

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Modelo predictivo de desembolsos de libranzas basado en datos operativos del Banco de Bogotá

Trabajo de Grado Modalidad Pasantía

Autor: Camilo Andres Hernandez Riveros
Director: Luis Alejandro Másmela Caita
Profesional designado: Lina Paola Bocanegra Chaparro

Bogotá, DC
Septiembre de 2025

Resumen

El presente trabajo desarrolla una serie de modelos *SARIMA* orientados a la detección temprana de alertas y a la predicción de los desembolsos en libranzas correspondientes a los principales convenios del Banco de Bogotá. Esta iniciativa busca fortalecer la toma de decisiones en el área de Impulso Comercial, al permitir anticiparse a comportamientos atípicos o caídas significativas en los flujos de desembolsos. Para ello, se lleva a cabo un proceso de depuración y análisis temporal de los datos, teniendo en cuenta la estacionalidad y los patrones de comportamiento de cada convenio. Los modelos construidos permiten generar pronósticos ajustados que constituyen un insumo clave para la planeación comercial y la optimización de estrategias operativas futuras.

Palabras Clave

Modelo *SARIMA*, Libranza, Estacionariedad

Agradecimientos

A mi madre, una mujer humilde, valiente, fuerte y amorosa, quien con sus lecciones de vida forjó la persona que soy hoy. Sé que, con su amor, me bendijo, protegió y cuidó durante todo este camino.

A mi padre, quien me brindó su apoyo incondicional y me impulsó a superarme día tras día, presionándome con cariño para alcanzar siempre una mejor versión de mí mismo.

Al profesor Alejandro Másmela, quien desde los primeros semestres fue un referente para mí en el campo de la matemática aplicada, y que como docente director me guió con paciencia en un proceso que al inicio parecía oscuro, pero en el que, gracias a su orientación, pude encontrar claridad y propósito.

A mis profesores del Proyecto Curricular de Matemáticas, porque cada uno aportó un valioso granito de arena semestre a semestre, entregándome herramientas y conocimientos que hoy valoro profundamente.

A mi grupo de amigos de la carrera, quienes más allá de ser ejemplos a seguir y fuente constante de motivación, fueron compañía y apoyo en cada etapa: entre parciales, quices, talleres y los interminables huecos de la universidad.

Gracias a todos ustedes, porque al mirar el camino recorrido, reconozco que fueron un pilar fundamental en todo lo que se construyó a lo largo de estos años.

Índice

1. Introducción **4**

2. Objetivos **4**

2.1. Objetivo General 4

2.2. Objetivos Específicos 4

3. Marco Teórico **5**

3.1. Banco de Bogotá 5

3.1.1. Historia y Fundación 5

3.1.2. Evolución y Fusión 6

3.2. Fundamentos teóricos 6

3.2.1. Series de Tiempo 6

3.2.2. Modelos 7

3.2.3. Estacionariedad 8

3.2.4. Prueba de Dickey-Fuller Aumentada (ADF) 9

3.2.5. Método Box-Jenkins 10

3.3. Contextualización de los datos 11

3.3.1. Base de datos 12

3.3.2. Análisis y limpieza de datos 13

4. Creación y Desarrollo de los Modelos **13**

4.1. Análisis preliminar y tratamiento de valores atípicos 14

4.2. Identificación de parámetros del modelo $SARIMA(p,d,q)(P,D,Q)_S$ 15

4.2.1. Orden de diferenciación 16

4.2.2. Orden de Autoregresión 16

4.2.3. Orden de Media Móvil 17

4.3. Búsqueda de los modelos óptimos 18

5. Análisis de resultados de los modelos *SARIMA* **20**

5.1. Policía Nacional 20

5.2. Colpensiones 20

6. Conclusiones **21**

7. Referencias **22**

1. Introducción

El área de Impulso Comercial del Banco de Bogotá, responsable de la gestión y seguimiento de las libranzas ofrecidas a través de diversos convenios con entidades y empresas, ha operado históricamente mediante metodologías y herramientas de análisis tradicionales. Aunque estos métodos han permitido un manejo adecuado de la información, la creciente complejidad del mercado, la variabilidad en el comportamiento de los clientes y la necesidad de respuestas ágiles han puesto de manifiesto la importancia de incorporar modelos estadísticos más avanzados que fortalezcan la toma de decisiones.

En este contexto, el presente trabajo de grado propone la implementación de modelos de series temporales, específicamente *ARIMA* y *SARIMA*, para proyectar los desembolsos de libranzas asociadas a los distintos convenios vigentes del banco. Estos modelos, ampliamente reconocidos por su capacidad para capturar patrones estacionales y tendencias en datos históricos, se constituyen en un pilar fundamental para anticipar comportamientos futuros, optimizar la planeación operativa y facilitar el diseño de estrategias comerciales más efectivas.

La aplicación de estas técnicas representa un cambio significativo en el enfoque analítico del área, que tradicionalmente no ha contado con herramientas predictivas de este nivel. Al proporcionar estimaciones más precisas sobre el comportamiento de cada convenio, el banco puede anticiparse a fluctuaciones en la demanda, asignar recursos de manera más eficiente y diseñar planes de acción acordes con la realidad proyectada. De este modo, el desarrollo de este proyecto no solo responde a una necesidad interna, sino que también se alinea con las tendencias globales del sector financiero, en las cuales la analítica avanzada se consolida como un elemento clave para la competitividad y sostenibilidad de las instituciones.

2. Objetivos

2.1. Objetivo General

Realizar una recopilación de datos históricos de créditos de libranza del Banco de Bogotá y efectuar un análisis de su comportamiento, organizándolos como series de tiempo. Este estudio considera aspectos relevantes con el propósito de construir modelos *SARIMA* que permitan pronosticar el comportamiento de los desembolsos en libranzas para días futuros.

2.2. Objetivos Específicos

1. **Recopilación de información**: Reunir la información histórica de libranzas disponible en el área de Impulso Comercial, con el propósito de consolidar una base de datos que permita

almacenar de manera organizada tanto los registros pasados como los futuros.

