



Modelos de forecasting aplicados a precio base de subastas ganaderas

Giovanny Reales Rodríguez

Ingeniero Agropecuario

Asesor

Edimer David Jaramillo, Especialista en Estadística.

Universidad de Antioquia

Facultad de Ciencias Agrarias, Escuela de Producción Agropecuaria
Ingeniería Agropecuaria

Caucasia

2022

Cita	(Reales Rodríguez, 2022)
Referencia	Reales Rodríguez, G. (2022). <i>Modelos de forecasting aplicados a precio base de subastas ganaderas</i> [Trabajo de grado profesional]. Universidad de Antioquia, Cauca, Colombia.
Estilo APA 7 (2020)	



Biblioteca Seccional Bajo Cauca (Caucasia)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A mi familia, amigos y seres queridos, quienes desde el principio me han apoyado incondicionalmente en mi proceso académico. Gracias por su paciencia, comprensión y amor en todas las etapas de mi vida, y en especial durante mi carrera universitaria y en la realización de esta tesis. También, quisiera agradecer a mis profesores y asesores de investigación por su guía y enseñanzas que me permitieron llegar hasta aquí. Este trabajo está dedicado a todos aquellos que han sido mi inspiración y motivación para seguir adelante en esta travesía. ¡Gracias por ayudarme a hacer realidad este logro!

Agradecimientos

Agradezco a todas y cada una de las personas que tuvieron la paciencia de esperar este momento, por cada café regalado, las palabras de aliento que me brindaron, a mi familia por confiar en mí, a mis profesores Mayra Mercado, Edimer David Jaramillo y Marcos Julio Cañas por todo su esfuerzo y los conocimientos que me transmitieron para hacer esto realidad. A la empresa Cencogan por brindarme la información necesaria.

Tabla de contenidos

Resumen	9
Abstract.....	10
Introducción.....	11
Metodología.....	13
Ubicación de subasta CENCOGAN	13
Fuentes de información	13
Descripción de datos.	14
Web Scraping	15
Análisis Descriptivo y Exploratorio.....	15
Modelación.....	16
Software	17
Tamaño muestral.....	18
Series temporales.....	19
Precio base.....	19
Distribución del precio base.....	20
promedio del Precio base por categoría.....	21
Comportamiento promedio de variables exógenas.....	22
Resultados.....	23
Series temporales normalizadas.....	23
Coeficiente de variación del precio base.....	24
Variación mensual Pre-Covid.....	25
Variación mensual Post-Covid.....	26
Distribución de las variables.....	27
Matriz de correlación.	28
Diagrama de dispersión	29
Análisis de componentes principales (PCA).....	30
Selección de componentes principales.....	30
Importancia de las componentes	31
Biplot.....	32
Procesamiento de texto.....	33
Modelación	34
Comparación de modelos.....	34
Precisión global.....	34

Precisión local.....	34
Predicciones en test.....	36
Pronósticos con nuevos datos	37
Conclusiones	40
Referencias	41

Listado de Tablas

Tabla 1 Variables de la subasta	14
Tabla 2 Categorías en la subasta.....	14
Tabla 3 Variables precio del maíz	15
Tabla 4 Variables del ganado gordo en la bolsa mercantil de Colombia	15
Tabla 5 Bibliotecas Usadas.....	17
Tabla 6 Cantidad de datos por año.	18
Tabla 7 Cantidad de datos por Categoría.....	18
Tabla 8 Retención de variabilidad ACP	30
Tabla 9 precisión global	34
Tabla 10 precisión local.....	34

Listado de Ilustración

Ilustración 1 Ubicación de CENCOGAN.....	13
Ilustración 2 Serie temporal del promedio precio base	19
Ilustración 3 Distribución del precio base	20
Ilustración 4 promedio del Precio base por categoría.....	21
Ilustración 5 Comportamiento del promedio de las variables exógenas	22
Ilustración 6 Comportamiento de las variables por categoría, promedio normalizado a través del tiempo.....	23
Ilustración 7 variación del precio base por categoría por año	24
Ilustración 8 variación del precio base por categoría por año Pre Covid	25
Ilustración 9 variación del precio base por categoría por año Post covid	26
Ilustración 10 variación del precio base por categoría por año	27
Ilustración 11 matriz de correlación	28
Ilustración 12 Dispersión de las variables vs precio base	29
Ilustración 13 Representación de las variables en las dimensiones.....	31
Ilustración 14 Representación simultanea	32
Ilustración 15 Nube de 3 palabras de la variable observaciones	33
Ilustración 16 Modelos regresión lineal múltiple	36
Ilustración 17 Modelos ARIMA.....	37
Ilustración 18 Predicción Modelos de regresión lineal múltiples.....	38
Ilustración 19 Predicción Modelos ARIMA.....	39

Siglas, acrónimos y abreviaturas

HL	Hembra de levante
HV	Hembra de vientre
MC	Macho de ceba
ML	Macho de levante
TO	Toro
VE	Vaca escotera
VP	Vaca parida

Resumen

Con el objetivo de entender el comportamiento de los precios base que se aplican en la subasta CENCOGAN, se llevó a cabo un proceso de extracción de datos utilizando la técnica de Web Scraping en las páginas oficiales de CENCOGAN, Fedegan y Datos Abiertos. Luego, la información recolectada fue analizada en detalle mediante el uso del programa RStudio, aplicando una serie de técnicas de análisis estadístico descriptivo, exploratorio univariado, bivariado y multivariado. Una vez completado el análisis descriptivo, se aplicaron técnicas de estadística inferencial y de pronóstico (forecasting) para modelar la variación de los precios a lo largo del tiempo. En este sentido, se utilizaron tanto modelos de regresión lineal múltiple como modelos Autorregresivos (ARIMA), con el fin de identificar patrones y tendencias a lo largo del tiempo. En conclusión, este estudio proporciona una visión completa y detallada del comportamiento de los precios base en la subasta CENCOGAN, utilizando una variedad de técnicas de análisis estadístico y herramientas de modelado predictivo. Los resultados obtenidos pueden ser de gran utilidad para los inversores y demás actores del mercado ganadero, permitiéndoles tomar decisiones más informadas y precisas en cuanto a la gestión de sus operaciones y estrategias comerciales.

Palabras clave: Web Scraping, Subastas, Series Temporales, Ganadería

Abstract

With the aim of understanding the behavior of the base prices applied in the CENCOGAN auction, a data extraction process was carried out using the Web Scraping technique on the official websites of CENCOGAN, Fedegan, and Open Data. Subsequently, the collected information was analyzed in detail using the RStudio program, applying a series of techniques including descriptive, exploratory univariate, bivariate, and multivariate statistical analysis. Once the descriptive analysis was completed, inferential statistics and forecasting techniques were applied to model the variation of prices over time. In this regard, both multiple linear regression models and Autoregressive Integrated Moving Average (ARIMA) models were used to identify patterns and trends over time. In conclusion, this study provides a comprehensive and detailed insight into the behavior of base prices in the CENCOGAN auction, using a variety of statistical analysis techniques and predictive modeling tools. The obtained results can be highly valuable for investors and other stakeholders in the livestock market, enabling them to make more informed and precise decisions regarding the management of their operations and business strategies.

Keywords: Web Scraping, Auctions, Time Series, Livestock.

Introducción

Se le llama ganado a todas las especies de mamíferos, aves, reptiles, peces y artrópodos que se han domesticado y mantenido en cautividad para obtener de ellos alimentos y otros productos, trabajar con ellos y otros usos. (Stellman 1998).

Según el Instituto Colombiano Agropecuario (ICA), Colombia cuenta con 29.301.392 cabezas de ganado bovino para el año 2022, el 68,6% del total de ganado bovino se concentra en diez departamentos, Antioquia (11,2%), Córdoba (7,8%), Meta (7,8%), Caquetá (7,5%), Casanare (7,5%), Santander (5,7%), Magdalena (5,5%), Cesar (5,5%), Bolívar (4,9%) y Cundinamarca (5,0%) (Instituto Colombiano Agropecuario ICA s. f.).

Las subastas son espacios donde se intercambian bienes y servicios, se tienen en cuenta las pujas que realizan los demandantes y un conjunto de reglas para que el desarrollo de la misma sea de forma efectiva, correcta y transparente (Montaño García 2018).

Las subastas ganaderas en Colombia se crearon hace más de 30 años, como una solución a la difícil situación que en los años 80 y 90 se vivía por falta de garantías de seguridad y el mal estado de las vías de acceso a las zonas rurales para la compra y venta del ganado bovino («Las subastas ganaderas» 2014).

En Colombia, las subastas ganaderas que se realizan son de tipo inglesa o ascendente, su característica principal radica en que el precio va incrementando sucesivamente hasta que queda un único comprador, este es quien define el precio final del bovino (Durá Juez 2003).

Estudios recientes demostraron que existen otras variables, además de la oferta y demanda que influyen en el precio del ganado al momento de ser comercializado; dentro de los factores analizados se encuentra el sexo, peso, temporada del año, ubicación geográfica, entre otros (McCabe et al. 2019) (Smith et al. 2021) (Denogean Ballesteros 2004).

La tecnología ha jugado un papel importante en la modernización de las empresas y en la búsqueda continua del aumento de los rendimientos y beneficios económicos y sociales, las subastas ganaderas no han sido la excepción, como las construcciones de una tecnología de red para las subastas de ganado (Graham 1998), sistemas de información para la identificación de los animales en las subastas (Bolte, Dhuyvetter, y Schroeder 2008), construcción de App móviles para las subastas ganaderas (Sarmiento Ortiz 2016) o sistemas de subastas Online (ANDRADE CACERES 2017) todos en pro de aumentar la cantidad de clientes a los cuales llegar.

También ha cobrado importancia el uso de las herramientas brindadas por la estadística para entender los fenómenos que suceden en el mercado de las subastas ganaderas, movimientos de los precios y su comercialización, ya que estos impactan de forma directa al beneficio de las producciones bovinas, por ejemplo, el uso de modelos lineales aditivos (Grajales Cedeño et al. 2021), media móvil multiplicativa, la tasa de crecimiento sobre medias anuales, modelos auto-regresivos heterocedásticos condicionales (Castillo N 2007) (Castillo 2007) (Castillo Nuñez 2008), modelo econométrico lineal auto - regresivo con retardos distribuidos (CASTILLO NUNEZ 2022).

