

Modelos RFR para turista internacional.

Recopilación de datos:

[INE. Instituto Nacional de Estadística](#)

[World Bank Group - International Development, Poverty, & Sustainability](#)

[INICIO - SEGITTUR | Segittur.es](#)

Modelo de predicción que analiza series temporales, entrenando los modelos de ARIMA y ARIMAX, siendo estos no representativos para los valores dados como resultado.

El modelo que ha dado mejores métricas con un error más representativo ha sido un random forest regressor.

Introducción al modelo:

Un modelo de Random Forest Regressor es un meta-estimador que ajusta una serie de árboles de decisión en varias submuestras del conjunto de datos y utiliza el promedio para mejorar la precisión predictiva y controlar el sobreajuste.

Datos de entrenamiento:

El modelo puede ser entrenado con datos históricos de turismo en Madrid, como el número de llegadas de turistas, el país de origen, la duración de la estancia, el gasto por turista, entre otros.

Predicción por áreas:

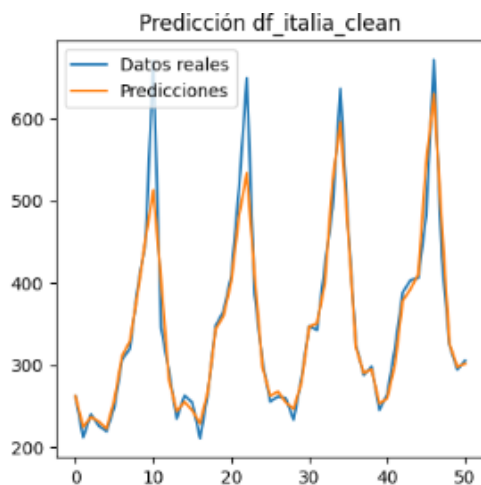
Para predecir la afluencia de turistas por áreas, se pueden utilizar datos específicos de cada área, como el número de atracciones turísticas, la capacidad hotelera, la oferta gastronómica, entre otros.

Una vez entrenado, el modelo puede predecir la afluencia futura de turistas a Madrid por áreas basándose en patrones y tendencias observados en los datos históricos.

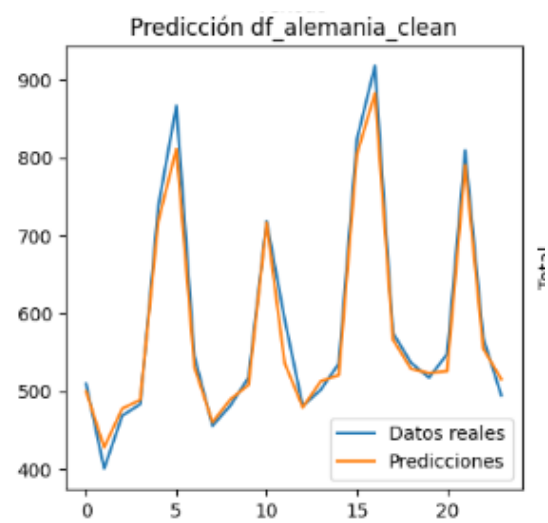
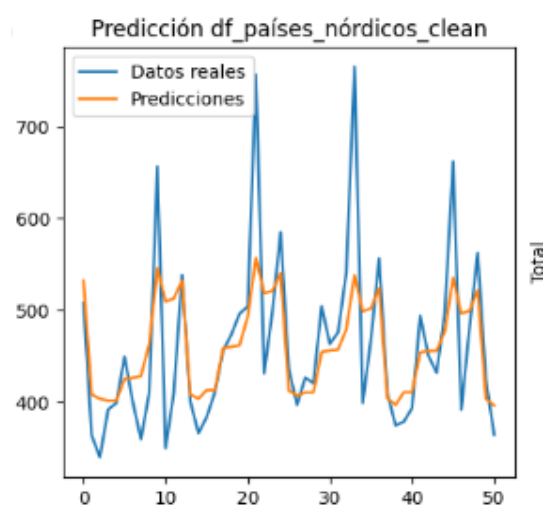
Países con mayor volumen de afluencia.

