



**FACULTAD DE ESTUDIOS ESTADÍSTICOS
GRADO EN ESTADISTICA APLICADA**

Curso 2023/2024

Trabajo de Fin de Grado

**ANÁLISIS TEMPORAL DE MULTAS DE TRÁFICO
EN MADRID: PATRONES Y PREDICCIONES**

LUCÍA ORIZA ROYO

SUPERVISADO POR

D^a. Juana María Alonso Revenga

JUNIO 2023



UNIVERSIDAD COMPLUTENSE
MADRID

ÍNDICE

Resumen	1
1-Introducción	2
1.1-Contexto y Relevancia del Estudio	2
1.2-Objetivos del Estudio	3
1.3-Estructura del Trabajo	4
1.4-Contribuciones Esperadas	4
2 Marco teórico	5
2.1 Análisis de Datos de Multas de Tráfico	5
3- Metodología	6
4. Estudio y depuración de la base	9
4.1 Estructura y Variables	9
4.2 Preprocesamiento de Datos	10
4.2.1 Análisis de Datos Faltantes	10
4.2.2 Análisis de Outliers	10
4.3 Análisis Exploratorio de Datos	11
4.3.1 Descripción General de los Datos	11
4. Creación de las nuevas variables a través del análisis factorial	13
4.1 Introducción	13
4.2 Determinación del número de factores	14
4.3 Interpretación de los factores	15
5. Análisis y predicción de las series temporales	18
5.1 Serie estacionamiento	18
5.2 Serie Regulación Vial y Control de Tráfico	25
5.3 Serie Regulación de Velocidad y Seguridad Vial	31
5.4 Predicciones de las variables originales	37
6. Conclusiones	39
7. Bibliografía	41
8. Anexo	42

Resumen

El presente Trabajo de Fin de Grado analiza las multas de tráfico en la ciudad de Madrid, identificando patrones y tendencias para posteriormente realizar predicciones basadas en datos. Se recopiló y depuró datos proporcionados por las autoridades de tráfico de Madrid. Se incluye una revisión de la literatura sobre análisis temporal de datos de tráfico, limpieza y preprocesamiento de datos, y un análisis.

Posteriormente, se aplicaron métodos de análisis de series temporales, incluyendo modelos de suavizado exponencial y ARIMA, para identificar patrones y realizar predicciones. También se utilizó el análisis factorial para reducir la dimensionalidad de los datos y facilitar la interpretación de los factores que influyen en las multas de tráfico. Finalmente, se realizaron predicciones sobre las variables originales utilizando los modelos desarrollados.

Las conclusiones ofrecen recomendaciones para mejorar la gestión y regulación del tráfico en la ciudad, y las predicciones pueden apoyar la planificación futura.

Abstract

This work analyzes traffic fines in the city of Madrid, identifying patterns and trends to subsequently make data-based predictions. Data provided by Madrid's traffic authorities were collected and cleaned. This project includes a literature review on temporal data analysis of traffic, data cleaning and preprocessing, and exploratory data analysis.

Subsequently, time series analysis methods, including exponential smoothing and ARIMA models, were applied to identify patterns and make predictions. Factor analysis was also used to reduce data dimensionality and facilitate the interpretation of factors influencing traffic fines. Finally, predictions of the original variables were made using the developed models.

The conclusions offer recommendations to improve traffic management and regulation in the city, and the predictions can support future planning.

1-Introducción

En las últimas décadas, el aumento constante de la densidad del tráfico ha sido una tendencia notable en las ciudades, y Madrid, como una urbe relevante en Europa, no ha escapado a este fenómeno. Esta situación ha llevado a un incremento significativo en la emisión de multas de tráfico, reflejando la necesidad de mantener el orden y la seguridad vial en entornos urbanos cada vez más congestionados.

La congestión vehicular continua y las crecientes preocupaciones sobre la seguridad han impulsado a las autoridades a implementar medidas más rigurosas para hacer cumplir las normativas de tráfico y estacionamiento. Esta evolución en las políticas de aplicación de la ley de tráfico ha generado un interés creciente en comprender en detalle las tendencias, patrones y factores que influyen en la emisión de multas en una ciudad como Madrid.

El objetivo de este proyecto es investigar y analizar en profundidad las multas de tráfico emitidas en Madrid, centrándose en diversos tipos de infracciones y utilizando técnicas de análisis de series temporales. A través de este análisis, se pretende obtener una comprensión más amplia de los patrones temporales, las tendencias y los factores asociados con las multas de tráfico en la ciudad.

1.1-Contexto y Relevancia del Estudio

La gestión eficiente del tráfico y la aplicación efectiva de las normativas de tráfico son aspectos críticos para mejorar la movilidad urbana y garantizar la seguridad de los ciudadanos. Las multas de tráfico no solo sirven como medidas disuasorias para comportamientos peligrosos en la carretera, sino que también generan ingresos que pueden reinvertir en infraestructuras viales y programas de seguridad vial.

El análisis de series temporales aplicado a datos de multas de tráfico ofrece una perspectiva para comprender cómo evolucionan los patrones de infracción a lo largo del tiempo.

1.2-Objetivos del Estudio

El objetivo principal de esta investigación es realizar un análisis detallado de las multas de tráfico emitidas en Madrid, utilizando técnicas avanzadas de series temporales. Se busca profundizar en el comportamiento y la evolución de estas multas para obtener conocimientos significativos que respalden la gestión del tráfico y la seguridad vial en la ciudad.

En este estudio, se persiguen varios objetivos específicos.

Identificar y categorizar los tipos más frecuentes de multas de tráfico en Madrid, incluyendo infracciones como exceso de velocidad, violaciones de semáforos, estacionamiento indebido y otras faltas comunes. Este análisis permitirá comprender mejor las áreas críticas de la seguridad vial en la ciudad.

Analizar la evolución temporal de las multas a lo largo de varios años, examinando detalladamente patrones estacionales, tendencias a largo plazo y variaciones interanuales. Este enfoque temporal proporcionará información valiosa sobre la dinámica de las infracciones en diferentes períodos del año y años sucesivos.

Utilizar modelos avanzados de series temporales, como ARIMA (Autoregressive Integrated Moving Average), para prever la emisión futura de multas y comprender cómo podrían evolucionar los patrones de interacción a lo largo del tiempo. Estos modelos permitirán desarrollar pronósticos precisos sobre las tendencias futuras de las infracciones, lo que será fundamental para la planificación y la toma de decisiones en materia de seguridad vial.

En resumen, este estudio se enfoca en examinar en profundidad el comportamiento de las multas de tráfico en Madrid mediante el uso de herramientas estadísticas y modelos avanzados de series temporales, con el objetivo de ofrecer conocimientos relevantes para mejorar la gestión del tráfico y la seguridad en las vías de la ciudad.

1.3-Estructura del Trabajo

La estructura de este trabajo se organiza en diferentes secciones que abarcan desde el marco teórico hasta las conclusiones:

En el **marco teórico**, se realiza una revisión de la literatura relacionada con el análisis de datos de multas de tráfico, destacando estudios previos y conceptos clave en el campo.

En la **metodología** utilizada, se describen detalladamente las técnicas estadísticas empleadas.

El **preprocesamiento de los datos** analiza desde la estructura y variables de los datos hasta el preprocesamiento y análisis exploratorio.

El análisis **factorial** identificado en el estudio se presenta en una sección dedicada, donde se analizan en profundidad las relaciones entre las variables originales y los factores.

El análisis de **series temporales** constituye otro aspecto central del estudio, donde se identifican tendencias, estacionalidades y se aplican métodos predictivos de suavizado para comprender mejor la evolución temporal de las multas de tráfico.

Las **predicciones** profundizan en la comprensión de los hallazgos obtenidos a partir de los análisis realizados. Se examinan las relaciones identificadas entre variables, se analizan los patrones de comportamiento a lo largo del tiempo y se destacan las principales conclusiones derivadas del estudio.

Por último, en la sección de **conclusiones** se destacan los resultados principales y se dan recomendaciones para futuras investigaciones.

1.4-Contribuciones Esperadas

Se espera que este estudio contribuya al conocimiento existente sobre el análisis de multas de tráfico desde una perspectiva temporal. Los resultados obtenidos podrían ser útiles para las autoridades de tráfico y planificadores urbanos en la toma de decisiones relacionadas con la gestión del tráfico y la implementación de políticas de seguridad vial más efectivas.

En resumen, este estudio busca ofrecer una visión integral de las multas de tráfico en Madrid, explorando su dinámica temporal y proporcionando información relevante para mejorar la movilidad urbana y promover conductas seguras en las carreteras de la ciudad.

2 Marco teórico

El análisis de datos de multas de tráfico es fundamental para comprender y mejorar la gestión urbana y la seguridad vial. Este apartado se centra en examinar las tendencias y patrones de las infracciones a lo largo del tiempo, utilizando técnicas estadísticas y de series temporales para obtener insights significativos.

2.1 Análisis de Datos de Multas de Tráfico

Los datos de multas de tráfico incluyen información detallada sobre las infracciones cometidas, como exceso de velocidad, violaciones de señales, o estacionamiento indebido.

Las series temporales son secuencias de datos ordenadas temporalmente. En el contexto de las multas de tráfico, estas técnicas son útiles para entender patrones estacionales, identificar cambios a lo largo del tiempo y prever futuras infracciones.

Estudios Relevantes

Un estudio relevante que aplicó técnicas de series temporales al análisis de datos de multas de tráfico es el trabajo de Daza et al. (2017) titulado "Predicting Traffic Accidents using Time-Series Data". En este estudio, los investigadores utilizaron modelos de series temporales para prever la ocurrencia de accidentes de tráfico en función de datos históricos de multas y otros factores relacionados con el tráfico.

Daza et al. demostraron que, al utilizar técnicas de series temporales, como modelos ARIMA y suavizado exponencial, era posible identificar patrones y tendencias en las multas de tráfico a lo largo del tiempo.

3- Metodología

Para el análisis de las multas de tráfico en Madrid, se emplearán dos técnicas estadísticas avanzadas: el Análisis Factorial y el Análisis de Series Temporales. El Análisis Factorial permitirá identificar las dimensiones subyacentes que influyen en la emisión de multas, mientras que el Análisis de Series Temporales ayudará a detectar tendencias temporales y estacionales, proporcionando una base sólida para hacer predicciones futuras.

El Análisis Factorial es una técnica multivariante ampliamente utilizada en diversas disciplinas, incluyendo la psicología, sociología y economía, para identificar estructuras subyacentes en los datos. Según *Hair et al. (2010)*, el objetivo principal del Análisis Factorial es reducir la dimensionalidad de un conjunto de datos observando las relaciones entre variables múltiples y agrupándolas en factores latentes que explican la mayor parte de la variabilidad observada. Esta técnica se basa en la premisa de que múltiples variables observadas pueden estar influenciadas por un número menor de factores subyacentes. Estos factores latentes no se observan directamente, pero se infieren a través de las correlaciones entre las variables observadas.

Antes de proceder con el análisis factorial, se debe evaluar la adecuación de la matriz de correlaciones utilizando el índice de Kaiser-Meyer-Olkin (KMO). Este índice mide el grado de correlación entre las variables.

Para extraer los factores, se pueden utilizar varios métodos:

Método de Componentes Principales: Este método se basa en la descomposición espectral de la matriz de correlaciones.

Método del Factor Principal: Similar al anterior, pero con una formulación distinta.

Método de Máxima Verosimilitud: Este método requiere la hipótesis adicional de normalidad en las variables originales.

La elección del número de factores a retener se basa en el criterio de los autovalores (Kaiser) y en la varianza explicada acumulada.

Los factores se interpretan examinando las cargas factoriales, que son los coeficientes de correlación entre las variables observadas y los factores. Las variables con cargas altas en un mismo factor se consideran como representantes de un mismo constructo subyacente.

La validación del modelo factorial implica verificar la adecuación del modelo a los datos. Se evalúan la comunalidad (la proporción de varianza explicada por los factores) y la especificidad (la parte de varianza específica no explicada por los factores comunes).

Las puntuaciones factoriales representan los valores individuales en cada factor y se obtienen a través de técnicas como la regresión múltiple.

El Análisis de Series Temporales es una técnica estadística que se utiliza para analizar datos que se recopilan secuencialmente en el tiempo. Este método permite identificar patrones, tendencias y estacionalidades en los datos temporales. *Box et al. (2015)* es una referencia fundamental en este campo, proporcionando una base comprensiva para el modelado y análisis de series temporales. Su aplicación en el análisis de las multas de tráfico permite entender cómo estas varían a lo largo del tiempo y prever tendencias futuras.

Los métodos de suavizado son técnicas diseñadas para filtrar el "ruido" en los datos y destacar las tendencias subyacentes. En este estudio, se aplicarán métodos de suavizado para obtener una visión clara de las tendencias a lo largo del tiempo. El Suavizado Exponencial Doble, también conocido como el modelo de Holt, es una técnica de suavizado diseñada para abordar series temporales con una tendencia lineal. Este método extiende el suavizado exponencial simple al considerar tanto el nivel (valor promedio de la serie) como la tendencia (cambio en el nivel a lo largo del tiempo).

El suavizado exponencial doble de Holt se define mediante ecuaciones de suavizado, una para el nivel y otra para la tendencia:

1. Ecuación de Nivel:

$$L_t = \alpha x_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

2. Ecuación de Tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

A diferencia del suavizado exponencial simple, el modelo de Holt puede capturar y pronosticar series temporales con una tendencia lineal.

Los modelos ARIMA son ampliamente utilizados para analizar series temporales debido a su capacidad para manejar diversas estructuras de datos temporales, incluyendo componentes estacionales y no estacionales.

La identificación de un modelo ARIMA adecuado comienza con la transformación de la serie en estacionaria, lo cual se logra mediante diferenciación. A continuación, se analizan los correlogramas de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los posibles valores de los parámetros p (orden autoregresivo) y q (orden de media móvil).

Una vez identificado el modelo ARIMA adecuado, se procede a la estimación de sus parámetros utilizando métodos como la Máxima Verosimilitud. El modelo ARIMA se define por los componentes p , d y q , que representan los órdenes de autoregresión, integración (diferenciación) y media móvil, respectivamente.

La validación del modelo implica varios pasos para asegurar que el modelo ajustado es adecuado para los datos. Esto incluye el diagnóstico de residuales, que verifica que los errores del modelo se comporten como ruido blanco, y el uso de criterios de información como AIC (Akaike Information Criterion) o BIC (Bayesian Information Criterion) para comparar diferentes modelos y seleccionar el más apropiado.

Con el modelo ARIMA ajustado y validado, se procede a realizar predicciones sobre la serie temporal. Estas predicciones se comparan con los datos reales para evaluar la precisión y la capacidad del modelo para capturar la estructura subyacente de la serie.

La combinación del Análisis Factorial y el Análisis de Series Temporales proporciona una aproximación holística al estudio de las multas de tráfico en Madrid. Mientras que el Análisis Factorial ofrece una visión profunda de las relaciones subyacentes entre múltiples variables, el Análisis de Series Temporales proporciona una perspectiva longitudinal y dinámica, capturando cómo estas relaciones y patrones evolucionan con el tiempo. Juntos, permiten una comprensión más completa de los datos.

4. Estudio y depuración de la base

La base de datos utilizada para este análisis consiste en registros mensuales de multas de tráfico emitidas en la ciudad de Madrid, abarcando un período desde enero de 2015 hasta junio de 2023. Estos datos son recopilados por las autoridades encargadas de la gestión del tráfico y la seguridad vial en la ciudad.