2. **Construcción de modelos**: Seleccionar, entre diferentes enfoques como *ARIMA*, *SARIMA* o redes neuronales, aquellos modelos que mejor se ajusten a los datos y que posibiliten obtener predicciones con el menor margen de error.

3. **Cualidades matemáticas**: Analizar, desde un enfoque matemático y estadístico, el comportamiento de los modelos construidos, así como el proceso metodológico que conlleva su desarrollo.

4. **Innovación**: Incorporar una herramienta poco explorada en el área de Impulso Comercial del Banco de Bogotá, mediante el uso de programas como *Python*, que permitan desarrollar procesos predictivos y generar resultados orientados al fortalecimiento y optimización del área.

5. **Soporte para la toma de decisiones estratégicas**: Proporcionar resultados confiables que sirvan de apoyo a los especialistas del área, facilitando el desarrollo de sus labores y respaldando la toma de decisiones de gran relevancia para su gestión.

3. Marco Teórico

3.1. Banco de Bogotá

El Banco de Bogotá es una de las principales entidades financieras de Colombia y forma parte del Grupo Aval, el conglomerado financiero más grande del país. Con presencia en 12 naciones y una amplia red de atención conformada por 1.477 oficinas, 3.375 cajeros automáticos y 10.093 corresponsales bancarios, la entidad ofrece servicios tanto a nivel nacional como internacional. Su modelo de Banca Universal se orienta a interpretar y atender las necesidades de diversos segmentos de clientes, promoviendo la inclusión financiera y desarrollando soluciones innovadoras a través de canales presenciales y digitales.[2]

3.1.1. Historia y Fundación

Fundado el 15 de noviembre de 1870, el Banco de Bogotá se convirtió en la primera institución financiera establecida en Colombia. Desde sus inicios, ha desempeñado un papel fundamental en el progreso económico del país, apoyando a empresarios, financiando grandes obras y facilitando el acceso a servicios bancarios. A lo largo de más de 145 años, la entidad ha evolucionado y expandido sus operaciones más allá de las fronteras nacionales, consolidándose como un referente del sector

financiero en América Latina y el Caribe, con presencia en países como Panamá, Costa Rica, Nicaragua, Honduras, El Salvador, Guatemala, México, Estados Unidos, Bahamas, Islas Caimán y Nassau.[2]

3.1.2. Evolución y Fusión

La historia del Banco de Bogotá es, en muchos sentidos, también la historia de Colombia: una narrativa de desafíos superados, de constante evolución y de un compromiso inquebrantable con el progreso, dentro de la cual resaltan acontecimientos como los siguientes:[3]

- **1870:** El Banco de Bogotá inicia operaciones como la primera institución financiera creada en el país, con la facultad de emitir billetes.
- **1967:** Se establece una oficina en Panamá, convirtiéndose en el primer banco colombiano en operar en el exterior.
- **1974:** Se abre una oficina en Nueva York, la cual posteriormente sería reconocida como *First Bank of the Americas*.
- **1987:** El Banco pasa a formar parte del grupo de empresas propiedad de la Organización Luis Carlos Sarmiento Angulo.
- **1991:** Se adquiere el 76 % del Banco del Comercio y, al año siguiente, se concreta la fusión, incorporando nuevas filiales al grupo.
- **2010:** Se adquiere *BAC Credomatic Inc.*, un conglomerado centroamericano que provee servicios financieros a través de sus filiales en distintos países de Centroamérica.

3.2. Fundamentos teóricos

3.2.1. Series de Tiempo

Una serie de tiempo es un conjunto de datos estadísticos recopilados en intervalos regulares (diarios, mensuales, anuales, entre otros), cuyo propósito es analizar cómo los valores pasados influyen en el comportamiento futuro. Su estudio resulta fundamental en diversas áreas, como la economía, la meteorología y la ingeniería.[5]

Sus observaciones pueden ser denotadas como:

$$Y_1, Y_2, \ldots, Y_{t-1}, Y_t$$

donde Y_t es el valor tomado por el proceso en el instante t con $t \in \mathbb{N}$.

Además, una serie de tiempo puede descomponerse en los siguientes componentes teóricos:

- **Tendencia:** Movimiento de largo plazo de la serie de tiempo, caracterizado por la ausencia de efectos de calendario y de regularidad periódica.
- **Estacionalidad:** Fluctuaciones que se repiten con periodicidad fija, generalmente asociadas a patrones del calendario (estaciones del año, meses o días de la semana), las cuales influyen de manera sistemática en la serie.
- **Componente cíclica:** Conjunto de fluctuaciones de mediano o largo plazo (por lo general, superiores a un año), originadas por variaciones en condiciones económicas, sociales o naturales.
- **Variaciones irregulares:** También llamadas ruido o variación aleatoria; agrupan movimientos impredecibles que no pueden explicarse mediante la tendencia, la estacionalidad o el ciclo, y suelen modelarse como un proceso estocástico de media cero.[5]

3.2.2. Modelos

Modelo autorregresivo de orden *p*
Denotado como *AR(p)*, es un proceso estocástico $\{Y_t\}_{t\in\mathbb{Z}}$ en el cual la observación en el tiempo t se expresa como una combinación lineal de un número finito de sus valores pasados más un término aleatorio. Se define como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t,$$

donde:

- Y_t es la variable dependiente en el tiempo t,
- $\phi_1, \phi_2, \ldots, \phi_p \in \mathbb{R}$ son los coeficientes autorregresivos,
- ε_t es un término de error o innovación, comúnmente supuesto como un ruido blanco gaussiano:

$$\varepsilon_t \sim \mathcal{N}(0, \sigma^2), \quad \mathbb{E}[\varepsilon_t] = 0, \quad \mathrm{Var}(\varepsilon_t) = \sigma^2, \quad \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0 \text{ para } t \neq s.$$

En este modelo, las variables explicativas corresponden a los rezagos $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}$, lo que refleja el carácter autorregresivo del proceso.

Modelo de media móvil de orden *q*
Denotado como *MA(q)*, es un proceso estocástico $\{Y_t\}_{t\in\mathbb{Z}}$ en el cual la observación en el tiempo t se expresa como una combinación lineal finita entre el término de error actual y sus q errores pasados. Se define como:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}, \quad t \in \mathbb{Z},$$

donde:

- $\theta_1, \theta_2, \ldots, \theta_q \in \mathbb{R}$ son los coeficientes de medias móviles,
- $\{\varepsilon_t\}$ es un proceso de ruido blanco, comúnmente supuesto como gaussiano, tal como se definió en el modelo anterior.

