



# **Universidad Nacional Mayor de San Marcos**

**Universidad del Perú. Decana de América**

**Facultad de Ciencias Matemáticas**

**Escuela Profesional de Estadística**

## **Pronóstico del gasto medio por turista en España mediante el uso de series temporales**

### **TRABAJO DE SUFICIENCIA PROFESIONAL**

Para optar el Título Profesional de Licenciada en Estadística

#### **AUTOR**

Evelyn Yanela NOLASCO PALOMINO

#### **ASESOR**

Dr. Helfer Joel MOLINA QUIÑONES

Lima, Perú

2022



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Nolasco, E. (2022). *Pronóstico del gasto medio por turista en España mediante el uso de series temporales*. [Trabajo de suficiencia profesional de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Matemáticas, Escuela Profesional de Estadística]. Repositorio institucional Cybertesis UNMSM.

---

## Metadatos complementarios

<b>Datos de autor</b>	
Nombres y apellidos	Evelyn Yanela Nolasco Palomino
Tipo de documento de identidad	DNI
Número de documento de identidad	74202449
URL de ORCID	<a href="https://orcid.org/0000-0001-5805-4202">https://orcid.org/0000-0001-5805-4202</a>
<b>Datos de asesor</b>	
Nombres y apellidos	Helfer Joel Molina Quiñones
Tipo de documento de identidad	DNI
Número de documento de identidad	40014631
URL de ORCID	<a href="https://orcid.org/0000-0003-1307-7253">https://orcid.org/0000-0003-1307-7253</a>
<b>Datos del jurado</b>	
<b>Presidente del jurado</b>	
Nombres y apellidos	Zoraida Judith Huamán Gutierrez
Tipo de documento	DNI
Número de documento de identidad	09890094
<b>Miembro del jurado 1</b>	
Nombres y apellidos	Ofelia Roque Paredes
Tipo de documento	DNI
Número de documento de identidad	06243124
<b>Datos de investigación</b>	
Línea de investigación	A.3.2.6. Análisis de Datos y Modelamiento de Problemas de la Sociedad (Empresa, Instituciones, Poblaciones locales, regionales y nacionales)

Grupo de investigación	No aplica
Agencia de financiamiento	Sin financiamiento
Ubicación geográfica de la investigación	Edificio: UNMSM País: Perú Departamento: Lima Provincia: Lima Distrito: Cercado de Lima Centro poblado: - Urbanización: - Manzana y lote: - Calle: - Latitud: -12.056423 Longitud: -77.0843327
Año o rango de años en que se realizó la investigación	Enero 2022 - mayo 2022
URL de disciplinas OCDE	Estadísticas, Probabilidad <a href="https://purl.org/pe-repo/ocde/ford#1.01.03">https://purl.org/pe-repo/ocde/ford#1.01.03</a>



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

Universidad del Perú. Decana de América  
FACULTAD DE CIENCIAS MATEMÁTICAS  
ESCUELA PROFESIONAL DE ESTADÍSTICA

**ACTA DE SUSTENTACIÓN DEL TRABAJO DE SUFICIENCIA PROFESIONAL EN  
LA MODALIDAD VIRTUAL PARA LA OBTENCIÓN DEL TÍTULO  
PROFESIONAL DE LICENCIADA EN ESTADÍSTICA  
(PROGRAMA DE TITULACIÓN PROFESIONAL 2022-I)**

En la UNMSM – Ciudad Universitaria – Facultad de Ciencias Matemáticas, siendo las 11:00 horas del sábado 23 de julio del 2022, se reunieron los docentes designados como Miembros del Jurado Evaluador (PROGRAMA DE TITULACIÓN PROFESIONAL 2022-I): Dra. Zoraida Judith Huamán Gutiérrez (PRESIDENTE), Dra. Ofelia Roque Paredes (MIEMBRO) y el Dr. Helfer Joel Molina Quiñones (MIEMBRO ASESOR), para la sustentación del Trabajo de Suficiencia Profesional titulado: “**PRONÓSTICO DEL GASTO MEDIO POR TURISTA EN ESPAÑA MEDIANTE EL USO DE SERIES TEMPORALES**”, presentado por la señorita **Bachiller Evelyn Yanela Nolasco Palomino**, para optar el Título Profesional de Licenciada en Estadística.

Luego de la exposición del Trabajo de Suficiencia Profesional, la Presidente invitó a la expositora a dar respuesta a las preguntas formuladas.

Realizada la evaluación correspondiente por los Miembros del Jurado Evaluador, la expositora mereció la aprobación de **BUENO**, con un calificativo promedio de **DIECISÉIS (16)**.

A continuación, los Miembros del Jurado Evaluador dan manifiesto que la participante **Bachiller Evelyn Yanela Nolasco Palomino** en vista de haber aprobado la sustentación de su Trabajo de Suficiencia Profesional, será propuesta para que se le otorgue el Título Profesional de Licenciada en Estadística.

Siendo las 11:30 horas se levantó la sesión firmando para constancia la presente Acta.

Dra. Zoraida Judith Huamán Gutiérrez  
**PRESIDENTE**

Dra. Ofelia Roque Paredes  
**MIEMBRO**

Dr. Helfer Joel Molina Quiñones  
**MIEMBRO ASESOR**



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS  
Universidad del Perú. Decana de América  
FACULTAD DE CIENCIAS MATEMÁTICAS  
ESCUELA PROFESIONAL DE ESTADÍSTICA

**INFORME DE EVALUACIÓN DE ORIGINALIDAD**

El Director de la Escuela Profesional de Estadística, Dr. Roger Pedro Norabuena Figueroa, informa lo siguiente:

1. Operador del programa informático de similitudes: Dr. Roger Pedro Norabuena Figueroa
2. Documento evaluado: PRONÓSTICO DEL GASTO MEDIO POR TURISTA EN ESPAÑA MEDIANTE EL USO DE SERIES TEMPORALES
1. Autor de la tesis: EVELYN YANELA NOLASCO PALOMINO
2. Fecha de recepción de la tesis: 11/10/2022
3. Fecha de aplicación del programa informático de similitudes: 11/10/2022
  - Software utilizado: Turnitin
4. Configuración del programa detector de similitudes:
  - Excluye textos entrecorridos
  - Excluye bibliografía
  - Excluye cadenas menores a 40 palabras
5. Porcentaje de similitudes según programa detector de similitudes: diez por ciento (10%)
6. Fuentes originales de las similitudes encontradas:
  - Fuentes de internet: 11 %
  - Publicaciones: 1 %
7. Calificación de originalidad:
  - Documento cumple criterios de originalidad, sin observaciones

Lima, 11 de octubre del 2022



Firmado digitalmente por  
NORABUENA FIGUEROA Roger  
Pedro FAU 20148092282 soft  
Motivo: Soy el autor del documento  
Fecha: 12.10.2022 09:45:02 -05:00

Dr. Roger Pedro Norabuena Figueroa  
Director

## RESUMEN

La empresa está dedicada a brindar consultoría en diferentes proyectos, aquí nos centraremos en el proyecto TurismEsp en específico, el cual está orientado al turismo en España, pero algo que aún no se ha implementado es la vista del gasto por turista en España, y esto es de suma importancia ya que se pueden implementar diversas promociones al indagar más sobre ello. Por ello, este trabajo tiene como objetivo encontrar un modelo apropiado para realizar los pronósticos del gasto medio por turista en España, se extrajo los datos de la Encuesta de Gasto Turístico Egatur de enero 2016 a marzo 2020. Para todo el procedimiento se utilizará la metodología de Box Jenkins, y se hizo la división de la serie de enero 2016 a diciembre 2018 (train) y de enero 2019 a marzo 2020 (test). Cabe resaltar que la codificación se realizó con el software RStudio versión 4.0.2. Después de haber seguido todos los pasos de la metodología de Box Jenkins, se identificó que el mejor modelo para predecir la serie de tiempo fue un modelo SARIMA(0,1,1)(1,1,0)<sub>12</sub>, se realizó la validación del modelo, es decir se aplicaron las pruebas de hipótesis para verificar la aleatoriedad, normalidad, independencia y homocedasticidad. En primera instancia, se pronosticó el gasto medio de abril 2019 a marzo 2020, y con ello se hizo la evaluación de los pronósticos (test), en la cual se obtuvo el Error Medio Cuadrático (EMC) de -16.11, el Porcentaje de Error Medio Absoluto (PEMA) de 1.63% y el Porcentaje Medio del Error (PME) de -1.48%. Palabras clave: Series de tiempo, Gasto por Turista, Box Jenkins, SARIMA, autocorrelaciones.



## ABSTRACT

The company is dedicated to providing consultancy on different projects, here we will focus on the TurismEsp project specifically, which is oriented to tourism in Spain, but something that has not yet been implemented is the view of spending per tourist in Spain, and this is of the utmost importance since various promotions can be implemented by inquiring more about it. Therefore, this work aims to find an appropriate model to forecast the average spending per tourist in Spain, data from the Egatur Tourist Expenditure Survey from January 2016 to March 2020 was extracted. For the entire procedure, it will be used the Box Jenkins methodology, and the series was divided from January 2016 to December 2018 (train) and from January 2019 to March 2020 (test). It should be noted that the coding was carried out with the RStudio version 4.0.2 software. After having followed all the steps of the Box Jenkins methodology, it was identified that the best model to predict the time series was a SARIMA(0,1,1)(1,1,0)<sub>12</sub> model, the model was validated, that is, the hypothesis tests were applied to verify randomness, normality, independence and homoscedasticity. In the first instance, the average expenditure from April 2019 to March 2020 was forecast, and with this the evaluation of the forecasts (test) was made, in which the Mean Square Error (MSE) of -16.11 was obtained, the Error Percentage Absolute Mean (PEMA) of 1.63% and the Mean Error Percentage (PME) of -1.48%. Keywords: Time series, Expenditure per Tourist, Box Jenkins, SARIMA, autocorrelations.

## Tabla de contenido

I.	Introducción.....	9
II.	Descripción de la Actividad.....	11
2.1.	Datos de la empresa o institución.....	11
2.1.1.	Nombre de la institución .....	11
2.1.2.	Periodo de duración del TSP .....	11
2.1.3.	Razón Social.....	11
2.1.4.	Dirección .....	11
2.1.5.	Correo electrónico .....	11
2.1.6.	Organigrama de la empresa .....	12
2.2.	Finalidad y Objetivos de la empresa .....	13
2.2.1.	Finalidad.....	13
2.2.2.	Objetivos de la empresa .....	13
2.3.	Descripción de la actividad .....	14
2.3.1.	Organigrama del área .....	14
2.3.2.	Finalidad del presente TSP.....	15
2.3.3.	Objetivos del TSP.....	15
2.3.4.	Problemática.....	15
2.3.5.	Metodología.....	16
III.	Marco teórico .....	17
3.1.	Series de Tiempo.....	17
3.2.	Antecedentes .....	33
3.2.1.	Antecedentes Nacionales.....	33

3.2.2. Antecedentes Internacionales .....	35
IV. Metodología .....	37
4.1. Tipo y diseño de investigación.....	37
4.2. Variables de investigación .....	37
4.3. Análisis Exploratorio.....	37
4.4. Ajuste de Modelos.....	43
4.5. Evaluación de los supuestos de los errores .....	45
V. Conclusiones .....	52
VI. Recomendaciones .....	54
VII. Referencias.....	55
VIII. Anexos.....	57

## Tabla de Figuras

<b>Figura 1</b> Organigrama de la Empresa .....	12
<b>Figura 2</b> Organigrama del área .....	14
<b>Figura 3</b> Modelo Aditivo y Modelo Multiplicativo .....	18
<b>Figura 4</b> Metodología de Box Jenkins .....	23
<b>Figura 5</b> Autocorrelaciones y Autocorrelaciones parciales AR(1) y AR(2).....	27
<b>Figura 6</b> Autocorrelaciones y Autocorrelaciones parciales MA(1) y MA(2).....	28
<b>Figura 7</b> Autocorrelaciones y Autocorrelaciones parciales ARMA(1) y ARMA(2)..	29
<b>Figura 8</b> Gastos medios por turista de enero 2016 – marzo 2020.....	38
<b>Figura 9</b> Diagrama de Cajas Simples.....	38
<b>Figura 10</b> Diagrama de Cajas Agregadas .....	39
<b>Figura 11</b> Descomposición de la serie gasto medio por turista .....	40
<b>Figura 12</b> Autocorrelación Simple y Autocorrelación Parcial.....	40
<b>Figura 13</b> ACF y PACF de la serie diferenciada .....	42
<b>Figura 14</b> ACF y PACF y Ljung-Box Q Test de los errores .....	45
<b>Figura 15</b> Distribución Normal de los Errores.....	46
<b>Figura 16</b> Pronósticos y valores reales del gasto medio por turista en España .....	49
<b>Figura 17</b> Pronósticos del gasto medio por turista para abril 2020 – marzo 2021 .....	50

## Lista de Tablas

<b>Tabla 1</b> Estadísticas Descriptivas del gasto medio por turista .....	37
<b>Tabla 2</b> Prueba de Estacionariedad (Raíz Unitaria) - Dickey Fuller Aumentada .....	41
<b>Tabla 3</b> Prueba de Estacionariedad (Raíz Unitaria) - Dickey Fuller Aumentada .....	42
<b>Tabla 4</b> Análisis de los Errores de Estimación .....	43
<b>Tabla 5</b> Coeficientes del modelo seleccionado SARIMA(0,1,1)(1,1,0) <sub>12</sub> .....	43
<b>Tabla 6</b> Coeficientes del modelo seleccionado SARIMA(2,1,1)(0,1,0) <sub>12</sub> .....	44
<b>Tabla 7</b> Coeficientes del modelo seleccionado SARIMA(1,1,1)(0,1,0) <sub>12</sub> .....	44
<b>Tabla 8</b> Coeficientes del modelo seleccionado SARIMA(1,1,0)(0,1,0) <sub>12</sub> .....	44
<b>Tabla 9</b> Test de Ljung-Box para la prueba de aleatoriedad de los errores.....	46
<b>Tabla 10</b> Test de normalidad de los errores (Shapiro Wilk).....	47
<b>Tabla 11</b> Test de homocedasticidad de los errores (Breush Pagan Test).....	47
<b>Tabla 12</b> Pronósticos (Test), valores reales e Intervalos de Confianza .....	48
<b>Tabla 13</b> Comparación de los errores (Train y Test) .....	48

**Tabla de Anexos**

<b>Anexo 1</b> Dashboard de TurismEsp .....	57
---	----

## I. Introducción

Instalada en Madrid y Barcelona, la empresa española conformada por profesionales con experiencia en el mundo de los sistemas de información, la tecnología relacionada con el Business Intelligence, destaca entre los datos y el manejo de la información una forma competente y sostenible para el crecimiento de sus clientes. Es de aclarar, que dadas las características de complejidades técnicas y del alto valor de la información para los clientes, los sistemas BI requieren de un personal altamente capacitado, con experiencia y dinámico, que garanticen el desarrollo y con ello el éxito de un proyecto.

