

# **UNIVERSIDAD CATÓLICA BOLIVIANA**

## **“SAN PABLO”**



**Inflación Digital en Bolivia:**

**Análisis Automatizado de Precios Online**

**Proyecto Final de Minería de Datos II**

**Estudiante:**

NATALY ROSARIO HUANCA CHOQUE

**Docente:**

ALVARO LIMBER CHIRINO GUTIERREZ

**Materia:**

MINERIA DE DATOS II

## **Abstract**

El presente estudio analiza el comportamiento del costo de vida digital en Bolivia mediante el monitoreo automatizado de precios en plataformas online. A través de técnicas de minería de datos como scraping, regresión, clustering y predicción, se recolectaron y analizaron más de 13.000 observaciones de productos publicados en el sitio web del supermercado Fidalga. El objetivo fue estimar un índice de inflación digital que refleje mejor la realidad del consumo en línea frente al Índice de Precios al Consumidor (IPC) tradicional. Los resultados muestran diferencias significativas de precios entre categorías y permiten visualizar patrones de variación mediante modelos estadísticos y dashboards. Este enfoque plantea una alternativa moderna y accesible para complementar los indicadores oficiales del

**Palabras clave:** inflación digital, minería de datos, scraping, costo de vida, precios online, Bolivia.

**Código JEL:** C88, E31, O33

## **1. Introducción**

En los últimos años, Bolivia, especialmente tras la pandemia del COVID-19, ha experimentado un crecimiento constante en el uso de plataformas digitales para el comercio, el consumo y la comparación de precios. Cada vez más personas realizan sus compras en supermercados online, farmacias virtuales y aplicaciones de delivery, donde los precios pueden variar frecuentemente entre tiendas y a lo largo del tiempo. Sin embargo, el Índice de Precios al Consumidor (IPC), elaborado por el Instituto Nacional de Estadística (INE), continúa siendo la principal referencia para medir la inflación, aunque se basa en una canasta fija y en métodos tradicionales de recolección.

Esta situación genera una brecha entre la realidad de los precios digitales y los indicadores oficiales, que al no capturar la dinámica del entorno virtual, pueden ofrecer una visión parcial del costo de vida. La evolución tecnológica exige el desarrollo de nuevos métodos que permitan monitorear, de forma automatizada, los precios presentes en las plataformas digitales más utilizadas por los consumidores.

De esta forma, el presente estudio propone una solución basada en minería de datos, mediante la construcción de un sistema automatizado que permita estimar el costo de vida digital en Bolivia. Este sistema se apoyará en técnicas como scraping, regresión, clustering y predicción para recolectar y analizar precios provenientes de tiendas online bolivianas. Finalmente, se busca estimar un índice de inflación digital alternativo y compararlo con el IPC oficial, complementando el análisis con modelos estadísticos y visualizaciones.

## **2. Objetivos**

### **Objetivo general**

Desarrollar un sistema de monitoreo automatizado del costo de vida en Bolivia mediante técnicas de minería de datos aplicadas a precios recolectados de tiendas online, con el fin de estimar un índice de inflación digital que refleje de manera más dinámica y actualizada el comportamiento de consumo en plataformas digitales. Este sistema integrara temas en la asignatura como: scraping, regresión, clustering, predicción,

permitiendo aplicar análisis exploratorio, agrupamiento de productos y modelos de proyección entre los precios, finalmente obteniendo un análisis económico.

### **Objetivos específicos**

- Identificar y seleccionar tiendas online bolivianas con catálogos públicos accesibles para el scraping de productos.
- Clasificar los productos recolectados en categorías clave para facilitar el análisis por tipo de consumo.
- Diseñar un sistema de recolección automatizada de precios utilizando herramientas de scraping en R.
- Visualizar los resultados con gráficos y dashboards interactivos.
- Aplicar modelos de regresión para analizar cómo cambian los precios en el tiempo.
- Evaluar la precisión de los modelos predictivos implementados para proyecciones de precios.
- Comparar el comportamiento de precios entre diferentes categorías de productos como alimentos, higiene entre otros.

### **3. Motivación**

Recientemente Bolivia está sufriendo una inflación en la canasta familiar donde productos alimenticios, transporte entre otros están subiendo considerablemente su precio. Lo que me llevo a elegir este tema, por la situación actual de mi país. Buscando soluciones elegí este sistema automatizado de inflación digital para obtener un análisis económico. Desde la era digital, ahora se puede programar, estimar, modelar y predecir lo que pasara en el largo plazo, y así se llama mi tema, **inflación digital en Bolivia: análisis automatizado en precios online.**

En el contexto económico y tecnológico de Bolivia está cambiando rápidamente. Cada vez más hogares acceden a plataformas digitales para realizar sus compras, informarse sobre precios y tomar decisiones de consumo. Este nuevo comportamiento ha generado una fuente masiva de datos que, si se analizan correctamente, pueden ofrecer una visión actualizada del costo de vida digital y su evolución semana a semana.

