

TRÁFICO EN LA WEB

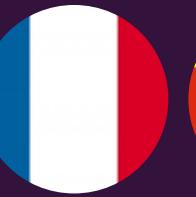
2019-2 | SIMULACIÓN DIGITAL
ESCUELA DE INGENIERÍA DE SISTEMAS
UNIVERSIDAD INDUSTRIAL DE SANTANDER
MARZO 2020

Contenido

Objetivo
Dataset
Métodos a utilizar
Implementación | Resultados
Conclusiones
Información

Objetivo

Predecir el comportamiento de las páginas web más visitadas en Wikipedia durante los últimos meses para los ocho(8) idiomas de los cuales tenemos información.



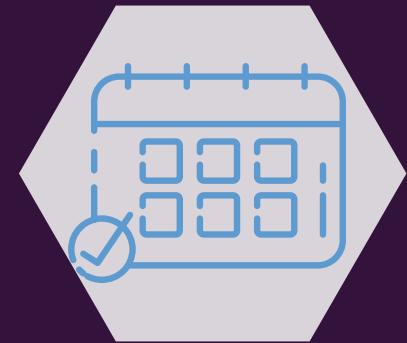
04

Dataset

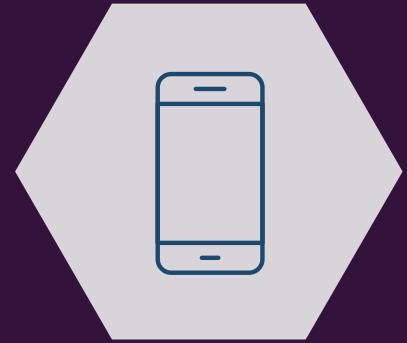


WIKIPEDIA

145.000 DATOS



1 JUL 2015
31 DIC 2016

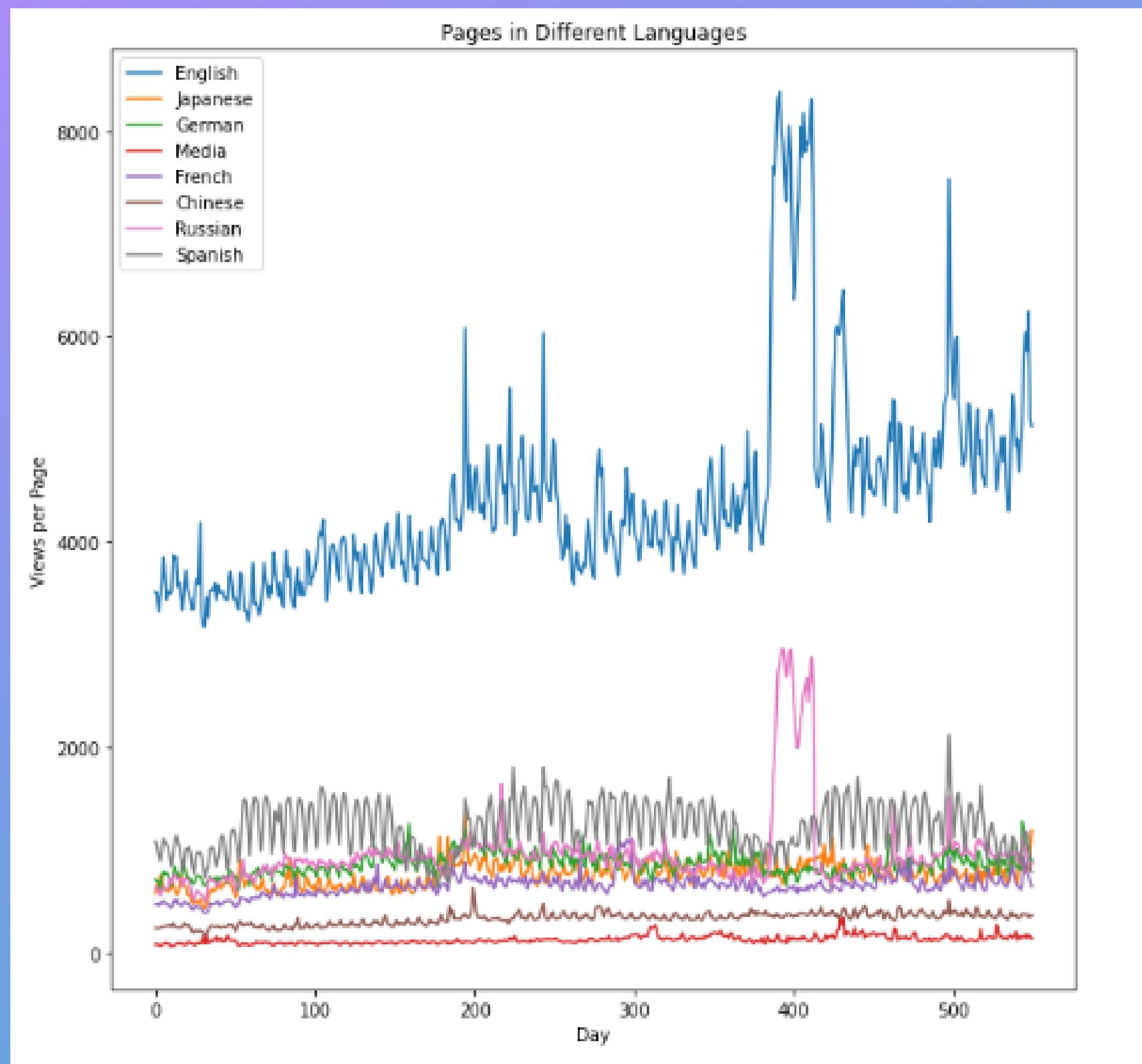


MOVIL
ESCRITORIO
ARAÑA

05

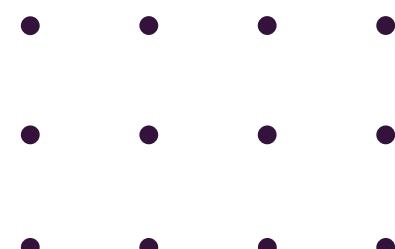
Dataset

06



Métodos a utilizar

| TEST ESTACIONARIDAD | ARIMA |
| SARIMA | REDES NEURONALES |



07

Implementación

| RESULTADOS |

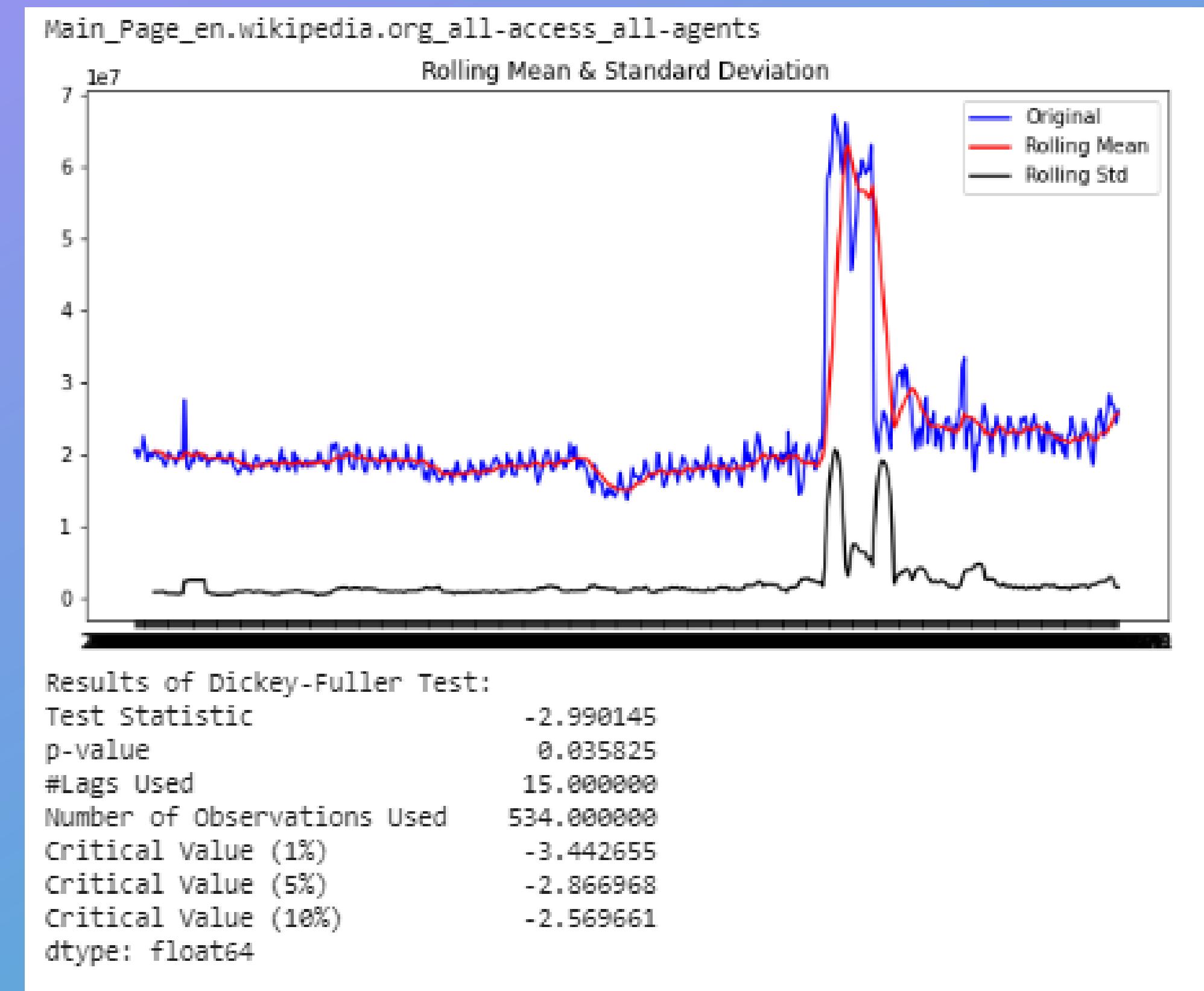
01

**TEST ESTACIONARIO
SIN SHIFTING**

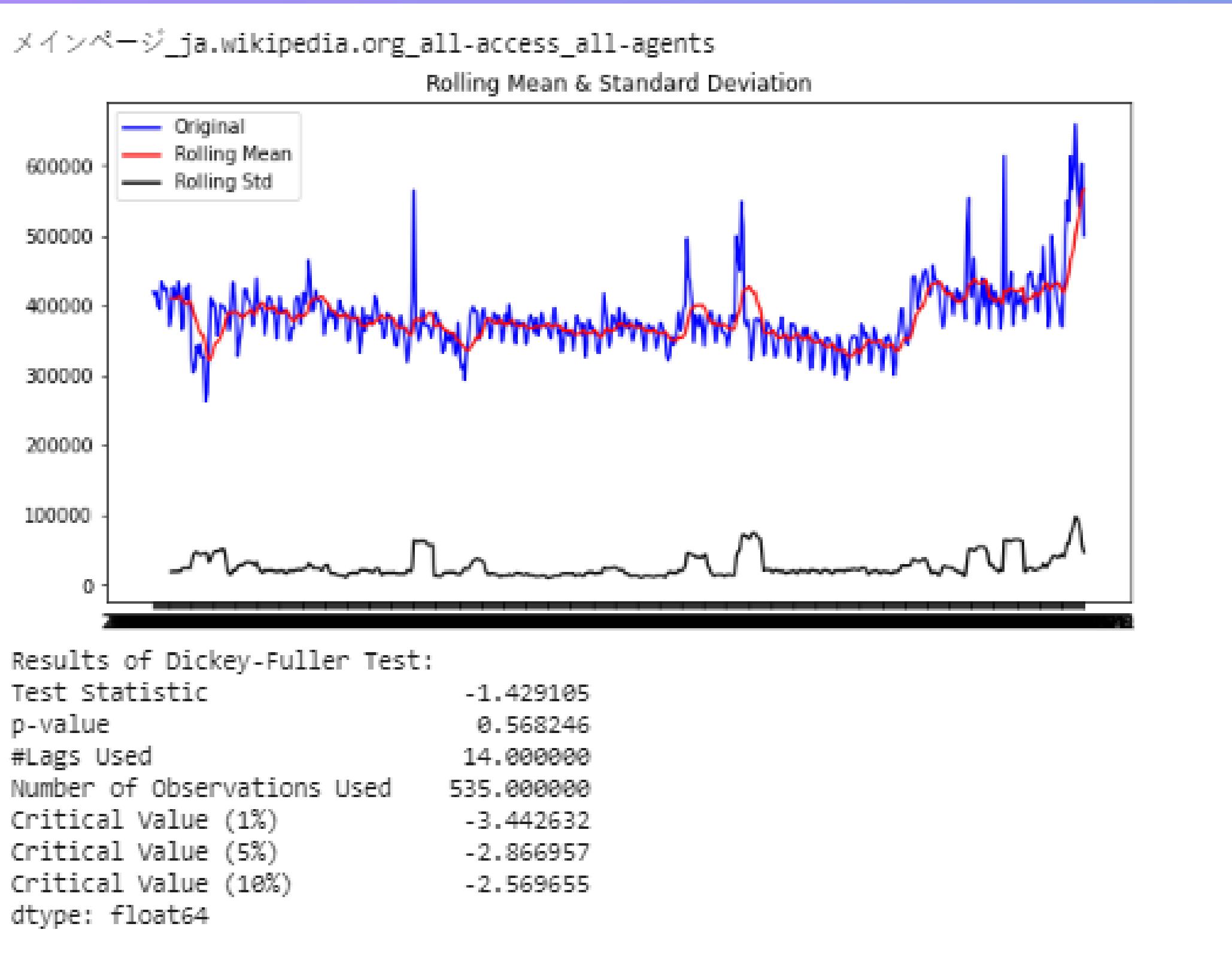
Test Estacionario

INGLÉS |
ESTACIONARIA

09



Test Estacionario



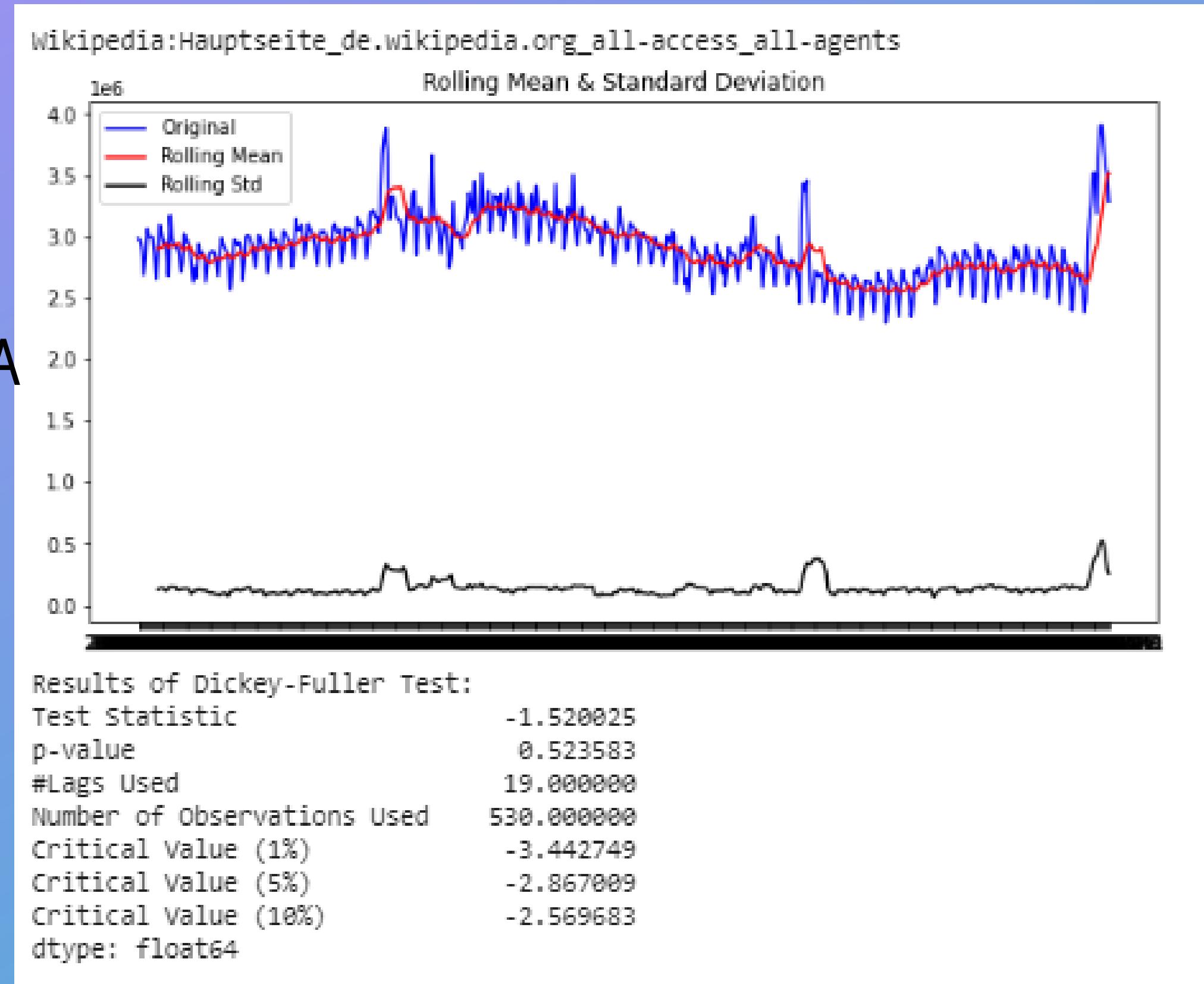
JAPONÉS |
