Stock Price Forecasting on C3.AI (March 2023- Sept 2023)

Nur Najla Nabila bt Azman (u2001004)

16/11/2023

a) Data Preparation

The stock price of C3 AI was downloaded from the Yahoo Finance website from 1 March 2023 to 21 September 2023 and imported into R using the code below.

```
library(fpp3)
library(readr)
AI <- read_csv("C:/Users/Asus/Downloads/C3.AI.csv")
```

Then, the data has been changed into the tsibble format with index= day where day is the number of the trading days.

```
AI <- AI |> mutate (day= row_number()) |> as_tsibble(index=day)
```

Next, the data has been extracted into 2 data frames. The mac_aug data frame was for the stock price from 1 March 2023 to 31 August 2023. The sept data frame was for the stock price from 1 September 2023 to 21 September 2023 that will be used for forecasting later.

```
mac_aug <- AI|> filter(month(Date)<09)
head(mac_aug)
```

```
## # A tsibble: 6 x 8 [1]
##   Date        Open  High   Low Close `Adj Close`   Volume   day
##   <date>     <dbl> <dbl> <dbl> <dbl>       <dbl>    <dbl> <int>
## 1 2023-03-01  22.8  22.9  20.3  20.7        20.7 14634400     1
## 2 2023-03-02  20.5  21.8  20.3  21.3        21.3 19332000     2
## 3 2023-03-03  25.3  28.9  24.0  28.5        28.5 75575200     3
## 4 2023-03-06  28.5  30.0  26.7  27.4        27.4 50682900     4
## 5 2023-03-07  27.2  28.4  25.8  26.1        26.1 26616100     5
## 6 2023-03-08  26.1  26.6  24.6  24.8        24.8 16689800     6
```

```
sept <- AI |> filter(month(Date)==09)
head(sept)
```

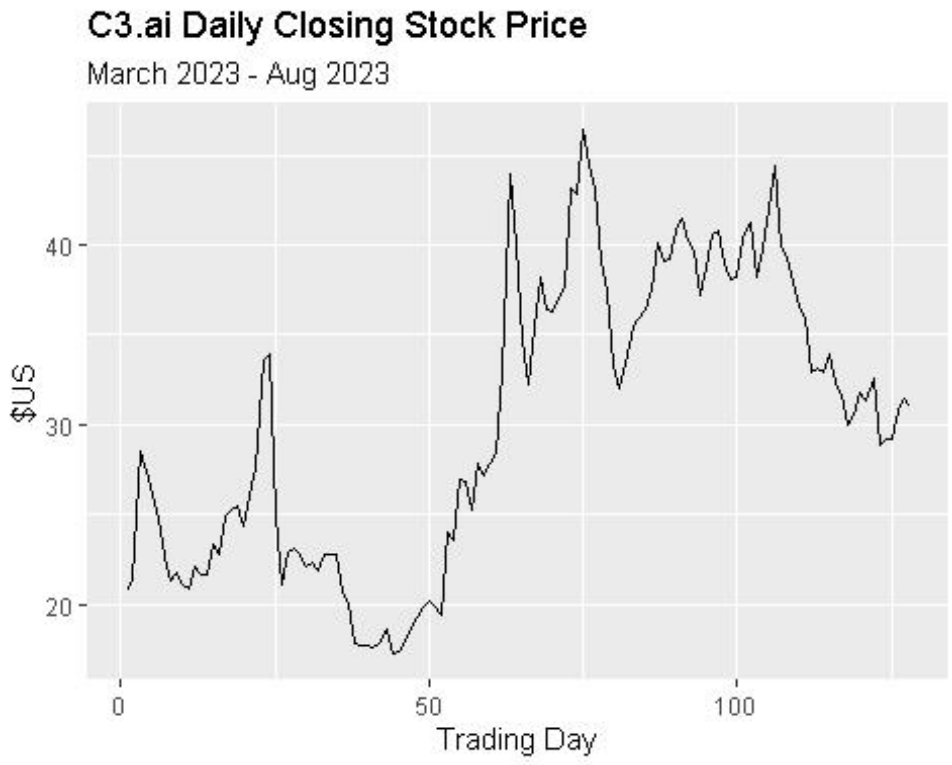
```
## # A tsibble: 6 x 8 [1]
##   Date        Open  High   Low Close `Adj Close`   Volume   day
##   <date>     <dbl> <dbl> <dbl> <dbl>       <dbl>    <dbl> <int>
## 1 2023-09-01  31.6  31.8  30.7  31.2        31.2  6757100   129
## 2 2023-09-05  31.2  31.5  30.0  31.4        31.4  8459200   130
## 3 2023-09-06  32.0  32.7  30.9  31.5        31.5 19546900   131
```

```
## 4 2023-09-07  27.9  28.2  25.5  27.6         27.6 26065400   132
## 5 2023-09-08  27.5  28.7  27.4  28.0         28.0 11619000   133
## 6 2023-09-11  28.2  28.5  27.3  28.2         28.2  8947800   134
```

