**Forecasting Monthly Sales – Video games Company**

**Overview.**

The supply chain department of a video games company has been tasked to forecast monthly sales data in order to synchronize supply with demand, aid in decision making that will help build a competitive infrastructure and measure company performance. The supply chain analyst has been assigned to help the manager run the numbers through a time series forecasting model.

A forecast for the next 4 months of sales is needed and findings reported.

**Analysis planning**.

**Does the dataset meet the criteria of a time series?**

Four defining attributed to a time series:

Is the data over a continuous time interval?

**Yes** – The data is ordered over a continuous period and in YYYY-MM format.

Are there sequential measurements across that interval?

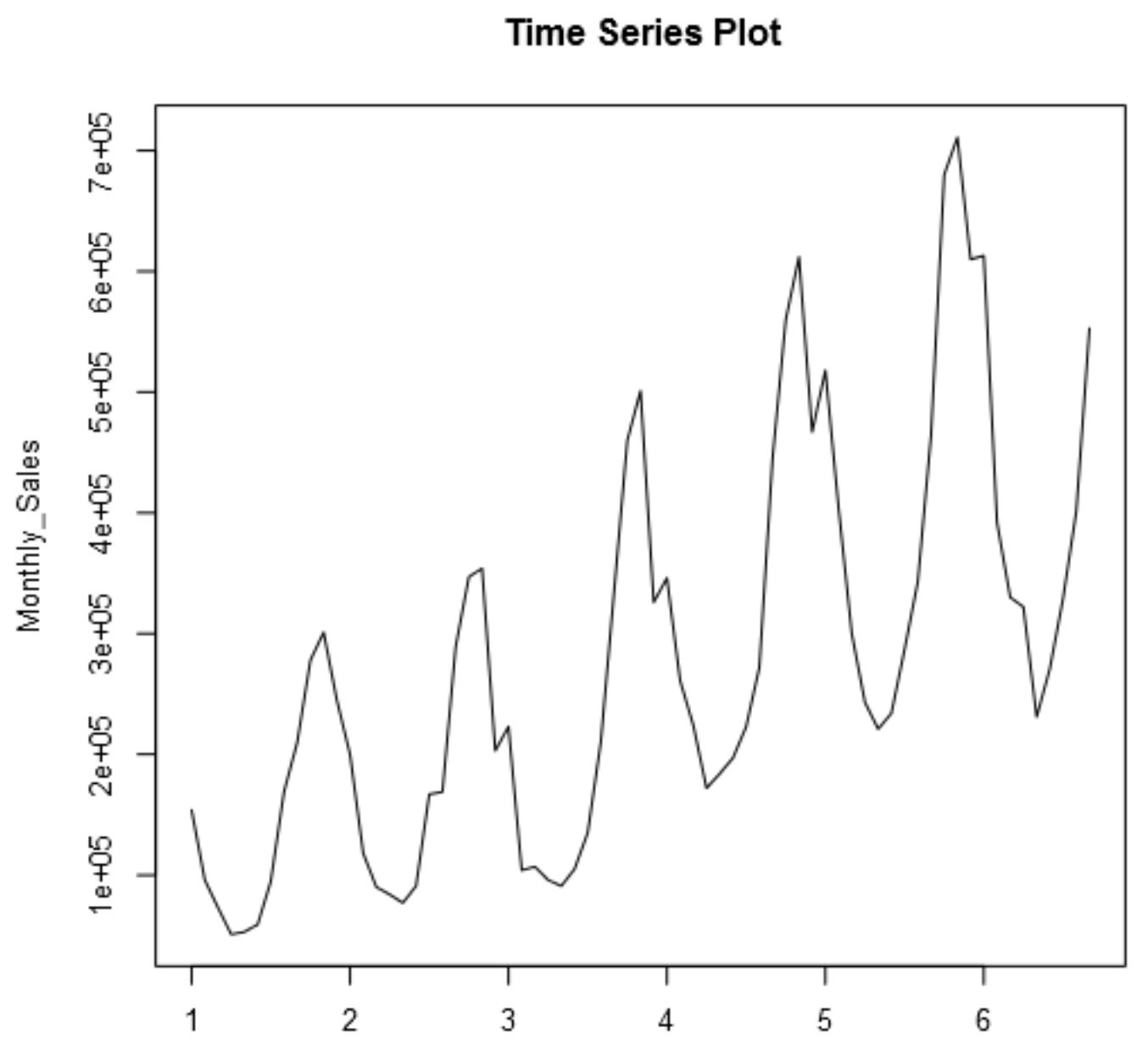
**Yes** – The date is ordered chronologically by month and year.