- Países Nórdicos: max\_depth=10, min\_samples\_split=10, n\_estimators=300, random\_state=0
- Italia: max\_depth=10, n\_estimators=200, random\_state=0
- Alemania: max\_depth=10, n\_estimators=400, random\_state=0
- Francia: max\_depth=10, n\_estimators=300, random\_state=0
- Reino Unido: max\_depth=10, min\_samples\_split=5, n\_estimators=400, random\_state

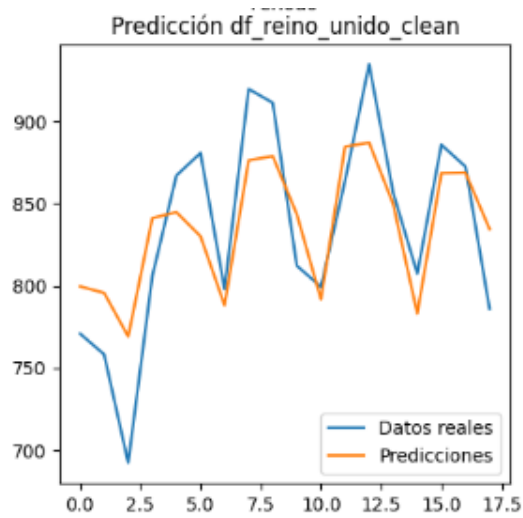
Visualizaciones del rendimiento del modelo. Precisión del rendimiento visualizando comparativa entre predicción y datos reales de test.

