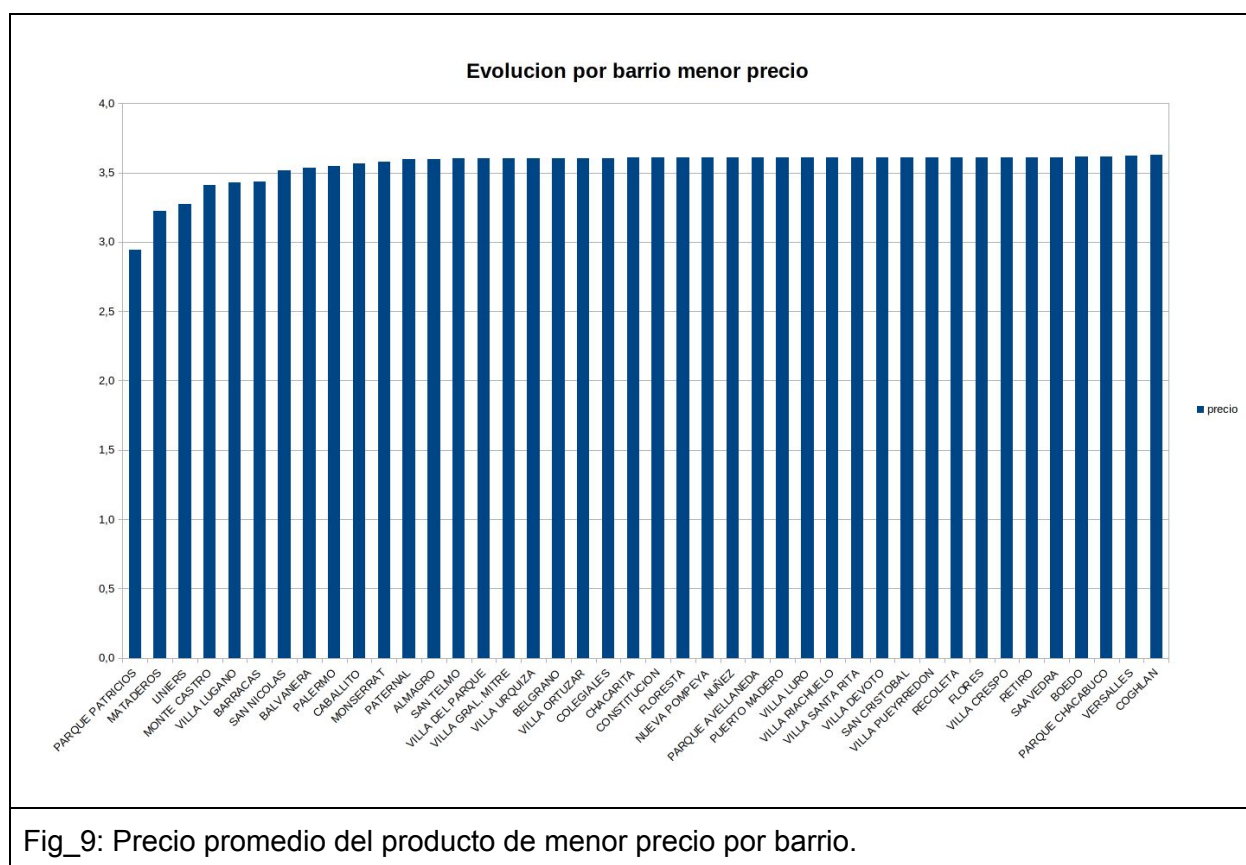
Un campo que nos interesó estudiar fue la variación de precios en los distintos barrios. Un primer análisis rápido que propusimos fue el de estudiar la variación de los precios promedios. Un primer análisis que propusimos fue el de estudiar cuál era el valor del promedio de todos los precios en la totalidad de los datos agrupados por barrio, el resultado que obtuvimos fue el siguiente