Is there equal spacing between every two consecutive measurements?

**Yes** – Each data point is separated by monthly intervals, with 12 intervals per year.

Does each time unit within the time interval has at most one data point?

**Yes** – There are data points for each time interval.

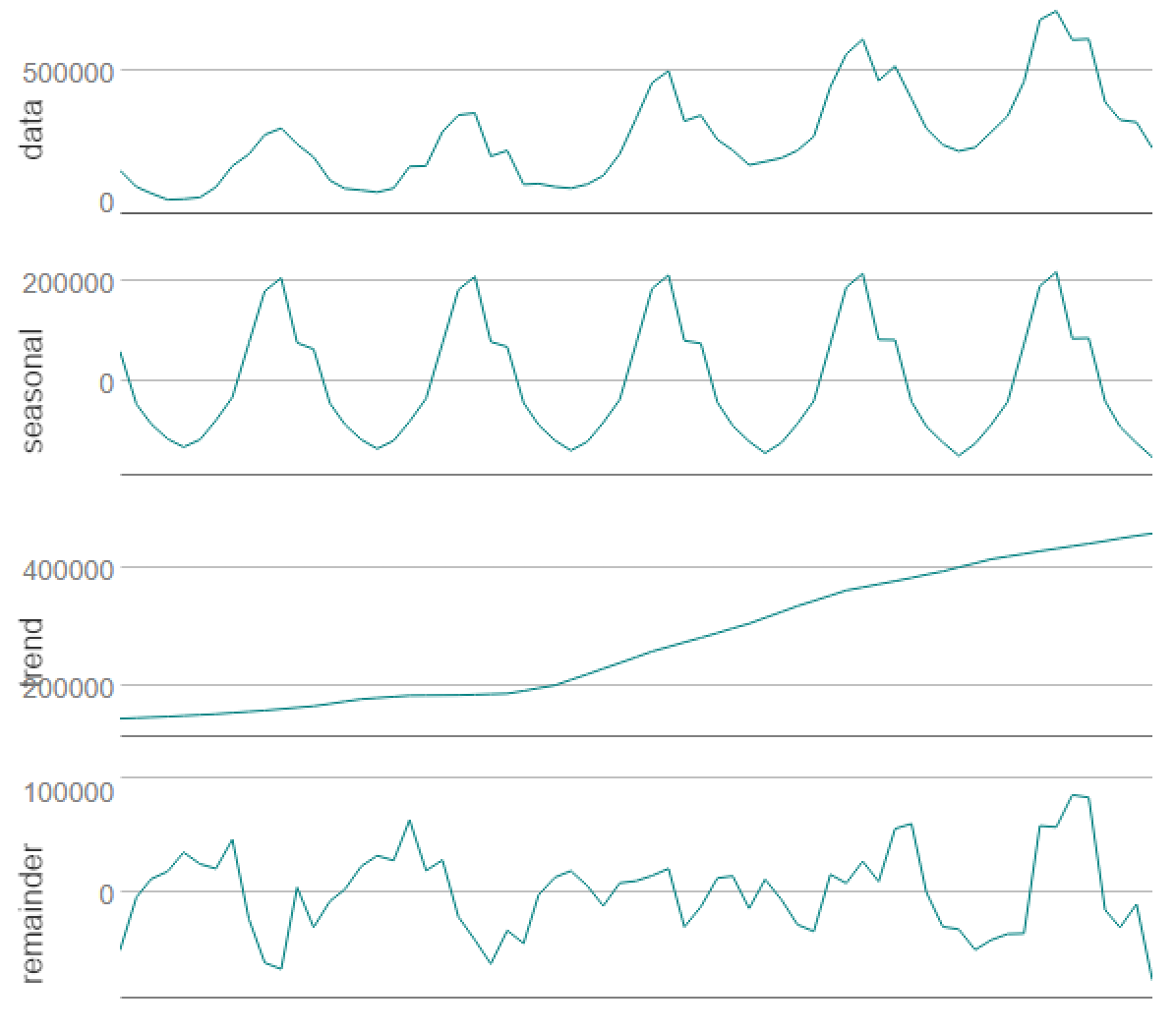


**Which records should be used for the holdout period?**

The final forecast will be to predict sales for the next 4 months, therefore a holdout period of at least 4 months will be used in our model for validation. The holdout period will be the final four months of the dataset, ie from 2013-06 to 2013-09.

