

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG[PROGRAMAÇÃO](#)[FRONT-END](#)[DATA SCIENCE](#)[INTELIGÊNCIA ARTIFICIAL](#)[DEVOPS](#)[UX & DESIGN](#)[MOBILE](#)[INOVAÇÃO & GESTÃO](#)[Artigos > Data Science](#)

## Melhorando a análise com o Boxplot

**Daniel Siqueira**

02/05/2021

Atualizado em 14 de janeiro

**COMPARTILHE**

30%

[Compre com desconto](#)

TÁ ACABANDO!

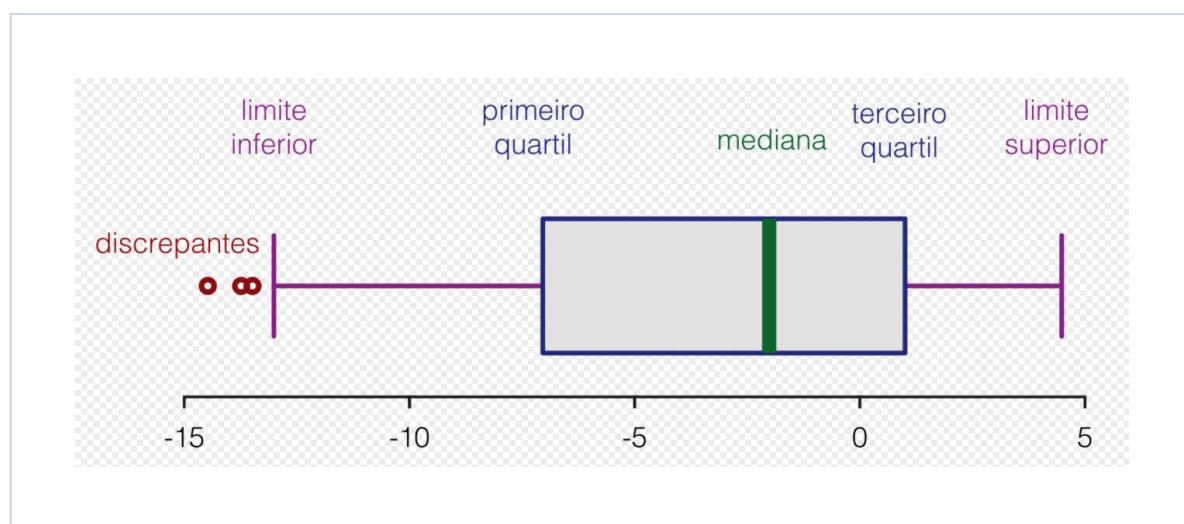
02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Um boxplot (ou diagrama de caixa, numa tradução livre) mostra a distribuição quantitativa dos dados de um jeito que facilita a comparação entre as variáveis, ou através dos níveis categóricos das variáveis.

**Confira neste artigo:**

- [Quer saber mais sobre o assunto?](#)

Essa caixa ("box") mostra os quartis do dataset enquanto os "whiskers" mostram o resto da distribuição, exceto os pontos que são chamados de **outliers**.



30% [Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

---

A linha no centro da figura em verde, no nosso caso, representa a mediana. A linha azul que está na aresta esquerda da figura representa 25% da minha distribuição, e a linha que está na aresta direita representa 75% da distribuição dos dados.

Nas linhas roxas que estão nas extremidades de ambos lados, saem da figura central e mostram uma delimitação, temos o início e o fim da área dos quartis. Agora, repare que há alguns pontos após essa delimitação no lado esquerdo. Esses pontos são chamados de outliers, ou seja, são pontos “isolados” na nossa distribuição.

