

Projeto: Previsão de vendas

Preencha todas as seções. Quando estiver pronto, salve seu arquivo como um documento PDF e o envie aqui: <https://classroom.udacity.com/nanodegrees/nd008/parts/edd0e8e8-158f-4044-9468-3e08fd08cbf8/project>

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.

R: Conjunto de dados atende sim aos critérios, onde está em conformidade para cada uma das quatro principais características da série temporal. O conjunto de dados está num intervalo de tempo contínuo de janeiro de 2008 a setembro de 2013, cada medição ocorre em sequência e existe um intervalo igual de um mês para outro. Cada unidade, mês, tem no máximo um ponto de dados.

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

R: Foi solicitado uma previsão para os próximos quatro meses. Portanto os últimos quatro registros, os períodos mais recentes, devem ser a amostra de validação. Esses quatro registros ocorrem num período entre junho de 2013 e setembro de 2013.

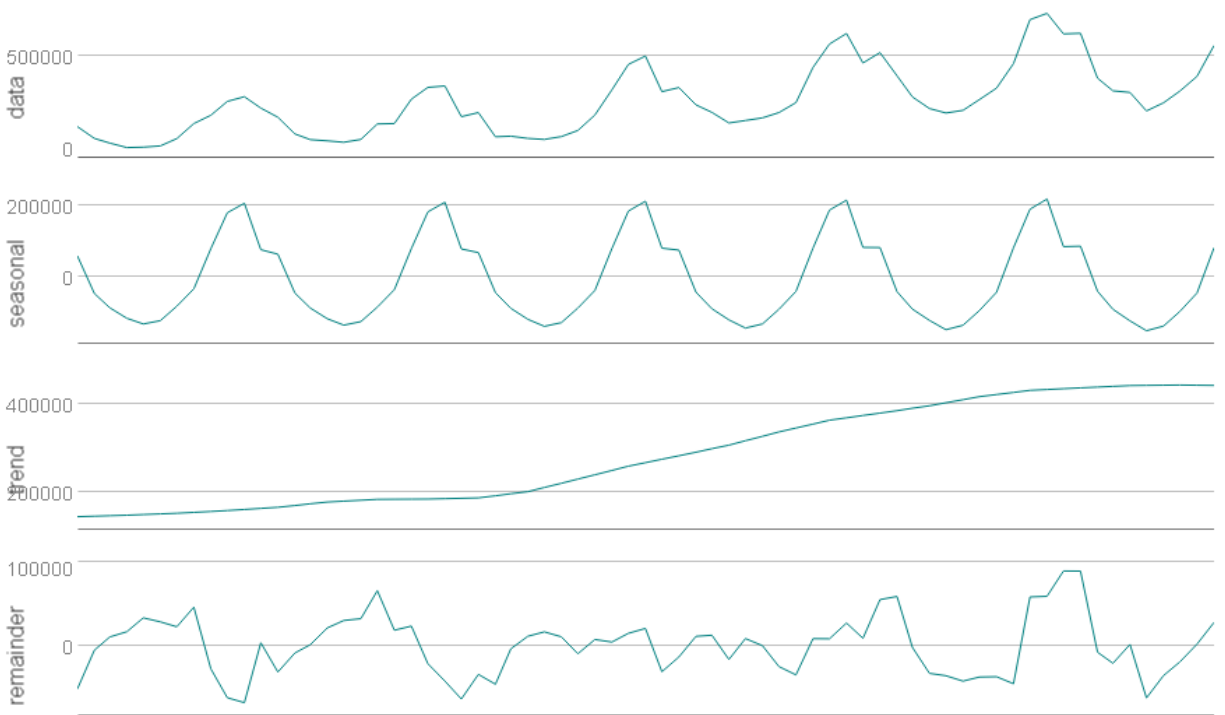
Passo 2: Determine os componentes tendência, sazonalidade e erro

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.

R: De acordo com o gráfico de decomposição abaixo, a tendência é relativamente constante e muda de forma linear ao longo do tempo, o que significa que é aditiva. Diferença de sazonalidade cresce de modo exponencial sendo assim é multiplicativa. Erro varia à medida que a série temporal se move, então é multiplicativa.