**Determine the trend, seasonal and error components.**

Below is the decomposition plot for the dataset.



**Trend.**

Trend here looks to be upward and looks to be linear.

**Seasonality.**

The seasonal chart looks to have peaks and troughs over the same intervals. The peaks and troughs also look to be increasing as time increases.

**Error.**

The remainder chart looks to have a pretty irregular pattern over time. There doesn’t seem to be any trend here.

**Build the model.**

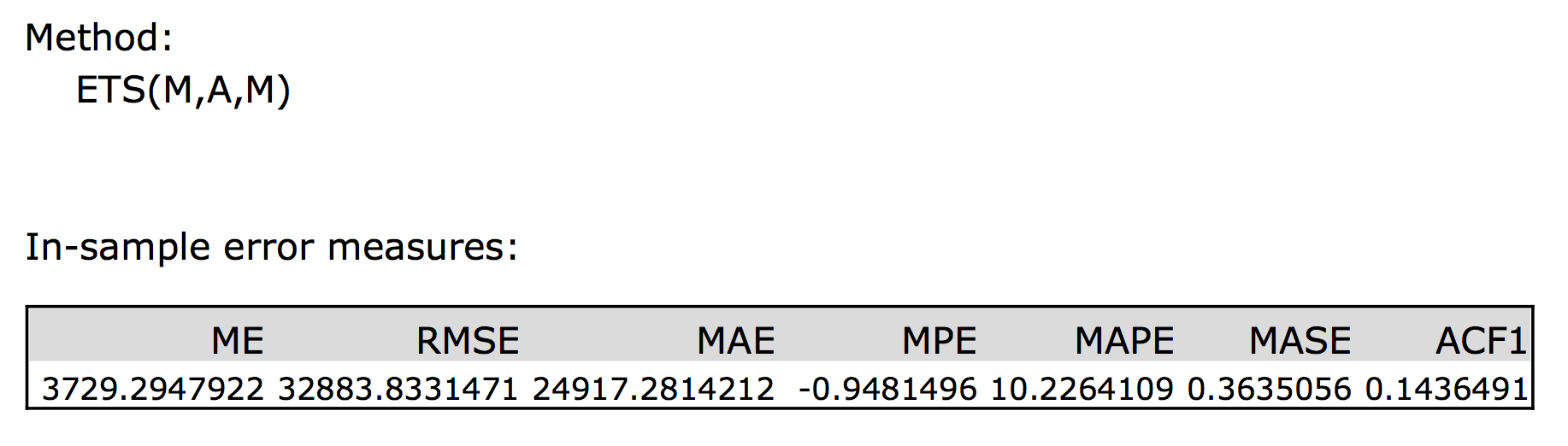
**Model Terms for ETS.**

Using the decomposition charts above, below are my model terms for ETS.

|  |  |  |
| --- | --- | --- |
| **Term** | **Component** | **Reason** |
| Error | Multiplicative | Inconsistent chart patterns. |
| Trend | Additive | Upward and linear |
| Seasonality | Multiplicative | Increasing peaks and troughs. |

**What are the in-sample errors?**

Below is the in-sample error.

