Project: Forecasting Sales

Complete each section. When you are ready, save your file as a PDF document and submit it here: https://classroom.udacity.com/nanodegrees/nd008/parts/edd0e8e8-158f-4044-9468-3e08fd08cbf8/project

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.
   The four characteristics are: The data has to be sequential (1) and equal (2) intervals, each time unit having at most one data point (3) and the ordering matters (4) and we're having a dependencies on time.
   In our project we're having sales data per month the intervals are having the same size, the data is ordered via date and for each month we're having only one point.
2. Which records should be used as the holdout sample?
   Ideally the holdout sample should have the size of the forecast (4 month). So we should use the June 2013 to September 2013 as our holdout sample.

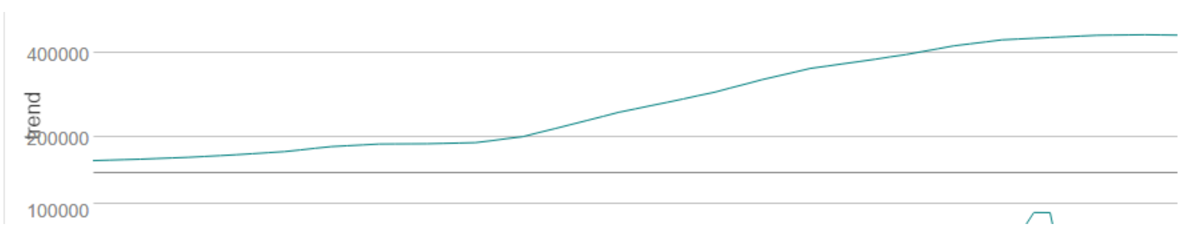
Step 2: Determine Trend, Seasonal, and Error components

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

Answer this question:

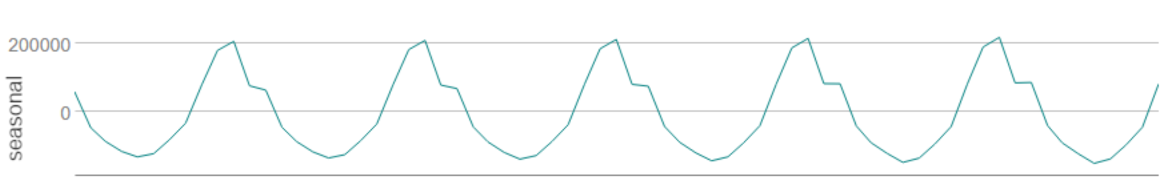
1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.

Trend:



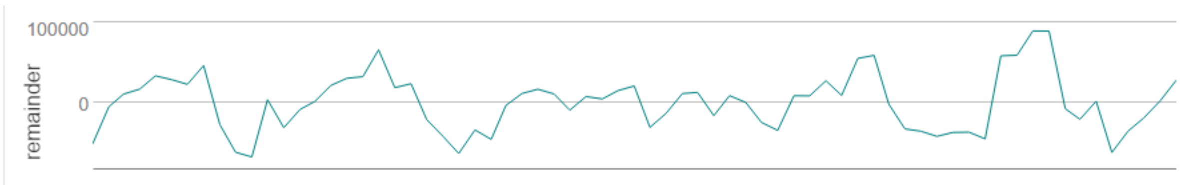
*As we can see the trend line is an uplifting line and linear so we're having an **additively** trend line.*

Seasonality:



We can see that the seasonality grows or shrinks over time so we’re having a ***multiplicative*** *seasonality.*

Error:



As we can see the error growing or shrinking over the time so would apply the error ***multiplicative****.*

