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 dataset meets the criteria of a time series dataset, since it has sequential measurements across a continuous time interval, equal spacing between every two consecutive measurements, and a data point for each time unit within the time interval.

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

We need to forecast sales for the next 4 months, hence we should use a holdout sample contains records from June, 2013 to September, 2013.

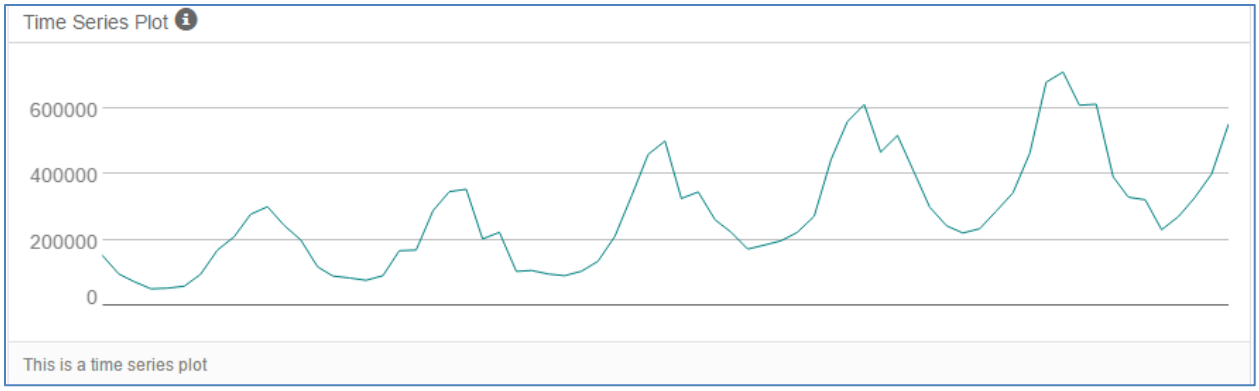
Step 2: Determine Trend, Seasonal, and Error components

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

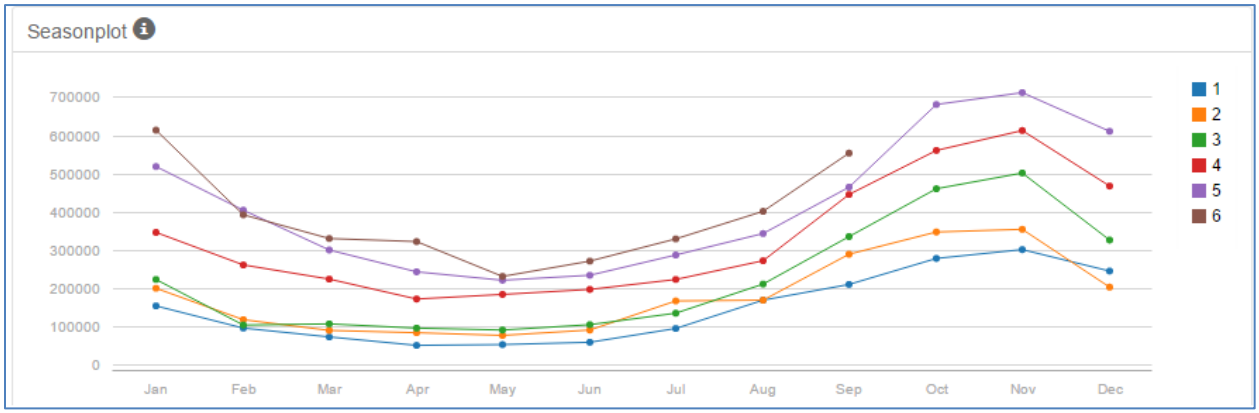
Answer this question:

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.

