



Instituto Politécnico Nacional

Escuela Superior de Computo



Carrera

Licenciatura en Ciencia de Datos

Alumno

Aguilar Ramirez Carlos Francisco

Arista Romero Juan Ismael

Jiménez Flores Luis Arturo

Vazquez Martin Marlene Gabriela

Profesor

Daniel Jiménez Alcantar

Materia

Análisis de Series de Tiempo

Grupo

6AV1

Practica 1

Explica cómo se aplica la metodología BOX-Jenkins en su serie de tiempo.



Practica 1.1 Naturaleza de las series de tiempo

Instrucciones

1.- Elija un Dataset de su elección, deberá acondicionarla de tal manera que pueda realizar el análisis de una serie de tiempo. Desarrolle un reporte técnico que permita observar el trabajo en los siguientes puntos:

- Construir gráfico de la serie de tiempo con los datos del dataset. Explicar el comportamiento.
- Hallar promedio, media, mediana, moda, desviación estándar y varianza.
- Identificar granularidad, ciclos, tendencia, estacionalidad, máximo, mínimos, picos, valles, razones de crecimiento y razones de reducción.
- Explica cómo se aplica la metodología BOX-Jenkins en su serie de tiempo.

Introducción

En esta práctica se analizará una serie de tiempo utilizando un dataset seleccionado y se realizará un estudio detallado del comportamiento de la serie. Además, se aplicarán conceptos estadísticos básicos y se explicará la metodología BOX-Jenkins para su aplicación en la serie de tiempo.

Dataset Seleccionado

El dataset utilizado para esta práctica contiene datos históricos del precio de las acciones de Walmart Inc. (WMT) desde el 25 de agosto de 1972 hasta el 21 de febrero de 2025. Se han realizado las siguientes modificaciones para acondicionar los datos y poder llevar a cabo el análisis de la serie de tiempo:

1. Conversión de la columna 'date' a tipo datetime con UTC.
2. Creación de un dataframe limpio que contenga solo las columnas 'date' y 'adj_close'.
3. Eliminación de la hora, dejando solo la fecha.
4. Verificación de valores faltantes y ordenamiento cronológico de los registros.

```
# Convertir a datetime con UTC para asegurar que todo esté correcto
df['date'] = pd.to_datetime(df['date'], utc=True)

# Crear dataframe limpio con fecha y adj_close únicamente
walmart_clean_df = df[['date', 'adj_close']].copy()

# Eliminar la hora, dejando solo la fecha
walmart_clean_df['date'] = walmart_clean_df['date'].dt.date

# Verificar el resultado
print(walmart_clean_df.head())
```

	date	adj_close
0	1972-08-25	0.011639
1	1972-08-28	0.011595
2	1972-08-29	0.011463
3	1972-08-30	0.011463
4	1972-08-31	0.011286

Figura 1. Transformación del conjunto de datos



Gráficos de la Serie de Tiempo

Se ha construido un gráfico de la serie de tiempo a partir de los datos del dataset, lo cual permite observar el comportamiento general de la serie.

