Project: Forecasting Sales

Step 1: Plan Your Analysis

Look at your data set and determine whether the data is appropriate to use time series models. Determine which records should be held for validation later on (250 word limit).

Answer the following questions to help you plan out your analysis:

1. Does the dataset meet the criteria of a time series dataset? Make sure to explore all four key characteristics of a time series data.

Attributes of a Time Series Data set:

1. It's over a continuous time interval

   Yes, the data is over a continuous period with YYYY- MM.

2. There are sequential measurements across that interval

   The Data is ordered by YYYY-MM and is sequential across the time interval.

3. There is equal spacing between every two consecutive measurements

   Yes, there is an equal spacing between two consecutive measurements which is equal to a month and there are 12 data point per year from 2008 – 2012.

4. Each time unit within the time interval has at most one data point

   All data points are populated with a single value corresponding to a month. There are no missing values.

2. Which records should be used as the holdout sample?

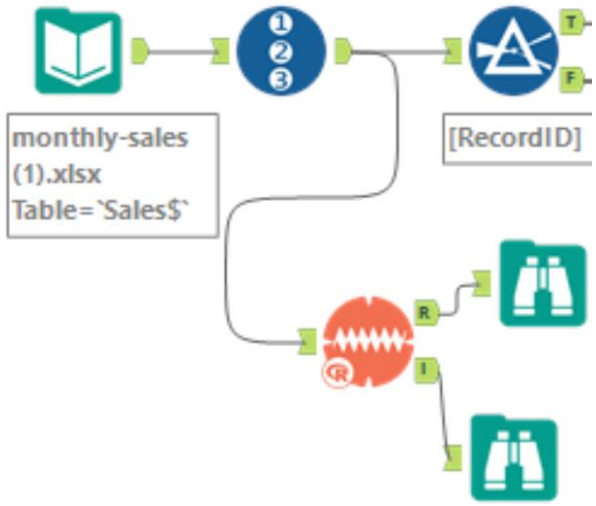
   Since the models are going to be forecasting 4 periods, we only need to use a holdout sample of 4 periods. In this case we will be using time periods from 2013-06 to 2013-09.

Step 2: Determine Trend, Seasonal, and Error components

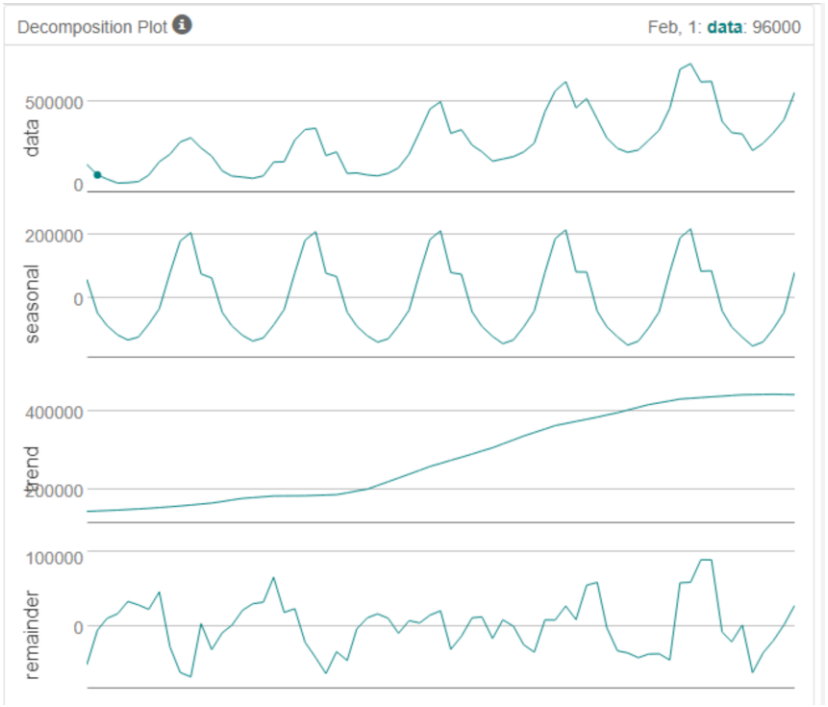
Graph the data set and decompose the time series into its three main components: trend, seasonality, and error. *(250 word limit)*

Answer this question:

1. What is the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.



ALTERYX WORK FLOW ETS.



