

Análisis Descriptivo De la Serie de Tiempo del Café

Angel Granados, Wilson Jerez, Santiago Montejo

Junio 2024

Nuestro objetivo en este análisis es determinar si los precios del café a lo largo del tiempo son estacionarios o no, utilizando la prueba de Dickey-Fuller. Para lograr esto, exploraremos los datos de precios del café desde 1980 hasta 2023, prestando especial atención al comportamiento en el año 2000. Identificaremos las bajas significativas en los precios según los registros de la bolsa de valores y evaluaremos la confiabilidad de la inversión en este sector. Esto lo haremos analizando la mayor caída en los precios del café durante este período y explorando las oportunidades que podrían surgir en futuras caídas.

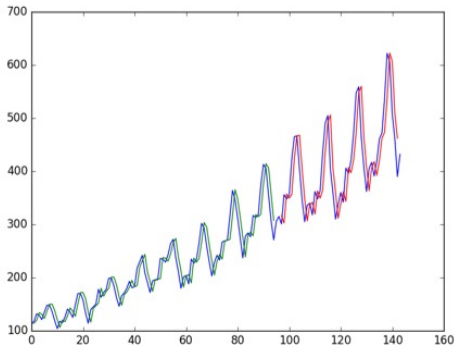
Utilizando conceptos estadísticos como las series de tiempo y la prueba de Dickey-Fuller, derivada de la teoría de series de tiempo, determinaremos si los precios del café son estacionarios. La hipótesis nula (H_0) en la prueba de Dickey-Fuller será que los precios no son estacionarios, mientras que la hipótesis alternativa (H_1) será que los precios son estacionarios.

Adicionalmente, evaluaremos la estacionalidad presente en los datos, que se manifiesta en patrones recurrentes a lo largo del tiempo.

A través de este análisis, plantearemos y probaremos la hipótesis nula de no estacionariedad y confirmaremos la presencia de estacionalidad.

Serie de Tiempo

Una serie de tiempo es una secuencia de datos recolectados y ordenados cronológicamente, generalmente a intervalos regulares, como diarios, mensuales, trimestrales o anuales. Estas series se utilizan para analizar cómo una variable cambia a lo largo del tiempo y son fundamentales en campos como la economía, la meteorología, la ingeniería, y muchas otras disciplinas.



Conceptos Importantes

- 1 Estacionariedad:
- 2 Tendencia
- 3 Estacionalidad
- 4 Ruido
- 5 Prueba de Dickey-Fuller
- 6 Modelos ARIMA (Autoregressive Integrated Moving Average)

Descripcion de la Base de Datos

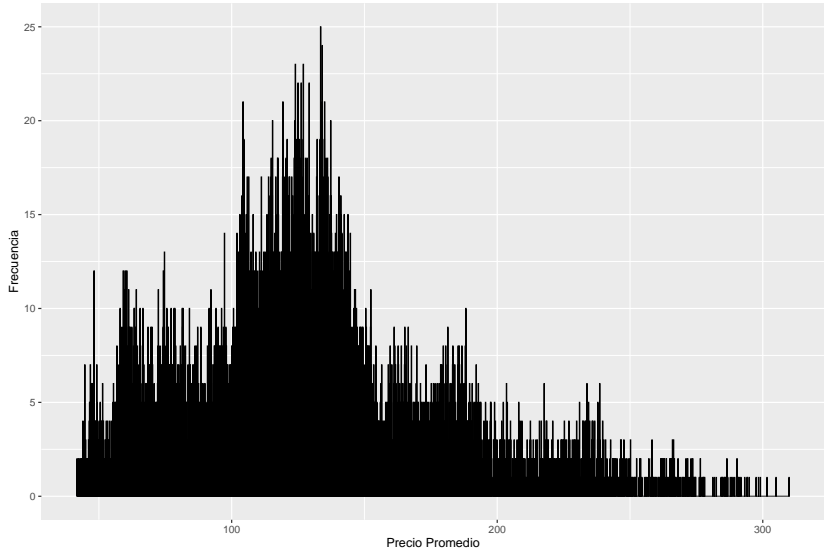
La base de datos utilizada para este análisis contiene información sobre los precios del café, obtenida del sitio web Investing. Los datos abarcan el período comprendido entre los años 1980 y 2023. Son datos diarios que no abordan los fines de semana (Sabado y Domingo) y se consideran dos Variables que son el tiempo (Expresado en Dias, Meses y años) y el precio promedio diario.