Esta compañía, posee vasta experiencia con empresas de diversos sectores, dando lugar al entendimiento de las problemáticas de una gran gama de situaciones y áreas, que permite la planeación de soluciones e implementación de herramientas que solucionen las carencias que se encuentran. Actualmente, me encuentro desempeñando labores desde febrero del 2022 en el puesto de Analytics Consultant, destacando en el área de análisis de datos y diseño de dashboards para los proyectos en asignación. En esta área nos asignan diferentes proyectos, en las cuales se maneja desde los análisis de datos, elaboración de dashboards, hasta la aplicación de ambas funciones. Cuando se habla de dashboards, en primera instancia se realiza la gestión de la información y los análisis de datos, con el fin de brindar la mejor visualización de los indicadores clave de desempeño (KPI), sus métricas y los datos fundamentales que posteriormente nos permiten un seguimiento del estado del proyecto de una manera óptima y eficiente.

Para este caso, se optará por un proyecto en específico, en el cual se realizará un análisis estadístico con sus respectivas técnicas, con el objetivo de generar un valor agregado a nuestro caso en estudio. En el proyecto se ha realizado una vista general del turismo en España mediante dashboards en Power BI, cabe resaltar que el dashboard realizado está estructurado en diferentes páginas, que van desde una vista general, viajeros, ocupación,

oferta, ADR y RexPar. En cada una de estas páginas se estableció un filtro que permite la visualización de dichos datos por hoteles, apartamentos o por el total, para ello se emplearon los datos de la Encuesta de Ocupación Hotelera y de Apartamentos Turísticos por el Instituto Nacional de Estadística con últimos datos disponibles en febrero del 2022. Algo que aún no se implementó en el dashboard es el análisis de gasto medio por turista, esto ayudará a tener una vista de cuánto gasta en promedio un turista en España, en qué meses gasta más y con ello ofrecer promociones para los turistas. Para ello se utilizó los datos de la Encuesta de gasto turístico Egatur, con el cual se realizará predicciones de abril 2020 – marzo 2021 mediante series temporales, se tendrá como objetivo encontrar el mejor modelo para predecir el gasto medio por turista.



## II. Descripción de la Actividad

### 2.1. Datos de la empresa o institución

#### 2.1.1. *Nombre de la institución*

Stratebi

#### 2.1.2. *Periodo de duración del TSP*

Del 01 de abril al 30 de junio

#### 2.1.3. *Razón Social*

Stratebi Data Engineering S.A.C

#### 2.1.4. *Dirección*

No aplica por términos de confidencialidad

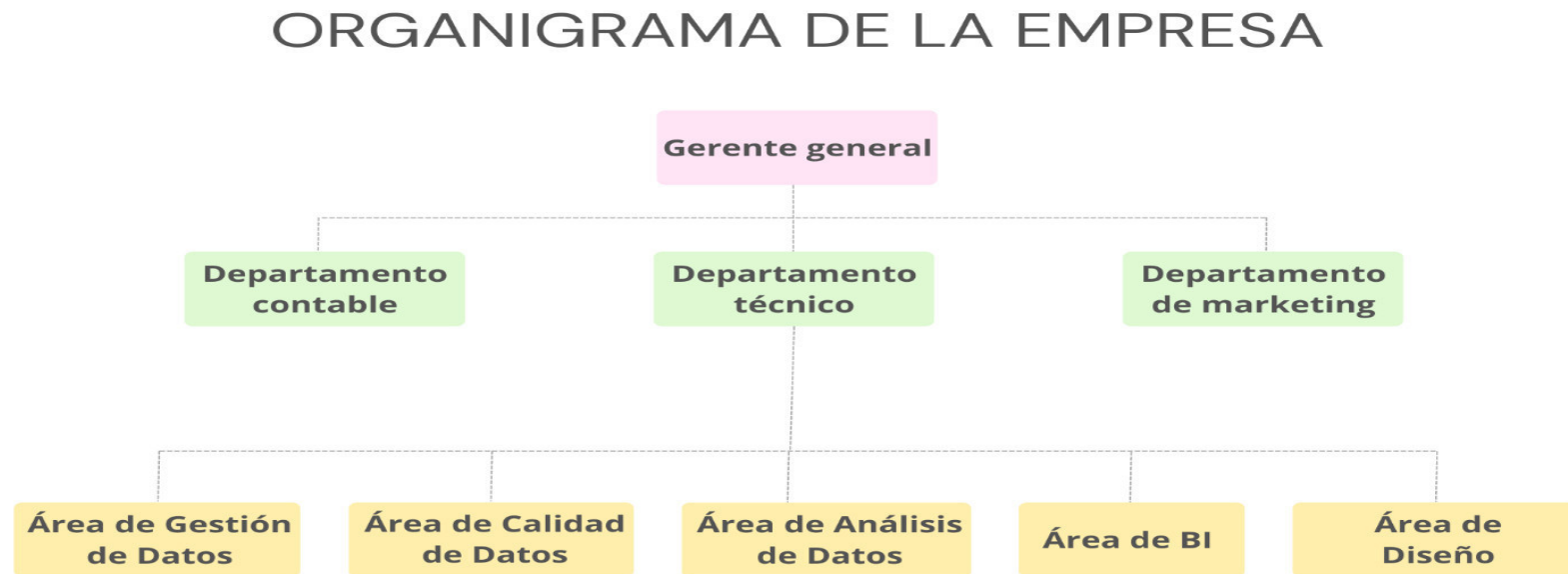
#### 2.1.5. *Correo electrónico*

No aplica por términos de confidencialidad

### 2.1.6. Organigrama de la empresa

**Figura 1**

*Organigrama de la Empresa*



*Nota.* Organigrama de las áreas que se tiene en la empresa.

## **2.2. Finalidad y Objetivos de la empresa**

### **2.2.1. Finalidad**

La finalidad de la empresa está en brindar soluciones de negocio de acuerdo a lo requerido por el cliente, en la cual se desempeñan profesionales de diversas carreras y con una experiencia de amplia gama, participando en proyectos estratégicos en compañías reconocidas mundialmente y dispuestos a brindar una consultoría de calidad a los clientes.

### **2.2.2. Objetivos de la empresa**

#### **a) Objetivo**

Proveer a las empresas de servicios profesionales, confiables y adaptados a las necesidades requeridas, que formen una estrategia de inteligencia de negocios, capaz de rentabilizar la información disponible en la empresa.

#### **b) Misión**

Dar las mejores soluciones a los clientes y con ello ser líderes en Big Data y Business Intelligence.

#### **c) Visión**

Llegar a ser considerados como una parte esencial para la creación de sistemas inteligentes que conlleve a tomar las mejores decisiones a través de nuestras herramientas y servicios profesionales.

### 2.3. Descripción de la actividad

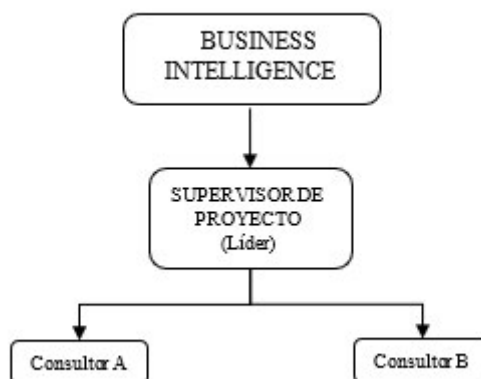
La empresa cuenta con el área de BI (Business Intelligence), cuyo objetivo es:

- Lenguaje dax para la creación de KPI's
- Elaboración de reportes
- Creación de visualizaciones con R
- Creación de visualizaciones con Python
- Medidas rápidas de agrupación /forecasting
- Unión de fuentes de datos: SAP HANA, MySQL, Teradata, IBM DB2, Dynamics Navision, CRM, SQL. Archivos de Excel, .CSV, JASON o descargar datos procedentes de servicios en línea como Facebook, Google Analytics,
- Automatización de reportes
- Revisar y validad los datos

#### 2.3.1. Organigrama del área

**Figura 2**

*Organigrama del área*



### **2.3.2. Finalidad del presente TSP**

La finalidad del presente trabajo es de pronosticar el gasto medio turístico de los visitantes extranjeros a España para facilitar una vista estratégica y complementar la visualización de datos realizada en Power BI con RStudio para el proyecto TurismEsp.

### **2.3.3. Objetivos del TSP**

#### **a) Objetivo general**

Pronosticar el gasto medio turístico de los visitantes extranjeros a España con el fin de evaluar los meses con más gasto.

#### **b) Objetivos específicos**

- Analizar de forma exploratoria los gastos medios turísticos de los visitantes a España.
- Identificar el mejor modelo para predecir los gastos medios de abril 2020 a marzo 2021.
- Verificar las pérdidas que hubo de abril 2020 – marzo 2021 por motivo de la pandemia

### **2.3.4. Problemática**

La empresa actualmente ha elaborado un vista de datos para el proyecto TurismEsp en el programa Power BI, en la cual se observan diferentes indicadores de actividad turística, pero algo que aún no se ha implementado es la elaboración de proyecciones, por lo que tener una vista del gasto promedio de turistas en España, facilitará a tener una visualización general de cómo evoluciona a medida que el tiempo pasa, tanto mensualmente como anualmente, y así ofrecer promociones según el flujo de turistas en un determinado mes, es decir proponer estrategias de marketing.

### 2.3.5. Metodología

Para el análisis de datos, se usará como herramienta estadística de apoyo el software

RStudio:

- Análisis exploratorio de los datos
- Analizar componentes de una serie
- Evaluación de las autocorrelaciones simples y parciales
- Definir el modelo
- Elección de modelo
- Verificación de supuestos de los errores
- Pronósticos para el periodo que va de abril 2020 a marzo 2021
- Interpretación

### III. Marco teórico

#### 3.1. Series de Tiempo

Según Palma, (2016)  $y_t$  denota la serie de tiempo y  $t$  denota el momento en que se tomó la observación. Por lo general,  $t \in Z$ , donde  $Z = \{ \dots, -2, -1, 0, 1, 2, \dots \}$  representa el conjunto de valores enteros positivos y negativos. En la práctica, sin embargo, solo se tiene una cantidad finita de datos. En tales casos, podemos escribir  $\{y_1, y_2, \dots, y_n\}$  para denotar la serie temporal. Una serie temporal  $\{y_t\}$  corresponde a un proceso estocástico que se compone de variables aleatorias observadas en el del tiempo.

##### 3.1.1. *Descomposición de la Serie*

Según Bee Dagum (2010), Un objetivo fundamental del estudio de series de tiempo es descomponer una serie en un conjunto de componentes no observables (latentes) asociados con diversos tipos de cambios en el tiempo. Una serie temporal consta de 4 tipos de fluctuaciones:

- 1) Tendencia secular o una tendencia a largo plazo.
- 2) Movimientos cíclicos superpuestos a la tendencia a largo plazo, es decir ascensos y caídas de la línea de tendencia.
- 3) Un desplazamiento estacional que se da en cada año, cuya forma es dependiente de la naturaleza de la serie.
- 4) Variaciones residuales debidas a cambios que afectan variables independientes u otros eventos relevantes como guerras y catástrofes nacionales que afectan diversas variables.

Generalmente, se supone que estas cuatro componentes son independientes entre sí y se expresan mediante el modelo de descomposición aditiva:

$$Y_t = T_t + C_t + S_t + I_t \quad (1)$$

donde  $Y_t$  es la serie observada,  $T_t$  es la componente tendencia,  $C_t$  es la componente

cíclica,  $S_t$  es la componente estacional y  $I_t$  es la componente irregular.

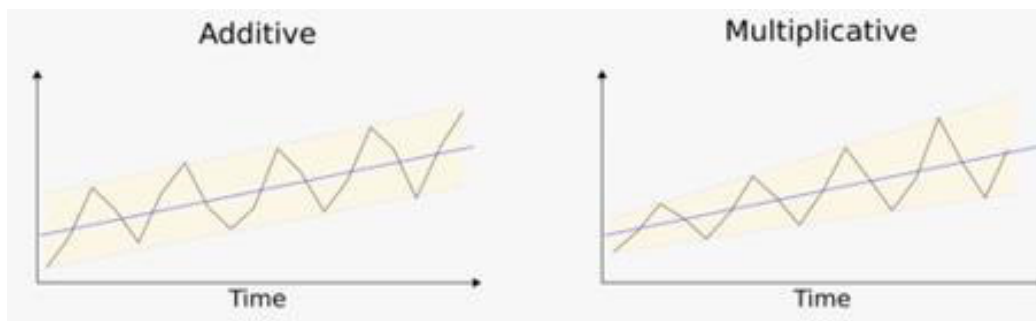
Si existe dependencia entre los componentes existentes, esta relación se representa mediante un modelo multiplicativo

$$Y_t = T_t \times C_t \times S_t \times I_t \quad (2)$$

En la cual  $S_t$  y  $I_t$  son expresados en proporción a la tendencia-ciclo  $C_t T_t$ . También se utilizan modelos mixtos aditivo-multiplicativo en algunos casos.

