

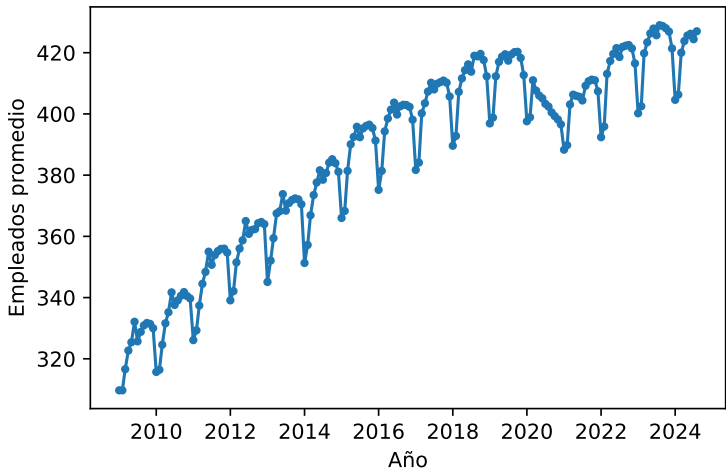
Tesis

Empleados promedio mensuales en Argentina, rubro: Enseñanza

(Personas con empleo asalariado registrado en el sector privado, según rama de actividad de la ocupación principal. Con estacionalidad. Total país. En miles. INDEC)

	fecha	empleados		fecha	empleados
0	2009-01-01	309.7			
1	2009-02-01	309.7			
2	2009-03-01	316.6			
185	2024-06-01	426.2			
186	2024-07-01	424.4			
187	2024-08-01	427.0			

Empleados promedio en Rosario en la industria de enseñanza en el sector privado (en miles)