b) Time Series Plot

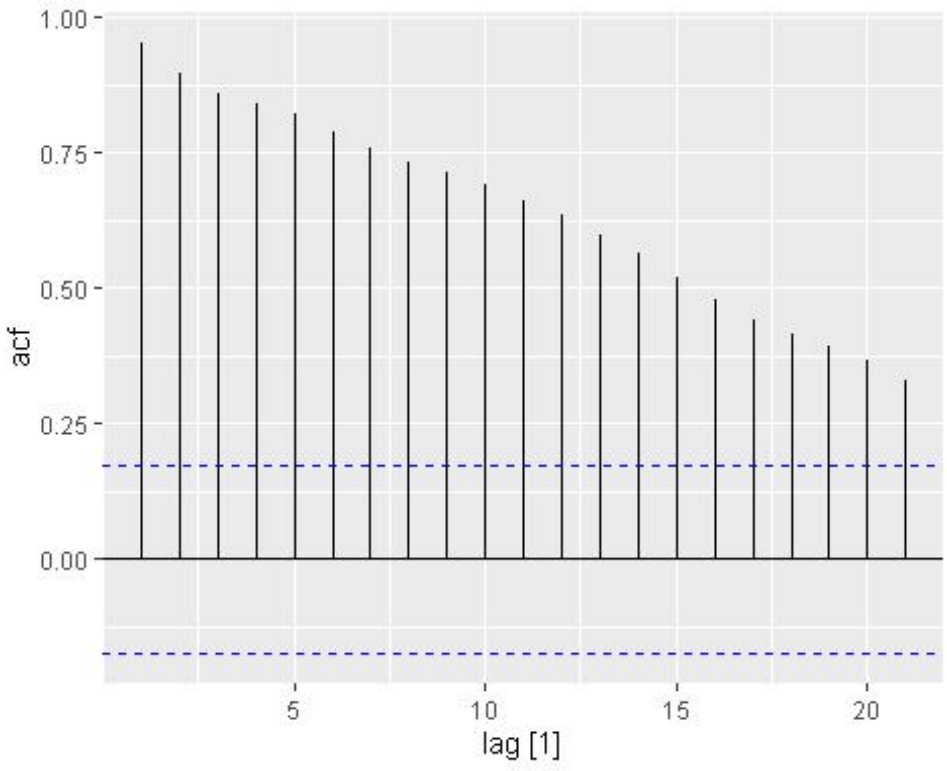
The closing price for C3.AI's stock from March to August has been plot using the autoplot function. From the plot, we can see an increasing trend with a strong drop at day 44. The plot also shows cyclicity and no seasonality.

```
plot_mac_aug <- autoplot(mac_aug,Close) + labs(y="$US", x="Trading Day",
 title="C3.ai Daily Closing Stock Price", subtitle="March 2023 - Aug 20
23")
plot_mac_aug
```



The ACF plot has also been plotted to help identify the existence of trend from the data. From the plot, it shows that the data does have strong trend since the ACF is decreasing as lag increasing with no seasonality since the plot does not shows "scalloped" pattern.

```
mac_aug |> ACF(Close)|>autoplot()
```



c) Analyzing Historical Data of NYSE

The historical data of NYSE from 1 January 2015 to 31 August 2023 has been imported into R using the code below.

```
NYSE <- read_csv("C:/Users/Asus/Downloads/^NYA.csv")
```