4.1 Estructura y Variables

Los datos están estructurados en un formato tabular, donde cada fila representa una multa de tráfico individual y cada columna corresponde a una variable específica que describe diferentes aspectos de las infracciones. Las variables incluyen:

Variables Exógenas

1. Velocidad: La velocidad promedio registrada en zonas específicas puede afectar la frecuencia de multas de tráfico. Altas velocidades suelen asociarse con más infracciones.
2. Semaforo: El estado y el funcionamiento de los semáforos pueden influir en la ocurrencia de ciertos tipos de multas, como pasarse la luz roja.
3. A_P_R : Esta variable podría indicar multas relacionadas por entrar en zonas protegidas.
4. Móvil: Multas asociadas con el uso de dispositivos móviles mientras se conduce.
5. Otras Dinámicas: Otros comportamientos dinámicos al volante que pueden generar multas.

Variables Endógenas

1. Ser_Sin_Tique: Multas relacionadas con estacionar sin tique.
2. Ser_Tique_Invalido: Multas asociadas con el uso indebido o inválido de tickets de estacionamiento.
3. Carga_Descarga: Multas relacionadas con la carga y descarga de mercancías.
4. Carril_Bus: Infracciones relacionadas con el uso indebido de carriles reservados para autobuses.

5. Prohibido_Aparcar: Multas por estacionar en lugares prohibidos.
6. Doble_Fila: Multas por estacionar en doble fila.

Variables Totales

1. Multas_Totales_Dinámicas: Suma de multas asociadas con comportamientos dinámicos al volante.
2. Multas_Totales_Estáticas: Suma de multas asociadas con estacionamiento y otras infracciones estáticas.
3. Multas_Totales: Suma total de todas las multas registradas.

La recopilación mensual de datos desde enero de 2015 hasta junio de 2023 ofrece la oportunidad de realizar un análisis detallado de la evolución de las multas de tráfico a lo largo de más de ocho años. Esta extensión temporal permite observar cambios significativos en el comportamiento de los conductores y en las políticas de tráfico durante este período.

4.2 Preprocesamiento de Datos

4.2.1 Análisis de Datos Faltantes

En primer lugar, se realizó un análisis exhaustivo para identificar cualquier dato faltante en las variables numéricas utilizando el procedimiento PROC FREQ en SAS.

Este procedimiento examinó todas las variables numéricas en el conjunto de datos de multas (MULTAS) y generó una tabla de frecuencias que mostraba la presencia de valores faltantes (missing values). El resultado de este análisis indicó que no se encontraron datos faltantes en ninguna de las variables numéricas, lo cual es fundamental para asegurar la integridad de los datos utilizados en el estudio.

4.2.2 Análisis de Outliers

Posteriormente, se llevó a cabo un análisis de outliers utilizando el procedimiento PROC UNIVARIATE en SAS sobre las variables de interés relacionadas con las multas de tráfico.

Este análisis univariado permitió identificar posibles valores atípicos (outliers) en las variables seleccionadas. Sin embargo, el resultado del análisis indicó que no se encontraron outliers significativos en ninguna de las variables, lo que sugiere que los datos de las multas de tráfico están dentro de rangos esperados y coherentes con el contexto de estudio.

El análisis realizado demostró que los datos de multas de tráfico utilizados están completos y libres de valores faltantes y outliers significativos. Este proceso de limpieza de datos es crucial para garantizar la fiabilidad y validez de los análisis posteriores realizados en el marco de la investigación.

4.3 Análisis Exploratorio de Datos

El análisis exploratorio de datos es una fase crucial en cualquier estudio, ya que permite comprender mejor las características y las relaciones dentro de un conjunto de datos.

4.3.1 Descripción General de los Datos

Se comienza con una descripción general del conjunto de datos para tener una visión general de las variables y sus estadísticas descriptivas.

Variable	N	Media	Desviación	Mínimo	Máximo
VELOCIDAD	102	43627,73	10095,40	17145	71506
SEMAFORO	102	4352,25	3415,34	718	13654
A_P_R	102	42815,98	26767,36	0	120305
MOVIL	102	561,51	233,02	116	1121
OTRAS_DINAMICAS	102	3312,17	1183,26	931	7004
SER_SIN_TIQUE	102	34036,96	10666,11	0	50721
SER_TIQUE_INVALIDO	102	30179,66	11347,54	3	47701
CARGA_DESCARGA	102	2460,46	1225,36	489	7256
CARRIL_BUS	102	1006,87	39685,01	21	2135
PROHIBIDO_APARCAR	102	14196,54	3376,44	6288	19149
DOBLE_FILA	102	2079,33	965,31	388	4479
multas_totales_dinamicas	102	94669,64	32497,26	33055	176812
MULTAS_TOTALES_ESTATICAS	102	83959,82	25664,18	7533	123077
multas_totales	102	178629,46	44101,29	40588	268637

Tabla 1. Estadísticos

Se realizará el estudio de las variables a lo largo del tiempo gráficamente, a través de un gráfico de líneas.

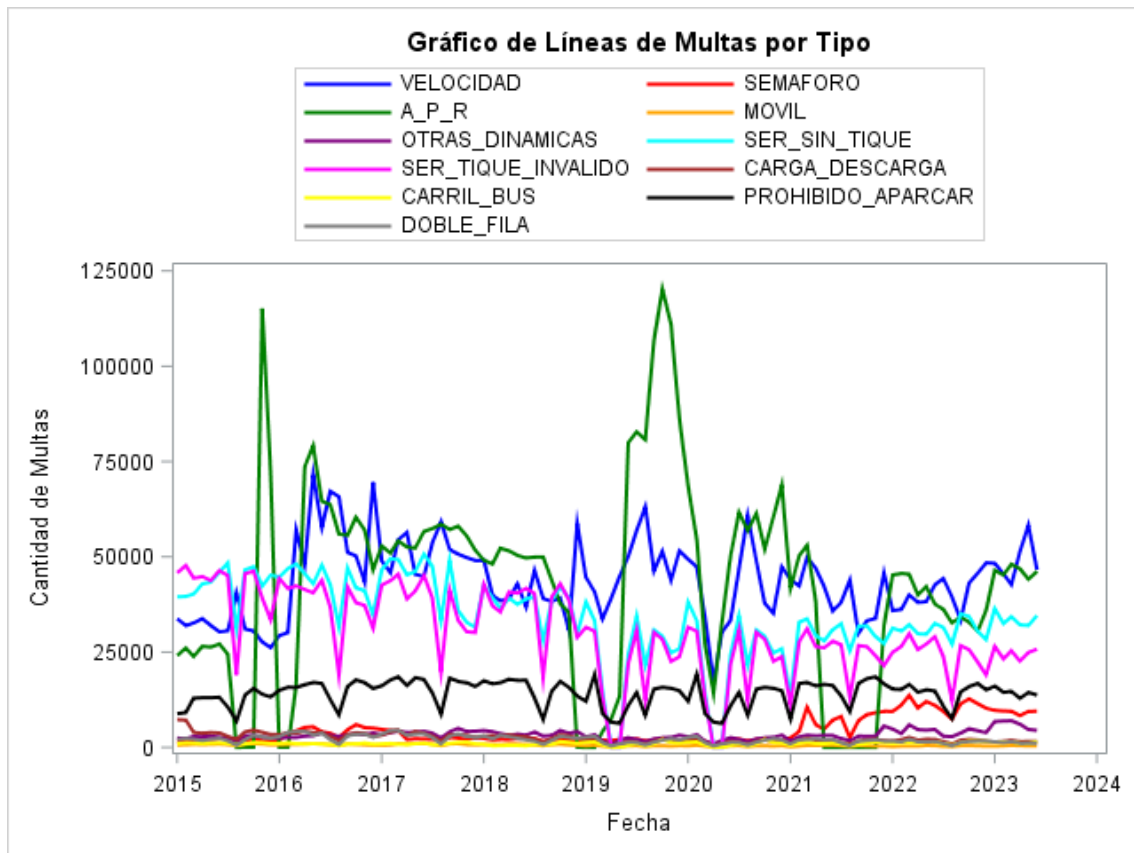


Imagen 1. Gráfico de líneas por tipo de multa

Se observa cómo la variable A_P_R es la que obtiene los valores más altos, alcanzados a finales de 2015 y finales de 2019, seguido de la variable velocidad, esto indica que son el tipo de multas más frecuentes, mientras que variables como doble fila, móvil o carril bus son menos frecuentes ya que alcanzan valores más bajos.

Se observa como algunas de las variables muestran un comportamiento similar alcanzando sus máximos y mínimos relativos en las mismas fechas, lo que da a entender que para facilitar el análisis se podrán agrupar.

Al ser muchas variables y algunas con comportamiento similar se estudiará reducir tal número y trabajar con menos pero que engloben a las descritas anteriormente.

4. Creación de las nuevas variables a través del análisis factorial

4.1 Introducción

El análisis factorial es una técnica estadística utilizada para identificar las estructuras subyacentes en un conjunto de datos multidimensionales. Este método es especialmente útil cuando se trabaja con grandes cantidades de variables que pueden estar interrelacionadas, como en el caso de los datos de multas de tráfico en Madrid. A través del análisis factorial, es posible reducir la complejidad de los datos y extraer los factores que mejor explican las variaciones observadas en las diferentes infracciones.

Se basa en la idea de que múltiples variables observadas pueden ser explicadas por un menor número de variables no observadas llamadas "factores". Este proceso incluye varias etapas clave.

Se comienza calculando la matriz de correlaciones entre todas las variables originales para entender cómo se relacionan entre sí. Este mapa de calor ayuda a identificar grupos de variables que están altamente correlacionadas.

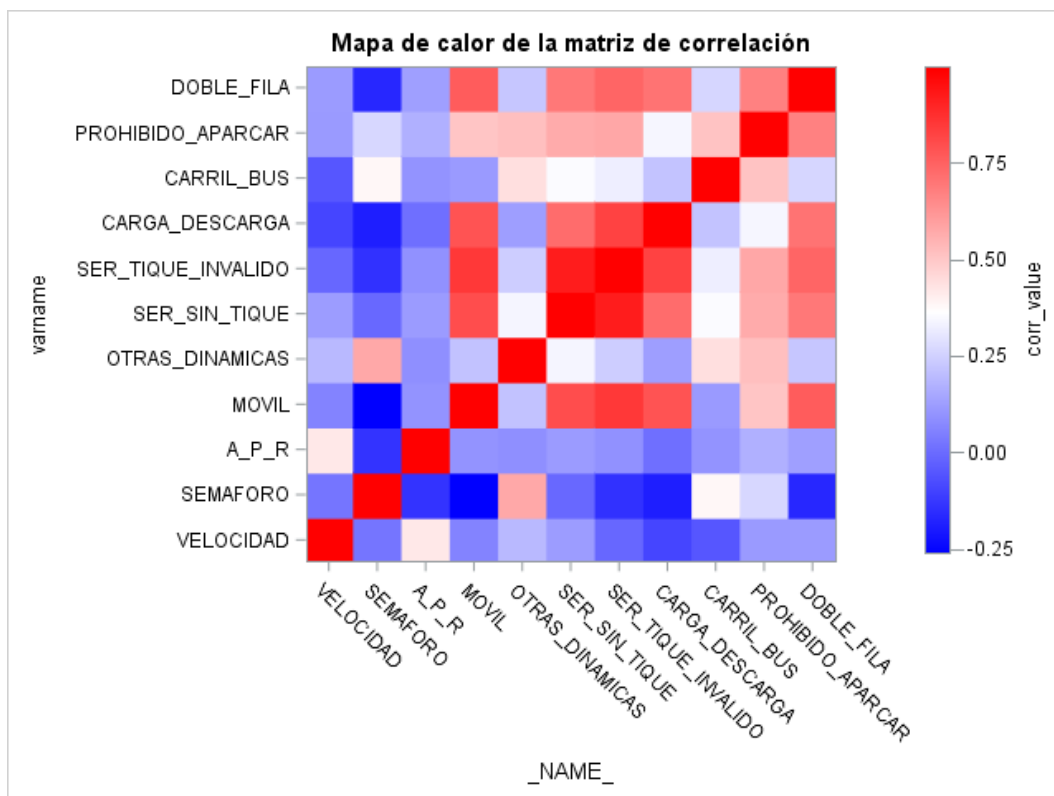


Imagen 2. Coeficientes de correlación

La identificación de correlaciones fuertes puede ayudar a las autoridades a entender cuáles infracciones tienden a ocurrir juntas y podrían indicar áreas problemáticas o comportamientos específicos de los conductores que necesitan ser abordados simultáneamente.

Efectivamente se observa como algunas variables están correlacionadas ya que tienen valores altos (color rojo) por lo que podremos reducir la dimensión.

Utilizando métodos como el análisis de componentes principales (PCA) o el análisis factorial exploratorio (EFA), se extraen los factores subyacentes. Estos factores son combinaciones lineales de las variables originales y están diseñados para explicar la mayor parte de la varianza en los datos con el menor número de factores posible. En este caso se trabajará con el análisis factorial para una vez analizado el mejor modelo de predicción a través de las series temporales poder deshacer la descomposición factorial y poder hacer predicciones para las variables originales.

4.2 Determinación del número de factores

Determinar el número de factores en un análisis factorial es una de las etapas cruciales para garantizar que el modelo proporcione una representación precisa y significativa de los datos.

El índice MSA de Kaiser-Meyer-Olkin es una medida global del grado de correlación, con nuestros datos se obtiene un 0.70164594, por lo que indica que es una correlación suficiente.

El criterio de Kaiser es uno de los métodos más sencillos y ampliamente utilizados para decidir el número de factores a retener, se calculan los autovalores de la matriz de correlaciones (un autovalor representa la cantidad de varianza explicada por cada factor), según este criterio, solo se retienen los factores con autovalores mayores que 1, ya que un autovalor mayor que 1 indica que el factor explica más varianza que una variable original individual.

Eigenvalues of the Correlation Matrix: Total = 11 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	4.89644087	2.79078820	0.4451	0.4451
2	2.10565267	0.65217753	0.1914	0.6366
3	1.45347515	0.71554044	0.1321	0.7687
4	0.73793470	0.22406021	0.0671	0.8358

Tabla 2. Autovalores de los factores

Como los tres primeros autovalores son mayores que 1 y explican más de un 75% de la varianza se acepta tomar 3 factores para explicar las variables originales.

El primer factor domina claramente, explicando cerca del 58% de la varianza total. Esto sugiere que hay una fuerte dimensión subyacente en los datos que aglutina la mayor parte de la variabilidad observada. Aunque no tan predominantes como el primer factor, los factores 2 y 3 juntos explican aproximadamente el 42.1% de la varianza total, indicando que hay múltiples dimensiones relevantes en los datos de las multas de tráfico.

4.3 Interpretación de los factores

A continuación, se muestran los gráficos "Gráfico de Patrón de Factores Iniciales" (Initial Factor Pattern) que se utiliza comúnmente en el análisis factorial para visualizar las cargas factoriales de las variables en los factores extraídos, y así poder determinar cada factor a que variables va a representar, se han recuadrado las variables que se asocian principalmente a cada factor representado en cada imagen por un eje.

El gráfico ayuda a identificar las dimensiones subyacentes que explican las correlaciones entre las variables observadas. Cada factor puede interpretarse como una dimensión latente que agrupa variables con patrones de comportamiento similares.

