TIME SERIES

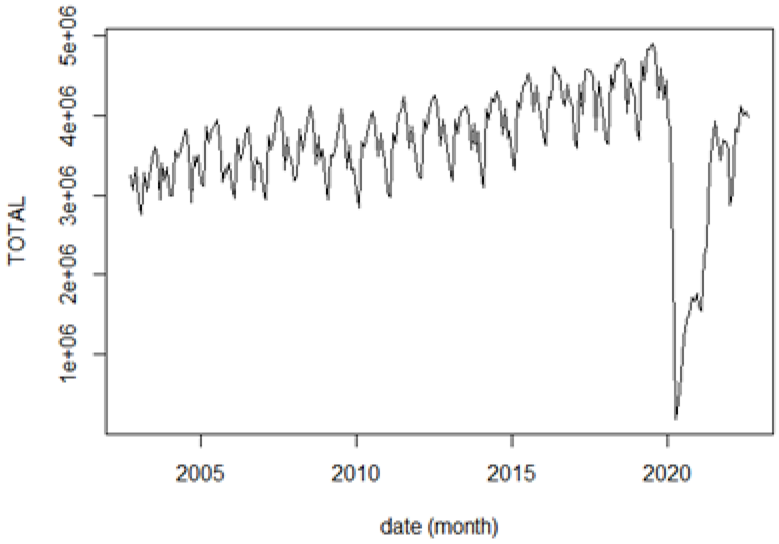
ZARKAN Soumia
DAFRANE Safiya
KNOUZI Mohamed-Amine

Description of the time series

```
> acf(ts(traf, frequency=1))
> pacf(ts(traf, frequency=1))
> str(traf)
 Time-Series [1:239, 1] from 2003 to 2023: 2002 2002 2002 2003 2003 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr "Year"
> class(traf)
[1] "ts"
> start(traf)
[1] 2002   10
> end(traf)
[1] 2022    8
> frequency(traf)
[1] 12
```

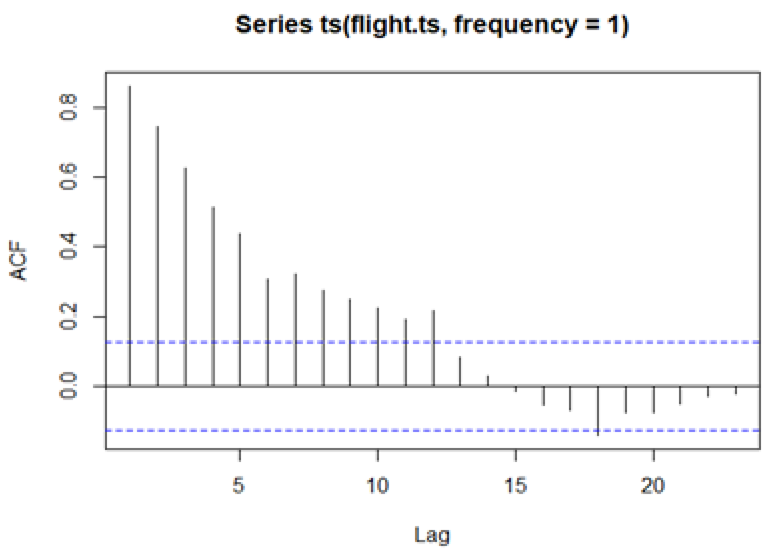
We can assume that our series isn't stationary because of a trend, random components and seasonality.