Then, the daily and monthly returns has been calculated to check the seasonality.

```
library(tidyquant)

#Calculate Daily Returns
nyse_daily_return <- NYSE |> tidyquant::tq_transmute(select = Close,
                                                     mutate_fun=periodR
eturn,
                                                     period="daily",
                                                     col_rename="daily_
returns")

#Calculate monthly returns
nyse_monthly_return <- NYSE |> tidyquant::tq_transmute(select=Close,
                                                       mutate_fun=perio
dReturn,
                                                       period="monthly",
```

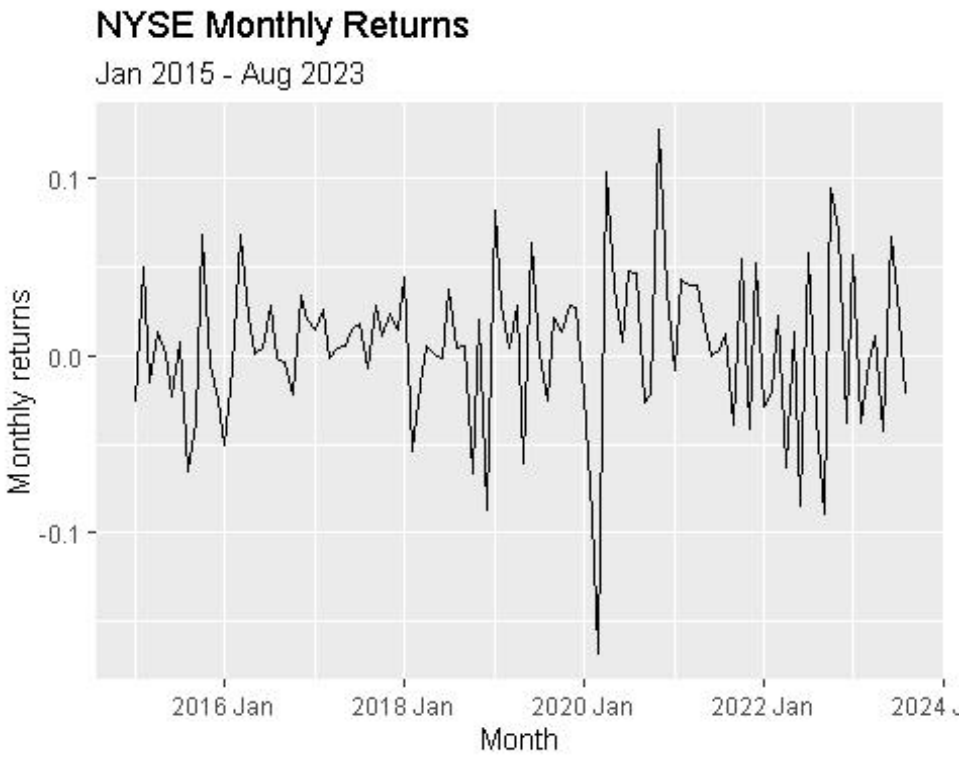
```
                                                    col_rename="mont
hly_returns")
#change to tsibble
nyse_return <- nyse_monthly_return |> mutate(Month=yearmonth(Date)) |>
select(-Date,Month,monthly_returns)|> as_tsibble(index=Month)

head(nyse_return)
```