Modelo autorregresivo de media móvil de orden *(p,q)*
Denotado como *ARMA(p,q)*, es un proceso estocástico $\{Y_t\}_{t\in\mathbb{Z}}$ que combina componentes autorregresivas y de medias móviles. Se define como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q},$$

donde los coeficientes ϕ_t y θ_t, así como el proceso de ruido blanco ϵ_t para $t \in \mathbb{Z}$, ya han sido definidos en los modelos anteriores.

Modelo autorregresivo integrado con media movil de orden *(p,d,q)*
Denotado como *ARIMA(p,d,q)*, es un proceso estocástico en el cual la serie $\{Y_t\}_{t\in\mathbb{Z}}$, diferenciada *(d)* veces para inducir estacionariedad, puede representarse como una combinación lineal de sus p valores pasados (componente autorregresiva) y de los q errores pasados (componente de medias móviles).

El modelo *ARIMA* se expresa como

$$Y_t^{(d)} = \phi_1 Y_{t-1}^{(d)} + \phi_2 Y_{t-2}^{(d)} + \cdots + \phi_p Y_{t-p}^{(d)} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q},$$

donde los coeficientes ϕ_t y θ_t, así como el proceso de ruido blanco ϵ_t para $t \in \mathbb{Z}$, ya han sido definidos en los modelos anteriores. [6]

3.2.3. Estacionariedad

Sea $\{Y_t\}_{t\in\mathbb{Z}}$ un proceso estocástico. Este puede ser estacionario en sentido estricto o presentar estacionariedad débil, condición fundamental en la construcción de modelos *ARIMA* o *SARIMA*. A continuación, se expone el significado de cumplir con cada una de estas formas de estacionariedad.

Estacionariedad estricta. El proceso $\{Y_t\}$ se dice estrictamente estacionario si, para todo entero $k \geq 1$, para todo conjunto de instantes $t_1, t_2, \ldots, t_k \in \mathbb{Z}$, y para todo desplazamiento temporal $h \in \mathbb{Z}$, se cumple que:

$$F_{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_k}}(y_1, y_2, \ldots, y_k) = F_{Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_k+h}}(y_1, y_2, \ldots, y_k),$$

donde F denota la función de distribución conjunta. Es decir, la distribución conjunta de cualquier subconjunto finito de la serie es invariante ante traslaciones en el tiempo.

Estacionariedad débil o en sentido amplio. El proceso $\{Y_t\}$ se dice estacionario en sentido débil si cumple:

- **Media constante:**
$$\mathbb{E}[Y_t] = \mu, \quad \forall t \in \mathbb{Z}.$$
- **Varianza constante y finita:**
$$\mathrm{Var}(Y_t) = \sigma^2 < \infty, \quad \forall t \in \mathbb{Z}.$$
- **Autocovarianza dependiente solo del rezago:**
$$\gamma(s,t) = \mathrm{Cov}(Y_s, Y_t) = \gamma(t-s), \quad \forall s,t \in \mathbb{Z}.$$

En este caso, la dependencia entre dos observaciones depende únicamente de la distancia temporal entre ellas, y no del instante absoluto en el que ocurren. [6]

3.2.4. Prueba de Dickey-Fuller Aumentada (*ADF*)

La prueba de Dickey-Fuller Aumentada (*ADF*) se utiliza para determinar si una serie de tiempo es estacionaria o no, aspecto de vital importancia en la construcción de modelos *ARIMA* y *SARIMA*. Dicha prueba se formula a partir de las siguientes hipótesis:

- H_0: La serie de tiempo presenta una raíz unitaria, por lo que no es estacionaria.
- H_1: La serie de tiempo no presenta una raíz unitaria, por lo que se considera estacionaria.

Tomando en cuenta que el **estadístico de prueba** funciona como un indicador de la fuerza de la evidencia respecto a la presencia o ausencia de una raíz unitaria en la serie de tiempo, su interpretación se basa en el análisis de los rezagos que la componen.

Condición para rechazar H_0
Si, al comparar el estadístico de prueba con los valores críticos de la prueba de *Dickey-Fuller* (al 1 %, 5 % y 10 %), este resulta menor y, adicionalmente, el *p-valor* es inferior a 0,05 (valor comúnmente adoptado), se rechaza la hipótesis nula H_0.

Por el contrario, si el estadístico de prueba es mayor que alguno de los valores críticos, no se rechaza la hipótesis nula y, en consecuencia, la serie de tiempo se considera no estacionaria en el nivel de significancia correspondiente.[4]

3.2.5. Método Box-Jenkins

El *método Box-Jenkins* es una metodología desarrollada para identificar y determinar el modelo probabilístico que mejor represente el comportamiento de un proceso a lo largo del tiempo, considerando tanto series estacionarias como no estacionarias. En este último caso, se recurre a la diferenciación para inducir estacionariedad. Asimismo, se asume que los errores aleatorios (ε_t) son independientes, tienen media cero y presentan varianza constante.

El procedimiento Box-Jenkins se organiza en cinco etapas fundamentales:

1. Estacionariedad
2. Identificación del modelo
3. Estimación de parámetros
4. Evaluación del ajuste
5. Pronóstico

Gracias a su carácter sistemático, el método Box-Jenkins se ha consolidado como una de las herramientas más utilizadas en la modelación de series temporales.[6]

