



Análisis predictivo de la morbilidad hospitalaria, buscando definir estrategias para el mejoramiento de la eficiencia en la atención en salud de una IPS de mediana complejidad de la ciudad de Buga

PROYECTO DE GRADO

Mario Germán Castro Saldarriaga

Asesor

Christian Camilo Urcuqui López

**FACULTAD DE INGENIERÍA
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI
2023**

Análisis predictivo de la morbilidad hospitalaria, buscando definir estrategias para el mejoramiento de la eficiencia en la atención en salud de una IPS de mediana complejidad de la ciudad de Buga

Mario Germán Castro Saldarriaga

**Trabajo de grado para optar al título de
Magister en Ciencia de Datos**

**Asesor
Christian Camilo Urcuqui López**



**FACULTAD DE INGENIERÍA
MAESTRÍA EN CIENCIA DE DATOS
SANTIAGO DE CALI
2023**

CONTENIDO

	pág.
RESUMEN	14
1. INTRODUCCIÓN	15
1.1 Contexto y Antecedentes	15
1.2 Planteamiento del Problema	16
1.3 Objetivo General	17
1.4 Objetivos Específicos	17
2. ANTECEDENTES	18
2.1 Marco Teórico	18
2.1.1 Proceso de atención hospitalaria	18
2.1.2 Prestación de servicios de salud	18
2.1.3 Morbilidad hospitalaria	19
2.1.4 Patologías clínicas	19
2.1.5 Modelos predictivos de datos	20
2.1.6 Evaluación de modelos predictivos	20
2.1.7 Técnicas de Machine Learning	22
2.1.8 Aprendizaje supervisado	23
2.1.9 Análisis de Series de tiempo	23
2.1.9.1 HoltWinters	26
2.1.9.2 ARIMA	26
2.2 Estado del arte	27
2.2.1 Criterios de Búsqueda y Clasificación	27
2.2.2 Trabajos seleccionados	27
2.2.3 Matriz resumen de criterios y trabajo	29
3. METODOLOGÍA	30
3.1 Fases de la Metodología CRISP-DM	31
3.1.1 Fase I. Business Understanding. Definición de necesidades del cliente (comprensión del negocio)	31
3.1.2 Fase II. Data Understanding. Estudio y comprensión de los datos	31
3.1.3 Fase III. Data Preparation. Análisis de los datos y selección de características	31
3.1.4 Fase IV. Modeling. Modelado	32
3.1.5 Fase V. Evaluation. Evaluación (obtención de resultados)	32
3.1.6 Fase VI. Deployment. Despliegue (puesta en producción)	32
4. PRESENTACIÓN DE LA PROPUESTA	33
4.1 Recolección de los datos	34
4.2 Limpieza y preparación de los datos	34
4.3 Descripción de los datos	35
4.4 Definición de variables	38
4.5 Conjunto de datos final	40

4.6	Construcción de nuevas variables	40
4.7	Modelamiento predictivo	76
4.9	Evaluación del modelo	92
5.	DISEÑO DE EXPERIMENTO DE VALIDACIÓN	94
6.	RESULTADOS OBTENIDOS	121
7.	CONCLUSIONES Y TRABAJO FUTURO	124
8.	BIBLIOGRAFÍA	125
9.	ANEXOS	127

LISTA DE TABLAS

Tabla 1. Estado del arte	29
Tabla 2. Diagnósticos representativos año 2022	36
Tabla 3. Diagnósticos representativos año 2019	37
Tabla 4. Diagnósticos representativos año 2019 – 2022	37
Tabla 5. Resumen estadístico	42
Tabla 6. Resumen estadístico del Cólico	46
Tabla 7. Resumen estadístico del dolor abdominal	51
Tabla 8. Resumen estadístico del EPOC	55
Tabla 9. Resumen estadístico de Fiebre	59
Tabla 10. Resumen estadístico de infección de vías urinarias	64
Tabla 11. Resumen estadístico de Otros dolores abdominales	68
Tabla 12. Resumen estadístico de Traumatismos múltiples	73
Tabla 13. Resultados del MAPE(%) con toda la data	121
Tabla 14. Resultados del MAPE(%) eliminando Outliers	122
Tabla 15. Comparativo del MAPE(%) de los 2 experimentos	122

LISTA DE FIGURAS

Figura 1. Comportamiento de una serie de tiempo	24
Figura 2. Ciclo de vida del proyecto de minería de datos	30
Figura 3. Pacientes ingresados en la institución 2019-2022	34
Figura 4. Diagnóstico de ingreso 2021	34
Figura 5. Diagnóstico de ingreso 2019	35
Figura 6. Diagnóstico de ingreso 2022	36
Figura 7. Total de pacientes ingresados intrahospitalariamente diariamente Octubre 2018 – 2019	40
Figura 8. Total de pacientes ingresados intrahospitalariamente diariamente 2022	41
Figura 9. Total de pacientes ingresados intrahospitalariamente diariamente 2018-2022	41
Figura 10. Diagrama de caja	42
Figura 11. Pacientes que ingresan a la institución Octubre 2018-2022 sin Tendencia	43
Figura 12. Autocorrelación	43
Figura 13. Suavización de la serie temporal	44
Figura 14. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2019	45
Figura 15. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria año 2022	45
Figura 16. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2022	46
Figura 17. Diagrama de caja por Cólico	47

Figura 18. Pacientes que ingresan a la institución por cólico Octubre 2018-2022 sin tendencia	47
Figura 19. Autocorrelación por cólico	48
Figura 20. Suavización de una serie temporal del cólico	48
Figura 21. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2019	49
Figura 22. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria año 2022	50
Figura 23. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022	50
Figura 24. Diagrama de caja del Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022	51
Figura 25. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	52
Figura 26. Autocorrelación por Dolor Abdominal	52
Figura 27. Suavización de una serie temporal por Dolor Abdominal	53
Figura 28. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2019	54
Figura 29. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria año 2022	54
Figura 30. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022	54
Figura 31. Diagrama de caja por EPOC	55
Figura 32. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	56
Figura 33. Autocorrelación por EPOC	56
Figura 34. Suavización de una serie temporal por EPOC	57

Figura 35. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2019	58
Figura 36. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria año 2022	58
Figura 37. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022	59
Figura 38. Diagrama de caja por FIEBRE	60
Figura 39. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	60
Figura 40. Autocorrelación por FIEBRE	61
Figura 41. Suavización de una serie temporal por FIEBRE	61
Figura 42. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2019	62
Figura 43. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria año 2022	63
Figura 44. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022	63
Figura 45. Diagrama de caja para infección de vías urinarias	64
Figura 46. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	65
Figura 47. Autocorrelación para infección de vías urinarias	65
Figura 48. Suavización de la serie temporal para infección de vías urinarias	66
Figura 49. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2019	67
Figura 50. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria año 2022	67

Figura 51. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022	68
Figura 52. Diagrama de caja para Otros dolores abdominales	69
Figura 53. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	70
Figura 54. Autocorrelación para Otros dolores abdominales	70
Figura 55. Suavización de la serie temporal para Otros dolores abdominales	71
Figura 56. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2019	72
Figura 57. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria año 2022	72
Figura 58. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022	73
Figura 59. Diagrama de caja por Traumatismos múltiples	74
Figura 60. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia	75
Figura 61. Suavización de la serie temporal para Traumatismos múltiples	75
Figura 62. Resultado modelo ARIMA	77
Figura 63. Residuales	78
Figura 64. Densidad	78
Figura 65. Comportamiento real vs predicción	78
Figura 66. Resultado modelo ARIMA para Cólico	79
Figura 67. Residual para Cólico	80
Figura 68. Densidad para Cólico	80

Figura 69. Comportamiento real vs predicción para Cólico	80
Figura 70. Resultado modelo ARIMA para dolor abdominal	81
Figura 71. Residual para Dolor abdominal	82
Figura 72. Densidad para Dolor abdominal	82
Figura 73. Comportamiento real vs predicción para dolor abdominal	82
Figura 74. Resultado modelo ARIMA para EPOC	83
Figura 75. Residual para EPOC	84
Figura 76. Densidad para EPOC	84
Figura 77. Comportamiento real vs predicción para EPOC	84
Figura 78. Resultado modelo ARIMA para fiebre	85
Figura 79. Residual para Fiebre	86
Figura 80. Densidad para Fiebre	86
Figura 81. Comportamiento real vs predicción para Fiebre	86
Figura 82. Resultado modelo ARIMA para infección de vías urinarias	87
Figura 83. Residual para Infección de vías urinarias	88
Figura 84. Densidad para Infección de vías urinarias	88
Figura 85. Comportamiento real vs predicción para Infección de vías urinarias	88
Figura 86. Resultado modelo ARIMA para otros dolores abdominales	89
Figura 87. Residual para otros dolores abdominales	90
Figura 88. Densidad para otros dolores abdominales	90
Figura 89. Comportamiento real vs predicción para otros dolores abdominales	90
Figura 90. Resultado modelo ARIMA para traumatismos múltiples	91

Figura 91. Residual para Traumatismos múltiples	92
Figura 92. Densidad para Traumatismos múltiples	92
Figura 93. Comportamiento real vs predicción para Traumatismos múltiples	92
Figura 94. Comportamiento predicción vs real	95
Figura 95. Estandarización residual por “v”	95
Figura 96. Histograma con estimación de densidad	95
Figura 97. Normal Q-Q	96
Figura 98. Correlograma	96
Figura 99. Estimación final de predicciones futuras	96
Figura 100. Comportamiento predicción vs real para cólico	98
Figura 101. Estandarización residual por “v” para cólico	98
Figura 102. Histograma con estimación de densidad para cólico	98
Figura 103. Normal Q-Q para cólico	99
Figura 104. Correlograma para cólico	99
Figura 105. Estimación final de predicciones futuras para cólico	99
Figura 106. Comportamiento predicción vs real para dolor abdominal	101
Figura 107. Estandarización residual por “v” para dolor abdominal	101
Figura 108. Histograma con estimación de densidad para dolor abdominal	101
Figura 109. Normal Q-Q para dolor abdominal	102
Figura 110. Correlograma dolor abdominal	102
Figura 111. Estimación final de predicciones futuras para dolor abdominal	102
Figura 112. Comportamiento predicción vs real para EPOC	104

Figura 113. Estandarización residual por “v” para EPOC	104
Figura 114. Histograma con estimación de densidad para EPOC	104
Figura 115. Normal Q-Q para EPOC	105
Figura 116. Correlograma para EPOC	105
Figura 117. Estimación final de predicciones futuras para EPOC	105
Figura 118. Comportamiento predicción vs real para Fiebre	108
Figura 119. Estandarización residual por “v” para Fiebre	108
Figura 120. Histograma con estimación de densidad para Fiebre	108
Figura 121. Normal Q-Q para Fiebre	109
Figura 122. Correlograma para Fiebre	109
Figura 123. Estimación final de predicciones futuras para Fiebre	109
Figura 124. Comportamiento predicción vs real para Infección de vías urinarias	111
Figura 125. Estandarización residual por “v” para Infección de vías urinarias	111
Figura 126. Histograma con estimación de densidad para Infección de vías Urinarias	112
Figura 127. Normal Q-Q para Infección de vías urinarias	112
Figura 128. Correlograma para Infección de vías urinarias	112
Figura 129. Estimación final de predicciones futuras par Infección de vías Urinarias	113
Figura 130. Comportamiento predicción vs real para Otros dolores abdominales	114
Figura 131. Estandarización residual por “v” para Otros dolores abdominales	115
Figura 132. Histograma con estimación de densidad para Otros dolores abdominales	115

Figura 133. Normal Q-Q para Otros dolores abdominales	115
Figura 134. Correlograma para Otros dolores abdominales	116
Figura 135. Estimación final de predicciones futuras para Otros dolores Abdominales	116
Figura 136. Comportamiento predicción vs real para Traumatismos múltiples	117
Figura 137. Estandarización residual por “v” para Traumatismos múltiples	118
Figura 138. Histograma con estimación de densidad para Traumatismos múltiples	118
Figura 139. Normal Q-Q para Traumatismos múltiples	118
Figura 140. Correlograma para Traumatismos múltiples	119
Figura 141. Estimación final de predicciones futuras para Traumatismos Múltiples	119

RESUMEN

En este trabajo de grado, se formuló una propuesta desde la ciencia de datos para abordar la ineficiencia en la prestación de servicios de salud a los pacientes atendidos en una IPS de mediana complejidad de la ciudad de Buga.

Para hacerlo, se propuso la metodología CRISP-DM que consideraba las siguientes fases: a) Definición de necesidades del cliente, b) Estudio y comprensión de los datos, c) Análisis de los datos y selección de variables, d) implementación del modelo de series de tiempo a través del método ARIMA y e) evaluación y resultados.

El modelo de ciencia de datos usado para abordar la solución al problema fue el método de serie de tiempo ARIMA. Los resultados obtenidos muestran que para la predicción de la morbilidad hospitalaria de las patologías más críticas utilizando como métrica el MAPE arrojó resultados entre 22% - 40%.

La validación a la que fue sometida la propuesta consistió en exponer el modelo dividiendo los datos en un set de entrenamiento y set de prueba correspondiente en un 75%:25% respectivamente.

Finalmente, después del trabajo desarrollado y de la validación realizada, se puede afirmar que el enfoque de solución propuesta y la metodología empleada para obtenerla resultan apropiados para dar respuesta a la pregunta de investigación ya que se logró una predicción de pacientes a atender por cada una de las patologías críticas.

1. INTRODUCCIÓN

1.1 Contexto y Antecedentes

La IPS de mediana complejidad de la ciudad de Guadalajara de Buga, es una institución dedicada a prestación de servicios de salud, sin ánimo de lucro. Su compromiso es lograr los mejores resultados en salud para la población, de manera eficiente para el sistema, con altos estándares de calidad y seguridad en la atención, brindando la mejor experiencia para el usuario y su familia, con sostenibilidad financiera para nuestra institución, contribuyendo al bienestar de nuestros clientes, colaboradores, proveedores y comunidad en general, soportados en un talento humano íntegro y competente y acorde con los valores de la institución.

Teniendo en cuenta lo anterior, es de suma importancia lograr consolidar unos procesos consistentes que permitan alcanzar el cumplimiento de los objetivos institucionales, es por esta razón que en la institución es importante cambiar el enfoque en su direccionamiento estratégico, y para esto es importante hacerlo bajo otro modelo, ya que actualmente la institución no cuenta con procesos sólidos enfocados en análisis de información que les permita predecir sucesos en la atención, y mucho menos proyectar los costos en la atención, que permitan desarrollar estrategias para mejorar la calidad en la atención, siendo eficiente en la utilización de recursos. La institución se ha dedicado a ser reactivos en la solución de las problemáticas enfocadas en el mejoramiento de los resultados en salud.

Por lo anterior, la institución tiene la oportunidad de aplicar los modelos estadísticos y de análisis de datos para planear la atención, enfocándolo en el mejoramiento de las condiciones de salud, analizando de una forma predictiva la morbilidad hospitalaria y como se puede calcular los costos asociados a esto, con el fin de impactar de forma positiva con decisiones estratégicas para mejorar la calidad en la salud de los pacientes y ahorrando en costos hospitalarios. Es importante que estos modelos puedan tener como resultados, indicadores de control (KPI) para la monitorización y seguimiento en tiempo real a través de tableros de control con visualizaciones específicas.

En este proyecto de trabajo de grado se va a abordar como situación de interés la prestación de servicios de salud hospitalaria, el problema de esta situación de interés es la ineficiencia en la prestación de los servicios de salud, esto es un problema que involucra al personal asistencia (Medico – Enfermería), pacientes, comunidad en general y al sistema de salud, lo anterior es un problema porque genera sobre costos en la atención en salud, disminución de la rentabilidad del negocio y malos resultados clínicos en la atención de los pacientes y las

consecuencias de este problema son la generación de diagnósticos tardíos que generan complicaciones en la atención del paciente, genera sobre costos en la atención que impacta en el aumento de desperdicios en el actuar médico y disminución de la rentabilidad del negocio, se generan re- ingresos de los pacientes ocasionando sobre ocupación de pacientes, y por ultimo genera estancias prolongadas ocasionando demoras en la recuperación física, mental y social de nuestros pacientes.

1.2 Planteamiento del Problema

Actualmente en la institución se está presentando una creciente demanda de servicios que sumado a la poca capacidad de oferta para responder a la demanda está causando inoportunidad en la atención, adicionalmente no se planea de forma proactiva la atención de los paciente, hay fallas en las capacitaciones y entrenamiento al personal y sobre todo no se cuenta con herramientas tecnológicas y personal calificado para el análisis de datos que influyen en que no se tengan procesos de análisis de datos predictivos que permitan promover la toma de decisiones en la atención de salud, todo esto genera **ineficiencia en la prestación de servicios de salud a los pacientes atendidos en nuestra institución**, que trae como consecuencia la generación de diagnósticos tardíos que generan complicaciones en la atención del paciente, genera sobre costos en la atención que impacta en el aumento de desperdicios en el actuar médico y disminución de la rentabilidad del negocio, se generan re- ingresos de los pacientes ocasionando sobre ocupación de pacientes, y por ultimo genera estancias prolongadas ocasionando demoras en la recuperación física, mental y social de nuestros pacientes.

¿Es posible predecir el comportamiento de la morbilidad hospitalaria de las patologías más críticas utilizando modelos de análisis de datos?

1.3 Objetivo General

Formular y validar un modelo de análisis de datos para predecir el comportamiento de la morbilidad hospitalaria de las patologías más críticas en el proceso de atención hospitalaria en una IPS de mediana complejidad de la ciudad de Guadalajara de Buga.

1.4 Objetivos Específicos

- Identificar el modelo de análisis de datos que permita predecir de manera eficiente el comportamiento de la morbilidad hospitalaria de las patologías más críticas.
- Construir el modelo de análisis de datos que permita predecir la morbilidad hospitalaria teniendo en cuenta los criterios determinantes de las patologías más críticas.
- Evaluar el modelo de análisis de datos que permita predecir de manera eficiente el comportamiento de la morbilidad hospitalaria.
- Predecir el comportamiento de la morbilidad hospitalaria de una IPS de mediana complejidad de Buga.

2. ANTECEDENTES

2.1 Marco Teórico

2.1.1 Proceso de atención hospitalaria

Con frecuencia se utiliza el término proceso de atención médica para hacer referencia a aquella actividad mediante la cual paciente y médico establecen una relación dirigida a la identificación y solución de uno o más problemas de salud presentes en dicho paciente. Mediante este proceso, el médico lleva a cabo su función profesional fundamental: la labor asistencial, en su dimensión básicamente curativa (Corona & Fonseca,2010).

2.1.2 Prestación de servicios de salud

Está relacionada con la garantía al acceso a los servicios de salud, el mejoramiento permanente de la calidad de la atención en salud, y la búsqueda y generación de eficiencia en la prestación de los servicios de salud a la población colombiana (Ministerio de salud,2010).

En Colombia la Política Nacional de Prestación de Servicios de Salud fue construida por el Ministerio de Salud y Protección Social, para determinar los lineamientos estratégicos sobre los cuales se debe regir la prestación de los servicios en Colombia. La política detalla las estrategias y líneas de acción establecidas para el logro de los objetivos propuestos. La página web del Ministerio de Salud y Protección Social, permite conocer con detalle, avances en materia de política y de normatividad para la prestación de los servicios de salud.

La política tiene como propósito "garantizar el acceso, optimizar el uso de los recursos y mejorar la calidad de los servicios que se prestan a la población". Además, establece tres ejes:

- En el eje de acceso, se destacan las estrategias de consolidación de las redes de prestación de servicios de salud, el mejoramiento de la capacidad resolutive en la prestación de servicios de baja complejidad y de urgencias, y el aumento de cobertura de aseguramiento a la población pobre y vulnerable.
- El eje de calidad está centrado en la adecuada implementación del sistema obligatorio de garantía de calidad y el fomento al desarrollo del talento humano.
- El eje de eficiencia se refiere a la red de prestadores públicos de servicios de salud, de tal manera, que las estrategias se dirigen al mejoramiento de las

condiciones de carácter estructural y funcional de estas instituciones para garantizar su viabilidad (Ministerio de salud,2010)

2.1.3 Morbilidad hospitalaria

La morbilidad es un estado enfermo, de discapacidad, o mala salud debido a cualquier causa. El término puede ser usado para referirse a la existencia de cualquier forma de enfermedad, o al grado en que esa condición de salud afecta al paciente. También se entiende como la proporción de seres vivos que enferman en un sitio y tiempo determinado. Minoritariamente también se usa como sinónimo de morbilidad, que etimológicamente es correcto.

Cabe mencionar también que una condición comórbida es una condición médica concomitante o concurrente con otra enfermedad primaria, por ejemplo COVID-19 con insuficiencia cardíaca o diabetes (Wikipedia).

2.1.4 Patologías clínicas

El término “patología”, sinónimo de anatomía patológica, tiene dos raíces, la palabra deriva del griego pathos y logos, que significan “enfermedad” y “estudio de”, respectivamente. La enfermedad es un estado anormal de la vida y la anatomía patológica, es el campo de la medicina que trata de conocer y explicar, con base racional, las condiciones bajo las que se da, sustentándose en el estudio de todos sus aspectos: 1) las causas, 2) los mecanismos de producción o patogénesis, 3) los cambios estructurales en células, tejidos y órganos y 4) las consecuencias funcionales de dichos cambios, expresados como síntomas y signos.

Desde sus orígenes como disciplina (siglo XVIII) la patología es el fundamento científico de la medicina, y tiene como fin entender la vida en su estado anormal. Se puede decir que, a pesar de los avances tecnológicos, en pruebas diagnósticas y la identificación de factores causales, sigue siendo piedra angular de la práctica médica y su papel aún es inapelable para el diagnóstico de muchas enfermedades. Es muy importante que el estudiante de medicina esté consciente de que el aprendizaje de la anatomía patológica debe ir unido a la práctica clínica. (Alicia Rodríguez Velasco).

2.1.5 Modelos predictivos de datos

Los Modelos Predictivos son un grupo de técnicas que, mediante los campos del aprendizaje automático, la recolección de datos históricos, el Big Data y el reconocimiento de patrones, pretende dar una predicción de resultados futuros; con el objetivo de precisar la toma de decisiones mediante técnicas de análisis de datos.

En los últimos años el área predictiva ha tomado gran protagonismo en los negocios, la medicina, los servicios financieros, las políticas gubernamentales, la publicidad, la mercadotecnia, las redes sociales y gran cantidad de campos de aplicación.

2.1.6 Evaluación de modelos predictivos

Imagina que ya has recopilado datos, los has limpiado y usado para generar un modelo predictivo usando Machine Learning. ¿Cómo demuestras que tu modelo funciona y es útil para usarlo en producción? Necesitas una forma de convencer a tu jefa/cliente de que puede usar tu modelo con cierta confianza, (Lopez,2021).

No basta con decirle que tome tu palabra, ¿cierto? Es aquí donde las métricas de evaluación de modelos entran en juego. Nos permiten calcular de forma *numérica* el desempeño de nuestra solución. Además, las métricas nos permiten tener un registro del avance de nuestros modelos (Lopez,2021).

Matriz de Confusión

Por definición, la matriz de confusión C se construye de forma que el elemento $C_{i,j}$ es igual al número de observaciones *verdaderas* del grupo i , y clasificadas en el grupo j .

Puedes pensar en la matriz de confusión como la forma de saber qué tanto se confunde tu clasificador/predictor.

Imagina que quieres un modelo que prediga si un cliente comprará un producto sugerido. Establezcamos las etiquetas: Sí y No. Cuando el modelo predice un Sí significa una predicción positiva; se predice que el cliente comprará el producto. Cuando el modelo predice un No significa una predicción negativa; se predice que el cliente no comprará el producto.

Exactitud

En inglés esta métrica se llama *Accuracy*. La exactitud nos indica la habilidad de un clasificador para predecir de forma correcta cada clase.

Matemáticamente la exactitud se ve así:

$$exactitud(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i)$$

Precisión

Puedes pensar en la precisión como la habilidad que tiene un clasificador de *NO ASIGNAR* falsos positivos (ejemplos clasificados como positivos que en *realidad* son negativos). En otras palabras, la métrica mide qué tan *preciso* es el clasificador para encontrar verdaderos positivos (True Positives).

Matemáticamente la precisión se ve así:

$$P = \frac{T_p}{T_p + F_p}$$

Con:

$T_p = Verdaderos\ positivos$

$F_p = Falsos\ positivos$

Recall (Sensibilidad)

Intuitivamente esta métrica nos indica la habilidad del clasificador para encontrar todos los ejemplos verdaderos positivos.

Matemáticamente se ve así.

$$R = \frac{T_p}{T_p + F_n}$$

Con:

$T_p = Verdaderos\ positivos$

$F_n = Falsos\ negativos$

¿Qué métrica usar para la Evaluación de Modelos Predictivos?

Si estás intentando determinar qué tan bien funciona tu sistema, la forma más obvia de hacer esto es calcular cuán exacto es. Para esto tenemos la exactitud o en inglés,

el Accuracy. Recapitulando, todas las métricas se calculan usando un conjunto de datos de prueba. Para realizar la Evaluación de modelos predictivos, haces que tu modelo prediga usando datos de tu conjunto de prueba. Después, comparas los resultados que obtuvo con las etiquetas reales de tu conjunto de prueba.

La exactitud responde: ¿Qué porcentaje clasificó el modelo correctamente? Por ejemplo, digamos que tenemos un conjunto de datos de prueba de 1000 personas y estás prediciendo si comprarán el producto o no; 990 de las personas no lo harán y 10 lo harán.

2.1.7 Técnicas de Machine Learning

Las técnicas de Machine Learning permiten a los algoritmos identificar patrones complejos entre gran cantidad de datos, infiriendo así sus propias reglas para detectar patrones similares en nuevos conjuntos de datos.

Se crean sistemas inteligentes que mejoran de forma autónoma viendo datos. Así pueden aprender a predecir comportamientos, detectar similitudes o anomalías automáticamente o tomar las decisiones adecuadas.

La Inteligencia Artificial aprende con ejemplos. Los modelos de Machine Learning analizan muestras de datos reales que reflejan el proceso que se quiere mejorar o automatizar. Hablamos de entrenar algoritmos: el sistema extrae conclusiones relevantes o aprende a realizar una tarea sin ser programado previamente, solo a partir de lo que observa en los datos.

Tres pasos de un proyecto de Machine Learning:

- Revisión de los datos

Revisión de aquellos datos que pueden ser útiles, su almacenamiento y formatos. Hay que tratar los datos, eliminar los antiguos, los incompletos o los erróneos.

- Organización de los datos

Organizar los datos facilita el aprendizaje automático. Se eligen los datos que reflejen las decisiones que después tendrá que tomar el sistema automáticamente.

➤ Entrenamiento y validación del modelo

Viendo los datos, el modelo descubrirá patrones y relaciones entre ellos, configurando su propia combinación de parámetros para detectarlos automáticamente en nueva información.

2.1.8 Aprendizaje supervisado

La clasificación supervisada es una de las tareas que más frecuentemente son llevadas a cabo por los denominados Sistemas Inteligentes. Por lo tanto, un gran número de paradigmas desarrollados bien por la Estadística (Regresión Logística, Análisis Discriminante) o bien por la Inteligencia Artificial (Redes Neuronales, Inducción de Reglas, Árboles de Decisión, Redes Bayesianas) son capaces de realizar las tareas propias de la clasificación.

Paso previo a aplicar un método de clasificación, es la partición del conjunto de datos en dos conjuntos de datos más pequeños que serán utilizadas con los siguientes fines: entrenamiento y test. El subconjunto de datos de entrenamiento es utilizado para estimar los parámetros del modelo y el subconjunto de datos de test se emplea para comprobar el comportamiento del modelo estimado. Cada registro de la base de datos debe de aparecer en uno de los dos subconjuntos, y para dividir el conjunto de datos en ambos subconjuntos, se utiliza un procedimiento de muestreo: muestreo aleatorio simple o muestreo estratificado. Lo ideal es entrenar el modelo con un conjunto de datos independiente de los datos con los que realizamos el test.

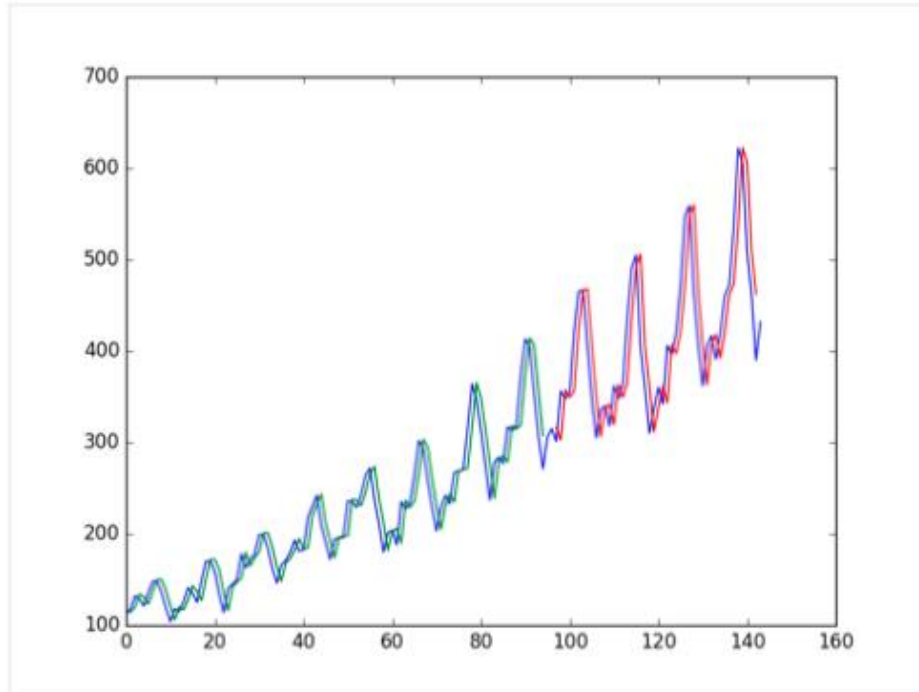
Como resultado de aplicar un método de clasificación, se cometerán dos errores, en el caso de una variable binaria que toma valores 0 y 1, habrá ceros que se clasifiquen incorrectamente como unos y unos que se clasifiquen incorrectamente como ceros.

2.1.9 Análisis de Series de tiempo

Rojas-Jimenez, K. (2022) define una serie de tiempo como una secuencia de datos u observaciones medidos en determinados momentos, en intervalos iguales o desiguales, y ordenados cronológicamente.

El análisis de series de tiempo se refiere al proceso de analizar los datos disponibles para descubrir el patrón o la tendencia en los datos. Permite extraer y modelar las relaciones entre datos a lo largo del tiempo, sea extrapolando (hacia futuro) o interpolando (hacia el pasado) el comportamiento de datos no observados.

Figura 1. Comportamiento de una serie de tiempo



Fuente: Rojas-Jimenez, K. (2022)

Rojas-Jimenez, K. (2022) , propone que los análisis de series de tiempo tienen los siguientes tres pasos:

- Análisis exploratorio
- Escogencia y ajuste del modelo
- Diagnóstico

Además, existen dos técnicas de modelado principales para hacer análisis de series de tiempo:

- Holt Winters
- ARIMA models

Análisis exploratorio

Dentro del análisis exploratorio de series de tiempo se pueden realizar los siguientes procedimientos:

- **Estimación de tendencias y descomposición:** Se utiliza para el ajuste estacional. Busca construir, a partir de una serie temporal observada, una serie de subcomponentes con diferentes características.
- **Análisis de autocorrelación:** permite estimar qué valor en el pasado tiene una correlación con el valor actual. Proporciona la estimación de los factores p , d , q para los modelos ARIMA.
- **Análisis espectral:** Describe cómo la variación en una serie de tiempo puede ser explicada por componentes cíclicos. Permite separar los componentes periódicos en un ambiente ruidoso.

Componentes de las series de tiempo

- **Tendencia:** Es el patrón subyacente en los datos a lo largo del tiempo. No es necesariamente lineal.
- **Estacionalidad:** Cuando una serie está influenciada por factores estacionales de periodo fijo como el día, mes, trimestre.
- **Ciclicidad:** Cuando los datos muestran subidas y caídas que no son del período fijo. Las fluctuaciones suelen ser de al menos 2 años.
- **Aleatoriedad:** parte inexplicable de los datos

Otras definiciones incluyen el **ruido blanco** (White noise) referente a la suposición de que los valores en una serie de tiempo son aleatorios, independientes entre sí (no están correlacionados), tienen una media de cero y varianza constante. Cuando estos supuestos no se cumplen se deben aplicar modelos autorregresivos (AR) y de media móvil (MA) para corregir las infracciones de esta suposición.

Los **time Lags** se refieren a que los valores de y_t en una serie de tiempo se ven afectados por los valores de y en el pasado. Las regresiones que no contabilizan estos lags fallan en establecer (sobrestiman) las relaciones entre las variables dependientes e independientes.

Corrección de la estacionariedad (stationarity) y estacionalidad (seasonality)

Algunas técnicas para corregir la no estacionariedad (stationarity) son las siguientes:

- **Ajuste por diferencial:** calcula las diferencias entre observaciones consecutivas
- **Cálculo de log o raíz cuadrada:** para estabilizar la varianza no constante

- **Unit root test (Prueba de raíz unitaria):** esta prueba se usa para descubrir la primera diferencia o regresión que se debe usar para hacerla estacionaria. En la prueba Kwiatkowski-Phillips-Schmidt-Shin (KPSS), los valores pequeños de p sugieren que se requiere una diferenciación.

Escogencia y ajuste del modelo:

2.1.9.1 HoltWinters

Rojas-Jimenez, K. (2022), define el método como una de las técnicas para hacer suavizado y predicción en series de tiempo. Se puede usar la función HoltWinters() del programa **forecasts** si la serie de tiempo puede describirse utilizando un modelo aditivo con tendencia creciente o decreciente y sin estacionalidad.

Se controla mediante el parámetro *alfa*, para la estimación del nivel en el punto de tiempo actual; *beta* para la estimación de la pendiente b del componente de tendencia en el punto de tiempo actual y *gamma* para el componente estacional en el momento actual.

Los parámetros tienen valores entre 0 y 1 donde los valores cercanos a 0 significan que se coloca poco peso en las observaciones más recientes.

2.1.9.2 ARIMA

Rojas-Jimenez, K. (2022), define **ARIMA** es una abreviación de AutoRegressive Integrated Moving Average. En sí, es una combinación de las propiedades de diferentes tipos de modelos:

(AR) Auto Regressive Son modelos donde el valor de una variable en un periodo se relaciona con los valores de periodos previos. Se refiere a los lags p de las diferencias de los valores entre las series.

(I) se refiere al orden del número diferencial que se usan para hacer las series de tiempo estacionarias.

(MA) Moving Average Consideran la posibilidad de una relación entre la variable y los residuales de periodos previos. Se refiere a los lags q de los errores,

Supuestos del ARIMA

- **Los datos deben ser estacionarios:** las propiedades de la serie no dependen del momento en que se capturan. Un proceso estacionario tiene

una media y una variación que no cambian con el tiempo y no tiene una tendencia.

- **Los datos deben ser univariados:** ARIMA trabaja con una sola variable. La regresión automática tiene que ver con la regresión con los valores pasados.

2.2 Estado del arte

2.2.1 Criterios de Búsqueda y Clasificación

A continuación, se presentan las ecuaciones de búsqueda utilizadas, en donde las palabras claves utilizadas son Modelos predictivos de datos y el segundo criterio es nuestro dominio de negocio, la morbilidad hospitalaria:

Ecuaciones de Búsqueda
Modelos predictivos de datos + morbilidad hospitalaria
Machine learning + patologías clínicas
Modelos predictivos de datos + análisis en la morbilidad hospitalaria

Sobre el conjunto de proyectos resultantes se definieron las siguientes dimensiones de caracterización para su correspondiente clasificación y comparación.

No.	Dimensión de comparación
1	Sector
2	País
3	Temporalidad
4	Proceso
5	Técnicas y herramientas

2.2.2 Trabajos seleccionados

Los criterios anteriormente mencionados nos arrojaron una serie de trabajos, para lo cual se seleccionó los siguientes trabajos que aportan en cuanto a metodología y enfoque más información:

Modelo predictivo de minería de datos de apoyo a la gestión hospitalaria sobre la morbilidad de pacientes hospitalizados, (Edwin John Saldaña Valqui,2015)

El objetivo de este estudio fue mostrarnos que el modelo predictivo es parte fundamental de la minería de datos, que ayuda a comprender el patrón de comportamiento de los pacientes y de esta manera permite llegar a una toma de

decisión rápida y óptima. En la mayoría de los centros hospitalarios existe una gestión hospitalaria deficiente, no utilizándose en forma adecuada tanto los recursos humanos como materiales. Esto trae consigo que se acentúe aún más el desbalance entre la capacidad de oferta y demanda.

Comparación de modelos predictivos para la selección de pacientes de alta complejidad, (Marcos Estupiñán-Ramírez, Rita Tristancho-Ajamil, María Consuelo Company, Sancho Hilda Sánchez-Janáriz, 2019)

Este trabajo nos muestra cómo se compara la concordancia de los pesos de complejidad entre los estratificadores Clinical Risk Groups (CRG) y los grupos de morbilidad ajustada (GMA), para determinar cuál de ellos es el mejor predictor de ingreso hospitalario y optimizar el método para seleccionar el 0,5% de pacientes de más alta complejidad que se incluirán en un protocolo de intervención. El método utilizado fue un estudio analítico transversal en 18 zonas de salud de Canarias, con una población a estudio de 385.049 personas, usando variables sociodemográficas procedentes de la tarjeta sanitaria, los diagnósticos y el uso de los recursos asistenciales obtenidos de la historia electrónica de salud de atención primaria (HSAP) y del conjunto mínimo básico de datos hospitalario, el estado funcional registrado en la HSAP y los fármacos prescritos en el sistema de receta electrónica. A partir de esos datos se estimó la concordancia entre estratificadores, se evaluó la capacidad de cada estratificador para predecir ingresos y se construyeron modelos para optimizar la predicción.

Aplicación de analítica de datos para predicción de infección respiratoria aguda en Colombia, (Mayra Alejandra Jiménez Manjarres, 2019)

Este trabajo nos muestra como la predicción de los casos de IRA a nivel nacional en el mundo se encuentra en proceso de consolidación en países como Estados Unidos y China. Sin embargo, en Colombia no se ha iniciado el abordaje de la predicción analítica de este evento por parte del Instituto Nacional de Salud. En la actualidad, el Instituto Nacional de Salud solo cuenta con análisis descriptivos de los datos recopilados, teniendo en cuenta variables de persona, tiempo y lugar para las enfermedades, y de grupos de edad y gravedad para la circulación de virus respiratorios. Después de integrar información adicional a los eventos relacionada con clima y navegadores de internet teniendo como base el estado del arte, este proyecto pretende iniciar los análisis predictivos de IRA con estimación de casos semanales de IRA utilizando minería de datos. Análisis predictivos con los cuales se pueda encaminar a la planeación y asignación de forma anticipada de procedimientos y de recursos del Sistema de Salud ante eventuales propagaciones de la Infección Respiratoria Aguda en Colombia, ya que los análisis descriptivos realizados en la actualidad tienen restricciones para la generación de alertas tempranas.

Aplicación de machine learning en la gestión de las personas con enfermedades crónicas. Francesc Güell y Martí Zamora de Amalfi Analytics. (2020).

Este proyecto tiene como objetivo crear soluciones que ayuden a la gestión de servicios de salud de la forma más informada, a hospitales, servicios sociosanitarios, centros de salud y centros de atención primaria a mejorar en estas áreas de forma fácil y rápida con herramientas de gestión basadas en avanzados algoritmos de Machine Learning. En relación a la necesidad de mejorar la atención a los pacientes pluripatológicos se ha diseñado una herramienta que permite identificar grupos de pacientes con condiciones similares para aplicar medidas diferentes, optimizar recursos y minimizar costes, centrándose en los aspectos clínicos de la gestión.

2.2.3 Matriz resumen de criterios y trabajo

Teniendo en cuenta el estado del arte anteriormente mencionado, se presenta a continuación en la tabla 1 el resumen del estado del arte, donde se resalta en nuestro estudio la importancia de la utilización de algoritmos de series de tiempo para lograr la predicción de la morbilidad hospitalaria de las principales patologías.

Tabla 1. Estado del arte

Criterio	Referencia (1)	Referencia (2)	Referencia (3)	Referencia (4)	Trabajo de Grado
Sector	Sector Sanitario	Sector Sanitario	Sector Sanitario	Sector Sanitario	Sector Sanitario
País	Perú	España	Colombia	España	Guadalajara de Buga
Dimensión temporal	2008-2014	2013 - 2014	2013 - 2018	2019 - 2020	2019 – 2022
Proceso	Gestión hospitalaria en la morbilidad hospitalaria	Atención de pacientes de alta complejidad	Infección respiratoria aguda	Gestión de enfermedad des crónicas	Morbilidad hospitalaria en patologías críticas
Técnicas y herramientas	Análisis de series de tiempo	Modelos analíticos de predicción	Modelos analíticos de predicción	Machine learning	Análisis de series de tiempo

Fuente: Construcción propia

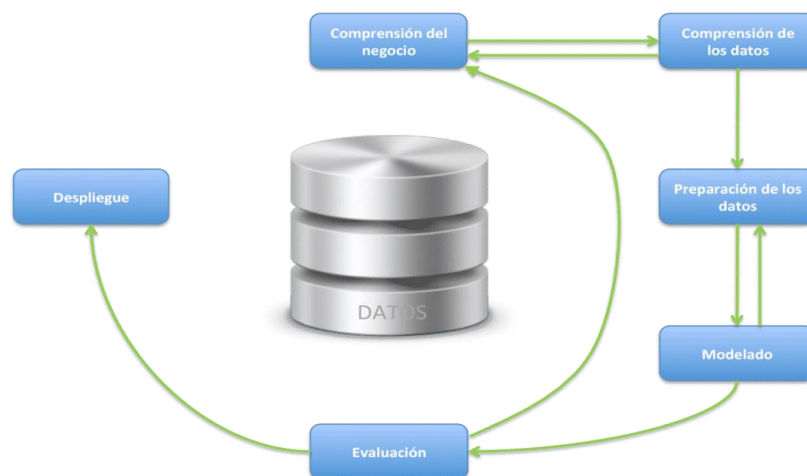
3. METODOLOGÍA

Para la ejecución del proyecto se utilizará la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), la cual proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software. El modelo CRISP-DM cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. En este nivel de descripción no es posible identificar todas las relaciones; las relaciones podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos. (Azevedo, Ana; Zantos, Manuel Filipe. KDD, SEMMA and CRISP-DM: a parallel overview. 2008).

La metodología CRISP-DM contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. Este contexto tiene en cuenta la existencia de un cliente que no es parte del equipo de desarrollo, así como el hecho de que el proyecto no sólo no acaba una vez se halla el modelo idóneo (ya que después se requiere un despliegue y un mantenimiento), sino que está relacionado con otros proyectos, y es preciso documentarlo de forma exhaustiva para que otros equipos de desarrollo utilicen el conocimiento adquirido y trabajen a partir de él.

El ciclo de vida del proyecto de minería de datos consiste en seis fases mostradas en la figura siguiente.

Figura 2. Ciclo de vida del proyecto de minería de datos



Fuente: Azevedo, Ana; Zantos, Manuel Filipe. KDD.(2008).

La secuencia de las fases no es rígida: se permite movimiento hacia adelante y hacia atrás entre diferentes fases. El resultado de cada fase determina qué fase, o qué tarea particular de una fase, hay que hacer después. Las flechas indican las dependencias más importantes y frecuentes.

El círculo externo en la figura simboliza la naturaleza cíclica de los proyectos de análisis de datos. El proyecto no se termina una vez que la solución se despliega. La información descubierta durante el proceso y la solución desplegada pueden producir nuevas iteraciones del modelo. Los procesos de análisis subsecuentes se beneficiarán de las experiencias previas.

A continuación, vamos a describir brevemente cada una de las fases.

3.1 Fases de la Metodología CRISP-DM

3.1.1 Fase I. Business Understanding. Definición de necesidades del cliente (comprensión del negocio)

Esta fase inicial se enfoca en la comprensión de los objetivos de proyecto de analítica. Después se convierte este conocimiento de los datos en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos.

