



Colegio Universitario de Estudios Financieros

Máster Universitario en Ciencia de Datos

MODELOS DE INTERVENCIÓN DE SERIES TEMPORALES: APLICACIÓN AL COVID-19

Alumno: Gómez Rivas, Alejandro José

Tutor: Queralt Sánchez de las Matas, Ricardo A.

Madrid, 9 de septiembre de 2022

AGRADECIMIENTOS

En primer lugar, quiero agradecer al doctor Queralt por guiarme en el desempeño de este trabajo, sobre todo por la formación crítica dentro de este campo de la ciencia de datos.

En segundo lugar, agradecer a todos los profesores y a todos mis compañeros del Máster que me ayudaron y estuvieron dispuestos para ayudarme a sacar adelante un máster tan completo como éste.

Por último, y no menos importante, deseo agradecer de corazón a mis padres, a mi prima, a mis amigos y a mi pareja por el apoyo recibido durante este periodo importante de mi vida.

1. RESUMEN

A principios del 2020, cuando en China se estaban preparando para conmemorar la Fiesta de la Primavera, también conocida como el Año Nuevo Chino, surge una nueva enfermedad respiratoria aguda denominada COVID-19 que azotó al país por completo y también, en cuestión de pocos meses, al mundo entero.

Como consecuencia de esta pandemia, se han cerrado muchas empresas, con una situación de incertidumbre para muchos datos socioeconómicos y bursátiles, por lo que un pronóstico preciso ante una posible futura pandemia o de una situación atípica, se vuelve crucial en tales situaciones.

El principal objetivo de este Trabajo de Fin de Máster es realizar un análisis e implantación de distintos modelos predictivos tradicionales del impacto del COVID-19 en distintas series temporales y determinar cuáles de ellos tienen un pronóstico más certero del futuro para que, de esta manera, las empresas y los países de todo el mundo puedan estar preparados para controlarlo.

Palabras clave

Análisis de series temporales, estacionariedad, ETS, ARIMA, Prophet

ABSTRACT

At the beginning of 2020, when China was preparing to celebrate the Spring Festival, also known as the New Chinese, a new respiratory disease variant called COVID-19 hit the country completely and also, in few months, the whole world. The pandemic has caused the closure of many companies and a situation of uncertainty for many socioeconomic and stock market data.

An accurate forecast of a possible future pandemic or an unusual situation becomes crucial in these situations. The main objective of this Final Project is to make an analysis and an implementation of different traditional predictive models about the impact of COVID-19 in different time series and determine which one of them have more accurate forecast of the future and, in this way, companies and countries around the world may be prepared to control it.

Keywords

Time-series analysis, stationary time-series, ETS, ARIMA, Prophet

ACRÓNIMOS Y ABREVIACIONES

AI	Inteligencia Artificial
ARIMA	Modelo Autorregresivo Integrado de Media Móvil
EDA	Análisis Exploratorio de Datos
ETS	Modelo de Suavizado Exponencial
IDE	Entorno de Desarrollo Integrado
MAPE	Error Porcentual Absoluto Medio
MSE	Media de los Errores al Cuadrado o Error Cuadrático Medio
RMSE	Raíz del Error Cuadrático Medio
SARIMAX	Modelo Exógeno de Media Móvil Integrada Autorregresivo Estacional
TS	Serie Temporal

REPOSITORIO GIT

Para facilitar la información del presente informe a los lectores, se incluye una carpeta comprimida con los notebooks de Python usados para el Trabajo de Fin de Máster.

El repositorio GIT se ubica en este [hipervínculo](#), y el contenido de éste se desglosa de la siguiente manera:

- En la carpeta “data” se almacenan los ficheros CSV con los datos de las series temporales analizadas.
- La carpeta “doc” contiene el presente informe y el resumen ejecutivo.
- La carpeta “pics” guarda las gráficas obtenidas de las series temporales para los análisis.
- La carpeta “predictions” almacena los ficheros CSV con los resultados finales obtenidos de cada serie temporal predichos por el modelo que mejor se ajusta a cada una.
- Por último, en la carpeta “notebooks” se encuentran todos los notebooks de las series temporales con los gráficos, los análisis y los modelos empleados para determinar la predicción.

Tabla de contenido

AGRADECIMIENTOS	2
1. RESUMEN	3
1.1. CONTEXTO DEL COVID-19	11
1.2. ESTADO DEL ARTE	12
1.3. OBJETIVO Y DESARROLLO.....	12
2. MARCO TEÓRICO.....	13
2.1. SERIES TEMPORALES	14
2.1.1. Definición	14
2.1.2. Componentes	14
2.1.3. Estacionariedad.....	15
2.1.4. Descomposición.....	16
2.2. MODELOS DE PREDICCIÓN DE SERIES TEMPORALES	17
2.2.1. ETS	17
2.2.2. ARIMA	18
2.2.3. PROPHET	19
3. MARCO PRÁCTICO	20
3.1. SERIE TEMPORAL 1: DESEMPLEO.....	20
3.1.1. Desempleo sin intervención.....	20
3.1.1.1. Modelo a batir: ETS	22
3.1.1.2. Modelo a batir: ARIMA	22
3.1.1.3. Modelo a batir: PROPHET	22
3.1.2. Desempleo con intervención.....	23
3.1.2.1. Modelo a batir: ETS	23
3.1.2.2. Modelo a batir: ARIMA	24
3.1.2.3. Modelo a batir: PROPHET	24
3.2. SERIE TEMPORAL 2: TURISMO.....	24
3.2.1. Turismo sin intervención	25
3.2.1.1. Modelo a batir: ETS	25
3.2.1.2. Modelo a batir: ARIMA	26
3.2.1.3. Modelo a batir: PROPHET	26
3.2.2. Turismo con intervención	26
3.2.2.1. Modelo a batir: ETS	27
3.2.2.2. Modelo a batir: ARIMA	27
3.2.2.3. Modelo a batir: PROPHET	28

3.3. SERIE TEMPORAL 3: MATRICULACIONES DE VEHÍCULOS	28
3.3.1. <i>Matriculaciones sin intervención</i>	28
3.3.1.1. <i>Modelo a batir: ETS</i>	29
3.3.1.2. <i>Modelo a batir: ARIMA</i>	29
3.3.1.3. <i>Modelo a batir: PROPHET</i>	30
3.3.2. <i>Matriculaciones con intervención</i>	30
3.3.2.1. <i>Modelo a batir: ETS</i>	31
3.3.2.2. <i>Modelo a batir: ARIMA</i>	31
3.3.2.3. <i>Modelo a batir: PROPHET</i>	32
3.4. SERIE TEMPORAL 4: HIPOTECAS	32
3.4.1. <i>Hipotecas sin intervención</i>	32
3.4.1.1. <i>Modelo a batir: ETS</i>	33
3.4.1.2. <i>Modelo a batir: ARIMA</i>	33
3.4.1.3. <i>Modelo a batir: PROPHET</i>	34
3.4.2. <i>Matriculaciones con intervención</i>	34
3.4.2.1. <i>Modelo a batir: ETS</i>	35
3.4.2.2. <i>Modelo a batir: ARIMA</i>	35
3.4.2.3. <i>Modelo a batir: PROPHET</i>	35
4. RESULTADOS DE LAS PREDICCIONES	36
4.1. SERIE TEMPORAL 1: DESEMPLEO	36
4.1.1. <i>Resultado sin la intervención</i>	36
4.1.2. <i>Resultado con la intervención</i>	36
4.2. SERIE TEMPORAL 2: TURISMO	37
4.2.1. <i>Resultado sin la intervención</i>	37
4.2.2. <i>Resultado con la intervención</i>	37
4.3. SERIE TEMPORAL 3: MATRICULACIONES DE VEHÍCULO	38
4.3.1. <i>Resultado sin la intervención</i>	38
4.3.2. <i>Resultado con la intervención</i>	39
4.4. SERIE TEMPORAL 4: HIPOTECAS	39
4.4.1. <i>Resultado sin la intervención</i>	39
4.4.2. <i>Resultado con la intervención</i>	40
5. CONCLUSIONES	41
APÉNDICE	42
REFERENCIAS	50

Índice de ilustraciones

Ilustración 1. Series estacionarias. Fuente: Francisco Parra, 2019.....	16
Ilustración 2: Descomposición de una serie temporal. Fuente: Elaboración propia.....	17
Ilustración 3: Evolución del desempleo en España. Fuente: Elaboración propia.	21
Ilustración 4. Comparativa ETS-ARIMA del desempleo. Fuente: Elaboración propia.	22
Ilustración 5. Evolución del desempleo con la intervención del COVID-19. Fuente: elaboración propia.....	23
Ilustración 6. Comparación ETS-ARIMA del desempleo con la intervención. Fuente: Elaboración propia. .	24
Ilustración 7: Evolución del turismo en España. Fuente: Elaboración propia.....	25
Ilustración 8. Comparativa ETS-ARIMA del turismo. Fuente: Elaboración propia.	26
Ilustración 9. Evolución del turismo en España con la intervención. Fuente: elaboración propia.	27
Ilustración 10. Comparación ETS-ARIMA del turismo con la intervención. Fuente: Elaboración propia.	27
Ilustración 11. Evolución del número de matriculaciones de vehículos en España. Fuente: Elaboración propia.....	29
Ilustración 12. Comparativa ETS-ARIMA de las matriculaciones. Fuente: Elaboración propia.....	30
Ilustración 13. Evolución del número de matriculaciones de vehículos en España con la intervención. Fuente: elaboración propia.	31
Ilustración 14. Comparación ETS-ARIMA de las matriculaciones con la intervención. Fuente: Elaboración propia.....	32
Ilustración 15. Evolución del número de hipotecas del total de fincas en España. Fuente: Elaboración propia.	33
Ilustración 16. Comparativa ETS-ARIMA de las hipotecas. Fuente: Elaboración propia.....	33
Ilustración 17. Evolución del número de hipotecas del total de fincas en España con la intervención. Fuente: elaboración propia.	34
Ilustración 18. Comparación ETS-ARIMA de las hipotecas con la intervención. Fuente: Elaboración propia.	35
Ilustración 19. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.	36
Ilustración 20. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.	36
Ilustración 21. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.	37
Ilustración 22. Predicción del turismo intervenido. Fuente: Elaboración propia.	38
Ilustración 23. Predicción de las matriculaciones sin intervenir. Fuente: Elaboración propia.	38
Ilustración 24. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.	39
Ilustración 34. Predicción de las hipotecas sin intervenir. Fuente: Elaboración propia.	40
Ilustración 26. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.	40

Índice de tablas

Tabla 1. MAPEs modelos para desempleo. Fuente: Elaboración propia.....	23
Tabla 2. MAPEs modelos para desempleo con la intervención. Fuente: Elaboración propia.	24
Tabla 3. MAPEs modelos para turismo. Fuente: Elaboración propia.	26
Tabla 4. MAPEs modelos para turismo con intervención. Fuente: Elaboración propia.....	28
Tabla 5. MAPEs modelos para matriculaciones. Fuente: Elaboración propia.	30
Tabla 6. MAPEs modelos para matriculaciones con la intervención. Fuente: Elaboración propia.....	32
Tabla 7. MAPEs modelos para hipotecas. Fuente: Elaboración propia.....	34
Tabla 8. MAPEs modelos para hipotecas con la intervención. Fuente: Elaboración propia.	35
Tabla 9. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.	36
Tabla 10. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.	37
Tabla 11. Predicción del turismo sin intervenir. Fuente: Elaboración propia.	37
Tabla 12. Predicción del turismo intervenido. Fuente: Elaboración propia.	38
Tabla 13. Predicción de las matriculaciones sin intervenir. Fuente: Elaboración propia.	38
Tabla 14. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.	39
Tabla 14. Predicción de las hipotecas sin intervenir. Fuente: Elaboración propia.	40
Tabla 16. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.	41

1. INTRODUCCIÓN

En este apartado se da una concepción general del proyecto para una mejor comprensión del estudio para los lectores. Se explica primero el contexto y la situación del COVID-19, a continuación, se detalla el estado del arte de los métodos actuales de análisis y predicción de series temporales y, por último, se explican los propósitos y objetivos del trabajo, seguidos de la metodología de investigación adoptada en el estudio.

1.1. CONTEXTO DEL COVID-19

La enfermedad por coronavirus 2019 (COVID-19) fue identificada por primera vez en la ciudad de Wuhan, provincia de Hubei, China, en diciembre de 2019 y es causada por el coronavirus 2 del síndrome respiratorio agudo severo (SRAS) -CoV-2). Los pacientes mostraban un cuadro clínico de una neumonía grave, provocando insuficiencia respiratoria e incluso la muerte. Se contagia a través de aerosoles que producen cuando el paciente tose o estornuda, por lo que las personas pueden infectarse con facilidad si no usan protección facial.

Debido a su alto nivel de contagio, pronto se convirtió en una pandemia que ha afectado a todo el mundo y que, a día de hoy continúa, aunque a niveles más bajos que antes atrás gracias a los avances médicos y científicos. En el caso de España, el virus ha tenido un impacto bastante notorio.

El COVID-19 ha impactado tanto en la salud de las personas como a la economía global. Los gobiernos de todos los países impusieron distintas políticas estrictas para hacer frente a la pandemia y “aplanar la curva” mediante el confinamiento domiciliario, restricciones de movilidad y el cierre temporal de muchas empresas y centros de distintos campos a excepción de aquellos que eran declarados como esenciales, como el sector sanitario o el de consumo, haciendo que muchas de estas empresas tuvieran que adaptarse a estas nuevas normas como buenamente pudieran.

Ante esta situación, se ha decidido estudiar para este trabajo el impacto del COVID-19 en distintas series temporales. Una serie temporal consiste en una secuencia de observaciones ordenadas de manera cronológica sobre una serie. Sus valores van ligados en el momento de tiempo en el que son medidos, de esta manera el análisis de la serie de tiempo implica el manejo en conjunto de dos variables: la variable objetivo (la que se va a estudiar) y la variable de tiempo. Los datos históricos de algunos activos empresariales o los datos macroeconómicos son claros datos típicos de series temporales.

El análisis de estas series temporales engloba un conjunto vasto de técnicas que permiten obtener la información y así observar el comportamiento que tuvieron dichas series temporales en el pasado y, de esta manera, poder tener opciones y mecanismos para poder predecir sus valores en un

futuro estipulado. A día de hoy existen muchos métodos de análisis distintos para descubrir el patrón o predecir hacia dónde van a tender los datos temporales. Los métodos estadísticos, como el modelo ARIMA, los modelos de aprendizaje automático o los de redes neuronales, son métodos que han demostrado ser eficientes para los análisis de series temporales, y algunos de ellos ya han proporcionado información relevante para saber hacer frente a situaciones atípicas.

1.2. ESTADO DEL ARTE

Los modelos de análisis de series temporales que se han aplicado para estudiar el seguimiento y la predicción de enfermedades infecciosas han obtenido, por lo general, unos resultados muy óptimos, pero todavía es un desafío encontrar una forma adecuada de analizar hacia dónde puede ir la tendencia del COVID-19, ya que se trata de una situación totalmente novedosa y de las que más ha impactado en la sociedad en lo que se lleva de siglo.

