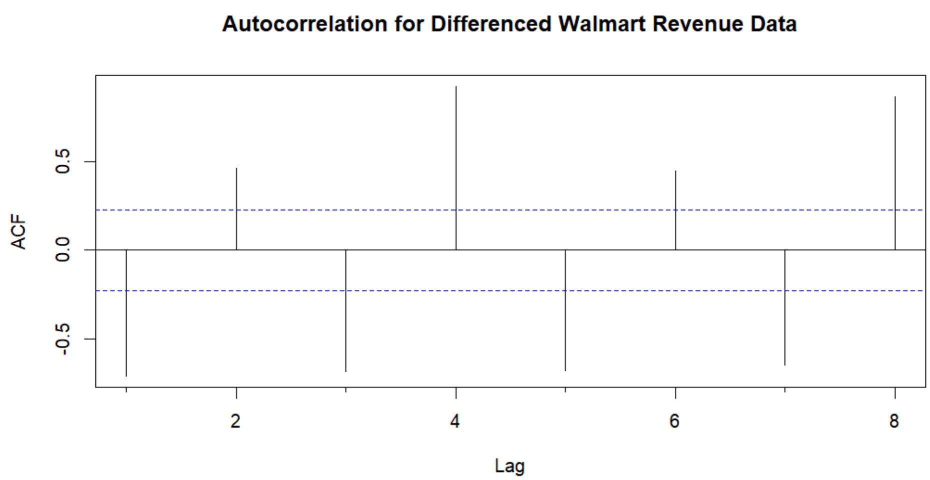
Case Study #3 - ARIMA Models

1. Identify time series predictability.
    a. Based on the p-value test, the historical data is not predictable since the null hypothesis is accepted.

```
Series: revenue.ts
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1      mean
      0.9449  120267.7
s.e.  0.0444   16666.3

sigma^2 = 95387002:  log likelihood = -806.13
AIC=1618.27   AICc=1618.6   BIC=1625.26

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE     MASE       ACF1
Training set 1022.498 9637.262 8133.299 0.2702847 6.990888 1.684678 -0.6532536
```

```
[1] "Accept null hypothesis"
```

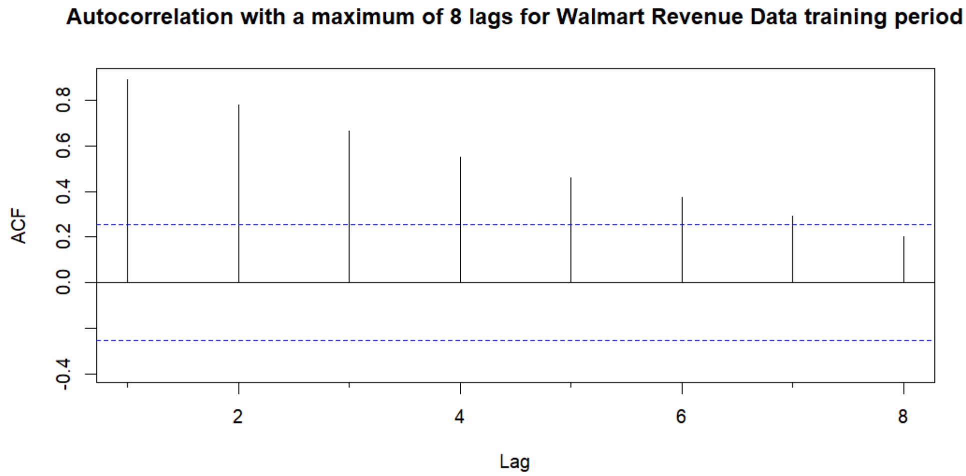
    b. Based on the autocorrelation plot of the historical data, we can see that the historical data is predictable since the lag 1 differencing exceeds the horizontal lines for all lags. This means that there is useful information in the dataset that can be modeled and predicted upon.



2. Apply the two-level forecast with a regression model and AR model for residuals.
    a. Validation revenue forecast using a regression model with linear trend and seasonality.

