

Predicción de Precios y Retornos sobre Bitcoin

Reconocimiento de Patrones

Santiago Zafra Rodríguez

Agenda



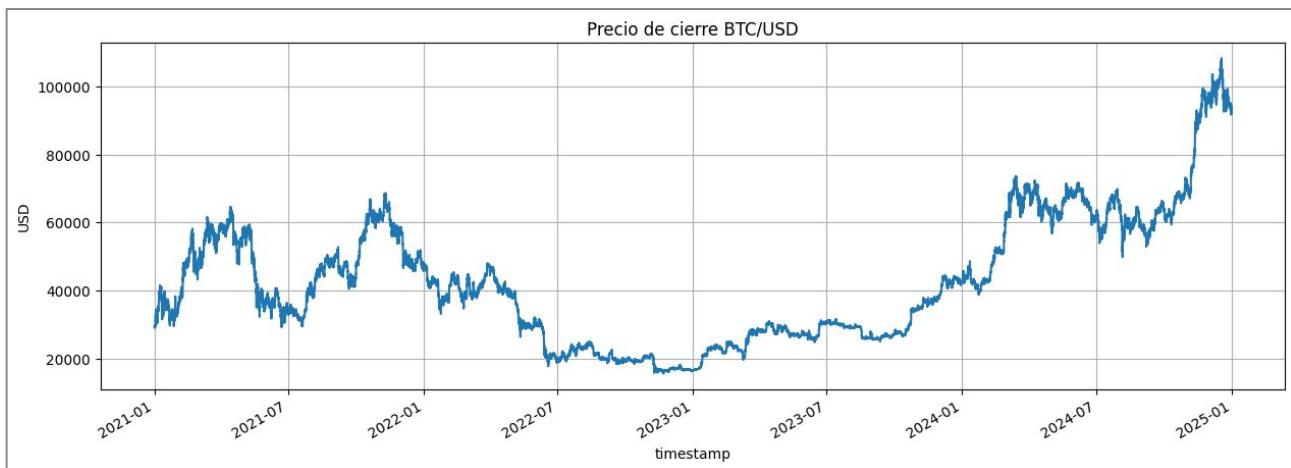


35.023
Registros

7
Características
Iniciales

1h
Temporalidad
de los datos

symbol	timestamp	open	high	low	close	volume	trade_count	vwap
BTC/USD	2021-01-01 06:00:00+00:00	29255.71	29338.25	29153.55	29234.15	42.244289	1243.0	29237.240312
BTC/USD	2021-01-01 07:00:00+00:00	29235.61	29236.95	28905.00	29162.50	34.506038	1070.0	29078.423185
BTC/USD	2021-01-01 08:00:00+00:00	29162.50	29248.52	28948.86	29076.77	27.596804	1110.0	29091.465155
BTC/USD	2021-01-01 09:00:00+00:00	29075.31	29372.32	29058.05	29284.92	20.694200	880.0	29248.730924
BTC/USD	2021-01-01 10:00:00+00:00	29291.54	29400.00	29232.16	29286.63	16.617646	742.0	29338.609132



- El dataset no contiene valores nulos.
- No hay saltos temporales. *Para todos los días del año se tiene información disponible*

Ventana de Tiempo

Los datos abarcan información histórica **horaria** del precio del bitcoin desde el **2021-01-01 hasta 2024-12-01**

Características

1. **Timestamp:** Registro temporal del precio
2. **Open:** Precio de apertura al inicio del período
3. **High:** Precio máximo alcanzado durante el período
4. **Low:** Precio Mínimo alcanzado durante el período
5. **Close:** Precio de cierre al final del período
6. **Volume:** Cantidad de unidades del activo negociadas
7. **Trade_Count:** Número de operaciones realizadas
8. **VWAP:** Volume Weighted Average Price: promedio ponderado por volumen