Descripcion de la Base de Datos

```
## 'data.frame':    11152 obs. of  8 variables:
## $ Fecha          : Date, format: "1979-12-27" "1979-12-28" .
## $ Último         : num  183 182 182 178 172 ...
## $ Apertura       : num  188 183 182 182 176 ...
## $ Máximo         : num  188 186 183 182 176 ...
## $ Mínimo         : num  183 181 181 176 170 ...
## $ Vol.           : chr   "" "" "" "1,14K" ...
## $ X..var.        : chr   "-2,23%" "-0,90%" "0,12%" "-2,29%" .
## $ precio_promedio: num  186 183 182 179 174 ...
```

Descripcion de la Base de Datos

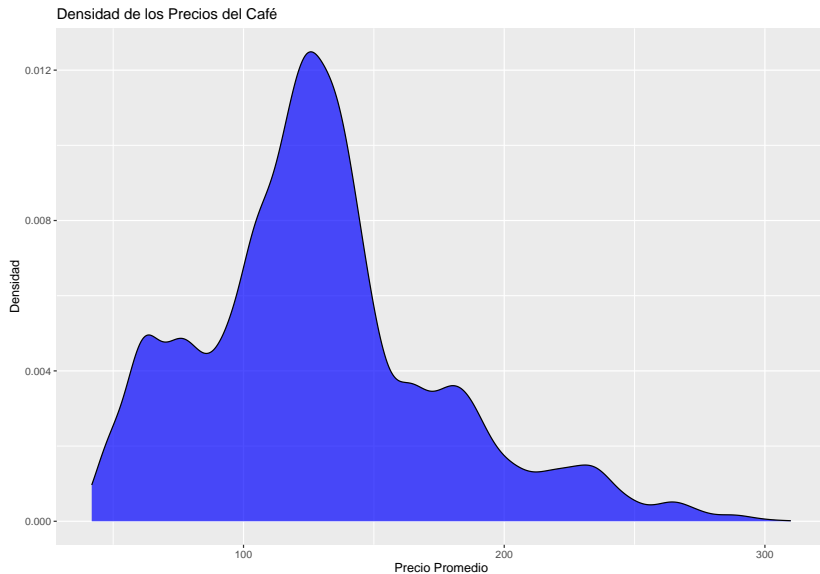
Distribución de los Precios del Café



Descripcion de la Base de Datos

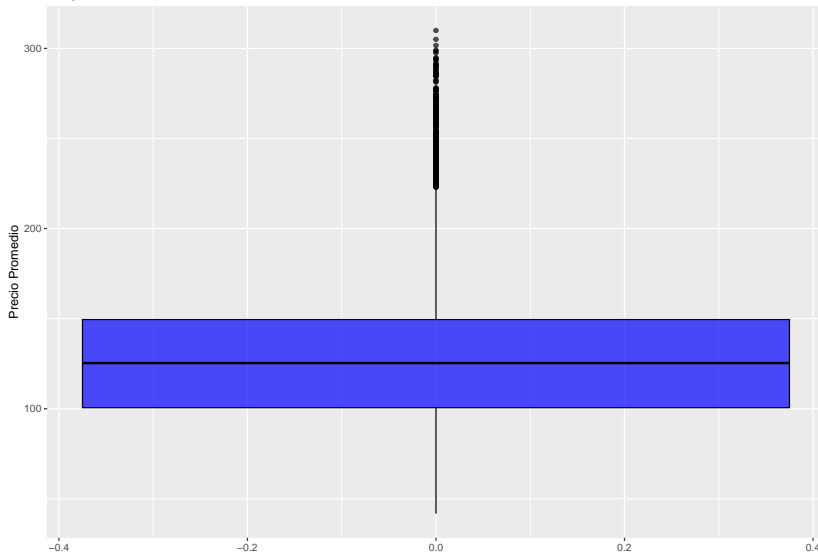
- Forma de la Distribución: Observando la forma del histograma, se puede identificar si los precios del café están distribuidos simétricamente, sesgados a la izquierda.
- Centralidad: El centro de la distribución se puede estimar visualmente observando dónde se concentran los valores más frecuentes.
- Dispersión: La dispersión de los datos puede ser evaluada observando la amplitud de los bins en el histograma y también considerando la varianza o el rango intercuartil (IQR). Un sesgo hacia la izquierda típicamente implica una menor dispersión hacia los valores más altos.

