

Projeto: Previsão de vendas

Planejando a análise

A gerência da cadeia de suprimentos da empresa necessita da previsão de vendas mensais (quatro meses) para planejar a produção de videogames de acordo com a demanda dos clientes. Para tanto, foi disponibilizado um conjunto de dados com as vendas da empresa de janeiro de 2008 a setembro de 2013.

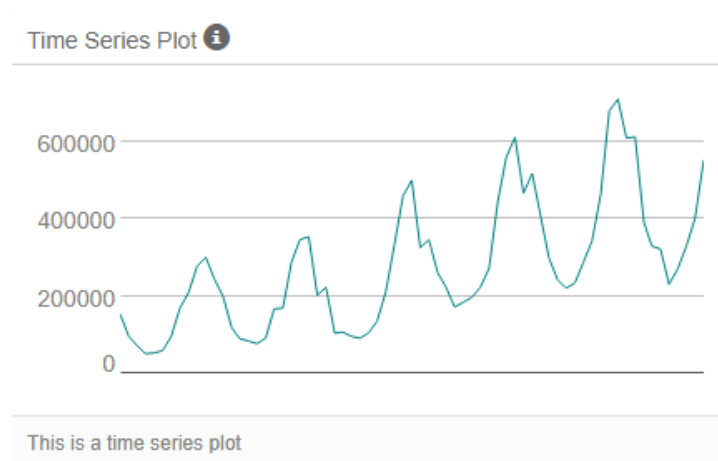
O problema consiste em prever valores futuros com base em resultados históricos (modelo de previsão rico em dados). Por outro lado, a variável-alvo (vendas mensais), além de ser numérica, é baseada em tempo. Logo, o modelo estatístico indicado é a previsão de série temporal.

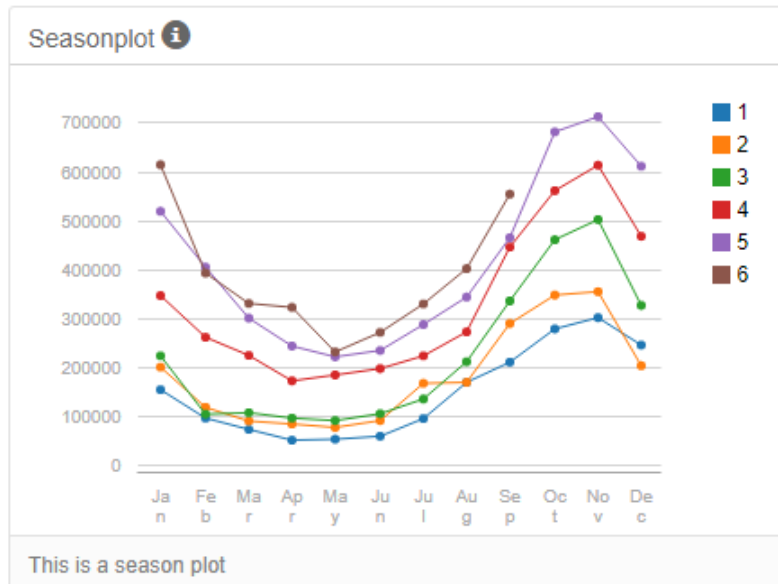
O conjunto de dados disponibilizado – vendas de janeiro/2008 a setembro/2013 - atende aos requisitos exigidos pelo modelo, visto que o intervalo é contínuo (não há interrupções entre o período inicial e o final), as medidas são sequenciais (períodos consecutivos) e com igual espaçamento (mensais) e cada unidade de tempo tem apenas um dado correspondente.

Como o modelo deverá fazer a previsão de quatro meses, é necessário usar uma amostra de validação de quatro períodos; portanto, serão separados os meses de junho a setembro de 2013 para validação posterior do modelo. Para a previsão final, no entanto, serão utilizados todos os dados disponíveis: 69 meses de 2008 a 2013.

