**Time series do nosso dia a dia**

Aprendemos que time series é uma coleção de observações ou dados, geralmente feita através de um tempo sequencial, que pode ser segundos, minutos, dias, meses, anos, entre outros.

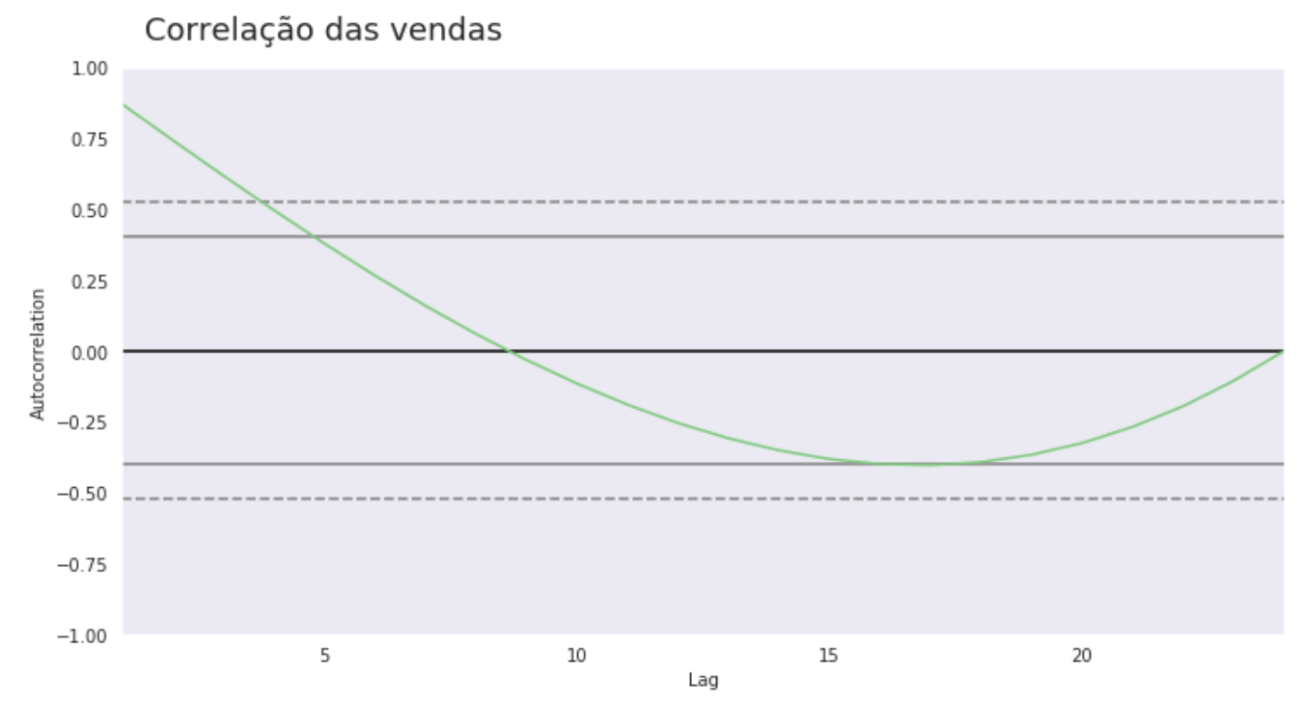
Tipicamente, esse tempo é espaçado entre cada ponto de dados, que podemos medir o crescimento das vendas, como visto nesta primeira aula.

Há um post no blog da Alura que podemos aprender sobre [Crescimento proporcional entre time series utilizando pandas](https://blog.alura.com.br/crescimento-proporcional-entre-linhas-time-series-utilizando-pandas/).

**Autocorelação**

A função de autocorrelação é a principal ferramenta de diagnóstico para análise de time series.

Autocorrelação significa a correlação de valores de uma mesma variável ordenados no tempo.



