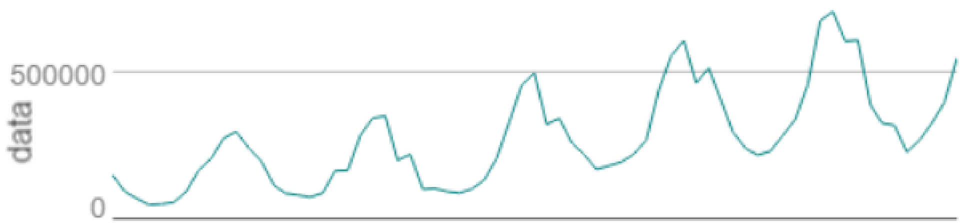
Step 1: Plan Your Analysis

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 dataset meets the criteria of time series dataset.

   By looking at data component of decomposition plot, I can observe:
   - The series is over a continuous time interval
   - Sequential measurements across that interval
   - There is equal spacing between every two consecutive measurements
   - Each time unit within the time interval has at most one data point

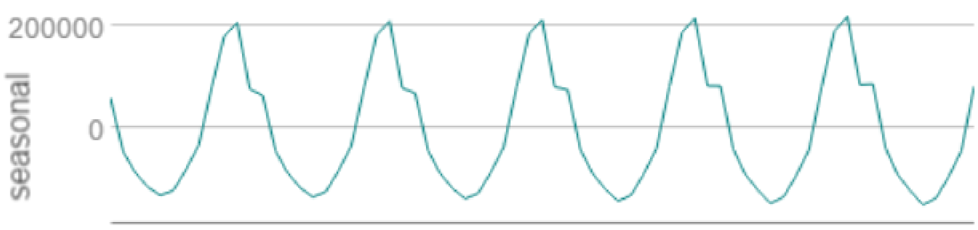


2. Which records should be used as the holdout sample?

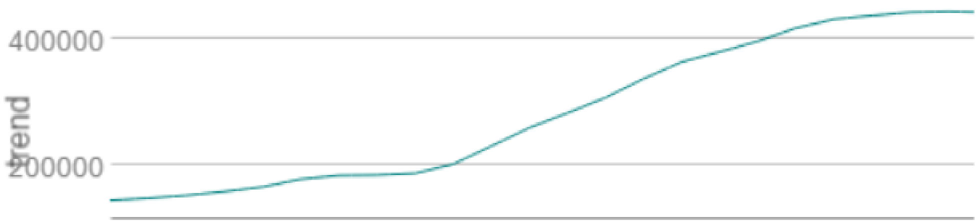
   Last 4 records should be used as holdout sample as they are the most recent and holdout out sample size should be at least amount of period I'm forecasting.

Step 2: Determine Trend, Seasonal, and Error components

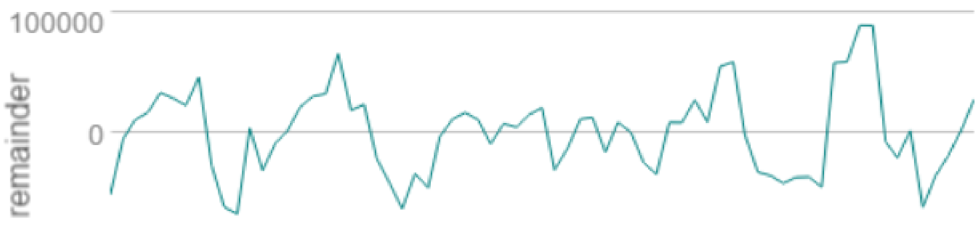
1. What are the trend, seasonality, and error of the time series? Show how you were able to determine the components using time series plots. Include the graphs.



It contains seasonality. It has peak electricity in November and repeats every year and increases over time from 207467.35 to 219237.38.



The pattern trend exhibits linear and uptrend.



The error is increasing over time.

Step 3: Build your Models

1. What are the model terms for ETS? Explain why you chose those terms.

   With observations from Step 2, I came up with ETS(M,A,M).

   It has increasing error (Multiplicatively), linear and uptrend (Additive), and increasing seasonal components (Multiplicatively).