Step 3: Build your Models

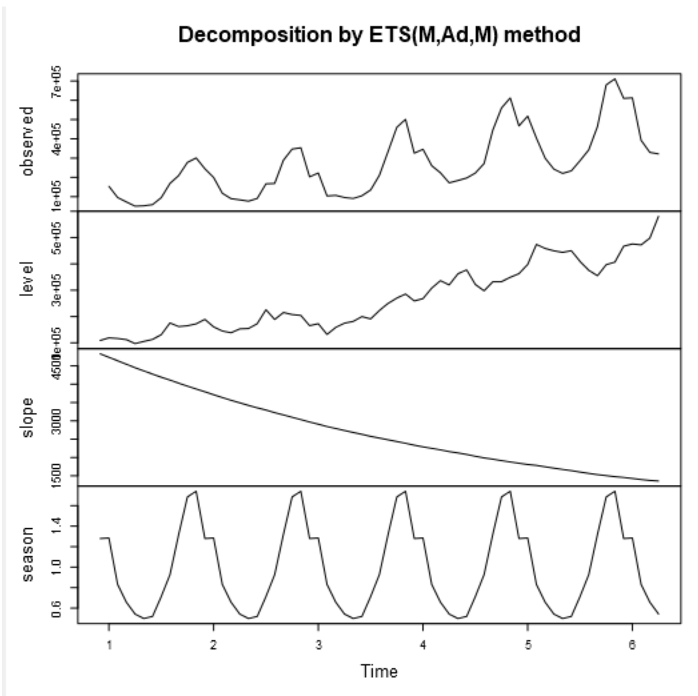
Analyze your graphs and determine the appropriate measurements to apply to your ARIMA and ETS models and describe the errors for both models. (500 word limit)

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

Thanks to our observation in step 2 I used the ETS Model MAM.

1. Trend dampening = “Yes”



In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5572.6821018	33302.042717	25725.4553044	0.1900065	10.54361	0.3752957	0.100576

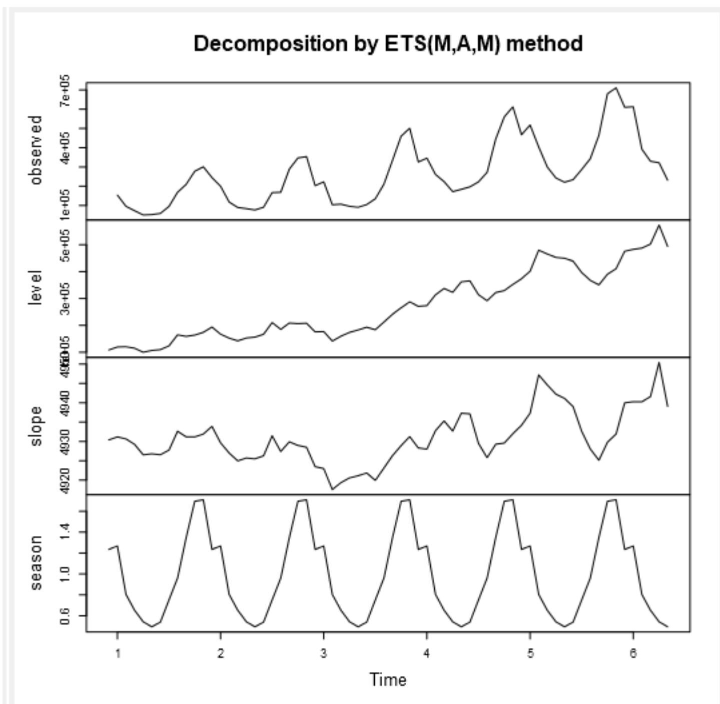
Information criteria:

AIC	AICc	BIC
1636.5328	1649.554	1673.4973

The AIC value is: 1636.5328 for the trend dampening ETS model, the MASE value is: 0.37 and the RMSE value is: 33302.042717. The MASE (Mean Absolute Scaled Error) value is ‘ideally its value will be significantly less than 1 but is relative to comparison across other models for the same series’. A low MASE value indicates a good model.

The RMSE (Root Mean Squared Error) represents the standard deviation of the differences between an actual and a forecasted value.