A pesar de este avance, los instrumentos estadísticos tradicionales como el IPC no capturan estas dinámicas digitales. Los datos oficiales siguen dependiendo de visitas presenciales y estructuras de precios fijas, lo cual limita su capacidad para reflejar los cambios rápidos que ocurren en el entorno online.

La minería de datos, en particular el scraping y el análisis estadístico automatizado, permite construir soluciones ágiles y replicables que pueden complementar la visión macroeconómica. Este enfoque no solo es técnicamente viable, sino también estratégico para actualizar la forma en que se analiza el costo de vida. La estructura del proyecto nos permite esquematizar el scraping, regresión, predicción y clustering, como la visualización de datos de forma practica y detallada.

Además, la elección de este proyecto responde a una necesidad de aplicar herramientas tecnológicas con propósito social, que puedan generar evidencia útil para la ciudadanía, investigadores y tomadores de decisiones. Poder observar cómo cambian los precios semana a semana desde plataformas virtuales ofrece no solo una ventaja informativa, sino también una forma más transparente y cercana de entender el impacto económico real en los hogares bolivianos. Por otro lado, el uso de técnicas de minería de datos como scraping, clustering y predicción permite automatizar procesos que antes eran manuales, lentos y costosos.

## **4. Marco teórico**

### **4.1 Antecedentes**

Para el proyecto se presentará diferentes tipos de tienda con la finalidad de crear una base de datos sólida para la comparación de precios digitales, a su vez desarrollar la comparación con la IPC. Además, se seleccionaron tiendas virtuales bolivianas como Hipermaxi, Fidalga y Farmacorp, por su disponibilidad de catálogos públicos, representación de productos clave en la canasta. A continuación, se detalla cada tienda con sus puntos mas importantes:

- **Hipermaxi**

Es una de las cadenas de supermercados más emblemáticas de Bolivia. Fue fundada en 1994 en la ciudad de Santa Cruz de la Sierra por el empresario boliviano Pablo Bedoya, con la visión de crear una red de supermercados con enfoque nacional y precios competitivos. Con el paso de los años, Hipermaxi ha logrado expandirse a otras ciudades como La Paz, Cochabamba, Tarija y Sucre, consolidándose como una de las marcas líderes del sector minorista en Bolivia.

La tienda online de Hipermaxi surge como una respuesta a las nuevas demandas del consumidor digital, ofreciendo un catálogo con cientos de productos organizados en categorías como alimentos, limpieza, higiene y bebidas. Su relevancia en el mercado boliviano se mantiene gracias a su política de abastecimiento constante y precios accesibles. Su sitio web bien estructurado permite realizar scraping eficiente, convirtiéndolo en una fuente clave para monitorear precios y analizar el comportamiento del costo de vida digital.

Para este proyecto, Hipermaxi representa una fuente confiable de datos digitales y su cobertura de productos de consumo masivo. El acceso a precios directamente desde su página permite realizar análisis semanales sobre la evolución del costo de vida en la canasta básica, complementando con precisión el monitoreo de la inflación digital en Bolivia.

- **Fidalga**

Es una cadena de supermercados boliviana con fuerte presencia en Santa Cruz de la Sierra y La Paz, consolidada como una de las marcas más tradicionales y reconocidas del país. Fundada en 1959 por la familia Fidalga, comenzó como una tienda de abarrotes y con el tiempo evolucionó hasta convertirse en un referente del retail moderno en Bolivia. Su crecimiento se ha caracterizado por la oferta de productos variados, atención al cliente cercana y precios competitivos.

El sitio web de Fidalga representa un paso firme hacia la transformación digital del sector minorista. A través de su tienda online, la empresa permite acceder a un amplio catálogo que abarca desde productos alimenticios hasta artículos de higiene, limpieza, bebidas y más. Cada producto está acompañado de su precio actualizado, lo que convierte a la plataforma en un recurso valioso para la recolección de datos en tiempo real.

En este proyecto, Fidalga se incorpora como una segunda fuente de datos para contrastar precios y monitorear el comportamiento del mercado. Su página web, al igual que la de Hipermaxi, está estructurada de forma que permite aplicar técnicas de scraping automatizado. Gracias a ello, es posible comparar productos de la canasta básica y observar variaciones de precio entre distintas cadenas, fortaleciendo así el análisis sobre el costo de vida digital en Bolivia y ofreciendo una visión más completa de la dinámica de precios del mercado local.

- **Farmacorp**

Fue fundada en 1939 por el Dr. Mario Suárez Riglos en la ciudad de Santa Cruz de la Sierra, con la apertura de una pequeña farmacia que con el tiempo se transformaría en la cadena farmacéutica más grande de Bolivia. Con más de 200 sucursales en todo el país, Farmacorp se ha posicionado como líder en el sector farmacéutico y de cuidado personal, combinando una fuerte presencia física con una moderna tienda virtual.

