**DSBA/MBAD 6211 Assignment 4**

Due Date: 11:59pm @ 8/2/2020

**Data description**

Source: *Data Mining for Business Analytics* (R Edition 2017), by Galit Shmueli, Peter Bruce, Inbal Yahav, Nitin Patel, Kenneth Lichtendahl. (ISBN-10: 1118879368)

The dataset records Australian red wine sales data between 1980 and 1994.

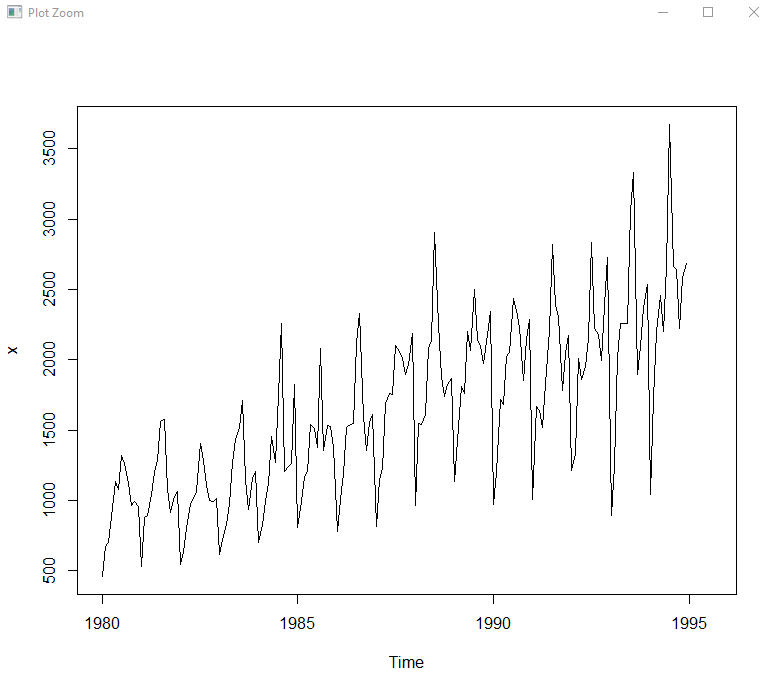
* Please use the last 2 years of data as the validation dataset (hold-out sample).
* Please run different forecasting models and compare their model performance.
  + Linear trend
  + Seasonality
  + Linear trend and seasonality
  + Simple exponential smoothing model

**Variable and model naming requirements:**

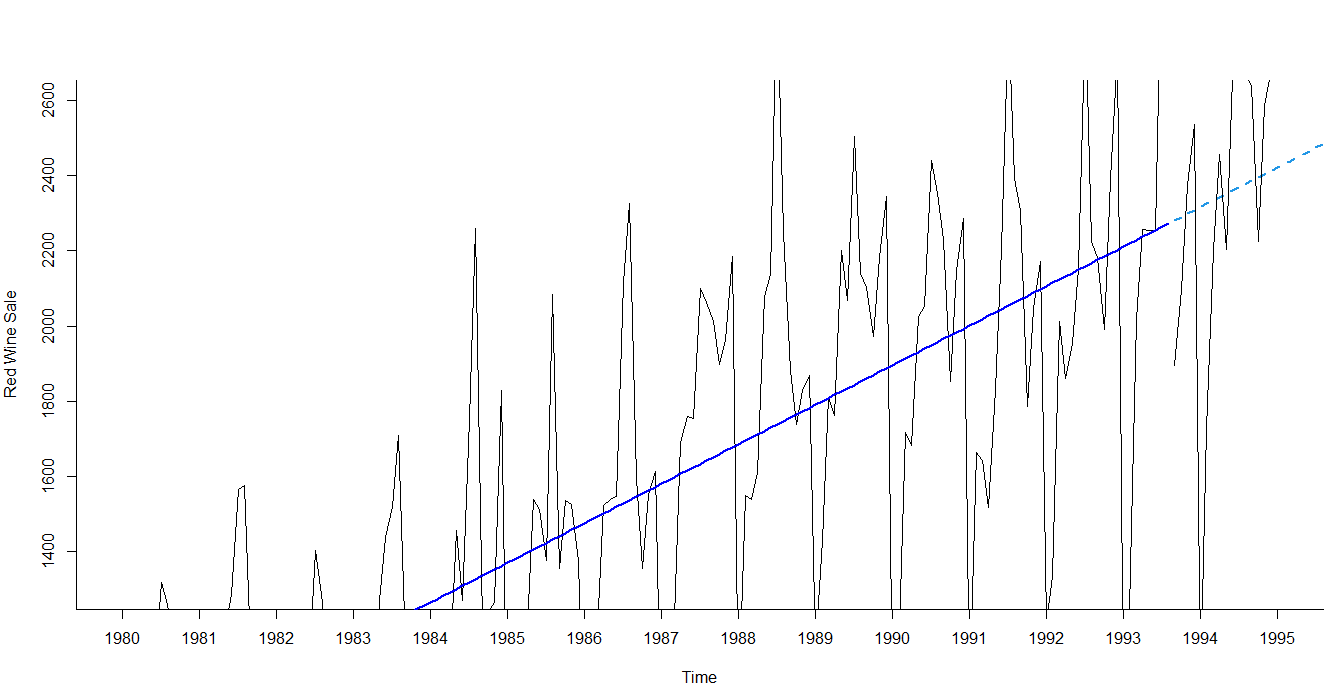
* + Please include your ***name initials*** to the data frame names as well as model names in your R coding.
  + Please instance, in my coding, I would name the data frames as ***dfKZ, dfKZ.train***, and ***dfKZ.valid.*** I would also name the models as ***train.lmKZ***, etc.