We can now compare the trend dampening ETS model with the none trend dampening ETS model



In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
3729.2947922	32883.8331471	24917.2814212	-0.9481496	10.2264109	0.3635056	0.1436491

Information criteria:

AIC	AICc	BIC
1634.6435	1645.9768	1669.4337

The AIC value is: 1634.6435 for the none trend dampening ETS model, the MASE value is: 0.36 and the RMSE value is: 32883.8331.

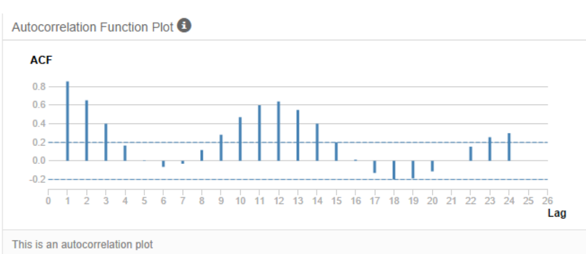
A small reminder from above the smaller value of the MASE or RMSE variable will win.

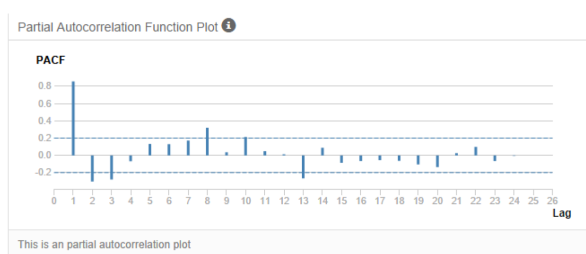
Now let's compare: 1634.6435 < 1636.5328 || 0.36 < 0.37 || 32883.8331 < 33302.042717

So for every value of the none dampening ETS model we're getting slightly better values. ***So we should use the damped ETS model.***

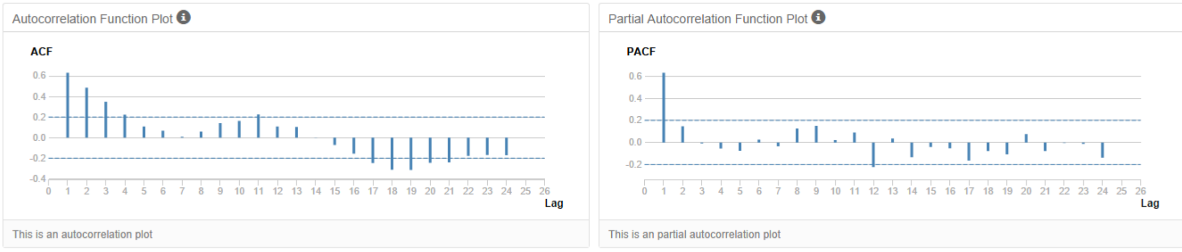
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.

   Let's check the ACF and the PACF plot. We notice a high correlation at the ACF diagram (Lag 1-3, Lag 9-14 for example) and at the PACF plot there are some gaps.

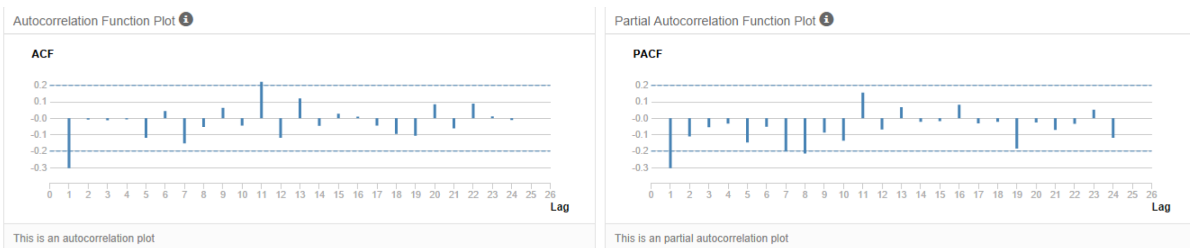




After smoothing the values via: [Monthly Sales]-[Row-12:Monthly Sales] we're having still a correlation inside (Lag 1-4) the ACF plot.