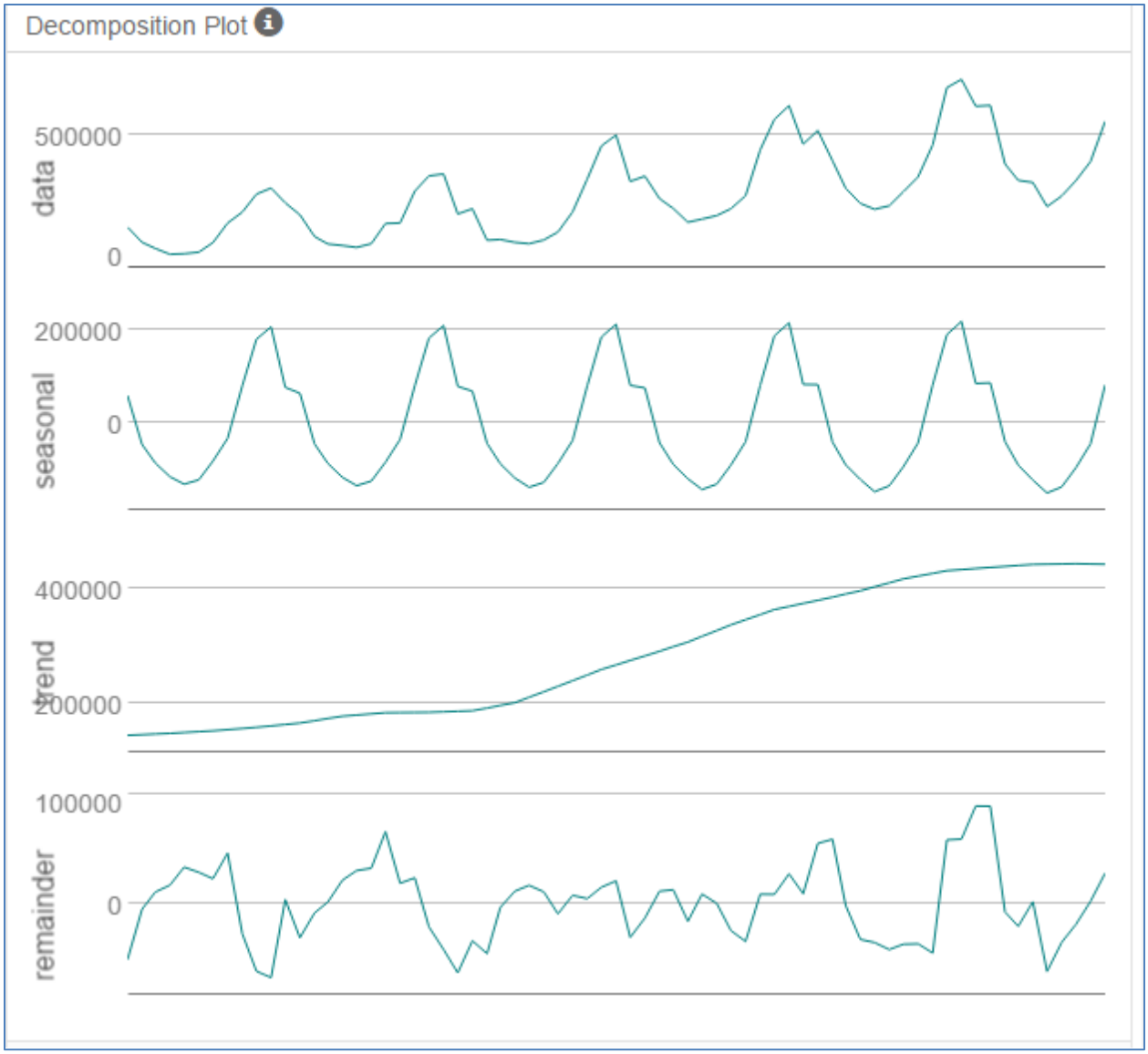
- The TS plot tool was used to determine the time series components.
    - The Time Series Plot is Uptrend



- The Time Series Plot contains seasonality, where the November month is significantly higher than other months.



- The Decomposition Plot shows how each of the trend, seasonal and error components should be applied.
  - The trend is linear, it will be applied additively.
  - The Seasonality increases over time, it will be applied multiplicatively.
  - The error increases over time, it will be applied multiplicatively.



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:

ETS Model

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

Step 1 - Select ETS model terms:

- As a result of the Decomposition Plot, error and seasonality are applied multiplicatively, while trend is applied additively.
- So, the ETS(M,A,M) is chosen.