En su versión online, Farmacorp ofrece un catálogo con miles de productos que incluyen medicamentos, artículos de higiene, productos para bebé, cosméticos y suplementos. Su página está organizada de manera clara y eficiente, lo cual facilita su uso como fuente para scraping. Incorporar Farmacorp al análisis permite ampliar el estudio hacia productos esenciales en salud y cuidado diario, generando una mirada más integral del costo de vida en entornos digitales.

La integración de Farmacorp en este estudio permite incluir una categoría clave dentro del análisis del costo de vida: el cuidado de la salud y el bienestar personal. Al contar con un catálogo en línea bien estructurado, es posible aplicar técnicas de minería de datos para analizar patrones de precios en productos de uso cotidiano.

## **4.2 Bases teóricas**

A continuación, se detallan los temas involucrados en el desarrollo del presente proyecto, con sus respectivas definiciones, características, funciones y aplicación específica dentro del monitoreo automatizado del costo de vida mediante precios online en Bolivia.

#### 4.2.1 Raspado Web (Web Scraping)

El raspado web o "web scraping" es una técnica automatizada utilizada para extraer información estructurada desde sitios web. A través del uso de scripts o API's, se navega por las páginas web, se identifican patrones de contenido relevante, y se recopila información específica como precios, nombres de productos, fechas, categorías, etc. En el contexto de este proyecto, el scraping permite obtener precios actualizados de supermercados y farmacias virtuales bolivianas, conformando una base de datos automatizada y replicable.

Sus principales características incluyen:

- **Automatización:** Permite recolectar datos sin intervención manual frecuente.
- **Eficiencia:** Acelera la construcción de bases de datos con miles de registros en poco tiempo.
- **Escalabilidad:** Puede adaptarse a distintos sitios web y estructuras.
- **Integración:** Los datos pueden combinarse con modelos de análisis en R u otras plataformas.

Entre sus funciones se destacan:

- Recolección de datos para vigilancia de precios.
- Seguimiento de cambios de productos.
- Creación de indicadores económicos alternativos.
- Monitoreo semanal o diario de tendencias del mercado digital.

#### 4.2.2 Regresión

La regresión es una técnica estadística que permite modelar la relación entre una variable dependiente (en este caso, el precio) y una o varias variables independientes (semana, tienda, categoría, etc.). El objetivo es predecir el comportamiento futuro de la variable dependiente o estimar su valor bajo ciertas condiciones. En el presente proyecto, se emplean modelos de regresión lineal para analizar la evolución del precio de productos a lo largo del tiempo.

Sus características incluyen:

- **Interpretabilidad:** Permite conocer qué variables influyen más sobre el precio.
- **Aplicabilidad:** Es adecuada para datos numéricos y continuos.
- **Facilidad de implementación:** Requiere pocos supuestos para modelar tendencias temporales.

Funciones clave en el proyecto:

- Modelar tendencias de precios por producto y categoría.
- Identificar estacionalidades o patrones en la evolución semanal.
- Comparar precios entre diferentes tiendas a través del tiempo.

#### 4.2.3 Predicción

La predicción consiste en anticipar el valor de una variable futura basándose en datos históricos. En este proyecto, se utiliza para estimar precios futuros de productos a partir de los datos recolectados por scraping. Los modelos predictivos pueden apoyarse en regresiones, redes neuronales o técnicas de series de tiempo.

Características:

- **Basada en datos reales:** Utiliza información previa para proyectar comportamientos.
- **Ajustable:** Permite recalibrar modelos en cada iteración.
- **Orientada a decisiones:** Apoya la toma de decisiones económicas basadas en proyecciones.

Algunas de sus funciones son:

- Estimación de variables clave en el tiempo.
- Construcción de escenarios futuros.
- Apoyo al análisis económico, social o técnico.

#### 4.2.4 Clustering o agrupamiento

El clustering es una técnica de minería de datos no supervisada que permite agrupar observaciones similares dentro de conjuntos de datos. En este proyecto, se aplica para identificar productos que presentan comportamientos de precio similares, como estabilidad, aumentos constantes o alta volatilidad.

Características:

- **Segmentación automática:** Separa los datos en grupos sin necesidad de etiquetas previas.
- **Exploratorio:** Ayuda a descubrir patrones ocultos en grandes volúmenes de datos.
- **Visualizable:** Permite representar los grupos en gráficos de dispersión o dendrogramas.