3.1.2 Fase II. Data Understanding. Estudio y comprensión de los datos

La fase de entendimiento de datos comienza con la recolección de datos inicial y continúa con las actividades que permiten familiarizarse con los datos, identificar los problemas de calidad, descubrir conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.

3.1.3 Fase III. Data Preparation. Análisis de los datos y selección de características

La fase de preparación de datos cubre todas las actividades necesarias para construir el conjunto final de datos (los datos que se utilizarán en las herramientas de modelado) a partir de los datos en bruto iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

3.1.4 Fase IV. Modeling. Modelado

En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema (cuantas más mejor), y se calibran sus parámetros a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de los datos. Por lo tanto, casi siempre en cualquier proyecto se acaba volviendo a la fase de preparación de datos.

3.1.5 Fase V. Evaluation. Evaluación (obtención de resultados)

En esta etapa en el proyecto, se han construido uno o varios modelos que parecen alcanzar calidad suficiente desde una perspectiva de análisis de datos.

Antes de proceder al despliegue final del modelo, es importante evaluarlo a fondo y revisar los pasos ejecutados para crearlo, comparar el modelo obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no haya sido considerada suficientemente. Al final de esta fase, se debería obtener una decisión sobre la aplicación de los resultados del proceso de análisis de datos.

3.1.6 Fase VI. Deployment. Despliegue (puesta en producción)

Generalmente, la creación del modelo no es el final del proyecto. Incluso si el objetivo del modelo es de aumentar el conocimiento de los datos, el conocimiento obtenido tendrá que organizarse y presentarse para que el cliente pueda usarlo. Dependiendo de los requisitos, la fase de desarrollo puede ser tan simple como la generación de un informe o tan compleja como la realización periódica y quizás automatizada de un proceso de análisis de datos en la organización.

4. PRESENTACIÓN DE LA PROPUESTA

4.1 Recolección de los datos

Para resolver la pregunta de investigación se utiliza como fuente de datos los reportes de los ingresos de pacientes al hospital, los cuales se encuentran almacenados en el sistema de información de la institución. Este reporte se escoge por que en su estructura muestran todas las variables relevantes requeridas para nuestro estudio. Al tener un volumen tan grande de información almacenada en estos reportes se procede a su generación con periodicidad mensual desde el mes de octubre 2018 a diciembre 2022 para un total de 51 meses de información.

El reporte inicialmente contaba con un total de 316615 registros los cuales contaban con ingresos de todos los servicios por donde puede ingresar un paciente (Ambulatorios e intrahospitalarios), pero como nuestro estudio está orientado al ámbito intrahospitalario se realiza la depuración respectiva obteniéndose finalmente una base de datos de 299149 registros.

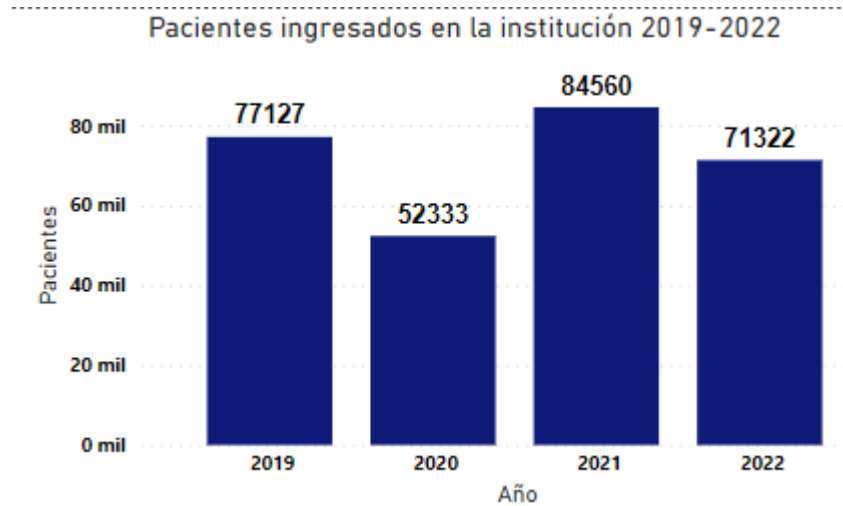
4.2 Limpieza y preparación de los datos

Luego de la recolección de los datos, se procede con la limpieza y preparación de los datos, para obtener la base final que permite estimar los modelos.

En esta etapa es importante destacar lo siguiente:

Definición del periodo de análisis: Como anteriormente se ha mencionado la data tiene como periodo de tiempo octubre 2018 – diciembre 2022, por lo que es importante destacar que en este periodo vivimos la pandemia del SARS-CoV-2, causa del COVID-19, la cual nos afectó considerablemente durante los años 2020 – 2021, por lo cual estos dos años se excluyen del proyecto de investigación quedándonos con 162398 registros, lo anteriormente se puede ver justificado en la siguiente información:

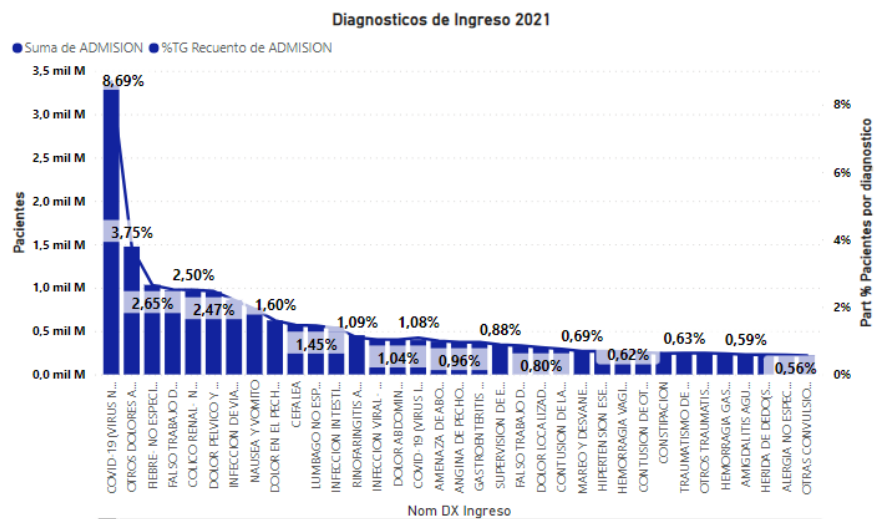
Figura 3. Pacientes ingresados en la institución 2019-2022



Fuente: Construcción propia

En la gráfica anterior se puede observar que para el año 2020 el decremento de pacientes atendido fue significativo (**32%**), y a pesar que en el año 2021 las atenciones incrementaron considerablemente, el comportamiento del COVID-19 continuaba predominando, por lo que la información del año 2021 no está completamente ajustada a la realidad ya que el diagnostico de mayor participación con respecto a los demás fue producto del COVID-19.

Figura 4. Diagnóstico de ingreso 2021



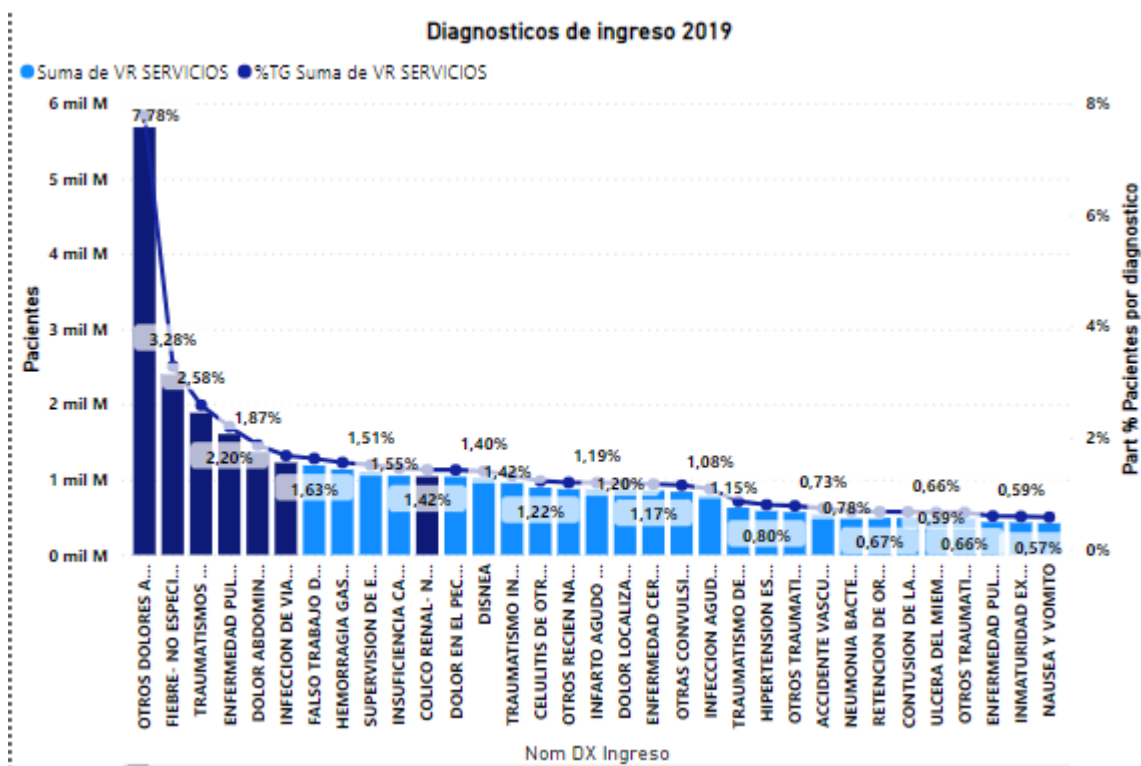
Fuente: Construcción propia

A partir de lo anterior, se obtiene un data set con el periodo octubre 2018 a diciembre 2019 y año 2022 con 162398 registros con 13 variables.

4.3 Descripción de los datos

A continuación, se presenta los siguientes análisis descriptivos de la base de datos, que permitirán conocer la información más relevante para el estudio. Inicialmente se incluye el comportamiento de los diagnósticos de ingreso para el año 2019 y 2022 en términos de su valor de facturación, indicándonos un dato relevante de importancia en la gestión medica:

Figura 5. Diagnóstico de ingreso 2019



Fuente: Construcción propia

A partir de esta gráfica, se puede evidenciar que para el año 2019 los diagnósticos (patologías clínicas) con mayor participación fueron los siguientes:

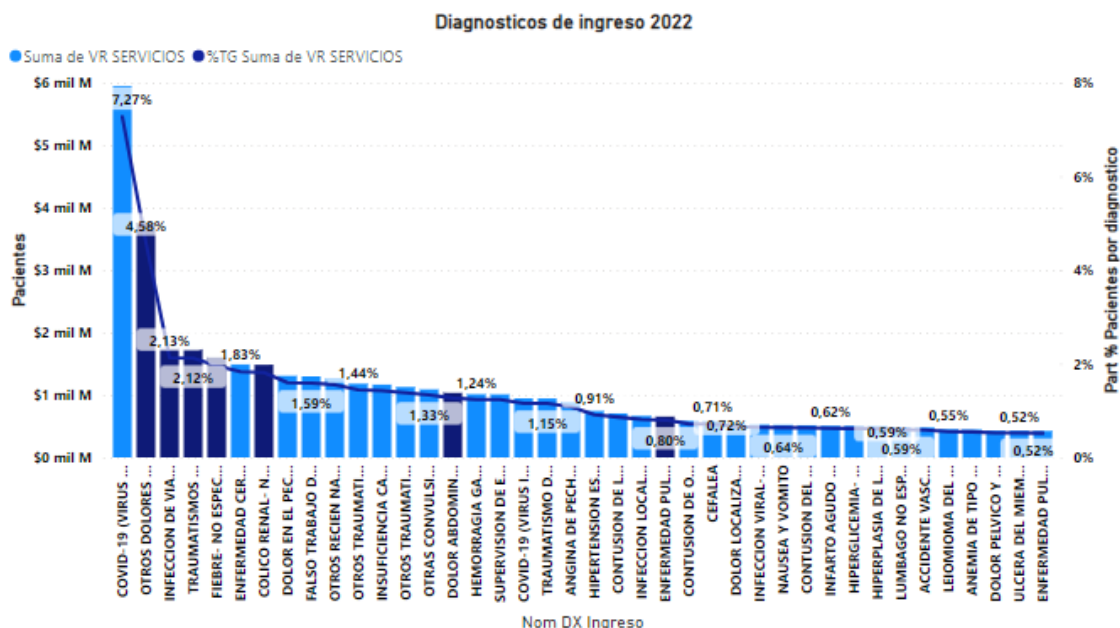
Tabla 2. Diagnósticos representativos año 2022

Nom DX Ingreso	Pacientes	VR SERVICIOS	% part VR SERVICIOS
OTROS DOLORES ABDOMINALES Y LOS NO ESPECIFICADOS	3751	\$ 5.682.954.447	7,78%
FIEBRE- NO ESPECIFICADA	4826	\$ 2.397.604.214	3,28%
TRAUMATISMOS MULTIPLES- NO ESPECIFICADOS	222	\$ 1.886.595.296	2,58%
ENFERMEDAD PULMONAR OBSTRUCTIVA CRONICA CON EXACERBACION AGUDA- NO ESPECIFICADA	427	\$ 1.605.646.780	2,20%
DOLOR ABDOMINAL LOCALIZADO EN PARTE SUPERIOR	846	\$ 1.364.359.024	1,87%
INFECCION DE VIAS URINARIAS- SITIO NO ESPECIFICADO	1103	\$ 1.224.406.879	1,68%

Fuente: Construcción propia

Para el año 2022, a pesar de que el COVID-19 no es tan crítico, continúa liderando en el comportamiento de los diagnósticos de ingreso, a pesar de ello, es importante analizar este año ya que se vuelve a una nueva normalidad donde la tendencia del COVID disminuyó considerablemente a lo largo del año. Por lo anterior se presenta el siguiente comportamiento:

Figura 6. Diagnóstico de ingreso 2022



Fuente: Construcción propia

A partir de esta gráfica, se puede evidenciar que para el año 2019 los diagnósticos (patologías clínicas) con mayor participación fueron los siguientes:

Tabla 3. Diagnósticos representativos año 2019

Nom DX Ingreso	Pacientes	VR SERVICIOS	% part VR SERVICIOS
COVID-19 (VIRUS NO IDENTIFICADO)	2932	\$ 5.949.346.477	7,27%
OTROS DOLORES ABDOMINALES Y LOS NO ESPECIFICADOS	2776	\$ 3.750.200.934	4,58%
INFECCION DE VIAS URINARIAS- SITIO NO ESPECIFICADO	1489	\$ 1.738.774.417	2,13%
TRAUMATISMOS MULTIPLES- NO ESPECIFICADOS	212	\$ 1.732.600.461	2,12%
FIEBRE- NO ESPECIFICADA	3286	\$ 1.600.416.209	1,96%
ENFERMEDAD CEREBROVASCULAR- NO ESPECIFICADA	277	\$ 1.496.074.410	1,83%
COLICO RENAL- NO ESPECIFICADO	1545	\$ 1.484.432.257	1,81%

Fuente: Construcción propia

Teniendo en cuenta lo anterior, se procede a definir las patologías más representativas para realizar el análisis predictivo utilizando modelos de análisis de datos, las cuales son aquellas que fueron las más representativas en el año 2019 y 2022:

Tabla 4. Diagnósticos representativos año 2019 - 2022

Nom DX Ingreso	2019			2022		
	Pacientes	VR SERVICIOS	% part VR SERVICIOS	Pacientes	VR SERVICIOS	% part VR SERVICIOS
OTROS DOLORES ABDOMINALES Y LOS NO ESPECIFICADOS	3751	\$ 5.682.954.447	7,78%	2776	\$ 3.750.200.934	4,58%
FIEBRE- NO ESPECIFICADA	4826	\$ 2.397.604.214	3,28%	3286	\$ 1.600.416.209	1,96%
TRAUMATISMOS MULTIPLES- NO ESPECIFICADOS	222	\$ 1.886.595.296	2,58%	212	\$ 1.732.600.461	2,12%

ENFERMEDAD PULMONAR OBSTRUCTIVA CRONICA CON EXACERBACION AGUDA- NO ESPECIFICADA	427	\$ 1.605.646.780	2,20%	135	\$ 653.552.085	0,80%
DOLOR ABDOMINAL LOCALIZADO EN PARTE SUPERIOR	846	\$ 1.364.359.024	1,87%	635	\$ 1.036.287.415	1,27%
INFECCION DE VIAS URINARIAS- SITIO NO ESPECIFICADO	1103	\$ 1.224.406.879	1,68%	1489	\$ 1.738.774.417	2,13%
COLICO RENAL- NO ESPECIFICADO	1256	\$ 1.040.062.475	1,42%	1545	\$ 1.484.432.257	1,81%

Fuente: Construcción propia

4.4 Conjunto de datos final

A partir del conjunto de datos seleccionados y variables definidas, se segmentó los datos por cada una de las diferentes patologías seleccionadas, y se crean 7 diferentes data set, cada una del periodo de tiempo octubre 2018 – diciembre 2022 (Omitiendo 2020 – 2021).

4.5 Definición de variables

Como la base de datos inicial cuenta con un total de 39 variables, se procede a seleccionar las variables más significativas para el estudio, las cuales se describen brevemente a continuación:

Admisión: número único de identificación de un paciente que ingresa a la institución y queda como registro en el sistema de información. Para efectos de la base de datos es para identificar cada registro de manera independiente.

Ciudad: Ciudad de residencia del paciente. Variable descriptiva para ampliar información de la procedencia de los pacientes que ingresan a la institución.

Edad: Edad en años del paciente al momento de generación del reporte. Variable

descriptiva para ampliar información de la edad de los pacientes que ingresan a la institución.

Género: Género Masculino o femenino del paciente. Variable descriptiva para ampliar información del género de los pacientes que ingresan a la institución.

Fecha ingreso: Fecha de ingreso del paciente a la institución. Este dato es fundamental en el estudio ya que es la variable independiente que nos mostrara el comportamiento de los ingresos intrahospitalarios.

Fecha egreso: Fecha de egreso del paciente a la institución. Variable descriptiva para ampliar información de la fecha de egreso de los pacientes que ingresan a la institución.

Dx ingreso: Código del diagnóstico principal definido para el ingreso del paciente. Este dato es fundamental en el estudio ya que es la variable dependiente que nos mostrara el comportamiento de los ingresos intrahospitalarios por diagnóstico de ingreso.

Nom Dx ingreso: Descripción del diagnóstico principal definido para el ingreso del paciente. Este dato es fundamental en el estudio ya que es la variable dependiente que nos mostrara el comportamiento de los ingresos intrahospitalarios por diagnóstico de ingreso.

Dx egreso: Código del diagnóstico principal definido para el egreso del paciente. Variable descriptiva para ampliar información del código del diagnóstico de egreso de los pacientes que ingresan a la institución.

Nom Dx egreso: Descripción del diagnóstico principal definido para el egreso del paciente. Variable descriptiva para ampliar información de los diagnósticos de egreso de los pacientes que ingresan a la institución.

Vr servicios: Es la suma total de todos los servicios prestados al paciente durante su hospitalización en la institución. Variable fundamental para analizar el comportamiento de cada diagnóstico de ingreso respecto a cuanto impacta financieramente los ingresos de la institución.

Pabellón: Servicio donde estuvo hospitalizado el paciente antes de su egreso de la institución. Variable descriptiva para ampliar información del pabellón de egreso de los pacientes que ingresan a la institución.

Empresa: Entidad a la que está afiliado el paciente al momento del ingreso a la institución. Variable descriptiva para ampliar información de la entidad responsable

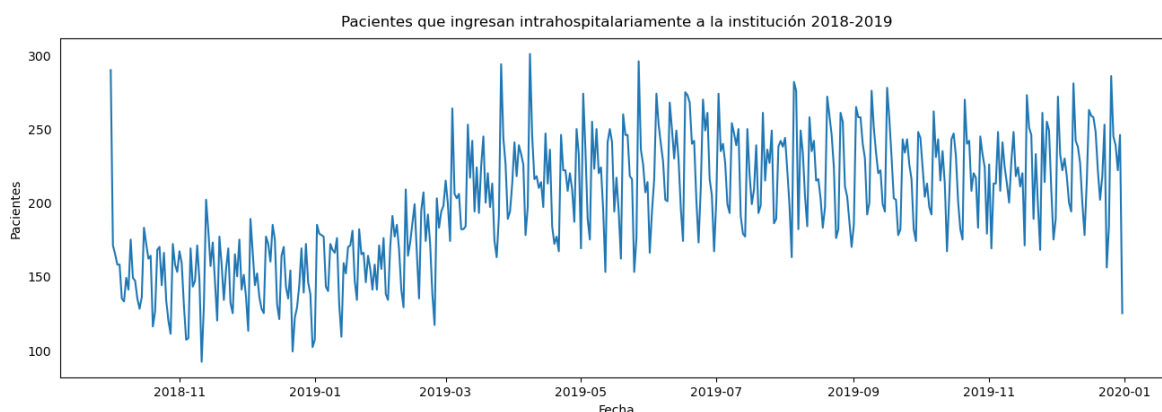
de pago de los pacientes que ingresan a la institución.

4.6 Construcción de nuevas variables

Con el objetivo de realizar el modelamiento de los datos, se procede a crear para cada patología un set de datos final, por ello se estructuran los datos en dos variables claves que son (Fecha ingreso y cantidad de pacientes ingresados por día). A partir de estas dos variables se procederá a estudiar el comportamiento de la morbilidad intrahospitalaria de las patologías críticas.

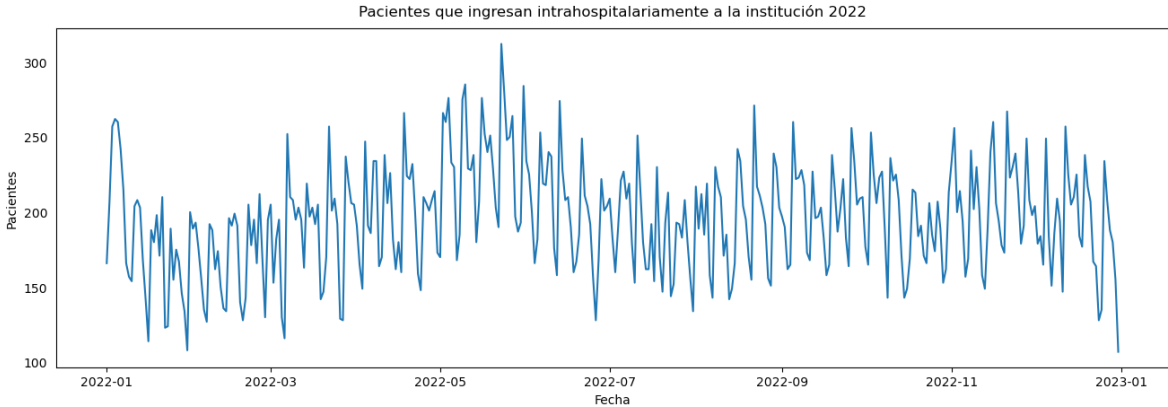
A partir de la definición de estas 7 patologías más críticas, paso a realizar un análisis descriptivo y de comportamiento en el tiempo de pacientes ingresados del total de patologías y por cada una de ellas:

Figura 7. Total de pacientes ingresados intrahospitalariamente diariamente Octubre 2018 – 2019



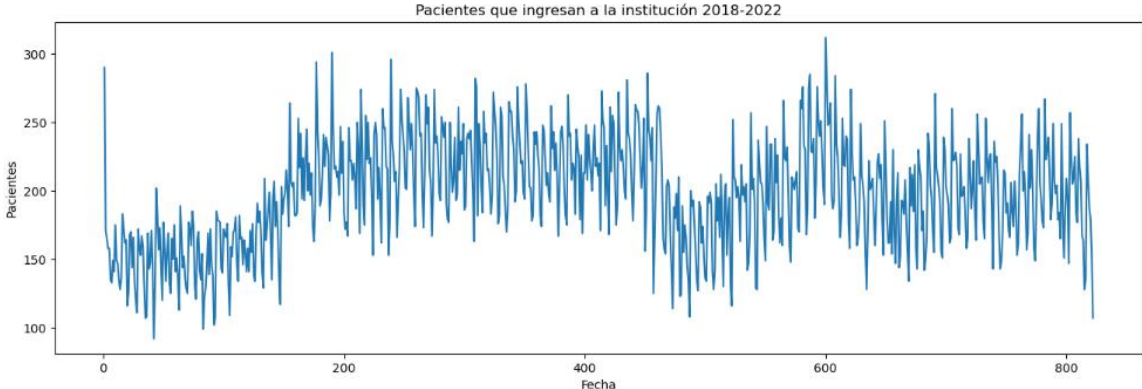
Fuente: Construcción propia

Figura 8. Total de pacientes ingresados intrahospitalariamente diariamente 2022



Fuente: Construcción propia

Figura 9. Total de pacientes ingresados intrahospitalariamente diariamente 2018-2022



Fuente: Construcción propia

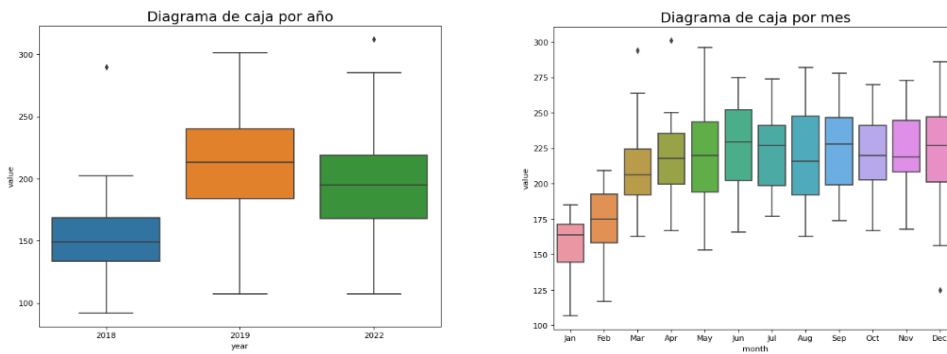
Tabla 5. Resumen estadístico

	value
count	822.000000
mean	197.559611
std	39.800363
min	92.000000
25%	169.000000
50%	197.500000
75%	226.000000
max	312.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 4 se obtiene una serie temporal con 822 datos, media de 197 pacientes diarios, con un mínimo de 92 pacientes y máximo de 312 pacientes diarios.

Figura 10. Diagrama de caja



Fuente: Construcción propia

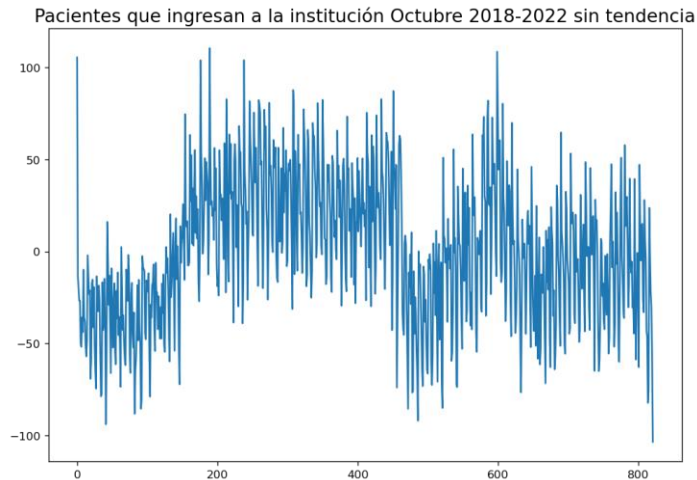
Probando estacionalidad

ADF Statistic: -2.2011023071243483
 p-value: 0.205845626241973
 Critical Values:
 1%, -3.438550508267578
 Critical Values:
 5%, -2.8651595688281253
 Critical Values:
 10%, -2.5686973890625

A probar estacionalidad en la serie se obtuvo que p-value es $>0,05$ por lo tanto se acepta la hipótesis nula lo que indica que la serie no es estacionaria

Eliminando tendencia

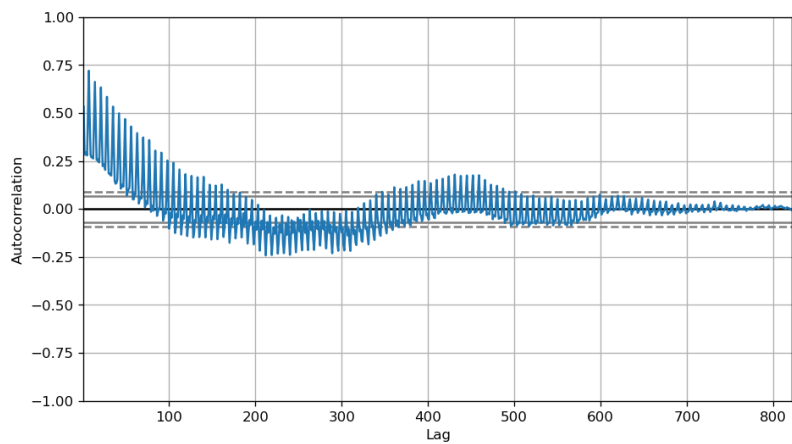
Figura 11. Pacientes que ingresan a la institución Octubre 2018-2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

Figura 12. Autocorrelación

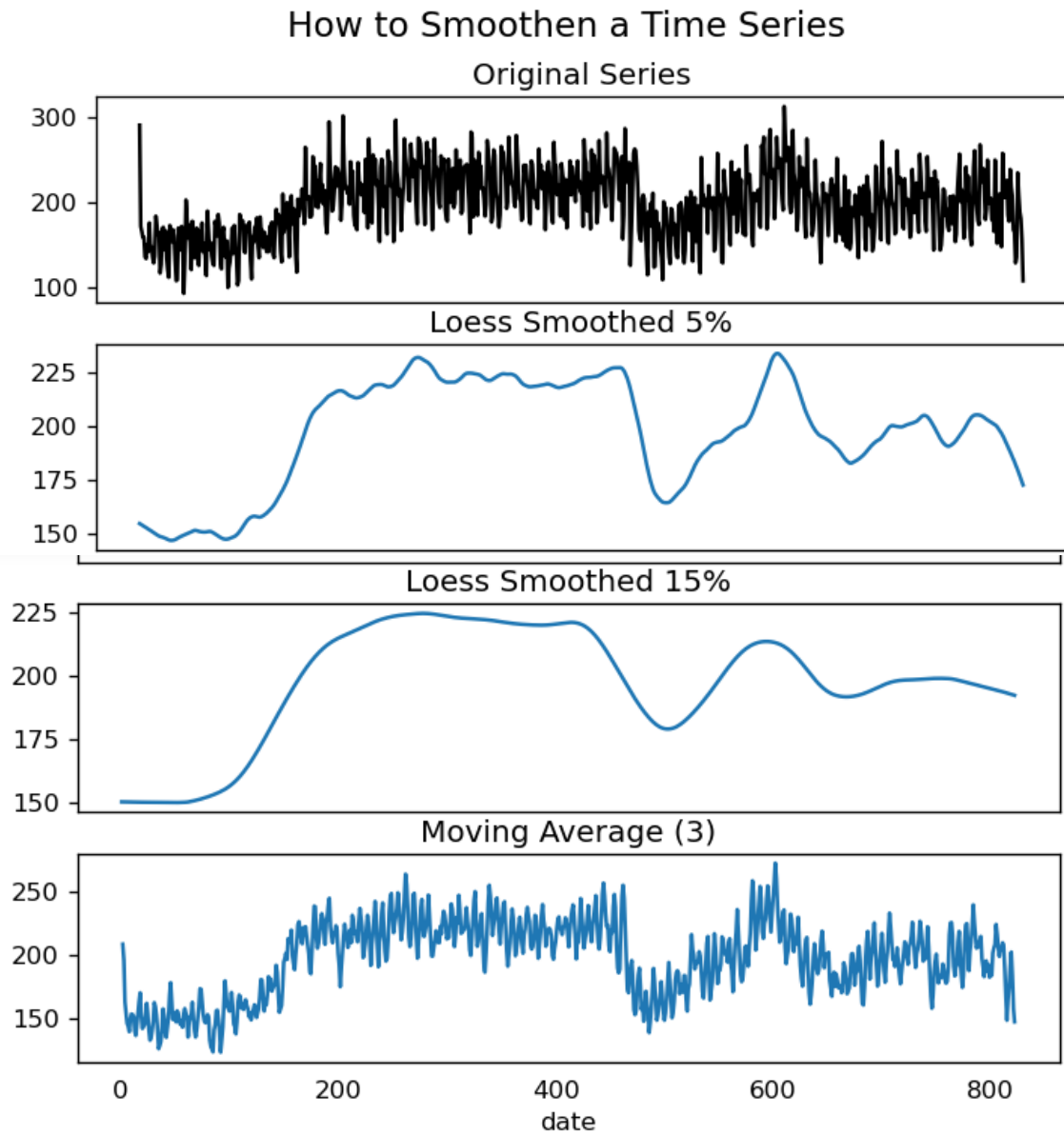


Fuente: Construcción propia

Función de autocorrelación

Suavización de una serie temporal

Figura 13. Suavización de la serie temporal

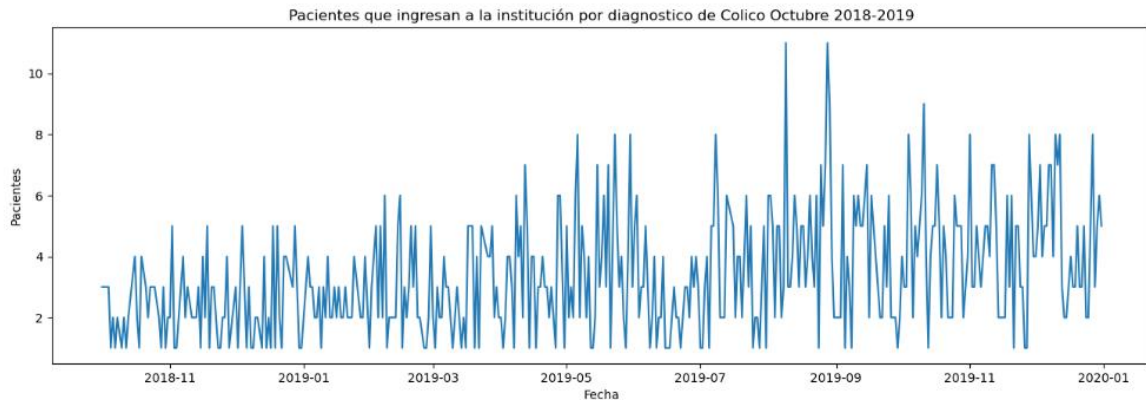


Fuente: Construcción propia

A continuación, se analizará cada uno de los diagnósticos:

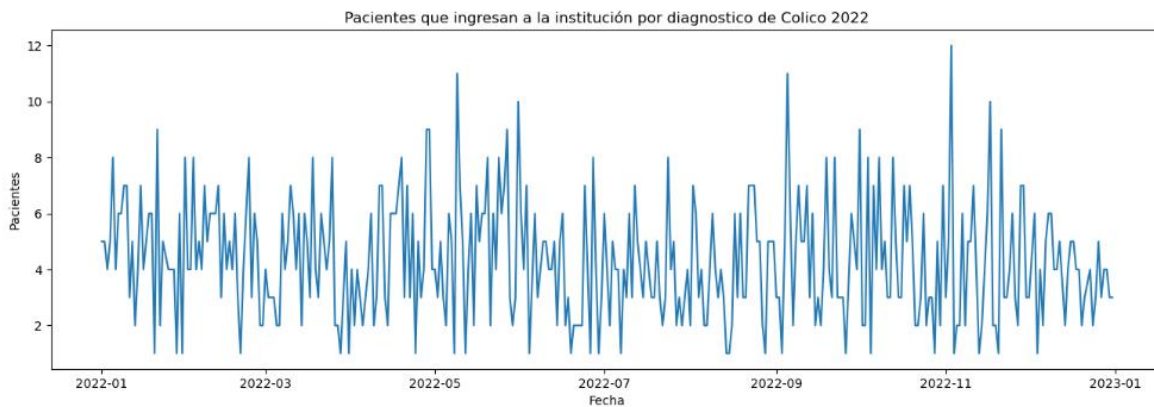
Cólico

Figura 14. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2019



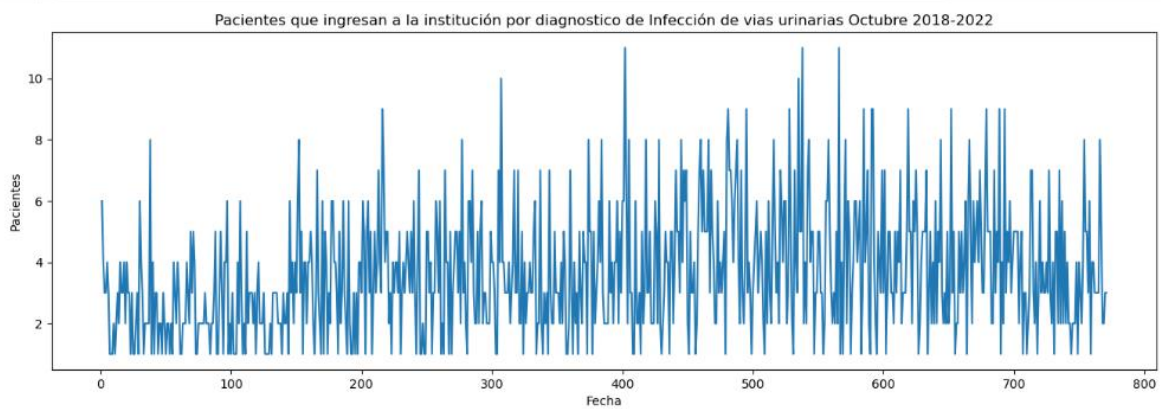
Fuente: Construcción propia

Figura 15. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 16. Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

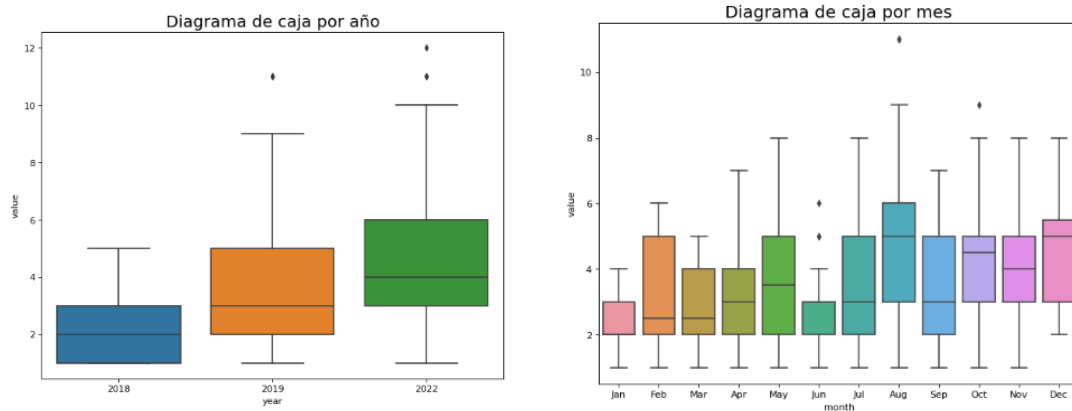
Tabla 6. Resumen estadístico del Cólico

	value
count	779.000000
mean	3.818999
std	2.059987
min	1.000000
25%	2.000000
50%	4.000000
75%	5.000000
max	12.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 5 se obtiene una serie temporal con 779 datos, media de 4 pacientes diarios, con un mínimo de 1 paciente y máximo de 12 pacientes diarios.

