

Proyecto Final

N. Pérez - Medina

2025-05-28

Introducción

El presente análisis se centra en estudiar la evolución temporal de dos aspectos del cine: las calificaciones críticas y la duración de las películas. Se busca identificar tendencias, patrones y posibles cambios estructurales a lo largo del tiempo, con el fin de proponer un modelo que describa correctamente la variabilidad de los datos y permita predecir tanto las calificaciones como la duración de películas en fechas futuras.

Contextualización de la base de datos

La base de datos `16k_Movies` contiene información sobre más de 16,000 películas estrenadas entre los años 1910 y 2024. Entre las variables disponibles se incluyen el título de la película, fecha de estreno, descripción, calificación promedio en *Metacritic*, número de personas que votaron, así como datos sobre la dirección, guión, duración y géneros asociados a cada película.

Su estructura permite aplicar modelos estadísticos y técnicas de análisis para buscar posibles patrones en el comportamiento de las variables a lo largo del tiempo.

Construcción de series de tiempo utilizadas

Serie 1: Duración de las películas de 1910 a 2024

```
# Cargamos la base de datos
pelis <- read.csv("16k_Movies.csv")

# Primer subconjunto: título, fecha de estreno y duración
pelis_duracion <- pelis[, c("Title", "Duration", "Release.Date")]
head(pelis_duracion, 1)

##           Title Duration Release.Date
## 1 Dekalog (1988) 9 h 32 m Mar 22, 1996

# Creamos columna con formato correcto de fecha
pelis_duracion$Fecha <- as.Date(pelis_duracion$Release.Date, format = "%b %d, %Y")
head(pelis_duracion, 1)

##           Title Duration Release.Date      Fecha
## 1 Dekalog (1988) 9 h 32 m Mar 22, 1996 1996-03-22

# Función para tener la duracion total en minutos
convertir_a_minutos <- function(duracion) {
  if (is.na(duracion) || duracion == "") return(0)

  # Si no contiene ni horas ni minutos válidos
  if (!grepl("\\d+\\s*h|\\d+\\s*m", duracion)) return(0)

  # Extraer horas
  if (grepl("\\d+\\s*h", duracion)) {
    h <- as.numeric(sub(".*?(\\d+)\\s*h.*", "\\1", duracion))
  } else {
    h <- 0
  }

  # Extraer minutos
  if (grepl("\\d+\\s*m", duracion)) {
    m <- as.numeric(sub(".*?(\\d+)\\s*m.*", "\\1", duracion))
  } else {
    m <- 0
  }

  return(h * 60 + m)
}

# Aplicamos la función a cada fila y creamos una nueva columna con la duración correcta
pelis_duracion$Duracion <- sapply(pelis_duracion$Duration, convertir_a_minutos)
head(pelis_duracion, 1)

##           Title Duration Release.Date      Fecha Duracion
## 1 Dekalog (1988) 9 h 32 m Mar 22, 1996 1996-03-22      572
```

```

# Extraemos los datos
pelis_duracion$Año <- as.numeric(format(pelis_duracion$Fecha, "%Y"))
pelis_duracion$Mes <- as.numeric(format(pelis_duracion$Fecha, "%m"))

# Agrupamos y ordenamos por año y mes
duracion_mensual <- aggregate(Duracion ~ Año + Mes, data = pelis_duracion, FUN = mean)
duracion_mensual <- duracion_mensual[order(duracion_mensual$Año, duracion_mensual$Mes), ]

# Lo hacemos serie de tiempo
pelis_duracion.ts <- ts(duracion_mensual$Duracion,
                        start = c(min(duracion_mensual$Año), min(duracion_mensual$Mes)),
                        frequency = 12)

```

Serie 2: Rating de las películas de 1910 a 2024

```
# Segundo subconjunto: título, calificación y fecha de estreno
pelis_rating <- pelis[, c("Title", "Rating", "Release.Date")]
head(pelis_rating, 1)
```

```
##           Title Rating Release.Date
## 1 Dekalog (1988)    7.4 Mar 22, 1996
```

```
# Creamos columna con formato correcto de fecha
pelis_rating$Fecha <- as.Date(pelis_rating$Release.Date, format = "%b %d, %Y")
head(pelis_rating, 1)
```

```
##           Title Duration Release.Date      Fecha Duracion  Año Mes
## 1 Dekalog (1988) 9 h 32 m Mar 22, 1996 1996-03-22      572 1996   3
```