Como pronosticar la tendencia del virus sigue siendo todo un reto, la predicción de cualquier serie temporal teniendo en cuenta la situación actual con el COVID-19 hace que sea una tarea ardua. Además, realizando un recorrido por internet, como se ha comentado anteriormente, existen muchos estudios sobre el pronóstico que tienen los modelos de la enfermedad por COVID-19, pero hay escasas predicciones que puede tener el impacto de la misma en distintas series temporales de distintos campos.

Por tanto, resulta cautivador realizar un estudio empleando estos modelos predictivos en distintas series temporales y si éstos son válidos para el escenario actual y determinantes para detectar un posible evento no común, para poder evitarlo o hacer frente a ello mediante una correcta toma de decisiones. Por eso, se ha tomado la decisión de realizar este estudio bajo la tutela del director Queralt Sánchez de las Matas.

1.3. OBJETIVO Y DESARROLLO

En este caso, la idea principal de este proyecto consiste en analizar la situación de distintas series temporales tras la pandemia por COVID-19 y estudiar su pronóstico mediante el entrenamiento y el testeo de distintos modelos predictivos y determinar si son capaces de realizar detecciones de situaciones temporales atípicas y con cuánto margen de error.

“Las técnicas matemáticas para evaluar la extensión y el impacto de una epidemia y ayudar a su control son muy variadas.” (León, 2020). Para el desarrollo de este estudio, se ha seguido el mismo orden que tiene cada uno de los notebooks que se encuentran alojados en el repositorio. Una vez que se obtienen los ficheros donde se hallan los datos históricos de cada una de las series temporales, se realizan las siguientes técnicas:

- Preprocesamiento: se aplican distintos métodos de limpieza y tratamiento del dato, como el cambio del formato de la fecha, la detección de datos faltantes o nulos, el recorte del conjunto de datos para analizar una muestra, entre otros mecanismos.
- EDA: consiste en un análisis exploratorio de los datos. *“Es utilizado por los científicos de datos para analizar e investigar conjuntos de datos y resumir sus principales características, empleando a menudo métodos de visualización de datos”* (IBM, 2020). Se realizan una serie de cálculos estadísticos y de representaciones gráficas que ofrecen una comprensión los datos de un solo vistazo y que permite la identificación de distintos patrones comunes para, posteriormente, ver si es necesario realizar un tratamiento o modificación de los datos para adaptarlo a los modelos seleccionados para su testeo.
- Modelización: en esta fase se aplican los distintos modelos seleccionados y se realiza un entrenamiento y test con la serie temporal incluyendo el momento COVID-19 para generar un resultado y un margen de error.
- Elección del modelo: de dichos resultados se selecciona el modelo que menos error haya generado y que más se haya aproximado a los datos reales.
- Previsión: con el modelo ganador, se le realiza un pronóstico para finales del año 2022, se guardan en un archivo CSV para contrastarlo cuando llegue la fecha y comprobar el margen de error que puede dar.

Además, a cada una de las series temporales se les realiza una intervención eliminando el momento del COVID-19 para observar el resultado final y compararlo con la serie original, con el objetivo de ver cuánta diferencia de error y de predicción dan cuando no existe un evento pandémico.

Por último, se da una conclusión final sobre los resultados con los modelos utilizados.

Para que fuera posible realizar toda esta metodología, se ha desplegado el proyecto mediante el uso del lenguaje de programación **Python** (versión 3.9.7), y el IDE en el que se ha apoyado para su desarrollo ha sido **Jupyter Notebook** (versión 6.4.5).

2. MARCO TEÓRICO

Este apartado proporciona los conocimientos más importantes para que los lectores comprendan el desarrollo del proyecto. Primero se explica qué es una serie temporal, los componentes que la constituyen, la estacionariedad y la descomposición de la misma, y luego se enumeran distintos modelos de predicción tradicionales.

2.1. SERIES TEMPORALES

2.1.1. Definición

Por lo general, “una serie temporal es una sucesión de observaciones de una variable (p. ej. ingresos, precios, etc.) tomadas en varios instantes de tiempo” (Queralt Sánchez de las Matas, 2021). Son una secuencia que tiene intervalos constantes entre puntos de tiempo, como los datos que se toman diariamente y a la misma hora. Los datos de estas series se usan habitualmente en la estadística, en la economía, en las finanzas y en las predicciones y se aplican en gran medida en dominios con información temporal involucrada. Los análisis de las series de tiempo se conocen como métodos analíticos, los cuales desglosan los componentes de la serie temporal y ayudan a determinar variables estadísticas y patrones significativos. (Madridmasd, 2019)

2.1.2. Componentes

Las series temporales presentan normalmente una o varias características que se denominan como componentes, que ayudan a explicar el comportamiento que tienen las series en el tiempo. Dichos componentes son los siguientes:

- **Tendencia.** Este componente hace referencia al comportamiento o el movimiento de la serie temporal a largo plazo. El movimiento que tiende a hacer es creciente o decreciente y no tiene por qué ser lineal. A veces incluye el ciclo, que es el componente que se va a definir a continuación.
- **Componente cíclico.** En este se refleja los comportamientos recurrentes, aunque no tienen por qué ser exactamente periódicos, con un periodo superior a un año. Muestran, habitualmente, cómo se suceden las etapas de, por ejemplo, la bonanza económica con las de las crisis, o al menos, desaceleración.

A menudo, los ciclos económicos resultan de la superposición de distintos efectos, con periodos diferentes, más cortos o más largos, como dos, cinco o diez años. Por ello, son difícilmente reconocibles y, a veces, no se separa de la tendencia. En este caso, al componente se le denomina como "*ciclo-tendencia*".

- **Componente estacional.** Consiste en movimientos de oscilación de un periodo de tiempo. La estacionalidad no se presenta únicamente cuando el periodo amplio es el año, a veces hay estacionalidades mensuales o semanales en series diarias, o estacionalidades diarias en series horarias, como son las series de cotizaciones bursátiles, por ejemplo. La estacionalidad siempre es de un periodo fijo y conocido. Es importante que las estacionalidades tengan un periodo no superior al anual para que no se confundan con las componentes cíclicas.

- **Componente irregular.** También llamado como “ruido”, recoge las alteraciones de la serie, pequeñas en su incidencia, y sin una pauta periódica ni tendencial reconocible. Se considera que está ocasionada por múltiples factores, de pequeña entidad y diferentes ritmos temporales, que no se pueden estudiar individualmente.

2.1.3. Estacionariedad

En Matemáticas, un proceso estacionario es un proceso estocástico cuya distribución de probabilidad en un instante de tiempo fijo o una posición fija es la misma para todos los instantes de tiempo o posiciones, es decir, si no presenta cambios sistemáticos en su media (no hay tendencia) ni en su varianza, y sus variaciones periódicas han sido eliminadas. (Criado, s.f.)

Aplicando el concepto anterior a las series temporales, una serie se considera estacionaria cuando es estable a lo largo de un periodo de tiempo, esto es, cuando las media y la varianza son constantes en el tiempo.

La estacionariedad puede ser más o menos restrictiva en torno a su definición, por lo que se hablaría de dos tipos de estacionariedad:

- Serie estacionaria en sentido estricto o fuerte: se dice que una serie tiene esa definición si todos sus momentos, independientemente del orden, son invariantes en el tiempo.
- Serie estacionaria en sentido amplio o débil: se caracteriza así una serie estacionaria cuando todos sus momentos de primer y segundo orden son invariantes en el tiempo. Su ejemplo más simple es el de “ruido blanco”, es ahí donde la media y la varianza son siempre cero.

Por otro lado, una serie no se va a considerar estacionaria cuando no es estable a lo largo del tiempo, es decir, que su media y su varianza no son constantes en el tiempo.

Es interesante que las series temporales sean estacionarias porque, de esta manera, se pueden obtener predicciones fácilmente por el patrón repetitivo que tienen.

Otra manera de estudiar la estacionariedad de las series temporales y que se aplica en este proyecto es la aplicación del test de Dickey-Fuller aumentado. Éste realiza una prueba de raíz unitaria que detecta estadísticamente la presencia de conducta tendencial estocástica en una serie de tiempo mediante un contraste de hipótesis. Esta prueba nos proporciona una verificación rápida y evidente de si la serie es estacionaria o no (Ortiz, 2020). Las hipótesis que se aplican para el test son las siguientes:

- **Hipótesis nula (H_0)** de la prueba determina que la serie temporal se puede representar mediante una raíz unitaria. En otras palabras, indica que la serie temporal no es estacionaria ya que tiene alguna estructura dependiente del tiempo.
- **Hipótesis alternativa (H_1)** rechaza la hipótesis nula e indica que la serie temporal sí es estacionaria.

La interpretación del resultado del test se basa en el valor que dé el valor ‘p-value’. Un valor ‘p’ por debajo del umbral de significancia (0,05) significa que la hipótesis nula se rechaza.

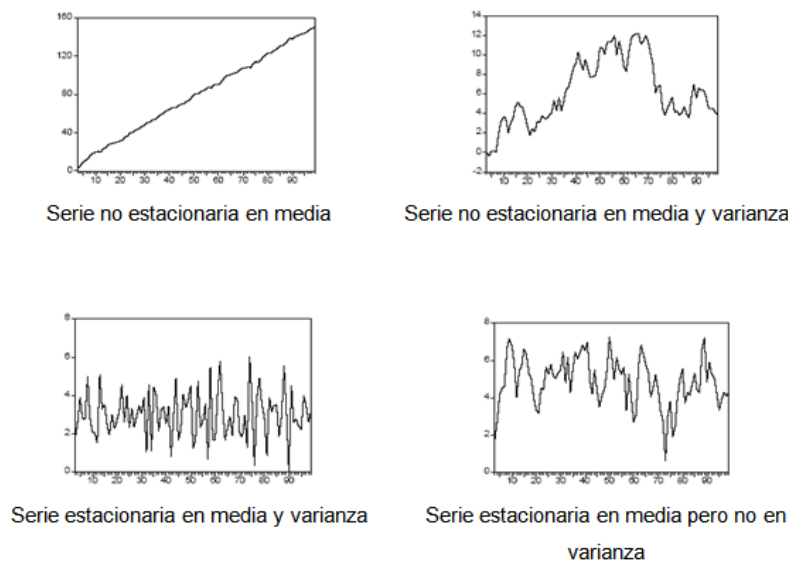


Ilustración 1. Series estacionarias. Fuente: Francisco Parra, 2019.

2.1.4. Descomposición

Siguiendo el enfoque comentado en el apartado referente a la serie temporal y sus componentes, en el procedimiento de análisis exploratorio de las series temporales que se han empleado en el trabajo se le han incluido la descomposición de las mismas. “*En pocas palabras, es una técnica de análisis que divide una serie temporal en los tres componentes ya descritos: tendencia, estacionalidad y residuos*”. (Lewinson, s.f.)

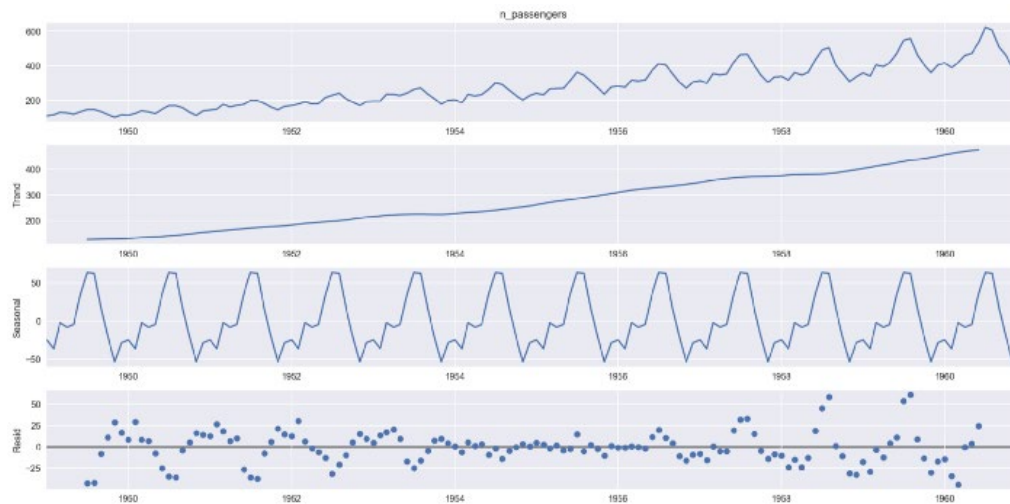


Ilustración 2: Descomposición de una serie temporal. Fuente: Elaboración propia

2.2. MODELOS DE PREDICCIÓN DE SERIES TEMPORALES

La predicción de las series de tiempo se basa de la siguiente forma: una vez que se conoce el comportamiento pasado de un sistema de datos, es posible realizar pronósticos sobre el comportamiento que ésta obtuvo a corto, medio o largo plazo.

Los métodos tradicionales estadísticos empleados para el pronóstico de las series temporales se basan en la construcción de funciones lineales a partir de las observaciones pasadas más recientes para proporcionar las predicciones futuras. Para este proyecto se han elegido tres modelos tradicionales, las cuales son ETS, ARIMA y PROPHET.

2.2.1. ETS

El modelo ETS, por sus siglas en inglés significa “*Exponential Smoothing*”, modelo de Suavizado Exponencial, es uno de los métodos de predicción de series temporales más antiguos y estudiados. Es muy eficaz cuando los valores de las series temporales siguen una tendencia gradual y presentan un comportamiento estacional marcado en el cual los valores siguen un patrón cíclico repetido a lo largo de un número determinado de periodos de tiempo (ArcGIS, s.f.).

Este modelo se emplea a menudo en series temporales con las características descritas anteriormente, como para predecir la temperatura de un municipio durante una ola de frío para prepararse para las enfermedades relacionadas con esa estación, o bien aplicados en datos macroeconómicos-demográficos como es la tasa de desempleo.

El funcionamiento del modelo ETS se divide en dos partes: en una parte de previsión y otra parte que es la de validación. La primera se emplea para realizar la previsión de los valores de futuros periodos de tiempo, y la segunda se utiliza para determinar la precisión de la predicción de valores

futuros que el modelo puede realizar para cada serie temporal. También se emplea para verificar la precisión del modelo dividiendo la serie temporal en entrenamiento y en test y, a través de la precisión y del error absoluto medio porcentual (MAPE), que es un indicador de desempeño del pronóstico que mide el tamaño del error, se comprueba la efectividad del modelo frente a los datos reales.

Cuanto más pequeño sea el MAPE, mucho más preciso va a ser el modelo aplicado a la serie temporal. Es muy útil este indicador cuando los modelos predictivos son muy similares entre sí y se quiere saber con exactitud la efectividad de uno u otro.