Figura 17. Diagrama de caja por Cólico



Fuente: Construcción propia

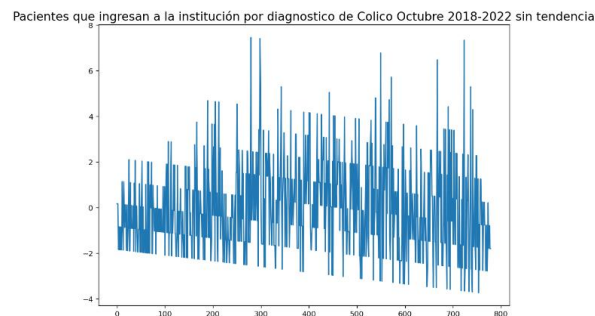
Probando estacionalidad

ADF Statistic: -4.512155307562413
 p-value: 0.00018693088035815224
 Critical Values:
 1%, -3.438926964986094
 Critical Values:
 5%, -2.8653254941943174
 Critical Values:
 10%, -2.5687857802554572

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza hipótesis nula lo que indica que la serie es estacionaria

Eliminando tendencia

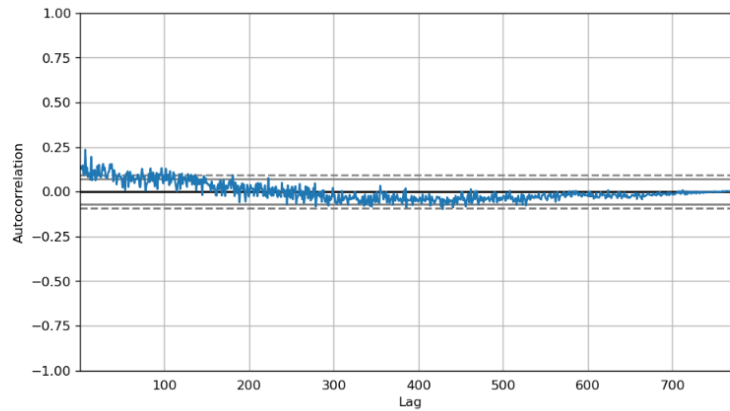
Figura 18. Pacientes que ingresan a la institución por cólico Octubre 2018-2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

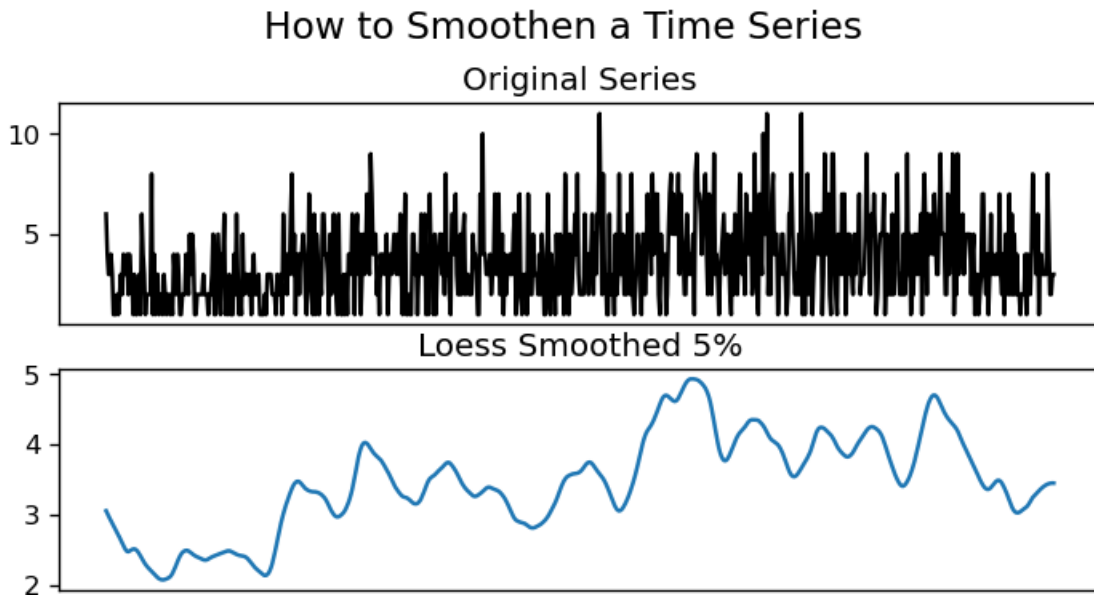
Figura 19. Autocorrelación por cólico

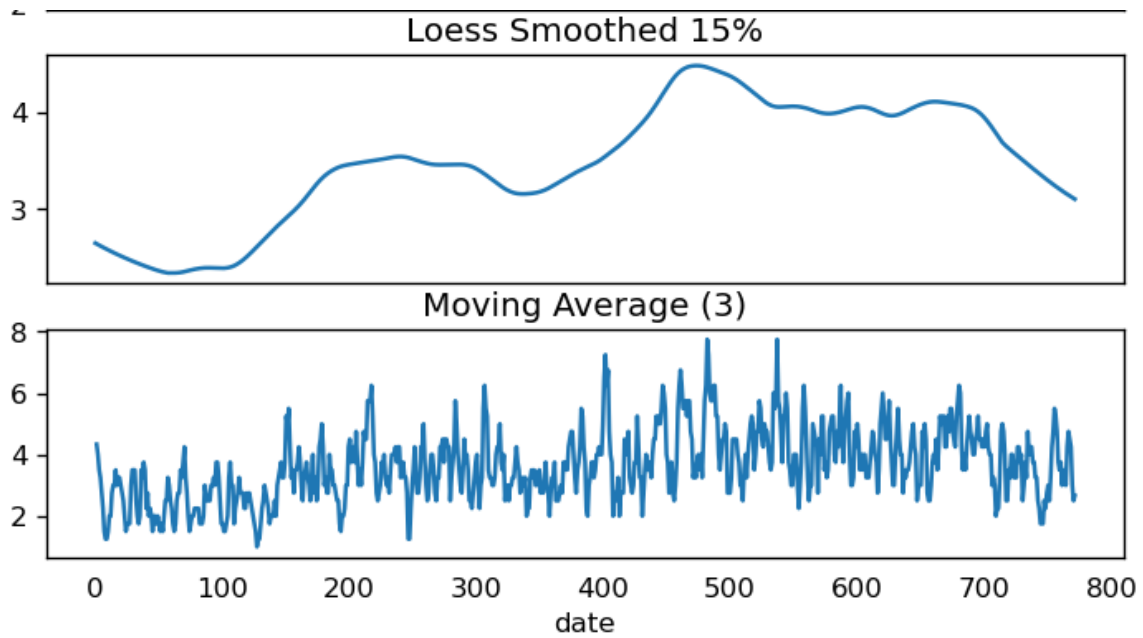


Fuente: Construcción propia

Suavización de una serie temporal

Figura 20. Suavización de una serie temporal del cólico

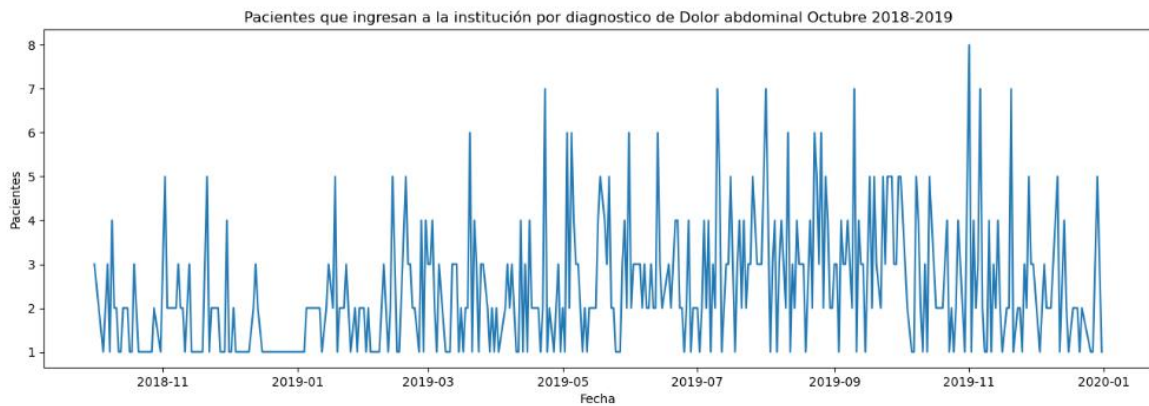




Fuente: Construcción propia

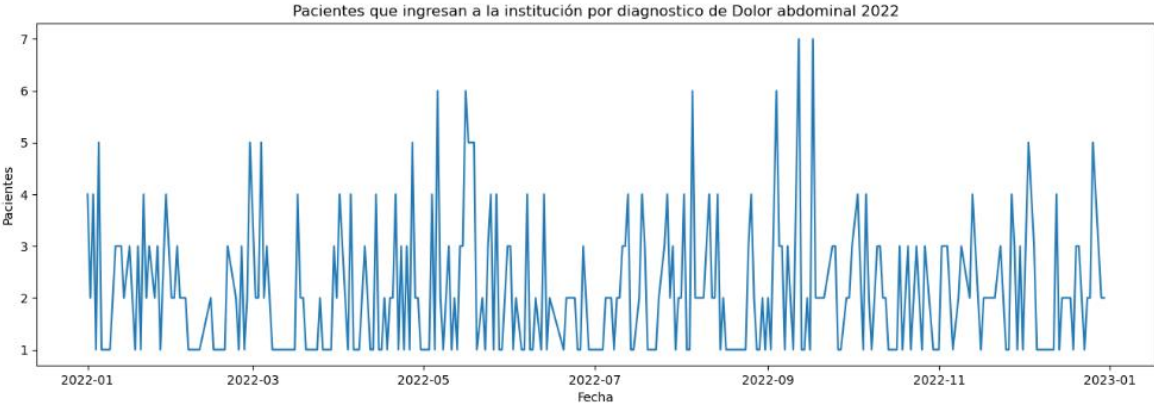
Dolor Abdominal

Figura 21. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2019



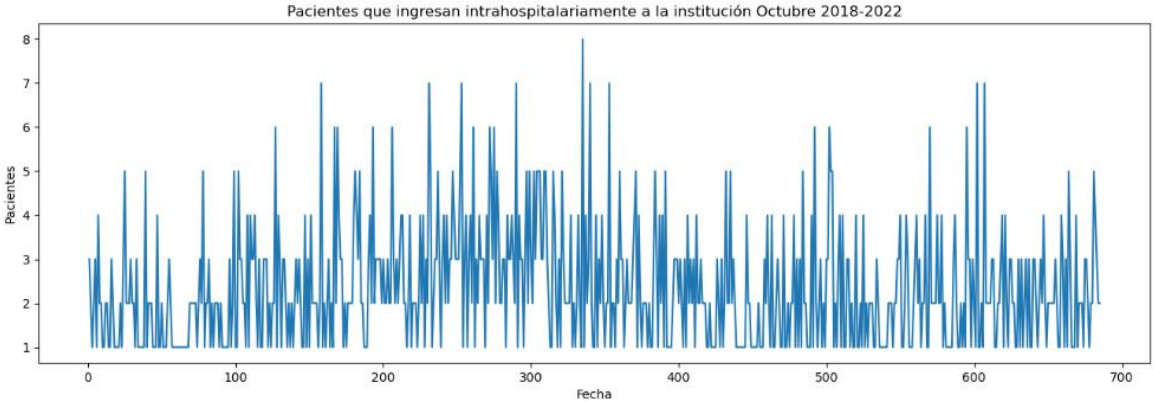
Fuente: Construcción propia

Figura 22. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 23. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

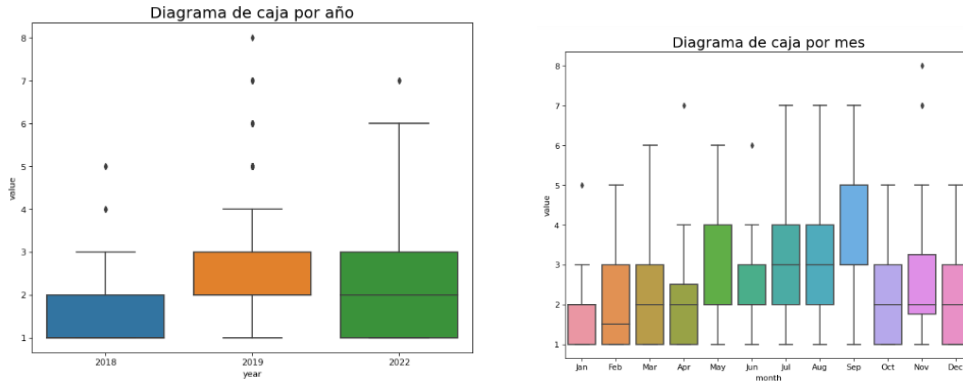
Tabla 7. Resumen estadístico del dolor abdominal

	value
count	685.000000
mean	2.325547
std	1.362833
min	1.000000
25%	1.000000
50%	2.000000
75%	3.000000
max	8.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 6 se obtiene una serie temporal con 685 datos, media de 2 pacientes diarios, con un mínimo de 1 paciente y máximo de 8 pacientes diarios.

Figura 24. Diagrama de caja del Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

Probando estacionalidad

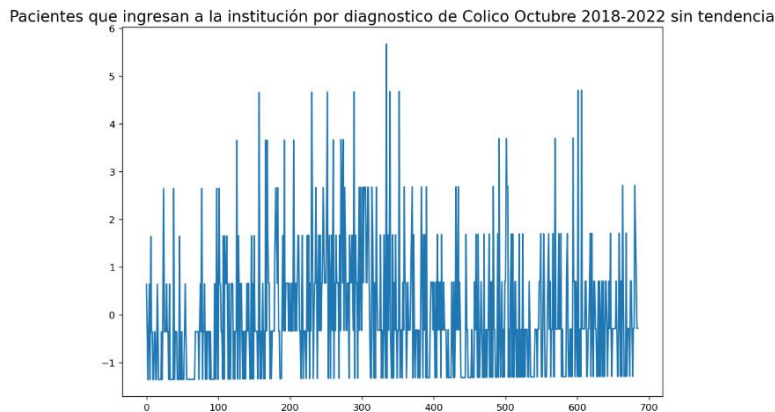
ADF Statistic: -4.045473126572237
 p-value: 0.0011902520381606285
 Critical Values:
 1%, -3.4401331487581754
 Critical Values:
 5%, -2.865856987951189

Critical Values:
10%, -2.569068936347423

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza hipótesis nula lo que indica que la serie es estacionaria.

Eliminando tendencia

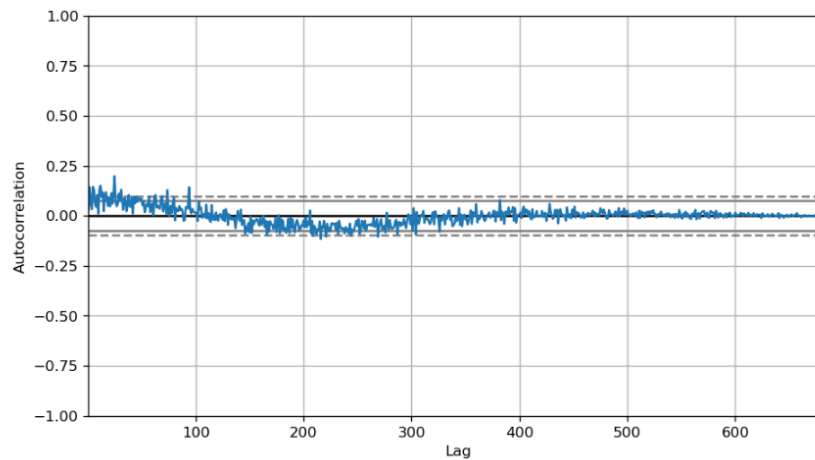
Figura 25. Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

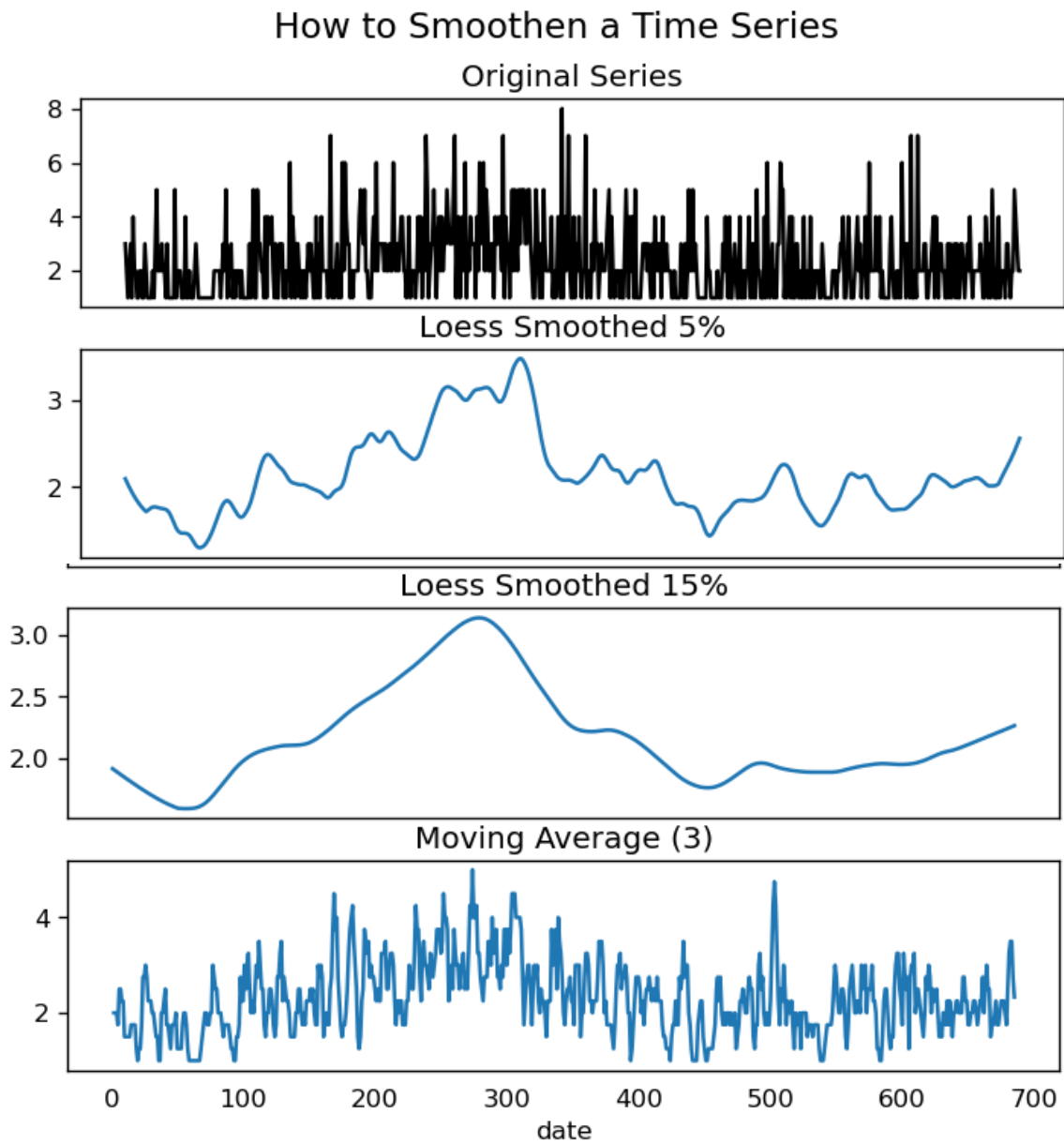
Figura 26. Autocorrelación por Dolor Abdominal



Fuente: Construcción propia

Suavización de una serie temporal

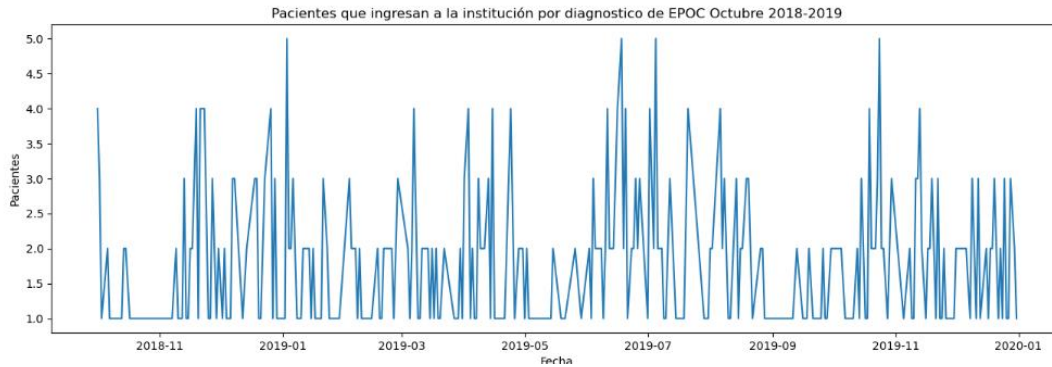
Figura 27. Suavización de una serie temporal por Dolor Abdominal



Fuente: Construcción propia

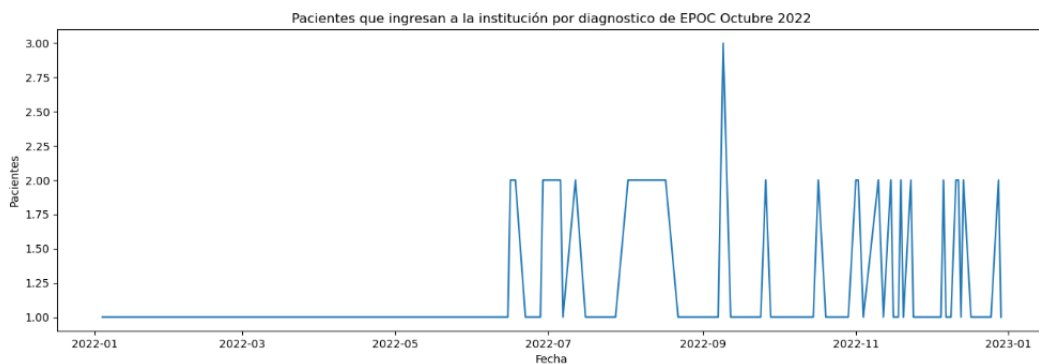
EPOC

Figura 28. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2019



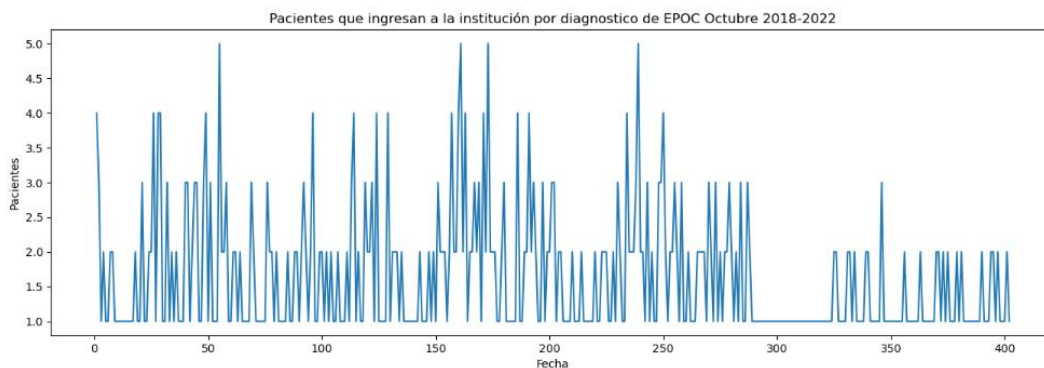
Fuente: Construcción propia

Figura 29. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 30. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

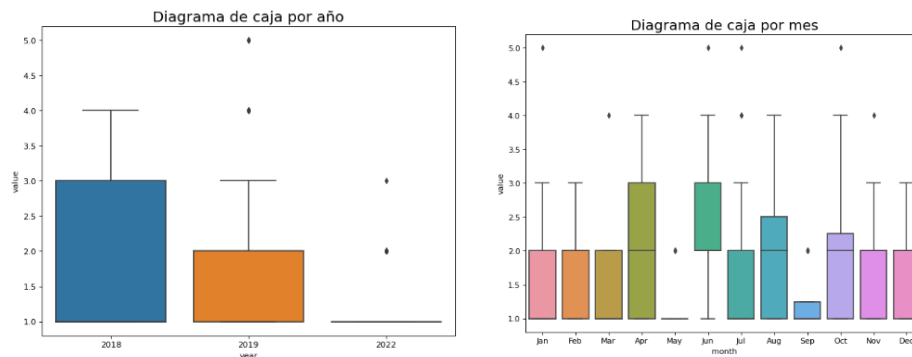
Tabla 8. Resumen estadístico del EPOC

	value
count	402.000000
mean	1.631841
std	0.889832
min	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
max	5.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 7 se obtiene una serie temporal con 402 datos, media de 1 paciente diario, con un mínimo de 1 paciente y máximo de 5 pacientes diarios.

Figura 31. Diagrama de caja por EPOC



Fuente: Construcción propia

Probando estacionalidad

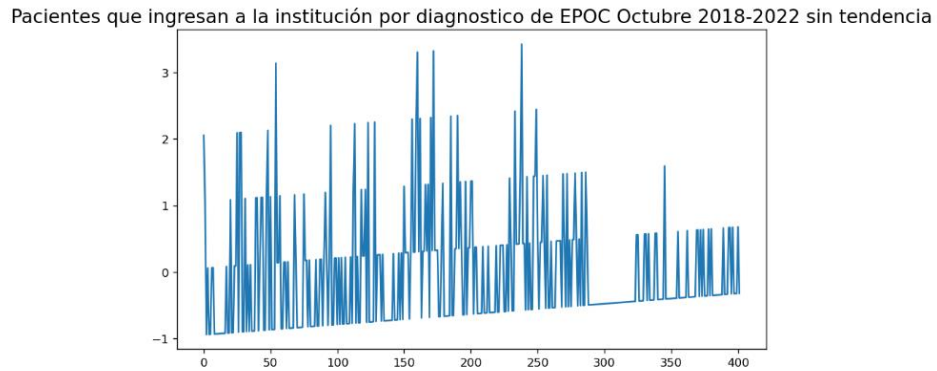
ADF Statistic: -4.24050682948909
 p-value: 0.0005622503849485984
 Critical Values:
 1%, -3.4470566389664703
 Critical Values:
 5%, -2.8689037160476016
 Critical Values:

10%, -2.570692663557422

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza hipótesis nula lo que indica que la serie es estacionaria.

Eliminando tendencia

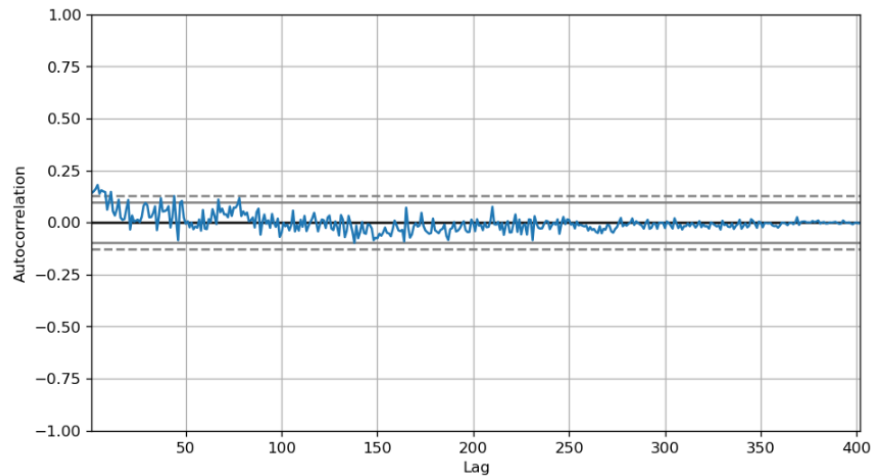
Figura 32. Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia



Fuente: Construcción propia

Figura 33. Autocorrelación por EPOC

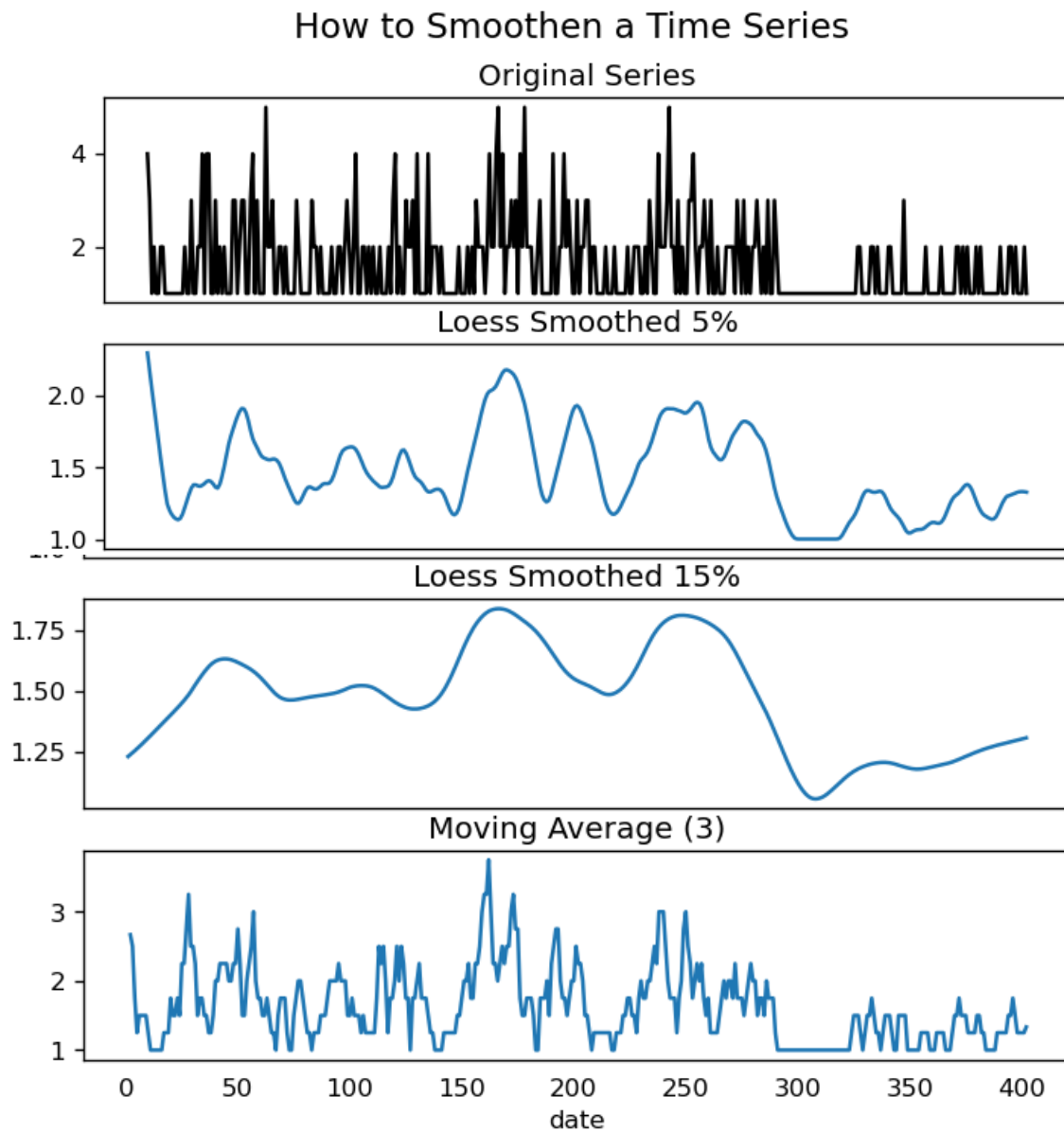
Autocorrelación



Fuente: Construcción propia

Suavización de una serie temporal

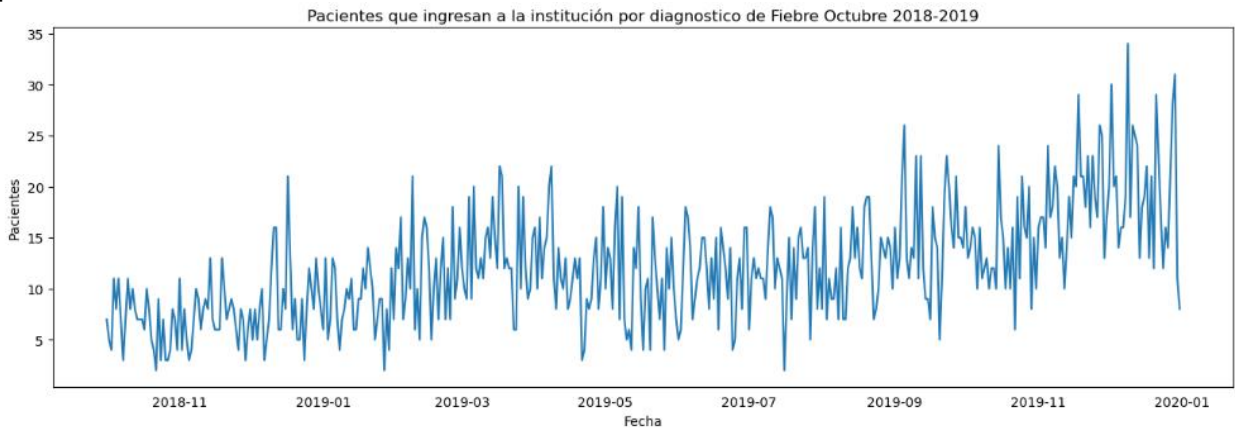
Figura 34. Suavización de una serie temporal por EPOC



Fuente: Construcción propia

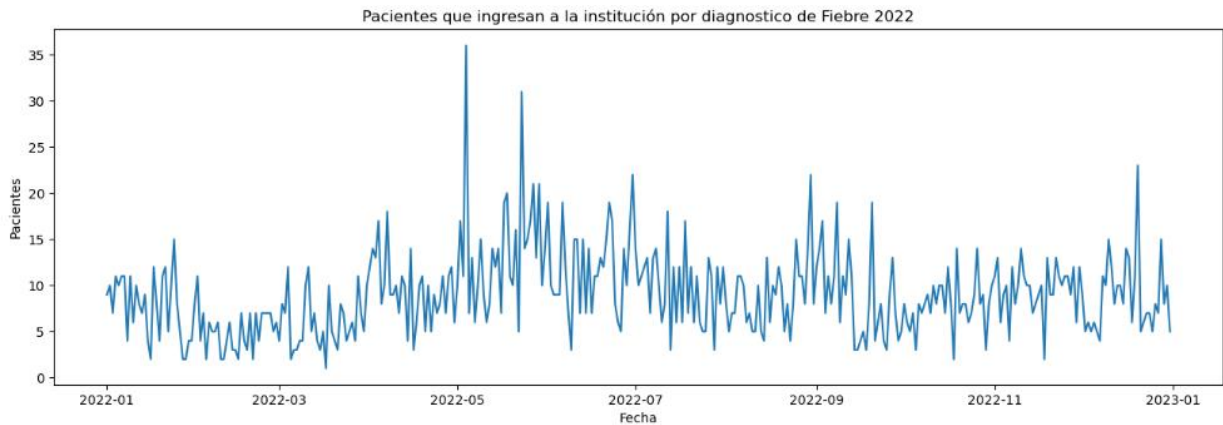
Fiebre

Figura 35. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2019



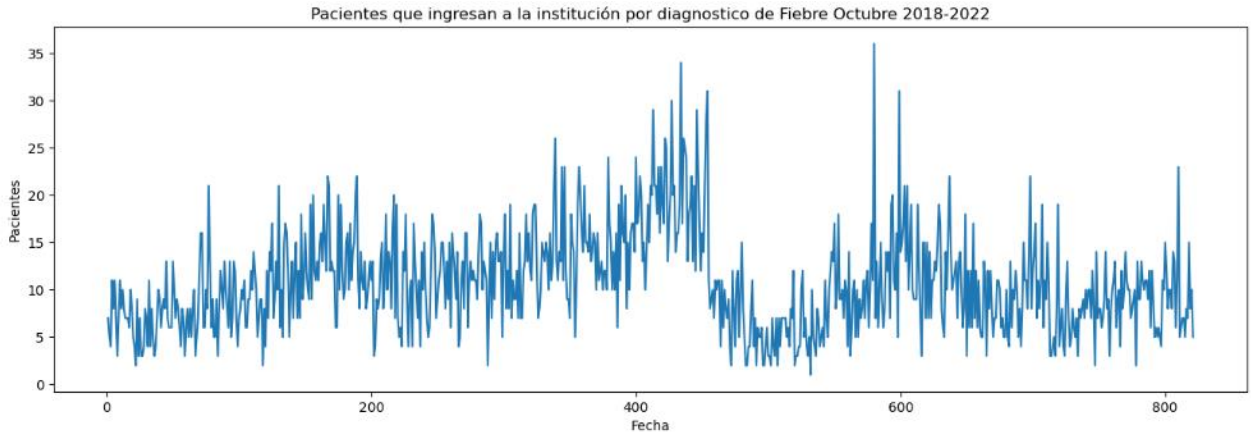
Fuente: Construcción propia

Figura 36. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 37. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

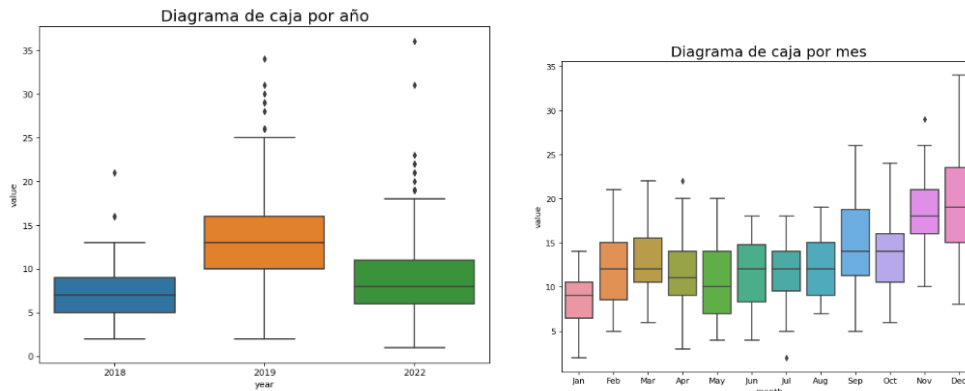
Tabla 9. Resumen estadístico de Fiebre

	value
count	821.000000
mean	10.718636
std	5.310597
min	1.000000
25%	7.000000
50%	10.000000
75%	14.000000
max	36.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 8 se obtiene una serie temporal con 821 datos, media de 11 paciente diario, con un mínimo de 1 paciente y máximo de 36 pacientes diarios.

Figura 38. Diagrama de caja por FIEBRE



Fuente: Construcción propia

Probando estacionalidad

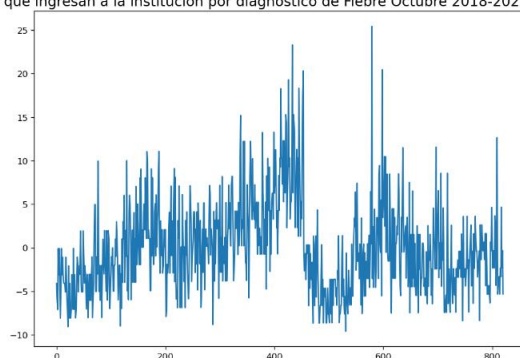
ADF Statistic: -2.5462775718865216
 p-value: 0.10461654214939137
 Critical Values:
 1%, -3.4385299921076995
 Critical Values:
 5%, -2.8651505256266865
 Critical Values:
 10%, -2.568692571687987

A probar estacionalidad en la serie se obtuvo que p-value es $>0,05$ por lo tanto se acepta hipótesis nula lo que indica que la serie no es estacionaria.

Eliminando tendencia

Figura 39. Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia

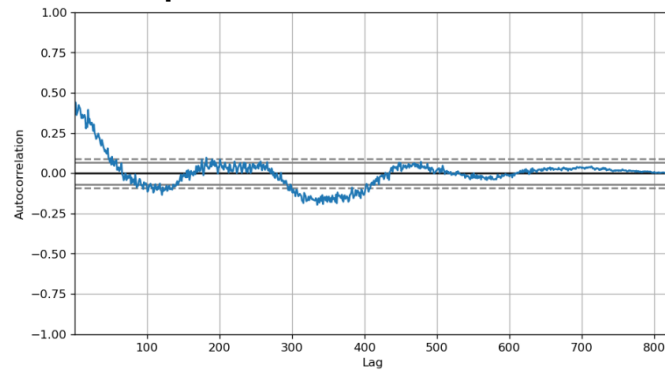
Pacientes que ingresan a la institución por diagnostico de Fiebre Octubre 2018-2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

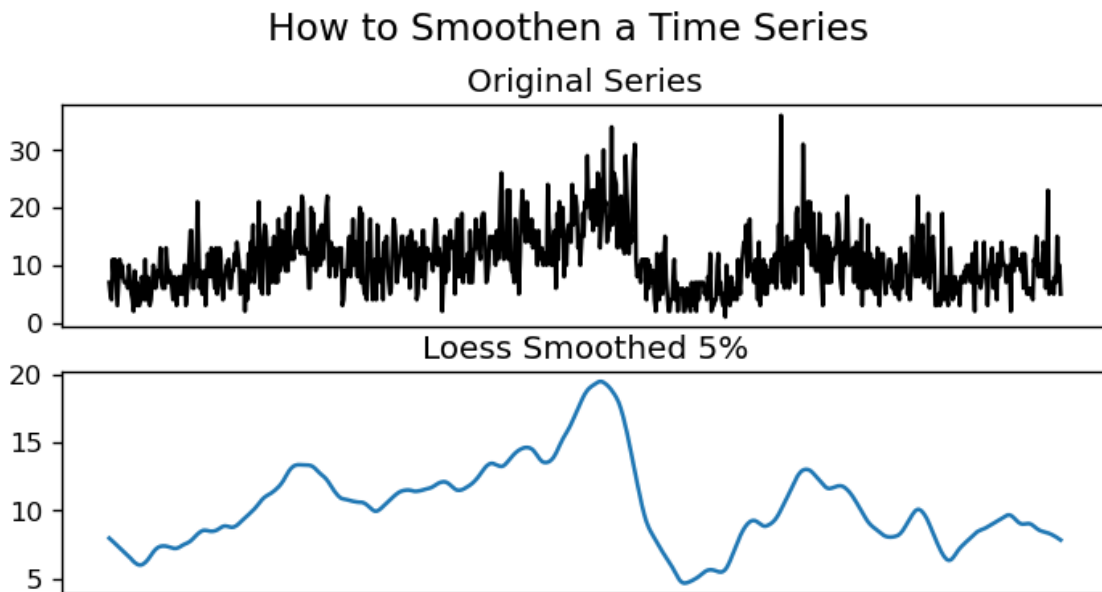
Figura 40. Autocorrelación por FIEBRE

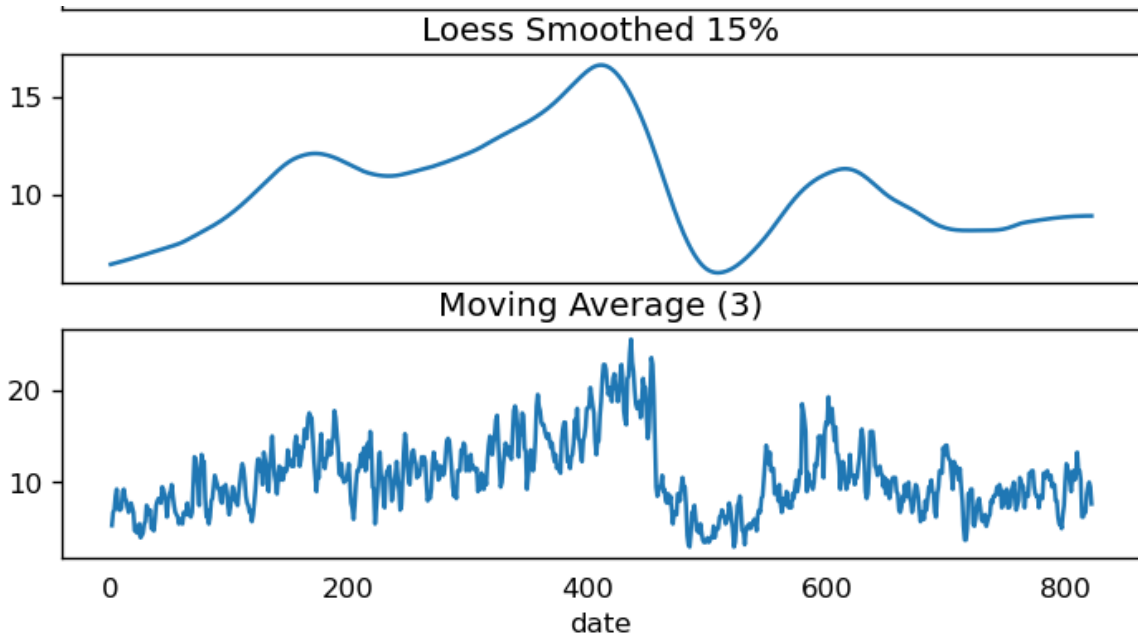


Fuente: Construcción propia

Suavización de una serie temporal

Figura 41. Suavización de una serie temporal por FIEBRE

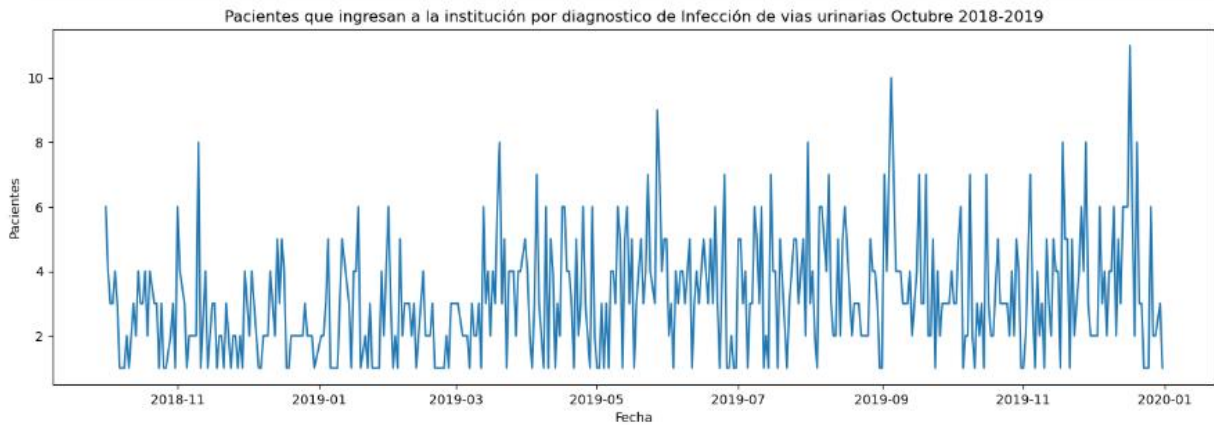




Fuente: Construcción propia

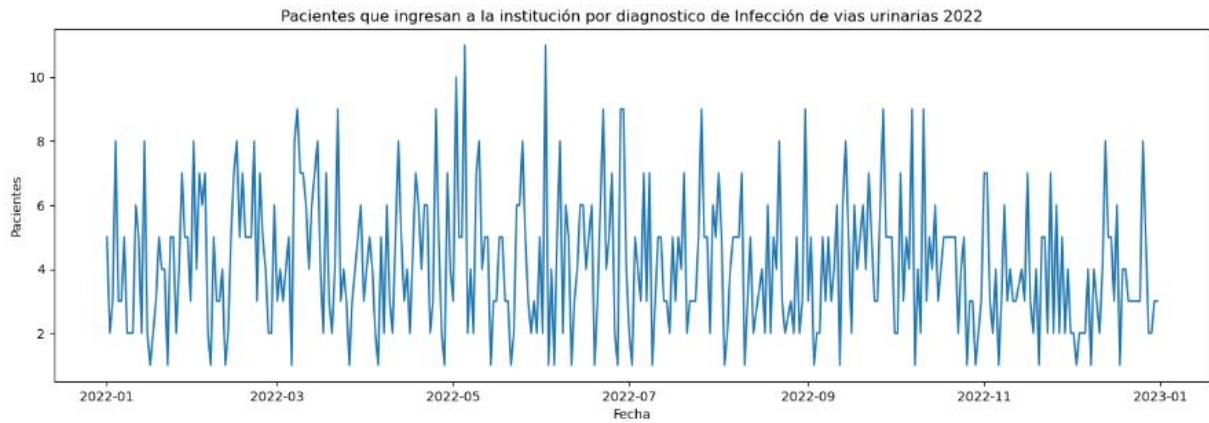
Infección de vías urinarias

Figura 42. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2019



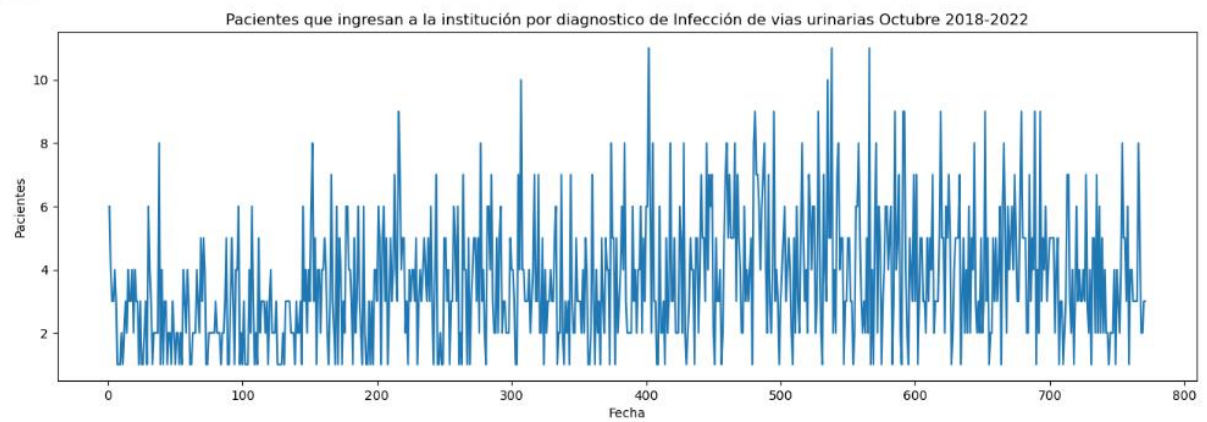
Fuente: Construcción propia

Figura 43. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 44. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

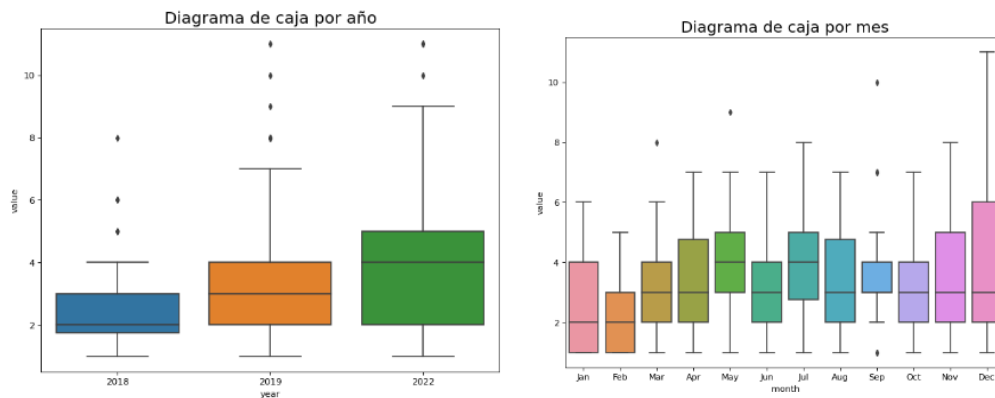
Tabla 10. Resumen estadístico de infección de vías urinarias

	value
count	771.000000
mean	3.631647
std	2.023687
min	1.000000
25%	2.000000
50%	3.000000
75%	5.000000
max	11.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 9 se obtiene una serie temporal con 771 datos, media de 3 paciente diario, con un mínimo de 1 paciente y máximo de 11 pacientes diarios.