La pandemia COVID-19 tuvo efectos negativos en todo el mundo, en todos los mercados, incluyendo el comercio de ganado bovino y las subastas, afectando los precios, sistema

logístico, la producción, la demanda de la carne bovina, mercados internacionales, cadenas de valor y con todo ello, afectando a los productores y a las subastas (Burkart et al. 2020).

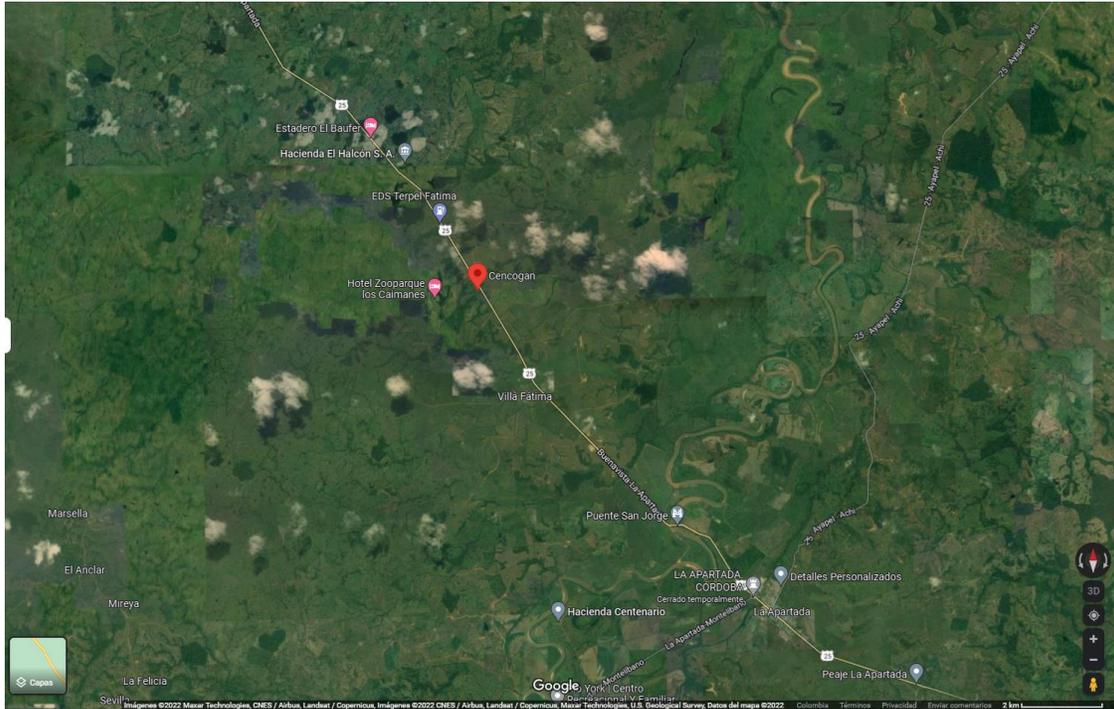
En virtud de lo expuesto anteriormente, este trabajo tuvo como propósito utilizar el software libre “R” a través del Entorno de Desarrollo Integrado (IDE) Rstudio, con la finalidad de implementar técnicas de Web Scraping para extracción de información de la subasta CENCOGAN y uso de métodos estadísticos para identificar factores que influyen en el precio base del ganado bovino.

Metodología

Ubicación de subasta CENCOGAN

Ilustración 1

Ilustración 1 Ubicación de CENCOGAN



Nota. Fuente <https://www.google.com/maps>

Ubicada en latitud 8.120008 con longitud -75.402306 en la zona rural del municipio de Buenavista Córdoba, con una altitud de 60 m.s.n.m. y con temperaturas medias de 30°C.

Fuentes de información

El proyecto se desarrolló con la información suministrada por la empresa Centro Comercial Ganadero S.A.S. (CENCONGAN) por medio de su página web <https://cencogan.com/loten-linea/>, en la cual, suben los resúmenes de las subastas que realiza la empresa (CENCOGAN, s. f.).

Se extrajeron los datos de la Tasa Representativa del Mercado de la página Datos Abiertos del Gobierno de Colombia <https://www.datos.gov.co/Econom-a-y-Finanzas/TRM/ceyp-9c7c/data> (Superintendencia financiera de Colombia).

En la página de FEDEGAN, se obtuvo el precio internacional del maíz y el precio del ganado gordo en la bolsa mercantil de Colombia <https://www.fedegan.org.co/estadisticas/precios> (Federación Colombiana de Ganaderos, FEDEGAN).

Descripción de datos.

Para la extracción de la información se utilizó la técnica de raspado de información en la red “Web Scraping” (Murillo y Saavedra 2017) generando múltiples archivos en formato CSV, estos contienen los resúmenes de cada subasta que realizó en el horizonte de la investigación, el precio internacional del maíz y el precio del ganado gordo en la bolsa mercantil de Colombia publicados por FEDEGAN en su página web y la tasa representativa del mercado publicada en la página de datos abiertos del gobierno nacional Colombiano.

Los datos comprenden las siguientes fechas: desde el 08 febrero del año 2018 hasta el 10 de octubre del 2022.

Los datos recolectados de la subasta tienen las siguientes variables:

Tabla 1.

Tabla 1 Variables de la subasta

Variables	Descripción	Tipo
Categoría	Categoría del ganado a comercializar	Nominal
Cantidad	Numero de animales en el lote	Numérico
Peso total	Peso total de los animales en el lote	Numérico
Peso Promedio	Promedio del peso de los animales en el lote	Numérico
Procedencia	Finca o municipio de salida del animal	Nominal
Observaciones	Notas sobre el lote	Nominal
Entrada	Hora de entrada del lote a la subasta	Date
Precio base	Precio de inicio para la subasta	Numérico
Precio final	Precio al que fue comprado el lote	Numérico

Cada categoría se subdivide en diferentes tipos y estos se relacionan según la condición de los animales, se tendrán en cuenta las siguientes categorías.

Tabla 2

Tabla 2 Categorías en la subasta

Abreviatura	Significado	Característica
HL	Hembra de levante	Pesa Menos de 250 Kg
HV	Hembra de vientre	Pesa Mas de 250 Kg
MC	Macho de ceba	Pesa Mas de 250 Kg
ML	Macho de levante	Pesa Menos de 250 Kg
TO	Toro	Pesa Mas de 500 Kg
VE	Vaca escotera	Despues del primer parto
VP	Vaca parida	Esta en lactancia

Solo se tuvo en cuenta los datos de ganado bovino comercial.

Los datos procedentes de fedegan con respecto al precio del maíz tienen las siguientes variables:

Tabla 3

Tabla 3 Variables precio del maíz

Variable	Descripción	Tipo
fecha	Fecha de registro	Date
maiz_fob_argentino	Precio internacional del maiz (USD/TON)	Numerico
maiz2_fob_usa	Precio internacional del maiz (USD/TON)	Numerico
maiz3_fob_usa	Precio internacional del maiz (USD/TON)	Numerico

Los datos procedentes de fedegan con respecto al precio del ganado gordo en la bolsa mercantil de Colombia tienen las siguientes variables:

Tabla 4

Tabla 4 Variables del ganado gordo en la bolsa mercantil de Colombia

Variable	Descripción	Tipo
depto	Departamentos	Carácter
fecha	fecha de registro	Carácter
precio_kilo_en_pie	Precio kilo en pie en pesos Colombianos	Numérico
fecha	fecha de registro	Date

La fecha es la conversión de la fecha suministrada por FEDEGAN para hacer el análisis en el programa Rstudio.

Las bases de datos se unieron en una sola base de datos general para mejor manejo de la información.

Web Scraping

Técnicas de *Web Scraping* fueron implementadas para extraer información desde la página web oficial de Cencogan.. Diferentes componentes de la página web fueron accedidos a través de la biblioteca rvest para extracción de información relacionada con los lotes de subastas ganaderas. Todo el proceso se automatizó a través del software R y es funcional para la obtención periódica de nueva información (nuevos lotes).

Análisis Descriptivo y Exploratorio

Resúmenes descriptivos fueron calculados para las variables bajo análisis. Gráficos para detección de atipicidades en los datos fueron construidos como parte del análisis exploratorio. Relaciones entre variables fueron analizadas a través de coeficientes de correlación no paramétrica (*Spearman*) y gráficos de dispersión. Patrones distribucionales fueron analizados con histogramas, densidades y diagramas de cajas y bigotes. Análisis de Componentes Principales (ACP) fue implementado para las variables numéricas y se incorporó la variable categórica relacionada con el *COVID* como variable suplementaria

cualitativa. El criterio de *Kaiser-Guttman* Ultsch y Löttsch (2015), Taherdoost, Sahibuddin, y Jalaliyoon (s. f.), el cual sugiere que los componentes principales con *valor propio* mayor a 1 deben ser considerados, fue utilizado para determinar el número de componentes principales óptimo. Nubes de palabras se utilizaron para representación de frecuencias en variables cualitativas con alto número de niveles.

Modelación

Se compararon métodos estadísticos de pronóstico (*forecasting*) para modelar la variación a través del tiempo del precio base del ganado en las subastas. El modelo de regresión lineal múltiple permitió estudiar combinaciones lineales entre las variables y el precio del ganado, también fueron ajustados modelos *Autoregresivos* (Rob J. Hyndman y Athanasopoulos 2021), cuyo interés radica en la combinación lineal de **valores pasados** (*rezagos*) de la variable precio base. Los modelos de regresión (1) y autoregresivo (2) pueden ser expresados de forma genérica como se muestra a continuación:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon_t \quad (1)$$

$$y = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (2)$$

Donde y_t es la variable dependiente o respuesta (*precio del ganado*), las variables X_1, X_2, \dots, X_p y $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ son variables independientes o predictorias, β_0 y C son constantes (representan el intercepto) y ϵ representa el error aleatorio del modelo.