Existen distintos tipos de suavizado exponencial, pero todos se basan en separar las series temporales en varios componentes. Los valores de cada componente se estiman ponderando exponencialmente los componentes de los periodos de tiempo anteriores, de manera que la influencia de cada periodo de tiempo disminuya exponencialmente a medida que se avanza en el tiempo. Cada componente se define recursivamente a través de un planteamiento de modelo de estado-espacio, y cada componente depende de todos los demás componentes. La estimación de todos los parámetros se realiza mediante una estimación de máxima probabilidad.

En esta herramienta, todos los componentes son aditivos, de modo que el modelo de predicción es la suma de los componentes individuales. Si se utiliza un componente estacional, la herramienta utiliza el método estacional amortiguado de Holt-Winters. Si no se utiliza ningún componente estacional, la herramienta utiliza el método de tendencia amortiguada (Parra, 2019).

2.2.2. *ARIMA*

El método ARIMA, por sus siglas en inglés significa “Autoregressive Integrated Moving Average”, modelo de Autorregresión Integrada de Media Móvil, es otro algoritmo de predicción clásico diseñado para ser empleado en series temporales estacionarias. (Athanasopoulos, 2018) Este método se basa en la combinación de tres parámetros (p , d , q) que son:

- AR: de Autorregresión (p), el cual quiere decir que el pasado predice el futuro de la serie temporal, por lo que un valor de la serie temporal en el tiempo es una función de los valores pasados.
- MA: de Media Móvil (q), significa que un punto de la serie temporal puede definirse en función de los errores de los valores pasados. Dichos errores son los obtenidos en el ajuste de la media ponderada respecto a la función original.
- Diferenciaciones (d), es el parámetro encargado de que logre que la serie temporal sea estacionaria.

“La correlación es a la regresión, lo que la autocorrelación es a la autorregresión” (Iglesias, 2021). Esto significa que, en una variable con autocorrelación, pueden utilizarse los valores pasados como predictores de valores futuros. De ahí la razón de ser del parámetro (p) del modelo ARIMA, que refleja el número de valores de observaciones anteriores, o retardos (*lags*), que se utilizan para predecir una presente. Un ARIMA (1,0,0) utilizaría la observación inmediatamente anterior para explicar la presente.

El componente de Media Móvil (q) indica el número de valores anteriores cuyos errores se utilizan para predecir un valor presente. Para comprender este componente hay que explicar la razón de ser de las diferenciaciones (d). Los ARIMA se usan para explicar o modelizar procesos estacionarios (o estocásticos) los cuales sus propiedades no dependen del tiempo en el que se observan. Diferenciar la serie (restar a los valores presentes el inmediatamente anterior) puede servir para eliminar tendencias que alteren esas propiedades, indicando con la Diferenciación “ d ” el número de veces que se repite ese proceso (o el número de diferenciaciones que se aplican).

En consecuencia, los procesos estocásticos están determinados por dos componentes o fuerzas principales: sus propios valores previos (explicados por su componente autorregresivo (AR) y las variaciones impredecibles producidas por multitud de variables que cambian e interaccionan a lo largo del tiempo de tal manera que son ostensiblemente aleatorias. Es el efecto de estas variaciones impredecibles lo que busca explicar el componente de medias móviles (Moving Averages, MA). Conceptualmente los modelos ARIMA (0, 0, q) o MA (q) buscan explicar un valor presente a través de una función que lo relacione con las variaciones impredecibles de “ q ” valores anteriores, o, en otras palabras, con el error de los términos de “ q ” tiempos anteriores, de tal manera que un ARIMA (0,0,1) utilizaría el error de la observación inmediatamente anterior para explicar el valor presente y vendría definido por la ecuación.

2.2.3. *PROPHET*

El modelo PROPHET es una herramienta de código abierto de Facebook utilizada para pronosticar datos de series de tiempo que ayuda a los científicos y estadísticos a comprender y predecir las series temporales.

Se basa en un modelo aditivo descomponible donde las tendencias no lineales se ajustan a la estacionalidad, donde se tiene en cuenta los efectos de los festivos y otras variables regresoras. Este método de pronóstico va analizando y cambiando los parámetros en función de lo óptimo que haya sido el ajuste, permitiendo de este modo un ajuste automático de los parámetros del modelo. Así

consigue una parametrización de las variables de forma automática gracias al aprendizaje realizado del comportamiento de la serie.

3. MARCO PRÁCTICO

En este apartado se lleva a cabo todo el marco teórico de manera pragmática con datos reales. El procedimiento que se aplica para el análisis es el mismo para todas las series temporales: primero se hace un sencillo análisis exploratorio de las series temporales a través de sus componentes y se aplican los modelos a batir para comprobar cuál de ellas es la que más se aproxima a los datos reales. Además, para comprobar el impacto que ha tenido el COVID-19 en las series temporales, se le realiza a cada serie temporal una intervención suprimiendo dicha etapa y sustituyendo sus datos con los obtenidos en la predicción de datos pasados mediante entrenamiento-test del modelo predictivo que mejor precisión tuvo en el modelo original.

Se han seleccionado para el proyecto cuatro series temporales, las cuales se han cogido los datos que se alojan en las páginas oficiales del gobierno español. Estas series temporales son el número de desempleados, el número de turistas, la matriculación de vehículos y el número de hipotecas.

3.1. SERIE TEMPORAL 1: DESEMPLEO

Para la obtención de los datos del desempleo español, se ha dirigido a la página web del Servicio Público de Empleo Estatal y se han seleccionado los datos registrados desde el año 2013 hasta el 2022 que, en la opinión del autor del proyecto, consideraba que ésta era una cantidad de tiempo óptima ya que hay suficientes datos para realizar el entrenamiento y el test del modelo, incluyendo el inicio de la pandemia, momento clave para determinar si los modelos empleados para esta serie son efectivos para predecir estas situaciones.

Cada serie, como se ha dicho antes, se desglosa en dos partes: una parte en la que no se intervienen en los datos del momento por COVID-19 para ver cómo predicen los modelos, y otra parte en la que sí se intervienen en los datos del desempleo, para ver la diferencia de pronósticos.

3.1.1. Desempleo sin intervención

Comienza el análisis tomando los datos mensuales. En total hay 115 meses y, como se puede apreciar en la Ilustración 3, existe una tendencia decreciente hasta la raya roja que es marzo del 2020, momento en el que inicia la pandemia por COVID-19. También se puede comprobar un

comportamiento estacional, pues a lo largo de los meses del verano se genera empleo temporal y baja significativamente el paro. Pasa lo mismo con el mes de diciembre, pero a menor escala.

En el momento en el que comienzan las restricciones por COVID-19, muchos de los empleados con contratos fueron despedidos y otros pasados a ERTes, lo que hace subir el número de parados. Este fenómeno se ha prolongado hasta el inicio del año 2021, momento en el que vuelve a haber una tendencia decreciente.

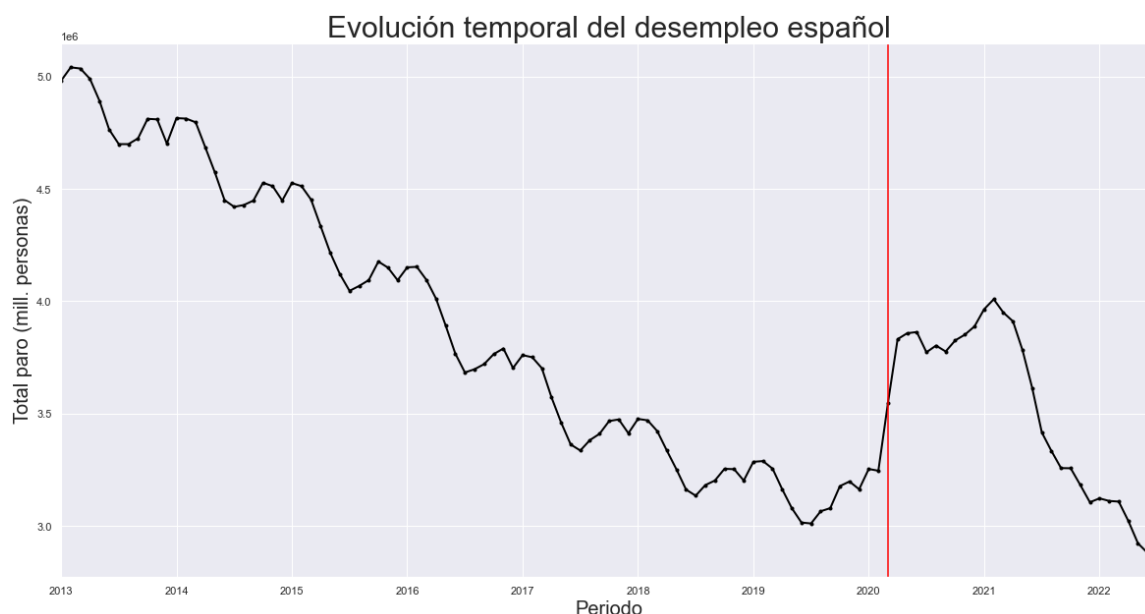


Ilustración 3: Evolución del desempleo en España. Fuente: Elaboración propia.

Para analizar la estacionariedad de la serie temporal, se realiza un cálculo de la media y de la varianza de todos los números de desempleados del conjunto de datos y se representa gráficamente. Sus gráficas se encuentran en el apéndice. A simple vista la media posee unos valores constantes decrecientes en el tiempo, lo que nos indica que la serie es estacionaria, pero en varianza no se aprecia bien. Para asegurar que una serie es estacionaria o no, se aplica el test de Dickey-Fuller aumentado, el cual indica que el valor 'p' está por debajo del umbral, y se concluye que la serie es estacionaria.

Por último, se hace una observación en la descomposición de la serie temporal para comprobar que todo lo descrito anteriormente se cumple. Ésta resume, de manera gráfica, los componentes que se han analizado, y se ubica su ilustración en el apéndice. Cabe destacar en la misma la presencia de ruido y no se observa una pauta periódica ni un patrón reconocible del motivo de la misma.

3.1.1.1. Modelo a batir: ETS

Como los datos son estacionales, comienza la batida de modelos con el modelo ETS, modelo de Suavizado Exponencial. Tras realizar un autoETS se obtiene un ETS(MAM), es decir, un modelo multiplicativo de Hol-Winters con errores multiplicativos.

3.1.1.2. Modelo a batir: ARIMA

La serie, al ser estacionaria, no se aplica la transformación de los datos a logaritmos. Tras realizar el autoARIMA para encontrar el mejor modelo, se obtiene SARIMAX (Seasonal ARIMA) $(1, 1, 0) \times (1, 1, 0, 12)$, es decir, con auto-regresión de orden 1, la serie depende del último retardo y media móvil de orden 1, es decir, la serie depende del error ocurrido en el retardo anterior. El 12 es la periodicidad, que son 12 meses.

Realizando una comparativa gráfica se observa que ambos modelos no predicen en absoluto el momento COVID-19, pero ambos predicen con bastante exactitud el número de desempleados a partir de diciembre del 2021, lo que quiere decirse que, de una manera estadística, el efecto COVID-19 termina en esa fecha.



Ilustración 4. Comparativa ETS-ARIMA del desempleo. Fuente: Elaboración propia.

Realizando el cálculo del MAPE, se comprueba que el modelo ETS da un error ligeramente inferior (12,08% frente al 12,27% de error que predice ARIMA).

3.1.1.3. Modelo a batir: PROPHET

Modelando la serie temporal con este método se observa que hace el mismo efecto que con ETS y ARIMA, no predice el momento por COVID-19, pero luego consigue una mejor predicción más adelante. Su MAPE da un error del 13,98%, lo que indica que el modelo ETS es el que menos error da. Tiene sentido, pues el modelo de Suavizado Exponencial es mejor cuando la tendencia estacional es de la misma magnitud a lo largo de todos los datos, como es este caso.

Modelo	MAPE
ETS	12,08%
ARIMA	12,27%
PROPHET	13,98%

Tabla 1. MAPEs modelos para desempleo. Fuente: Elaboración propia.

3.1.2. Desempleo con intervención

Para realizar la intervención, se ha realizado un entrenamiento-test con el modelo que menos error dio en el conjunto de datos del desempleo original hasta febrero del 2020, el mes antes del comienzo de la pandemia. Dicho modelo es el ETS. Después se realiza una predicción con los datos hasta el mes de diciembre del 2021, que se mencionó anteriormente que era el punto donde la predicción y los datos reales coincidían, concluyendo el momento por COVID-19. Por último, los datos obtenidos se añaden al dataset y, al representarlo, se obtiene esta serie temporal de desempleo intervenido. Las rayas rojas dividen el momento que se ha intervenido con el resto del conjunto de datos.



Ilustración 5. Evolución del desempleo con la intervención del COVID-19. Fuente: elaboración propia.

La serie cumple exactamente con las mismas características que la serie original: tendencia decreciente que se aplana al final del periodo, es estacionaria y tiene estacionalidad marcada.

3.1.2.1. Modelo a batir: ETS

La predicción obtiene muy buena precisión y un error MAPE de un 1,19%. El modelo que se obtiene es el ETS(MAM), el mismo que se obtuvo en el desempleo sin intervención.

3.1.2.2. Modelo a batir: ARIMA

Con ARIMA pasa exactamente lo mismo: obtiene una muy buena predicción con un error del 1,36% y un modelo SARIMAX (1, 1, 0)x(1, 1, 0, 12)

Haciendo un contraste gráfico, no se ven apenas diferencias. Se podría predecir perfectamente con cualquiera de los dos, ya que apenas fallan para este tipo de series temporales con estacionalidad marcada y repetitiva.

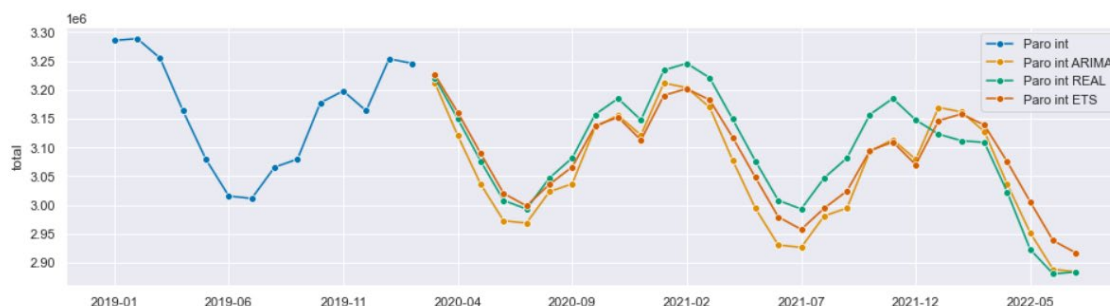


Ilustración 6. Comparación ETS-ARIMA del desempleo con la intervención. Fuente: Elaboración propia.