Using the Alteryx work flow we get the ETS charts from the TS plot module. The following observations are made.

Error Plot – Indicates a trend line characteristic that has no unique pattern but rather has an inconsistent behavior. We will apply this as a Multiplicative component in the Error component.

Trend Plot – The trend line is a linear increase over time hence we will apply this as an additive component in the Trend component.

Seasonality Plot – The Seasonality plot shows an incremental trend over a period with Highs and Lows that are consistent hence we apply this is as a multiplicative component.

Using the plots, we will apply an ETS (MAM) non-damped and an ETS (MAdM) damped.

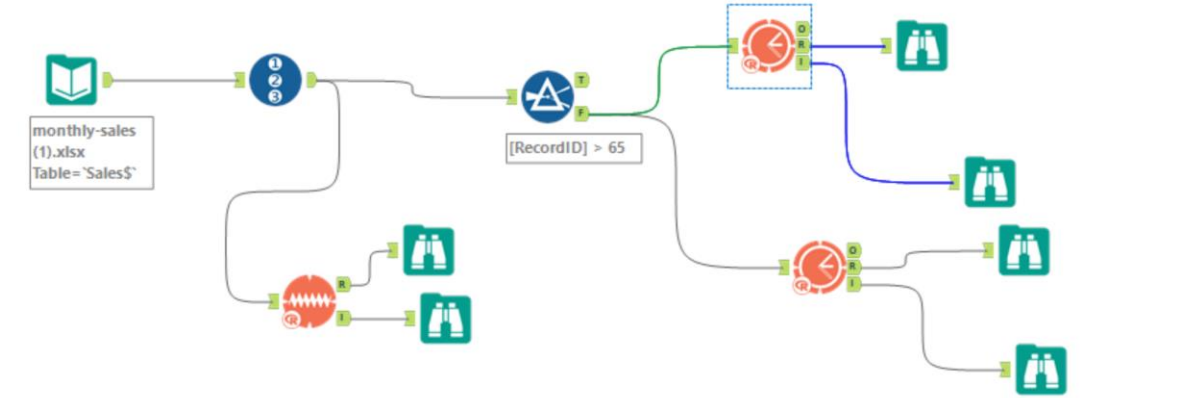
Step 3: Build your Models

Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

Answer these questions:

1. What are the model terms for ETS? Explain why you chose those terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

From the completed ETS analysis we will utilize an ETS (MAM) non-damped and a TS (MAdM) damped for our analysis and compare the performance of both.

