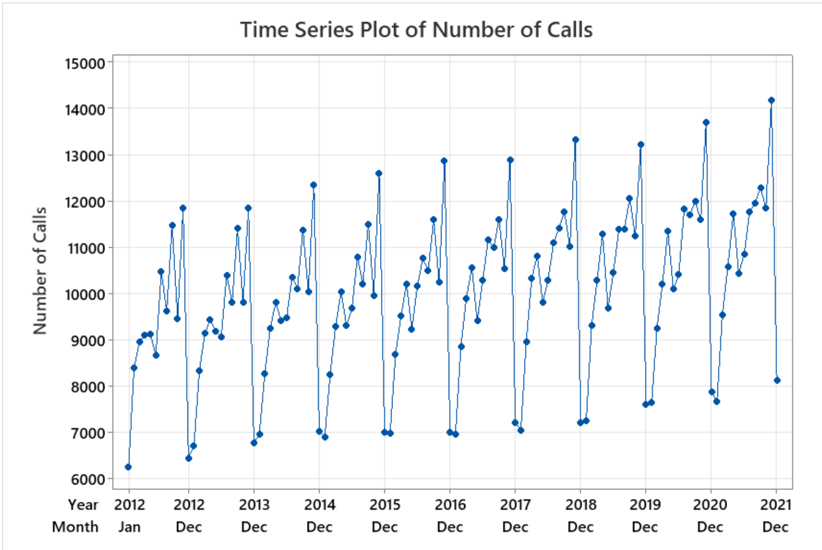
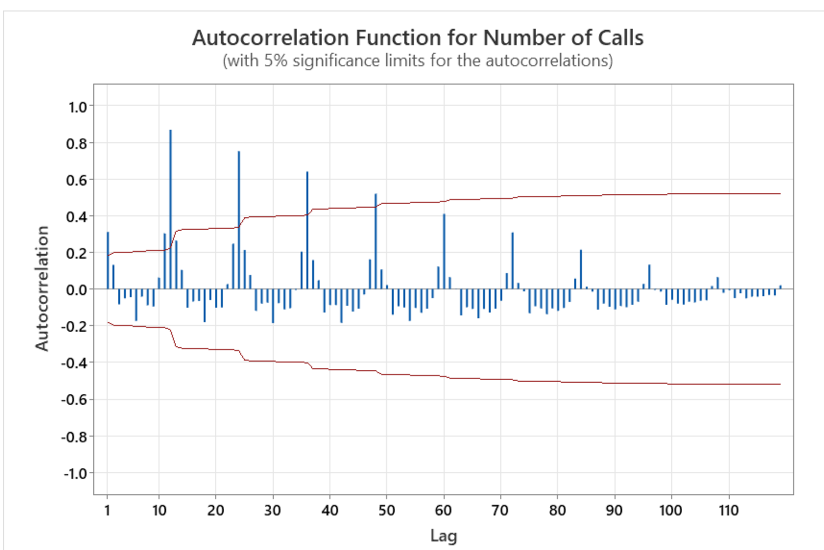
1. The time series plot for the data is presented below. The series does show a trend as the number of calls is generally increasing over the years.



2. The plot below depicts the autocorrelation function for number of calls for 120 lags. We observe significant autocorrelation coefficients at lags 12, 24, 36, etc., which are multiples of 12, the seasonal lag for monthly data. We can, therefore, conclude that there is monthly seasonality in the data at hand.