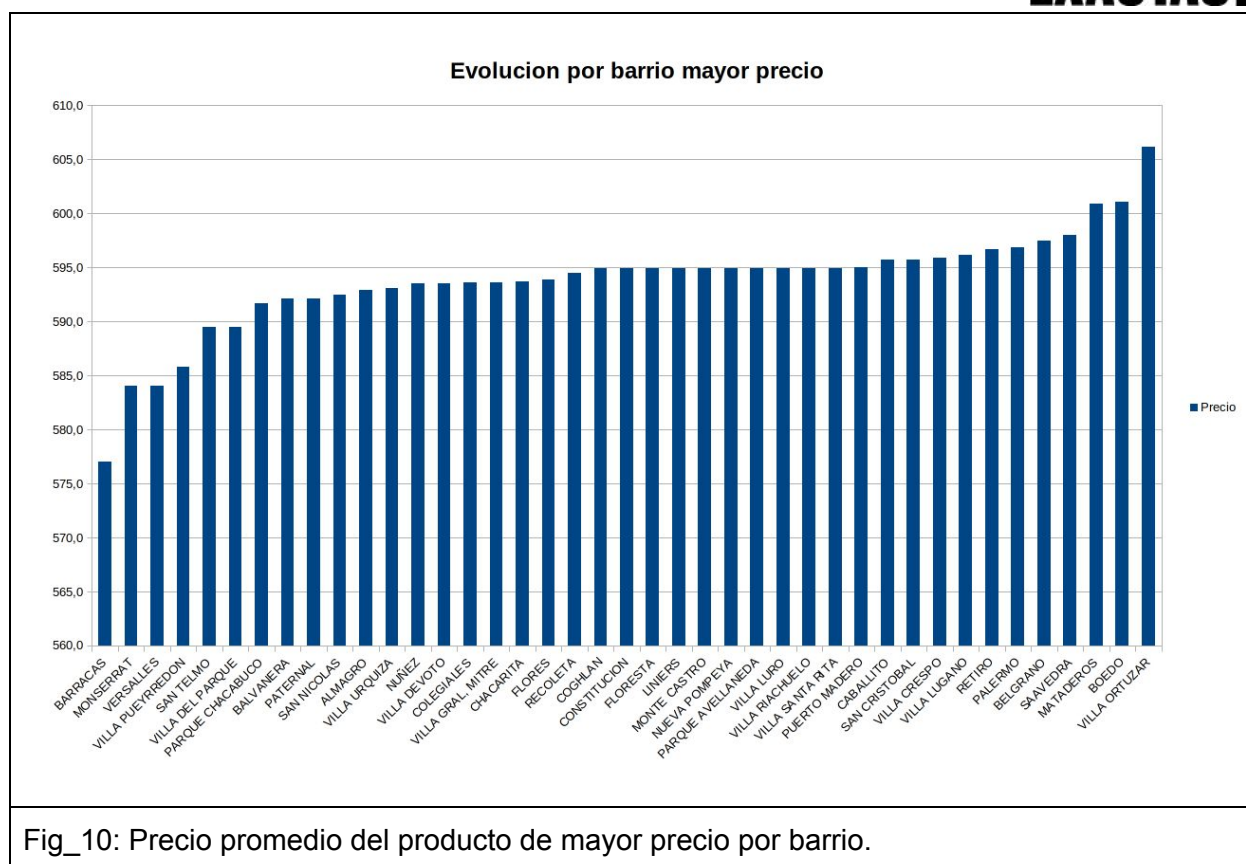


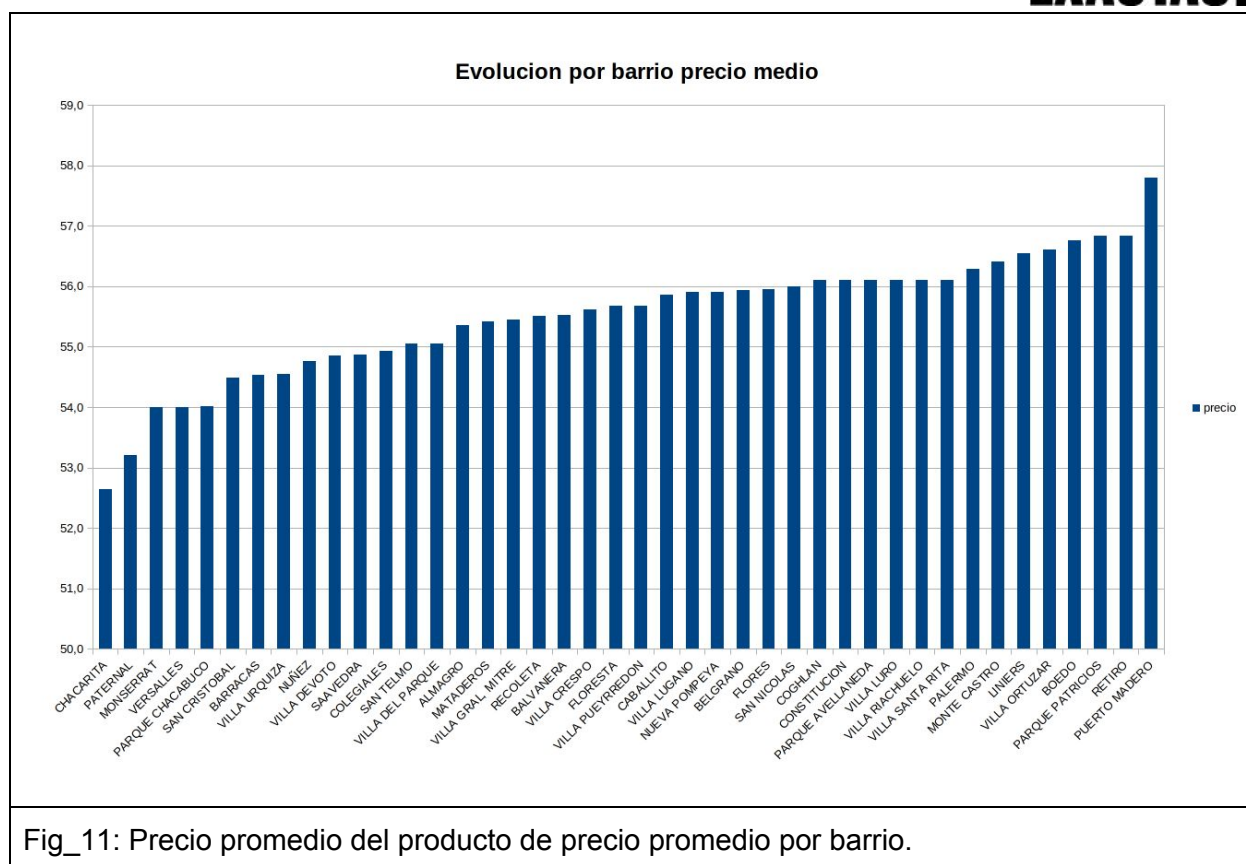
Observamos que si bien hay algunos barrios que muestran un promedio de precios significativamente más alto (como ser el caso de Puerto Madero o Retiro) o mas bajo

(Mataderos y Parque Patricio), el resto de los barrios muestran comportamientos en bloque donde el avance de los precios es bastante suave y una variación punta a punta del orden de los 5 Pesos.

Una vez que realizamos este análisis, nos pareció interesante tomar 3 valores de la muestra de datos y realizar el mismo estudio. Estos son el producto con valor más bajo, valor más alto y el de valor valor medio con mayor cantidad de apariciones, los resultados fueron los siguientes:





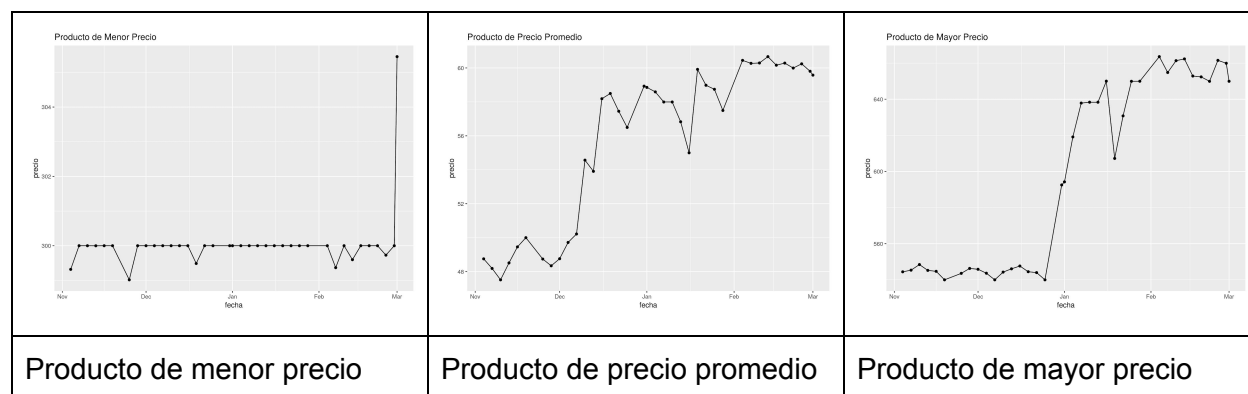


Lo que concluimos con este análisis es que si bien no se mantuvieron fijos los barrios del análisis de todos los productos, para el caso de los precios más altos no se invirtieron sino que mantuvieron una cierta concordancia, por ejemplo Puerto Madero en los 3 casos estudiados se ubicó entre el 50% de los barrios que acumularon el mayor precio. Pero para el caso de los precios menores si encontramos diferencia por ejemplo en el caso del producto de mayor precio Mataderos se encuentra en el top 3 de barrios más caros cuando en la foto general estaba como el más barato.

Un punto interesante es ver que el producto más bajo mostró una variación de precio aprox al 25% pero el precio mayor y el de valor medio mostraron variaciones de menos del 10%.



En el siguiente gráfico exploramos el análisis de cada uno de los valores de estos 3 grupos de productos:



Como se observa el precio del artículo de menor precio se mantiene bastante constante en el tiempo, no así el de precio promedio o el de mayor precio que incrementan su valor constantemente en especial en los meses de Diciembre y Enero.

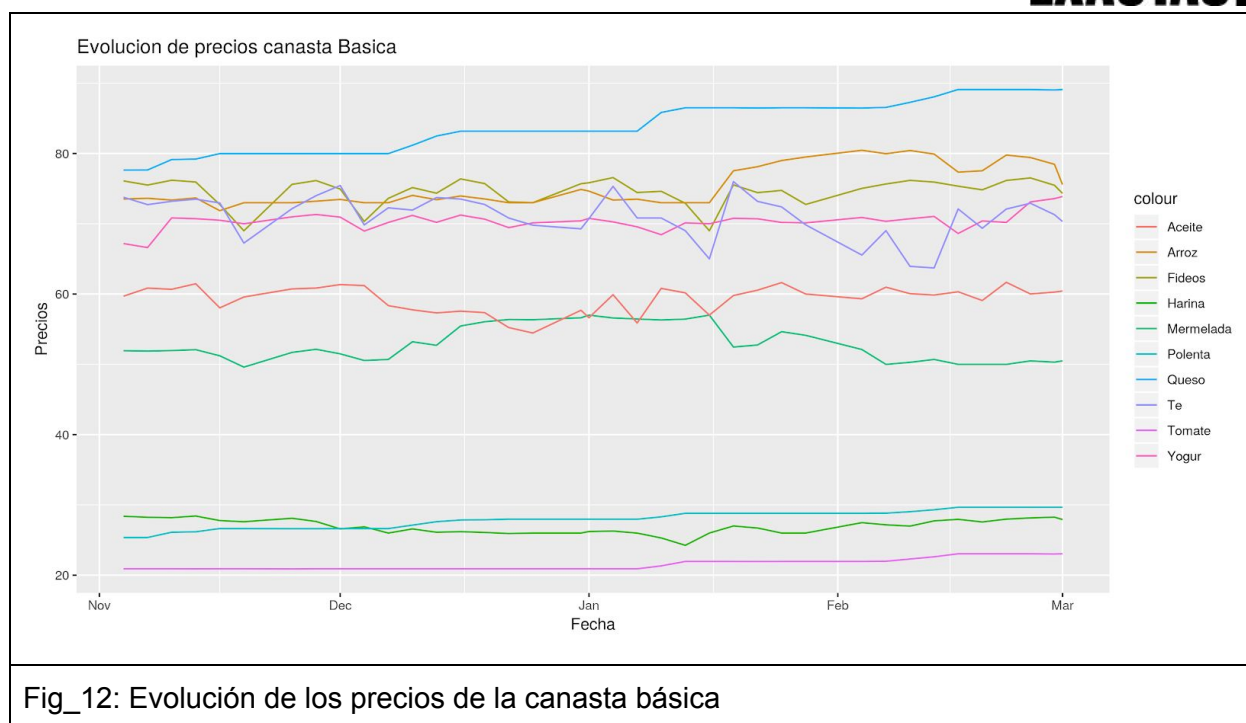
### 4.3 Canasta Básica

Dentro del estudio de los productos, una partición de ellos que nos pareció importante analizar fue el de un grupo de productos que componen la canasta básica en 2018, lo mismo fueron:

**Fideos Tallarin Don Vicente 500 Gr - Arroz Gallo Oro 1 Kg - Harina de Trigo 000 Cañuelas 1 Kg - Polenta Instantanea Magica Quaker 500 Gr - Queso Crema Untable Casancrem Clasico 500 Gr - Tomate Pelado Perita en Lata Arcor 400 Gr - Mermelada de Durazno Frasco Arcor 454 Gr - Aceite de Girasol Cocinero 900 MI - Yogur Bebible Entero Frutilla Yogurissimo 1 Lt - Te en Saquitos Green Hills 50 Un**

Elegimos estos productos de todos los que conforman la canasta básica ya que comparten una franja de precio similar, así al realizar los análisis evitamos que un producto con precio muy alto o muy bajo sesgue la información representada en los gráficos.

Al graficar la evolución de los precios tomando el promedio de precios por semana de cada artículo lo que obtuvimos fue:



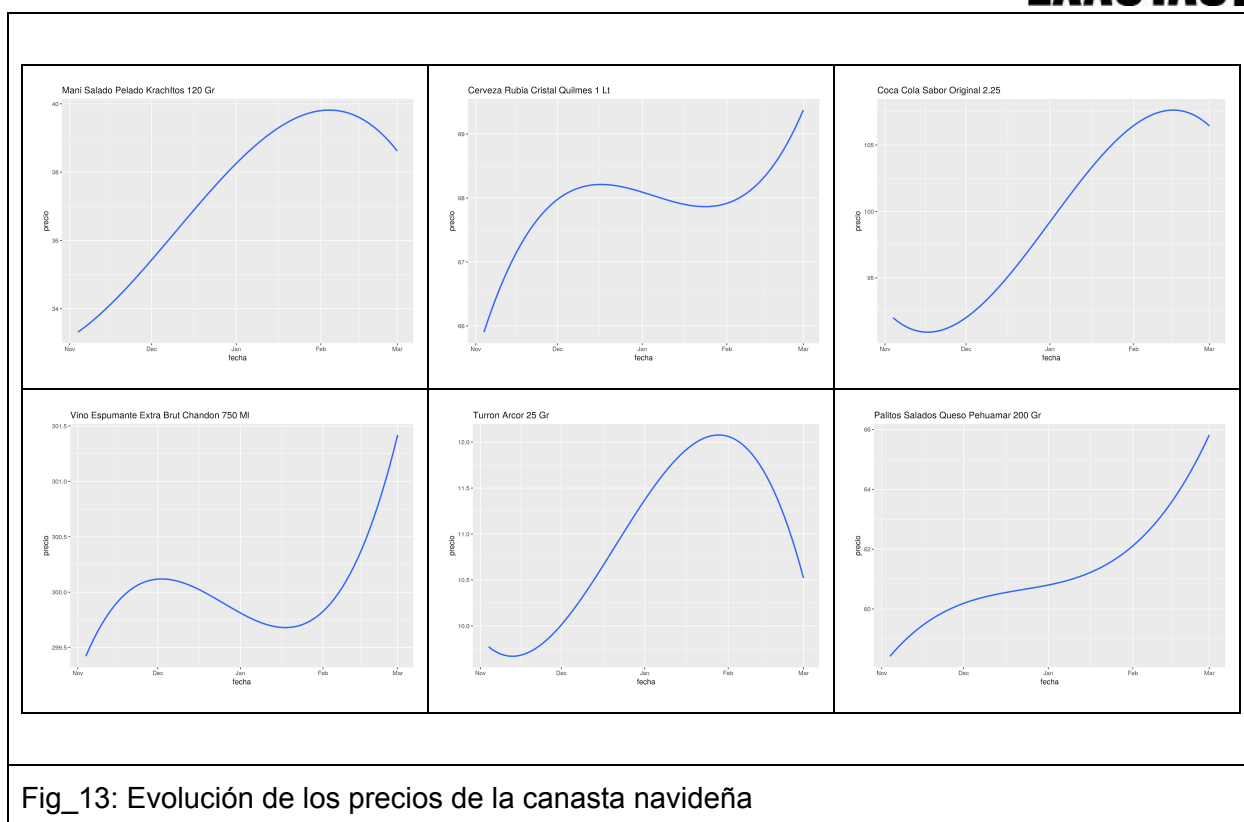
Evidenciamos que en la franja de los 3 precios más bajos los valores se encuentran casi constantes en el periodo analizado, en la franja de valores intermedios si bien hay variación en especial en los meses de Diciembre y Enero el precio final termina muy similar al inicial inclusive en algunos casos menor, como por ejemplo el Té. El producto que sí mostró un incremento sostenido en su valor y que terminó el periodo con un precio significativamente mayor fue el queso.