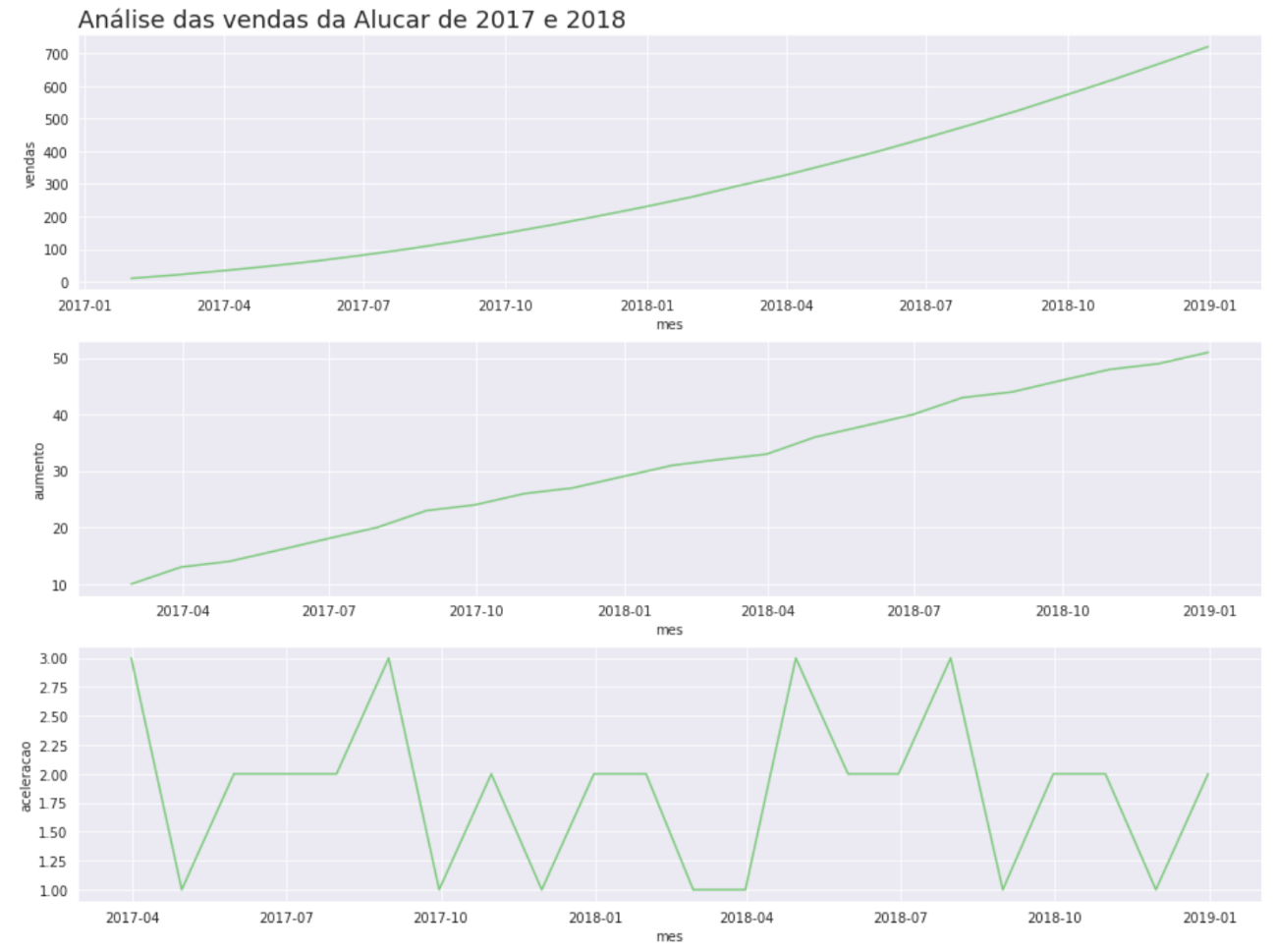
Para melhor interpretação do gráfico, podemos analisar a tabela abaixo:

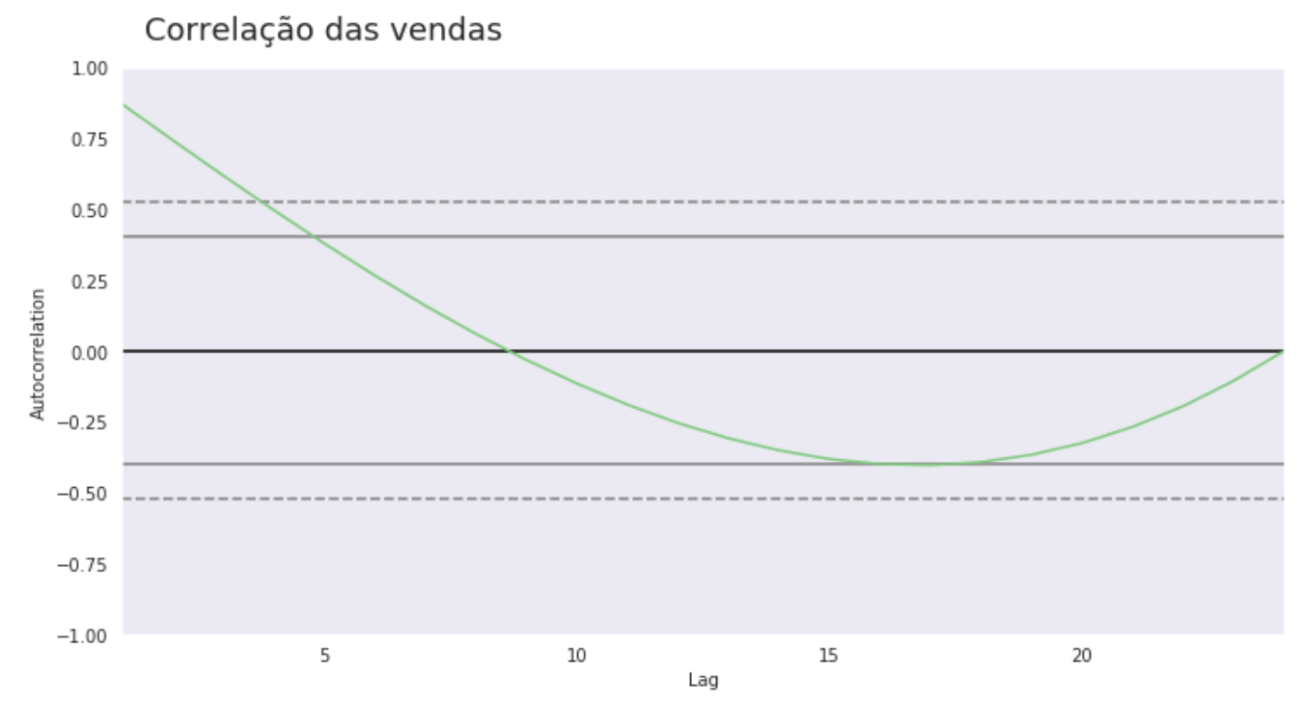
| **Autocorrelação** | **Interpretação** |
| --- | --- |
| 1 | correlação perfeita positiva |
| 0,7 a 0,9 | correlação forte |
| 0,4 a 0,7 | correlação moderada |
| 0,2 a 0,4 | correlação fraca |
| 0 | correlação nula |
| < 0 | correlação negativa |
| -1 | correlação perfeita negativa |

[Neste link](https://pt.wikipedia.org/wiki/Autocorrela%C3%A7%C3%A3o), você encontra mais informações sobre autocorrelação.

Revisão da aula

* Aplicamos a técnica de Decomposição de uma time series, para mensurar o crescimento mês a mês
* Utilizamos a função diff() para decomposição das vendas para criar o aumento, e do aumento para descobrir a aceleração
* Executamos a função de Autocorrelação para descobrir o nível de correlação das vendas, do aumento e da aceleração
* Criamos uma função chamada plotar() e plot\_comparacao() para geração de gráficos padronizados, evitando código duplicado



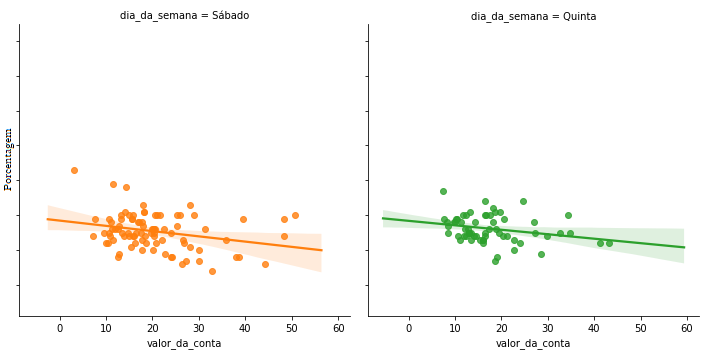


**Quando não aplicar análise de time series**

Sabemos que time series ou series temporais são uma forma de organizar no tempo as informações quantitativas e que podemos extrair muitas informações deste tipo de dado. Porém, nem sempre podemos aplicar a análise de time series.