Decomposition Plot 

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.

R: Termos para o modelo ETS é MAM. Pois erro é multiplicativo, tendência é aditiva e sazonalidade é multiplicativo.

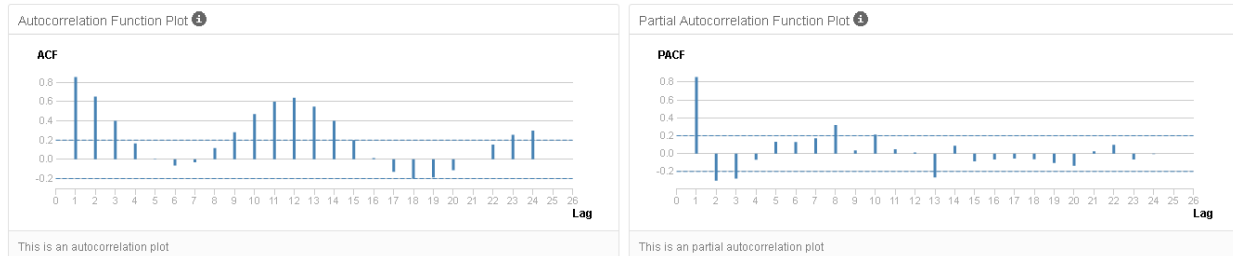
a) Descreva os erros na amostra. Use pelo menos RMSE e MASE ao examinar os resultados.

R: Erros na amostra está descrito no gráfico abaixo. RMSE representa o desvio padrão da amostra nas diferenças entre os valores previstos e observados. RMSE deste modelo ETS é 33153.5267713. MASE tem um valor de 0.3675478. Erros no MASE é bem menor que 1, onde o ideal seria estar mais próximo a 1 ou igual. Esses erros de amostra serão comparados com o modelo ARIMA para determinar o melhor modelo.

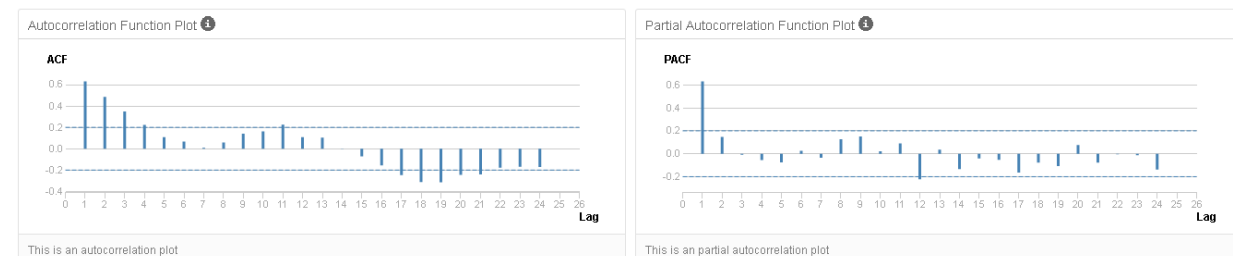
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277

2. Quais são os termos modelo para o ARIMA? Explique por que você escolheu esses termos. Crie um gráfico com a função de correlação automática (Auto-Correlation Function - ACF) e lotes de função de autocorrelação parcial (Partial Autocorrelation Function Plots - PACF) para as séries temporais e o componente sazonal e use esses gráficos para justificar a escolha dos termos do modelo.

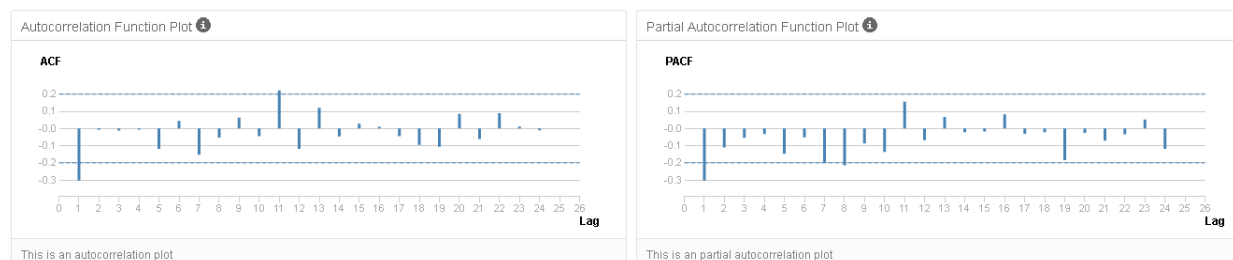
R: Segue representação do ACF e o PACF da série temporal. O gráfico ACF mostra a diferença da sazonalidade. O modelo ARIMA é representado da forma ARIMA (p, d, q) (P, D, Q) [período] devido a essa sazonalidade. Onde o período é 12 pois a série temporal é mensal.