Figura 45. Diagrama de caja para infección de vías urinarias



Fuente: Construcción propia

Probando estacionalidad

ADF Statistic: -3.0185638193443736
 p-value: 0.033196766636863495
 Critical Values:
 1%, -3.439099096730074
 Critical Values:
 5%, -2.8654013553540745
 Critical Values:

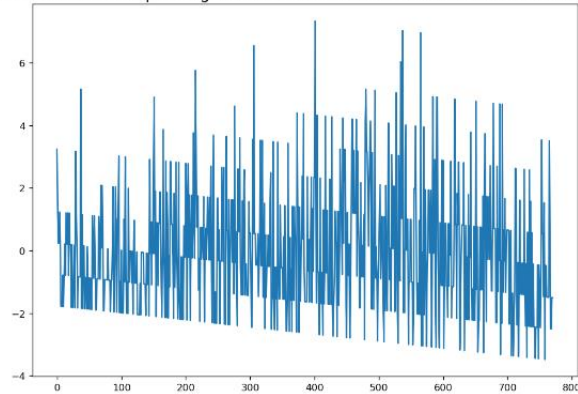
10%, -2.568826193777778

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza la hipótesis nula lo que indica que la serie es estacionaria.

Eliminando tendencia

Figura 46. Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia

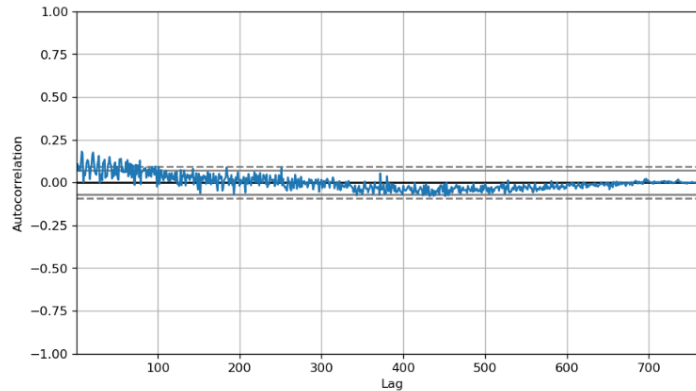
Pacientes que ingresan a la institución por diagnostico de Infección de vías Urinarias Octubre 2018-2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

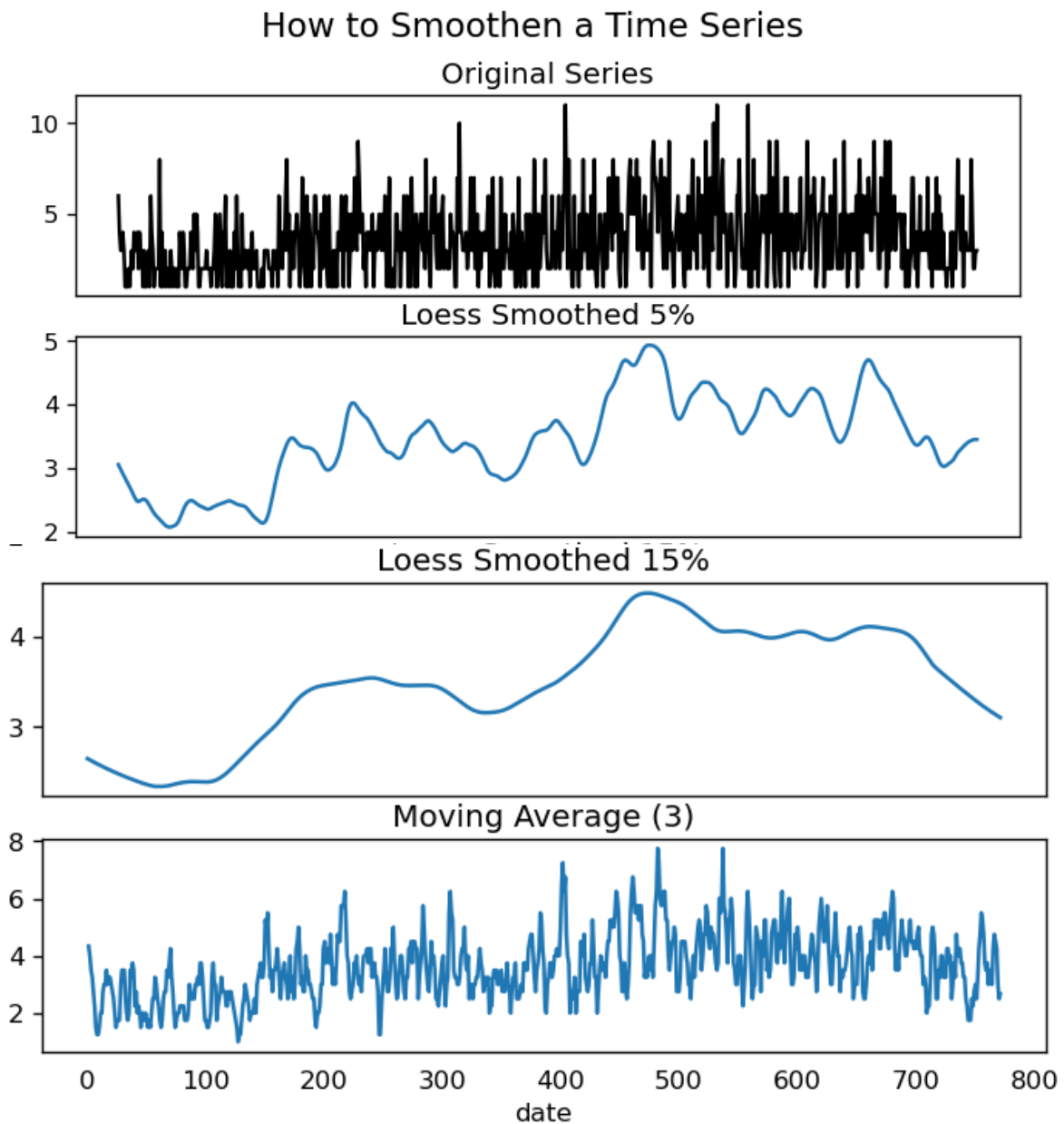
Figura 47. Autocorrelación para infección de vías urinarias



Fuente: Construcción propia

Suavización de la serie temporal

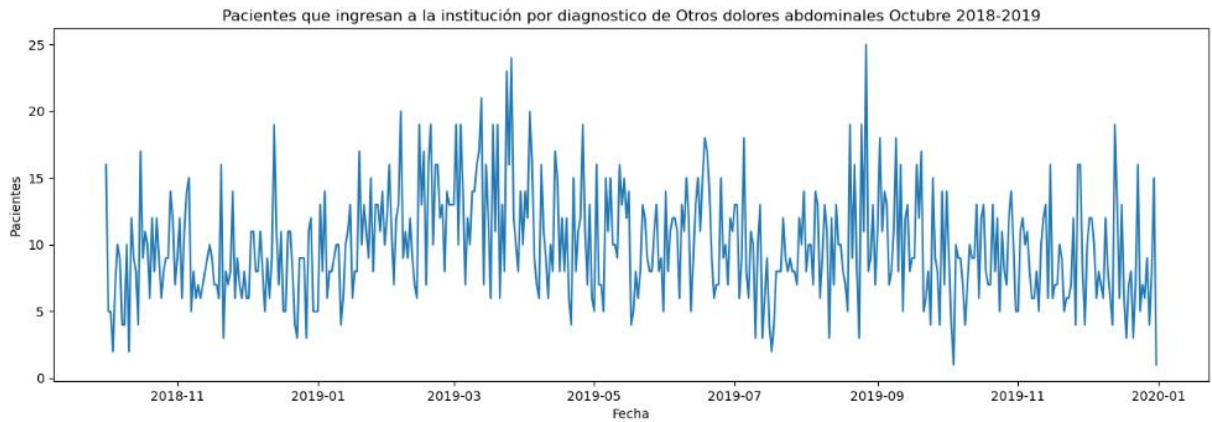
Figura 48. Suavización de la serie temporal para infección de vías urinarias



Fuente: Construcción propia

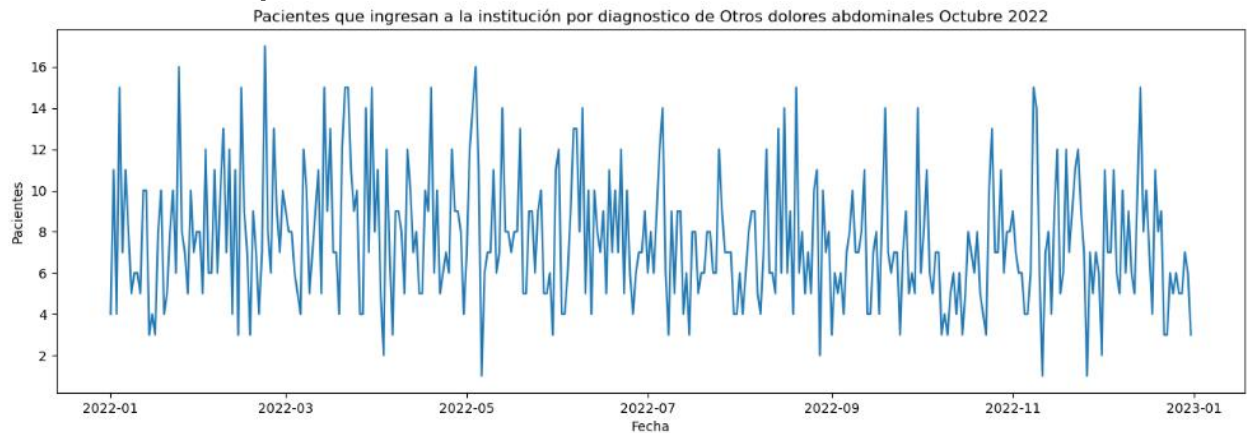
Otros dolores abdominales

Figura 49. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2019



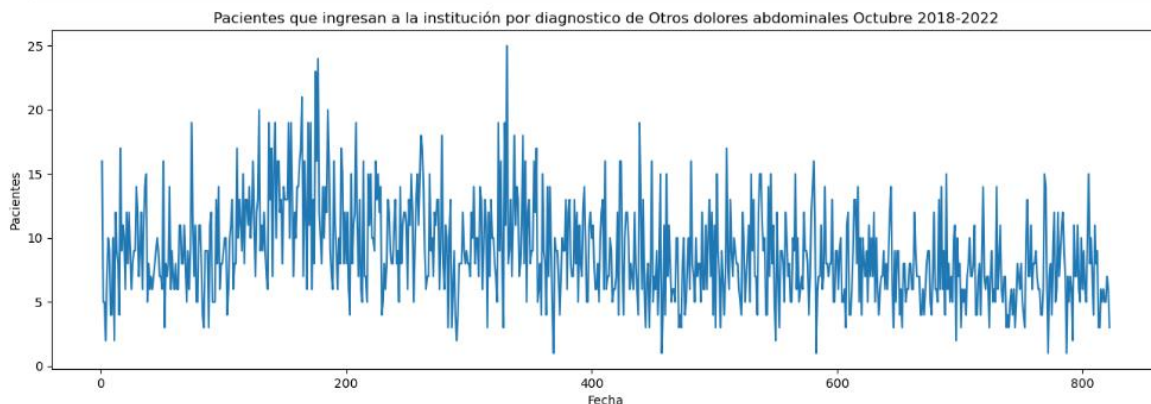
Fuente: Construcción propia

Figura 50. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 51. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

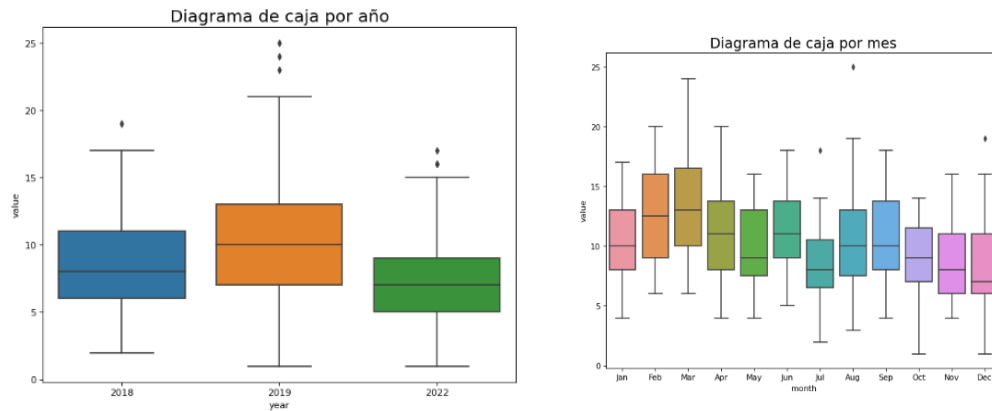
Tabla 11. Resumen estadístico de Otros dolores abdominales

	value
count	822.000000
mean	8.884428
std	3.847270
min	1.000000
25%	6.000000
50%	8.000000
75%	11.000000
max	25.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 10 se obtiene una serie temporal con 822 datos, media de 9 paciente diario, con un mínimo de 1 paciente y máximo de 25 pacientes diarios.

Figura 52. Diagrama de caja para Otros dolores abdominales



Fuente: Construcción propia

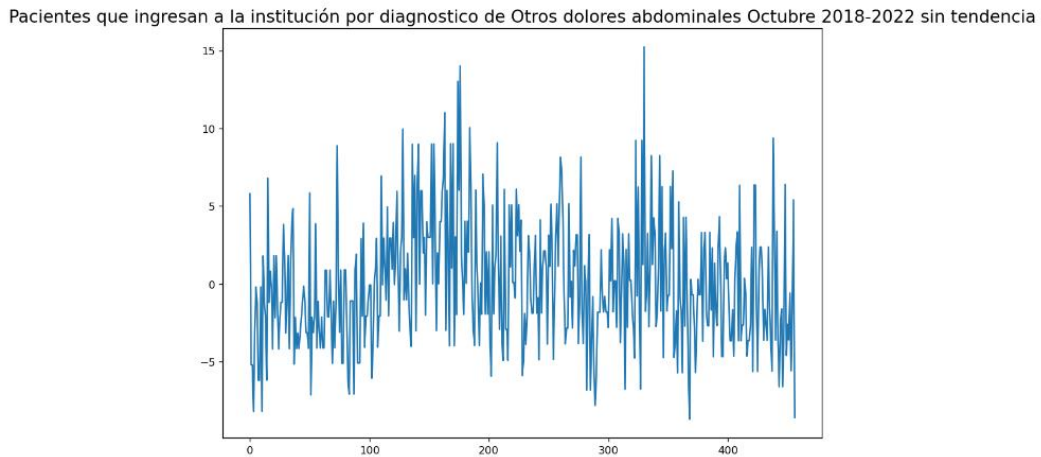
Probando estacionalidad

ADF Statistic: -2.9164199045268555
p-value: 0.043471633711728574
Critical Values:
1%, -3.4384791480370605
Critical Values:
5%, -2.865128114094396
Critical Values:
10%, -2.5686806329529874

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza la hipótesis nula lo que indica que la serie es estacionaria.

Eliminando tendencia

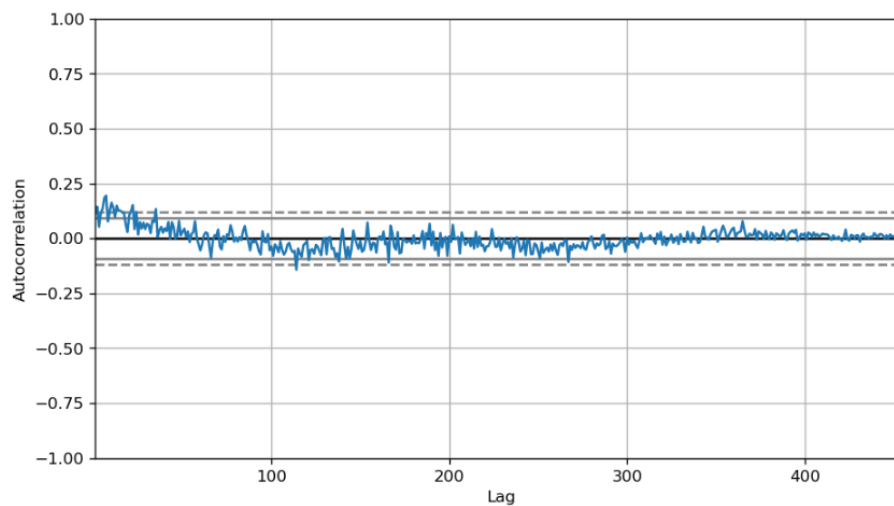
Figura 53. Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia



Fuente: Construcción propia

Autocorrelación

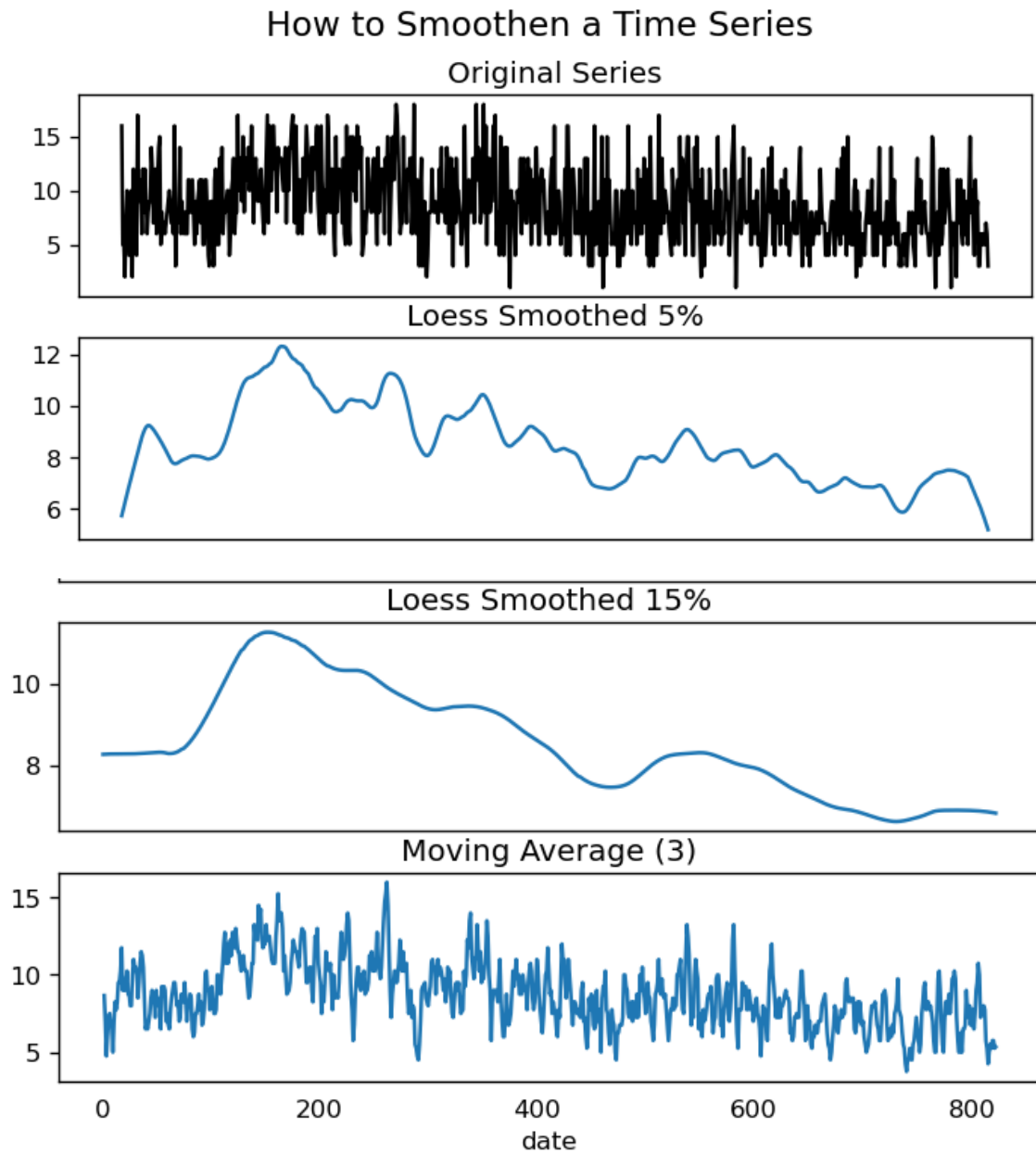
Figura 54. Autocorrelación para Otros dolores abdominales



Fuente: Construcción propia

Suavización de la serie temporal

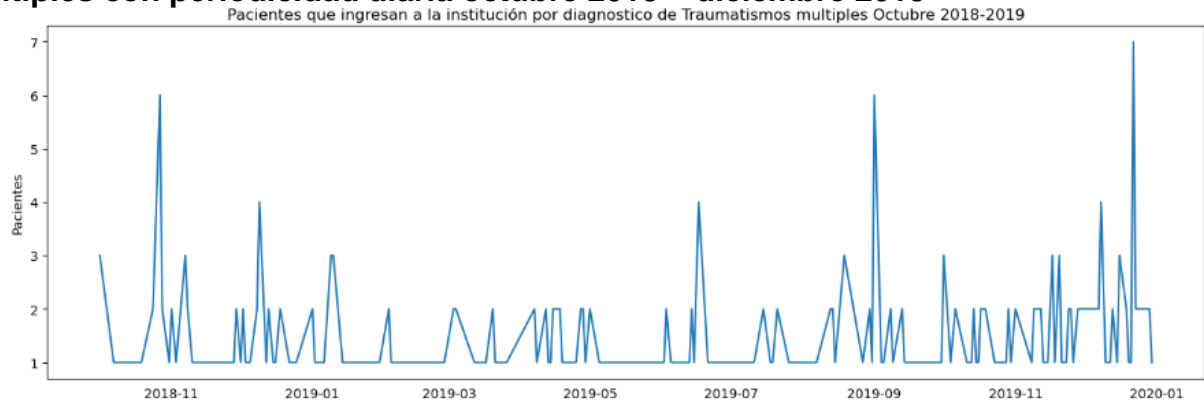
Figura 55. Suavización de la serie temporal para Otros dolores abdominales



Fuente: Construcción propia

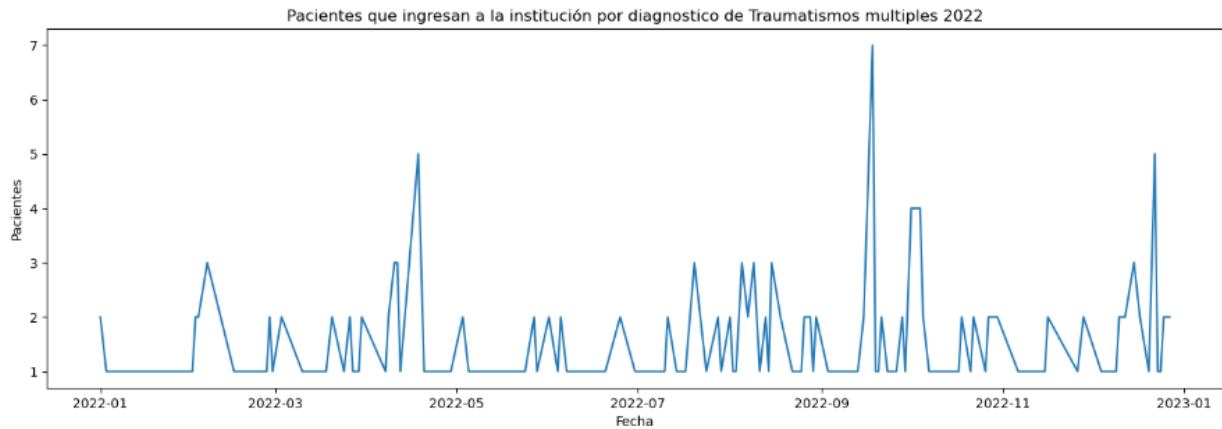
Traumatismos múltiples

Figura 56. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2019



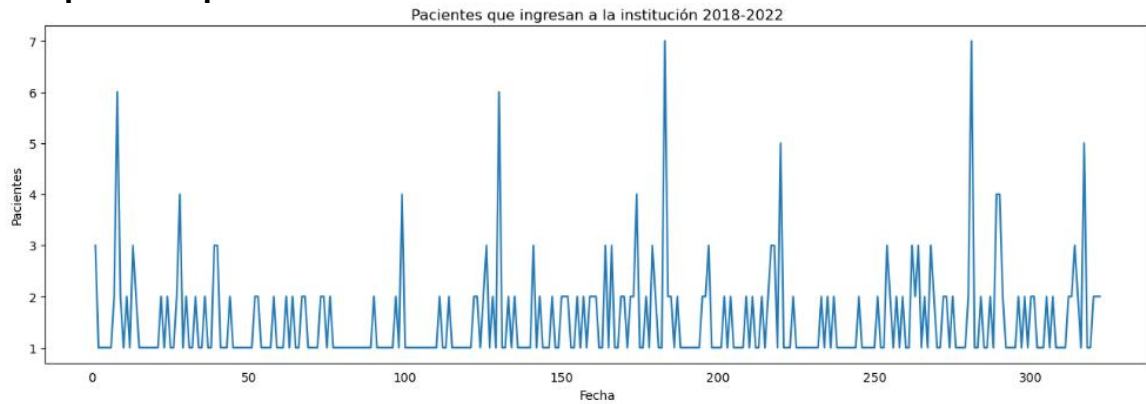
Fuente: Construcción propia

Figura 57. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria año 2022



Fuente: Construcción propia

Figura 58. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022



Fuente: Construcción propia

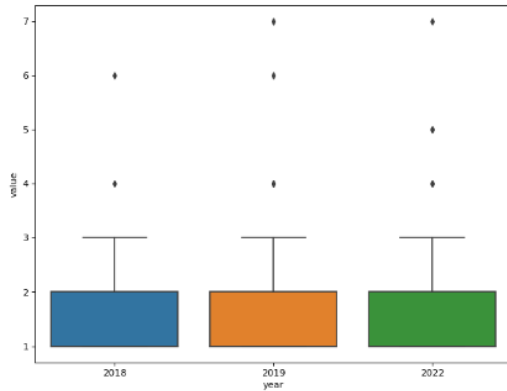
Tabla 12. Resumen estadístico de Traumatismos múltiples

	value
count	322.000000
mean	1.521739
std	0.914316
min	1.000000
25%	1.000000
50%	1.000000
75%	2.000000
max	7.000000

Fuente: Construcción propia

De acuerdo a la tabla no. 11 se obtiene una serie temporal con 322 datos, media de 1 paciente diario, con un mínimo de 1 paciente y máximo de 7 pacientes diarios.

Figura 59. Diagrama de caja por Traumatismos múltiples



Fuente: Construcción propia

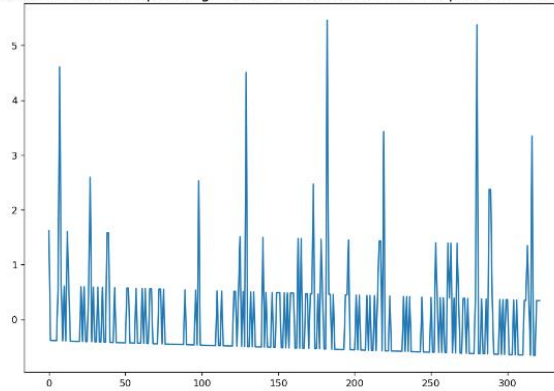
Probando estacionalidad

ADF Statistic: -17.88365411856515
p-value: 2.9958229957580167e-30
Critical Values:
1%, -3.450886958636161
Critical Values:
5%, -2.870586350823483
Critical Values:
10%, -2.5715897843576827

A probar estacionalidad en la serie se obtuvo que p-value es $<0,05$ por lo tanto se rechaza la hipótesis nula lo que indica que la serie es estacionaria.

Figura 60. Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022 sin tendencia

Pacientes que ingresan a la institución por diagnostico de Traumatismos multiples Octubre 2018-2022 sin tendencia



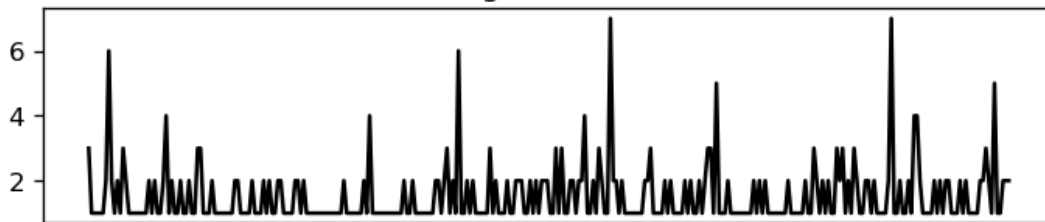
Fuente: Construcción propia

Suavización de la serie temporal

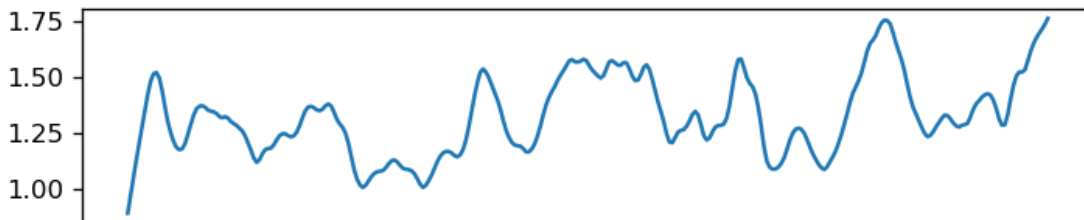
Figura 61. Suavización de la serie temporal para Traumatismos múltiples

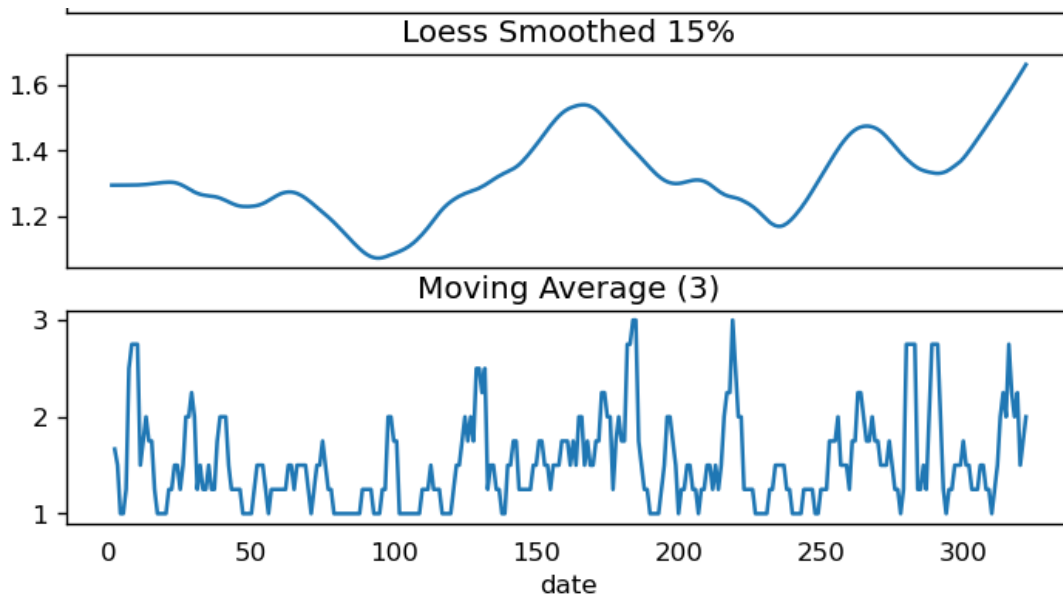
How to Smoothen a Time Series

Original Series



Loess Smoothed 5%





Fuente: Construcción propia

Los resultados hasta este momento evidencian para el total de patologías, como para las mas criticas un panorama descriptivo importante, enfocados principalmente en el comportamiento de cada serie de tiempo, el resumen estadístico de la serie, la variabilidad en el comportamiento de cada patología por año, si la serie es estacionaria o no y la autocorrelación de los datos de la serie, información que permitirá continuar a la siguiente fase que es el modelamiento.

4.7 Modelamiento

Para resolver el objetivo de este proyecto se propone utilizar el algoritmo de serie de tiempo, ya que lo que se busca es predecir la evolución del comportamiento de pacientes atendidos de cada patología critica en el tiempo, utilizando los datos históricos descritos anteriormente. Como método de serie de tiempo se trabajará con el ARIMA abreviatura de 'Promedio Móvil Integrado AutoRegresivo', el cual es un algoritmo de pronóstico basado en la idea de que la información en los valores pasados de la serie de tiempo puede usarse solo para predecir los valores futuros.

Para todas las patologías se obtiene los siguientes resultados:

Pacientes ingresados intrahospitalariamente con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -2.201102
p-value: 0.205846

d = 1, ya que es estacionaria en 1 diferenciación.
p= 3, es un lags bastante significativo
q= 2, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (3,1,2) obteniéndose un AIC =7901, que al compararlo con un ARIMA (3,1,1) (AIC =7942) es mucho mejor.

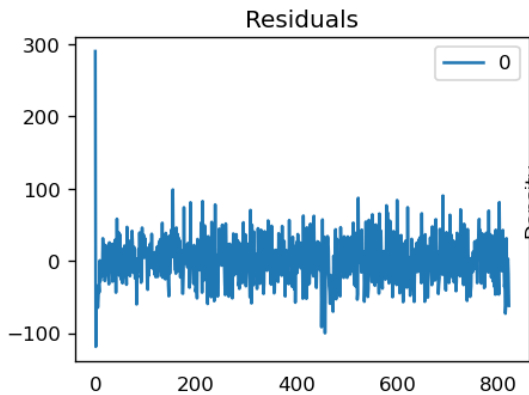
modelo ARIMA (p,d,q) = (3,1,2)

Figura 62. Resultado modelo ARIMA

SARIMAX Results						
Dep. Variable:	value	No. Observations:	822			
Model:	ARIMA(3, 1, 2)	Log Likelihood	-3944.987			
Date:	Fri, 19 May 2023	AIC	7901.974			
Time:	18:53:35	BIC	7930.237			
Sample:	0	HQIC	7912.818			
	- 822					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.7705	0.100	7.736	0.000	0.575	0.966
ar.L2	-0.3743	0.059	-6.346	0.000	-0.490	-0.259
ar.L3	-0.1021	0.061	-1.669	0.095	-0.222	0.018
ma.L1	-1.4881	0.092	-16.174	0.000	-1.668	-1.308
ma.L2	0.5872	0.082	7.118	0.000	0.426	0.749
sigma2	871.1122	44.168	19.723	0.000	784.544	957.680
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	5.89			
Prob(Q):	0.99	Prob(JB):	0.05			
Heteroskedasticity (H):	1.28	Skew:	0.20			
Prob(H) (two-sided):	0.04	Kurtosis:	3.12			

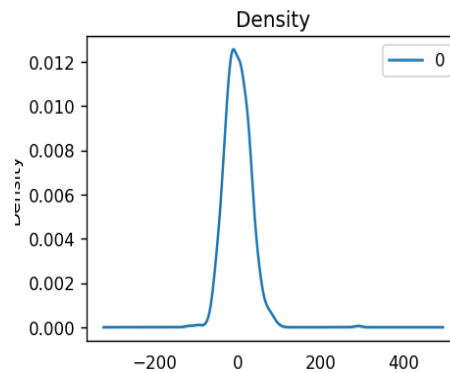
Fuente: Construcción propia

Figura 63. Residuales



Fuente: Construcción propia

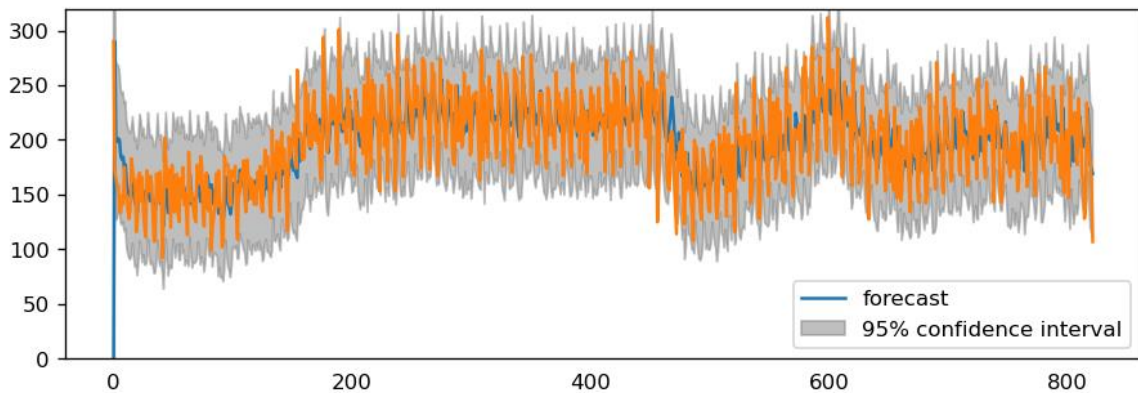
Figura 64. Densidad



Fuente: Construcción propia

Vemos que a partir de un ARIMA (3,1,2) se obtiene un comportamiento de la serie en torno a media =0, y una densidad con una distribución normal con media = 0.