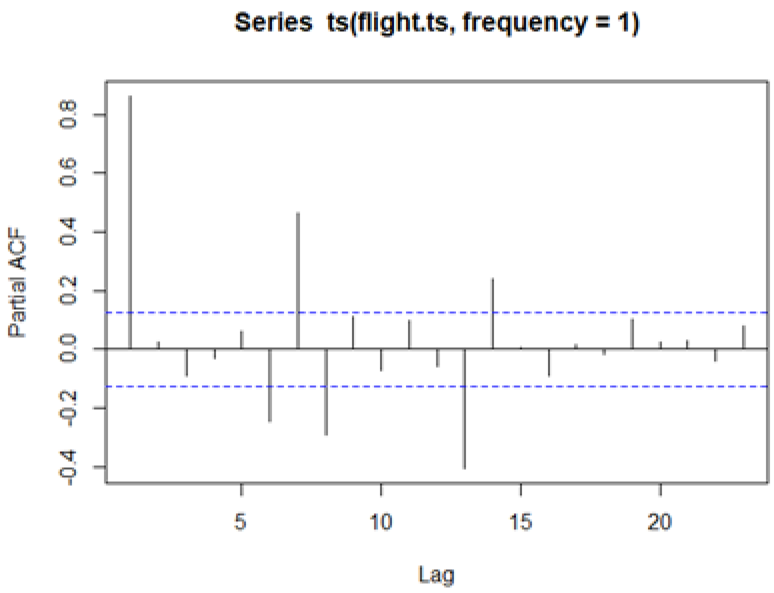
Monthly traffic of Atlanta Airport from October 2002 to August 2022 with 239 observations.



We have a variance not constant.

Stationarity: Is our series stationary?



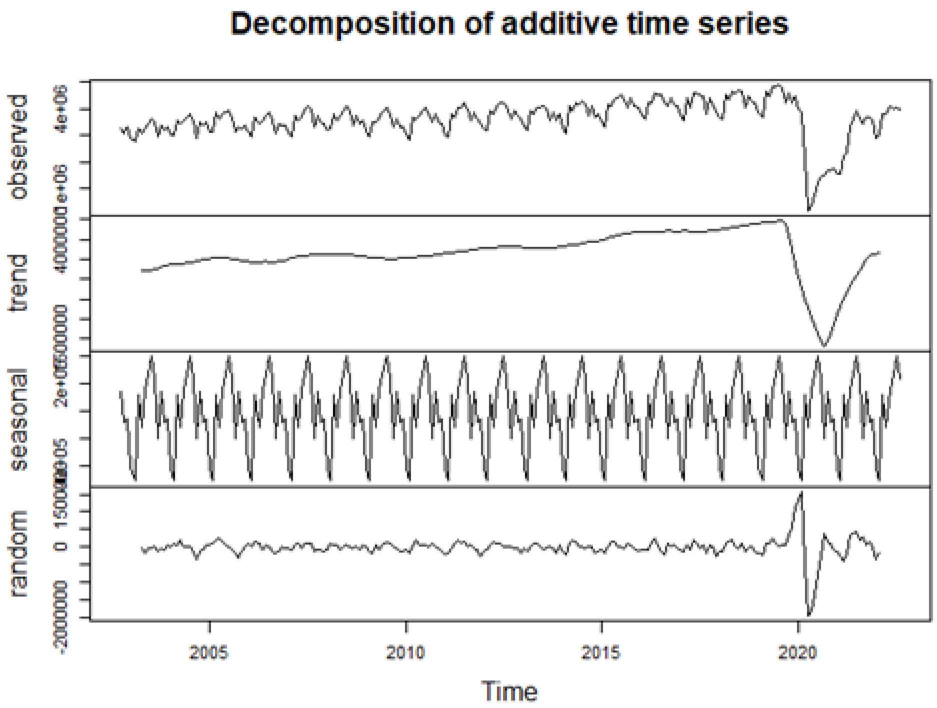


<u>For ACF :</u>
Significant coefficients at lags 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

<u>For PACF :</u>
The significant coefficients at lags 1, 6, 7, 8, 13, 14 .

These significant coefficients shows that our serie isn't stationary

Stationarity: Is our series stationary?