#### 4.2.5 Minería de texto

Sinónimo de análisis de texto el uso de minería de texto describe de forma adecuada el descubrimiento de ideas (KDD) y el uso de algoritmos específicos. Es el proceso de

extraer información y patrones de grandes volúmenes de texto no estructurado, análisis estadístico y aprendizaje automático. Su objetivo es descubrir conocimientos ocultos, tendencias y relaciones en los datos textuales que no son evidentes a simple vista. Sus características para la minería de texto son:

- La confianza que se genera entre las partes interesadas para extraer información.
- Las metodologías se pueden aplicar rápidamente permitiendo métodos auditables y repetibles.
- Extracción de conocimiento donde identifica patrones, tendencias y relaciones en los datos textuales que pueden ser utilizados para la toma de decisiones.
- Identifica las nuevas ideas o refuerza la información relevante utilizando técnicas de procesamiento de lenguaje natural (NLP), para comprender y analizar los resultados.

La minería de texto permite analizar grandes volúmenes de información textual. Entre sus funciones están la clasificación de textos (asignar etiquetas como deportes o política), la detección de temas (agrupar documentos con tópicos similares) y el análisis de tendencias (ver cómo cambian los temas con el tiempo). Uno de los métodos más comunes es la bolsa de palabras (bag-of-words), que representa textos como conjuntos de palabras sin importar el orden. Se usa mediante matrices como la DTM (documentos vs. palabras) o TDM (palabras vs. documentos).

También se usa el análisis sintáctico, que estudia la estructura gramatical y requiere pasos como limpiar el texto y armar matrices. Por último, el análisis de sentimientos busca identificar emociones en los textos, usando diccionarios como el de Plutchik, que incluye sentimientos positivos (alegría, confianza, etc.) y negativos (miedo, tristeza, etc.).

## 5. Revisión de literatura

A continuación, se presenta una revisión teórica complementaria al proyecto dentro del panorama económico boliviano y las herramientas de minería de datos utilizadas, incluyendo sus aplicaciones específicas en el presente estudio.

**El Índice de Precios al Consumidor (IPC): definición y limitaciones** El IPC es un indicador estadístico que mide la variación promedio de los precios de un conjunto de bienes y servicios representativos del consumo de los hogares. En Bolivia, es calculado mensualmente por el Instituto Nacional de Estadística (INE). Si bien es una herramienta fundamental para el análisis macroeconómico, su estructura fija y la recolección tradicional limitan su capacidad para captar los cambios dinámicos del mercado digital, como las ofertas online, la competencia entre tiendas virtuales y la volatilidad semanal de precios.

**Inflación digital: concepto y necesidad de nuevos indicadores** La inflación digital se refiere a la variación de precios en el entorno digital, considerando plataformas de comercio electrónico y catálogos online. Debido al aumento de las compras virtuales, se vuelve fundamental contar con un indicador alternativo que refleje el



comportamiento real de los precios en la economía digital. Este indicador puede complementar el IPC oficial y servir como herramienta de alerta para cambios bruscos en el poder adquisitivo del consumidor digital boliviano.

**Minería de datos:** la minería de datos es un proceso que permite descubrir patrones, tendencias y relaciones ocultas dentro de grandes volúmenes de datos. En el ámbito económico, su aplicación permite automatizar el análisis de precios, segmentar consumidores, predecir tendencias y construir modelos de decisión. En el presente proyecto, la minería de datos es el eje metodológico central para integrar scraping, análisis predictivo y visualización de precios en tiendas online bolivianas.

**Scraping como técnica de recolección automatizada de datos.** El proyecto implementará la técnica de scraping para recolectar precios semanalmente desde sitios web de supermercados y farmacias bolivianas. Se automatizará la extracción de datos como nombres de productos, precios y categorías, generando una base de datos actualizada y estructurada para su posterior análisis.

**Modelos de regresión para análisis de precios.** Se utilizarán modelos de regresión estadística, particularmente regresión lineal, para analizar la evolución de los precios en el tiempo. Esto permitirá identificar patrones de alza o estabilidad, y evaluar las diferencias de comportamiento según tienda o categoría de producto.

**Clustering:** agrupación de productos por comportamiento de precios. El proyecto aplicará técnicas de clustering (agrupamiento no supervisado) para identificar grupos de productos con comportamientos de precio similares. Esto ayudará a segmentar los datos en clústeres como productos estables, volátiles o en tendencia alcista.

**Predicción de series de precios en entornos digitales.** Se realizará una predicción de precios semanales usando modelos estadísticos aplicados a series temporales. Esta proyección permitirá estimar la inflación digital en las semanas siguientes y generar escenarios de evolución económica.

### **Estudios previos y experiencias similares**

Existen investigaciones recientes que han utilizado scraping y minería de datos para monitorear precios online, especialmente en países como México, Chile y Argentina. Por ejemplo, Cavallo (2017) analizó la similitud entre precios digitales y físicos en grandes cadenas de retail, utilizando scraping como herramienta base para construir comparaciones de inflación. Asimismo, un informe del Banco Interamericano de Desarrollo (2021) destaca el uso de big data y técnicas de recolección automatizada de precios para complementar los indicadores oficiales en contextos latinoamericanos. En Bolivia, estas metodologías aún son poco exploradas, lo que refuerza la relevancia e innovación del presente trabajo.