Step 2 - Build the model:

- After running damped and non-damped ETS models with 4 months as a period to include in the forecast plot.
  - The result of non-damped ETS(M,A,M) model
    - RMSE = 32992.7261011
    - MASE= 0.372685
    - AIC= 1639.7367

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

  - The result of damped ETS(M,Ad,M) model
    - RMSE = 33153.5267713
    - MASE= 0.3675478
    - AIC= 1639.465

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

Step 3 - Validate model:

- After running TS Compare tool to calculate the accuracy measures
  - The result of non-damped ETS(M,A,M) model
    - RMSE = 74101.16
    - MASE= 1.0066

Actual and Forecast Values:

Actual	ETS_Sales_No_D
271000	248063.01908
329000	351306.93837
401000	471888.58168
553000	679154.7895

Accuracy Measures:

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

  - The result of damped ETS(M,Ad,M) model
    - RMSE = 60176.47
    - MASE= 0.8116

Actual and Forecast Values:

Actual	ETS_Sales_D
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

Step 4 - Select the best model

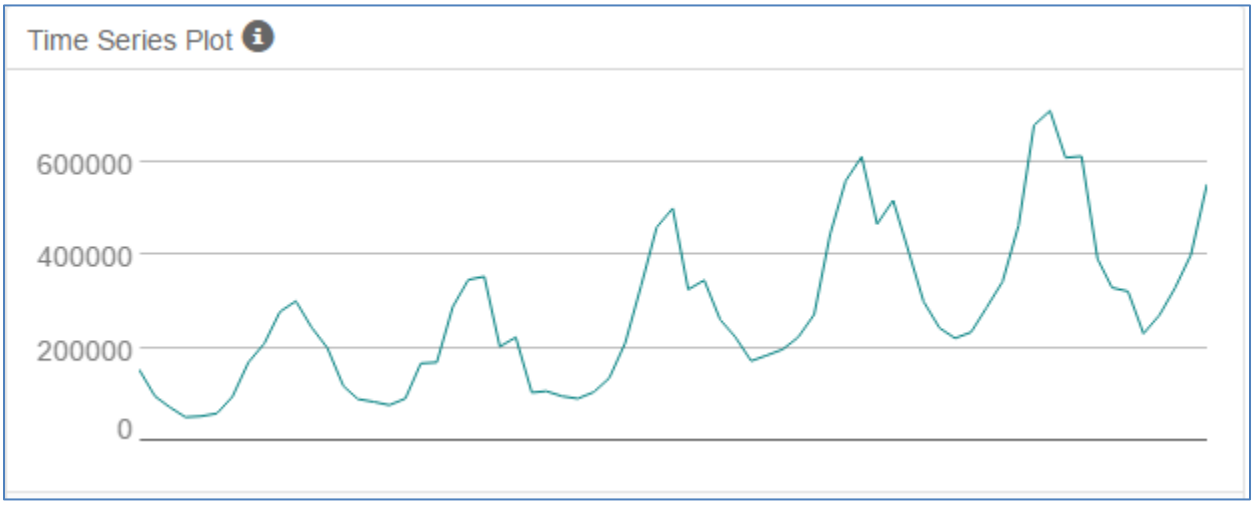
- The dampened ETS model is selected, since it has lower values of AIC, RMSE and MASE. Also, it tends to predict values more accurately.

ARIMA Model

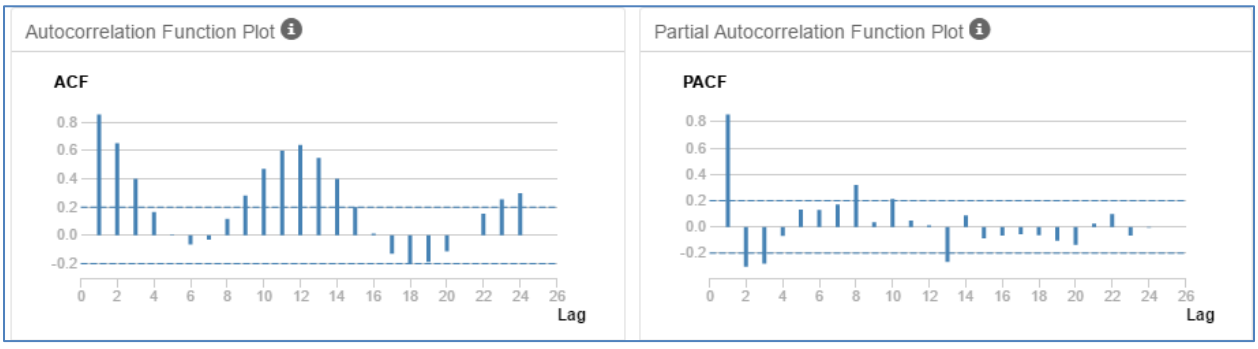
2. What are the model terms for ARIMA? Explain why you chose those terms. Graph the Auto-Correlation Function (ACF) and Partial Autocorrelation Function Plots (PACF) for the time series and seasonal component and use these graphs to justify choosing your model terms.
    a. Describe the in-sample errors. Use at least RMSE and MASE when examining results
    b. Regraph ACF and PACF for both the Time Series and Seasonal Difference and include these graphs in your answer.