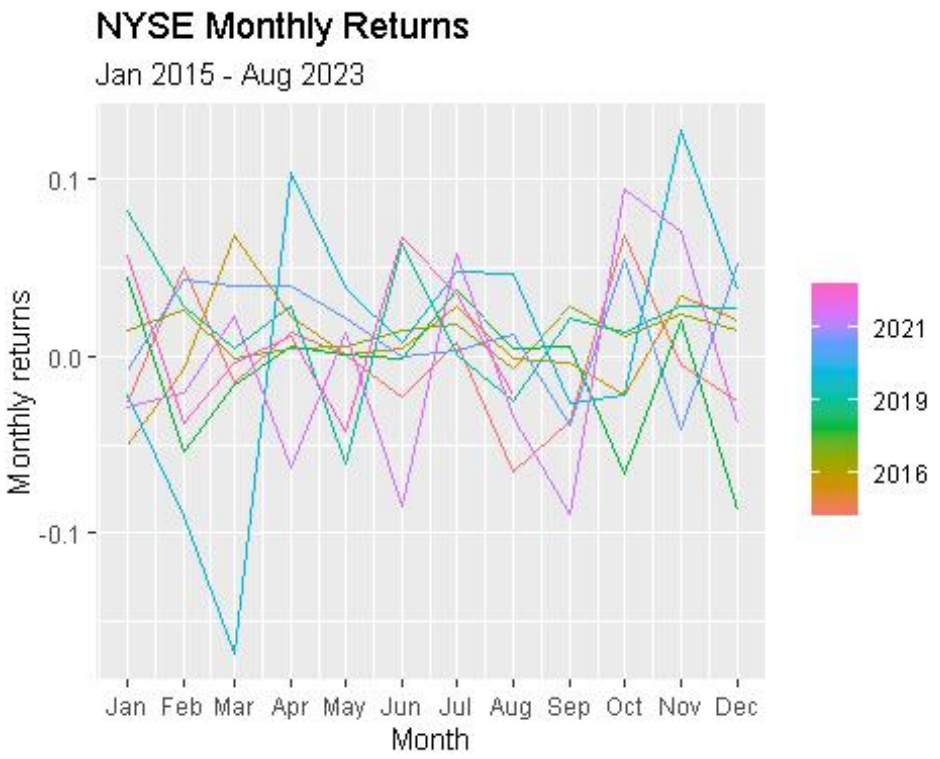
```
## # A tsibble: 6 x 2 [1M]
##   monthly_returns    Month
##             <dbl>    <mth>
## 1       -0.0271   2015 Jan
## 2        0.0499   2015 Feb
## 3       -0.0148   2015 Mar
## 4        0.0138   2015 Apr
## 5        0.000594 2015 May
## 6       -0.0227   2015 Jun
```

The following plots was produces to check the seasonality

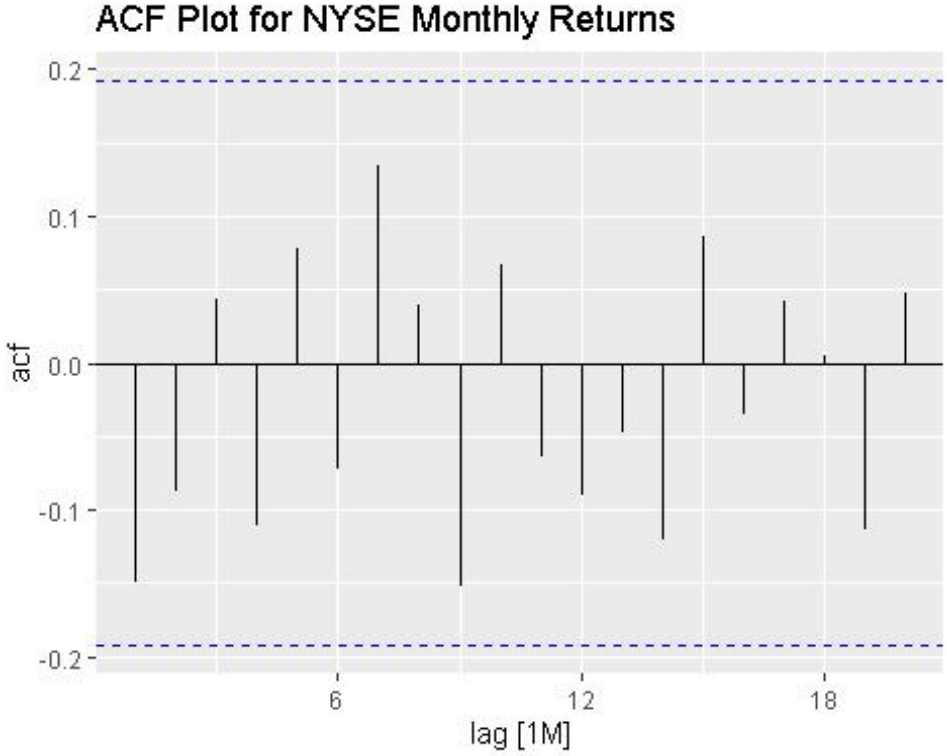
```
#time series plot
autoplot(nyse_return,monthly_returns) + labs(title="NYSE Monthly Return
s",subtitle= "Jan 2015 - Aug 2023", y="Monthly returns",x="Month")
```




```
#ggseason
gg_season(nyse_return,monthly_returns) +labs(title="NYSE Monthly Return
s",subtitle= "Jan 2015 - Aug 2023", y="Monthly returns",x="Month")
```




```
#ACF Plot
nyse_return |> ACF(monthly_returns) |> autoplot()+ labs(title="ACF Plot
 for NYSE Monthly Returns")
```