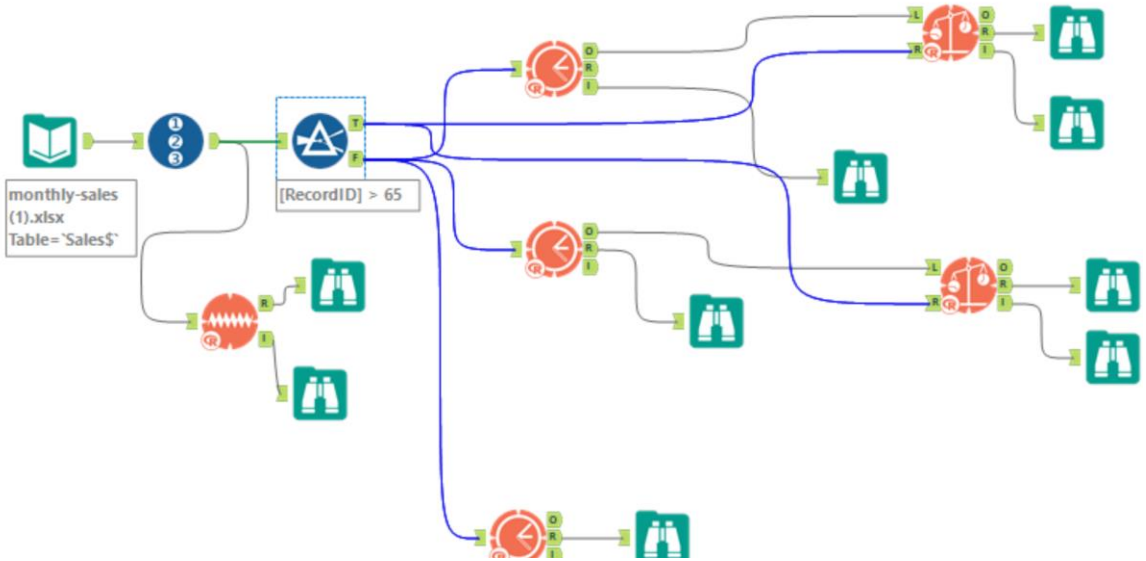


ALTERYX WORK FLOW FOR ETS

Error Plot – Indicates a trend line characteristic that has no unique pattern but rather has an inconsistent behavior. We will apply this as a Multiplicative component in the Error component.

Trend Plot – The trend line is a liner increase over time hence we will apply this as an additive component in the Trend component.

Seasonality Plot – The Seasonality plot shows an incremental trend over a period of time with Highs and Lows that are consistent hence we apply this is as multiplicative component.



ALTERYX WORK FLOW FOR ETS AND TS COMPARE

Non-Damped Model:

Summary of Time Series Exponential Smoothing Model MIM

Method:
ETS(M,A,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
2818.2731122	32992.7261011	25546.503798	-0.3778444	10.9094683	0.372685	0.0661496

Information criteria:

AIC	AICc	BIC
1639.7367	1652.7579	1676.7012

Damped Model:

Summary of Time Series Exponential Smoothing Model MAdM

Method:
ETS(M,Ad,M)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

Information criteria:

AIC	AICc	BIC
1639.465	1654.3346	1678.604

Non- Damped Model : Model Comparision.

Report

Comparison of Time Series Models

Actual and Forecast Values:

Actual	MIM
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
MIM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066

Damped Model Comparison:

Report

Comparison of Time Series Models

Actual and Forecast Values:

Actual	MAdM
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
MAdM	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116

OBSERVATIONS:

The Following observations are made
Non – damped Model:
Running the ETS (MAM) the RMSE – 32992.72, MASE – 0.3726, AIC – 1639.73
Non – damped Model:
Running the ETS (MAdM) the RMSE – 33153.52, MASE – 0.3675, AIC – 1639.46

MODEL comparison: ETS (MAM) RMSE – 74101.16 / ETS (MAdM) RMSE – 6017.47
ETS (MAM) MASE – 1.006 / ETS (MAdM) RMSE— 0.811

From the above observations we will select the ETS (MAdM) model for our time series forecasting model. Based on the values of RMSE, MASE which are lower and the model forecasts more accurately.

Auto Model: For Comparison purposes was run after analysis as a verification of our observation which it confirms.
We see that Alteryx choose the ETS (M,Ad,M) model which confirms our analysis.

Summary of Time Series Exponential Smoothing Model MAdM

Method:
ETS(M,Ad,M)

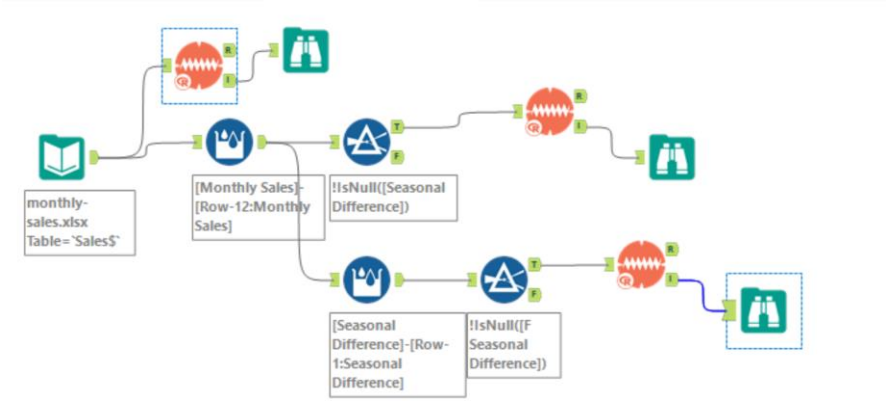
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