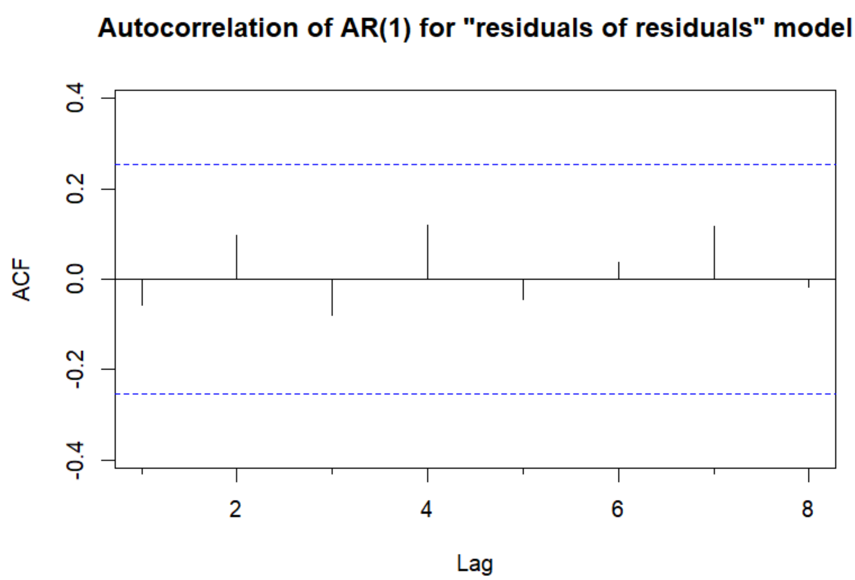
```
        Point Forecast
2020 Q1       131681.2
2020 Q2       136857.2
2020 Q3       135273.8
2020 Q4       148512.4
2021 Q1       135075.7
2021 Q2       140251.8
2021 Q3       138668.4
2021 Q4       151907.0
2022 Q1       138470.2
2022 Q2       143646.3
2022 Q3       142062.9
2022 Q4       155301.5
2023 Q1       141864.7
2023 Q2       147040.8
2023 Q3       145457.4
2023 Q4       158696.0
```

b. Since it looks like there is still information available that would help us to better predict, I believe it is a good idea to add an AR model for residuals.



c. The model equation for AR(1) model for regression residuals is as follows: $e_t = -\ 2438.784\ +\ 0.9502\ e_{t-1}$ which expresses the linear trend component of residuals.

```
Series: train.lin.season$residuals
ARIMA(1,0,0) with non-zero mean

Coefficients:
         ar1       mean
      0.9502  -2438.784
s.e.  0.0380   3254.527

sigma^2 = 2304385:  log likelihood = -524.79
AIC=1055.59   AICc=1056.02   BIC=1061.87

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE    MASE        ACF1
Training set 181.7393 1492.505 1117.175 -17.85092 104.4523 0.44047 -0.05713674
```

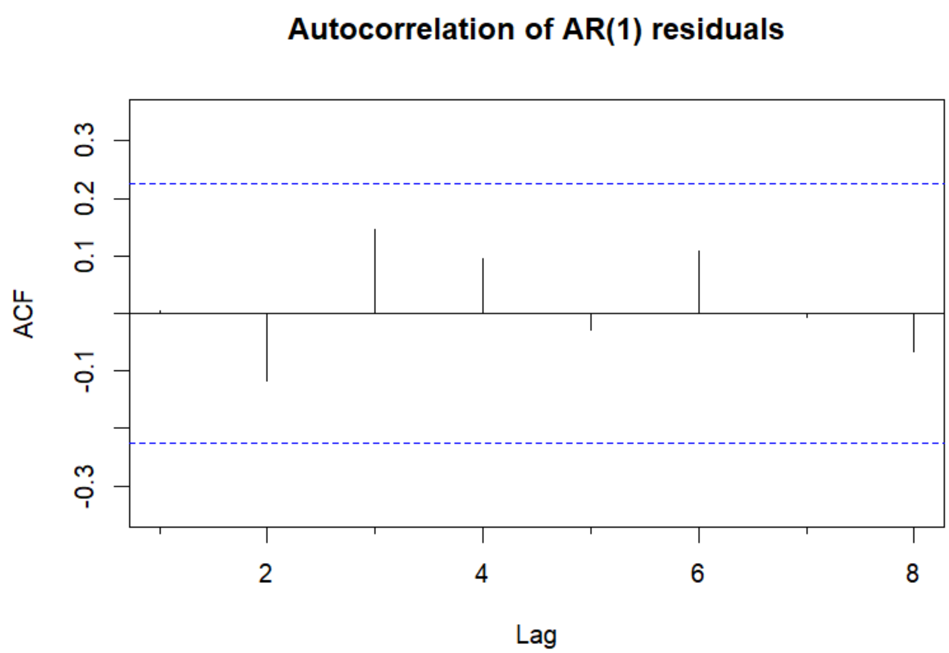


Based on the Autocorrelation plot, it seems that all the useful information of the residuals has been captured by the model since all lags are below the significance threshold.

d. Table with validation data, regression forecast for the validation data, AR(1) forecast for the validation data, and combined forecast for the validation period.

	Revenue	Reg.Forecast	AR(1)Forecast	Combined.Forecast
1	134622	131681.2	-3396.695	128284.5
2	137742	136857.2	-3348.981	133508.2
3	134708	135273.8	-3303.644	131970.2
4	152079	148512.4	-3260.564	145251.9
5	138310	135075.7	-3219.631	131856.1
6	141048	140251.8	-3180.737	137071.0
7	140525	138668.4	-3143.780	135524.6
8	152871	151907.0	-3108.663	148798.3
9	141569	138470.2	-3075.296	135394.9
10	152859	143646.3	-3043.591	140602.7
11	152813	142062.9	-3013.466	139049.4
12	164048	155301.5	-2984.840	152316.6
13	152301	141864.7	-2957.641	138907.1
14	161632	147040.8	-2931.796	144109.0
15	160804	145457.4	-2907.239	142550.2
16	173388	158696.0	-2883.905	155812.1

e. Autocorrelation chart for the AR(1) model's residuals shows that all the useful information of the residuals has been captured by the model since all lags are below the significance threshold. Table of forecasts below.