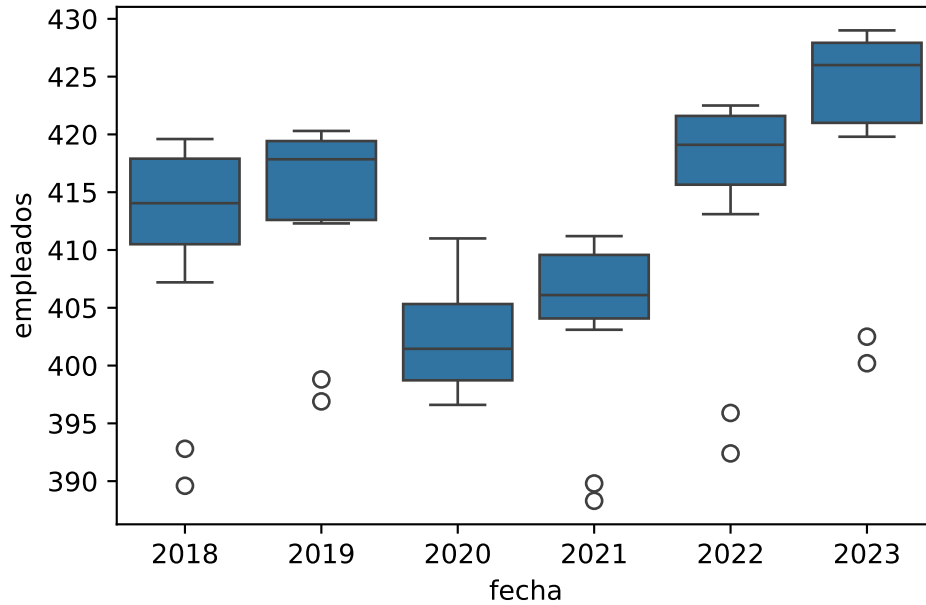


Tiene estacionalidad y tendencia

Hay para muchos mas rubros

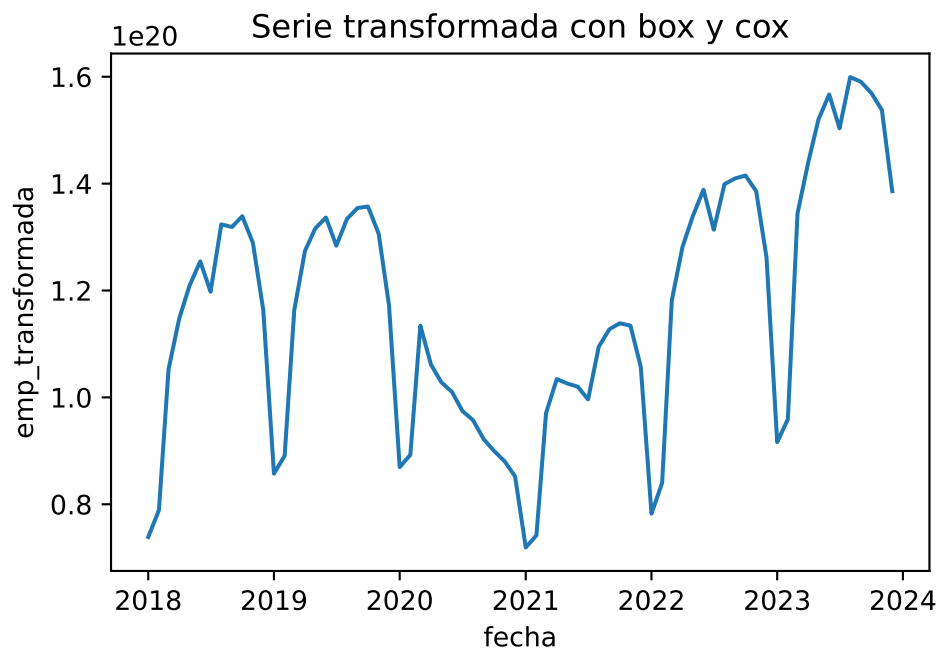
Solo tengo datos hasta agosto

Inicialmente habia tomado datos solo hasta febrero para pronosticar 6 meses, pero al tener solo 2 observaciones en 2024 afectaba mucho la variancia de este año. Además, la serie es muy larga, asi que voy a trabajar con 6 años nomás (2018-2023)



No parece haber necesidad de transformar la variable

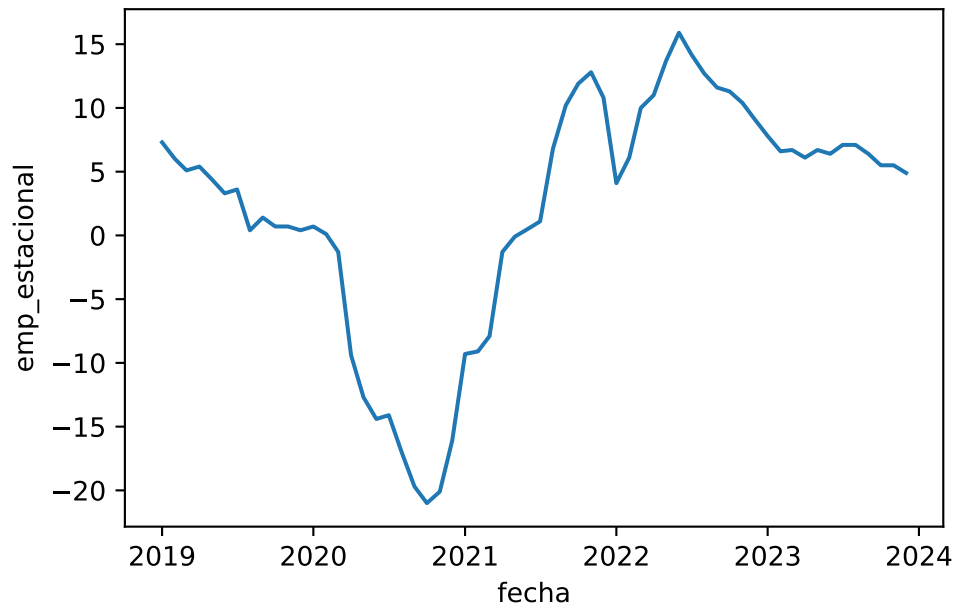
```
Text(2020-03-01 00:00:00, 3.1988977331307737e+19, 'P  metro de\n Box y Cox: 8.02')
```