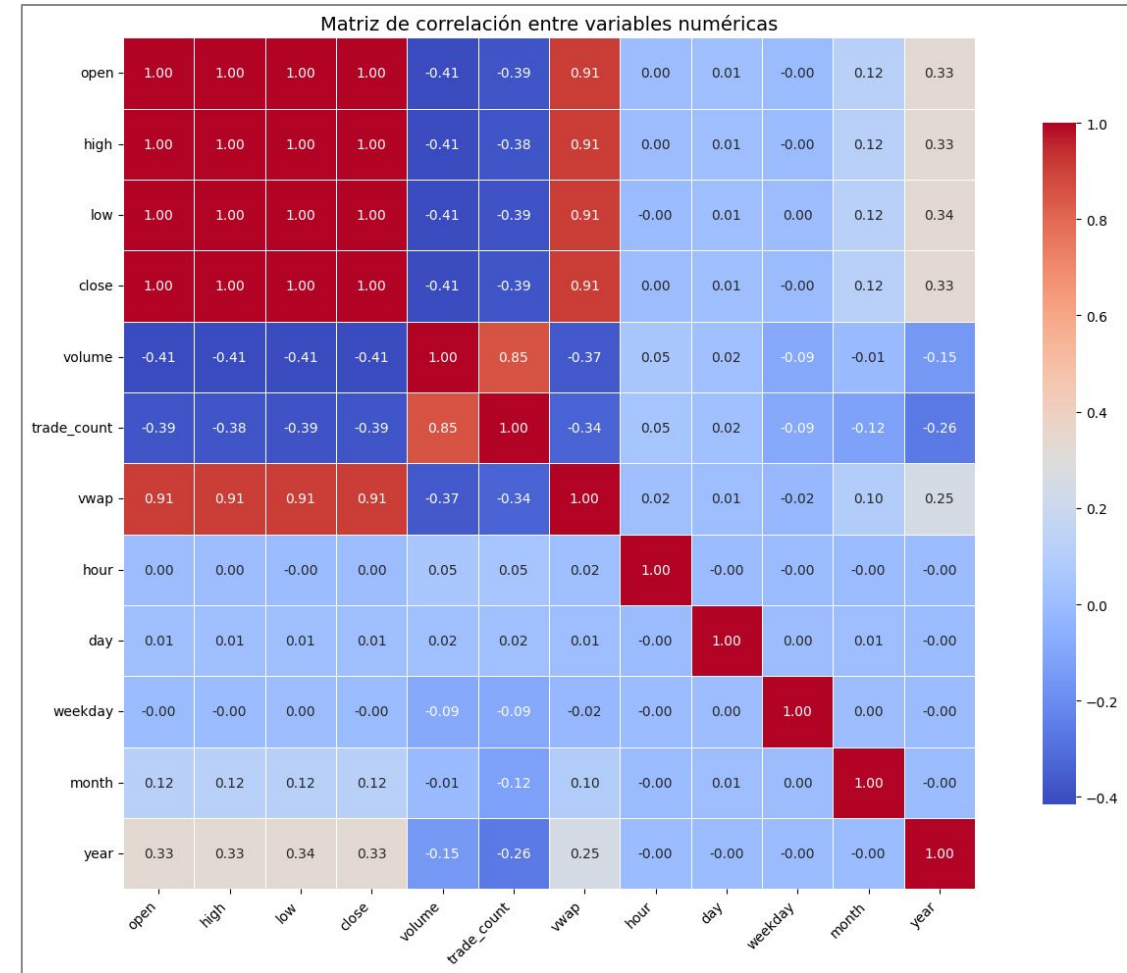
Avances

Exploración de Datos

Distribución de características

	count	mean	std	min	25%	50%	75%	max	skew	kurtosis
open	35023.0	42568.224980	18764.824908	15627.65	27288.150000	39740.820000	57339.705000	108249.820000	0.787299	0.352234
high	35023.0	42756.786597	18834.653267	15750.44	27415.827250	39940.330000	57594.250000	108401.732000	0.783361	0.346725
low	35023.0	42364.413575	18693.592789	8200.00	27195.120154	39520.880000	57031.430000	107152.527500	0.791938	0.358877
close	35023.0	42568.605657	18766.747970	15631.84	27285.585546	39744.700000	57333.420000	108287.880000	0.787380	0.352202
volume	35023.0	94.817083	194.472915	0.00	0.036017	25.832720	97.211113	5213.685947	5.551978	62.229957
trade_count	35023.0	2143.935500	3444.282708	0.00	6.000000	1243.000000	2828.500000	110487.000000	6.036933	96.498802
vwap	35023.0	41446.972214	19478.473852	0.00	26809.810187	39013.971195	56542.748049	107667.463077	0.632688	0.436964
hour	35023.0	11.502727	6.922364	0.00	6.000000	12.000000	18.000000	23.000000	-0.000662	-1.204030
day	35023.0	15.721098	8.793960	1.00	8.000000	16.000000	23.000000	31.000000	0.007000	-1.194121
weekday	35023.0	3.002541	2.001290	0.00	1.000000	3.000000	5.000000	6.000000	-0.002595	-1.251320
month	35023.0	6.520886	3.446767	1.00	4.000000	7.000000	10.000000	12.000000	-0.009120	-1.207728
year	35023.0	2022.500642	1.117833	2021.00	2022.000000	2023.000000	2024.000000	2024.000000	-0.000689	-1.359525