#### 4.4 Productos navideños

Nos propusimos ver la estacionalidad de algunos productos, fue por ello que decidimos estudiar la evolución de los precios de 6 artículos que se consumen con frecuencia durante las fechas de las fiestas de diciembre, estos son:

**Turron Arcor 25 Gr - Vino Espumante Extra Brut Chandon 750 MI - Coca Cola Sabor Original 2.25 Lt - Cerveza Rubia Cristal Quilmes 1 Lt - Mani Salado Pelado Krachltos 120 Gr - Palitos Salados Queso Pehuamar 200 Gr.**

Al graficar la evolución de los precios tomando el promedio de precios cada 3 días de cada artículo lo que obtuvimos fue un grafico donde lo dispares de los precios de los artículos no permitían estudiar la estacionalidad, fue por ello que graficamos los 5 artículos por separado:



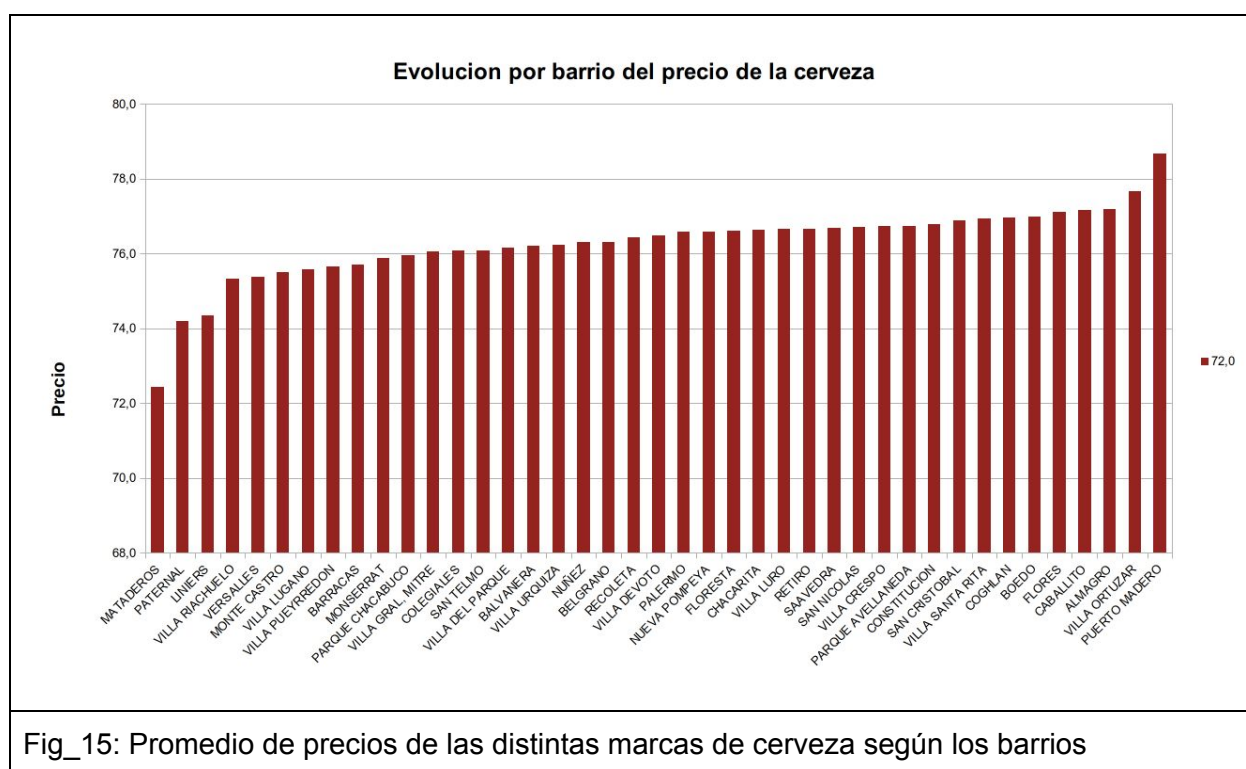
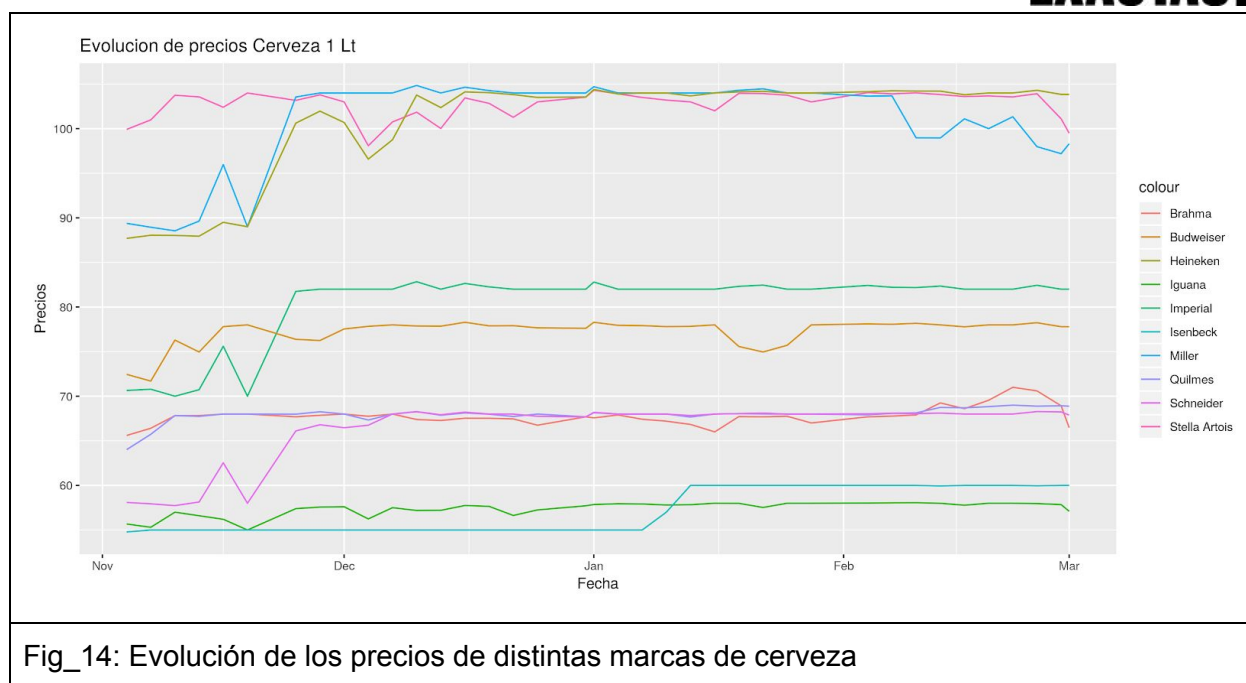
Fig\_13: Evolución de los precios de la canasta navideña

Lo que se observa es que estos productos tienen un aumento fuerte en sus precios comenzando en Diciembre y finalizando a mediados de Enero, donde algunos o bien comienzan a decrecer en sus precios o siguen incrementándose pero más lentamente.