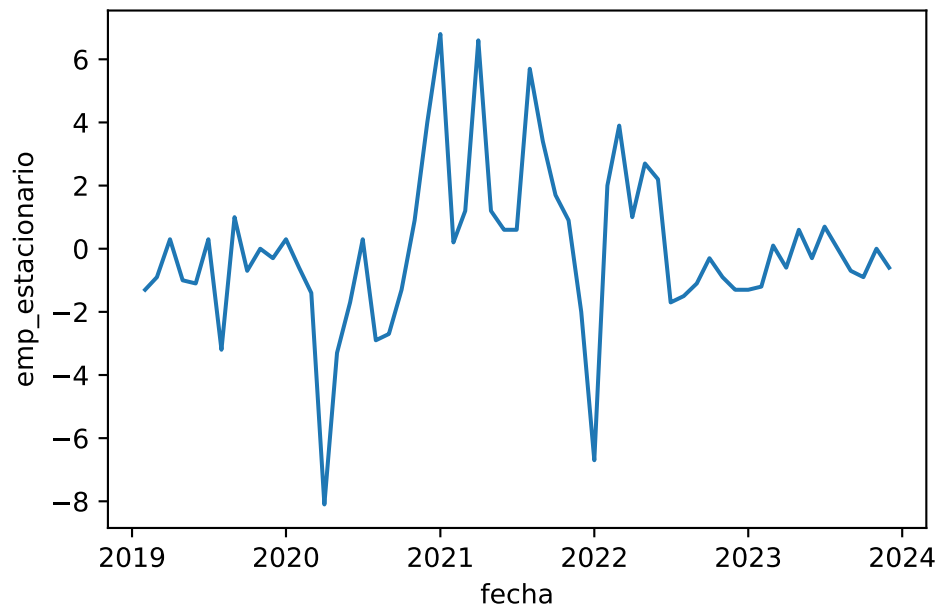
Párametro de
Box y Cox: 8.02

Finalmmente decido no usar la transformacion

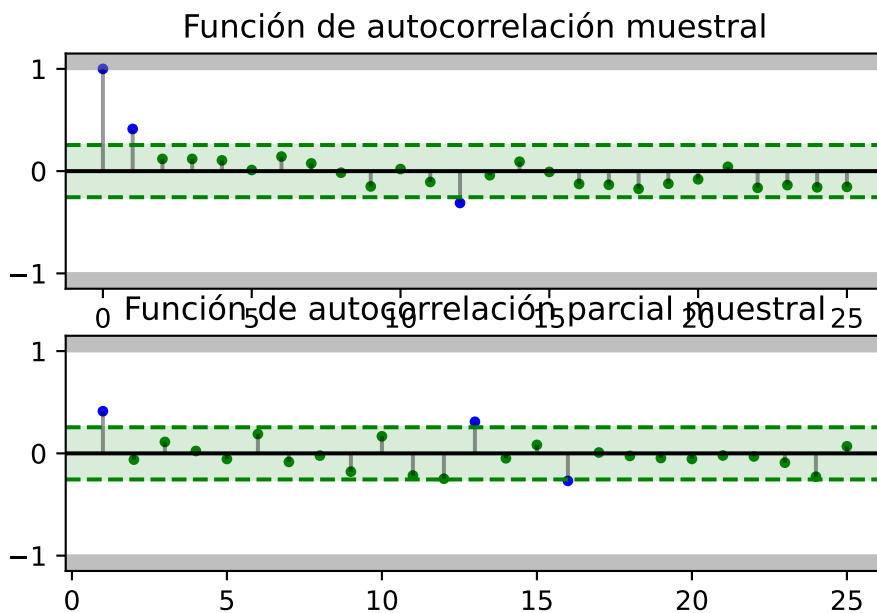
Diferencio estacionalmente la serie:



diferencio la serie en la parte estacionaria



Parece ser estacionaria pero con grandes outliers



La serie parece tener una componente MA tanto en la parte estacional como en la estacionaria

Selección manual

Dep. Variable:	empleados	No. Observations:	72
Model:	ARIMA(0, 1, 1)x(0, 1, 1, 12)	Log Likelihood	-123.501
Date:	Mon, 06 Jan 2025	AIC	253.003
Time:	10:58:45	BIC	259.235
Sample:	0	HQIC	255.436
	- 72		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	0.3880	0.133	2.926	0.003	0.128	0.648
ma.S.L12	-0.9983	49.836	-0.020	0.984	-98.676	96.679
sigma2	2.6811	133.424	0.020	0.984	-258.826	264.188