- Las variables de precios (open, high, low, close, vwap) tienen una **distribución simétrica y ligeramente sesgada a la derecha**, con baja kurtosis, lo que indica dispersión moderada y pocos extremos.
- Presentan **alta volatilidad** (mínimos de ~15 mil y máximos casi 10 veces mayores).
- Volume y trade_count están **fuertemente sesgados a la derecha**, con kurtosis extremadamente alta, lo que evidencia una gran cantidad de valores pequeños y algunos valores atípicamente grandes.



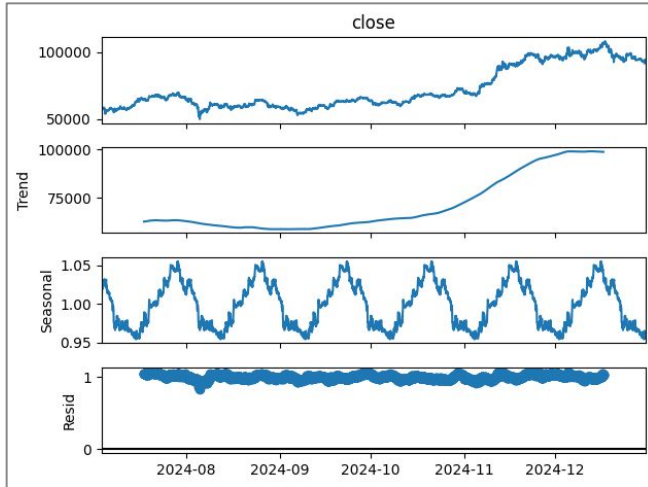
- Las variables de precios presentan problemas de multicolinealidad, lo que sugiere la **simplificación de estas características. (Revisar PCA)**
- Dependiendo del método, es importante realizar **escalado y normalización** de las variables.

Avances

Análisis Temporal

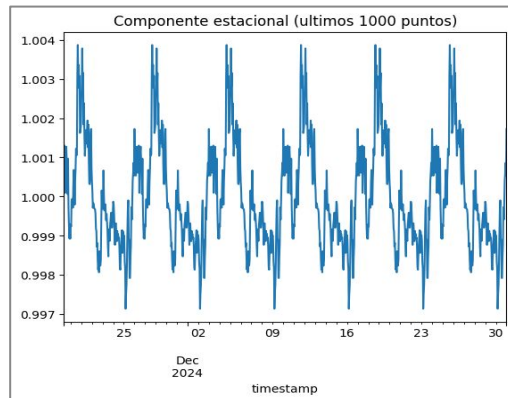
Descomposición Estacional Multiplicativa

Últimos 6 meses - Ciclo Mensual



- Hay un componente estacional interesante mensual, donde se evidencian picos y valles repetitivos.
- Esta frecuencia es interesante para explorar características que capturen **patrones mensuales** que enriquezcan el dataset.

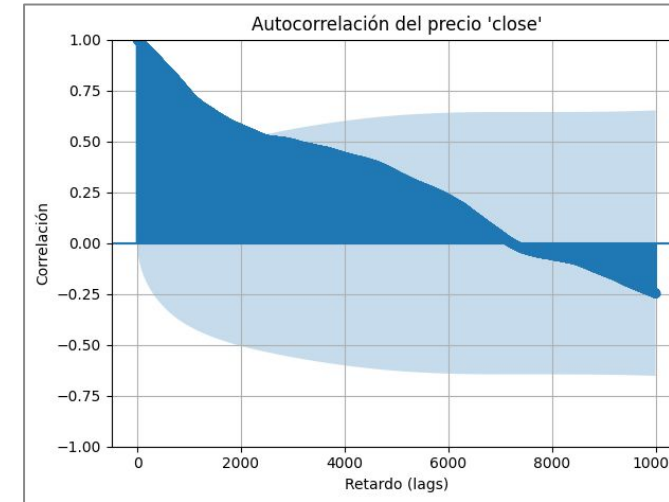
Ciclo Semanal



- **Semanalmente** se ven patrones que indican comportamientos similares en días y horas específicos de la semana y capturan parte del cambio en el precio
- La amplitud del componente estacional es **pequeña** pero consistente.
- Incluir variables **semanales y mensuales** para comprender patrones más complejos