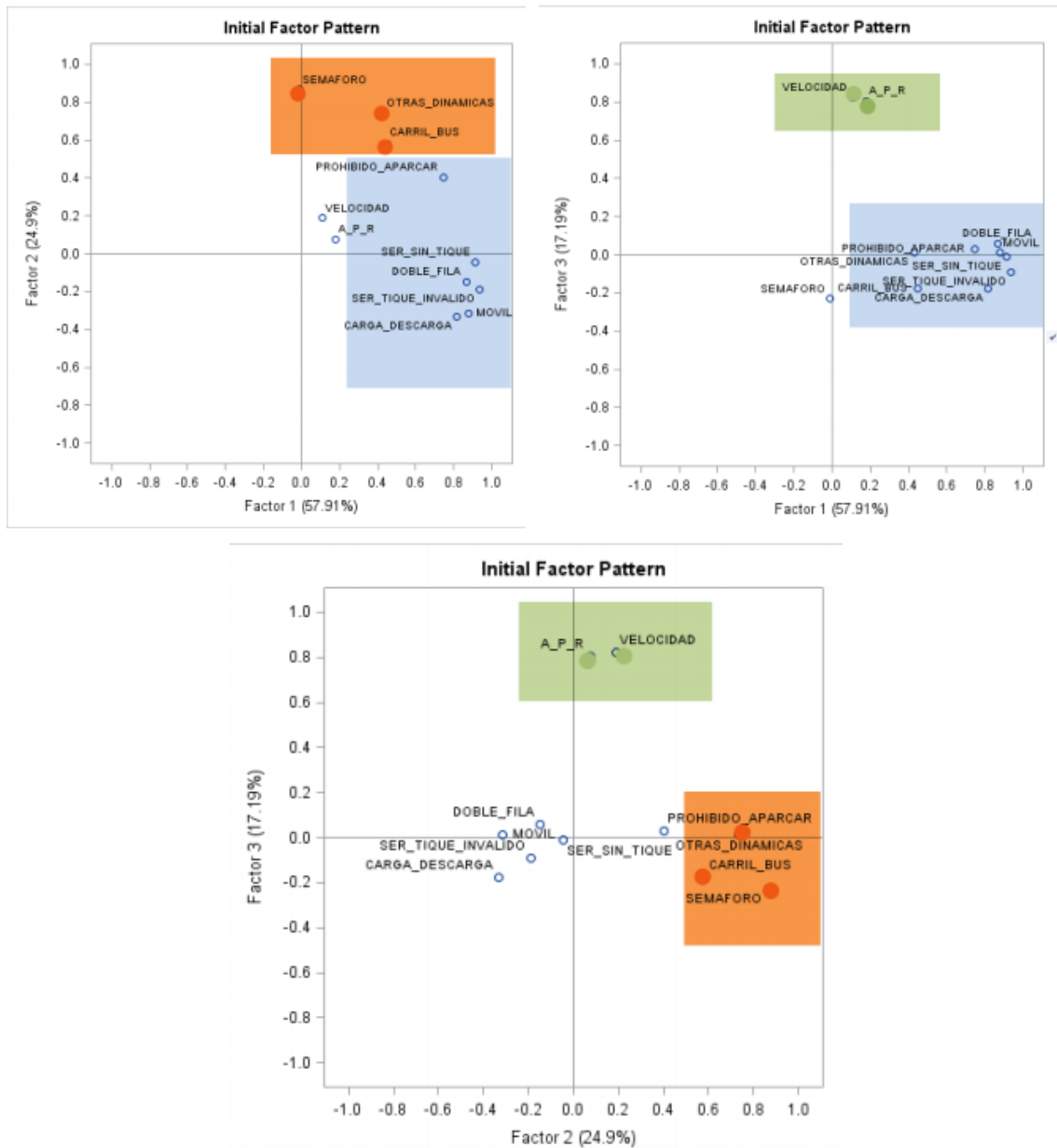


Imagen 3. Factor Pattern

A partir de la imagen 3 (Factor Pattern) con los gráficos de las cargas factoriales, en este apartado, se analizarán dichos gráficos que se corresponden a cada par de factores, ofreciendo una visión detallada de las conexiones entre las variables y los tres factores obtenidos. Los recuadros representan las variables con altas cargas para cada factor, el color azul para el factor 1, el color naranja para el factor 2 y por último el verde para el factor 3, gracias a esto se ofrece una visión más clara.

En la primera imagen, el **FACTOR 1** tiene cargas positivas altas para variables que parecen estar relacionadas con las normativas de estacionamiento (SERV_TIQUE_INVALIDO, MOVIL, PROHIBIDO_APARCAR, DOBLE_FILA, SERV_SIN_TIQUE, CARGA_DESCARGA).

Interpretación: Este factor podría representar una dimensión de "**Estacionamiento**", indicando un conjunto de variables de estacionado por el ticket, por el impedimento de aparcar en ciertos lugares y por la forma de aparcar en ellos.

En la primera y tercera imagen, el **FACTOR 2** se asocia principalmente con (SEMAFORO, OTRAS_DINAMICAS y CARRIL_BUS).

Interpretación: Este factor representa aspectos relacionados con la regulación vial y el control del tráfico, incluyendo variables como semáforos, otras dinámicas de tráfico y el uso del carril bus). "**Regulación Vial y Control de Tráfico**".

Refleja la influencia de medidas en las infracciones de tráfico.

En la segunda y tercera imágenes, el **FACTOR 3** tiene una fuerte correlación con las variables (APR y VELOCIDAD).

Interpretación: El factor 3 está asociado con la regulación de la velocidad y las cuestiones de seguridad vial, representadas por variables como multas por entrar en zonas protegidas (APR) y por el exceso de velocidad. "**Regulación de Velocidad y Seguridad Vial**".

Destaca la importancia de las normativas relacionadas con la velocidad y la seguridad en las infracciones registradas.

Al resumir bien estos factores a todos los tipos de multas se realizará el estudio con estas nuevas variables, factores.

5. Análisis y predicción de las series temporales

En primer lugar, se dividió el conjunto de datos en conjuntos de entrenamiento y prueba, reservando el último medio año de datos (de enero a junio de 2023) como conjunto de prueba para hacer las predicciones.

5.1 Serie estacionamiento

A continuación, se ajustó el modelo de suavizado de Holt-Winters utilizando el conjunto de entrenamiento, lo que permitió capturar las tendencias y estacionalidades presentes en los datos históricos de multas.

El siguiente gráfico muestra la serie original y suavizada de estacionamiento.

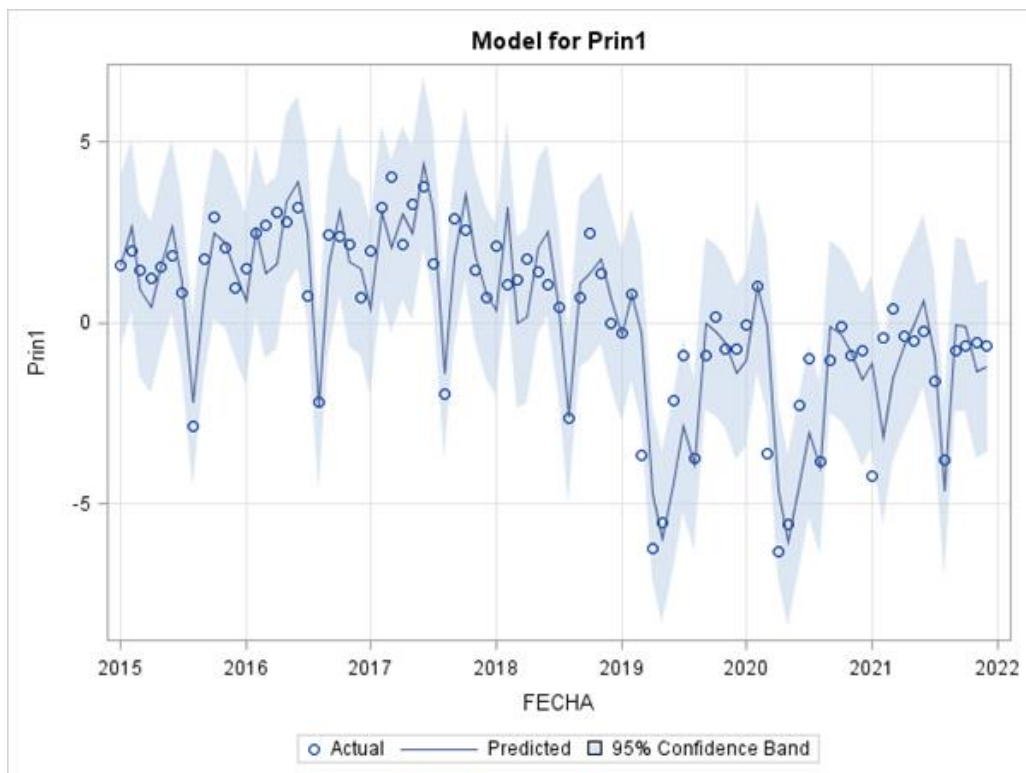


Imagen 4. Serie estacionamiento original y suavizada

Winters Method (Additive) Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.08050	12.41	<.0001
Trend Weight	0.0010000	0.03033	0.03	0.9738
Seasonal Weight	0.0010000	30.68397	0.00	1.0000

Tabla 3. Parámetros de Holt Winters para estacionamiento

Ecuación 1

$$Y_t = 0.999L_{t-1} + 0.001T_{t-1} + 0.001S_{t-1}$$

El coeficiente de nivel tiende a 1, esto indica que los datos del pasado no influyen en los cambios de nivel, por lo que la serie mantiene su nivel basado únicamente en el período estacional del instante t, sin tener en cuenta los niveles anteriores.

Un trend weight cercano a 0 en el método de Holt-Winters indica que el modelo le está dando muy poca importancia a la tendencia al realizar las predicciones.

Por otro lado, considerando que el coeficiente de estacionalidad tiende a cero indica que los datos del pasado si influyen en la estacionalidad y que todos los coeficientes de estacionalidad son iguales a lo largo del tiempo.

Se genera una gráfica a continuación que mostrará los valores observados de la variable de estacionamiento junto con los valores suavizados y las predicciones obtenidas a partir de un modelo específico de series temporales.

Además, en la gráfica se incluirán los intervalos de confianza para las predicciones, lo que nos permitirá visualizar la incertidumbre asociada a las estimaciones futuras. Esta representación nos brindará una herramienta visual para evaluar la capacidad predictiva del modelo y entender mejor cómo se comportan los datos observados en comparación con las proyecciones realizadas.

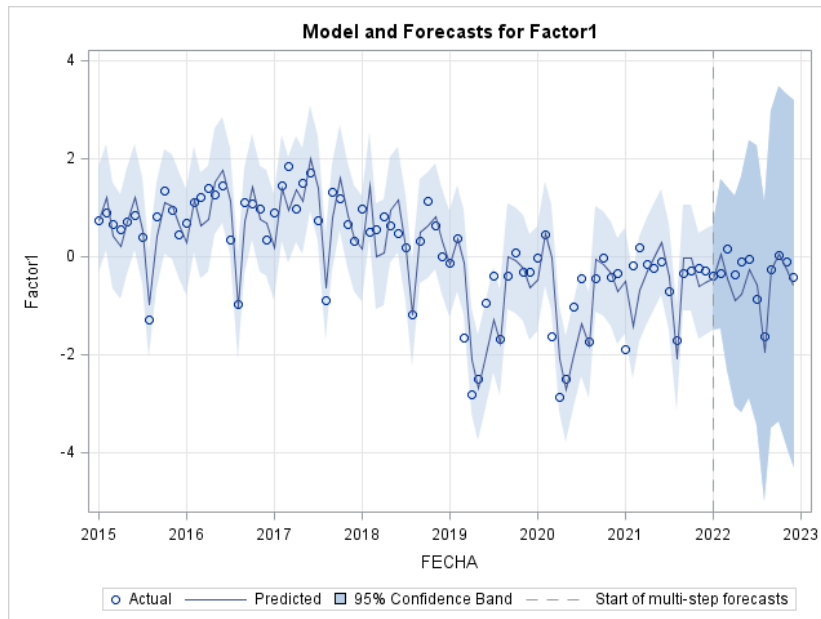


Imagen 5. Observados y predicciones estacionamiento

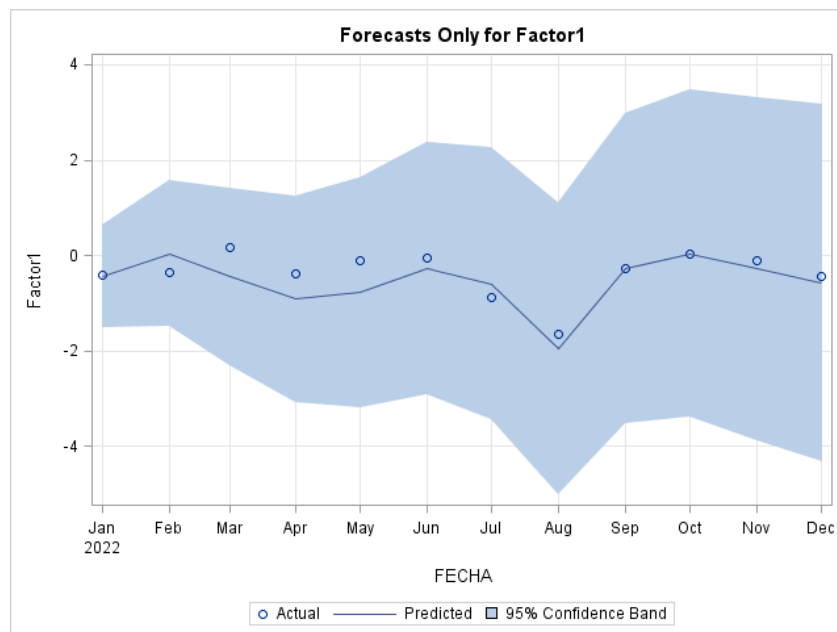


Imagen 6. Observados y predicciones estacionamiento 2022

A medida que nos desplazamos hacia el futuro a partir de los datos actuales, los intervalos de confianza se amplían, lo que indica un aumento en el error estándar. Este fenómeno refleja una disminución en la precisión del modelo a lo largo del tiempo.

Con el modelo de suavizado se obtiene un Mean Square Error 0.12728688.

Tras haber analizado el método de suavizado se estudiará cuál es el modelo ARIMA más adecuado para la variable estacionamiento.

Dado que en el correlograma de la variable de estacionamiento se observan valores significativos en los primeros cinco rezagos en el análisis de correlación (ACF), esto sugiere una fuerte autocorrelación en la serie temporal. Además, si en el análisis de correlación parcial (PACF) e inversa también se observan valores significativos para los primeros rezagos, indica que hay autocorrelación significativa incluso después de considerar la correlación en los rezagos anteriores. También se muestra una autocorrelación grande en el retardo 12, por lo que se debe ajustar un modelo estacional.

Por lo tanto, un modelo SARIMA(p, d, q)(P, D, Q) s (si se requiere diferenciación para estacionalizar la serie) podría ser una opción.