Abaixo os gráficos sazonais ACF e PACF. A série temporal ainda exibe correlação e ainda não está estacionária. Termo de diferença sazonal é D (1). Os termos ARIMA são agora ARIMA (p, d, q) (0,1,0) [12].



Abaixo, estão representados os gráficos ACF e PACF do Seasonal First Difference. Os gráficos mostram que a série temporal está agora estacionária. A primeira diferença indica um termo d(1) e não há necessidade de mais diferenciação. Lag 1 no gráfico ACF é negativo, o que indica um modelo MA e um termo q (1). Como é raro ter um termo q (1) e um p (1), então p (0). Isto é confirmado pela falta de correlação significativa nas defasagens sazonais (12, 24, ...). O modelo ARIMA é agora ARIMA (0,1,1) (0,1,0) [12].



Segue abaixo representação ACF e PACF da série temporal após escolha dos termos do modelo final ARIMA.



a) Descreva os erros na amostra. Use pelo menos RMSE e MASE ao examinar os resultados.
R: Erros na amostra está descrito no gráfico abaixo. RMSE representa o desvio padrão da amostra nas diferenças entre os valores previstos e observados. RMSE deste modelo ARIMA é 36761.5281724. MASE tem um valor de 0.3646109. Erros no MASE é bem menor que 1, onde o ideal seria estar mais próximo a 1 ou igual. Esses erros de amostra serão comparados com o modelo ETS para determinar o melhor modelo.

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

b) Refaça os gráficos ACF e PACF tanto para a série temporal como para a diferença sazonal e inclua esses gráficos em sua resposta.
R: Pergunta respondida em gráficos anteriores.

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.

R: Escolhido o modelo ARIMA (0,1,1) (0,1,0) [12].

Abaixo os valores da 'Actual' da amostra de validação comparados com os valores de previsão dos modelos.

Actual	ETS_Auto	ARIMA
271000	255966.17855	263228.48013
329000	350001.90227	316228.48013
401000	456886.11249	372228.48013
553000	656414.09775	493228.48013

Abaixo um gráfico mostrando as medidas de erro na amostra. O modelo ARIMA tem valores absolutos menores na maioria das métricas em comparação com o modelo ETS. O modelo ARIMA produzir menos erros preditivos do que o modelo ETS.

Name	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ARIMA	-356.2665104	36761.5281724	24993.041976	-1.8021372	9.824411	0.3646109	0.0164145
ETS	5597.130809	33153.5267713	25194.3638912	0.1087234	10.3793021	0.3675478	0.0456277
Best_Model	ARIMA	ETS	ARIMA	ETS	ARIMA	ARIMA	ARIMA

Foi criado um gráfico onde que mostra as diferenças absolutas e relativas entre os valores 'Actual' e os valores previstos dos modelos ETS e ARIMA. O campo final mostra o melhor modelo, conforme determinado pelo menor valor absoluto da diferença relativa de cada modelo. O modelo ARIMA previu com mais precisão.