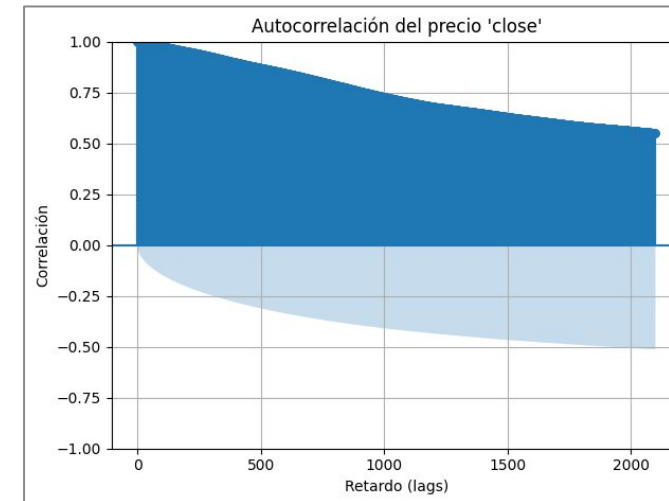
Autocorrelación del precio

Últimos 18 meses



- El precio está fuertemente correlacionado consigo mismo en el **corto plazo** (20 días mayor a 0.8).
- Se ven comportamientos claros de tendencia pero no de estacionalidad. Esto sugiere explorar variables como los **retornos**, que pueden tener **comportamientos estacionales** más claros.

Últimos 3 meses



Avances Modelamiento

Modelo Base

Regresión Lineal Simple

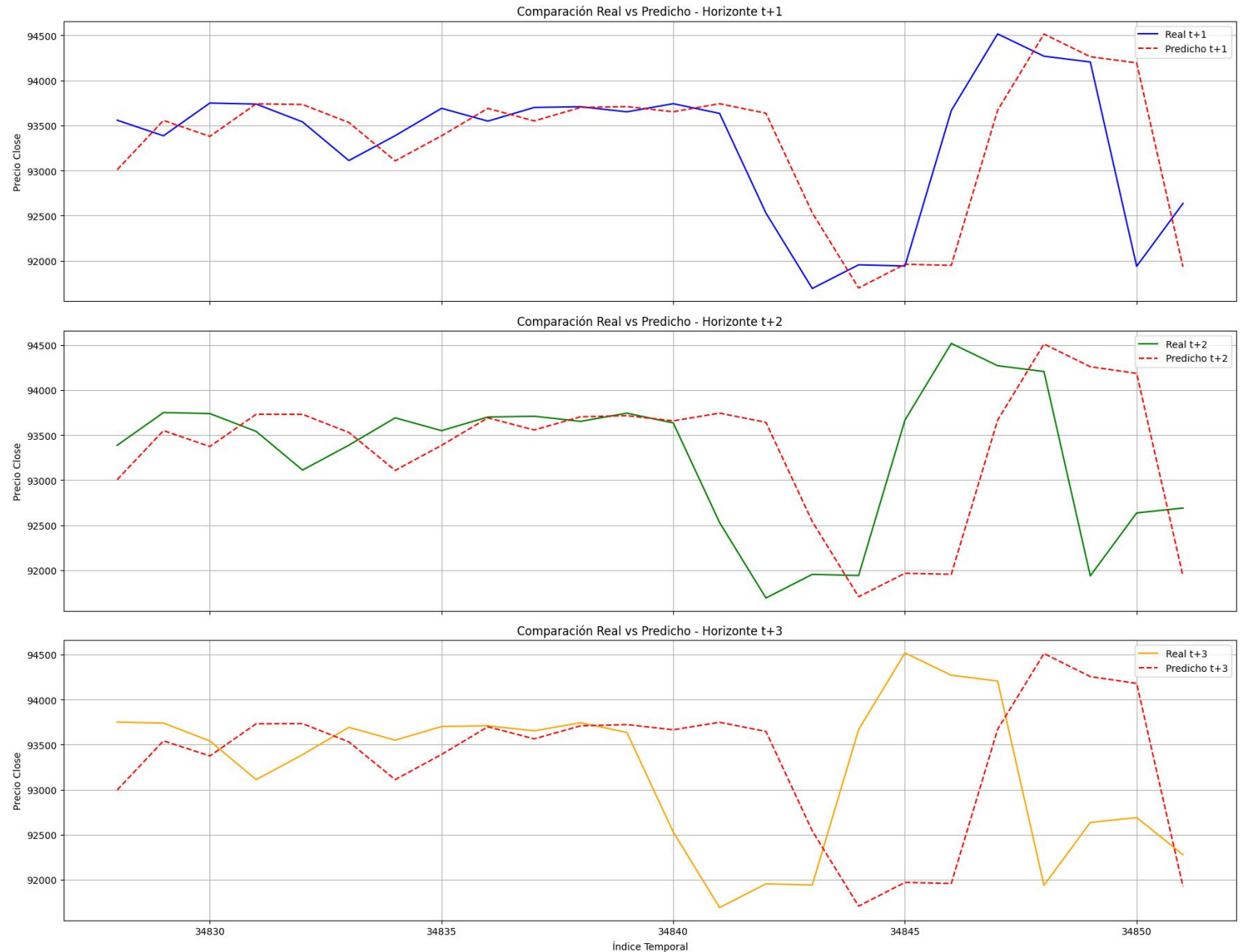
```
# --- 1. Crear nuevos targets desplazados ---  
df['close_target_1h'] = df['close'].shift(-1) # +1 hora  
df['close_target_2h'] = df['close'].shift(-2) # +2 horas  
df['close_target_3h'] = df['close'].shift(-3) # +3 horas
```