Information criteria:

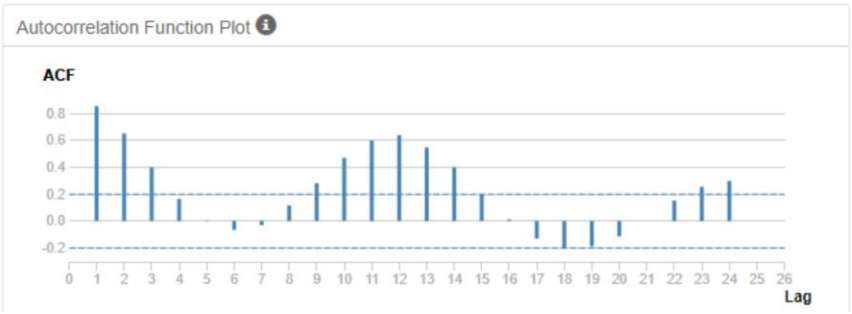
AIC	AICc	BIC
1639.465	1654.3346	1678.604

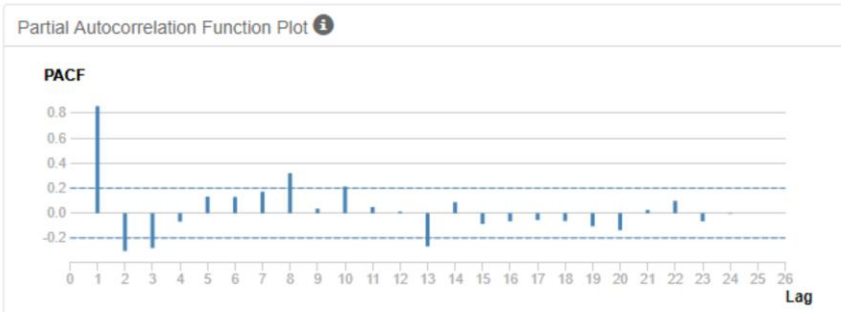
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.



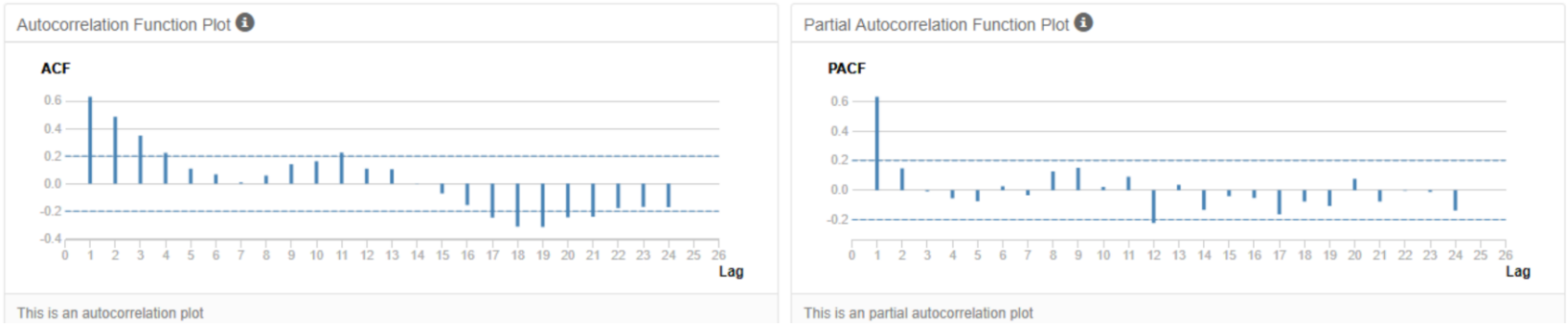
ALTERYX WORKFLOW FOR DIFFERENCING

Non-Differenced ACF/PACF – We see a strong correlation in the ACF plot which shows a positive value at lag1 and gradually decay towards zero and is considered non-stationary hence will require differencing. The PACF has a positive value at lag 1 and a spike drop at lag 2. The ACF also shows seasonality components which will need to be eliminated by differencing.