Figura 65. Comportamiento real vs predicción



Fuente: Construcción propia

Por lo tanto, con un ARIMA (3,1,2) se puede observar en la figura 81. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Las patologías más críticas se detallan a continuación:

Cólico

Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -4.512155
p-value: 0.000187

d = 0, ya que es estacionaria sin diferenciarla
p= 1, es un lags bastante significativo
q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (1,0,1) obteniéndose un AIC =3261, que al compararlo con un ARIMA (1,0,2) (AIC =3263) es mucho mejor.

modelo ARIMA (p,d,q) = (1,0,1)

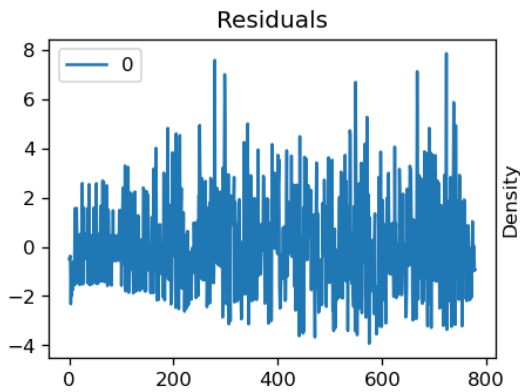
Figura 66. Resultado modelo ARIMA para Cólico

```
=====
SARIMAX Results
=====
Dep. Variable:          value    No. Observations:          779
Model:                 ARIMA(1, 0, 1)  Log Likelihood             -1626.822
Date:                  Fri, 28 Apr 2023  AIC                        3261.645
Time:                  06:13:53       BIC                        3280.277
Sample:                0             HQIC                       3268.811
                   - 779
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
const          3.5042    0.800         4.382    0.000     1.937     5.072
ar.L1          0.9972    0.003       321.170    0.000     0.991     1.003
ma.L1         -0.9680    0.011      -91.838    0.000    -0.989    -0.947
sigma2         3.8085    0.190        20.040    0.000     3.436     4.181
=====
Ljung-Box (L1) (Q):                0.23  Jarque-Bera (JB):                53.32
Prob(Q):                            0.63  Prob(JB):                        0.00
Heteroskedasticity (H):              1.94  Skew:                             0.61
Prob(H) (two-sided):                 0.00  Kurtosis:                         3.37
=====

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

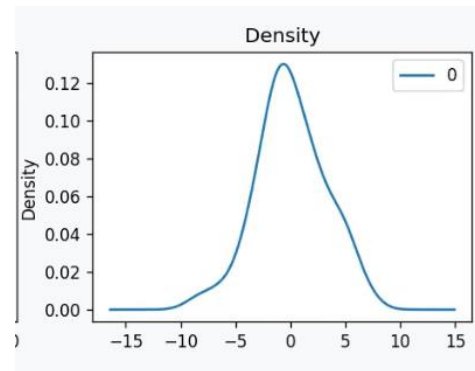
Fuente: Construcción propia

Figura 67. Residual para Cólico



Fuente: Construcción propia

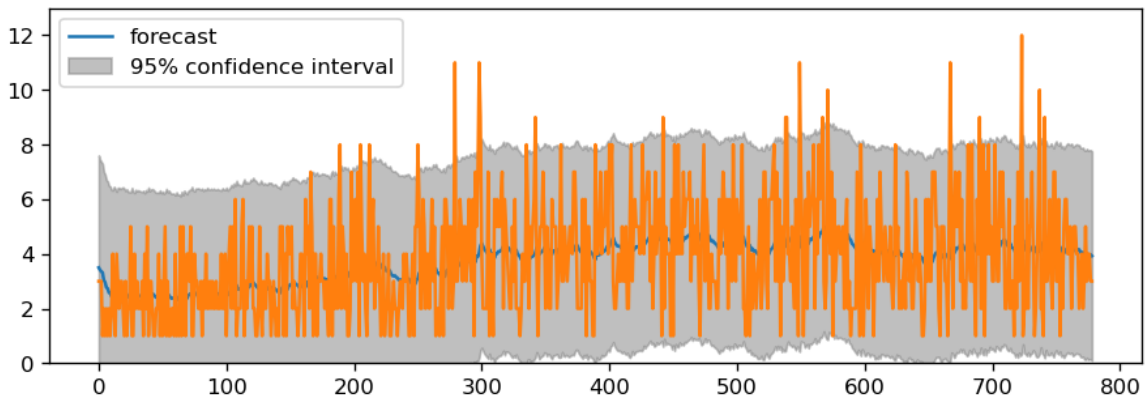
Figura 68. Densidad para Cólico



Fuente: Construcción propia

Vemos que a partir de un ARIMA (1,0,1) se obtiene un comportamiento de la serie en torno a media =0 , y una densidad con una distribución normal con media = 0.

Figura 69. Comportamiento real vs predicción para Cólico



Fuente: Construcción propia

Por lo tanto, con un ARIMA (1,0,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Dolor Abdominal

Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -4.045473
 p-value: 0.001190

d = 0, ya que es estacionaria sin diferenciarla
 p= 2, es un lags bastante significativo
 q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (2,0,1) obteniéndose un AIC =2328, que al compararlo con un ARIMA (1,0,1) (AIC =2330) es mucho mejor.

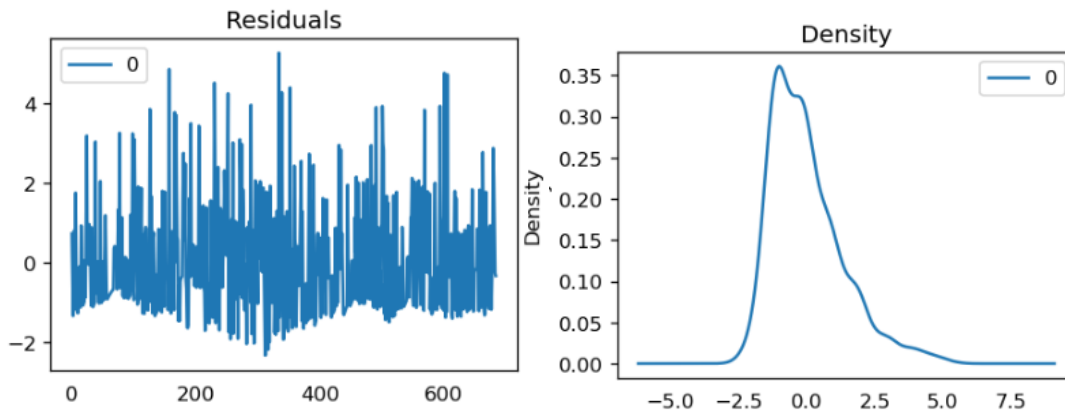
modelo ARIMA (p,d,q) = (2,0,1)

Figura 70. Resultado modelo ARIMA para dolor abdominal

SARIMAX Results						
Dep. Variable:	value	No. Observations:	685			
Model:	ARIMA(2, 0, 1)	Log Likelihood	-1159.194			
Date:	Sat, 29 Apr 2023	AIC	2328.388			
Time:	07:42:22	BIC	2351.035			
Sample:	0	HQIC	2337.151			
	- 685					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
const	2.2786	0.291	7.824	0.000	1.708	2.849
ar.L1	0.9129	0.043	21.067	0.000	0.828	0.998
ar.L2	0.0773	0.041	1.872	0.061	-0.004	0.158
ma.L1	-0.9527	0.017	-55.497	0.000	-0.986	-0.919
sigma2	1.7258	0.091	18.900	0.000	1.547	1.905
Ljung-Box (L1) (Q):		0.01	Jarque-Bera (JB):	202.36		
Prob(Q):		0.93	Prob(JB):	0.00		
Heteroskedasticity (H):		1.12	Skew:	1.15		
Prob(H) (two-sided):		0.39	Kurtosis:	4.34		

Fuente: Construcción propia

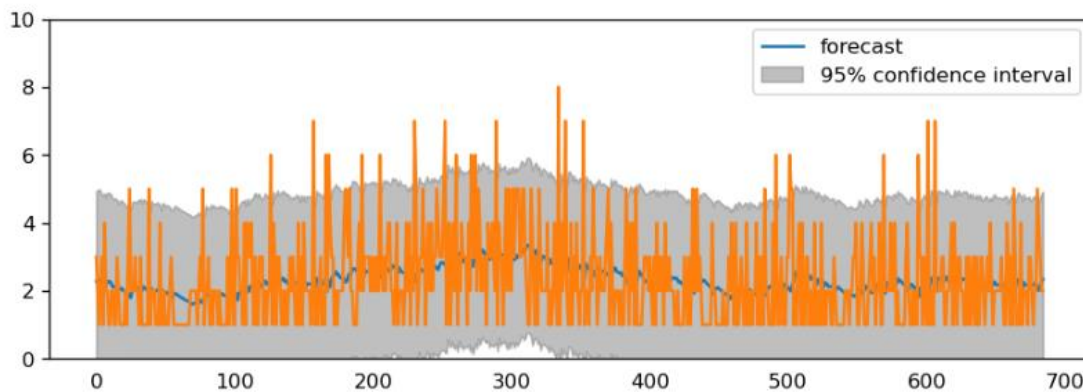
Figura 71. Residual para Dolor abdominal **Figura 72. Densidad para Dolor abdominal**



Fuente: Construcción propia

Fuente: Construcción propia

Figura 73. Comportamiento real vs predicción para dolor abdominal



Fuente: Construcción propia

Por lo tanto, con un ARIMA (2,0,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

EPOC

Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -4.240507

p-value: 0.000562

d = 0, ya que es estacionaria sin diferenciarla
p= 1, es un lags bastante significativo
q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (1,0,1) obteniéndose un AIC =1018, que al compararlo con un ARIMA (1,0,2) (AIC =1019) es mucho mejor.

modelo ARIMA (p,d,q) = (1,0,1)

Figura 74. Resultado modelo ARIMA para EPOC

```
=====
                        SARIMAX Results
=====
Dep. Variable:          value      No. Observations:          402
Model:                 ARIMA(1, 0, 1)  Log Likelihood            -505.400
Date:                 Sat, 29 Apr 2023  AIC                       1018.800
Time:                 14:22:49        BIC                       1034.785
Sample:               0              HQIC                      1025.129
                        - 402
Covariance Type:      opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
const          1.6332      0.163       10.005     0.000       1.313       1.953
ar.L1           0.9400      0.037       25.379     0.000       0.867       1.013
ma.L1          -0.8371      0.051      -16.492     0.000      -0.937      -0.738
sigma2          0.7232      0.053       13.674     0.000       0.620       0.827
=====
Ljung-Box (L1) (Q):          0.59  Jarque-Bera (JB):          155.05
Prob(Q):                    0.44  Prob(JB):                  0.00
Heteroskedasticity (H):     0.30  Skew:                      1.31
Prob(H) (two-sided):        0.00  Kurtosis:                   4.54
=====
```

Fuente: Construcción propia

Figura 75. Residual para EPOC

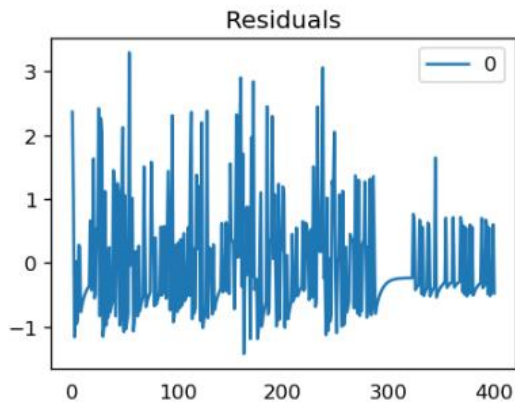
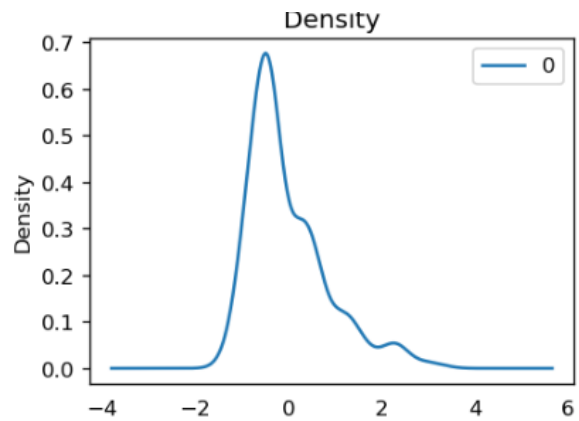
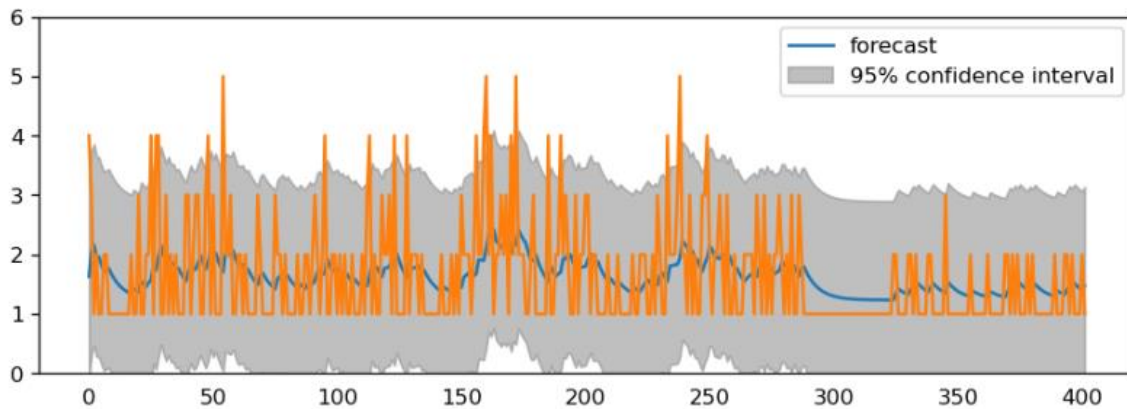


Figura 76. Densidad para EPOC



Fuente: Construcción propia

Figura 77. Comportamiento real vs predicción para EPOC



Fuente: Construcción propia

Por lo tanto, con un ARIMA (1,0,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Fiebre

Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -2.546278

p-value: 0.104617

d = 1, ya que se vuelve estacionaria con la primera diferenciación

p= 1, es un lags bastante significativo

q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (1,1,1) obteniéndose un AIC =4728, que al compararlo con un ARIMA (1,1,2) (AIC =4731) es mucho mejor.

modelo ARIMA (p,d,q) = (1,1,1)

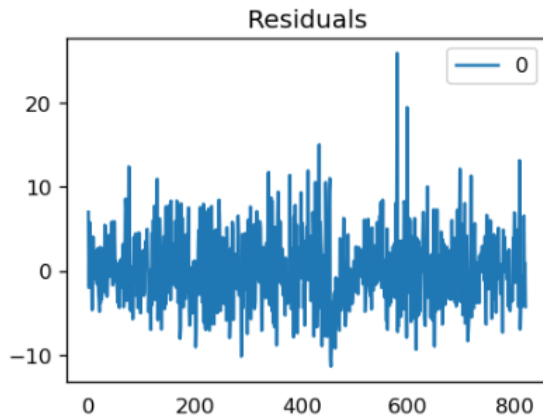
Figura 78. Resultado modelo ARIMA para fiebre

SARIMAX Results						
=====						
Dep. Variable:	value	No. Observations:	821			
Model:	ARIMA(1, 1, 1)	Log Likelihood	-2361.475			
Date:	Sat, 29 Apr 2023	AIC	4728.951			
Time:	15:39:09	BIC	4743.078			
Sample:	0	HQIC	4734.371			
	- 821					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	0.0388	0.037	1.060	0.289	-0.033	0.110
ma.L1	-0.8755	0.019	-45.197	0.000	-0.913	-0.837
sigma2	18.5436	0.674	27.504	0.000	17.222	19.865
=====						
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	165.40			
Prob(Q):	0.94	Prob(JB):	0.00			
Heteroskedasticity (H):	1.29	Skew:	0.63			
Prob(H) (two-sided):	0.04	Kurtosis:	4.80			
=====						

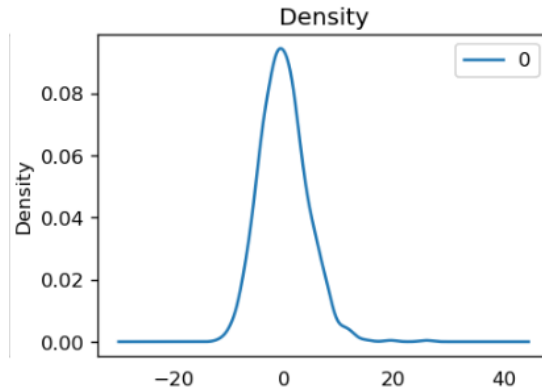
Fuente: Construcción propia

Figura 79. Residual para Fiebre



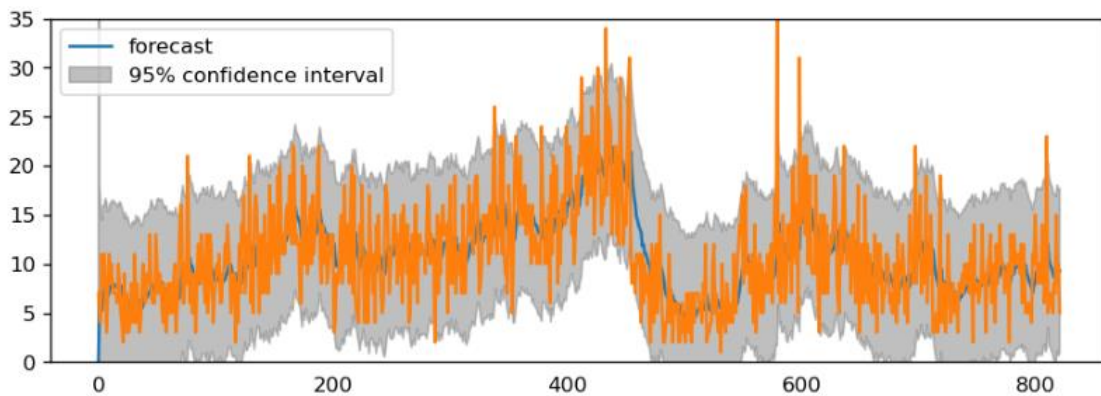
Fuente: Construcción propia

Figura 80. Densidad para Fiebre



Fuente: Construcción propia

Figura 81. Comportamiento real vs predicción para Fiebre



Fuente: Construcción propia

Por lo tanto, con un ARIMA (1,1,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Infección de vías urinarias

Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -3.018564
p-value: 0.033197

d = 0, ya que es estacionaria sin diferenciarla
p= 1, es un lags bastante significativo
q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (1,0,1) obteniéndose un AIC =3221, que al compararlo con un ARIMA (2,0,1) (AIC =3223) es mucho mejor.

modelo ARIMA (p,d,q) = (1,0,1)

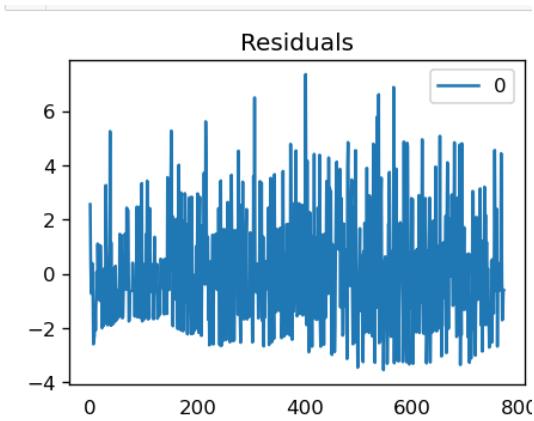
Figura 82. Resultado modelo ARIMA para infección de vías urinarias

SARIMAX Results						
=====						
Dep. Variable:	value	No. Observations:	771			
Model:	ARIMA(1, 0, 1)	Log Likelihood	-1606.883			
Date:	Sat, 29 Apr 2023	AIC	3221.766			
Time:	16:08:43	BIC	3240.357			
Sample:	0	HQIC	3228.920			
	- 771					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

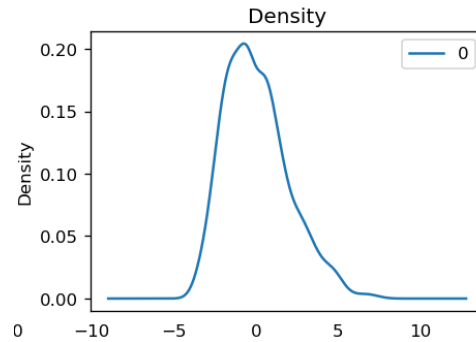
const	3.4438	0.536	6.420	0.000	2.393	4.495
ar.L1	0.9961	0.004	257.909	0.000	0.988	1.004
ma.L1	-0.9701	0.011	-85.764	0.000	-0.992	-0.948
sigma2	3.7783	0.197	19.156	0.000	3.392	4.165
=====						
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	65.43			
Prob(Q):	0.94	Prob(JB):	0.00			
Heteroskedasticity (H):	1.77	Skew:	0.70			
Prob(H) (two-sided):	0.00	Kurtosis:	3.29			
=====						

Fuente: Construcción propia

Figura 83. Residual para Infección de vías urinarias **Figura 84. Densidad para Infección de vías urinarias**

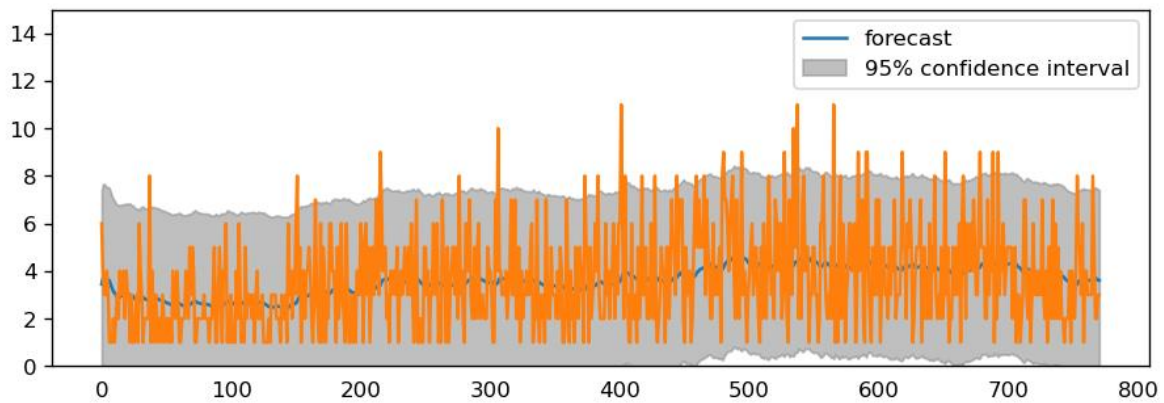


Fuente: Construcción propia



Fuente: Construcción propia

Figura 85. Comportamiento real vs predicción para Infección de vías urinarias



Fuente: Construcción propia

Por lo tanto, con un ARIMA (1,1,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Otros dolores abdominales

Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -2.916420

p-value: 0.043472

d = 0, ya que es estacionaria sin diferenciarla

p= 1, es un lags bastante significativo

q= 1, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (1,0,1) obteniéndose un AIC =4431, que al compararlo con un ARIMA (2,0,1) (AIC =4433) es mucho mejor.

modelo ARIMA (p,d,q) = (1,0,1)

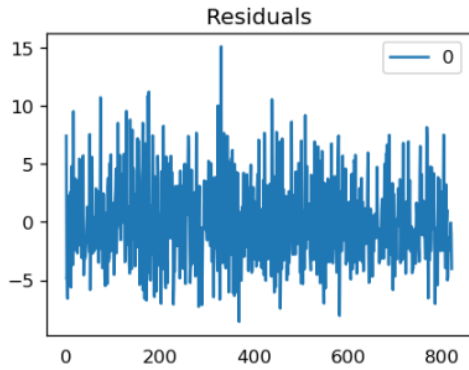
Figura 86. Resultado modelo ARIMA para otros dolores abdominales

SARIMAX Results						
=====						
Dep. Variable:	value	No. Observations:	822			
Model:	ARIMA(1, 0, 1)	Log Likelihood	-2211.908			
Date:	Sat, 29 Apr 2023	AIC	4431.815			
Time:	16:39:40	BIC	4450.662			
Sample:	0	HQIC	4439.046			
	- 822					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	8.5814	0.844	10.171	0.000	6.928	10.235
ar.L1	0.9917	0.006	174.989	0.000	0.981	1.003
ma.L1	-0.9394	0.016	-59.211	0.000	-0.971	-0.908
sigma2	12.7156	0.626	20.307	0.000	11.488	13.943
=====						
Ljung-Box (L1) (Q):	0.11	Jarque-Bera (JB):	37.78			
Prob(Q):	0.74	Prob(JB):	0.00			
Heteroskedasticity (H):	0.62	Skew:	0.51			
Prob(H) (two-sided):	0.00	Kurtosis:	3.22			
=====						

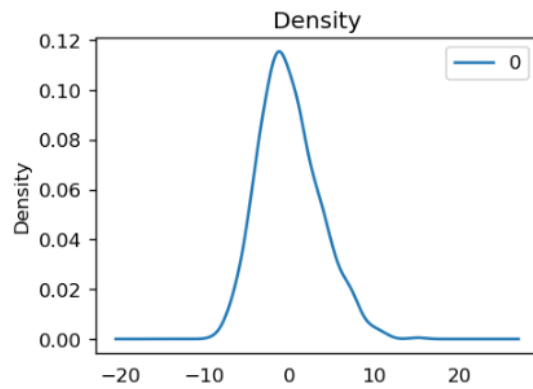
Fuente: Construcción propia

Figura 87. Residual para otros dolores abdominales



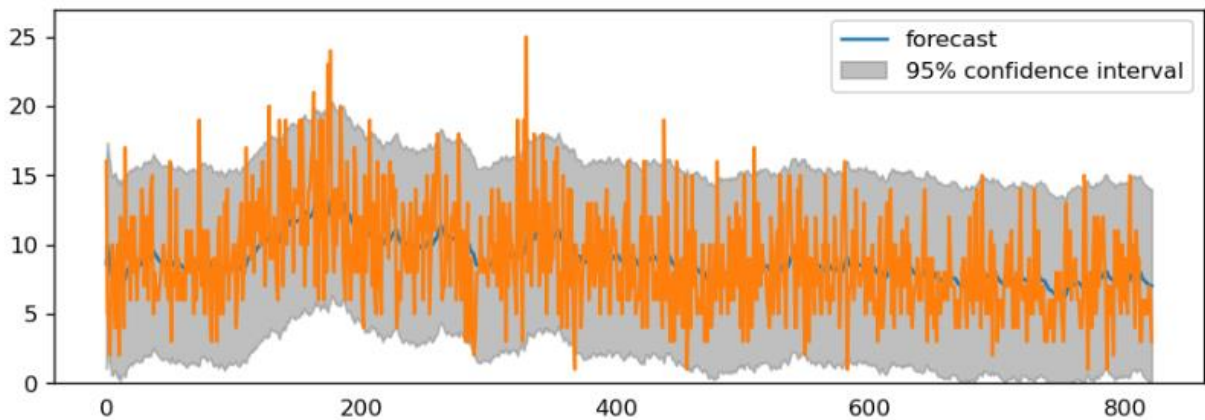
Fuente: Construcción propia

Figura 88. Densidad para otros dolores abdominales



Fuente: Construcción propia

Figura 89. Comportamiento real vs predicción para otros dolores abdominales



Fuente: Construcción propia

Por lo tanto, con un ARIMA (1,0,1) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

Traumatismos múltiples

Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022

Definición del modelo ARIMA

ADF Statistic: -17.883654
p-value: 0.000000

d = 0, ya que es estacionaria sin diferenciarla
p = 0, es un lags bastante significativo
q = 0, es un lags bastante significativo

Al realizar la determinación del orden (p,d,q) se obtiene un valor de (0,0,0) obteniéndose un AIC =859, que al compararlo con un ARIMA (1,0,1) (AIC =863) es mucho mejor.

modelo ARIMA (p,d,q) = (0,0,0)

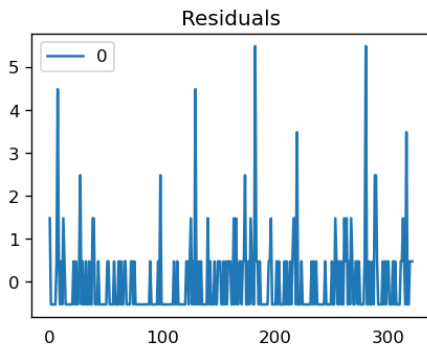
Figura 90. Resultado modelo ARIMA para traumatismos multiples

SARIMAX Results						
=====						
Dep. Variable:	value	No. Observations:	322			
Model:	ARIMA	Log Likelihood	-427.553			
Date:	Sat, 29 Apr 2023	AIC	859.106			
Time:	17:13:06	BIC	866.655			
Sample:	0	HQIC	862.120			
	- 322					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

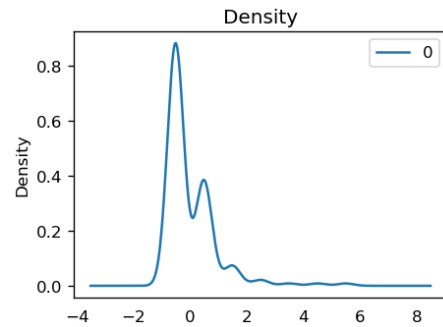
const	1.5217	0.083	18.355	0.000	1.359	1.684
sigma2	0.8334	0.042	20.045	0.000	0.752	0.915
=====						
Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	2143.35			
Prob(Q):	0.97	Prob(JB):	0.00			
Heteroskedasticity (H):	1.58	Skew:	2.87			
Prob(H) (two-sided):	0.02	Kurtosis:	14.26			
=====						

Fuente: Construcción propia

Figura 91. Residual para Traumatismos múltiples **Figura 92. Densidad para Traumatismos múltiples**

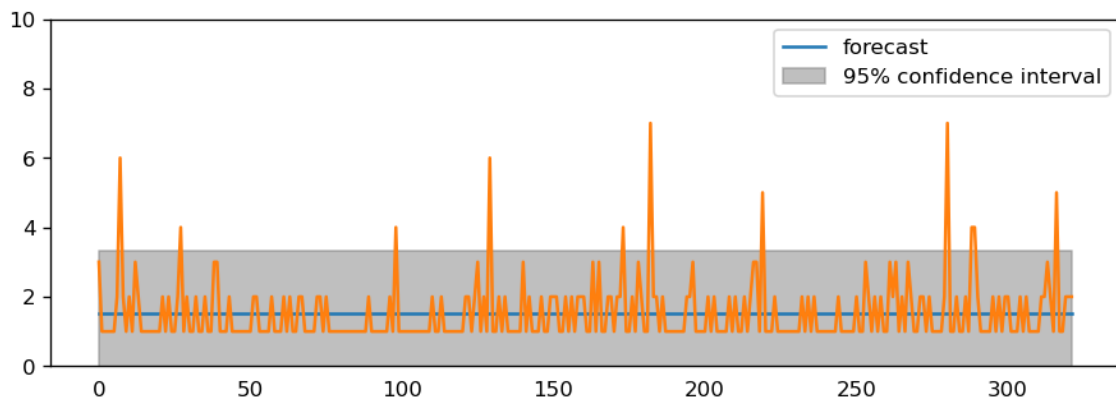


Fuente: Construcción propia



Fuente: Construcción propia

Figura 93. Comportamiento real vs predicción para Traumatismos múltiples



Fuente: Construcción propia

Por lo tanto, con un ARIMA (0,0,0) se puede observar en la figura 84. el comportamiento de la serie (Línea naranjada) comparándola en el mismo periodo con su predicción (Línea azul). Por lo anterior, se obtiene un comportamiento que permitirá realizar la validación del modelo.

4.8 Evaluación del modelo

Para la validación real se necesita ahora la validación cruzada fuera del tiempo. En la validación cruzada Out-of-Time, se retrocede unos pasos en el tiempo y pronostica en el futuro tantos pasos como retrocedió. Luego se compara el pronóstico con los datos reales. Para realizar una validación cruzada fuera de tiempo, se debe crear el conjunto de datos de entrenamiento y prueba dividiendo la

serie temporal en 2 partes contiguas en una proporción de aproximadamente 75:25 o una proporción razonable basada en la frecuencia temporal de la serie

Para series de tiempo, el mejor método de evaluación del modelo es aplicar los indicadores de MSE, RMSE, MAPE y R^2 , quedándonos con el MAPE como el indicador de validación más idóneo ya que nos muestra el error % de la distribución. Por lo anterior para cada una de las patologías críticas se procedió a hacer 3 tipos de comparación para el modelo, utilizando un test y train completo (85% - 15%), un test y train completo (75% - 15%) y por último partiendo de un test y train (85% - 15% o 75% - 15%) se decide utilizar los primeros 15 datos del test para observar la capacidad de predicción del modelo, obteniéndose los siguientes resultados:

5. DISEÑO DE EXPERIMENTO DE VALIDACIÓN

A continuación, se expone los resultados de las validaciones realizadas para encontrar el mejor resultado de las métricas seleccionadas para validar el modelo:

Total de pacientes

Pacientes ingresados intrahospitalariamente con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la métrica:-

MSE es:990.583905852989
MSE es:25.678905306677787
RMSE es:31.473542950436784
MAPE es:12.834928638748295
R2 es:-240.9309528441043

Con test completo 75:25

Resultado de evaluación de la métrica:-

MSE es:1206.7146787916822
MSE es:28.183492701517984
RMSE es:34.73779899175655
MAPE es:13.367242160073674
R2 es:-10040.824974174602

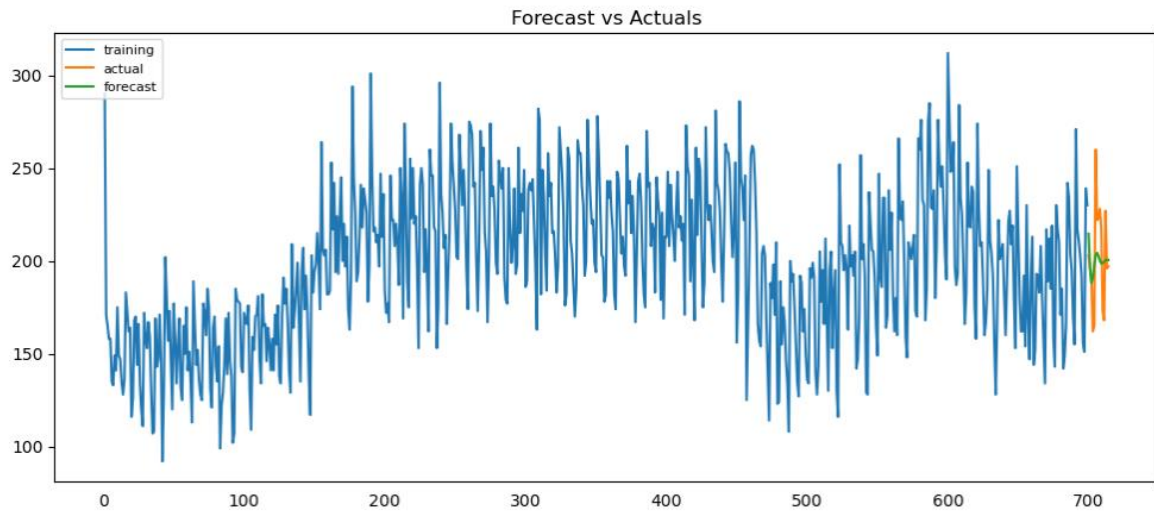
Con test [:15]

Resultado de evaluación de la métrica:-

MSE es:636.059389331458
MSE es:20.44697675490897
RMSE es:25.220217868437576
MAPE es:10.222618615096986
R2 es:-6.032365728298834

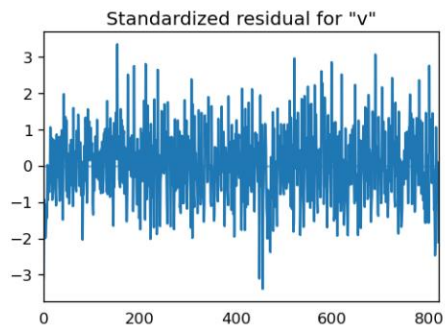
Con test completo 75:25 Se obtiene un mejor **MAPE**

Figura 94. Comportamiento predicción vs real

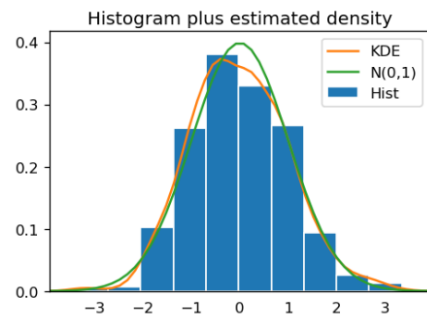


Fuente: Construcción propia

Figura 95. Estandarización residual por "v" Figura 96. Histograma con estimación de densidad

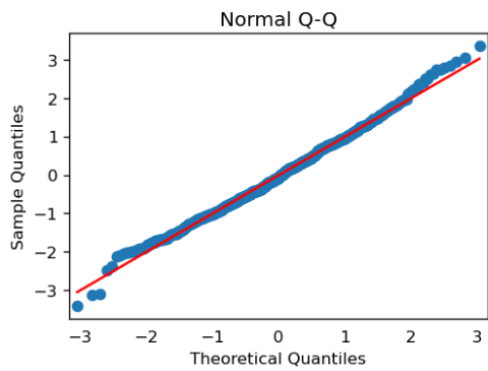


Fuente: Construcción propia



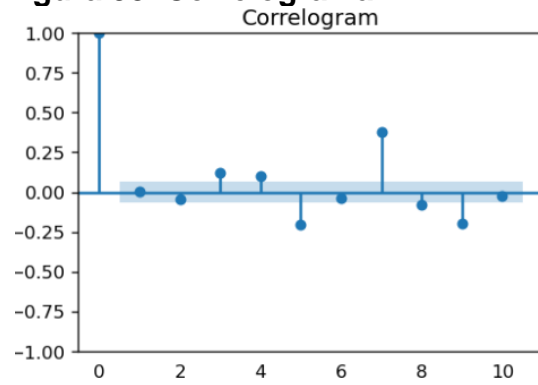
Fuente: Construcción propia

Figura 97. Normal Q-Q



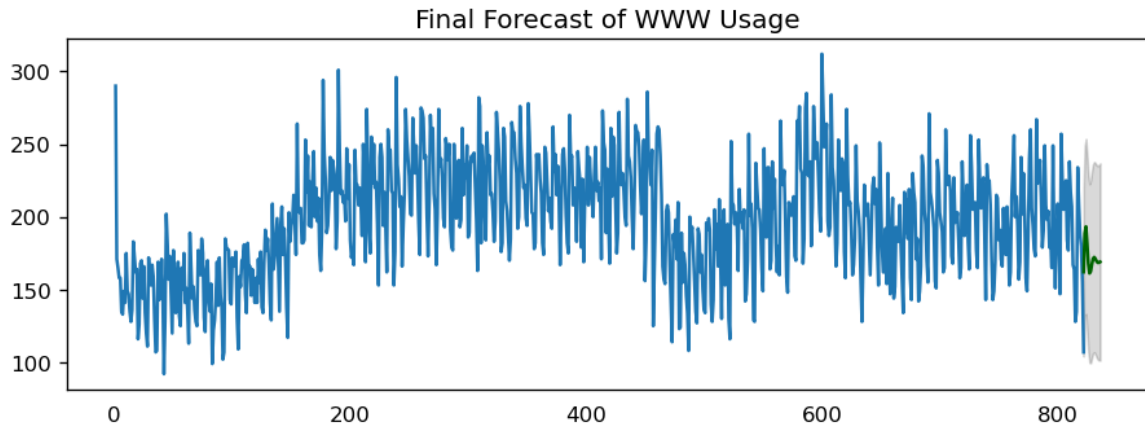
Fuente: Construcción propia

Figura 98. Correlograma



Fuente: Construcción propia

Figura 99. Estimación final de predicciones futuras



Fuente: Construcción propia

```
822    162.134045
823    188.643224
824    193.332584
825    181.393374
826    167.732423
827    161.197048
828    162.494150
829    167.334665
830    171.245979
831    172.315285
832    171.180877
833    169.507215
834    168.533125
835    168.524904
836    169.054072
dtype: float64
```

Cólico

Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

MSE es:5.1425971052211725
MSE es:1.7643949738312035
RMSE es:2.2677295044209247
MAPE es:45.58742669200082
R2 es:-2222.2113524080587

Con test completo 75:25

Resultado de evaluación de la metrica:-

MSE es:4.746667905660148
MSE es:1.816724653011709
RMSE es:2.17868490279346
MAPE es:39.50258281641741
R2 es:-452.82071967495364

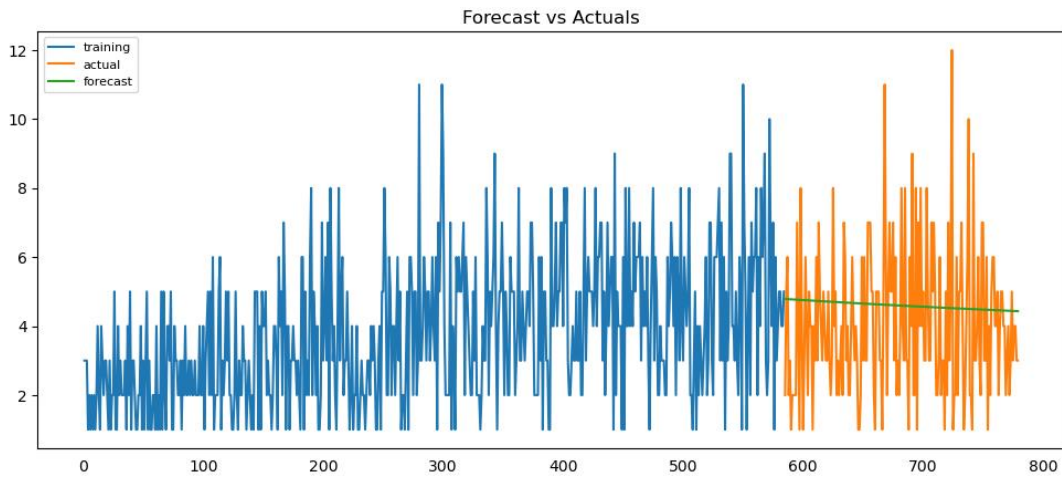
Con test [:15]

Resultado de evaluación de la metrica:-

MSE es:6.523238976209851
MSE es:2.0220299955010455
RMSE es:2.554063228702424
MAPE es:51.31994450889804
R2 es:-117747.22112571275

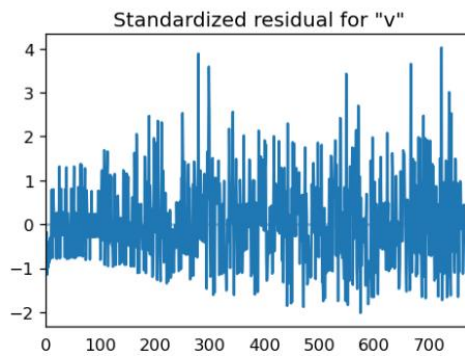
Con test completo 75:25 Se obtiene un mejor **MAPE**

Figura 100. Comportamiento predicción vs real para cólico



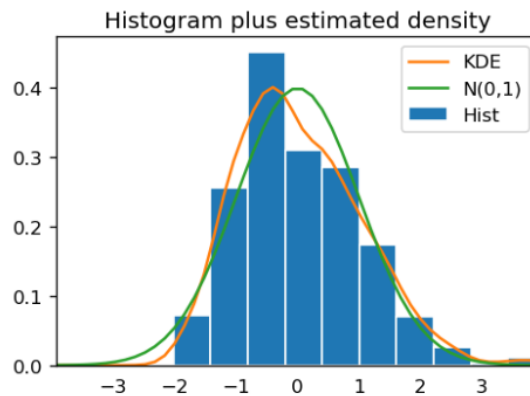
Fuente: Construcción propia

Figura 101. Estandarización residual por "v" para cólico



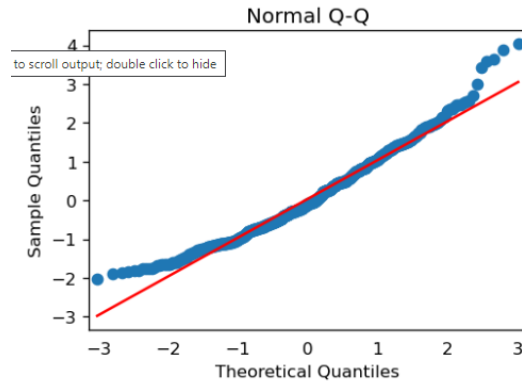
Fuente: Construcción propia

Figura 102. Histograma con estimación de densidad para cólico



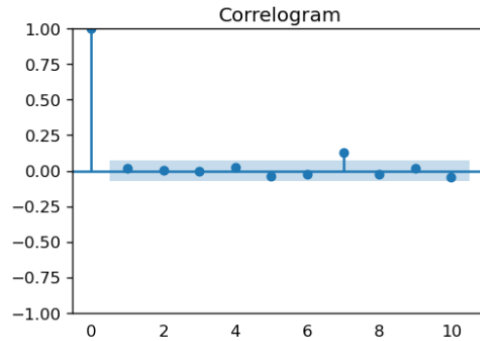
Fuente: Construcción propia

Figura 103. Normal Q-Q para cólico



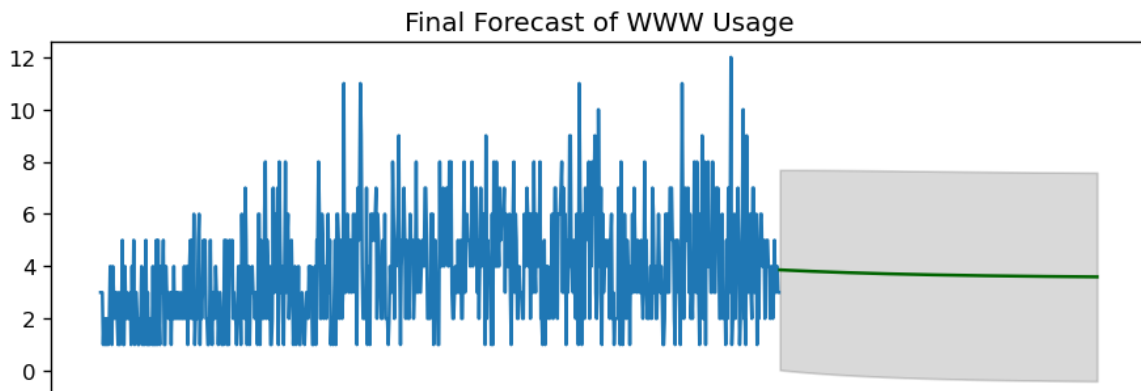
Fuente: Construcción propia

Figura 104. Correlograma para cólico



Fuente: Construcción propia

Figura 105. Estimación final de predicciones futuras para cólico



Fuente: Construcción propia

```
779    3.863291
780    3.861642
781    3.860002
782    3.858371
783    3.856748
...
969    3.664908
970    3.664302
971    3.663700
972    3.663101
973    3.662505
Length: 195, dtype: float64
```

Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

```
MSE es:1.6061945783229001
MSE es:0.958398789489902
RMSE es:1.2673573206964561
MAPE es:43.52814041853359
R2 es:-4304.767968725975
```

Con test completo 75:25

Resultado de evaluación de la metrica:-

```
MSE es:1.525157227081506
MSE es:1.006584605078241
RMSE es:1.2349725612666484
MAPE es:42.6542907048087
R2 es:-9171.164702242846
```

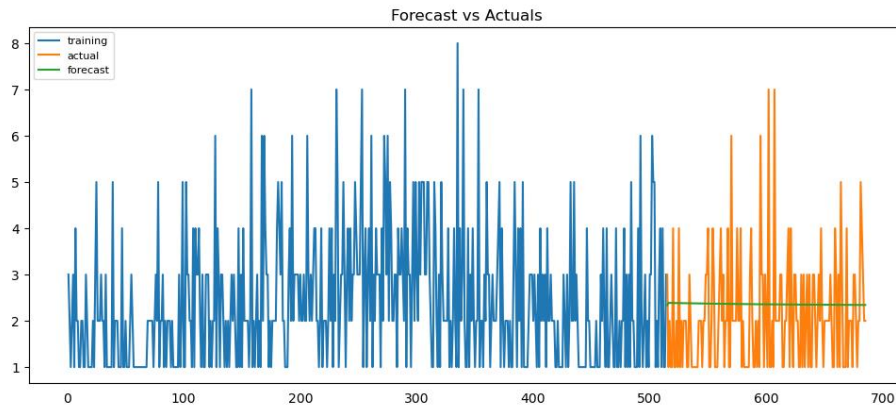
Con test [:15]

Resultado de evaluación de la metrica:-