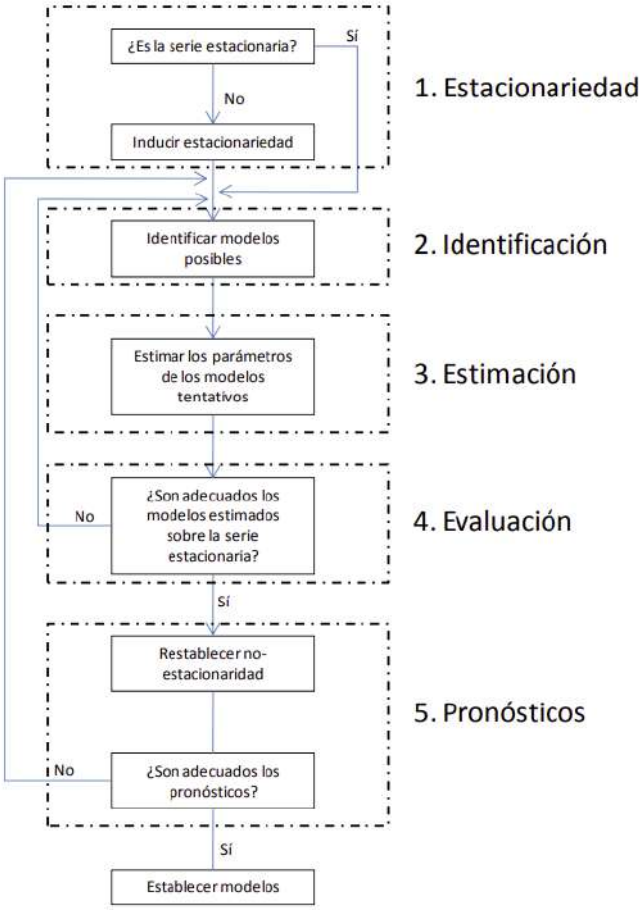


Figura 1: Diagrama de flujo del método **Box-Jenkins**, detalla las cinco etapas del procedimiento para identificar y determinar el modelo probabilístico. El esquema ilustra la secuencia del proceso y la retroalimentación necesaria cuando los modelos no cumplen con los criterios de adecuación.

3.3. Contextualización de los datos

En esta sección se presentan los recursos y la información empleados como base para la elaboración del presente trabajo. Del mismo modo, se describe de manera general el proceso inicial de construcción de los modelos de interés que constituyen el objeto de estudio.

Con el propósito de contextualizar el análisis, resulta pertinente precisar la definición de un crédito en libranza:

Crédito en libranza
Es un mecanismo de financiación mediante el cual una entidad financiera otorga un crédito a un

trabajador activo, pensionado o contratista, cuyo pago se garantiza a través de descuentos directos y periódicos sobre su salario, mesada pensional o ingresos contractuales.

Este esquema opera a partir de un *convenio de libranza* suscrito entre la entidad financiera y el empleador, fondo de pensiones o pagador autorizado, quien actúa como agente de retención. Dicho agente se encarga de realizar los descuentos correspondientes a las cuotas del crédito y transferirlos directamente a la entidad acreedora.[7]

3.3.1. Base de datos

El Banco de Bogotá dispone de información limitada en el área de libranzas, dado que los registros diarios no se almacenan de manera sistemática y, en muchos casos, son eliminados. Como resultado, únicamente se contó con 243 observaciones en frecuencia diaria (días hábiles), correspondientes al período comprendido entre junio de 2024 y mayo de 2025. Es importante señalar que los datos fueron modificados de manera leve con el fin de preservar la confidencialidad de la entidad.

Organización de la información La información se encontraba consolidada en informes acumulativos en formato Excel. En consecuencia, fue necesario llevar a cabo un proceso de extracción de los registros diarios a partir de los informes de cierre correspondientes a cada mes.

Estructura de los datos Las libranzas están asociadas a convenios suscritos entre las entidades empleadoras o de pensiones de los clientes y el Banco de Bogotá. En este sentido, la información puede organizarse de dos maneras principales:

1. Por región, zona, distrito u oficina en la cual se efectuó la apertura de la libranza.
2. Por convenio establecido entre la entidad empleadora (o de pensiones) y el banco.

En el presente trabajo se optó por la segunda alternativa, dado que la primera presentaba dificultades significativas: los cambios estructurales en la organización de los datos hacían inviable realizar un cruce consistente entre la información histórica y la futura, lo cual habría implicado un proceso excesivamente complejo en la etapa de preparación.

Convenios de interés El Banco de Bogotá dispone de una amplia variedad de convenios; sin embargo, dos de ellos resultan particularmente relevantes: **Policía Nacional** y **Colpensiones**. Estos concentran la mayor proporción de desembolsos de libranzas, razón por la cual se decidió construir modelos predictivos independientes para cada uno.

3.3.2. Análisis y limpieza de datos

Se llevó a cabo una consolidación de todos los registros históricos por convenio, a partir de la cual se organizaron los valores en frecuencia diaria. Posteriormente, se realizaron análisis de estacionalidad, tendencia y patrones de comportamiento en determinados periodos, los cuales, con el apoyo de los especialistas del área, permitieron identificar las causas subyacentes de dichos fenómenos.

Asimismo, se eliminaron los valores iguales a cero, dado que afectaban el desempeño de los modelos en la estimación de métricas de error, como el $MAPE$ (que se estudiará más adelante). La presencia de estos valores se explicó, en el caso de la **Policía Nacional**, por días en los que las plataformas no estuvieron operativas, y en el caso de **Colpensiones**, por cierres no especificados en los datos o fallas en el funcionamiento de su sistema.

Finalmente, se aplicó un proceso de suavización de *outliers*, tras determinar que estos valores atípicos correspondían, en su mayoría, a errores manuales en el registro de información o a efectos transitorios ocasionados por fallas en la plataforma.

4. Creación y Desarrollo de los Modelos

Tras aplicar diversas pruebas con la metodología de Box-Jenkins, así como otros procedimientos complementarios, se determinó que los modelos que ofrecían un mejor ajuste a las series de tiempo correspondían a la familia *SARIMA*. En consecuencia, los modelos seleccionados para el análisis se enmarcan en distintas especificaciones de esta clase. Ahora bien, surge la pregunta: ¿qué es un modelo *SARIMA*? A continuación, se presenta su definición.

Modelo autorregresivo integrado de medias móviles estacional de orden *(p,d,q)*×*(P,D,Q,S)*
Denotado como $SARIMA(p,d,q)(P,D,Q)_S$, es un proceso estocástico en el cual la serie de tiempo $\{Y_t\}_{t\in\mathbb{Z}}$, diferenciada d veces y diferenciada estacionalmente D veces con periodo S para inducir estacionariedad, puede representarse como la combinación lineal de:

1. p valores pasados no estacionales (componente autorregresiva no estacional),
2. q errores pasados no estacionales (componente de medias móviles no estacional),
3. P valores pasados estacionales (componente autorregresiva estacional),
4. Q errores pasados estacionales (componente de medias móviles estacional).

