Question 4

The dataset here gives you the profit made in each quarter by Dr Coad selling second-hand books in his shop in New York City over the past 10 years (data from the first quarter in 2012 to the fourth quarter in 2021).

(a) Load the data into R and plot the time series. [5 marks]

(b) Fit a SARIMA model using the R function `auto.arima`. [5 marks]

(c) Forecast the time series for 2022. [5 marks]

(d) Write a paragraph (300 words maximum) commenting on the model fit and the forecast. [25 marks]

Include R code used to carry out the analysis and forecast. Please upload your answer below as a single pdf file.

Dataset assigned:
62,51,94,139,39,153,105,51,29,163,202,80,30,19,187,337,227,245,249,476,398,608,515,524,521,608,504,307,332,277,387,171,339,219,308,337,323,352,370,232

Answer 4

A) Load the data into R and plot the time series.

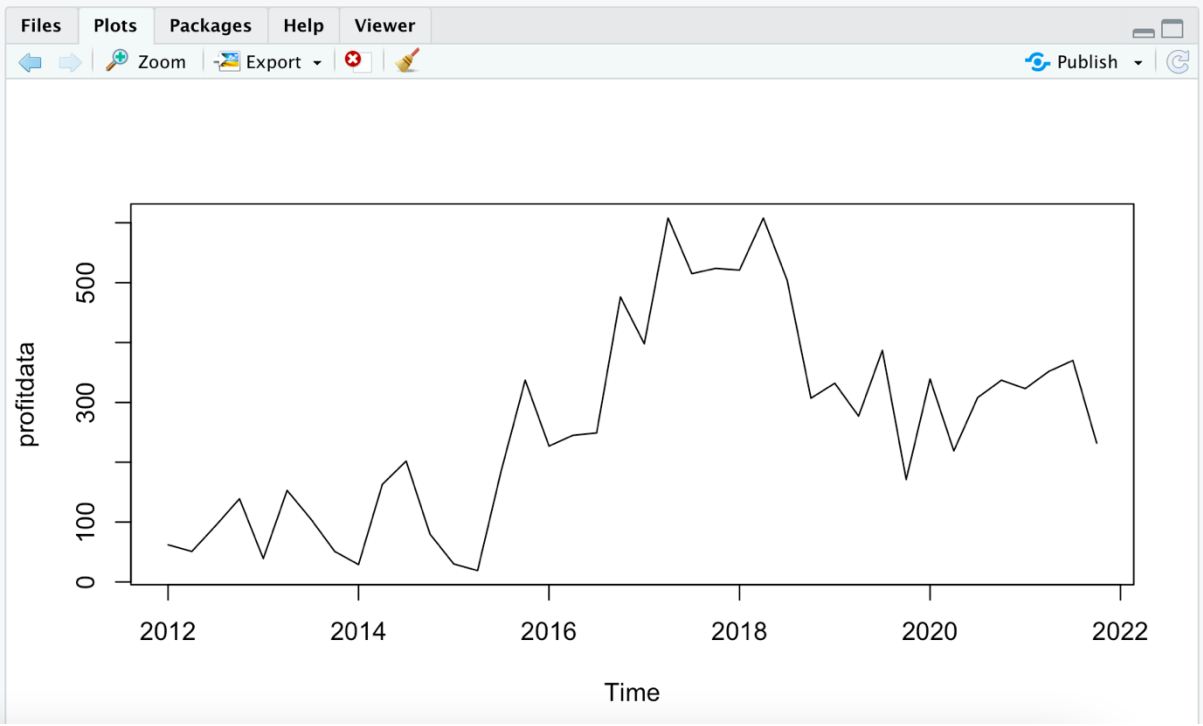
First, we have loaded the libraries needed to time series i.e.
Output:

```
#loading the required libraries
library(timeSeries)
library(fma)
```

Then we load the dataset which gives you the profit made in each quarter by Dr Coad selling second-hand books in his shop in New York City over the past 10 years:
Output:

```
#(a)Loading the time series
> booksalesprofit <-
ts(c(62,51,94,139,39,153,105,51,29,163,202,80,30,19,187,337,227,245,249,476,398,608,515,524,521,608,504,307,332,277,387,171,339,219,308,337,323,352,370,232))
> #Plotting the time series
> plot(profitdata)
```

Solution:



B) Fit a SARIMA model using the R function `auto.arima`.