NO ESTACIONARIA

10

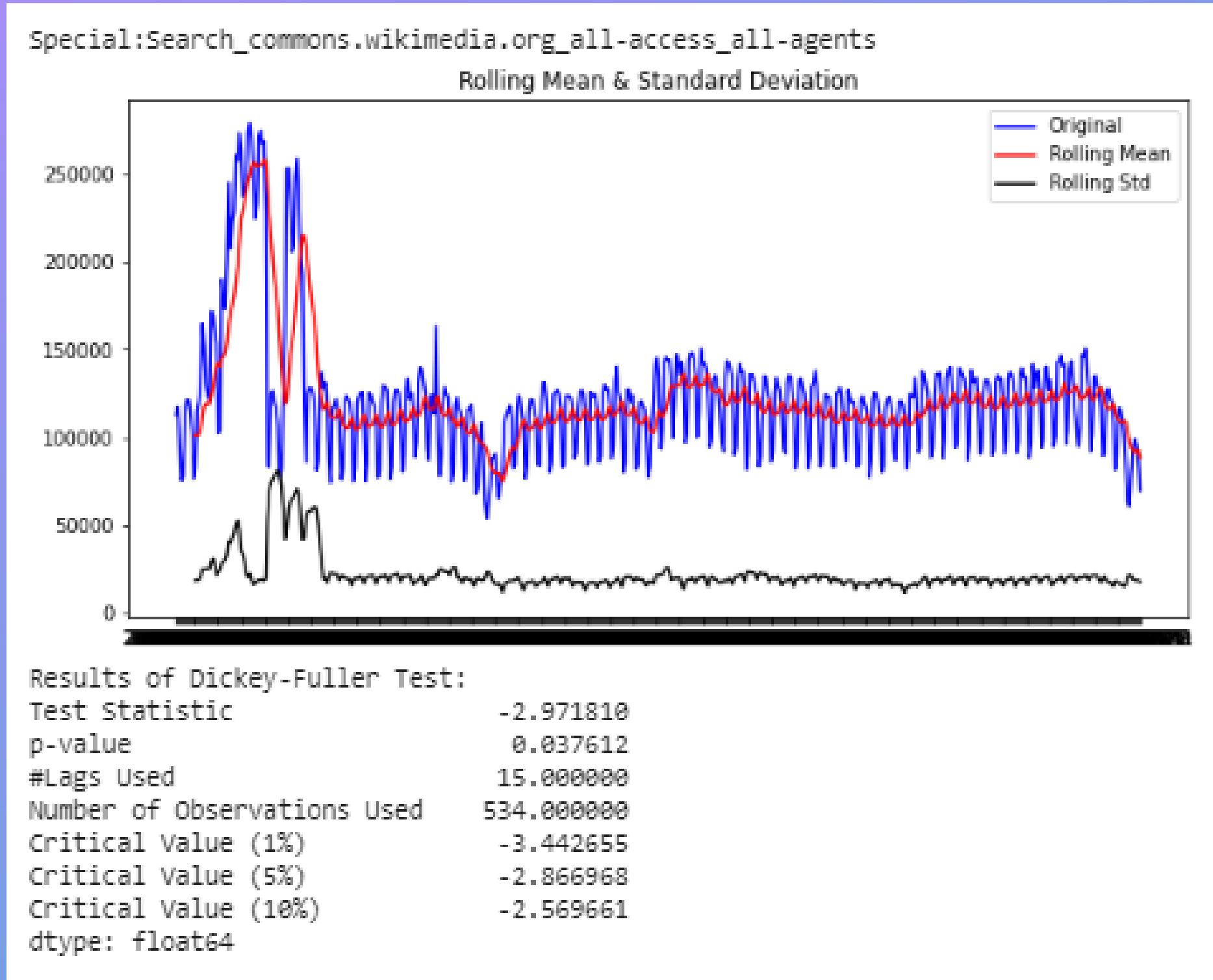
Test Estacionario

ALEMÁN |
NO ESTACIONARIA

11



Test Estacionario

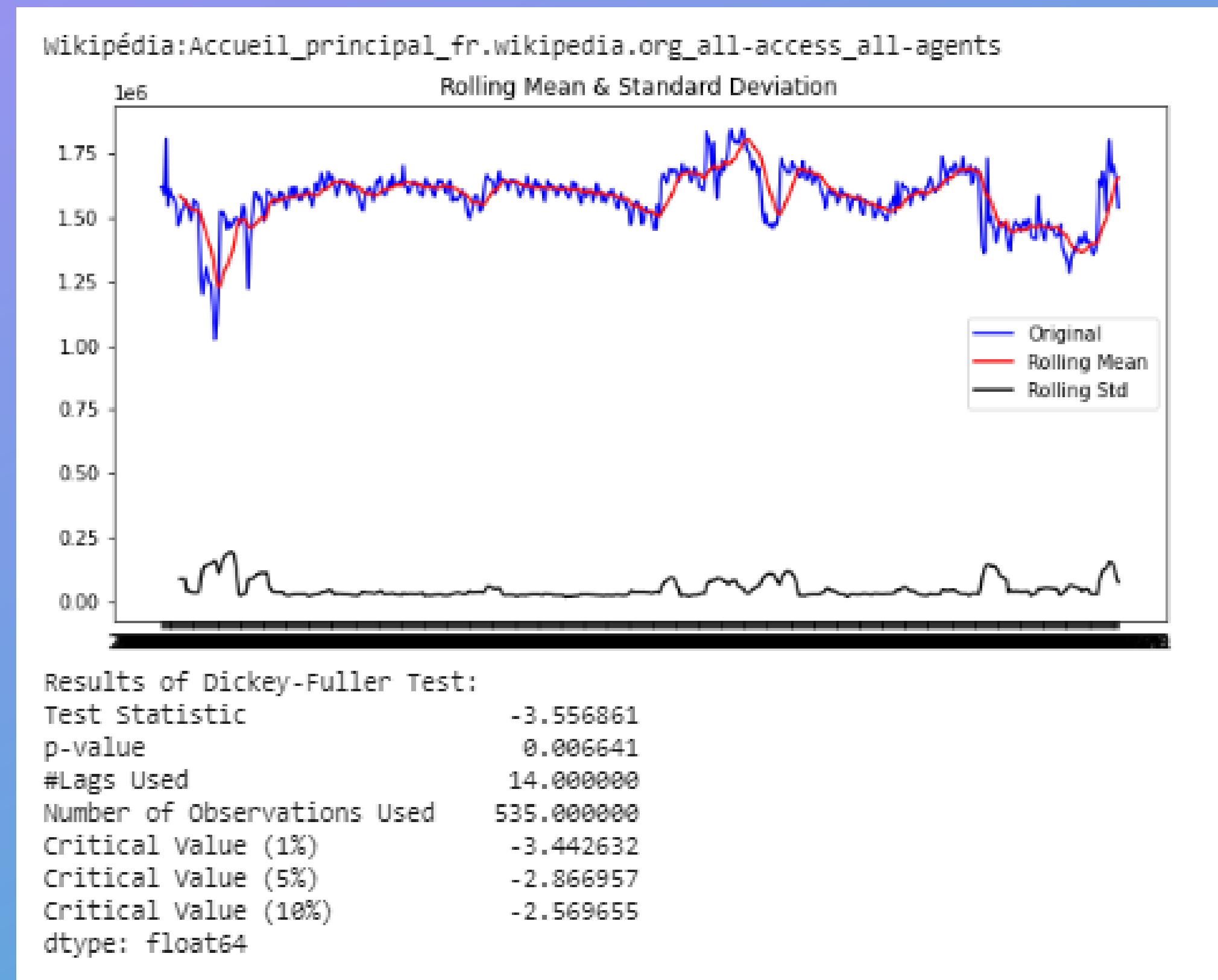


COMÚN |
ESTACIONARIA

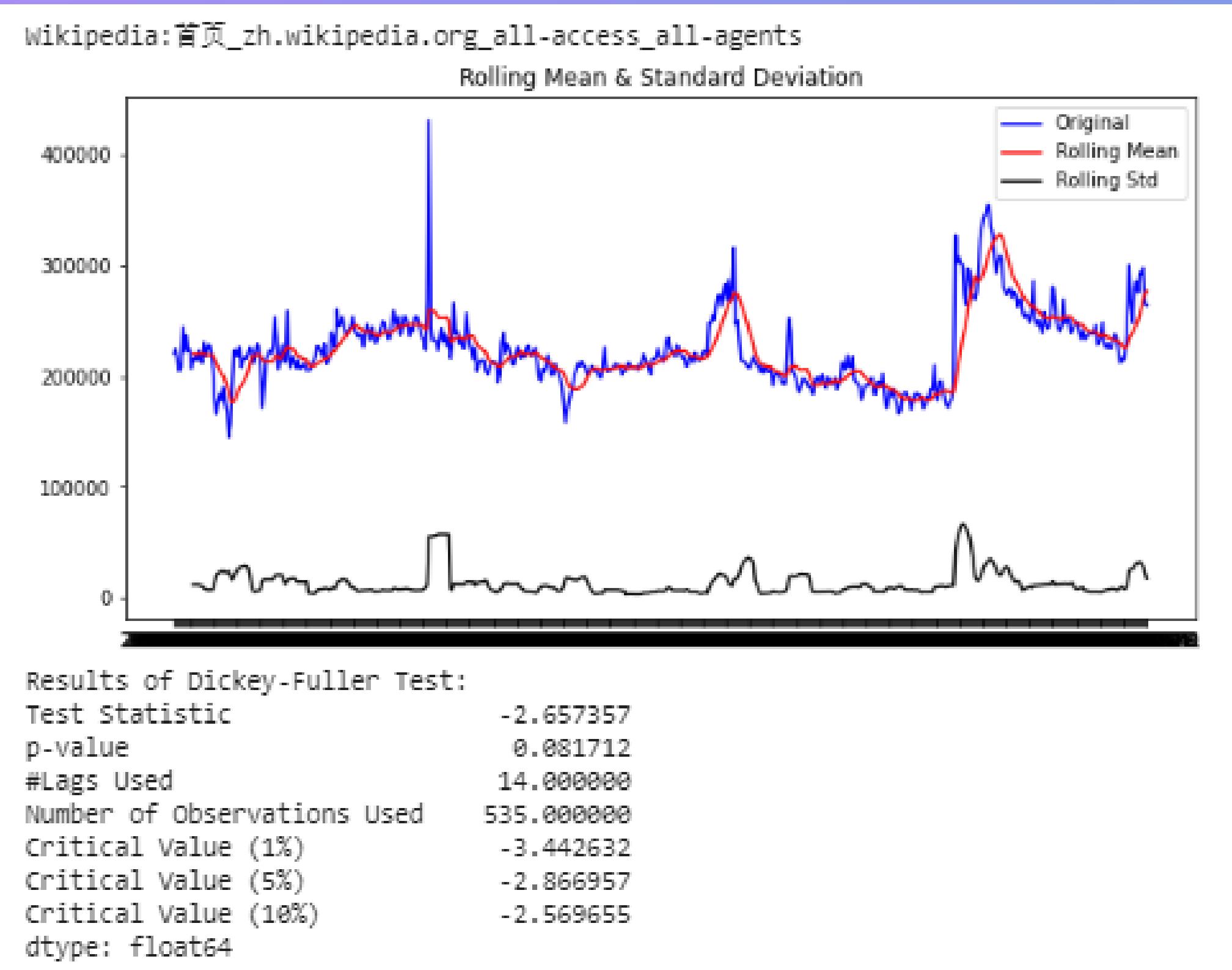
Test Estacionario

FRANCÉS |
ESTACIONARIA

13



Test Estacionario

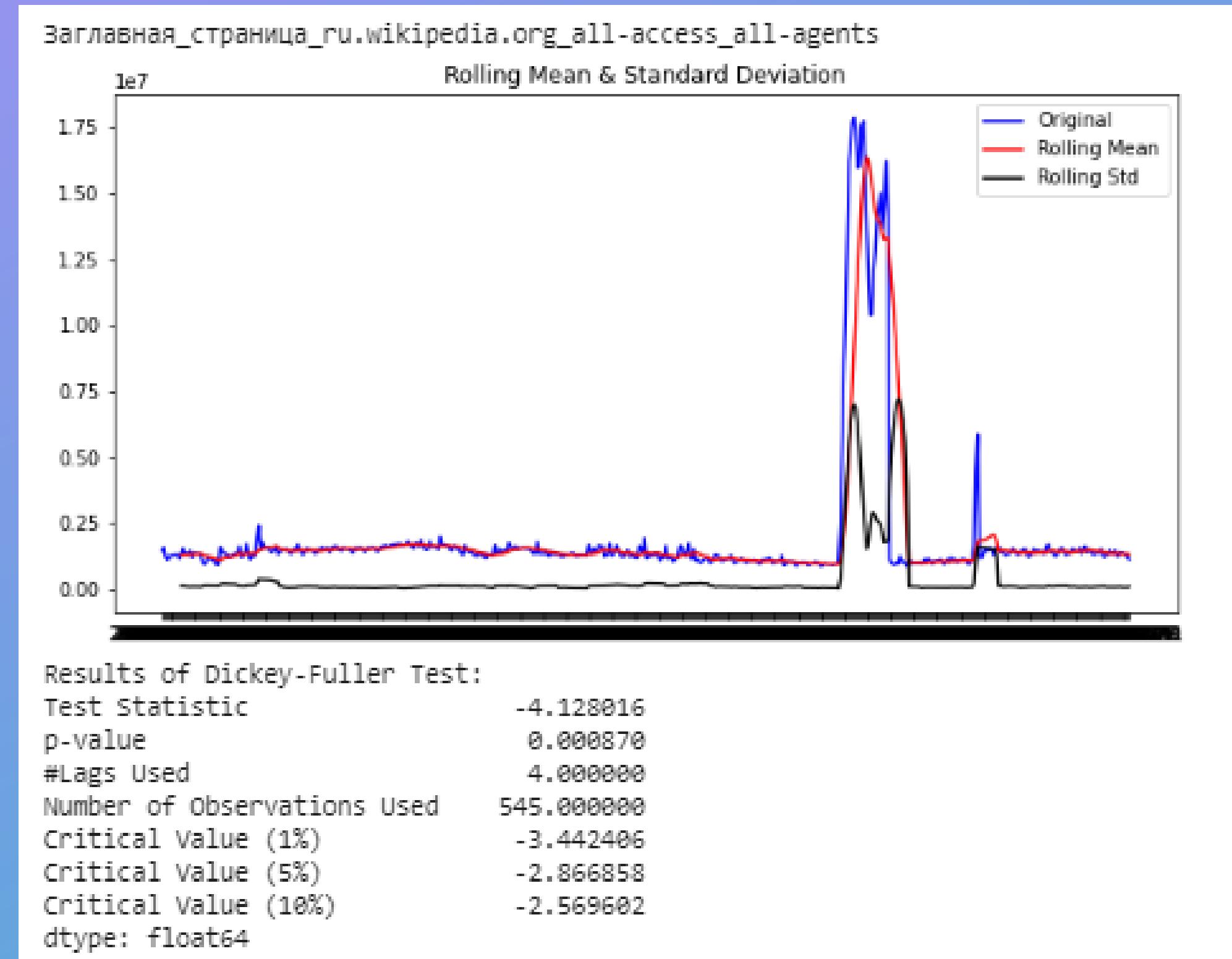


CHINO |
NO ESTACIONARIA

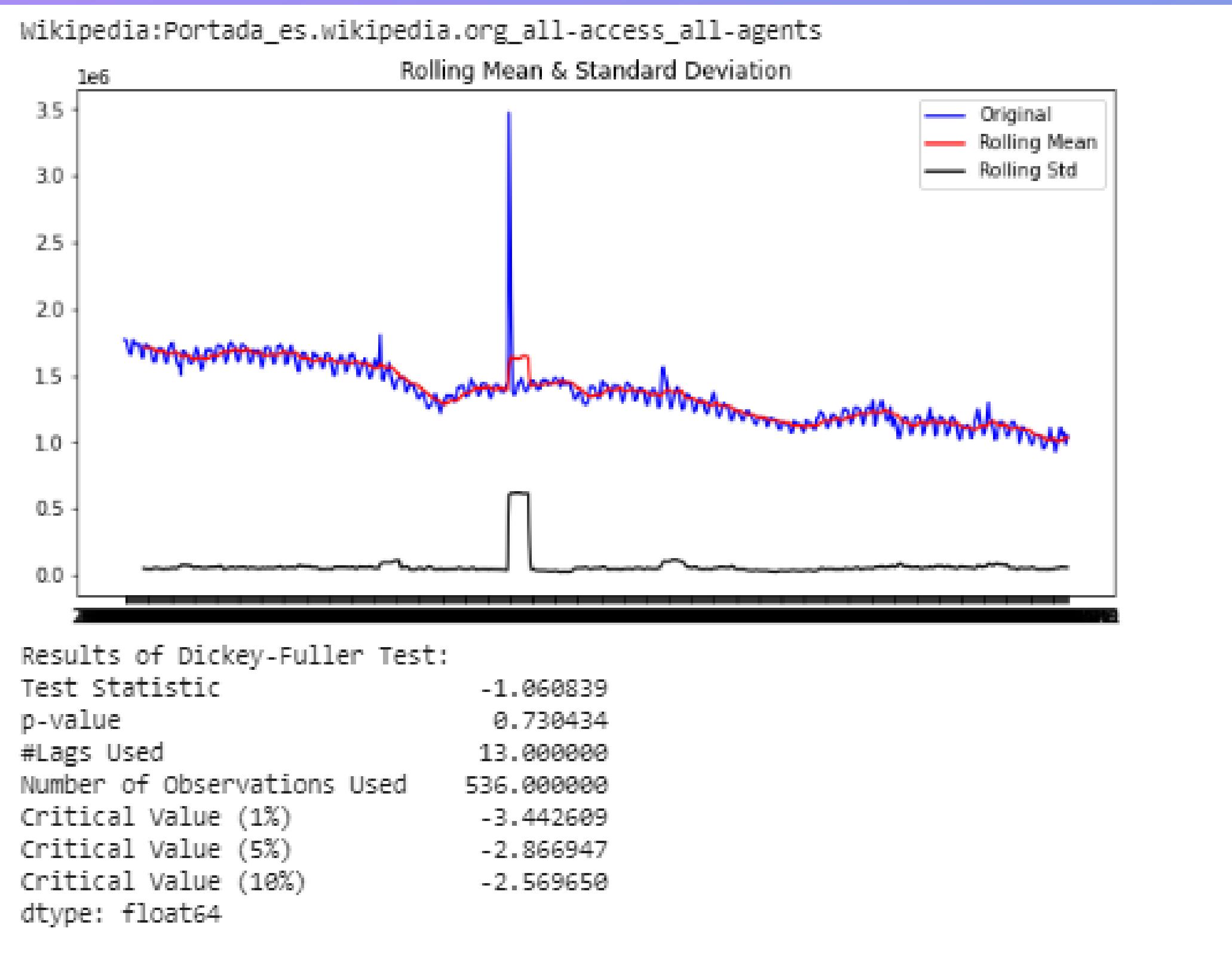
Test Estacionario

RUSO |
ESTACIONARIA

15



Test Estacionario



ESPAÑOL |
NO ESTACIONARIA

17

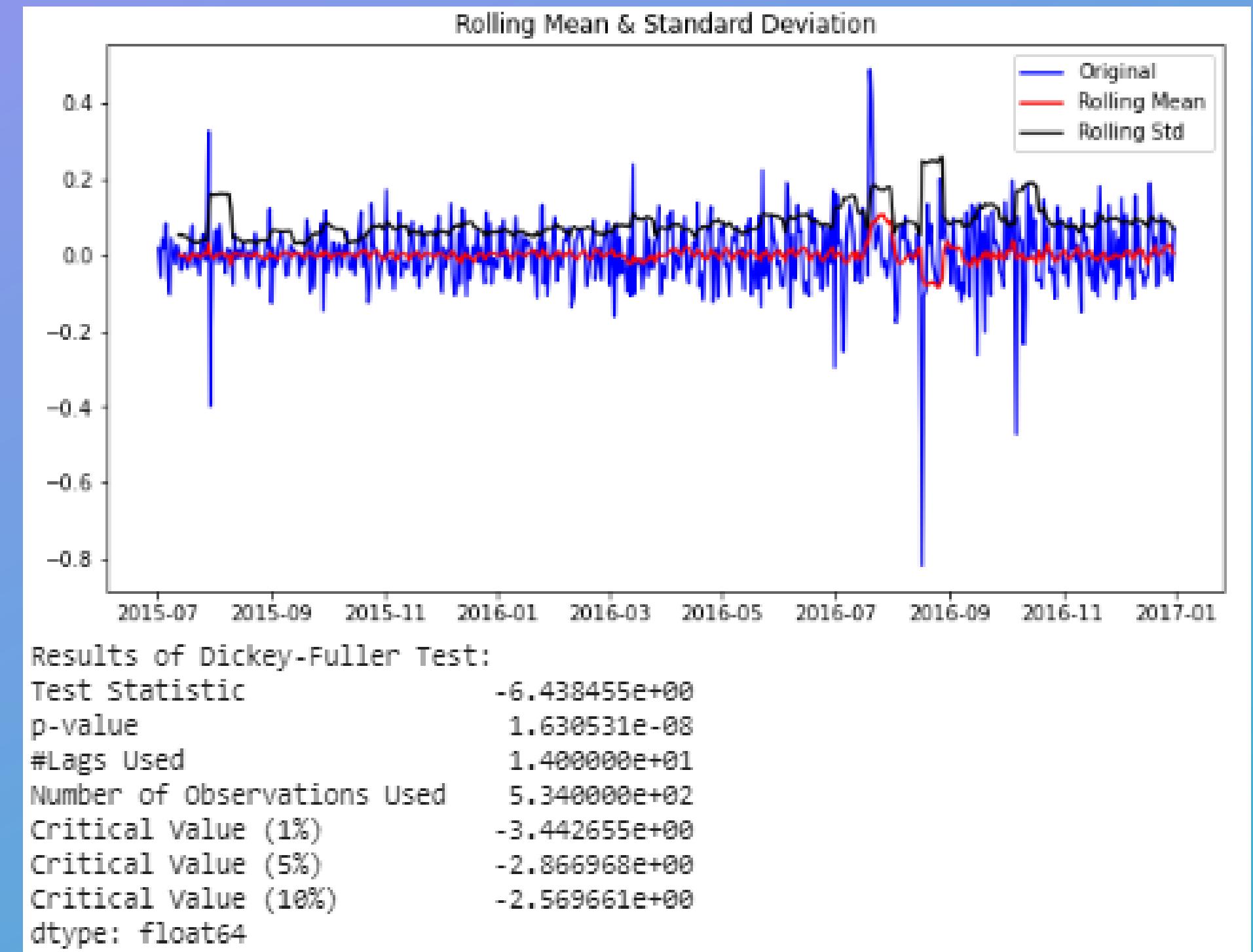
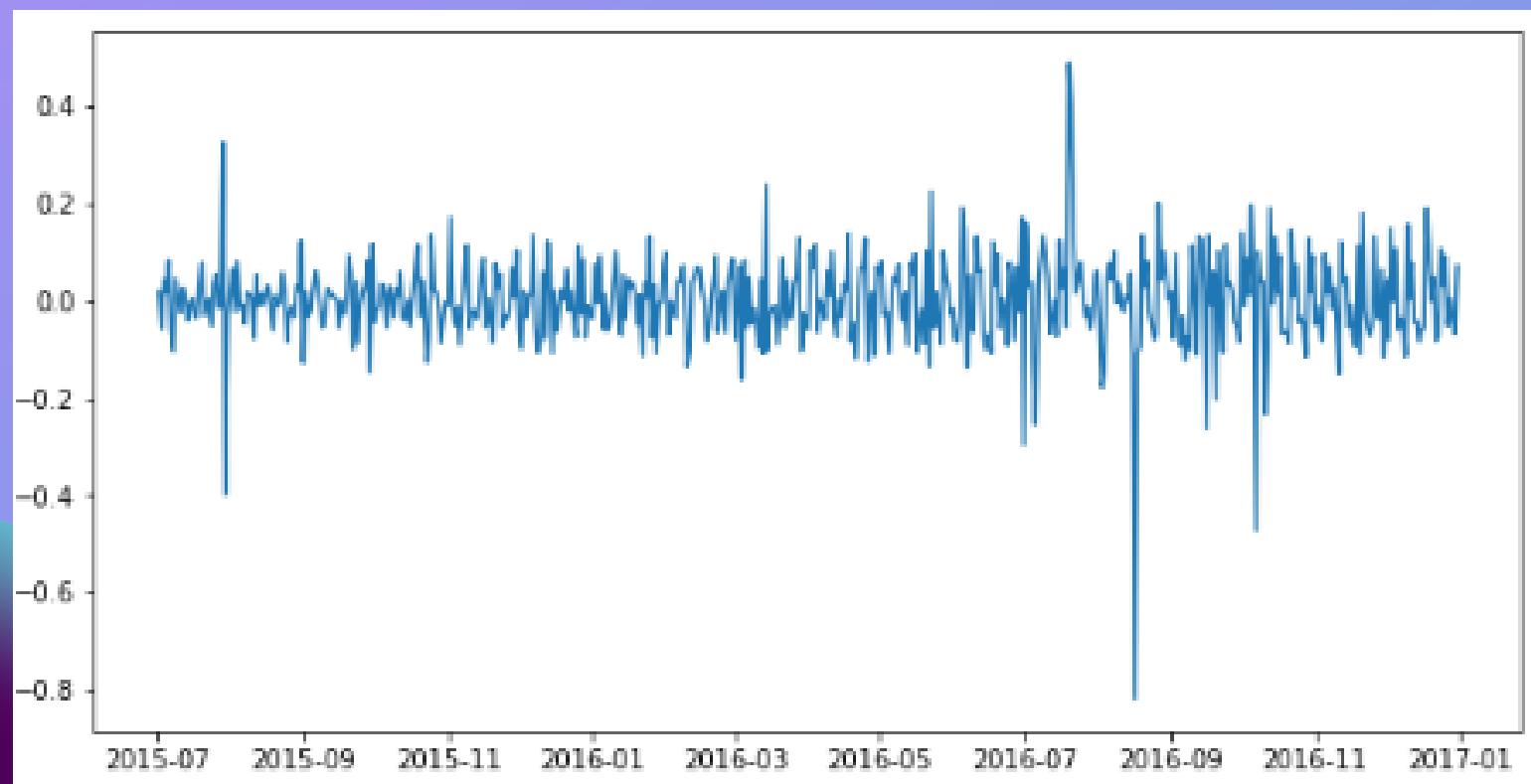
Implementación

| RESULTADOS |

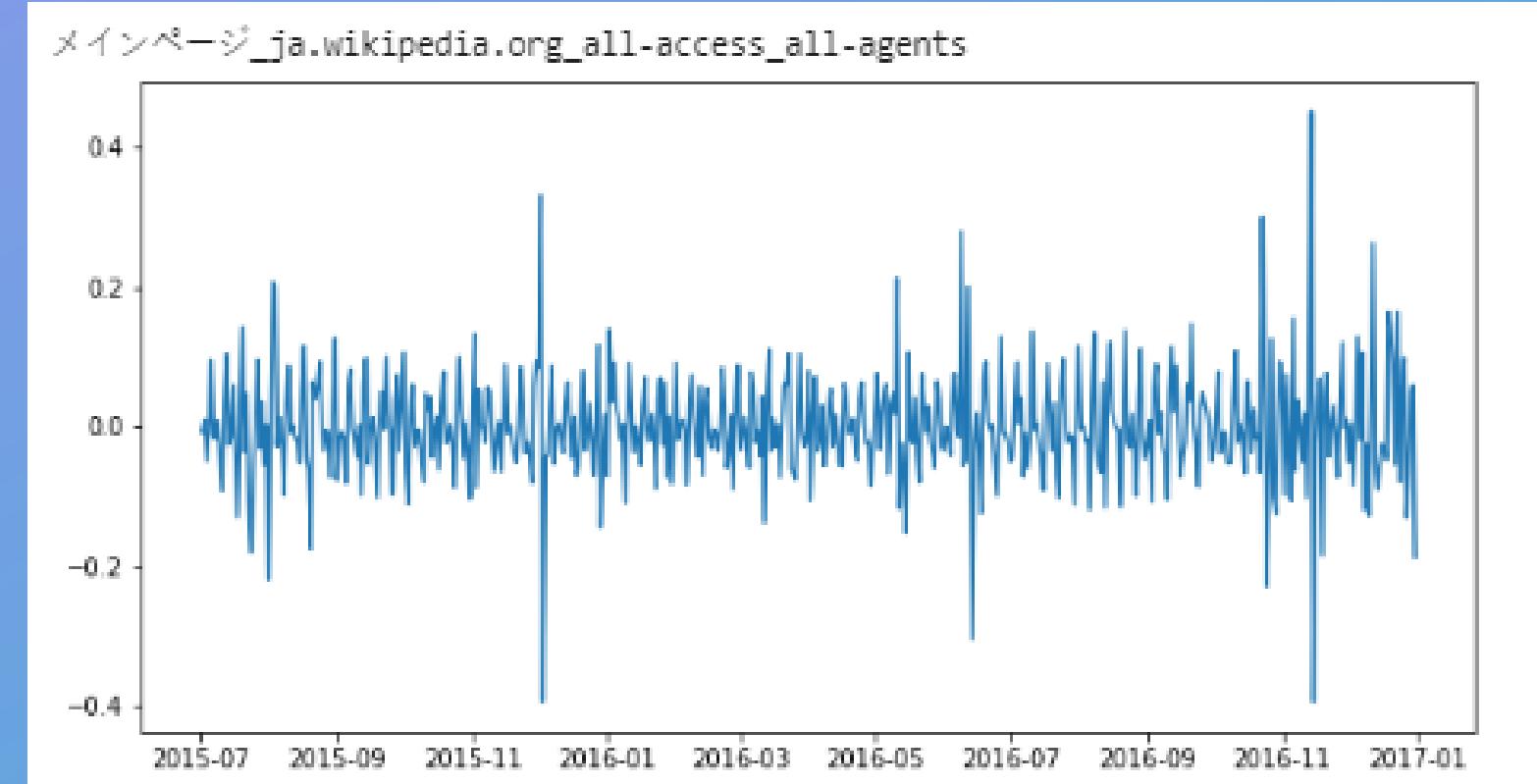
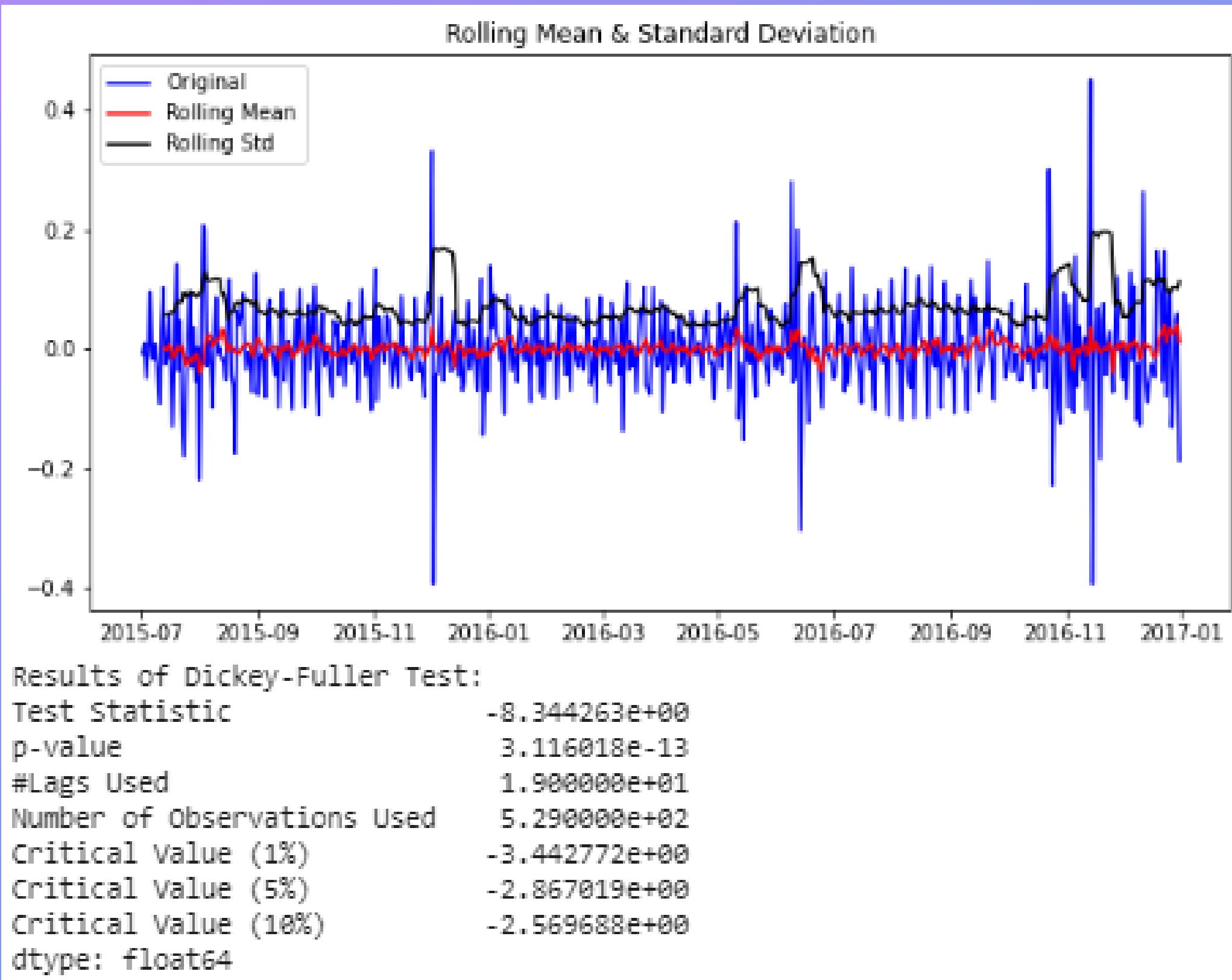
01

**TEST ESTACIONARIO
CON SHIFTING**

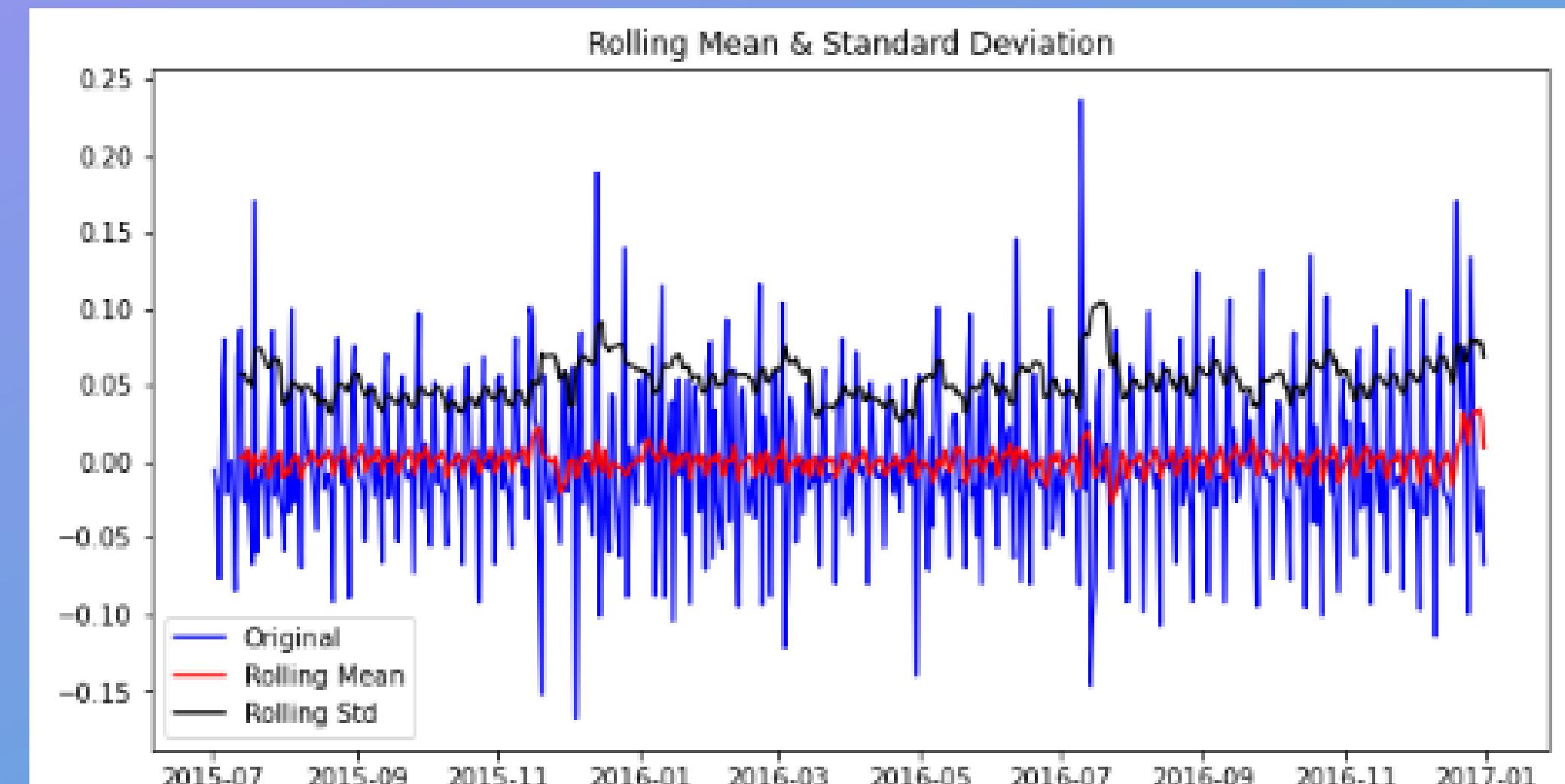
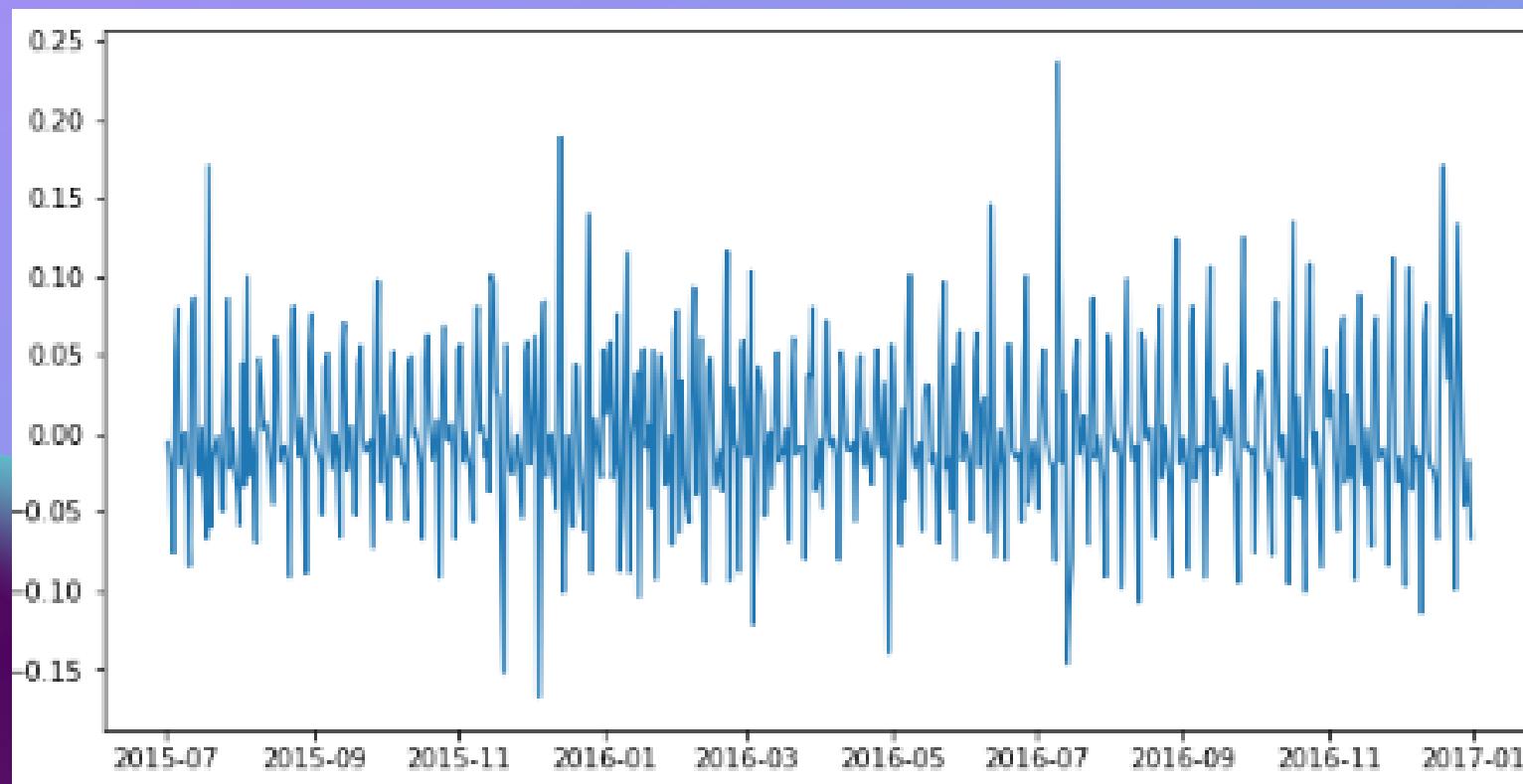
Test Estacionario



Test Estacionario



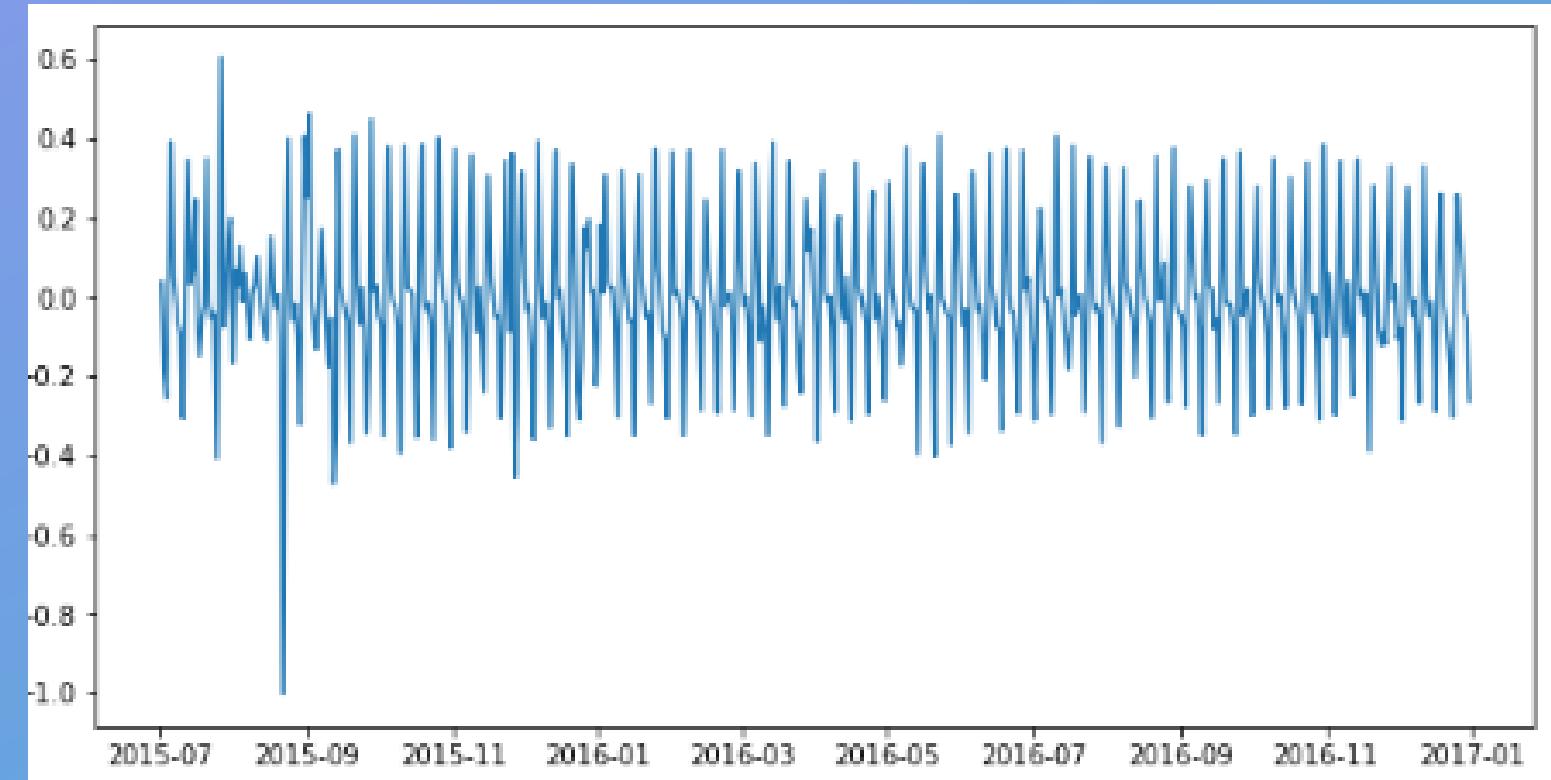
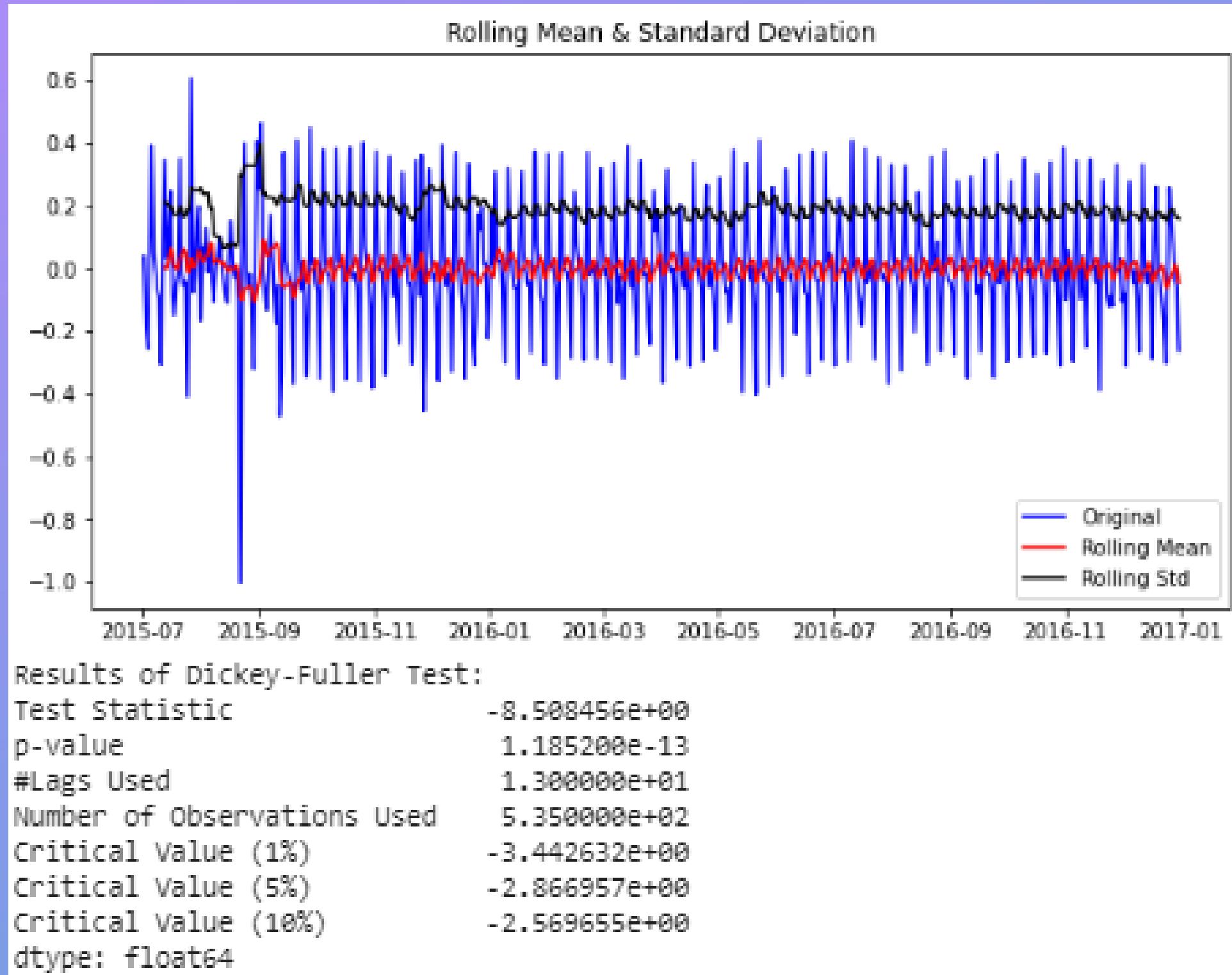
Test Estacionario



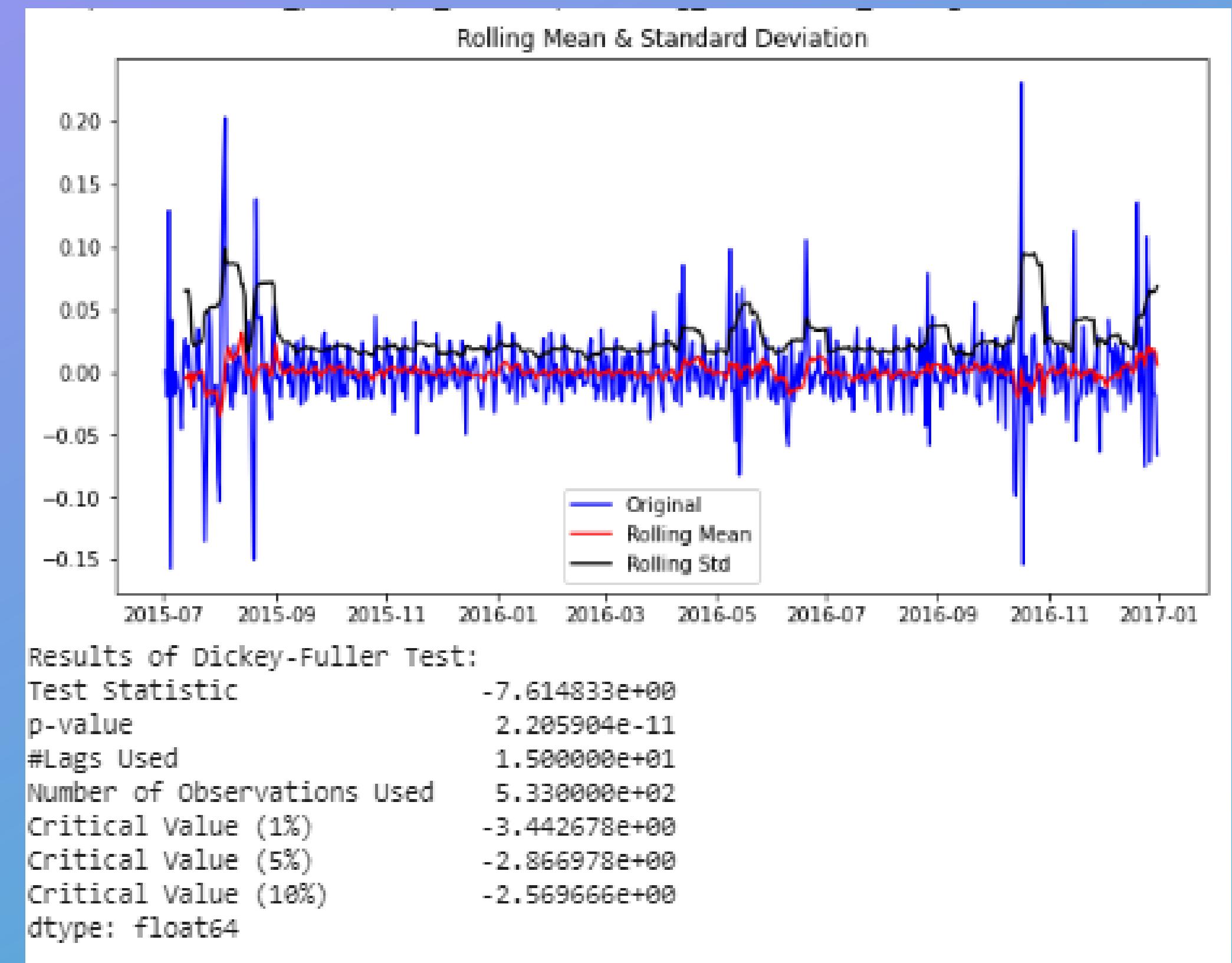
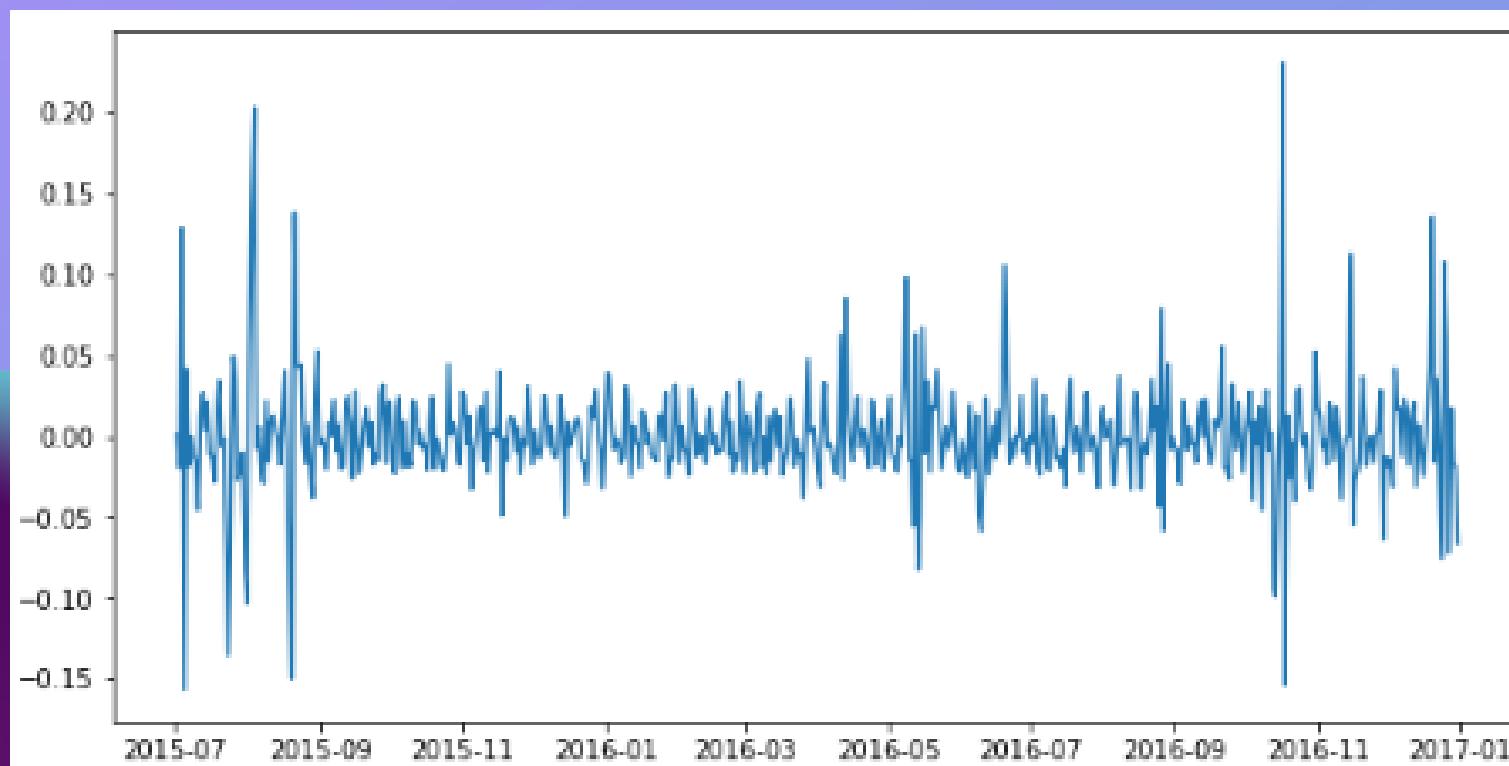
Results of Dickey-Fuller Test:

Test Statistic	-7.045289e+00
p-value	5.707684e-10
#Lags Used	1.900000e+01
Number of Observations Used	5.290000e+02
Critical Value (1%)	-3.442772e+00
Critical Value (5%)	-2.867019e+00
Critical Value (10%)	-2.569688e+00
dtype: float64	

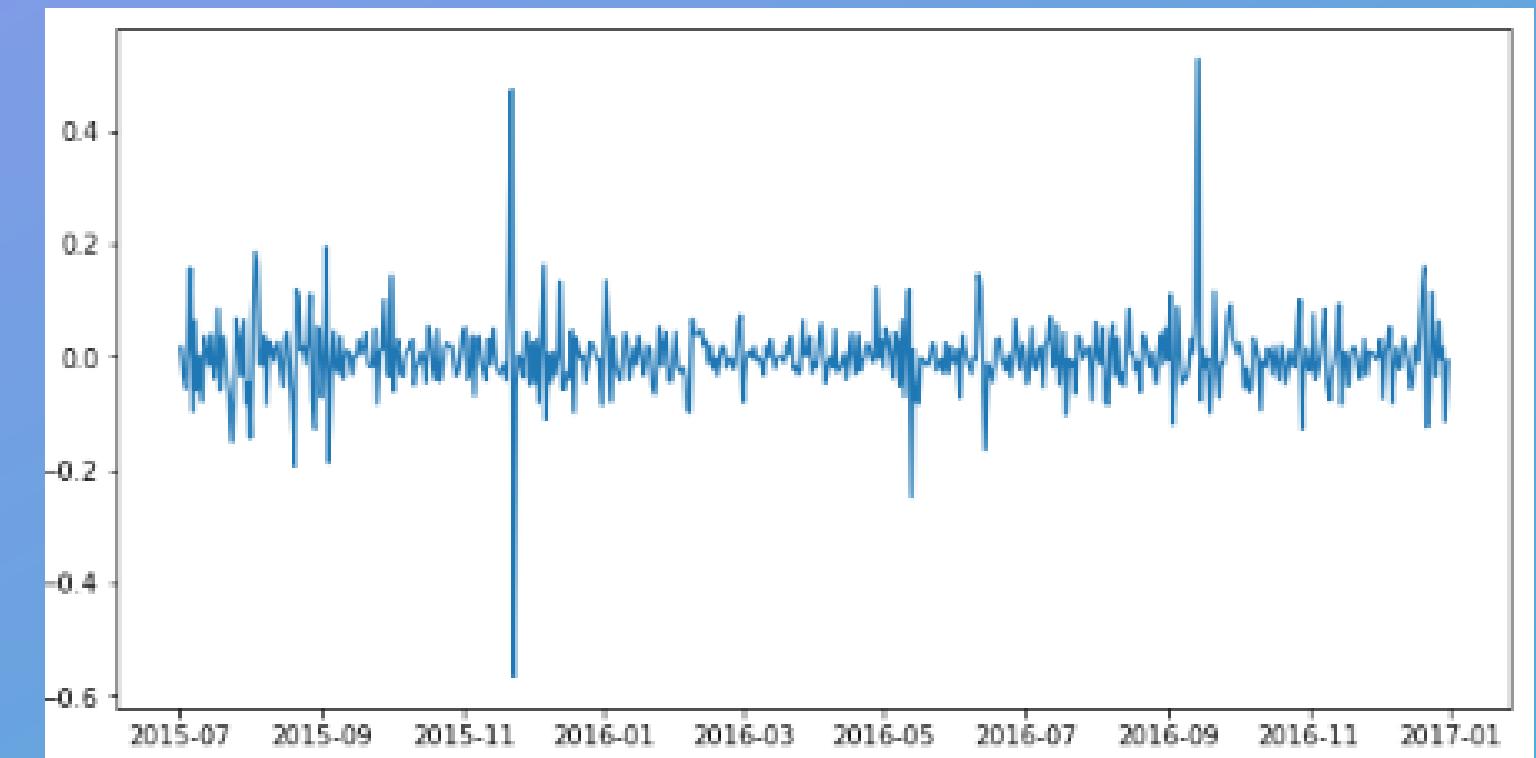
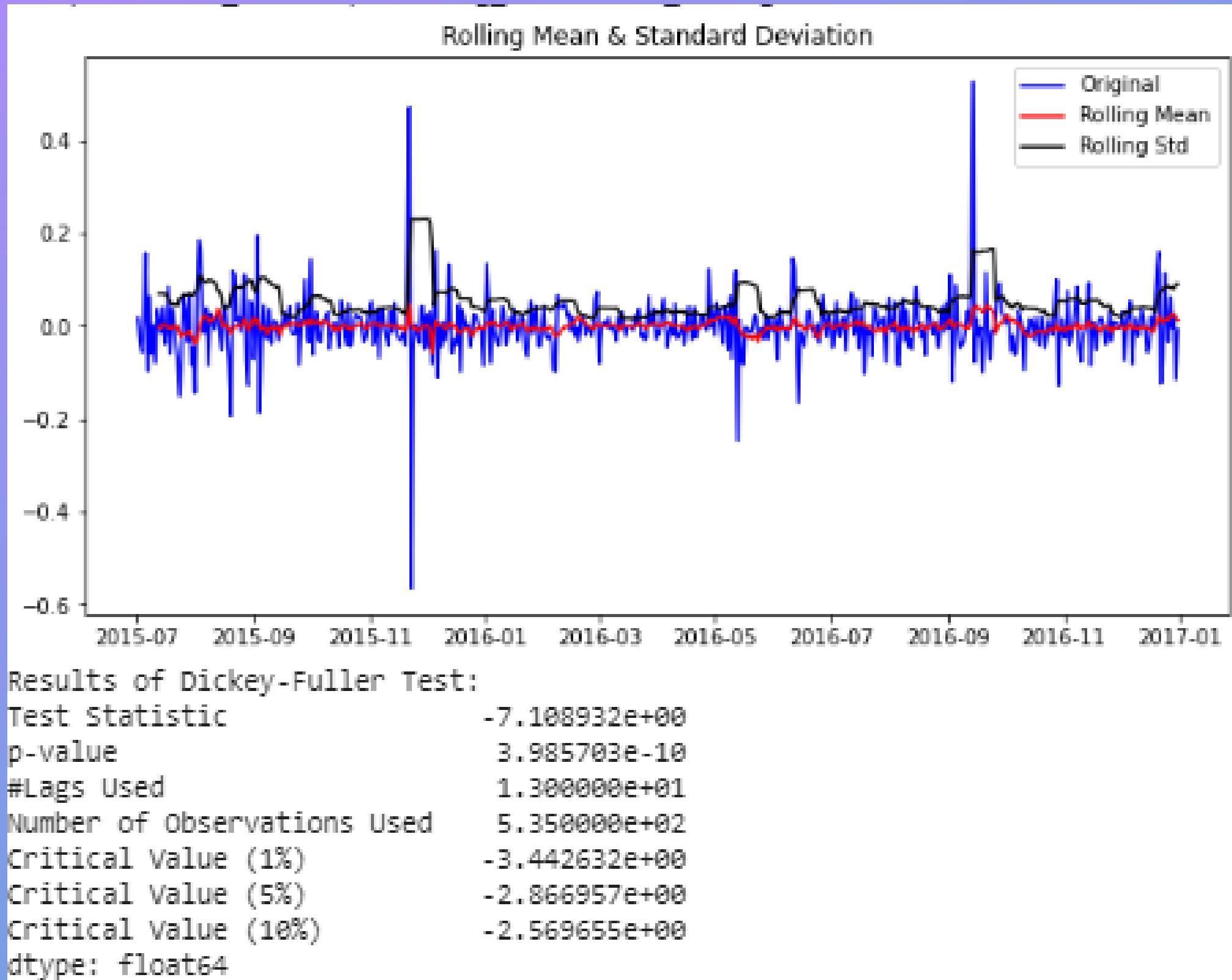
Test Estacionario



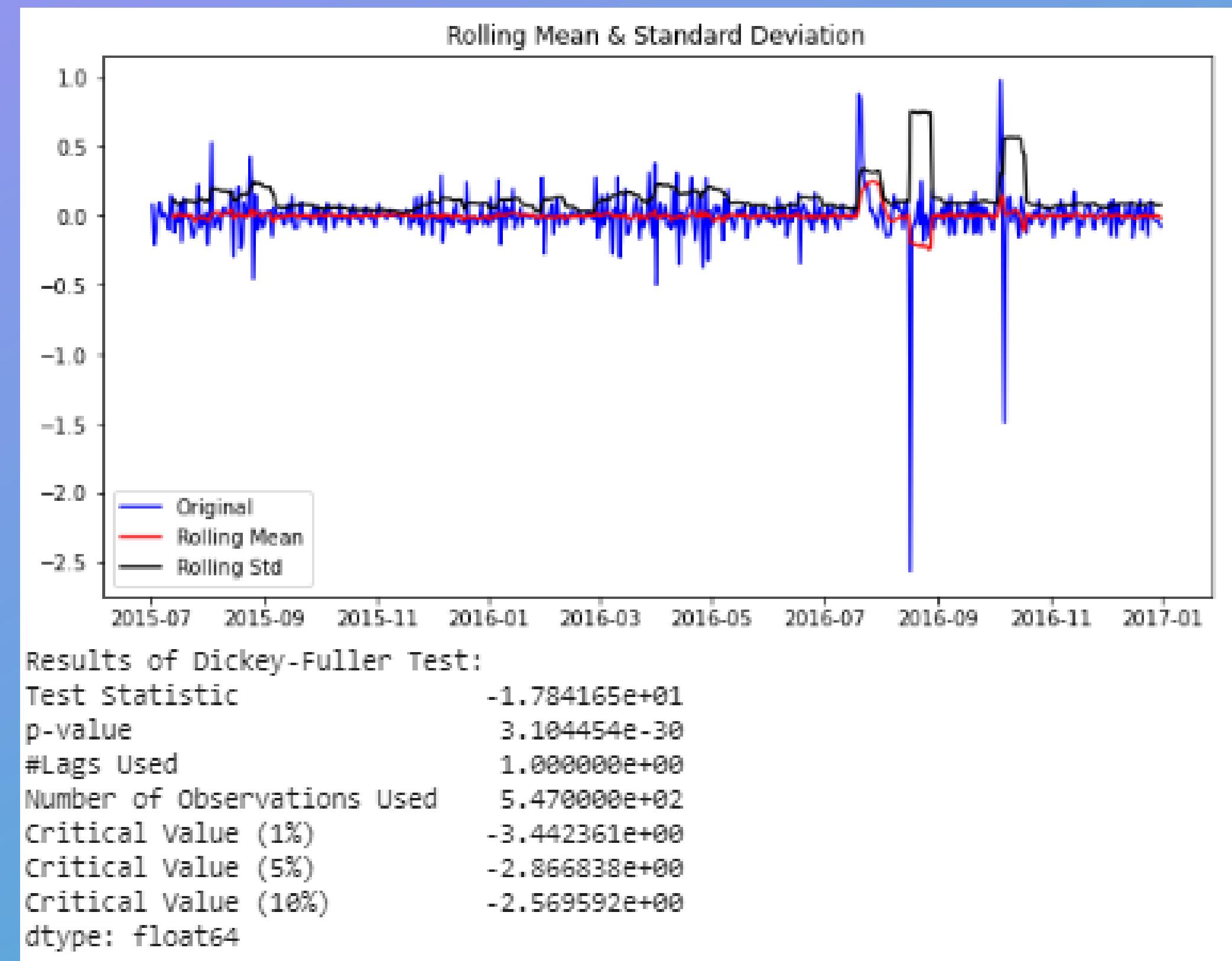
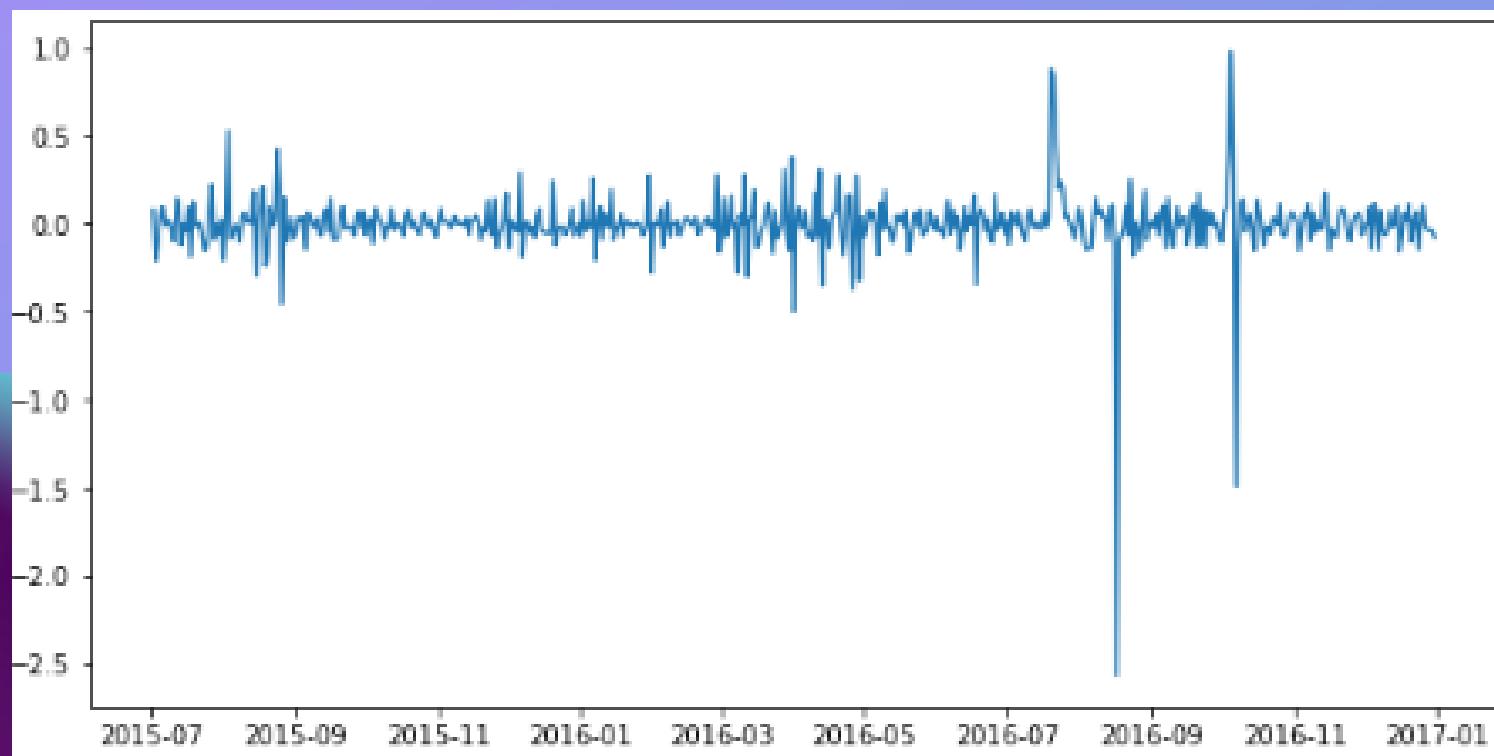
Test Estacionario



Test Estacionario

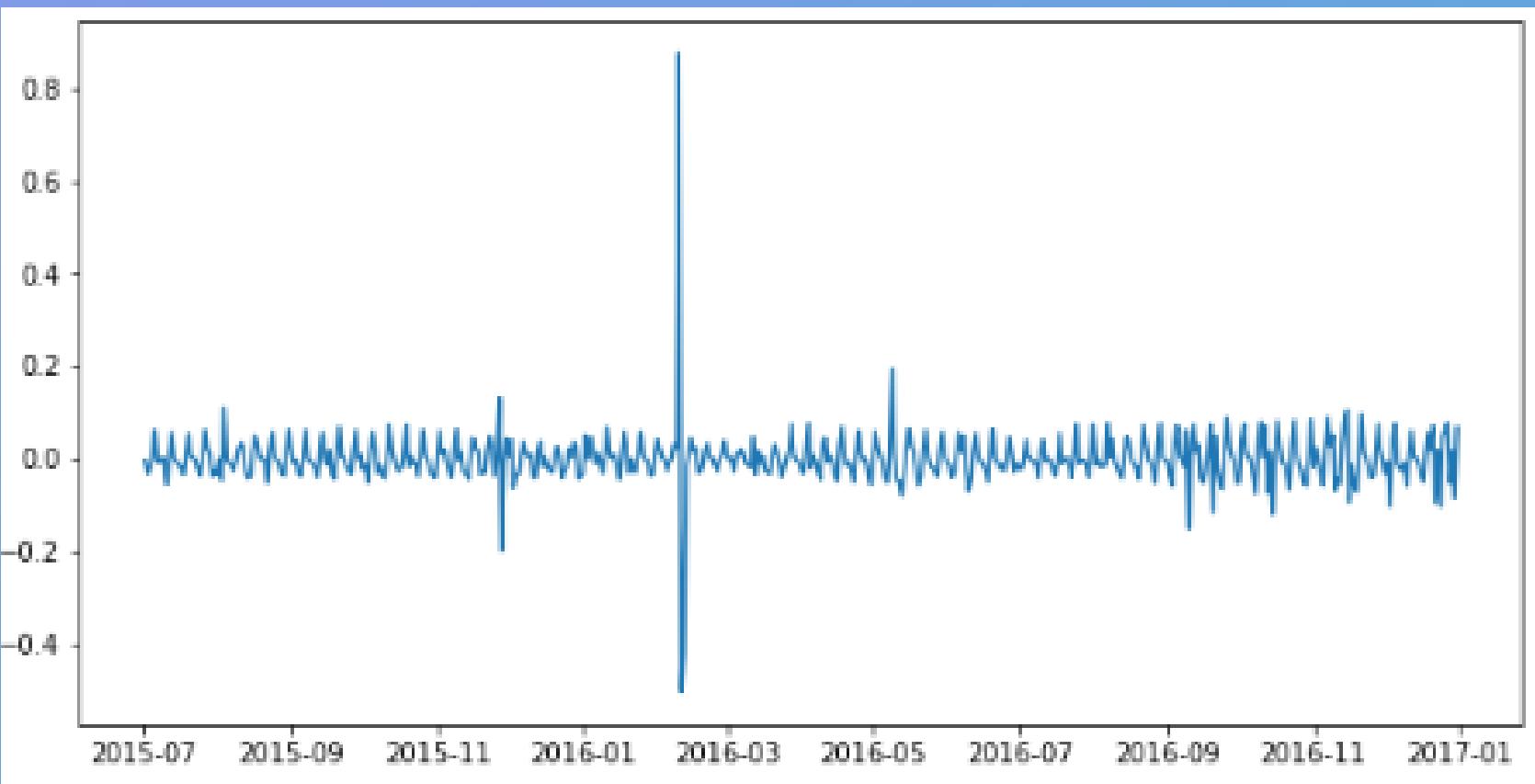
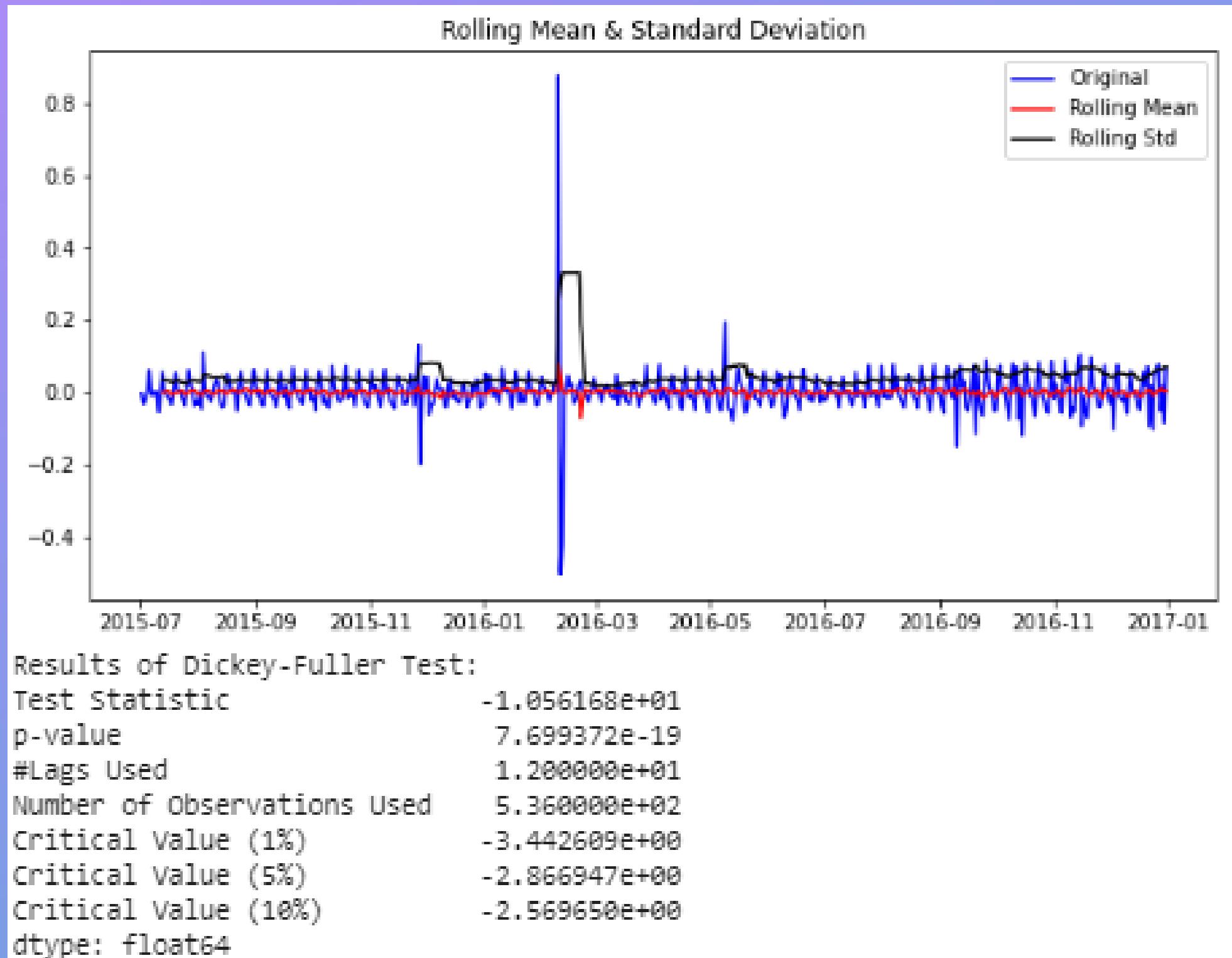