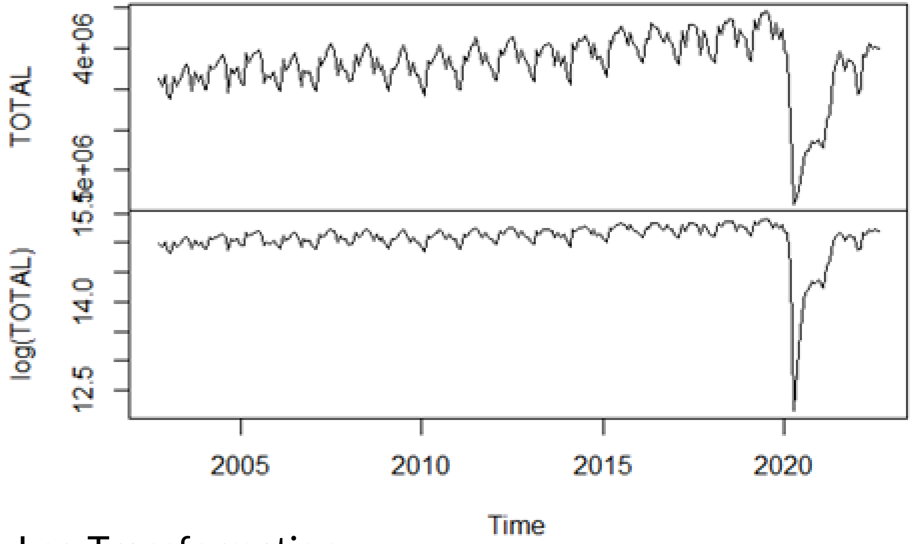


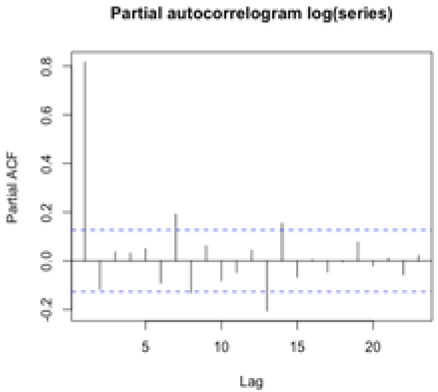
That shows that our time series is not stationary.

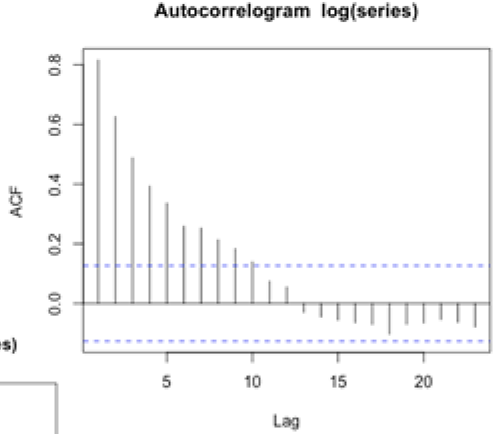
<u>Decomposition of time series :</u>

- For the trend, We can see a fairly steady crescent over time, until 2019.
- We have a seasonality that repeats itself over time.
- For the random component , We have a regular curve, until 2019.