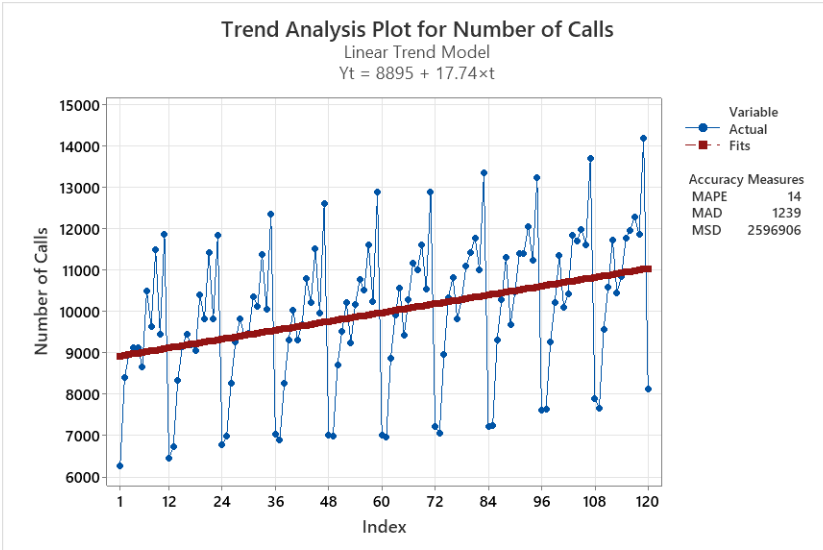


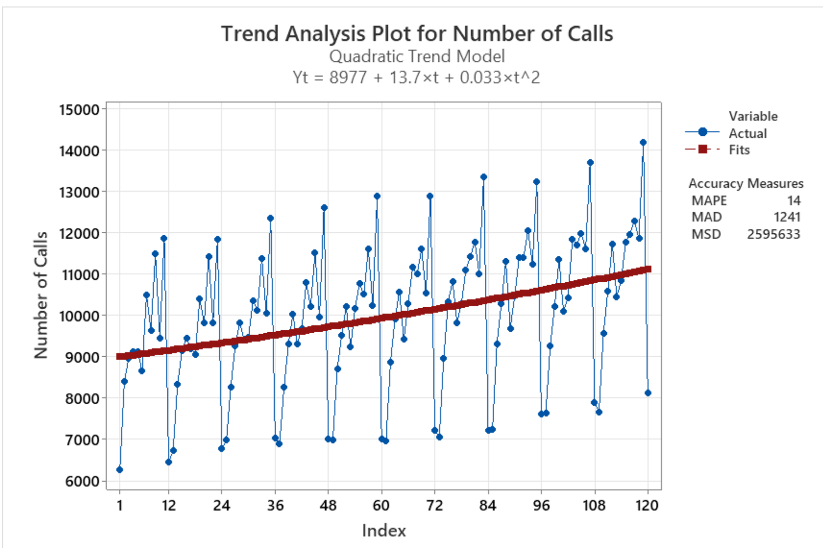
3. Winter’s method was run on this data in the accompanying Minitab file. The smoothing constants were estimated via trial and error to minimize Mean Squared Error (MSE). We obtain α=0.3185, β=0.15, and γ=0.624 for an MSE of 82812.3. Forecasts for the next 12 months using this model are tabulated below:

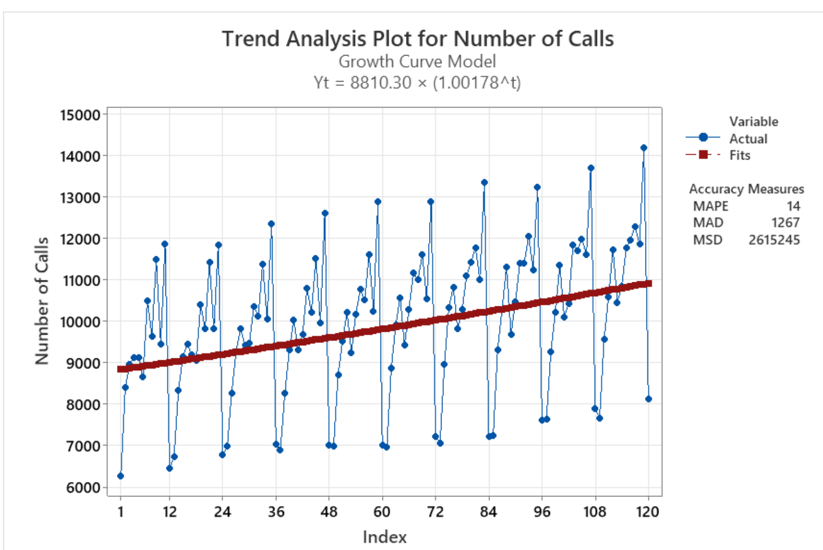
Month	Forecast (Rounded)
January, 2022	7862
February,2022	9797.1 (9797)
March, 2022	10915.5 (10916)
April, 2022	12046.3 (12046)
May, 2022	10689.4 (10689)
June, 2022	11192
July, 2022	12354.1 (12354)
August, 2022	12460.4 (12460)
September, 2022	12916.2 (12916)
October, 2022	12359.8 (12360)
November, 2022	14697.4 (14697)
December, 2022	8361

The forecasts seem reasonable. They follow the general pattern in the data in that the number of calls is minimum for January, December, and February, maximum for November, and rather stable across the remaining months.