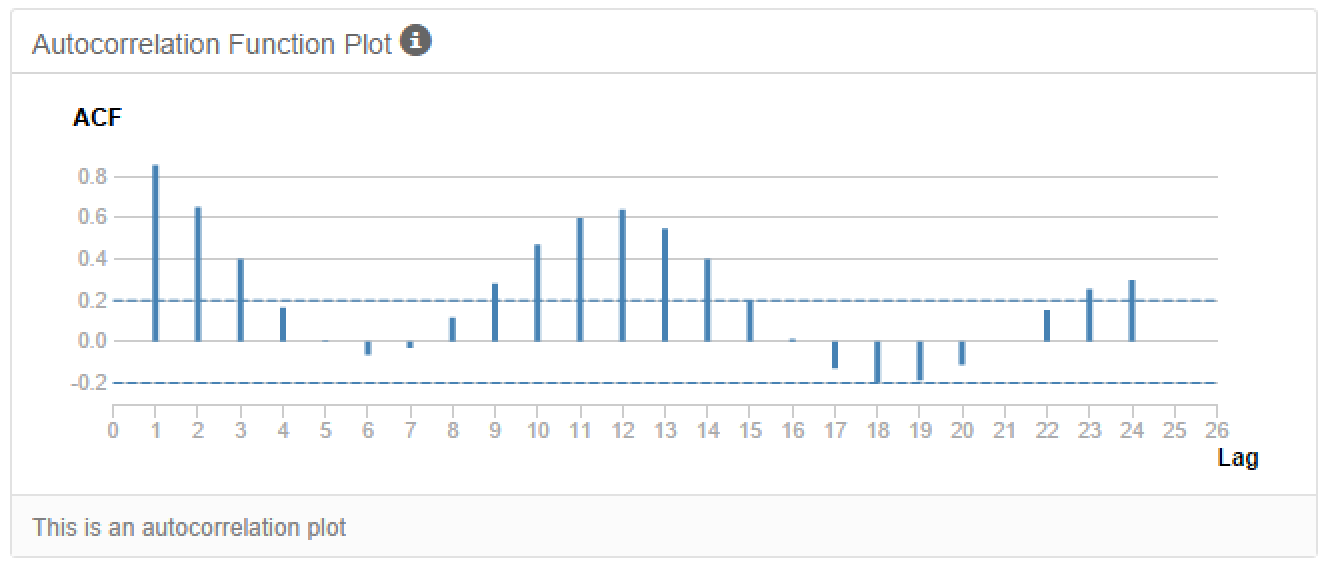


The ETS model shows an RMSE or variance of the error of 32883.83 and a MASE 0.3635. MASE value is under the 1.0 threshold which is good.

**ARIMA Model.**

The original time series plot showed signs of seasonality and so differencing for the 12 monthly sales figures will be done. This is done to try to remove seasonality from the sales figures.

Below is the ACF and PACF plot for the time series.