Test Estacionario



ESPAÑOL

Test Estacionario



26

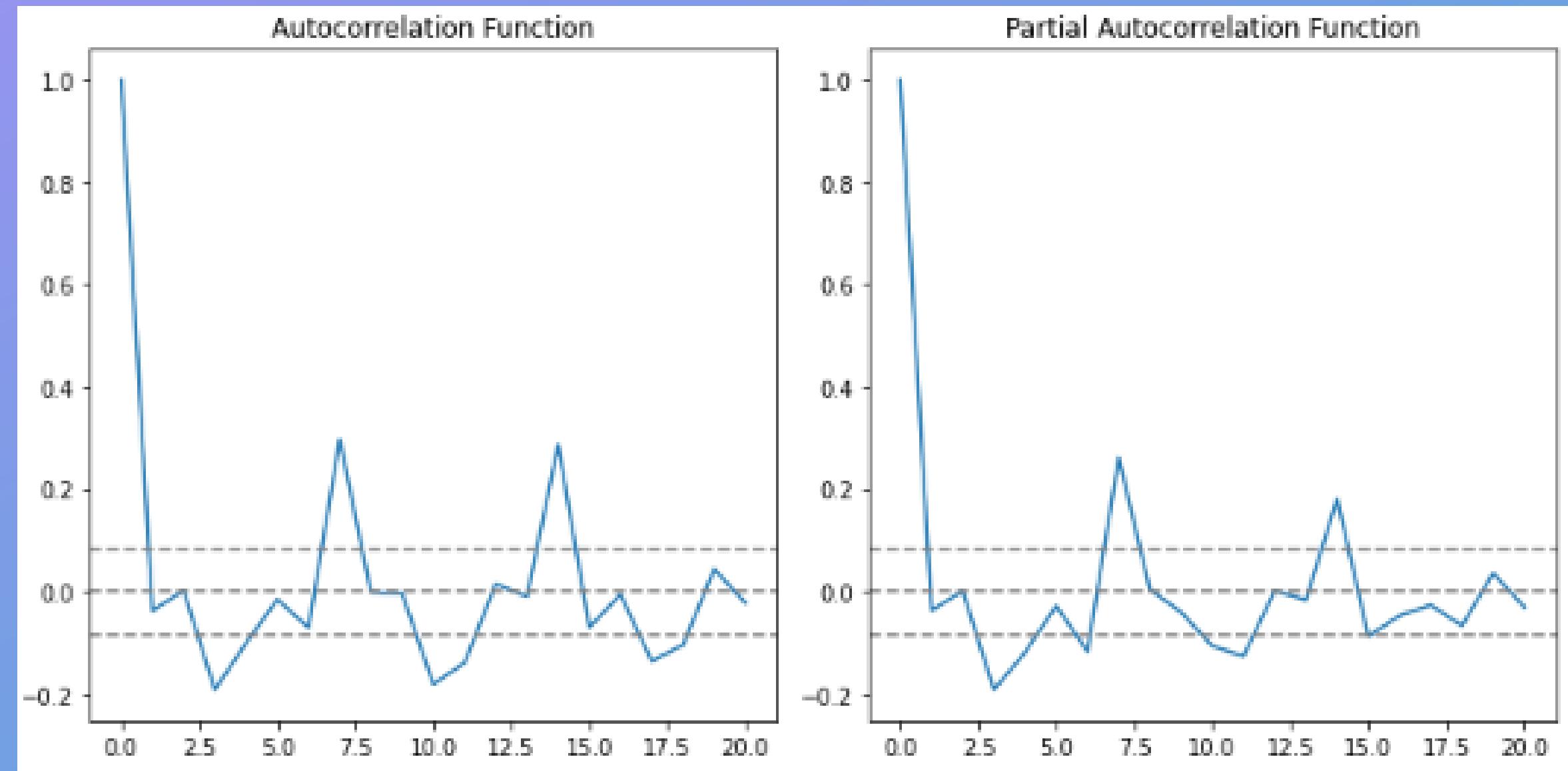
02

ARIMA

Implementación

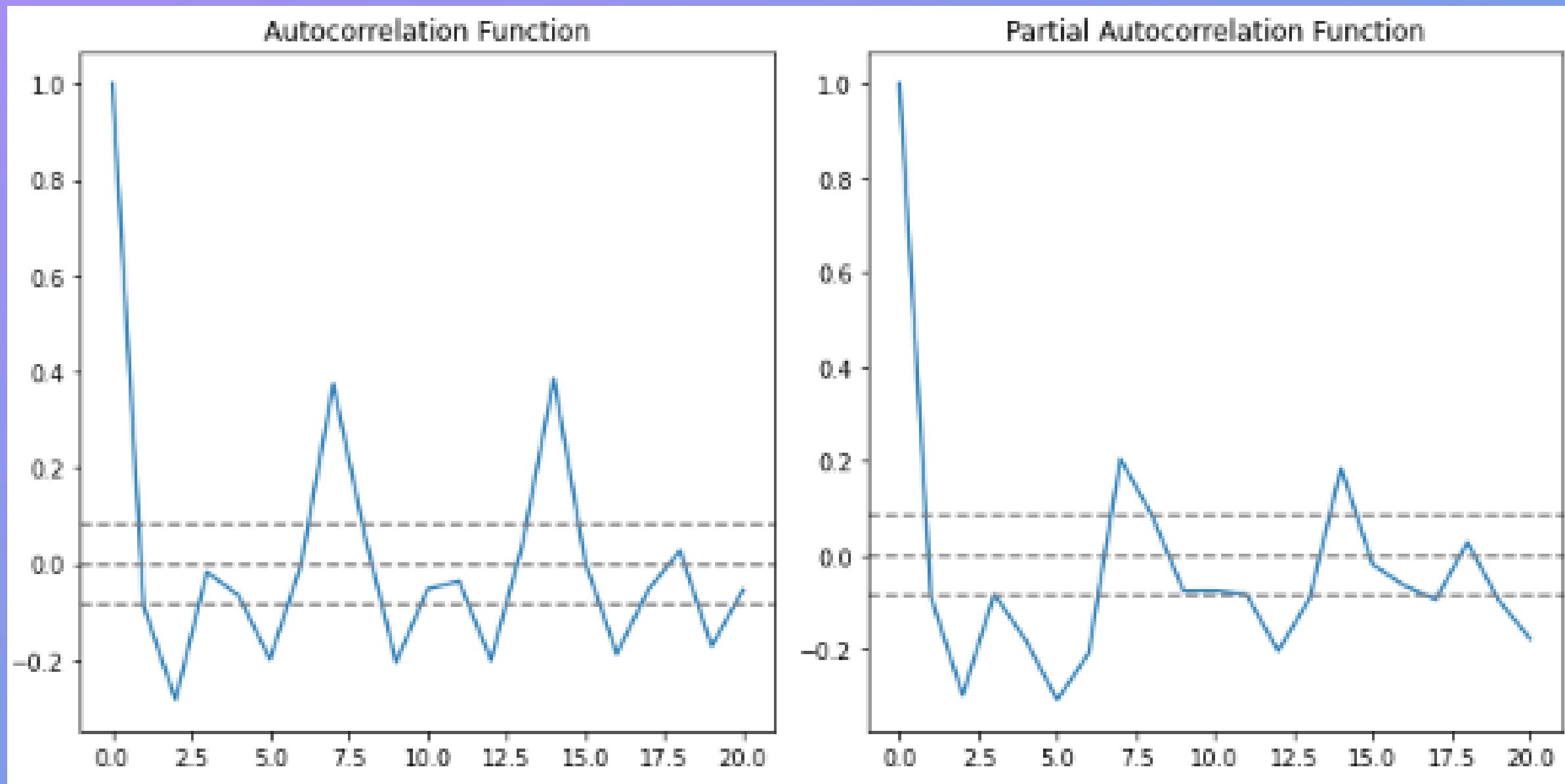
| RESULTADOS |

$p = 2$
 $q = 2$



JAPONÉS

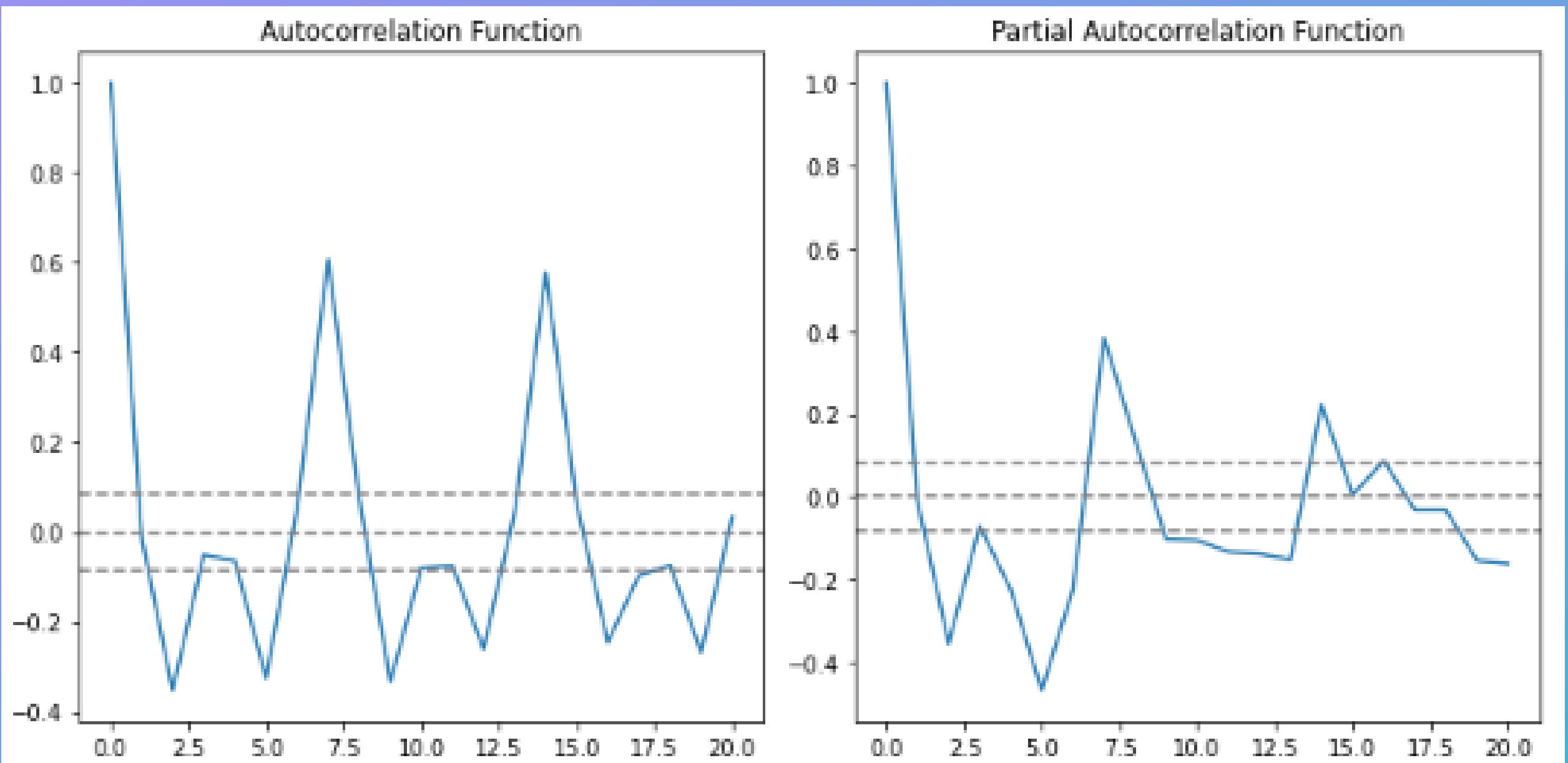
ARIMA



$p = 2$
 $q = 2$

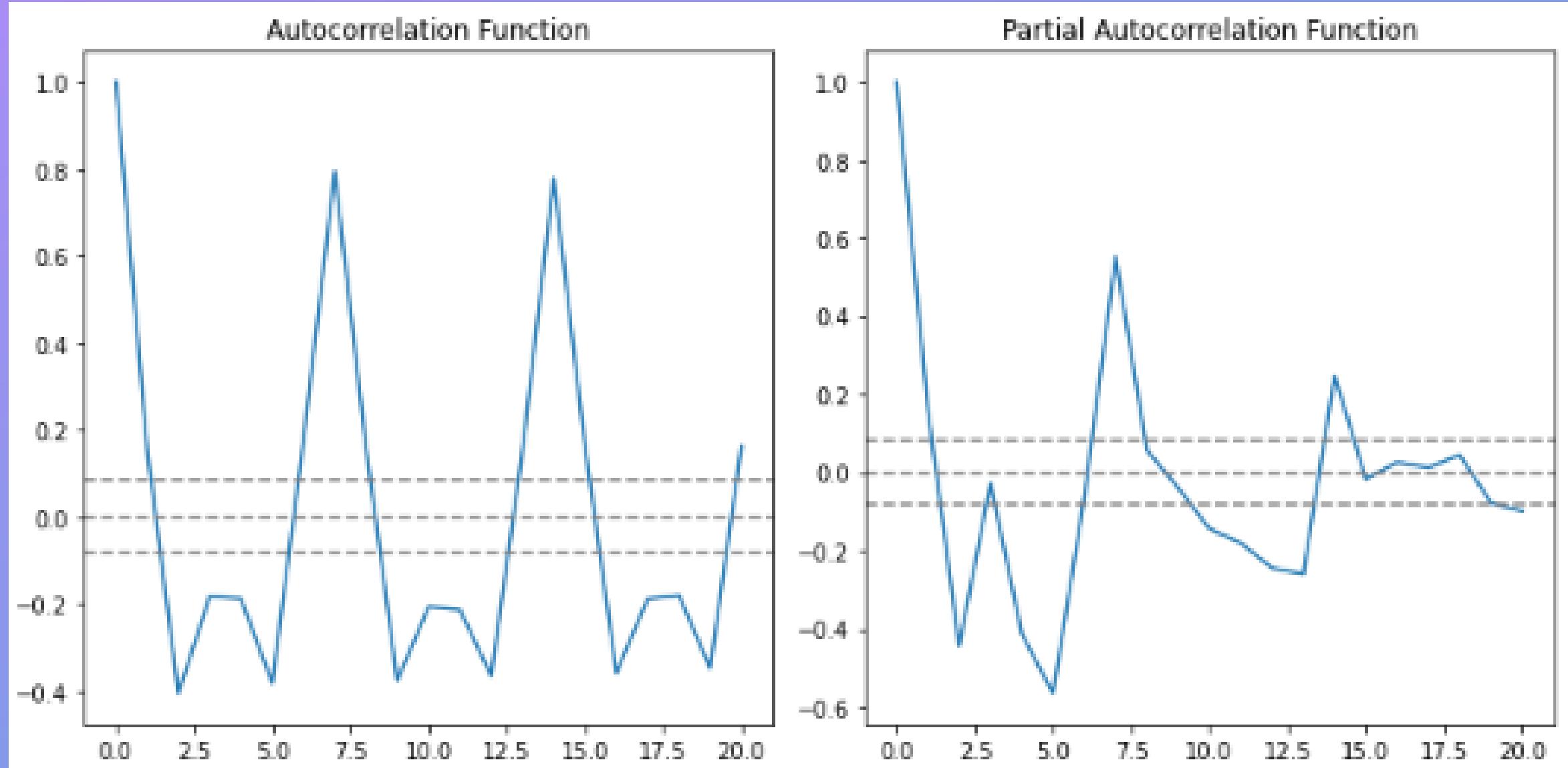
28

$p = 2$
 $q = 2$



COMÚN

ARIMA



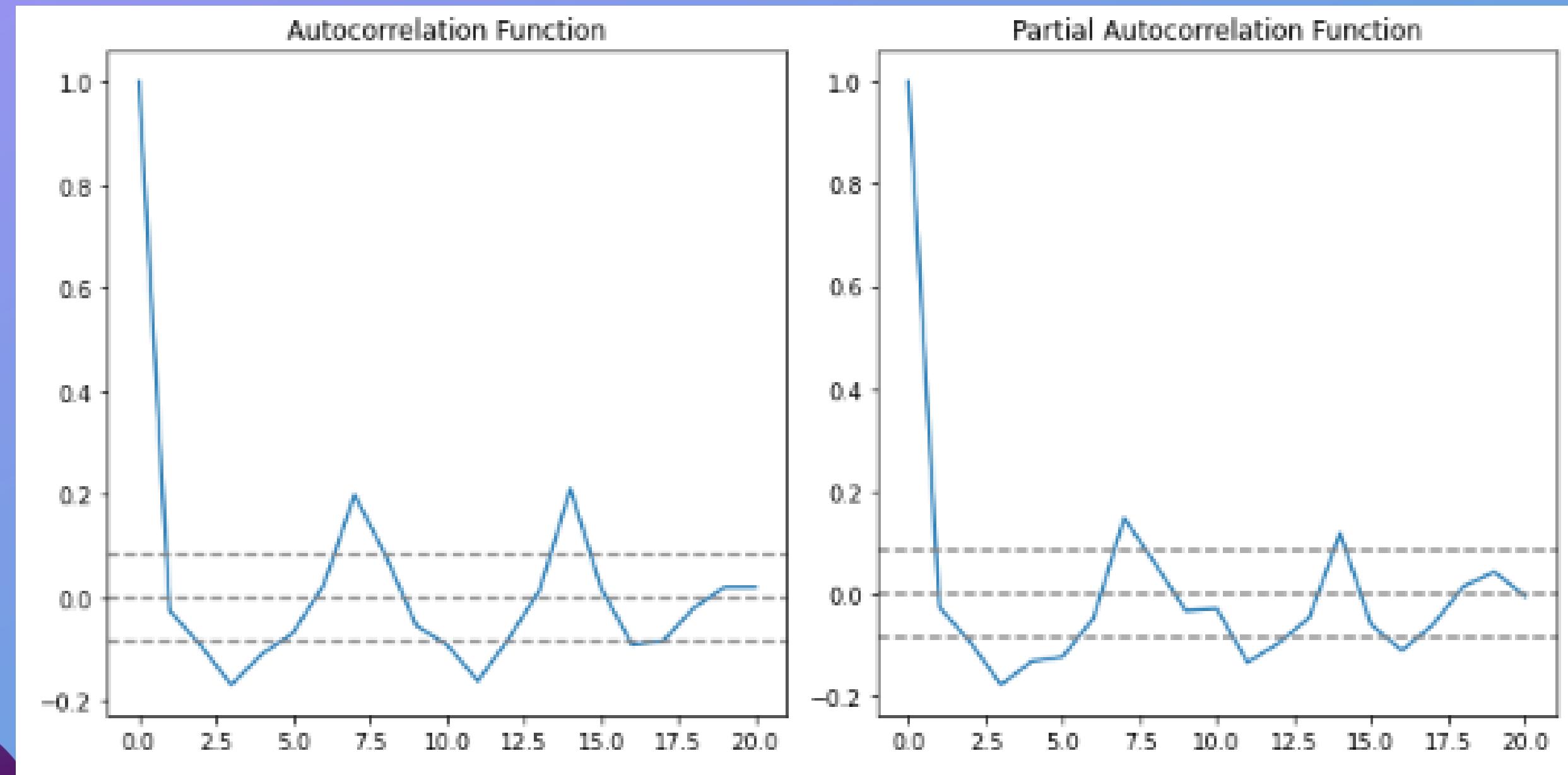
$p = 4$
 $q = 4$

30

FRANCÉS

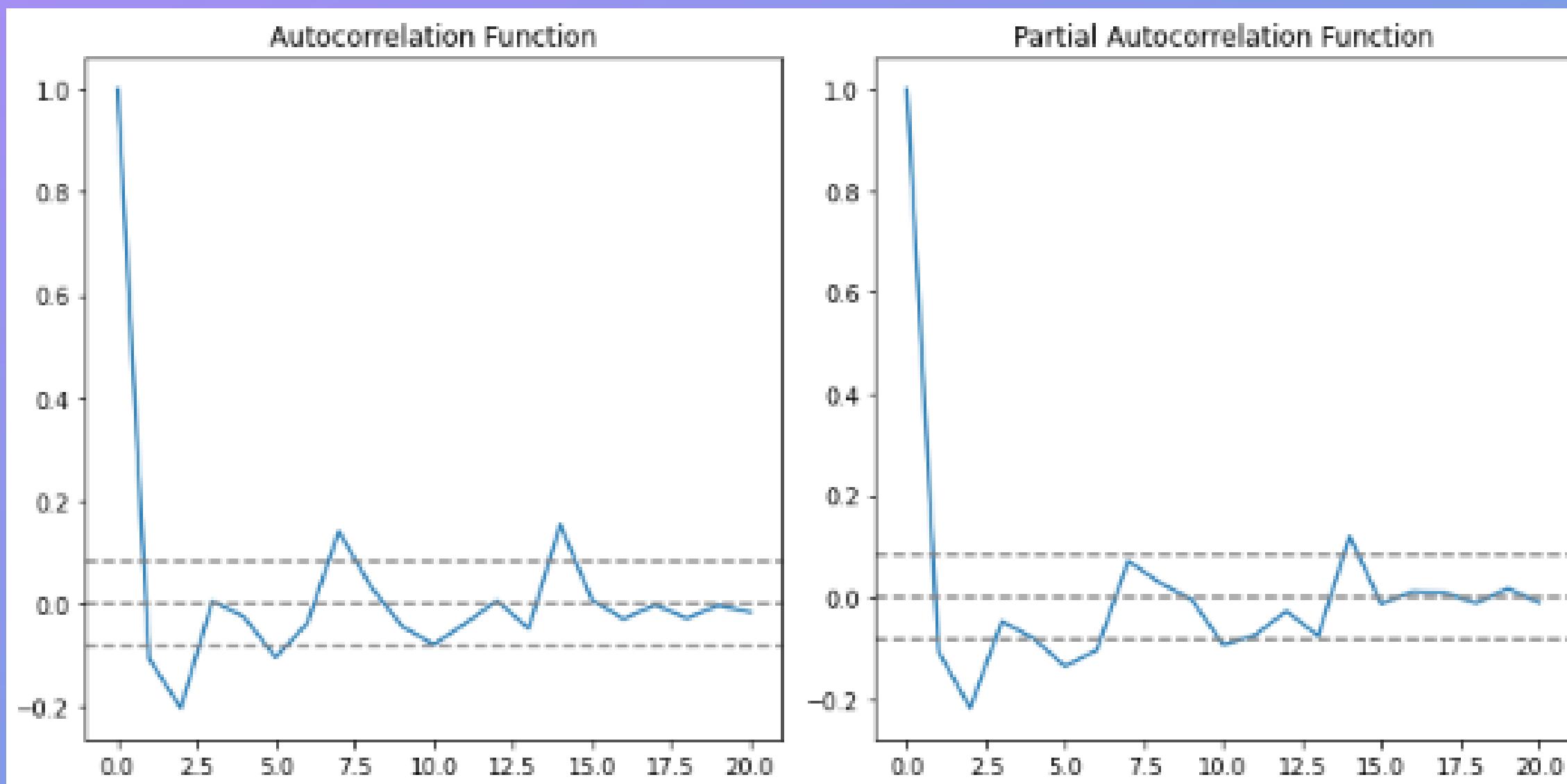
ARIMA

$p = 4$
 $q = 4$



CH1NO

ARIMA

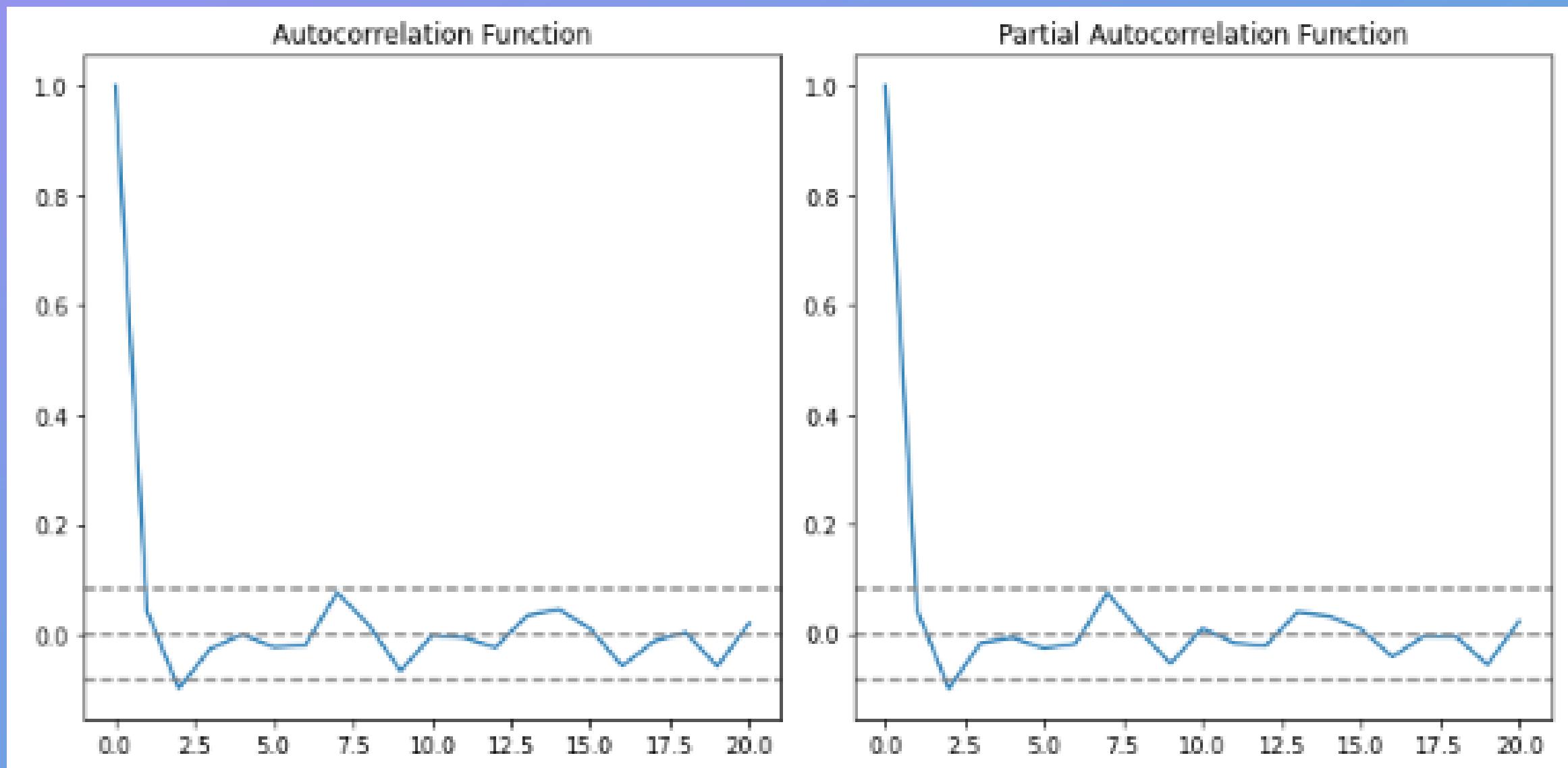


$p = 2$
 $q = 1$

32

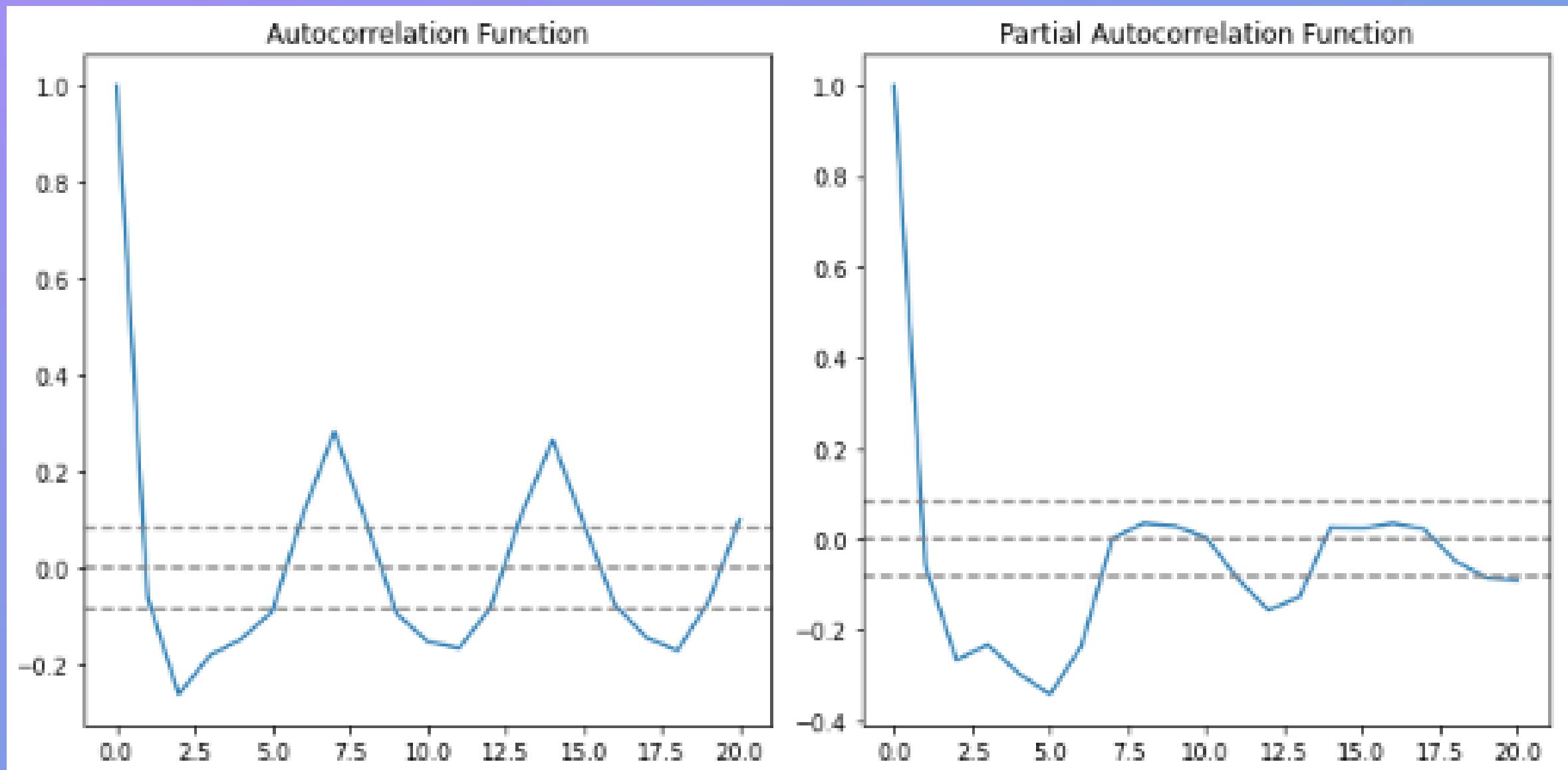
ARIMA

$p = 2$
 $q = 2$



ESPAÑOL

ARIMA



$p = 2$
 $q = 2$

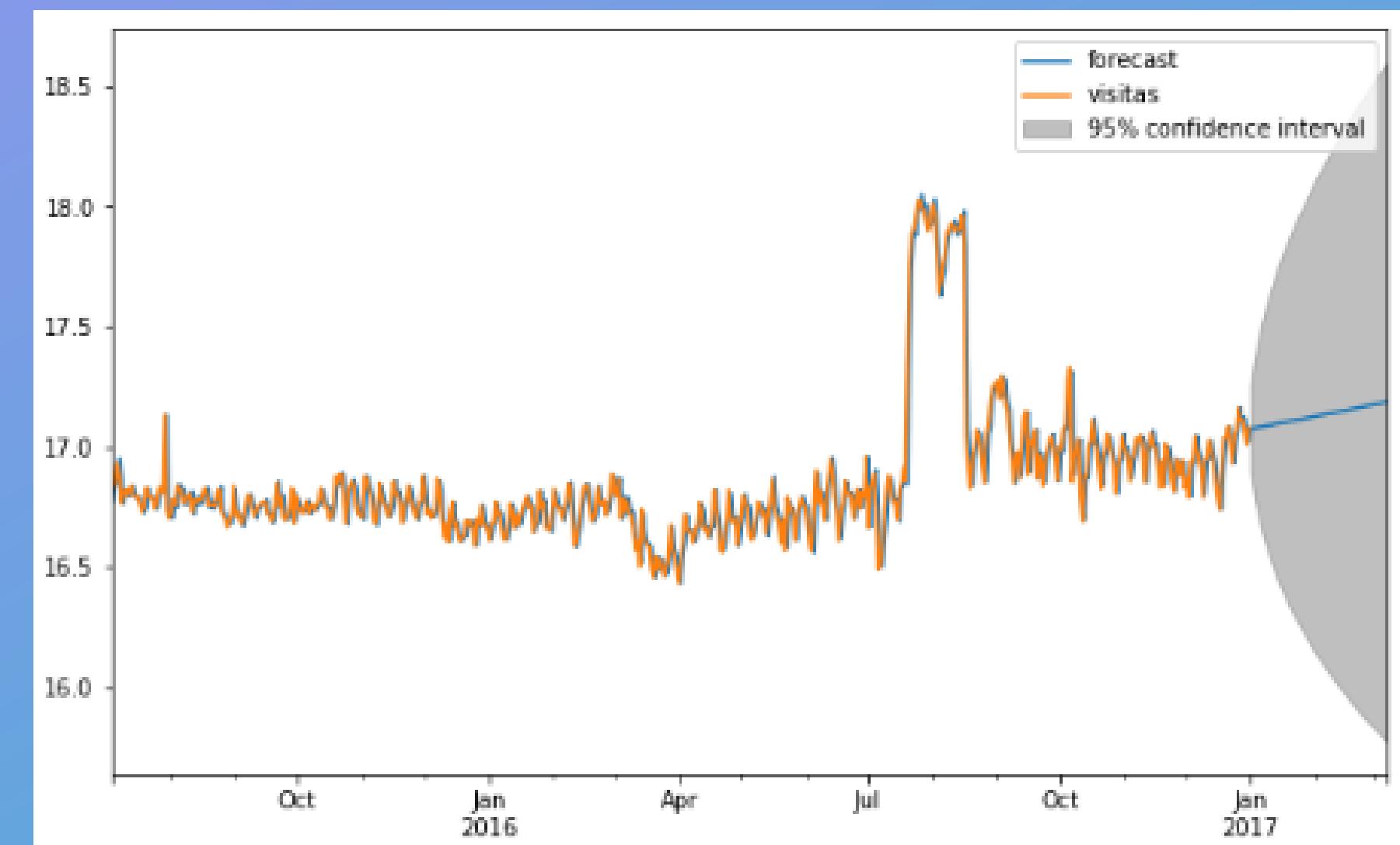
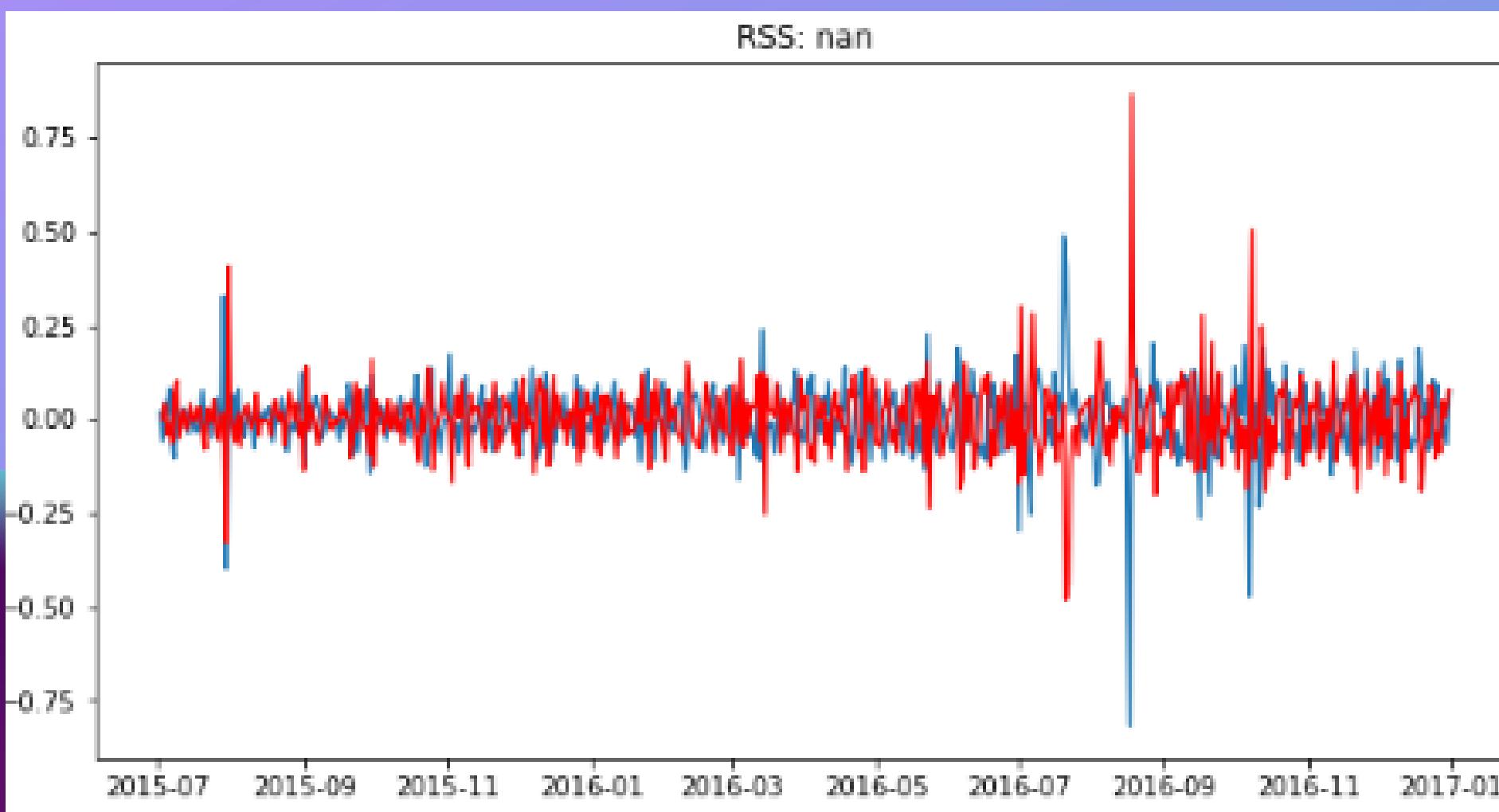
35

02

ARIMA
PREDICCIÓN

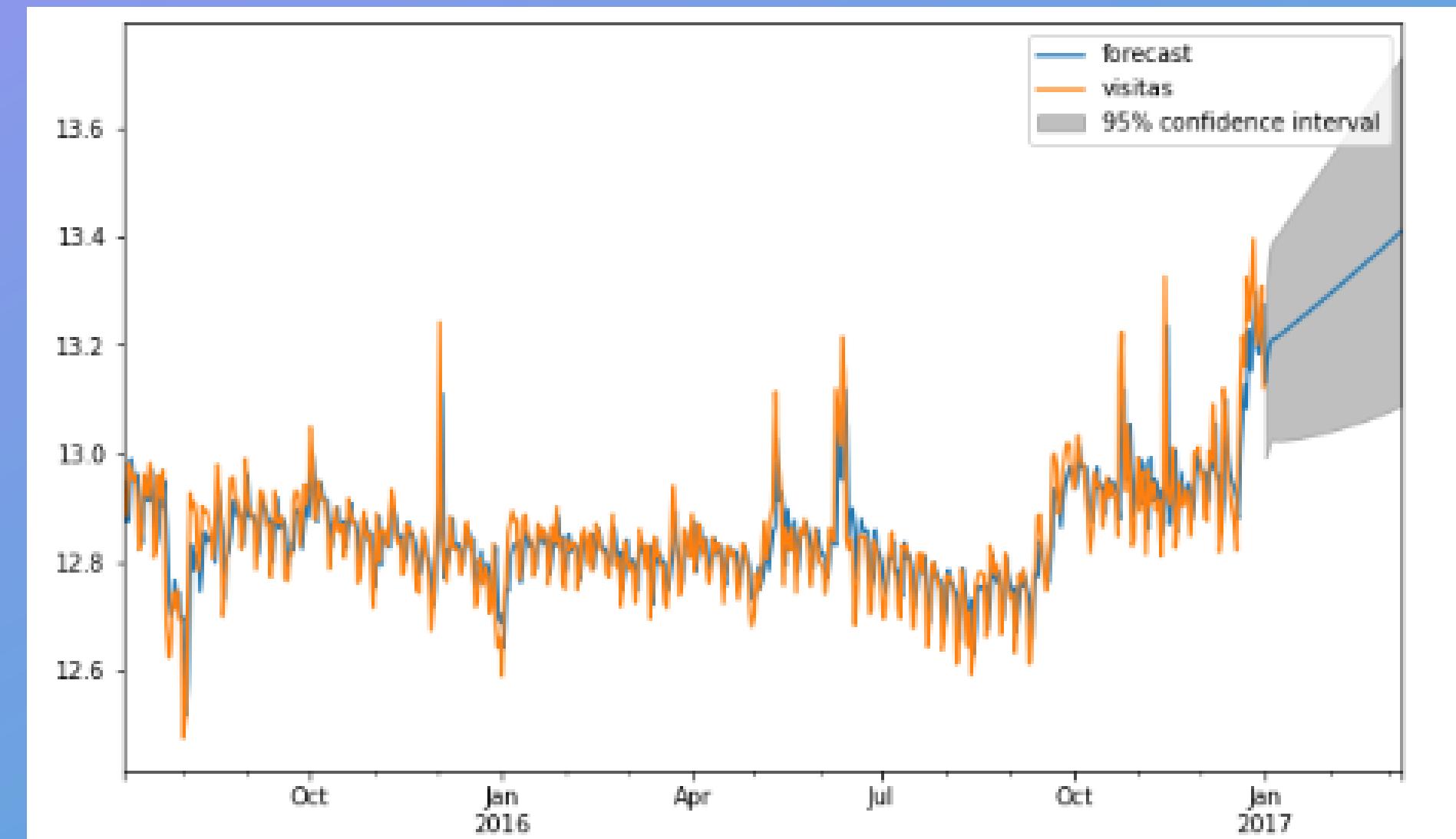
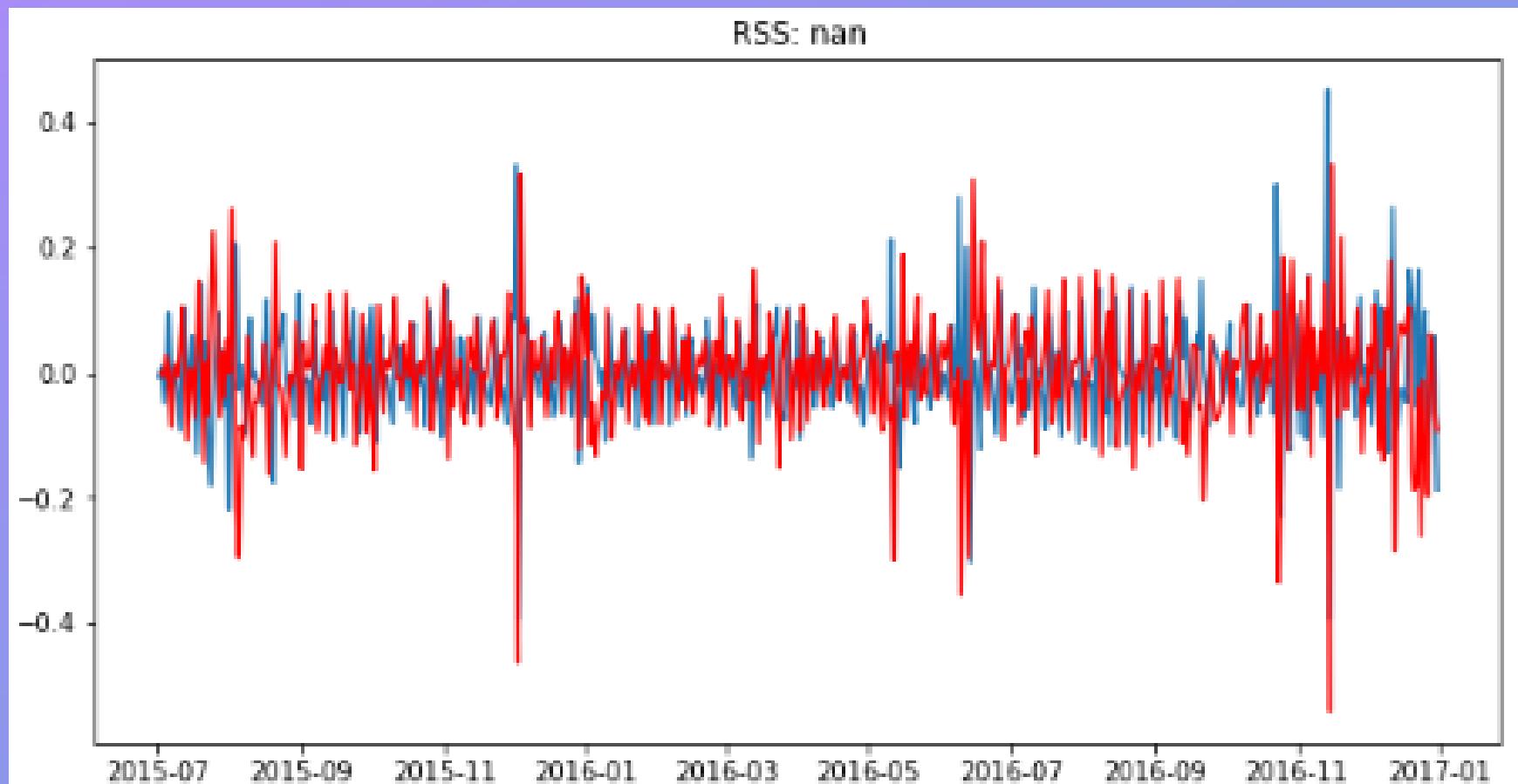
Implementación

| RESULTADOS |



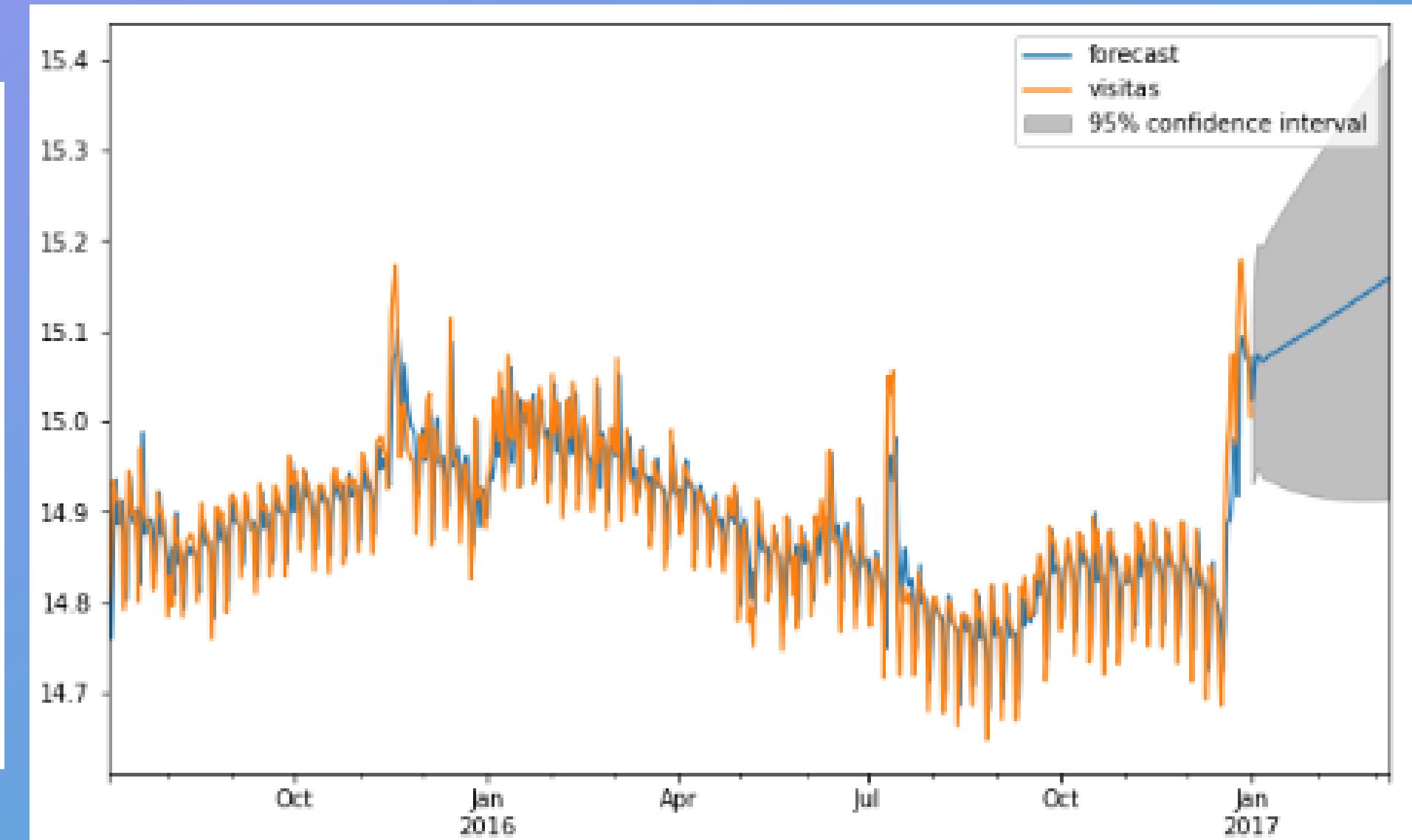
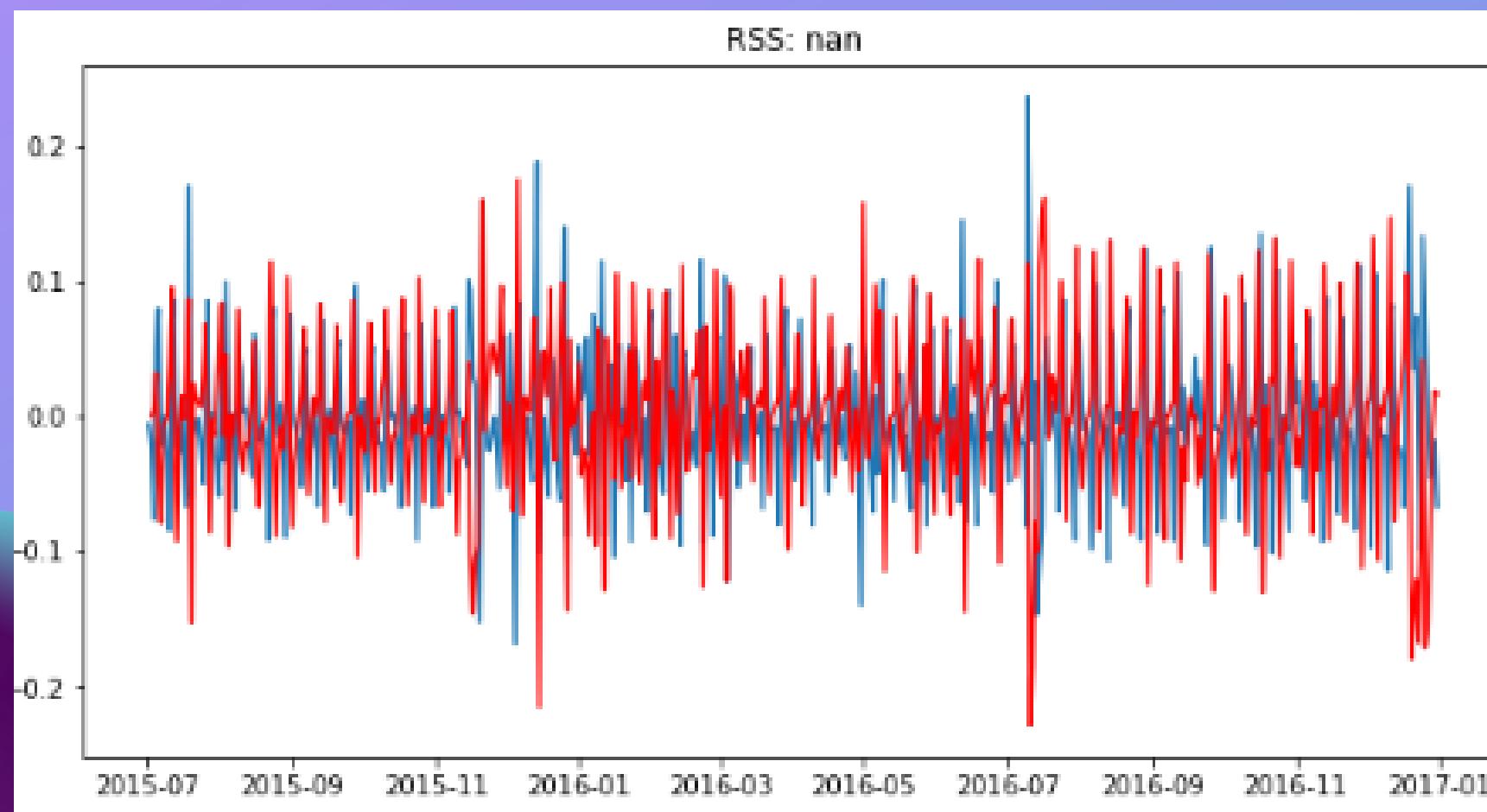
JAPONÉS