4. Before decomposing our data, we ran a trend analysis on the data using a linear, quadratic, and exponential trend. The results are displayed below:

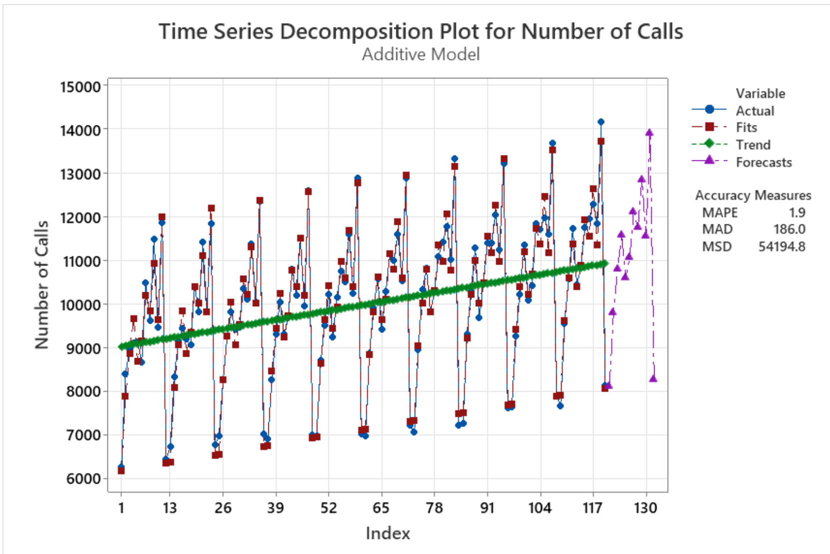






The plots above reveal that the exponential model has the highest errors out of all three. We also observe that the quadratic model has slightly better performance than the linear model. However, because the improvement is miniscule, and in line with the principle of parsimony, we choose the linear model.

We then run a decomposition model on the data. Because the variability seems rather constant, we go with an additive model. The results are displayed below:



The forecasts for 2022, i.e., the next 12 months are tabulated below:

Month	**Forecast (Rounded)**
January, 2022	8086.12 (8086)
February, 2022	9794.94 (9795)
March, 2022	10781.96 (10782)
April, 2022	11567.78 (11568)
May, 2022	10588.06 (11588)
June, 2022	11064.42 (11064)
July, 2022	12112.74 (12113)
August, 2022	11743.93 (11744)
September, 2022	12834.96 (12835)
October, 2022	11543.69 (11544)
November, 2022	13910.18 (13910)
December, 2022	8255.20 (8255)

The forecasts seem reasonable. They follow the general pattern in the data in that the number of calls is minimum for January, December, and February, maximum for November, and rather stable across the remaining months.

5. For this question, we consider the 2 types of regression models for forecasting with seasonality, the ***linear*** model and the ***exponential*** model. We run both on Minitab using Number of Calls as the dependent variable for the linear model, and a new variable, LogNumCalls as the dependent variable for the exponential model. For independent variables, we create 11 dummy variables for the 12 months of a year, with January being the base case. The dummy variables are run as categorical predictors, with time included as a continuous predictor.
   Results for the two models are presented below:

   - **Linear Model:**

     Regression Equation

     Number of Calls = 6158.6 + 15.962 t + 0.0 Feb_0
     + 1733 Feb_1 + 0.0 Mar_0 + 2674 Mar_1
     + 0.0 Apr_0 + 3346 Apr_1 + 0.0 May_0
     + 2467 May_1 + 0.0 Jun_0 + 2810 Jun_1
     + 0.0 Jul_0 + 3866 Jul_1 + 0.0 Aug_0
     + 3618 Aug_1 + 0.0 Sep_0 + 4535 Sep_1
     + 0.0 Oct_0 + 3390 Oct_1 + 0.0 Nov_0
     + 5680 Nov_1 + 0.0 Dec_0 + 14 Dec_1

     Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
238.620	98.29%	98.10%	97.84%

   - **Exponential Model:**

     Regression Equation

     LogNUMCalls = 3.80823 + 0.000699 t + 0.0 Feb_0
     + 0.09580 Feb_1 + 0.0 Mar_0
     + 0.13987 Mar_1
     + 0.0 Apr_0 + 0.16815 Apr_1 + 0.0 May_0
     + 0.13093 May_1 + 0.0 Jun_0
     + 0.14565 Jun_1 + 0.0 Jul_0
     + 0.18997 Jul_1 + 0.0 Aug_0
     + 0.17934 Aug_1
     + 0.0 Sep_0 + 0.21576 Sep_1 + 0.0 Oct_0
     + 0.16997 Oct_1 + 0.0 Nov_0
     + 0.25544 Nov_1 + 0.0 Dec_0
     + 0.00359 Dec_1