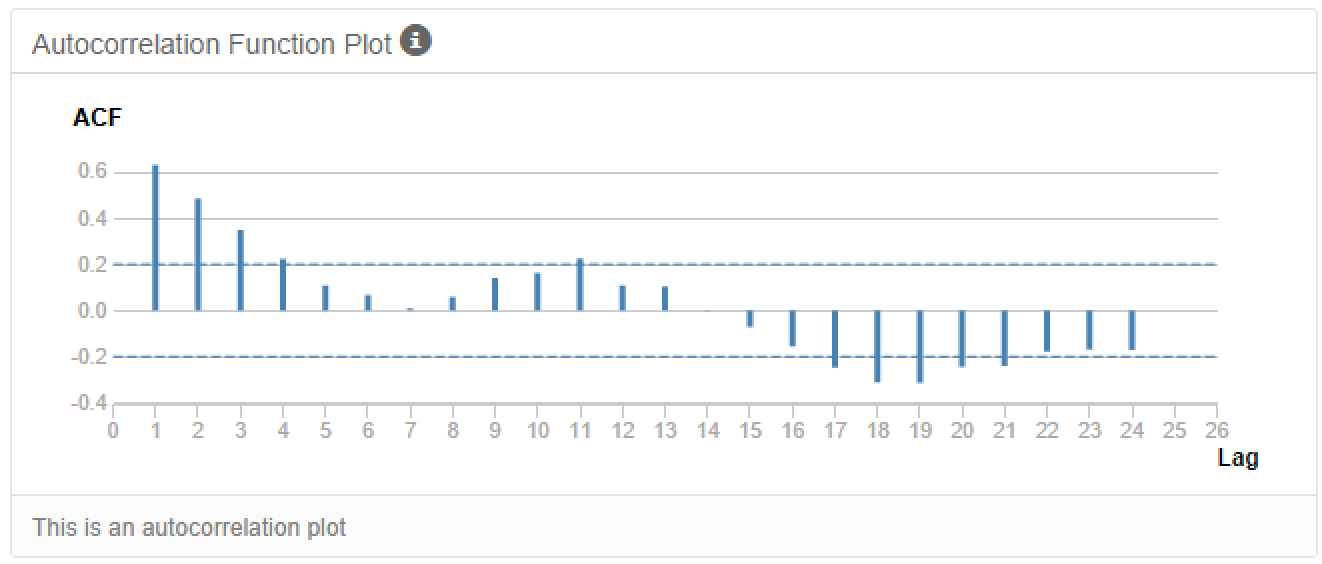


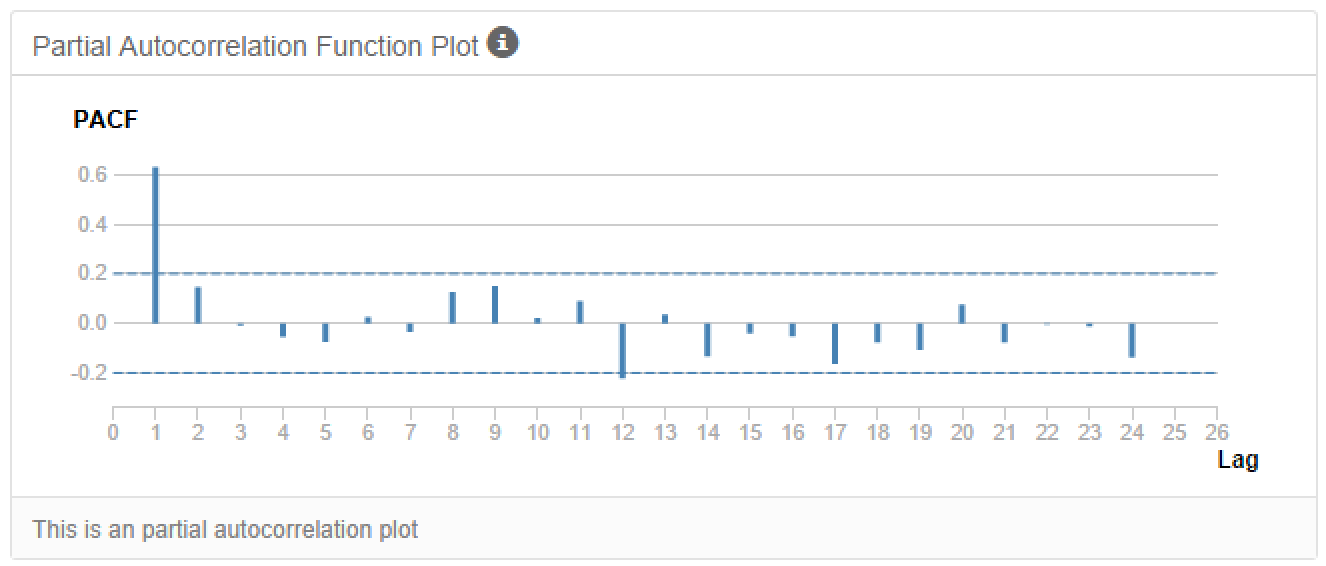
The ACF plots shows a positive correlation at lag 1 and then gradually drops towards zero, with increases again around seasonal patterns.

The PACF plot shows a positive correlation at lag 1 and then suddenly drops off.

Due to high serial correlation, seasonal differencing will be done.

Below is the ACF and PACF plot for the time series adjusted for seasonality (12 months differencing)