3.1.2.3. Modelo a batir: PROPHET

Con este modelo se obtienen unas predicciones ligeramente peores, pues este modelo es más preciso para modelos que tengan puntos de datos atípicos grandes, eventos irregulares o con tendencias de crecimiento no lineal que se aproximen a un límite. El MAPE ha dado un resultado del 5,73%, lo que indica que el modelo ETS vuelve a triunfar para este conjunto de datos en el que se ha intervenido el momento COVID-19.

Modelo	MAPE
ETS	12,08%
ARIMA	12,27%
PROPHET	13,98%

Tabla 2. MAPEs modelos para desempleo con la intervención. Fuente: Elaboración propia.

3.2. SERIE TEMPORAL 2: TURISMO

Para la obtención de los datos del número de turistas en España, se ha dirigido a la página web del Instituto Nacional de Estadística. El CSV recopila datos del número de turistas a nivel nacional y a nivel regional por Comunidades Autónomas. Para este trabajo se han seleccionado solamente los datos a nivel nacional.

3.2.1. Turismo sin intervención

Al realizar la lectura del CSV se observa que hay más de 2000 filas y muchos valores, por lo que se realiza un visualizado de los datos para determinar con cuáles se van a realizar el análisis. Haciendo las cribas pertinentes, se queda con un DataFrame de 75 meses correspondientes a los meses comprendidos entre marzo del 2016 y junio del 2022 (momento en el que se desarrolla este proyecto). Como se aprecia en la Ilustración 7, no se aprecia tendencia hasta la raya roja del inicio de la pandemia por COVID-19, donde los meses de abril y mayo obtienen cero números de turistas por el cierre de fronteras y el confinamiento domiciliario provocado por el inicio de la pandemia. En ese momento, la tendencia toma un claro comportamiento decreciente pero que se recupera ligeramente en la época del verano. La serie posee una estacionalidad muy marcada antes del COVID-19 ya que, en España, gracias al clima y a la oferta y demanda turística que hay, es un foco de atracción para turistas en la época de verano.

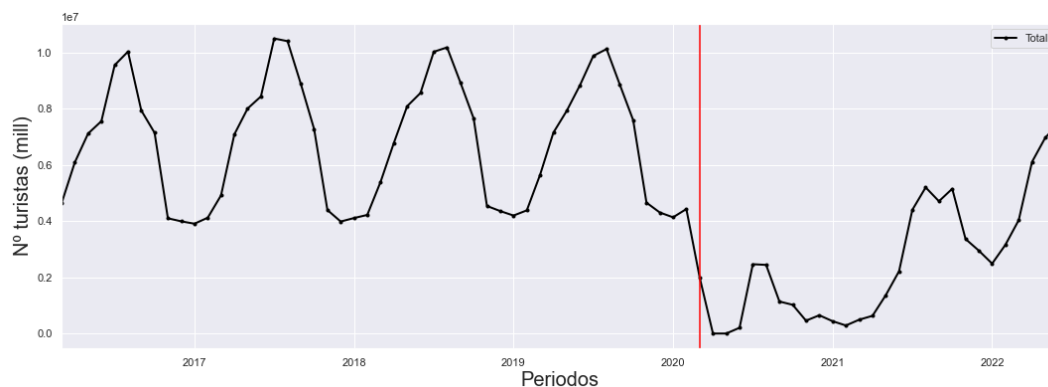


Ilustración 7: Evolución del turismo en España. Fuente: Elaboración propia.

En cuanto a la estacionariedad de la serie temporal, en la gráfica de la media y de la varianza ubicada en el apéndice no queda claro si cumple con el requisito de estacionariedad o no, por lo que, realizando el test de Dickey-Fuller aumentado, se cumple la hipótesis nula, lo que significa que la serie temporal no es estacionaria y habría que realizarles transformaciones a los datos para adaptarlos al modelo ARIMA.

La descomposición de la serie, también reflejada en el apéndice, resume todo el EDA que se acaba de describir.

3.2.1.1. Modelo a batir: ETS

Comienza la batida de modelos con el modelo ETS, y tras realizar un autoETS se obtiene un ETS(MNM), es decir, un modelo con error y estacionalidad multiplicativos y sin tendencia. Es lógico,

pues anteriormente se había descrito que esta serie no tiene tendencia y tiene una estacionalidad muy marcada.

3.2.1.2. Modelo a batir: ARIMA

La serie, al ser no estacionaria, se aplica la transformación de los datos a logaritmos. Tras realizar el autoARIMA para encontrar el mejor modelo, se obtiene SARIMAX (Seasonal ARIMA) $(1, 1, 0) \times (0, 1, 0, 12)$.

Realizando una comparativa gráfica se observa que ambos modelos no predicen en absoluto el momento COVID-19, pero ambos empiezan a predecir con bastante aproximación el número de turistas a partir de noviembre del 2021, lo que quiere decirse que, de una manera estadística, el efecto COVID-19 termina a partir esa fecha.

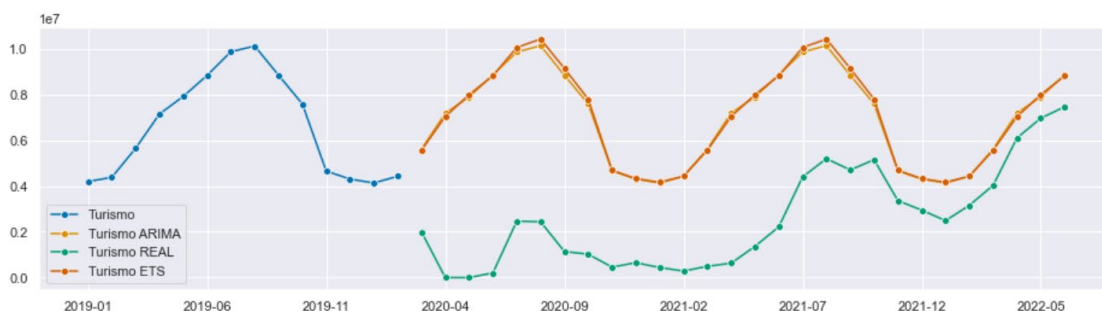


Ilustración 8. Comparativa ETS-ARIMA del turismo. Fuente: Elaboración propia.

3.2.1.3. Modelo a batir: PROPHET

Modelando la serie temporal con este método se observa que hace el mismo efecto que con ETS y ARIMA, es decir, no predice el momento por COVID-19, pero luego consigue una mejor predicción más adelante. Analizando sus MAPEs, el modelo ETS es el que vuelve a predecir mejor, aunque por poco.

Modelo	MAPE
ETS	24,17%
ARIMA	24,25%
PROPHET	24,31%

Tabla 3. MAPEs modelos para turismo. Fuente: Elaboración propia.

3.2.2. Turismo con intervención

Para esta intervención se “cortan” los datos hasta noviembre del 2021, que fue cuando los datos de la predicción por ETS se asemejaban a los datos reales.

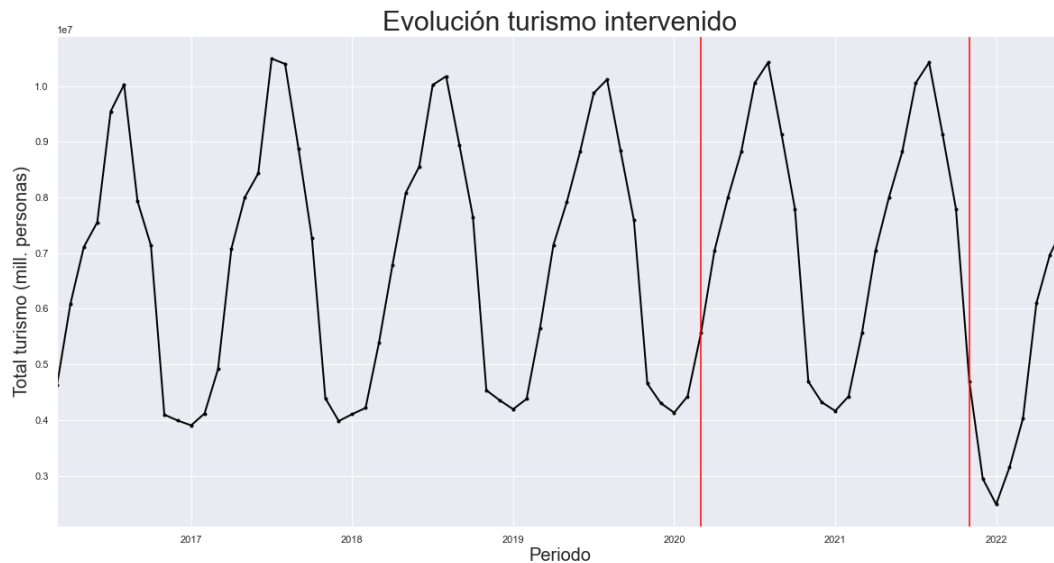


Ilustración 9. Evolución del turismo en España con la intervención. Fuente: elaboración propia.

La serie cumple de manera similar con la original: no existe tendencia, tiene estacionalidad marcada y hay ausencia de estacionariedad.

3.2.2.1. Modelo a batir: ETS

La predicción obtiene muy buena precisión y un error MAPE de un 8,6%, lo que vuelve a repetirse lo mencionado anteriormente, y es que estos modelos predicen con bastante aproximación en series temporales en las que no hay saltos temporales irregulares y con un patrón de estacionalidad repetitivo. El modelo que se obtiene es el ETS(MNM), el mismo que se obtuvo en el turismo sin intervención.

3.2.2.2. Modelo a batir: ARIMA

Aplicando los logaritmos para transformar los datos, se comprueba que con ARIMA pasa exactamente lo mismo: obtiene una muy buena predicción con un error del 9,81% y un modelo SARIMAX (1, 1, 0)x(0, 1, 0, 12)

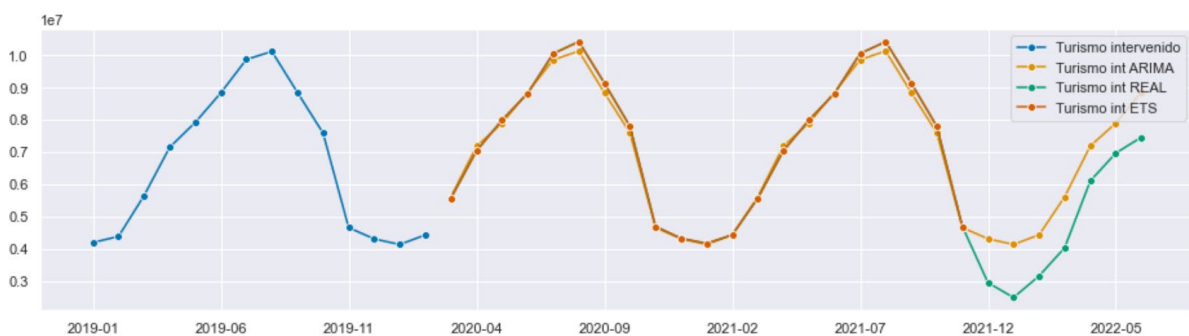


Ilustración 10. Comparación ETS-ARIMA del turismo con la intervención. Fuente: Elaboración propia.

3.2.2.3. Modelo a batir: PROPHET

Este modelo obtiene una predicción pésima aun realizando una intervención para que complete el patrón repetitivo de la estacionalidad, dando un resultado del MAPE del 62,23%. El motivo del fracaso de este modelo es por lo siguiente: el modelo PROPHET estudia mejor el impacto de los componentes de una serie temporal cuando ésta tiene una tendencia y una estacionalidad marcada. La estacionalidad la cumple, pero falla debido a la ausencia de la tendencia. “*PROPHET alcanza sus mejores resultados con series estacionarias o series con una tendencia y estacionalidad marcadas*” (Arribas, 2022).

El modelo ETS sigue siendo factible para este conjunto de datos.

Modelo	MAPE
ETS	8,6%
ARIMA	9,81%
PROPHET	62,2%

Tabla 4. MAPEs modelos para turismo con intervención. Fuente: Elaboración propia.

3.3. SERIE TEMPORAL 3: MATRICULACIONES DE VEHÍCULOS

Para la obtención de los datos del número de matriculaciones de vehículos en España, se ha dirigido a Datos Macro, una página web perteneciente al diario económico “Expansión”. Para el proyecto se ha seleccionado los datos de la columna “Venta mensual de vehículos”, que engloba a los vehículos comerciales como los de pasajeros. Se entienden por coches comerciales aquellos que son diseñados para empresas y autónomos, furgonetas y camiones destinados al trabajo, al transporte de mercancías e incluso desempeñan funciones especiales como vehículos policiales, de emergencias o ambulancias.

Como se ha realizado con los otros conjuntos de datos, primero se analizará la serie temporal sin intervenir el periodo por COVID-19 y otra parte que sí se interviene en los datos.

3.3.1. Matriculaciones sin intervención

El DataFrame con el que se trabaja tiene un total de 103 datos, que son los datos mensuales de las ventas de vehículos en España desde enero del 2014 hasta julio del 2022. Al realizar una representación de sus datos, se aprecia que al principio tiene una tendencia creciente hasta principios del año 2018, pierde esa tendencia y se mantiene constante hasta la llegada de la pandemia por COVID-19, en la que el mes de abril del 2020 se matricularon solamente alrededor de 7000 vehículos, la cifra más baja del conjunto de datos con el que se está trabajando. Por otra parte, las matriculaciones

en España antes de la pandemia poseían una estacionalidad marcada, en la que hay más matriculaciones de vehículos en los siete primeros meses del año, especialmente en marzo, mayo y junio.

Estudiando la estacionariedad, se observa que su media y varianza es constante a lo largo del tiempo, lo que indica que es una serie estacionaria.

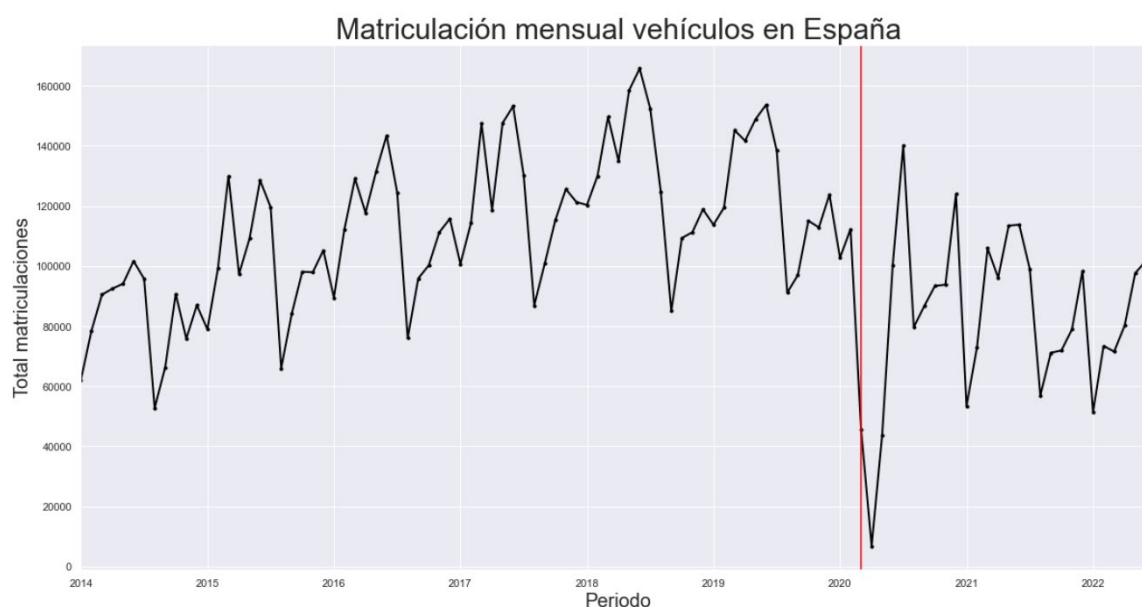


Ilustración 11. Evolución del número de matriculaciones de vehículos en España. Fuente: Elaboración propia.