With the second smoothing step: [Season_1]-[Row-1:Season_1] we're getting ride of the correlations and the GAP.



So thanks to our observations we can built our ARIMA model: ARIMA (0,1,1)(0,1,0)12. At the ARIMA model we're having the following structure: ARIMA (p, d, q) or for the seasonal ARIMA model (p, d, q)(P, D, Q) m.

So p is the **AR** order, d is the **i**ntegration order, q the **MA** order and m is the period. After we have stationary the values ones we have to set d = 1(none stationary time series plot = 0), p has to be set to = 0, because we're having a negative lag_1 and q has set to 1 because of the negative correlation at lag_1. The period (m) will set to 12 (it's monthly data).

The seasonal first difference of the series has removed most of the significant lags from the ACF and PACF model so there is no further differencing. The remaining correlation can be accounted for using autoregressive and moving average terms and the differencing terms will be D(1).

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

Information Criteria:

AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

In-sample error measures:

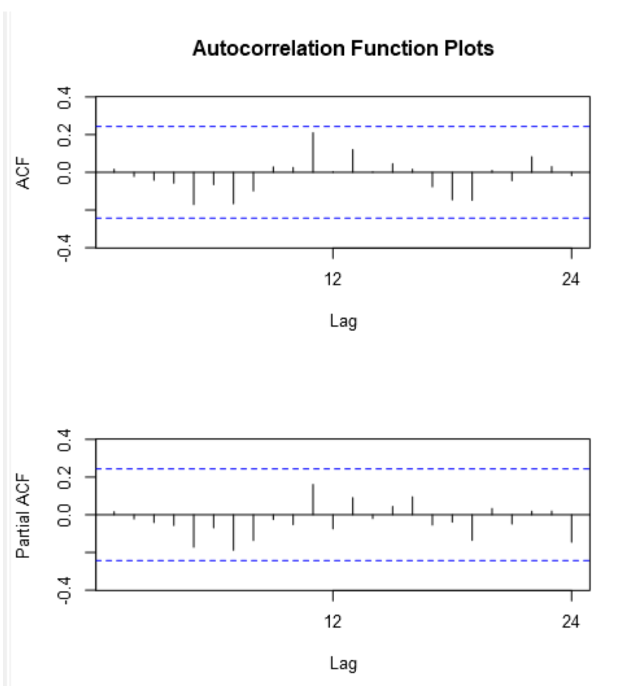
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145

The MASE (Mean Absolute Scaled Error) value is ‘ideally its value will be significantly less than 1 but is relative to comparison across other models for the same series’. A low MASE value indicates a good model.

The RMSE (Root Mean Squared Error) represents the standard deviation of the differences between an actual and a forecasted value.

The AIC value is: 1256.5967 for the ARIMA model, the MASE value is: 0.36 and the RMSE value is: 36761.53.

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



As we see at the ACF and the PACF there are no *correlations*.

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.

   ETS model and dampening (with holdout sample)

In-sample error measures:

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5572.6821018	33302.042717	25725.4553044	0.1900065	10.54361	0.3752957	0.100576

Information criteria:

AIC	AICc	BIC
1636.5328	1649.554	1673.4973

ARIMA model

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

Now we can compare the ETS model with the ARIMA model. We can to this via comparing the MASE value AMIRA_VALUE = 0.3464 < ETS_VALUE= 0.3752 The ARIMA MASE value is smaller. We can also compare the RMSE value of each model: AMIRA_VALUE = 36761.52 > ETS_VALUE= 33302.04 so as a third variable we can compare the AIC value of both models AMIRA_VALUE=1256.5967 < 1256.84 the AIC (a smaller AIC value indicates a better model).

RMSE = The standard deviation of the differences between predicted values and observed value is smaller at the ARIMA model. So we will use the AMIRA model for calculate the forecast.