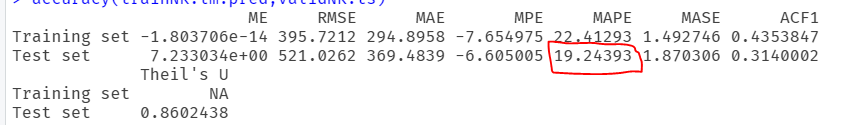
**Questions**

1. Please show the plot of the entire dataset.

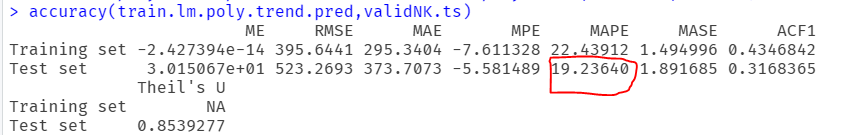


1. Please provide results for each model estimation
   1. Linear Trend:

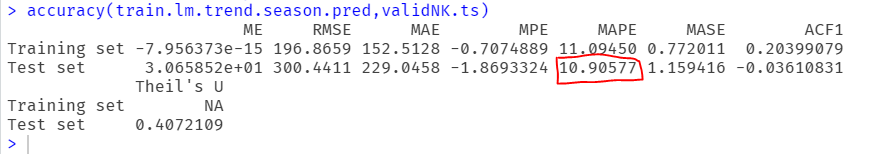




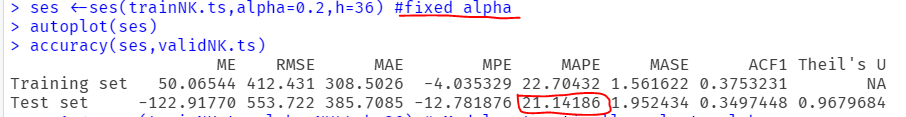
* 1. Poly Trend

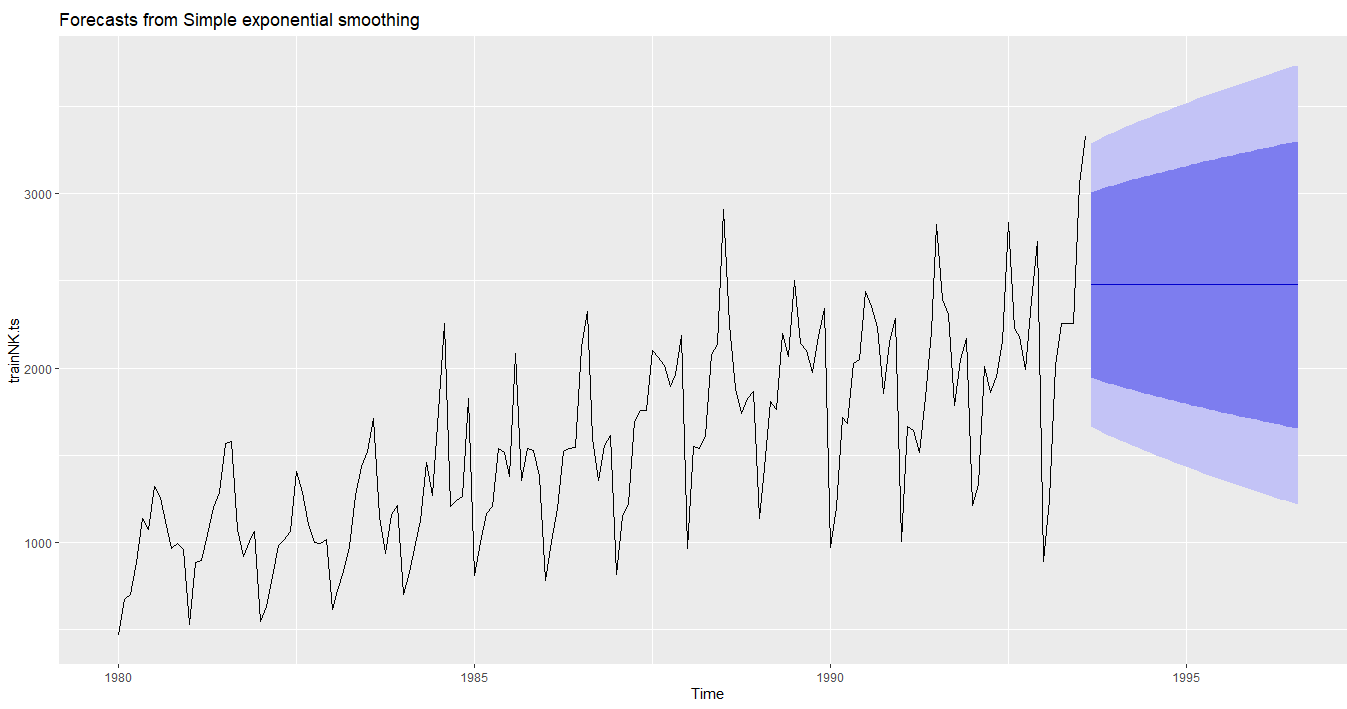


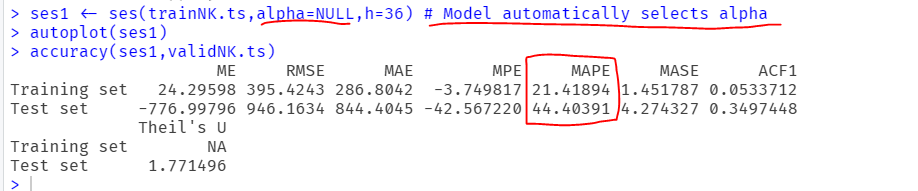
* 1. Linear trend and seasonality

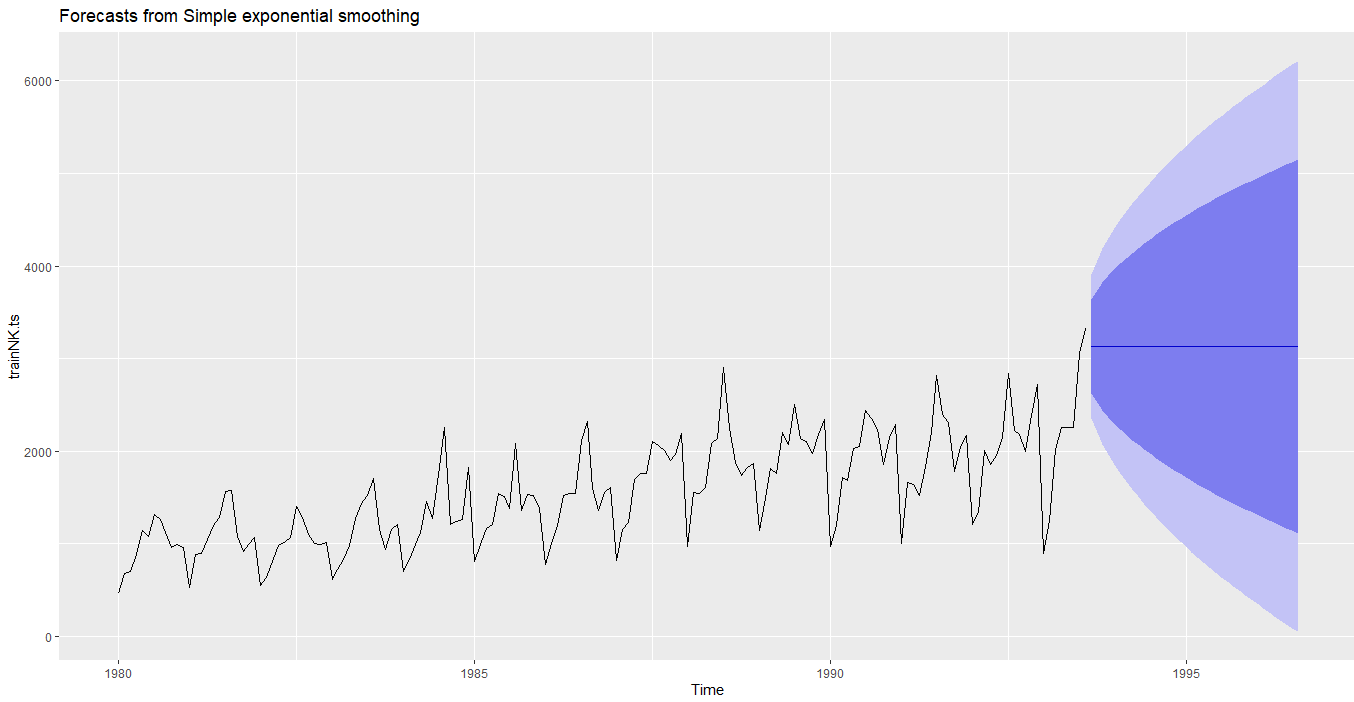


* 1. Simple exponential smoothing model









1. Which model is the best forecasting model? Please provide your evidence

MAPE is the accuracy measure for the time series model. Lower the MAPE better the model is the formula to check the model performance and based on the results **Linear trend and seasonality** has the lowest value so this model is best forecasting model.

1. Please copy and paste your R codes in your WORD submission.

#install.packages('forecast')

library(forecast)

getwd()