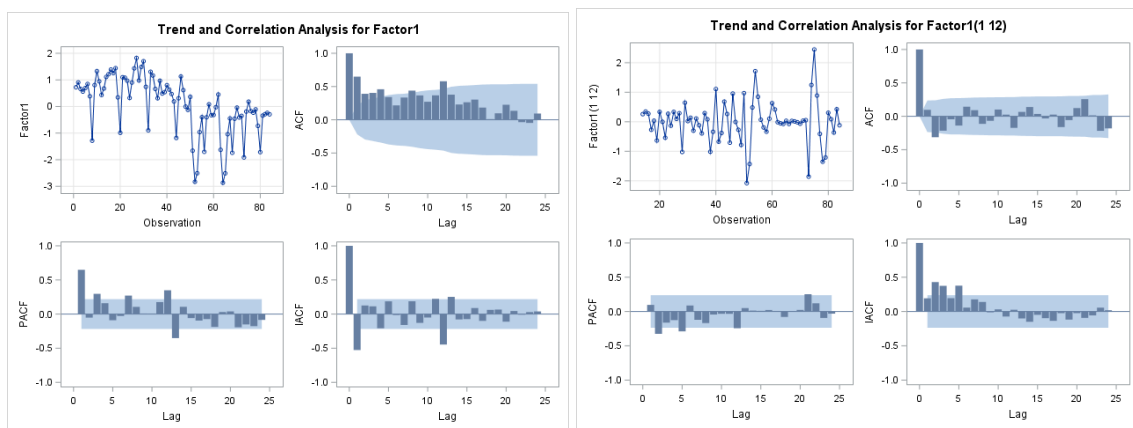


Imagen 7. Análisis autocorrelación estacionamiento

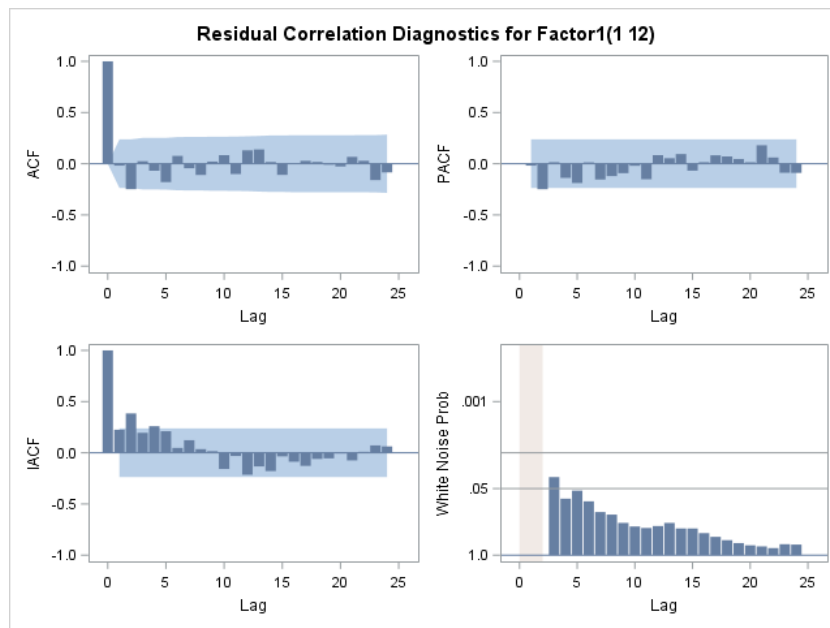
En la representación de la serie con una diferenciación estacional podemos observar que el rango de valores en los que se mueve la serie es constante, exceptuando algunos valores atípicos.

Debido a que fue necesario realizar una diferenciación estacional para que la serie sea estacionaria entonces ajustaremos un modelo ARIMA estacional con $s=12$. Además, se muestra significativo el tercer retardo, $p=3$.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	0.00002756	0.01897	0.00	0.9988	0
MA1,1	0.87922	0.11641	7.55	<.0001	12
AR1,1	-0.29279	0.11699	-2.50	0.0147	3

Tabla 4. Estimación parámetros SARIMA(3,1,0)(0,1,1)12

A pesar de que MU no resulta significativo según su valor p, su exclusión no es viable dado que representa la constante del modelo. Por consiguiente, se opta por ajustar un modelo con $p=3$ y $q=(12)$ para verificar que sus residuales no sean correlacionados.



Autocorrelation Check of Residuals										Correlations of Parameter Estimates			
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations						Parameter	MU	MA1,1	AR1,1
6	8.05	4	0.0899	-0.018	-0.248	0.025	-0.068	-0.179	0.077	MU	1.000	-0.050	0.013
12	12.20	10	0.2719	-0.044	-0.110	0.021	0.084	-0.100	0.131	MA1,1	-0.050	1.000	-0.025
18	15.18	16	0.5115	0.140	0.018	-0.109	0.004	0.029	0.018	AR1,1	0.013	-0.025	1.000
24	19.35	22	0.6235	-0.008	-0.026	0.067	0.029	-0.160	-0.085				

Imagen 8. Análisis residuos SARIMA(3,1,0)(0,1,1)12

Se ha comprobado que los residuales no están correlados basándonos en las siguientes afirmaciones resultantes de la información analizada en este apartado:

- En el correlograma simple (ACF) se observa que todos los coeficientes de autocorrelación están dentro de las bandas de confianza, por lo que los mismos se consideran cero o cercanos a cero.
- En la tabla “Autocorrelation Check of Residuals” se observa que los Pr (p-valor) son grandes, es decir, mayores a 0,005.
- En la tabla “Correlations of Parameter Estimates” se observan bajas correlaciones (Menor a 0,6) entre los estimadores de los parámetros del modelo.

El modelo ajustado en el apartado anterior: SARIMA(3,1,0)(0,1,1)₁₂ tiene la siguiente expresión algebraica:

Ecuación 2. SARIMA(3,1,0)(0,1,1)₁₂

$$(1 + 0.29279B^3)(1 - B^{12})x_t = (1 - 0.87922B^{12})z_t$$

La representación gráfica exhibe tanto las predicciones obtenidas como los intervalos de confianza asociados. Esta visualización permite apreciar de manera clara y concisa las estimaciones futuras y la incertidumbre.

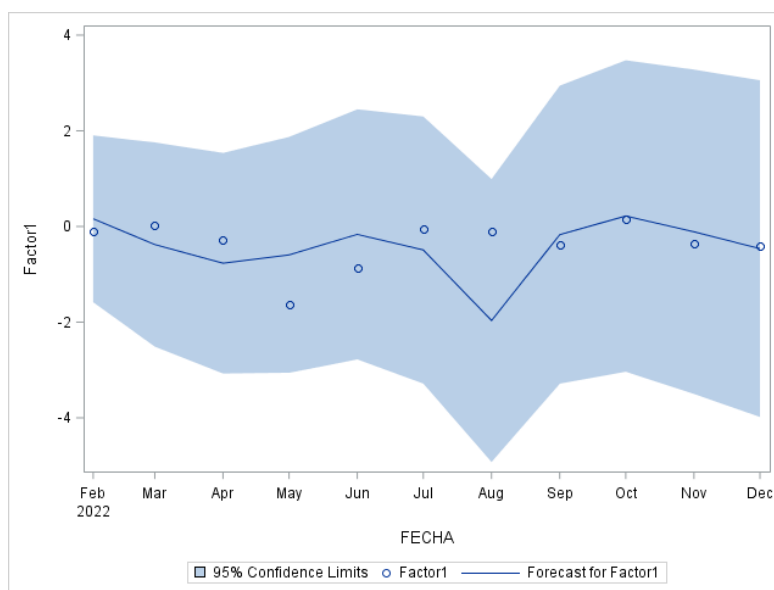


Imagen 9. Predicciones y observaciones estacionamiento

Con el modelo SARIMA seleccionado se obtiene un MSE de -0.0067276. Por lo que en comparación con el método de suavizado es más recomendable trabajar con el modelo SARIMA(3,1,0)(0,1,1)₁₂ ya que tiene un menor error y hará mejores predicciones.

Una vez tenemos el mejor modelo para predecir las futuras multas de estacionamiento, se calculan dichos valores para 2023, es importante recordar que los factores representan combinaciones lineales de las variables originales y no tienen una interpretación directa en términos del dominio del problema, por lo tanto, deshacer el cambio y convertir las predicciones de los factores de nuevo en predicciones de las variables originales nos permite obtener resultados más intuitivos y útiles desde la perspectiva del problema real.

Este cambio se realizará en el siguiente apartado tras haber analizado los modelos y las predicciones para los tres factores seleccionados.

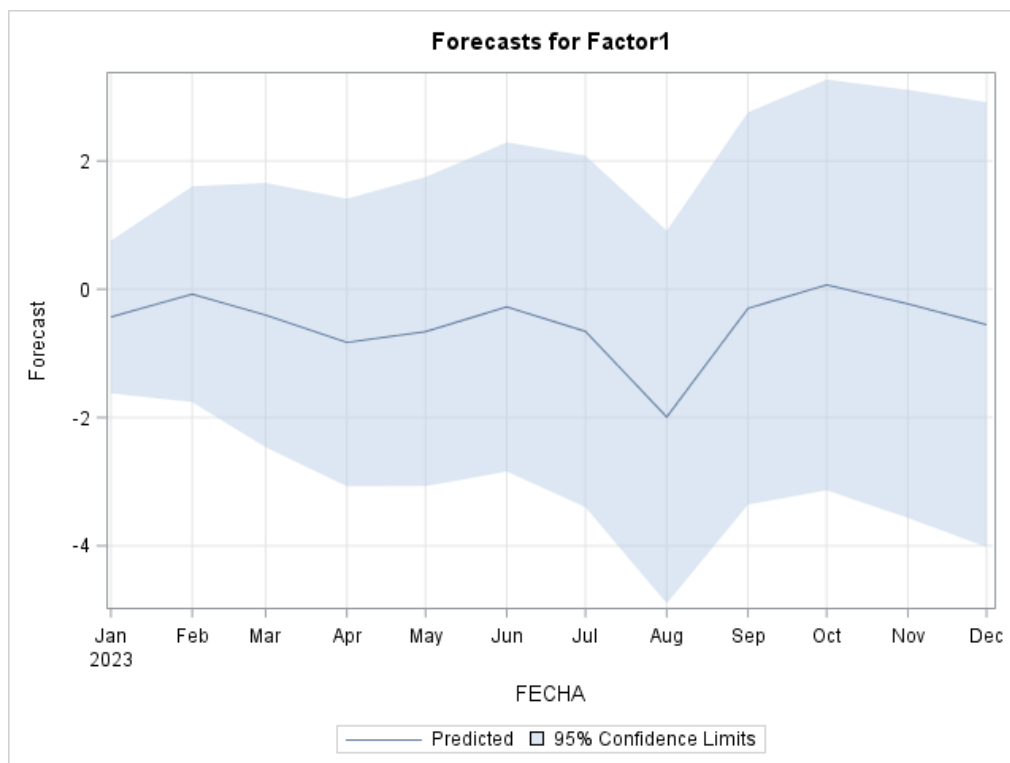


Imagen 10. Predicciones estacionamiento

5.2 Serie Regulación Vial y Control de Tráfico

Se muestra con el siguiente gráfico el modelo que nos muestra la serie original Regulación Vial y Control de Tráfico además de la serie suavizada.

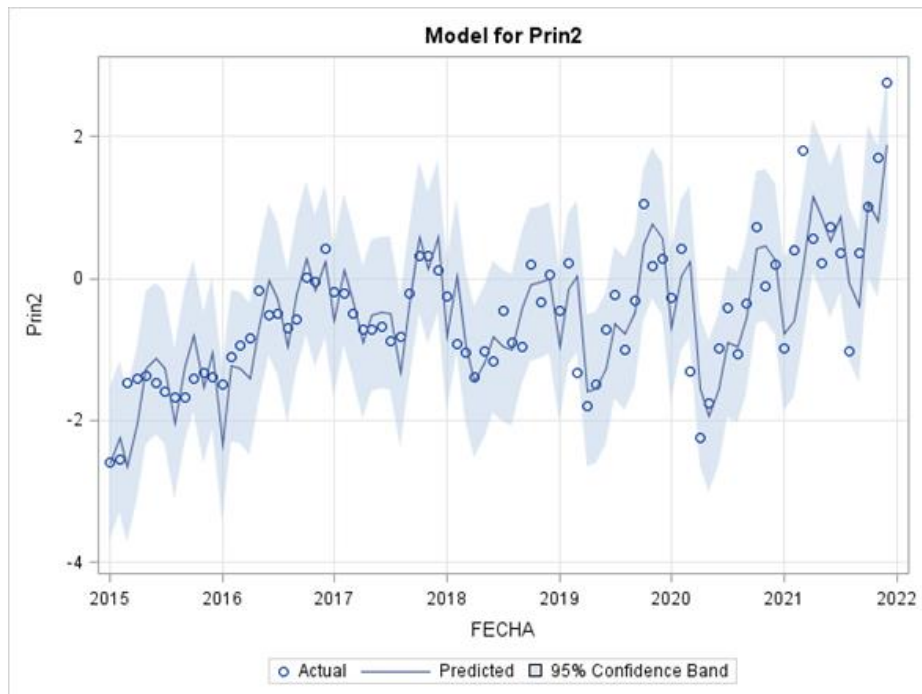


Imagen 11. Serie regulación vial y control de tráfico original y suavizada

Winters Method (Additive) Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.87325	0.07986	10.93	<.0001
Trend Weight	0.0010000	0.0098843	0.10	0.9197
Seasonal Weight	0.0010000	0.16743	0.01	0.9952

Tabla 5. Parámetros de Holt Winters para regulación vial y control de tráfico

Ecuación 3

$$Y_t = 0.87325L_{t-1} + 0.001T_{t-1} + 0.001S_{t-1}$$

El coeficiente de nivel tiende a 1, lo que sugiere que los datos del pasado no influyen en los cambios de nivel. Esto implica que la serie mantiene su nivel basado únicamente en el período estacional del instante actual, sin considerar los niveles anteriores.

Por otro lado, un trend weight cercano a 0 en el método de Holt-Winters indica que el modelo otorga muy poca importancia a la tendencia al realizar las predicciones. Esto puede deberse a varios factores, como la falta de una tendencia clara en los datos históricos o la presencia de incertidumbre en la dirección de la tendencia.

En cuanto al coeficiente de estacionalidad, un valor que tiende a cero indica que los datos del pasado sí influyen en la estacionalidad y que todos los coeficientes de estacionalidad son iguales a lo largo del tiempo.

Se generará una gráfica que mostrará los valores observados de la variable junto con los valores suavizados y las predicciones obtenidas a partir de un modelo específico de series temporales. Además, se incluirán los intervalos de confianza para las predicciones, lo que permitirá visualizar la incertidumbre asociada a las estimaciones futuras.