Descripcion de la Base de Datos



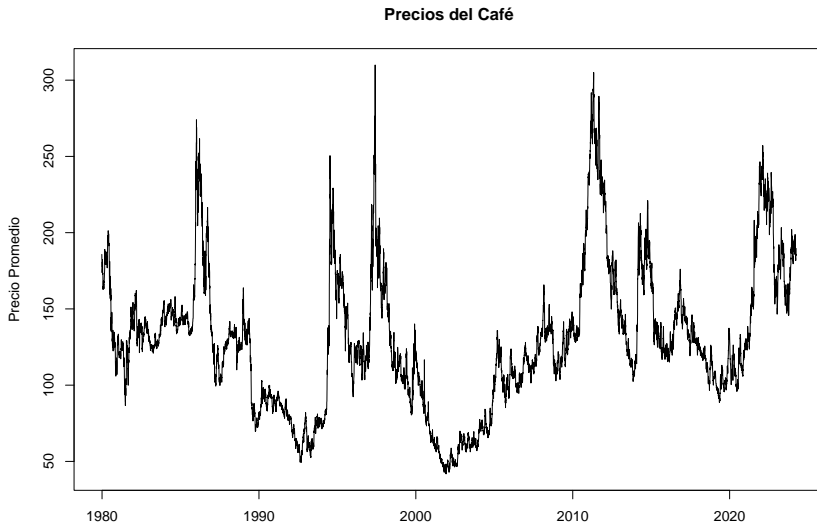
Descripcion de la Base de Datos

Diagrama de Caja de los Precios del Café



Serie de tiempo de todos los datos

```
KCN4 <- ts(datosHist$precio_promedio,freq=12)  
plot(datosHist$Fecha, KCN4, type = "l", xlab = "Fecha",  
      ylab = "Precio Promedio", main = "Precios del Café")
```

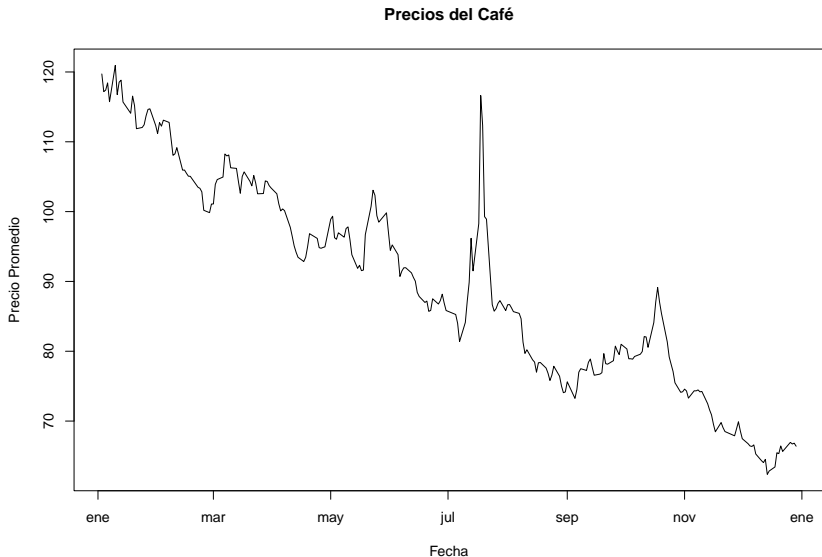


Serie de tiempo por año

```
# Crear un bucle for para particionar los datos por año
for (i in 1980:2023) {
  start_date <- as.Date(paste(i, "-01-01", sep = ""))
  end_date <- as.Date(paste(i + 1, "-01-01", sep = ""))

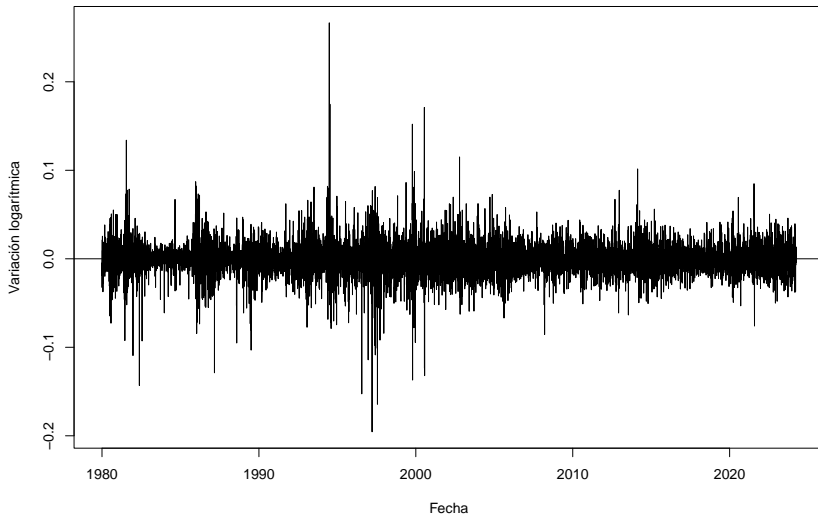
  datosporAño <- subset(datosHist,
                        Fecha >= start_date
                        & Fecha < end_date)
  KCN4porAño <- ts(datosporAño$precio_promedio,
                  frequency = 12)
  assign(paste0("Datos_Hist_", i), datosporAño)
  assign(paste0("KCN4_", i), KCN4porAño)
}
```

```
plot(Datos_Hist_2000$Fecha, KCN4_2000, type = "l",  
     xlab = "Fecha",  
     ylab = "Precio Promedio", main = "Precios del Café")
```