Seis modelos fueron ajustados para la serie temporal del precio base, estos modelos se describen a continuación:

- **Modelo 1:** modelo de regresión lineal múltiple con inclusión sólo de variables derivadas de la fecha.
- **Modelo 2:** modelo de regresión lineal múltiple con inclusión variables derivadas de la fecha y la variable exógena TRM.
- **Modelo 3:** modelo de regresión lineal múltiple con inclusión variables derivadas de la fecha y las variables exógenas TRM, precio de ganado nacional y precio del maíz en bolsa.
- **Modelo 4:** modelo *Autoregresivo Integrado de Medias Móviles - ARIMA* con inclusión sólo de variables derivadas de la fecha.
- **Modelo 5:** modelo *Autoregresivo Integrado de Medias Móviles - ARIMA* con inclusión variables derivadas de la fecha y la variable exógena TRM.
- **Modelo 6:** modelo *Autoregresivo Integrado de Medias Móviles - ARIMA* con inclusión variables derivadas de la fecha y las variables exógenas TRM, precio de ganado nacional y precio del maíz en bolsa.

Todos los modelos fueron ajustados utilizando estrategias de validación cruzada simple para series temporales, es decir, fraccionando los datos en conjuntos de *entrenamiento* (*train*) y *prueba* (*test*), la ventana de tiempo utilizada para el conjunto de prueba fue de tres meses.

El desempeño predictivo de los modelos fue obtenido a través de las métricas de error *Error Porcentual Absoluto Medio - MAPE* y la *Raíz del Cuadrado Medio del Error - RMSE* R. J. Hyndman y Athanasopoulos (2023) Bianco et al. (2010).

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}$$

Donde y_t es el valor real y \hat{y}_t es el valor predicho o estimado por el modelo. Para ambas métricas un valor más bajo será interpretado como mejor ajuste del modelo.

Software

Todos los análisis fueron desarrollados con el software estadístico R (R Core Team) a través del *Entorno de Desarrollo Integrado (IDE)* R Core Team Bibliotecas especializadas en diferentes procedimientos de acceso, transformación, visualización y modelamiento de datos fueron utilizadas para extracción de información, análisis descriptivo y exploratorio y comparación de modelos estadísticos. Algunas de las bibliotecas de uso frecuente en este trabajo se muestran en la tabla 5.

Tabla 5

Tabla 5 Bibliotecas Usadas

Biblioteca	Descripción	Sitio Web - URL
rvest	Técnicas de <i>Web Scraping</i> con R	https://rvest.tidyverse.org/
tidyverse	Colección de bibliotecas para <i>Ciencia de Datos</i> con R	https://www.tidyverse.org/
FactoMineR	Análisis estadístico multivariado con R	http://factominer.free.fr/
tidymodels	Colección de bibliotecas para modelamiento estadístico con R	https://www.tidymodels.org/
modeltime	Ajuste de modelos de pronóstico (<i>forecasting</i>) con R	https://business-science.github.io/modeltime/

Tamaño muestral.

En la tabla 6 se muestran la cantidad de datos de la subasta extraído por año hasta el horizonte del proyecto de investigación, Observamos que el año 2019 tiene la mayor cantidad de datos con un 25.12%, y el año 2020 el que menos datos tiene con un 10.48%, esto se explica por el cierre de operaciones para cumplir con las exigencias del gobierno para combatir la COVID 19.

Tabla 6

Tabla 6 Cantidad de datos por año.

Año	N	Proporción (%)
2018	23266	22.5
2019	25978	25.12
2020	10839	10.48
2021	22709	21.96
2022	20633	19.95

En la tabla 7 se ve la cantidad de datos que se tiene de cada categoría de ganado bovino vendido en la subasta, en los 5 años de subasta que se tiene reportado en la investigación, las Vacas Escoterías son las que más se han vendido con un 35% de los datos, además, el 67.05% de los datos están concentrados en las categorías Vaca Escotera (VE) y Hembra de Levante (HL) y Hembra de Vientre (HV) mostrando que en la subasta CENCOGAN se comercializa más ganado bovino hembra que machos.

Tabla 7

Tabla 7 Cantidad de datos por Categoría.

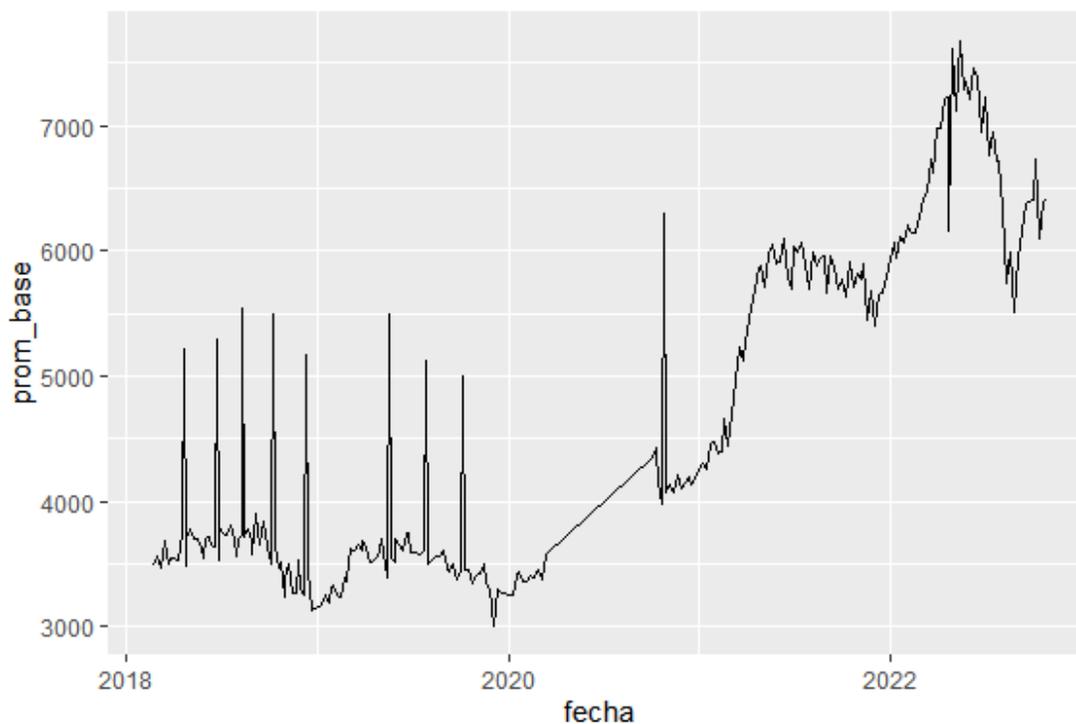
Categoría	N	Proporción (%)
VE	36202	35
HL	17730	17.14
ML	17398	16.82
HV	15421	14.91
MC	6890	6.662
VP	5644	5.457
TO	4140	4.003

Series temporales

Precio base

Ilustración 2

Ilustración 2 Serie temporal del promedio precio base

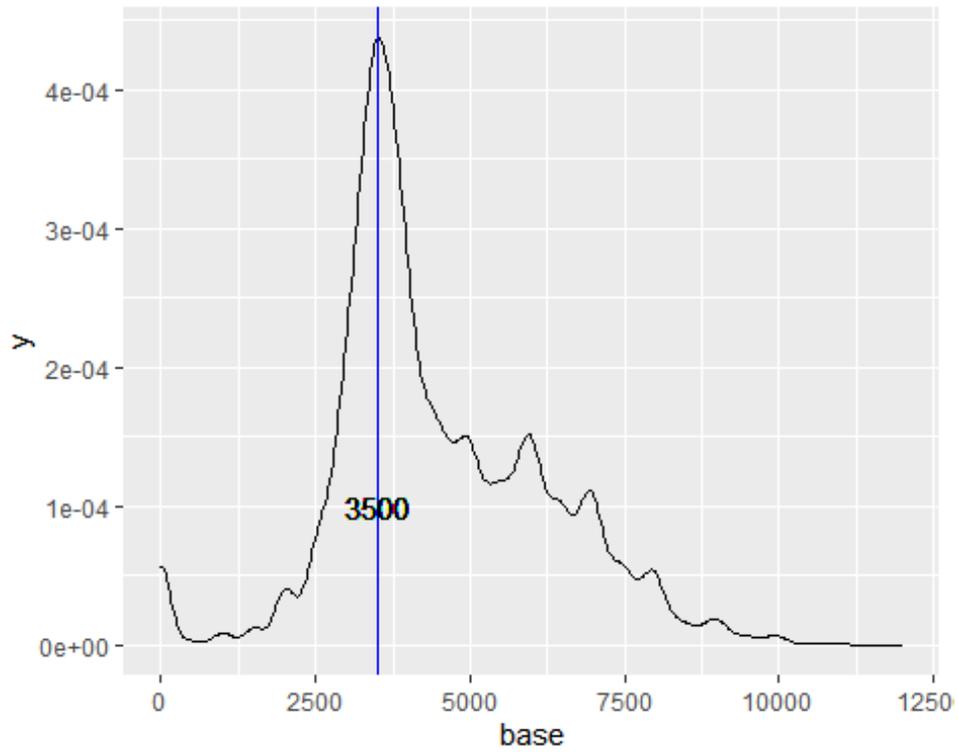


En la Ilustración 2 se observa la serie temporal del promedio de la variable precio base objetivo de la investigación, se omitieron los individuos con precio base igual a 0, el comportamiento entre 2018 y 2020 es no tiene similitudes al comportamiento desde el 2021 hasta octubre del 2022, notándose una disrupción en el mercado.