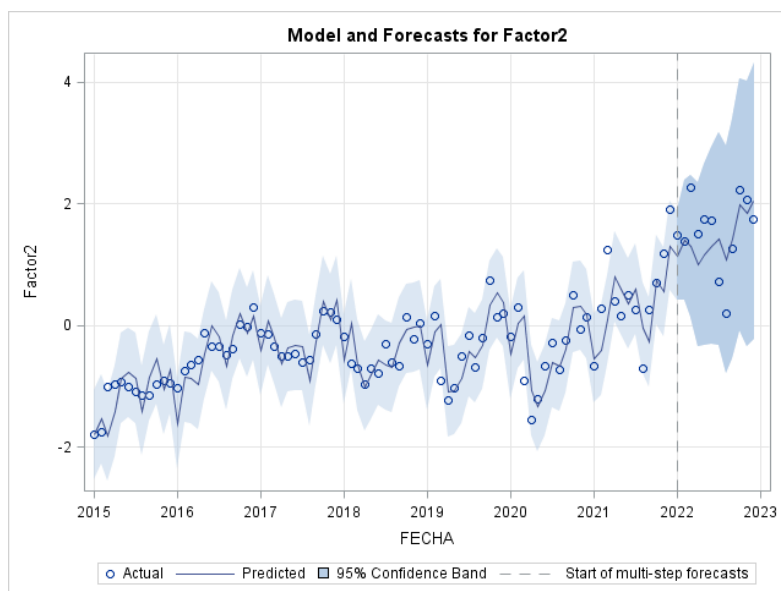


Imagen 12. Observados y predicciones regulación vial y control de tráfico

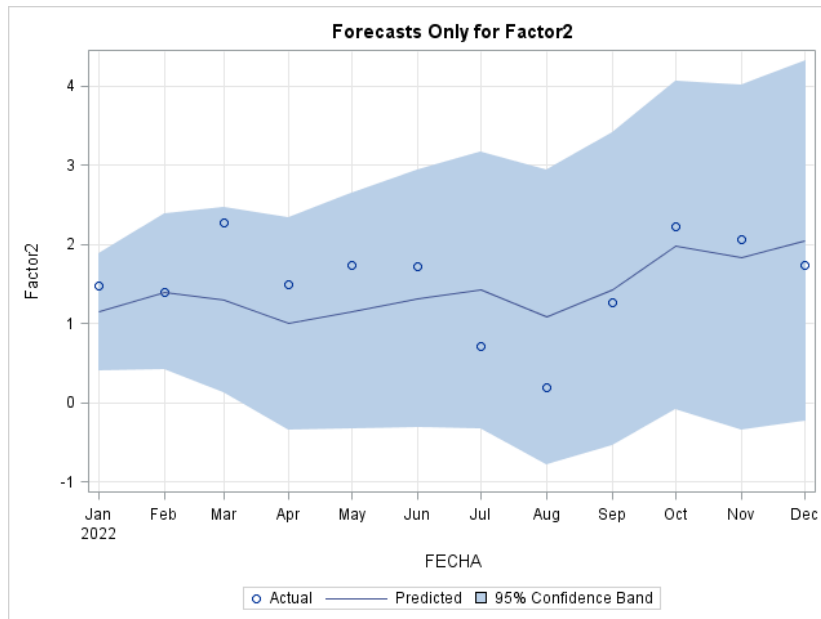


Imagen 13. Observados y predicciones regulación vial y control de tráfico 2022

Con este método de suavizado se obtiene un MSE (Mean Absolute Error) de 0.41033061.

Para seleccionar el ARIMA adecuado nos fijaremos que en el correlograma de la variable de Regulación Vial y Control de Tráfico se evidencia un decrecimiento notable en su análisis de correlación y la presencia de una correlación significativa solo para el sexto retardo en el análisis de correlación parcial, además se observa tendencia por lo que se diferenciará la serie.

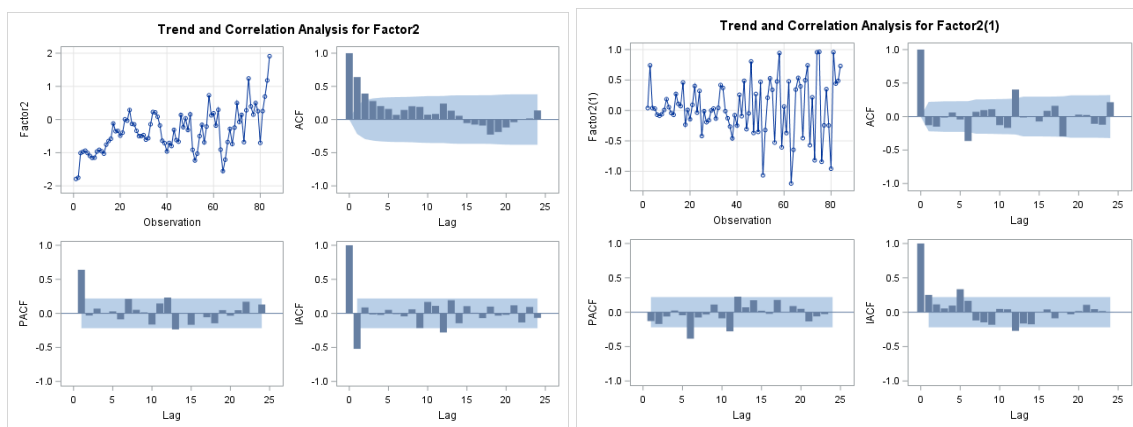


Imagen 14. Análisis de autocorrelaciones de regulación vial y control de tráfico

Tras diferenciar se observa una significación en el retardo 6 por lo que se supondrá un modelo con $p=6$ y un $q=12$.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	0.04827	0.04886	0.99	0.3261	0
MA1,1	-0.44515	0.14906	-2.99	0.0037	12
AR1,1	-0.27410	0.12433	-2.20	0.0303	6

Tabla 6. Estimación parámetros ARIMA(6,1,12)

En la tabla previa se aprecia que los parámetros son significativos, por lo que se decide que el modelo a ajustar para comprobar que sus residuales no están correlados es con $p=(6)$ y $q=(12)$.

Autocorrelation Check of Residuals										Correlations of Parameter Estimates			
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations						Parameter	MU	MA1,1	AR1,1
6	7.16	4	0.1277	-0.160	-0.097	0.023	-0.075	-0.198	0.021	MU	1.000	-0.054	0.035
12	10.91	10	0.3646	0.055	0.055	0.147	-0.093	-0.033	0.043	MA1,1	-0.054	1.000	-0.298
18	13.43	16	0.6409	0.039	-0.001	-0.043	0.034	0.109	-0.086	AR1,1	0.035	-0.298	1.000
24	17.42	22	0.7399	-0.041	0.038	0.004	-0.081	-0.132	0.085				

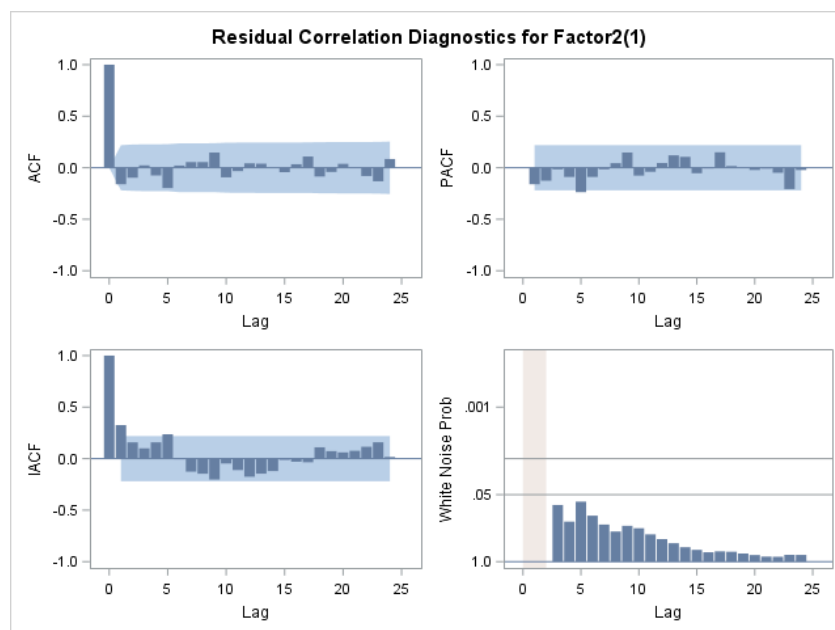


Imagen 15. Análisis de residuos de regulación vial y control de tráfico

Se ha comprobado que los residuales están incorrelados basados en las siguientes afirmaciones resultantes de la información analizada en este apartado:

- En el correlograma simple (ACF) se observa que todos los coeficientes de autocorrelación están dentro de las bandas de confianza, por lo que los mismos se consideran cero o cercanos a cero.
- En la tabla “Autocorrelation Check of Residuals” se observa que los Pr (p-valor) son grandes, es decir, mayores a 0,05.
- En la tabla “Correlations of Parameter Estimates” se observan bajas correlaciones (Menor a 0,6) entre los estimadores de los parámetros del modelo.

Modelo ajustado en el apartado anterior: ARIMA(6,1,12)

Ecuación 4. ARIMA(6,1,12)

$$(1 + 0.45562B^6)(1 - B)x_t = 0.03666 + (1 - 0.25047B^{12})z_t$$

Esta visualización permite apreciar de manera clara y concisa las estimaciones futuras y la incertidumbre relacionada con dichas predicciones en forma de intervalo de confianza que se muestra sombreado en el gráfico, el cuál va aumentando con el paso del tiempo para las estimaciones.

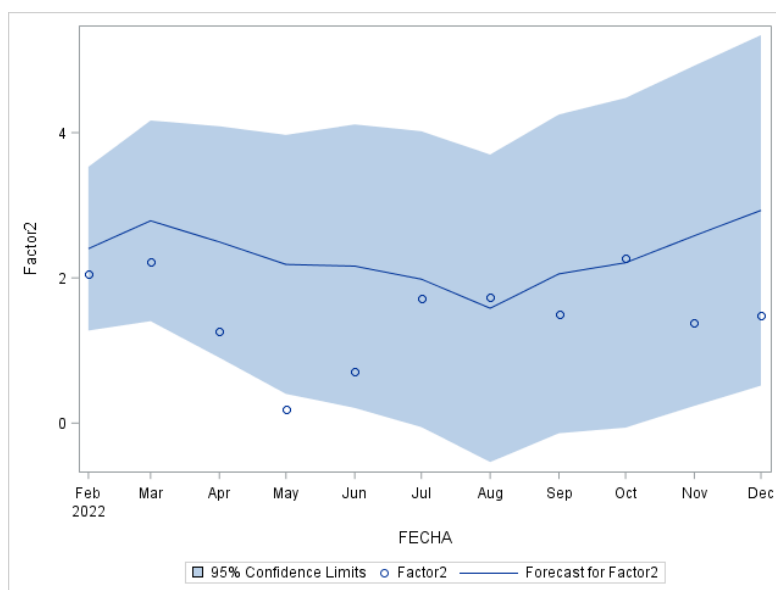


Imagen 16. Predicciones y observaciones de regulación vial y control de tráfico

Para el modelo ARIMA seleccionado obtenemos un MSE de 0.000341818, por lo que será mucho más recomendable trabajar con este modelo en vez de con el de suavizado para hacer las predicciones.

Una vez identificado el mejor modelo para predecir las futuras multas de regulación vial y control de tráfico, se calculan estos valores para los meses de 2023, de los cuales no se tenían datos para la segunda mitad de año. Es importante recordar que los factores representan combinaciones lineales de las variables originales y no tienen una interpretación directa en términos del dominio del problema. Por lo tanto, revertir esta transformación y convertir las predicciones de los factores nuevamente en predicciones de las variables originales nos permite obtener resultados más intuitivos y útiles desde la perspectiva del problema real.

Este proceso se llevará a cabo en el apartado 5.4, después de haber analizado los modelos y las predicciones para los tres factores seleccionados.

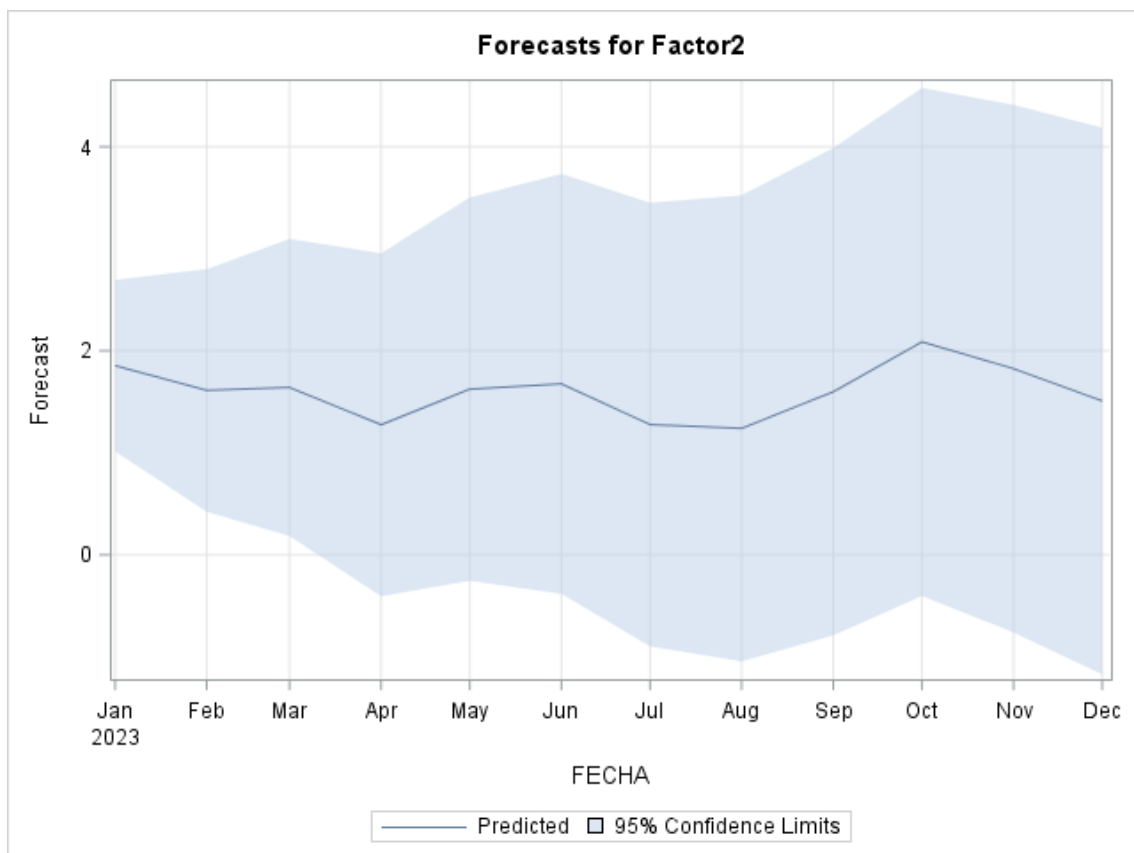


Imagen 17. Predicciones regulación vial y control de tráfico

5.3 Serie Regulación de Velocidad y Seguridad Vial

Por último, se analiza la serie de Regulación de Velocidad y Seguridad Vial.

Se muestra al igual que para las variables anteriores el gráfico de la serie original junto a la suavizada con el modelo de Holt-Winters.

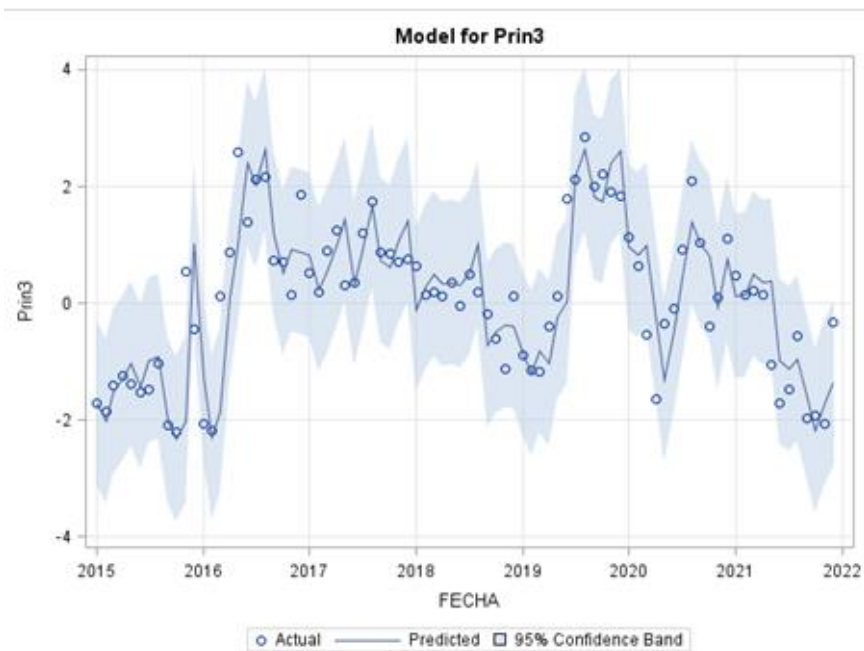


Imagen 18. Serie regulación de velocidad y seguridad vial original y suavizada

Winters Method (Additive) Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.91388	0.08207	11.14	<.0001
Trend Weight	0.0010000	0.02148	0.05	0.9630
Seasonal Weight	0.0010000	0.54642	0.00	0.9985

Tabla 7. Parámetros de Holt Winters de regulación de velocidad y seguridad vial

Ecuación 5

$$Y_t = 0.91388L_{t-1} + 0.001T_{t-1} + 0.001S_{t-1}$$

El coeficiente de nivel, al tender a 1, señala que los datos pasados tienen un escaso impacto en los cambios de nivel. Esto implica que la serie mantiene su nivel basado únicamente en el período estacional del instante t , sin considerar los niveles previos.

