

## Projeto: Previsão de vendas

### Passo 1: Planeje sua análise

Confira seu conjunto de dados e determine se os dados são apropriados para usar modelos de séries temporais. Determine quais registros devem ser mantidos para validação posteriormente (limite de 250 palavras).

Responda às perguntas a seguir para ajudá-lo a planejar 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.

Relembrando as características de séries temporais:

- É ao longo de um intervalo de tempo contínuo
- Existem medidas sequenciais ao longo desse intervalo
- Existe espaçamento igual entre cada duas medidas consecutivas
- Cada unidade de tempo dentro do intervalo tem no máximo um ponto de dados

Com o intuito de verificar a não existência de datas repetidas ou ausentes, a coluna *Month* foi separada em colunas de anos e meses para posterior contagem de combinações em uma tabela dinâmica, que pode ser vista ao lado.

Já a sequencialidade foi verificada através da formatação condicional utilizando a fórmula  $=A2<>(A1+1)$  na coluna de período agrupado, não indicando nenhuma entrada fora de sequência.

Com estes dois passos, podemos confirmar que todas as 4 características estão presentes.

	2008	2009	2010	2011	2012	2013	Total
01	1	1	1	1	1	1	6
02	1	1	1	1	1	1	6
03	1	1	1	1	1	1	6
04	1	1	1	1	1	1	6
05	1	1	1	1	1	1	6
06	1	1	1	1	1	1	6
07	1	1	1	1	1	1	6
08	1	1	1	1	1	1	6
09	1	1	1	1	1	1	6
10	1	1	1	1	1	0	5
11	1	1	1	1	1	0	5
12	1	1	1	1	1	0	5
Total	12	12	12	12	12	9	69

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

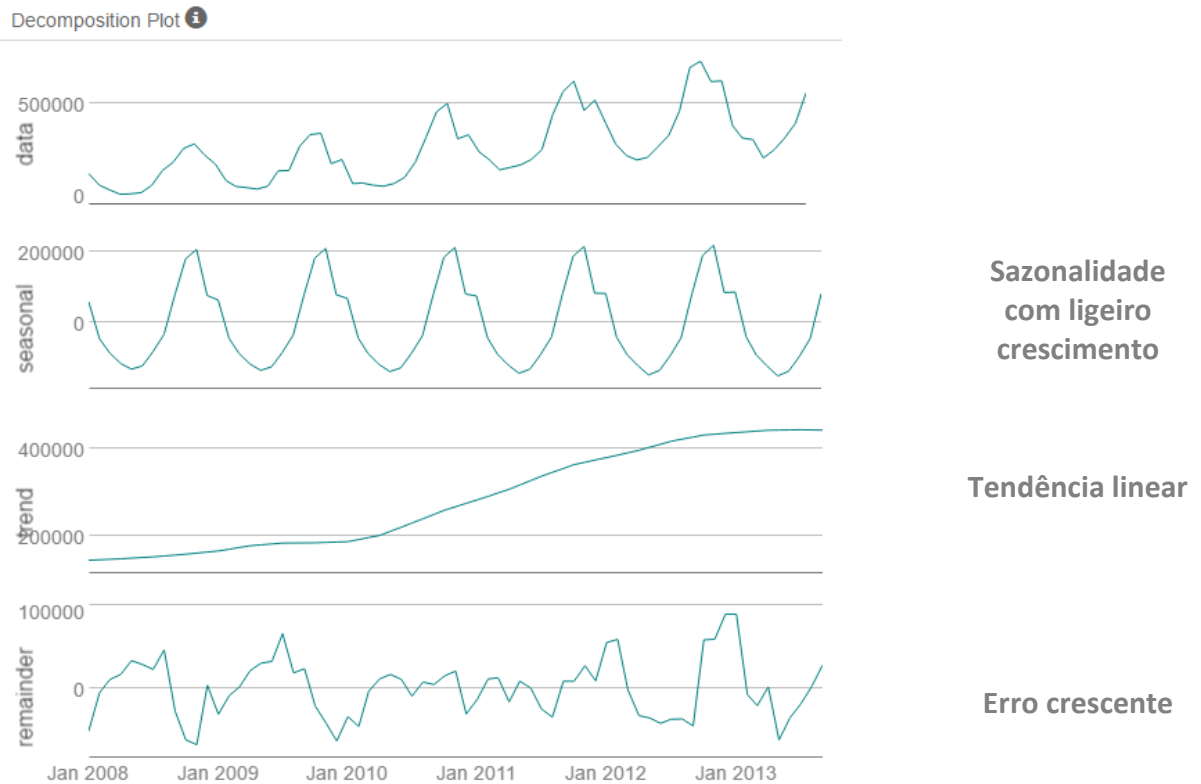
Nossa amostra de retenção deverá ter o mesmo número de períodos que desejamos prever, ou seja, devemos usar os 4 últimos meses na nossa amostra de retenção.