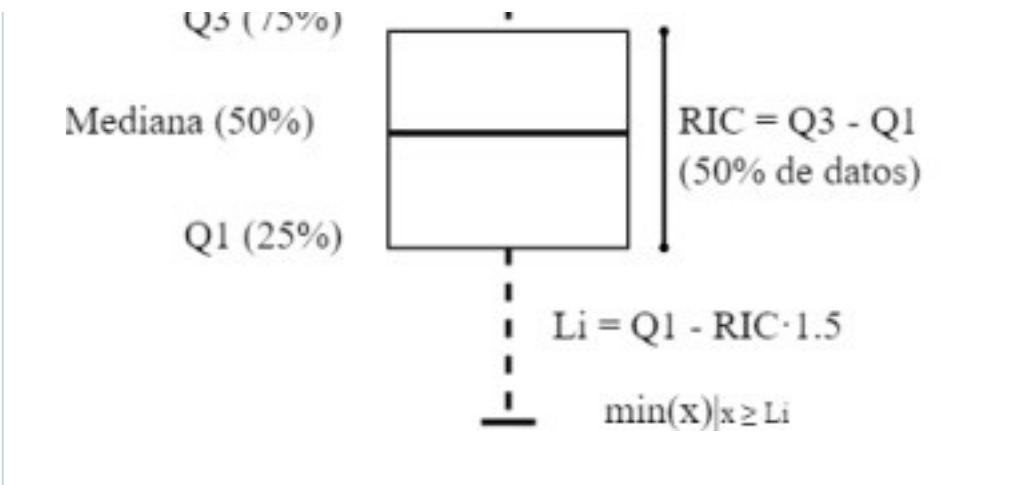
Para sabermos até onde vão os whiskers, que são esses “bigodinhos de gato” que delimitam os quartis e marcam onde se inicia os outliers, vamos fazer um cálculo. Uma vez que encontramos a mediana, o ponto central da distribuição, sabemos que temos 50% da distribuição à esquerda e 50% à direita. Após isso, vamos dividir novamente os lados na metade, para obtermos duas partes de 25% à direita e 25% à esquerda.

Agora que temos os 4 quartis, cada um com 25%, podemos desenhar a caixa, o “box”, e delimitarmos com o 2º e o 3º quartis. Para encontrarmos então a delimitação dos whiskers vamos multiplicar a distância do 2º quartil até a mediana por 1,5. Fazemos o mesmo procedimento com a distância do 3º quartil até a mediana: vamos multiplicá-la por 1,5.

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Em suma, o boxplot nos auxilia a visualizar a distribuição de dados divididos nos quartis. Ademais, mostra onde os dados estão mais concentrados e se há outliers fora dos nossos quartis.

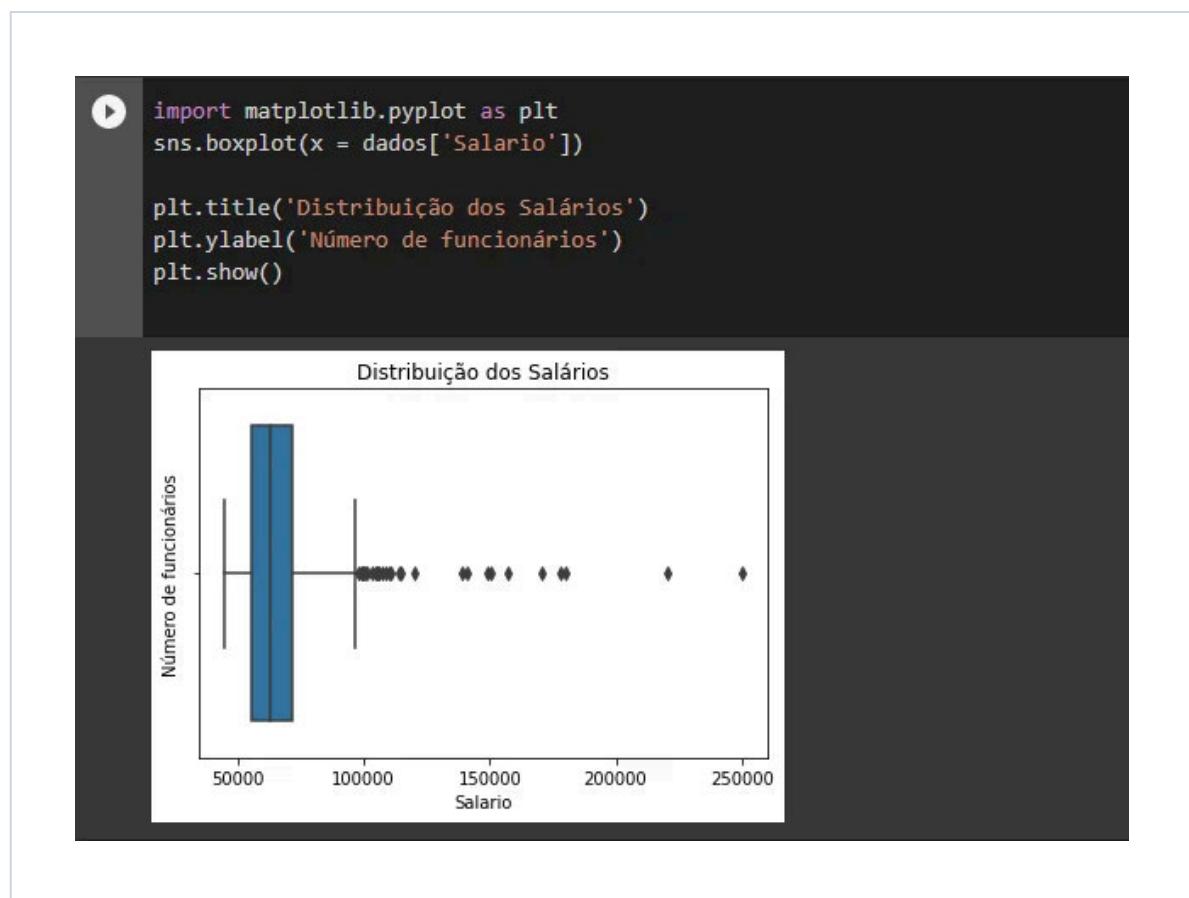
### Como gerar um BoxPlot usando Python?