Por otro lado, un trend weight cercano a 0 en el método de Holt-Winters indica que el modelo subestima la importancia de la tendencia al realizar las predicciones.

Asimismo, un coeficiente de estacionalidad que tiende a cero sugiere que los datos pasados influyen en la estacionalidad, con todos los coeficientes de estacionalidad siendo iguales a lo largo del tiempo.

En cuanto a la visualización, se genera una gráfica que exhibe los valores observados de la variable de estacionamiento junto con los valores suavizados y las predicciones derivadas de un modelo específico de series temporales. Además, se incorporan intervalos de confianza para las predicciones, lo que facilita la evaluación de la incertidumbre asociada a las estimaciones futuras. Esta representación visual proporciona una herramienta efectiva para evaluar la capacidad predictiva del modelo y comprender mejor cómo se comportan los datos observados en comparación con las proyecciones realizadas.

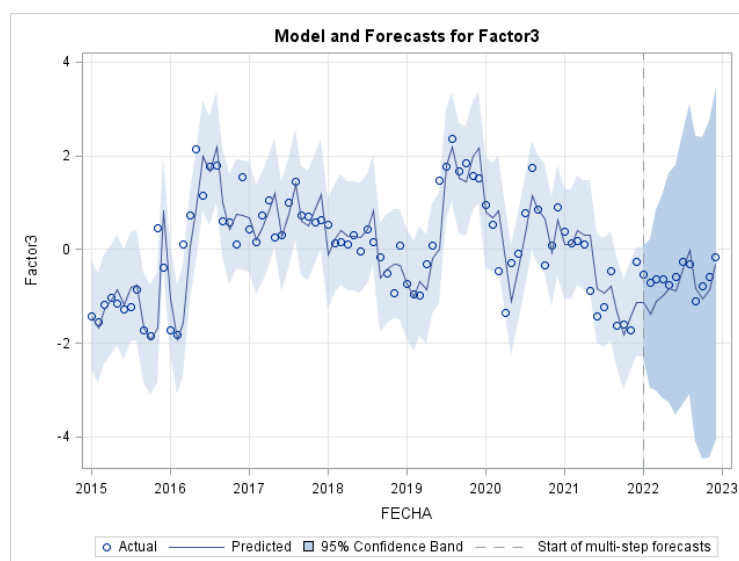


Imagen 19. Observados y predicciones regulación de velocidad y seguridad vial

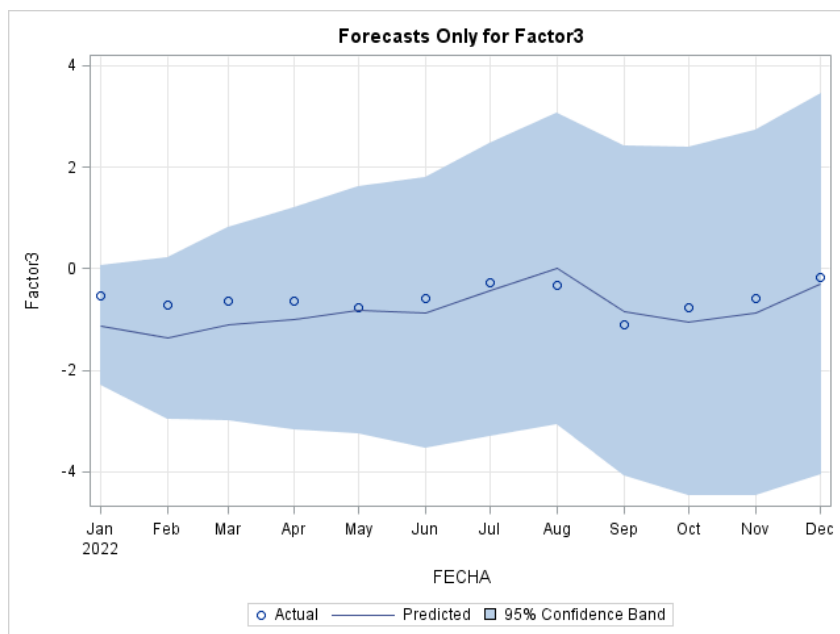


Imagen 20. Observados y predicciones regulación de velocidad y seguridad vial 2022

Al igual que en las anteriores series, el intervalo de confianza va aumentando junto al error, obteniendo un Mean Square Error de 0.34404492.

Tras observar el método de suavizado se hará la selección del modelo ARIMA, para la selección del más adecuado nos fijaremos que en el correlograma de la variable de Regulación de Velocidad y Seguridad Vial se presencia un decrecimiento en su análisis de correlación simple, además es necesario diferenciar la serie.

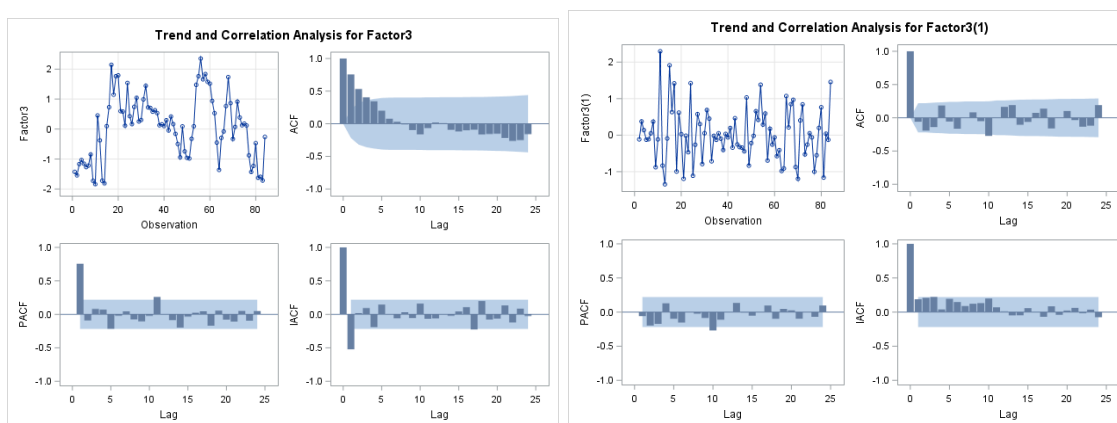


Imagen 21. Análisis de autocorrelaciones regulación de velocidad y seguridad vial

Tras diferenciar se observa una significación en el retardo 10 únicamente por lo que se supondrá un modelo con $p = (10)$.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MU	0.01163	0.06053	0.19	0.8481	0
AR1,1	-0.31083	0.11355	-2.74	0.0076	10

Tabla 7. Estimación parámetros ARIMA(10,1,0)

Se aprecia que los parámetros son significativos, por lo que se trabaja con el modelo seleccionado para comprobar que sus residuales no están relacionados

Autocorrelation Check of Residuals										Correlations of Parameter Estimates		
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations						Parameter	MU	AR1,1
6	7.13	5	0.2110	-0.089	-0.146	-0.088	0.129	-0.110	-0.121	MU	1.000	-0.013
12	8.56	11	0.6621	0.015	-0.024	-0.072	0.013	-0.033	0.086	AR1,1	-0.013	1.000
18	16.72	17	0.4732	0.151	-0.019	-0.119	0.061	0.120	-0.149			
24	23.00	23	0.4609	-0.018	0.043	-0.031	-0.091	-0.096	0.181			

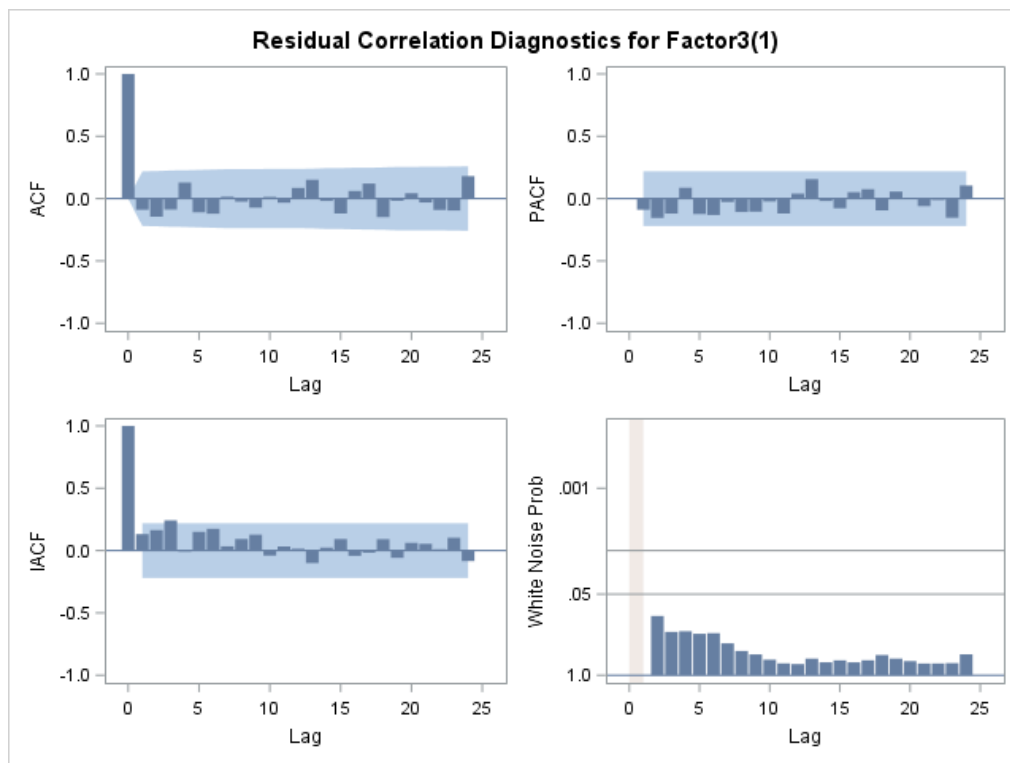


Imagen 22. Análisis de residuos de ARIMA(10,1,0)

Se ha comprobado que los residuales están incorrelados basados en las siguientes afirmaciones resultantes de la información analizada en este apartado:

- En el correlograma simple (ACF) se observa que todos los coeficientes de autocorrelación están dentro de las bandas de confianza, por lo que los mismos se consideran cero o cercanos a cero.
- En la tabla “Autocorrelation Check of Residuals” se observa que los Pr (p-valor) son grandes, es decir, mayores a 0,05.
- En la tabla “Correlations of Parameter Estimates” se observan bajas correlaciones (Menor a 0,6) entre los estimadores de los parámetros del modelo.

Modelo ajustado en el apartado anterior: ARIMA(10,1,0)

Ecuación 6. ARIMA(10,1,0)

$$(1 + 0.31083B^{10})(1 - B)x_t = 0.01163 + z_t$$

Esta visualización permite apreciar de manera clara y concisa las estimaciones futuras y la incertidumbre relacionada con dichas predicciones en forma de intervalo de confianza que se muestra sombreado en el gráfico, el cuál va aumentando con el paso del tiempo para las estimaciones.

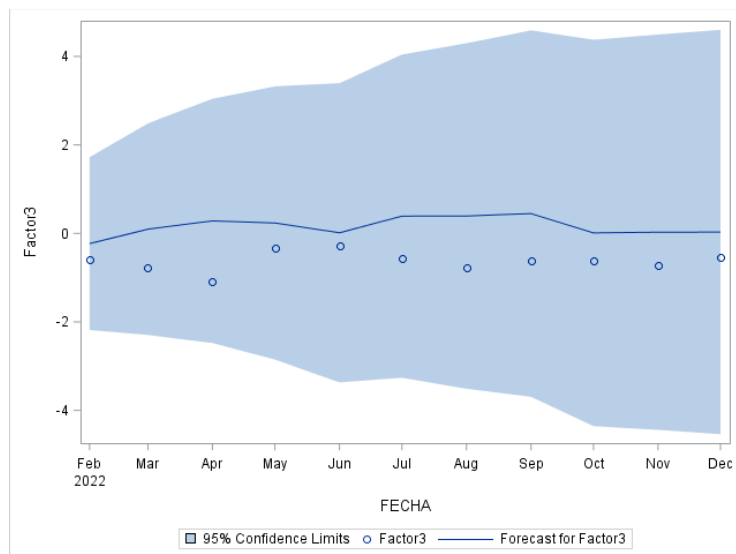


Imagen 23. Predicciones y observaciones de regulación de velocidad y seguridad vial

Para el modelo ARIMA(10,1,0) calculamos el error medio obteniendo un 0.0050441 menor que el obtenido por el método de suavizado, este será el modelo a utilizar para las predicciones.

Una vez que se haya determinado el modelo más adecuado para prever las posibles sanciones relacionadas con el cumplimiento de límites de velocidad y seguridad vial, se procederá a calcular estos valores para el año 2023, donde no se contaba con la segunda mitad de los datos.

Al igual que en las series anteriores estos valores no representan a las variables originales pero este cambio se hará en el siguiente apartado.