Stationarity: Log - 1st transformation



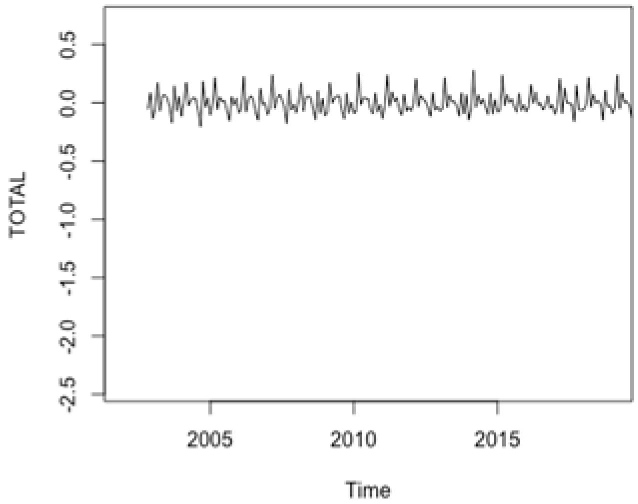


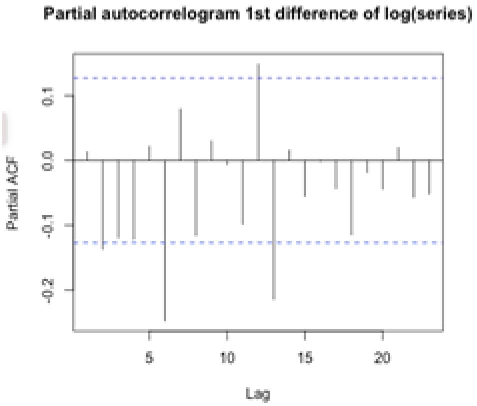


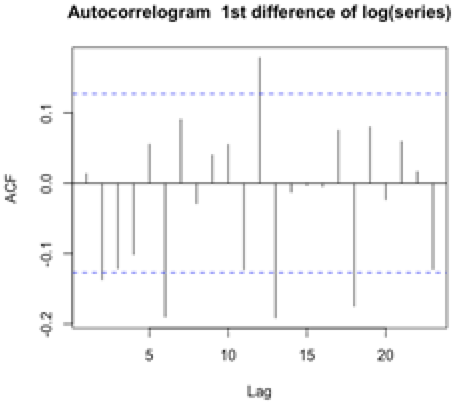
Log Transformation :

The log remove the variance effect and change it into a constant one as we can see in the graph log(TOTAL).
We see in ACF the same patterns of non stationarity thanks to a persistence of significant coefficients, no fast decay to 0.
We see in PACF, the log(series) is still non stationary.

Stationarity: 1st order difference - 2nd transformation



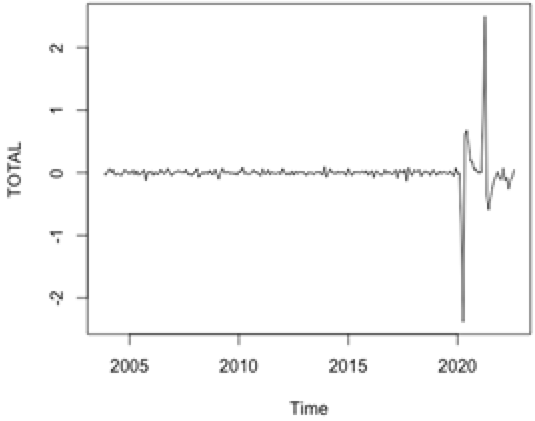


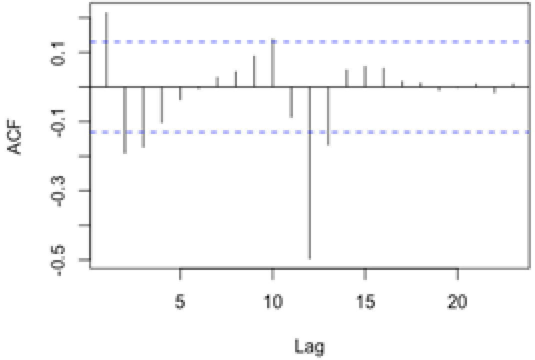


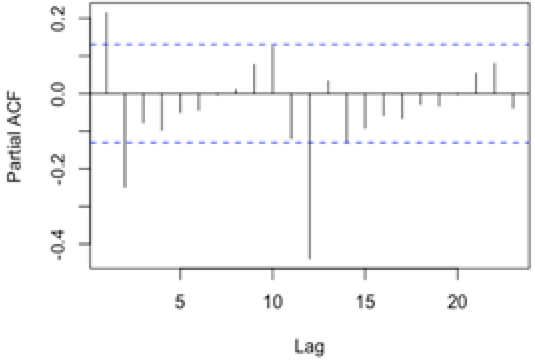
<u>1st order difference :</u>

There is no more increasing variance and no more trend thanks to the 1st order difference transformation.
We can assume that there is still a potential seasonal effects.
ACF and PACF show us that there is a persistence of significant coefficients with seasonality s=12
So The series is still non stationary.

Stationarity: Difference of order 12 - 3rd transformation







<u>Difference of order 12 :</u>

The difference of order 12 remove the seasonal effects as we can see in the graph TOTAL.
We can identify some perturbations in the traffic after 2020.
On ACF, we see significant coefficients at lags 1,2,3,9,11,12.
On PACF, we see significant coefficients at lags 1,2,12.