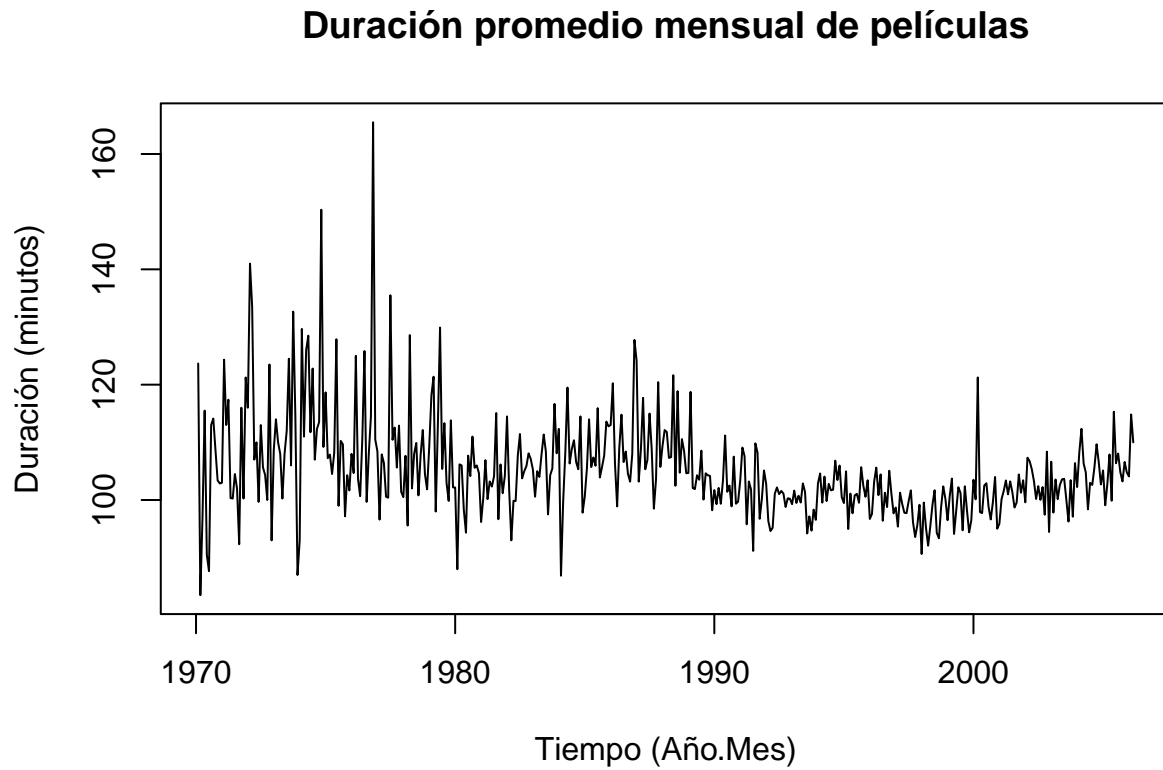
```
# Extraemos los datos
pelis_rating$Año <- as.numeric(format(pelis_rating$Fecha, "%Y"))
pelis_rating$Mes <- as.numeric(format(pelis_rating$Fecha, "%m"))

# Agrupamos y ordenamos por año y mes
rating_mensual <- aggregate(Rating ~ Año + Mes, data = pelis_rating, FUN = mean)
rating_mensual <- rating_mensual[order(rating_mensual$Año, rating_mensual$Mes), ]

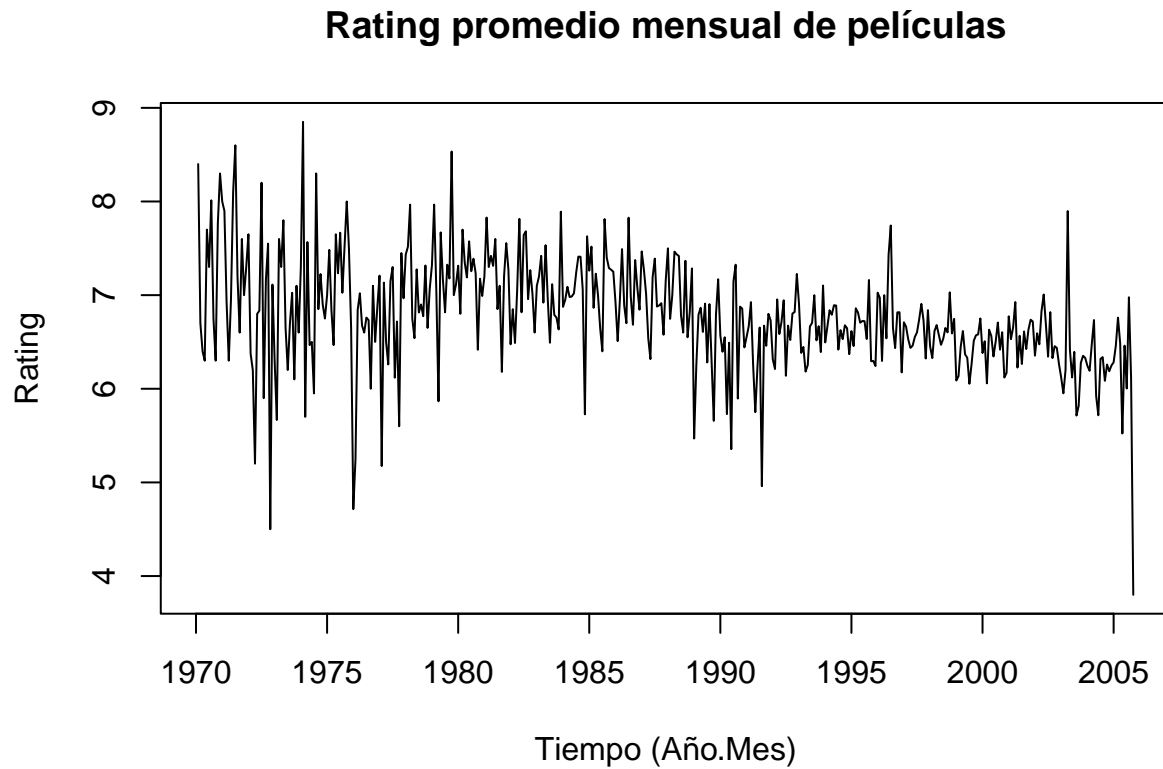
# Lo hacemos serie de tiempo
pelis_rating.ts <- ts(rating_mensual$Rating,
                      start = c(min(rating_mensual$Año), min(rating_mensual$Mes)),
                      frequency = 12)
```