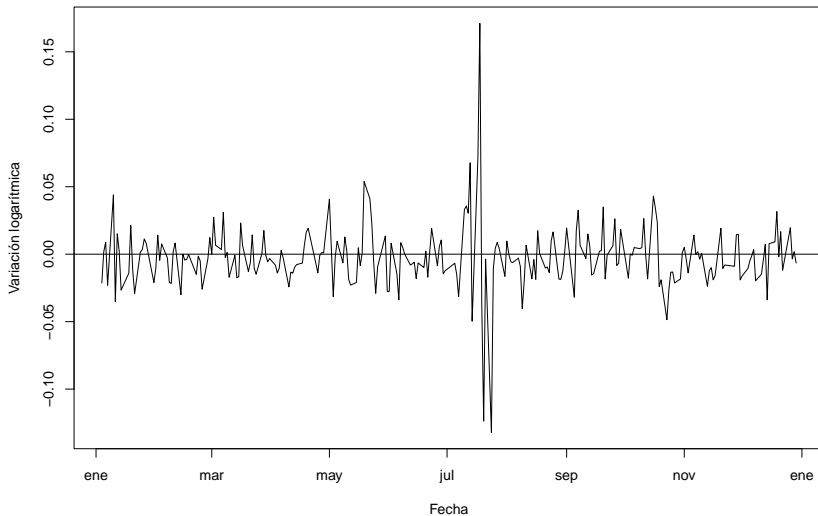


Cálculo de la variación de las series de tiempo

Variación en cada año

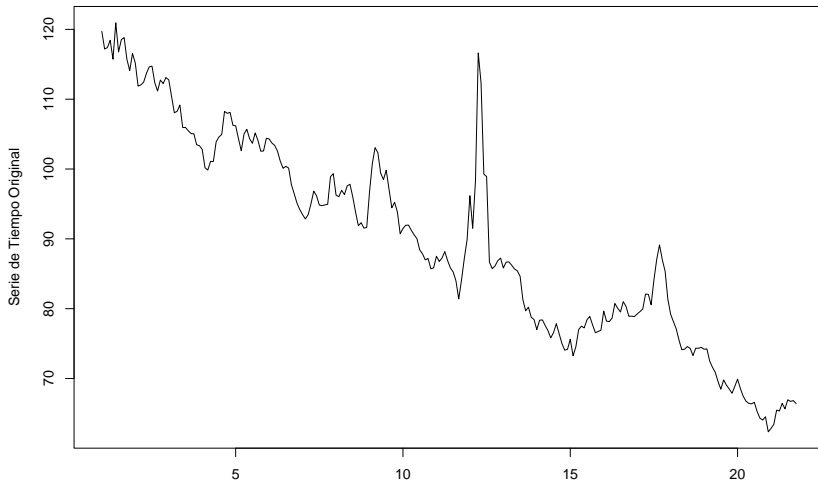


Variación en el año 2000

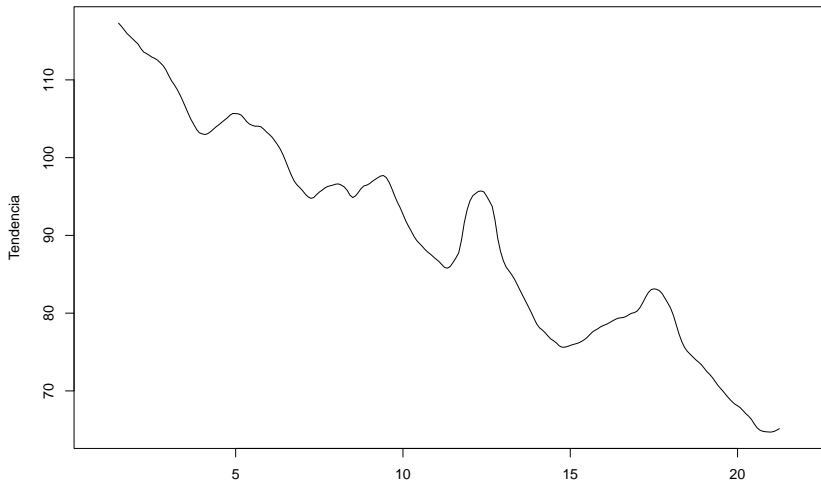


Descomposición de la serie de tiempo

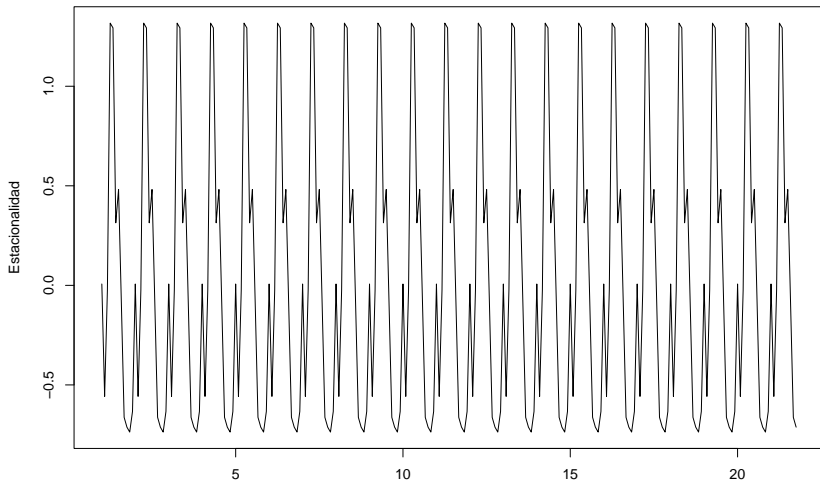
```
decomposition_2000 <- decompose(KCN4_2000,  