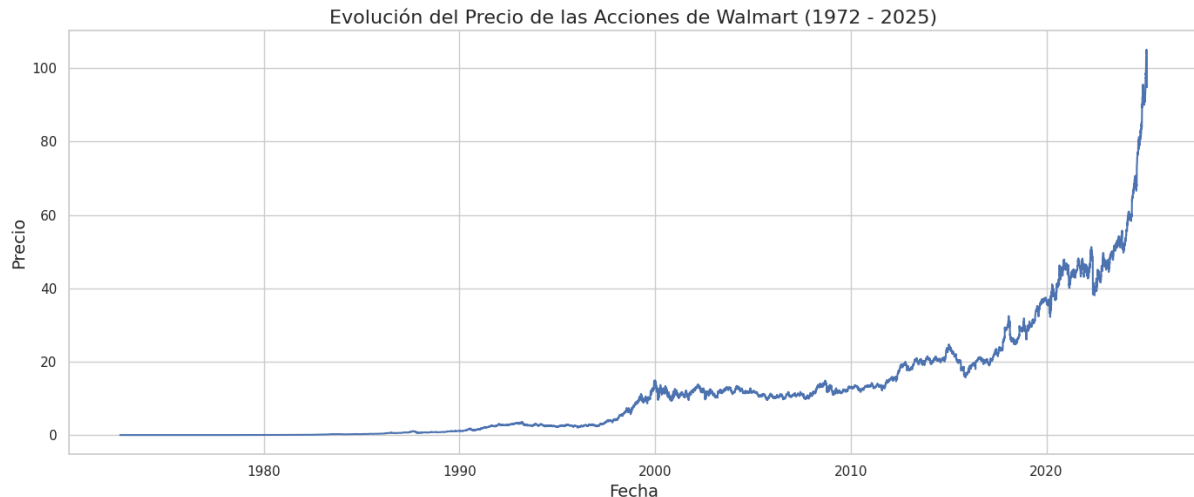


Gráfico 1. Comportamiento de la Serie de Tiempo

- **Observaciones:**
 - En el gráfico se observa una tendencia alcista general.
 - Se identifican patrones de estacionalidad y ciclos de crecimiento a largo plazo.
 - Los puntos máximos se alcanzan en los últimos años.
 - Los puntos mínimos se observan en los primeros años de la serie.

Cálculos Estadísticos

Se calcularon las siguientes medidas estadísticas para analizar el comportamiento general de la serie de tiempo:

- **Promedio:** 11.9461
- **Mediana:** 7.3507
- **Moda:** 0.0069
- **Desviación Estándar:** 15.8654
- **Varianza:** 251.7129

Estas medidas proporcionan una visión general sobre la distribución de los datos y su dispersión.

```
# Calcular estadísticas básicas
mean_value = walmart_clean_df['adj_close'].mean() # Promedio
median_value = walmart_clean_df['adj_close'].median() # Mediana
mode_value = walmart_clean_df['adj_close'].mode()[0] # Moda (puede haber múltipl
std_dev = walmart_clean_df['adj_close'].std() # Desviación estándar
variance = walmart_clean_df['adj_close'].var() # Varianza

# Resultados
stats_results = {
    "Promedio (Media)": mean_value,
    "Mediana": median_value,
    "Moda": mode_value,
    "Desviación Estándar": std_dev,
    "Varianza": variance
}

stats_results

Python

{'Promedio (Media)': 11.946184229469727,
 'Mediana': 7.350796699523926,
 'Moda': 0.0068982178345322,
 'Desviación Estándar': 15.865461507019829,
 'Varianza': 251.71286883072787}
```

Figura 2. Código utilizado para obtener la estadística de los datos



Análisis Detallado

1. **Granularidad:** La mayoría de los registros tienen una granularidad diaria, con observaciones de 1 día de diferencia. También se observan saltos típicos en fines de semana y días feriados.
2. **Ciclos:** Se identificaron ciclos que se repiten anualmente.
3. **Tendencia:** La tendencia observada en la serie es alcista.
4. **Estacionalidad:** Se presenta un patrón estacional anual, lo que indica que ciertos patrones se repiten cada año.
5. **Máximos y Mínimos:**
 - Máximo absoluto: \$105.05 el 13 de febrero de 2025.
 - Mínimo absoluto: \$0.0028 el 10 de diciembre de 1974.
6. **Picos y Valles:**
 - Número de picos locales identificados: 54.
 - Número de valles locales identificados: 52.