d) Forecasting
e) Training and Test Set

To select the best forecasting method, the data set has been split into training and test set using 8:2 ratio. The training data set was set from 1 March 2023 to 26 July 2023 which is from day 1 to day 102 while the test set was from 27 July 2023 to 31 August 2023.

```
train <- mac_aug |> filter(day<=102)
head(train)
```

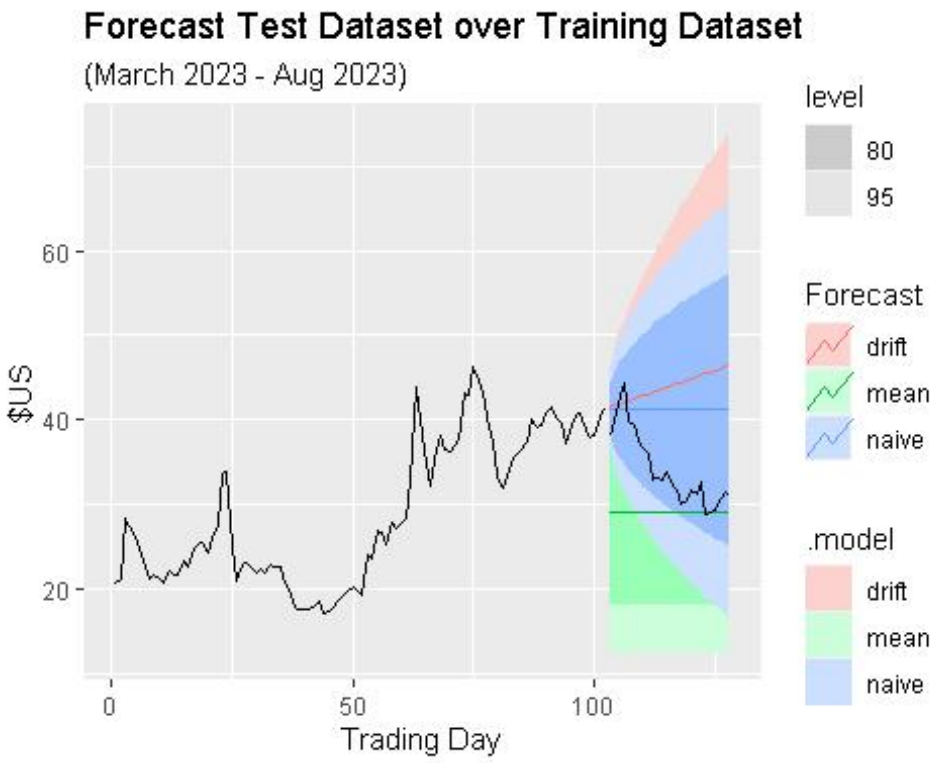
```
## # A tsibble: 6 x 8 [1]
##   Date        Open  High   Low Close `Adj Close`   Volume   day
##   <date>     <dbl> <dbl> <dbl> <dbl>       <dbl>    <dbl> <int>
## 1 2023-03-01  22.8  22.9  20.3  20.7        20.7 14634400     1
## 2 2023-03-02  20.5  21.8  20.3  21.3        21.3 19332000     2
## 3 2023-03-03  25.3  28.9  24.0  28.5        28.5 75575200     3
## 4 2023-03-06  28.5  30.0  26.7  27.4        27.4 50682900     4
## 5 2023-03-07  27.2  28.4  25.8  26.1        26.1 26616100     5
## 6 2023-03-08  26.1  26.6  24.6  24.8        24.8 16689800     6
```

```
test <- mac_aug |> filter(day>102)
head(test)
```

```
## # A tsibble: 6 x 8 [1]
##   Date        Open  High   Low Close `Adj Close`   Volume   day
##   <date>     <dbl> <dbl> <dbl> <dbl>       <dbl>    <dbl> <int>
## 1 2023-07-27  42.3  43.8  37.8  38.2        38.2 30001900   103
## 2 2023-07-28  39.0  40.0  38.3  39.8        39.8 15823600   104
## 3 2023-07-31  40    42.2  39.8  42          42   18478800   105
## 4 2023-08-01  41.3  44.9  40.2  44.4        44.4 28084100   106
## 5 2023-08-02  42.3  42.4  38.4  39.9        39.9 29427600   107
## 6 2023-08-03  39.5  40.5  38.7  39.3        39.3 13596200   108
```