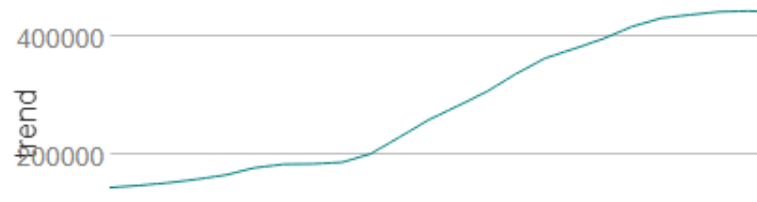
Determinando os componentes tendência, sazonalidade e erro

O relatório interativo da ferramenta “TS Plot” do software Alteryx mostra os seguintes gráficos para a série temporal em análise:

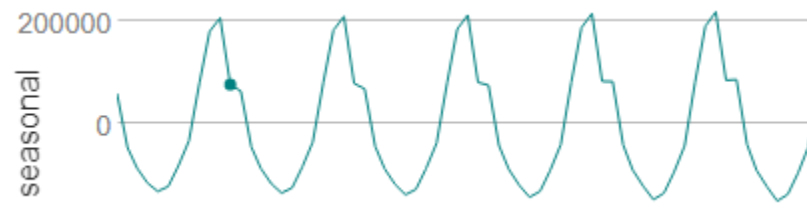




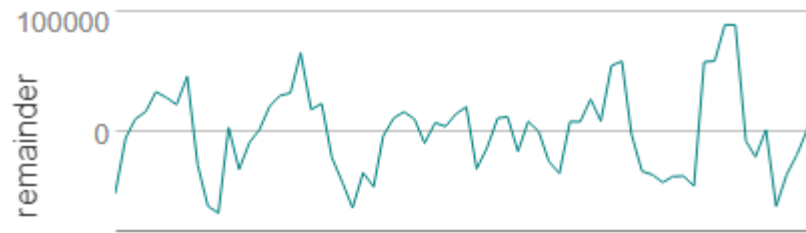
O “Decomposition Plot” desse gráfico mostra uma tendência linear, de maneira que esse componente deve ser aplicado no modelo na forma aditiva:



O “Decomposition Plot” evidencia também um crescimento da sazonalidade ao longo do tempo, embora bem sutil, o que resulta na necessidade de aplicar esse componente no modelo de forma multiplicativa:



Quanto ao componente erro, o “Decomposition Plot” não apresenta uma variância constante no tempo (picos e vales não são uniformes no tempo), motivo pelo qual deve ser aplicado no modelo de forma multiplicativa:



Dessa forma, o modelo a ser construído será ETS (M, A, M).

Construindo os modelos

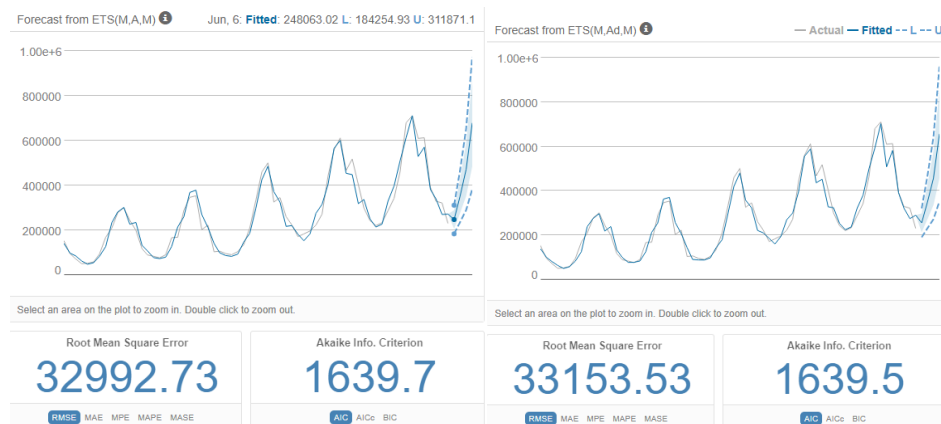
Como os componentes erro, tendência e sazonalidade devem ser aplicados no modelo ETS de forma multiplicativa, aditiva e multiplicativa, respectivamente, o modelo construído foi ETS (M, A, M).

Já a comparação entre os modelos ETS sem e com o componente “dampening” revela um desempenho melhor do ETS_damp, visto que os valores de previsão estão mais próximos dos reais:

Actual and Forecast Values:

Actual	ETS_MAM	ETS_MAM_damp
271000	248063.01908	255966.17855
329000	351306.93837	350001.90227
401000	471888.58168	456886.11249
553000	679154.7895	656414.09775

Por outro lado, como os AIC (Akaike Info. Criterion) de ambos modelos ETS – sem e com “dampening” – são muito similares (1.639,7 e 1.639,5) – é necessária uma análise das medidas dos erros para definir a melhor opção. Destaque-se que o AIC mede a qualidade relativa de um modelo estatístico, na medida em que equilibra a qualidade de ajustamento do modelo e a sua complexidade.