An auto.arima function is used to find fit best ARIMA model to univariate time series. It returns best ARIMA model according to either AIC, AICc or BIC value. .
Auto ARIMA considers the AIC and BIC values generated (as shown in the code) while determining the ideal parameter combination. The estimators AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are used to compare models. The model is better if these values are lower.

Output:

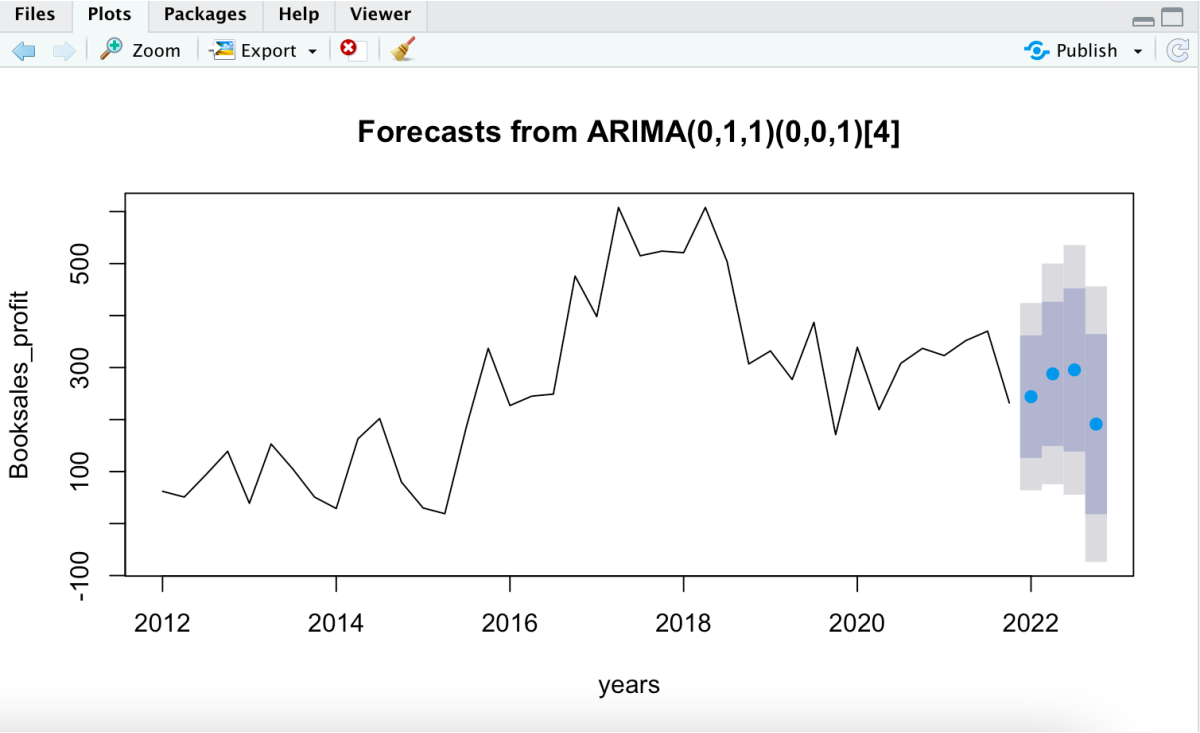
```
> auto.arima(profitdata,seasonal = TRUE)
Series: profitdata
ARIMA(0,1,1)(0,0,1)[4]

Coefficients:
         ma1    sma1
      -0.3770  0.5455
s.e.   0.1664  0.1721

sigma^2 = 8442:  log likelihood = -231.38
AIC=468.77   AICc=469.45   BIC=473.76
```

C) Forecast the time series for 2022.

Output:

```
> #(c) Forecast the time series
> model1 <- auto.arima(profitdata, seasonal = TRUE)
> forecast_2021 <- forecast(model1,h=4)
> plot(forecast_2021,xlab="years",ylab="Booksales_profit")
```

Solution:



D) Write a paragraph (300 words maximum) commenting on the model fit and the forecast.

In this project we are given a dataset about profit made in each quarter by Dr Coad selling books in New York City.