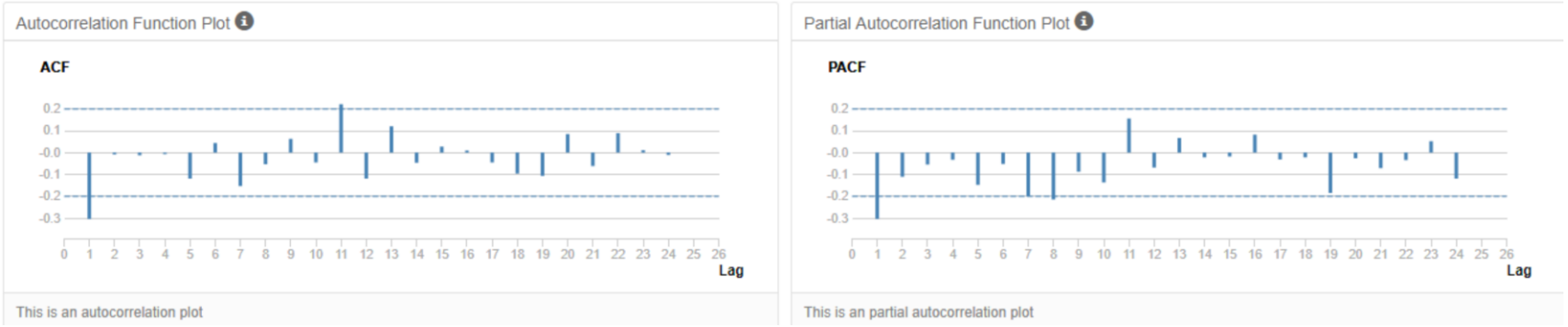




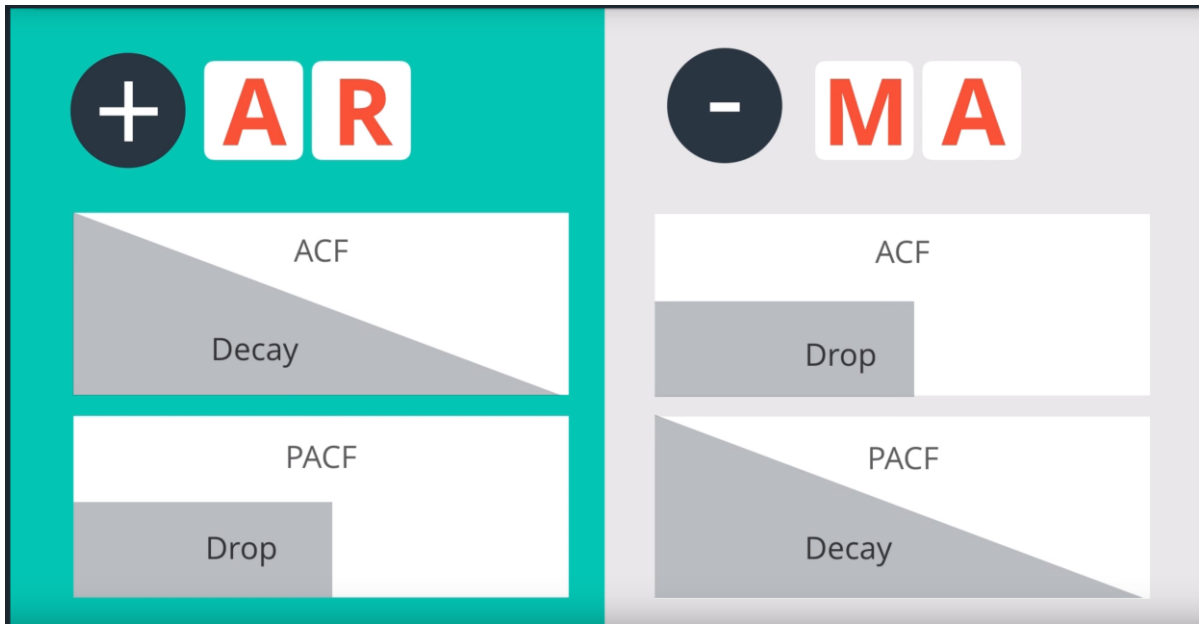
Differenced ACF/PACF – We see a strong correlation in the ACF plot which shows a positive value at lag1 and gradually decay towards zero and is considered non-stationary hence will require differencing. The PACF has a positive value at lag 1 and a spike drop at lag 12 and 24. The ACF also shows seasonality components which will need to be eliminated by further differencing.