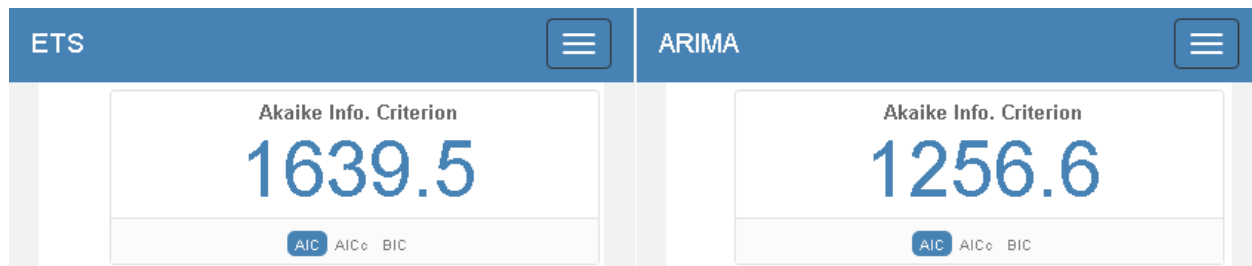
Actual	ETS	ARIMA	ETS_Abs_Diff	ETS_Rel_Diff	ARIMA_Abs_Diff	ARIMA_Rel_Diff	Best_Model
271000	255966.17855	263228.48013	15033.82145	0.055475	7771.51987	0.028677	ARIMA
329000	350001.90227	316228.48013	-21001.90227	-0.063836	12771.51987	0.038819	ARIMA
401000	456886.11249	372228.48013	-55886.11249	-0.139367	28771.51987	0.071749	ARIMA
553000	656414.09775	493228.48013	-103414.09775	-0.187006	59771.51987	0.108086	ARIMA

Abaixo estão as medidas de precisão dos modelos de ETS e ARIMA. Modelo ARIMA supera o modelo ETS em cada medida, possuindo valores absolutos menores. Isso mostra que o modelo ARIMA prevê melhor em praticamente todas as métricas.

Accuracy Measures:

Model	ME	RMSE	MAE	MPE	MAPE	MASE	NA
ETS_Auto	-41317.07	60176.47	48833.98	-8.3683	11.1421	0.8116	NA
ARIMA	27271.52	33999.79	27271.52	6.1833	6.1833	0.4532	NA

Validado as medidas da AIC para cada modelo também. Melhor modelo é geralmente aquele com menor pontuação na AIC. Neste caso, o modelo ARIMA tem a melhor pontuação e isso se enquadra com o restante das comparações que apontam para o modelo ARIMA.

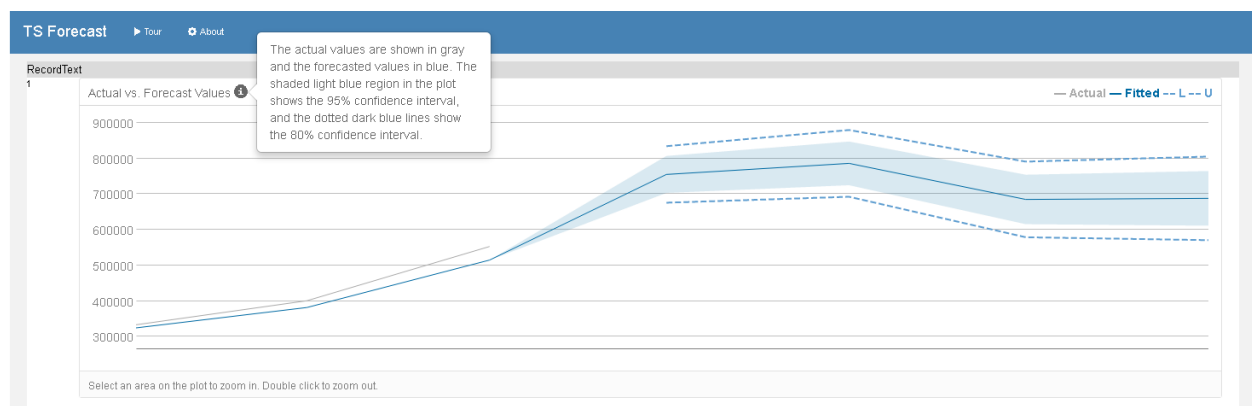


2. Qual é a previsão para os próximos quatro períodos? Crie um gráfico com os resultados, usando intervalos de confiança de 95% e 80%.

Segue abaixo a previsão para os próximos quatro períodos, além dos intervalos de confiança de 80% e 95%.

Period	Sub_Period	Forecast	Forecast_high_95	Forecast_high_80	Forecast_low_80	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

Abaixo gráfico com previsão com intervalo de confiança de 80% e 95%.



Antes do envio

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