Para gerar o Boxplot no [Python](#) vamos usar a biblioteca do Seaborn. Nós poderíamos também gerar o Boxplot através de outros métodos, porém uma das vantagens de se utilizar o Seaborn é que ele ficará mais bonito, claro e apresentável.



Para alterar os eixos x e y do nosso diagrama podemos utilizar a biblioteca matplotlib, através do código:

```
import matplotlib.pyplot as plt
sns.boxplot(x = dados['Salario'])
plt.ylabel('Número de funcionários')
plt.show()
```



30%

[Compre com desconto](#)

TÁ ACABANDO!

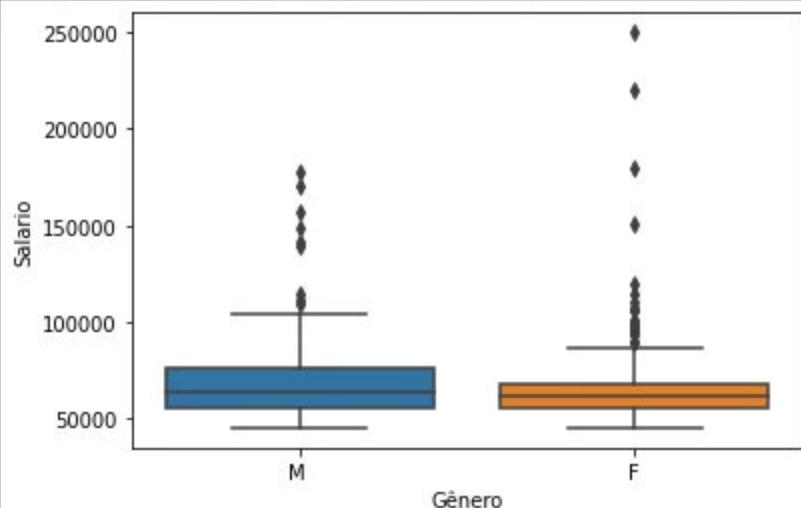
02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Boxplot com diferentes categorias (salários) para fazer análises de **comparação** entre elas.

