DatasetGuidelines

-

2022/4/9

Introduction

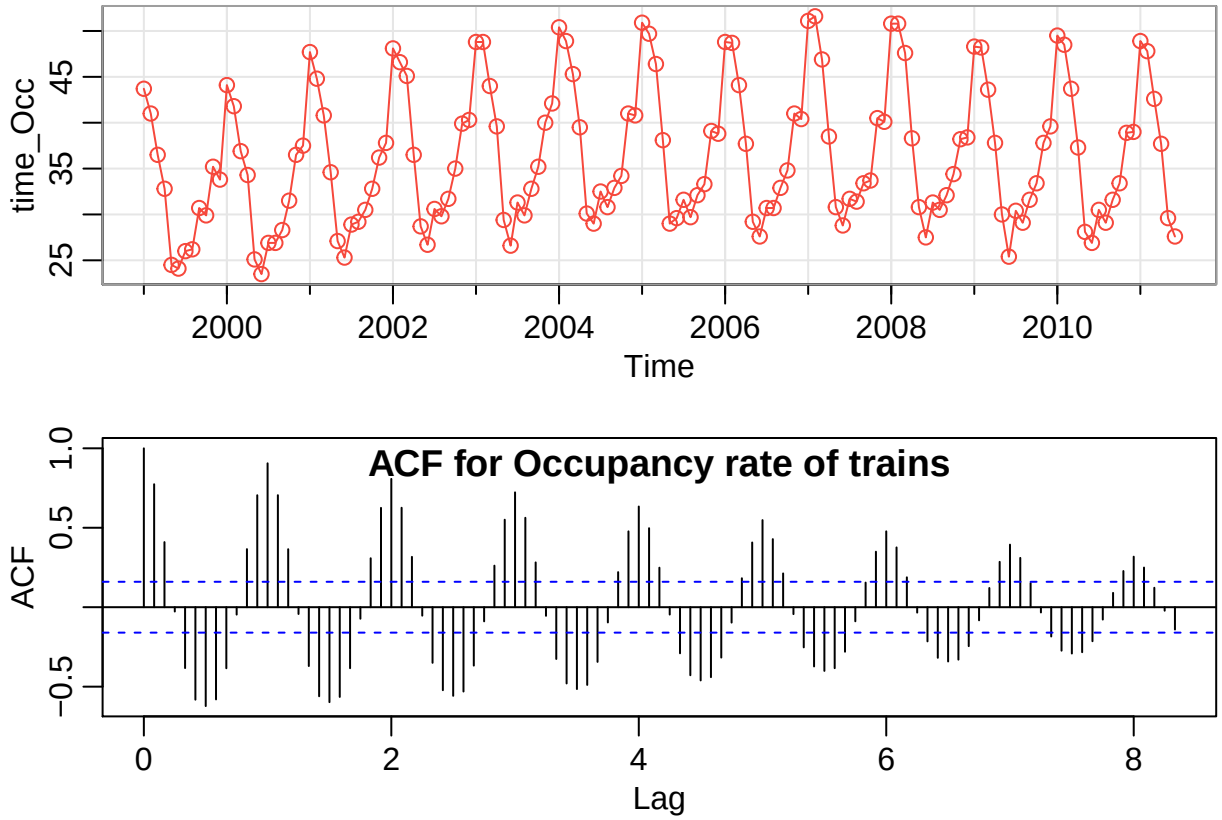
This data set is chosen from Kaggle,website: https://www.kaggle.com/datasets/gajjadarahul/predict-train-occupancy-time-series

This data set includes data ranges over the years from 1999 to 2011(monthly). It records the percentage of occupancy of the trains in that particular month.

We first import the data to Rstudio and convert the data to a time series.

```
## Rows: 150 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Time
## dbl (1): Total Occupancy rate (percent)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
par(mfrow = c(2,1))
tsplot(time_Occ,type = 'o',col = 2)
acf(time_Occ,lag.max = 100)
title("ACF for Occupancy rate of trains",line=-1)
```



The graph shows that variance is not constant and an obvious seasonal trend for this time series.The ACF shows that differencing is needed to make the series stable.We can apply Box-Cox method transformation and differencing to make the series stationary.This data set has 150 observations,which is big enough to perform time series analysis.