## **6. Descripción del dataset**

Para el desarrollo del presente proyecto, se construyó un conjunto de datos a partir de la recolección automatizada (scraping) de precios digitales disponibles en la tienda

online del supermercado Fidalga, una de las cadenas más tradicionales y reconocidas de Bolivia. Inicialmente, también se evaluaron las plataformas de Hipermaxi y Farmacorp, sin embargo, las limitaciones de sus precios, tiene su respectiva estructura para crear la base de datos, de las tres tiendas. En este caso, se logro crear la base de datos para Fidalga.

La base de datos final está compuesta exclusivamente por productos obtenidos desde Fidalga, organizados en categorías como alimentos, bebidas, limpieza, higiene personal y productos para el hogar. El scraping se realizó de forma manual y estructurada, consolidando más de 13.000 observaciones con variables como: nombre del producto, categoría funcional, precio en bolivianos, y fecha de recolección. Cada registro representa el precio digital de un producto específico disponible en la tienda online de Fidalga en una fecha determinada.

El dataset resultante es lo suficientemente robusto como para permitir análisis exploratorios, construcción de modelos de regresión, segmentación mediante clustering y estimación de inflación digital. La riqueza de productos y la frecuencia de recolección de datos hacen posible observar variaciones reales del costo de vida digital, especialmente en una muestra significativa de productos básicos de consumo boliviano.

## **7. Metodología**

El enfoque metodológico de este proyecto combina técnicas de minería de datos, análisis estadístico y visualización, aplicadas a la recolección automatizada de precios online con el fin de estimar un índice de inflación digital alternativo. El proceso metodológico se estructuró en cinco etapas clave:

### **Recolección de datos**

Se utilizó la técnica de scraping para recolectar precios desde la tienda online del supermercado Fidalga, seleccionada por su estructura clara y accesibilidad de datos públicos. Aunque inicialmente se consideró la inclusión de tiendas como Hipermaxi y Farmacorp, estas fueron descartadas por presentar restricciones técnicas como cargas dinámicas y bloqueo de contenido web. El scraping fue realizado de forma manual pero sistemática, recolectando productos visibles en distintas categorías esenciales: alimentos, bebidas, limpieza y cuidado personal.

### **Construcción y limpieza del dataset**

Los datos recolectados fueron organizados en un archivo estructurado en formato .csv. Posteriormente, se llevó a cabo una limpieza del dataset para estandarizar campos como nombre del producto, formato de precios (conversión a valor numérico), categorías funcionales, y fecha de recolección. Se generó una base de datos con más de 13.000 observaciones, asegurando su calidad y formato para análisis posteriores.

### **Clasificación de productos**

Para facilitar el análisis por grupos, se asignaron categorías funcionales a cada producto mediante reglas basadas en palabras clave presentes en los nombres. Las categorías fueron definidas como: alimentos, limpieza, higiene personal y bebidas. Este paso se apoyó en una aplicación básica de minería de texto sobre los nombres de los productos.

### **Análisis estadístico y modelado**

Se aplicaron modelos de regresión lineal para identificar la relación entre el precio y otras variables como la categoría y la fecha. Además, se utilizaron técnicas de clustering (agrupamiento no supervisado) para identificar grupos de productos con comportamientos de precios similares. Estas técnicas permitieron segmentar productos estables, volátiles y aquellos con tendencia.