Box-Jenkins methodology : Identification of the orders p and q

```
Call:
arima(x = ltraf, order = c(2, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12),
    method = "ML")

Coefficients:
         ar1      ar2      ma1     sma1
      1.0705  -0.2691  -0.9751  -0.8834
s.e.  0.0665   0.0653   0.0314   0.0533

sigma^2 estimated as 0.0309:  log likelihood = 61.48,  aic = -112.97
```

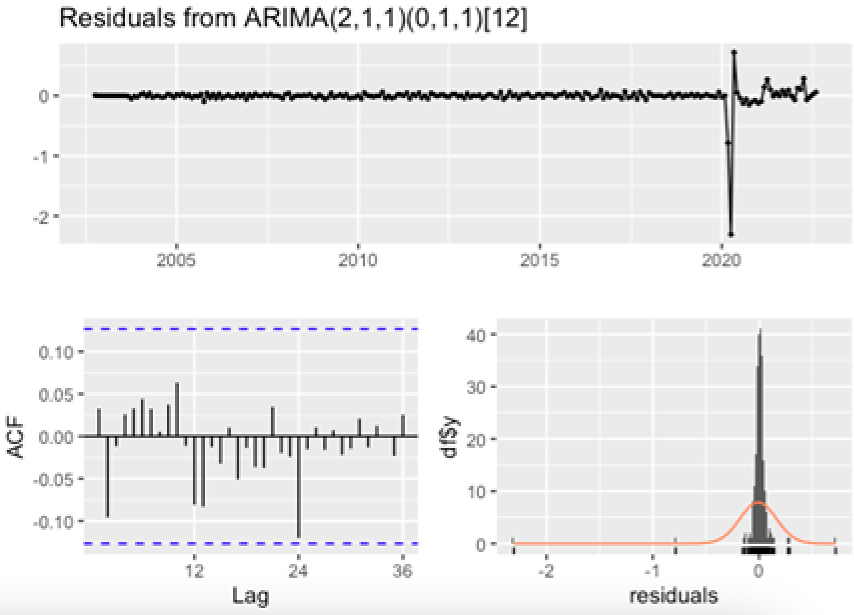
So ARMA model : p,q (2, 1)

```
> pvalue
         ar1          ar2          ma1         sma1
0.000000e+00 3.733478e-05 0.000000e+00 0.000000e+00
```

To determine our ARIMA model, we had to observe the lags of ACF and PACF. So we determined p, d, q: (2, 1, 1) and P, D, Q: (0, 1, 1). The presence of d = 1 is justified by the difference order of 1.
The presence of D = 1 is justified by the difference order of 12.
We observe an AIC of -112.97, the lower the better so it's good.

We checked p-values of ar1, ar2, ma1 and sma1. All are lower than 5% so they are significant.
We can accept the model.

Residual diagnostic



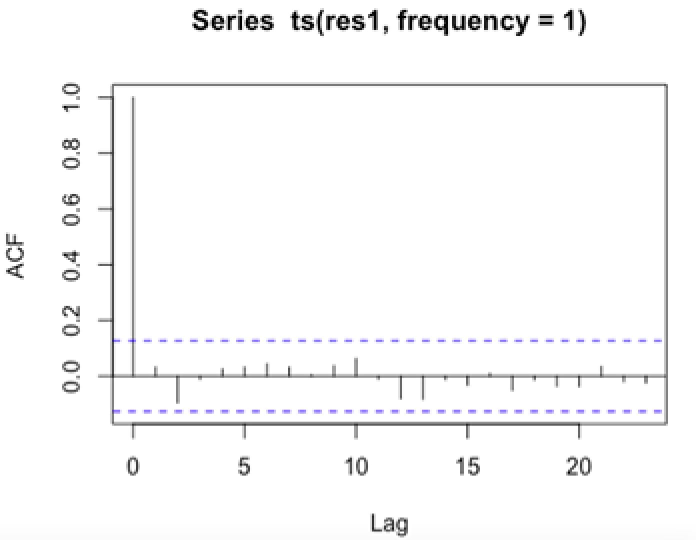

```
        Ljung-Box test

data:  Residuals from ARIMA(2,1,1)(0,1,1)[12]
Q* = 14.752, df = 20, p-value = 0.7904

Model df: 4.   Total lags used: 24
```

We can identify one extreme value: the residuals are less than -2.0

In conclusion, the Ljung-Box test is used to verify the time series model's assumption of residual independence, as we can see there is some degree of autocorrelation because chi^2 is not equal to 0.