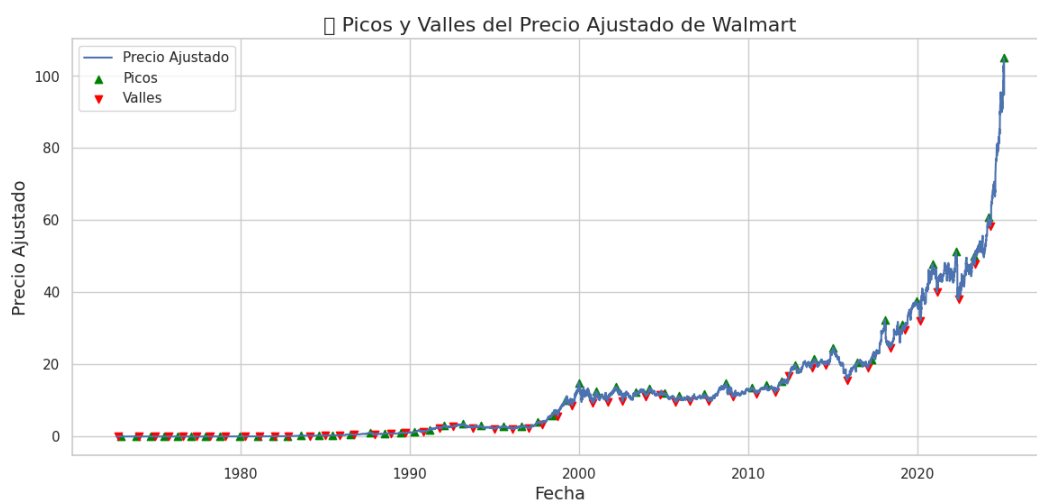


Gráfico 2. Picos y Valles de la Serie de Tiempo

Análisis de Crecimiento y Reducción

Se calculó el crecimiento anual del precio ajustado, mostrando un comportamiento variable a lo largo de los años.

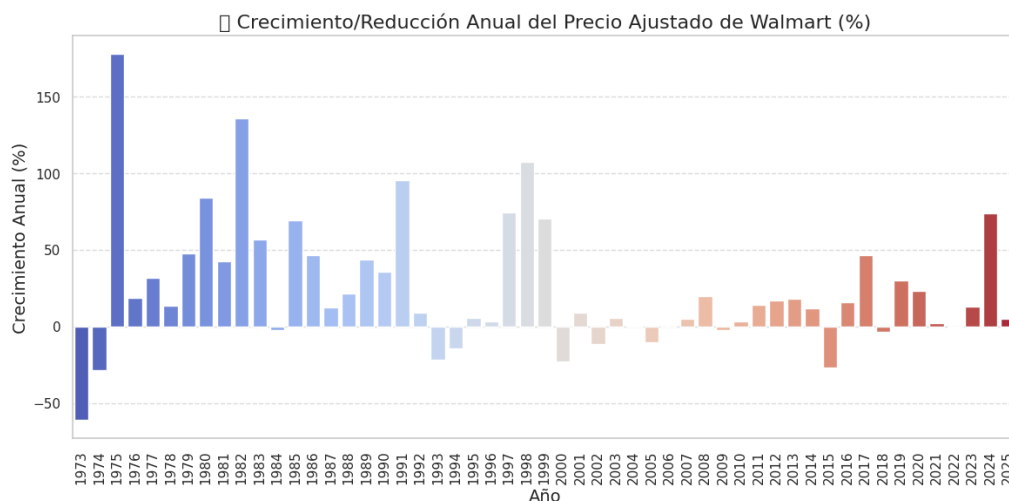


Gráfico 2. Crecimiento/Reducción Anual del Precio Ajustado de Walmart



Observaciones:

- Se observan años con crecimientos significativos, especialmente en los primeros años de la serie.
- Algunos años presentan una disminución significativa del precio ajustado.

Análisis de Tendencia

Se realizó un análisis de la tendencia general utilizando un promedio móvil anual (window=365 días):

- El promedio móvil muestra un crecimiento constante en los últimos años.

