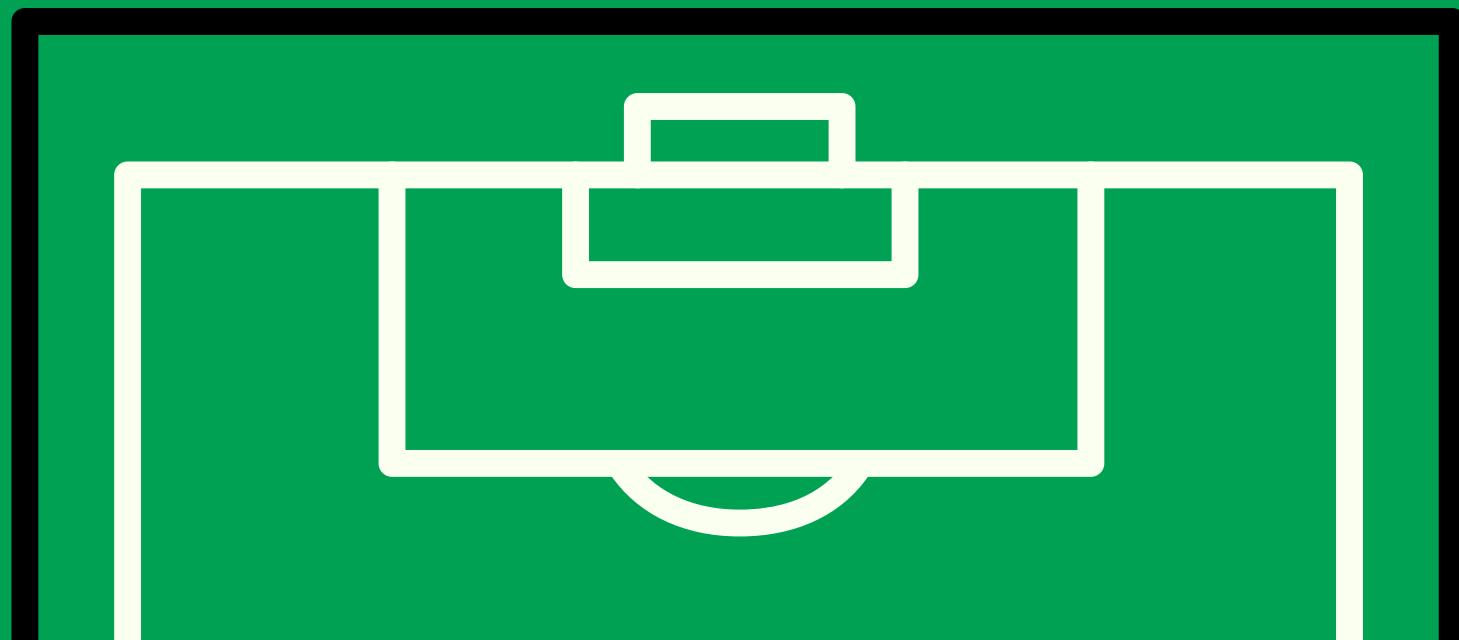


¿EL FUTBOL SE PUEDE PREDECIR?





LA RESPUESTA ES? DEPENDE...

PREDECIR EL RESULTADO Y ESTADÍSTICAS DE UN PARTIDO DE FÚTBOL ES UN DESAFÍO COMPLEJO DEBIDO A LA NATURALEZA ALTAMENTE IMPREDECIBLE Y MULTIFACÉTICA DEL DEPORTE. SIN EMBARGO, HOY EN DÍA SE PUEDEN UTILIZAR MODELOS ESTADÍSTICOS Y DE APRENDIZAJE AUTOMÁTICO PARA HACER PREDICCIONES BASADAS EN DATOS HISTÓRICOS Y CARACTERÍSTICAS ESPECÍFICAS DE LOS PARTIDOS Y EQUIPOS.

OBJETIVOS

OBTENCION DE LOS DATOS

- [HTTPS://FBREF.COM/ES](https://fbref.com/es)
- [HTTPS://WWW.FOOTBALL-DATA.CO.UK/](https://www.football-data.co.uk/)
- [HTTPS://WWW.SOFASCORE.COM](https://www.sofascore.com)

EXPLORACION, LIMPIEZA Y INGENIERIA DE DATOS

- [HTTPS://GITHUB.COM/19972024?TAB=REPOSITORIES](https://github.com/19972024?tab=repositories)

ANALISIS DESCRIPTIVO MEDIANTE VISUALIZACIONES

- [HTTPS://PUBLIC.TABLEAU.COM/APPPROFILE/MATTEO.STEFANO.ORTA/LLI/VIZ/PROYECTOFINALFINAL_1720267821790/RESULTADOSDEMEDIOTIEMPOYFINALESDEPARTIDO](https://public.tableau.com/app/profile/matteo.stefano.orta/visualizations/proyectofinalfinal_1720267821790/resultadosdemediotiempoynfinalesdepartido)

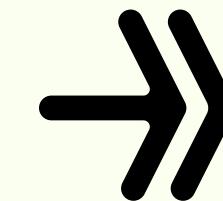
APLICACION DE MODELOS

- [HTTPS://GITHUB.COM/19972024?TAB=REPOSITORIES](https://github.com/19972024?tab=repositories)

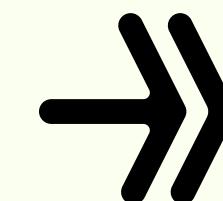


OBTENCION DE LOS DATOS:

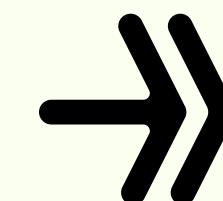
Obtuvimos 5 data sets de 5 diferentes temporadas (2019-2024) por 5 ligas diferentes (Bundesliga, Ligue 1, La Liga, Serie A y Premier League)