1st Differenced ACF/PACF – We see a no correlation in the ACF plot which shows a negative value at lag1 with a spike drop to zero at lag 2. The PACF has a negative value at lag 1 . We can consider the series sufficiently differenced to compensate for both correlation and seasonality.



ARIMA term selection form the 1st difference: ARIMA (pdq)(PDQ)M



NON – Seasonal component:
AR (p) – ACF negative correlation at lag 1 with a spike drop / PACF negative correlation at lag 1 with a gradual decay -- 0
I(d) -- First difference is used – 1
MA(q) -- ACF negative correlation at lag 1 with a spike drop / PACF negative correlation at lag one with a gradual decay – 1

Seasonal component:
AR (P) – ACF non-significant negative correlation at lag 12 and 24 / PACF non-significant negative correlation at lag 12 and 24-- 0
I(D) -- First difference is used – 1
MA(Q) -- ACF non-significant negative correlation at lag 12 and 24 / PACF non-significant negative correlation at lag 12 and 24 – 0

M – 12 since monthly data is being used and the period is 12 months.

ARIMA (pdq)(PDQ)M – ARIMA (011) (010)12

a. Describe the in-sample errors. Use at least RMSE and MASE when examining results

Report

Summary of ARIMA Model ARIMA

Method: ARIMA(0,1,1)(0,1,0)[12]

Call:
Arima(Monthly.Sales, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 0), period = 12))

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

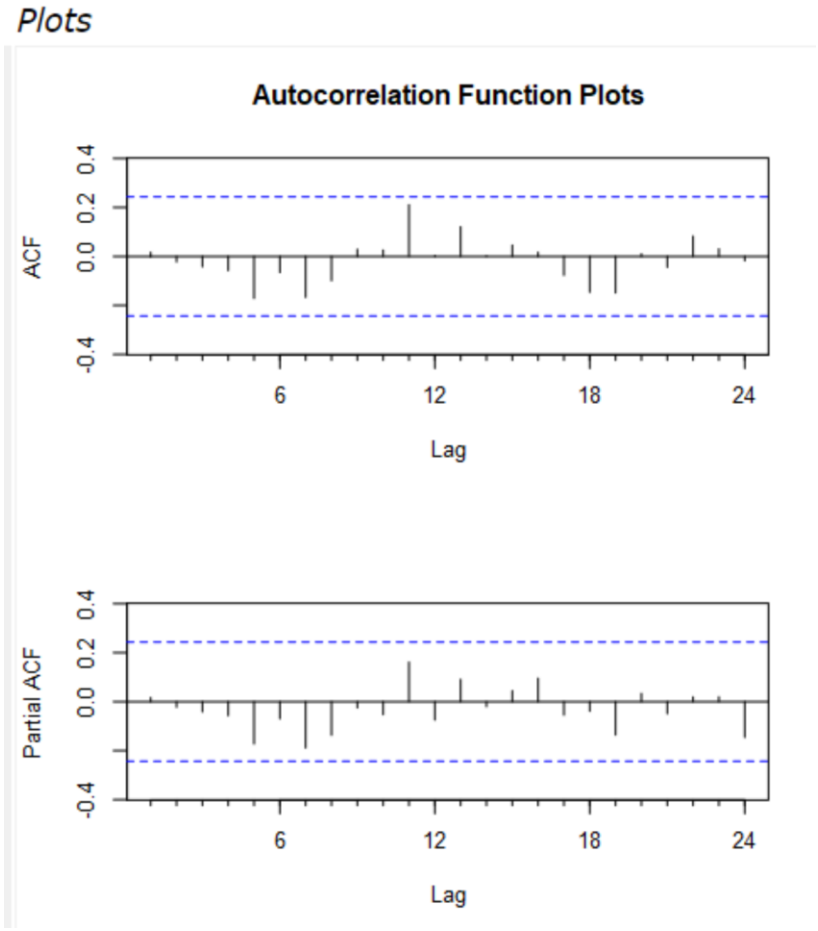
In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