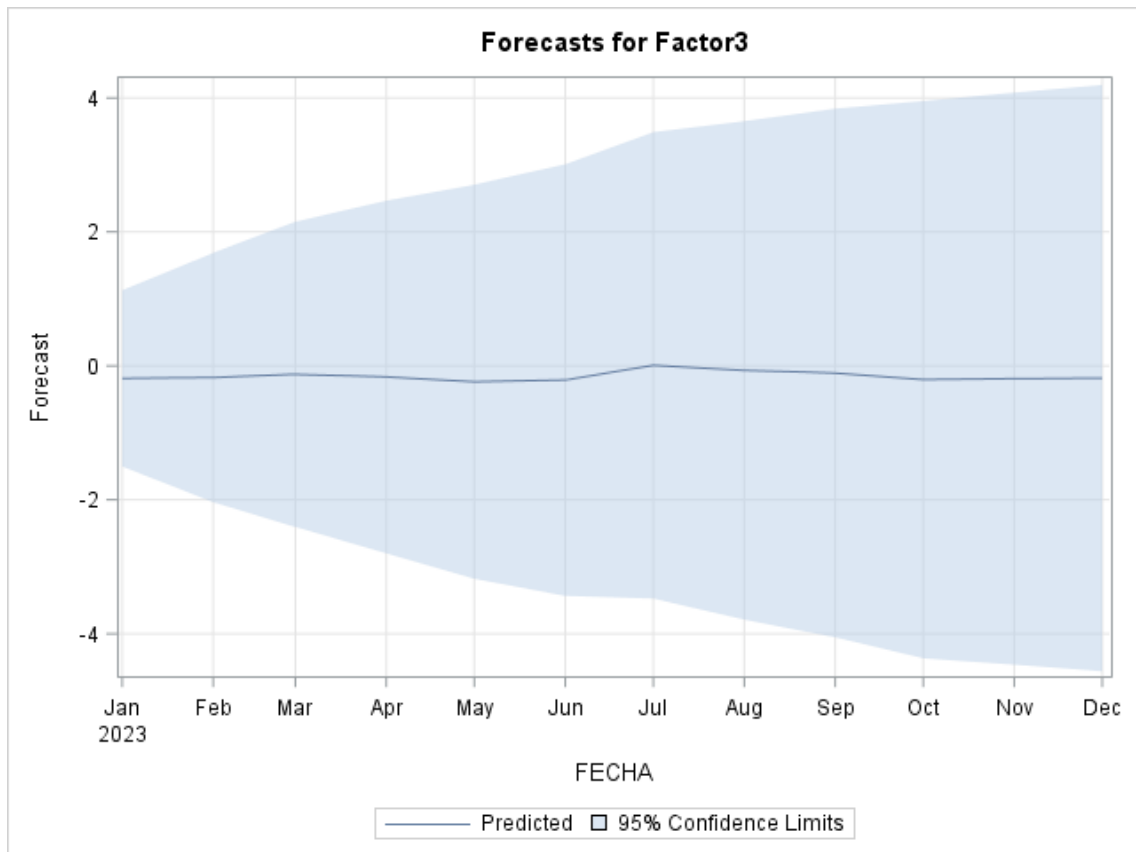


Imagen 24. Predicciones regulación de velocidad y seguridad vial

5.4 Predicciones de las variables originales

En el análisis factorial, al pronosticar los factores subyacentes en lugar de las variables originales, simplificamos el proceso de modelado al trabajar con un conjunto reducido de dimensiones. Sin embargo, es crucial recordar que estos factores carecen de una interpretación directa.

Al deshacer el cambio y volver a las variables originales, conservamos la capacidad de realizar análisis detallados sobre cada variable individualmente, lo que es esencial para comprender mejor los factores que influyen en el comportamiento de las multas de tráfico. Esto nos permite tomar decisiones informadas basadas en insights específicos de cada variable, en lugar de depender únicamente de las predicciones de los factores.

Para obtener dichos valores originales se trabajará con la matriz de cargas indica cómo cada variable original está asociada con cada uno de los factores extraídos.

Una vez que tenemos las predicciones de los factores y las cargas factoriales, simplemente multiplicamos las predicciones de los factores por las cargas factoriales inversas para obtener las predicciones de las variables originales, pero sin estandarizar, para obtener valores reales se multiplicará por la desviación típica de la variable original y se le sumará su media.

Este proceso permite transformar las predicciones de los factores de nuevo a las predicciones de las variables originales, manteniendo la misma relación entre las variables y los factores que se estableció durante el análisis factorial.

Una vez realizado dicho proceso obtenemos el siguiente gráfico que nos muestra los valores de dichas variables para 2023 cuya segunda mitad desconocíamos los valores reales.

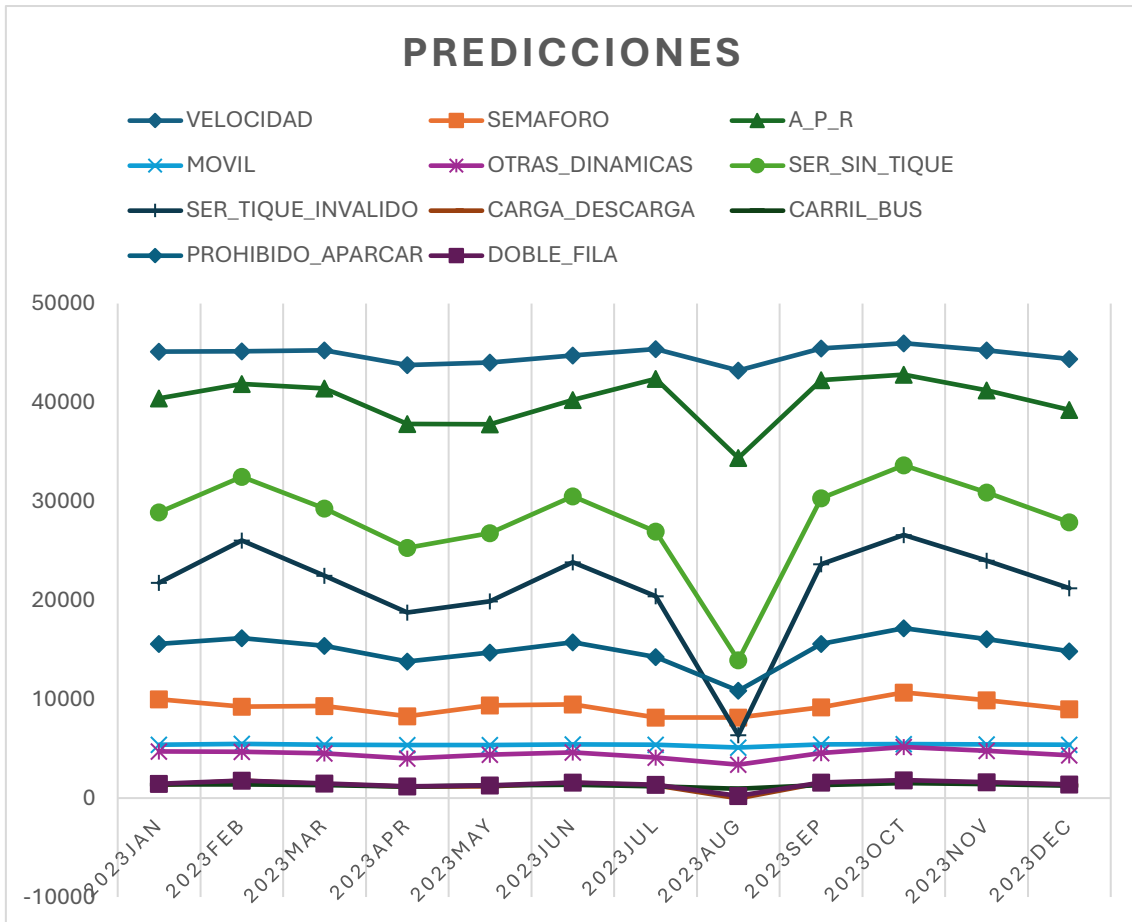


Imagen 25. Predicciones variables originales

Varias infracciones presentan fluctuaciones mensuales marcadas, especialmente en junio y mayo, donde se observa un aumento significativo en el número de infracciones seguido de una disminución. Estos picos pueden ser el resultado de intervenciones específicas de control de tráfico o variaciones estacionales en el comportamiento de los conductores.

Infracciones como "SEMÁFORO", "CARGA_DESCARGA", "CARRIL_BUS", "PROHIBIDO_APARCAR" y "DOBLE_FILA" muestran menor variabilidad y se mantienen bastante constantes a lo largo del tiempo, sugiriendo una incidencia regular y controlada, mientras que "SERSINTIQUE" y "SERTIQUEINVALIDO" muestran mayor variabilidad

Las fluctuaciones significativas en ciertos meses sugieren que las intervenciones específicas, como campañas de vigilancia intensificadas, tienen un impacto notable en el número de infracciones, lo cual es reflejado en las predicciones.

6. Conclusiones

El análisis temporal de multas de tráfico en Madrid ha revelado patrones significativos en la emisión de infracciones. Los datos muestran variaciones estacionales claras, con picos durante ciertos meses del año, lo que sugiere una relación con factores estacionales y eventos específicos de la ciudad. Las multas se agrupan en varias categorías predominantes, como exceso de velocidad, violaciones de semáforos y estacionamiento indebido. Cada categoría muestra un comportamiento distinto a lo largo del tiempo, permitiendo identificar áreas críticas para la intervención.

Un aspecto destacado de este estudio ha sido la aplicación de modelos de series temporales para analizar y prever la evolución de las multas. Utilizando modelos ARIMA y Holt-Winters, se ha logrado capturar tanto las tendencias a largo plazo como las variaciones estacionales y cíclicas en los datos. El modelo ARIMA (AutoRegressive Integrated Moving Average) ha sido particularmente útil para identificar patrones en los datos históricos y proyectar estos patrones en el futuro. Por otro lado, el modelo Holt-Winters ha sido eficaz en la descomposición de la serie temporal en componentes de nivel, tendencia y estacionalidad, proporcionando una visión más detallada de las fluctuaciones a lo largo del tiempo.

Los hallazgos pueden ayudar a las autoridades a desarrollar estrategias de control de tráfico más focalizadas y eficientes, como aumentar la vigilancia en períodos de alta incidencia de ciertas infracciones. Además, la identificación de patrones temporales y factores subyacentes permite diseñar políticas de seguridad vial más efectivas, como campañas de concienciación específicas durante los meses con mayor número de infracciones.

Se recomienda complementar el análisis temporal con un análisis espacial para identificar zonas específicas de la ciudad con mayores incidencias de multas. También sería útil investigar el impacto de factores externos como condiciones meteorológicas, eventos especiales y cambios en las normativas de tráfico sobre la emisión de multas. Integrar tecnologías avanzadas como el aprendizaje

automático y la inteligencia artificial podría mejorar la precisión de los modelos predictivos y explorar patrones más complejos en los datos.

Las limitaciones del estudio incluyen la disponibilidad y calidad de los datos, que pueden haber restringido algunos aspectos del análisis. Se recomienda futuras investigaciones con conjuntos de datos más amplios y detallados. Además, los resultados obtenidos son específicos para la ciudad de Madrid y pueden no ser generalizables a otras ciudades con características de tráfico diferentes.

Este estudio proporciona una comprensión profunda de la dinámica temporal de las multas de tráfico en Madrid, ofreciendo una base sólida para futuras investigaciones y aplicaciones prácticas en la gestión del tráfico y la seguridad vial. Las herramientas y modelos desarrollados pueden ser adaptados y aplicados en otros contextos urbanos para mejorar la movilidad y la seguridad en las vías.

7. Bibliografía

Ghanim, Khan, A. A., Hamed, & Al-Maniri, A. (2019). *Diagnostic time series models for road traffic accidents data*.

Montgomery, D., & Jennings, C. (2008). *Introduction to Time Series Analysis and Forecasting*.

Brocklebank, J. C., & Dickey, D. A. (2003). *SAS® for Forecasting Time Series*. SAS Institute Inc.

Alonso Revenga, J. M. (2012). *Series Temporales. Análisis Practico Con SPSS y SAS*. Eae Editorial Academia Espanola.

Peña, D. (2002). *Análisis de Datos Multivariantes*. McGraw-Hill.

BOAM no 8358/496 (15/03/2019) - Decreto de 14 de marzo de 2019 de la Delegada del Área de Gobierno de Medio Ambiente y Movilidad por el que se establecen medidas en ejecución del Acuerdo de la Junta de Gobierno de la ciudad de Madrid, de 29 de octubre de 2018, por el que se desarrolla el régimen de gestión y funcionamiento de la zona de bajas emisiones "Madrid Central". - SEDE ELECTRÓNICA. Disponible en:

<https://sede.madrid.es/portal/site/tramites/menuitem.b4c91589e7f6a5d829da39e5a8a409a0/?vqnextoid=7d9a44c351c79610VqnVCM1000001d4a900aRCRD&vqnextchannel=741d814231ede410VqnVCM1000000b205a0aRCRD&vqnextfmt=pda>

El Gobierno decreta el estado de alarma para hacer frente a la expansión de coronavirus COVID-19. Disponible en:

https://www.lamoncloa.gob.es/consejodeministros/resumenes/Paginas/2020/1403_2020_alarma.aspx

8. Anexo

Código SAS:

```
DATA MULTAS;
DATE2='30DEC2014'D;
T+1;
FECHA=INTNX("MONTH",DATE2,T);
FORMAT FECHA YYMON7.;
INPUT
VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS SER_SIN_TIQUE
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR
DOBLE_FILA;
multas_totales_dinamicas=
velocidad+semaforo+A_P_R+MOVIL+OTRAS_DINAMICAS;
MULTAS_TOTALES_ESTATICAS=SER_SIN_TIQUE+SER_TIQUE_INVALIDO+CARGA_DESCAR
GA+CARRIL_BUS+PROHIBIDO_APARCAR+DOBLE_FILA;
multas_totales=multas_totales_dinamicas+multas_totales_estaticas;
CARDS;
33747 1444 24102 622 2440 39530 45814 7256 1042 8848 2112
31941 1558 26077 714 2290 39587 47701 7147 1247 9146 2340
32655 1966 23851 835 2844 40233 44380 3894 1149 12943 2138
33779 2265 26535 773 2767 42902 44828 3700 1044 13047 1812
31985 2097 26291 810 2977 43349 43778 3876 1140 13056 2128
30313 2260 27126 840 2661 45965 46354 3805 1238 13162 2398
30585 2536 24452 766 2396 48435 45112 2868 1069 10599 1711
40543 1754 14 386 1796 31181 19013 2119 630 6890 584
31012 1808 37 865 2995 46619 45607 4092 995 13810 2054
30575 2206 48 961 3200 47428 46215 4522 1278 15439 3212
27638 1767 115141 896 2655 42340 39098 3957 1294 13928 2724
26170 1822 72215 684 2528 45197 33520 3376 1035 13250 2366
29374 1842 36 680 2502 44699 44282 3342 1054 14993 2836
30102 3827 43 779 2639 46930 41661 3832 1244 15816 3968
57026 4338 16309 819 2633 48125 42372 3720 997 15691 4167
47956 5250 73684 909 2989 45538 41425 4248 895 16467 4165
71506 5402 79133 961 3084 42923 40537 3838 985 17035 3233
57715 4077 64566 921 3798 47811 43778 3963 892 16772 3728
67186 3706 63712 692 3155 42753 36894 2824 770 12535 2008
65780 2591 55966 460 2880 31561 18913 1837 461 8504 733
51153 4649 55623 908 3335 47078 42011 3443 891 16001 3064
50237 6003 60373 800 3590 41847 37953 3812 930 17736 3416
43458 5175 57044 712 3375 41152 37262 3660 1155 16987 3514
69560 5044 46599 587 3793 34853 31304 2886 827 15477 2613
48935 4717 52824 561 3406 46558 42598 3203 936 16159 3521
45836 4466 50978 641 3901 49534 43728 4023 1021 17657 4479
54494 4333 54155 969 4272 49274 45422 4522 832 18549 4478
56411 2068 52567 846 3879 45269 38959 3199 743 16233 3151
45381 2274 52281 1013 4032 46135 41054 4157 1009 18269 3478
44855 2198 56675 1121 4173 50721 45068 3715 1130 17804 3584
53738 2139 57427 802 3747 47479 39097 2974 684 14295 2574
59216 1911 58548 463 2465 33315 19830 1696 723 8727 924
51861 2608 57144 917 4082 49052 41165 3398 1083 18187 2948
50726 2418 58060 980 4998 35913 33351 3302 1316 17302 3542
49772 2129 55462 721 4179 32761 30298 3224 1376 16944 2888
49015 2626 51744 674 4246 31562 30133 2442 955 15989 2651
49096 2386 49111 768 4437 42508 42742 2955 816 17464 3016
39848 1948 48107 739 4108 39170 37088 3074 396 16671 2241
38441 2159 52247 695 3594 36521 35496 2964 470 16969 3233
39023 2116 51567 895 3152 39184 40801 3286 491 17883 2754
42875 2280 50381 836 3485 37536 40466 2981 510 17590 2398
36804 2431 49694 765 3213 38941 41748 2781 415 17614 1980
```