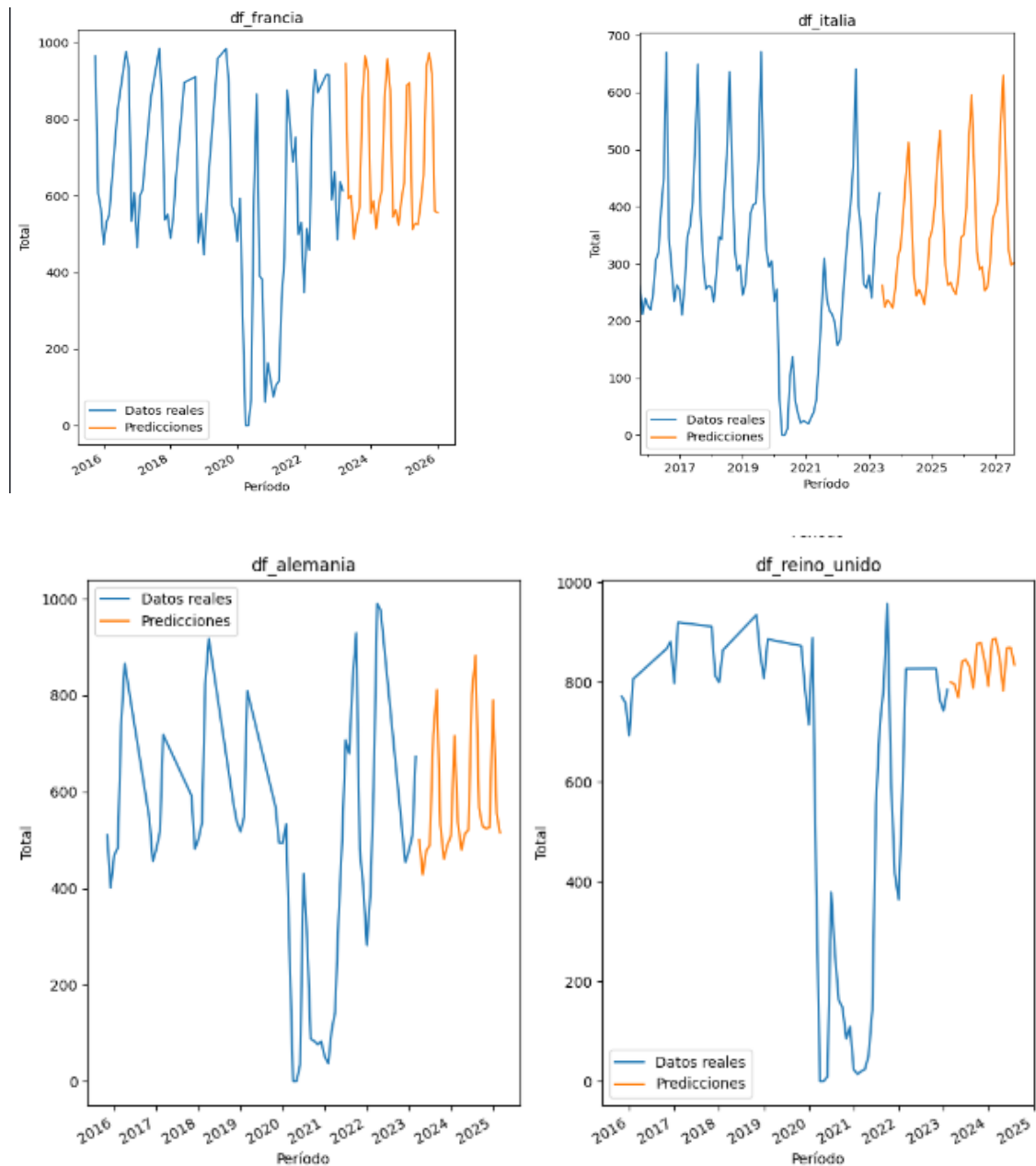


Aquí se visualiza los puntos de predicción y la cercanía a los datos reales guardados para test. Los modelos han sido entrenados para buscar los mejores parámetros que generalizan mejor, siendo reentrenados de nuevo con estos mejores estimadores.



\*Nota: El modelo no ha sido reentrenado en la fase de validación con todos los datos ya que partimos de una sucesión de crisis como el COVID, por lo que tener en cuenta esos datos aporta ruido a la predicción además de no ser fiel a los resultados, entendiendo estos parámetros como ruido en una situación de estabilización continuada desde 2023.





Resto del mundo:

Tiene un MSE de 1974.48, RMSE de 44.44,  $R^2$  de 0.89 y MAE de 37.23, con valores de afluencia turística que oscilan entre 300k y 600k. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva aceptable.

Los Países Bajos:

MSE de 347.74, RMSE de 18.65,  $R^2$  de 0.99 y MAE de 16.17, con valores de afluencia turística que oscilan entre 150k y 550k. Estos resultados indican que el modelo tiene un excelente ajuste a los datos y una capacidad predictiva muy alta.

Rusia:

MSE de 71.85, RMSE de 8.48,  $R^2$  de 0.98 y MAE de 7.02, con valores de afluencia turística que oscilan entre 50k y 175k. Estos resultados indican que el modelo tiene un excelente ajuste a los datos y una capacidad predictiva muy alta.

Los EEUU:

Tienen un MSE de 1640.23, RMSE de 40.50,  $R^2$  de 0.88 y MAE de 31.78, con valores de afluencia turística que oscilan entre 100k y 450k. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva aceptable.

Los Países Nórdicos:

Tienen un MSE de 11108.20, RMSE de 105.40,  $R^2$  de 0.32 y MAE de 82.09, con valores de afluencia turística que oscilan entre 350k y 800k. Estos resultados indican que el modelo no tiene un buen ajuste a los datos y su capacidad predictiva es baja.

Bélgica:

Tiene un MSE de 861.78, RMSE de 29.36,  $R^2$  de 0.92 y MAE de 18.36, con valores de afluencia turística que oscilan entre 100k y 400k. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva aceptable.

Irlanda:

Tiene un MSE de 153.68, RMSE de 12.40,  $R^2$  de 0.98 y MAE de 11.10, con valores de afluencia turística que oscilan entre 50k y 500k. Estos resultados indican que el modelo tiene un excelente ajuste a los datos y una capacidad predictiva muy alta.

	Country	MSE	RMSE	R^2	MAE
0	Resto_del_mundo	1974.48	44.44	0.89	37.23
1	Países_bajos	347.74	18.65	0.99	16.17
2	Rusia	71.85	8.48	0.98	7.02
3	EEUU	1640.23	40.50	0.88	31.78
4	P. Nordicos	11108.20	105.40	0.32	82.09
5	Belgica	861.78	29.36	0.92	18.36
6	Suiza	954.91	30.90	0.79	25.27
7	Italia	4308.95	65.64	0.83	43.63
8	Resto de América	448.04	21.17	0.94	18.37
9	Irlanda	153.68	12.40	0.98	11.10
10	Resto de europa	1641.43	40.51	0.96	34.38
11	Portugar	768.87	27.73	0.87	20.90
12	Francia	1866.68	43.21	0.90	37.84
13	Alemania	807.49	28.42	0.81	21.80
14	Reino_unido	650.72	25.51	0.80	21.78

El Resto de Europa:

Tiene un MSE de 1641.43, RMSE de 40.51,  $R^2$  de 0.96, y MAE de 34.38, con valores de afluencia turística que oscilan entre 700k y 300k. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva aceptable.