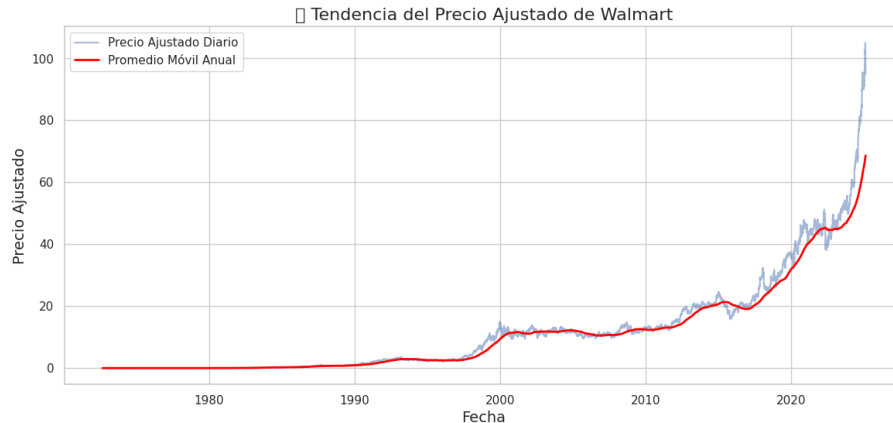


Gráfico 3. Análisis de Tendencia del Precio Ajustado de Walmart

Descomposición Estacional

Se realizó un análisis de descomposición estacional de la serie temporal, separando los componentes de tendencia, estacionalidad y residuos.

Descomposición Estacional del Precio Ajustado de Walmart

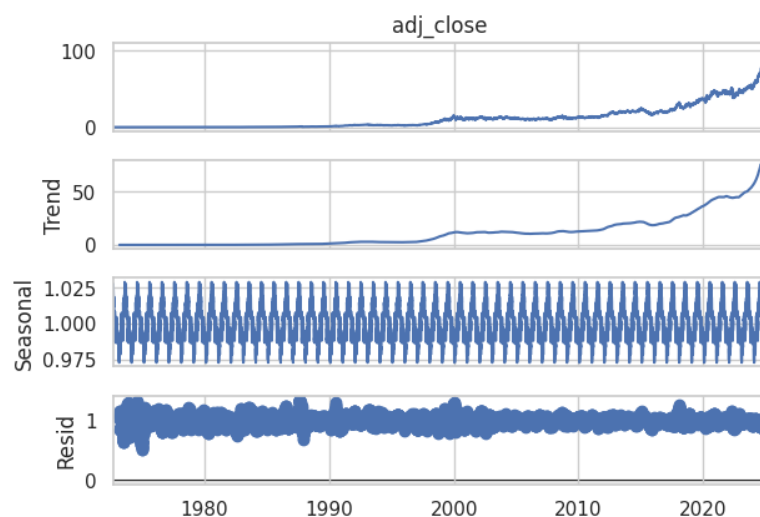


Gráfico 4. Descomposición Estacional del Precio Ajustado de Walmart

- **Componente de tendencia:** Refleja un crecimiento sostenido.
- **Componente estacional:** Patrón que se repite anualmente.
- **Componente de residuos:** Variaciones no explicadas por los patrones anteriores.

Metodología BOX-Jenkins



Figura 3. Proceso metodológico de Box-Jenkins

La metodología Box-Jenkins, también conocida como enfoque Box-Jenkins, es un método sistemático desarrollado en la década de 1970 por los estadísticos George E.P. Box y Gwilym M. Jenkins. Se utiliza principalmente para pronosticar y comprender datos dependientes del tiempo, enfatizando la selección y validación de modelos para representar con precisión los procesos de generación de datos en series temporales.

Se divide generalmente en tres etapas y en ocasiones cuenta con una cuarta:

1. **Identificación del modelo:** Se utilizan funciones de autocorrelación y autocorrelación parcial para determinar el orden apropiado del modelo ARIMA.
2. **Estimación de parámetros:** Se ajusta el modelo seleccionado a la serie temporal observada.
3. **Diagnóstico del modelo:** Se valida el modelo para asegurar que representa con precisión el proceso de generación de datos.
4. **Uso (pronóstico):** Esta etapa se trata de realizar pronósticos de la serie en el futuro.

En la realización de esta práctica simplemente se hizo un análisis exploratorio de los datos donde únicamente aplicamos la primera parte de la metodología de Box-Jenkins, la cual se trata de identificar el modelo. Básicamente se analizan los datos para identificar que modelo sería más adecuado para la serie de tiempo.

De la fase 1 lo que hicimos fue:

- Detectar si es estacionaria o no (en media, varianza y autocorrelaciones).
- Identificar posibles transformaciones (log, diferencias, etc.) para volverla estacionaria.
- Distinguir si hay estacionalidad y de qué tipo.
- Explorar la presencia de tendencias, picos, valles, y comportamientos cíclicos.