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0098321	98.61%	98.46%	98.24%

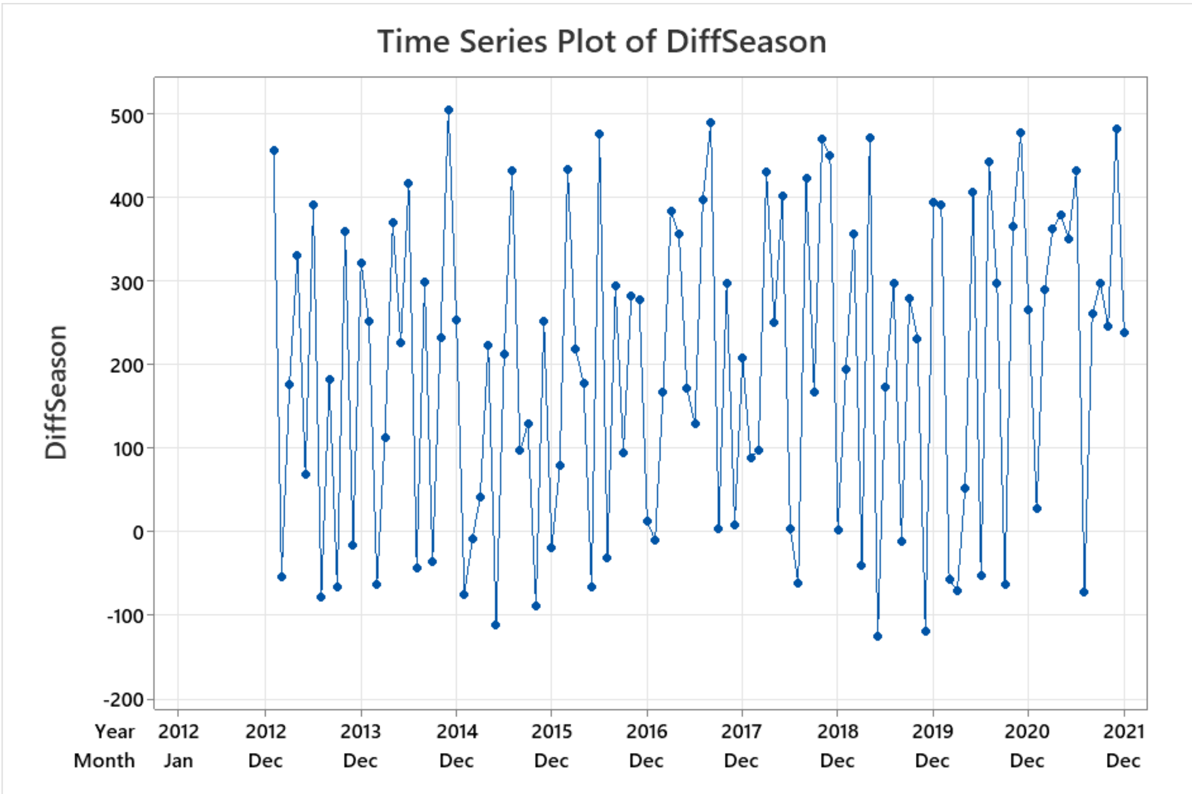
The results above show that the exponential model represents a better fit with higher R^2 and smaller standard error of estimate, S. Moreover, it has a larger predicted R^2, implying better predictive power. With that in mind, ***we proceed with the exponential model***.

We use the ***predict*** platform under regression to make forecasts for the next 12 months. The exponential model yields predictions for LogNumCalls, so to extract predictions for Number of Calls, we use Excel to obtain the values 10 to the power of LogNumCalls. The forecasts are tabulated below:

Month	**LogNumCalls**	**Number of Calls (Rounded)**
January, 2022	3.892819	7813.026 (7813)
February, 2022	3.98932	9757.083 (9757)
March, 2022	4.03409	10816.58 (10817)
April, 2022	4.06307	11562.99 (11562)
May, 2022	4.02655	10630.41 (10630)
June, 2022	4.04197	11014.63 (11015)
July, 2022	4.08698	12217.43 (12217)
August, 2022	4.07706	11941.53 (11942)
September, 2022	4.11417	13006.79 (13007)
October, 2022	4.06908	11724.11 (11724)
November, 2022	4.15525	14297.17 (14297)
December, 2022	3.9041	8018.627 (8019)

The forecasts seem reasonable. They follow the general pattern in the data in that the number of calls is minimum for January, December, and February, maximum for November, and rather stable across the remaining months.