setwd("C:\\Users\\P2190101\\Desktop\\NK Personal\\NK Study\\UNCC\\6211 - Advanced BI\\Assignment 4") # Set working directory

data <- read.csv('AustralianWines.csv')

x<- ts(data$Red,

start=c(1980,1),

frequency = 12) # what is the cycle

x

plot(x)

AustralianWines.lm <- tslm(x~trend)

summary(AustralianWines.lm)

nValid <-24

nTrain <- length(x) -nValid

trainNK.ts <-window(x,start=c(1980,1),end=c(1980,nTrain))

validNK.ts <- window(x,start=c(1980,nTrain+1),end=c(1980,nTrain+nValid))

#----------- Linear Trend -------------------

trainNK.lm<- tslm(trainNK.ts~trend)

summary(trainNK.lm)

trainNK.lm.pred <- forecast(trainNK.lm, h=nValid,level=0) #level=0 - does not need to provide confidence interval

accuracy(trainNK.lm.pred,validNK.ts)

# Visualize the linear trend model

par(mfrow = c(1, 1)) # Create single graph

plot(trainNK.lm.pred, ylim = c(1300, 2600), ylab = "Red Wine Sale", xlab = "Time",

bty = "l", xaxt = "n", xlim = c(1980,1995),main = "", flty = 2)

axis(1, at = seq(1980,1995, 1), labels = format(seq(1980,1995, 1)))

lines(trainNK.lm.pred$fitted, lwd = 2, col = "blue")

lines(validNK.ts)