```
MSE es:1.9926515953117945
MSE es:1.1281547754444186
RMSE es:1.41161311814243
MAPE es:51.79271384231691
R2 es:-4501.925414166578
```

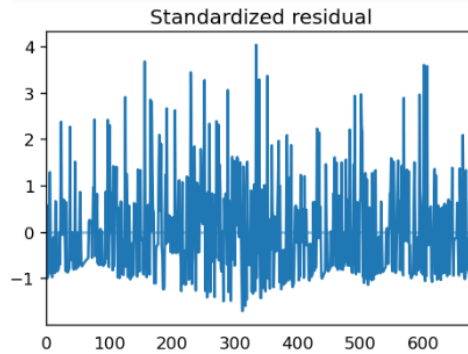
Con test completo 75:25 Se obtiene un mejor **MAPE**

Figura 106. Comportamiento predicción vs real para dolor abdominal



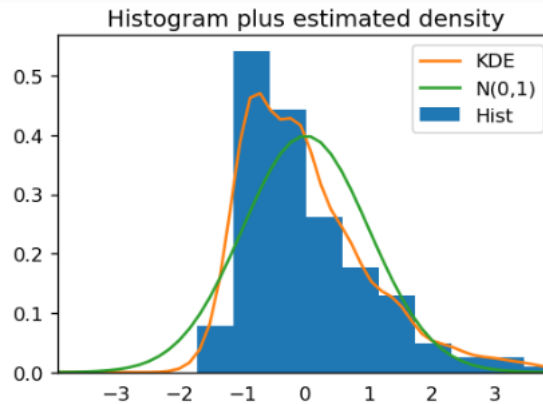
Fuente: Construcción propia

Figura 107. Estandarización residual por "v" para dolor abdominal



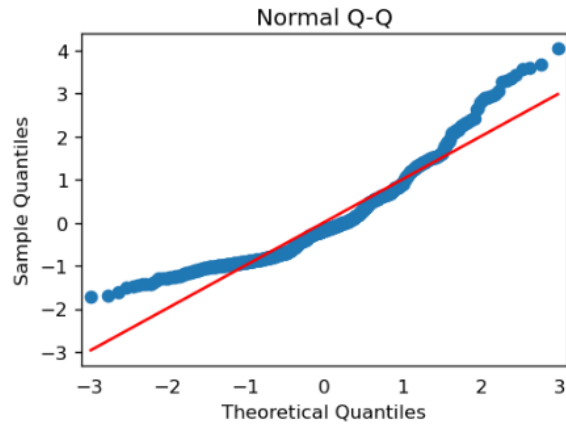
Fuente: Construcción propia

Figura 108. Histograma con estimación de densidad para dolor abdominal



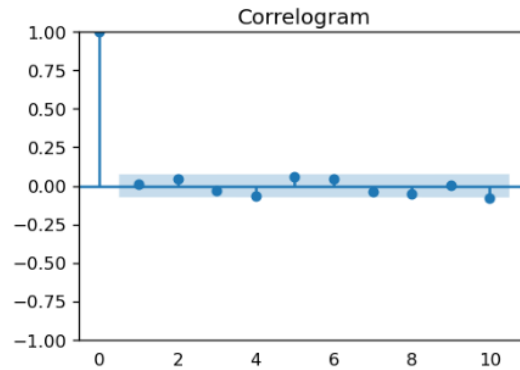
Fuente: Construcción propia

Figura 109. Normal Q-Q para dolor abdominal



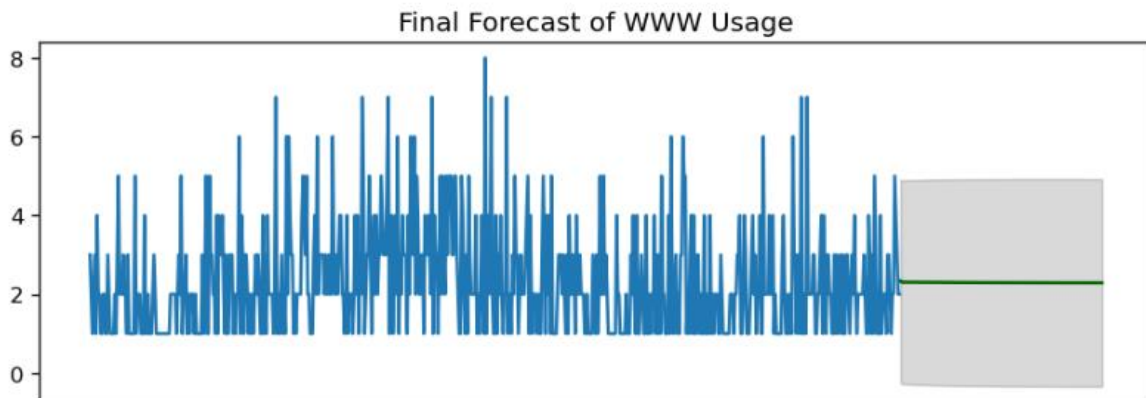
Fuente: Construcción propia

Figura 110. Correlograma dolor abdominal



Fuente: Construcción propia

Figura 111. Estimación final de predicciones futuras para dolor abdominal



Fuente: Construcción propia

```
685 2.332092
686 2.305252
687 2.307117
688 2.306690
689 2.306450
...
851 2.290118
852 2.290091
853 2.290065
854 2.290039
855 2.290013
Length: 171, dtype: float64
```

EPOC

Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

```
MSE es:0.38447766122480725
MSE es:0.5917861216063921
RMSE es:0.6200626268570033
MAPE es:35.83969721144405
R2 es:-126.76386463911268
```

Con test completo 75:25

Resultado de evaluación de la metrica:-

```
MSE es:0.45812311968172637
MSE es:0.6415611394345455
RMSE es:0.6768479295098171
MAPE es:36.82713887729172
R2 es:-139.84424061521344
```

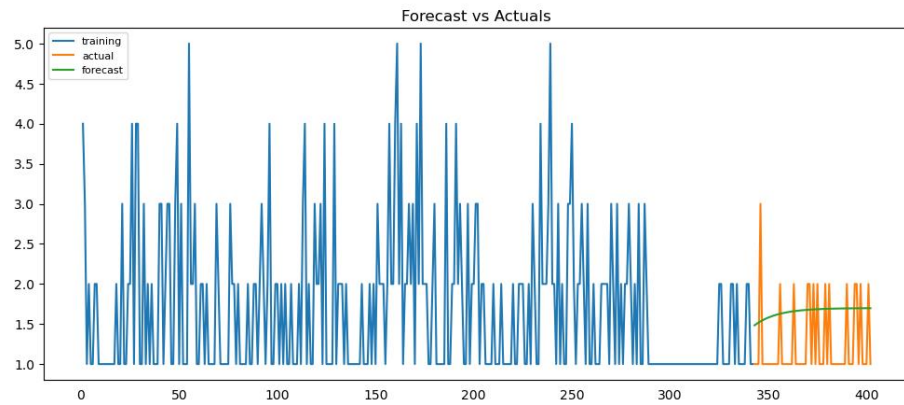
Con test [:15]

Resultado de evaluación de la metrica:-

```
MSE es:0.43729094669254354
MSE es:0.6176130565519109
RMSE es:0.6612797794372239
MAPE es:39.36541586454711
R2 es:-206.32959057583767
```

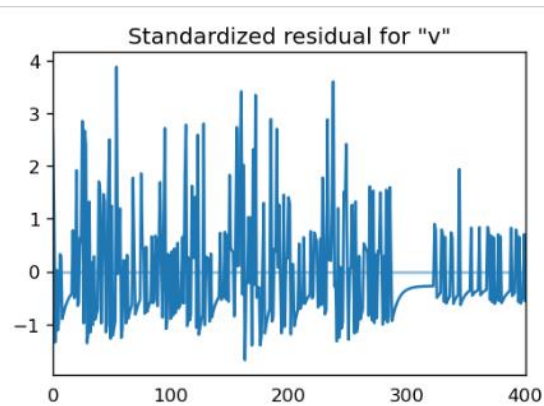
Con test completo 85:15 Se obtiene un mejor **MAPE**

Figura 112. Comportamiento predicción vs real para EPOC



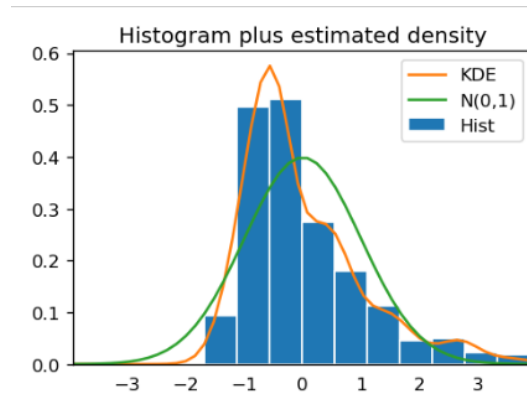
Fuente: Construcción propia

Figura 113. Estandarización residual por "v" para EPOC



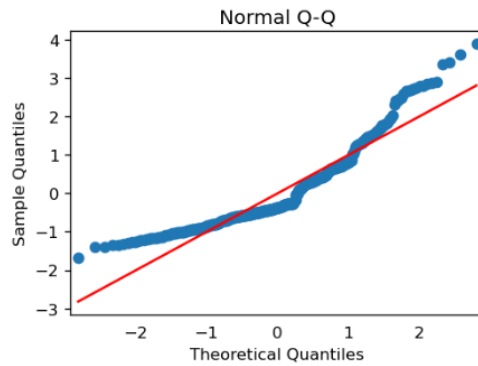
Fuente: Construcción propia

Figura 114. Histograma con estimación de densidad para EPOC



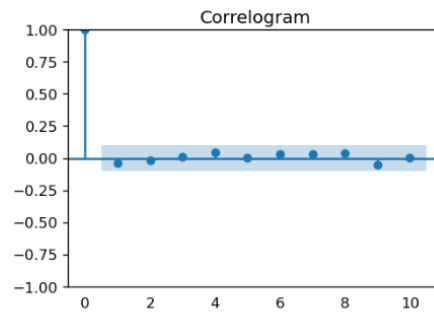
Fuente: Construcción propia

Figura 115. Normal Q-Q para EPOC



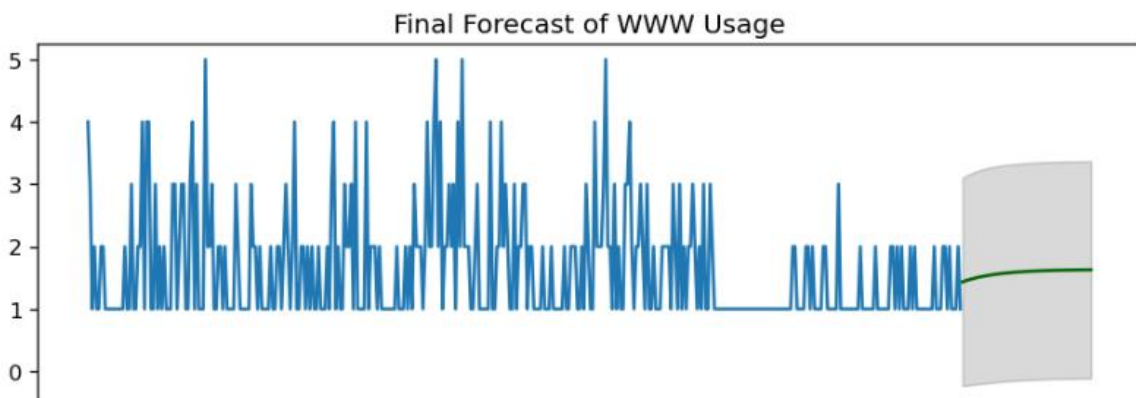
Fuente: Construcción propia

Figura 116. Correlograma para EPOC



Fuente: Construcción propia

Figura 117. Estimación final de predicciones futuras para EPOC



Fuente: Construcción propia

```
402    1.434644
403    1.446523
404    1.457690
405    1.468188
406    1.478058
407    1.487338
408    1.496062
409    1.504263
410    1.511974
411    1.519223
412    1.526038
413    1.532446
414    1.538469
415    1.544132
416    1.549456
417    1.554462
418    1.559168
419    1.563592
420    1.567751
421    1.571661
422    1.575337
423    1.578794
424    1.582043
425    1.585098
426    1.587969
427    1.590669
428    1.593208
429    1.595594
430    1.597838
431    1.599947
432    1.601930
433    1.603794
434    1.605547
435    1.607195
436    1.608744
437    1.610200
438    1.611569
439    1.612857
440    1.614067
441    1.615205
442    1.616274
443    1.617280
444    1.618225
445    1.619114
446    1.619950
447    1.620735
448    1.621474
449    1.622168
dtype: float64
```

Fiebre

Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

MSE es:20.103478814460072
MSE es:3.7178242025664074
RMSE es:4.48369031205993
MAPE es:32.74896340276028
R2 es:-12406.33379498068

Con test completo 75:25

Resultado de evaluación de la metrica:-

MSE es:19.439955928851774
MSE es:3.62730096990668
RMSE es:4.409076539237187
MAPE es:32.20916978698213
R2 es:-22948.52581734094

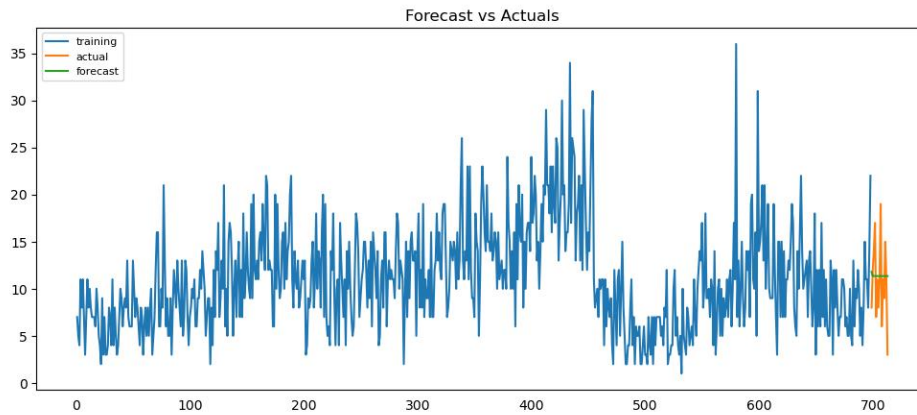
Con test [:15]

Resultado de evaluación de la metrica:-

MSE es:17.33945125098389
MSE es:3.27818427221456
RMSE es:4.164066672254887
MAPE es:28.80008251024012
R2 es:-1393.6550085070128

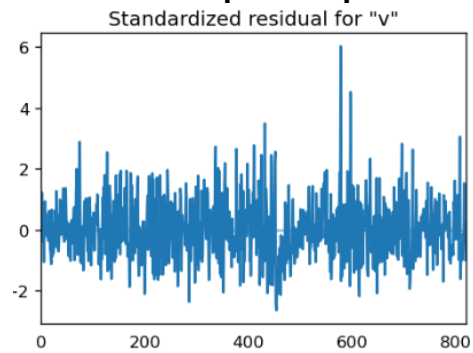
Con test [:15] Se obtiene un mejor **MAPE**

Figura 118. Comportamiento predicción vs real para Fiebre



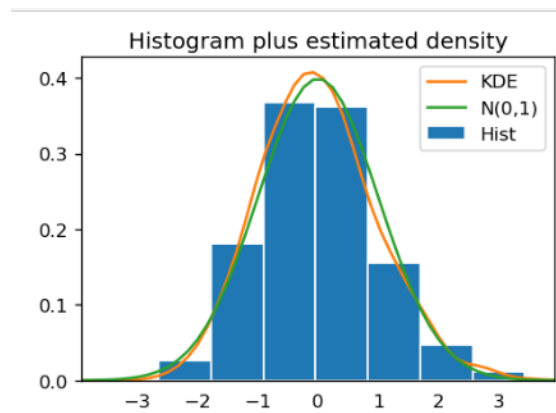
Fuente: Construcción propia

Figura 119. Estandarización residual por "v" para Fiebre



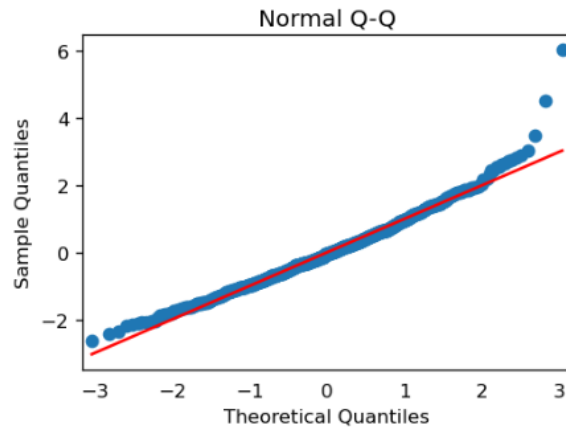
Fuente: Construcción propia

Figura 120. Histograma con estimación de densidad para Fiebre



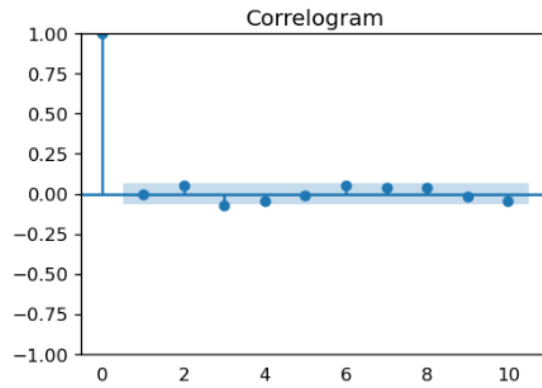
Fuente: Construcción propia

Figura 121. Normal Q-Q para Fiebre



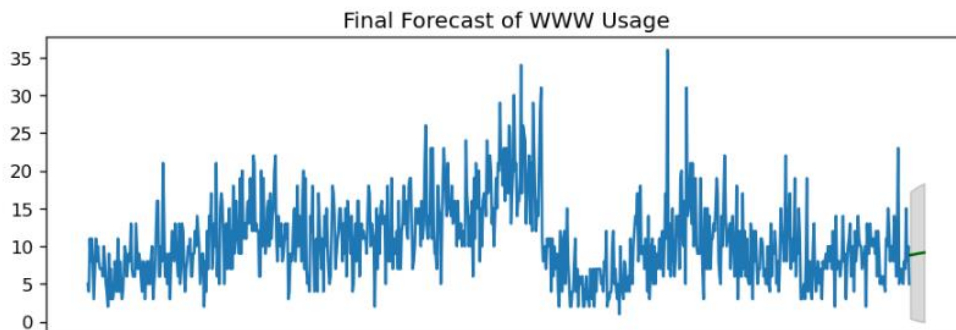
Fuente: Construcción propia

Figura 122. Correlograma para Fiebre



Fuente: Construcción propia

Figura 123. Estimación final de predicciones futuras para Fiebre



Fuente: Construcción propia

```
820    8.815001
821    8.841579
822    8.867712
823    8.893406
824    8.918669
825    8.943508
826    8.967930
827    8.991943
828    9.015553
829    9.038767
830    9.061592
831    9.084033
832    9.106098
833    9.127793
834    9.149124
dtype: float64
```

Infeción de vías urinarias

Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022 Con test completo 85:15

Resultado de evaluación de la metrica:-

```
MSE es:3.643078086130086
MSE es:1.5452155593164663
RMSE es:1.9086849101226966
MAPE es:39.65065223470109
R2 es:-1605.9198309866556
```

Con test completo 75:25

Resultado de evaluación de la metrica:-

```
MSE es:4.014720257413168
MSE es:1.6236845499204637
RMSE es:2.0036766848504195
MAPE es:41.05155493686455
R2 es:-359.80001941533516
```

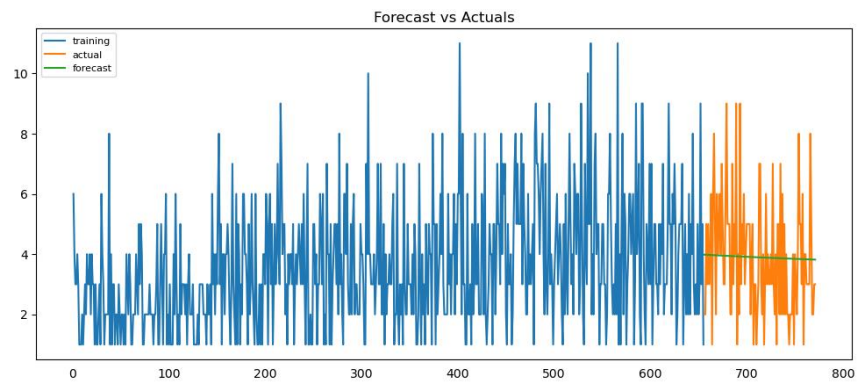
Con test [:15]

Resultado de evaluación de la metrica:-

```
MSE es:3.618371037325644
MSE es:1.6087651780396126
RMSE es:1.9022016289882742
MAPE es:40.519643934162104
R2 es:-68836.84051728019
```

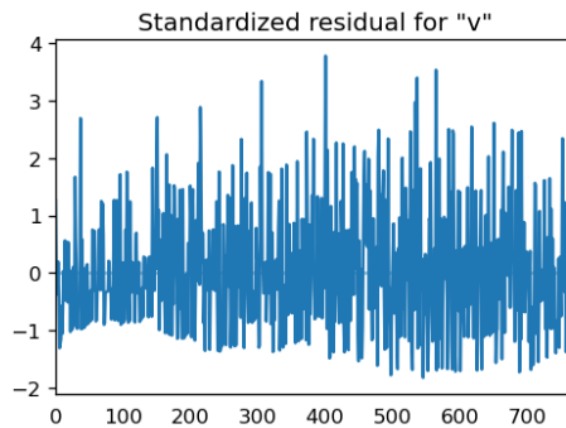
Con test completo 85:15 Se obtiene un mejor **MAPE**

Figura 124. Comportamiento predicción vs real para Infección de vías urinarias



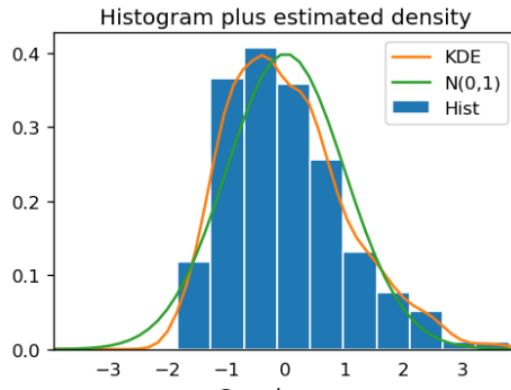
Fuente: Construcción propia

Figura 125. Estandarización residual por "v" para Infección de vías urinarias



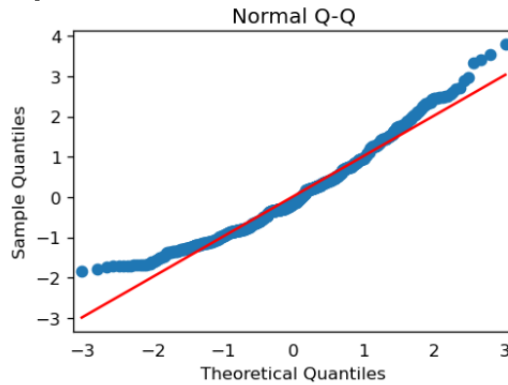
Fuente: Construcción propia

Figura 126. Histograma con estimación de densidad para Infección de vías ur inarias



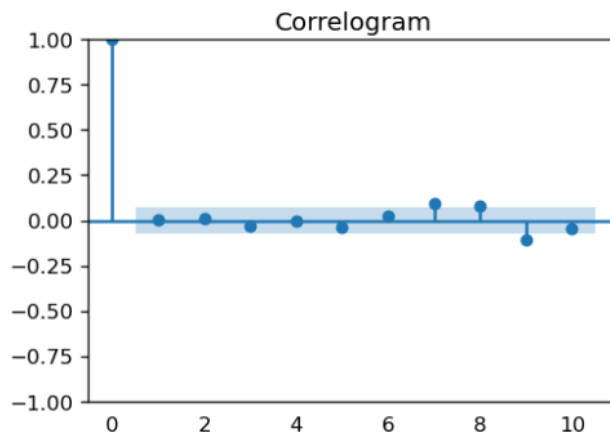
Fuente: Construcción propia

Figura 127. Normal Q-Q para Infección de vías urinarias



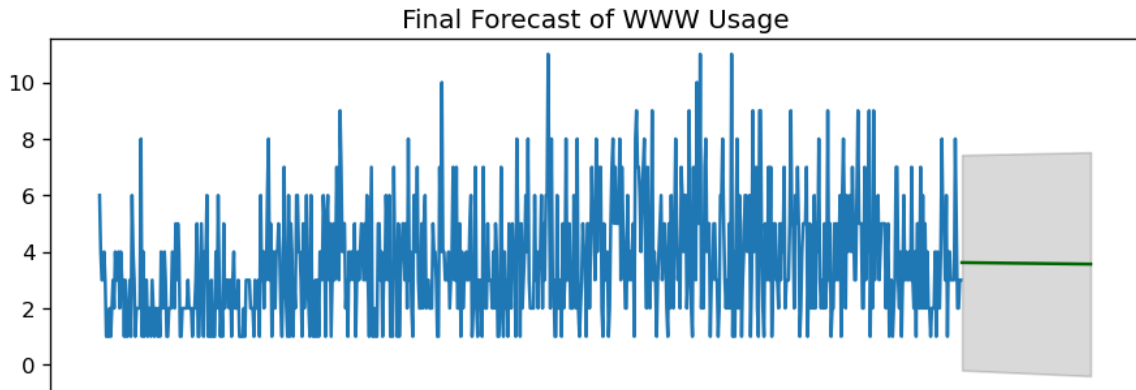
Fuente: Construcción propia

Figura 128. Correlograma para Infección de vías urinarias



Fuente: Construcción propia

Figura 129. Estimación final de predicciones futuras par Infección de vías urinarias



Fuente: Construcción propia

```

771    3.617611
772    3.617139
773    3.616667
774    3.616195
775    3.615722
      ...
882    3.565563
883    3.565097
884    3.564632
885    3.564166
886    3.563701
Length: 116, dtype: float64
    
```

Otros dolores abdominales

Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

```

MSE es:10.094010394243165
MSE es:2.6413834918382912
RMSE es:3.1771072368182924
MAPE es:31.618403150404916
R2 es:-145.3460891004158
    
```

Con test completo 75:25

Resultado de evaluación de la metrica:-

MSE es:12.052122414838038

MSE es:2.968170874623078

RMSE es:3.471616686046724

MAPE es:32.441989865606246

R2 es:-559.1293462560272

Con test [:15]

Resultado de evaluación de la metrica:-

MSE es:6.46742817490093

MSE es:2.1816698845755407

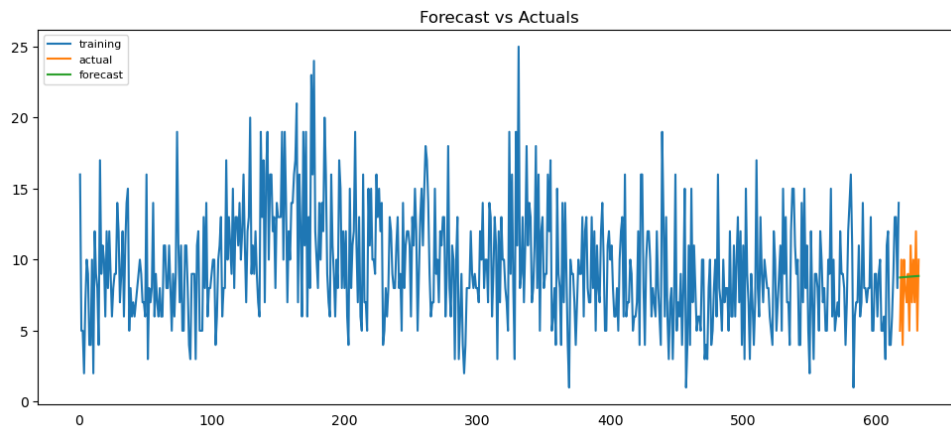
RMSE es:2.5431138737581

MAPE es:24.792016692471908

R2 es:-5833.202900192476

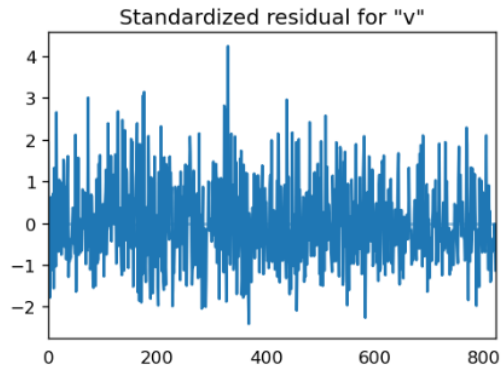
Con test completo 75:15 [:15] Se obtiene un mejor **MAPE**

Figura 130. Comportamiento predicción vs real para Otros dolores abdominales



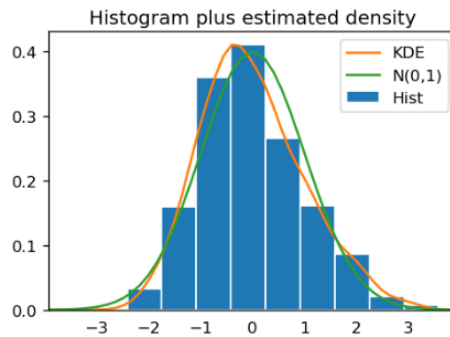
Fuente: Construcción propia

Figura 131. Estandarización residual por "v" para Otros dolores abdominales



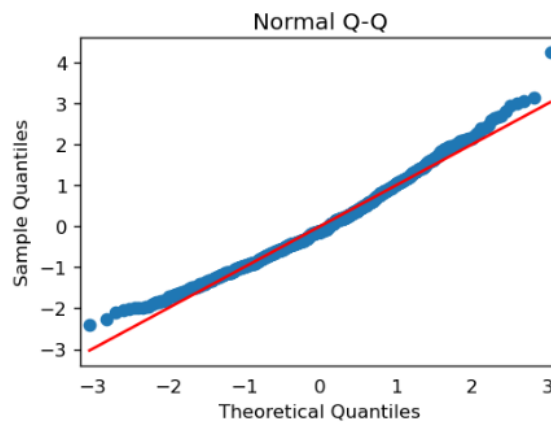
Fuente: Construcción propia

Figura 132. Histograma con estimación de densidad para Otros dolores abdominales



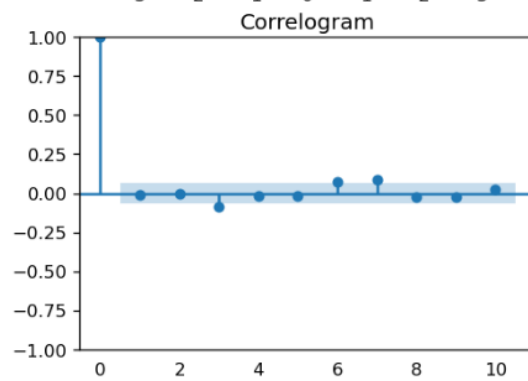
Fuente: Construcción propia

Figura 133. Normal Q-Q para Otros dolores abdominales



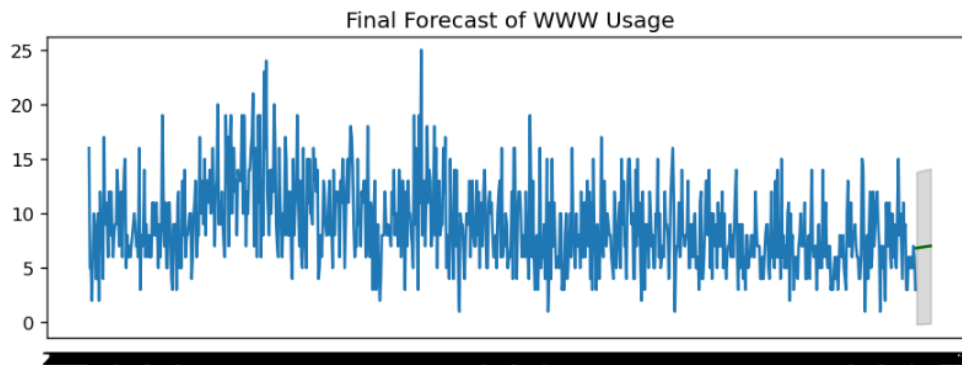
Fuente: Construcción propia

Figura 134. Correlograma para Otros dolores abdominales



Fuente: Construcción propia

Figura 135. Estimación final de predicciones futuras para Otros dolores abdominales



Fuente: Construcción propia

```
822    6.825700
823    6.840211
824    6.854602
825    6.868872
826    6.883024
827    6.897059
828    6.910976
829    6.924778
830    6.938465
831    6.952039
832    6.965499
833    6.978848
834    6.992086
835    7.005213
836    7.018232
dtype: float64
```

Traumatismos múltiples

Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022

Con test completo 85:15

Resultado de evaluación de la metrica:-

MSE es:1.4235732169558364
MSE es:0.7696163720526225
RMSE es:1.1931358753117083
MAPE es:51.812185526075424
R2 es:0.0

Con test completo 75:25

Resultado de evaluación de la metrica:-

MSE es:1.0586744313640835
MSE es:0.7100201624769238
RMSE es:1.0289190596757762
MAPE es:48.13038099710544
R2 es:0.0

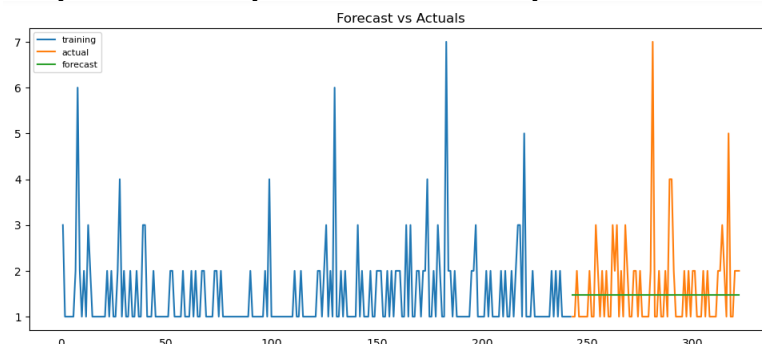
Con test [:15]

Resultado de evaluación de la metrica:-

MSE es:2.6609225225155155
MSE es:0.9637459595929606
RMSE es:1.631233435935984
MAPE es:64.8813958118628
R2 es:-1.349248013131043e+31

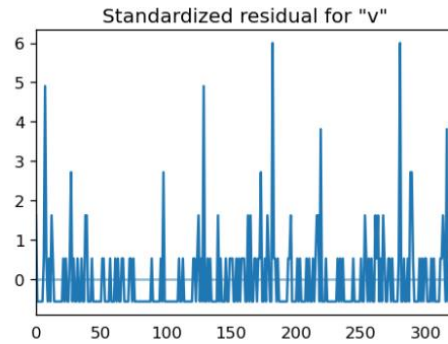
Con test completo 75:25 Se obtiene un mejor **MAPE**

Figura 136. Comportamiento predicción vs real para Traumatismos múltiples



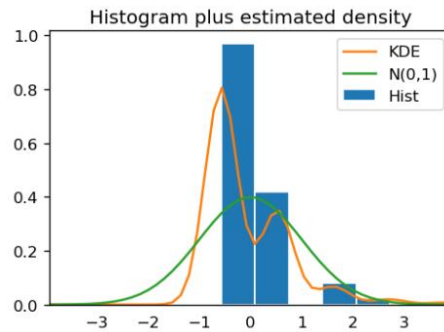
Fuente: Construcción propia

Figura 137. Estandarización residual por "v" para Traumatismos múltiples



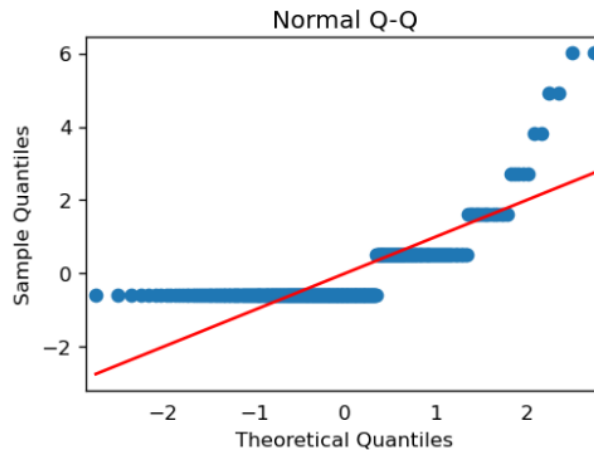
Fuente: Construcción propia

Figura 138. Histograma con estimación de densidad para Traumatismos múltiples



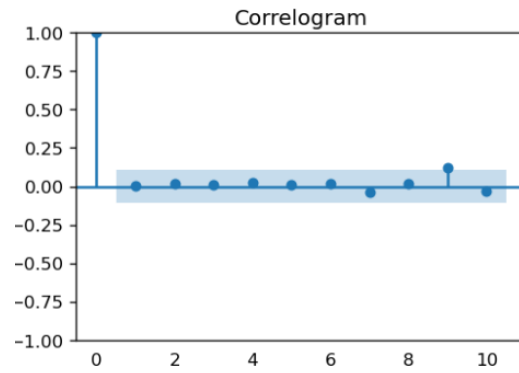
Fuente: Construcción propia

Figura 139. Normal Q-Q para Traumatismos múltiples



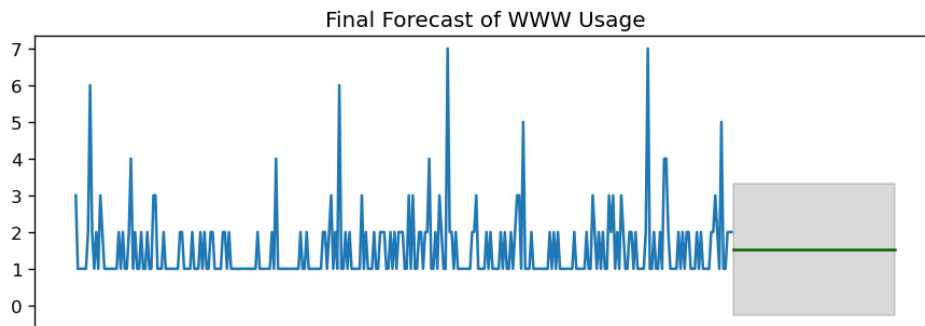
Fuente: Construcción propia

Figura 140. Correlograma para Traumatismos múltiples



Fuente: Construcción propia

Figura 141. Estimación final de predicciones futuras para Traumatismos múltiples



Fuente: Construcción propia