ARIMA



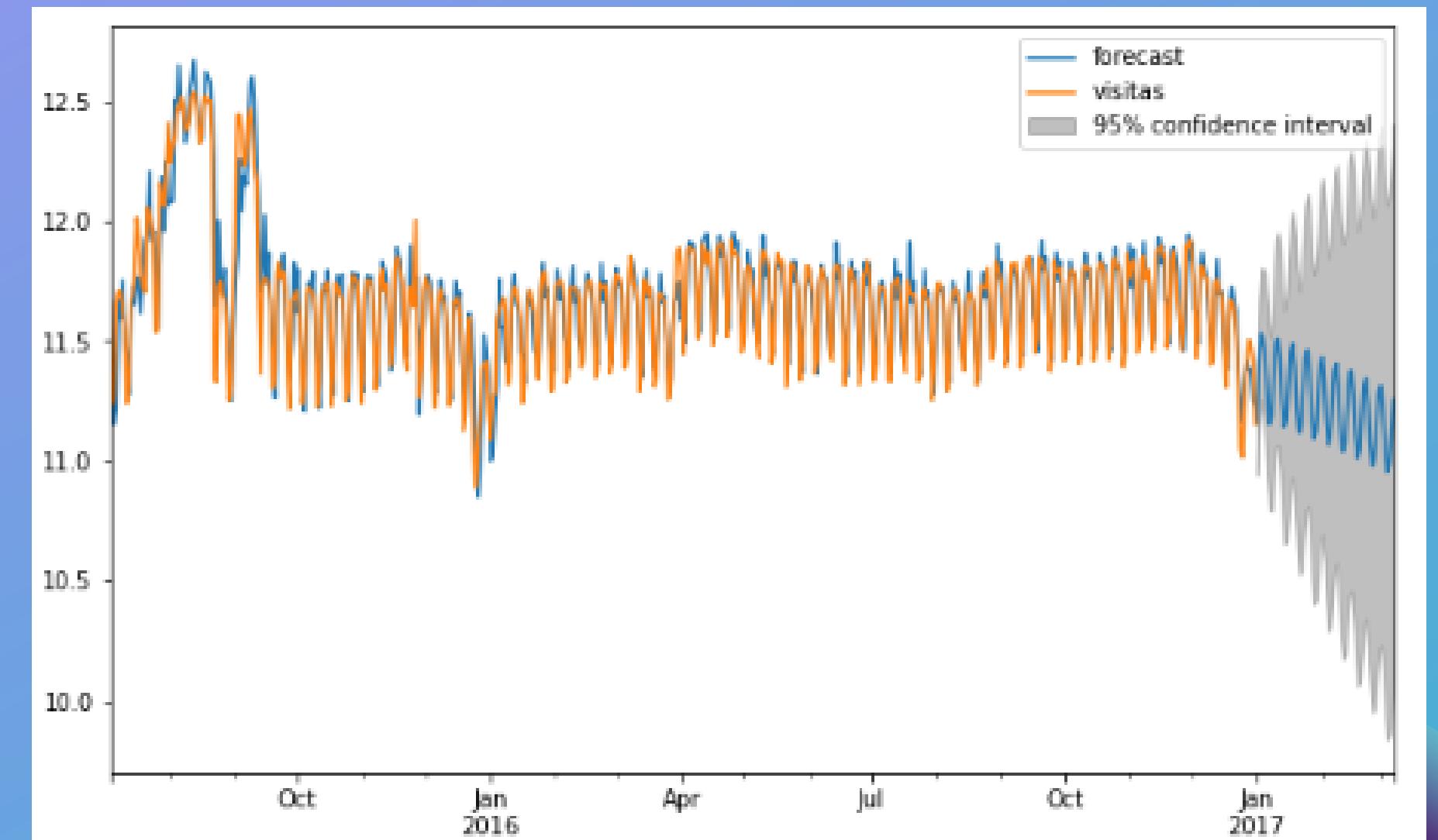
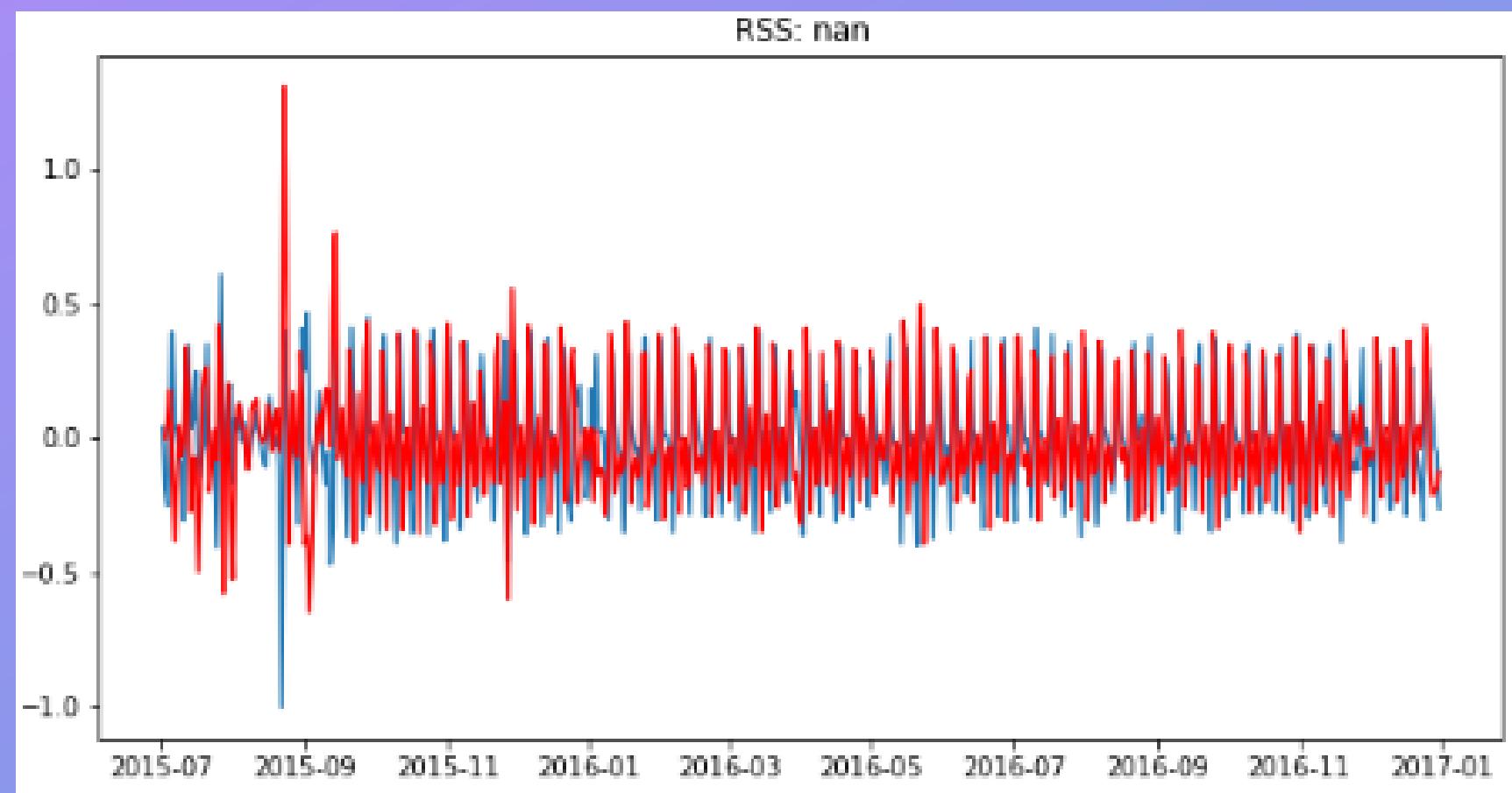
ALEMÁN

ARIMA



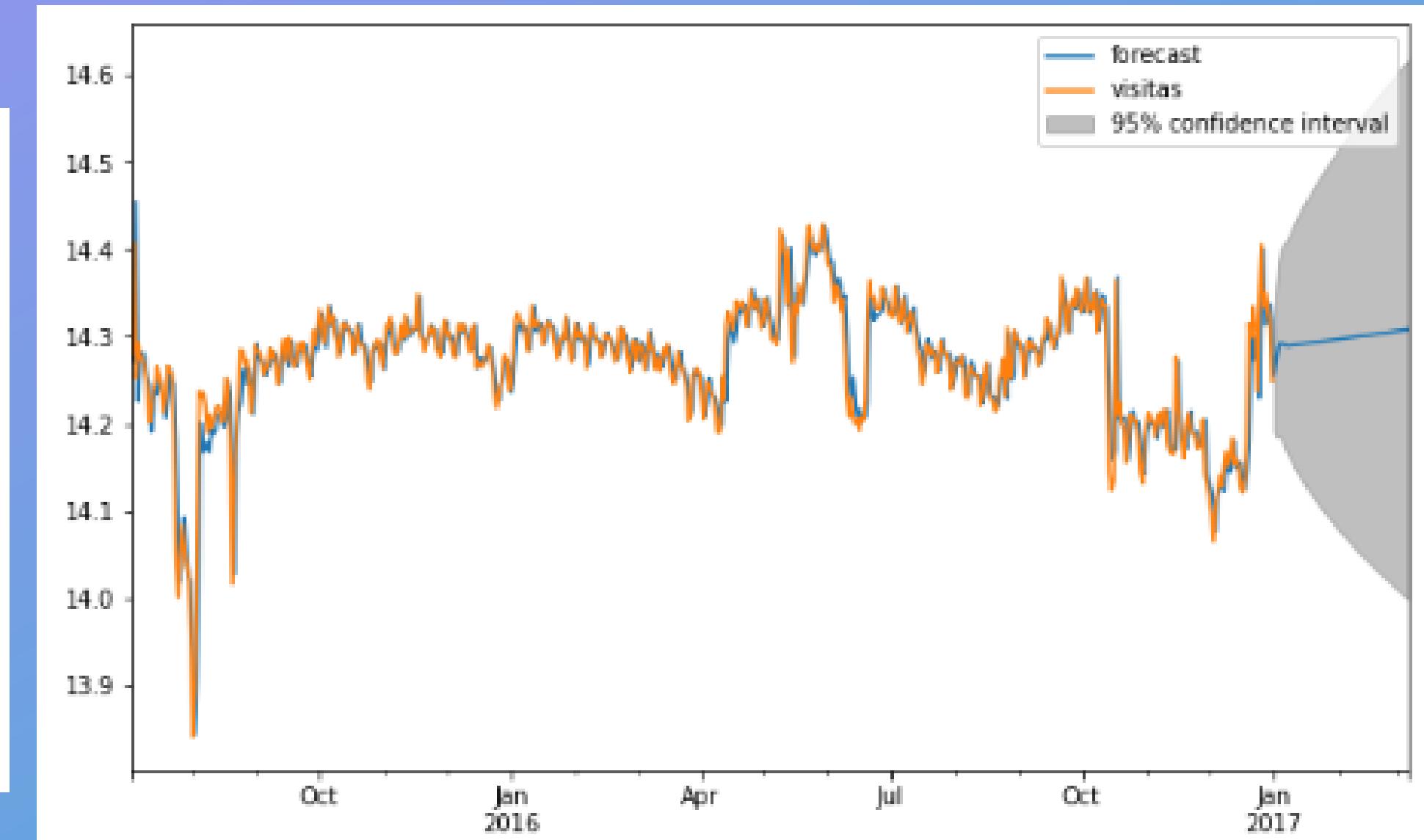
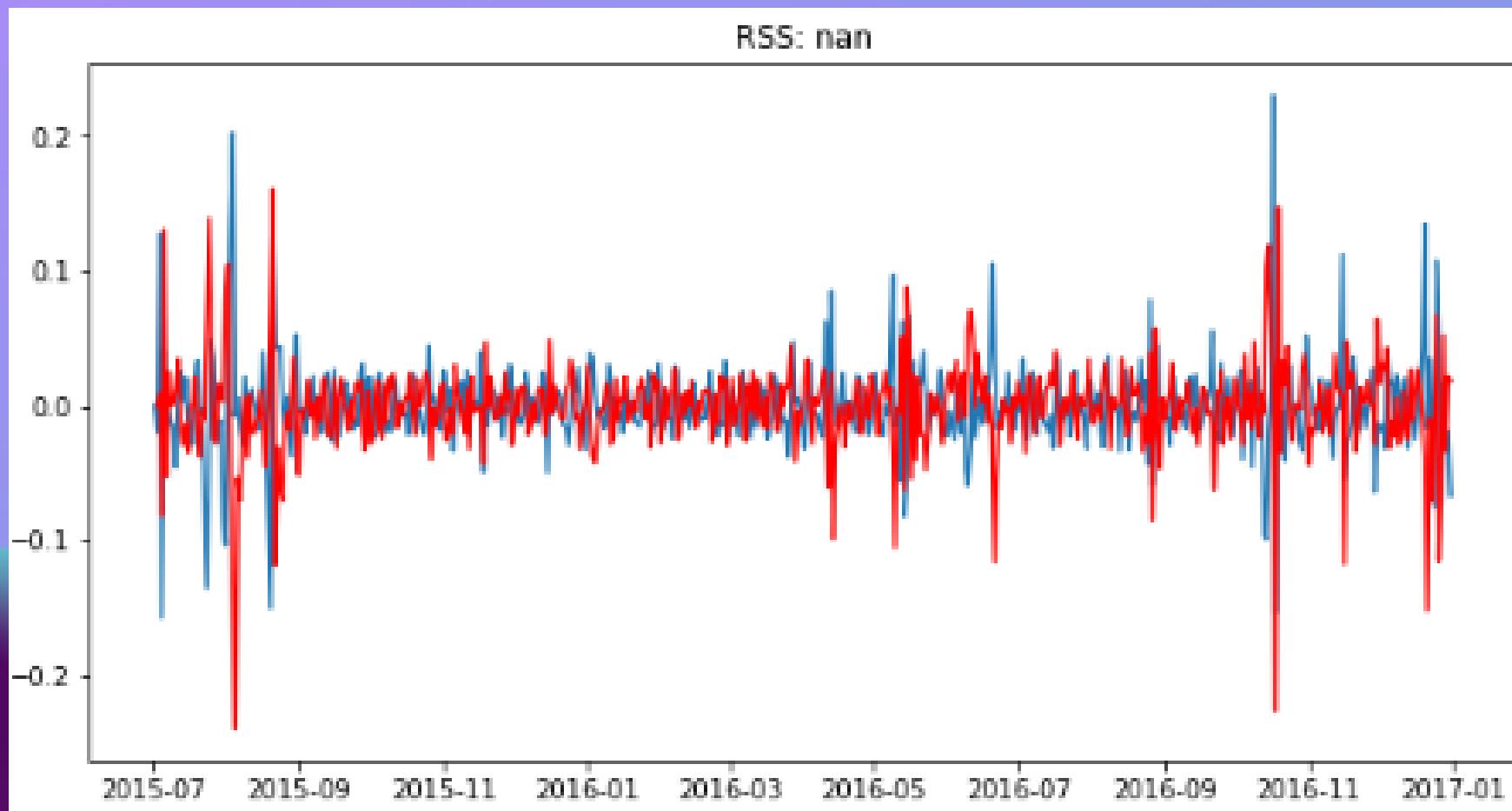
COMÚN

ARIMA



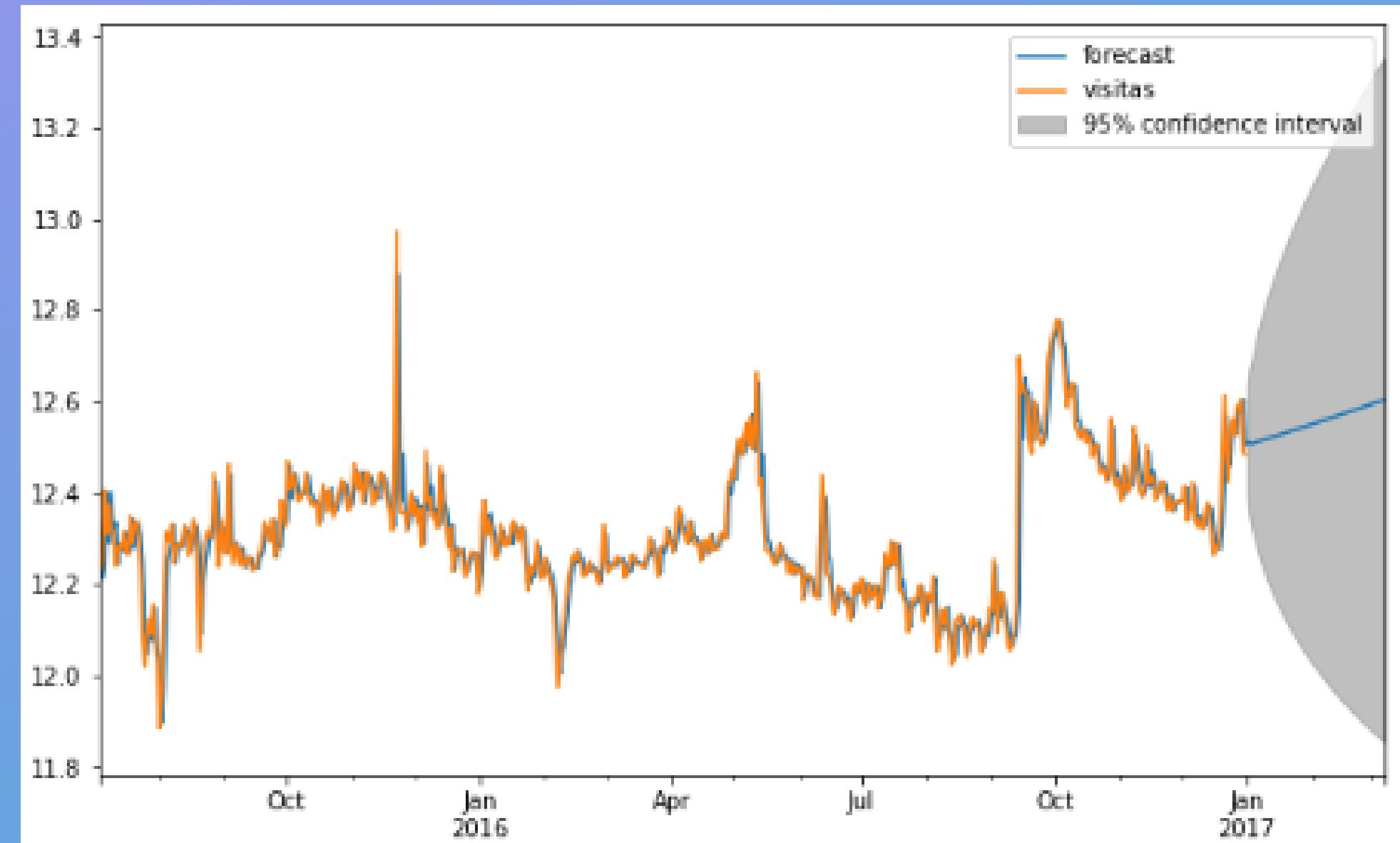
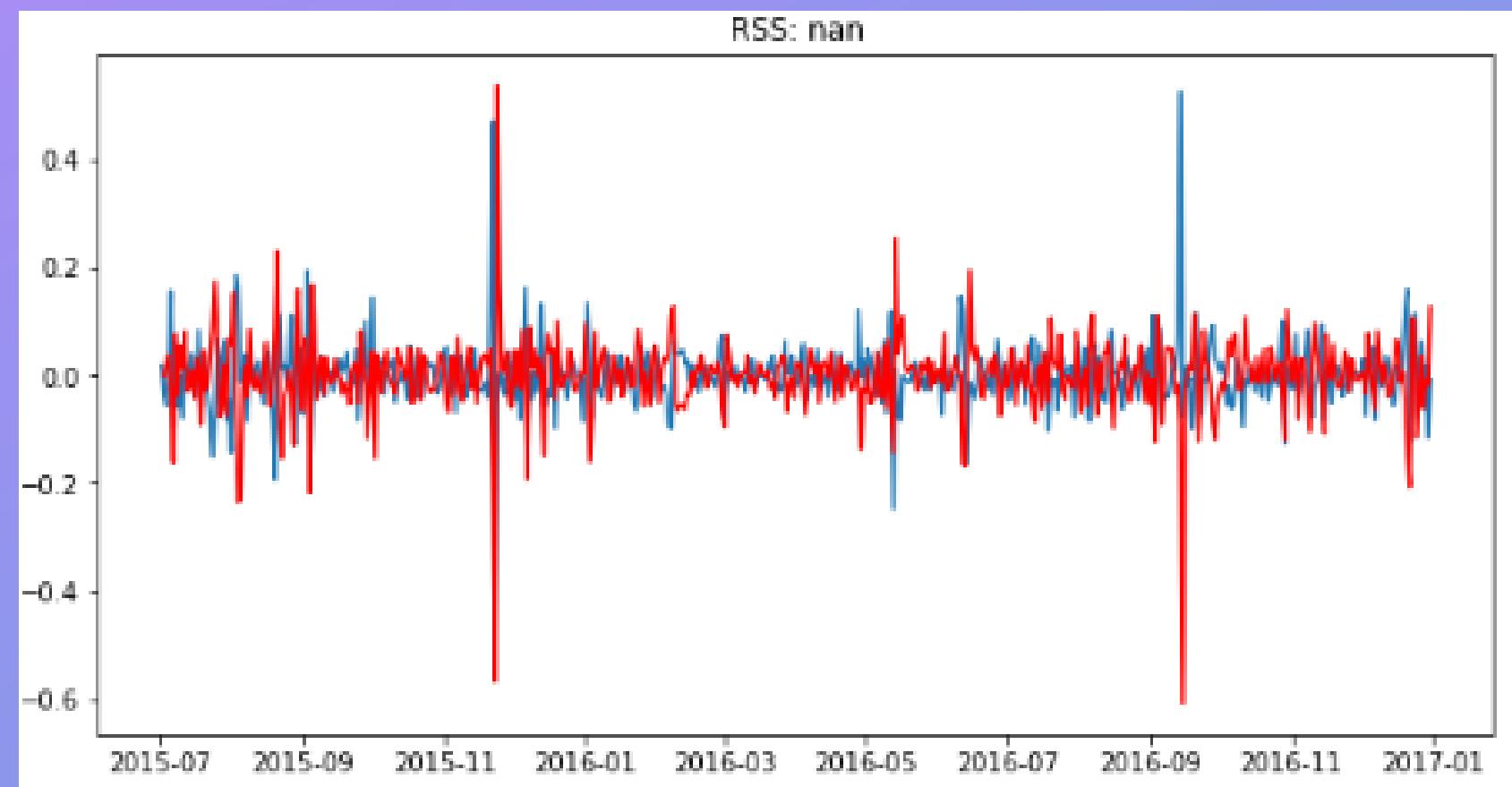
FRANCÉS

ARIMA



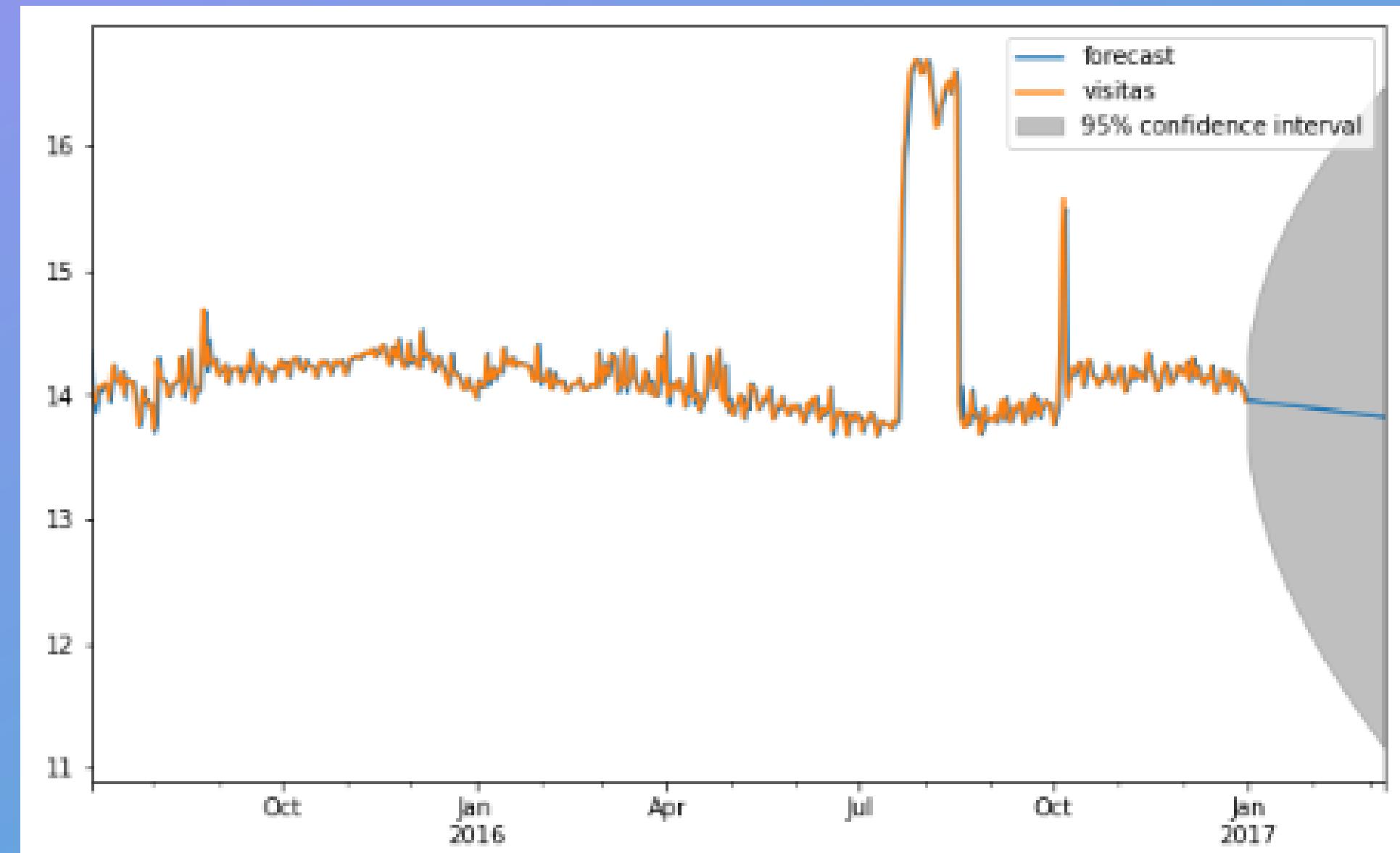
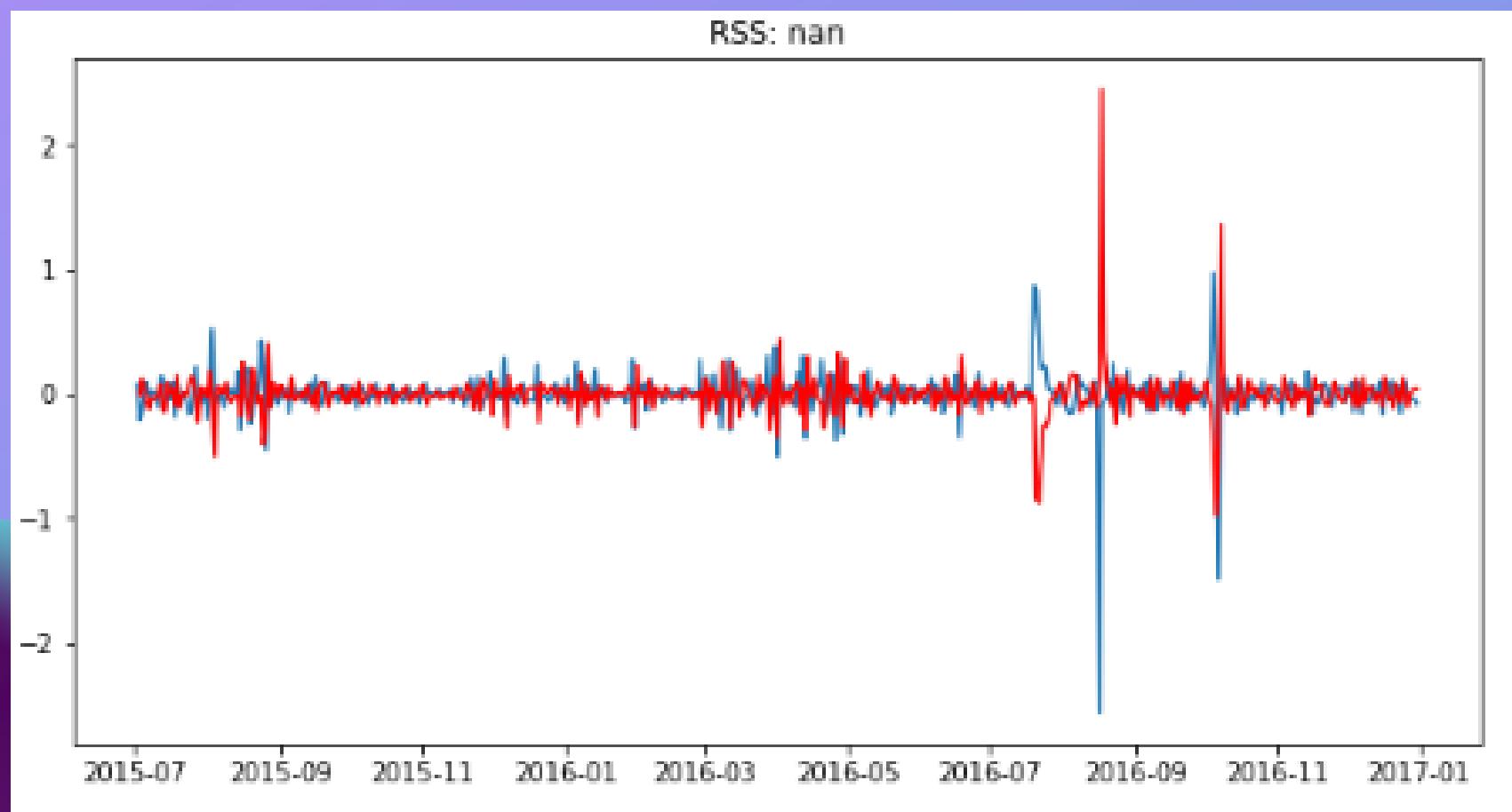
CHINO

ARIMA



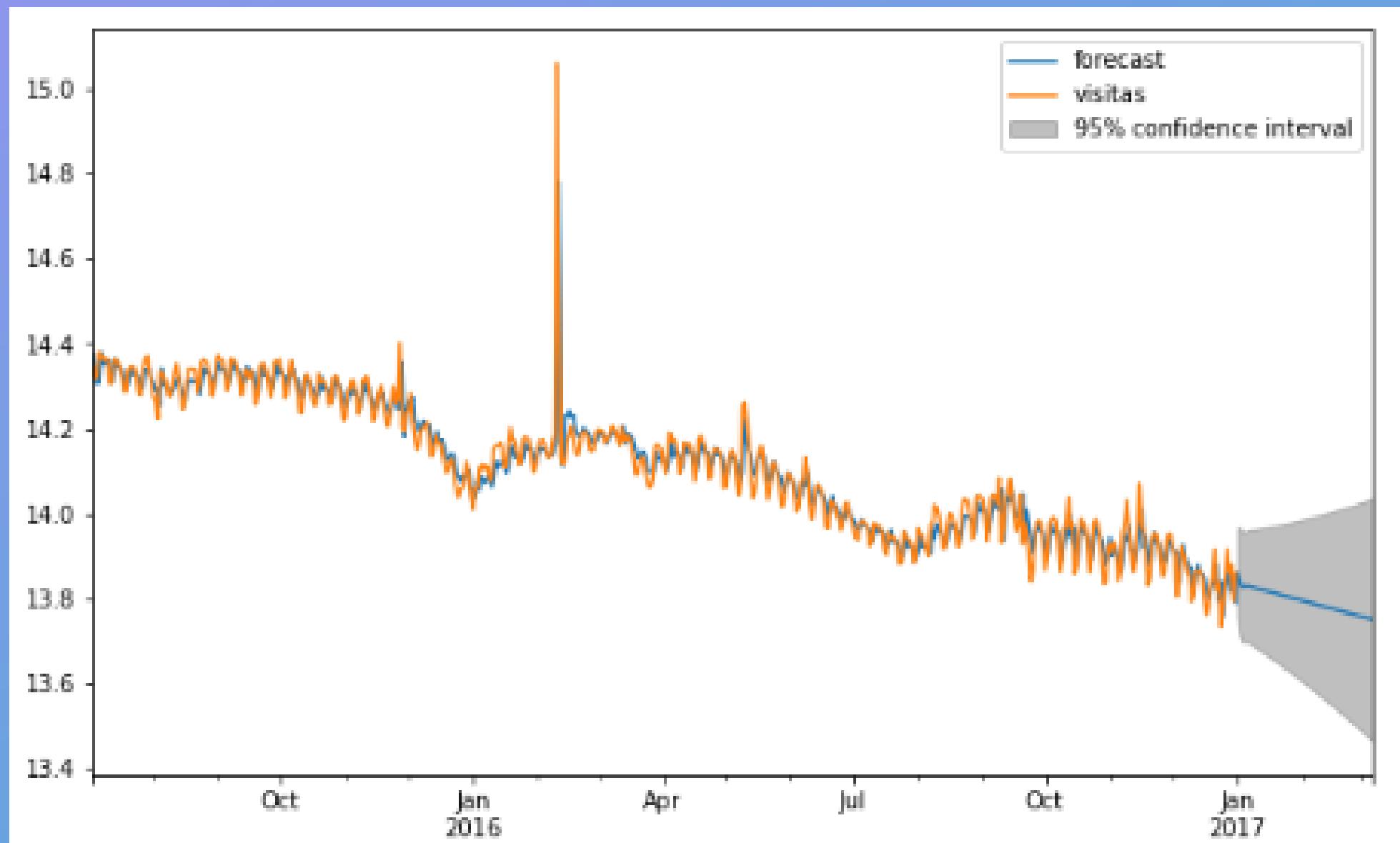
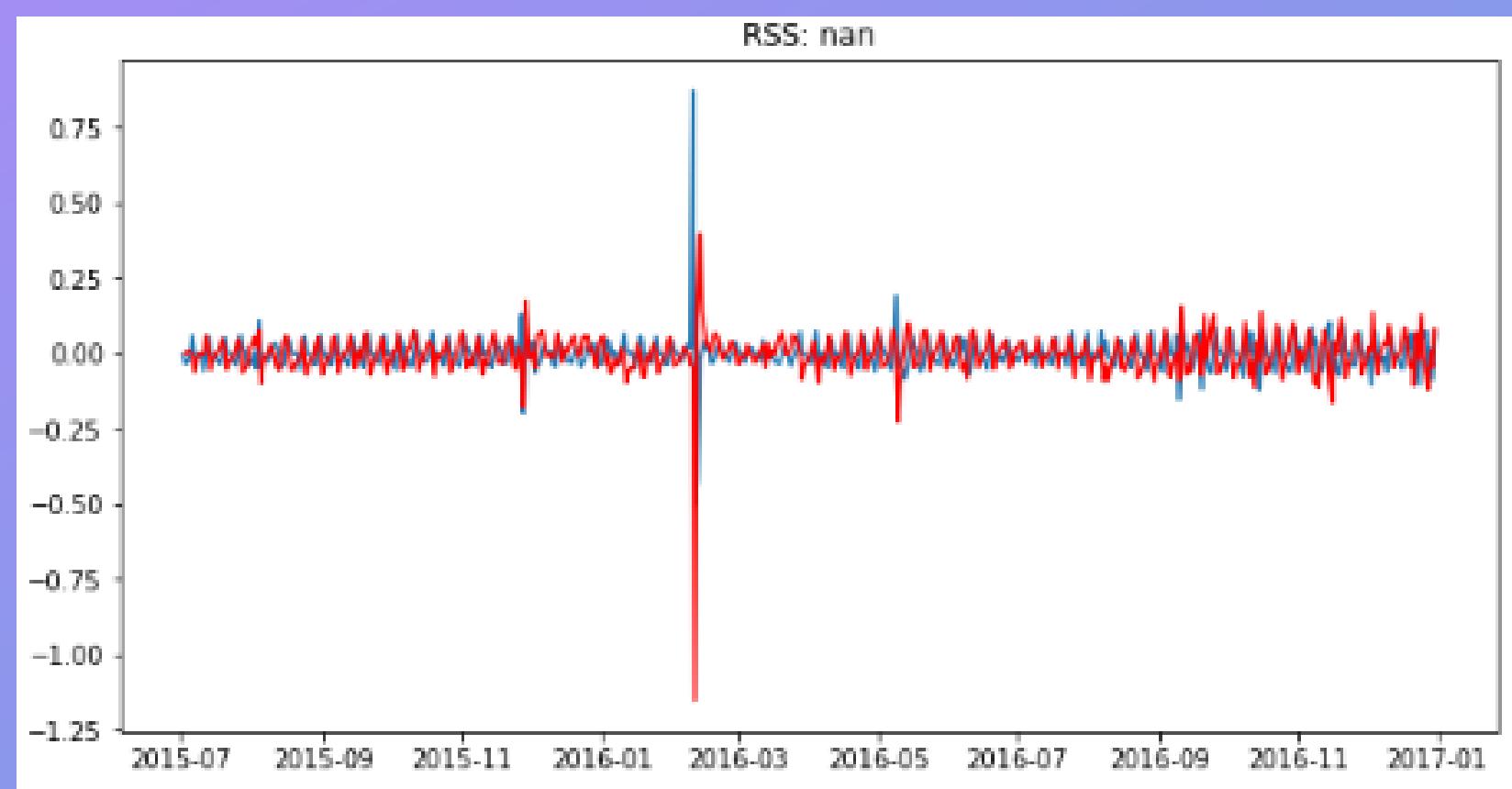
RUSO