El modelo se expresa como:

$$\Phi_P(Y_{t-s},\ldots,Y_{t-Ps}) + \phi_p(Y_{t-1},\ldots,Y_{t-p})\ \nabla^d\nabla_s^D Y_t = \Theta_Q(\varepsilon_{t-s},\ldots,\varepsilon_{t-Qs}) + \theta_q(\varepsilon_{t-1},\ldots,\varepsilon_{t-q}),$$

donde:

- $\phi_p(\cdot)$ representa la combinación lineal de los términos autorregresivos no estacionales,
- $\theta_q(\cdot)$ representa la combinación lineal de los términos de medias móviles no estacionales,
- $\Phi_P(\cdot)$ representa la combinación lineal de los términos autorregresivos estacionales,
- $\Theta_Q(\cdot)$ representa la combinación lineal de los términos de medias móviles estacionales,
- $\nabla^d\nabla_s^D$ denota las diferenciaciones ordinarias y estacionales aplicadas a la serie,
- $\{\varepsilon_t\}$ es un proceso de ruido blanco con media cero y varianza constante. [8]

Ejemplo. Modelo $SARIMA(1,0,1)(2,1,0)_{12}$
Tomando en cuenta que hay una diferencia estacional entonces:

$$y_t - y_{t-12} = W_t$$

Dado esto entonces

$$W_t = \mu + \phi_1 W_{t-1} + \phi_2 W_{t-12} + \phi_3 W_{t-24} + \theta_1\varepsilon_{t-1} + \varepsilon_t$$

4.1. Análisis preliminar y tratamiento de valores atípicos

A continuación, se presenta la serie de tiempo correspondiente a los desembolsos asociados al convenio con la **Policía Nacional**. Este convenio fue seleccionado como ejemplo, dado que, por razones de espacio, resulta más adecuado para su exposición y presenta menos dificultades en las etapas de análisis y depuración de datos descritas anteriormente. Cabe señalar que el procedimiento aplicado al convenio con **Colpensiones** fue análogo en todas sus fases.

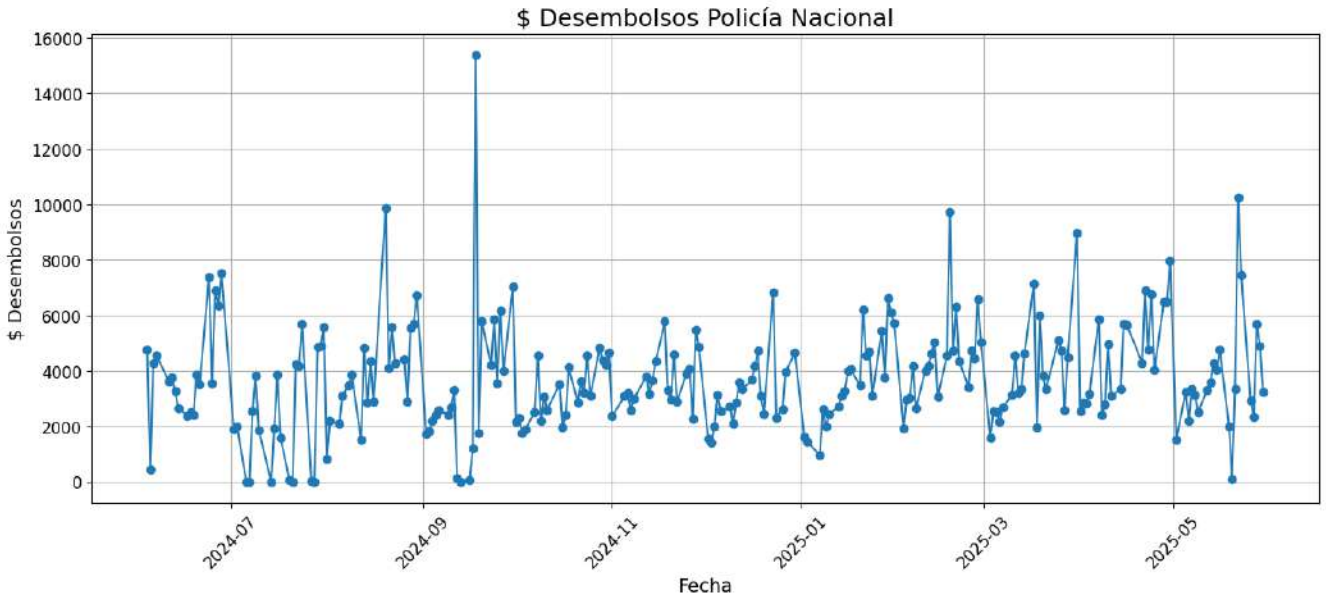


Figura 2: Serie de tiempo de los desembolsos correspondientes al convenio con la **Policía Nacional**. El eje horizontal muestra la *Fecha*, mientras que el eje vertical representa los *Montos desembolsados* en millones de pesos colombianos. El gráfico refleja la evolución diaria de los desembolsos durante el periodo analizado.

Se identificaron 5 registros con valor cero, asociados a fallas en el funcionamiento de la plataforma durante las fechas correspondientes. Asimismo, tal como se mencionó previamente, se detectaron 6 *outliers*, los cuales fueron suavizados mediante la aplicación de umbrales, con el propósito de mejorar el ajuste del modelo y garantizar una mayor precisión en las estimaciones.

Considerando que, para nuestro conjunto de valores $\{Y_t\}_{t\in\mathbb{Z}}$, un elemento Y_t se clasifica como un **valor atípico (*outlier*)** si cumple alguna de las siguientes condiciones:

$$Y_t < Q_1 - 1{,}5 \cdot IQR \qquad \text{o} \qquad Y_t > Q_3 + 1{,}5 \cdot IQR,$$

donde Q_1 y Q_3 representan el primer y tercer cuartil de $\{Y_t\}_{t\in\mathbb{Z}}$, respectivamente, y el *rango intercuartílico* se define como $IQR = Q_3 - Q_1$.