### Figura 3

*Modelo Aditivo y Modelo Multiplicativo*



*Nota.* Tipos de modelo. Elaboración propia

#### 3.1.2. Modelos de Series de Tiempo

##### 3.1.2.1. Modelos lineales estacionarios

Según Bowerman, B., Koehler, A., & O'Connell (2007), el modelo predice valores futuros de una serie temporal, su base radica en los datos del pasado de la misma serie temporal. Es decir, se debe analizar los valores pasados para identificar patrones en los datos cuando se utilice un modelo univariado. Entonces, suponiendo que esto continúe en el futuro, el patrón encontrado se generaliza para propósitos de pronósticos.

Otro autor afirma lo siguiente:

Una especificación relativamente fácil, pero efectiva, producida al representar una variable en función de su pasado se denomina autorregresiva y forma parte de la metodología de Box-Jenkins. Los modelos autorregresivos  $AR(p)$  y los



modelos de promedio móvil MA(q) son las dos especificaciones primordiales, diseñados para procesos estocásticos estacionarios univariados. (Álvaro, 2010, p.50)

### ***Modelo Autorregresivo (AR(p))***

Este modelo muestra el valor presente de una serie que es estacionaria ( $y_t$ ) en función a datos pasados, esto es de sus rezagos  $y_{t-1}, \dots, y_{t-p}$ . Se representa de la siguiente forma:

$$y_t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} = \varepsilon_t + c \quad (3)$$

En la cual  $p$  representa el orden o el mayor rezago del modelo,  $c$  es una constante y  $\varepsilon_t$  presenta el ruido blanco con varianza constante  $\sigma^2_\varepsilon$  y media igual a cero. Cuando se habla de series temporales, el  $\{\varepsilon_t\}$  es conocido como ruido, residuos, error, shocks o también como innovaciones. Otra manera de expresar la ecuación (3) es despejando  $y_t$ .

### ***Modelo de Promedio Móvil (MA(q))***

Es el segundo modelo en la metodología de Box Jenkins. En donde,  $y_t$  está dada en función del presente, y también el pasado de una serie de ruido blanco  $\{\varepsilon_t\}$  con media igual a cero y la varianza es finita:

$$y_t = c + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} = c + \sum_{s=0}^q \beta_s \varepsilon_{t-s} \quad (4)$$

En la cual, los coeficientes  $\beta$  no muestran limitaciones respecto a la estabilidad de la serie. Esto está dado, ya que una suma la cual es finita, que está compuesta de  $q$  términos y que tiene varianza finita, tal como los  $\varepsilon_t$ , no es capaz de ser explosiva. En (4) usualmente se supone que  $\beta_0 = 1$ . Se observa que si  $c = 0$  entonces  $Ey_t = 0$  porque  $E\varepsilon_t = 0$ , de otra forma  $Ey_t = c$ . La varianza de  $y_t$  será  $\sigma_\varepsilon^2 \sum_{s=0}^q \beta_s^2$ .

En series de tiempo y cuando se habla de los modelos de promedio móvil, también se conoce como innovaciones a  $\{\varepsilon_t\}$  ya que se entiende como información nueva que llega al sistema.

### ***Modelo Autorregresivo de Promedio Móvil (ARMA (p, q))***

Luego de haber presentado los modelos anteriores, se sabe que cualquier proceso débilmente estacionario, con media cero, a pesar de ser generado de manera no lineal o partiendo de errores autocorrelacionados, finalmente tendrá una representación a Wold. De forma que tenemos la posibilidad de usar como aproximaciones los modelos  $AR(p)$  y  $MA(q)$ . Además, tenemos la posibilidad de usar la combinación de los dos modelos y obtener el  $ARMA(p, q)$ , y con ese modelo comúnmente resulta dar mayor parsimonia, por consiguiente, se puede utilizar una cantidad menor de términos en comparación con el AR o el MA independientemente

El modelo  $ARMA(p, q)$  será dada de la siguiente forma:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (5)$$

También se puede expresar como:

$$y_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j} \quad (6)$$

En la cual, comúnmente  $\beta_0 = 1$ . Se puede añadir una constante si se observa que es necesario. El valor esperado de  $y_t$  se obtiene, como en los dos modelos anteriores, sin olvidar que la serie es estacionaria; si no existe constante este será igual a a cero y  $\varepsilon_t$  con media igual a cero. Por otro lado, con la solución de las ecuaciones de autocorrelación se obtiene la varianza. Se tiene la siguiente expresión en función del operador de rezago:

$$\alpha(B)y_t = \beta(B)\varepsilon_t$$

En la cual  $\alpha(B)$  y  $\beta(B)$  son polinomios en  $B$ .

Se debe tomar en cuenta que  $p$  y  $q$  hacen referencia al rezago último y no específicamente a la cantidad de rezagos de la parte promedio móvil o de la parte autorregresiva, esto se da porque varios rezagos tienen la posibilidad de ser no significativos.

Hay que tener en cuenta que la estacionariedad del  $ARMA$  no depende de la parte de

promedio móvil, pero sí es dependiente únicamente de la parte autorregresiva; esto conlleva a que las raíces que se obtiene de la ecuación auxiliar de la parte autorregresiva estén por fuera del círculo unitario.

Es más complejo establecer la ACF para un ARMA en comparación con los modelos anteriores. En general, se tendrá inicialmente unas autocorrelaciones significativas establecidas por la parte promedio móvil y seguidas, desde  $\tau > q$ , por un comportamiento parecido al de la parte autorregresiva.

### 3.1.2.2. Modelos No Estacionarios

#### *Modelos Autorregresivos de Promedio Móvil Integrado (ARIMA (p,d,q))*

El término ARIMA refiere a Modelos Autorregresivos Integrados de Medias Móviles (Mahía & De Arce, 2003).

Según De la Fuente Fernandez (2018), un modelo  $ARIMA(0,d,0)$  después de ser diferenciada  $d$  veces, se convierte en ruido blanco (el cual es un proceso netamente aleatorio). (p. 7)

El modelo  $(0,d,0)$  está expresado como:  $(1 - B)^d y_t = \varepsilon_t$

El modelo  $ARIMA(p,d,q)$ , el cual es nombrado como proceso autorregresivo integrado de medias móviles con orden  $p,d,q$ , se expresa como:

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)(1 - B)^d x_t = (1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q) \varepsilon_t \quad (7)$$

El modelo  $ARIMA(p,d,q)$  es la cual muestra una serie temporal luego de haber aplicado  $d$  diferencias, para así eliminar la no estacionariedad. Esto es factible aplicar también a otros modelos. En caso alguna componente  $p,d,q$ , resulte ser igual a cero, ésta se borra de la fórmula general. Las oscilaciones cíclicas o variaciones estacionales son lo que caracteriza a los modelos cíclicos o estacionales.

Los modelos  $ARIMA(p,d,q)(P,D,Q)$  pueden referir a series que tienen tendencia secular y variaciones en ciclos. Cabe recalcar que  $(p,d,q)$  se refiere a la parte regular de la serie y

$(P,D,Q)$  a la parte estacional, o cíclica de la serie temporal.

Por consiguiente, presentamos como ejemplo los cuatro modelos:

- $ARIMA(0,1,1)(0,1,1)_{12}$  :

$$(1 - B)(1 - B^{12})y_t = (1 - \beta_1 B^{12})(1 - \delta_{12} B^{12})$$

- $ARIMA(0, 1, 1)(0,0, 1)_{12}$  :

$$(1 - B)y_t = (1 - \beta_1 B^{12})(1 - \delta_{12} B^{12})$$

- $ARIMA(2, 1, 0)(1, 0, 0)_{12}$  :

$$(1 - \alpha_1 B - \alpha_2 B^{12})(1 - \Omega_1 B^{12})(1 - B) y_t = \varepsilon_t$$

- $ARIMA(1, 1, 1)(2, 1, 1)_{12}$  :

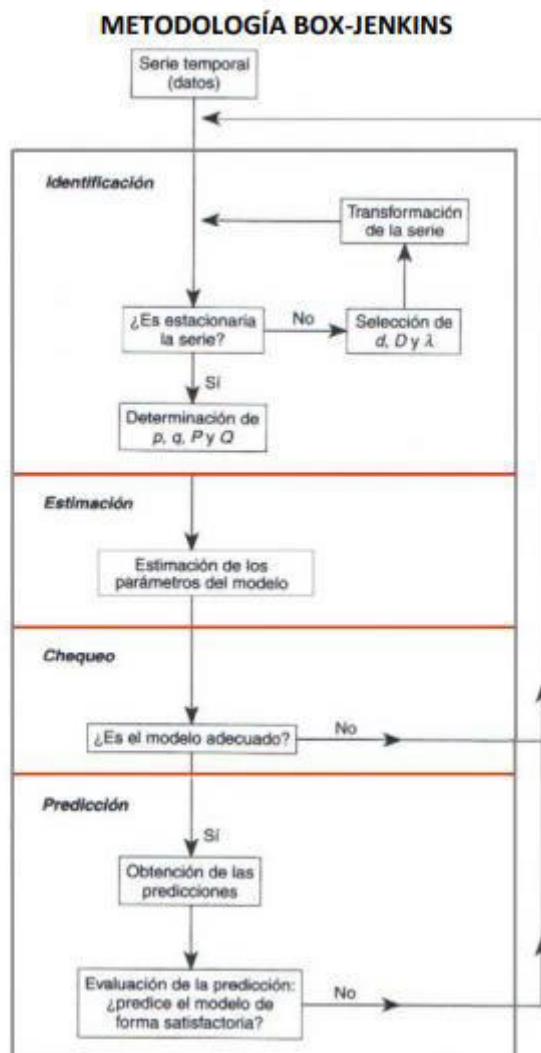
$$(1 - \alpha_1 B)(1 - \Omega_1 B^{12} - \Omega_2 B^{24})(1 - B^{12})(1 - B) y_t = (1 - \beta_1 B)(1 - \delta_{12} B^{12})\varepsilon_t$$

### 3.1.3. Metodología de Box – Jenkins

Esta metodología para realizar predicciones difiere de otros métodos, esto acontece porque los datos históricos de la serie que se realiza las predicciones no siguen ningún patrón en particular. Identifica posibles modelos de una categoría de modelo general basada en un enfoque iterativo. Después, el modelo que se ha escogido se compara con datos históricos para observar si describe exactamente a la serie. Se espera que los residuos sean pequeños, aleatorios y sin información relevante para que el modelo se ajuste bien. En caso el modelo elegido, no sea el más adecuado, el proceso se repite con modelo nuevo y por consiguiente, elegir un modelo más óptimo, y cuando se encuentre dicho modelo, se considera útil para realizar las predicciones. (E. Hanke & Dean, 2010)

Figura 4

Metodología de Box Jenkins



*Nota.* Adaptado de *Metodología de Box Jenkins* (p. 5), por S. De la Fuente, 2018, Metodología de Box – Jenkins.

Ahora se explicará la metodología de Box – Jenkins más detalladamente:

### 1. *Identificación*

El primer paso para identificar el modelo es establecer si la serie temporal es estacionaria, es decir cuando la serie es constante en un periodo determinado de tiempo. Para ello, se puede realizar la gráfica de los valores reales acompañado de la gráfica de

autocorrelación simple y parciales:

- Si observamos una tendencia positiva o negativa de la serie en el tiempo y las autocorrelaciones no decrecen rápidamente; entonces esto indica que la serie no es estacionaria.
- Si se da que la serie no es estacionaria, es posible hacer diferencias a la serie y así convertirla en estacionaria. Luego se halla un modelo ARIMA para la serie que ha sido diferenciada.
- La diferenciación se realiza hasta que al graficar la serie muestre que sea constante en el tiempo y que las autocorrelaciones decrezcan con rapidez.  $d$  denota la cantidad de diferencias requeridas para conseguir la estacionariedad. Los modelos no estacionarios están dados por ARIMA  $(p, d, q)$  y son nombrados modelos de promedio móvil integrados autorregresivos. Es importante mencionar que:  $p$  es el orden de la parte autorregresiva,  $d$  es la cantidad de diferencias, y  $q$  es el orden de la parte de promedio móvil.
- $d = 0$  cuando la serie temporal es estacionaria, es decir no necesita de ninguna diferencia, por lo que los modelos ARIMA ahora serán modelos ARMA. Ahora se utilizará la notación ARIMA  $(p, d, q)$  para los modelos de series que son estacionarias; es decir  $d = 0$ , así como también no estacionarias; es decir  $d > 0$ . Se sabe que un modelo ARIMA es cuando se hace diferencias a la serie.
- Existen pruebas estadísticas para ver si la serie es estacionaria:

**Prueba de Dickey – Fuller:**

$H_0$ : La serie tiene raíz unitaria (la serie no es estacionaria)

$H_1$ : La serie no tiene raíz unitaria (la serie es estacionaria)

$$\hat{t} = \frac{\hat{\rho} - 1}{S / \sqrt{\sum_{t=1}^T Y_{t-1}^2}}$$

En la cual:

$\rho$  : Es el coeficiente obtenido de la regresión  $Y_t$  sobre  $Y_{t-1}$

$S$  : Es el error estándar de la regresión

$Y_{t-1}$  : Es la variable desfasada

Si el valor resultante de  $t$  es mayor que los valores absolutos críticos de Dickey Fuller, entonces la serie es estacionaria. Por tanto, si el valor resultante es menor que el valor crítico, entonces la serie no es estacionaria y se aplicaría diferencias a la serie.

### **Prueba de Dickey – Fuller aumentado (ADF):**

Esta prueba es una versión extendida de la prueba de Dickey Fuller. Debido al término de error, no es muy probable que sea ruido blanco, por ello para remover el problema de la autocorrelación se extendió la prueba, por tanto, se incorporó rezagos adicionales en función de las variables que son dependientes. Comúnmente utilizamos la prueba de Dickey-Fuller aumentada, por las razones mencionadas anteriormente, y por último cabe resaltar que las hipótesis son las mismas que la prueba de Dickey Fuller simple. (Mushtaq, 2011)

Al obtener una serie que es estacionaria, se identificará el modelo a usar:

- Es posible que exista alguna dificultad al identificar un modelo ARIMA al observar las autocorrelaciones simples y parciales, ya que no es del todo claro. Por lo que, el modelo inicial solo se debe considerar como un modelo a primera vista, más no como el definitivo. Se debe realizar un análisis en los pasos 2 y 3 para ver si el modelo es el que más se adecúa, y en caso de no ser el mejor modelo, se debe encontrar un modelo alternativo.
- No olvidar que, cuando las autocorrelaciones simples se desvanecen exponencialmente hacia cero y las parciales se cortan de manera abrupta, el modelo requiere de términos autorregresivos (AR).