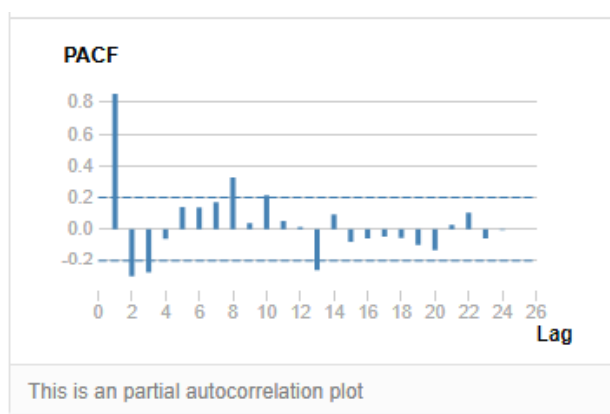
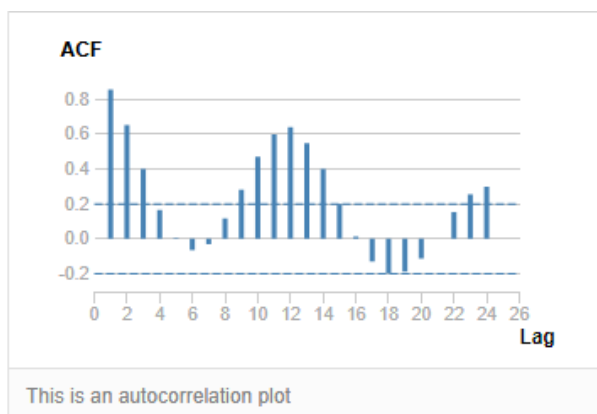
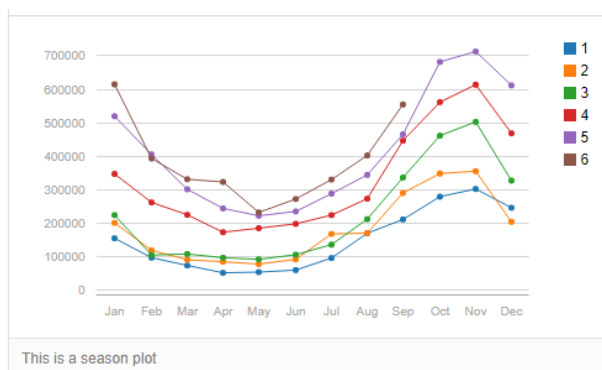
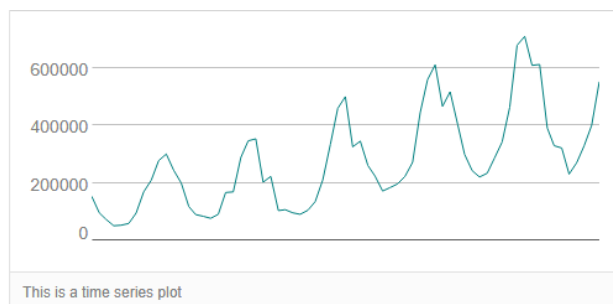


Accuracy Measures:

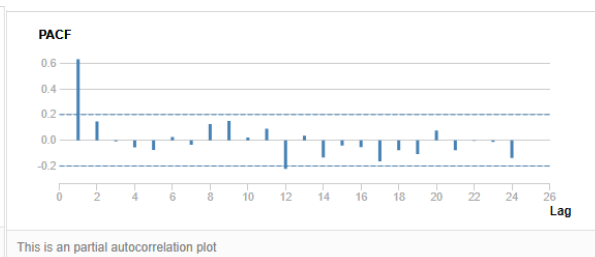
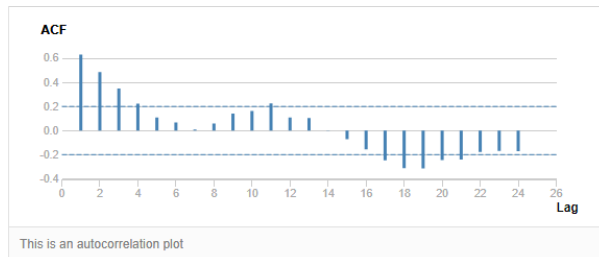
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_MAM	-49103.33	74101.16	60571.82	-9.7018	13.9337	1.0066
ETS_MAM_damp	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116

A medida MASE, entendida como o erro médio absoluto do modelo dividido pelo valor médio absoluto da primeira diferença da série, mede a redução relativa no erro em relação a um modelo naïve e deve ser menor que 1. Essa medida indica o melhor modelo como o ETS_damp. A medida RMSE, por seu turno, representa o desvio-padrão das diferenças entre os valores previstos e os realizados. Considerando que a medida para o modelo ETS com amortecimento é a menor, este é o melhor modelo para continuar a análise.