Distribución del precio base

Ilustración 3

Ilustración 3 Distribución del precio base

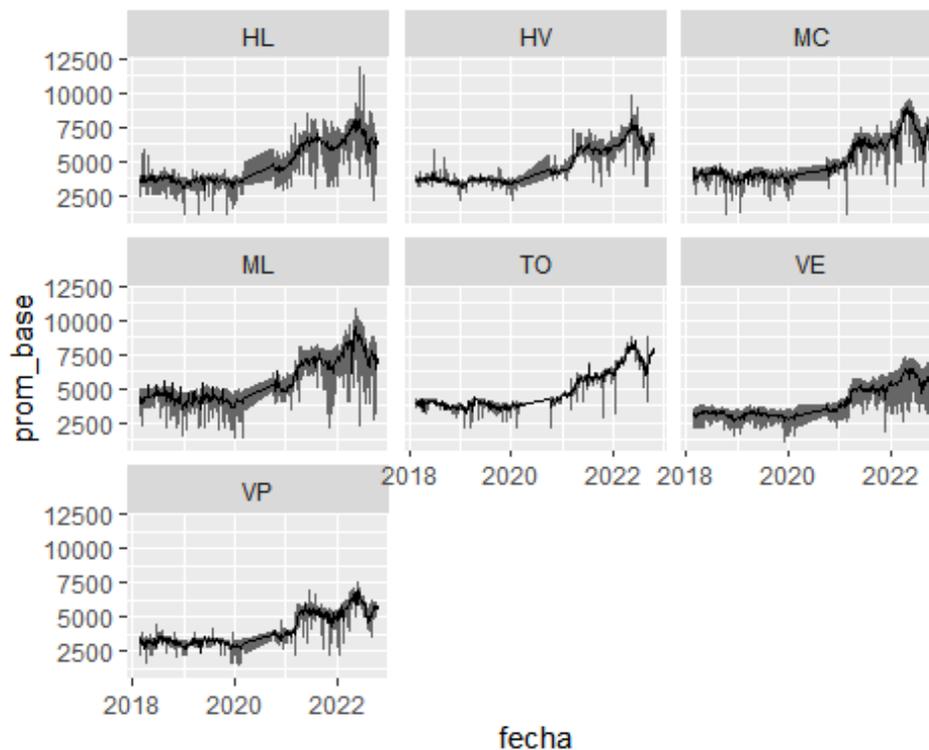


En la Ilustración 3 observamos que el precio con mayor densidad de datos en los 5 años de recolección de información está en los 3500 pesos.

promedio del Precio base por categoría

Ilustración 4

Ilustración 4 promedio del Precio base por categoría

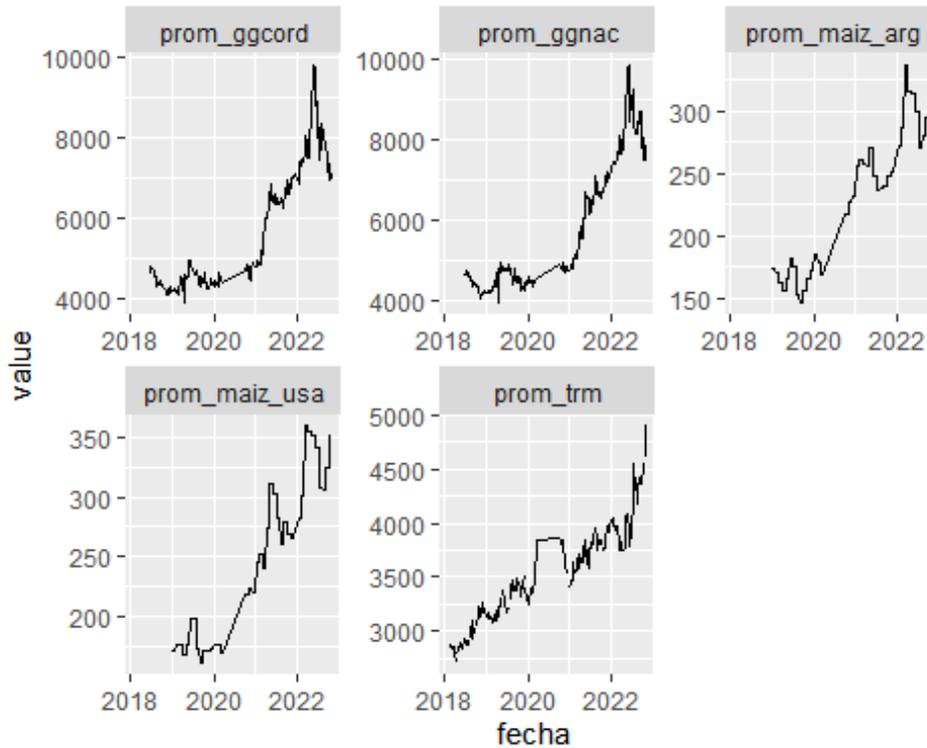


La Ilustración 4, contiene una gráfica por cada categoría de ganado bovino analizada. El comportamiento es similar (alcista) a través del tiempo, pero la categoría toro (TO) tiene un comportamiento más homogéneo, con menos ruido que las demás categorías, vale recordar que la categoría toro (TO) es la que me nos datos tiene.

Comportamiento promedio de variables exógenas

Ilustración 5

Ilustración 5 Comportamiento del promedio de las variables exógenas



En la Ilustración 5, se observa el comportamiento a través del tiempo de las variables ganado gordo córdoba (prom_ggcord), ganado gordo nacional (prom_ggnac), maíz argentino (prom_maiz_arg), maíz USA (prom_maiz_usa), TRM (prom_trm), omitiendo los individuos con precio base 0. El comportamiento de las variables es similar (alcista) a través del tiempo.

Resultados

Series temporales normalizadas

Ilustración 6

Ilustración 6 Comportamiento de las variables por categoría, promedio normalizado a través del tiempo

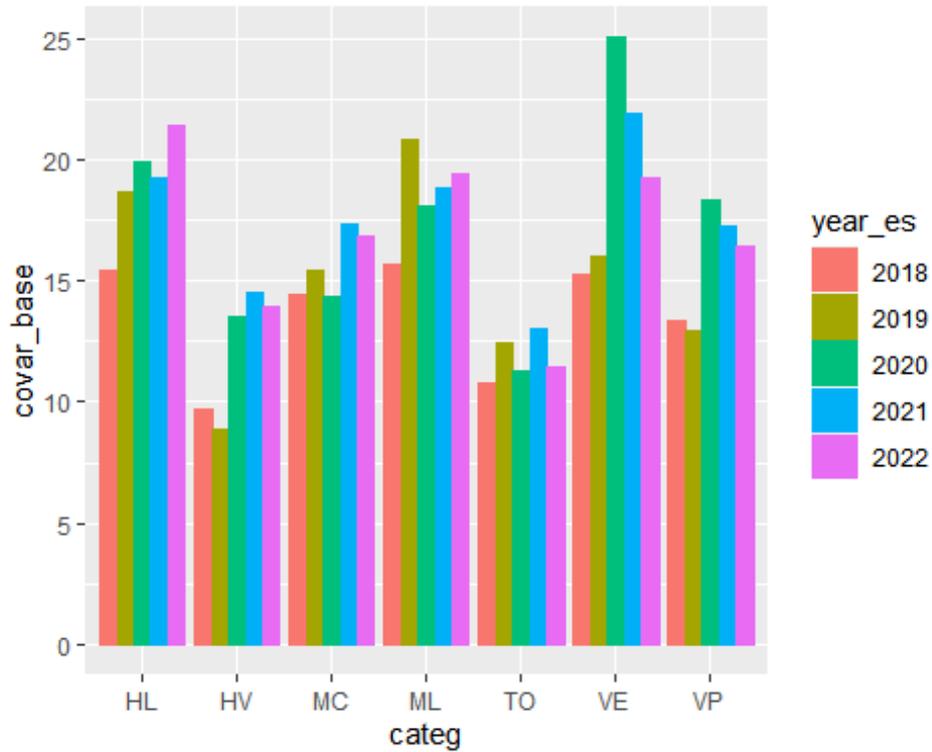


En la Ilustración 6, Unimos todas las variables numéricas tomadas por fuera de la información suministrada por CENCOGAN, las variables fueron normalizadas para compararlas a la misma escala por cada categoría analizada, confirmando que tienen el mismo comportamiento.

Coefficiente de variación del precio base

Ilustración 7

Ilustración 7 variación del precio base por categoría por año

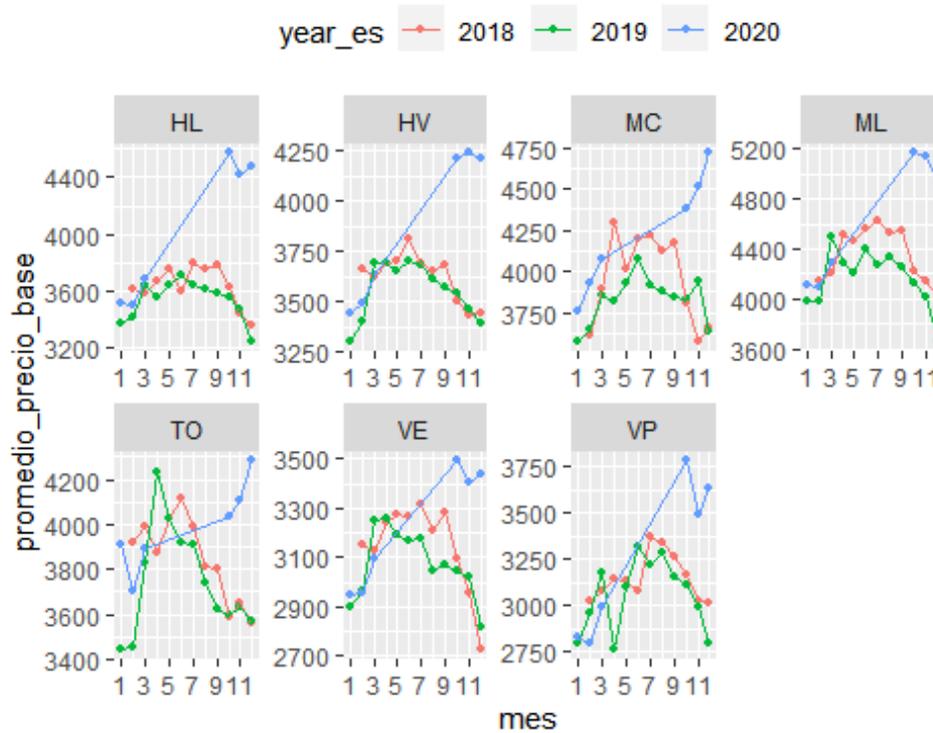


La Ilustración 7, muestra la variación del precio base en cada categoría a través del tiempo, Vemos que el toro (TO) tiene variaciones sostenidas y menores que las demás categorías a través del tiempo.