46425	2348	49960	665	4030	39630	40356	2167	779	12953	1624	
39016	1533	49956	428	2946	28031	18177	1439	759	7306	767	
38306	2116	43789	767	3111	38680	39498	2373	812	14793	1887	
39229	2052	37666	781	4459	42099	42960	3269	1541	17293	2938	
30778	1924	35437	647	3672	39209	38675	3066	1349	15747	2591	
58791	1633	37	565	4158	29857	28893	2389	1116	13446	2423	
44568	1502	0	388	2761	38121	31496	1809	1333	12086	2699	
40814	1726	0	556	3334	33230	30502	2199	1423	19149	2900	
33720	1895	4161	333	1606	14492	13158	1102	675	8867	1245	
39061	1816	7358	116	931	0	4	489	21	6631	388	
44817	2071	13369	239	1842	1	3	823	63	6288	926	
50068	2068	79974	394	2551	24442	21926	1365	557	10973	1326	
56859	2217	82800	436	2125	34451	30749	1184	963	14406	1301	
63103	1677	80758	239	1544	21630	12410	762	375	8419	633	
46252	2023	106816		374	2199	30833	30155	1598	919	15298	1392
51428	2137	120305		410	2532	29165	28329	2105	2135	15777	1889
43825	2023	110964		364	2625	24777	22567	1787	1250	15503	2343
51575	1753	86130	442	3226	25891	23815	1762	1175	14871	1957	
49344	1503	68558	388	2761	38121	31496	1809	1333	12086	2699	
47387	1804	54518	556	3334	33230	30502	2199	1423	19149	2900	
34405	1632	27332	333	1606	14492	13158	1102	675	8867	1245	
17145	718	14145	116	931	0	4	489	21	6631	388	
30232	1415	32764	239	1842	1	3	823	63	6288	926	
33321	2157	50371	394	2551	24442	21926	1365	557	10973	1326	
46886	2176	61617	436	2125	34451	30749	1184	963	14406	1301	
60857	1696	56926	239	1544	21630	12410	762	375	8419	633	
48853	2078	61378	374	2199	30833	30155	1598	919	15298	1392	
37807	1956	52049	410	2532	29165	28329	2105	2135	15777	1889	
35183	1561	60603	364	2625	24777	22567	1787	1250	15503	2343	
47317	1863	68784	442	3226	25891	23815	1762	1175	14871	1957	
43988	2486	41847	190	1648	13453	11344	864	579	7382	975	
42374	4182	50320	370	2750	32870	27746	1875	1141	16757	1998	
49794	10506	52967	498	3195	33745	31097	2363	1378	17032	2098	
47013	5939	38356	493	3043	29297	26400	1892	1016	16121	2258	
42255	4772	0	578	3124	27960	26108	1912	974	16553	1980	
35778	7204	0	582	3137	30888	27869	1887	1198	16229	2083	
38053	8105	0	437	2311	32463	26903	1185	866	13497	1173	
43896	2561	0	244	1734	25369	13108	822	458	9386	501	
29940	7075	0	526	2963	31816	26727	1711	827	16823	1650	
33262	8612	0	387	2928	31891	26428	2019	1005	18053	1779	
33844	9097	0	326	2933	29064	24005	2068	1512	18526	1999	
45801	9457	31887	353	5614	27222	21458	1743	1191	16663	1649	
35920	9444	45215	239	4817	31303	24998	1676	1061	15308	1644	
36217	11098	45618	331	3631	30515	26453	2040	1294	15195	1434	
39978	13654	45433	339	5915	32209	29797	2613	1223	16532	2143	
38037	10317	40040	342	4667	29822	25668	1912	1165	14497	1458	
38220	12047	42171	514	4526	29704	27105	2155	1436	15111	1435	
42691	11043	37575	466	4690	32550	29053	2087	1467	14790	1531	
44303	9431	36217	307	2925	31499	23894	1508	997	10807	1025	
39466	7395	32677	239	2832	27085	12378	929	590	7569	436	
33372	11439	34185	463	3747	35062	26698	1899	1207	14572	1318	
43139	12747	32856	451	4899	34370	25664	2105	1685	15955	1536	
45710	11454	30516	410	4344	30453	22354	1994	1706	16804	1679	
48439	10280	35949	354	3865	28288	19024	1553	1567	15098	1674	
48294	9692	46451	358	6797	36400	26488	1467	1410	16068	1433	
45532	9522	45248	419	6923	32457	23196	1577	1375	14451	1302	
42694	9477	48044	547	7004	34278	25357	1876	1474	14710	1507	
51388	8346	46857	474	6098	32151	22673	1534	1050	12893	1070	
58321	9424	44151	405	4707	31979	24910	1577	1038	14421	1121	
46614	9457	46176	411	4455	34566	25779	1538	1509	13732	1111	

;

```
PROC PRINT;RUN;
```

```
***
```

```
*depuracion
```

```
*****
```

```
*PASO 1: ANALIZAR DATOS MISSING;
```

```
PROC FREQ DATA=MULTAS;  
TABLES _NUMERIC_ / MISSING;  
RUN;
```

```
*PASO 2: ANALIZAR OUTLIERS;
```

```
PROC UNIVARIATE DATA=MULTAS;  
VAR VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS  
SER_SIN_TIQUE  
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR  
DOBLE_FILA  
multas_totales_dinamicas MULTAS_TOTALES_ESTATICAS multas_totales;  
OUTPUT OUT=OUTLIERS;  
RUN;
```

```
PROC SORT DATA=OUTLIERS;  
BY DESCENDING VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS  
SER_SIN_TIQUE  
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR  
DOBLE_FILA  
multas_totales_dinamicas MULTAS_TOTALES_ESTATICAS multas_totales; /*  
Lista todas las variables */  
RUN;
```

```
***
```

```
*analisis exploratorio
```

```
*****
```

```
/* Calcular estadísticos descriptivos */
```

```
PROC MEANS DATA=multas;  
VAR VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS SER_SIN_TIQUE  
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR  
DOBLE_FILA  
multas_totales_dinamicas MULTAS_TOTALES_ESTATICAS  
multas_totales;  
OUTPUT OUT=descriptivos;  
RUN;
```

```
proc sgplot data=multas;  
series x=fecha y=VELOCIDAD / lineattrs=(color=blue thickness=2)  
legendlabel="VELOCIDAD";  
series x=fecha y=SEMAFORO / lineattrs=(color=red thickness=2)  
legendlabel="SEMAFORO";  
series x=fecha y=A_P_R / lineattrs=(color=green thickness=2)  
legendlabel="A_P_R";  
series x=fecha y=MOVIL / lineattrs=(color=orange thickness=2)  
legendlabel="MOVIL";  
series x=fecha y=OTRAS_DINAMICAS / lineattrs=(color=purple  
thickness=2) legendlabel="OTRAS_DINAMICAS";  
series x=fecha y=SER_SIN_TIQUE / lineattrs=(color=cyan  
thickness=2) legendlabel="SER_SIN_TIQUE";
```

```

series x=fecha y=SER_TIQUE_INVALIDO / lineattrs=(color=magenta
thickness=2) legendlabel="SER_TIQUE_INVALIDO";
series x=fecha y=CARGA_DESCARGA / lineattrs=(color=brown
thickness=2) legendlabel="CARGA_DESCARGA";
series x=fecha y=CARRIL_BUS / lineattrs=(color=yellow thickness=2)
legendlabel="CARRIL_BUS";
series x=fecha y=PROHIBIDO_APARCAR / lineattrs=(color=black
thickness=2) legendlabel="PROHIBIDO_APARCAR";
series x=fecha y=DOBLE_FILA / lineattrs=(color=gray thickness=2)
legendlabel="DOBLE_FILA";

/* Añadir etiquetas y título */
xaxis label="Fecha";
yaxis label="Cantidad de Multas";
title "Gráfico de Líneas de Multas por Tipo";

/* Mostrar leyenda */
keylegend / position=right across=1;
run;

***
*análisis correlaciones;
***

PROC CORR DATA=MULTAS1 PLOTS=matrix outp=corr_matrix;
VAR VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS SER_SIN_TIQUE
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR
DOBLE_FILA;
RUN;

data corr_long;
set corr_matrix(where=( _TYPE_="CORR"));
array vars[11] VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS
SER_SIN_TIQUE
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS
PROHIBIDO_APARCAR DOBLE_FILA;

do i = 1 to 11;
varname = vname(vars[i]);
corr_value = vars[i];
output;
end;
keep _NAME_ varname corr_value;
run;

proc sgplot data=corr_long;
heatmapparm x=_NAME_ y=varname colorresponse=corr_value /
colormodel=(blue white red);
xaxis discreteorder=data;
yaxis discreteorder=data;
gradlegend;
title "Mapa de calor de la matriz de correlación";
run;

***
*análisis factorial;
.....

proc factor DATA= multas plots=ALL OUTSTAT=ESTEJ1 n=3 out=PROEJ1 MSA;

```

```

VAR VELOCIDAD SEMAFORO A_P_R MOVIL OTRAS_DINAMICAS SER_SIN_TIQUE
SER_TIQUE_INVALIDO CARGA_DESCARGA CARRIL_BUS PROHIBIDO_APARCAR
DOBLE_FILA;
pathdiagram fuzz=0.5;
run;

***
/* Dividir los datos en conjuntos de entrenamiento y prueba */
.....

data entrenamiento prueba;
  set proej1;
  if _n_ < 97 then output entrenamiento;
  else output prueba;
run;

***
/* HOLT WINTERS */
.....

/*suavizado FACTOR1*/
proc esm data=entrenamiento out=salida lead=12 back=12
print=statistics;
id fecha interval=month;
forecast Factor1 / model=addwinters;
run;
proc esm data=entrenamiento out=salida lead=12 back=12
prin=estimates;
id fecha interval=month;
forecast Factor1 / model=addwinters;
run;
proc esm data=entrenamiento out=salida lead=12 back=12
plot = MODELFORECAST;
id fecha interval=month;
forecast Factor1 / model=addwinters;
run;
proc esm data=entrenamiento out=salida lead=12 back=12
plot = FORECASTSONLY;
id fecha interval=month;
forecast Factor1 / model=addwinters;
run;

proc means data=prediccionestacionamiento mean;
var residual;
output out=mse_results mean=mse;
run;

/*suavizado FACTOR2*/
proc esm data=entrenamiento out=salida lead=12 back=12
prin=estimates;
id fecha interval=month;
forecast Factor2 / model=addwinters;
run;
proc esm data=entrenamiento out=salida lead=12 back=12
plot = forecastonly;
id fecha interval=month;
forecast Factor2 / model=addwinters;
run;

```

```

/*suavizado FACTOR3*/
proc esm data=entrenamiento out=salida lead=12 back=12
print=statistics;
id fecha interval=month;
forecast Factor3 / model=addwinters;
run;
proc esm data=entrenamiento out=salida lead=12 back=12
plot = forecastonly;
id fecha interval=month;
forecast Factor3 / model=addwinters;
run;

***
/* ARIMA*/
.....

*ARIMA FACTOR1;

data arima noarima;
set proej1;
if _n_ < 85 then output arima;
   else output noarima;
run;
proc arima data=arima;
identify var=factor1 nlag=24; run;

proc arima data=arima;
identify var=Factor1(1 12) nlag=24;
estimate p=(3) q=(12);
run;

proc arima data=arima;
identify var=Factor1(1 12) nlag=24;
estimate p=(3) q=(12);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamientol printall; run;

DATA prediccionestacionamientol;
SET prediccionestacionamientol;
ID=_N_;
RUN;
PROC SORT DATA=prediccionestacionamientol; by descending id;run;
DATA noarima;
SET noarima;
ID=_N_;
RUN;
data union1 no;
merge noarima prediccionestacionamientol(drop=Factor1);
if id<=85 then output no;
else output union1;
run;
proc sgplot data=union1;
band Upper=u95 Lower=l95 x=fecha
/ LegendLabel="95% Confidence Limits";
scatter x=fecha y=Factor1;
series x=fecha y=forecast; run;

proc means data=prediccionestacionamientol mean;
var residual;
output out=mse_results mean=mse;

```

```

run;

proc arima data=entrenamiento;
identify var=Factor1(1 12) nlag=24;
estimate p=(3) q=(12);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamiento1 printall; run;
data prediccionestacionamiento1;
set prediccionestacionamiento1;
p_factor1=forecast;
run;
proc sort data=prediccionestacionamiento1;by id;run;

*ARIMA FACTOR2;

proc arima data=arima;
identify var=factor2 nlag=24; run;

proc arima data=arima;
identify var=factor2(1) nlag=24; run;

proc arima data=arima;
identify var=factor2(1) nlag=24;
estimate p=(6) q=(12);
run;

proc arima data=arima;
identify var=Factor2(1) nlag=24;
estimate p=(6) q=(12);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamiento2 printall; run;

DATA prediccionestacionamiento2;
SET prediccionestacionamiento2;
ID=_N_;
RUN;

PROC SORT DATA=prediccionestacionamiento2; by descending id;run;

DATA noarima;
SET noarima;
ID=_N_;
RUN;

data union2 no;
merge noarima prediccionestacionamiento2(drop=Factor2);
if id<=85 then output no;
else output union2;
run;

proc sgplot data=union2;
band Upper=u95 Lower=l95 x=fecha
/ LegendLabel="95% Confidence Limits";
scatter x=fecha y=Factor2;
series x=fecha y=forecast;
run;

```

```

proc means data=prediccionestacionamiento2 mean;
var residual;
output out=mse_results mean=mse;
run;

proc arima data=entrenamiento;
identify var=Factor2(1) nlag=24;
estimate p=(6) q=(12);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamiento2 printall;
run;

data prediccionestacionamiento2;
set prediccionestacionamiento2;
p_factor2=forecast;
run;

proc sort data=prediccionestacionamiento2;by id; run;

*ARIMA FACTOR3;

proc arima data=arima;
identify var=factor3 nlag=24;
run;

proc arima data=arima;
identify var=factor3(1) nlag=24;
run;

proc arima data=arima;
identify var=factor3(1) nlag=24;
estimate p=(10);
run;

proc arima data=arima;
identify var=factor3(1) nlag=24;
estimate p=(10);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamiento3 printall;
run;

DATA prediccionestacionamiento3;
SET prediccionestacionamiento3;
ID=_N_;
RUN;

PROC SORT DATA=prediccionestacionamiento3; by descending id;run;

DATA noarima;
SET noarima;
ID=_N_;
RUN;

data union3 no;
merge noarima prediccionestacionamiento3(drop=Factor3);
if id<=85 then output no;
else output union3;
run;

```

```

proc sgplot data=union3;
band Upper=u95 Lower=l95 x=fecha
/ LegendLabel="95% Confidence Limits";
scatter x=fecha y=Factor3;
series x=fecha y=forecast; run;

proc means data=prediccionestacionamiento3 mean;
var residual;
output out=mse_results mean=mse;
run;

proc arima data=entrenamiento;
identify var=factor3(1) nlag=24;
estimate p=(10);
forecast lead=12 id=fecha interval=month
out=prediccionestacionamiento3 printall;
run;

data prediccionestacionamiento3;
set prediccionestacionamiento3;
p_factor3=forecast;
run;

proc sort data=prediccionestacionamiento3;by id;run;

```