The RMSE is at 36761.52 is comparable to ETS (MAdM) model.
and MASE is at 0.364 < 1.

b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

ACF and PACF shows no correlations and seasonality has been accounted for hence no additional terms are required for AR or MA.



Step 4: Forecast

Compare the in-sample error measurements to both models and compare error measurements for the holdout sample in your forecast. Choose the best fitting model and forecast the next four periods. (250 words limit)

Answer these questions.

1. Which model did you choose? Justify your answer by showing: in-sample error measurements and forecast error measurements against the holdout sample.

Actual and Forecast Measurement Comparison:

ETS (MAdM)

Actual and Forecast Values:

Actual	MAdM
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE
MAdM	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116

ARIMA(011)(010)12

Actual and Forecast Values:

Actual	ARIMA
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

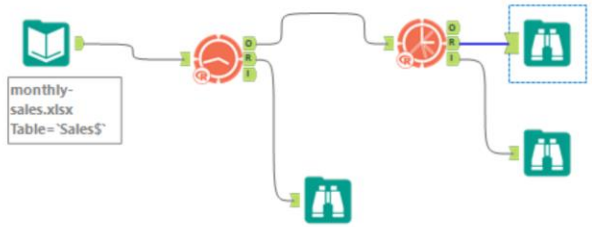
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532

Observation: We will utilize the ARIMA model for the sales forecast. We see from the measurement report that ARIMA has a better accuracy forecasting against the holdout sample as compared to the ETS (MAdM) model.

RMSE ETS vs RMSE ARIMA (011)(010)12 -- 60176.47 > 33999.79. ARIMA is a better choice from the RMSE value comparison.

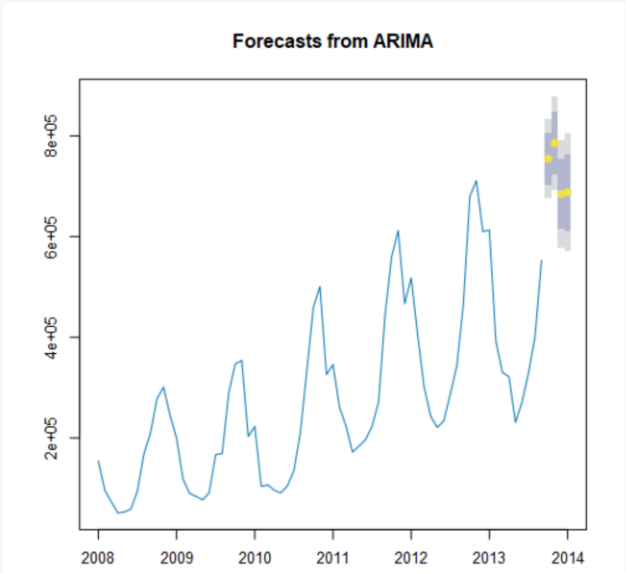
MASE ETS vs MASE ARIMA (011)(010)12 – 0.8116 > 0.4532 which are both < 1 but ARIMA is a better choice with a lower MASE value.

2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.



Report

4 Period Forecast from ARIMA



Period	Sub_Period	forecast	forecast_high_95	forecast_high_80	forecast_low_80	forecast_low_95
2013	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
2013	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
2013	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
2014	1	687854.460048	804889.286634	764379.419903	611329.500193	570819.633462

Before you Submit

Please check your answers against the requirements of the project dictated by the rubric here. Reviewers will use this rubric to grade your project.