Variación mensual Pre-Covid

Ilustración 8

Ilustración 8 variación del precio base por categoría por año Pre Covid

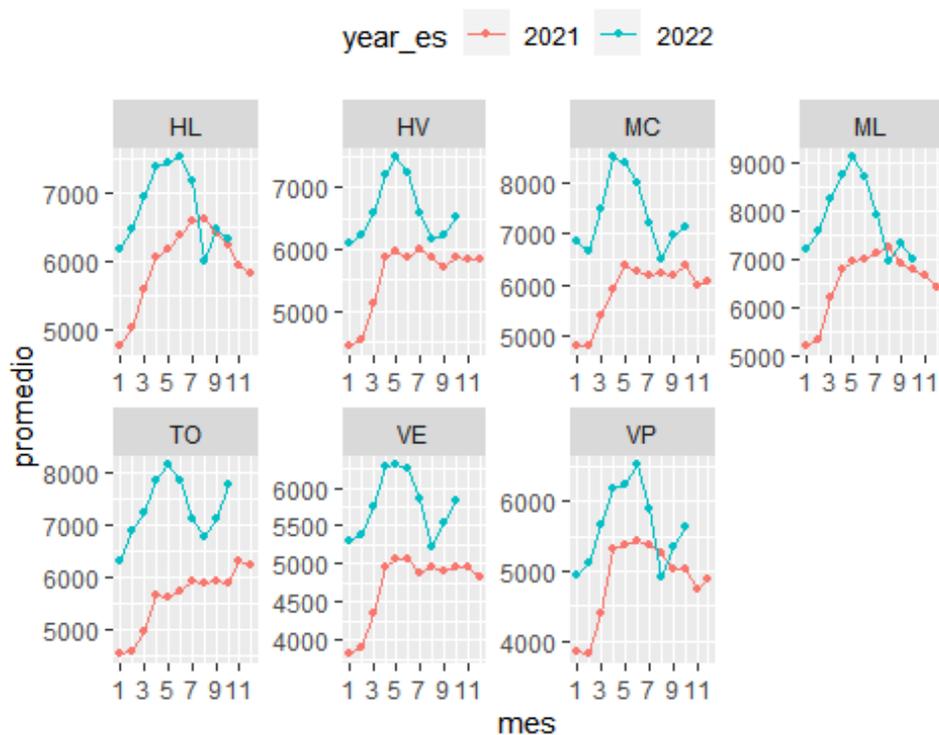


En la Ilustración 8 se visualiza el promedio del precio base por mes del año, cada línea indica un año distinto, en 2018 y 2019 entre los meses 8 y 9 (agosto y septiembre respectivamente) cambia la tendencia, al parecer, es el punto de inflexión para la disminución del precio base, en 2020 hay un cambio en la tendencia.

Variación mensual Post-Covid

Ilustración 9

Ilustración 9 variación del precio base por categoría por año Post covid

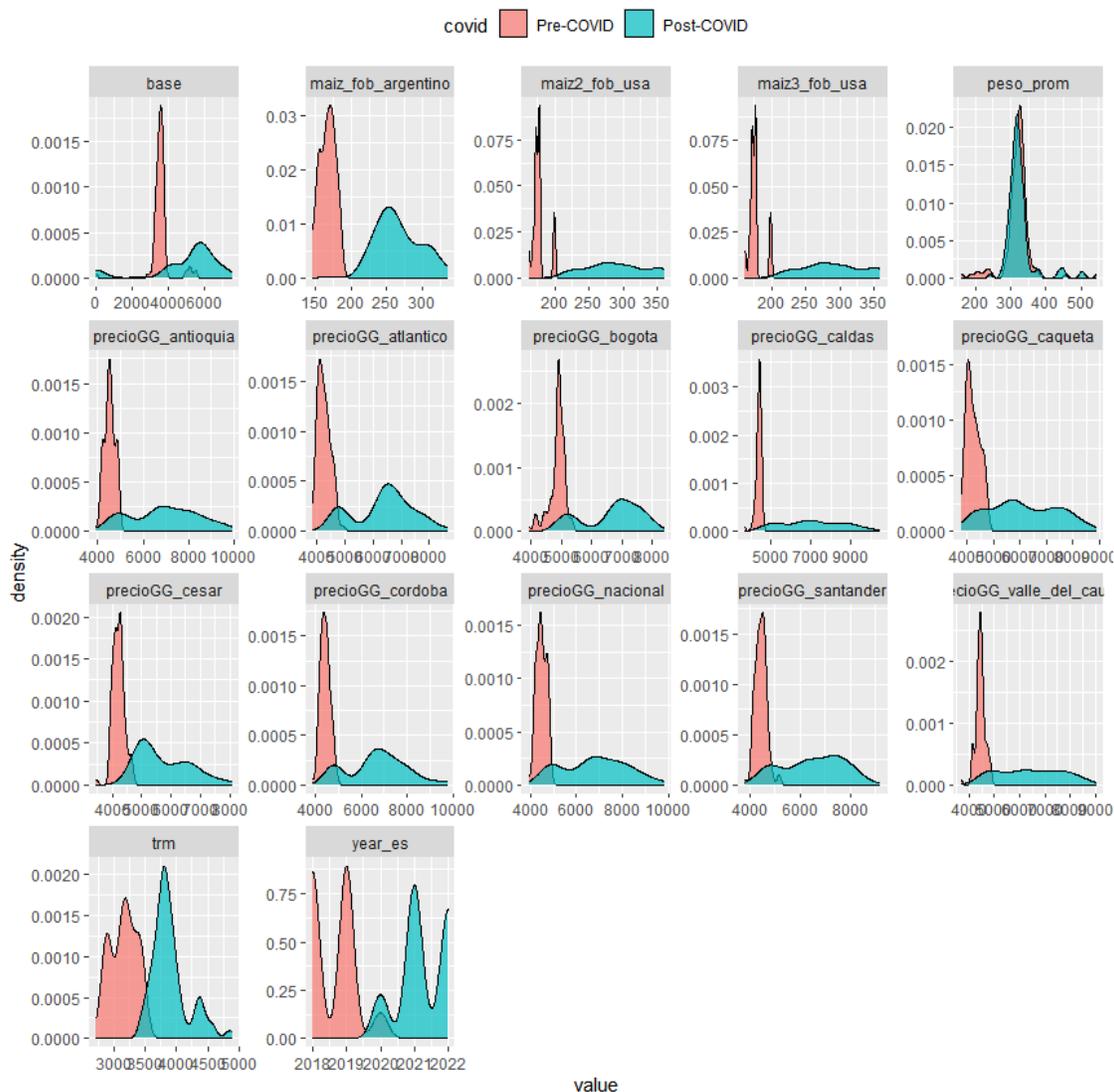


En la Ilustración 9 se visualiza el promedio del precio base por mes del año, cada línea indica un año distinto, en comparación con la Ilustración 8, las unidades promedio son mucho mayores, no hay un comportamiento marcado entre un año y otro, lo cual confirma el cambio en el promedio del precio base luego de la pandemia.

Distribución de las variables.

Ilustración 10

Ilustración 10 variación del precio base por categoría por año

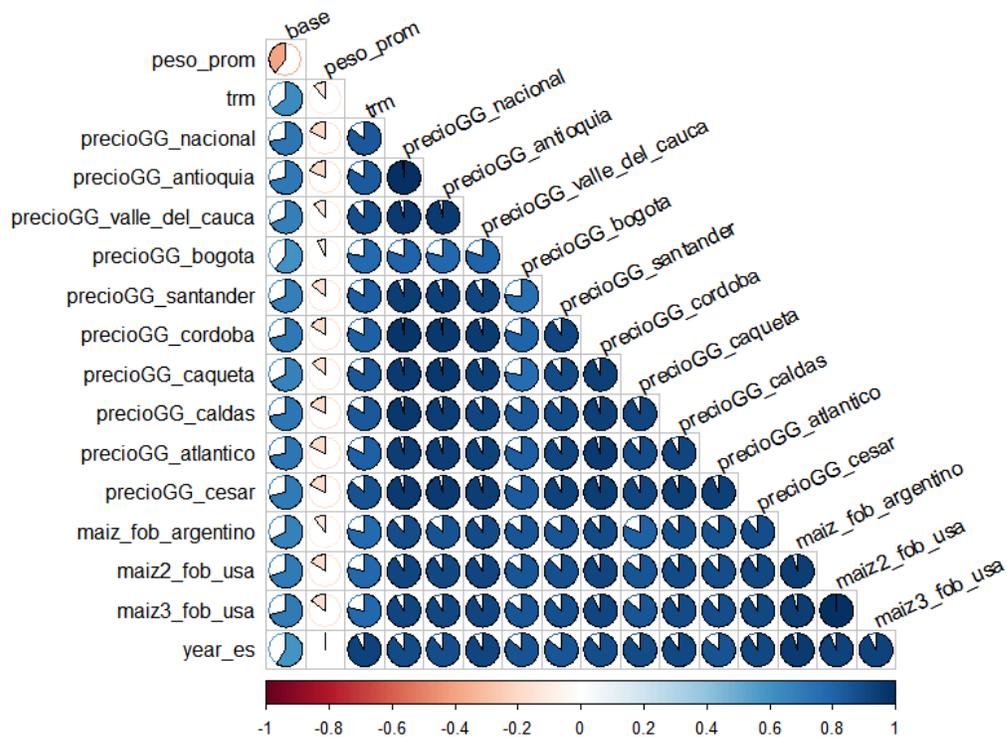


En la Ilustración 10, se observa el comportamiento de las variables antes y después del Covid-19, el peso promedio de los animales es casi idéntico, en todas las demás variables hay diferencias que marcan el cambio de comportamiento después del Covid-19.

Matriz de correlación.