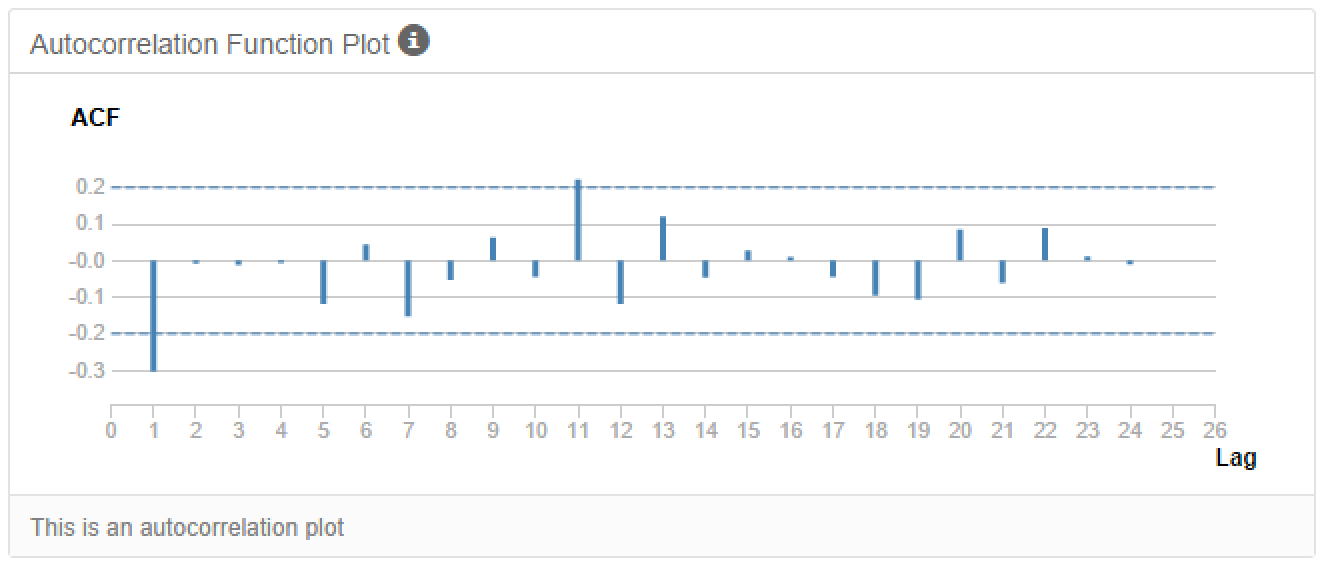


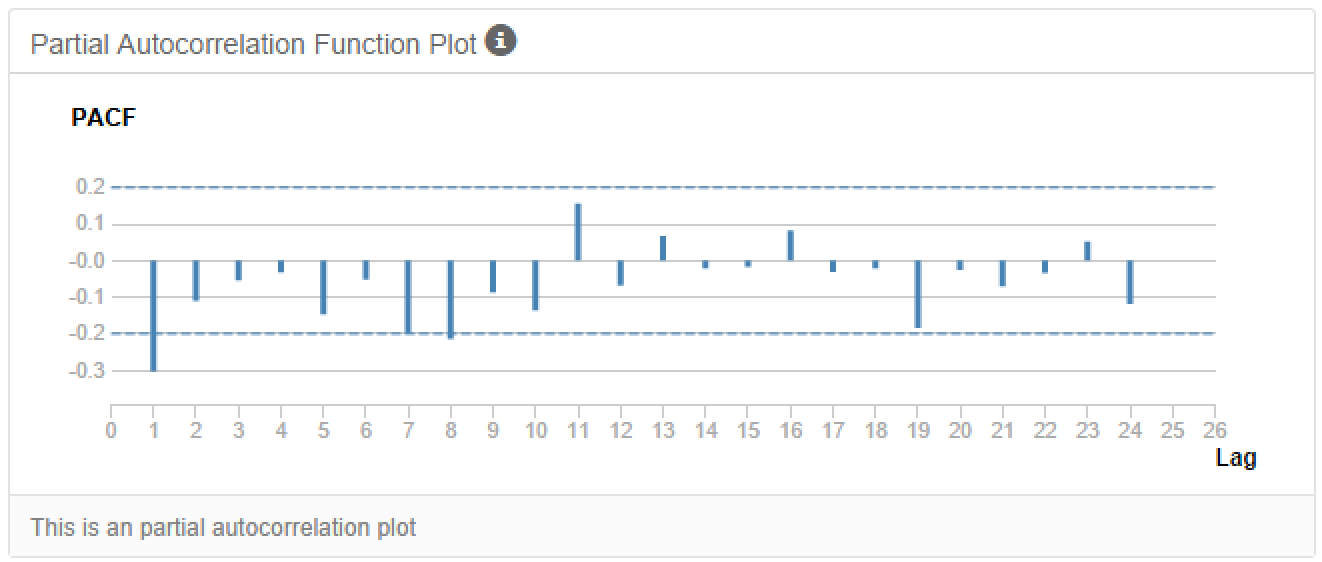
The seasonal ACF plot shows positive correlation at lag 1 and decreases towards 0 with further seasonal patterns.

The PACF shows positive correlation at lag 1 and then suddenly drops off.

Further differencing will need to be done in order to further remove any serial correlation.

Below is the seasonal ACF and PACF plots with 1st degree differencing.