The p-value is greater than 5%, it means that the residuals of the model appear to be uncorrelated

Autocorrelation of the residuals

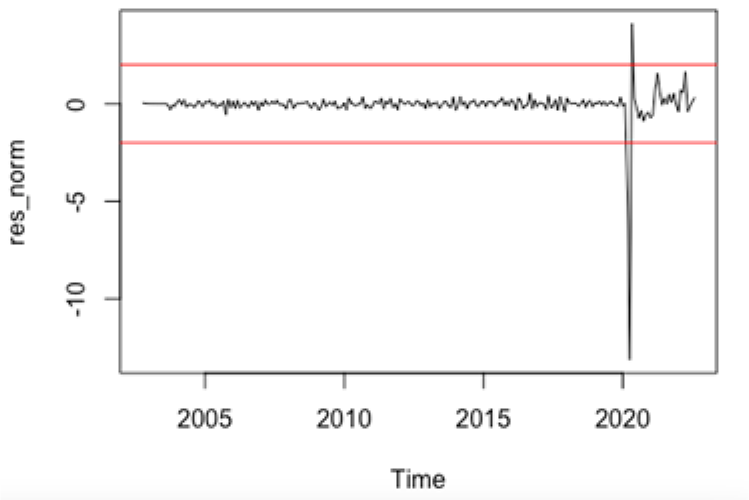



```
        Box-Ljung test

data:  res1
X-squared = 10.286, df = 20, p-value = 0.9627
```

There is no significant coeff. The autocorrelation of our residuals are really good.

Ho: all correlations are equal to 0
Our p-value=96% > 5%, we accept HO, so there’s no autocorrelation

Normality distribution assumption

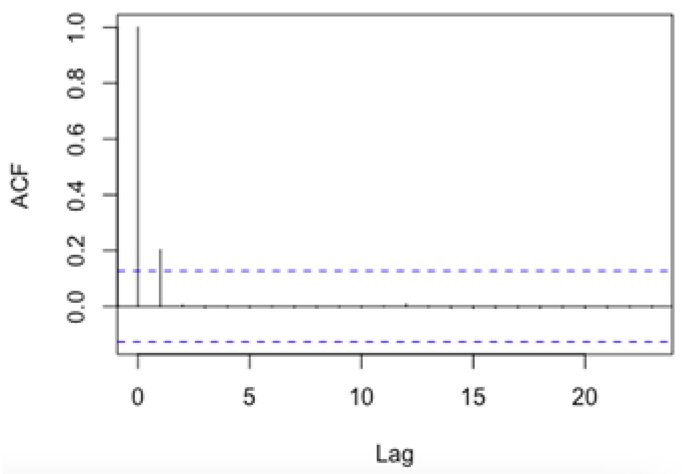



```
> res_norm <- res1/sqrt(mod1$sigma2) # normalized residuals
> summary(res_norm)
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-13.127363  -0.125985   0.002899  -0.046186   0.124514   4.102024

> out1 <- which(res_norm < -5) # identification number of the outlier
> out1 # the outlier corresponds ot observation n°211
[1] 211
```

```
> index(res_norm)[out1] #date of the outlier
[1] 2020.25
> # the outlier occurs at date 2020.25 (1 month= 1/12)
> traf[out1] # value of the outlier
[1] 188766
> traf[(out1-2):(out1+2)] # values around the outlier
[1] 3834870 2153479  188766  379346  746102
> traf[(out1-14):(out1-10)] # values 12 months before the outlier
[1] 3696212 4682013 4447875 4834781 4824531
```

If the residuals follow a Normal distribution, the values of res_norm should lie in between -2 and 2, with 95% of chance. We identified outliers in number 5 and this corresponds to the observation n°211.
The date of the outlier is 2020 according our model and the graph