- **No** se pueden usar las características del momento **t** para predecir el precio del mismo momento.
 - Estas características ya contienen información (low, min, max, vwap) que conocen el precio **close**.
- Es necesario que el modelo use las características del momento **t**. Para predecir el precio en t+1, t+2,...,t+n.

Hiperparámetros: Por defecto (Sin regularización)

Resultados:

Horizonte		MSE	MAE	R ²
0	t+1	156063.38	257.76	0.9991
1	t+2	307245.43	361.94	0.9982
2	t+3	456115.13	445.60	0.9973



Avances Modelamiento

Modelo SVR

Support Vector Regression

Hiperparámetros: Optimización con GridSearch

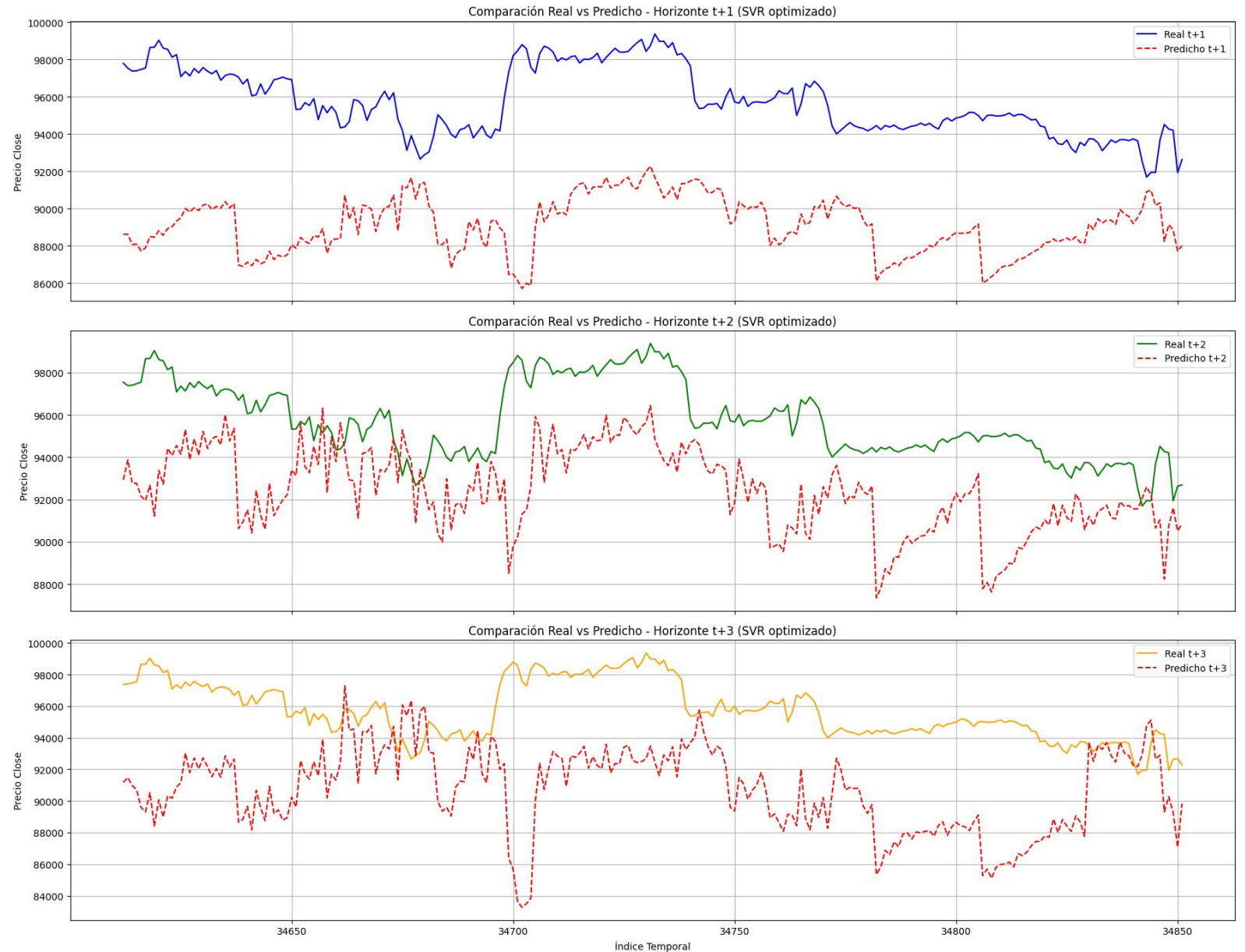
```
# --- 5. Grid search manual (C, epsilon, gamma) ---
param_grid = {
    'C': [100, 500, 1000],
    'epsilon': [0.1, 0.01, 0.001],
    'gamma': ['scale', 0.01, 0.001]
}
```

- Mejor modelo encontrado: C=1000, epsilon=0.1, gamma=0.01

Resultados:

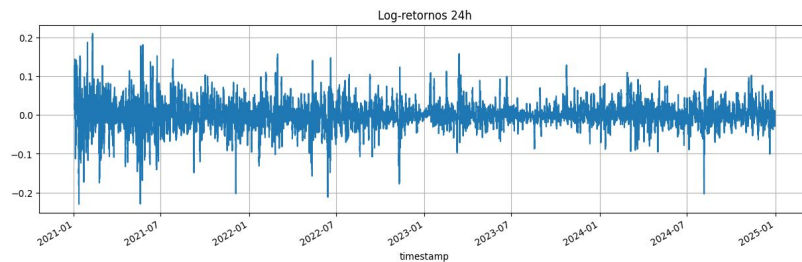
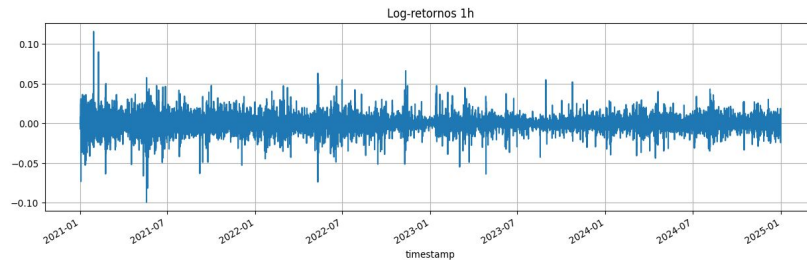
Horizonte		MSE	MAE	R ²
0	t+1	10480774.29	1915.72	0.9372
1	t+2	5977993.71	1512.93	0.9642
2	t+3	7894055.91	1767.94	0.9528

El modelo presenta dificultades computacionales para explorar hipótesis más complejas, lo que dificulta su modelamiento.