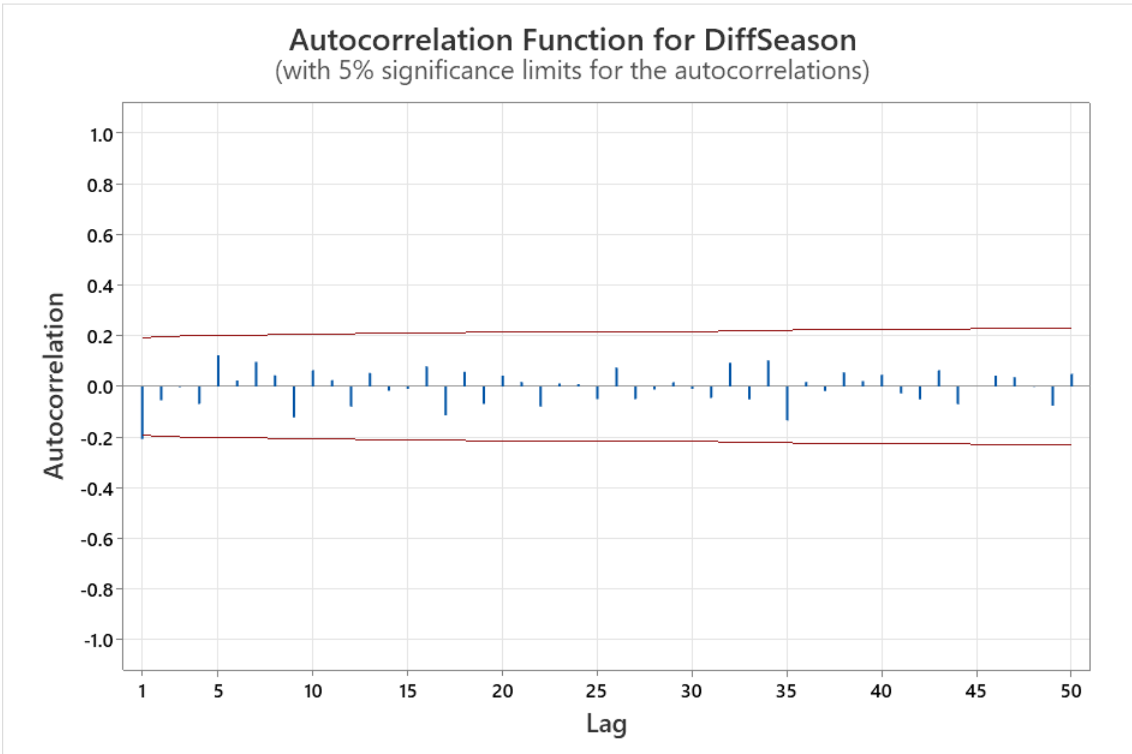
6. Partial autocorrelations, regular differences, and seasonal differences for Number of Calls are provided in the accompanying Minitab file. As part of model selection, because our data is seasonal, as proved in part 2, we will use seasonally different data for ARIMA model selection. We do not need to carry out a second regular differencing on the seasonally differenced data because the first differencing yields a stationary series, as evidenced by the time series plot below.