With the results of the TS compare we're getting this values:

Accuracy Measures:

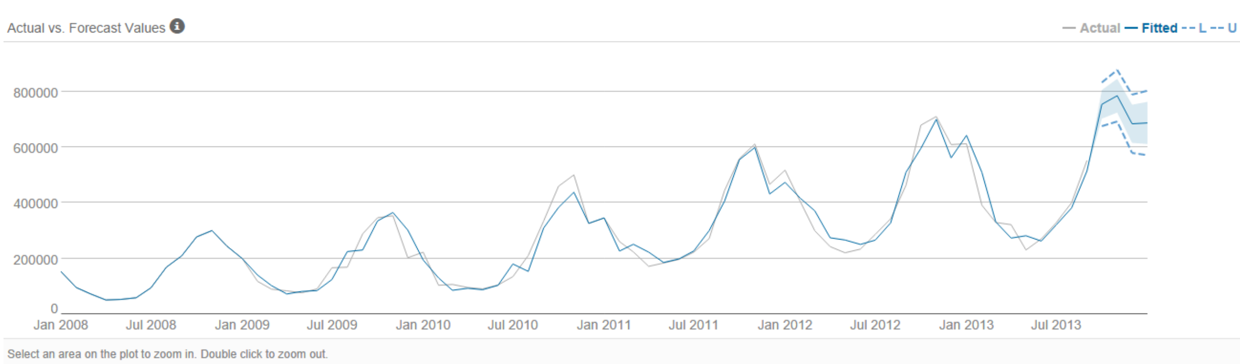
Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_Model	-53469.25	74839.3	61055.35	-11.1217	13.921	1.0147	NA

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA_Model	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

We can also see that the MASE value, the MAE value is smaller at the ARIMA model. So this is one more hint for using the ARIMA model.

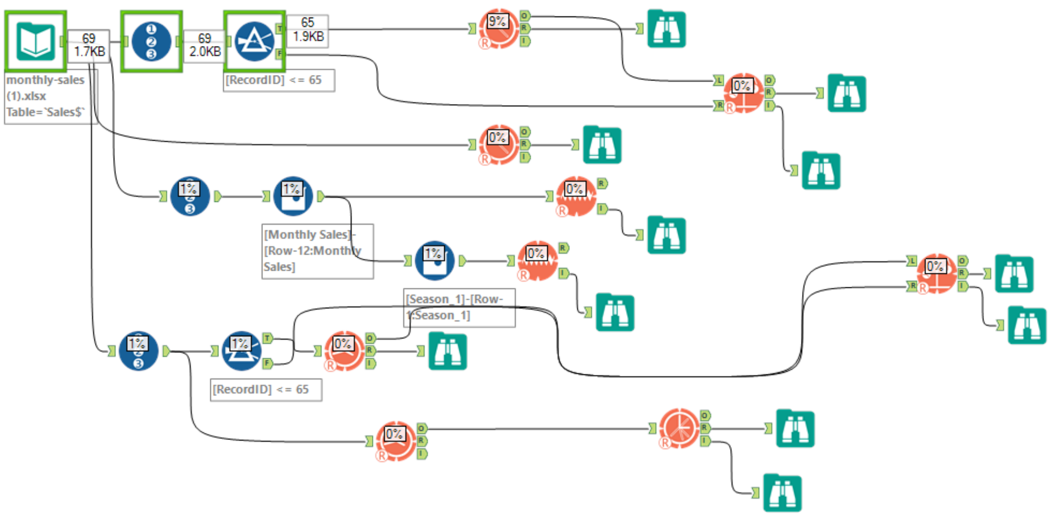
2. What is the forecast for the next four periods? Graph the results using 95% and 80% confidence intervals.



The forecast for the next four periods with our ARIMA model:

Period	Sub_Period	ARIMA_Model_Forecast	
2013	10	754854.460048	8
2013	11	785854.460048	8
2013	12	684854.460048	7
2014	1	687854.460048	8

The workflow:



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.