3.3.1.1. Modelo a batir: ETS

Tras aplicarle el autoETS a la serie temporal se obtiene un modelo ETS(MNM), que quiere decirse que es un modelo con un error y una estacionalidad multiplicativas y sin tendencia. Guarda la lógica de la tendencia debido a que al principio se comportaba de manera creciente y luego decreciente, y su estacionalidad es marcada por los siete primeros meses de matriculaciones. El error, para esta serie temporal, es relativamente alta, con un 42,37% de MAPE.

3.3.1.2. Modelo a batir: ARIMA

La serie, al ser estacionaria, no se le aplica la transformación de los datos, y tras realizarle el autoARIMA, se obtiene un modelo SARIMAX (0, 1, 1)x(0, 1, 1, 12) y un MAPE de casi el 43%.

Realizando la comparación gráfica de estos dos modelos, como ha pasado con las otras series temporales, el momento COVID-19 no lo predice, pero a partir de agosto del 2020 se aprecia una mayor aproximación de las predicciones a los datos reales, especialmente del modelo ARIMA.

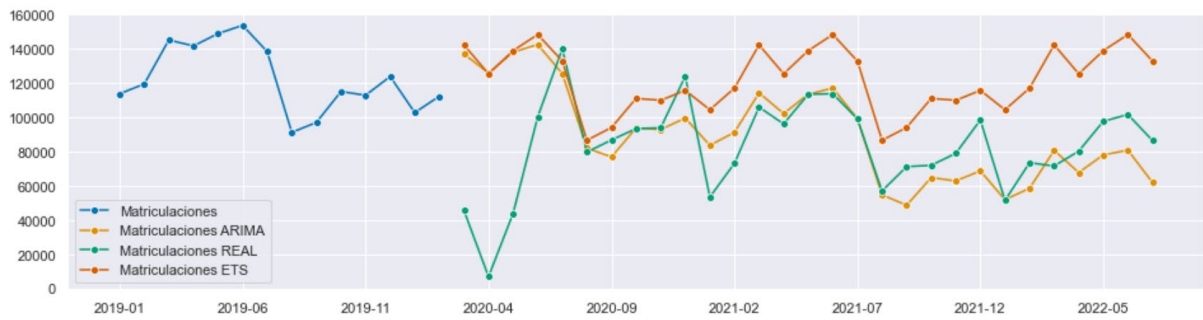


Ilustración 12. Comparativa ETS-ARIMA de las matriculaciones. Fuente: Elaboración propia.

3.3.1.3. Modelo a batir: PROPHET

Modelando la serie temporal con este método se observa que hace el mismo efecto que con ETS y ARIMA.

Analizando sus MAPEs, el modelo ETS es el que vuelve a predecir mejor, aunque por poco y con errores altos.

Modelo	MAPE
ETS	42,37%
ARIMA	42,99%
PROPHET	43,01%

Tabla 5. MAPEs modelos para matriculaciones. Fuente: Elaboración propia.

3.3.2. Matriculaciones con intervención

Se realiza la intervención de esta serie temporal se suprimen los datos comprendidos entre marzo del 2020 hasta agosto del mismo año, que fue cuando los datos de la predicción por ETS y ARIMA se asemejaban a los datos reales.

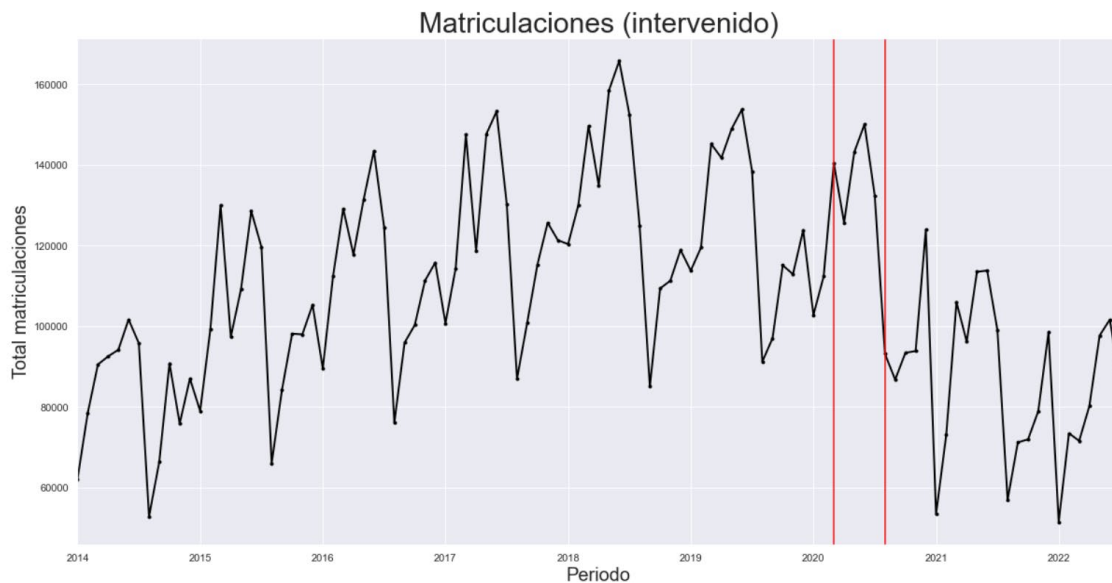


Ilustración 13. Evolución del número de matriculaciones de vehículos en España con la intervención. Fuente: elaboración propia.

La serie intervenida se comporta parecida a la original, donde la tendencia es similar, llegando incluso a llegar a la conclusión de que no la posee, y la estacionalidad es similar.

En cambio, estudiando la estacionariedad se comprueba que ahora no es estacionaria. El motivo de su cambio es porque en la serie original se estudia su estacionariedad hasta febrero del 2020, el mes antes del inicio de la pandemia, pero al estudiar la serie intervenida, se incluyen todos los datos (además de la parte intervenida), y se observa que la parte final de la serie temporal, su tendencia es decreciente y eso influye considerablemente en la constancia de la media y de la varianza de la serie temporal.

3.3.2.1. Modelo a batir: ETS

La predicción obtiene una buena precisión al principio, pero después empieza a perder la aproximación tomando unos datos predictivos superiores a las reales y el MAPE da un porcentaje del 35,42%, lo que indica que esta serie temporal, aun habiéndole intervenido el momento del COVID-19, el efecto de la misma no ha terminado. El modelo que se obtiene sigue siendo el ETS(MNM), el mismo que se obtuvo en el turismo sin intervención.

3.3.2.2. Modelo a batir: ARIMA

Aplicando los logaritmos para transformar los datos, se comprueba que con ARIMA mejora la predicción y obtiene la mitad del error que se obtuvo en la serie original (16,89%). El modelo ARIMA se comporta mejor en modelos que no son estacionarias, de ahí que obtenga mejores resultados que con el modelo ETS.

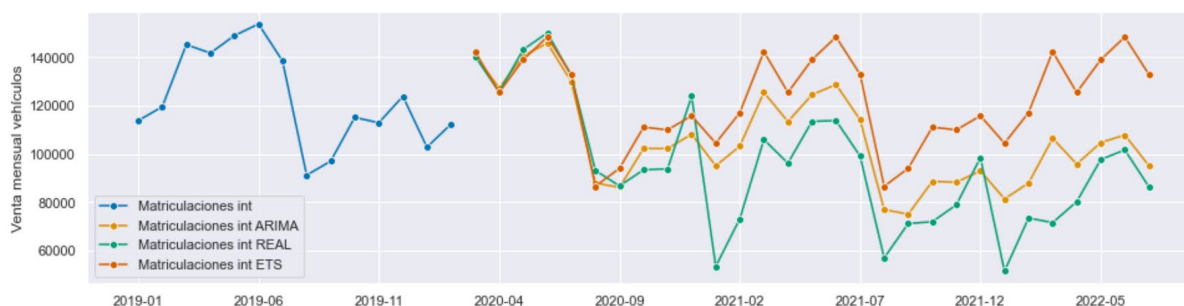


Ilustración 14. Comparación ETS-ARIMA de las matriculaciones con la intervención. Fuente: Elaboración propia.

3.3.2.3. Modelo a batir: PROPHET

Con este modelo se obtienen unas predicciones mejores que con los datos originales, pero el MAPE da un resultado del 27,51%. Finalmente, el mejor modelo predictivo para esta serie temporal es la ARIMA.

Modelo	MAPE
ETS	35,42%
ARIMA	16,89%
PROPHET	27,51%

Tabla 6. MAPEs modelos para matriculaciones con la intervención. Fuente: Elaboración propia.

3.4. SERIE TEMPORAL 4: HIPOTECAS

Para la obtención de los datos del número de hipotecas en España, se ha recurrido a los datos ofrecidos por la página web del Gobierno de España. Para el análisis, se han seleccionado los datos mensuales sobre el total de fincas, que engloba a todo tipo de inmuebles, y a nivel nacional.

3.4.1. Hipotecas sin intervención

El DataFrame de la serie de tiempo contiene un total de 229 datos mensuales que se desarrolla desde abril del año 2003 hasta mayo del 2022. Para mejorar su estudio, se reduce el tamaño y se empieza a analizar desde enero del 2015. Al realizar una representación de sus datos, se aprecia una tendencia creciente en todo momento a excepción del inicio de la pandemia, en la que se aprecia un descenso, aunque se recupera rápidamente a partir del inicio del año 2021. La serie temporal tiene un comportamiento mensual en la que los meses de enero y mayo son los que más hipotecan, pero observando la descomposición de la serie no cumple con la estacionalidad marcada, pues la gráfica desglosada se observa que no cumple un patrón repetitivo.

Por otra parte, no es una serie estacionaria debido a que su media y varianza no es constante a lo largo del tiempo, por lo que se le realiza ninguna transformación en los datos de esta serie temporal para modelarlo con el ARIMA.

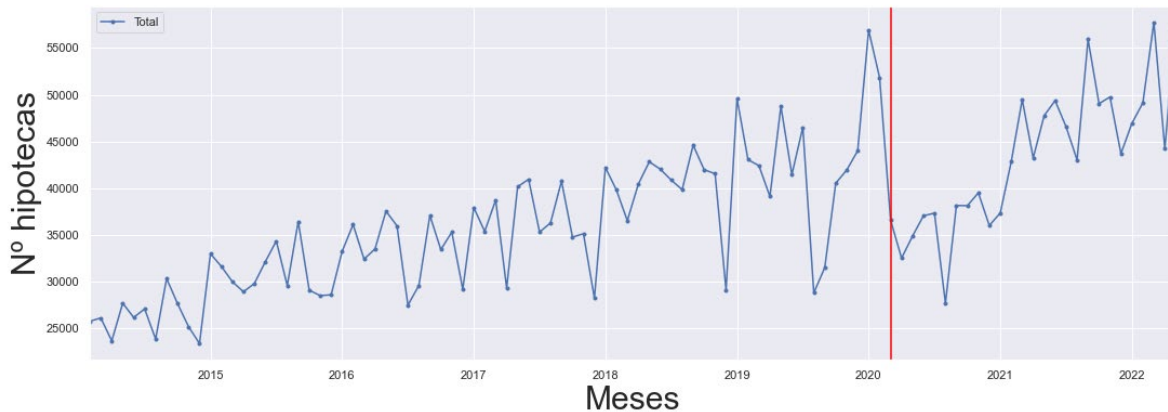


Ilustración 15. Evolución del número de hipotecas del total de fincas en España. Fuente: Elaboración propia.

3.4.1.1. Modelo a batir: ETS

Tras aplicarle el autoETS a la serie temporal se obtiene un modelo ETS(MAM), es decir, un modelo multiplicativo de Hol-Winters con errores multiplicativos. Al principio no predice el momento por COVID-19, pero a partir de marzo del 2021 se ve cómo el modelo autoETS predice con más exactitud los datos reales.

3.4.1.2. Modelo a batir: ARIMA

La serie, al ser no estacionaria, se le aplica la transformación de los datos, y tras realizarle el autoARIMA, se obtiene un modelo SARIMAX (2, 1, 0)x(0, 1, [1], 12), en este caso con auto-regresión de nivel 2. La predicción se ve que es peor que el realizado por el autoETS debido a la carencia de estacionalidad, que es una componente muy a tener en cuenta en estos modelos predictivos.

Realizando la comparación gráfica de estos dos modelos se observa lo siguiente:



Ilustración 16. Comparativa ETS-ARIMA de las hipotecas. Fuente: Elaboración propia

Como se ha comentado anteriormente, el modelo por autoETS es mucho más precisa y además tiene un error inferior (16,42% frente al 35,72% que da ARIMA).

3.4.1.3. Modelo a batir: PROPHET

Modelando la serie temporal con este método se observa que el modelo Prophet es más preciso aún que el ETS. La gráfica del modelo PROPHET de esta serie temporal se ubica en el apéndice, en la página 57.

Analizando sus MAPEs, el modelo PROPHET aquí es el que predice mejor y el que se va a emplear para el resultado de la predicción para finales del año 2022.

Modelo	MAPE
ETS	16,42%
ARIMA	35,72%
PROPHET	15,08%

Tabla 7. MAPEs modelos para hipotecas. Fuente: Elaboración propia.

3.4.2. Matriculaciones con intervención

Se realiza la intervención de esta serie temporal se suprimen los datos comprendidos entre marzo del 2020 hasta marzo del año siguiente.