MIRAMOS QUE LAS COLUMNAS COINCIDAN Y CONCATENAMOS CADA SET DE DATOS PARA FORMAR UNA LIGA



CONCATENAMOS LOS DATA SETS DE CADA LIGA PARA CREAR UN UNICO SET DE DATOS



GUARDAMOS EL DATA SET PARA PODER REALIZARLE UN EDA

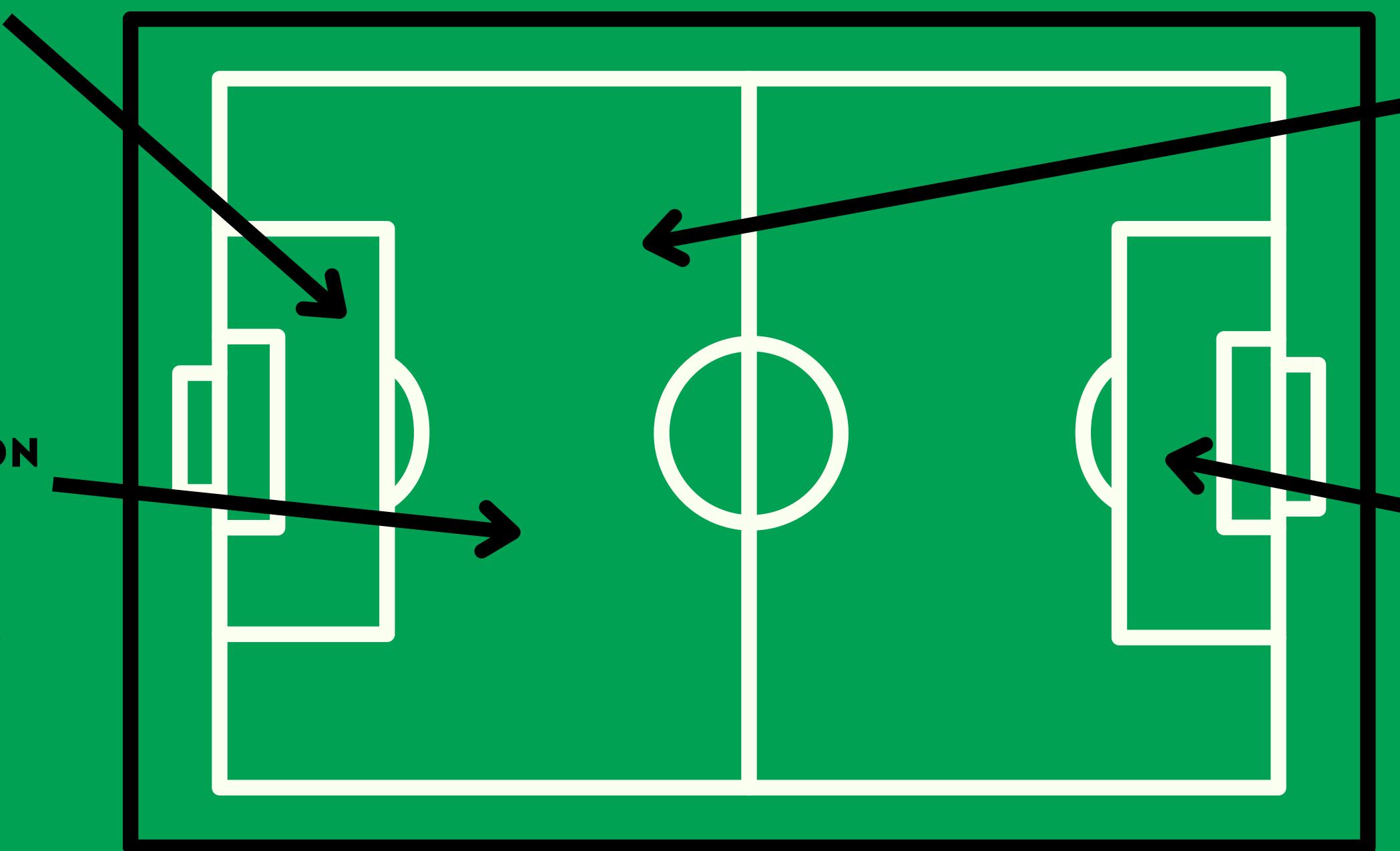
Exploracion, Limpieza y Ingenieria de datos

LIMPIEZA DE
NULOS Y
DUPLICADOS

ELIMINACION, CREACION
Y MODIFICACION DE
NUEVAS COLUMNAS.
(DIV, TOTAL MATCHES,
DAY TIME, TGG, THGG,
HOME Y AWAY TEAMS)

CONTROL DE
OUTLIERS Y
CREACION DEL
DATA SET FINAL

PREPARAMOS
NUESTROS DATOS
PARA LOS
DIFERENTES
MODELOS



NUMERICAL VS CATEGORICAL

ORDINAL ENCODER + CHI CUADRADO

#APLICAMOS UN ORDINAL ENCODER A NUESTRAS VARIABLES CATEGORICAS, NO QUEDA DEL TODO CONFORME CON LA TRANSFORMACION DE LOS EQUIPOS YA QUE NO CONSEGUI DARLE MUCHO PESO A LAS VARIABLES. (NO ELIMINAMOS VARIABLES, P VALUES CORRECTOS)

