

## Projeto: Previsão de vendas

### Passo 1: Planeje sua análise

1. O conjunto de dados atende aos critérios de um conjunto de dados da série temporal?  
Certifique-se de explorar as quatro principais características de um dado de séries temporais.

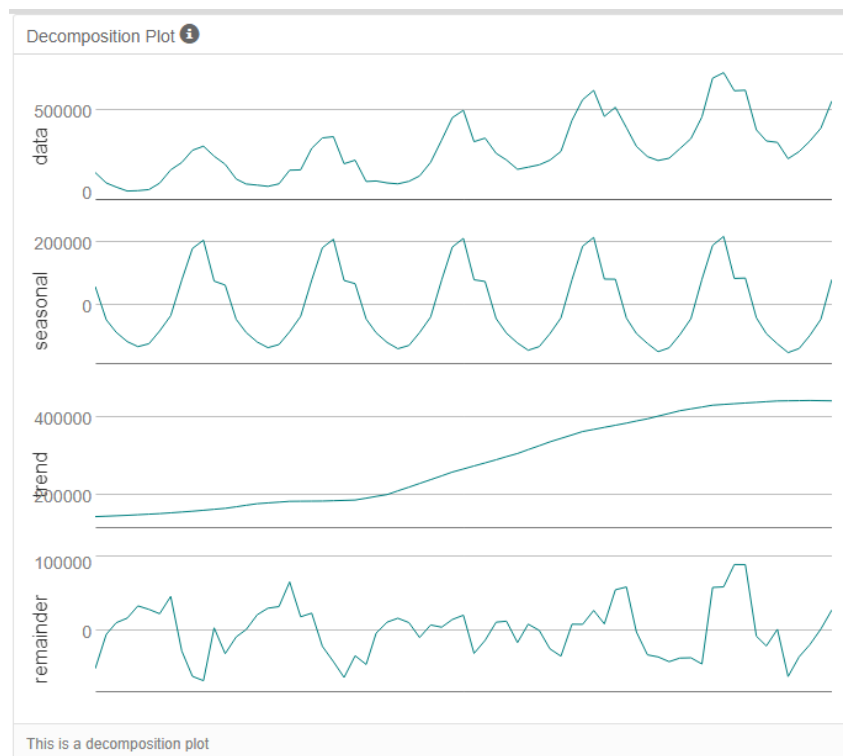
Podemos falar que o conjunto de dados atende sim aos critérios da série temporal, afirmamos isso pois os dados estão em ordem, 2008-01 à 2013-09, de forma sequencial e em intervalos de tempo iguais, além de cada unidade de tempo ter no máximo um pondo de dados.

2. Quais registros devem ser usados como amostra de retenção?

O tamanho da amostra de retenção depende do tamanho da nossa série tempos e de quantos tempo temos como expectativa de prever. Neste problema de negócio, temos como objetivo prever os próximos 4 meses de vendas, com isso podemos dizer a amostra de retenção é dos últimos 4 períodos da nossa base, ou seja, se 2013-06 à 2013-09.

### Passo 2: Determine os componentes tendência, sazonalidade e erro

1. Qual é a tendência, a sazonalidade e o erro da série temporal? Mostre como você conseguiu determinar os componentes usando gráficos de séries temporais. Inclua esses gráficos.



*Figura 1: Tendência, Sazonalidade e Erro Plot*

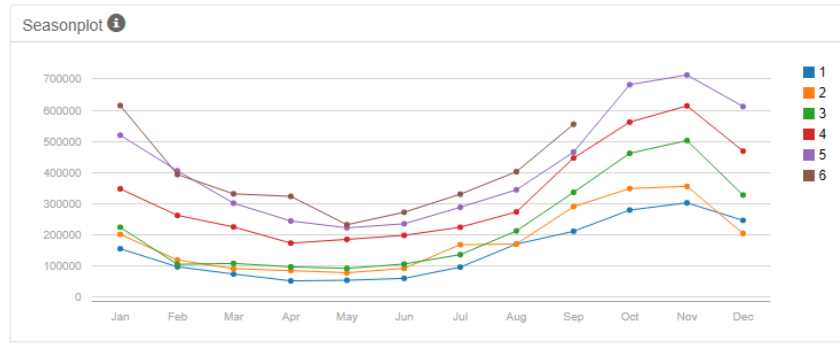


Figura 2: Erro Plot

Analisando os gráficos acima, podemos dizer que temos uma **Tendência Crescente**, pois como vemos o gráfico *Trend*, a nossa linha tem um padrão crescente.