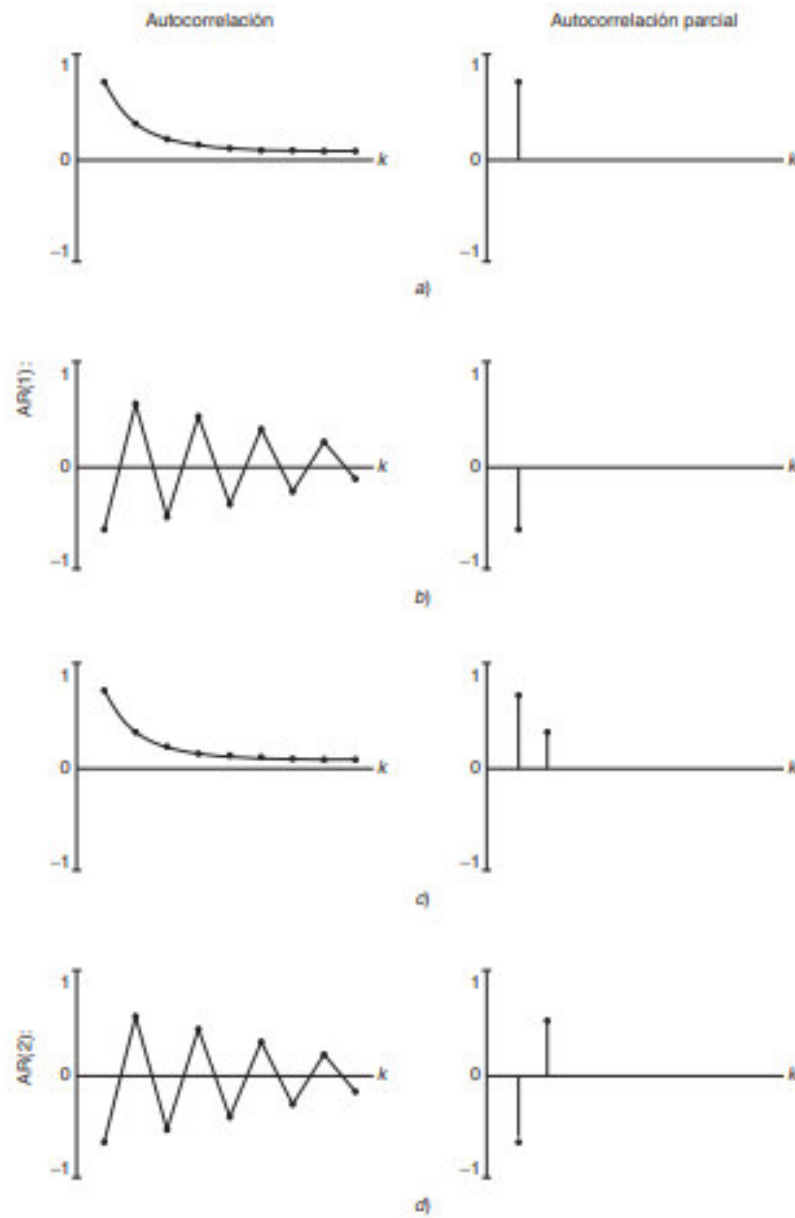
- Si las autocorrelaciones simples terminan de manera abrupta y las parciales se desvanecen lentamente, el modelo requiere términos de promedio móvil (MA)
- Si las autocorrelaciones simples y las parciales se desvanecen lentamente ambas, entonces se requieren términos tanto autorregresivas como de promedio móvil (ARMA)

A continuación, se muestra las autocorrelaciones simples y parciales:



**Figura 5**

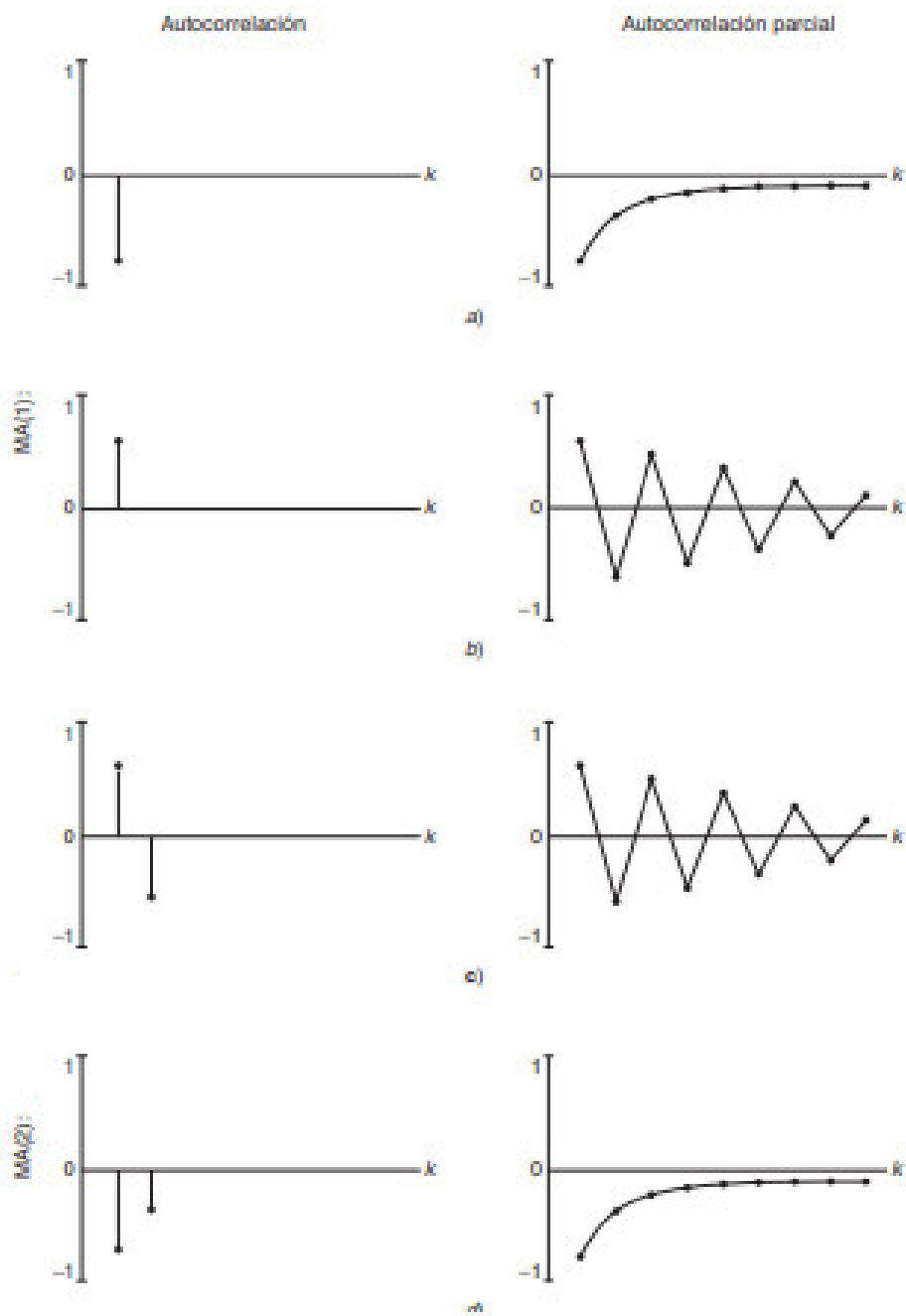
*Autocorrelaciones y Autocorrelaciones parciales AR(1) y AR(2)*



*Nota.* Adaptado de *Autocorrelaciones y Autocorrelaciones parciales AR(1) y AR(2)* (p. 401), por E. Hanke & Dean, 2010, *Pronósticos de los negocios*

**Figura 6**

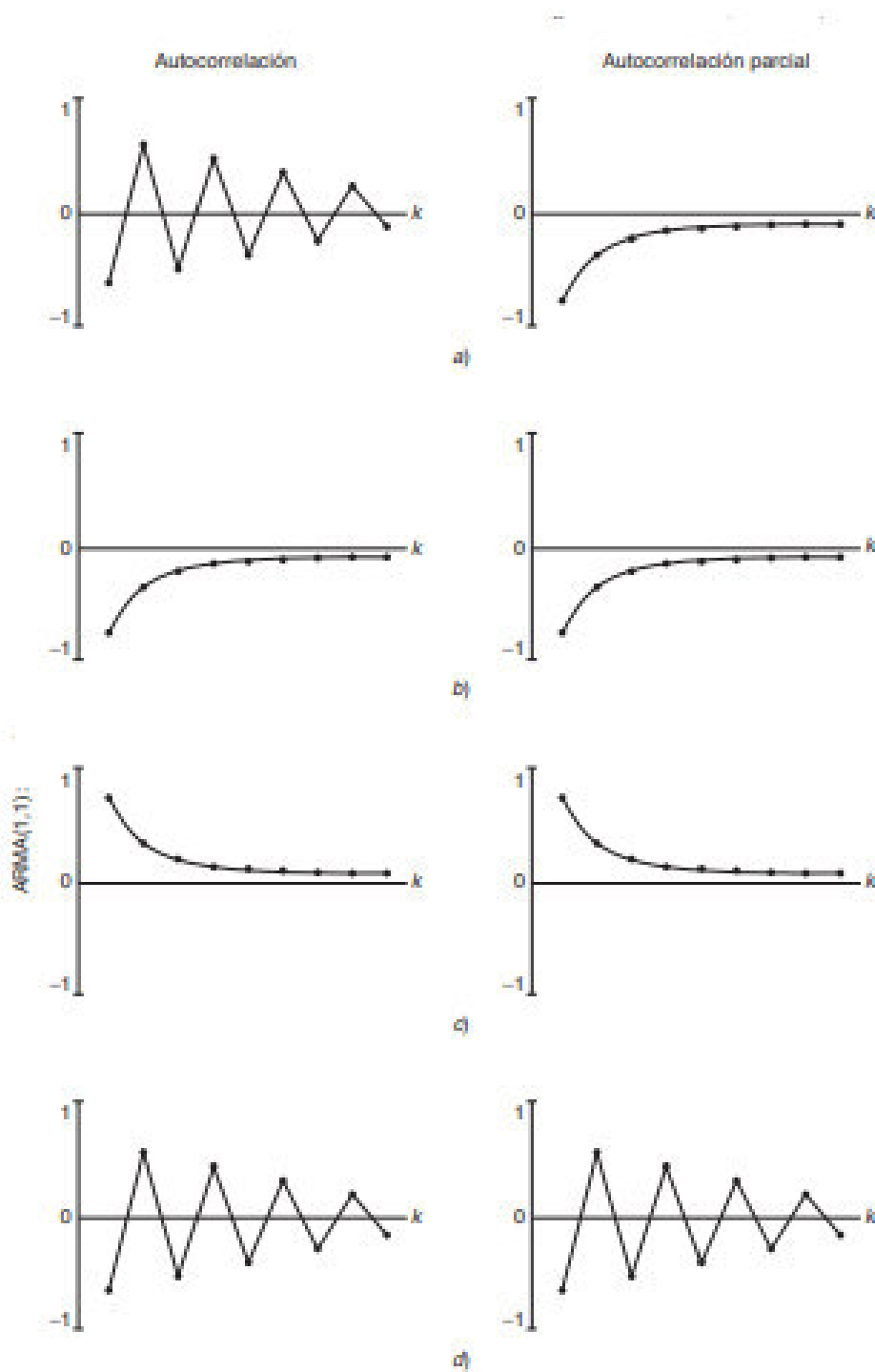
*Autocorrelaciones y Autocorrelaciones parciales MA(1) y MA(2)*



*Nota.* Adaptado de *Autocorrelaciones y Autocorrelaciones parciales MA(1) y MA(2)* (p. 402), por E. Hanke & Dean, 2010, *Pronósticos de los negocios*

**Figura 7**

*Autocorrelaciones y Autocorrelaciones parciales ARMA(1) y ARMA(2)*



*Nota.* Adaptado de *Autocorrelaciones y Autocorrelaciones parciales ARMA(1) y ARMA(2)* (p. 403), por E. Hanke & Dean, 2010, *Pronósticos de los negocios*

## 2. Estimación

Después de haber seleccionado un modelo (cabe resaltar que es un modelo tentativo), lo que prosigue es estimar los parámetros del modelo, estos se estiman al minimizar la suma de cuadrados de los errores de ajuste (para los modelos ARIMA). Mediante un procedimiento no lineal de mínimos cuadrados se obtienen los estimados de los mínimos cuadrados. Este es un proceso no lineal de mínimos cuadrados, el cual consiste en un método que busca el mínimo de la suma de la función de errores cuadrados. Luego de establecer los estimados por mínimos cuadrados y hallar los errores estándar, aquí ya los valores  $t$  pueden construirse y, por lo tanto, ser interpretados. Los parámetros que son distintos de cero, es decir son significativos, se incluyen en el modelo ajustado; pero los parámetros estimados que no son significativos se excluyen del modelo.

Luego de hacer la estimación, usaremos el criterio de AIC (Criterio de Información de Akaike), se espera que sea el mínimo posible.

### **Criterio de Akaike:**

Dado un modelo, el cual tiene  $l$  parámetros y se ha sido estimado con  $n$  observaciones, el AIC se expresa de la siguiente forma. (Mauricio, 2007)

$$AIC = \frac{2L^*}{n} + \frac{2l}{n}$$

En la cual  $L^*$  será el logaritmo neperiano de la función de verosimilitud y está relacionada con el modelo que se ha sido estimado, y a la vez evaluada en las estimaciones que se obtienen de los parámetros del modelo.