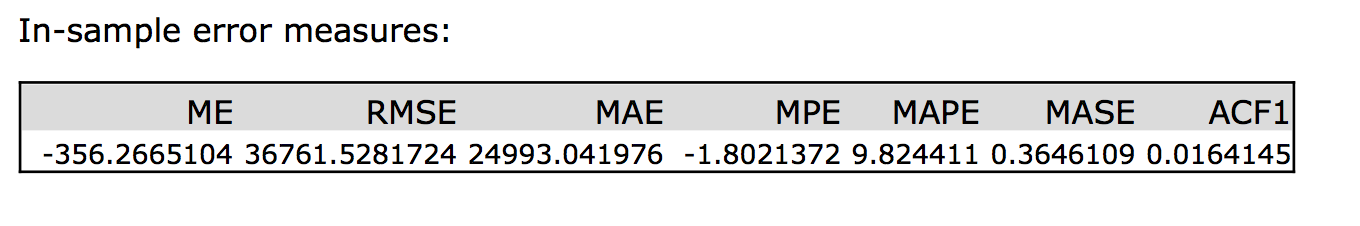
The ACF plot shows negative correlation at lag 1 then drops off.

The PACF shows negative correlation at lag 1 as well, however both the charts now look to have removed any serial correlation.

**Below are the model terms for the ARIMA model.**

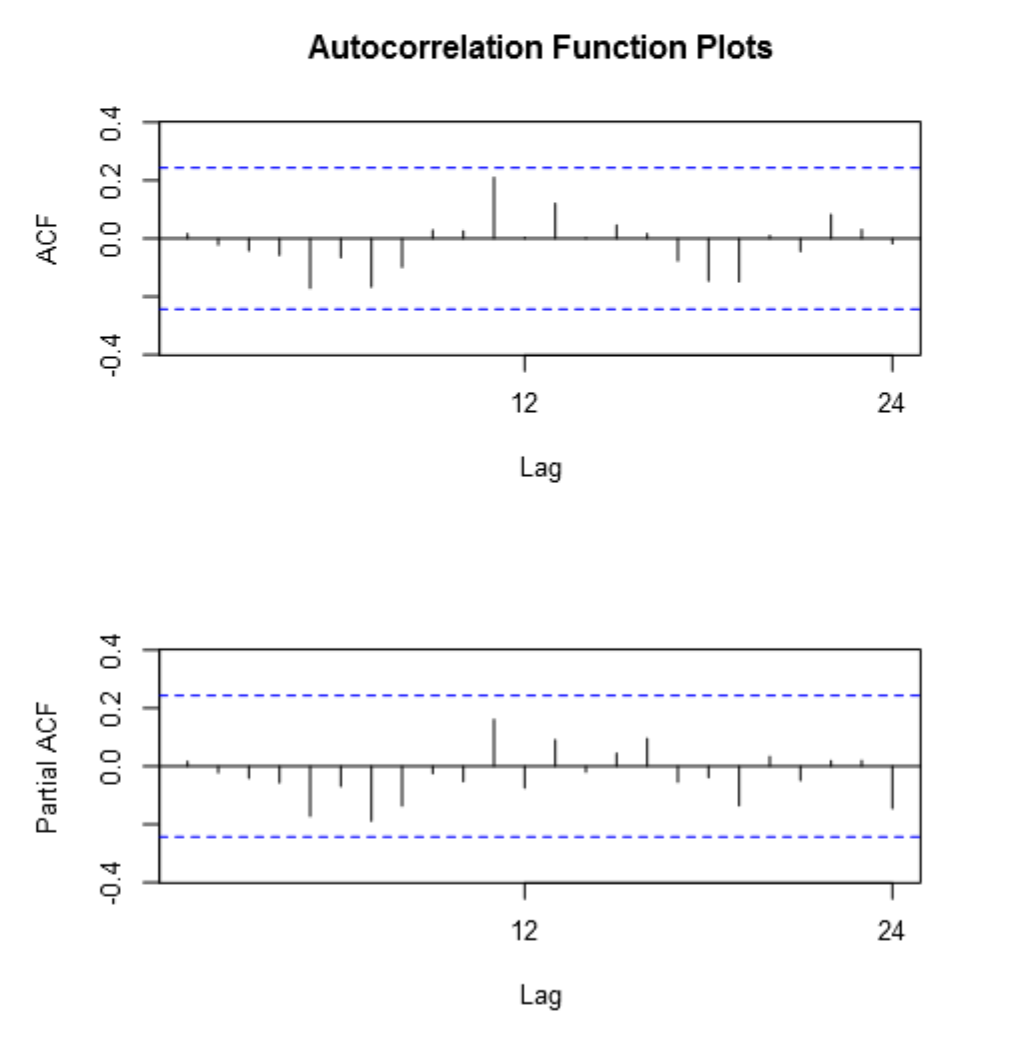
|  |  |  |
| --- | --- | --- |
| **Term** | **Component** | **Reason** |
| AR (p) | 0 | No positive correlation |
| I (d) | 1 | 1st degree differencing used. |
| MA (q) | 1 | Negative correlation at lag 1 |
|  |  |  |
| AR (P) | 0 | No real season correlation at lag 12, 24 |
| I (D) | 1 | Seasonal difference used |
| MA (Q) | 0 | No real season correlation at lag 12, 24 |
|  |  |  |
| m | 12 | Monthly data |

Below is the summary of the In-Sample Error.