Em relação à **Sazonalidade**, podemos ver claramente no gráfico *Seasonplot*, que temos uma queda de vendas entre Fev-Mai e um aumento de vendas entre Mai-Nov e isso ocorre em todos os anos analisados.

A variável **Erro** é demonstrado no gráfico *Remainder* e nele vemos claramente que temos uma variação ao longo do tempo. Quanto mais próximo de zero o *Remainder* estiver, o modelo interpretará os dados com maior acurácia.

### Passo 3: Construa seus modelos

1. Quais são os termos modelo para o ETS? Explique por que você escolheu esses termos.
  - a. Descreva os erros na amostra. Use pelo menos RMSE e MASE ao examinar os resultados.

Como vemos uma **sazonalidade** forte dentro dos anos, usamos o termo de modelo *multiplicativamente*. Vemos uma **tendência** de crescimento ano após ano, ou seja, uma tendência linear, então aplicamos o termo de modelo *aditivo*. Para a variável **erro**, por encontrarmos um erro que aumenta e diminui conforme o tempo, usamos o termo de modelo *multiplicativamente*. Por fim, executamos o modelo ETS *dampened* (amortecido) e não amortecido, conforme resultados abaixo.

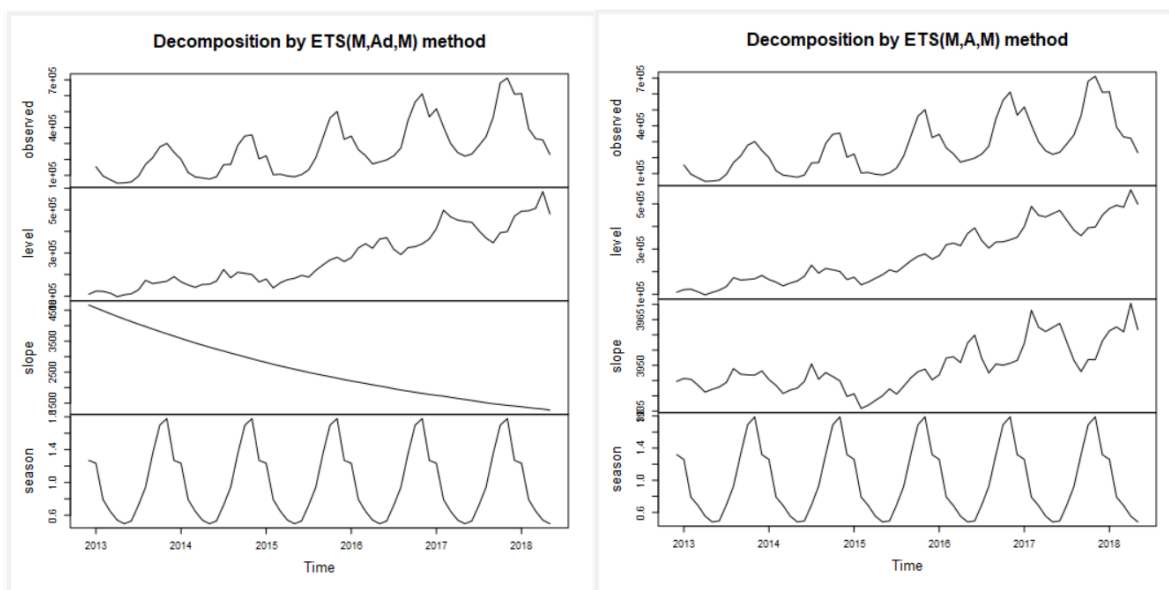


Figura 3: Decomposition Plot – Method Depended x Non-Depended

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

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

Figura 4: Error Summary – Method Depended x Non-Depended

Podemos analisar os indicadores de ambos os métodos na comparação abaixo:

Método	RMSE	MASE	AIC
Depended (M,Ad,M - Amortecido)	33153.52	0.367	1639.46
Non-Depended (M,A,M - Não Amortecido)	32992.72	0.372	1639.73

Figura 5: Comparação de Modelo ETS Amortecido e Não Amortecido

Como o MASE e o AIC são menores no modelo ETS Depended (Amortecido), podemos afirmar que este modelo é o que tem mais fit com os meus dados históricos, ou seja, esse método é o mais acurado para o nosso problema de negócio. Portanto, é ele que devemos usar para realizar nosso forecast. O MASE e o AIC podemos dizer que são algumas das principais métricas para análise de fit de modelo para séries temporais.

2. Quais são os termos modelo para o ARIMA? Explique por que você escolheu esses termos.
- Descreva os erros na amostra. Use pelo menos RMSE e MASE ao examinar os resultados.
  - Refaça os gráficos ACF e PACF tanto para a série temporal como para a diferença sazonal e inclua esses gráficos em sua resposta.

