



67%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODO
NOTURNO



15.4k xp

a

03

Statsmodel

Transcrição

Sempre que vamos estudar uma série temporal temos que ter em mente componentes como a tendência.

Por exemplo, na análise que fizemos das vendas de carros da Alucar, a tendência é o crescimento, ou seja, esse movimento da série de valores se torna mais alto. Ela se difere da análise que fizemos do aumento de assinantes da newsletter, que começou com uma tendência de crescimento, mas depois houve uma queda muito significativa.

Outro ponto muito importante e presente nas Time Series, além da tendência, é a sazonalidade. Nos casos que analisamos em aulas anteriores tivemos sempre dois pontos (ou meses) em que as vendas estouraram, ou vimos vendas maiores durante a semana do que aos finais-de-semana. Isso significa que a sazonalidade se refere a um padrão repetitivo dentro de um período fixo de tempo.

Também teremos que ter claro que os ruídos serão comuns nas séries temporais, e pensaremos em formas de normalizar esses ruídos, assim como normalizamos os picos de vendas por quantidade de dias de finais de semana se tratando das vendas da Chocolura, e tivemos um gráfico com uma linha mais homogênea.



67%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNOS



15.4k xp

a

Resumindo, sempre que analisarmos Time Series será necessário identificar componentes como a tendência, a sazonalidade e essas variações aleatórias, os ruídos. Esses elementos sempre serão presentes nas séries temporais.

Discutiremos agora: Será que existe alguma forma de passar os dados e eles nos mostrarem o que será a tendência, a sazonalidade e o ruído de uma só vez?

Para visualizar essas informações, criaremos uma célula do tipo texto clicando no botão "Text" na barra superior, colocaremos uma hashtag (#) e escreveremos o título "Statsmodels". Faremos a importação dessa biblioteca.

```
from statsmodels.tsa.seasonal import seasonal_decompose
```

Pressionaremos "Shift + Enter". Agora, para conseguir utilizar essa função, que nos mostrará nossa observação, o que será a tendência, a sazonalidade e o ruído, teremos que passar uma lista para ela. Faremos o passo a passo de atribuir a lista a uma variável chamada resultado.

Para esse exemplo, a lista que passaremos será a de vendas da Chocolura. Nesse código teremos que passar o parâmetro da frequência, que deverá ser um número inteiro a ser exibido. Atribuiremos esse resultado a `ax` e faremos um plot dele.

```
resultado = seasonal_decompose(chocolura['vendas'], freq=1)  
ax = resultado.plot()
```

COPIAR CÓDIGO

Mais uma vez pressionaremos "Shift + Enter" e receberemos os nossos próprios dados das vendas, assim como gráficos de tendência, sazonalidade e ruído. Vamos mudar a frequência para 2.



67%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNOS



15.4k xp

a

Quando executarmos, a observação continuará a mesma, mas nossa tendência será alterada. A sazonalidade seguirá um padrão de ascender e cair. E teremos bastante ruído.

Com a frequência 3, teremos o crescimento no ponto da observação. Na tendência, o valor aumentará, cairá, depois aumentará e cairá de novo. A sazonalidade segue o padrão e o ruído não traz o início. Isso indica que começamos a perder referência conforme o nível de frequência que colocamos.

Por exemplo, se colocarmos 10, não conseguiremos ver mais alguns dados na tendência nem no ruído. A sazonalidade ficará semelhante à própria observação apesar de distorções do gráfico.

Usaremos a frequência 3. Vamos exibir esses gráficos em tamanho maior utilizando nossa função `plot_comparacao`. Para isso, chamaremos as respectivas variáveis. Quando escrevermos "resultado." e dermos um "Tab", o programa nos mostrará uma lista de opções e escolheremos a que corresponder à nossa variável.

```
observacao = resultado.observado
tendencia = resultado.trend
sazonalidade = resultado.seasonal
ruído = resultado.resid
```

COPIAR CÓDIGO

Armazenamos a lista em todas as variáveis. Agora, criaremos um DataFrame com todos esses valores. Visualizaremos ele logo depois.



67%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNO



15.4k xp

a

```
data = ({
    'observacao':observacao,
    'tendencia':tendencia,
    'sazonalidade':sazonalidade,
    'ruído':ruído
})
resultado = pd.DataFrame(data)
resultado.head()
```

COPIAR CÓDIGO

Serão mostradas as colunas e os valores. Agora faremos o plot de comparação, passando os parâmetros. O x será o próprio index do resultado. `plot_comparacao(x, y1, y1, y3dataset, titulo)`

```
plot_comparacao(resultado.index, 'observacao', 'tendencia', 'sazonalidade', resultado,
'Exemplo de Statsmodels')
```

Pressionaremos "Shift + Enter" e veremos que serão criados os gráficos baseados nos dados que vimos anteriormente à frequência 3. Teremos a observação; a tendência, em que o gráfico sobe e depois sofre uma queda; e a sazonalidade, em que há aumento e queda repetidas vezes. Em vez de sazonalidade, podemos chamar esse parâmetro de ruído no código para visualizá-lo.

```
plot_comparacao(resultado.index, 'observacao', 'tendencia', 'ruído', resultado, 'Exemplo de
Statsmodels')
```

Então, teremos o valor do ruído dessa nossa observação. Mas deixaremos mesmo como a sazonalidade, pois criamos uma função para plotar apenas 3 gráficos.



67%

ATIVIDADES
3 de 6

DISCORD
ALURA

FÓRUM DO
CURSO

VOLTAR
PARA
DASHBOARD

MODOS
NOTURNO



15.4k xp

a

O Statsmodels, portanto, é um exemplo de como podemos utilizar uma função para visualizar todas as informações que desejamos. Vimos também que a frequência é muito importante para que os resultados sejam visualizados corretamente.

A documentação da frequência poderá ser acessada quando, no trecho do código em que ela aparece "freq=3,...", digitarmos uma vírgula e apertarmos "Enter". Esse documento dirá que essa será uma int opcional. No entanto, a partir da versão do Pandas que estamos utilizando no curso, o Colab não permitirá que ela seja opcional. Precisaremos passá-la para ter um entendimento mais claro, já que estamos trabalhando com Series.

Temos agora uma visão melhor do que se trata o Statsmodels. Com ele podemos criar também outros exemplos de gráficos de outros DataSets que já trabalhamos até aqui.