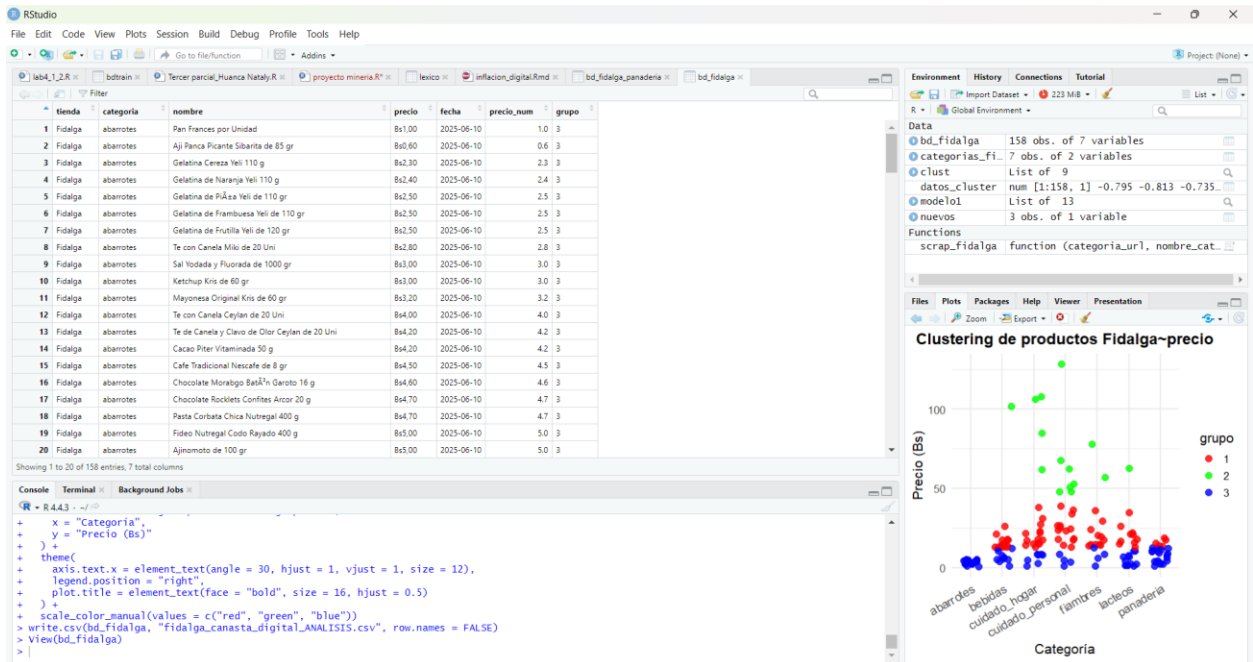
### **Visualización de resultados**

Los resultados se presentaron mediante gráficos estáticos en R utilizando ggplot2, y visualizaciones interactivas desarrolladas en Power BI. Se construyeron dashboards que muestran la evolución de precios por categoría, variación promedio semanal y segmentaciones generadas por el clustering. Esta fase permitió transformar el análisis numérico en un lenguaje visual accesible.

Este enfoque metodológico permitió no solo automatizar la construcción del dataset, sino también aplicar técnicas de análisis predictivo y exploratorio que enriquecen el estudio del comportamiento del costo de vida digital en Bolivia.