Ilustración 11

Ilustración 11 matriz de correlación

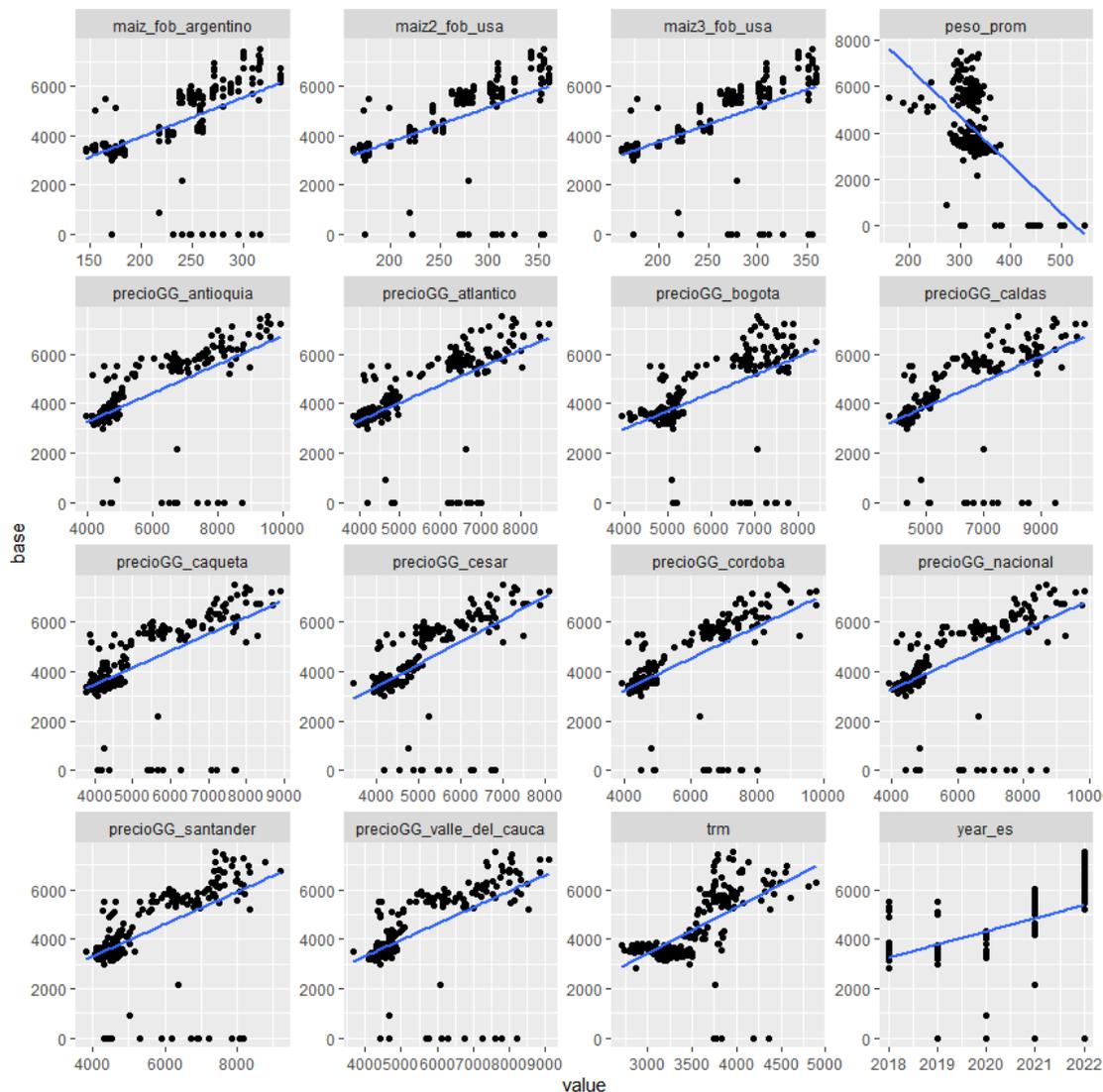


La Ilustración 11 muestra la matriz de correlaciones entre las variables numéricas analizadas, es notable la presencia de correlaciones entre las variables.

Diagrama de dispersión

Ilustración 12

Ilustración 12 Dispersión de las variables vs precio base



En la Ilustración 12, aparece el gráfico de dispersión de las variables vs el precio base en la subasta, omitiendo la variable años (year_es) y el peso promedio (peso_prom), todas las demás variables tienen una tendencia lineal positiva.

Análisis de componentes principales (PCA)

Selección de componentes principales.

Se aplica la técnica de selección de componentes principales por la correlación observada en los gráficos presentados anteriormente, el objetivo es obtener nuevas variables (componentes) no correlacionados y disminuir la dimensionalidad del conjunto de datos.

Para la selección de los componentes principales se tendrán en cuenta el criterio de káiser, este nos dice que, se debe conservar aquellos factores cuyos valores propios (eigenvalues) son mayores a la unidad. Braeken y Assen (2017)

Tabla 8

Tabla 8 Retención de variabilidad ACP

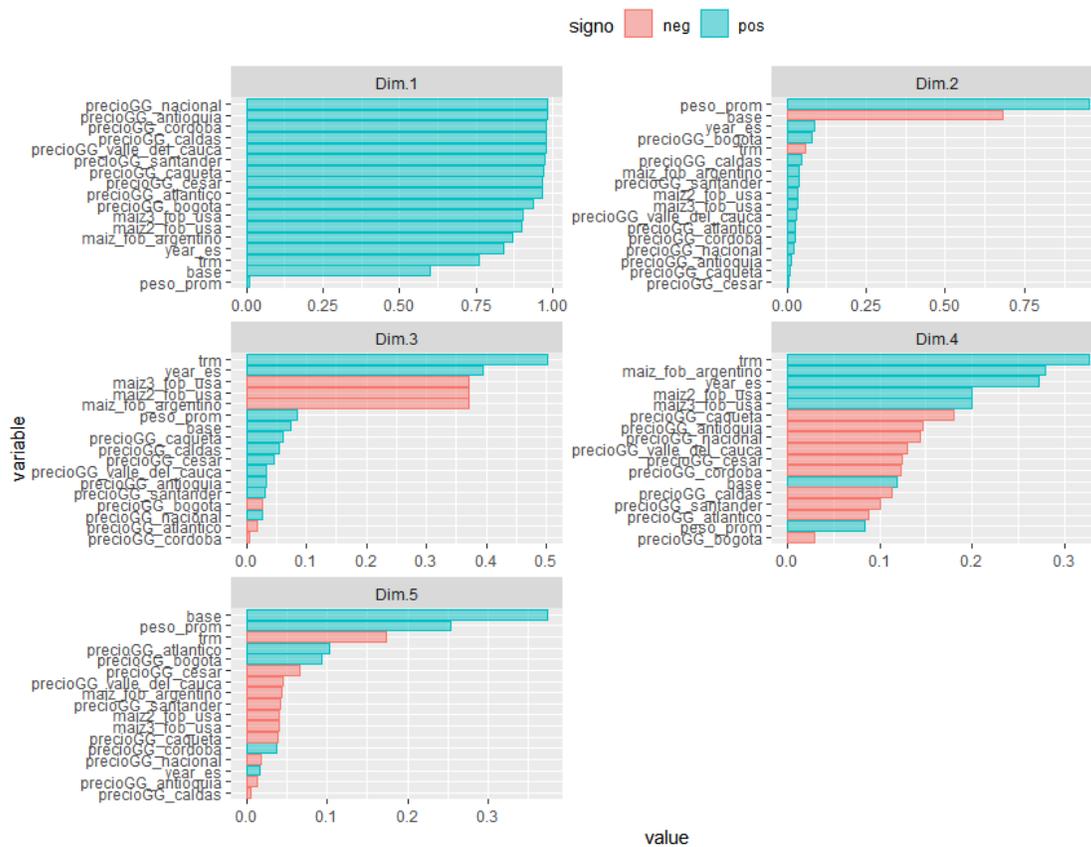
Componente	Valor Propio	Varianza Retenida	Varianza Acumulada
comp 1	13.48	79.27	79.27
comp 2	1.4	8.22	87.49
comp 3	0.85	4.98	92.47
comp 4	0.51	3.02	95.49
comp 5	0.27	1.59	97.08
comp 6	0.18	1.05	98.14
comp 7	0.11	0.64	98.78
comp 8	0.05	0.3	99.08
comp 9	0.04	0.24	99.32
comp 10	0.04	0.21	99.54
comp 11	0.03	0.15	99.69
comp 12	0.02	0.1	99.79
comp 13	0.01	0.08	99.87
comp 14	0.01	0.06	99.93
comp 15	0.01	0.05	99.98
comp 16	0	0.02	100
comp 17	0	0	100

En la tabla 8 se observa la distribución de las componentes, se tomarán las 2 primeras componentes (comp 1 y comp 2), se observa que las dimensiones seleccionadas por el criterio de kaiser representan el 88% de la variación.

Importancia de las componentes

Ilustración 13

Ilustración 13 Representación de las variables en las dimensiones.

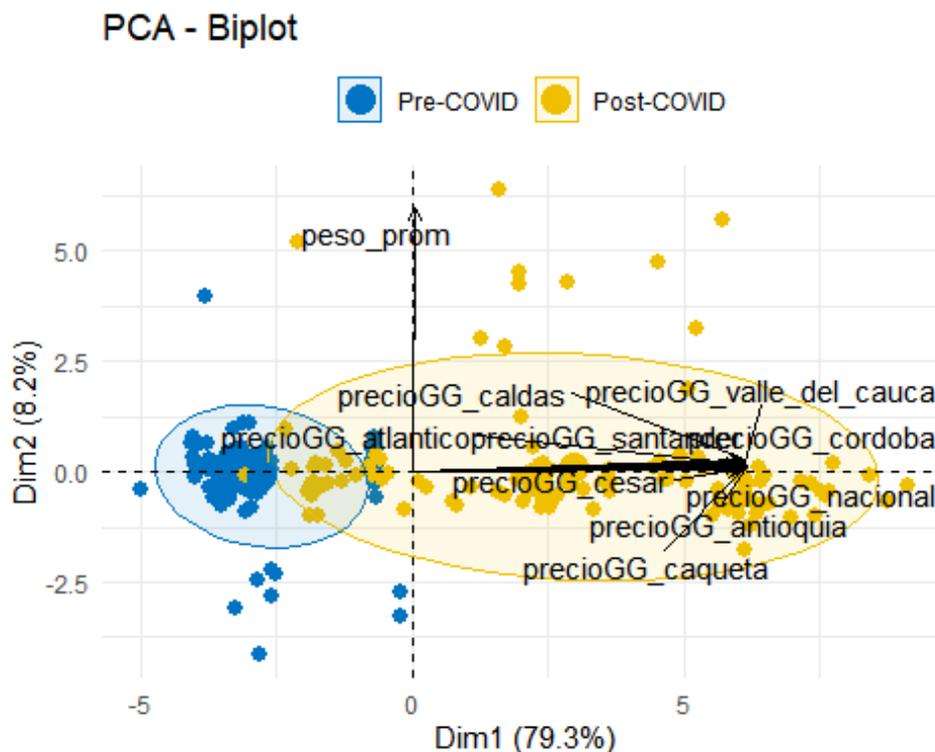


En la Ilustración 13 se observa que las dimensiones 1 y 2 captan las variables en escalas muy superiores que las siguientes dimensiones.