                                type = "additive")  
plot(decomposition_2000$x,  
     ylab = "Serie de Tiempo Original", xlab = "")
```



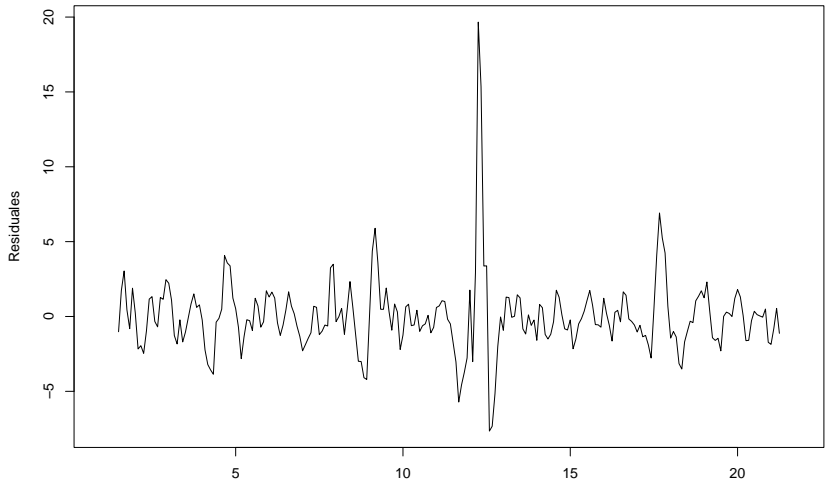
```
plot(decomposition_2000$trend,  
     ylab = "Tendencia", xlab = "")
```



```
plot(decomposition_2000$seasonal,  
     ylab = "Estacionalidad", xlab = "")
```