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

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

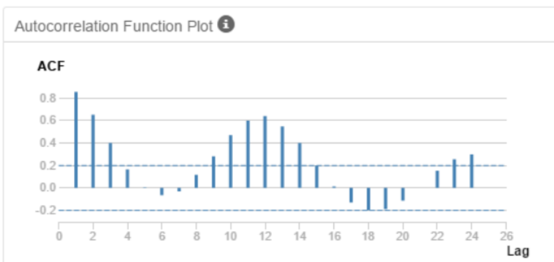
   RMSE (Root Mean Squared Error) can only be compared between models whose errors are measured in the same units. It indicates the absolute fit of the model to the data–how close the observed data points are to the model's predicted values. It is a great measurement to use when comparing models as it shows how many deviations from the mean the forecasted values fall. There is no absolute criterion for a "good" value of RMSE. For this model, there needs to be another set to compare the value to say what the value means.

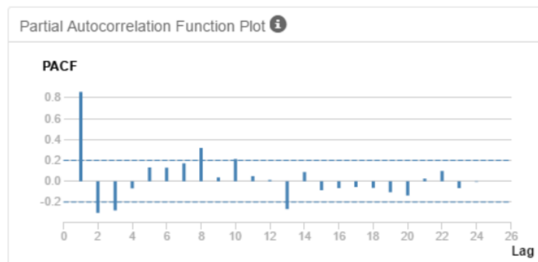
   MAPE (Mean Absolute Percentage Error) is scale independent, so it can be used to compare forecasts between different data series. MAPE is expressed in generic percentage terms and easy to understand than the other statistics. For this model, on average, the forecast is off by 10.91%.

MASE (Mean Absolute Scaled Error) is also scale independent, but it is only applicable for time series data. It measures the relative reduction in error compared to a naive model. Ideally its value will be significantly less than 1 but is relative to comparison across other models for the same series. For this model, MASE is 0.37. it is less than 1, but not significantly less than 1.

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.

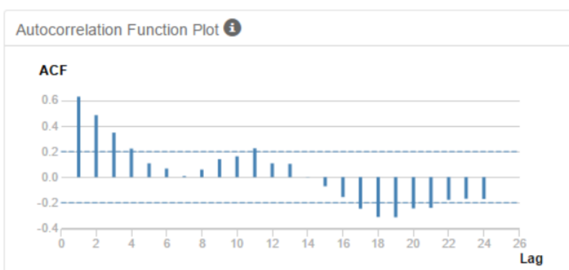
Time series:

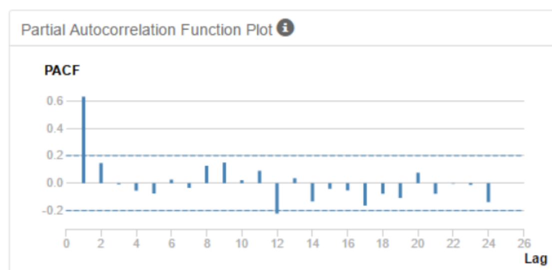




Looking at ACF plot, seasonal increases can be observed at 12 and 24 lags. The seasonal differencing was taken to make the data set stationary.

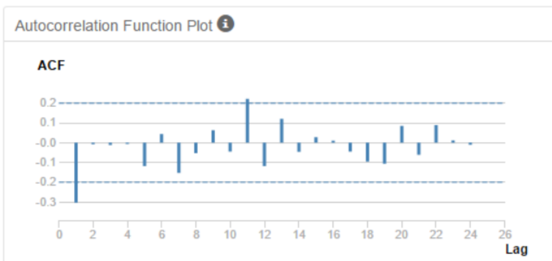
Seasonal difference:

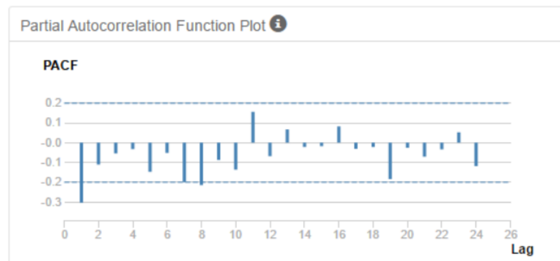




Seasonal increases are still observed. Seasonal First Difference was taken to correct seasonality before the data set is stationary.