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

Crie um gráfico do conjunto de dados e decompõe a série temporal em seus três componentes principais: tendência, sazonalidade e erro (limite de 250 palavras).

Responda à seguinte pergunta:

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.



### Passo 3: Construa seus modelos

*Analise seus gráficos, determine as medidas apropriadas para serem aplicadas aos seus modelos ARIMA e ETS e descreva os erros de ambos os modelos (limite de 500 palavras).*

*Responda à seguinte pergunta:*

1. Quais são os termos modelo para o ETS? Explique por que você escolheu esses termos.

Tendo em vista as características de cada um dos componentes, levantados na questão anterior, o modelo que melhor se encaixaria seria um ETS(M,A,M). Assim sendo, foram criados 4 modelos para posterior análise dos resultados, sendo os dois últimos apenas com intuito de confirmação:

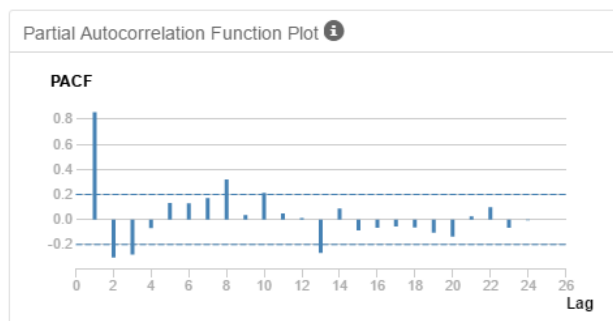
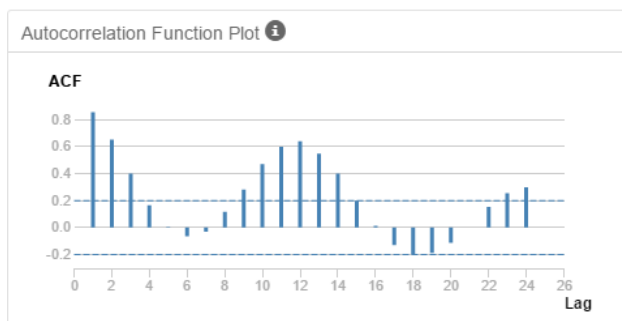
1. ETS (M,Ad,M)
2. ETS (M,A,M)
3. ETS (M,A,A)
4. ETS (Z,Z,Z) → ETS (M,Ad,M), idêntico ao primeiro

In-Sample Errors	ME	RMSE	MASE	AIC
ETS (M,Ad,M)	5597.13	33153.53	0.3675	1639.5
ETS (M,A,M)	2818.27	32992.73	0.3727	1639.7
ETS (M,A,A)	-3165.52	48206.68	0.5283	1673.4

Analisando as medidas de erro acima, concluímos o modelo (M,Ad,M) se mostra como melhor opção dentre os modelos ETS, pois, apesar de possuir um desvio padrão empírico ligeiramente acima do modelo sem *dampening*, acaba por apresentar uma redução relativa no erro - em relação à um modelo *naive* - maior e com um modelo menos complexo do que a opção sem *dampening*.