De este modo, se procedió a la estimación de los posibles parámetros para los modelos *SARIMA*.

4.2. Identificación de parámetros del modelo $SARIMA(p,d,q)(P,D,Q)_S$

Se procederá a la búsqueda de los parámetros previamente definidos, con el fin de identificar aquellos que resulten óptimos para la especificación de cada modelo *SARIMA*.

4.2.1. Orden de diferenciación

En un modelo *SARIMA*, la estacionariedad de la serie de tiempo constituye un requisito esencial. Por ello, el primer parámetro a determinar corresponde al orden de diferenciación necesario para que la serie satisfaga dicha condición.

Este orden se establece mediante la prueba de *Dickey-Fuller Aumentada*, definida anteriormente, cuyos resultados fueron obtenidos empleando la librería *statsmodels* de *Python*, ampliamente reconocida en el ámbito de la econometría y la modelación estadística.

Métrica	Valor
Estadístico ADF	−6,070
p-valor	0,000
Número de rezagos usados	9,000
Número de observaciones	228,000
Valor crítico (1%)	−3,459
Valor crítico (5%)	−2,874
Valor crítico (10%)	−2,575

Cuadro 1: Resultados de la prueba *ADF* de la **Policía Nacional**.

Métrica	Valor
Estadístico ADF	−7,443
p-valor	0,000
Número de rezagos usados	1,000
Número de observaciones	215,000
Valor crítico (1%)	−3,461
Valor crítico (5%)	−2,875
Valor crítico (10%)	−2,573

Cuadro 2: Resultados de la prueba *ADF* de **Colpensiones**.

Se observa que tanto el *p-valor* como el *estadístico ADF* resultan inferiores a los niveles de significancia establecidos. En consecuencia, se rechaza la hipótesis nula H_0 de presencia de raíz unitaria, lo que permite concluir que ambas series temporales son estacionarias y, por tanto, no requieren la aplicación de diferenciaciones adicionales. En este sentido, se determina que $d = 0$ y $D = 0$, dado que también se evidencia estacionariedad en el componente estacional de las series.

4.2.2. Orden de Autoregresión

Dado que el modelo *SARIMA* incorpora dependencias entre valores pasados de la serie de tiempo *(rezagos)*, es necesario determinar cuántos de estos valores previos resultan significativamente relevantes para explicar el valor actual. Identificar un número adecuado de rezagos permite mejorar la capacidad predictiva del modelo y, al mismo tiempo, evitar tanto el subajuste como el sobreajuste.

El gráfico de *Autocorrelación Parcial (PACF)* constituye una herramienta clave para estimar el número máximo de rezagos a considerar. En particular, aquellos coeficientes que se ubican fuera del intervalo de confianza y tras los cuales se observa un corte abrupto, indican el número de rezagos potencialmente significativos. A continuación, se presenta el gráfico de *PACF* correspondiente a la serie analizada:

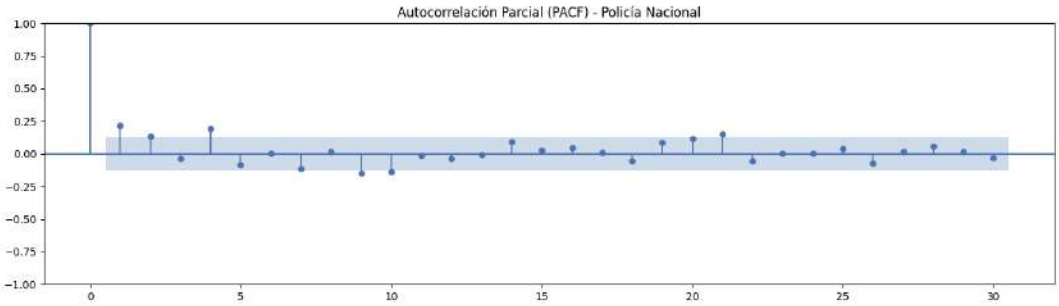


Figura 3: Función de autocorrelación parcial (**PACF**) de los desembolsos del convenio con la **Policía Nacional**. Los rezagos se representan en el eje horizontal, mientras que en el eje vertical se ilustran los coeficientes de autocorrelación parcial. El gráfico permite identificar qué rezagos presentan una relación estadísticamente significativa con los desembolsos analizados.

En el gráfico se observa que los tres primeros rezagos presentan valores significativos y que, a partir del tercero, los coeficientes descienden de manera abrupta hacia cero. Este comportamiento sugiere que los posibles valores del parámetro p se encuentran entre 1, 2 o 3. Asimismo, se identifica un rezago aislado significativo alrededor del rezago 20, lo cual constituye un indicio preliminar de que el parámetro estacional P podría tomar el valor de 1.

4.2.3. Orden de Media Móvil

El modelo *SARIMA* no solo incorpora la dependencia respecto a los valores pasados de la serie, sino que también considera la influencia de los errores cometidos en rezagos anteriores. Por esta razón, resulta necesario identificar el número de rezagos significativos de los errores pasados, con el objetivo de estimar un valor apropiado para el parámetro de media móvil del modelo.

El gráfico de *Autocorrelación (ACF)* ofrece una referencia visual para determinar dichos rezagos significativos. El procedimiento de identificación se lleva a cabo de manera análoga al utilizado para establecer los posibles valores del orden autorregresivo. A continuación, se presenta el gráfico de *Autocorrelación (ACF)*:

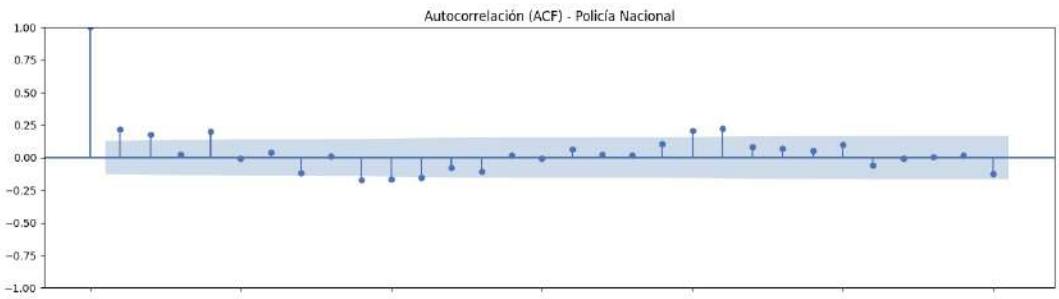


Figura 4: Función de autocorrelación (**ACF**) de los desembolsos del convenio con la **Policía Nacional**. Los rezagos se representan en el eje horizontal, mientras que en el eje vertical se ilustran los coeficientes de autocorrelación. El gráfico permite identificar qué rezagos presentan una relación estadísticamente significativa con los desembolsos analizados.