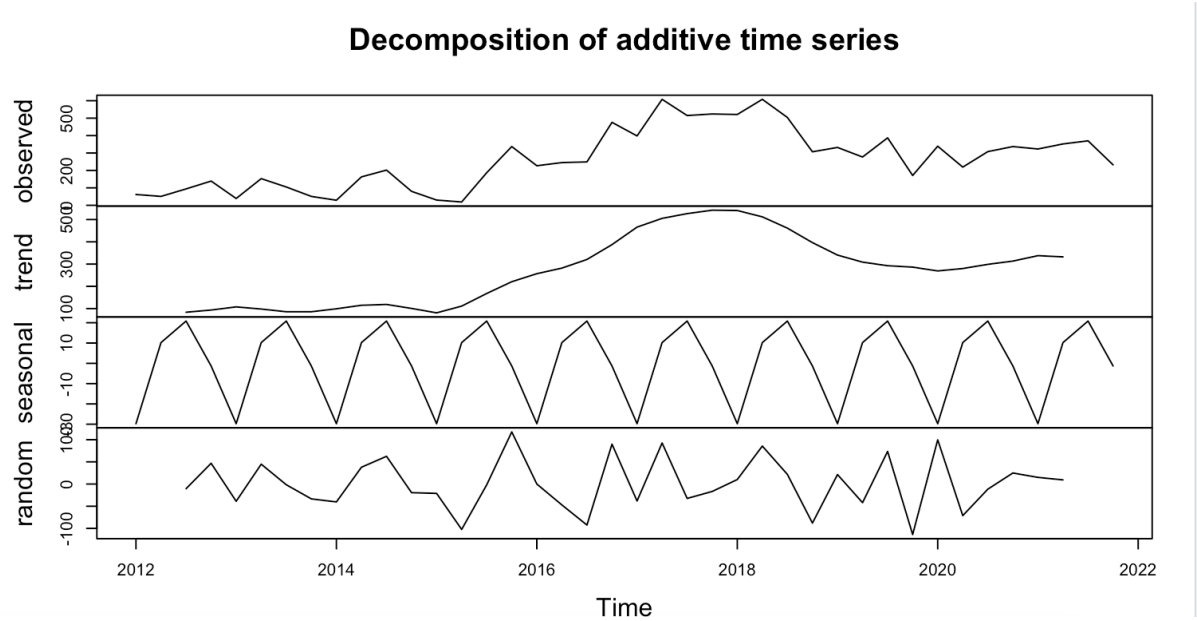
Firstly all libraries required were uploaded. The data was converted into time series. Subsequently pre-processed of the data was done before applying the model for prediction.

Secondly, step was to decompose the data to check trend and seasonality was there or not. Our data has a mix of an upward and downward trend as you can see in the graph attached below.

Output:

```
> profit_decompose<- decompose(profitdata)
> plot(profit_decompose)
```

Graph:



As discussed in the lectures we used the BoxCox.lambda function as it will choose a value of λ for you. A good value of λ is one which makes the size of the seasonal variation about the same across the whole series, as that makes the forecasting model simpler.

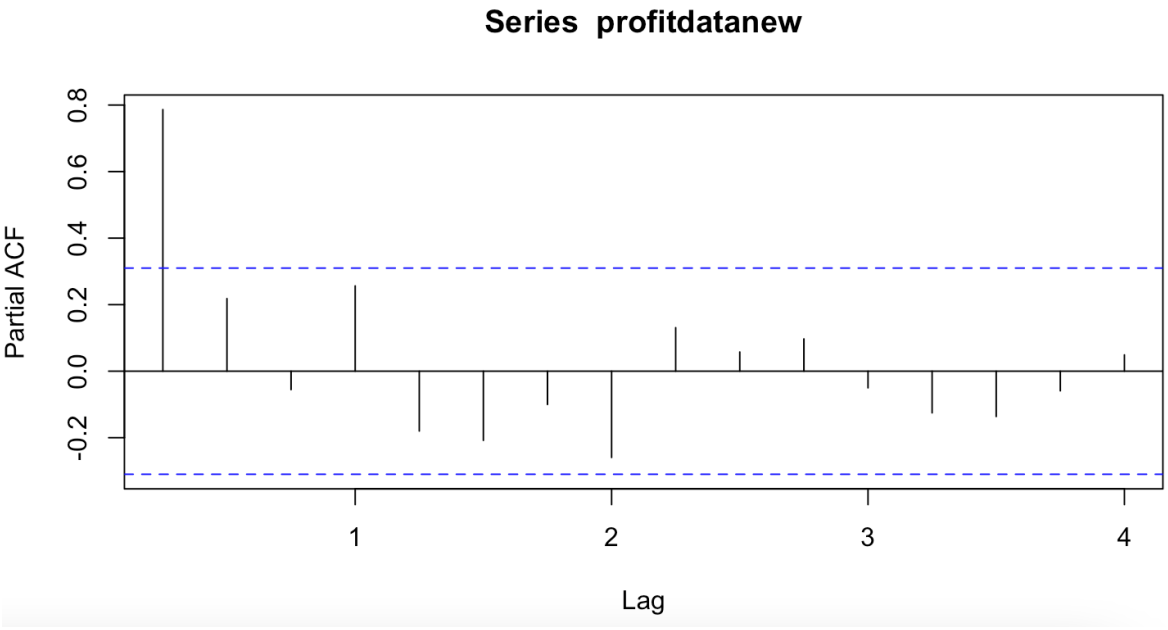
To have a model for comparison I have taken another model along with auto.arima model as mentioned in the question so as to have comparison grounds.