```
322    1.521739
323    1.521739
324    1.521739
325    1.521739
326    1.521739
...
397    1.521739
398    1.521739
399    1.521739
400    1.521739
401    1.521739
Length: 80, dtype: float64
```

Durante la experimentación, se realizó las validaciones para el total de pacientes y por cada una de las patologías críticas seleccionadas, logrando para cada uno de los análisis el mejor MAPE teniendo en cuenta los 3 tipos de validaciones realizadas, una vez logrado este resultado se graficó el comportamiento real vs la predicción y por ultimo se grafico el comportamiento a futuro de la predicción para los n periodos óptimos. Como los resultados obtenidos están susceptibles a mejoras, en el siguiente capitulo se procederá con la ejecución de una propuesta que permitirá el mejoramiento de los resultados obtenidos.

6. RESULTADOS OBTENIDOS

Como el estudio era la predicción de la morbilidad hospitalaria en la institución de salud, teniendo en cuenta las patologías más críticas, a continuación, se presentan los resultados del análisis de manera consolidada para el total y por diagnóstico:

Tabla 13. Resultados del MAPE(%) con toda la data

CON TODA LA DATA					
Patología	MAPE(%)			MEJOR OPCIÓN	Observaciones a predecir
	[85:15]	[75:25]	[:15]		
Todas	12,83	14,03	10,34	10,34	15
Cólico	45,59	39,50	51,32	39,50	195
Dolor abdominal	43,53	42,65	51,79	42,65	171
EPOC	35,84	36,83	39,37	35,84	60
Fiebre	32,75	32,21	28,80	28,80	15
Infección vías urinaria	39,65	41,05	40,52	39,65	116
Otros dolores abdominales	31,62	32,44	24,79	24,79	15
Traumatismos múltiples	51,81	48,13	64,88	48,13	80

Fuente: Construcción propia

Los resultados anteriores nos muestran claramente que para el análisis el total de patologías el MAPE es muy bueno (10,34%), mientras que al analizar las series de tiempo por separado se obtienen resultados que varían entre 24.8 % - 48,13%, por lo cual, nos lleva a replantear el modelo y realizar un nuevo experimento que nos permita mejorar el MAPE, teniendo en cuenta estos resultados, se procede a eliminación de la variabilidad de las series de tiempo eliminando los outliers. Con este nuevo experimento se obtuvo los siguientes resultados:

Tabla 14. Resultados del MAPE(%) eliminando Outliers

ELIMINANDO OUTLIERS					
Patología	MAPE(%)				Observaciones a predecir
	[85:15]	[75:25]	[:15]	MEJOR OPCIÓN	
Todas	12,83	14,06	10,32	10,32	15
Cólico	42,54	38,10	40,92	38,10	193
Dolor abdominal	39,99	40,86	46,16	39,99	101
EPOC	34,12	34,47	32,84	32,84	15
Fiebre	31,70	31,46	31,02	31,02	15
Infección vías urinaria	38,88	40,48	38,61	38,61	15
Otros dolores abdominales	30,83	30,52	22,44	22,44	15
Traumatismos múltiples	36,38	39,48	34,42	34,42	15

Fuente: Construcción propia

Para los anteriores resultados, se puede detallar mas a profundidad los resultados y análisis realizado a cada patología en los anexos del presente trabajo.

Los resultados anteriores nos muestran claramente que al analizar el total de patologías el MAPE es muy bueno (10,32%), mientras que al analizar las series de tiempo por separado se obtienen resultados que varían entre 22.4 % - 39.99% lo que nos muestra claramente que se presentó una mejora en los resultados comparado con el experimento anterior.

A continuación, se presenta el comparativo de los resultados del MAPE de ambos experimentos:

Tabla 15. Comparativo del MAPE(%) de los 2 experimentos

Patología	MEJOR MAPE	
	Toda la Data	Sin Outliers
Todas	10,34	10,32
Cólico	39,50	38,1
Dolor abdominal	42,65	39,99
EPOC	35,84	32,84
Fiebre	28,80	31,02

Infección vías urinaria	39,65	38,61
Otros dolores abdominales	24,79	22,44
Traumatismos múltiples	48,13	34,42

Fuente: Construcción propia

7. CONCLUSIONES Y TRABAJO FUTURO

El modelo desarrollado de series de tiempo ARIMA permitió predecir el comportamiento de la morbilidad hospitalaria de las patologías más críticas en un intervalo 60% - 77,5% por lo cual podemos decir que para ciertas patologías la capacidad predictora deberá calcularse con otras técnicas ya que los resultados no fueron los esperados como lo fue en dolor abdominal e infección de vías urinarias con un % de precisión del 60% y 61.39% respectivamente, y por el contrario para las patologías de Cólico, EPOC, fiebre, otros dolores abdominales y traumatismos múltiples se obtuvo % de precisión entre el 61.9% - 77.5%, lo cual es significativo y de relevancia para implementar y utilizar estas predicciones para la planeación de la atención y lograr la eficiencia en la prestación de los servicios de atención de los pacientes que lleguen por estas patologías, impactando en la oportunidad en la atención, en la operación de la institución y en el flujo de recursos.

Adicionalmente, con el objetivo de revisar como los modelos desarrollados impactaban en la predicción se contempló en el análisis el total de pacientes que ingresaron intrahospitalariamente, lo cual nos mejoro notablemente los resultados obteniéndose una capacidad predictora de 89,68%, lo cual nos da un punto importante de partida para realizar análisis de forma global en la atención tendiendo en cuenta el total de pacientes que ingresaran diariamente a la institución, y de esta forma planear la atención de forma integral, impactando en indicadores claves como la oportunidad y minimización de costos hospitalarios.

Teniendo en cuenta lo anterior, es importante dejar como precedente que estos modelos son susceptible de mejoras, por lo tanto, como trabajo futuro se pretenderá realizar este análisis predictivo utilizando otras técnicas de machine learning como lo es la utilización de redes neuronales, y adicionalmente al análisis visto de las diferentes patologías clínicas, la idea sería ampliar el análisis e incluir más patologías clínicas como sea posible.

8. BIBLIOGRAFÍA

Edwin John Saldaña Valqui. (2015). Modelo Predictivo De Minería De Datos De Apoyo A La Gestión Hospitalaria Sobre La Morbilidad De Pacientes Hospitalizados. <https://repositorio.upao.edu.pe>

Marcos Estupiñán-Ramírez , Rita Tristancho-Ajamil , María Consuelo Company Sancho, Hilda Sánchez-Janáriz. (2019). Comparación de modelos predictivos para la selección de pacientes de alta complejidad. <https://www.sciencedirect.com/science/article/pii/S0213911117301693>

Mayra Alejandra Jiménez Manjarrez. (2019). Aplicación De Analítica De Datos Para Predicción De Infección Respiratoria Aguda En Colombia. <https://repositorio.uniandes.edu.co>

Francesc Güell y Martí Zamora de Amalfi Analytics. (2020). Aplicación de machine learning en la gestión de las personas con enfermedades crónicas. <https://hospitecnia.com/tecnologia/inteligencia-artificial/aplicacion-machine-learning-gestion-personas-enfermedades-cronicas/>

Luis Alberto Corona Martínez, Mercedes Fonseca Hernández (2010). Un modelo simplificado del proceso de atención médica. Implicaciones asistenciales, docentes e investigativas.

Ministerio de Salud (2013). <https://www.dnp.gov.co/programas/desarrollo-social/subdireccion-de-salud/Paginas/prestacion-de-servicios-de-salud.aspx>

Alicia Rodríguez Velasco. Capítulo 2: Definición de patología y su campo de estudio

Dr. Alan López (2021). Evaluación de Modelos Predictivos. <https://machinelearningenespanol.com/2021/01/12/evaluacion-de-modelos-predictivos/>

Método de clasificación. <https://bookdown.org/content/2274/metodos-de-clasificacion.html>

Estadística y Machine Learning con R. <https://bookdown.org/content/2274/metodos-de-clasificacion.html>

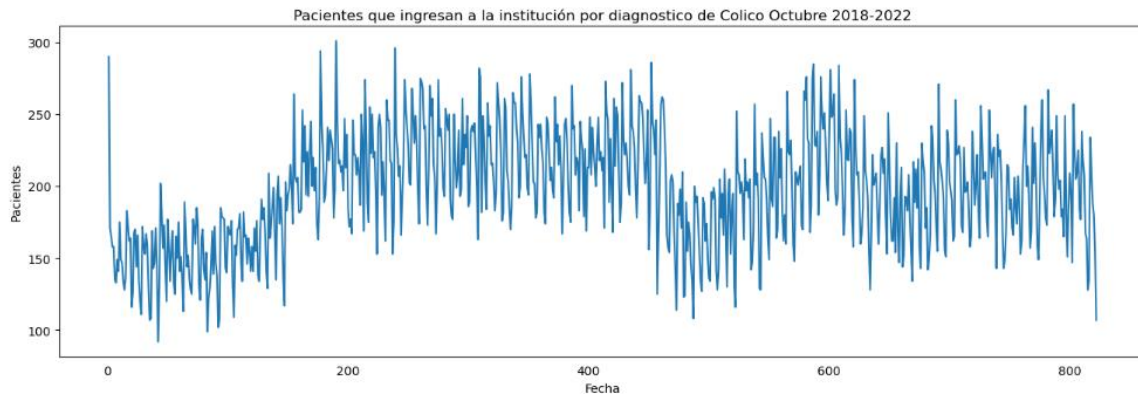
Azevedo, Ana; Zantos, Manuel Filipe. KDD, SEMMA and CRISP-DM: a parallel overview. (2008). CRISP-DM: La metodología para poner orden en los proyectos

Rojas-Jimenez, K. (2022). Ciencia de Datos para Ciencias Naturales. https://bookdown.org/keilor_rojas/CienciaDatos/

9. ANEXOS

Anexo 1. Total de pacientes que ingresan intrahospitalariamente

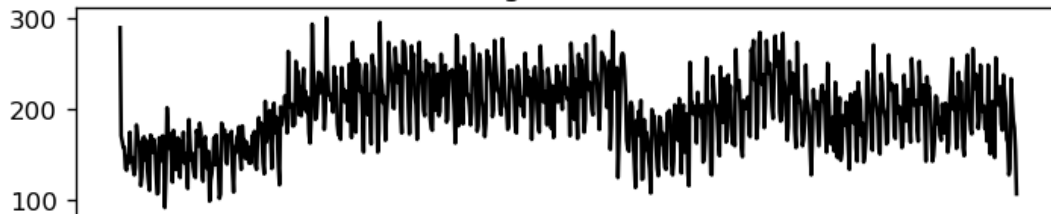
Pacientes ingresados intrahospitalariamente con periodicidad diaria octubre 2018 – diciembre 2022



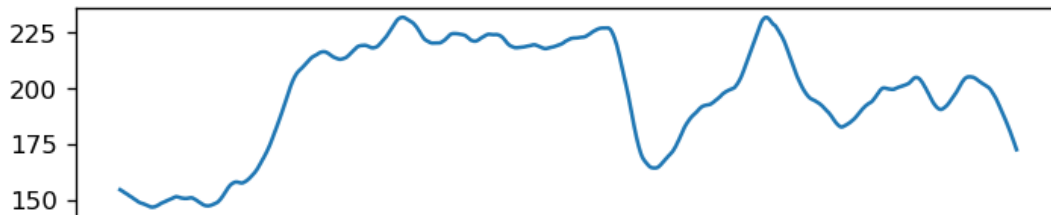
Fuente: Construcción propia

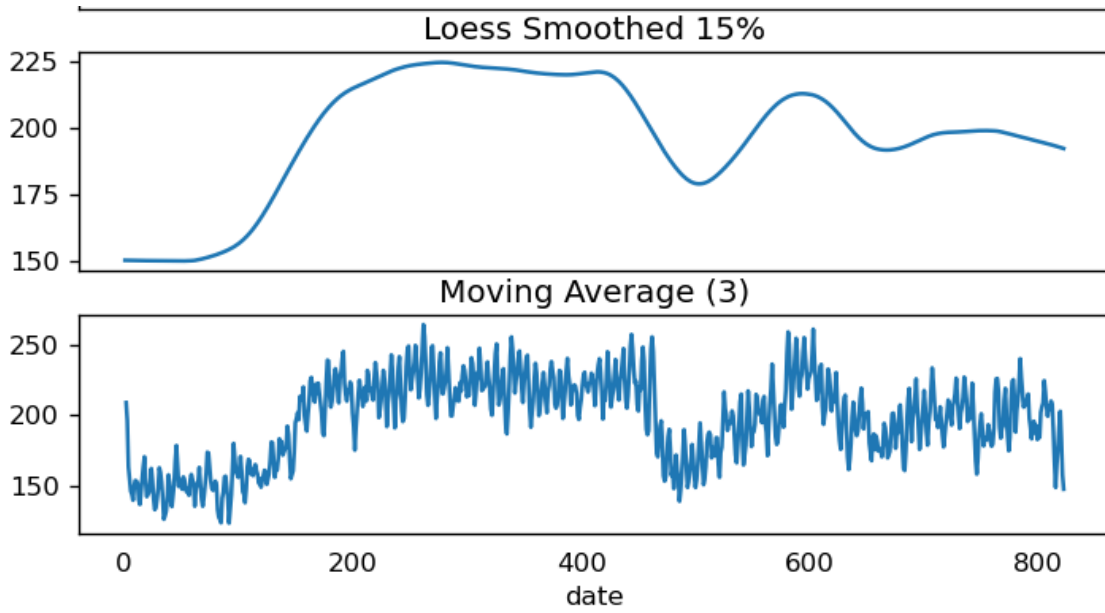
How to Smoothen a Time Series

Original Series

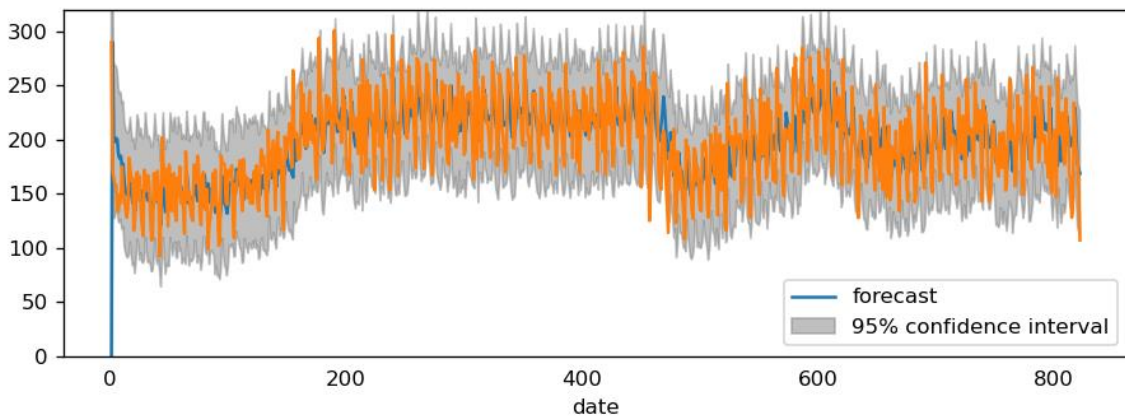


Loess Smoothed 5%

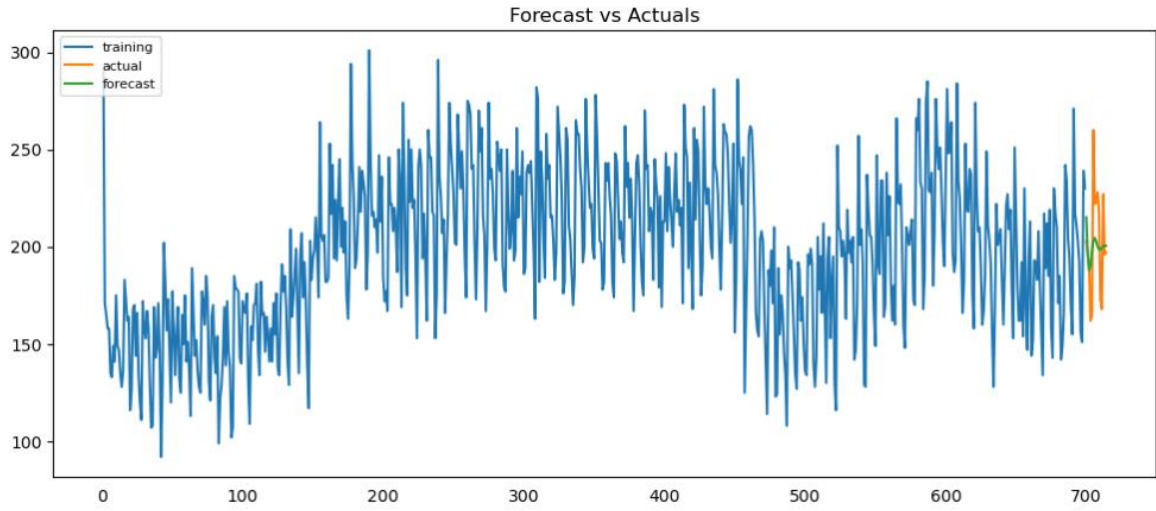




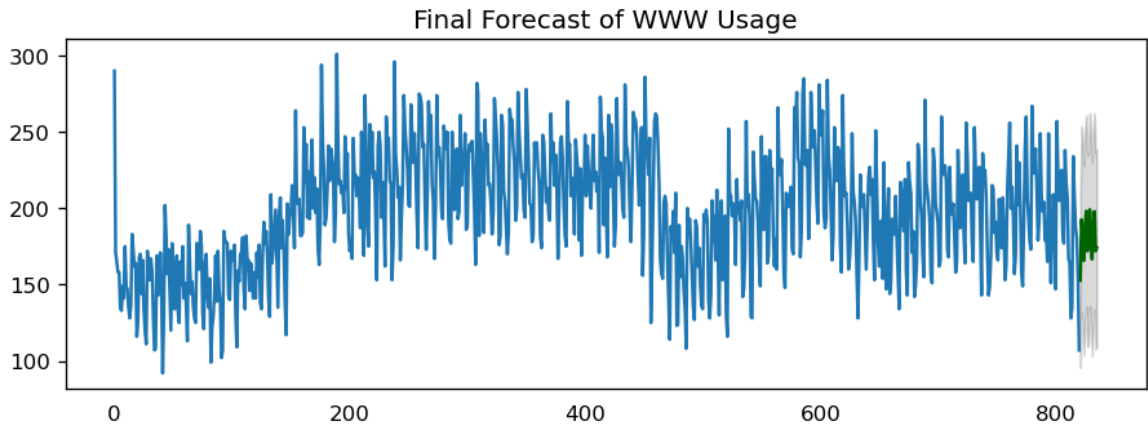
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

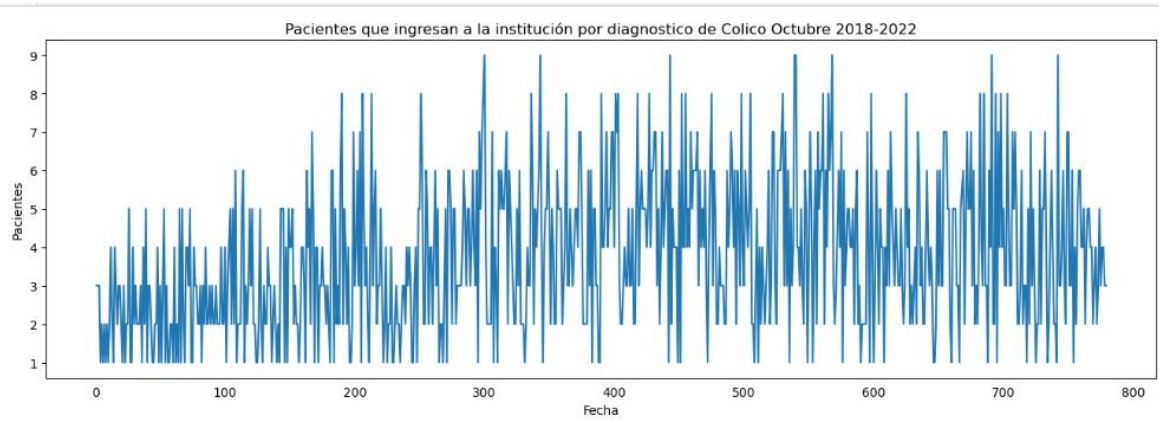


Fuente: Construcción propia

```
Out[110]: 821    152.731710
          822    192.604581
          823    184.073367
          824    165.745901
          825    188.074187
          826    198.152093
          827    171.979871
          828    173.829836
          829    199.135083
          830    185.994505
          831    166.667674
          832    188.403539
          833    197.937446
          834    172.020432
          835    174.112909
          dtype: float64
```

Anexo 2. Total de pacientes que ingresan intrahospitalariamente por Cólico

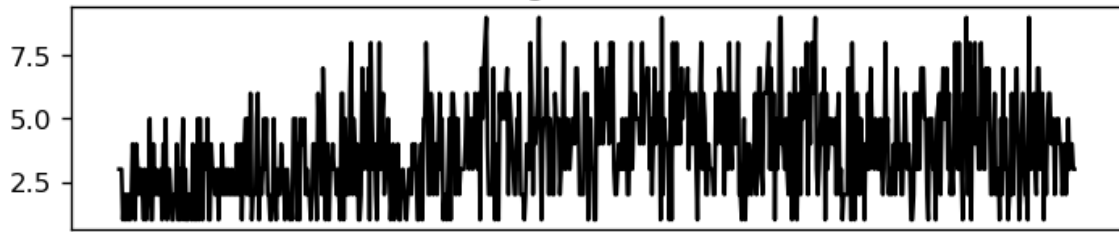
Pacientes ingresados intrahospitalariamente por Cólico con periodicidad diaria octubre 2018 – diciembre 2022



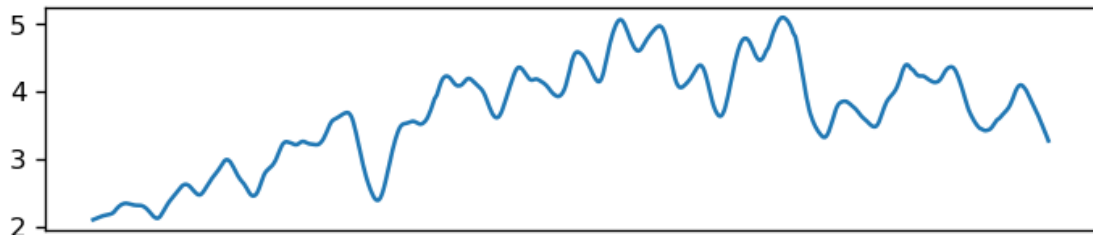
Fuente: Construcción propia

How to Smoothen a Time Series

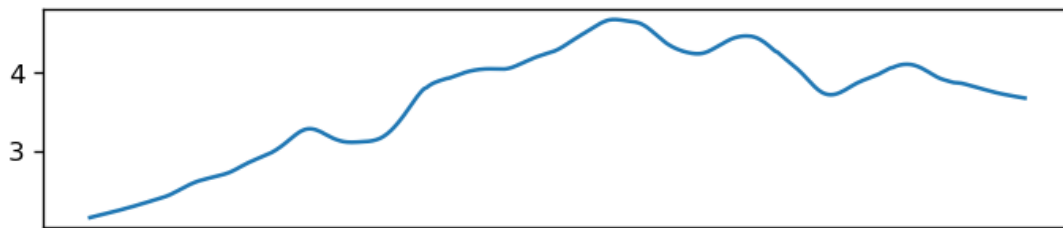
Original Series



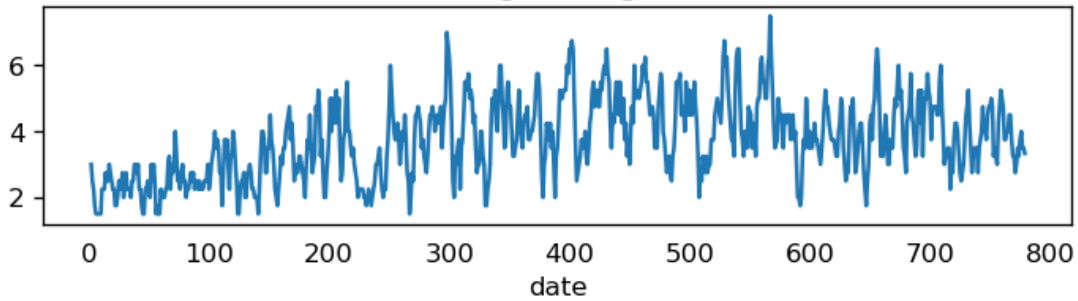
Loess Smoothed 5%



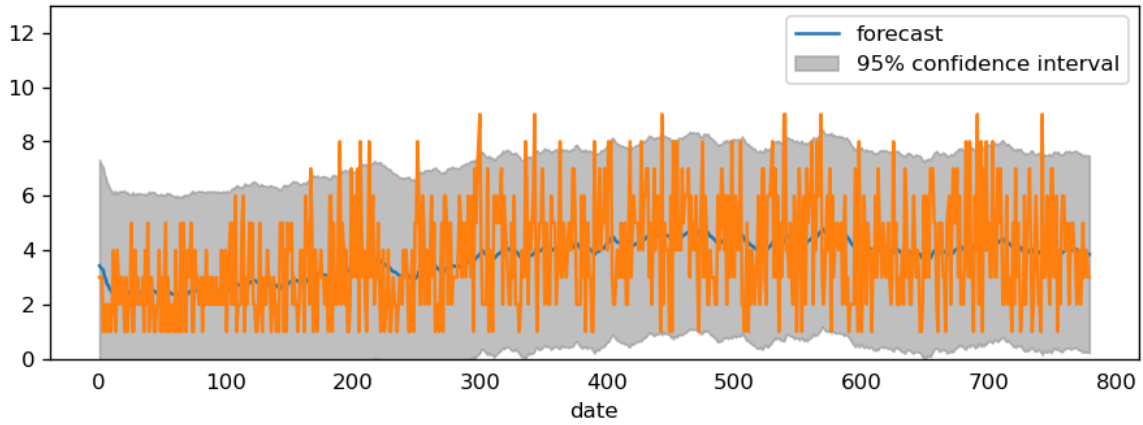
Loess Smoothed 15%



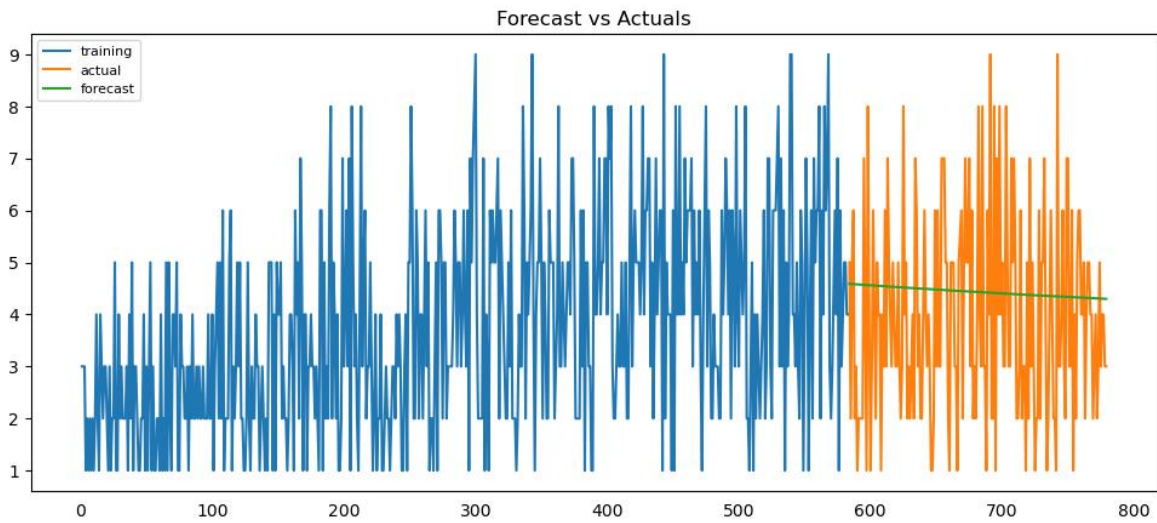
Moving Average (3)



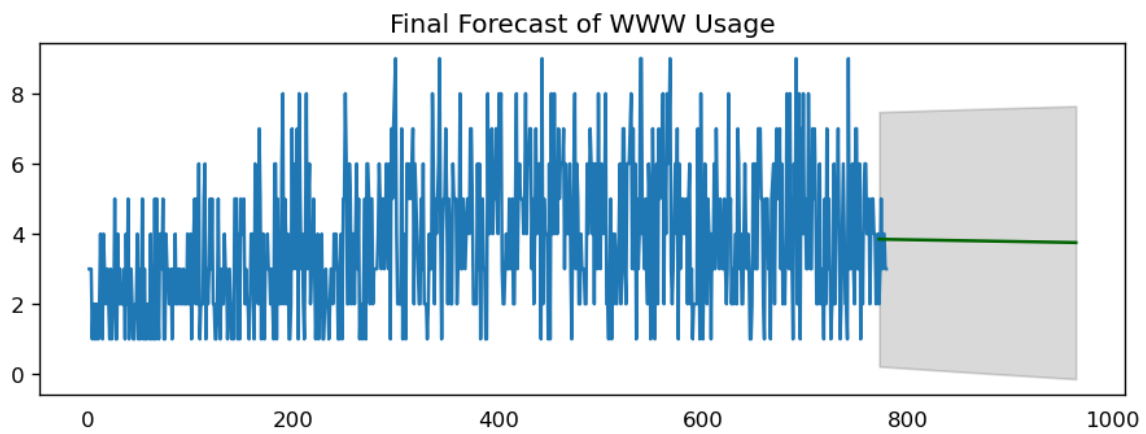
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

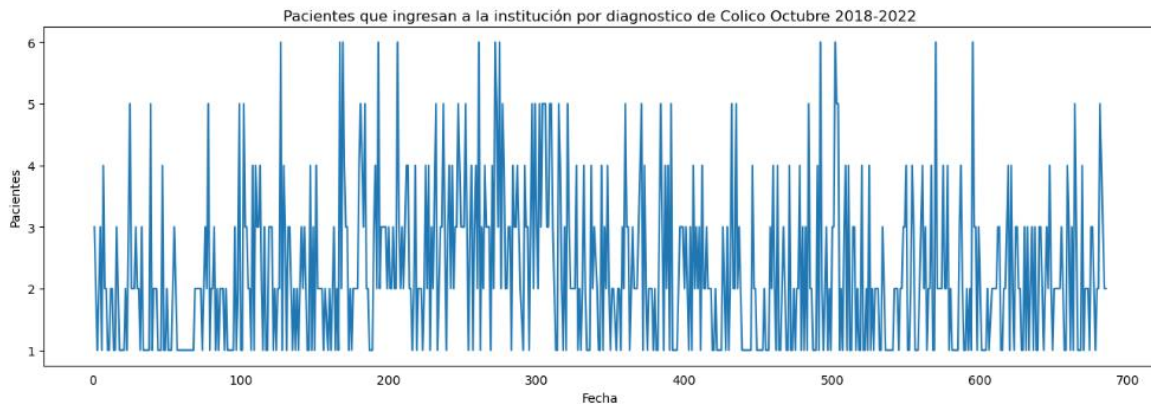


Fuente: Construcción propia

```
772  3.847153
773  3.846648
774  3.846144
775  3.845639
776  3.845135
...
960  3.753459
961  3.752967
962  3.752475
963  3.751983
964  3.751491
Length: 193, dtype: float64
```

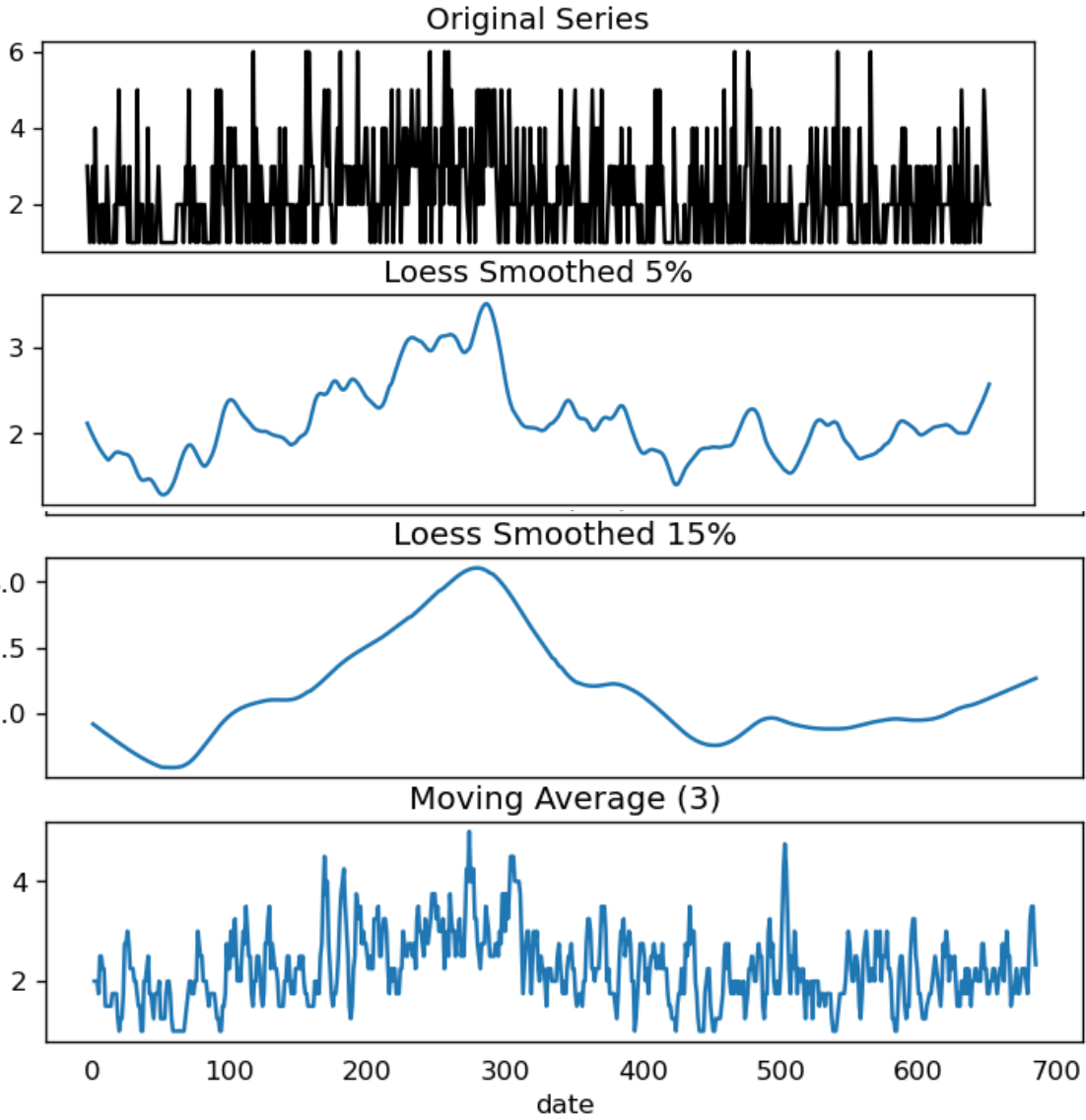
Anexo 3. Total de pacientes que ingresan intrahospitalariamente por Dolor abdominal

Pacientes ingresados intrahospitalariamente por Dolor Abdominal con periodicidad diaria octubre 2018 – diciembre 2022

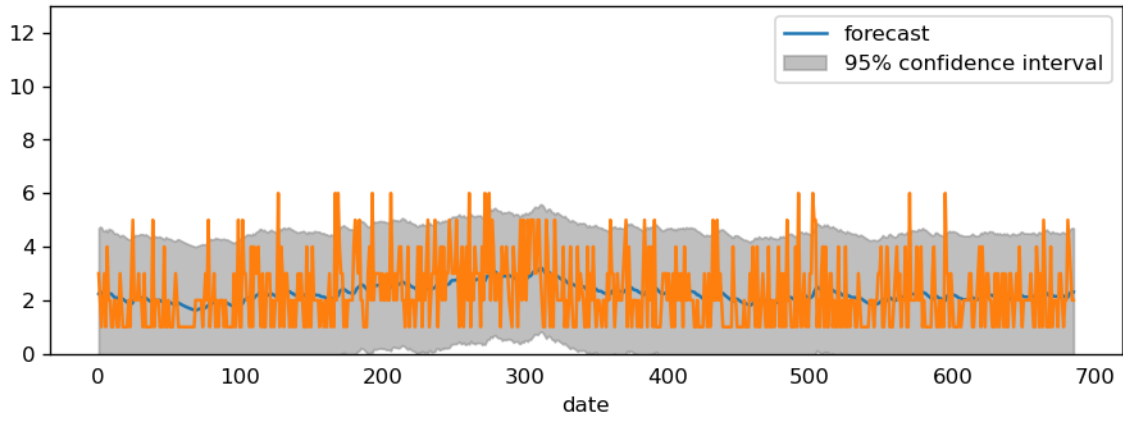


Fuente: Construcción propia

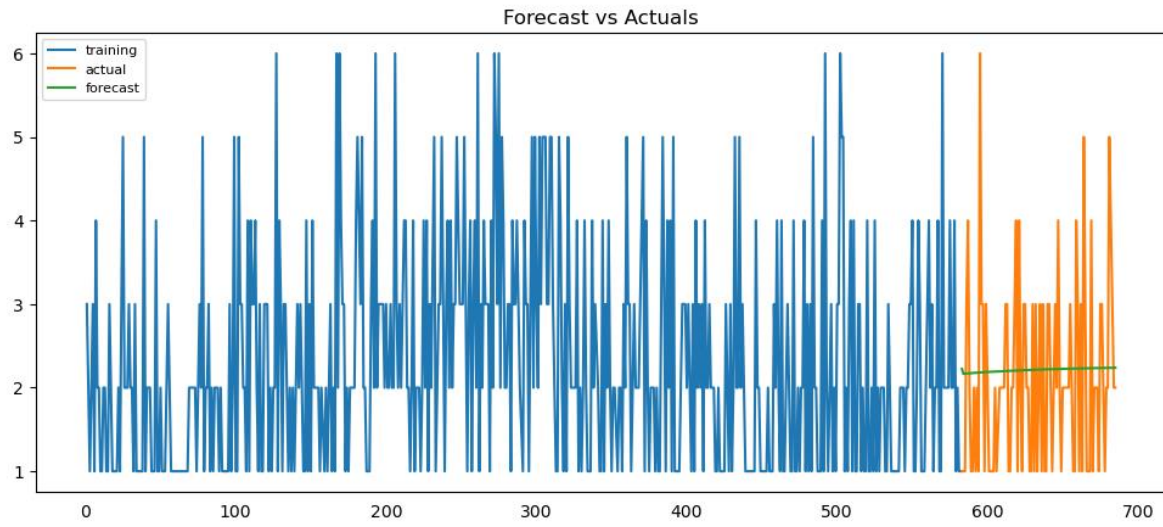
How to Smoothen a Time Series



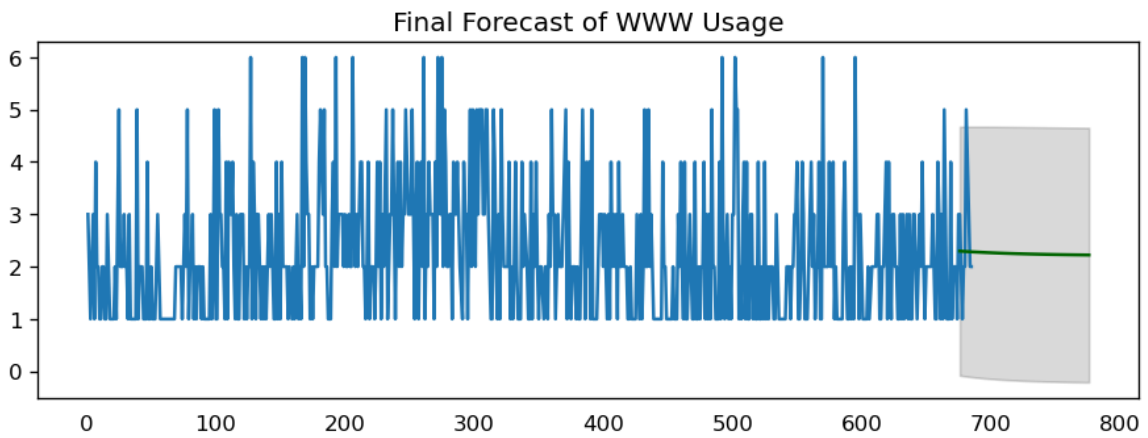
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

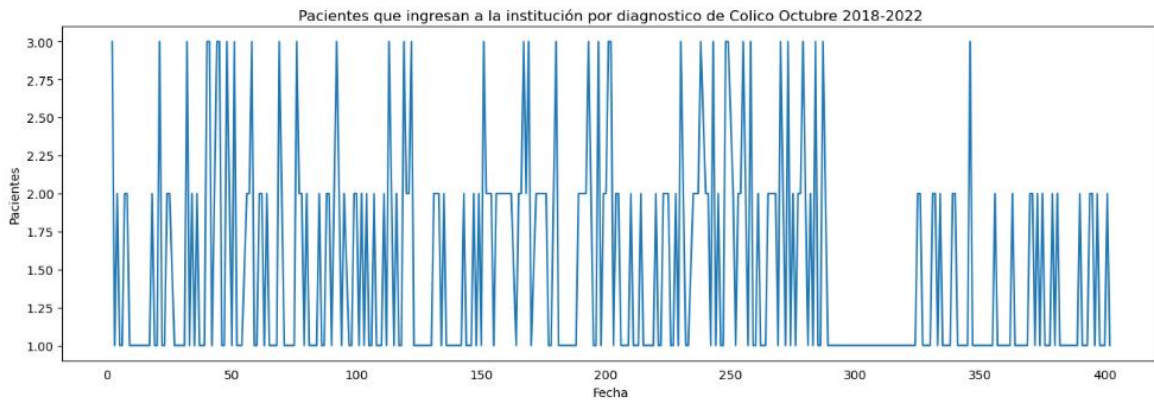


Fuente: Construcción propia

```
676 2.297227
677 2.295643
678 2.294086
679 2.292557
680 2.291055
...
772 2.223787
773 2.223501
774 2.223220
775 2.222945
776 2.222674
Length: 101, dtype: float64
```

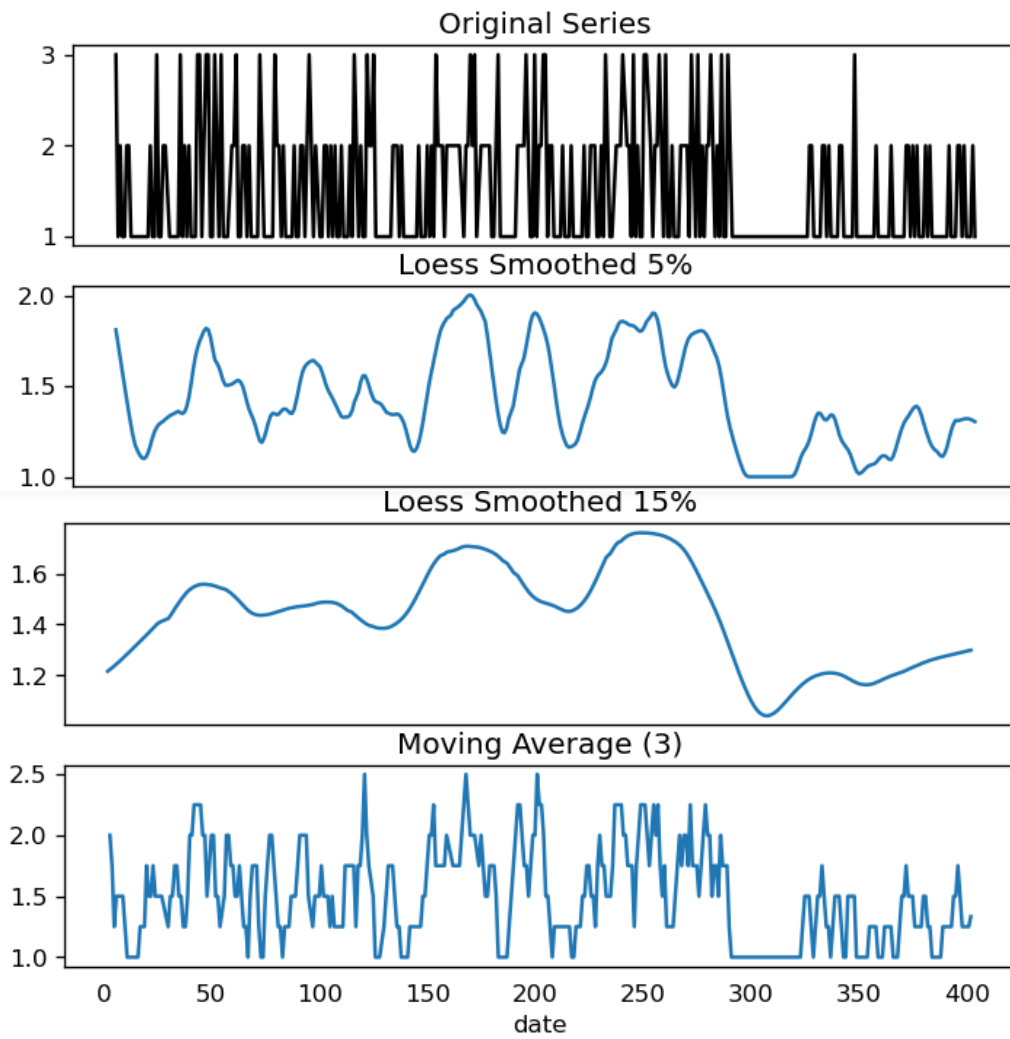
Anexo 4. Total de pacientes que ingresan intrahospitalariamente por EPOC

Pacientes ingresados intrahospitalariamente por EPOC con periodicidad diaria octubre 2018 – diciembre 2022

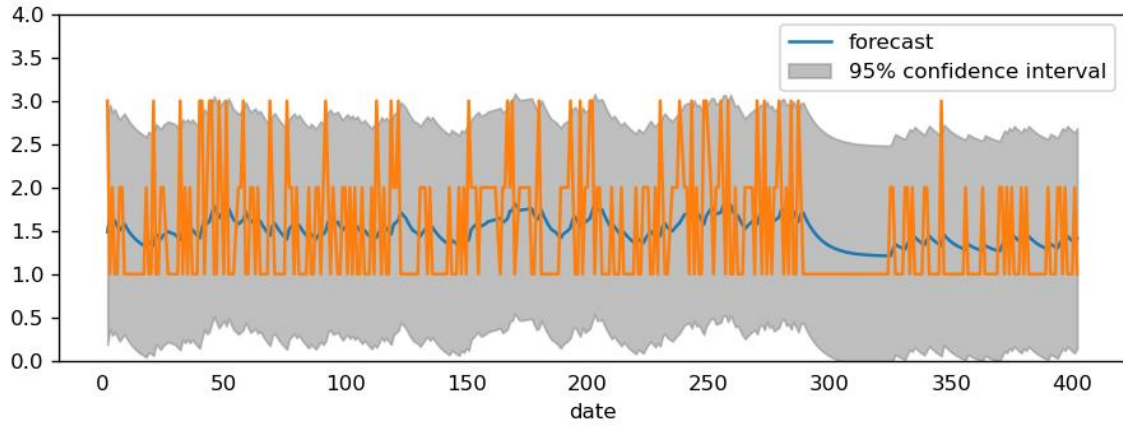


Fuente: Construcción propia

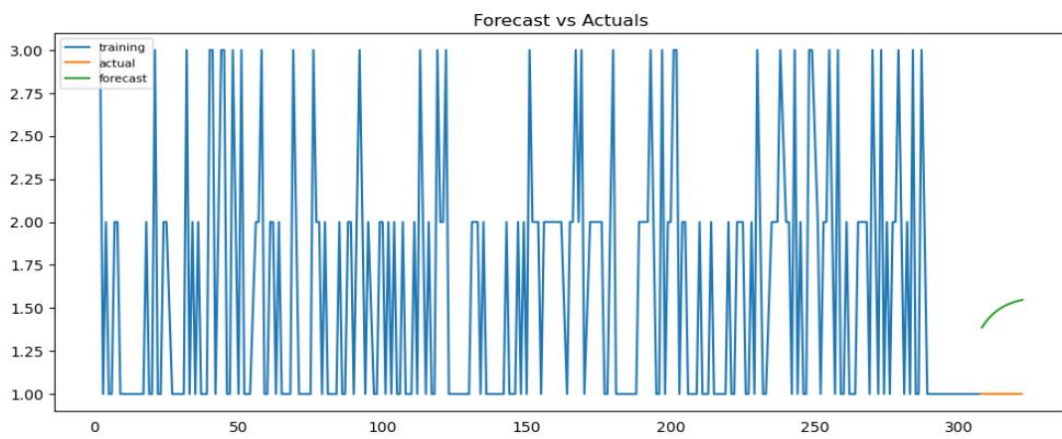
How to Smoothen a Time Series



Fuente: Construcción propia



Fuente: Construcción propia

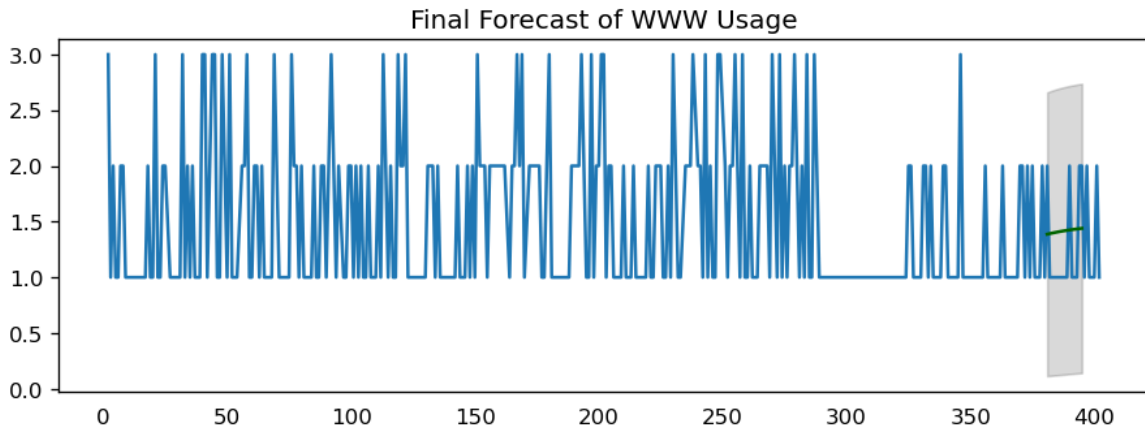


Fuente: Construcción propia