Ljung-Box (L1) (Q):	0.32	Jarque-Bera (JB):	20.26
Prob(Q):	0.57	Prob(JB):	0.00
Heteroskedasticity (H):	0.34	Skew:	-0.61
Prob(H) (two-sided):	0.02	Kurtosis:	5.60

Warnings:

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

Seleccion automatica

Dep. Variable:	y	No. Observations:	72
Model:	SARIMAX(1, 1, 0)x(3, 1, [1], 12)	Log Likelihood	-118.318
Date:	Mon, 06 Jan 2025	AIC	248.636
Time:	10:59:01	BIC	261.101
Sample:	0 - 72	HQIC	253.502
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.2680	0.148	1.807	0.071	-0.023	0.559
ar.S.L12	-0.9769	0.141	-6.945	0.000	-1.253	-0.701
ar.S.L24	-0.8867	0.103	-8.603	0.000	-1.089	-0.685
ar.S.L36	-0.8524	0.075	-11.433	0.000	-0.999	-0.706
ma.S.L12	0.4374	0.441	0.993	0.321	-0.426	1.301
sigma2	1.4344	0.573	2.505	0.012	0.312	2.557

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	4.20
Prob(Q):	0.99	Prob(JB):	0.12
Heteroskedasticity (H):	0.82	Skew:	-0.06
Prob(H) (two-sided):	0.66	Kurtosis:	4.30

Warnings:

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

Modelos seleccionados:

- $SARIMA(0, 1, 1)(0, 1, 1)_{12}$

Con $q = 0.3888$ y $Q = -0.9983$ | $AIC = 253.003$

Es admisible

- $SARIMA(1, 1, 0)(3, 1, 1)_{12}$

Con $p = 0.268$, $P_1 = -0.9769$, $P_2 = -0.8867$, $P_3 = -0.8524$, $Q = -0.4374$ | $AIC = 248.636$

No es admisible

Probamos el modelo $SARIMA(1, 1, 0)(2, 1, 1)_{12}$

Dep. Variable:	empleados	No. Observations:	72
Model:	ARIMA(1, 1, 0)x(2, 1, [1], 12)	Log Likelihood	-120.004
Date:	Mon, 06 Jan 2025	AIC	250.009
Time:	10:59:03	BIC	260.397
Sample:	0	HQIC	254.064
	- 72		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.4025	0.136	2.966	0.003	0.137	0.668
ar.S.L12	-0.1533	0.306	-0.501	0.617	-0.753	0.447
ar.S.L24	-0.3886	0.219	-1.771	0.077	-0.819	0.041
ma.S.L12	-0.9970	76.572	-0.013	0.990	-151.076	149.082
sigma2	1.9548	148.980	0.013	0.990	-290.041	293.951

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	9.18
Prob(Q):	0.94	Prob(JB):	0.01
Heteroskedasticity (H):	0.51	Skew:	-0.35
Prob(H) (two-sided):	0.14	Kurtosis:	4.80

Warnings:

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

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

Con $p = 0.4025$, $P_1 = -0.1533$, $P_2 = -0.3886$, $Q = -0.997$ | $AIC = 250$

Es admisible

Tambien voy a probar el mismo modelo anterior, pero con la componente ma en lugar de ar en la parte estacionaria, ya que es lo que se ve en los graficos de autocorrelacion

Dep. Variable:	empleados	No. Observations:	72
Model:	ARIMA(0, 1, 1)x(2, 1, 1, 12)	Log Likelihood	-120.204
Date:	Mon, 06 Jan 2025	AIC	250.407
Time:	10:59:04	BIC	260.795
Sample:	0	HQIC	254.462
	- 72		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	0.3989	0.148	2.697	0.007	0.109	0.689
ar.S.L12	-0.1333	0.273	-0.489	0.625	-0.668	0.401
ar.S.L24	-0.4221	0.195	-2.162	0.031	-0.805	-0.040
ma.S.L12	-0.9970	73.656	-0.014	0.989	-145.360	143.366
sigma2	1.9402	142.297	0.014	0.989	-276.956	280.836
<hr/>						
Ljung-Box (L1) (Q):			0.04	Jarque-Bera (JB):		9.92
Prob(Q):			0.84	Prob(JB):		0.01
Heteroskedasticity (H):			0.47	Skew:		-0.43
Prob(H) (two-sided):			0.10	Kurtosis:		4.81

Warnings:

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

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

Con $q = 0.3989$, $P_1 = -0.1333$, $P_2 = -0.4221$, $Q = -0.997$ | $AIC = 250.4$

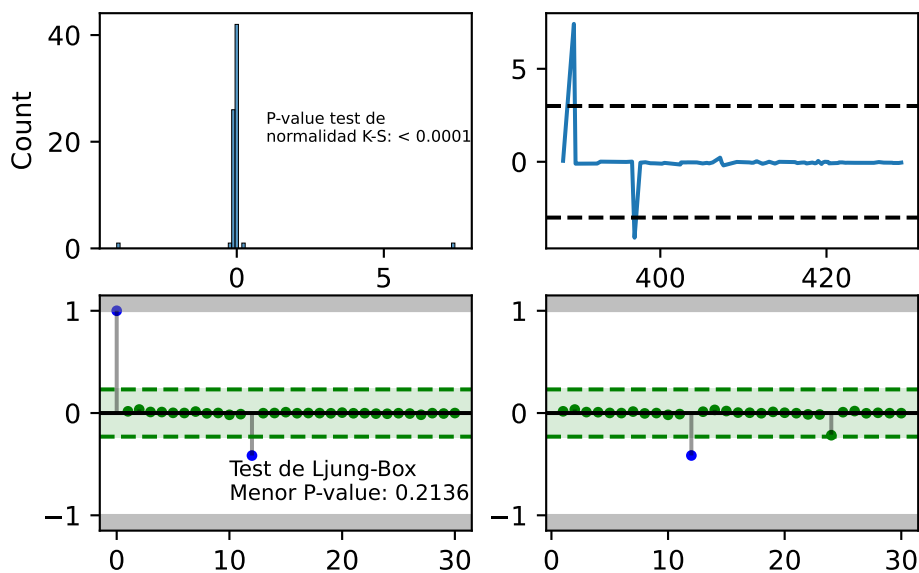
Es admisible

Comprobacion de supuestos

Hago la comprobacion de supuestos unicamente con los modelos admisibles

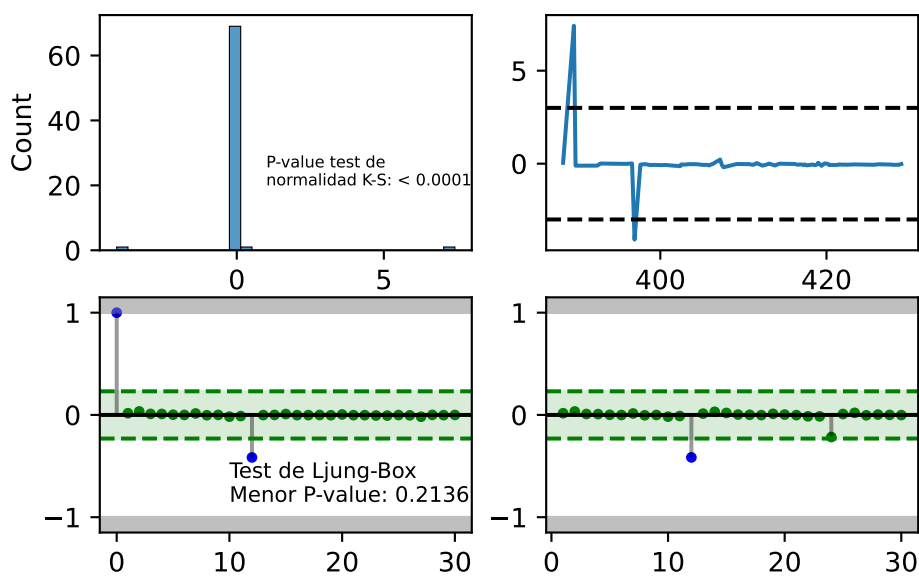
$SARIMA(0, 1, 1)(0, 1, 1)_{12}$

Text(0.5, 1.0, '')