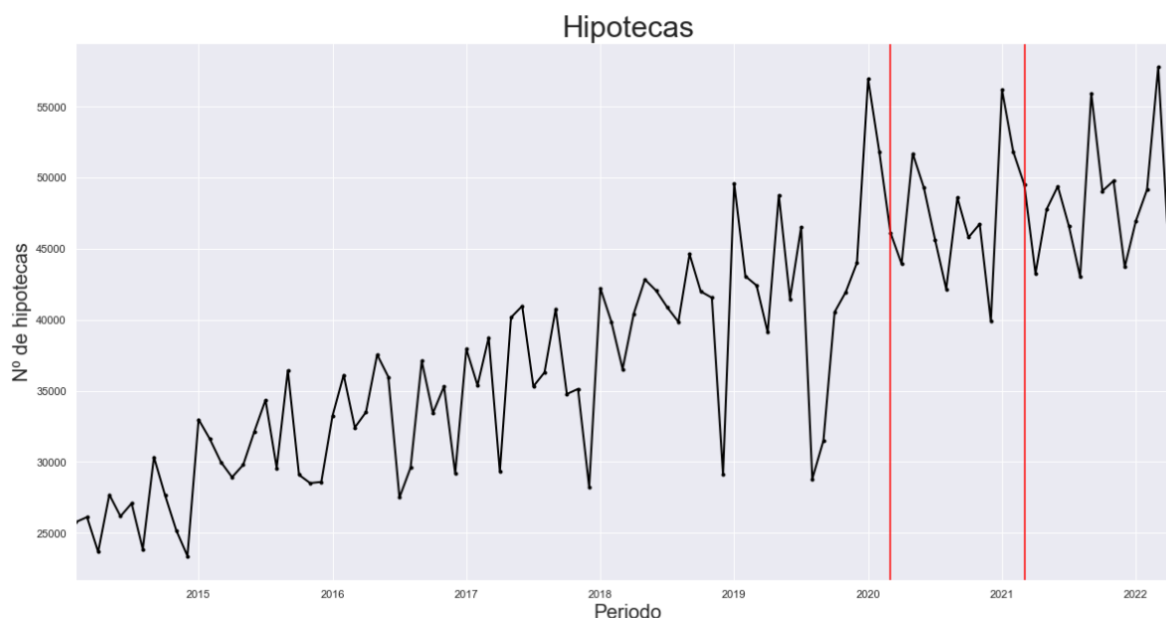


Ilustración 17. Evolución del número de hipotecas del total de fincas en España con la intervención. Fuente: elaboración propia.

Los componentes de la serie cumplen con las mismas características que los de la serie original, aunque mejorando la tendencia al haber ausencia por el momento del COVID-19.

3.4.2.1. Modelo a batir: ETS

La predicción obtiene una buena precisión y un MAPE del 7,17%. El modelo que se obtiene sigue siendo el ETS(MNM), el mismo que se obtuvo en el turismo sin intervención.

3.4.2.2. Modelo a batir: ARIMA

Aplicando los logaritmos para transformar los datos, se comprueba que con ARIMA también se obtiene una mejor predicción y un MAPE del 6,67%.

Haciendo un contraste gráfico, se aprecia la diferencia de los modelos y la fidelidad del modelo autoARIMA con los datos reales.

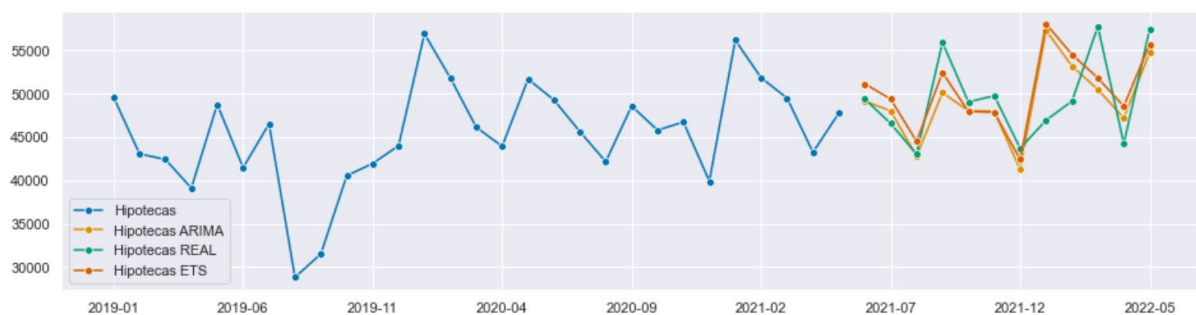


Ilustración 18. Comparación ETS-ARIMA de las hipotecas con la intervención. Fuente: Elaboración propia.

3.4.2.3. Modelo a batir: PROPHET

Con este modelo, sin embargo, se obtienen unos resultados peores incluso que si no se hubiese intervenido la serie temporal, lo cual resulta bastante curioso que haga este efecto.

Modelo	MAPE
ETS	7,17%
ARIMA	6,67%
PROPHET	24,42%

Tabla 8. MAPEs modelos para hipotecas con la intervención. Fuente: Elaboración propia.

Con los resultados obtenidos, se observa que, para esta ocasión, el mejor modelo de predicción para esta serie temporal es la ARIMA.

4. RESULTADOS DE LAS PREDICCIONES

4.1. SERIE TEMPORAL 1: DESEMPLEO

4.1.1. Resultado sin la intervención

Para la predicción de la serie temporal del desempleo se emplea el modelo que mejor predijo, en este caso fue el ETS. El pronóstico que toma la serie temporal para finales del año 2022 tiene una tendencia creciente y coincidente con el del resto de los finales de año, a excepción de la etapa de COVID-19. Tiene sentido, pues el modelo ETS al final es útil para conjunto de datos con estacionalidad como es este caso y hace un cálculo promedio móvil simple.

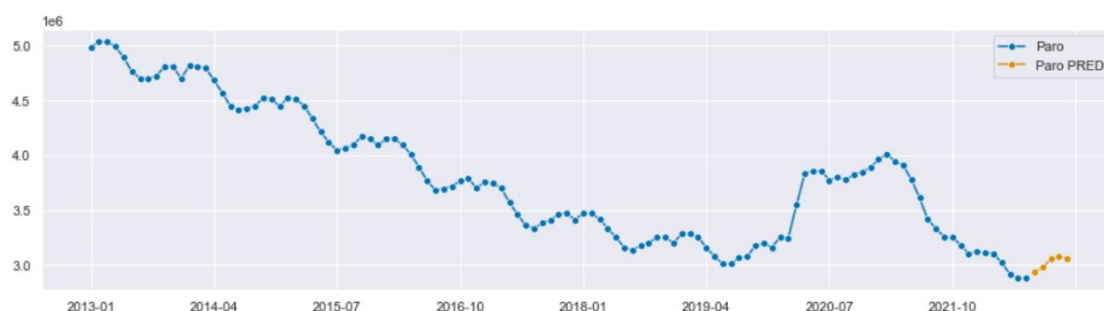


Ilustración 19. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.

A continuación, los datos obtenidos en números:

Periodo	Total desempleados
2022-08	2.938.509
2022-09	2.978.051
2022-10	3.057.594
2022-11	3.080.354
2022-12	3.058.159

Tabla 9. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.

4.1.2. Resultado con la intervención

Para la predicción se emplea de nuevo ETS, y lo que predice para finales del año 2022 es un pronóstico muy similar a la de la serie original.

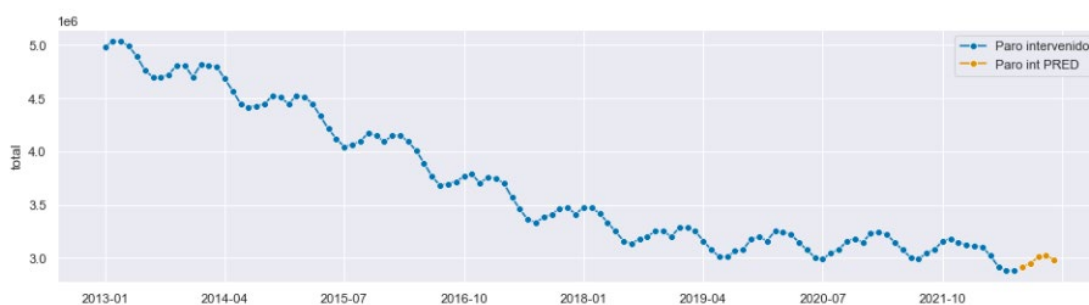


Ilustración 20. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.

Periodo	Total desempleados
2022-08	2.920.012
2022-09	2.947.077
2022-10	3.012.890
2022-11	3.026.518
2022-12	2.985.489

Tabla 10. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.

4.2. SERIE TEMPORAL 2: TURISMO

4.2.1. Resultado sin la intervención

La predicción se obtiene con el modelo ETS, y para finales del año 2022 los resultados obtenidos son más altos que los de los datos pasados, por encima de los 13 millones de turistas, y no cumple del todo con el patrón estacional.

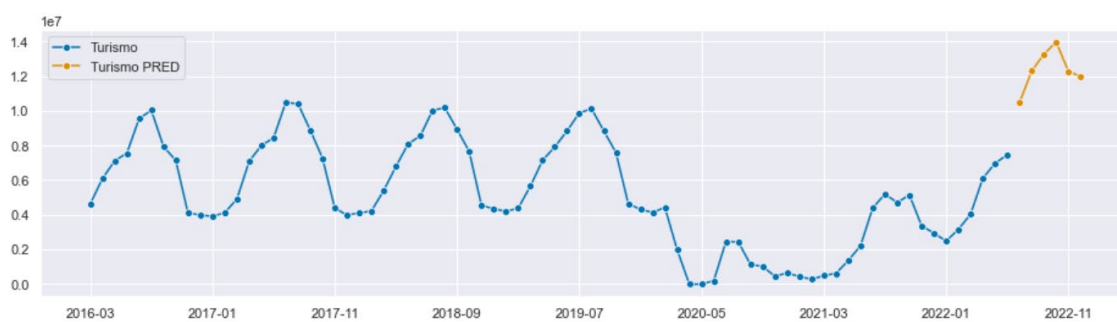


Ilustración 21. Predicción del desempleo sin intervenir. Fuente: Elaboración propia.

A continuación, los datos obtenidos en números:

Periodo	Total desempleados
2022-07	10.457.316
2022-08	12.292.119
2022-09	13.259.977
2022-10	13.966.227
2022-11	12.280.633
2022-12	11.972.648

Tabla 11. Predicción del turismo sin intervenir. Fuente: Elaboración propia.

4.2.2. Resultado con la intervención

La predicción que toma para finales del año 2022, realizada nuevamente con el modelo ETS, guarda más sentido que las obtenidas en la serie original, pues ésta tiene una tendencia decreciente y coincidente con el del resto de los finales de año, y tiene sentido: esta serie temporal tiene un componente estacional muy fuerte y cumple la predicción con ese patrón decreciente a finales de año, en la que no hay tanto turismo por ser la estación más fría del año, y eso no atrae al turismo. Está

claro que los datos obtenidos en la serie original están bastante alterados por el efecto del COVID-19, en la que los meses de abril y mayo los resultados eran cero.

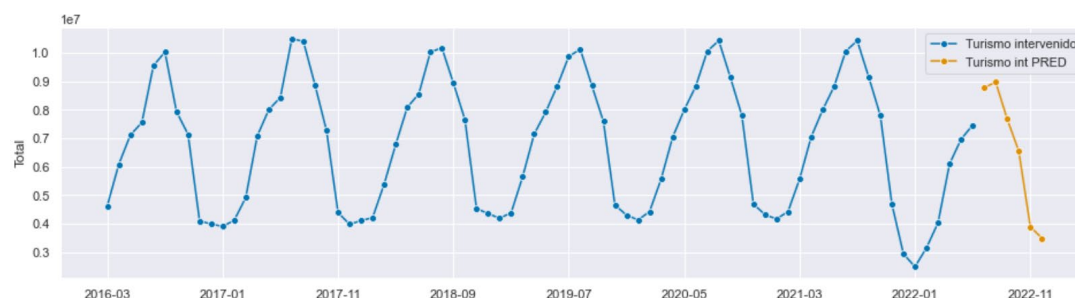


Ilustración 22. Predicción del turismo intervenido. Fuente: Elaboración propia.

Periodo	Total turistas
2022-07	8.777.526
2022-08	8.965.844
2022-09	7.670.754
2022-10	6.544.311
2022-11	3.896.953
2022-12	3.495.512

Tabla 12. Predicción del turismo intervenido. Fuente: Elaboración propia.

4.3. SERIE TEMPORAL 3: MATRICULACIONES DE VEHÍCULO

4.3.1. Resultado sin la intervención

La predicción por ETS que toma para finales del año 2022 tiene una tendencia creciente y sin ningún tipo de patrón repetitivo por datos pasados. La alteración es debido por la etapa COVID-19.



Ilustración 23. Predicción de las matriculaciones sin intervenir. Fuente: Elaboración propia.

Periodo	Total matriculaciones
2022-08	58.813
2022-09	64.028
2022-10	73.290
2022-11	73.714
2022-12	82.743

Tabla 13. Predicción de las matriculaciones sin intervenir. Fuente: Elaboración propia.

4.3.2. Resultado con la intervención

Para esta predicción se aplica ARIMA, y los resultados obtenidos para finales del año 2022 tiene una tendencia similar a la de la serie sin la intervención, es decir, una tendencia creciente y coincidente con el del resto de los finales de año, al tener esta serie temporal un componente estacional marcado cumple con la predicción gracias a ese patrón estacional.

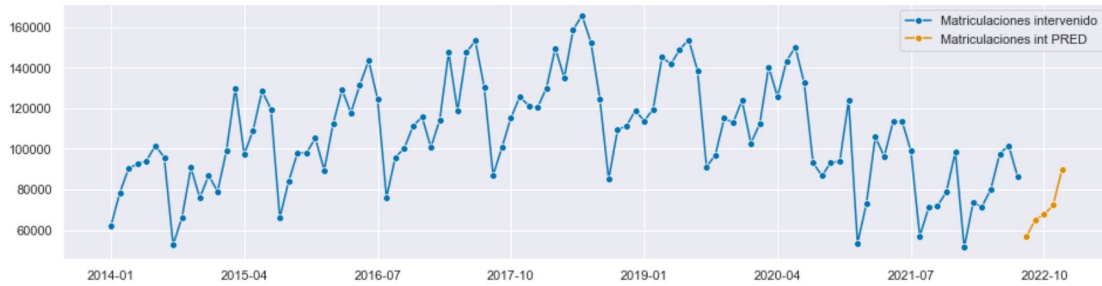


Ilustración 24. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.

Periodo	Total matriculaciones
2022-08	56.754
2022-09	65.244
2022-10	67.621
2022-11	72.332
2022-12	89.770

Tabla 14. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.

4.4. SERIE TEMPORAL 4: HIPOTECAS

4.4.1. Resultado sin la intervención

La predicción con el modelo PROPHET que toma para finales del año 2022 tiene una tendencia creciente sin patrón estacional debido a la alteración de la serie por el COVID-19 y por el carácter de esta serie temporal.