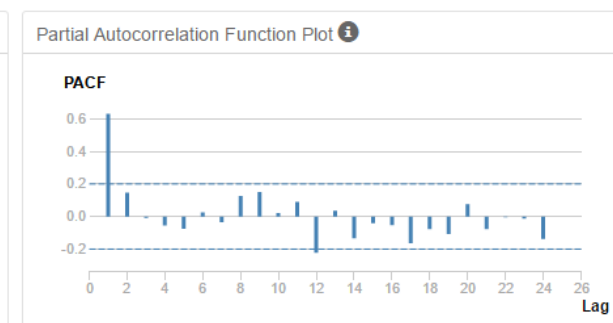
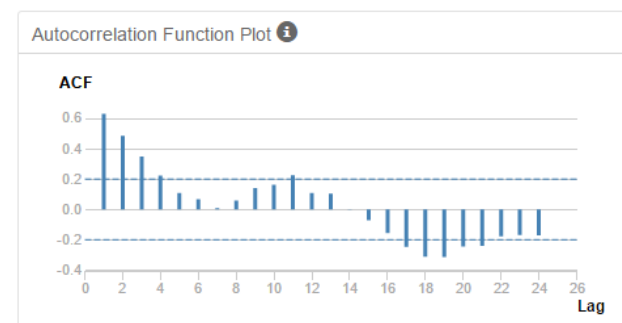
## 2. Quais são os termos modelo para o ARIMA? Explique por que você escolheu esses termos.

Ao analisar o gráfico da função de autocorrelação da série temporal original, pudemos perceber altos níveis de autocorrelação em diferentes *lag*, indicando que a série não era estacionária.



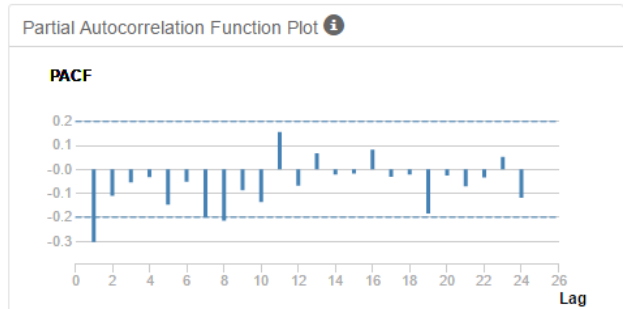
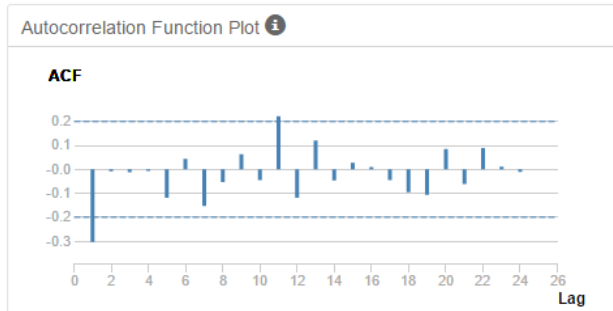
Sem diferenciação ( $D = 0, d = 0$ )

Com o intuito de tornar a série estacionária e por conta da alta autocorrelação nos lags sazonais, foi realizada uma diferenciação sazonal, com ACF e PACF ilustrados abaixo:

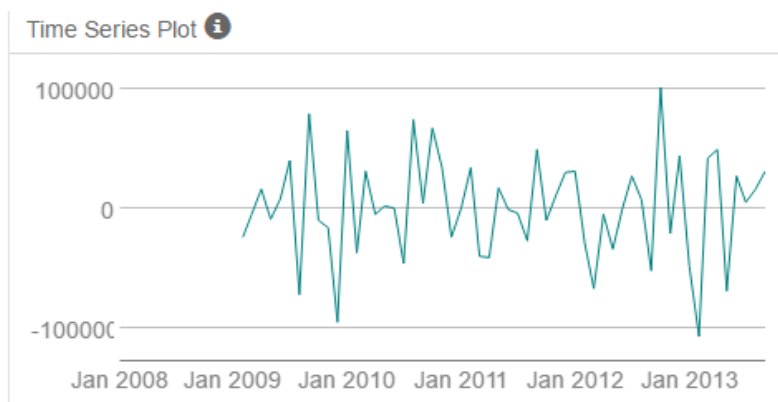


Diferenciação Sazonal ( $D = 1, d = 0$ )

Após primeira rodada de diferenciação sazonal, percebemos que a série não apresenta mais uma autocorrelação significativa nos lags sazonais (12 e 24), porém, por conta da existência de correlação nos lags sequenciais, será aplicada uma diferenciação normal na série.



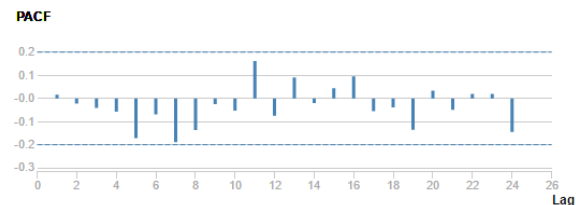
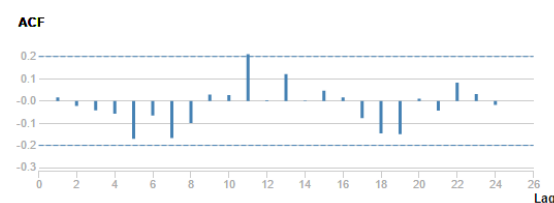
1ª diferenciação sazonal ( $D = 1, d = 1$ )



Série temporal após 1ª diferenciação sazonal ( $D = 1, d = 1$ )

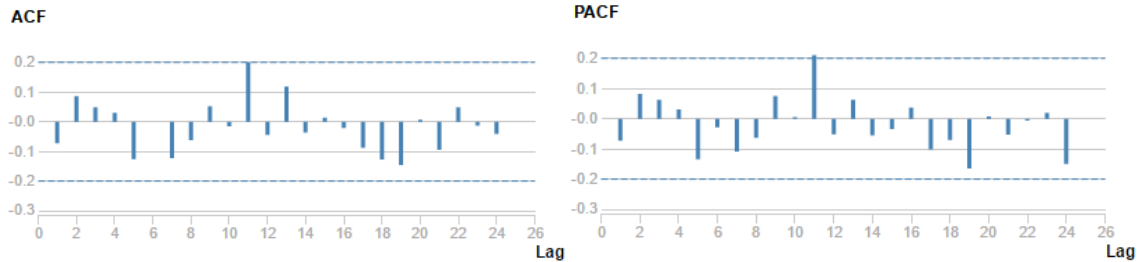
Tendo alcançado aparente estacionariedade e com as autocorrelações significativas da série eliminadas, partiu-se desta série para a definição dos termos autoregressivos e de média móvel do modelo.

Uma alta autocorrelação negativa com queda repentina pode ser observada no Lag1 do ACF e um PACF com queda gradual, indicando a necessidade de uso de um termo MA não sazonal ( $q = 1$ ).



ARIMA (0,1,1) (0,1,0) [12]

A inserção de um termo MA não sazonal foi capaz de eliminar a autocorrelação encontrada no Lag1, porém, por conta da existência de uma alta autocorrelação positiva em Lag11, testaremos acrescentar um termo não sazonal AR ( $p = 1$ ) ao modelo.



ARIMA (1,1,1) (0,1,0) [12]

Percebemos que a inserção de um termo AR foi capaz de trazer a autocorrelação presente em lag11 para um valor abaixo de 0.2, não havendo mais nenhuma correlação não sazonal significativa a ser tratada no conjunto de teste.