ARIMA



ESPAÑOL

ARIMA



44

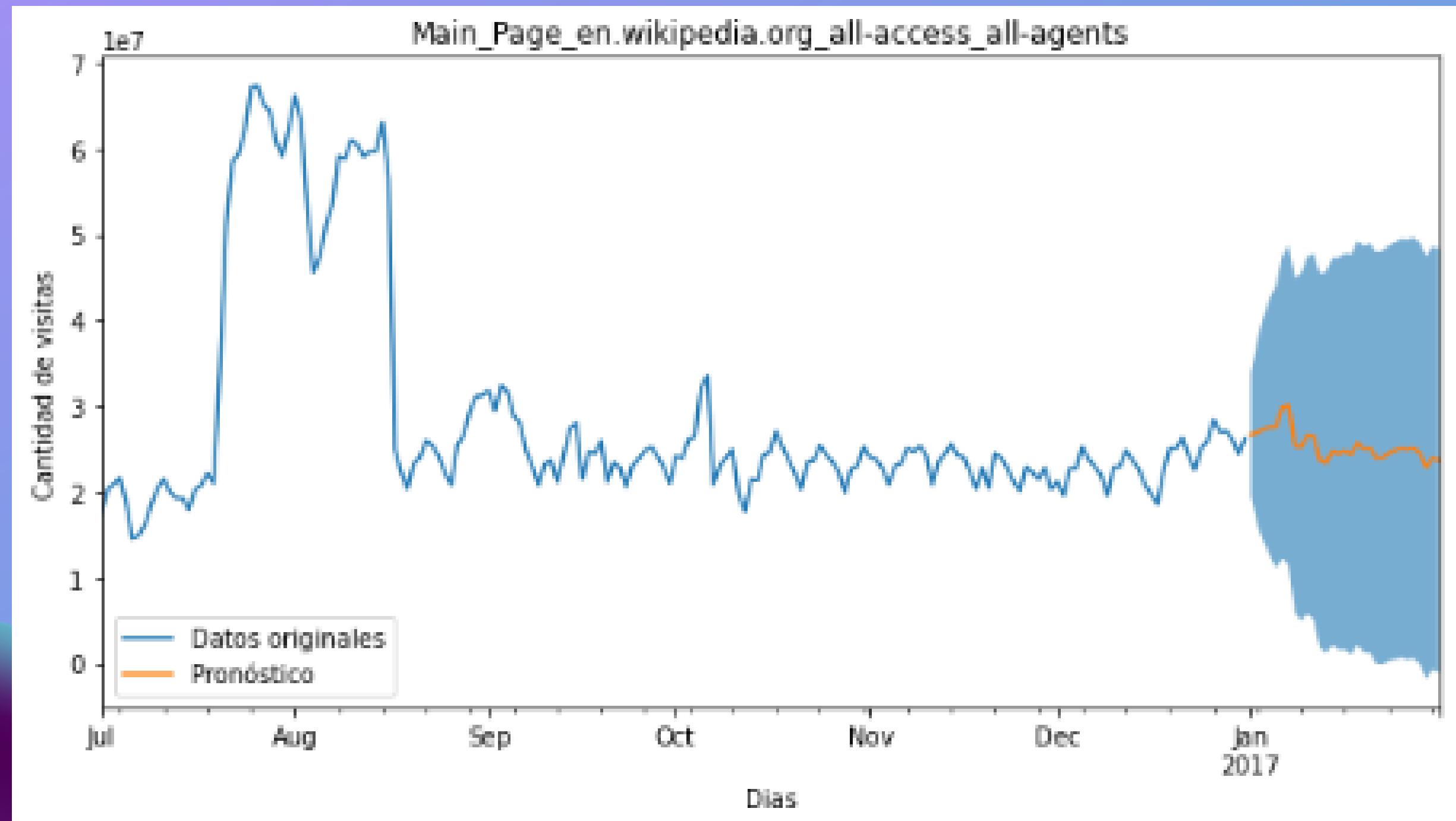
Implementación

| RESULTADOS |

03

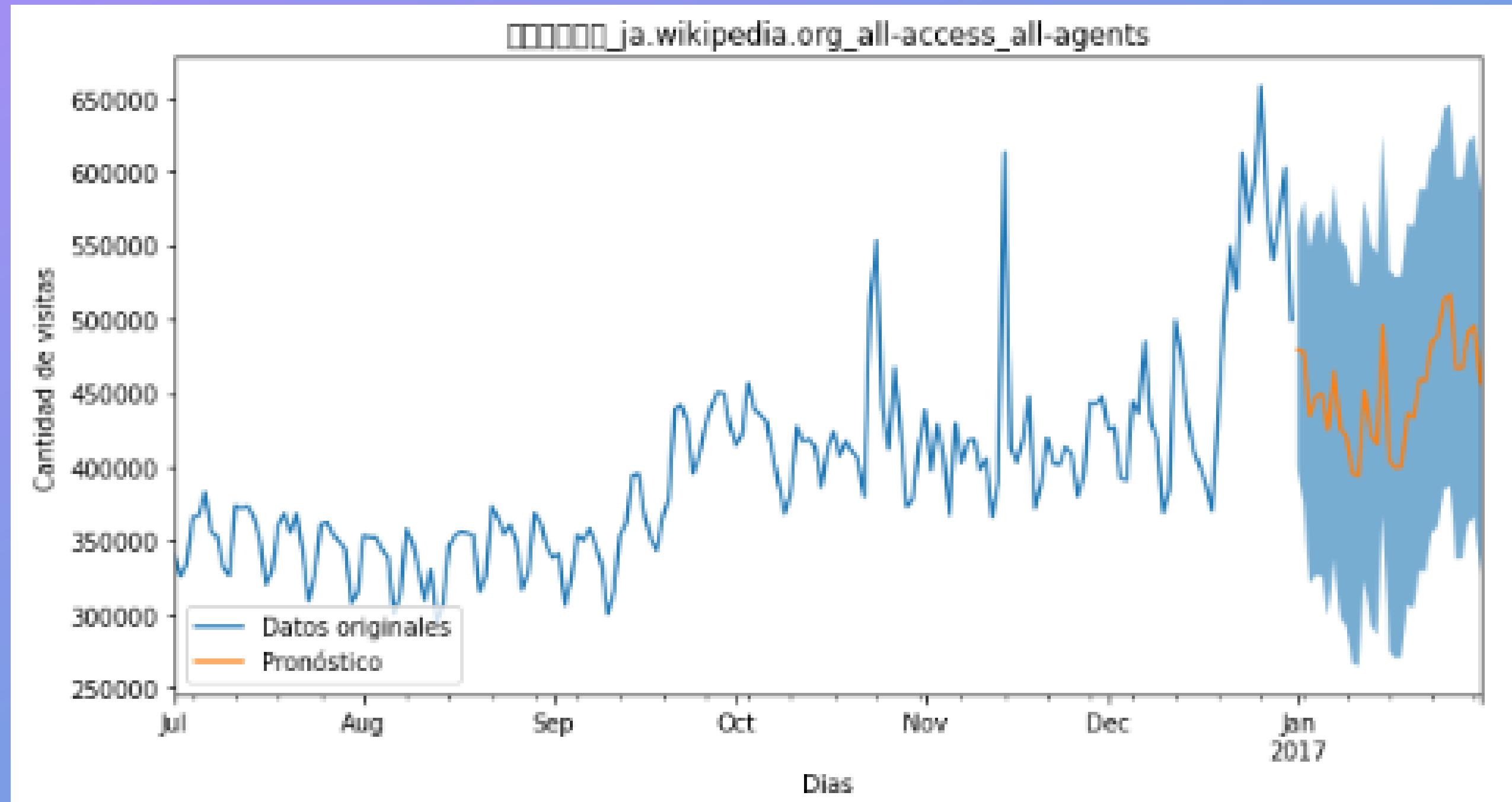
SARIMA

SARIMA

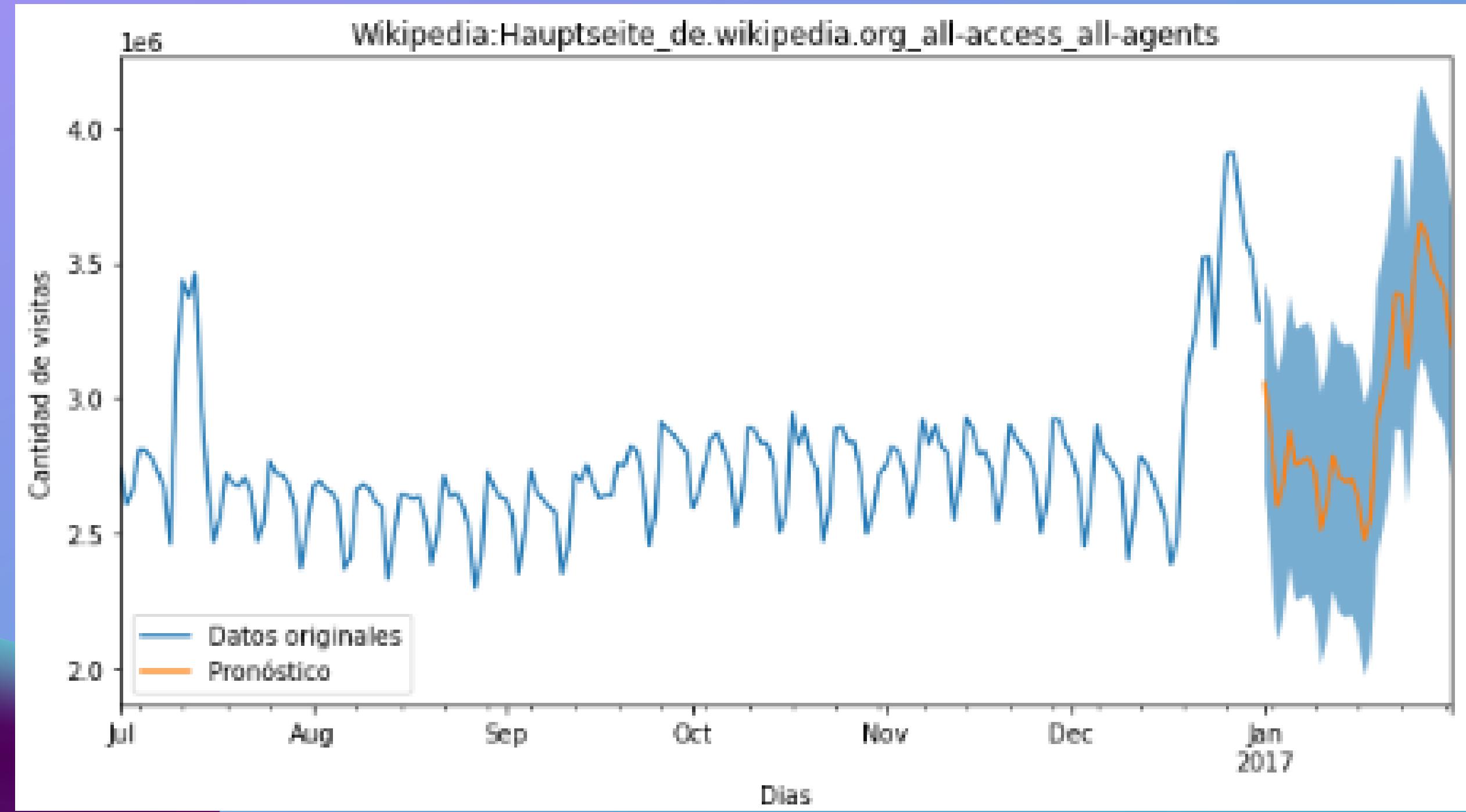


JAPONÉS

SARIMA

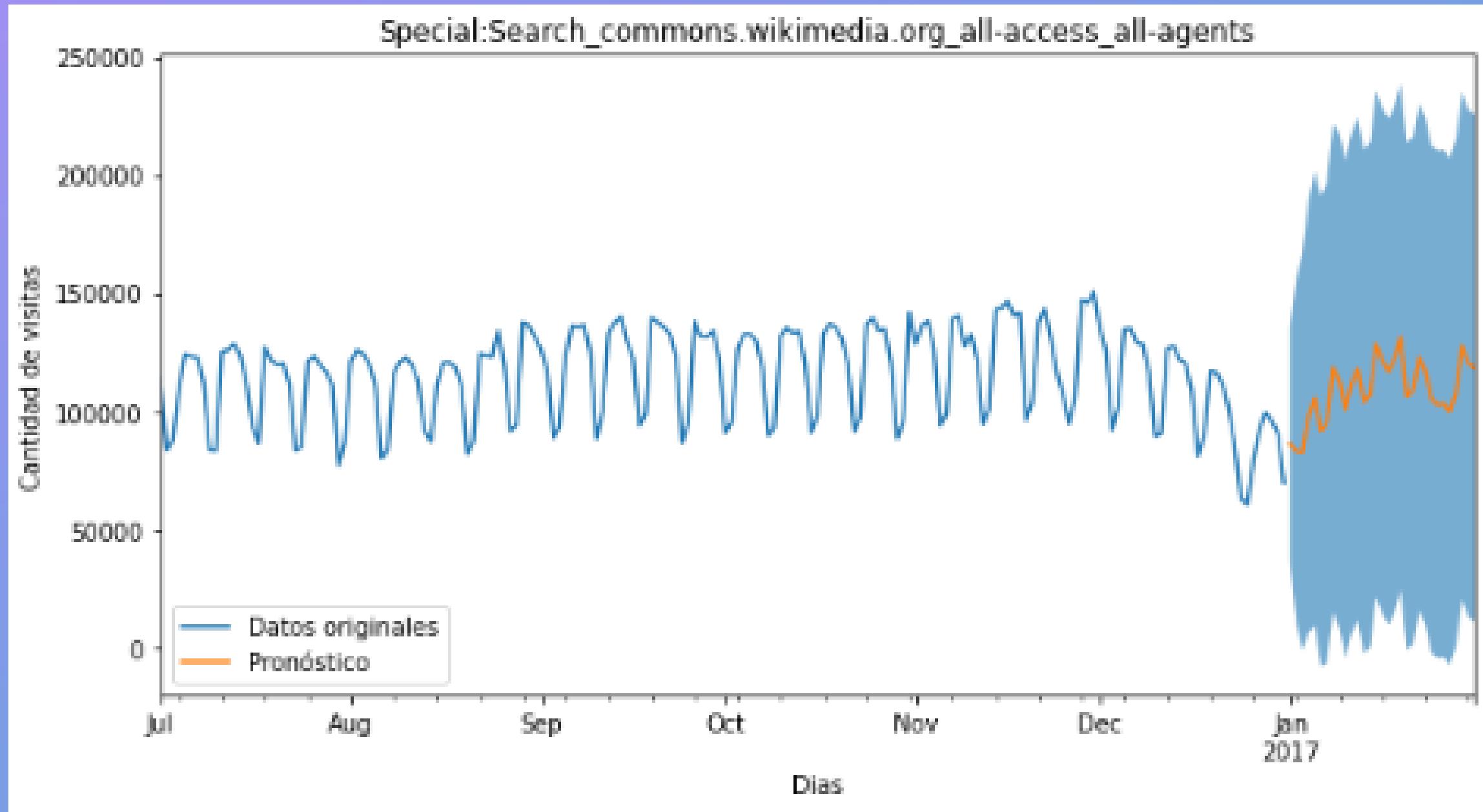


SARIMA



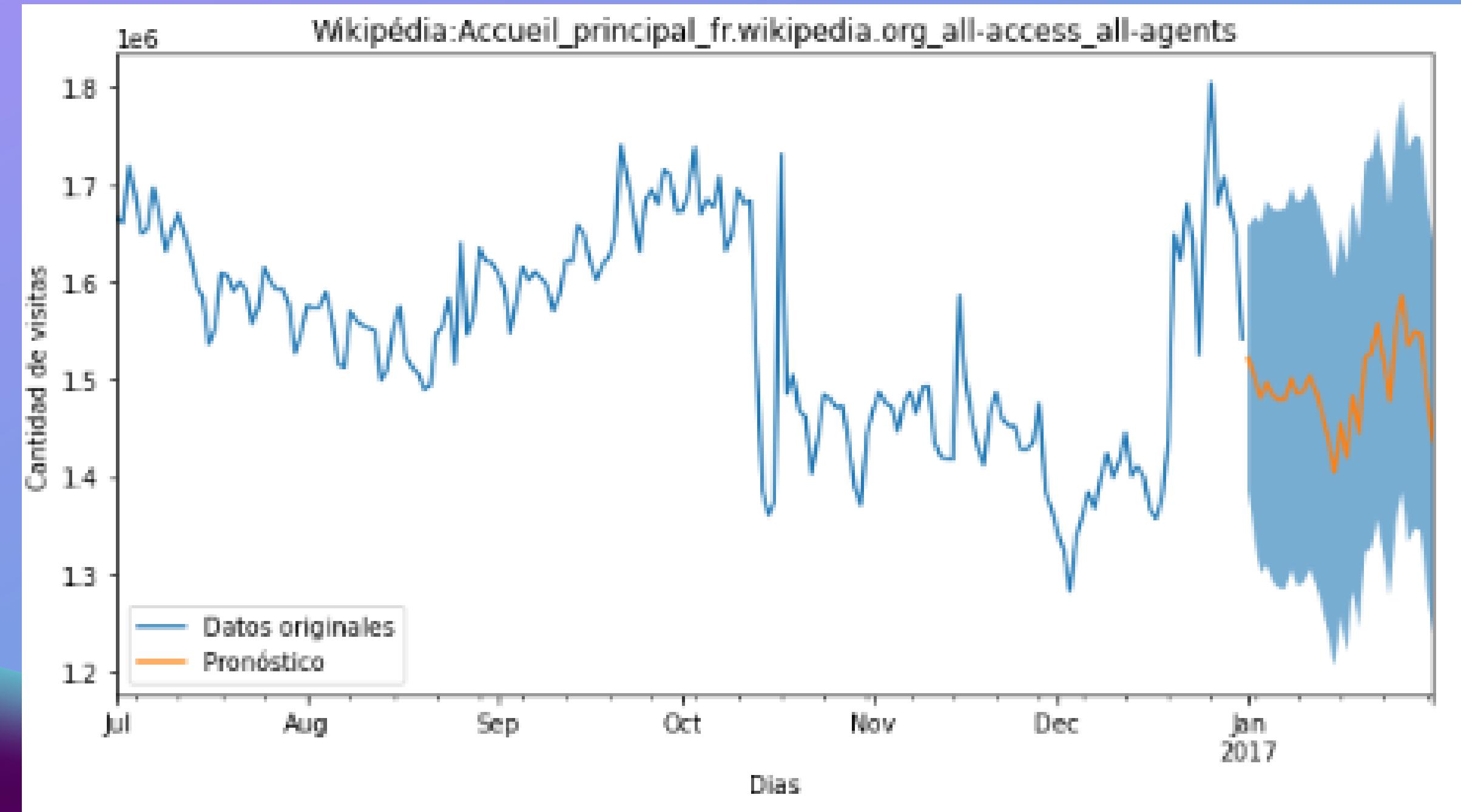
COMÚN

SARIMA



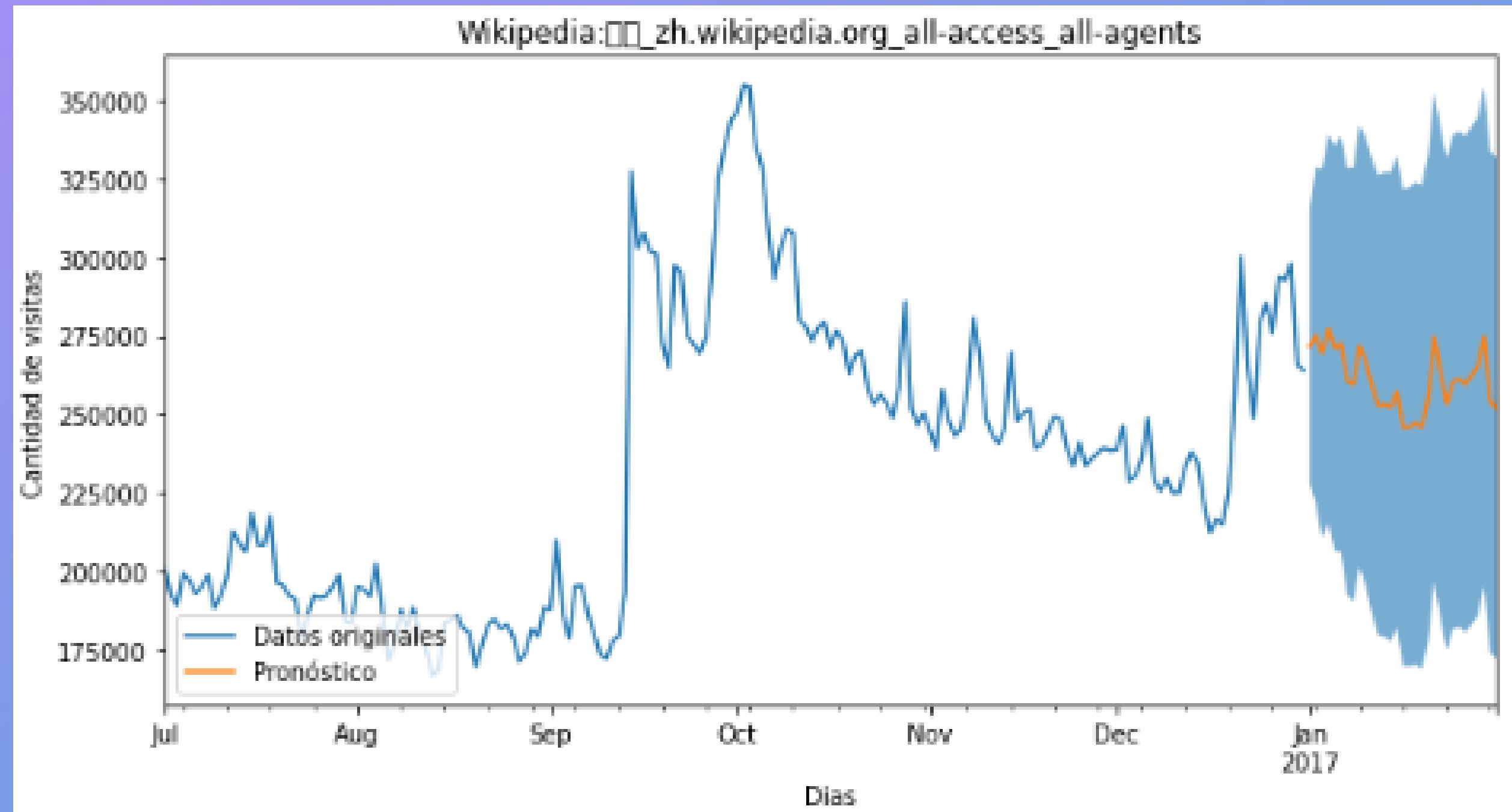
FRANCÉS

SARIMA



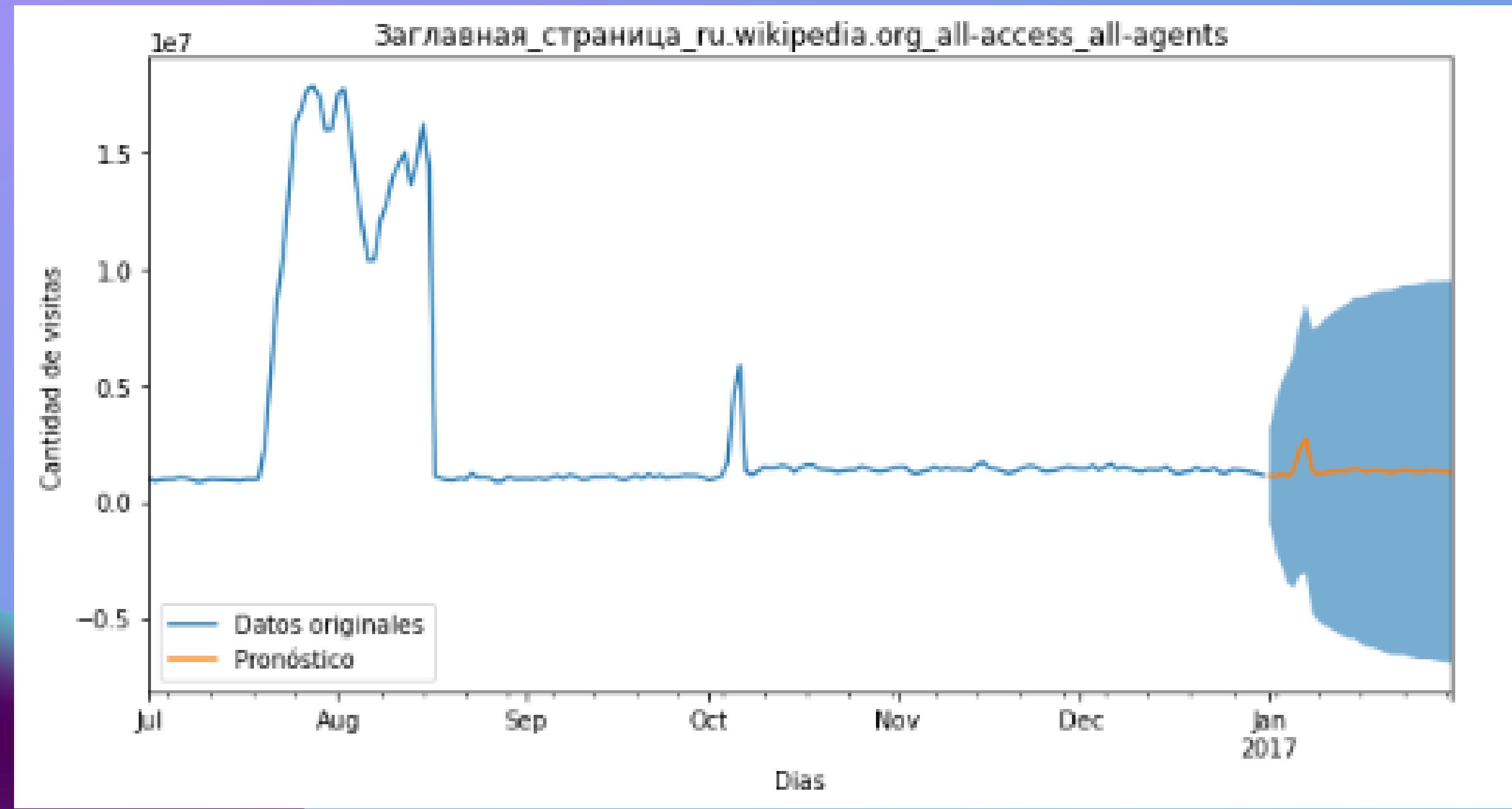
CHINO

SARIMA



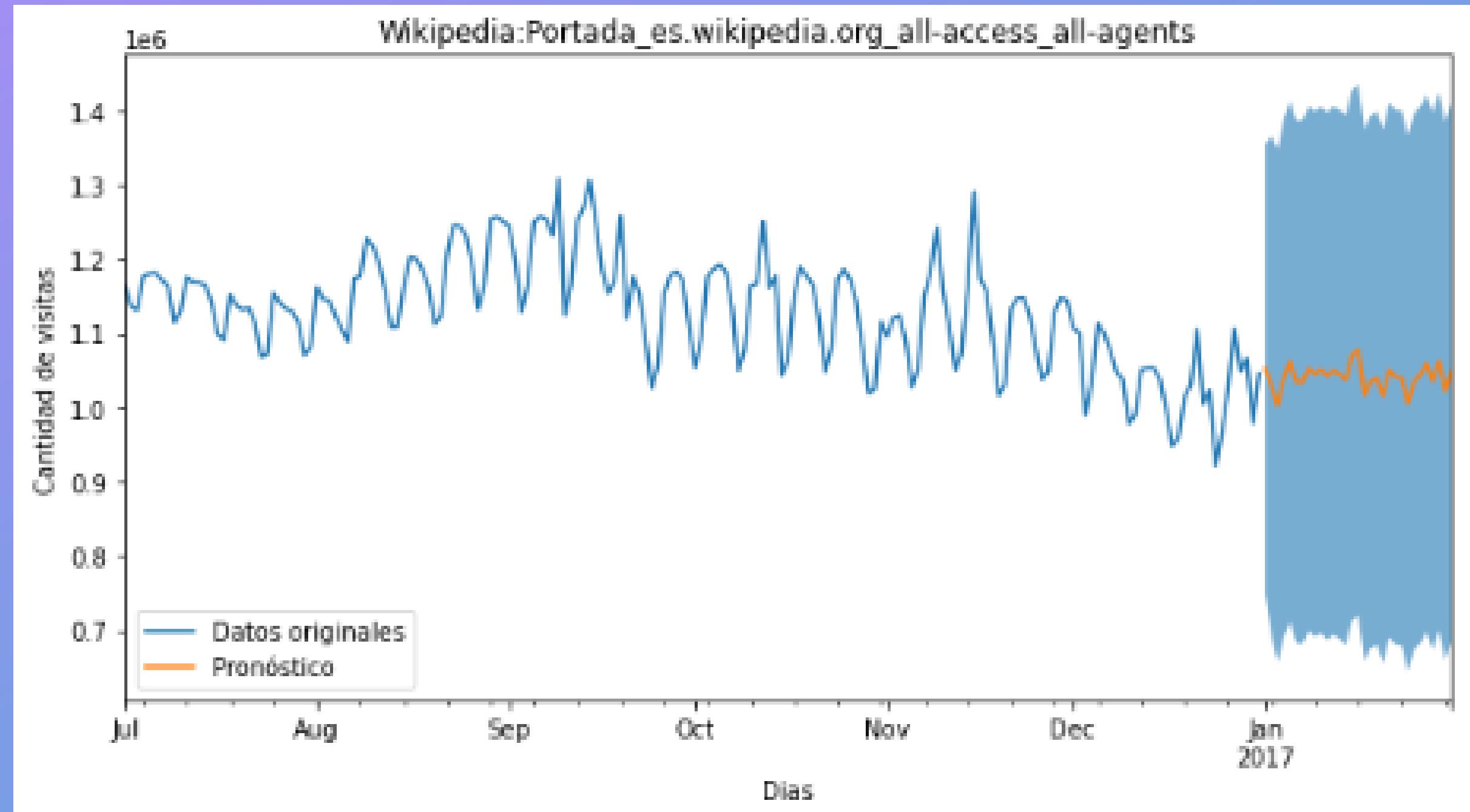
50

SARIMA



ESPAÑOL

SARIMA



53

04

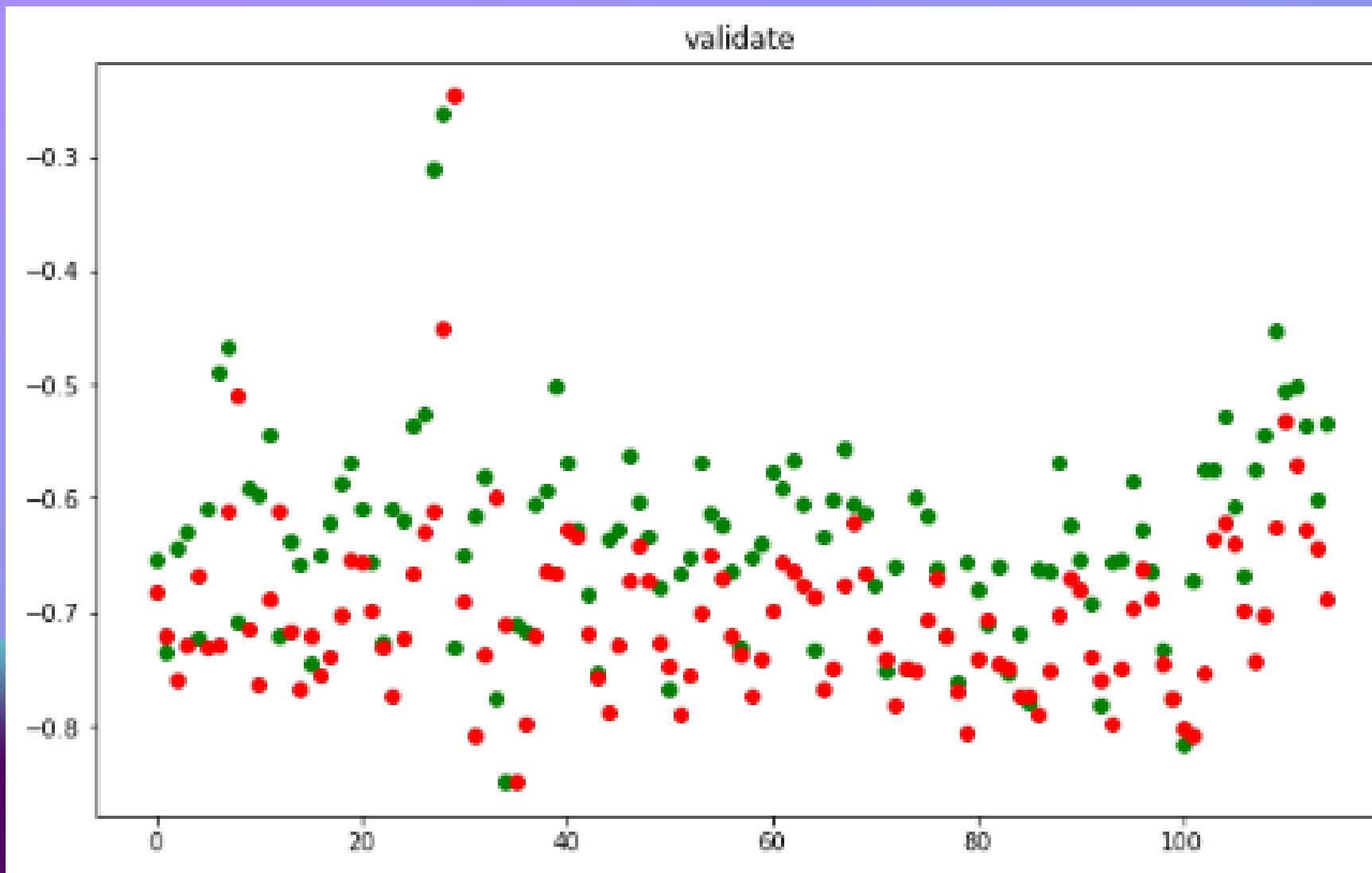
REDES NEURONALES