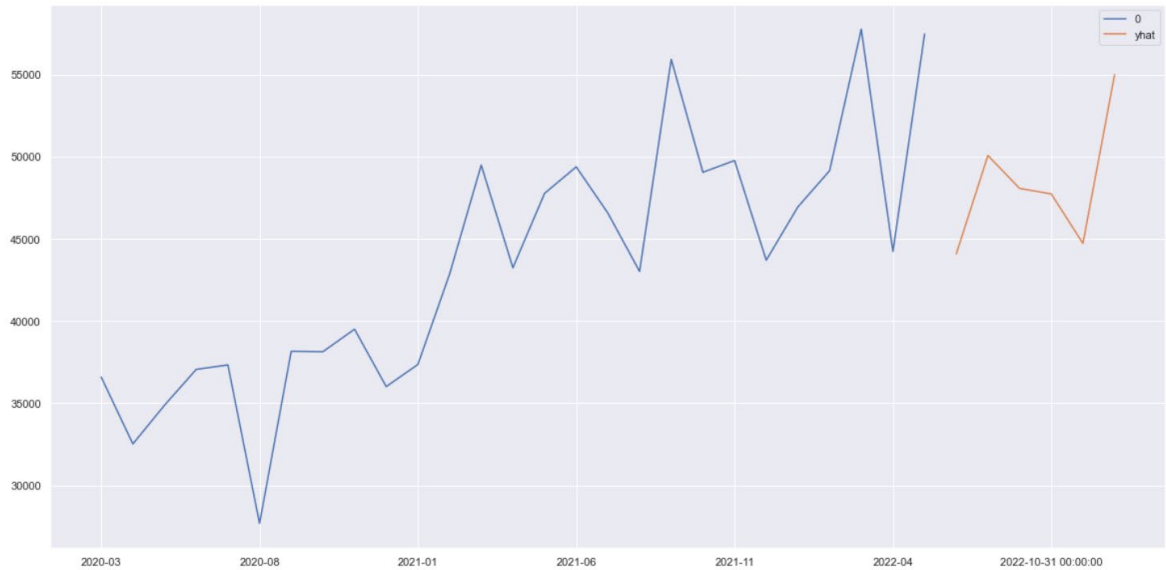


Ilustración 25. Predicción de las hipotecas sin intervenir. Fuente: Elaboración propia.

Periodo	Total matriculaciones
2022-06	49.934
2022-07	44.082
2022-08	50.074
2022-09	48.070
2022-10	47.736
2022-11	44.724
2022-12	54.993

Tabla 15. Predicción de las hipotecas sin intervenir. Fuente: Elaboración propia.

4.4.2. Resultado con la intervención

La predicción (ARIMA) que toma para finales del año 2022 obtiene unos resultados bastante parecidos a los obtenidos por el modelo PROPHET para la serie temporal original.



Ilustración 26. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.

Periodo	Total matriculaciones
2022-06	55.379
2022-07	53.242
2022-08	48.461
2022-09	56.086
2022-10	53.314
2022-11	53.657
2022-12	47.941

Tabla 16. Predicción del número de matriculaciones en España intervenido. Fuente: Elaboración propia.

5. CONCLUSIONES

Una vez analizado todo el proceso de análisis de las series temporales originales y las series con la intervención del COVID-19, los ajustes de cada uno de los modelos predictivos seleccionados para el proyecto y las predicciones obtenidas para finales del año 2022 usando los mejores modelos para cada serie, se han llegado a obtener varias conclusiones.

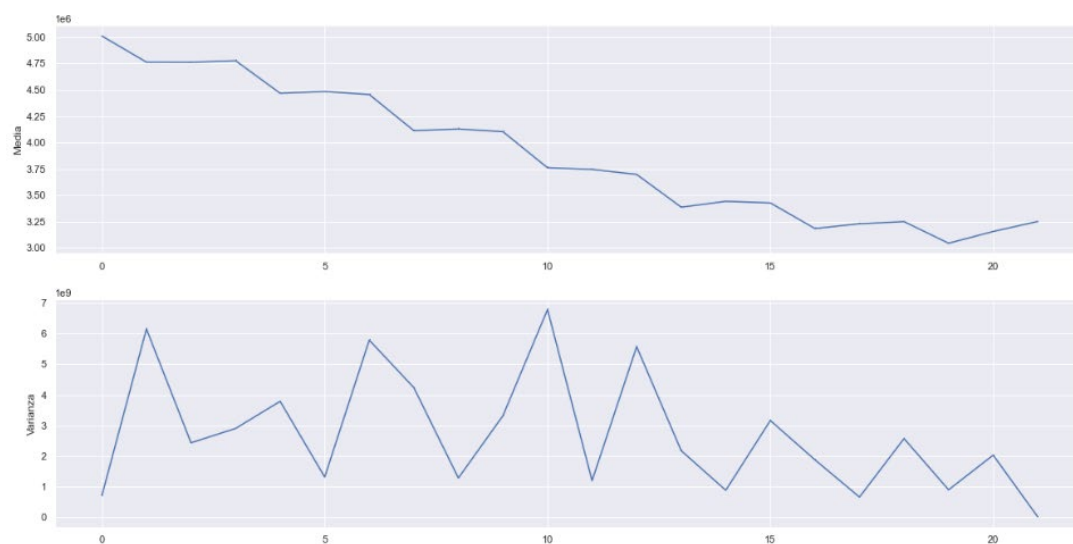
La primera conclusión a la que se llega es que el modelo ETS es el modelo predictivo estadístico que mejores resultados se han obtenido y más se ha repetido para realizar la predicción para finales del año vigente, ya que casi todas las series temporales que se han empleado cumplían con el requisito de una estacionalidad marcada. En el caso del modelo de ARIMA, se ha comportado mejor con series temporales que cumplen con la condición de no estacionariedad, como ha sido el caso de las matriculaciones de vehículos.

La segunda deducción es que la introducción del modelo PROPHET para predecir las series temporales, se han obtenido unos resultados inferiores a los que se esperaban, a excepción de los resultados obtenidos en la serie temporal original de las hipotecas. Por lo general, no han sido buenas predicciones porque este modelo es más preciso para modelos que tengan puntos de datos atípicos grandes, eventos irregulares o con tendencias de crecimiento no lineal que se aproximen a un límite.

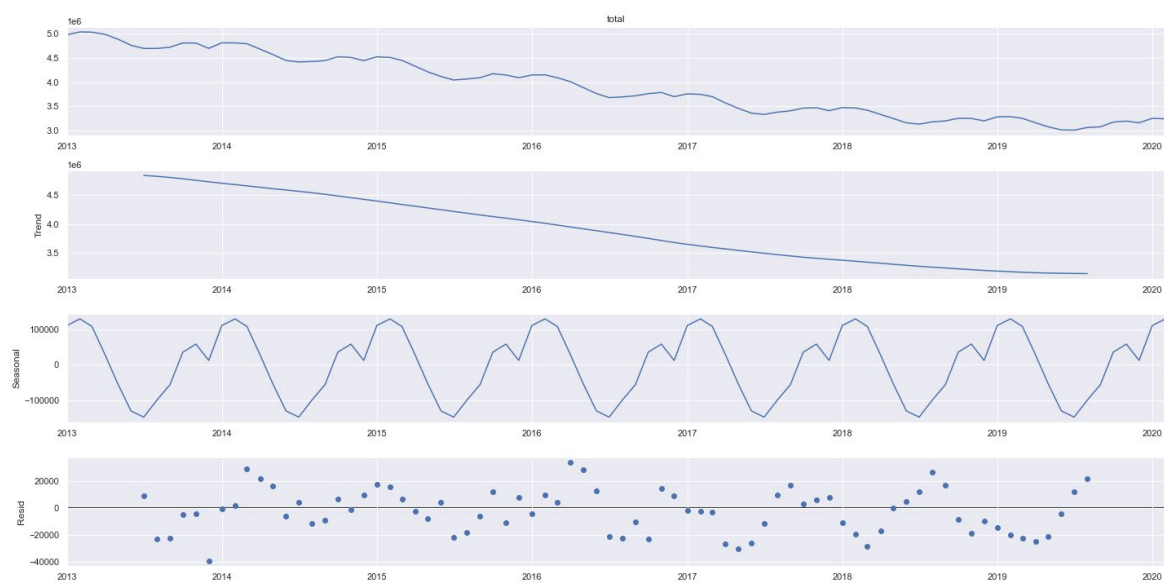
Como conclusión final y general, se ha comprobado que la intervención realizada en las series temporales eliminando la parte afectada por el efecto del COVID-19 ha obtenido unos resultados más precisos que en los resultados obtenidos en las series originales. Esto significa que los modelos predictivos tradicionales, basados en la estadística, fallan cuando hay un fenómeno temporal irregular, como es el COVID-19, pero son modelos óptimos para emplearlos en series temporales con una estacionalidad o una tendencia marcada. Para series temporales que no cumplan con estos requisitos, habría que recurrir al uso de modelos predictivos basados en redes neuronales o en Machine Learning.

APÉNDICE

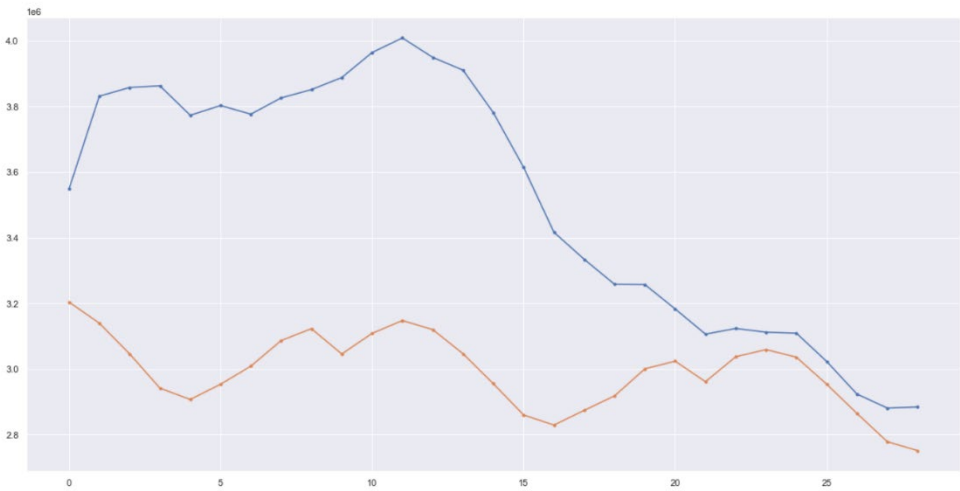
Gráfica de la media y de la varianza del desempleo



Gráfica de la descomposición del desempleo



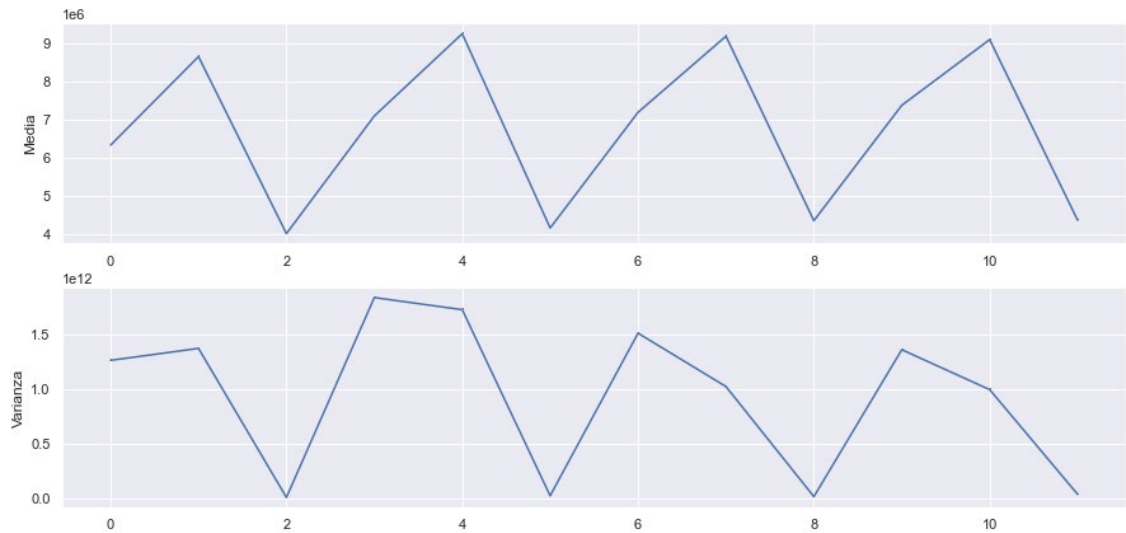
Gráfica del modelo *PROPHET* del desempleo



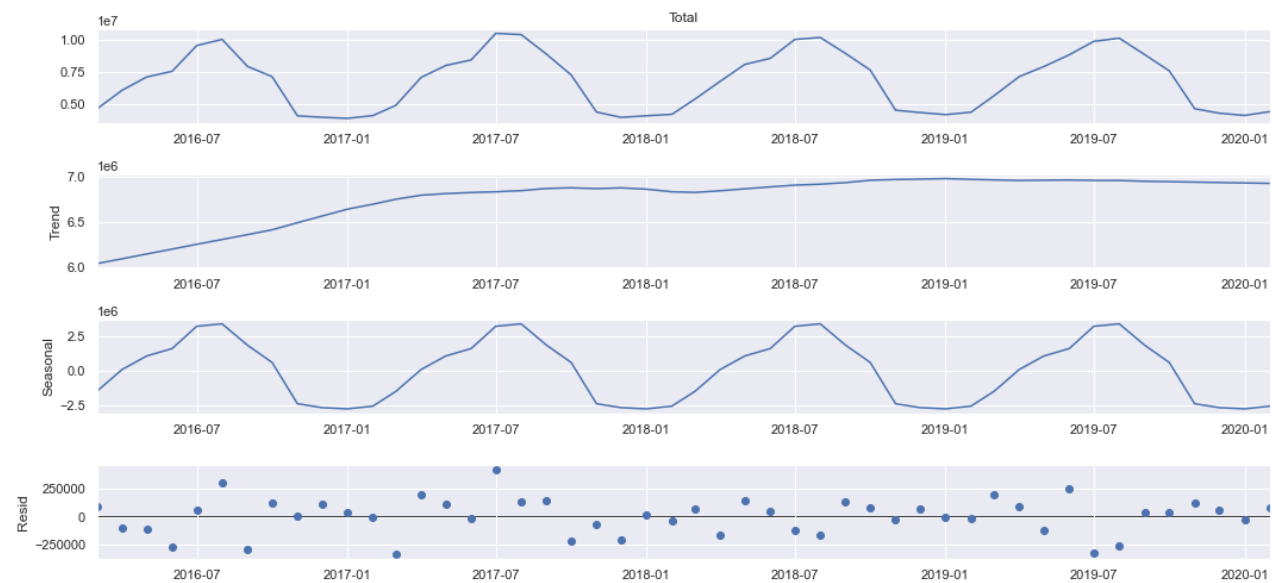
Gráfica del modelo *PROPHET* del desempleo con la intervención