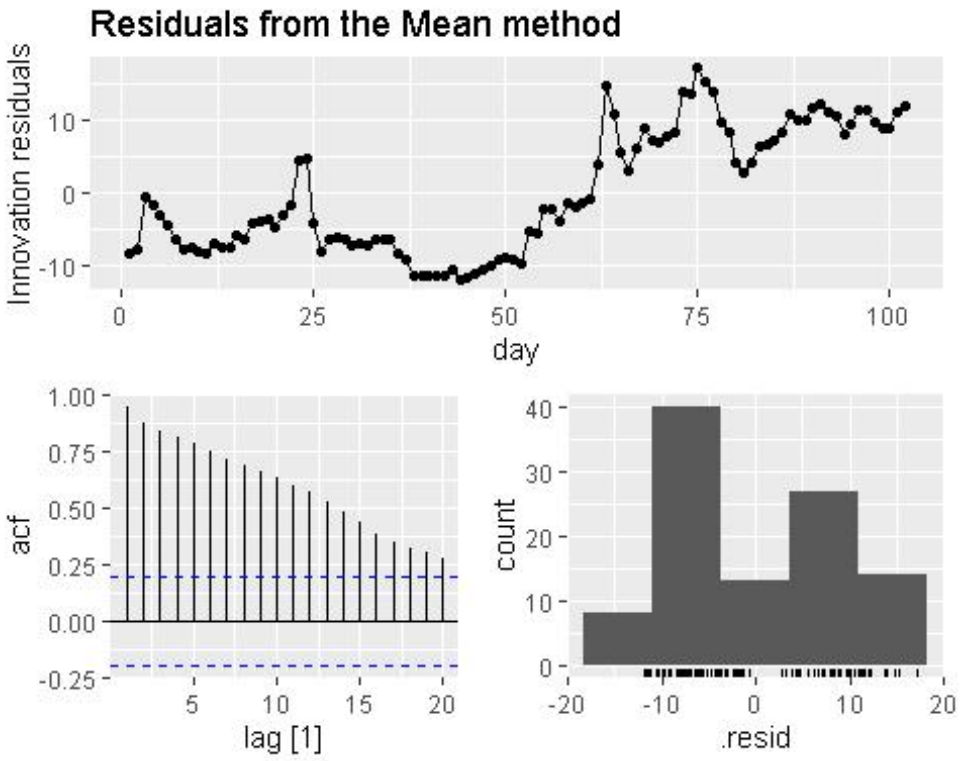
Training set was used to forecast the test set using the 3 basic forecasting method which are Mean, Naive, and Drift method. Snaive method was not suitable to used in this stock price dataset as this dataset does not have seasonality.

```
fit_AI <- train |> model(mean=MEAN(Close),naive=NAIVE(Close),drift =RW
(Close~drift()))
fc_AI <- fit_AI|> forecast(new_data=test)
fc_AI |> autoplot(train)+ autolayer(test,Close,colour="black") +
  labs(y = "$US", x="Trading Day",
       title = "Forecast Test Dataset over Training Dataset",
       subtitle = "(March 2023 - Aug 2023)") +
  guides(colour = guide_legend(title = "Forecast"))
```