Para a construção do modelo ARIMA, é necessário verificar inicialmente se a série temporal é estacionária, com média e variância constantes ao longo do tempo. No caso, como a série possui tendência e sazonalidade, ela deve ser transformada em estacionária primeiro. O relatório gerado pela ferramenta “TS Plot” a partir dos dados de venda confirma que a série não possui essa característica:

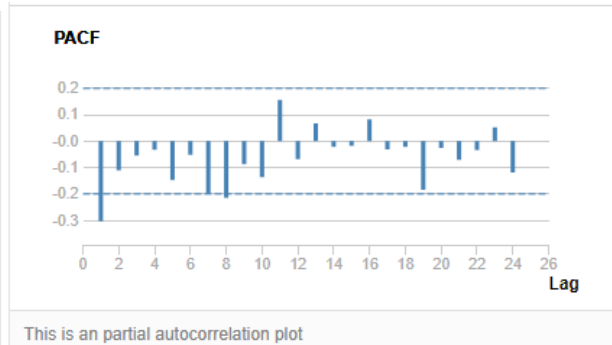
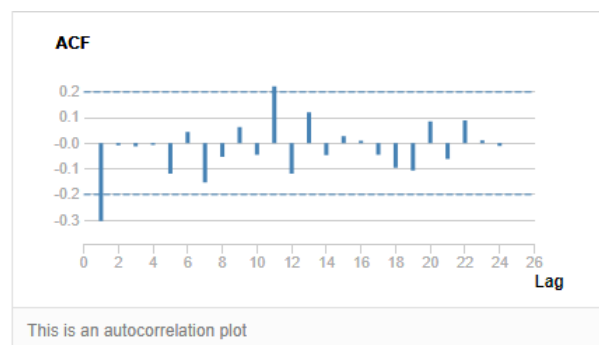


O resultado de uma primeira diferenciação – sazonal - pode ser observado nos seguintes gráficos:



O gráfico de autocorrelação ACF, porém, evidencia que a série ainda não é estacionária e que deve ser novamente diferenciada, pois a variância ainda não é constante. Além disso, o gráfico PACF dessa parte sazonal do modelo mostra um Lag 0 positivo, o que denota a inexistência do componente de média móvel. Assim, o modelo ARIMA, na parte sazonal, deve ter os termos “p”, “d”, “q” iguais a 0,1,0.

Após uma nova diferenciação, a série já pode ser considerada estacionária, de forma que o termo I ou “d” do modelo ARIMA deve ser 1 tanto para a parte sazonal quanto para a parte não sazonal:



Por outro lado, os gráficos acima mostram uma correlação negativa em Lag 1; além disso, o ACF trunca abruptamente após a primeira defasagem e o PACF cai de forma mais gradual. Esses são indicadores de que o termo a ser usado no modelo é MA ou componente de média móvel. O componente autorregressivo (AR), nesse caso, é 0. Assim, o modelo ARIMA, na parte não sazonal, deve ter os termos “p”, “d”, “q” iguais a 0,1,1.

O modelo ARIMA a ser utilizado, portanto, é ARIMA (0,1,1) (0,1,0) [12]. Por meio do software Alteryx, pode-se observar que suas principais medidas são: AIC de 1.256,5967, RMSE de 36.761,5281 e MASE de 0,3646:

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))

Coefficients:

	ma1
Value	-0.378032
Std Err	0.146228

sigma^2 estimated as 1722385234.94439: log likelihood = -626.29834

Information Criteria:

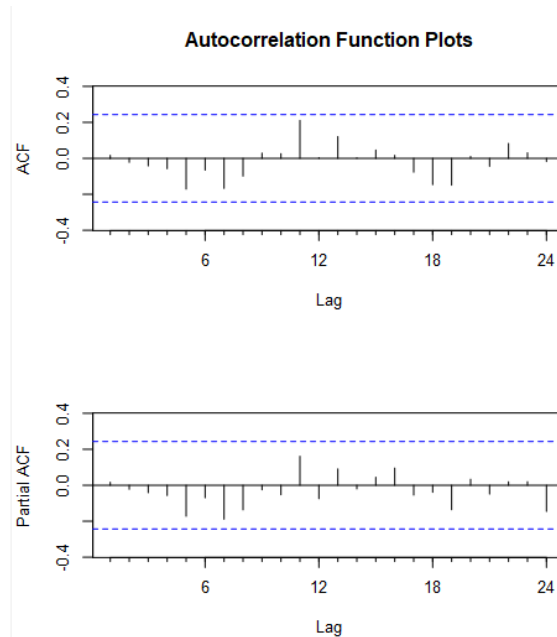
AIC	AICc	BIC
1256.5967	1256.8416	1260.4992

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

Ljung-Box test of the model residuals:

Chi-squared = 16.4458, df = 23, p-value = 0.83553



Prevendo as Vendas

O modelo ETS_damp (M,A,M) apresentou AIC de 1.639,5; já o modelo ARIMA apresentou AIC de 1.256,5967. Como o ARIMA possui AIC menor, ele deve ser considerado o melhor modelo de previsão de série temporal para o caso concreto.

Além disso, após a união de ambos os modelos e sua conexão com a ferramenta “TS Compare” do Alteryx, o respectivo relatório de comparação evidencia medidas menores para o modelo ARIMA, cabendo ressaltar que a medida MASE desse modelo (0,45) é bem inferior à do ETS (0,81):

Accuracy Measures:

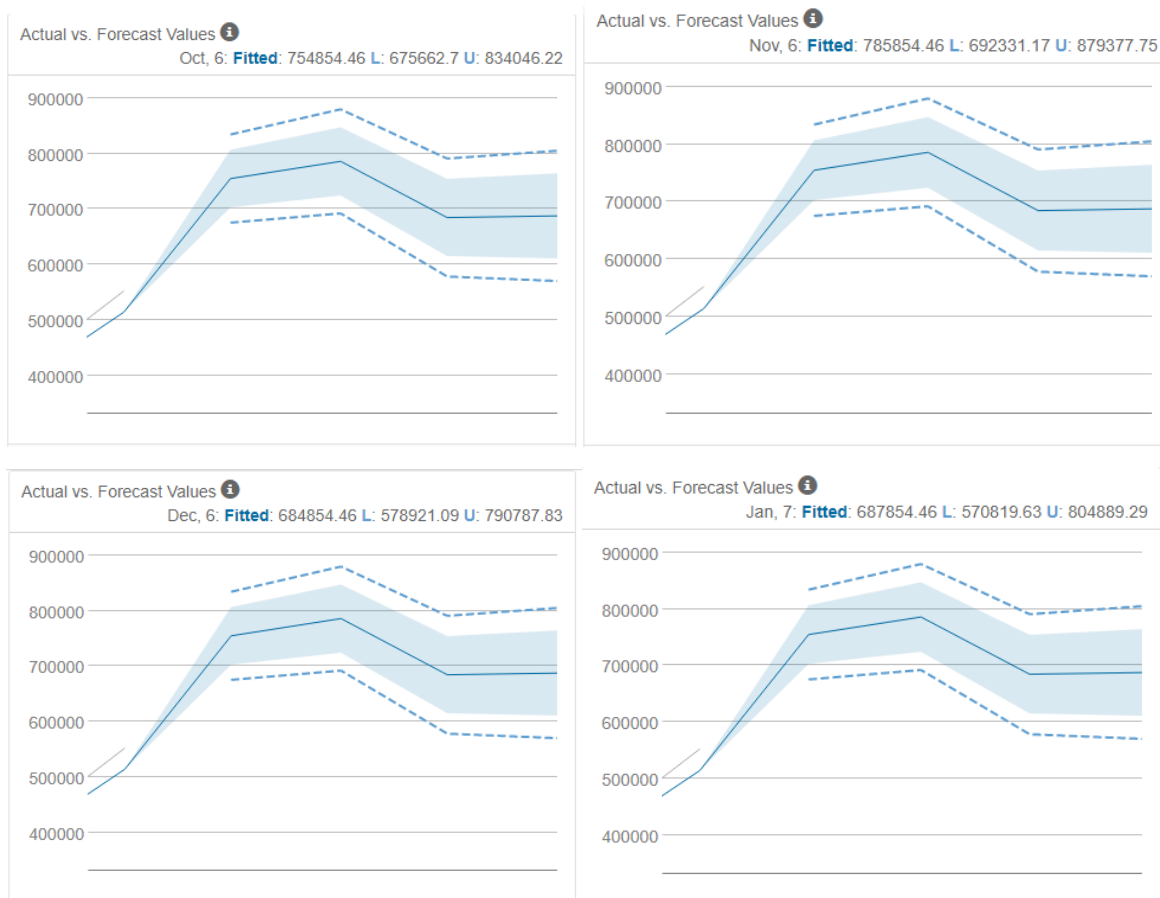
Model	ME	RMSE	MAE	MPE	MAPE	MASE
ETS_MAM_damp	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116
ARIMA_011_010	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532