Abaixo temos duas recomendações de quando não aplicar análise de time series:

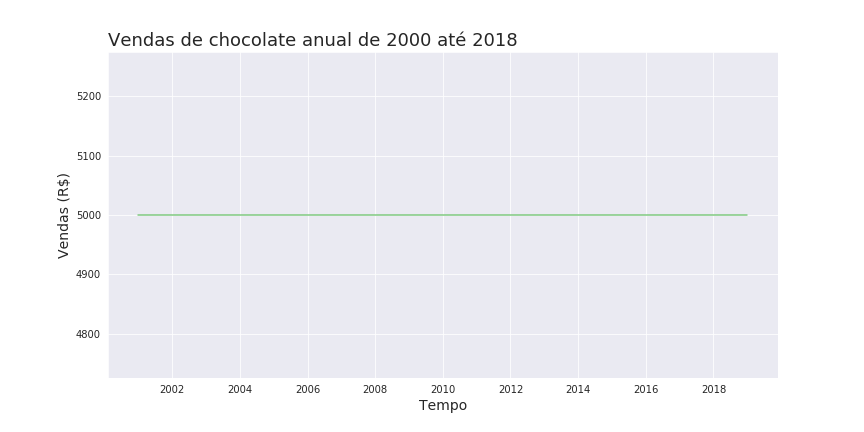
**Não temos a variável tempo**



Analisando os gráficos acima, não temos em nenhum dos eixos, a variável **tempo**.

A principal variável em uma time series é o tempo. Sendo assim, a ausência desta variável impossibilita análise de séries temporais.

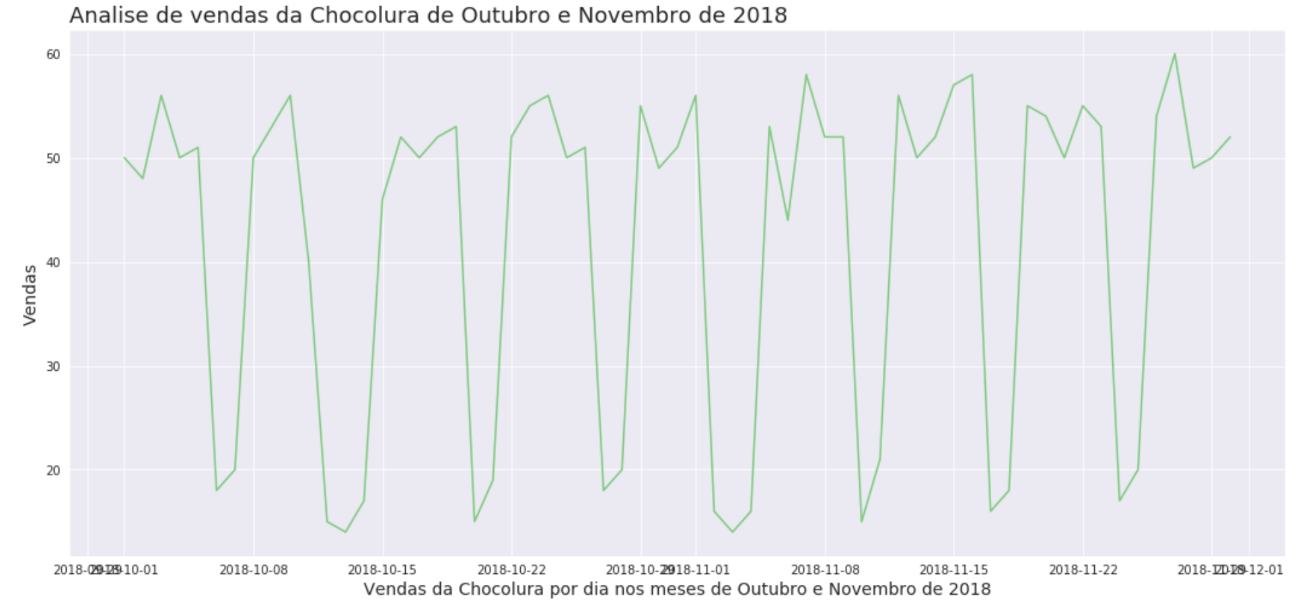
**As variáveis ou valores são constantes**