## 3. Validación

Box y Jenkins propusieron diversas pruebas para confirmar si el modelo escogido se ajusta adecuadamente a los datos de la serie. Entre ellos, se destacó la sobre parametrización, el cual consta de ajustar un modelo de orden preeminente al escogido, luego evaluar si los

parámetros son significativos, es decir diferentes de cero.

- Por otra parte, cuando el modelo se aproxima adecuadamente a la serie temporal, los residuos son ruido blanco, esto se puede corroborar observando los coeficientes de autocorrelación simple y parciales de los residuos. En la cual, estos coeficientes deben de ser nulas en todo su recorrido, salvo en el cero.
- En el caso que el modelo no se ajuste a la serie temporal, entonces los residuos tendrían el comportamiento de un ruido autocorrelacionado. Entonces, para verificar lo mencionado, se debe de utilizar pruebas estadísticas tal como la de Durbin-Watson (para la autocorrelación de primer orden) o la prueba de Wallis (para la de cuarto orden).

Existen otras pruebas para los residuos, estos se encargan de verificar si los residuos que se han obtenido cumplen el supuesto de ruido blanco (aleatorios):

A continuación, se presenta el estadístico propuesto por Box y Pierce:

$$Q = \sum_{k=1}^m r_k^2$$

Suponiendo que  $m$  es suficientemente grande, y sabiendo que los autores demostraron que el estadístico  $Q$  sigue una distribución Chi-Cuadrado con  $(m-p-q)$  grados de libertad(gl). Por lo tanto, si  $Q$  tiene valores muy altos, se concluye que los residuos no son ruido blanco. Otra manera para ver si es ruido blanco o no, es de la siguiente forma, se puede hallar la región crítica a un nivel de significancia  $\alpha$ , con el cual se calcula un valor  $I$  que cumpla  $P(Q > I) = \alpha$ . Cuando el valor de  $Q$  cae dentro de la región crítica, entonces los residuos no son un ruido blanco (rechazamos la hipótesis nula ( $H_0$ )). Por otra parte, si cae fuera de la región crítica, entonces los residuos son un ruido blanco (no se rechaza la hipótesis nula( $H_0$ )). A pesar de que  $m$  es arbitrario, es favorable tomar el valor más alto posible.

Ljung y Box proponen un estadístico alternativo para valores de  $m$  no muy grandes:

$$Q' = \frac{n(n+2) \sum_{k=1}^m r_k^2}{(n-k)} \approx X_{m-p-q}^2 \quad (8)$$

Si  $p$  que está asociado con el estadístico  $Q$  es pequeño ( $p < 0.05$ ), el modelo no se considera adecuado. Por lo que, se debe de buscar un modelo nuevo o un modelo modificado y continuar hasta encontrar un modelo óptimo.

#### 4. *Predicción*

El análisis de los errores es la parte fundamental cuando se elabora modelos predictivos (pronósticos) que nos permitirá diferenciar y elegir dentro de un conjunto de modelos. Este análisis está dividido en dos clases de errores: la primera consta de los errores de dispersión absolutos que son el error cuadrático medio (MSE) y la segunda es el error absoluto medio (MAE). La segunda clase representa a los errores relativos porcentuales. Estos, utilizan la dispersión absoluta del error como el MSE o MAE que se divide por la media de los datos reales para producir los valores normalizados. (Boland, 2011)

A continuación, se muestran algunos tipos de errores para evaluar el modelo al hacer pronósticos. (Montenegro García, 2010)

##### **Error cuadrado medio (MSE o Mean Square Error)**

Se expresa en unidades al cuadrado de la variable que se va a realizar la predicción:

$$MSE = E(y_{t+T} - \hat{y}_{t+T})^2 \quad (8)$$

##### **Raíz del error cuadrado medio (RMSE o Root Mean Square Error)**

Esto se obtiene al sacar la raíz cuadrada al  $MSE$ , está expresado en unidades de la variable que se va a realizar la predicción.

##### **Error medio absoluto (MAE o Mean Absolute Error)**

Está expresado en unidades de la variable que se va a realizar la predicción:

$$MAE = E|y_{t+T} - \hat{y}_{t+T}| \quad (9)$$

Cabe resaltar que, si se presenta una disminución o aumento en el tiempo en los índices presentados anteriormente, esto no quiere decir que las predicciones hayan empeorado o

mejorado. Más que ello, podría referirse a una disminución o aumento en la cantidad de la variable en el tiempo ( $y_t$ ).

### **Error medio porcentual absoluto (MAPE o Mean Absolute Percentage Error)**

Está expresado en porcentaje y es adimensional. Cuando la variable muestra valores cercanos a cero, se dice que el indicador no tiene mucho sentido.

$$MAPE = E \left| \frac{y_{t+T} - \hat{y}_{t+T}}{Y_{t+T}} \right| \quad (10)$$

## **3.2. Antecedentes**

### **3.2.1. Antecedentes Nacionales**

En el artículo elaborado por Divino & McAleer (2017) titulada “Modelling the Growth and Volatility in Daily International Mass Tourism to Peru” con el objetivo de modelar adecuadamente las tasas de crecimiento y la volatilidad (o la variabilidad en la tasa de crecimiento) en llegadas de turistas internacionales al Perú de 1997 a 2007. Los resultados empíricos muestran que las llegadas de turistas internacionales y sus tasas de crecimiento son estacionarias, y que los modelos de volatilidad condicional simétrica y asimétrica estimada, específicamente los modelos GARCH, GJR y EGARCH ampliamente utilizados, se ajustan extremadamente bien a los datos. Además, las estimaciones se asemejan a las que surgen de los datos de series temporales financieras, con persistencia de perturbaciones en la tasa de crecimiento de las llegadas de turistas internacionales tanto a corto como a largo plazo.

En el artículo elaborado por Barreto (2017) titulada “Estudio de la estacionariedad de la llegada mensual de turistas internacionales al santuario histórico de Machu Picchu en el periodo de enero de 2005 a diciembre de 2014” con el objetivo de determinar la estacionariedad de la serie de tiempo de la llegada mensual de turistas internacionales al Santuario Histórico de Machu Picchu en el periodo enero de 2005 a diciembre de 2014. Para ello, se utilizó o la prueba

de raíz unitaria de Dickey-Fuller aumentada con la finalidad de que posteriormente se use correctamente en la propuesta de un modelo de serie de tiempo univariante, tales como: AR, MA, ARIMA, SARIMA. Se obtuvo que la serie de tiempo en estudio es estacionaria transformando la serie original al logaritmo en base 10 y realizando las primeras diferencias regulares y estacionales, la cual servirá para evitar la obtención de regresiones espurias y construir un modelo de series de tiempo adecuado que permita realizar pronósticos.

En la tesis elaborada por Laurente Blanco (2021) titulada “Demanda de turismo internacional en tiempos de covid-19 en la región de Puno-Perú” con el objetivo de conocer el comportamiento de la demanda de turismo internacional en la región de Puno y proyectarla a futuro haciendo uso de modelos de series de tiempo con información mensual del período 2003 a 2019. Para ello, mediante la metodología de Box Jenkins se propusieron modelos ARIMA para la demanda de turismo internacional en la región de Puno-Perú, y para la selección de modelos eficientes se hace uso del Criterio de Información de Akaike (AIC) y el Criterio de Schwarz (SBIC). Después de la búsqueda de modelos de ajuste univariados *ARIMA*, y utilizando los menores valores AIC y SC, se seleccionó el modelo *SARIMA*(6, 1, 24)(1, 0, 1)<sub>12</sub> como el modelo de mayor ajuste dentro de un grupo de modelos para el arribo de turistas internacionales a Puno. Por otra parte, para demostrar la eficiencia del modelo seleccionado por AIC y SBIC, se hace uso de las pruebas estadísticas MAPE o medida del error, cuyo nivel de pronóstico a la información es “altamente preciso”, pues presenta un valor de MAPE igual a 0.85% el mínimo dentro del grupo de modelos estimados.

En el artículo elaborado por Tudela et al. (2022) titulada “Impacto del COVID-19 en la demanda de turismo internacional del Perú. Una aplicación de la metodología Box-Jenkins” con el objetivo de estimar y proyectar la demanda de turismo internacional en el Perú con datos



de serie de tiempo de frecuencia mensual comprendida entre enero 2003 a diciembre 2020. Para ello se analizó, a través de un modelo SARIMA o ARIMA estacional después de seguir la metodología de Box Jenkins, cabe recalcar que en el paso de la identificación del modelo SARIMA se realizó una suavización mediante la técnica de media móvil ponderada exponencial doble (DEWMA). Los resultados muestran que el modelo ARIMA estacional o SARIMA(1,1,1)(0,1,1)<sub>12</sub> resultó apropiado para la proyección, usando los criterios de Akaike (AIC) y Schwarz (SC). Por último, cabe resaltar que el modelo estima una recuperación cíclica parsimoniosa de la llegada de turistas internacionales al Perú.

### **3.2.2. Antecedentes Internacionales**

En la tesis elaborada por Boj (2017) titulada “Panorama del turismo español en el periodo 2004-2016. Análisis de la estacionalidad del gasto del turismo extranjero” con el objetivo de averiguar el impacto del turismo en España haciendo el análisis del gasto turístico extranjero del 2004-2016. Para ello primero se analizó la estacionalidad utilizando los contrastes de Kendall y Kruskal-Wallis, luego de ello, los modelos que se propusieron fueron un modelo autorregresivo ARIMA y un Modelo General Lineal con estimadores MCO, con el fin de comparar ambos. Los resultados obtenidos muestran que la serie del gasto turístico extranjero es estacionaria y estacional, usando un modelo ARIMA, las predicciones se ajustaron mucho a los datos reales, sin embargo, al hacer las predicciones con el modelo MCO se notó en algún mes que existe una ruptura estructural.

En la tesis elaborada por Lindes Castro (2017) titulada “Análisis del gasto por turista en Andalucía mediante el uso de series temporales” con el objetivo de conocer la evolución del gasto por turista mediante la Encuesta del Gasto por Turista. Para ello, se realizó un estudio de series de tiempo mediante la metodología Box y Jenkins, con el objetivo de realizar pronósticos de la serie, para realizar ello, se utilizó el software libre Gretl. Se propuso un modelo

ARIMA(1,1,0)(0,1,1) del gasto, con el cual se observó un buen ajuste y se utilizó ese modelo para realizar la predicción. Por último, cabe resaltar que el componente estacional que caracteriza la serie se repetirá en observaciones a futuro acompañado del crecimiento que viene dado por la tendencia.

En la tesis elaborada por Torres Diaz (2020) titulada “Estimación de la Demanda Turística en Puerto Rico: 1998 a 2018”, se tiene como objetivo modelar la demanda turística de los estadounidenses que visitan a Puerto Rico en el periodo 1998 a 2018. En la cual, las variables dependientes serán: El gasto turístico y el número de turistas. Para ello, se realizaron tres modelos, el de MCO, ARIMA y el VAR. Como resultados más notables, el número de turistas es afectado por el gasto en promociones y mercadeo mientras que el gasto turístico es influenciado por el ingreso per cápita de los Estados Unidos. La predicción más precisa fue generada por el modelo ARIMA con un error de 5% a 6 % según MAPE y RMSPE.

En el artículo elaborado por Turtureanu et al. (2022) titulada “A Projection Approach of Tourist Circulation under Conditions of Uncertainty”, se tiene como objetivo proyectar la demanda turística y ver los efectos que ha tenido el turismo por el COVID-19. Se sabe que es de suma importancia para el turismo y para el sector de hostelería, la previsión del flujo turístico relacionada con las llegadas, una previsión precisa de los flujos turísticos estacionales ayuda a realizar estrategias y con ello aumentar la eficiencia. Para ello, se utilizó la metodología de Box Jenkins, y con ello se determinó el mejor modelo de predicción siguiendo los pasos: identificación, estimación, diagnóstico y Proyecciones. Se aplicó el modelo ARMA en los datos mensuales de Rumania para el periodo de enero 2010 a septiembre 2021, por último se comparó dos modelos de pronóstico AR(1)MA(1) y AR(1)MA(2). Los resultados mostraron que el mejor modelo para predecir la demanda turística es AR(1)MA(1)

## IV. Metodología

### 4.1. Tipo y diseño de investigación

No experimental, Longitudinal

### 4.2. Variables de investigación

- Variable: Gasto medio por turista
- Tipo: Cuantitativa
- Unidades: Miles de euros
- Descripción: El gasto medio de los turistas internacionales que visitan España

### 4.3. Análisis Exploratorio

En esta sección analizaremos la variable de estudio, es decir el gasto medio por turista.

**Tabla 1**

*Estadísticas Descriptivas del gasto medio por turista*

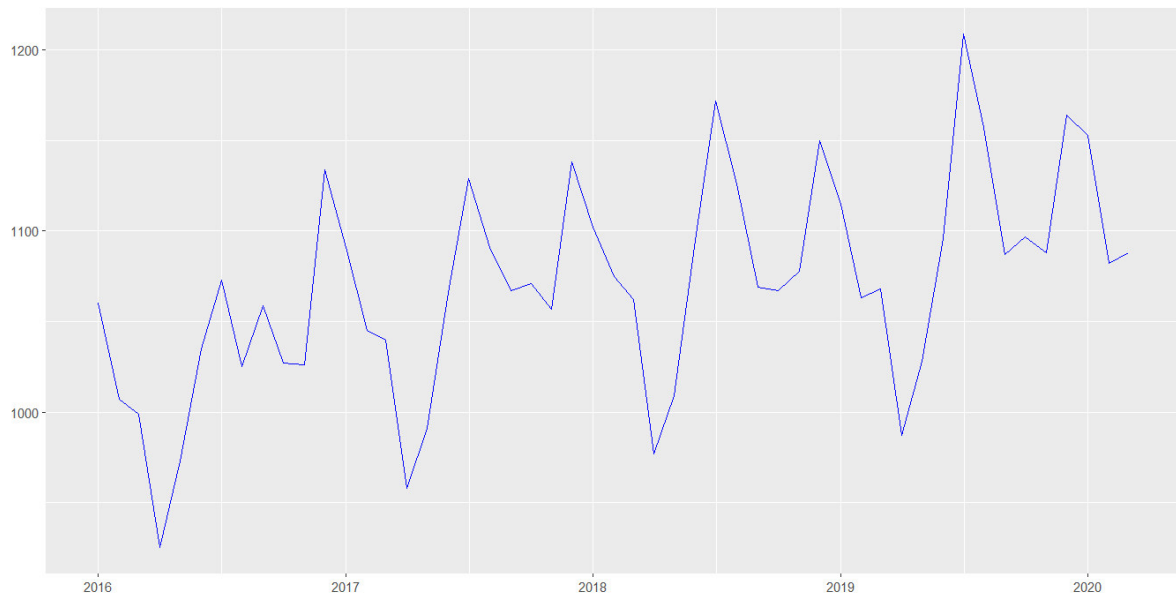
	Gasto Medio (miles de euros)
Valor mínimo	925
Primer Cuartil	1032
Mediana	1069
Media	1070
Tercer Cuartil	1096
Valor Máximo	1209
Desviación Estándar	58.35

*Nota.* Expresada en miles de Euros de enero 2016 – marzo 2020. Elaboración propia

En la tabla 1, observamos que, en promedio, el gasto por turista en el periodo 2016-2020, fue de  $1070 \pm 58.35$  euros. Mientras que el 50% de los gastos están por debajo de 1069, siendo el gasto mínimo de 925 euros, mientras que el máximo fue de 1209 euros.