Com relação aos termos sazonais, não há nenhuma correlação forte nos períodos de lag sazonal (Lag12 e 24), não havendo necessidade de inserção de termos AR ou MA.

In-Sample Errors	ME	RMSE	MASE	AIC
ARIMA (1,1,1) (0,1,0) [12]	1213.68	36039.30	0.3570	1256.9
ARIMA (0,1,1) (0,1,0) [12]	-356.27	36761.53	0.3646	1256.6

Os modelos ARIMA (1,1,1) (0,1,0) [12] e ARIMA (0,1,1) (0,1,0) [12] apresentam medidas de erro e AIC semelhantes e, portanto, ambos serão validados contra a amostra de retenção.

## Passo 4: Previsão

*Compare as medidas de erro da amostra em ambos os modelos e compare as medidas de erro da amostra de retenção na sua previsão. Escolha o modelo de melhor ajuste e preveja os próximos quatro períodos (limite de 250 palavras).*

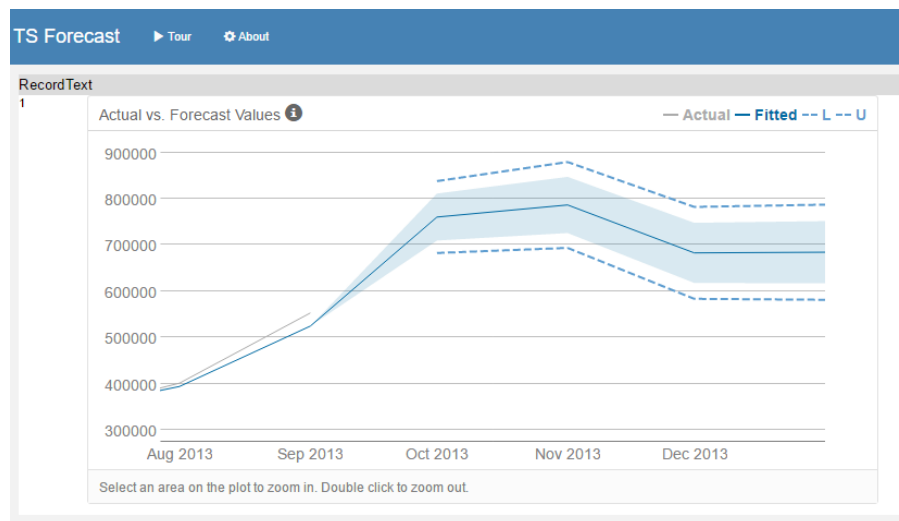
*Responda às seguintes perguntas:*

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.

Forecast Errors	ME	RMSE	MASE	AIC
ARIMA (1,1,1) (0,1,0) [12]	11,221.7	18060.35	0.2048	1256.9
ARIMA (0,1,1) (0,1,0) [12]	27,271.52	33999.79	0.4532	1256.6
ETS (M,Ad,M)	-41,317.07	60176.47	0.8116	1639.5
ETS (M,A,M)	-49,103.33	74101.16	1.0066	1639.7
ETS (M,A,A)	-38,267.74	46267.66	0.6529	1673.4

Ao comparar os 5 modelos criados acima contra a amostra de retenção, podemos observar que o modelo **ARIMA (1,1,1) (0,1,0) [12]** apresenta melhor acurácia, sendo o modelo com maior ganho em relação ao método *naive*, apesar do AIC ligeiramente inferior ao modelo sem o termo AR.

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



Period	Forecast	Forecast High 95	Forecast High 80	Forecast Low 80	Forecast Low 95
2013.10	760,617.15	838,430.10	811,496.30	709,738.00	682,804.21
2013.11	786,812.70	879,783.55	847,603.07	726,022.33	693,841.85
2013.12	683,059.13	782,556.81	748,117.17	618,001.09	583,561.45
2014.1	684,480.98	787,471.90	751,823.12	617,138.84	581,490.06