De manera análoga a lo observado con el parámetro p, en la Figura 4 se aprecia que los tres primeros rezagos son significativos, ya que, a partir del tercero, la autocorrelación desciende bruscamente hacia cero. En consecuencia, los posibles valores del parámetro q se sitúan en el rango de 1 a 3.

Asimismo, se observa que después del rezago 20 emergen dos valores significativos adicionales, lo cual constituye un indicio de la presencia de efectos estacionales. Por tanto, los valores plausibles para el parámetro estacional Q serían 1 o 2.

4.3. Búsqueda de los modelos óptimos

En este punto se aborda la cuarta etapa del método de *Box-Jenkins*, correspondiente a la evaluación de la capacidad de adaptación de distintos modelos a los datos. Este procedimiento se llevó a cabo en *Python* y, en el Cuadro 3, se presentan los cinco modelos que ofrecieron un mejor desempeño.

Modelo	MAE	RMSE	MAPE (%)
$SARIMA(2,0,2)(1,0,1,5)$	1877,810	2191,482	323,986
$SARIMA(1,0,1)(1,0,1,5)$	1840,664	2151,596	333,524
$SARIMA(2,0,1)(1,0,1,20)$	1276,513	1798,319	314,100
$SARIMA(1,0,2)(1,0,1,21)$	1589,932	1948,173	303,184
$SARIMA(3,0,3)(1,0,0,20)$	1587,148	1912,615	286,621

Cuadro 3: Resultados comparativos de modelos *SARIMA*, evaluados con base en las métricas *MAE*, *RMSE* y *MAPE*. Los valores reportados se expresan en millones de pesos colombianos.

Cabe señalar que el desempeño de los modelos fue evaluado a partir de las siguientes métricas de error:

Error Absoluto Medio (*MAE*): Mide la diferencia promedio entre los valores predichos y los reales, utilizando el valor absoluto para evitar cancelaciones entre errores positivos y negativos. Se calcula como:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Un *MAE* más bajo indica mayor precisión del modelo.

Raíz del Error Cuadrático Medio (*RMSE*): Dado que el Error Cuadrático Medio *(MSE)* se expresa en unidades cuadradas, se toma su raíz cuadrada para facilitar la interpretación en las unidades originales. Esta métrica se conoce como **Raíz del Error Cuadrático Medio (*RMSE*)** y se define como:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Un valor menor de *RMSE* indica una mejor capacidad predictiva del modelo.

Error Porcentual Absoluto Medio (*MAPE*): Métrica de precisión expresada como porcentaje, independiente de la escala y fácil de interpretar. Su desventaja es que puede generar valores indefinidos o extremos cuando los valores reales son cercanos a cero. Se calcula como:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

donde y_i son los valores reales y $\hat{y}_i$ los valores predichos.

Considerando los valores obtenidos en dichas métricas, se seleccionaron los modelos que presentaron el menor margen de error.

Finalmente, sobre los modelos seleccionados se aplicó la prueba de *Ljung-Box*, con el propósito preventivo de verificar que los residuos correspondieran a ruido blanco; es decir, que no presentaran autocorrelación y, por tanto, no contuvieran información remanente que pudiera ser modelada.

5. Análisis de resultados de los modelos *SARIMA*

En este punto se han seleccionado los modelos óptimos para ambas series de tiempo correspondientes a los convenios **Policía Nacional** y **Colpensiones**. Ambas series evidenciaron un comportamiento similar, dado que se analizó el desempeño de un mismo producto crediticio, la *libranza*, lo que permitió que, en gran medida, los procesos aplicados fueran equivalentes.

5.1. Policía Nacional

Tras el proceso desarrollado, se determinó que el modelo que mejor se ajustó a los datos del convenio con la **Policía Nacional** fue el $SARIMA(3,0,3)(1,0,0)_{20}$.

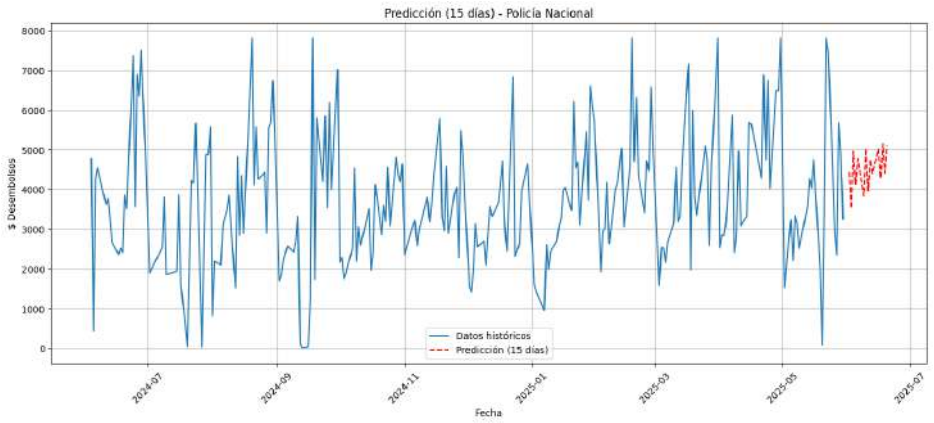


Figura 5: Serie de tiempo de los desembolsos correspondientes al convenio con la **Policía Nacional**, junto con la predicción para los siguientes 15 días hábiles. El eje horizontal representa la *Fecha*, mientras que el eje vertical muestra los *Desembolsos* en pesos colombianos. El gráfico permite observar tanto el comportamiento histórico como la tendencia proyectada en el corto plazo.