## Figura 8

*Gastos medios por turista de enero 2016 – marzo 2020*

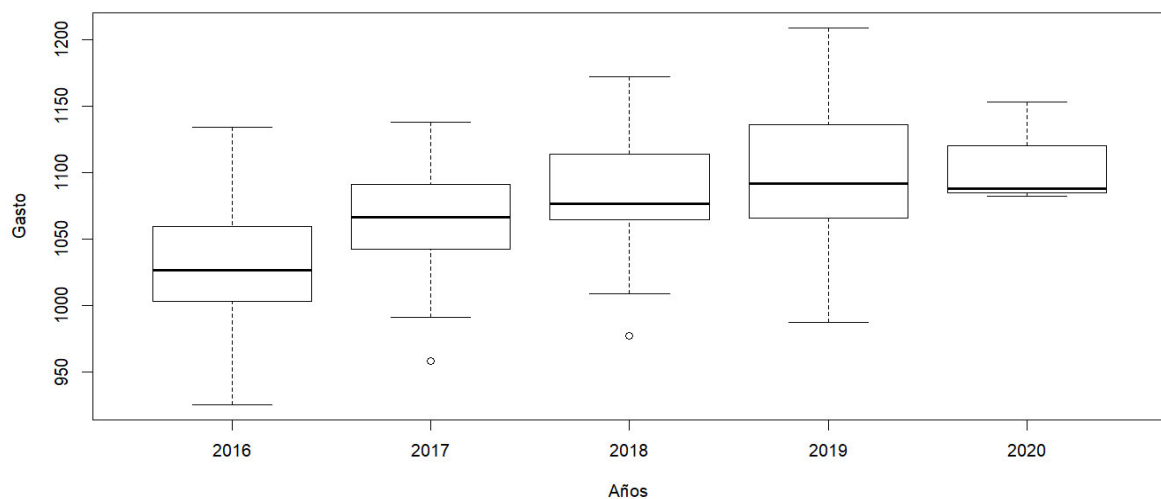


Nota. Elaboración propia en RStudio.

En la figura 8, se observa la serie del gasto medio por turista, en la cual se tiene picos que nos indicarían un patrón de estacionalidad en meses, también se observa una tendencia positiva en el tiempo que nos indica que posiblemente la serie no sea estacionaria.

## Figura 9

*Diagrama de Cajas Simples*



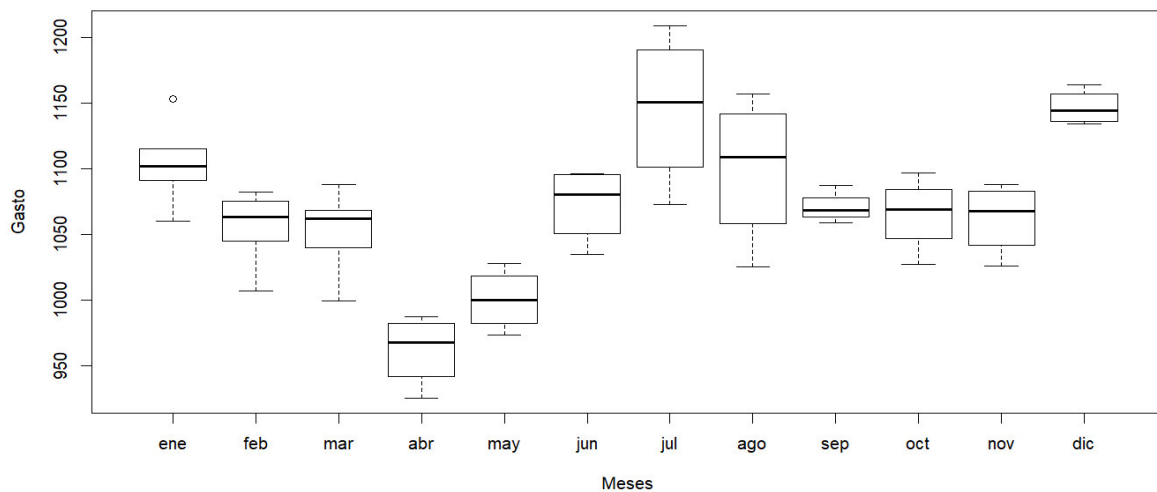
*Nota.* Gasto medio anual por turista en España de enero 2016 a marzo 2020.

Elaboración propia en RStudio.

En la figura 9, se observa que hay tendencia a medida que los años pasan, por ello se indicaría que no hay estacionariedad por nivel. Por otra parte, se observa una mayor dispersión para el año 2016 y 2019. Por último, para el año 2017 y 2018 se tienen datos atípicos.

## Figura 10

### Diagrama de Cajas Agregadas



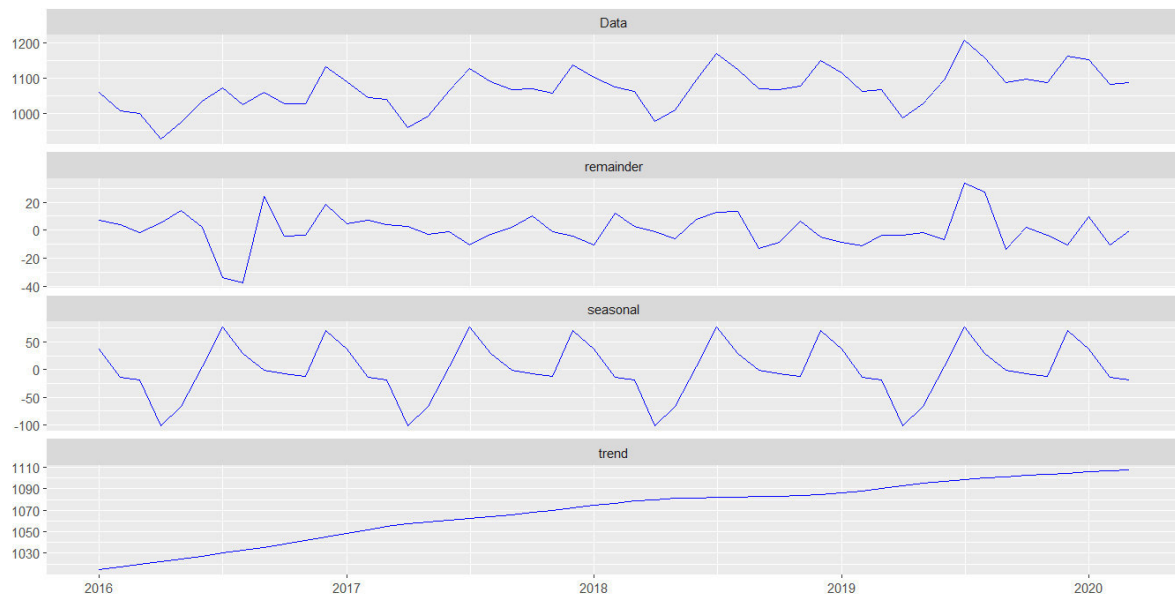
*Nota.* Gasto medio mensual por turista en España de enero 2016 a marzo 2020.

Elaboración propia en RStudio.

En la figura 10, se observa que la media no es constante por mes, por ello la serie no es estacionaria en media. Se observa un dato atípico en el mes de enero, además los gastos medios más altos por turista se dieron en el mes de julio y diciembre. Por último, se observa un patrón estacional, ya que, del mes de junio a julio, el gasto se va incrementando y luego se observa una caída, posteriormente del mes de noviembre a diciembre, el gasto se incrementa nuevamente.

## Figura 11

### Descomposición de la serie gasto medio por turista

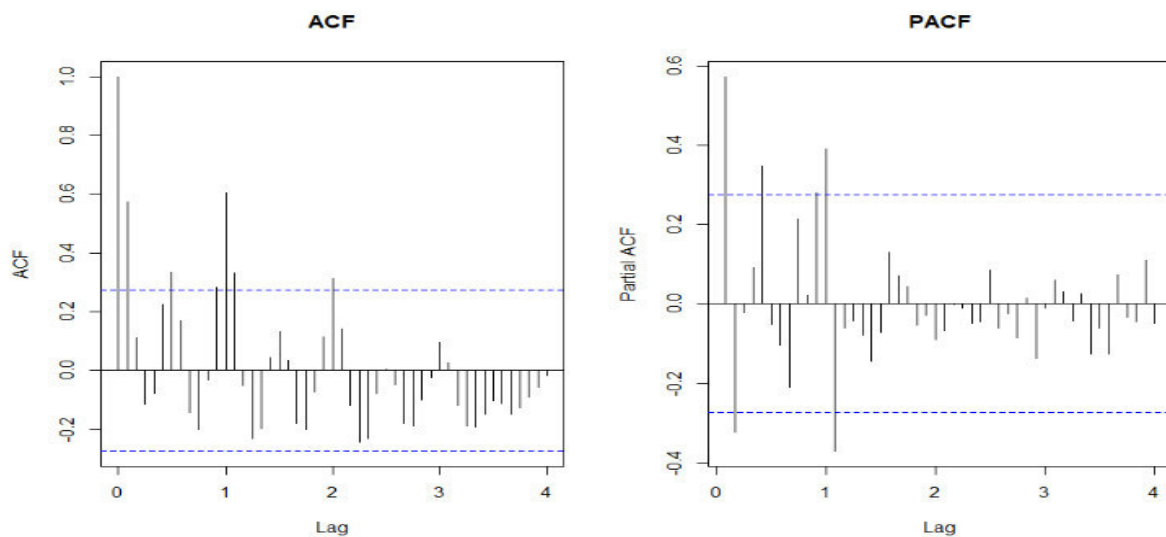


*Nota.* Componentes de la serie gasto medio por turista en España de enero 2016 a marzo 2020. Elaboración propia en RStudio.

En la figura 11, se observa picos en la serie, por lo tanto, se comprueba lo anterior, es decir la presencia de la componente estacional, además se tiene una tendencia positiva de la serie.

## Figura 12

### Autocorrelación Simple y Autocorrelación Parcial



*Nota.* Elaboración propia en RStudio.

En la figura 12, observamos como las autocorrelaciones simples descienden rápidamente que nos indica que la serie, posiblemente, sea estacionaria, es decir, que se mantiene constante en el tiempo, pero como habíamos mencionado anteriormente se observaba que no era estacionaria, por eso esto se verificará con el test de Dickey – Fuller posteriormente

## **Tabla 2**

*Prueba de Estacionariedad (Raíz Unitaria) - Dickey Fuller Aumentada*

	Estadístico t	P-Value
Test Estadístico de Dickey	-5.1381	0.01
Fuller Aumentada		

*Nota.* Elaboración propia

En la tabla 2, se realizó la prueba de Dickey – Fuller Aumentada, en la cual la hipótesis nula( $H_0$ ) es que la serie tiene raíz unitaria (no es estacionaria) y la hipótesis alterna( $H_1$ ) es que la serie no tiene raíz unitaria (es estacionaria). De acuerdo a los resultados y a un nivel de significancia del 5%, se obtuvo un p-value =  $0.01 < 0.05$ , por lo que se rechaza la hipótesis nula y se concluye que la serie es estacionaria.

### **Observación:**

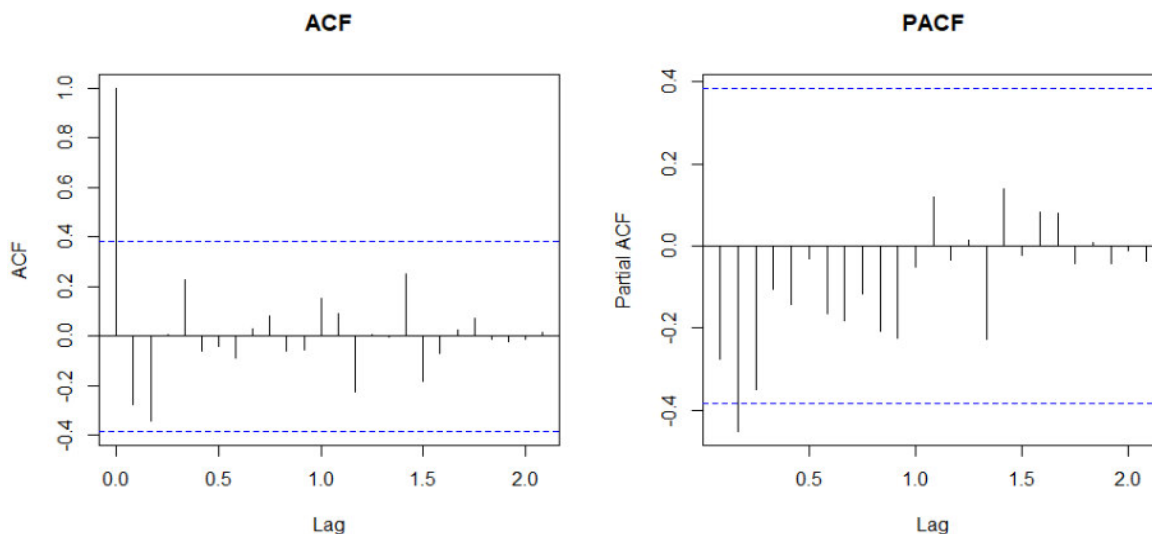
Si bien se ya se comprobó que la serie es estacionaria, guiándonos del software RStudio, el cual tiene funciones (ndiffs(yt) y nsdiffs(yt)), que nos ayuda a ver las diferencias que se deben de hacer a la serie, para obtener un modelo más óptimo. Se encontró que debemos realizar una diferencia en la parte regular, así como una diferencia en la parte estacional.