VISULIZAMOS SKEWNESS DE NUESTRAS VARIABLES NUMERICAS

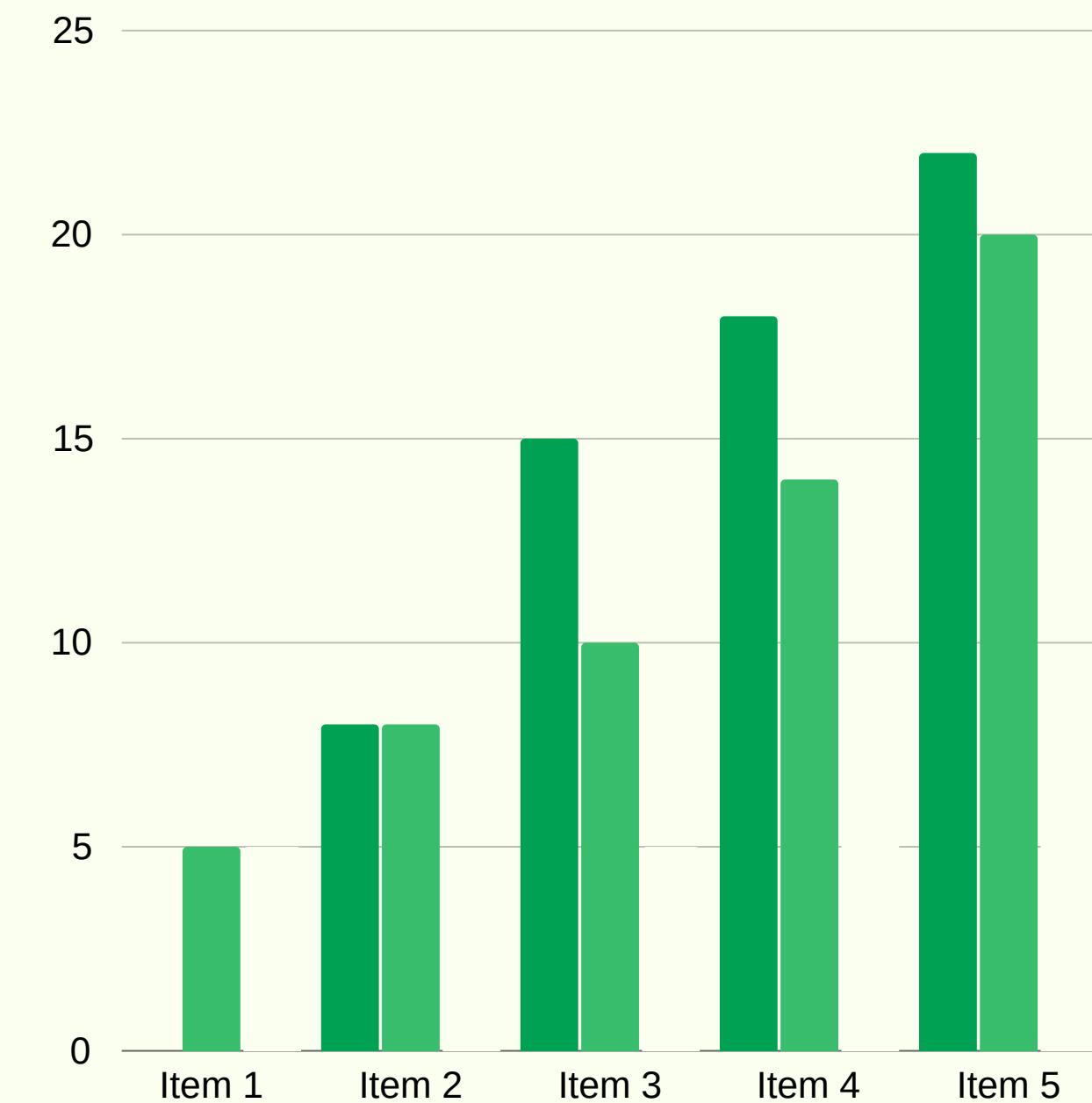
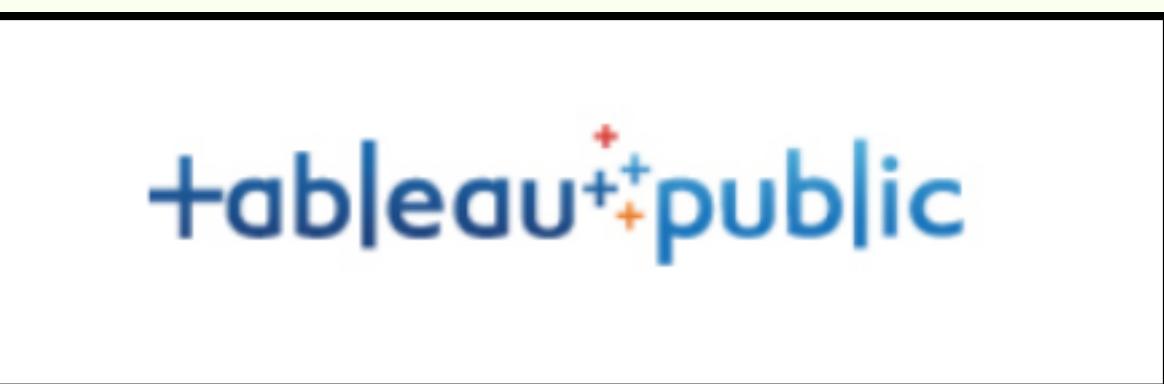
#DECIDIMOS NO NORMALIZAR NI STANDARIZAR YA QUE QUEREMOS PRESERVAR DE LA ESTRUCTURA DE LOS DATOS. ESTO TAMBIEN PARA MI GUSTO AGREGARIA COMPLEJIDAD A LA PREDICCION DEL MODELO.

APLICACION DE HEAT-MAPS PARA NUESTRAS VARIABLES NUMERICAS

#OBERVAMOS VARIABLES ALTAMENTE CORRELACIONADAS Y NO TANTO ENTRE SI, QUE LUEGO MIRAREMOS MAS DETALLADAMENTE CON LAS UTILIZACION DE VIF.