	Reg.Forecast	AR(1)Forecast	Combined.Forecast
1	149976.4	9326.569	159302.9
2	155456.4	8780.763	164237.1
3	153973.1	8268.571	162241.7
4	167242.3	7787.922	175030.2
5	153738.8	7336.875	161075.7
6	159218.8	6913.605	166132.4
7	157735.6	6516.403	164252.0
8	171004.8	6143.662	177148.5

3. Use ARIMA Model and Compare Various Methods.
   a. Model Equation which captures AR(1) and order 1 seasonality, order 1 differencing, order 1 moving average for error lags and quarterly seasonality.

$$y_t + y_{t-1} = -0.7543(y_{t-1} - y_{t-2}) + 0.6701(y_{t-2} - y_{t-3}) + 0.2202\varepsilon_{t-1} - 0.8295\varepsilon_{t-2}$$

```
Series: train.ts
ARIMA(1,1,1)(1,1,1)[4]

Coefficients:
          ar1     ma1    sar1     sma1
      -0.7543  0.6701  0.2202  -0.8295
s.e.   0.3046  0.3200  0.1994   0.1718

sigma^2 = 2687313:  log likelihood = -484.6
AIC=979.2   AICc=980.43   BIC=989.24

Training set error measures:
                    ME     RMSE      MAE        MPE      MAPE      MASE
Training set -301.8863 1511.362 1067.091 -0.2885631 0.9726523 0.2674402
                     ACF1
Training set -0.006241409
```

	Point Forecast
2020 Q1	127543.4
2020 Q2	133248.8
2020 Q3	131191.6
2020 Q4	144646.8
2021 Q1	130767.7
2021 Q2	136244.2
2021 Q3	134308.1
2021 Q4	147677.1
2022 Q1	133880.5
2022 Q2	139285.6
2022 Q3	137391.9
2022 Q4	150730.1
2023 Q1	136960.6
2023 Q2	142343.3
2023 Q3	140464.0
2023 Q4	153791.6

   b. Model Equation which captures no autogressiveness, order 1 differencing, order 1 seasonality, order 1 moving average for error lag and quarterly seasonality. $y_t - y_{t-1} = -0.6785\rho_{t-1}$

```
Series: train.ts
ARIMA(0,1,0)(0,1,1)[4]

Coefficients:
         sma1
      -0.6785
s.e.   0.1605

sigma^2 = 2698549:  log likelihood = -485.99
AIC=975.98   AICc=976.22   BIC=980

Training set error measures:
                   ME     RMSE      MAE        MPE     MAPE      MASE      ACF1
Training set -242.4378 1558.427 1146.186 -0.2410367 1.039209 0.2872633 -0.134278
```

	Point Forecast
2020 Q1	127552.1
2020 Q2	133115.3
2020 Q3	131035.1
2020 Q4	144424.5
2021 Q1	130305.6
2021 Q2	135868.8
2021 Q3	133788.6
2021 Q4	147178.0
2022 Q1	133059.1
2022 Q2	138622.3
2022 Q3	136542.1
2022 Q4	149931.6
2023 Q1	135812.7
2023 Q2	141375.9
2023 Q3	139295.6
2023 Q4	152685.1

   c. Based on MAPE and RMSE, the ARIMA(1,1,1)(1,1,1) would be the model to apply as it outperforms the auto ARIMA model.

	Model	MAPE	RMSE
1	ARIMA Seasonal	6.948	12111.80
2	Auto ARIMA	7.350	12807.88

d. ARIMA(1,1,1)(1,1,1) / Auto ARIMA

ARIMA(1,1,1)(1,1,1)

	Point Forecast
2024 Q1	161000.0
2024 Q2	167250.1
2024 Q3	165899.6
2024 Q4	179026.0
2025 Q1	166537.2
2025 Q2	172199.2
2025 Q3	170749.9
2025 Q4	183981.8

Auto ARIMA

	Point Forecast
2024 Q1	161241.9
2024 Q2	169400.0
2024 Q3	168846.9
2024 Q4	181029.5
2025 Q1	169197.9
2025 Q2	178445.2
2025 Q3	177950.9
2025 Q4	189995.2

e. Based on MAPE and RMSE the best model to apply for forecasting revenue is the ARIMA(1,1,1)(1,1,1) model. Runner ups are the two-level model with AR(1) for residuals and Auto ARIMA.

	Model	MAPE	RMSE
1	Reg Linear Seasonal	3.428	4700.120
2	Two-level model (with AR(1) model for residuals)	1.187	1899.501
3	ARIMA(1,1,1)(1,1,1) model	1.048	1840.715
4	Auto ARIMA	1.233	2109.946
5	Seasonal Naive	4.081	5863.128