The RSME value is 36761.5 and the MASE is 0.3646 which is below the 1.0 threshold.

**Below are the new ACF and PACF plots from the model.**

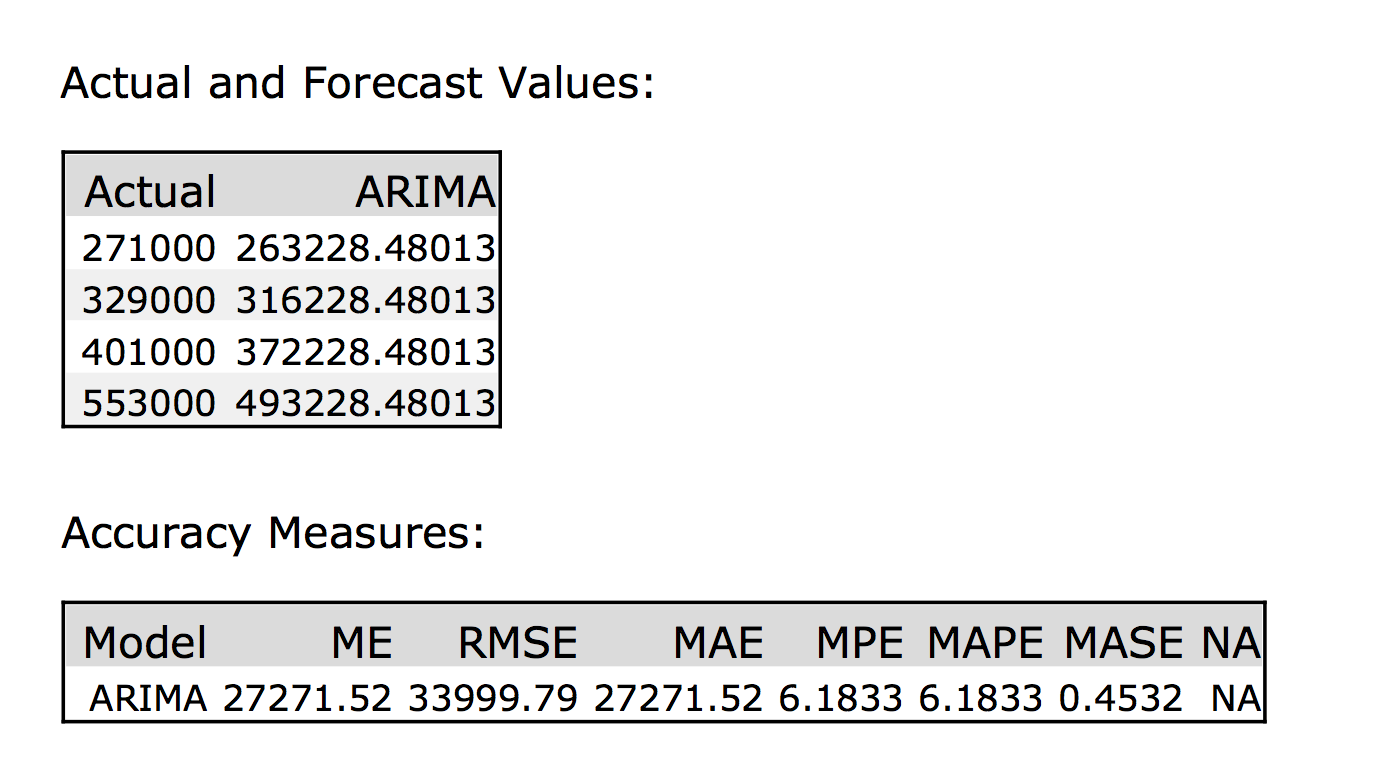


**Forecast.**

**Which model should I choose?**

Looking at the two models, I would choose the ARIMA model.

Below is the ARIMA measurements against the hold out periods.



**Forecast for the next four periods and the forecast graph at 95% and 80% confidence intervals.**