```

381    1.386890
382    1.392057
383    1.396958
384    1.401606
385    1.406016
386    1.410199
387    1.414166
388    1.417930
389    1.421500
390    1.424886
391    1.428098
392    1.431145
393    1.434035
394    1.436776
395    1.439376
dtype: float64

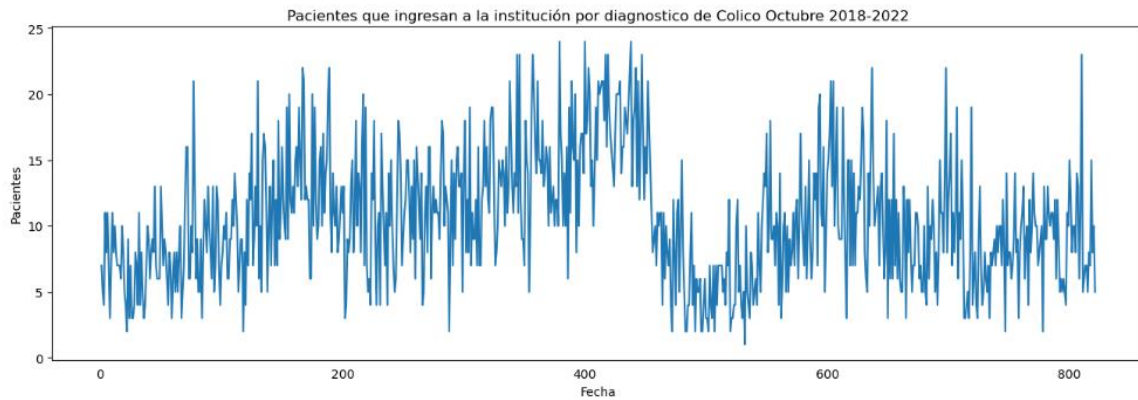
```



Fuente: Construcción propia

Anexo 5. Total de pacientes que ingresan intrahospitalariamente por Fiebre

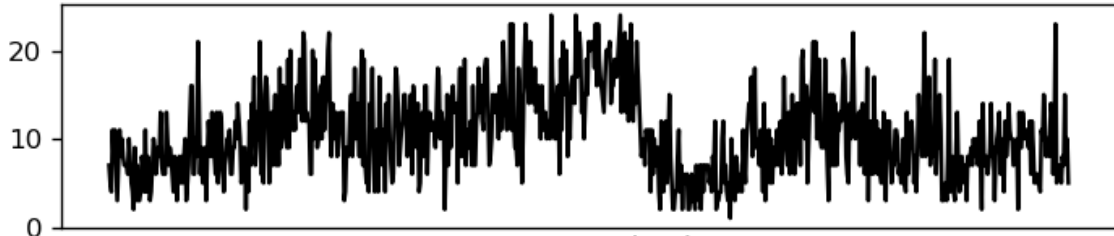
Pacientes ingresados intrahospitalariamente por FIEBRE con periodicidad diaria octubre 2018 – diciembre 2022



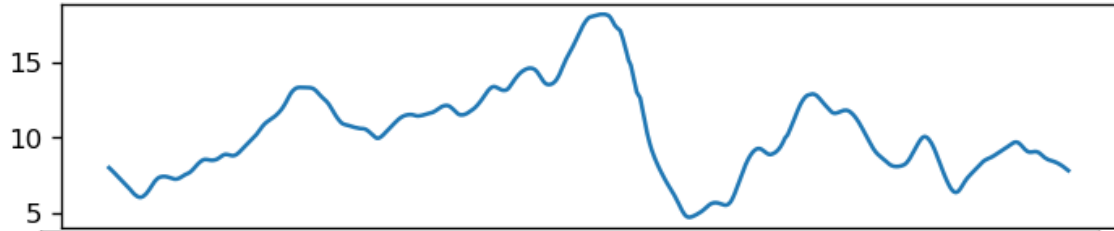
Fuente: Construcción propia

How to Smoothen a Time Series

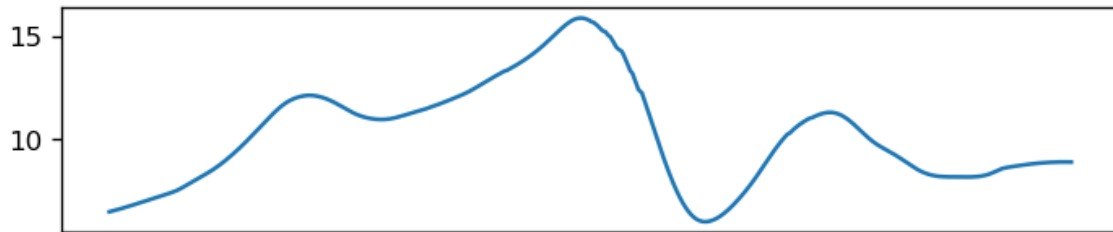
Original Series



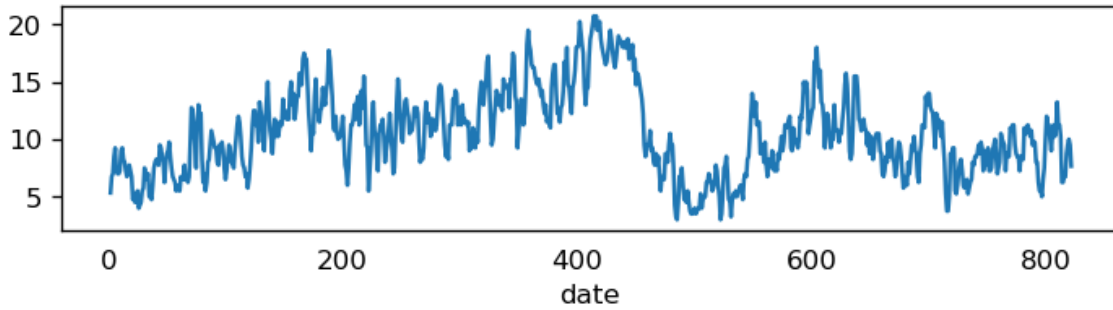
Loess Smoothed 5%



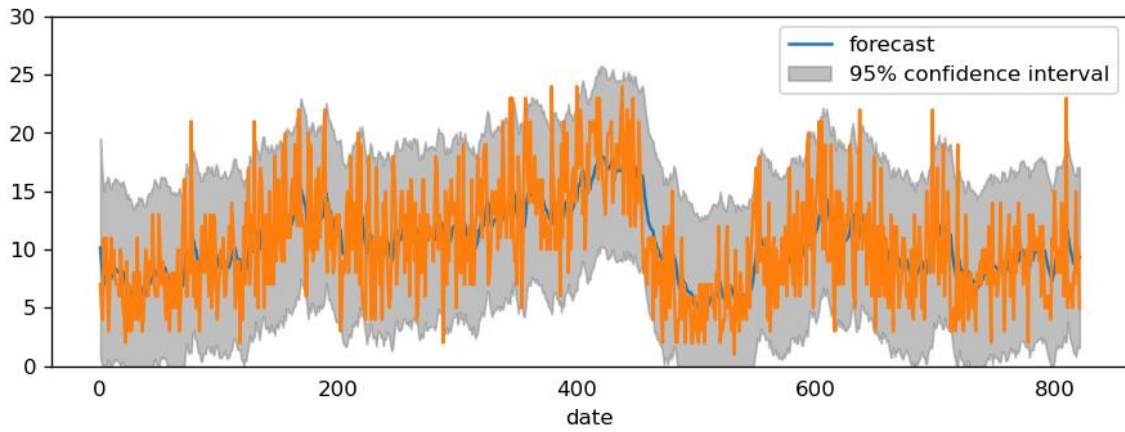
Loess Smoothed 15%



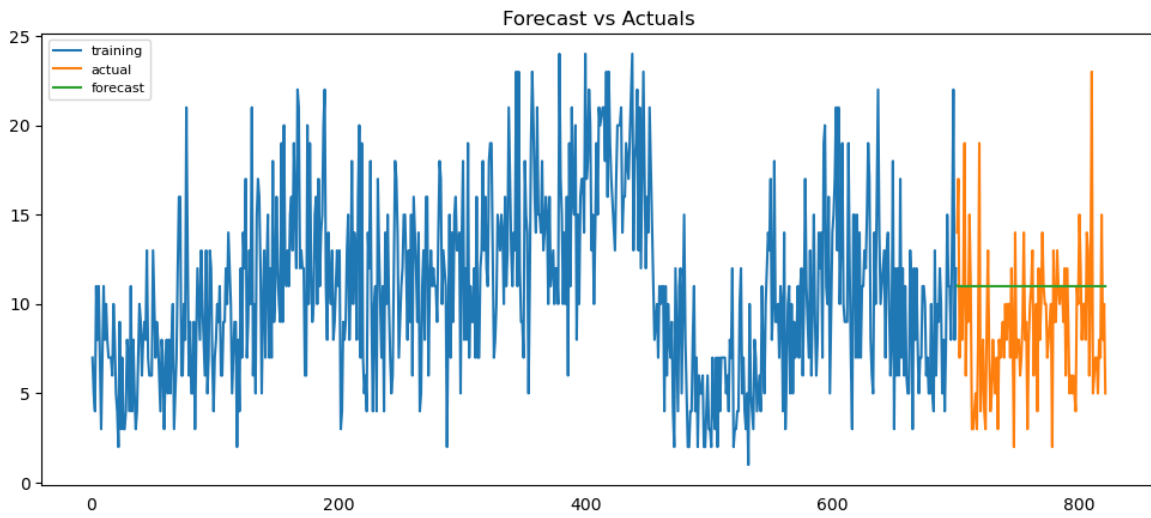
Moving Average (3)



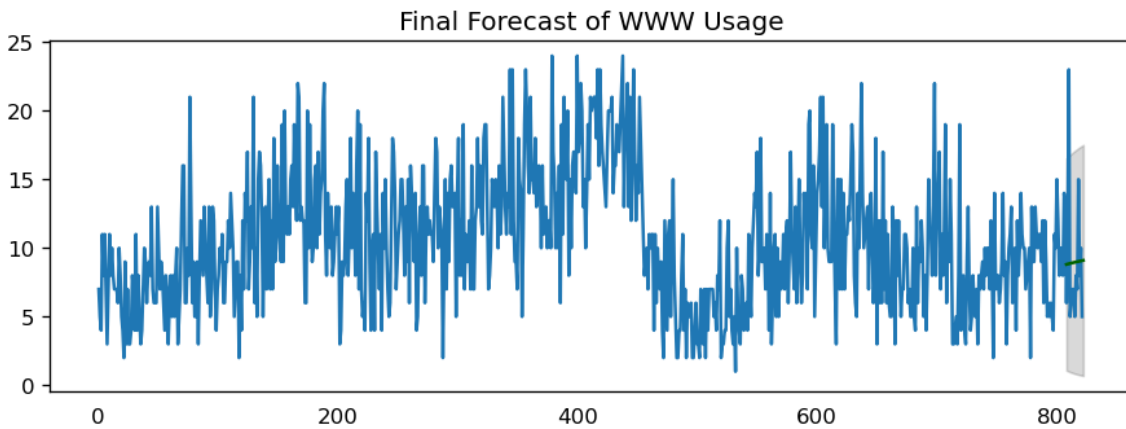
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

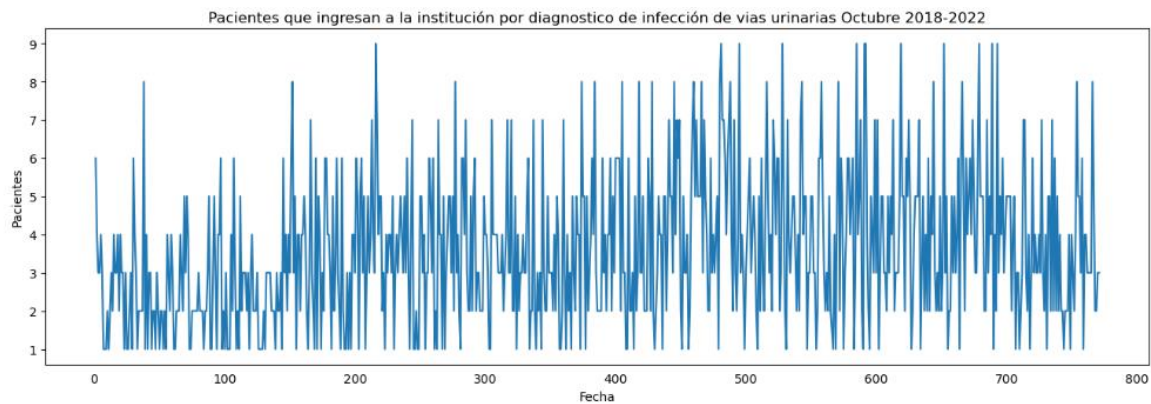


Fuente: Construcción propia

```
808    8.807341
809    8.830582
810    8.853419
811    8.875859
812    8.897908
813    8.919573
814    8.940861
815    8.961779
816    8.982333
817    9.002529
818    9.022374
819    9.041873
820    9.061033
821    9.079860
822    9.098359
dtype: float64
```

Anexo 6. Total de pacientes que ingresan intrahospitalariamente por Infección de vías urinarias

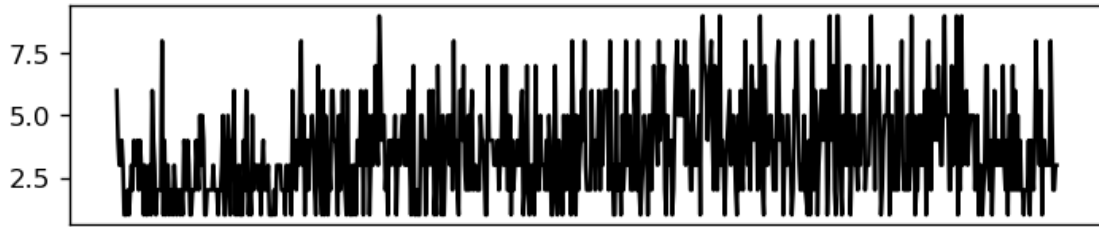
Pacientes ingresados intrahospitalariamente por infección de vías urinarias con periodicidad diaria octubre 2018 – diciembre 2022



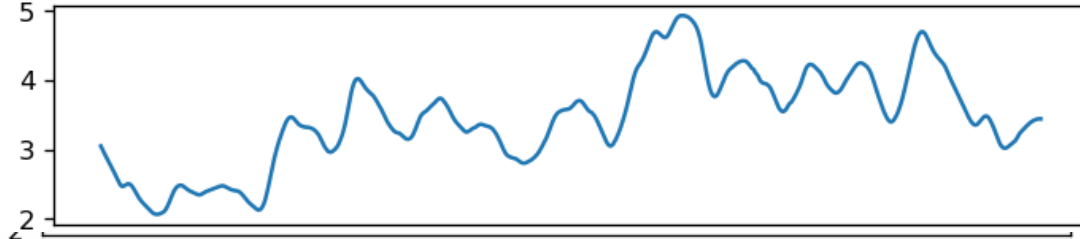
Fuente: Construcción propia

How to Smoothen a Time Series

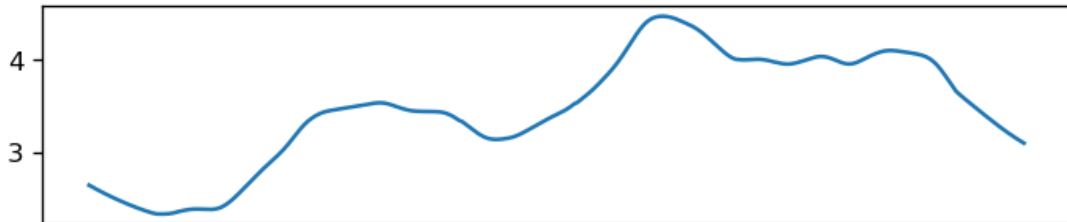
Original Series



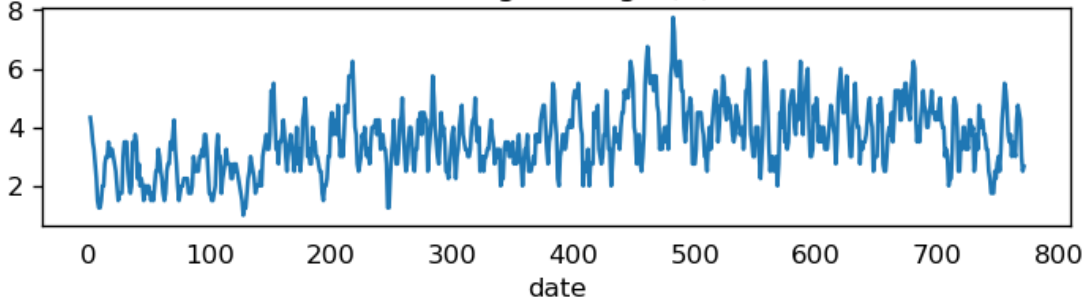
Loess Smoothed 5%



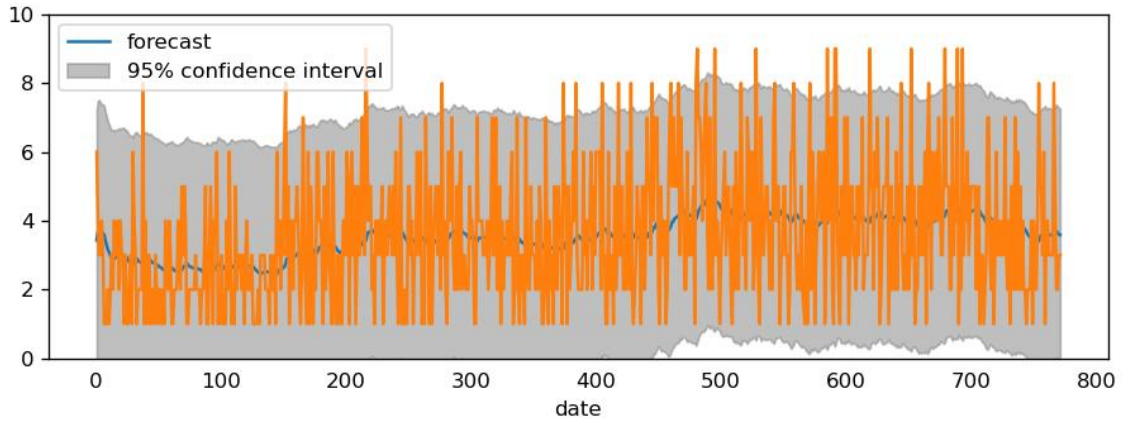
Loess Smoothed 15%



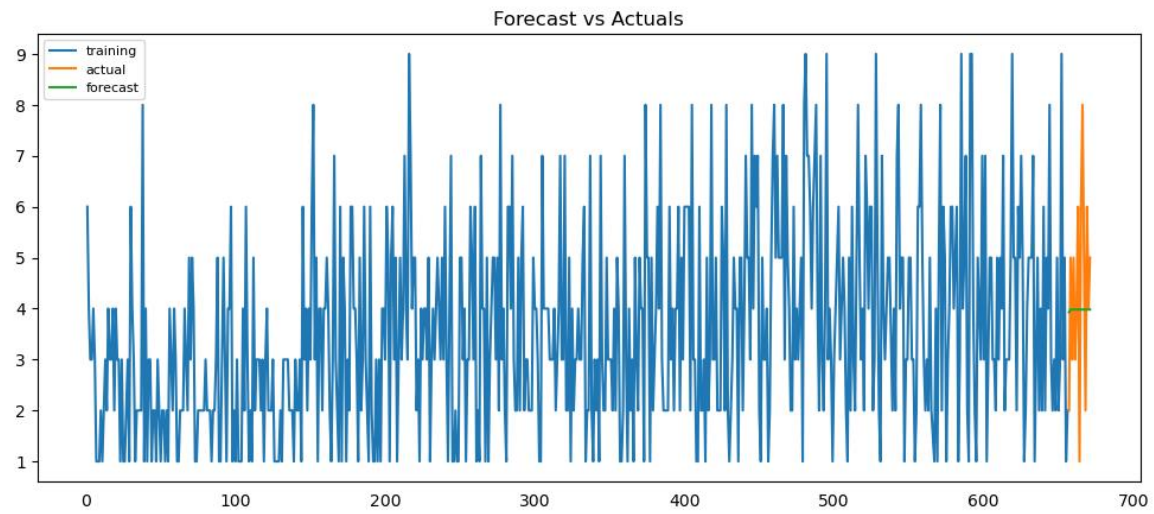
Moving Average (3)



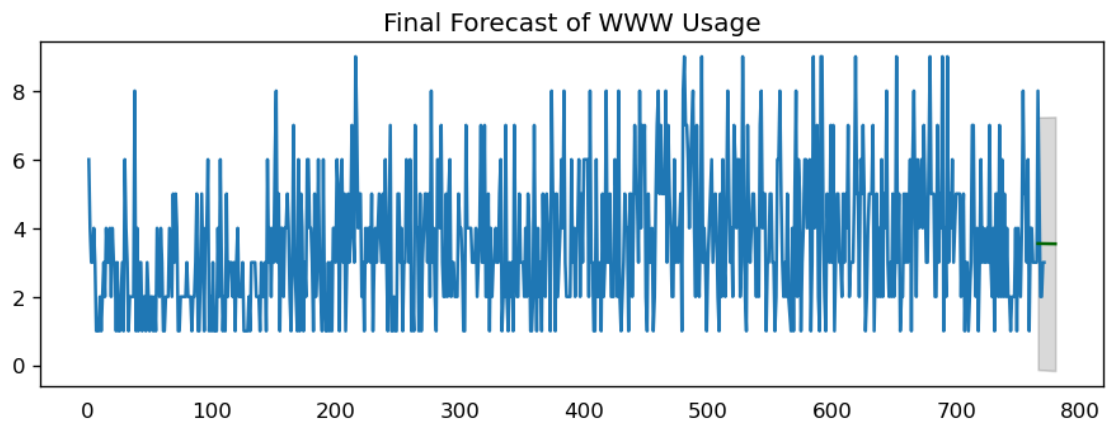
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

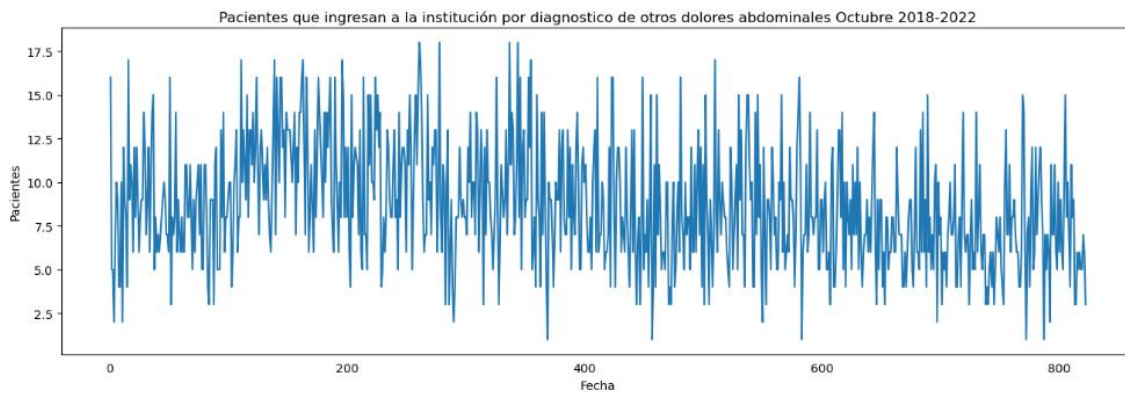


```
766 3.549535
767 3.548850
768 3.548170
769 3.547494
770 3.546823
771 3.546156
772 3.545494
773 3.544836
774 3.544182
775 3.543533
776 3.542888
777 3.542248
778 3.541611
779 3.540979
780 3.540352
dtype: float64
```

Fuente: Construcción propia

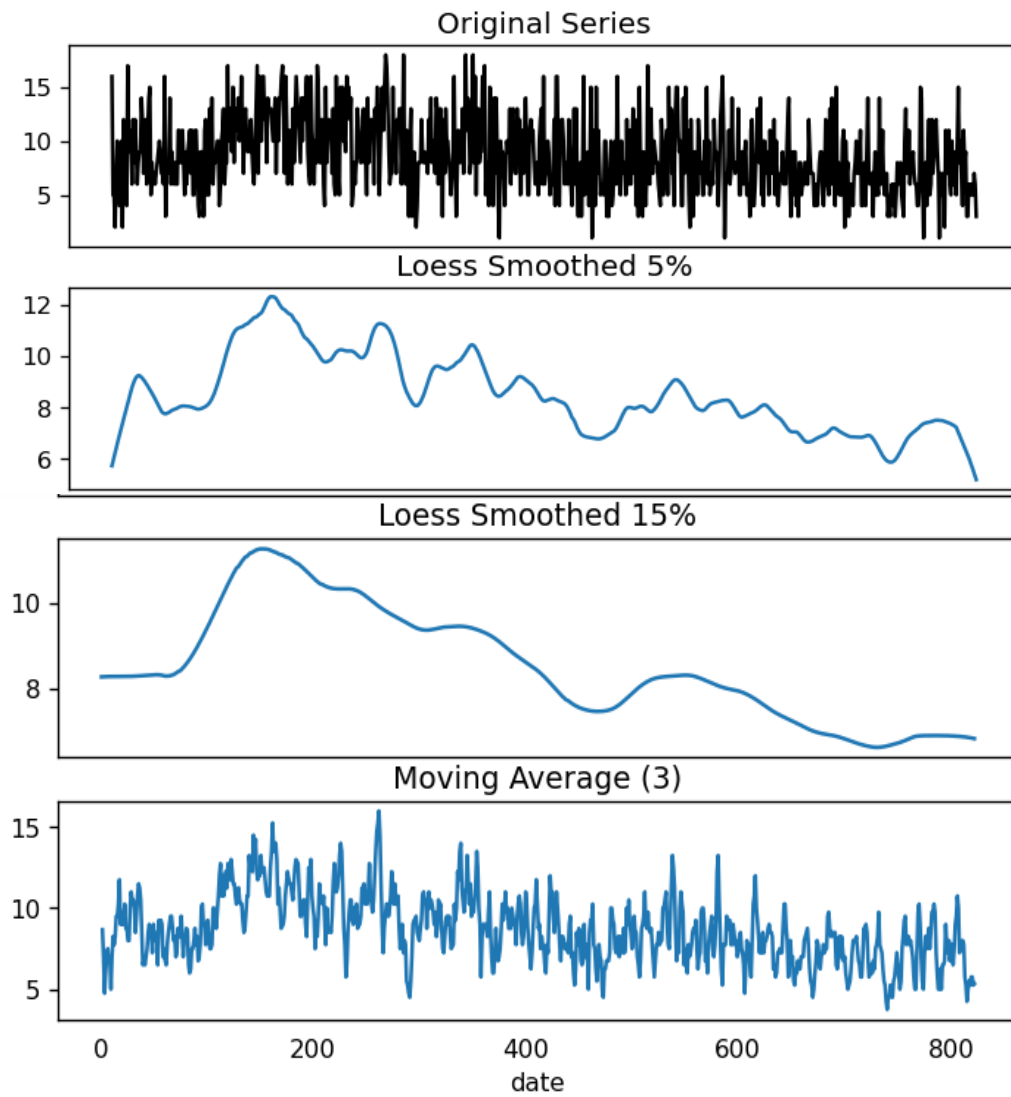
Anexo 7. Total de pacientes que ingresan intrahospitalariamente por Otros dolores abdominales

Pacientes ingresados intrahospitalariamente por Otros dolores abdominales con periodicidad diaria octubre 2018 – diciembre 2022

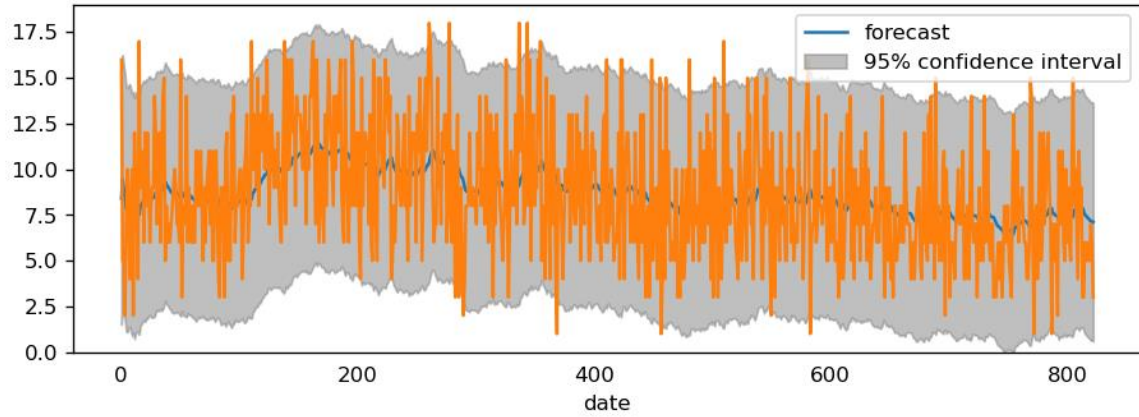


Fuente: Construcción propia

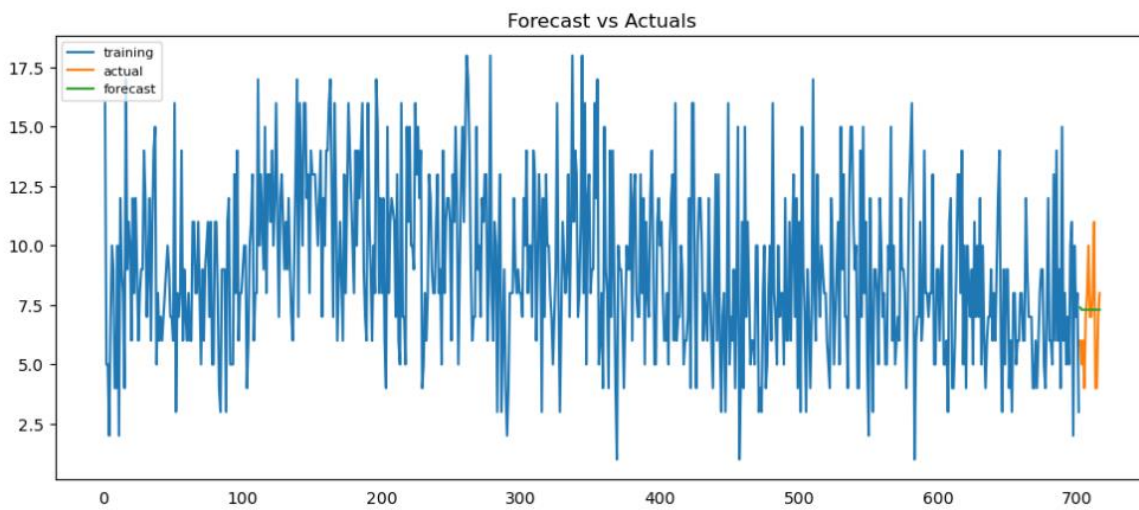
How to Smoothen a Time Series



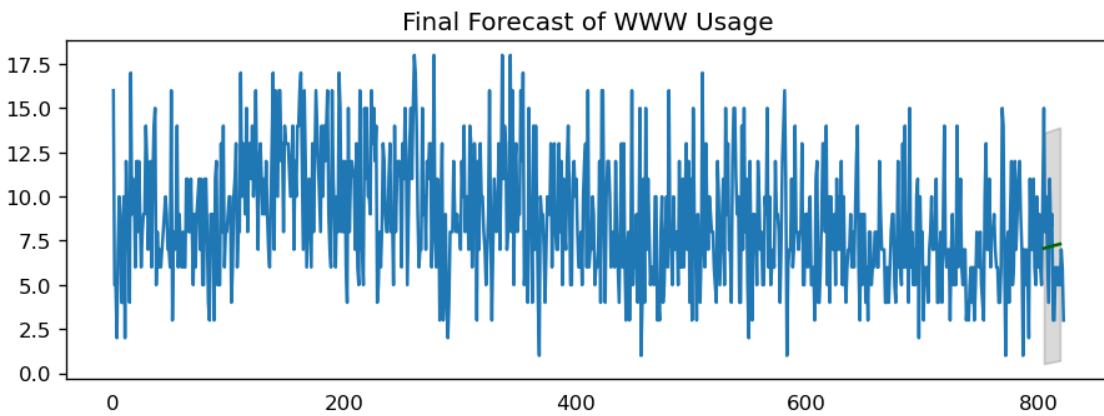
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

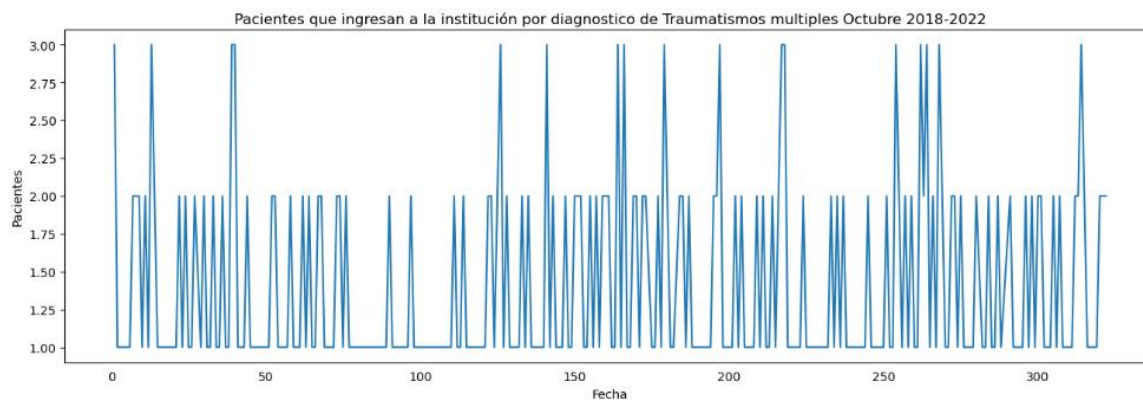


Fuente: Construcción propia

```
805    7.068332
806    7.087313
807    7.106038
808    7.124512
809    7.142737
810    7.160717
811    7.178455
812    7.195955
813    7.213219
814    7.230251
815    7.247053
816    7.263630
817    7.279984
818    7.296117
819    7.312034
dtype: float64
```

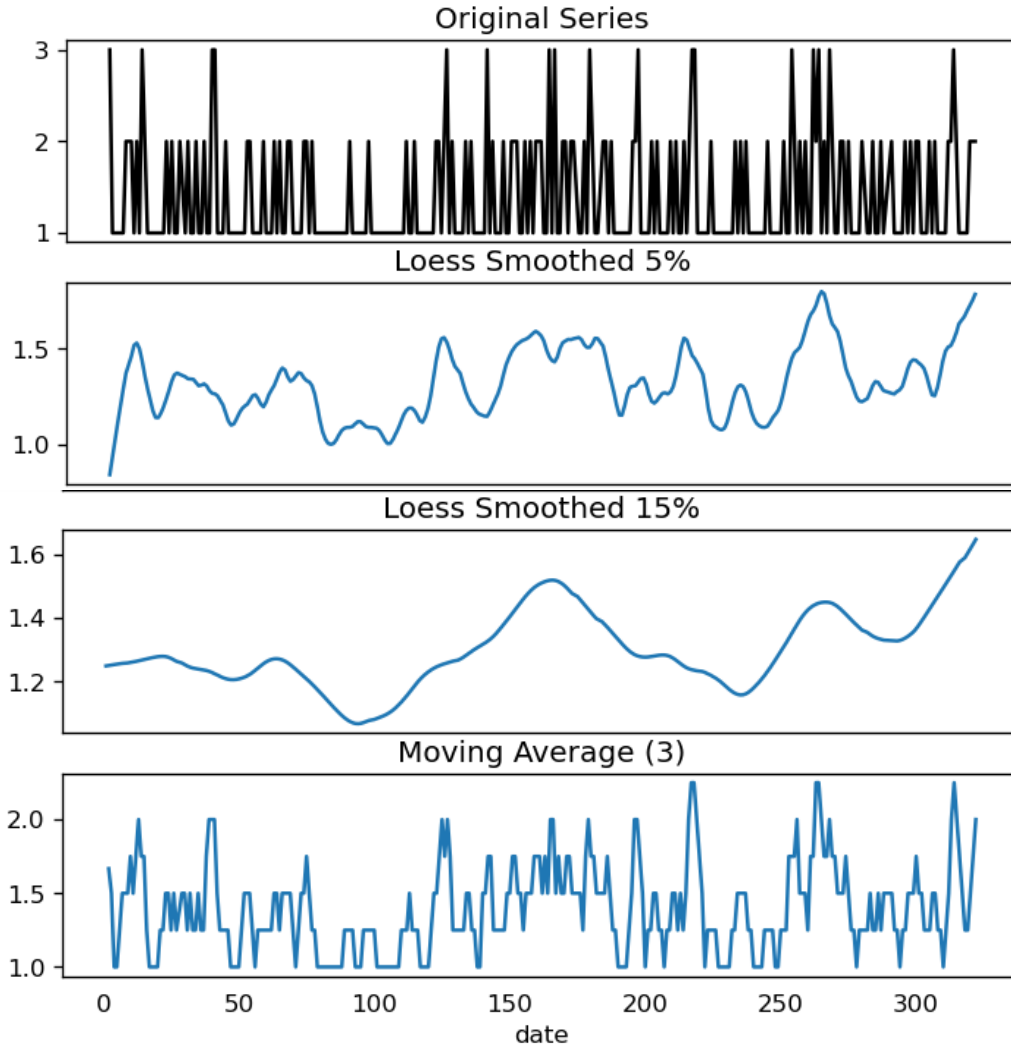
Anexo 8. Total de pacientes que ingresan intrahospitalariamente por traumatismos múltiples

Pacientes ingresados intrahospitalariamente por Traumatismos múltiples con periodicidad diaria octubre 2018 – diciembre 2022

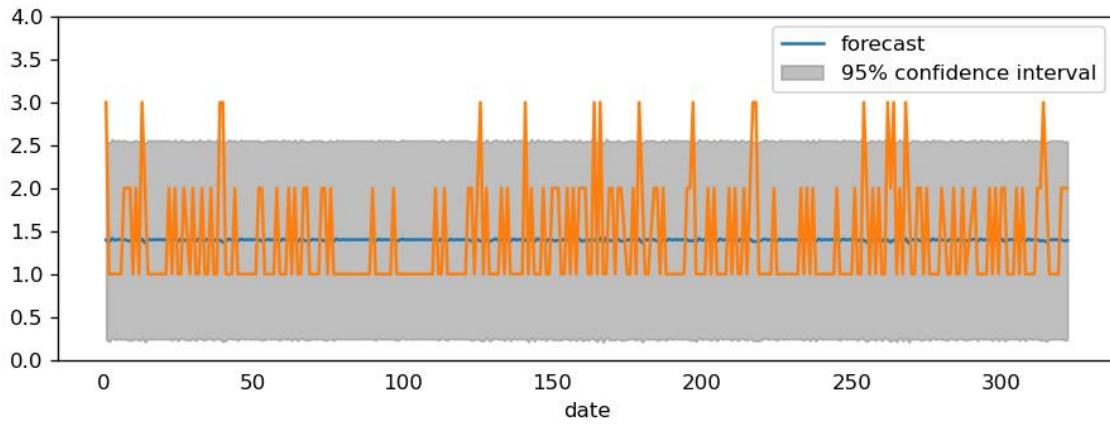


Fuente: Construcción propia

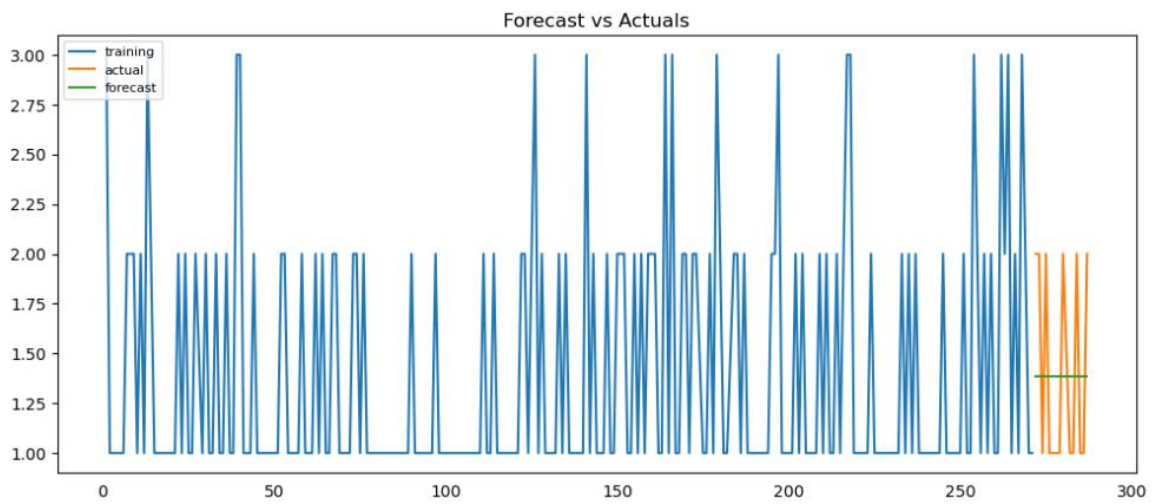
How to Smoothen a Time Series



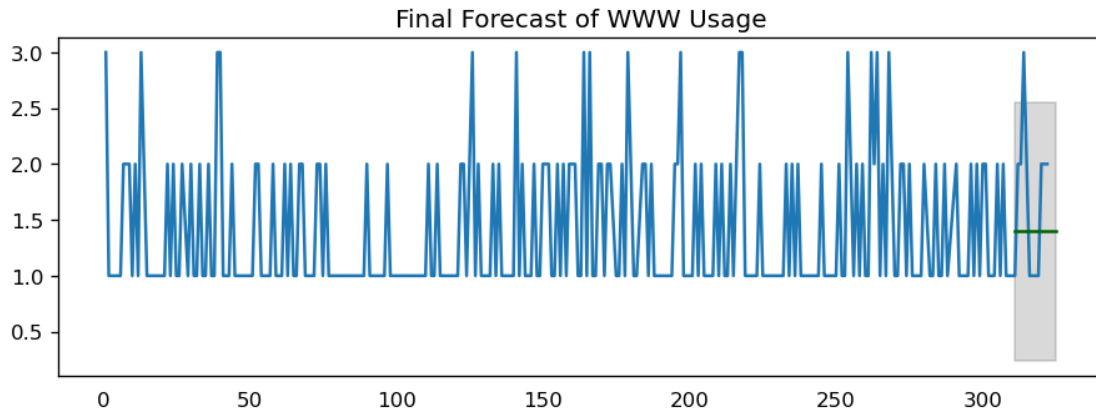
Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia



Fuente: Construcción propia

```

311    1.395498
312    1.395498
313    1.395498
314    1.395498
315    1.395498
316    1.395498
317    1.395498
318    1.395498
319    1.395498
320    1.395498
321    1.395498
322    1.395498
323    1.395498
324    1.395498
325    1.395498
dtype: float64

```