Aquí se dividió la serie de enero 2016 hasta marzo 2019 (train) y de abril 2019 hasta marzo 2020 (test).

Luego de realizar la diferencia en la parte estacional y la parte regular a la serie se obtuvo las siguientes autocorrelaciones simples y parciales:

**Figura 13**

*ACF y PACF de la serie diferenciada*



*Nota.* Autocorrelaciones simples y parciales después de hacer una diferencia en la parte regular y una diferencia en la parte estacional. Elaboración propia en RStudio.

En la figura 13, observamos como las autocorrelaciones simples descienden rápidamente que nos indica que la serie, es estacionaria, es decir, que se mantiene constante en el tiempo, cabe resaltar que ahora se observa que las autocorrelaciones descienden más rápido que las autocorrelaciones sin hacer alguna diferencia.

**Tabla 3**

*Prueba de Estacionariedad (Raíz Unitaria) - Dickey Fuller Aumentada*

	Estadístico t	P-Value
Test Estadístico de Dickey	-4.928	0.01
Fuller Aumentada		

*Nota.* Elaboración propia

En la tabla 3, se realizó la prueba de Dickey – Fuller Aumentada, en la cual la hipótesis nula( $H_0$ ) es que la serie tiene raíz unitaria (no es estacionaria) y la hipótesis alterna( $H_1$ ) es que la serie no tiene raíz unitaria (es estacionaria). De acuerdo a los resultados y a un nivel de



significancia del 5%, se obtuvo un  $p\text{-value} = 0.01 < 0.05$ , por lo que se rechaza la hipótesis nula y se concluye que la serie temporal es estacionaria en la parte regular y estacional.

#### 4.4. Ajuste de Modelos

**Tabla 4**

*Análisis de los Errores de Estimación*

MODELOS	AIC	MAPE
SARIMA(0,1,1)(1,1,0) <sub>12</sub>	224.86	0.71
SARIMA(2,1,1)(0,1,0) <sub>12</sub>	227.23	0.78
SARIMA(1,1,1)(0,1,0) <sub>12</sub>	227.23	0.82
SARIMA(1,1,0)(0,1,0) <sub>12</sub>	231.68	0.90

*Nota.* Elaboración propia

En la tabla 4, se realizó la evaluación de los modelos, en la cual se comparó el AIC y el MAPE, se sabe que el mejor modelo es el que tiene menor AIC y MAPE, por lo que al observar la tabla, para el modelo SARIMA(0,1,1)(1,1,0)<sub>12</sub>, se obtuvo el valor más pequeño de AIC = 224.86, de igual forma con el MAPE = 0.71, en comparación a los otros modelos.

**Tabla 5**

*Coefficientes del modelo seleccionado SARIMA(0,1,1)(1,1,0)<sub>12</sub>*

	Coefficiente	Error Est.	Valor Z	Pr(> z )
ma1	-0.78	0.11	-6.87	6.139e-12 ***
sar1	0.46	0.22	2.05	0.03 *

*Nota.* Elaboración propia

En la tabla 5, se deduce que a un nivel de significancia del 5%, el coeficiente ma1 y sar1 son significativos.

**Tabla 6***Coefficientes del modelo seleccionado SARIMA(2,1,1)(0,1,0)<sub>12</sub>*

	Coefficiente	Error Est.	Valor Z	Pr(> z )
ar1	0.07	0.27	0.27	0.78
ar2	-0.32	0.21	-1.49	0.13
ma1	-0.64	0.26	-2.42	0.01 *

*Nota.* Elaboración propia

En la tabla 6, se deduce que a un nivel de significancia del 5%, el coeficiente ma1 es significativo, y los coeficientes ar1 y ar2 no son significativos.

**Tabla 7***Coefficientes del modelo seleccionado SARIMA(1,1,1)(0,1,0)<sub>12</sub>*

	Coefficiente	Error Est.	Valor Z	Pr(> z )
ar1	0.18	0.23	0.78	0.43
ma1	-0.80	0.13	-5.86	4.403e-09 ***

*Nota.* Elaboración propia

En la tabla 7, se concluye que a un nivel de significancia del 5%, el coeficiente ma1 es significativo y el coeficiente ar1 no es significativo.

**Tabla 8***Coefficientes del modelo seleccionado SARIMA(1,1,0)(0,1,0)<sub>12</sub>*

	Coefficiente	Error Est.	Valor Z	Pr(> z )
ar1	-0.27	0.18	-1.44	0.14

*Nota.* Elaboración propia

En la tabla 8, se concluye que a un nivel de significancia del 5%, el coeficiente ar1 no es significativo.

#### 4.5. Evaluación de los supuestos de los errores

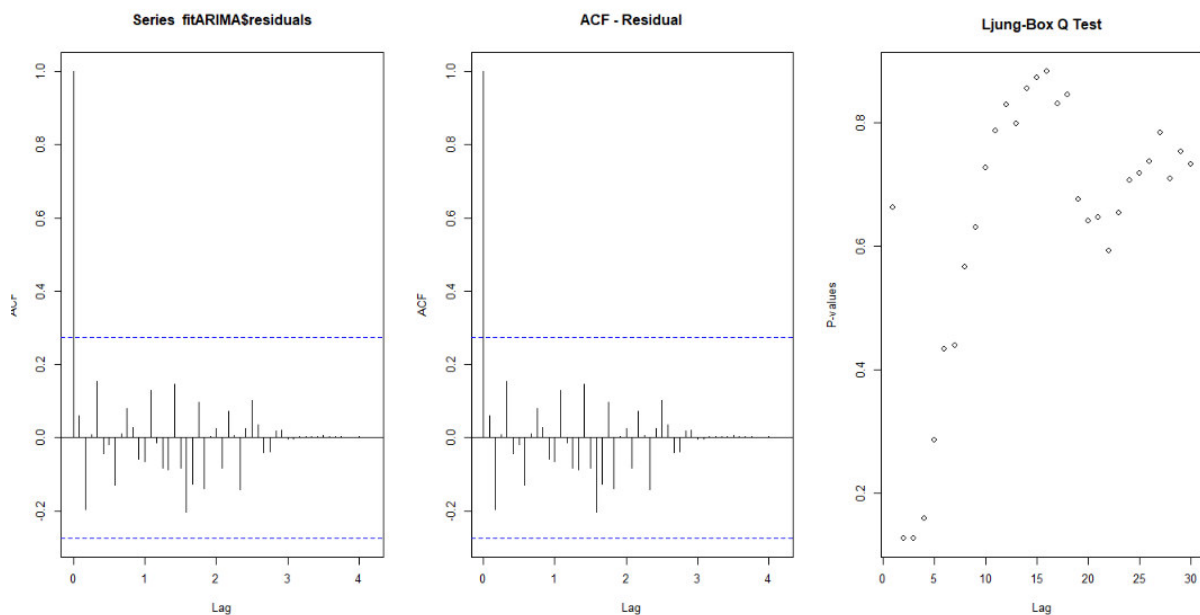
El ruido blanco presenta las siguientes características:

- *La varianza es constante*
- *Proviene de una distribución normal*
- *Son independientes y aleatorios en cualquier periodo del tiempo*

Por ello, se realizó las siguientes pruebas para probar la aleatoriedad, normalidad, la independencia, homocedasticidad.

#### Figura 14

*ACF y PACF y Ljung-Box Q Test de los errores*



*Nota.* Elaboración propia en RStudio.

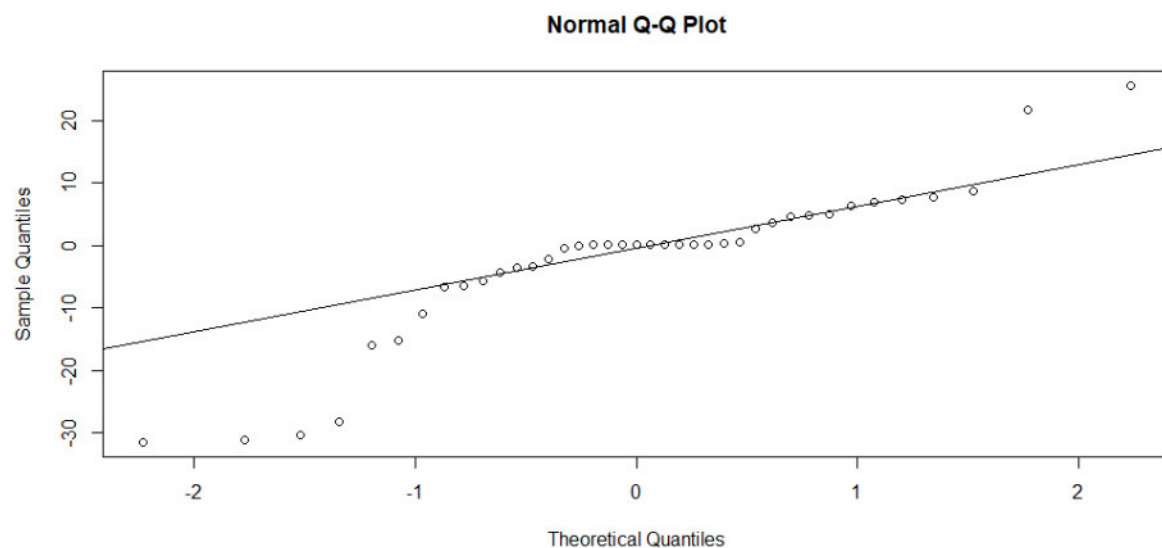
En la figura 14, observamos como las autocorrelaciones simples y autocorrelaciones parciales se encuentran dentro de las bandas de confianza. Por otro lado, se observa que los errores son aleatorios, pero esto lo comprobaremos con el Test de Ljung-Box.

**Tabla 9***Test de Ljung-Box para la prueba de aleatoriedad de los errores*

Estadístico	Valor
X-squared	0.02
df	1
p-value	0.88

*Nota.* Elaboración propia

En la tabla 9, se puede concluir que a un nivel de significancia del 5%, teniendo como  $H_0$ : Los errores son independientes y  $H_1$ : Los errores no son independientes, se concluye con  $p\text{-valor} = 0.88 > 0.05$ , por lo que no se rechaza la  $H_0$ , y se concluye que los errores son independientes.

**Figura 15***Distribución Normal de los Errores**Nota.* Elaboración propia en RStudio.

En la figura 15, observamos que gráficamente los errores no se siguen una distribución normal ya que no se ajustan a la línea, para corroborar ello se aplicará el test de Shapiro Wilk.

**Tabla 10***Test de normalidad de los errores (Shapiro Wilk)*

Estadístico	Valor
W	0.87
p-value	0.0004

*Nota.* Elaboración propia

En la tabla 10, podemos concluir que a un nivel de significancia del 5%, teniendo como Ho: Los errores siguen una distribución normal y H1: Los errores no siguen una distribución normal, se concluye con p-valor = 0.0004 < 0.05, por lo que se rechaza la H0, y se concluye que los errores no se comportan como una distribución normal.

**Tabla 11***Test de homocedasticidad de los errores (Breush Pagan Test)*

Estadístico	Valor
BP	0.96
df	1
p-value	0.32

*Nota.* Elaboración propia

En la tabla 11, podemos concluir que a un nivel de significancia del 5%, teniendo como Ho: Los errores son homocedásticos y H1: Los errores no son homocedásticos, se concluye con p-valor = 0.96 > 0.05, por lo que no se rechaza la H0, y se concluye que los errores son homocedásticos.

**Tabla 12***Pronósticos (Test), valores reales e Intervalos de Confianza*

Mes	Año	Predicciones	Valores Reales	IC 95%	
				LI	LS
Abril	2019	983.2697	1068	947.9832	1018.556
Mayo	2019	1014.8010	987	978.7199	1050.882
Junio	2019	1105.9566	1028	1069.0979	1142.815
Julio	2019	1189.5183	1096	1151.8981	1227.138
Agosto	2019	1139.7688	1209	1101.4022	1178.135
Septiembre	2019	1067.3019	1157	1028.2032	1106.401
Octubre	2019	1062.4897	1087	1022.6724	1102.307
Noviembre	2019	1085.2070	1097	1044.6838	1125.730
Diciembre	2019	1152.9888	1088	1111.7717	1194.206
Enero	2020	1118.5183	1164	1076.5580	1160.357
Febrero	2020	1054.7402	1153	1012.1693	1097.311
Marzo	2020	1068.1766	1082	1024.9448	111.409

*Nota.* Elaboración propia

En la tabla 12, se tienen las predicciones, valores reales de la serie y los intervalos de confianza para el pronóstico de abril 2019 a marzo 2020 (test).