Devido ao fato dos nossos dados conterem sazonalidade e o modelo de ARIMA trabalhar com dados estacionários, nos gráficos abaixo vemos uma alta variabilidade dos dados e alta correlação nos números, devido a sazonalidade. Diante disso precisamos ajustar a série temporal para estacionária, utilizando a metodologia de diferença sazonal.

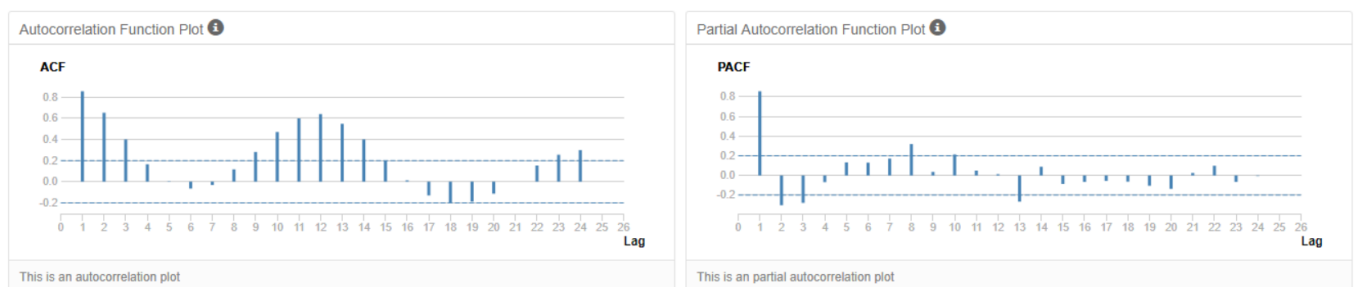


Figura 6: ACF & PACF sem diferenciação

Após fazermos a primeira transformação pela diferenciação, percebemos que os dados ficaram estacionários, ou seja, podemos considerar que a primeira diferença sazonal ajusta os nossos dados para a aplicação do modelo.

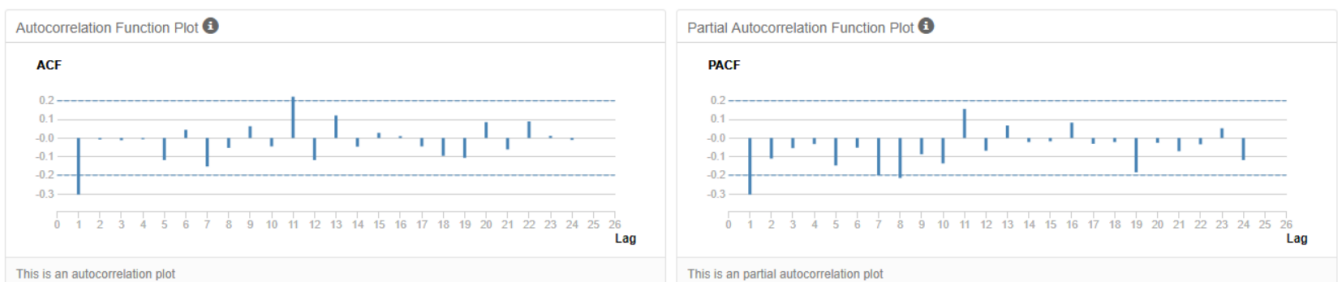


Figura 7: ACF & PACF primeira diferença sazonal

Agora que temos o modelo estacionário, podemos fazer a definição dos termos de AR e/ou MA para aplicação do modelo.

Para as séries não sazonais, como lag-1 é negativo e altamente correlacionado, podemos considerar os termos:  $AR = 0$ ,  $I = 1$  e  $MA = 1$ .

Enquanto as séries sazonais, percebemos que não há mais picos nos intervalos de 12 à 24 meses, portanto podemos assumir que os termos são:  $AR = 0$ ,  $I = 1$  e  $MA = 0$ .

Por fim, nosso modelo fica da seguinte forma:  $ARIMA(0,1,1)(0,1,0)[12]$ .