La predicción de los desembolsos del convenio con la **Policía Nacional** muestra una ligera tendencia al alza en los próximos 15 días hábiles, con valores que reflejan cierta estabilidad y menor variabilidad frente a los datos históricos. Si bien el modelo suaviza los cambios más abruptos, ofrece una proyección confiable que puede servir de apoyo en la planeación operativa y financiera del área de Movilización Comercial.

5.2. Colpensiones

Tras el proceso desarrollado, se determinó que el modelo que mejor se ajustó a los datos del convenio con la **Colpensiones** fue el $SARIMA(1,0,2)(1,0,1)_{21}$.

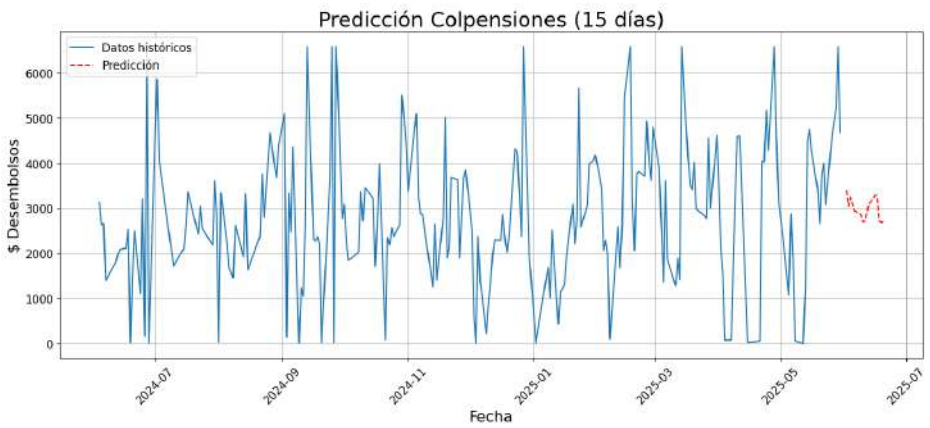


Figura 6: Serie de tiempo de desembolsos correspondientes al convenio con **Colpensiones**, junto con la predicción para los siguientes 15 días hábiles. El eje horizontal representa la *Fecha*, mientras que el eje vertical indica los *Desembolsos* en pesos colombianos. El gráfico permite apreciar tanto el comportamiento histórico como la proyección estimada en el corto plazo.

Aunque se mantiene cierta coherencia en los niveles estimados, el modelo suaviza en exceso las fluctuaciones, lo que limita su precisión. Esto evidencia que, para este convenio, la calidad y dispersión de los datos dificultan la construcción de predicciones más robustas.

6. Conclusiones

Este trabajo logró cumplir con su objetivo general, orientado a innovar en los procesos del área mediante el desarrollo de un procedimiento predictivo que respalde de manera efectiva las labores de los especialistas. La construcción de los modelos *SARIMA*, a partir de la información disponible, permitió establecer las siguientes conclusiones:

- **Fortalecimiento de la gestión de datos:** Se identificó una debilidad en el área relacionada con la ausencia de un sistema formal para almacenar y preservar históricos de información, lo que limita la capacidad de análisis a futuro. Asimismo, la escasez de registros disponibles restringe la precisión de los modelos; por ello, se recomienda ampliar la base de datos en al menos dos años adicionales, lo que permitiría mejorar el desempeño de los modelos e incluso evaluar metodologías más robustas, como *redes neuronales* o técnicas de *machine learning*.
- **Efectividad de los modelos:** Los modelos *SARIMA* desarrollados demostraron coherencia estadística y generaron predicciones razonables. No obstante, sus resultados deben ser interpretados en conjunto con el criterio de los especialistas del área de Impulso Comercial, a fin de respaldar de manera adecuada la toma de decisiones.

- **Innovación y automatización en el área de Impulso Comercial:** Se evidenció que el uso de *Python*, más allá de la construcción de modelos, ofrece un notable potencial para el área de Impulso Comercial, al posibilitar la innovación mediante la automatización de procesos cotidianos y la optimización de las tareas operativas.

7. Referencias

Referencias

[1] Universidad de Las Palmas de Gran Canaria, «Tema 4: Tratamiento de la estacionalidad. Modelos SARIMA,» Material docente, s.f. URL: https://www2.ulpgc.es/hege/almacen/download/29/29014/tema4.pdf.

[2] Banco de Bogotá, «Un banco que hace historia,» Banco de Bogotá, 2015. URL: https://www.bancodebogota.com/wps/portal/banco-de-bogota/bogota/nuestra-organizacion/sala-de-prensa/2015/un-banco-que-hace-historia.

[3] Banco de Bogotá Internacional, «History,» Banco de Bogotá Internacional, s.f. URL: https://www.bancodebogotainternacional.com/bbi/es/thebank/history.

[4] Santra, Ritus, «Tests for Stationarity in Time Series — Dickey Fuller Test & Augmented Dickey Fuller (ADF) Test,» Medium, 2023. URL: https://medium.com/@ritusantra/tests-for-stationarity-in-time-series-dickey-fuller-test-augmented-dickey-fuller-adf-test-d2e92e214360.

[5] C. Miranda Chinlli, «Modelización de series temporales,» Trabajo Fin de Máster en Estadística Aplicada, Universidad de Granada, 2019. URL: https://masteres.ugr.es/estadistica-aplicada/sites/master/moea/public/inline-files/TFM_MIRANDA_CHINLLI_CARLOS.pdf.

[6] Universidad Nacional Autónoma de México, «Modelos ARIMA,» Material académico, UNAM, s.f. URL: http://www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/363/7/A7.pdf.

[7] BBVA Colombia, «¿Qué es una libranza?», Blog Educación Financiera, BBVA Colombia, s.f. URL: https://www.bbva.com.co/personas/blog/educacion-financiera/prestamos/que-es-libranza.html.

[8] Ritu Santra, «Introduction to SARIMA Model», Medium, 8 junio 2023. URL: https://medium.com/@ritusantra/introduction-to-sarima-model-cbb885ceabe8.