Implementación

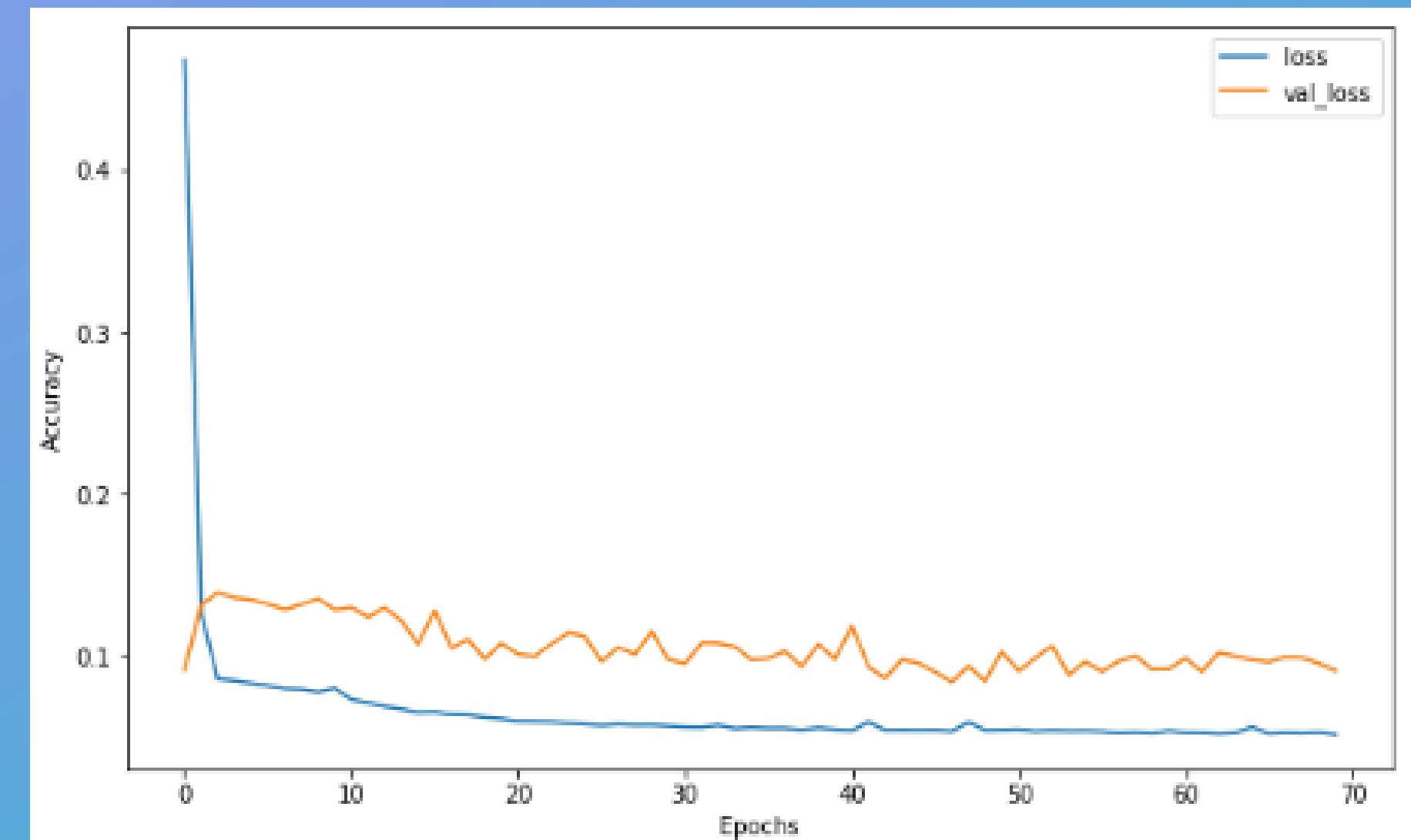
| RESULTADOS |

INGLÉS

RED NEURONAL

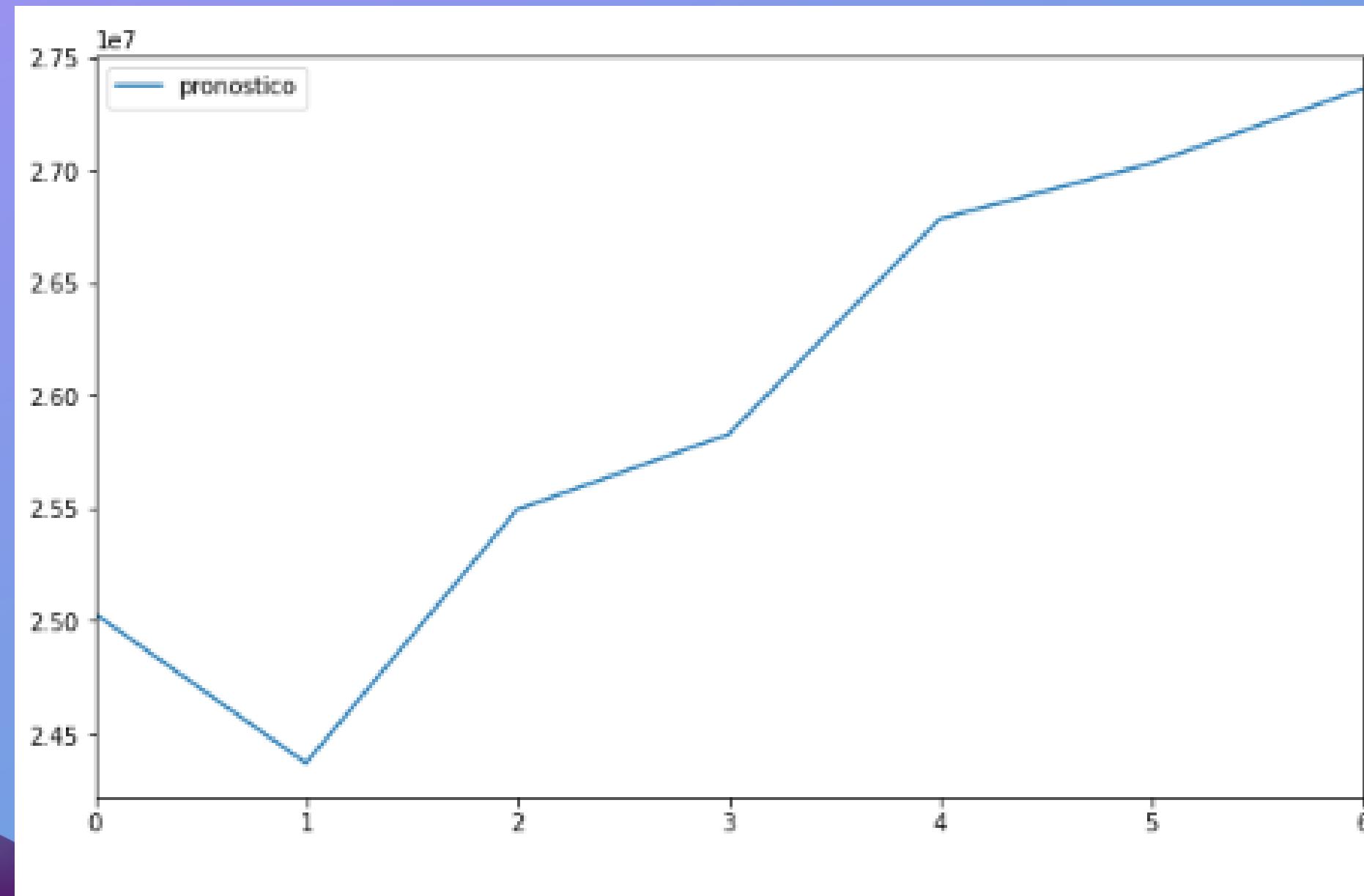


54



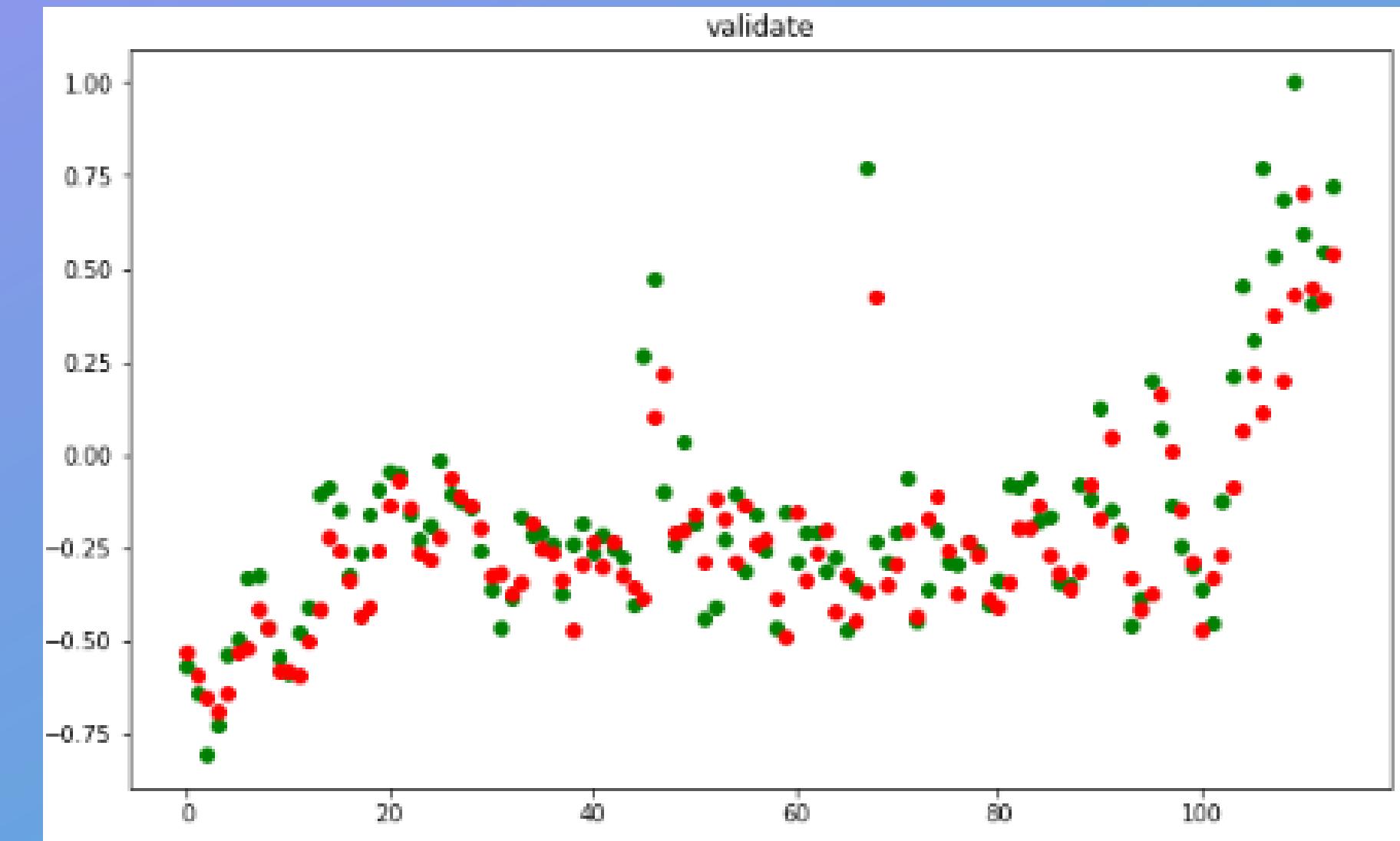
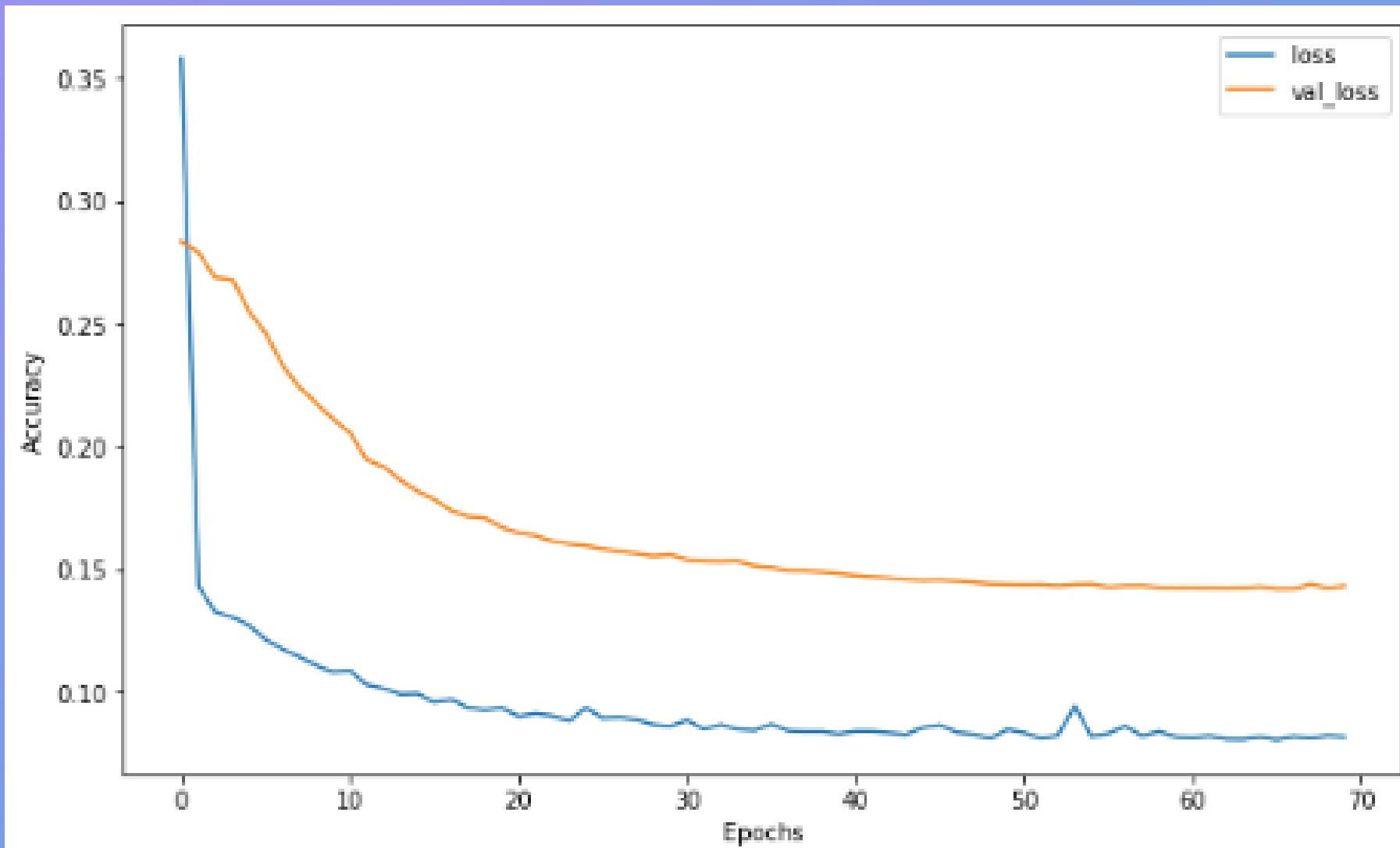
INGLÉS

RED NEURONAL



JAPONÉS

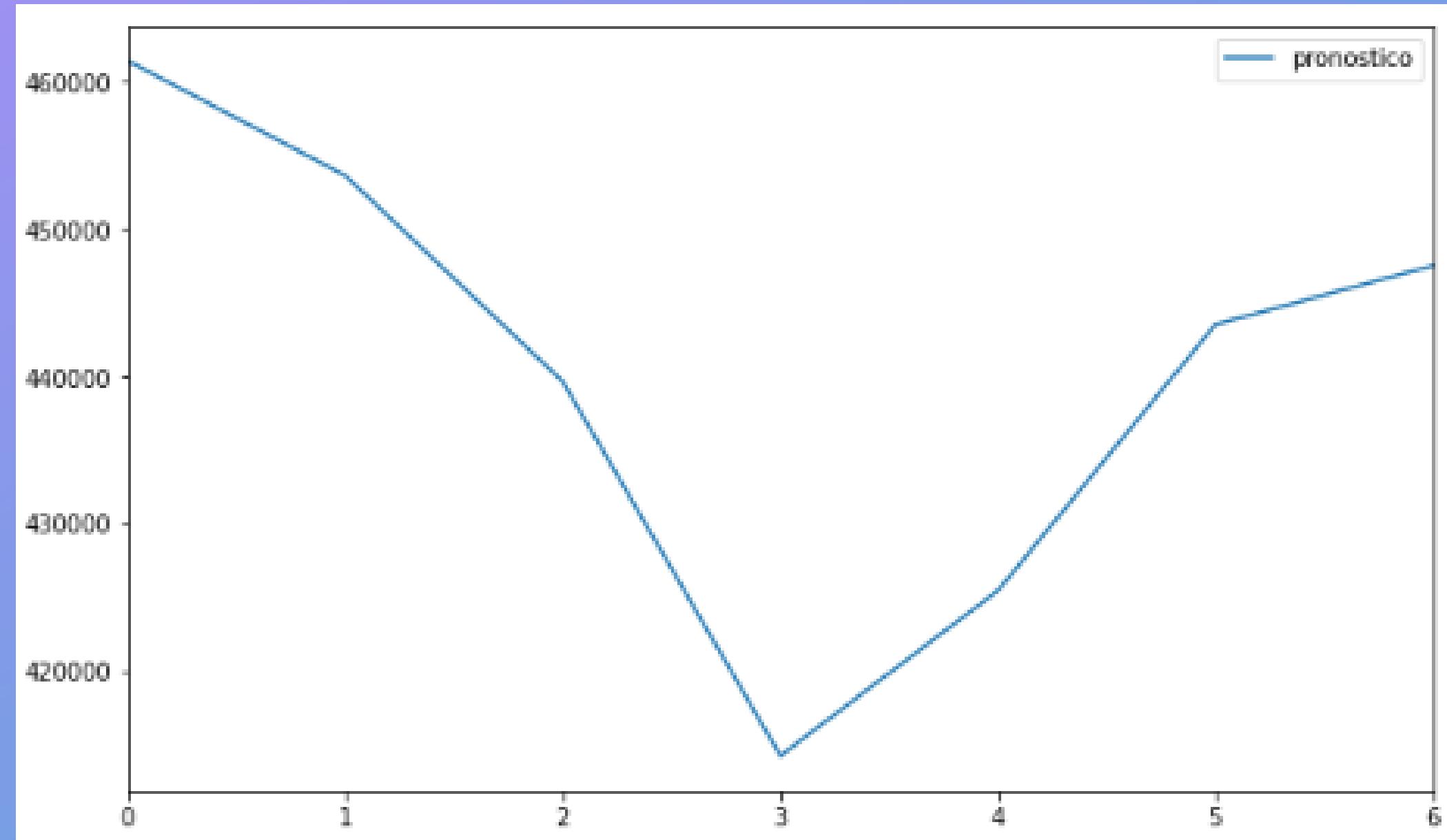
RED NEURONAL



56

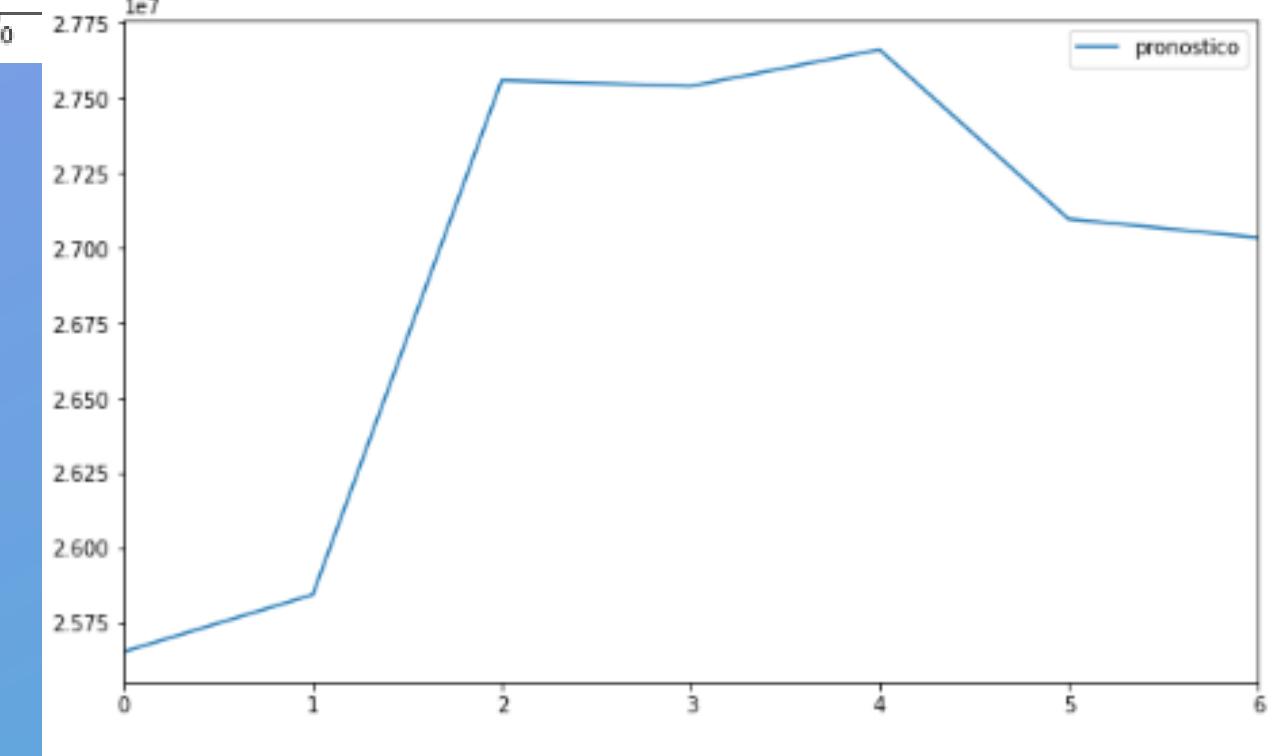
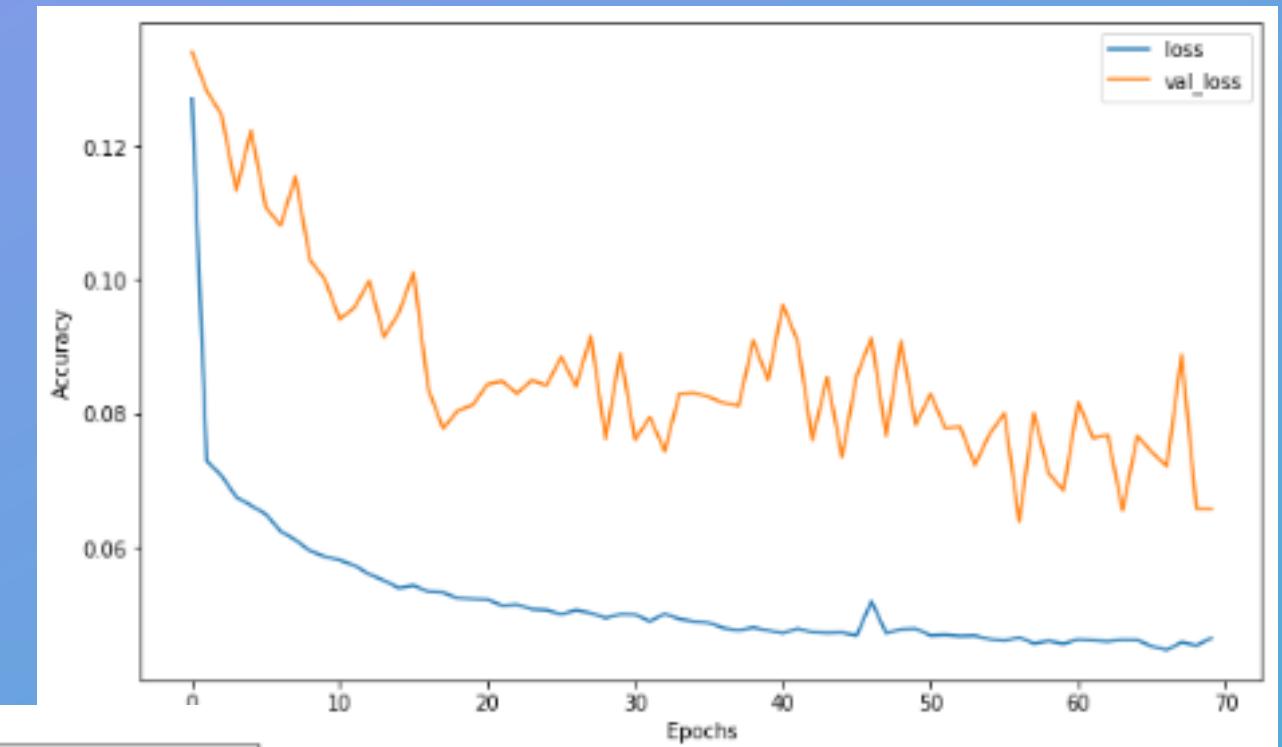
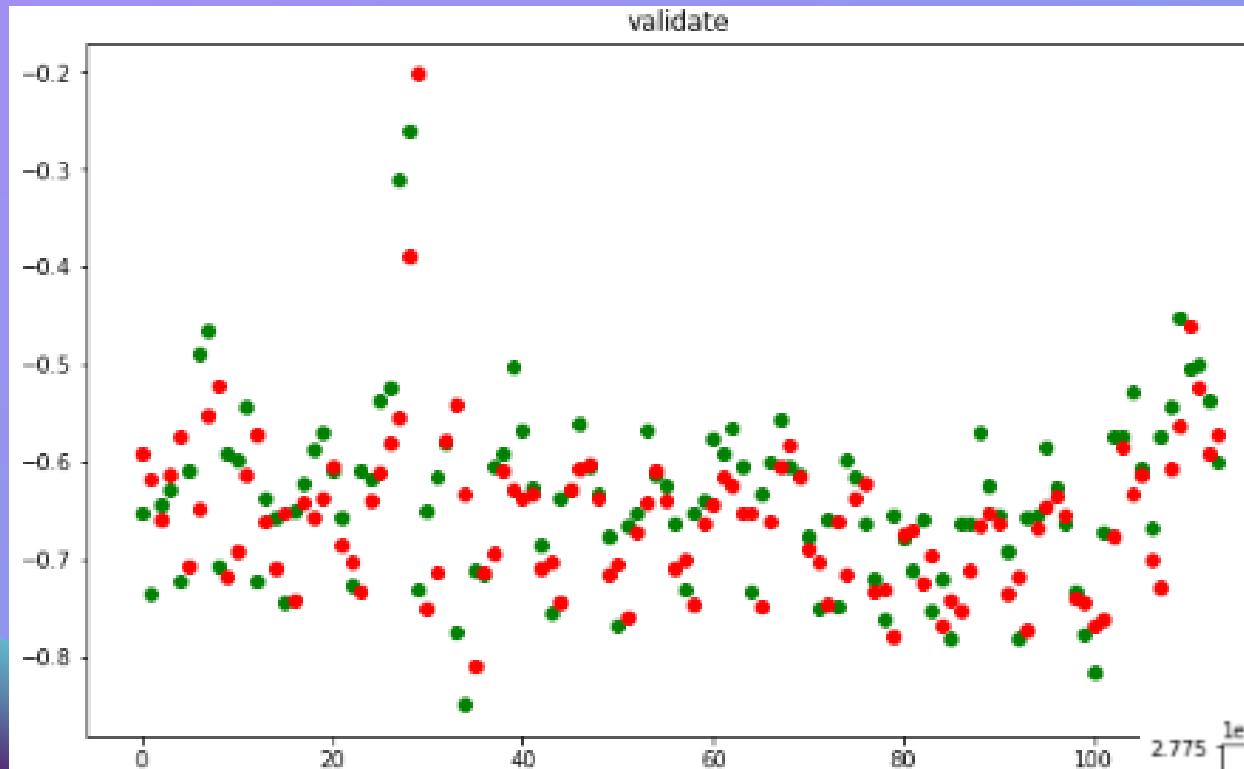
JAPONÉS

RED NEURONAL



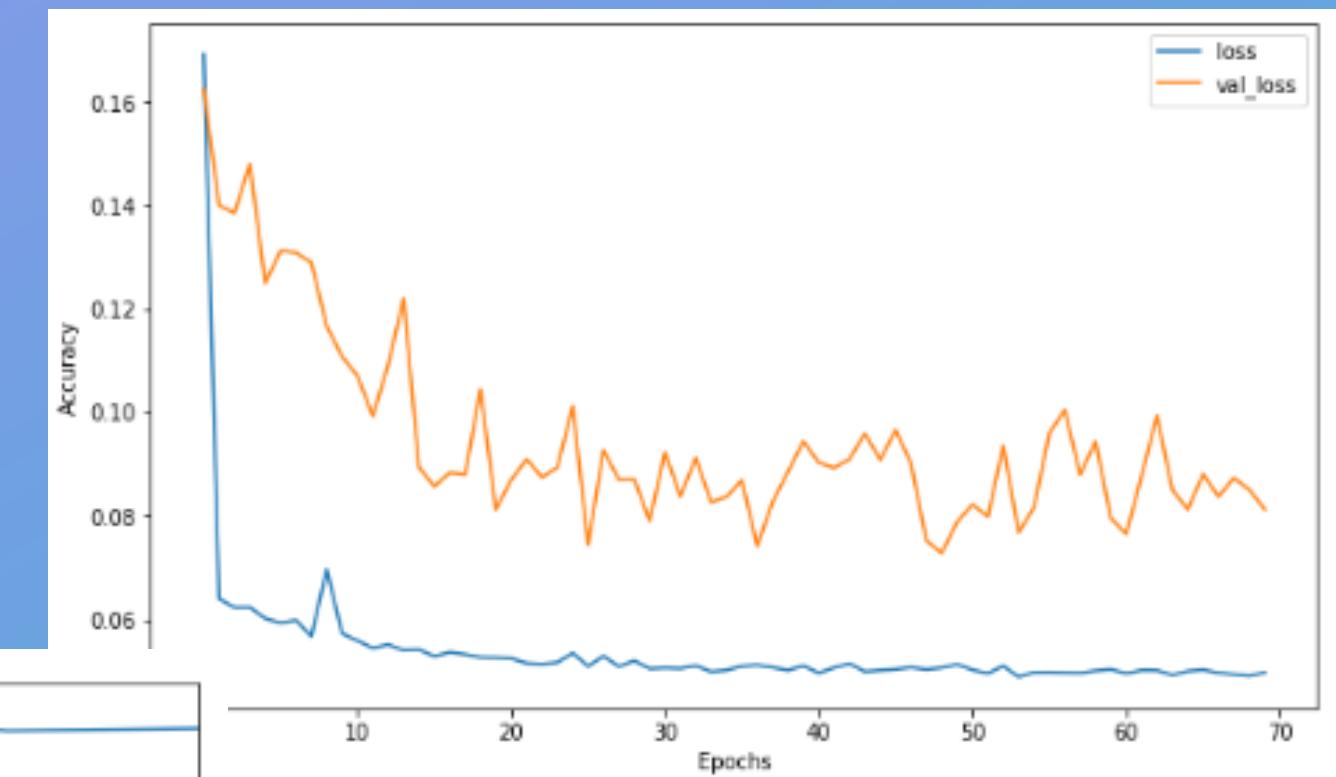
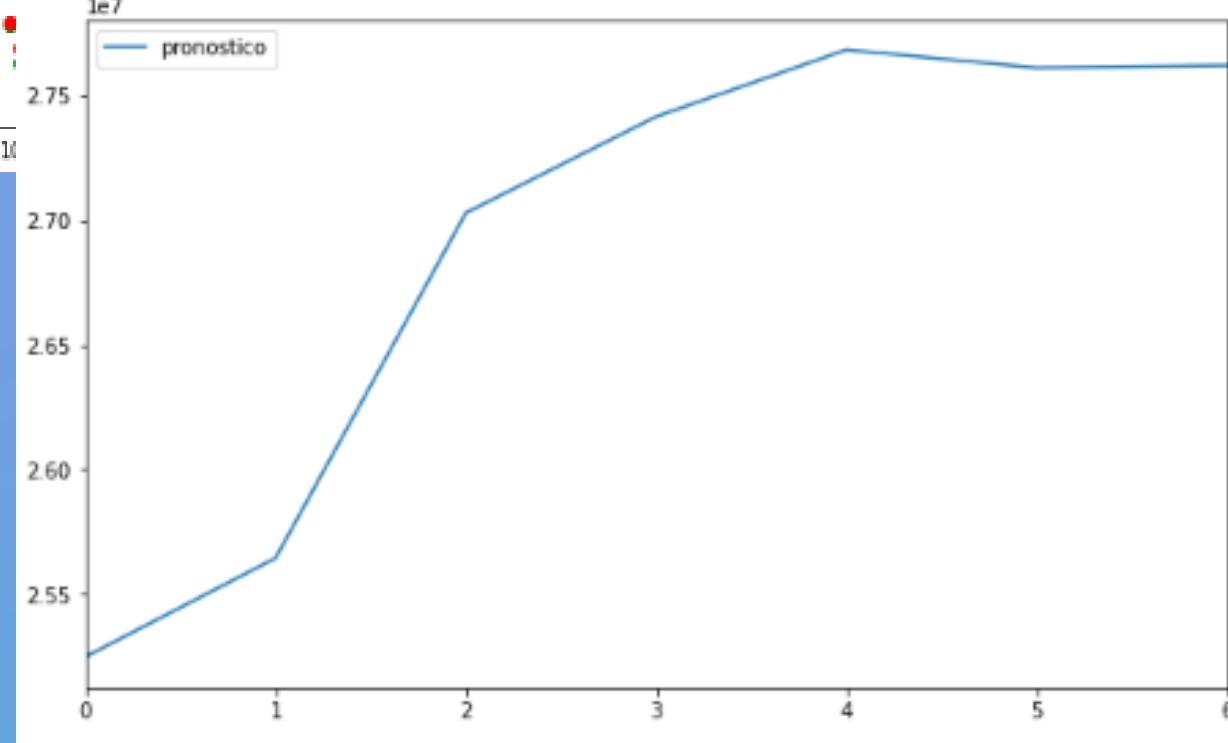
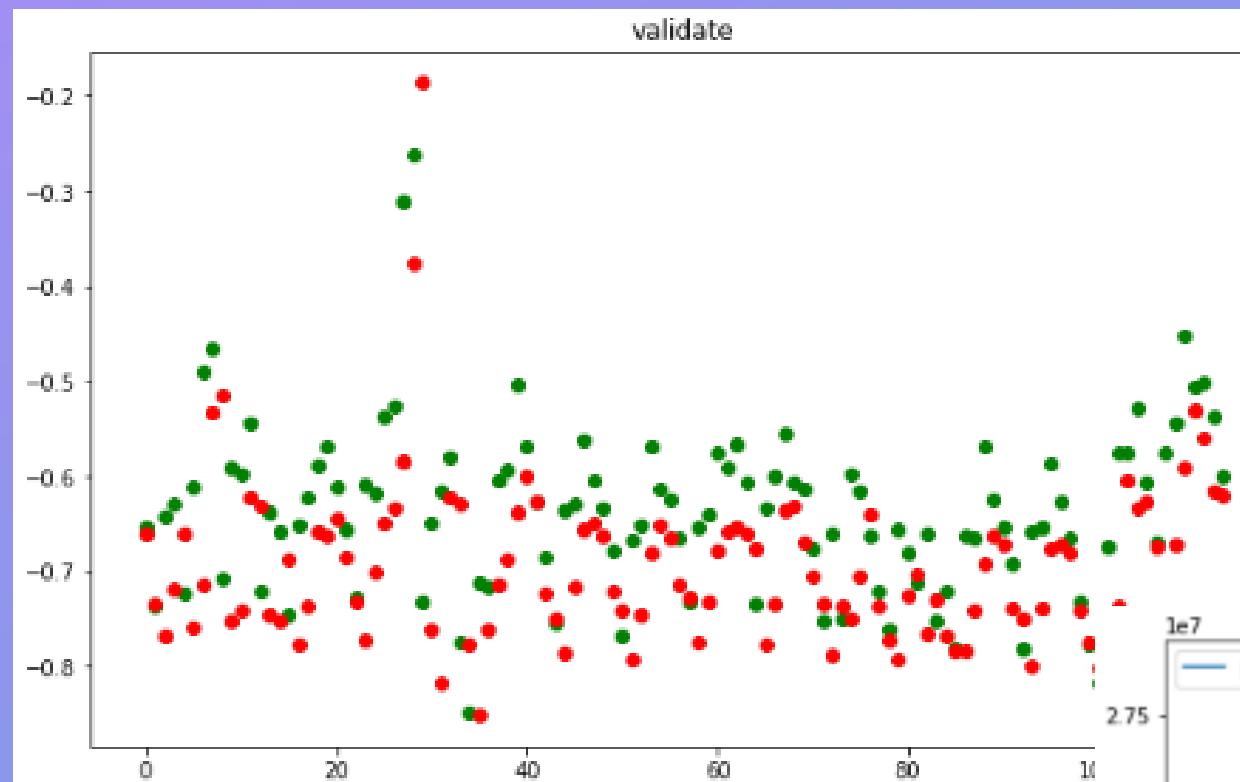
ALEMÁN