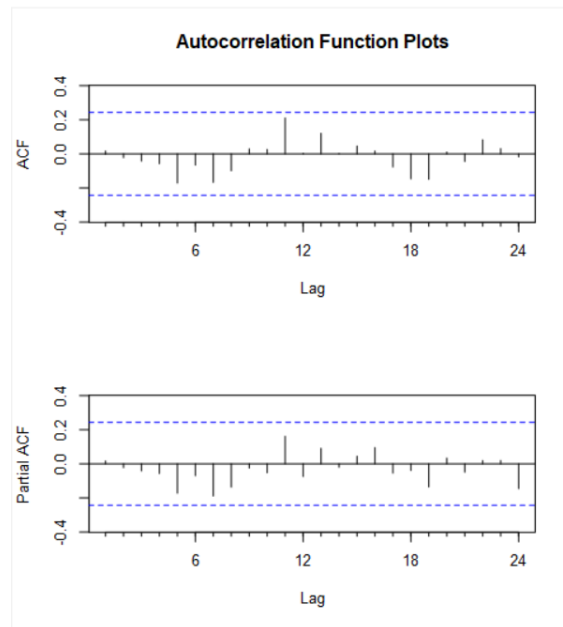
Como mostrado na tabela abaixo, os erros de previsão e precisão do modelo são: RMSE 36761,52 e MASE 0.3646.

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

*Figura 8: Erros dos Resultados*



*Figura 9: Autocorrelation Plot*

Analisando os gráficos acima, concluímos que os dados não mostram correlação, portando nosso transformamos nossa base sazonal em um modelo ARIMA estacionário.

## Passo 4: Previsão

1. Qual modelo você escolheu? Justifique sua resposta mostrando: medições de erro na amostra e medidas de erro de previsão contra a amostra de retenção.

Para fazer a comparação do melhor modelo, escolhemos o modelo ARIMA e o ETS Amortecido, visto que o modelo amortecido tinha mais acurácia que o modelo não amortecido.

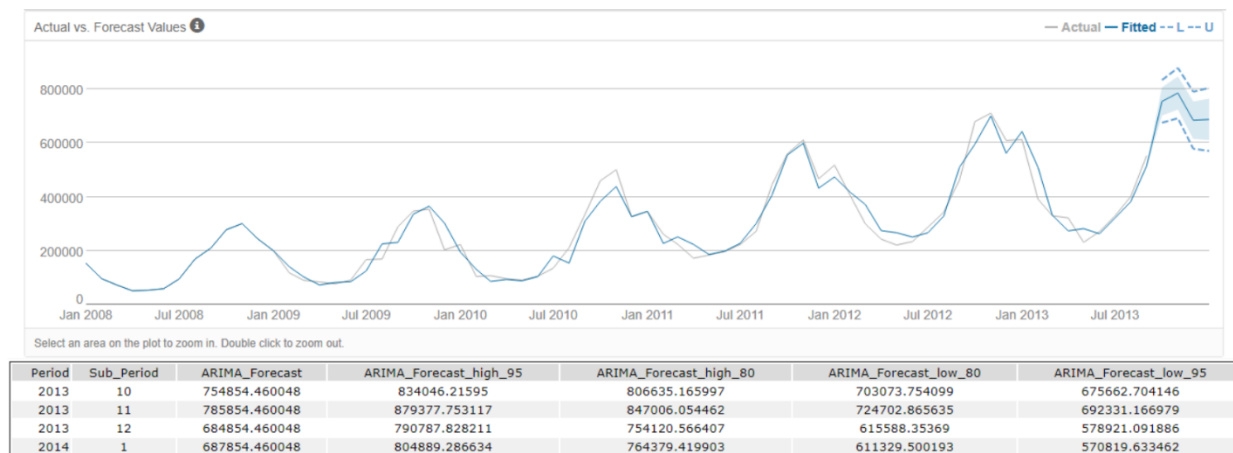
Ao analisar os resultados de ambos modelos, percebemos que o modelo ARIMA tem mais acurácia com os nossos dados, portanto é o que devemos usar para realizar o nosso Forecast. Constatamos isso, pois o RMSE para ARIMA é de 33999.79 enquanto para ETS é de 60176.47. Já para MASE, ARIMA tem acurácia de 0.4532 e ETS 0.8116.

Actual and Forecast Values:							Actual and Forecast Values:						
Actual ARIMA_Model							Actual ETS_Dempened						
271000	263228.48013						271000	255966.17855					
329000	316228.48013						329000	350001.90227					
401000	372228.48013						401000	456886.11249					
553000	493228.48013						553000	656414.09775					
Accuracy Measures:							Accuracy Measures:						
Model	ME	RMSE	MAE	MPE	MAPE	MASE	Model	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA_Model	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	ETS_Dempened	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116

Figura 10: Comparação de Modelos

2. Qual é a previsão para os próximos quatro períodos? Crie um gráfico com os resultados, usando intervalos de confiança de 95% e 80%.

A previsão para os próximos 4 meses (2013-10 à 2014-1) são: 754,854, 785,854, 684,854 e 687,854.



## Antes do envio

Compare suas respostas com os requisitos do projeto, de acordo com esta [rubrica](#). Os revisores a usarão para avaliar seu projeto.