ANALISIS DESCRIPTIVO MEDIANTE VISUALIZACIONES



PROPHET PRONOSTICO TGG

2020



Covid

2021



Vuelta del
futbol a la
normalidad

2022



Mundial Qatar

2023



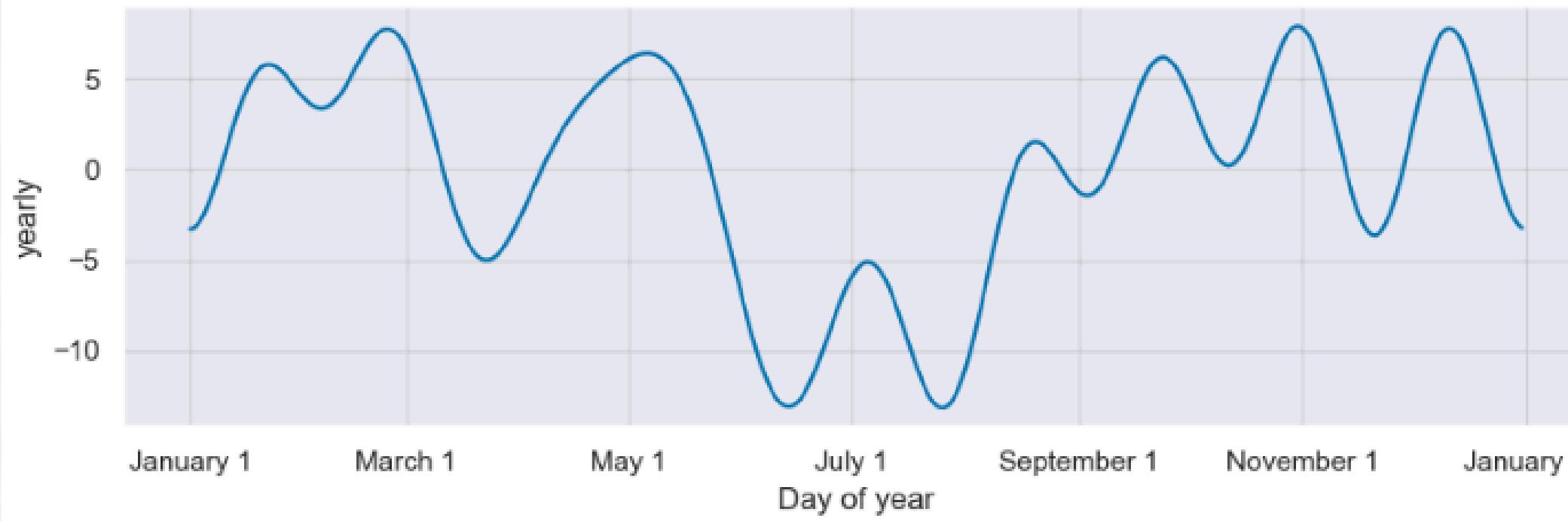
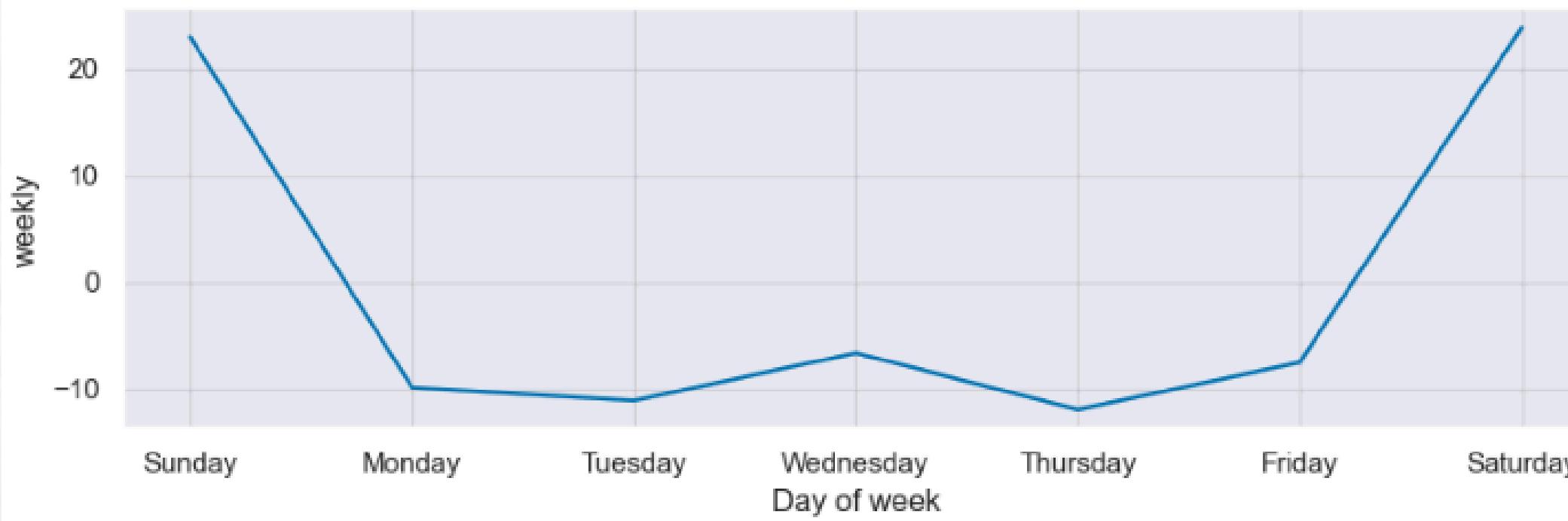
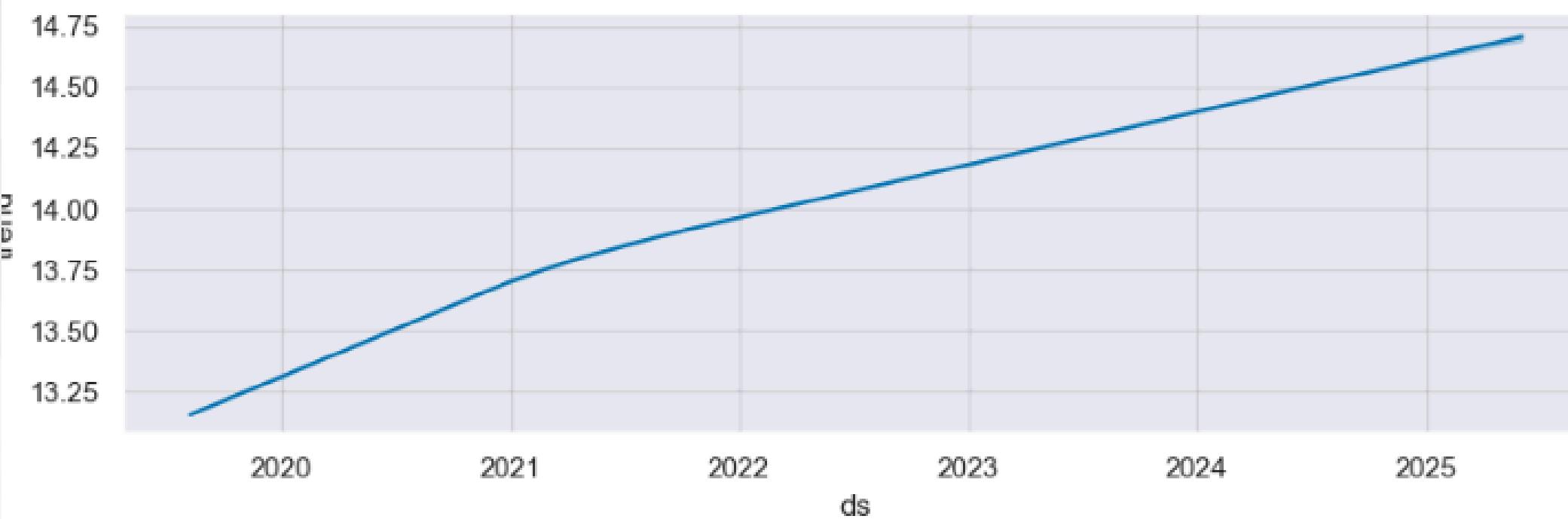
Vacaciones y
Fechas
internacionales