Biplot

Ilustración 14

Ilustración 14 Representación simultanea



En la Ilustración 14 está representado el biplot de las dimensiones seleccionadas luego del proceso de selección de componentes principales, diferenciando su comportamiento antes del Covid-19 y después del Covid-19, los vectores que representan las variables de los precios del ganado gordo en la bolsa mercantil tienen gran efecto en la variación, el vector peso promedio (peso_prom) no tiene vínculos con los demás.

Modelación

Comparación de modelos

A continuación, se observan los resultados obtenidos por los modelos, las métricas seleccionadas para evaluarlos fueron (mape) y (rmse) de forma global y por categoría.

Tabla 9

Tabla 9 precisión global

Precisión global				
id	variables	mape	rmse	
ARIMA1	Solo_fechas	9.15	679	
ARIMA2	TRM	9.36	683.7	
Regresion_Lineal_Multiple3	todasVar	10.02	740.6	
Regresion_Lineal_Multiple1	Solo_fechas	11.2	835.5	
ARIMA3	todasVar	11.6	831.6	
Regresion_Lineal_Multiple2	TRM	11.89	925.6	

En la tabla 9 se observa los resultados globales, el modelo ARIMA sin variables exógenas fue el de mejor comportamiento.

Precisión local

Tabla 10

Tabla 10 precisión local

id	variables	categ	mape	rmse
ARIMA1	Solo_fechas	ML	4.28	390.6
ARIMA2	TRM	ML	4.6	399.3
Regresion_Lineal_Multiple1	Solo_fechas	HV	5.2	413.3
Regresion_Lineal_Multiple1	Solo_fechas	HL	5.62	477.4
ARIMA2	TRM	MC	5.75	452.5
ARIMA1	Solo_fechas	HV	5.82	454.4
ARIMA1	Solo_fechas	MC	6.26	513.6
Regresion_Lineal_Multiple3	todasVar	MC	6.31	611.2
ARIMA2	TRM	HV	6.31	470.3
ARIMA2	TRM	TO	6.45	536.7
Regresion_Lineal_Multiple2	TRM	HV	7	521.1
ARIMA3	todasVar	TO	7.1	615.2
Regresion_Lineal_Multiple3	todasVar	TO	7.4	649.2
ARIMA1	Solo_fechas	HL	7.42	553.9

id	variables	categ	mape	rmse
ARIMA2	TRM	HL	7.57	574.2
Regresion_Lineal_Multiple1	Solo_fechas	MC	7.62	607.2
ARIMA1	Solo_fechas	TO	7.78	665.2
ARIMA3	todasVar	MC	7.85	705.8
Regresion_Lineal_Multiple3	todasVar	ML	7.92	655.7
Regresion_Lineal_Multiple2	TRM	HL	7.93	649.6
Regresion_Lineal_Multiple3	todasVar	HV	8.47	654.5
ARIMA3	todasVar	ML	9.54	778.3
Regresion_Lineal_Multiple3	todasVar	HL	9.84	711.1
ARIMA3	todasVar	HV	9.99	743.7
Regresion_Lineal_Multiple1	Solo_fechas	TO	10.12	882.1
Regresion_Lineal_Multiple1	Solo_fechas	ML	10.25	878.2
Regresion_Lineal_Multiple2	TRM	MC	11.28	910.2
ARIMA3	todasVar	HL	11.67	820.5
Regresion_Lineal_Multiple2	TRM	VE	11.88	778.6
Regresion_Lineal_Multiple3	todasVar	VE	12.05	754.9
ARIMA1	Solo_fechas	VE	12.86	788.5
Regresion_Lineal_Multiple2	TRM	TO	13.99	1203
ARIMA2	TRM	VE	14.22	857.8
ARIMA3	todasVar	VE	14.51	877.1
Regresion_Lineal_Multiple2	TRM	ML	14.85	1192
Regresion_Lineal_Multiple2	TRM	VP	16.3	999.6
Regresion_Lineal_Multiple1	Solo_fechas	VE	17.15	1008
Regresion_Lineal_Multiple3	todasVar	VP	18.13	1054
ARIMA1	Solo_fechas	VP	19.66	1111
ARIMA3	todasVar	VP	20.56	1168
ARIMA2	TRM	VP	20.65	1154
Regresion_Lineal_Multiple1	Solo_fechas	VP	22.42	1246

En la tabla 10 se observan los resultados de la evaluación de los modelos en cada categoría:

- para ML el modelo con mejor comportamiento es el ARIMA sin tener en cuenta las variables exógenas con un mape del 4.28 y un rmse de 390.6
- Para HV el modelo con mejor comportamiento es el Lineal sin tener en cuenta las variables exógenas, con un mape del 5.2 y un rmse de 413.3
- Para MC el modelo con mejor comportamiento es el ARIMA teniendo en cuenta la variable TRM, con un mape de 5.75 y un rmse 452.5

- Para HL el modelo con mejor comportamiento es el Lineal sin tener en cuenta las variables exógenas, con un mape de 5.62 y un rmse 477.4
- Para TO el modelo con mejor comportamiento es el ARIMA teniendo en cuenta la variable TRM, con un mape del 6.45 y un rmse 536.7
- Para VE el modelo con mejor comportamiento es el Lineal teniendo en cuenta la variable TRM, con un mape del 11.88 y un rmse 778.6
- Para VP el modelo con mejor comportamiento es el Lineal teniendo en cuenta la variable TRM, con un mape del 16.3 y un rmse 999.6

Predicciones en test

Los modelos se pusieron a prueba para predecir con valores que no les fueron mostrados, algunos resultados son los siguientes:

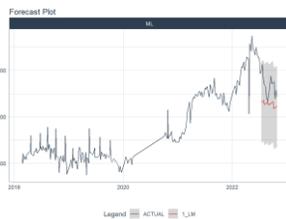
- Modelos regresión lineal múltiple

Ilustración 16

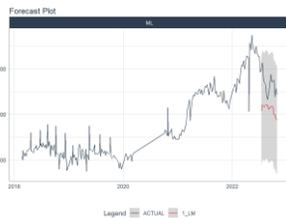
Ilustración 16 Modelos regresión lineal múltiple



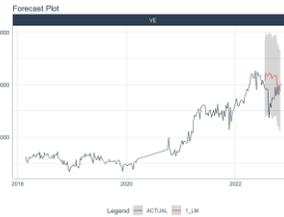
Modelo1_VE



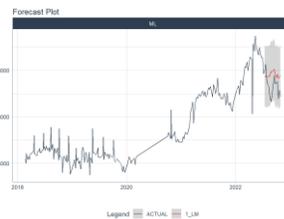
Modelo1_ML



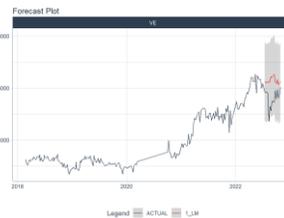
Modelo3_ML



Modelo2_VE



Modelo2_ML

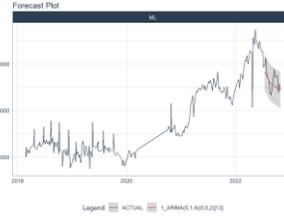


Modelo3_VE

- Modelos ARIMA

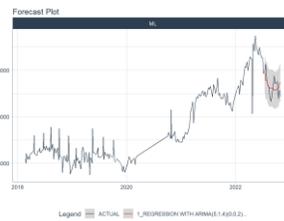
Ilustración 17

Ilustración 17 Modelos ARIMA



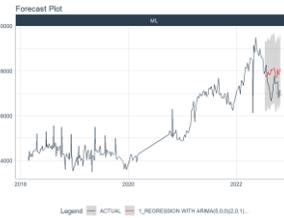
Modelo1ARIMA_VE

Modelo1ARIMA_ML



Modelo2ARIMA_VE

Modelo2ARIMA_ML



Modelo3ARIMA_VE

Modelo3ARIMA_ML

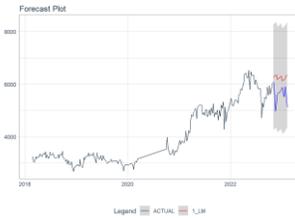
Pronósticos con nuevos datos

Se volvió hacer una extracción por medio de la técnica web scraping para comparar los datos reales de las subastas hasta el presente año y compararlos con los pronósticos hechos por los modelos, estos fueron los resultados:

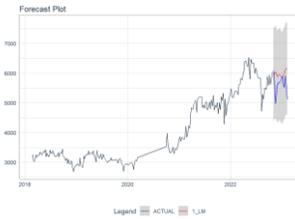
- Modelos de regresión lineal múltiples

Ilustración 18

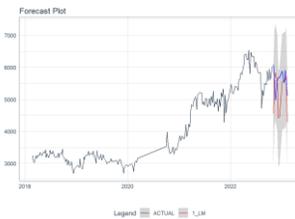
Ilustración 18 Predicción Modelos de regresión lineal múltiples



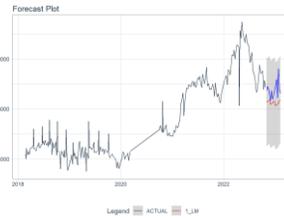
Pre_Modelo1_VE



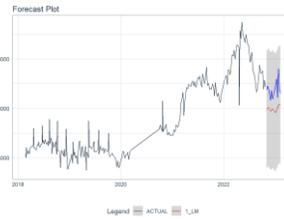
Pre_Modelo2_VE



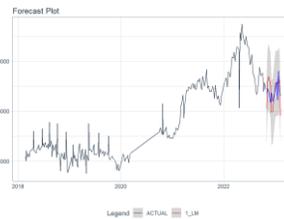
Pre_Modelo3_VE



Pre_Modelo1_ML



Pre_Modelo2_ML

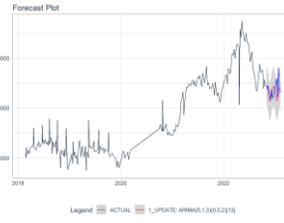
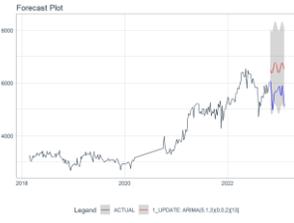


Pre_Modelo3_ML

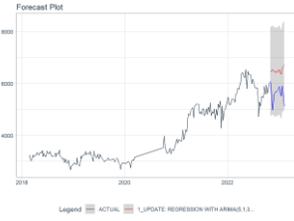
- Modelos ARIMA

Ilustración 19

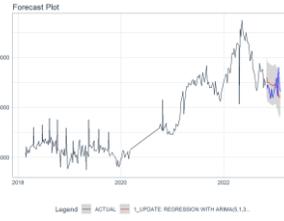
Ilustración 19 Predicción Modelos ARIMA



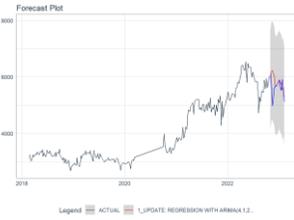
Pre_Modelo1ARIMA_VE



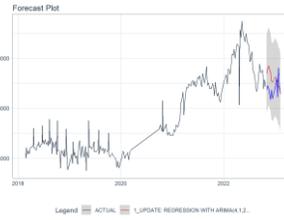
Pre_Modelo1ARIMA_ML



Pre_Modelo2ARIMA_VE



Pre_Modelo2ARIMA_ML



Pre_Modelo3ARIMA_VE

Pre_Modelo3ARIMA_ML

Conclusiones

Aplicar web scraping para la extracción de información libre y accesible te ahorra tiempo para implementarlo en otra actividad.