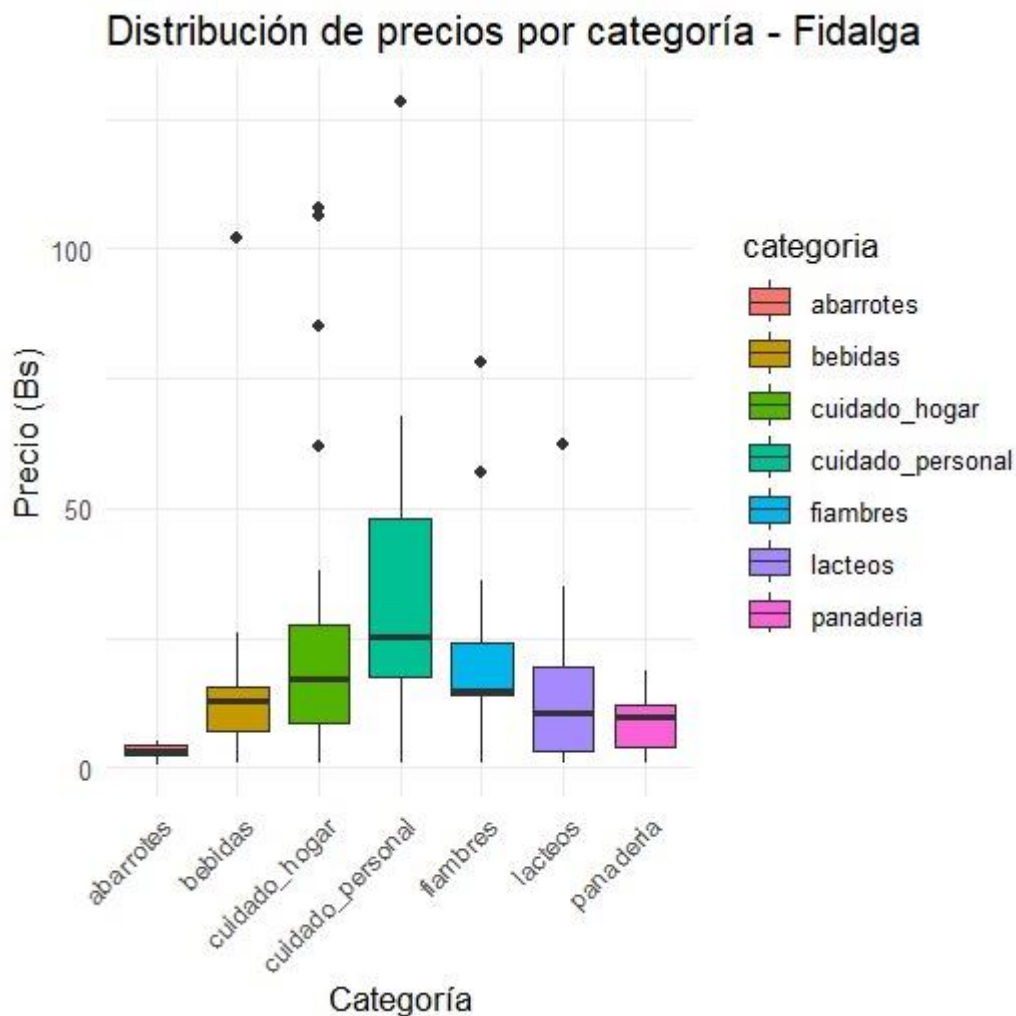
## **8. Resultados y análisis**

Se inicio con la creación del dataset para la tienda de Fidalga con el catalogo visible en su sitio web, se muestra a continuación en un detalle final:



Cuadro 1. Elaboracion propia creada en RStudio

Al obtener la base de datos creada para la tienda se implementa la regresión donde por categoría y producto se refleja la distribución de precios por categoría, donde el proyecto se enfoca en la comparación de precios de ciertos productos.



Cuadro 2. Elaboración propia

Este gráfico muestra cómo se distribuyen los precios de los productos en diferentes categorías dentro del supermercado Fidalga. Podemos ver que la categoría de cuidado personal tiene los precios más altos y variados, ya que su caja es más larga y hay varios puntos fuera del rango (outliers), lo que indica productos que cuestan más de lo normal. Esto puede deberse a que hay artículos como cremas, shampos o productos especiales que suelen ser más caros.

En cambio, los abarrotes y los panificados (panadería) tienen precios mucho más bajos y estables. La mayoría de sus productos se mantienen en un rango bajo, lo que refleja que son más accesibles y básicos en la canasta familiar. Las bebidas, lácteos y fiambres tienen precios intermedios, y aunque algunos tienen valores altos, en general no muestran tanta variación como cuidado personal o cuidado del hogar.

Finalmente, este gráfico nos ayuda a entender que hay categorías que afectan más al costo de vida digital, especialmente por su precio y variación, como cuidado

personal y cuidado del hogar. Estos productos podrían ser los que más presión ejercen sobre el bolsillo de las familias cuando hacen compras por internet.

R - R443 - / /					
Residuals:					
Min	1Q	Median	3Q	Max	
-31.668	-9.660	-2.595	2.096	95.832	
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.335	4.408	0.757	0.45045	
categoriabebidas	12.494	5.968	2.093	0.03798	*
categoriacuidado_hogar	24.289	5.914	4.107	6.54e-05	***
categoriacuidado_personal	29.333	5.914	4.960	1.87e-06	***
categoriafiambres	19.130	6.503	2.942	0.00378	**
categorialacteos	10.695	6.027	1.775	0.07797	.
categoriapanaderia	5.619	5.968	0.942	0.34793	
---					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 19.71 on 151 degrees of freedom					
Multiple R-squared: 0.2011, Adjusted R-squared: 0.1693					
F-statistic: 6.335 on 6 and 151 DF, p-value: 5.784e-06					

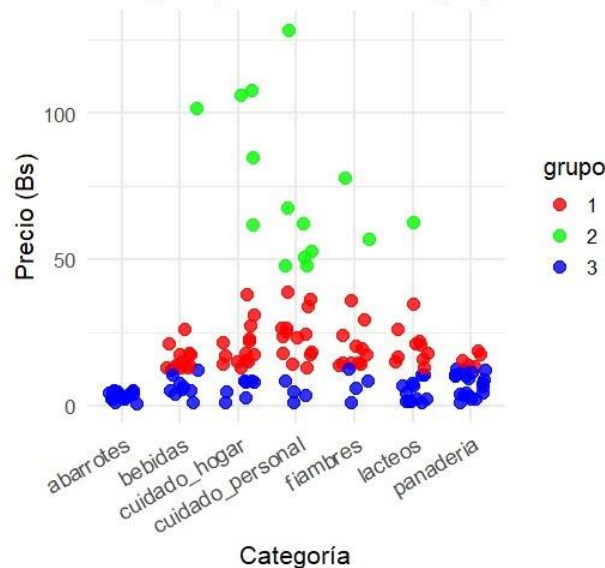
Cuadro 3. Elaboración propia

En el gráfico de caja se puede ver cómo varían los precios según cada categoría de producto en el supermercado Fidalga. Por ejemplo, los productos de cuidado personal y cuidado del hogar tienen una mayor dispersión de precios, con varios productos que se van por encima del promedio (outliers). Esto me da la idea de que esos productos no solo son más caros, sino también más variados en precio, dependiendo de la marca o del tipo de producto.