S

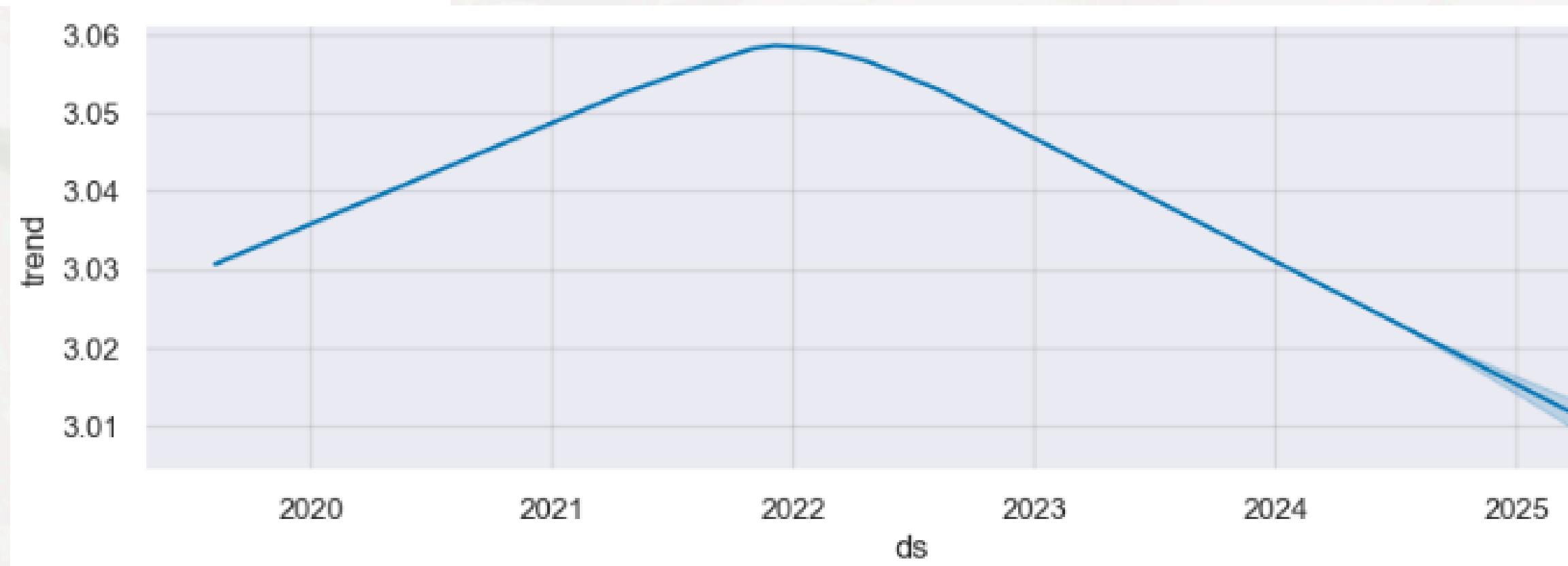
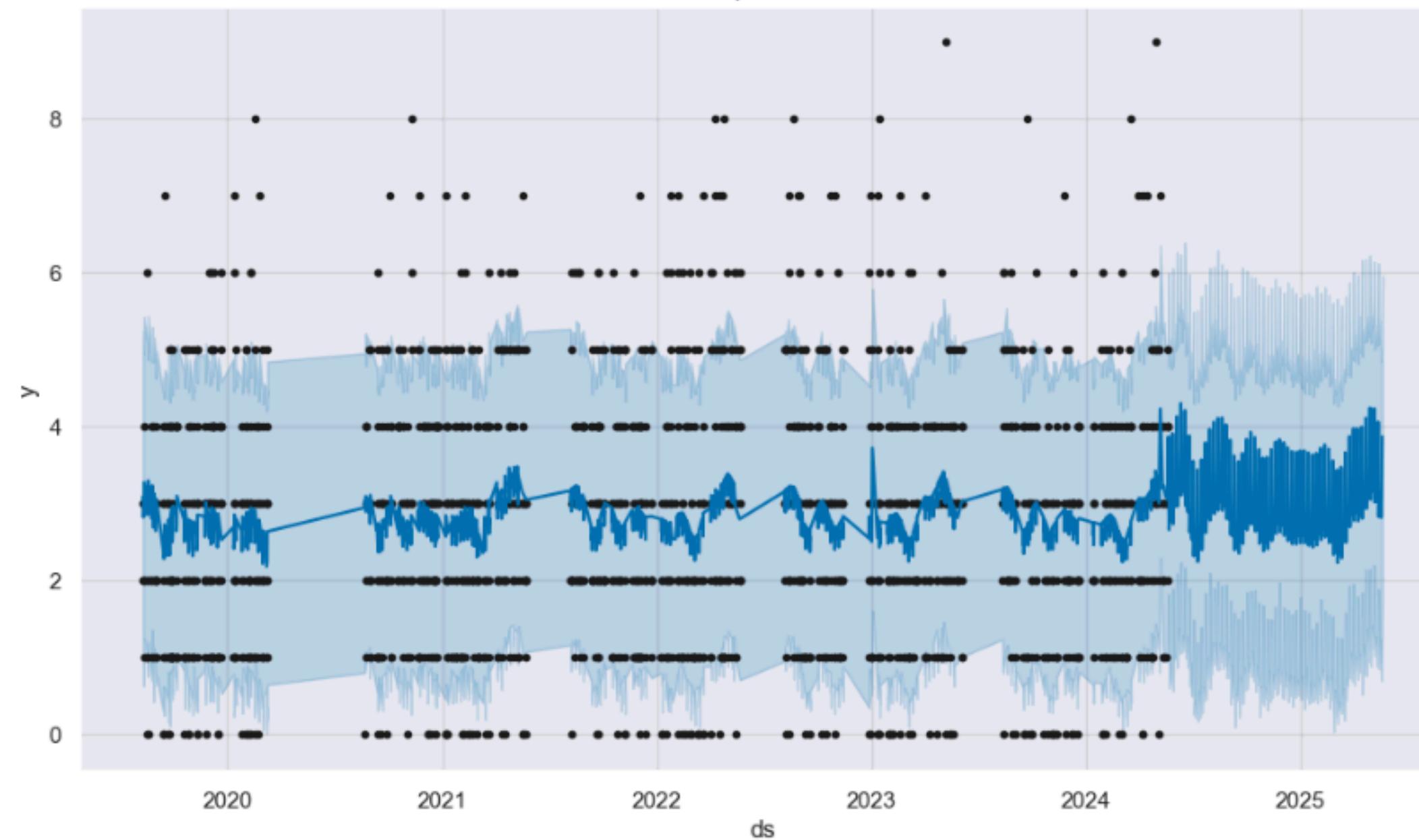
2024