Seasonal First Difference:





The data set is now stationalized and seasonal increases are eliminated.

Autocorrelation after the seasonal first differencing has now lowered to the point where I can add auto regressive and moving average terms to account for any remaining autocorrelation.

By observing at Seasonal First Difference plots, I came up with

d = 1 & D = 1 : Seasonal difference and Seasonal First Difference were taken.
q = 1 : Negative autocorrelation at lag 1 in ACF and PACF.
Q = 0 : No significant autocorrelation at lag 12 and 24 in ACF and PACF.
m = 12 : The data set is monthly data.

The final ARIMA model is ARIMA(0,1,1)(0,1,0)12

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

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

General description of each error can be found in the answer 1.a.

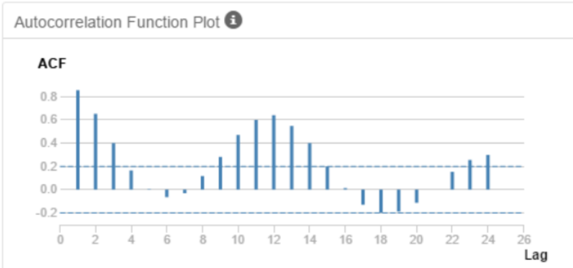
RMSE: The ETS model has smaller value 32992.73 and better than the ARIMA model (36761.53).

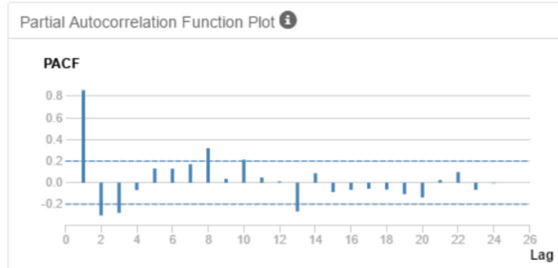
MAPE: For this model, on average, the forecast is off by 9.82%.

MASE: 0.36 is smaller and better than the ETS model (0.37).

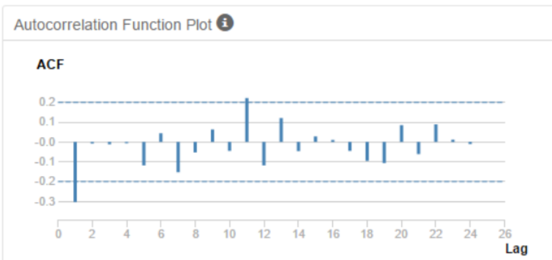
b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

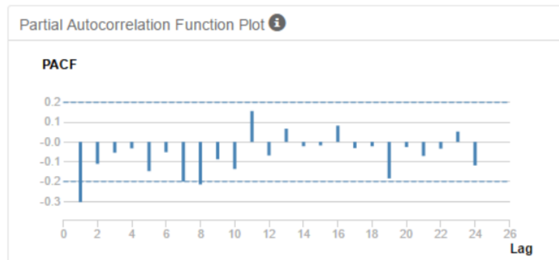
Time Series





Seasonal First Difference





Only a couple of lags have significant correlation in Seasonal First Difference ACF plot. It will not make much difference in the forecast. No more terms need to be added.

Step 4: Forecast

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

In-sample error measurements:

ETS

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

ARIMA

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

ARIMA model has smaller number for all errors. This means the error is smaller, and the ARIMA model is better model than the ETS model.

The below is the matrix showing the difference between Actual vs Forecasted for both models. Difference between Actual and Forecasted are smaller for the ARIMA model for all 4 months.

Actual	Forecasted		Difference	
	MAM	ARIMA	MAM	ARIMA
271000	248063.02	263228.48	-22936.98	-7771.52
329000	351306.94	316228.48	22306.94	-12771.52
401000	471888.58	372228.48	70888.58	-28771.52
553000	679154.79	493228.48	126154.79	-59771.52

I chose the ARIMA model to forecast monthly sales data for the next 4 months.

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

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