Analisando o gráfico acima de uma loja que vende 5000 caixas de chocolates há 18 anos, não podemos aplicar análises de séries temporais para extrair mais informações dos dados.

**Nessa aula:**

* Vimos a importância da técnica de Decomposição na análise dos assinantes da newsletter da Alucar
* Analisamos as vendas da Chocolura e descobrimos um padrão repetitivo no movimento das vendas dentro de um período de tempo fixo, na qual é chamado de Sazonalidade
* Examinamos as vendas de uma determinada loja da Chocolura nos meses de Outubro e Novembro de 2018, e descobrimos que também havia uma sazonalidade
* Investigamos o que causava a sazonalidade desta loja



**Encontrei dados nulos... e agora?**

Se tentarmos plotar um gráfico que possui dados nulos, uma mensagem de erro será exibida informando que não temos tipos numéricos para agregar.

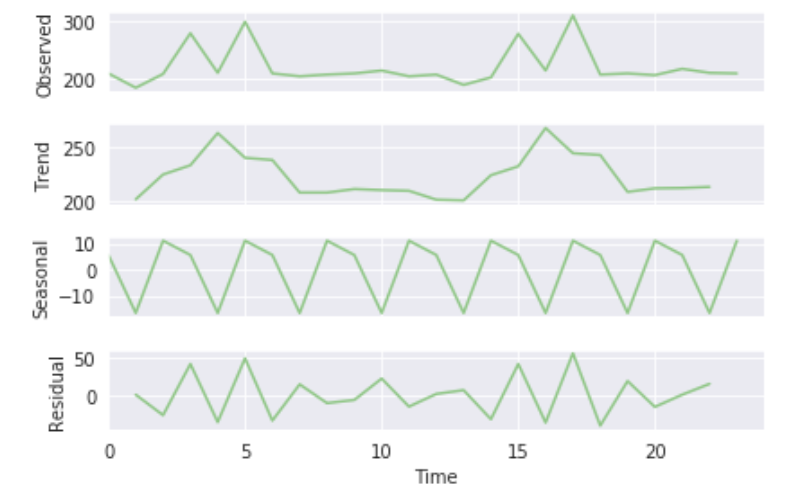
Pensando nisso, dentre as inúmeras possíveis soluções de contornar esse problema, [criei este caderno com 5 possíveis soluções para contornar a falta dos dados](https://colab.research.google.com/drive/1UZFN5Jke6gFgYLNMbsU7eBJ9Pk9TdpIm).

Dentre as soluções listadas no notebook acima, temos:

* Preenchendo os dados nulos com o valor médio da coluna
* Preenchendo os dados nulos com um valor específico
* Deletando os dados nulos do DataFrame
* Preenchendo um dado nulo com o valor anterior ou posterior
* Interpolando dados nulos com a média

## Nessa aula:

* Aplicamos a função de Autocorrelação nas vendas, no aumento e na aceleração das vendas diárias e vimos que havia uma correlação entre elas
* Executamos uma técnica de normalização de time series para minimizar as frequências pela quantidade de dias de finais de semana de cada mês
* Analisamos uma time series importando da biblioteca statsmodels.tsa.seasonal a função seasonal\_decompose, que nos mostra o que é a nossa observação, tendência, sazonalidade e ruído de uma só vez



## Nessa aula:

* Aprendemos que um componente presente na maioria das Time Series é o ruído
* Vimos que podemos minimizar os ruídos de uma time series aplicando a técnica da média móvel
* Criamos um gráfico com a média móvel de 7 e 21 dias e comparamos com nossa observação, conforme ilustra a imagem abaixo