**Tabla 13***Comparación de los errores (Train y Test)*

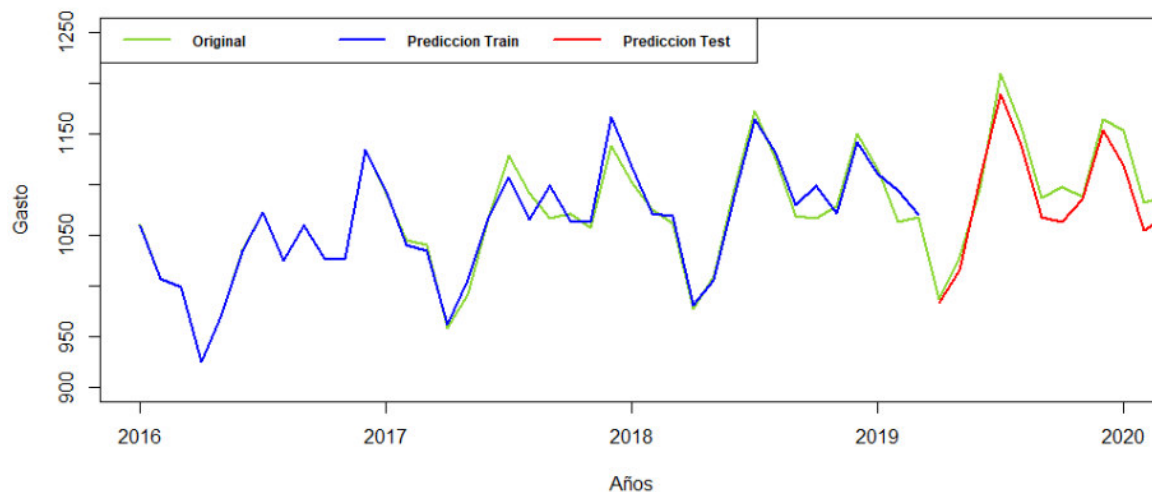
	ME	RMSE	MAE	MPE	MAPE	ACF1
Train	2.284181	12.34952	7.76133	0.2072992	0.7120019	0.02309774
Test	-16.11033	20.41443	17.76977	-1.480344	1.630389	0.03904257

*Nota .* Elaboración propia

En la tabla 13, se tienen los indicadores de errores para el train (enero 2016 a marzo 2019) y para el test (abril 2019 a marzo 2020). Se observa que el MAPE = 0.71% para el Train, sin embargo, para el Test MAPE = 1.63%, esto quiere decir que, si bien para los datos de entrenamiento se tiene un menor MAPE, es decir realizará mejores predicciones con menos errores, al hacer el Test, el MAPE se incrementa, es decir los errores son mayores.

**Figura 16**

*Pronósticos y valores reales del gasto medio por turista en España*



*Nota.* Del periodo enero 2016 – marzo 2020. *Elaboración propia en RStudio.*

En la figura 16, observamos la serie original y la serie usando el modelo elegido SARIMA(0,1,1)(1,1,0)<sub>12</sub>. (Train y Test)

**Tabla 14**

*Pronósticos e Intervalos de Confianza abril 2020 – marzo 2021*

Mes	Año	Predicciones	IC 95%	
			LI	LS
Abril	2020	983.5727	910.1405	1057.005
Mayo	2020	1014.8843	939.1352	1090.633
Junio	2020	1108.4564	1030.4591	1186.454
Julio	2020	1195.0935	1114.9111	1275.276
Agosto	2020	1143.5866	1061.2770	1225.896
Septiembre	2020	1063.8705	979.4873	1148.254
Octubre	2020	1057.7403	971.3333	1144.147
Noviembre	2020	1085.9494	997.5650	1174.334
Diciembre	2020	1151.7541	1061.4355	1242.073
Enero	2021	1117.4425	1025.2303	1209.655
Febrero	2021	1048.2333	954.1656	1142.301
Marzo	2021	1065.6239	969.7366	1161.511

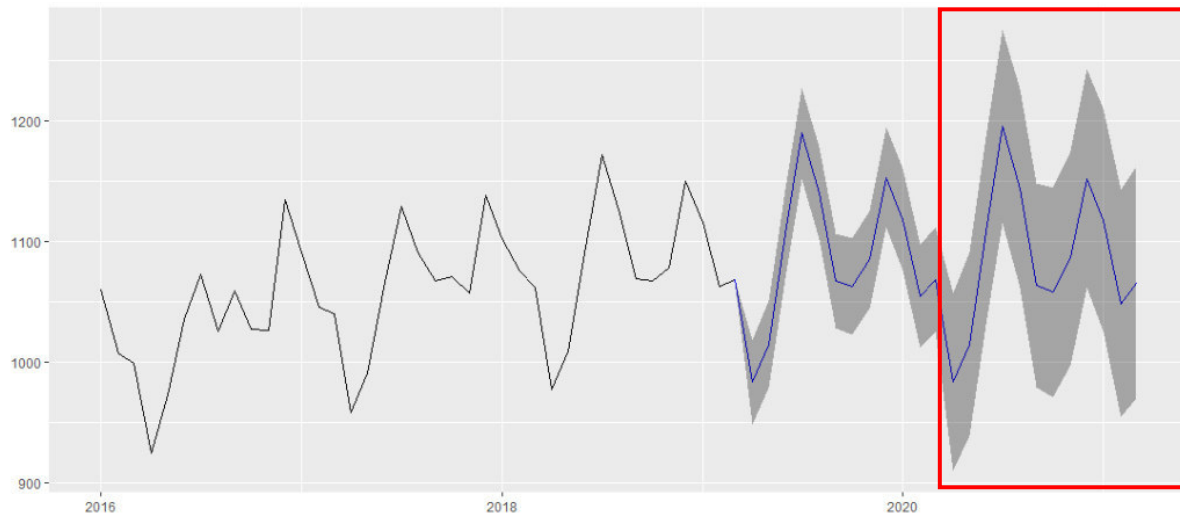
*Nota* . *Elaboración propia*

En la tabla 14, se tienen los pronósticos e intervalos de confianza de abril 2020 a marzo

2021.

### Figura 17

*Pronósticos del gasto medio por turista para abril 2020 – marzo 2021*



*Nota. Elaboración propia en RStudio.*

En la figura 16, observamos las predicciones usando el modelo elegido SARIMA(0,1,1)(1,1,0)<sub>12</sub>, se puede observar un buen ajuste y que el gasto medio por turista se sigue incrementando a media que pasan los meses, se observa la tendencia positiva en la serie, por otra parte se tiene la caída en el mes de abril 2020 como en los meses de abril para los años anteriores, es decir en el mes de abril es en donde el gasto medio por turista cae en el pico más bajo y los picos más altos en julio y diciembre del 2020, así como en los mismos meses para los años anteriores, es decir en julio y diciembre, el gasto medio por turista llega al pico más alto.



**Tabla 15***Pronósticos y Datos reales abril 2020 – marzo 2021*

Mes	Año	Predicciones	Datos reales	Pérdida
Abril	2020	983.5727	0	-983.57
Mayo	2020	1014.8843	0	-1014.88
Junio	2020	1108.4564	133,89	-974.57
Julio	2020	1195.0935	2.457,94	1262.85
Agosto	2020	1143.5866	2.456,85	1313.26
Septiembre	2020	1063.8705	964,49	-99.38
Octubre	2020	1057.7403	864,49	-193.25
Noviembre	2020	1085.9494	481,98	-603.97
Diciembre	2020	1151.7541	696,81	-454.94
Enero	2021	1117.4425	457,43	-660.01
Febrero	2021	1048.2333	323,82	-724.41
Marzo	2021	1065.6239	543,89	-521.73

*Nota* . Elaboración propia

En la tabla 15, se tienen los pronósticos y los valores reales de abril 2020 a marzo, en la cual se puede observar que para la gran parte de los meses hubo pérdidas en el gasto medio por turista, esto se debió a la pandemia que hubo a partir de esos meses, en la cual las fronteras fueron cerradas y no hubo turistas como en otros años anteriores sin pandemia.

## V. Conclusiones

- El modelo que se identificó para el gasto medio por turista en España de enero 2016 a diciembre 2018 (train) es una serie estacionaria y estacional, cabe resaltar que a la serie se le aplicó 1 diferencia en la parte regular y 1 diferencia en la parte estacional, esto fue ya que en el software RStudio, se tiene una función llamada  $ndiffs(yt)$  y  $nsdiffs(yt)$ , la cual nos dice si sería favorable aplicar alguna diferencia a la serie para encontrar un modelo más óptimo.
- El modelo elegido para pronosticar el gasto medio por turista de enero 2016 a diciembre 2018 (train) es un SARIMA(0,1,1)(1,1,0)<sub>12</sub>.
- Se sabe que es necesario realizar la validación de los errores, por lo que se hizo la prueba de shapiro wilk y se concluyó que los errores no siguen una distribución normal; también se realizó el Test de Ljung-Box y se concluyó que los errores son independientes, y por último, se realizó la prueba de Breush Pagan y se concluyó que los errores son homocedásticos.
- Se realizó el pronóstico (test) y se evaluó los errores del pronóstico del gasto medio por turista en España de enero 2019 a marzo 2020, se obtuvo el Porcentaje de Error Medio Absoluto (MAPE) de 1.63%, Porcentaje Medio del Error (PME) de -1.48% y el Error Medio Cuadrático (EMC) de -16.11.
- Los pronósticos que se han obtenido para el gasto medio por turista en España de abril 2020 a marzo 2021 son: 983.5727 euros en el mes de abril 2020, 1014.8843 euros en el mes de mayo 2020, 1108.4564 euros en el mes de junio 2020, 1195.0935 euros en el mes de julio 2020, 1143.5866 euros en el mes de agosto 2020, 1063.8705 euros en el mes de septiembre 2020, 1057.7403 euros en el mes de octubre 2020, 1085.9494 euros en el mes de noviembre 2020, 1151.7541 euros en el mes de diciembre 2020, 1117.4425

euros en el mes de enero 2021, 1048.2333 euros en el mes de febrero 2021 y 1065.6239 euros en el mes de marzo 2021.

- Al observar las predicciones para el gasto medio por turista, se observa que, para el mes de abril 2020, el gasto medio cae en comparación a los otros meses, y para el mes de julio y diciembre 2020, el gasto llega al pico más alto. Este patrón se ha repetido en los años anteriores.
- Por último, al observar los pronósticos y los valores reales de abril 2020 a marzo se puede concluir que para la gran parte de los meses hubo pérdidas en el gasto medio por turista, esto se debió a la pandemia que hubo a partir de esos meses, en la cual las fronteras fueron cerradas y no hubo turistas como en otros años anteriores sin pandemia.

## VI. Recomendaciones

- Implementar los pronósticos del gasto por turista en el dashboard del proyecto TurismEsp, ya que con ello se pueden ofrecer mejores servicios a los turistas, más promociones para el turista y con ello incrementar los ingresos de la empresa.
- Se recomienda tener en cuenta que, para los meses de julio y diciembre, es en donde el turista realiza el mayor gasto medio, aquí puede influir que en estos meses haya más llegada de turistas, o que haya eventos importantes y por ello se tienen más turistas, por lo que se podría realizar un estudio más a detalle de la afluencia de turistas en los meses del año. Por otro lado, teniendo en cuenta que, en estos meses los turistas gastan más, se pueden ofrecer diversas promociones, tours, etc.
- De igual forma, se recomienda tomar en cuenta que para el mes de abril se generó el gasto medio más bajo en comparación a los otros meses, por lo que se pueden ofrecer descuentos para el turista o precios que estén acorde al mercado.

## VII. Referencias

- Álvaro, M. G. (2010). *Análisis de serie de tiempo*. 389.
- Barreto, R. (2017). Estudio de la estacionariedad de la llegada mensual de turistas internacionales al santuario histórico de Machu Pichu en el periodo de enero de 2005 a diciembre de 2014. *In Crescendo*, 7(2), 77–90.  
<https://dialnet.unirioja.es/servlet/articulo?codigo=5758209>
- Bee Dagum, E. (2010). *View of Time series modeling and decomposition*. *Statistica*, 70(4).  
<https://doi.org/https://doi.org/10.6092/issn.1973-2201/3597>
- Boj, A. G. (2017). *Panorama del turismo español en el periodo 2004-2016. Análisis de la estacionalidad del gasto del turismo extranjero*.
- Boland, J. (2011). Box–Jenkins Time Series Models. *International Encyclopedia of Statistical Science*. [https://doi.org/https://doi.org/10.1007/978-3-642-04898-2\\_153](https://doi.org/https://doi.org/10.1007/978-3-642-04898-2_153)
- Bowerman, B., Koehler, A., & O’Connell, R. (2007). *Pronósticos, Series de Tiempo y Regresión: Un Enfoque Aplicado* (Cuarta).
- De la Fuente Fernandez, S. (2018). *METODOLOGÍA DE BOX - JENKINS*.
- Divino, J. A., & McAleer, M. (2017). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management*, 31(6), 846–854.  
<https://doi.org/10.1016/j.tourman.2009.09.002>
- E. Hanke, J., & Dean, W. W. (2010). *Pronósticos en los Negocios* (Novena).
- Laurente Blanco, L. F. (2021). Demanda De Turismo Internacional En Tiempos De Covid-19 En La Región De Puno- Perú. *Economía Coyuntural, Revista de Temas de Coyuntura y Perspectivas*, 6(1), 49–78. <https://doi.org/10.5281/zenodo.4657227>

- Lindes Castro, P. (2017). Análisis del gasto por turista en Andalucía mediante el uso de Series Temporales. *Universidad de Jaén*.
- Mahía, R., & De Arce, R. (2003). Modelos arima. *Academia*.
- Mauricio, J. A. (2007). *Análisis de Series Temporales*.
- Montenegro García, Á. (2010). *Analisis-de-Series-de-Tiempo* (p. 389).
- Mushtaq, R. (2011). *Testing time series data for stationarity. 1*, 1–19.
- Palma, W. (2016). *TIME SERIES ANALYSIS*.
- Torres Diaz, J. E. (2020). Estimación de la Demanda Turística en Puerto Rico: 1998 a 2018. *Universidad de Puerto Rico*.
- Tudela-Mamani, J. W., Cahui-Cahui, E., & Aliaga-Melo, G. (2022). Impacto del COVID-19 en la demanda de turismo internacional del Perú. Una aplicación de la metodología Box-Jenkins. *Revista de Investigaciones Altoandinas - Journal of High Andean Research*, 24(1), 27–36. <https://doi.org/10.18271/ria.2022.317>
- Turtureanu, A. G., Pripoaie, R., Cretu, C. M., Sirbu, C. G., Marinescu, E. Ș., Talaghir, L. G., & Chițu, F. (2022). A Projection Approach of Tourist Circulation under Conditions of Uncertainty. *Sustainability (Switzerland)*, 14(4). <https://doi.org/10.3390/su14041964>

## VIII. Anexos

## Anexo 1

## Dashboard de TurismEsp