Modelamiento

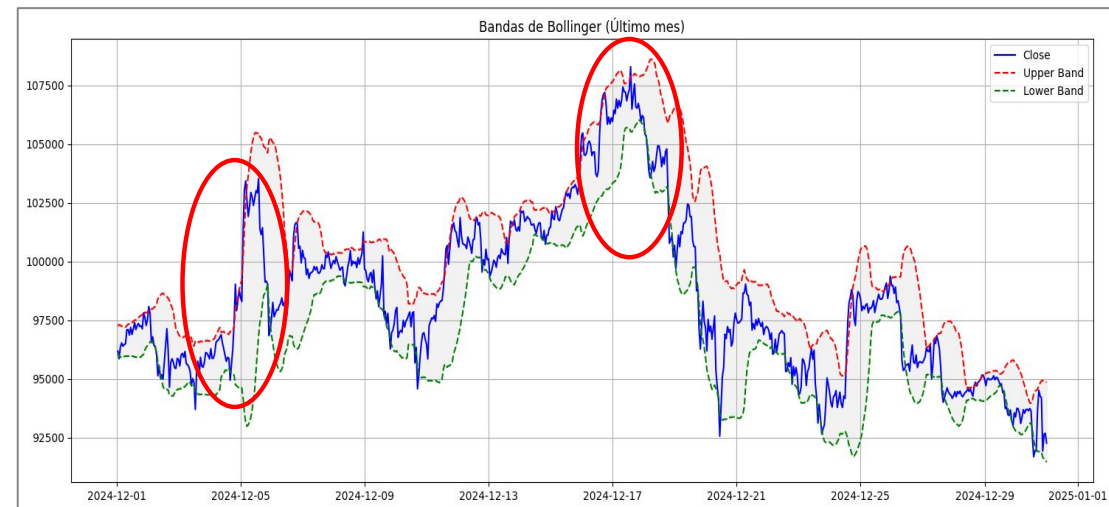
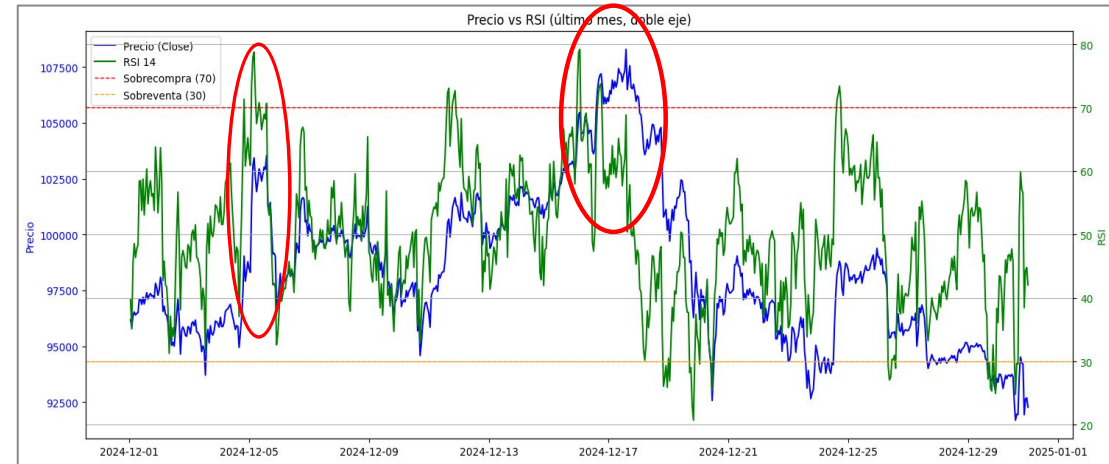
1. Predecir el precio más que complejo puede sesgar los resultados, **sin capturar cambios de tendencias adecuadamente**. Por lo que se busca cambiar la variable a predecir **precio** por **retornos logarítmicos** para identificar comportamientos y evitar que el modelo tome como predicción la tendencia.



2. Explorar modelos que **computacionalmente sean más efectivos** y puedan captar **patrones más complejos** XGBoost, RandomForest, ARIMA, SARIMA, para **explorar mejor el espacio de hipótesis**.

3. Agregar variables que capturen **la volatilidad y posibles cambios de tendencia**, que permitan que el modelo **se adelante al precio** y evite sesgarse por el comportamiento **pasado**.

Indicadores Técnicos a Revisar: RSI, Bandas de Bollinger, MACD, etc.



GRACIAS