Gráfica de la media y de la varianza del turismo



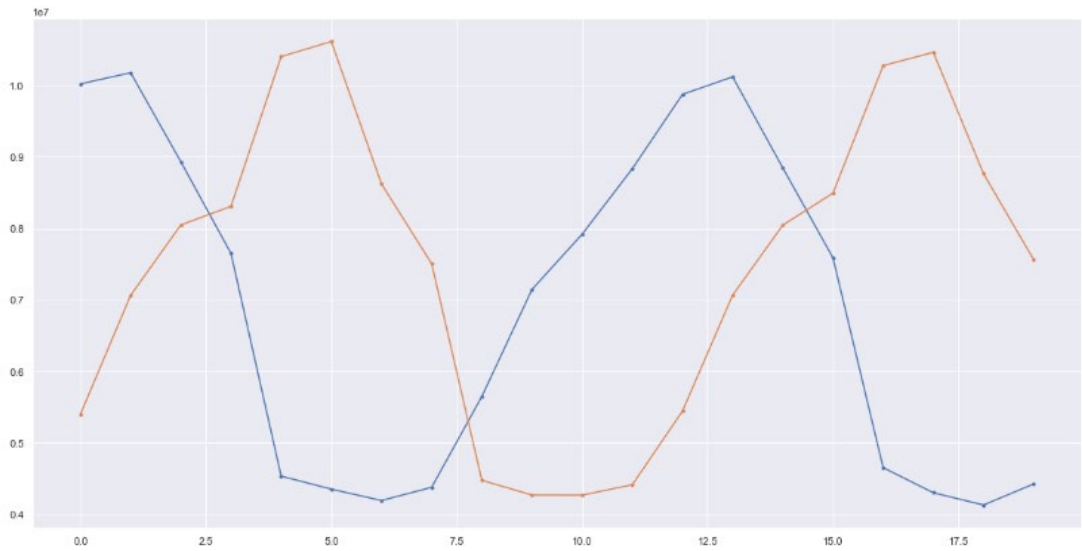
Gráfica de la descomposición del turismo



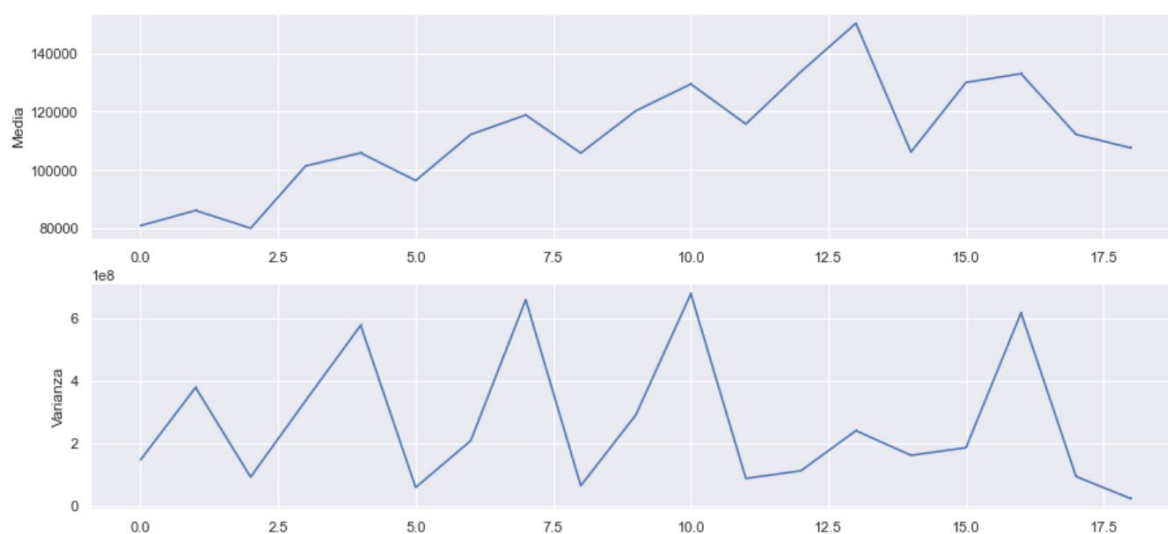
Gráfica del modelo *PROPHET* del turismo



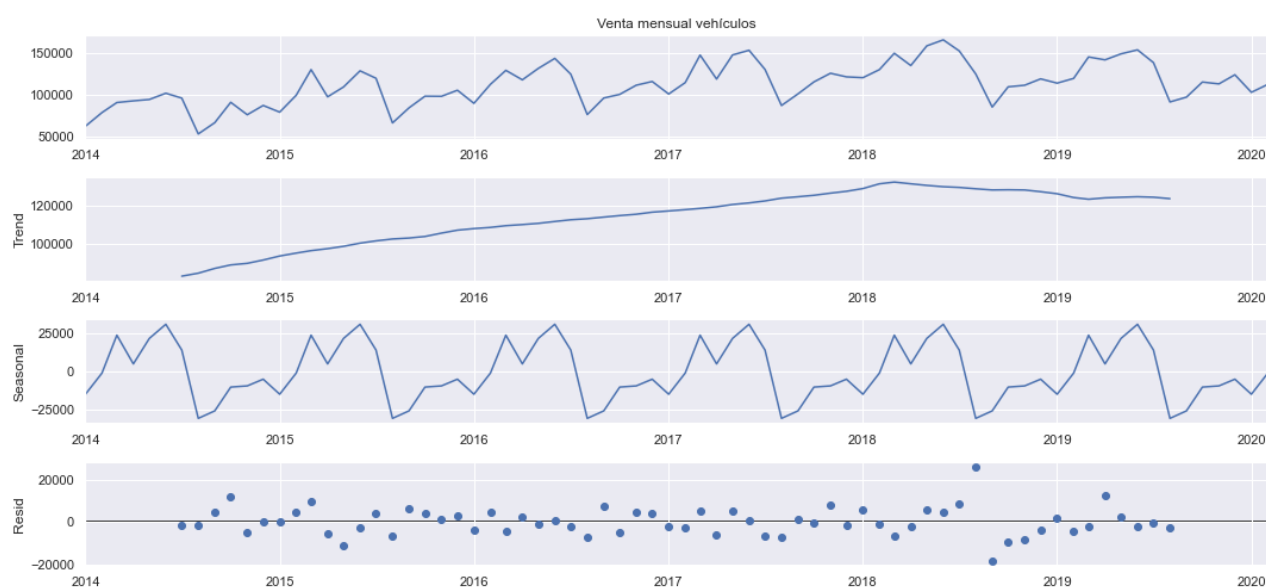
Gráfica del modelo *PROPHET* del turismo con la intervención



Gráfica de la media y de la varianza de las matriculaciones de vehículos



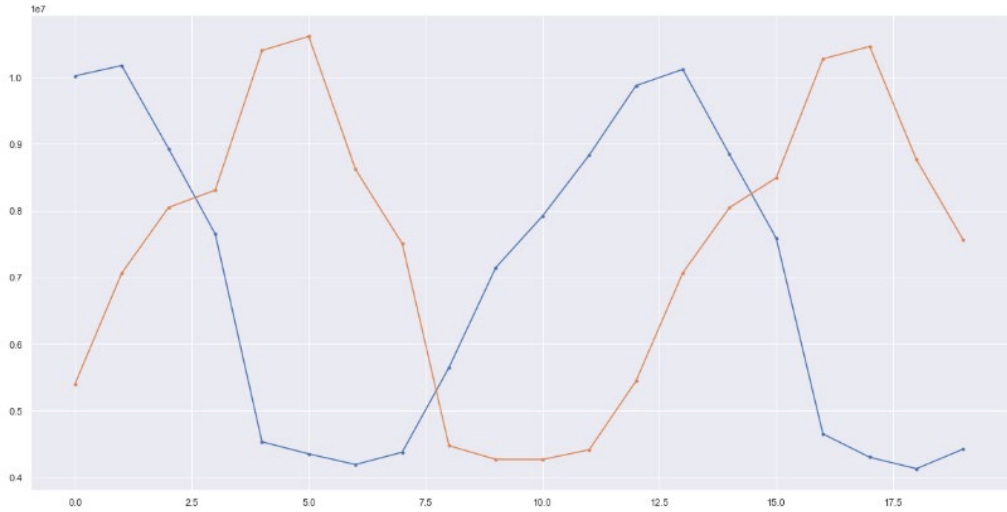
Gráfica de la descomposición de las matriculaciones de vehículos



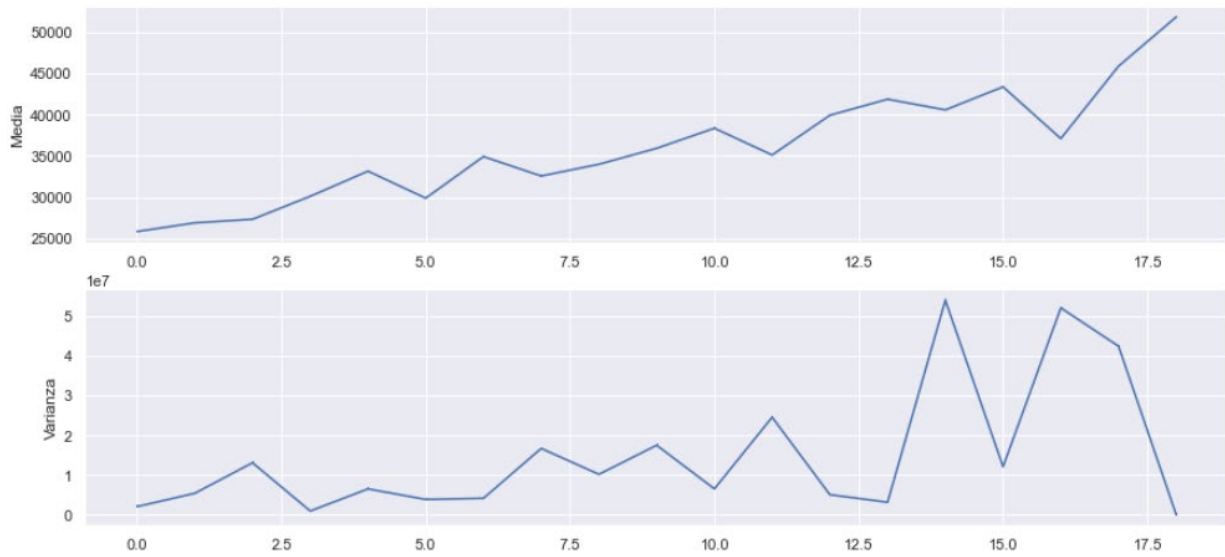
Gráfica del modelo *PROPHET* de las matriculaciones de vehículos



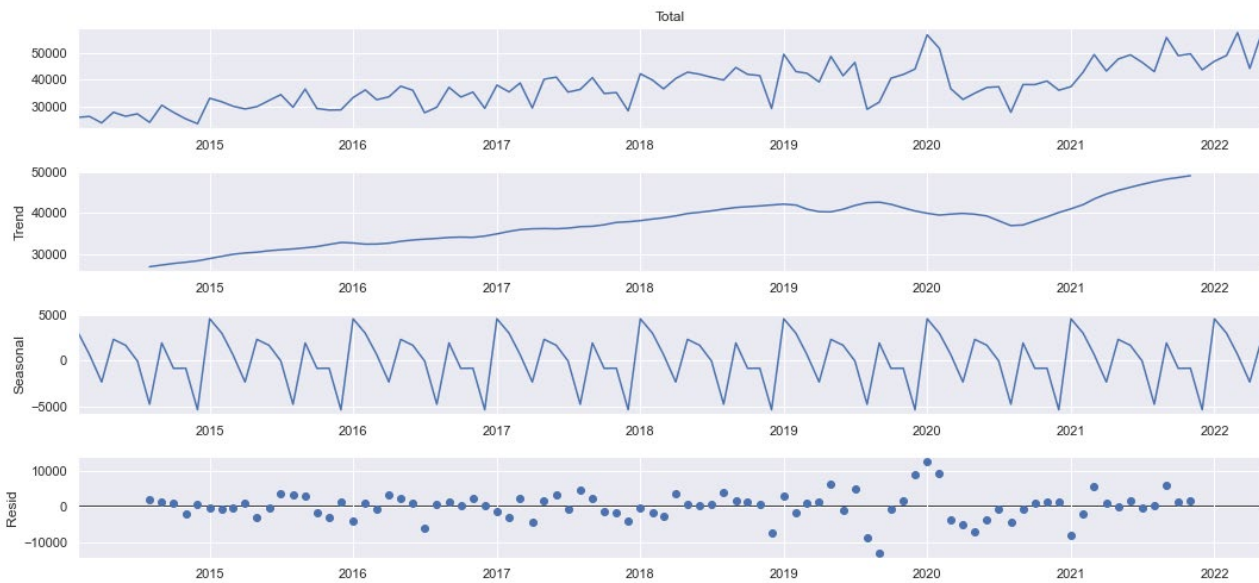
Gráfica del modelo *PROPHET* de las matriculaciones de vehículos con la intervención



Gráfica de la media y de la varianza de las hipotecas del total de fincas



Gráfica de la descomposición de las hipotecas



Gráfica del modelo *PROPHET* de las hipotecas



Gráfica del modelo *PROPHET* de las hipotecas con la intervención



REFERENCIAS

- ArcGIS. (s.f.). *ArcGIS Pro*. Obtenido de <https://pro.arcgis.com/es/pro-app/2.8/tool-reference/space-time-pattern-mining/learnmoreexponentialsmoothingforecast.htm>
- Arribas, M. A.-C. (26 de Abril de 2022). *Model Differently*. Obtenido de https://www.modeldifferently.com/2022/04/analisis_prediccion_ts_prophet/#:~:text=As%C3%AD%20se%20puede%20estudiar%20el,la%20aleatoria%20de%20la%20muestra
- Athanasopoulos, R. J. (2018). *Otexts*. Obtenido de <https://otexts.com/fpp2/arima.html>
- Criado, J. M. (s.f.). *analisisdedatos.net*. Obtenido de <https://analisisdedatos.net/analisis/ST/estacionariedad.php>
- IBM. (Agosto de 2020). *www.ibm.com*. Obtenido de <https://www.ibm.com/es-es/cloud/learn/exploratory-data-analysis>
- Iglesias, I. H. (Septiembre de 2021). *Universidad Complutense de Madrid*. Obtenido de <https://eprints.ucm.es/id/eprint/68280/1/hoyuela-iglesias-ivan-tfm.pdf>
- León, M. d. (1 de Junio de 2020). *madrимasd.org*. Obtenido de <https://www.madrимasd.org/blogs/matematicas/2020/06/01/148025>
- Lewinson, E. (s.f.). *Towards Data Science*. Obtenido de <https://towardsdatascience.com/time-series-diy-seasonal-decomposition-f0b469afed44>
- Madridmasd*. (1 de Junio de 2019). Obtenido de <https://www.madrимasd.org/blogs/matematicas/2020/06/01/148025>
- Ortiz, D. U. (1 de Agosto de 2020). *QuantSpace*. Obtenido de <https://quantspace.es/2020/08/01/analisis-de-series-temporales-con-python-parte-2/>
- Parra, F. (25 de Enero de 2019). *Bookdown*. Obtenido de <https://bookdown.org/content/2274/series-temporales.html#metodos-de-suavizado>
- Queralt Sánchez de las Matas, R. (2021). *Predicción de Series Temporales*. Madrid, Madrid, España.