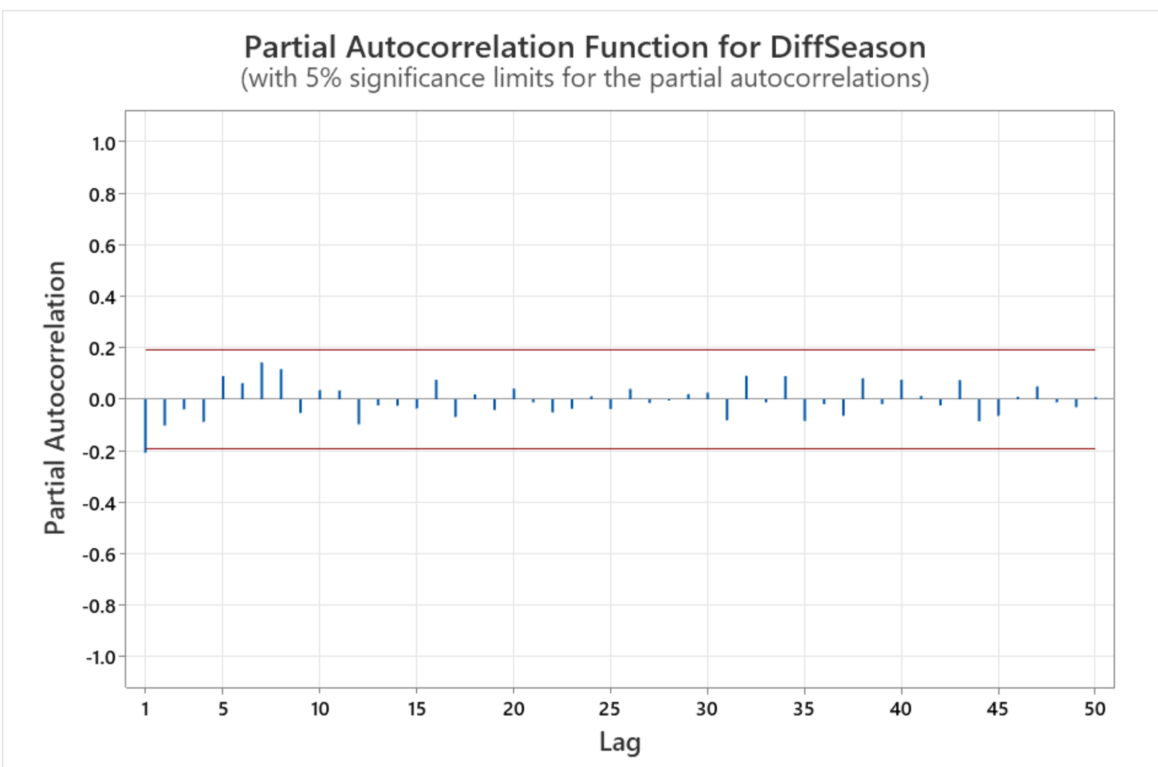


7. The ARIMA model is built as follows:

A. **Model Identification**:

We first present the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the seasonally differenced data.





Seasonality does not suggest itself, as evidenced by the lack of significant ACF or PACF at lags 12, 24, 36, 48, etc., i.e., at multiples of the monthly lag 12. We infer that for any model we choose, P=0, D=1, and Q=0.

We, therefore, proceed to search for patterns suggesting AR or MA models. The PACF is straightforward; there is a significant partial autocorrelation at lag 1, and the lags then seem to die out gradually. This eliminates the possibility for pure AR models.

For ACF, there are three patterns we can glean. Either:

-(1) There is significant autocorrelation at lag 1, after which the ACF cut off.

-(2) The ACF cuts off after the first 2 lags.

-(3) The ACF dies out gradually.

The above, coupled with the dying out PACF, suggests three different models we can evaluate: MA (1) for scenario 1, MA (2) for scenario 2, and ARMA (1,0,1) for scenario 3.

B. Model Building:

(1) ARIMA (0,0,1) $(0,1,0)_S$

Results for this model are displayed below:

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.2547	0.0945	2.70	0.008
Constant	194.7	13.1	14.86	0.000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 120, after differencing 108

Residual Sums of Squares

DF	SS	MS
106	3535405	33352.9

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	7.89	11.11	16.58	20.59
DF	10	22	34	46
P-Value	0.640	0.973	0.995	1.000

(2) ARIMA (0,0,2) (0,1,0)$_S$

Results for this model are displayed below:

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
MA 1	0.2376	0.0976	2.43	0.017
MA 2	0.0612	0.0993	0.62	0.539
Constant	194.5	12.4	15.72	0.000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 120, after differencing 108

Residual Sums of Squares

DF	SS	MS
105	3522717	33549.7

Back forecasts excluded

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	7.39	10.85	16.85	20.57
DF	9	21	33	45
P-Value	0.597	0.966	0.991	0.999

The variable MA2 is insignificant at the 95% confidence level with a p-value of 0.539. The model is deemed inadequate and is excluded from further analysis.