Series de tiempo resultantes

```
# Graficamos la serie 1
plot(pelis_duracion.ts,
     main = "Duración promedio mensual de películas",
     xlab = "Tiempo (Año.Mes)", ylab = "Duración (minutos)")
```

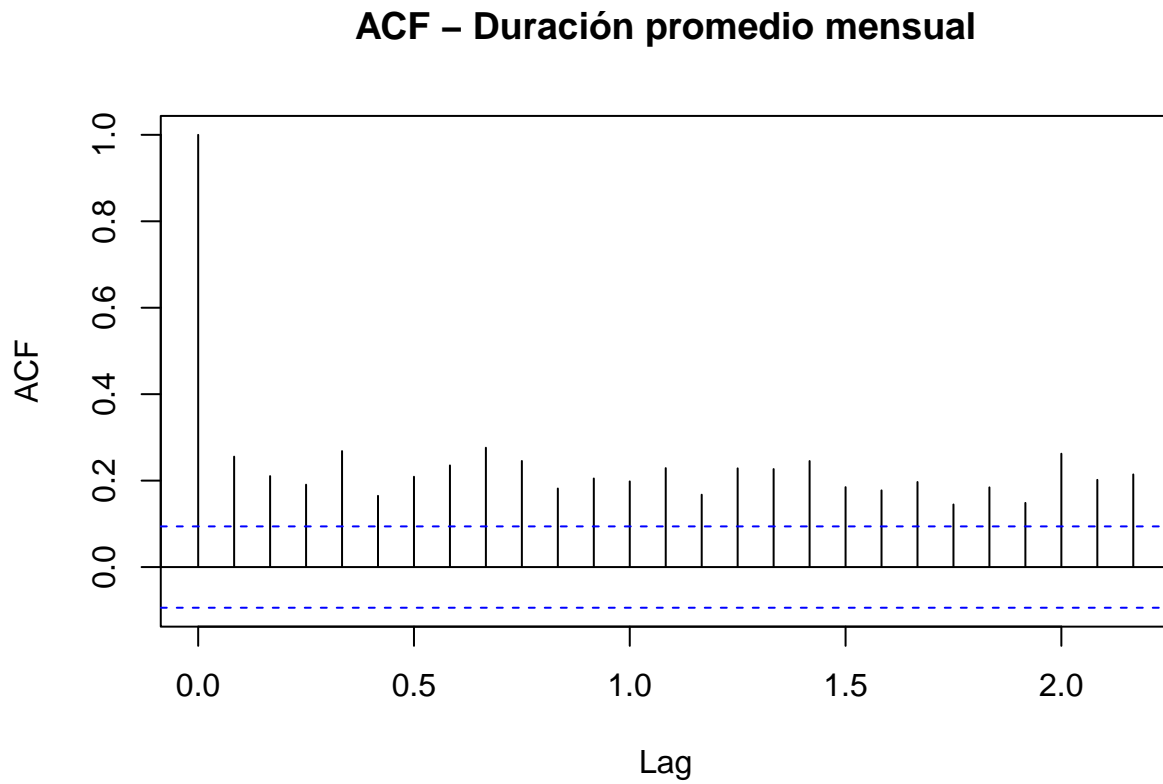


```
# Graficamos la serie 2
plot(pelis_rating.ts,
     main = "Rating promedio mensual de películas",
     xlab = "Tiempo (Año.Mes)", ylab = "Rating")
```



Modelos diseñados

```
acf(pelis_duracion.ts, main = "ACF - Duración promedio mensual")
```



```
pacf(pelis_duracion.ts, main = "PACF - Duración promedio mensual")
```

PACF – Duración promedio mensual