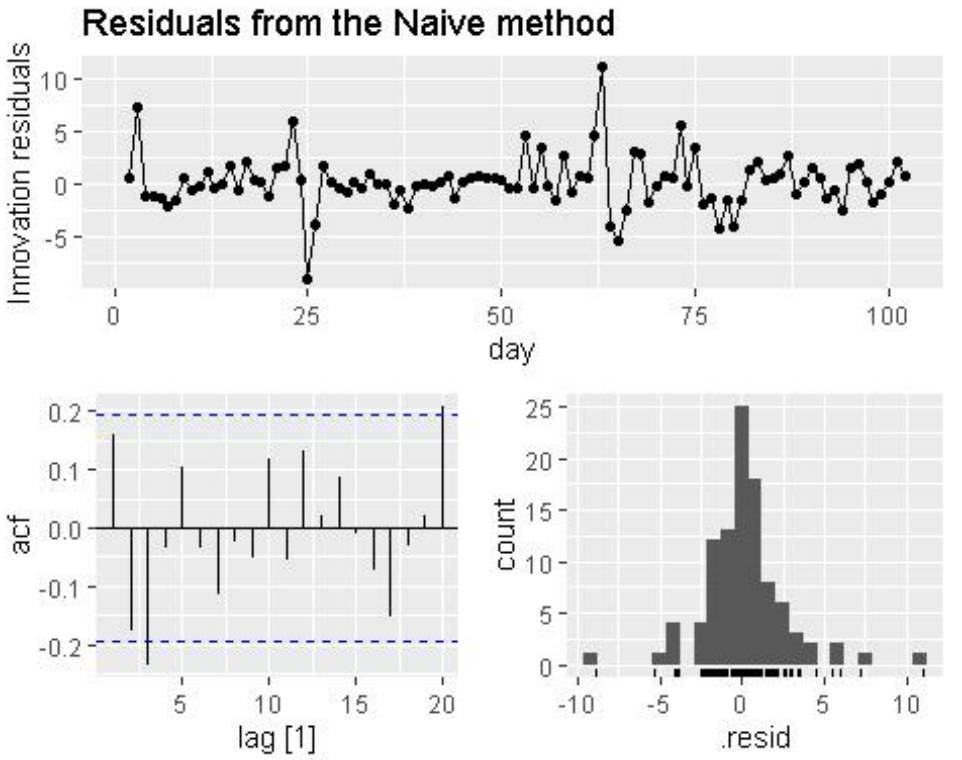


Then, I plot the residual plot to see the assumptions of the residual for Mean and Naive method. From the residual plot for mean method, the residual series are correlated and the mean is not zero. However, the residual plot for Naive method shows that residuals are uncorrelated and the mean is close to zero. Hence, Naive method might be the best forecasting method.

```
train |> model(MEAN(Close)) |> gg_tsresiduals()+ labs(title="Residuals
from the Mean method")
```




```
train |> model(NAIVE(Close)) |> gg_tsresiduals()+ labs(title="Residuals
 from the Naive method")
```



The ljung box test was used to test for the autocorelation of the residual for Naive method model.Since p-value is more than 0.05. we do not reject null hypothesis where the residuals shows white noise (uncorrelated). Hence, the Naive method is concluded as the best forecasting method.

```
aug <- train |> model(NAIVE(Close)) |> augment()
aug |> features(.innov,ljung_box,lag=10)
```

```
## # A tibble: 1 × 3
##   .model       lb_stat lb_pvalue
##   <chr>          <dbl>     <dbl>
## 1 NAIVE(Close)    16.6    0.0846
```

ii) The Naive forecasting method was used to forecast the stock price from 1 September 2023 to 21 September 2023 (14 trading days).

```
fit <- mac_aug |> model(NAIVE(Close))
fit |> forecast(h=14)|> autoplot(mac_aug) + autolayer(sept,Close,colour
="black") +
  labs(y = "$US", x= "Trading Day",
       title = "Forecast on 14 Trading Days of C3.AI Stock Price",
       subtitle = "(1 March 2023 - 21 Sept 2023)") +
  guides(colour = guide_legend(title = "Forecast"))
```