(3) ARIMA (1,0,1) (0,1,1)$_S$

Results for this model are presented below. The model was run without a constant term because the inclusion of one makes the variables insignificant.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.99968	0.00110	907.37	0.000
MA 1	0.996427	0.000277	3592.00	0.000

Differencing: 0 regular, 1 seasonal of order 12

Number of observations: Original series 120, after differencing 108

Residual Sums of Squares

DF	SS	MS
106	3763421	35504.0

Back forecasts excluded

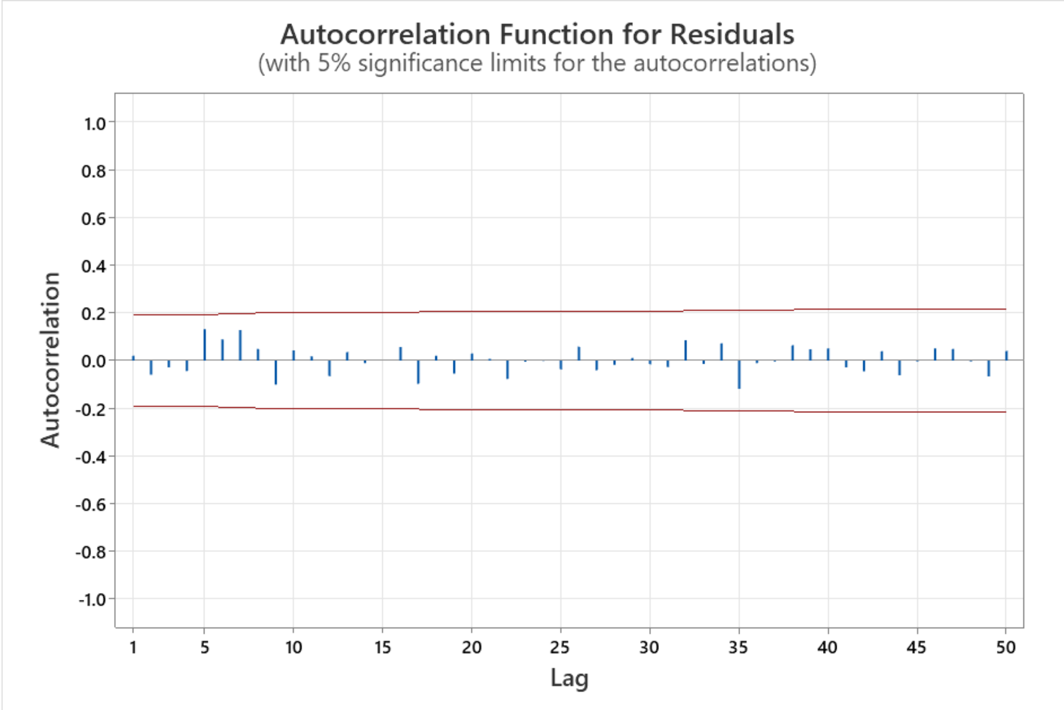
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	11.74	16.90	25.18	29.03
DF	10	22	34	46
P-Value	0.303	0.769	0.864	0.976