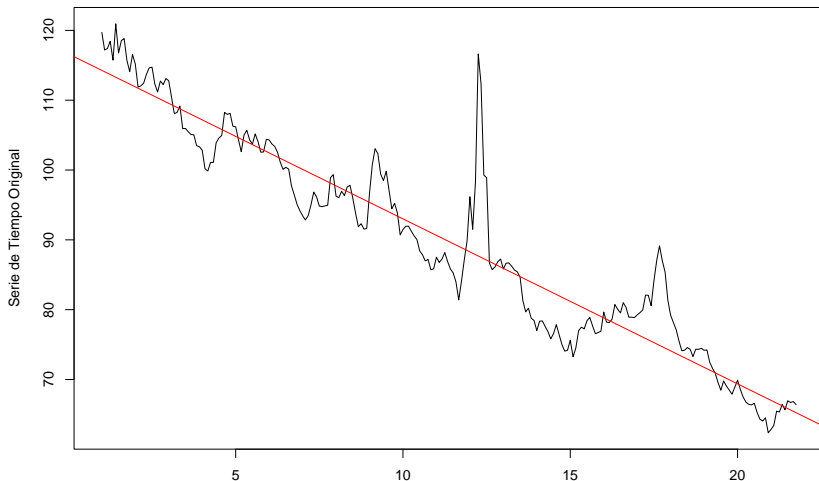


```
plot(decomposition_2000$random,  
     ylab = "Residuales", xlab = "")
```

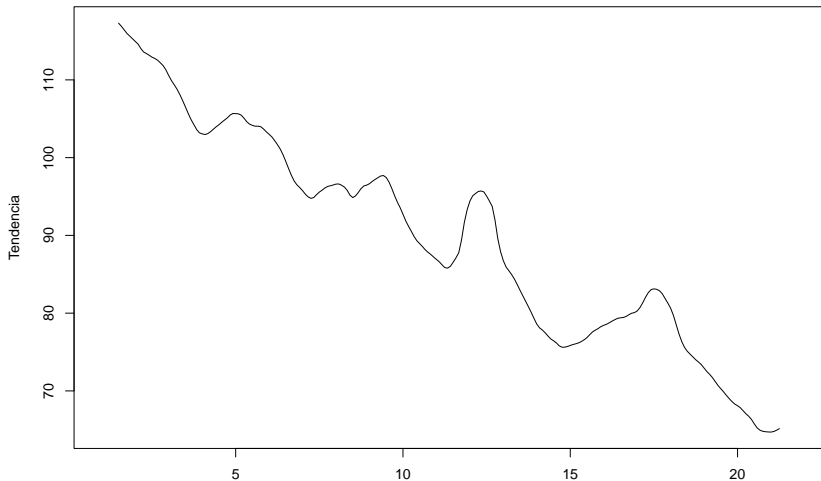


Análisis de tendencia

```
plot(decomposition_2000$x,  
      ylab = "Serie de Tiempo Original", xlab = "")  
abline(lm(decomposition_2000$x  
          ~ time(decomposition_2000$x)), col = "red")
```



```
plot(decomposition_2000$trend,  
     ylab = "Tendencia", xlab = "")
```



H_0 : No es estacionaria.

H_1 : Es estacionaria

```
##                                     Resumen
## 1  F-statistic: 8.203 on 3 and 244 DF,  p-value: 3.185e-05
## 2
## 3
## 4      Value of test-statistic is: -4.6517 7.7753 10.8352
## 5
## 6      Critical values for test statistics:
## 7                                     1pct  5pct 10pct
## 8      tau3 -3.99 -3.43 -3.13
## 9      phi2  6.22  4.75  4.07
## 10     phi3  8.43  6.49  5.47
## 11
```

Conclusión

Basado en el resumen del test de Dickey-Fuller aumentado (ADF) para verificar la estacionariedad de la serie de tiempo de los precios del café (KCN4_2000)

Planteamiento de la Hipótesis del Test de Dickey-Fuller:

- Hipótesis Nula (H_0): La serie de tiempo NO es estacionaria.
- Hipótesis Alternativa (H_1): La serie de tiempo es estacionaria.

Resultado del Test

El valor p obtenido del test de Dickey-Fuller Aumentado es $3.185e-05$, que es significativamente menor que cualquier nivel de significancia comúnmente utilizado (como 1%, 5% o 10%). Esto indica una fuerte evidencia en contra de la hipótesis nula (H_0) de que la serie de tiempo es estacionaria.

Esta conclusión sirve para el análisis y la predicción de los precios del café, ya que una serie estacionaria permite utilizar modelos adecuados de series temporales, como los modelos ARIMA, para realizar pronósticos más precisos y confiables.

- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2020). Probabilidad y estadística para ingeniería y ciencias (9.a ed., L. E. Pineda Ayala, Trad.). Pearson.
- Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2010). Estadística matemática con aplicaciones (7.a ed., J. H. Romo Muñoz, Trad.). Cengage Learning.
- Brockwell, P. J., & Davis, R. A. (2002). Introduction to Time Series and Forecasting (2nd ed.). Springer.