Com relação aos erros verificados na amostra, o gráfico a seguir, que compara os dados reais com os dados previstos pelos modelos, demonstra que valores de predição do ARIMA são mais próximos dos observados:

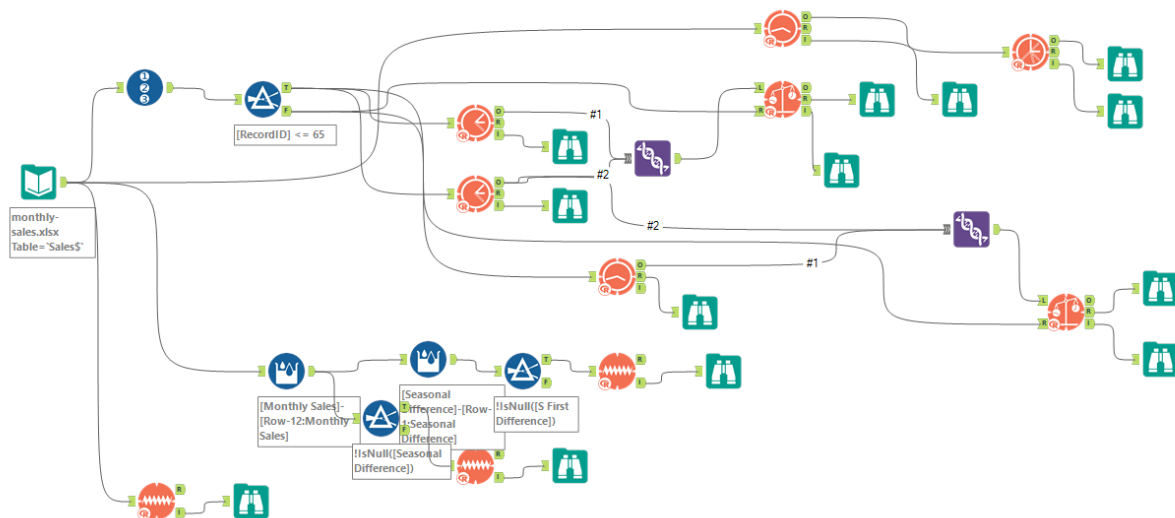
Actual and Forecast Values:

Actual	ETS_MAM_damp	ARIMA_011_010
271000	255966.17855	263228.48013
329000	350001.90227	316228.48013
401000	456886.11249	372228.48013
553000	656414.09775	493228.48013

A partir do modelo ARIMA (0,1,1)(0,1,0)[12] e considerando um intervalo de confiança entre 80% e 95%, a previsão de demanda para os meses de outubro, novembro e dezembro de 2013 e janeiro de 2014 é 754.854,46; 785.854,46, 684.854,46 e 687.854,46, respectivamente:



Fluxo no Alteryx:



Fontes:

Curso Nanodegree Análise de Dados da Udacity

Slack Udacity

<https://www.midomenech.com.br/lean-seis-sigma/downloads/artigos/157-identifica%C3%A7%C3%A3o-modelo-arma-p,d,q.html>

https://pt.wikipedia.org/wiki/S%C3%A9rie_temporal

<http://www.portalaction.com.br/series-temporais/35-medidas-de-acuracia>

<http://ucanalytics.com/blogs/arma-models-manufacturing-case-study-example-part-3/>

<https://www.youtube.com/watch?v=abBzvDijEnM>