C. **Model Checking:**

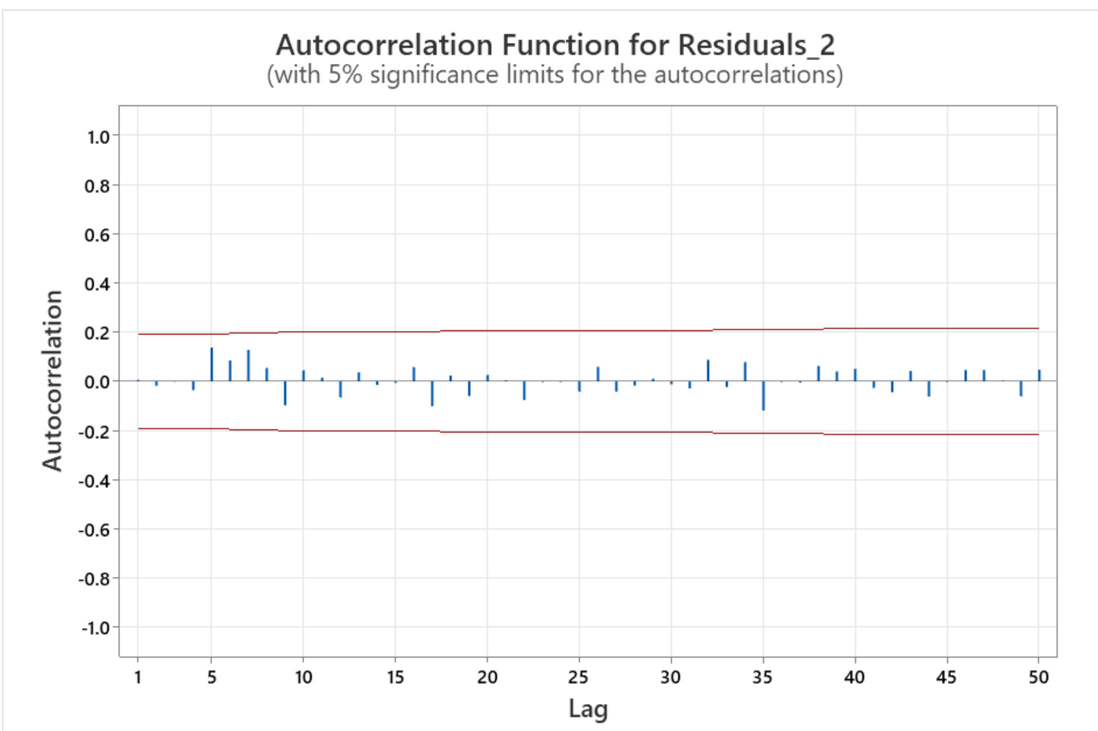
(1) ARIMA (0,0,1) (0,1,0)$_S$

- With p-values<0.05, estimated parameters are significant at the 95% confidence level.
- The p-values for the residuals as a group are sufficiently large, i.e., consistent with those produced by random errors.
- The check for individual residuals is done using the ACF carried out on residual. The function, displayed below, reveals that all errors are small, and there is no significant autocorrelation at any lag, which is consistent with independent errors.



(3) ARIMA (1,0,1) (0,1,0)s

- With p-values<0.05, estimated parameters are significant at the 95% confidence level.
- The p-values for the residuals as a group are sufficiently large, i.e., consistent with those produced by random errors.
- The check for individual residuals is done using the ACF carried out on residual. The function, displayed below, reveals that all errors are small, and there is no significant autocorrelation at any lag, which is consistent with independent errors.



Both models yield good results and are therefore adequate. However, we will limit our analysis to one model. Model 1 has a smaller MSE and is more parsimonious and is therefore our choice for this analysis.

The model we will be using to produce forecasts is the ARIMA (0,0,1) $(0,1,0)_S$.

D. Forecasting with the Model:

Forecasts for 2022, i.e., the next 12 months using our model of choice are tabulated below:

Month	**Forecast (Rounded)**
January, 2022	7826.417 (7826)
February, 2022	9738.694 (9739)
March, 2022	10767.69 (10768)
April, 2022	11912.69 (11913)
May, 2022	10633.69 (10634)
June, 2022	11032.69 (11033)
July, 2022	11951.69 (11952)
August, 2022	12149.69 (12150)
September, 2022	12470.69 (12471)
October, 2022	12040.69 (12041)
November, 2022	14365.69 (14366)
December, 2022	8311.694 (8312)

The forecasts seem reasonable. They follow the general pattern in the data in that the number of calls is minimum for January, December, and February, maximum for November, and rather stable across the remaining months.

8. Accuracy measures for Winter's smoothing, decomposition, and the ARIMA model are tabulated below. For the ARIMA model, Minitab results only provide the MSE, so we use fits to calculate MAPE and MAD manually using Excel.

Model	MAPE	MAD	MSE
Winter's Method	2.244%	216.2	82812.325
Decomposition	1.908%	186	54194.786
ARIMA $(0,0,1)(0,1,0)_S$	1.609%	158.691	33352.873

As evidenced by the values in the table, the ARIMA model has the smallest MAPE, MAD, and MSE out of the three. It is, therefore, our preferred model and the model we'll use for forecasting next year's monthly calls.

9. The table below shows the number of monthly calls for 2021 and the monthly forecasts for 2022 using our preferred model, the ARIMA model.

Month	**Value for 2021**	**Forecast for 2022 (Rounded)**
January	7663	7826
February	9544	9739
March	10573	10768
April	11718	11913
May	10439	10634
June	10838	11033
July	11757	11952
August	11955	12150
September	12276	12417
October	11846	12041
November	14171	14366
December	8117	8312

According to the forecasts, the number of calls is expected to grow for each month in 2022. If in 2021 employees are already overwhelmed by the number of calls, then they are expected to be even more overwhelmed in 2022 with the increase in demand.

With that in mind, there is a need for additional staff to handle the growing demand across all months of 2022.