Record de
goles historico

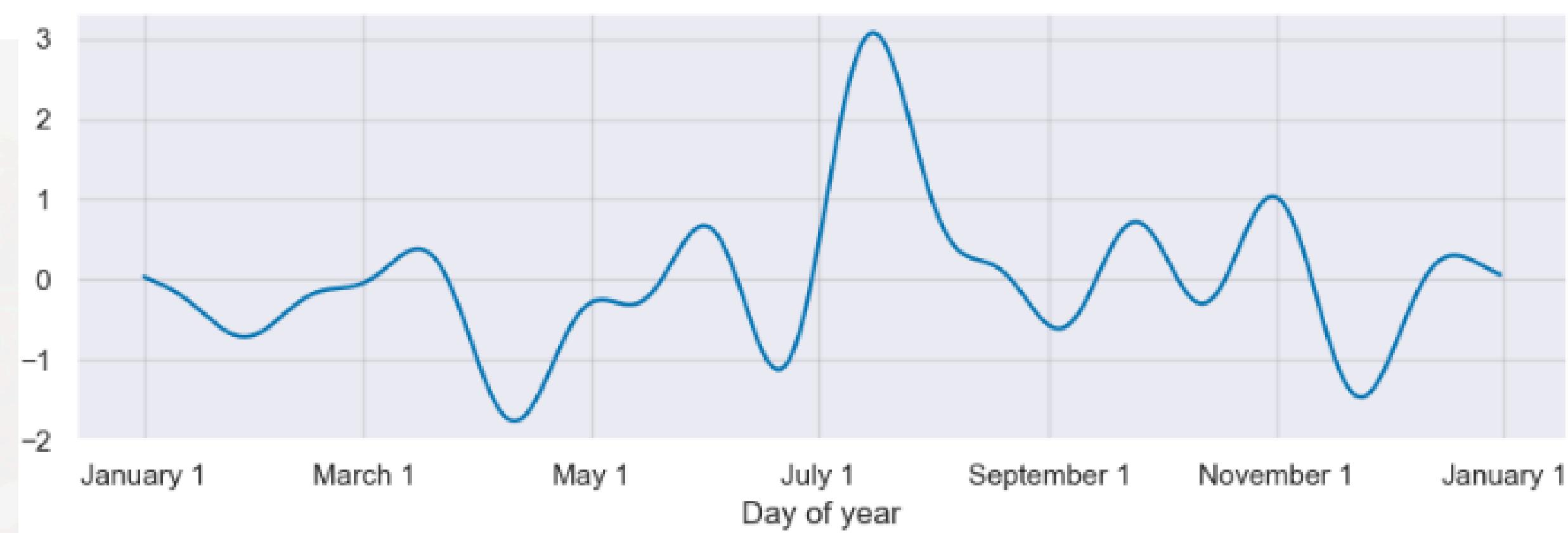
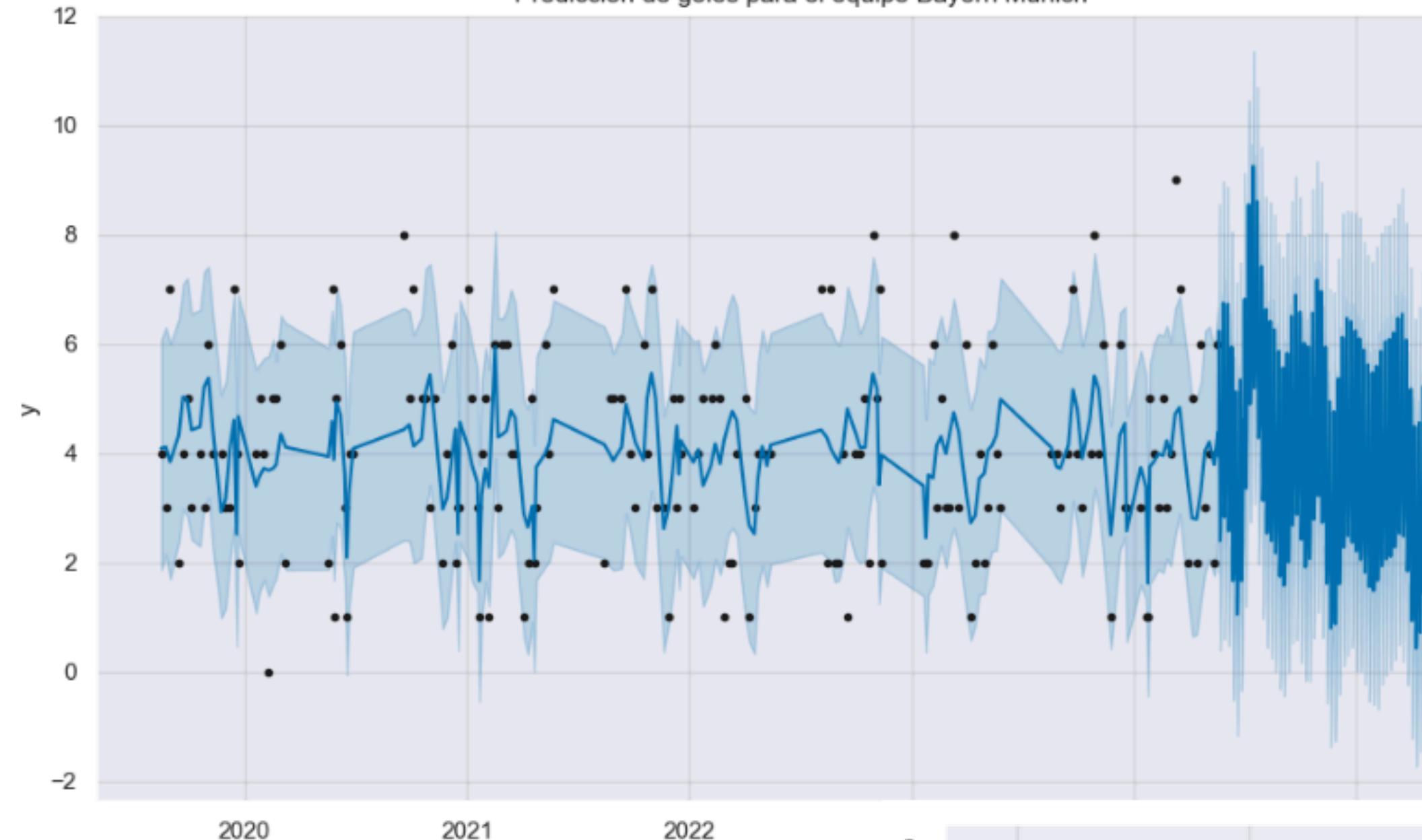