Homoscedasticity of the residuals




```
sq.res <- (res1)^2
acf(ts(sq.res, frequency=1))
# 2 highly significant coeffs: lag 0 and lag 1
# There is an issue of non constant variance

> Htest
$p.values
 [1] 0.002200570 0.009194704 0.024546455 0.052023603 0.094145331 0.152218306
 [7] 0.224690663 0.308431106 0.399423813 0.492160917 0.582478573 0.665029170
[13] 0.738720884 0.800742046 0.851742409 0.892209731 0.923540529 0.946854881
[19] 0.963929436 0.975954211 0.984303266 0.989990029 0.993741698

> #pvalue=0.0022 < 5% so we reject the constant variance assumption
```

The residuals are not homoscedastic, then the variance of the residuals is not constant. There is an issue of non constant variance. There is two highly significant coefficients : lag and lag 1
Applying McLeod test, The analysis of the 1st p-value is enough with a p-value= 0.0022 < 5% so we reject the constant variance assumption

Model validation

Model 2

```
> mod2 <- arima(ltraf, c(2,1,2), seasonal=list(order=c(0,1,1), period=12), method='ML')
> mod2

Call:
arima(x = ltraf, order = c(2, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12),
    method = "ML")

Coefficients:
         ar1      ar2      ma1      ma2     sma1
      0.7615  -0.0010  -0.6341  -0.3427  -0.8764
s.e.  0.1913   0.1752   0.1799   0.1803   0.0520

sigma^2 estimated as 0.03051:  log likelihood = 62.99,  aic = -115.98
```

```
> pvalue
         ar1          ar2          ma1          ma2         sma1
6.850313e-05 9.956312e-01 4.249337e-04 5.739047e-02 0.000000e+00
```

The model 2 has an AIC of -115.98 and only ar2 is non significant.

Model 3

```
> mod3

Call:
arima(x = ltraf, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12),
    method = "ML")

Coefficients:
         ma1     ma2     sma1
      0.1563  -0.187  -0.8854
s.e.  0.0757   0.081   0.0515

sigma^2 estimated as 0.03297:  log likelihood = 55.68,  aic = -105.37
```

```
> pvalue
       ma1        ma2       sma1
0.03896098 0.02089872 0.00000000
```

The model 3 has an AIC of 105.37 and only sma1 is significant.

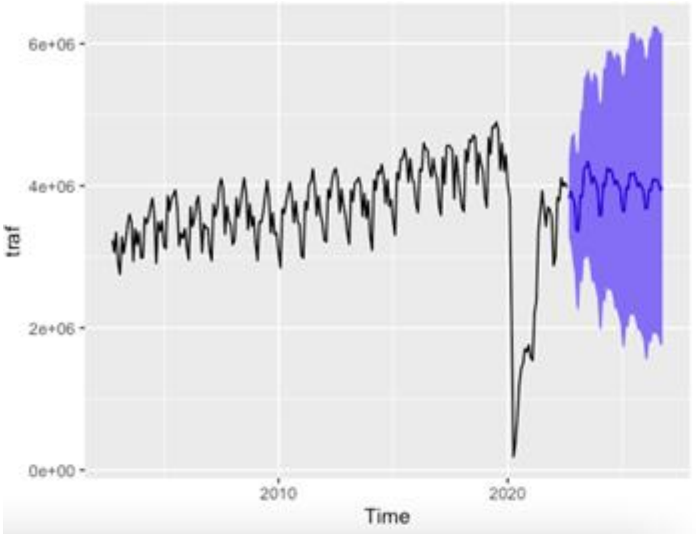
Model 4

```
> mod4

Call:
arima(x = ltraf, order = c(0, 1, 2), seasonal = list(order = c(0, 1, 1), period = 12),
    xreg = Altltraf$dum1, method = "ML")

Coefficients:
         ma1     ma2     sma1     xreg
      0.8924  0.3498  -0.9165  -1.1426
s.e.  0.0805  0.0720   0.0686   0.0462

sigma^2 estimated as 0.009516:  log likelihood = 194.09,  aic = -380.18
```

```
> pvalue
         ma1          ma2         sma1         xreg
0.000000e+00 1.183434e-06 0.000000e+00 0.000000e+00
```

The model 4 has the best AIC = -380 and all coefficient are significant.

Forecast



Here is the forecast for the next 4 periods.