RED NEURONAL



COMÚN

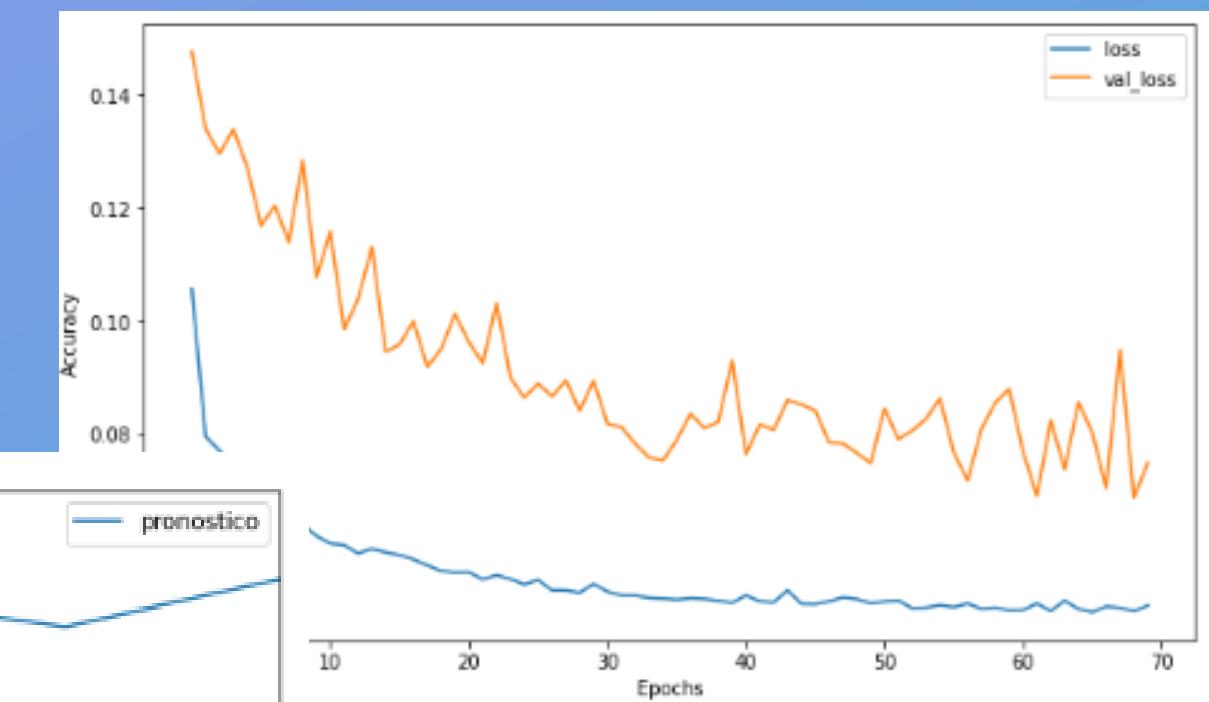
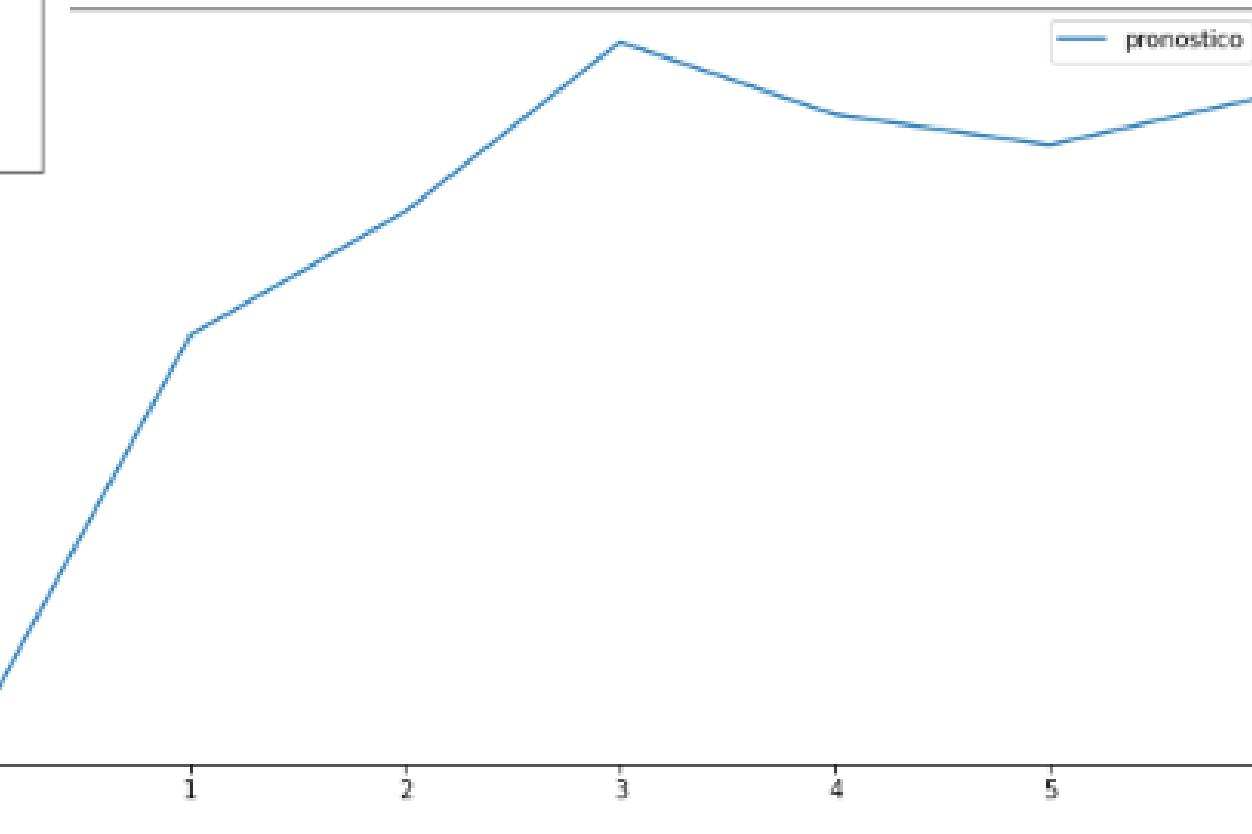
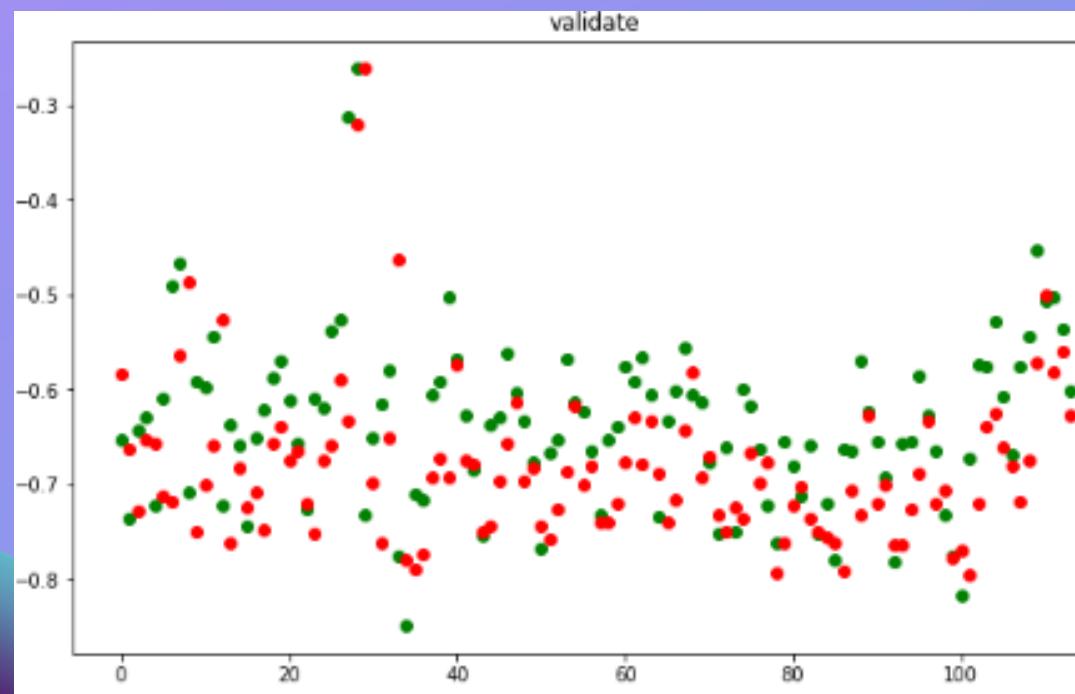
RED NEURONAL



59

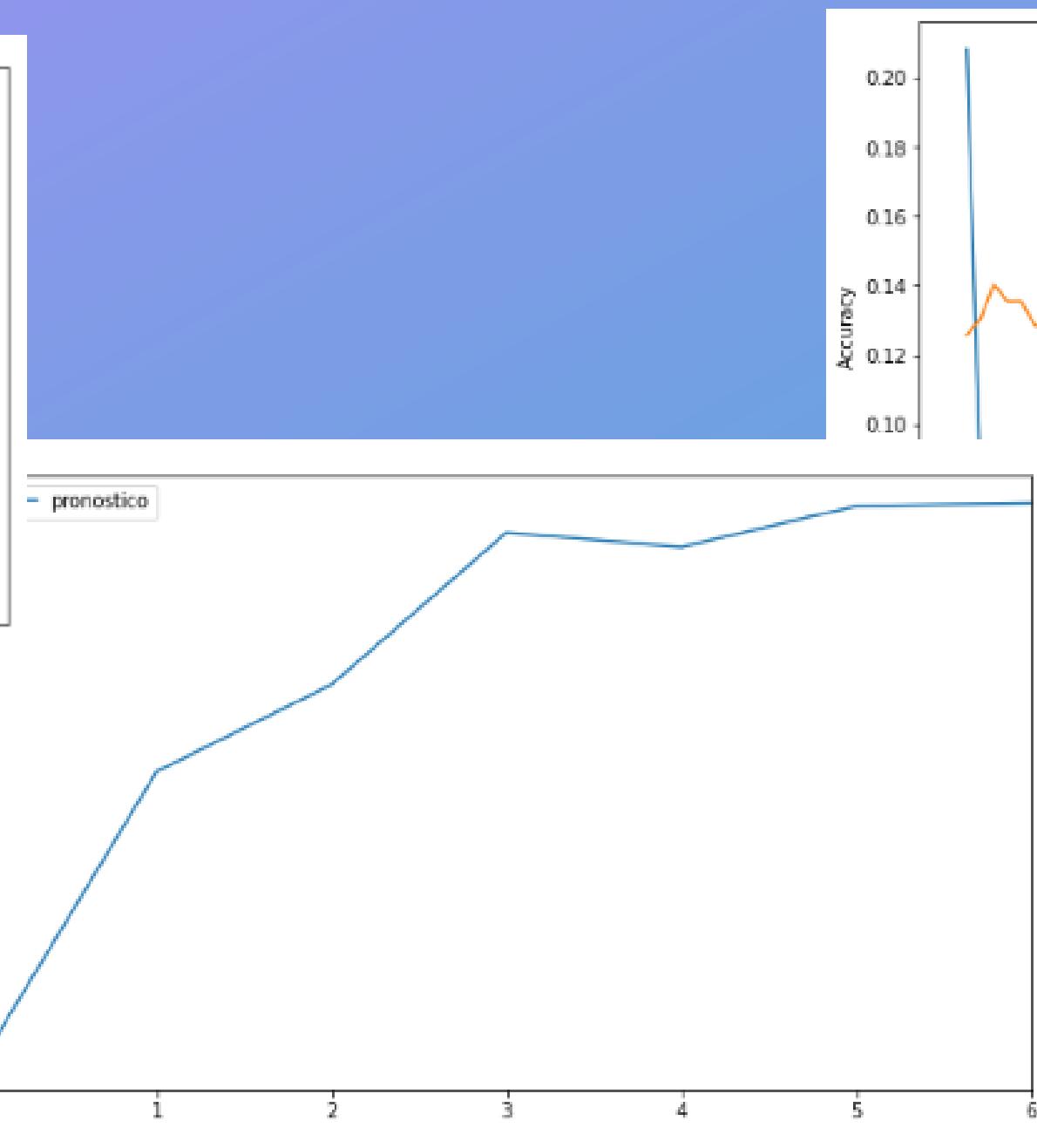
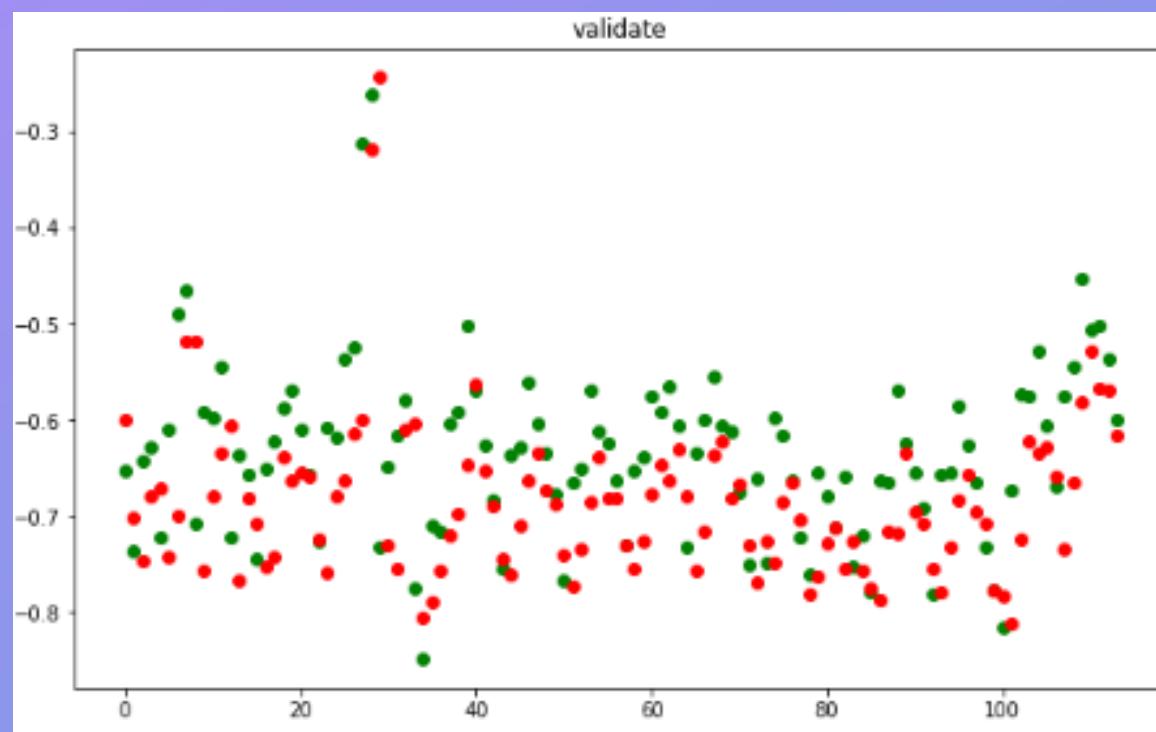
FRANCÉS

RED NEURONAL



CHINO

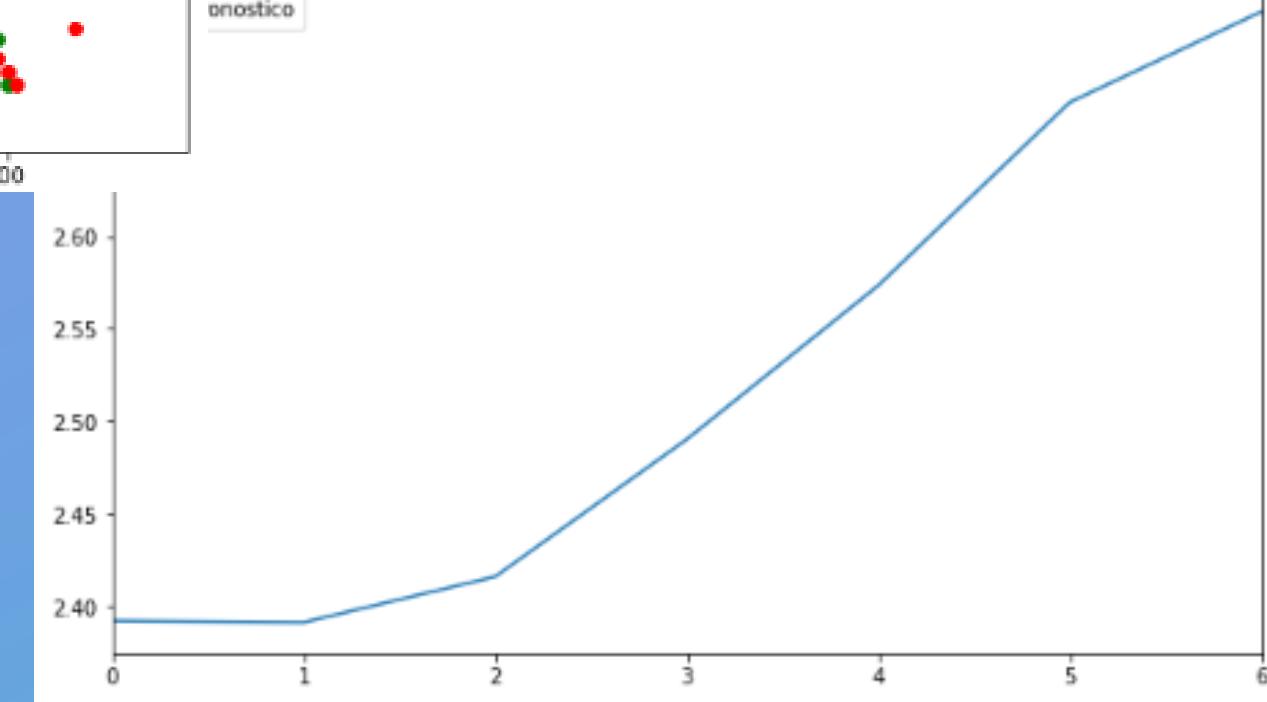
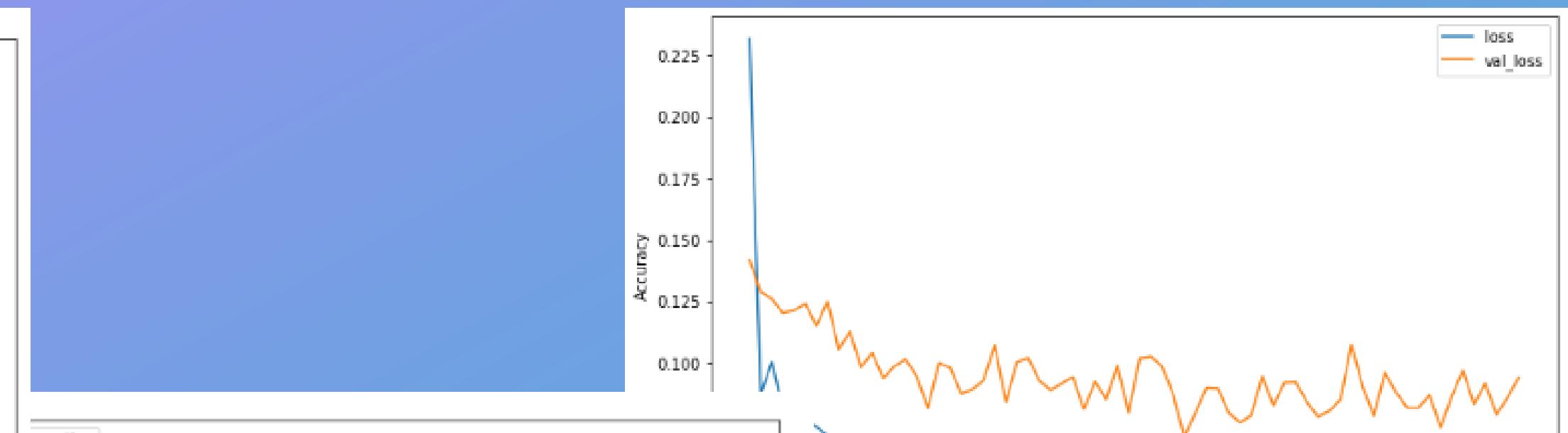
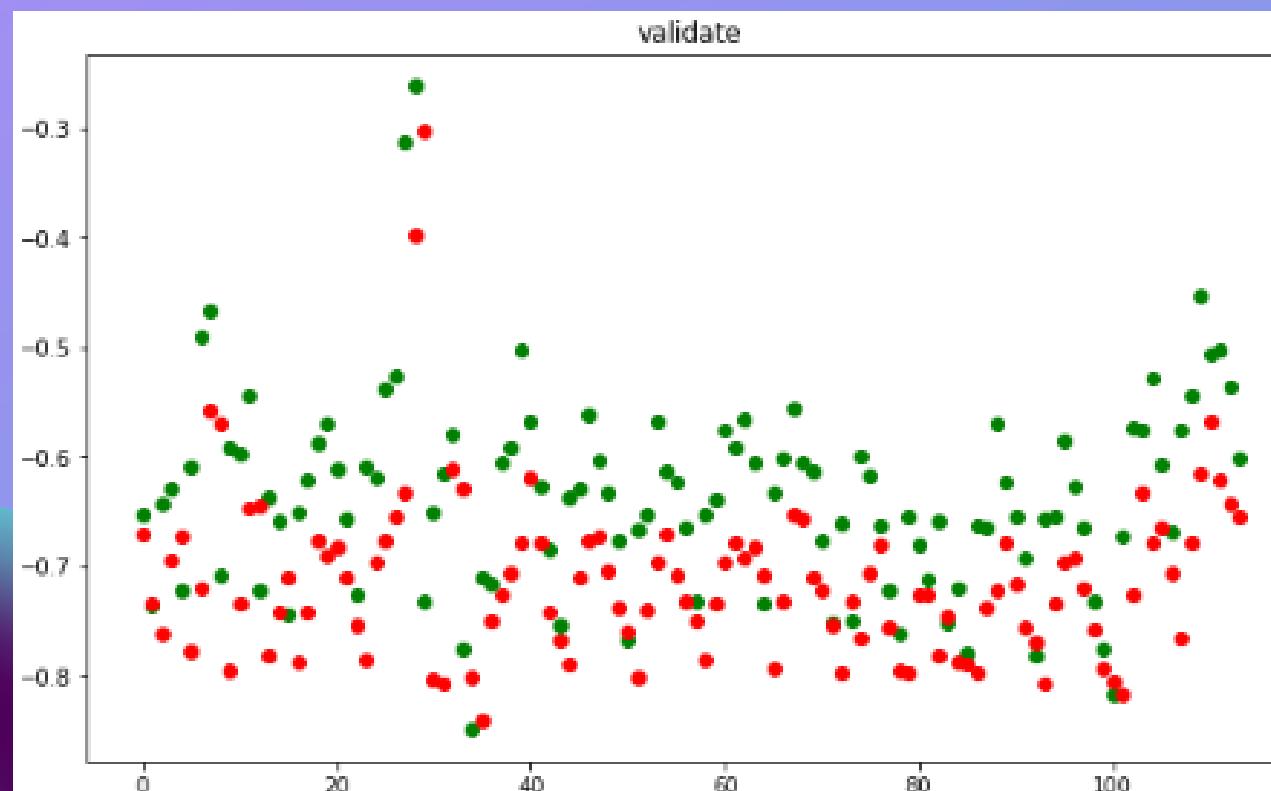
RED NEURONAL



61

RUSO

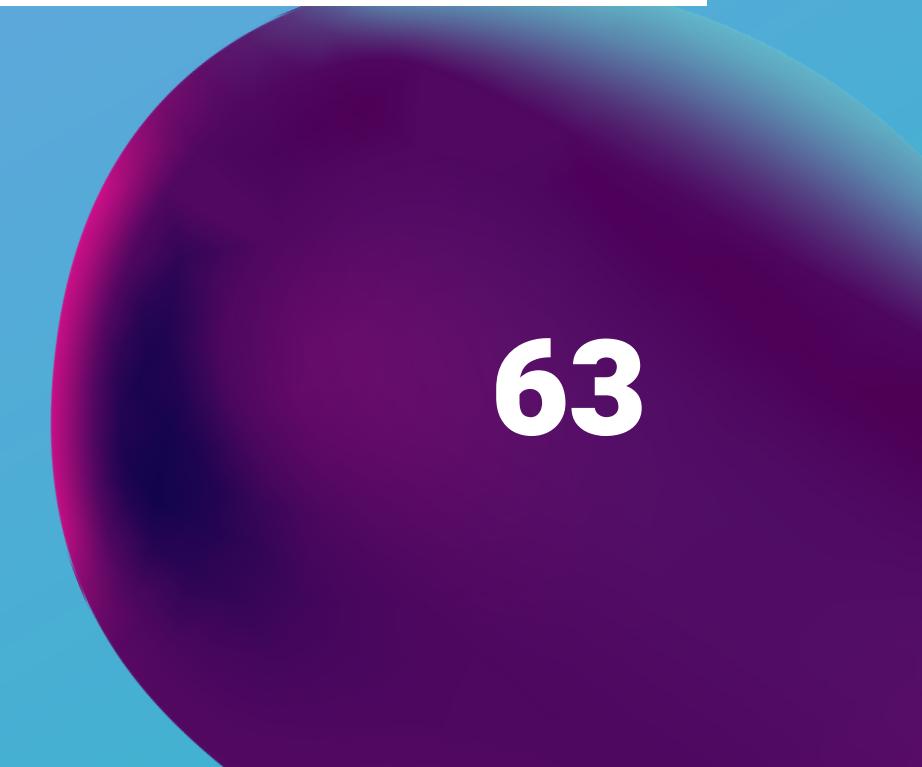
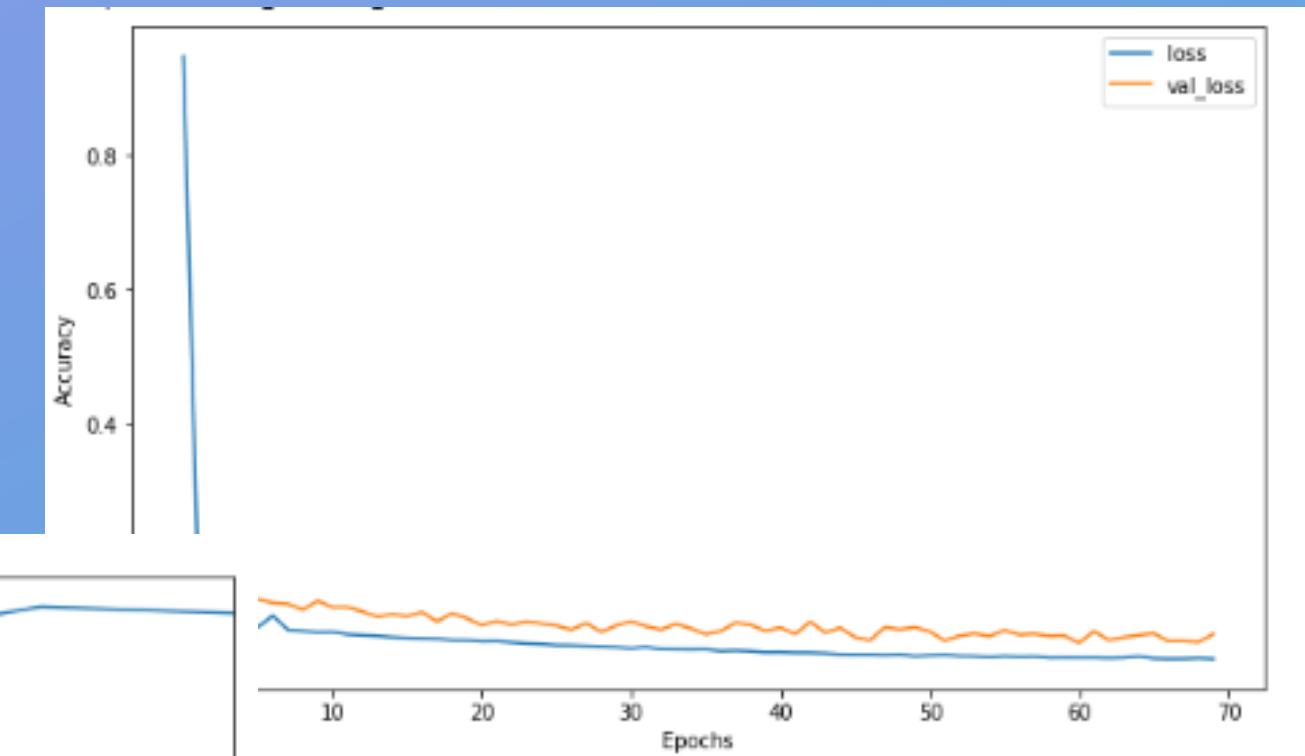
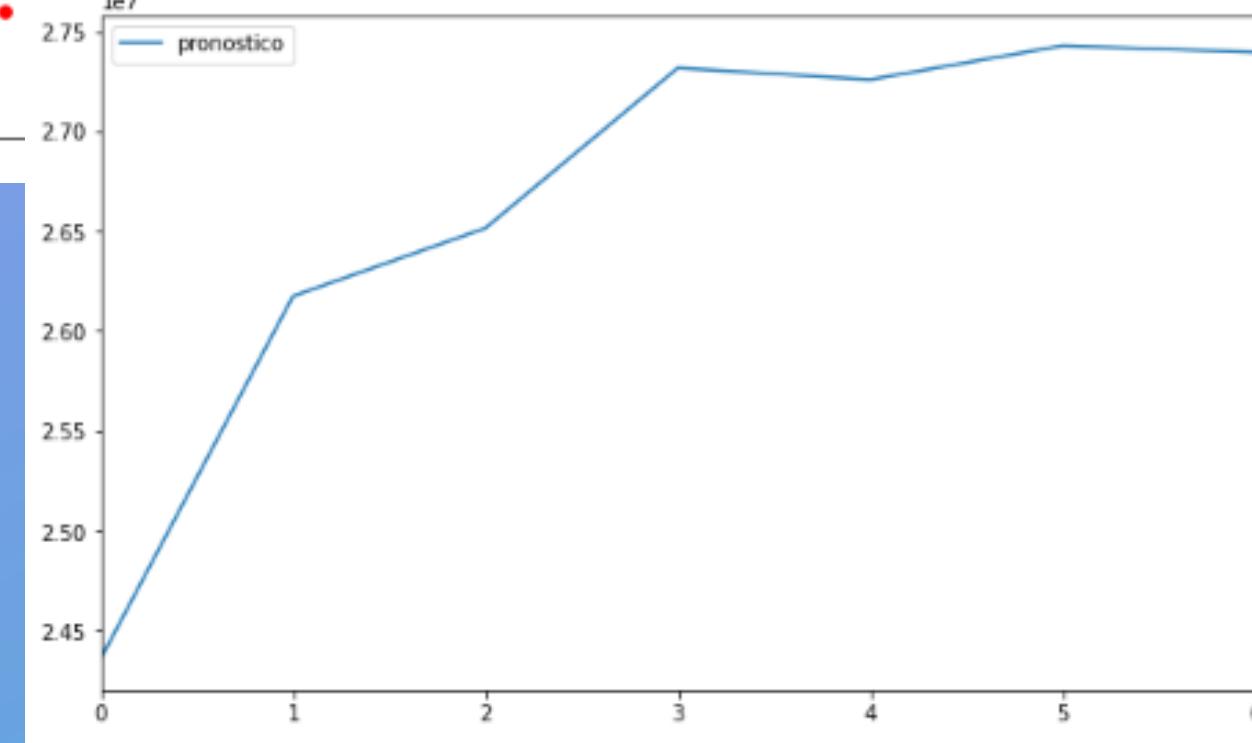
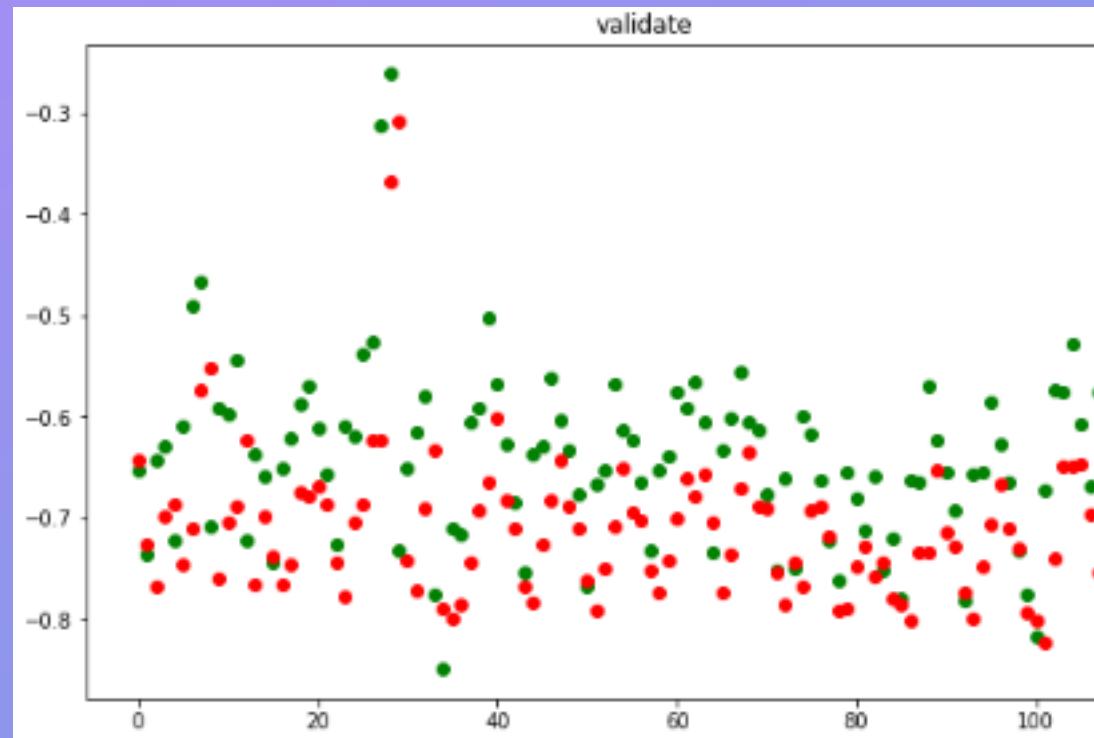
RED NEURONAL



62

ESPAÑOL

RED NEURONAL



Conclusión

Observamos que tratando la serie como un modelo de aprendizaje supervisado también se puede lograr una predicción muy buena con el uso de redes neuronales.

Al utilizar el modelo ARIMA nos dimos cuenta que al realizar la predicción se asemeja a los valores reales de la serie de tiempo, por lo que podemos concluir que ARIMA es un modelo ideal para aplicar a nuestro proyecto

Mediante el modelo de SARIMA se puede evidenciar que se obtiene una predicción de los datos más cercana a la realidad, pues este modelo tienen en cuenta la estacionalidad al momento de hacer sus predicciones de los datos.

Se puede concluir que la estacionaridad es un factor de gran importancia, ya que es uno de los requisitos primordiales para poder aplicar los métodos de predicción, porque garantiza que la media y la varianza sean constantes. Por eso al aplicar el test de estacionaridad podemos ver grafica y numericamente estos parametros. Ademas, si la serie no llega a ser estacionaria por medio del metodo shift garantizamos que todas lo fueran.

Si el Dataset contará con series de tiempo que tengan estacionalidad, se podría hacer una mejor predicción con respecto al comportamiento que tendrían las páginas con respecto a las visitas.

Información

GitHub

<https://github.com/mcamilaanova/Proyecto-Simulaci-n-Digital-.git>

Collaboratory Notebook

https://colab.research.google.com/drive/11YHrQ6_EqseAhOhU0oNNTka1CLXS1Yhr

Gracias

¿Alguna pregunta?