Predicción para División L1

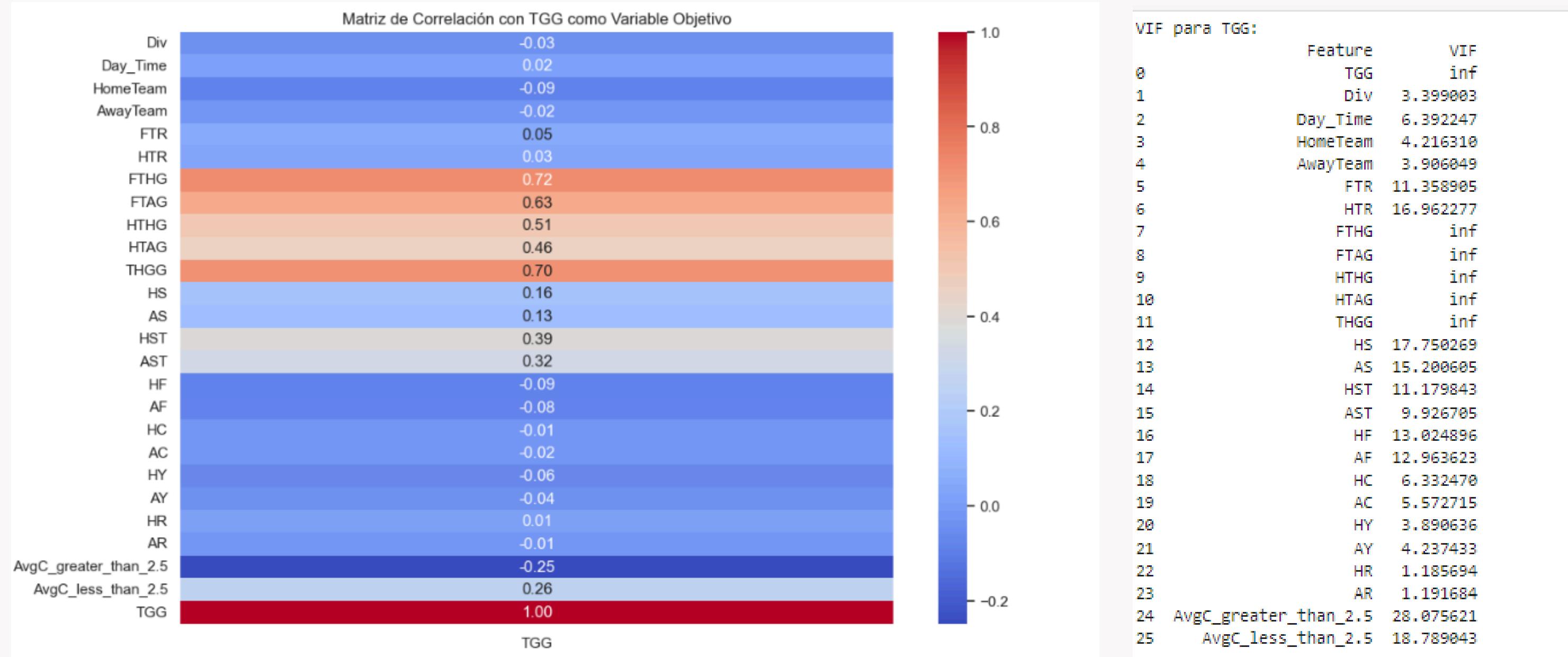


Media de 'yhat' para las predicciones del año 2025 para el equipo Bayern Munich: 3.3880191206355468