# Evaluate model performance

accuracy(trainNK.lm.pred,validNK.ts)

#------------------ Poly Trend -----------------

train.lm.poly.trend <- tslm(trainNK.ts~ trend+I(trend^2))

summary(train.lm.poly.trend)

train.lm.poly.trend.pred <- forecast(train.lm.poly.trend,h=nValid,level=0)

accuracy(train.lm.poly.trend.pred,validNK.ts)

#--------- lm with seasonality -----------------

train.lm.season <- tslm(trainNK.ts ~season)

summary(train.lm.season)

train.lm.trend.season <- tslm(trainNK.ts~trend+I(trend^2)+season)

summary(train.lm.trend.season)

train.lm.trend.season.pred <- forecast(train.lm.trend.season,h=nValid,level=0)

accuracy(train.lm.trend.season.pred,validNK.ts)

#----------- Moving Averages------------

library(zoo)

x

ma <- rollmean(x,k=12,align='right')

summary(ma)

ma

MAPE =mean(abs(ma-x)/x,na.rm=T)

MAPE

ses <-ses(trainNK.ts,alpha=0.2,h=36) #fixed alpha

autoplot(ses)

accuracy(ses,validNK.ts)

ses1 <- ses(trainNK.ts,alpha=NULL,h=36) # Model automatically selects alpha

autoplot(ses1)

accuracy(ses1,validNK.ts)