## 4.5 Estudio de variaciones

En esta sección nos propusimos estudiar la variación de la Cerveza en sus distintas marcas y en presentaciones de aprox 1 litro en CABA, la idea es ver en un conjunto cerrado de productos la variación de los precios en el tiempo, el conjunto a estudiar es el siguiente:

**Cerveza Rubia Iguana 970 Cc - Cerveza Rubia Stella Artois 970 Cc - Cerveza Rubia Chopp Brahma 1 Lt - Cerveza Rubia Cristal Quilmes 1 Lt - Cerveza Rubia Lager Imperial 1 Lt - Cerveza Rubia Miller Genuine Draft 1 Lt - Cerveza Rubia Retornable Budweiser 1 Lt - Cerveza Rubia Retornable Schneider 1 Lt - Cerveza Rubia Retornable Heineken 1 Lt - Cerveza Rubia Isenbeck 970 Cc**



Lo que observamos en el estudio de los precios de las distintas marca de cerveza es que tenemos un grupo del 70% de los casos estudiando que están entre 55 y 85 pesos y que dicho precio se sostiene bastante con el tiempo. Lo mismo ocurre con la Stella Artois que se mantiene como una de las de mayor precio, como caso interesantes están Miller y Heineken que en el periodo de tiempo analizado aumentan casi un 20% su valor e igualan el precio y el comportamiento de Stella Artois.

En el comparativo de los precios promedio por barrio, la diferencia entre los extremos de los barrios es del 8%.

## 5. Conclusiones

- Según el análisis que realizamos podemos concluir que el valor de los precios que analizamos tuvo una tendencia de crecimiento en el periodo seleccionado acompañado al comportamiento del Dólar y no tanto a la inflación en sí.
- El recorrido de las distintas pruebas que hicimos sobre los precios en los distintos barrios evidenció que no siempre el barrio que tiene el mayor precio promedio, lo tiene para productos o conjunto de productos en particular e inclusive esta tendencia en el promedio puede ser revertida para algunos productos en particular.
- Conclusión canasta básica, canasta navideña y precios extemo.
- Si bien entre supermercados e hipermercados no encontramos una diferencia de precios significativas si los encontramos entre las distintas razones sociales, claramente el precio promedio de los productos de la cadena Jumbo y Coto son superiores a los del resto y estos dos tienen una evolución de sus precios muy parecida. Este es un caso muy interesante de analizar en un trabajo a futuro ya que si bien las cadenas “más caras” también son las que más promociones bancarias tienen a diferencia por ejemplo de la cadena Día.
- Nos pareció muy interesante poder realizar un análisis de la evolución de los precios con datos reales, en un comienzo no teníamos idea de qué resultados podríamos encontrar o qué métricas podríamos extraer.

- Todo el análisis que se realizó fue sobre los precios de lista que informan los supermercados dejando de lado las promociones que ofrecen no solo de los bancos como mencionamos anteriormente sino las ya conocidas 2x1, 3x2, 70 u 80% de descuento en la segunda unidad o descuentos en la próxima compra que uno realiza. Tener en cuenta todas estas variables y sobre todo cambian día a día hace realmente muy difícil un análisis más exacta de precios.

## 5.1 Trabajos futuros

- Como pendiente nos quedó haber estudiado no solo el comportamiento del producto más barato o del más caro, sino el comportamiento del grupo del 5%, 10%, 15% de los precios más baratos y los 5%, 10%, 15% de los datos más caros a modo de poder suavizar cualquier singularidad que presente un solo producto en un intervalo dado.
- Nos faltó poder analizar en los conjuntos que definimos (canasta básica, canasta navideña y cervezas) cuál era la variación de los precios en los supermercados vs hipermercados y también entre las distintas razones sociales de los establecimientos para entender si había alguna relación interesante.
- Planteamos en un momento revisar si en momentos de promociones bancarias se evidenciaba algún cambio en los precios la semana anterior o posterior a la promoción.

## 6. Referencias bibliográficas

Cotización del dólar día a día: <http://www.pullman.com.ar/es/historico-dolar/>

Inflación mes a mes: <https://www.indec.gob.ar/>

Coordenadas de los barrios de Buenos Aires: <https://data.buenosaires.gob.ar/dataset/barrios>

Análisis de Canasta Básica:

<https://www.argentina.gob.ar/sites/default/files/cbayt-gba-mar18.pdf>