Predicción de goles para el equipo Bayern Munich



Random Forest



```

from sklearn.model_selection import GridSearchCV

# Definir la cuadrícula de hiperparámetros a explorar
param_grid = {
    'n_estimators': [50, 100, 200, 300],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10, 15]
}

# Inicializar el GridSearchCV con el modelo Random Forest
grid_search = GridSearchCV(estimator=RandomForestRegressor(random_state=42),
                           param_grid=param_grid,
                           cv=5, # Número de divisiones para la validación
                           scoring='r2', # Métrica de evaluación
                           verbose=1,
                           n_jobs=-1) # Utilizar todos los núcleos de CPU

# Entrenar GridSearchCV en los datos de entrenamiento
grid_search.fit(X_train, y_train)

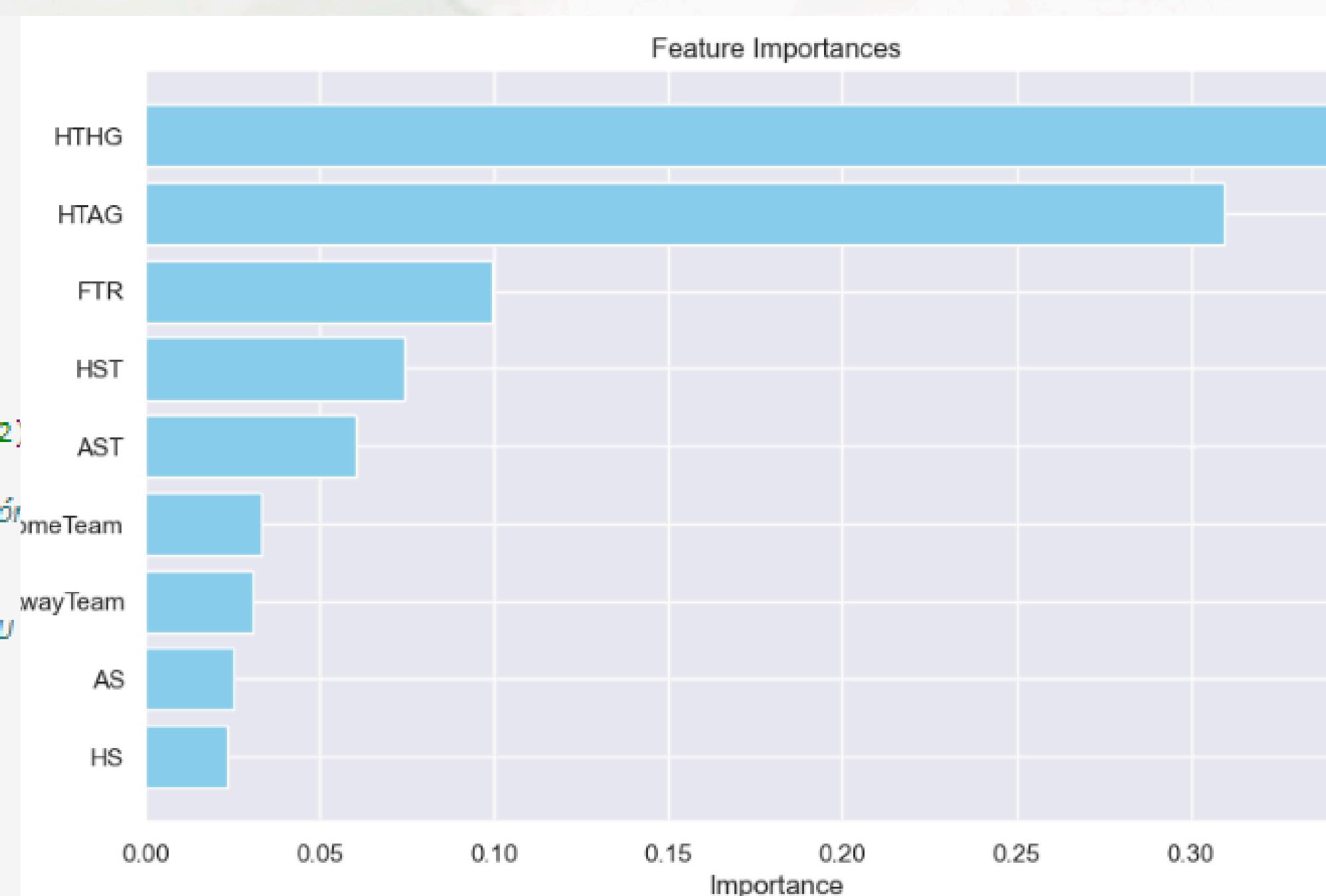
# Mostrar los mejores hiperparámetros encontrados
print("Mejores hiperparámetros encontrados:")
print(grid_search.best_params_)

# Obtener el mejor modelo entrenado
best_rf_model = grid_search.best_estimator_

# Hacer predicciones y evaluar el modelo ajustado
y_pred_grid = best_rf_model.predict(X_test)
r2_grid = r2_score(y_test, y_pred_grid)
print(f"R^2 Score después de ajuste de hiperparámetros: {r2_grid}")

Fitting 5 folds for each of 64 candidates, totalling 320 fits
Mejores hiperparámetros encontrados:
{'max_depth': 10, 'min_samples_split': 15, 'n_estimators': 300}
R^2 Score después de ajuste de hiperparámetros: 0.6174788842722858

```



Resultados R-cuadrado



TGG

0.71

0.17



HS

0.59

0.55



HST

0.58

0.57



AF

0.19

0.31



HF



AS



AST



HC

Predicciones



```
# Ajusta new_data para que coincida con las columnas de entrenamiento
new_data = pd.DataFrame({
    'HomeTeam': [1.10],
    'AwayTeam': [0.10],
    'FTR': [2.0],
    'HTHG': [2],
    'HTAG': [0],
    'HS': [13],
    'AS': [15],
    'HST': [8],
    'AST': [4],
})

# Hacer la predicción utilizando el modelo entrenado
predicted_tgg = rf_model.predict(new_data)
print(f"Predicción de TGG para el nuevo partido: {predicted_tgg}")
```

Predicción de TGG para el nuevo partido: [5.07]



```
: # Ajusta new_data para que coincida con las columnas de entrenamiento
new_data2 = pd.DataFrame({
    'HomeTeam': [1.10],
    'AwayTeam': [0.10],
    'FTR': [2.0],
    'FTHG': [5],
    'FTAG': [1],
    'HTHG': [2],
    'HTAG': [0],
    'HS': [13],
    'AST': [4],
    'AF': [14],
    'HC': [4],
    'AC': [2],
})

# Hacer la predicción utilizando el modelo entrenado
predicted_HST = rf_model.predict(new_data2)
print(f"Predicción de HST para el nuevo partido: {predicted_HST}")
```

Predicción de HST para el nuevo partido: [7.24]

Conclusiones

- MEJORAR EL MODELO, BUSCANDO Y AGREGANDO NUEVAS VARIABLES, EJEMPLO ARBITROS PARA LA PREDICCIÓN DE TARJETAS
- LAS PREDICCIONES TIENDEN A SER MÁS PRECISAS PARA LIGAS DONDE LOS EQUIPOS TIENEN UN RENDIMIENTO MÁS CONSISTENTE.
- LAS PREDICCIONES PUEDEN SER ÚTILES PARA APUESTAS DEPORTIVAS, ANÁLISIS TÁCTICOS Y ESTRATÉGICOS POR PARTE DE LOS EQUIPOS, Y PARA LOS FANÁTICOS INTERESADOS EN LAS PROBABILIDADES DE LOS RESULTADOS
- RESUMEN, AUNQUE LA PREDICCIÓN DEL FÚTBOL PUEDE PROPORCIONAR INFORMACIÓN VALIOSA Y PROBABILIDADES ÚTILES, SIEMPRE HABRÁ UN GRADO DE INCERTIDUMBRE DEBIDO A LA NATURALEZA DINÁMICA Y MULTIFACÉTICA DEL DEPORTE.

MUCHAS
GRACIAS