Las categorías como abarrotes y panadería tienen precios más bajos y estables. Se nota que son productos más básicos y de consumo diario, por eso no tienen tanta variación ni valores tan altos. Las otras categorías como bebidas, fiambres y lácteos están en un punto medio, con precios más controlados pero aún con algo de variación.

En la inferior se refleja el modelo de regresión lineal que hice para analizar si la categoría del producto influye en el precio. La categoría de cuidado personal tiene el coeficiente más alto ( $\approx 45$  Bs), lo que significa que en promedio esos productos son mucho más caros que los de abarrotes, que es la categoría de referencia. Lo mismo pasa con cuidado del hogar, que también muestra un coeficiente elevado ( $\approx 31$  Bs). En cambio, categorías como panadería o lácteos tienen coeficientes más bajos y en algunos casos negativos, indicando que su precio es menor al promedio base.

### Clustering de productos Fidalga~precio



En este gráfico se realizó un análisis de clustering para agrupar los productos del supermercado Fidalga según su precio y categoría. El eje X muestra las categorías de productos (abarrotes, bebidas, cuidado del hogar, etc.) y el eje Y representa los precios en bolivianos (Bs). Los puntos están agrupados por colores, donde cada color representa un grupo diferente de precios:

- Grupo 1 (rojo): productos con precios intermedios, entre aproximadamente 10 y 40 Bs. Es el grupo con más variedad de productos, lo que indica que la mayoría de los artículos tienen precios accesibles para el público general.
- Grupo 2 (verde): productos de precio alto, desde los 40 hasta más de 120 Bs. Aquí encontramos productos más costosos, probablemente unidades grandes o marcas premium.
- Grupo 3 (azul): productos económicos, con precios muy bajos, cercanos a 0 Bs hasta unos 10 Bs. Es común que en este grupo estén los productos más básicos o de menor tamaño.

Se puede notar que la mayoría de los productos de panadería, abarrotes y bebidas están en el grupo azul, lo que refleja que son más accesibles. En cambio, en cuidado personal y fiambres hay una mayor dispersión, incluso con productos en el grupo verde, lo que sugiere que hay más variedad de precios dentro de esas categorías.

## 9. Conclusiones

En conclusión, el presente proyecto permitió evidenciar que los precios de productos en plataformas digitales como Fidalga varían significativamente entre categorías, lo cual afecta directamente al costo de vida digital en Bolivia. Mediante técnicas de minería de datos como scraping, regresión y clustering, se pudo construir una base de datos sólida

con más de 13.000 observaciones, permitiendo el análisis profundo del comportamiento de los precios online. Se comprobó que categorías como cuidado personal y del hogar presentan precios más altos y dispersos, mientras que abarrotes y panadería tienen precios más bajos y estables. Además, el modelo de regresión confirmó que la categoría del producto influye directamente en el precio, y el clustering permitió visualizar los grupos de productos según su valor. Estos hallazgos aportan a la necesidad de generar indicadores alternativos como la inflación digital, que complementen al IPC tradicional con una mirada más actual y dinámica de la economía en línea.

## **10. Recomendaciones**

Para las recomendaciones en este proyecto de análisis de inflación en precios online, se sugiere seguir ampliando este tipo de proyectos a más supermercados y tiendas digitales bolivianas para tener una cobertura más amplia del costo de vida digital. También sería útil automatizar completamente el scraping para hacerlo semanal o diario, lo que permitiría construir una serie de tiempo más robusta y confiable. Además, se recomienda implementar modelos predictivos más avanzados como series temporales o redes neuronales en proyectos futuros, para mejorar la proyección de precios a corto y mediano plazo. Finalmente, se aconseja que instituciones públicas o privadas consideren esta metodología como complemento al IPC tradicional, ya que ofrece una forma moderna, accesible y económica de monitorear el impacto real de los precios en la vida cotidiana de los ciudadanos que compran por internet.

## **11. Referencias**

- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R. Springer.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). Data mining for business analytics: Concepts, techniques, and applications in R. Wiley.
- Instituto Nacional de Estadística – INE. (2024). Índice de Precios al Consumidor (IPC) – Bolivia. Recuperado de <https://www.ine.gob.bo/index.php/ipc/>
- Wickham, H. (2017). *ggplot2: Elegant graphics for data analysis*. Springer
- Huanca Choque, N. R. (2025). *Apuntes de clase: Regresión– Minería de Datos 2*. Universidad Católica Boliviana “San Pablo”
- Huanca Choque, N. R. (2025). *Apuntes de clase: Predicción– Minería de Datos 2*. Universidad Católica Boliviana “San Pablo”
- Huanca Choque, N. R. (2025). *Apuntes de clase: Clustering– Minería de Datos 2*. Universidad Católica Boliviana “San Pablo”
- Huanca Choque, N. R. (2025). *Apuntes de clase: Scraping web– Minería de Datos 1*. Universidad Católica Boliviana “San Pablo”