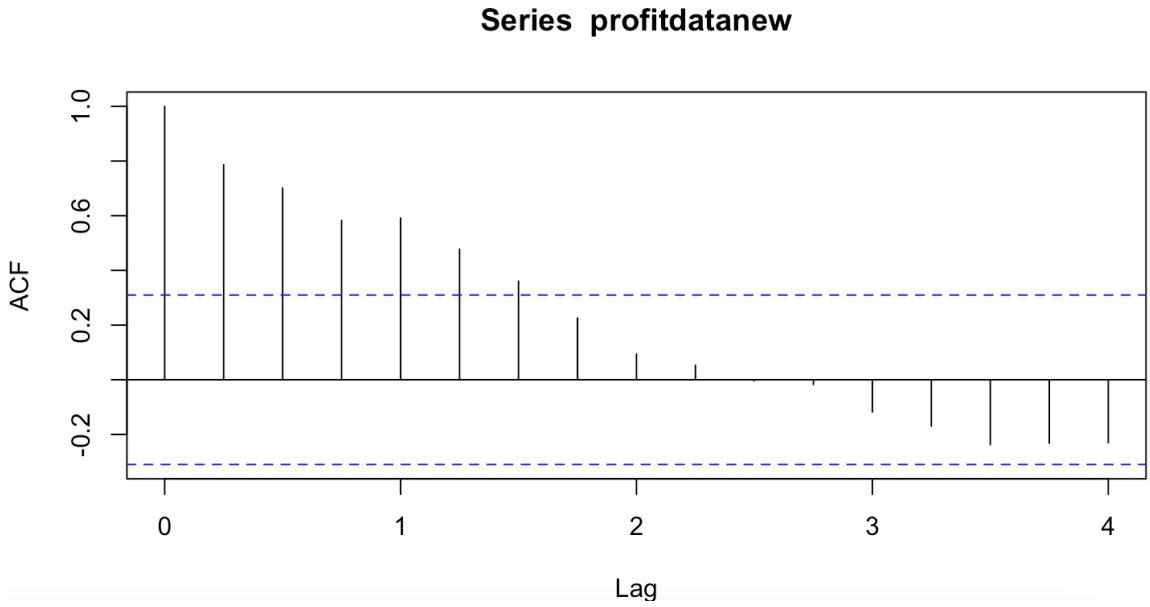
Output:
> lambda<- BoxCox.lambda(profitdata)
> lambda
[1] 0.8805541

Thirdly we are making the data stationary, using the differencing function(diff)

```
> #creating new profile for Box Cox
> profitdatanew<- BoxCox(profitdata, lambda)
> plot(profitdatanew)
> #Checking stationary of the data using diff function
> profitdatatstationary<- diff(profitdatanew)
> plot(profitdatatstationary)
```

We can also check for resisduals.Lastly we split the data into testing and training for ifnromation on our model comparison.





Model one featured an excellent residual plot with no pattern formation, an acf plot with all white noise below critical, which is good, and a density plot that was nearly perfectly normally distributed. Model 2 had a good residual value as well. In terms of AIC and BIC values, the second model was marginally better.