Comprender tus datos es la base fundamental para desarrollar tus objetivos.

El rigor científico, las bases matemáticas, estadísticas y el uso de herramientas tecnológicas si son de utilidad para entender los fenómenos que ocurren en las subastas ganaderas.

Los precios bases dados por los martillos en la subasta CENCOGAN van acordes a las tendencias del mercado, mostrando influencia por la temporada del año, la categoría del animal, la TRM, precio del maíz.

La subasta CENCOGAN también fue afectada por la Pandemia COVID-19, y por ello el cambio en el comportamiento de los precios bases.

Tener diferentes modelos de predicción nos ayuda a comprender que no siempre tener muchas variables los hace más preciso, “el poder de lo simple”.

Los modelos de predicción no son el espejo del futuro, pero si nos permiten disminuir la incertidumbre con sus pronósticos.

Como trabajos al futuro, se pueden construir otros modelos e involucrar otras variables para brindar una herramienta de evaluación del mercado y bajar el error en la modelación de la categoría de vacas paridas para beneficio de los participantes de las subastas ganaderas.

Referencias

- ANDRADE CACERES, WILLIAM ALEXIS. 2017. «IMPLEMENTACIÓN DE UN SISTEMA DE SUBASTAS ONLINE EN EL COMITÉ REGIONAL DE GANADEROS DEL MUNICIPIO DE TAME DEPARTAMENTO DE ARAUCA DESARROLLADO BAJO LA METODOLOGÍA DRA.» Tesis doctoral, UNIVERSIDAD COOPERATIVA DE COLOMBIA.
- Bianco, Vincenzo, Oronzio Manca, Sergio Nardini, y Alina A. Minea. 2010. «Analysis and forecasting of nonresidential electricity consumption in Romania». *Applied Energy* 87 (11): 3584-90. <https://doi.org/https://doi.org/10.1016/j.apenergy.2010.05.018>.
- Bolte, K., Kevin C. Dhuyvetter, y Ted C. Schroeder. 2008. «Information needs regarding the national animal identification system in the livestock auction market industry». *Kansas Agricultural Experiment Station Research Reports*, n.º 1 (enero): 33-36. <https://doi.org/10.4148/2378-5977.1511>.
- Braeken, Johan, y Marcel A. L. M. van Assen. 2017. «An empirical Kaiser criterion.» *Psychological Methods* 22: 450-66. <https://doi.org/10.1037/met0000074>.
- Burkart, Stefan, Manuel Francisco Díaz, Karen Enciso-Valencia, José Luis Urrea Benitez, Andrés Charry-Camacho, y Natalia Triana Ángel. 2020. «COVID-19 y el sector ganadero bovino en Colombia: Desarrollos actuales y potenciales, impactos y opciones de mitigación». Working {Paper}. International Center for Tropical Agriculture. <https://cgspace.cgiar.org/handle/10568/108370>.
- Castillo N, Omar. 2007. «Estacionalidad, ciclos y volatilidad en los precios del ganado macho de levante en Montería, Colombia». *Revista MVZ Córdoba* 12 (1). <https://doi.org/10.21897/rmvz.434>.
- CASTILLO NUNEZ, Omar Enrique. 2022. «Determinantes de la oferta primaria de ganado vacuno para ceba en el departamento de Córdoba, Colombia.» *Lecturas de Economía*, junio, 279-314.
- Castillo Nuñez, Omar Enrique. 2008. «COMPORTAMIENTO DE LOS PRECIOS DEL GANADO HEMBRA DE LEVANTE DE PRIMERA CLASE EN MONTERÍA Y SINCELEJO (COLOMBIA)». *Revista Facultad de Ciencias Económicas: Investigación y Reflexión* 16 (2): 105-16. <http://www.scielo.org.co/pdf/rfce/v16n2/v16n2a07.pdf>.
- Castillo, Omar E. 2007. «Comportamiento temporal de los precios del ganado macho de levante de primera en Sincelejo». *Temas agrarios* 12 (1): 22-36. <https://dialnet.unirioja.es/servlet/articulo?codigo=5002427>.
- CENCOGAN. s. f. *Compañía*. Buenavista, Colombia: Centro Comercial Ganadero SAS. <https://cencogan.com/compania/>.

- Denogean Ballesteros, Salomón, Francisco G. Moreno Medina. 2004. «Comportamiento a través del año de la comercialización de vacas de desecho en el Estado de Sonora». *Revista Mexicana de Agronegocios*.
<https://www.redalyc.org/articulo.oa?id=14101506> .
- Durá Juez, Pedro. 2003. *Teoría de Subastas y Reputación del Vendedor*.
<http://dx.doi.org/10.2139/ssrn.3364558>.
- Federacion Colombiana de Ganaderos, FEDEGAN. *Precios*. Federacion Colombiana de Ganaderos, FEDEGAN. <https://www.fedegan.org.co/estadisticas/precios>.
- Graham, Ian. 1998. «The Construction of a Network Technology: Electronic Livestock Auction Markets». *International Journal of Innovation Management* 02 (02): 183-99. <https://doi.org/10.1142/S1363919698000092>.
- Grajales Cedeño, Joseph Kaled, Javier Alexander Bethancourt García, Guadalupe Corrales, Enrique A. Sánchez-Galán, Reynaldo Vargas, Efraín R. Quintero Chanis, José Gallardo, y Jeffrey Geider. 2021. «Factores que intervienen en la variación del precio de comercialización de bovinos en subastas ganaderas de Panamá (2016 – 2020)». *Zootecnia Tropical* 39 (e5545844).
<https://doi.org/10.5281/zenodo.5545844>.
- Hyndman, R. J., y G Athanasopoulos. 2023. *Forecasting: Principles and Practice (3rd ed)*.
<https://otexts.com/fpp3/index.html>.
- Hyndman, Rob J, y George Athanasopoulos. 2021. *Forecasting: Principles and Practice*. 3rd ed. Melbourne, Australia: OTexts. <https://otexts.com/fpp3/>.
- Instituto Colombiano Agropecuario ICA. s. f. *Censos Pecuarios Nacional*. Bogota, Colombia: Instituto Colombiano Agropecuario. Accedido 17 de diciembre de 2022. <https://www.ica.gov.co/areas/pecuaria/servicios/epidemiologia-veterinaria/censos-2016/censo-2018>.
- «Las subastas ganaderas». 2014. ANEIA - Universidad de Los Andes. septiembre de 2014. <https://agronegocios.uniandes.edu.co/2014/09/las-subastas-ganaderas/>.
- McCabe, E. D., M. E. King, K. E. Fike, K. L. Hill, G. M. Rogers, y K. G. Odde. 2019. «Breed composition affects the sale price of beef steer and heifer calves sold through video auctions from 2010 through 2016». *Applied Animal Science* 35 (2): 221-26. <https://doi.org/https://doi.org/10.15232/aas.2018-01806>.
- Montaño García, Juan Carlos. 2018. «Introducción a las subastas y sus aplicaciones prácticas». Info:eu-repo/semantics/{bachelorThesis}.
<http://uvadoc.uva.es/handle/10324/34336>.
- Murillo, Danny, y Dalys Saavedra. 2017. «Web Scraping de los Perfiles y Publicaciones de una Afiliación en Google Scholar utilizando Aplicaciones Web e implementando un Algoritmo en R». *Memorias de Congresos UTP*, julio, 8-15.
<https://revistas.utp.ac.pa/index.php/memoutp/article/view/1465>.

- R Core Team. *Regresión lineal múltiple*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sarmiento Ortiz, Jonathan Efreem. 2016. «Prototipo de Aplicación Móvil para Subastas Ganaderas». Tesis doctoral, UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS. <http://repository.udistrital.edu.co/handle/11349/8271>.
- Smith, M. J., E. D. McCabe, K. E. Fike, M. E. King, G. M. Rogers, T. E. Wittum, y K. G. Odde. 2021. «Factors affecting the sale price of bred heifers and bred beef cows sold through video auctions». *Applied Animal Science* 37 (2): 217-24. <https://doi.org/https://doi.org/10.15232/aas.2020-02066>.
- Stellman, Jeanne Mager. 1998. *Enciclopedia de Salud y Seguridad en el Trabajo. Vol. 3*. Vol. 3. Madrid: Ministerio de Trabajo y Asuntos Sociales.
- Superintendencia financiera de Colombia. *Tasa de Cambio Representativa del Mercado - Historico*. Superintendencia financiera de Colombia. <https://www.datos.gov.co/Econom-a-y-Finanzas/Tasa-de-Cambio-Representativa-del-Mercado-Historic/mcec-87by>.
- Taherdoost, Hamed, Shamsul Sahibuddin, y Neda Jalaliyoon. s. f. «Exploratory Factor Analysis; Concepts and Theory».
- Ultsch, Alfred, y Jörn Lötsch. 2015. «Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data». *PLOS ONE* 10 (6): e0129767. <https://doi.org/10.1371/journal.pone.0129767>.