Portugal:

tiene un MSE de 768.87, RMSE de 27.73,  $R^2$  de 0.87, y MAE de 20.90, con valores de afluencia turística que oscilan entre 100k y 400k. Estos resultados indican que el modelo tiene un buena ajuste a los datos y una capacidad predictiva aceptable.

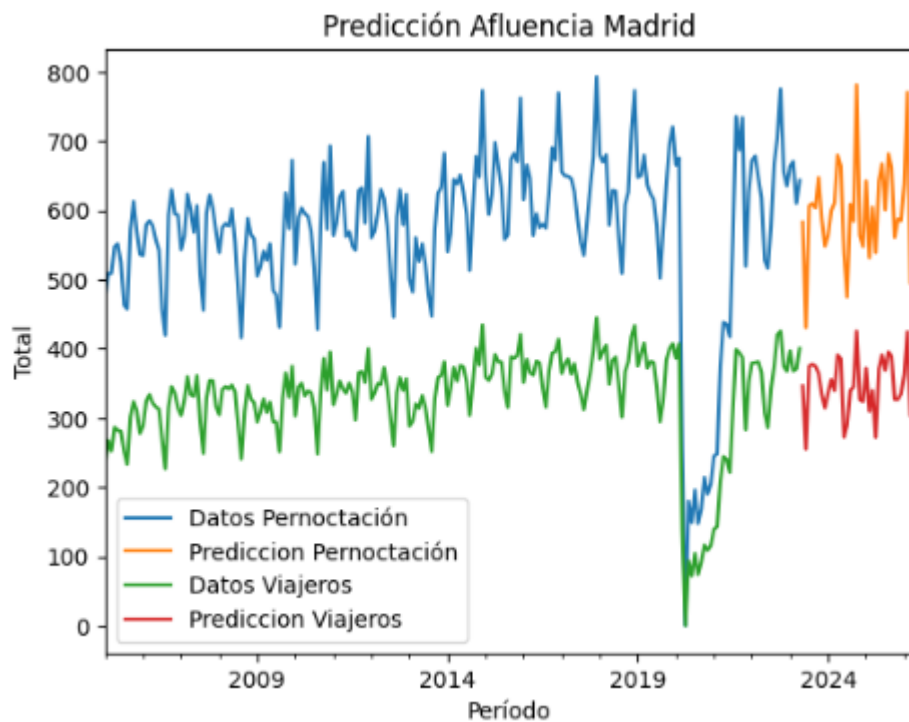
Francia:

Tiene un MSE de 1866.68, RMSE de 43.21,  $R^2$  de 0.90, y MAE de 37.84, con valores de afluencia turística que oscilan entre 450k y 1000k. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva aceptable.

Explicación métricas Turista nacional

Al igual que con los internacionales el modelo ganador RFR.

Aquí la visualización de las predicciones para los años siguientes en afluencia



Observamos que no has estacionalidad. Al realizar la prueba estadística y no dar por encima de 0.05 en la prueba adfuller.

Pernoctaciones

```
# valor de P_pernoc  
  
result_p = adfuller(df_pernoctaciones['Total'])  
result_p[1]  
  
✓ 0.0s  
0.0008187371111526778
```

Viajeros

```
# valor de Viajeros  
  
result_v = adfuller(df_viajeros['Total'])  
result_v[1]  
  
✓ 0.0s  
0.0021496386662418843
```

El valor de MAE (Error Absoluto Medio) de 253.27 indica que, en promedio, el modelo se equivoca en 253.27 pernoctaciones al hacer una predicción. Dado que el rango de pernoctaciones es de 400k a 800k, un error de 253.27 puede considerarse bajo en términos absolutos

Para los viajeros, cuya afluencia oscila entre 250k y 400k, el modelo tiene un MSE de 247.27, RMSE de 15.72,  $R^2$  de 0.87 y MAE de 13.25. Estos resultados indican que el modelo tiene un buen ajuste a los datos y una capacidad predictiva buena.