Step 1 - Check stationarity:

- We need to check stationarity of the time series plot using TS Plot tool.
    - Time Series is non-stationary



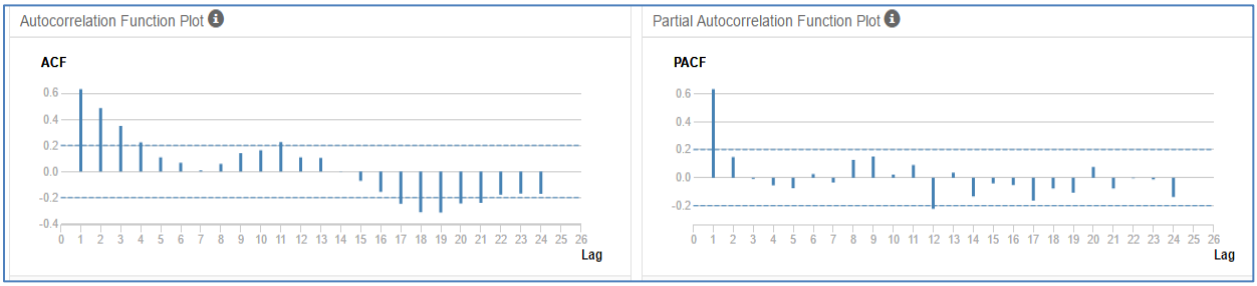
    - The Auto-Correlation Function (ACF) indicates a high correlation between points.
    - The Partial Autocorrelation Function Plots (PACF) displays a significant lag at point 13, because of seasonality.



Step 2 - Difference:

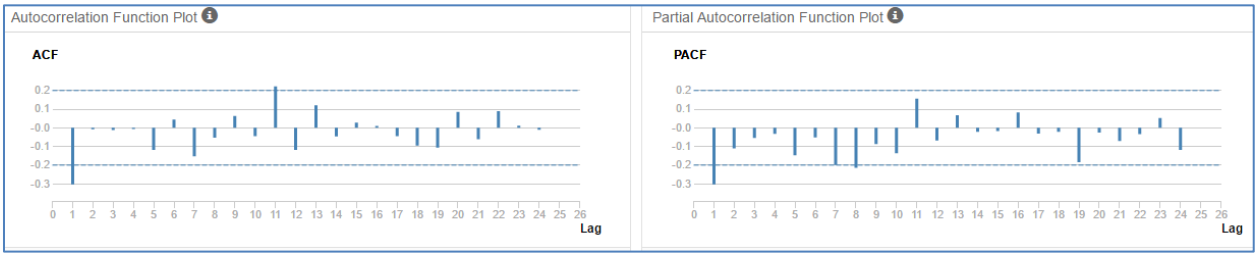
- Because the time series is non-stationary, we need to take a seasonal difference.
    - After applying the first seasonal difference, the time series is still non-stationary.

    - The correlation between points is high as shown in ACF
    - There's no strong correlation between the points as shown in PACF



    - After applying the second seasonal difference, the time series is now a stationary.

- The correlation between points is not high anymore as shown in ACF
- There's no strong correlation between the points as shown in PACF



Step 3 - Select AR and MA terms:

- We need to select the AR and MA terms
  - The ACF negatively correlated at Lag 1 and PACF decreases more gradually, we need to add one MA term for non-seasonal ARIMA model (p).
  - As we took seasonal differences, we need to add one differencing term for both the components of seasonal and non-seasonal ARIMA model (d, D)
  - There's no seasonal correlations, so no need to add the components of seasonal ARIMA model (P,Q)
  - As a result, the ARIMA model terms are: ARIMA(0,1,1)(0,1,0)12

Step 4 - Build the model:

- The result of in-sample errors is:
  - RMSE = 36761.5281724
  - MASE= 0.3646109

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

Step 5 - Validate model:

- After running TS Compare tool to calculate the accuracy measures
  - RMSE = 33999.79
  - MASE= 0.4532

Actual and Forecast Values:

Actual	ARIMA_Sales_not_auto
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

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

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

Step 1 - Choose the best model:

In-sample error measurements

ETS(M,Ad,M) model

- The result of in-sample errors is:
  - RMSE = 33153.5267713
  - MASE= 0.3675478

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

ARIMA(0,1,1)(0,1,0)12 Model

- The result of in-sample errors is:
  - RMSE = 36761.5281724
  - MASE= 0.3646109

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

Forecast error measurements

ETS(M,Ad,M) model

- The result of Forecast error measurements is:
  - RMSE = 60176.47
  - MASE= 0.8116

Actual and Forecast Values:

Actual	ETS_Sales_D
271000	255966.17855
329000	350001.90227
401000	456886.11249
553000	656414.09775

Accuracy Measures:

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

ARIMA(0,1,1)(0,1,0)12 Model

- The result of Forecast error measurements is:
  - RMSE = 33999.79
  - MASE= 0.4532

Actual and Forecast Values:

Actual	ARIMA_Sales_not_auto
271000	263228.48013
329000	316228.48013
401000	372228.48013
553000	493228.48013

Accuracy Measures:

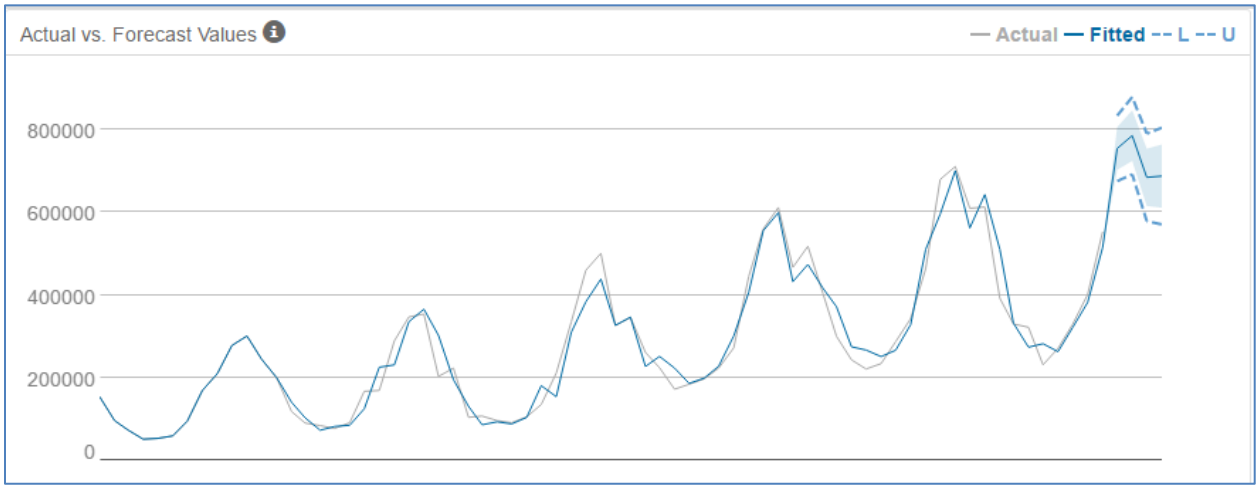
	Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ARIMA_Sales_not_auto		27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

The best model

- ARIMA Model is better than ETS Model, because of the following:
  - The values of RMSE and MASE are smaller
  - It tends to predict values more accurately

Step 2 – Forecast the next four periods:

After running TS Forecast tool, the predicted values of records from June, 2013 to September, 2013 are 754854.460048, 785854.460048, 684854.460048, and 687854.460048 respectively.



Period	Sub_Period	ARIMA_forecast	ARIMA_forecast_high_95	ARIMA_forecast_high_80	ARIMA_forecast_low_80	ARIMA_forecast_low_95
6	10	754854.460048	834046.21595	806635.165997	703073.754099	675662.704146
6	11	785854.460048	879377.753117	847006.054462	724702.865635	692331.166979
6	12	684854.460048	790787.828211	754120.566407	615588.35369	578921.091886
7	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.