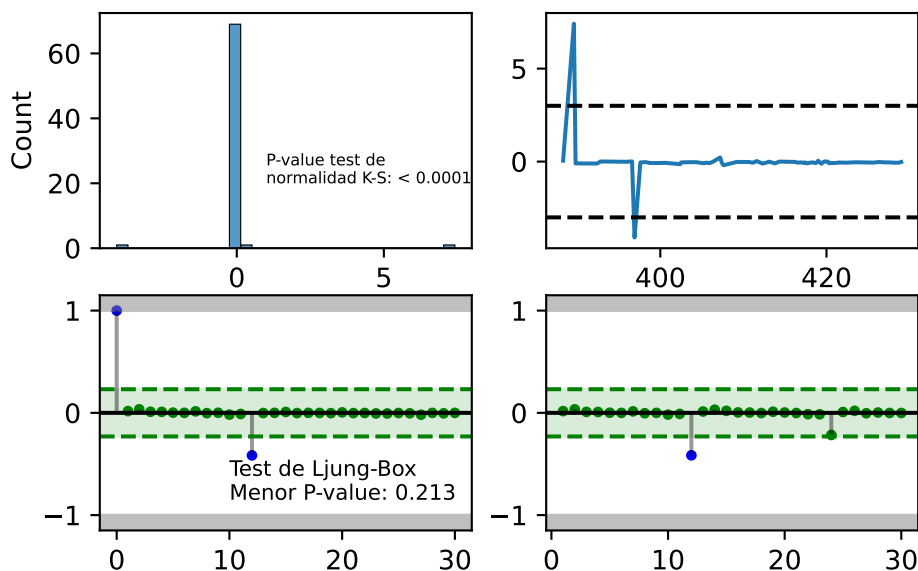
$SARIMA(1,1,0)(2,1,1)_{12}$

Text(0.5, 1.0, '')



$SARIMA(0, 1, 1)(2, 1, 1)_{12}$

Text(0.5, 1.0, '')



Comentarios en general:

Para todos los modelos se rechaza la normalidad de los residuos incluso sacando los 2 outlayers. La independencia de los residuos tambien se cumple para todos los modelos.

Pronosticos

```
# Pronostico Time gpt
pro_timeGPT = nixtla_client.forecast(
    df = empleo, h = 8, time_col= "fecha",
    target_col= "empleados",
    level=[90], finetune_steps=40, finetune_loss= "mape")
```

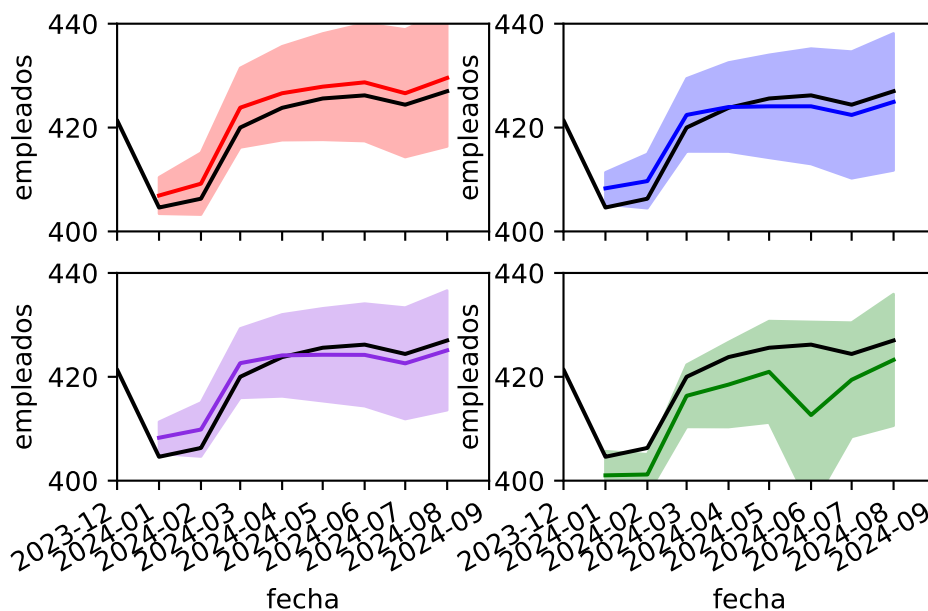
	fecha	TimeGPT	TimeGPT-hi-90	TimeGPT-lo-90
0	2024-01-01	401.02795	405.66992	396.38600
1	2024-02-01	401.19037	405.07877	397.30197
2	2024-03-01	416.33817	422.32013	410.35620
5	2024-06-01	412.63390	430.57077	394.69705
6	2024-07-01	419.42688	430.42750	408.42627

fecha TimeGPT TimeGPT-hi-90

7 2024-08-01 423.26140 435.93300 410.58980

Leyenda:

- $SARIMA(0, 1, 1)(0, 1, 1)_{12}$: Rojo
- $SARIMA(1, 1, 0)(2, 1, 1)_{12}$: Azul
- $SARIMA(0, 1, 1)(2, 1, 1)_{12}$: Violeta
- TimeGPT: Verde



El area marcada denota el intervalo de confianza del 90%

	Modelo	MAPE
0	$SARIMA(0,1,1)(0,1,1)_{12}$	0.637710
1	$SARIMA(1,1,0)(2,1,1)_{12}$	0.520113
2	$SARIMA(0,1,1)(2,1,1)_{12}$	0.516931
3	Time GPT	1.322634

Dado que los pronosticos con time gpt fueron muy malos voy a tratar con distintos parametros. Muy seguramente falla mucho por el sobreajuste, ya que se hicieron 40 pasos de tuneo de parametros.

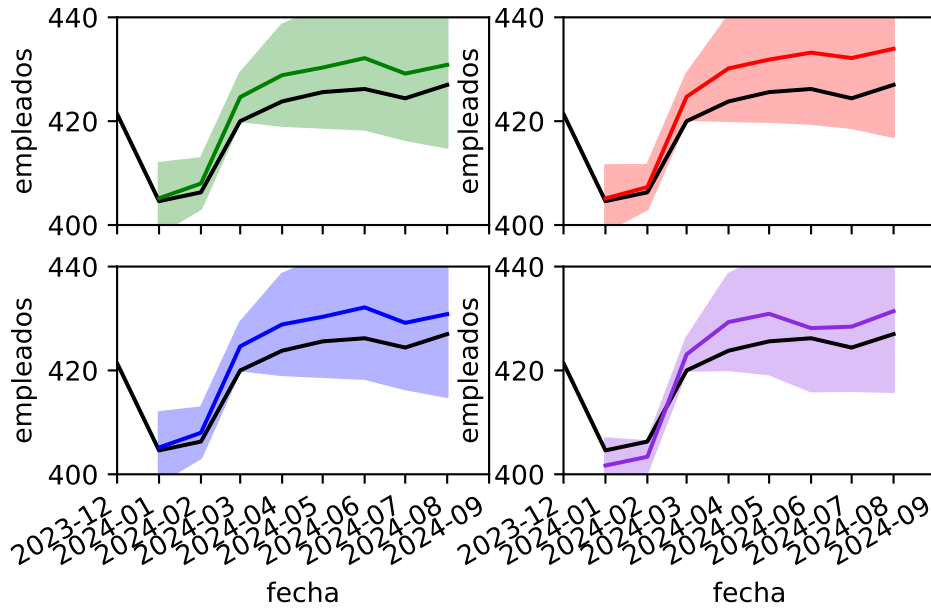
Distintos pronosticos con time gpt

```
# Pruebo las opciones por defecto
time_gpt1 = nixtla_client.forecast(
    df = empleo, h = 8, time_col= "fecha",
    target_col= "empleados",
    freq = 'M',
    level=[90])

# Pruebo usar 5 pasos de ajuste
time_gpt2 = nixtla_client.forecast(
    df = empleo, h = 8, time_col= "fecha",
    target_col= "empleados",
    level=[90], finetune_steps=5)

# Pruebo usar la funcion de perdida mape
time_gpt3 = nixtla_client.forecast(
    df = empleo, h = 8, time_col= "fecha",
    target_col= "empleados",
    level=[90], finetune_loss= "mape")

# Pruebo usar la funcion de perdida mape con 5 pasos de ajuste
time_gpt4 = nixtla_client.forecast(
    df = empleo, h = 8, time_col= "fecha",
    target_col= "empleados",
    level=[90], finetune_loss= "mape", finetune_steps=10)
```



	Modelo	MAPE
0	Time GPT conf 1	0.922305
1	Time GPT conf 2	1.195043
2	Time GPT conf 3	0.922305
3	Time GPT conf 4	0.895898

Los pronosticos mejoraron considerablemente pero siguen siendo peores que los metodos clasicos por ahora, hay que seguir viendo que problemas pueden tener y como solucionarlo.