Vamos fazer um experimento: analisar a que tipos de conclusões podemos chegar comparando o salário anual com o gênero do(a) funcionário(a). Para isso, vamos gerar o boxplot com o seguinte código:

```
sns.boxplot(data=dados, x='Gênero', y='Salario')
```

```
C: <matplotlib.axes._subplots.AxesSubplot at 0x7faaadd4f590>
```

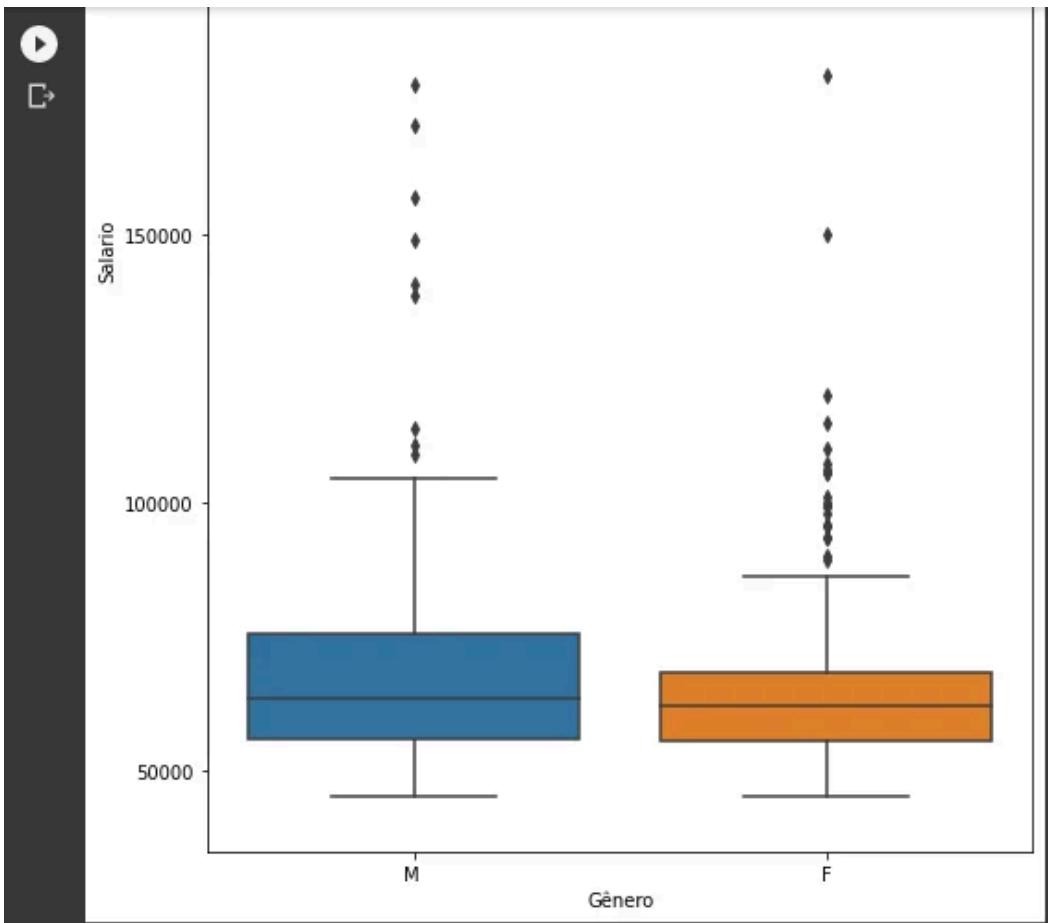


Como na categoria do gênero feminino temos mais outliers, o gráfico ficou um pouco achatado. Vamos utilizar o matplotlib para melhorar um pouco a visualização:

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Vamos analisar um pouco esses boxplots. Não conseguimos responder diretamente se os homens ganham mais que as mulheres, ao menos somente com este passo, já que percebemos que a mediana dos dois boxplots está bem próxima uma da outra. Mas conseguimos analisar que o 3º quartil do gênero feminino acaba bem antes do gênero masculino, o que pode nos dizer que conforme o salário anual aumenta temos uma maior concentração de pessoas do gênero masculino. Percebemos também que há muito mais outliers, dados atípicos na distribuição, de pessoas do gênero feminino que têm altos salários.

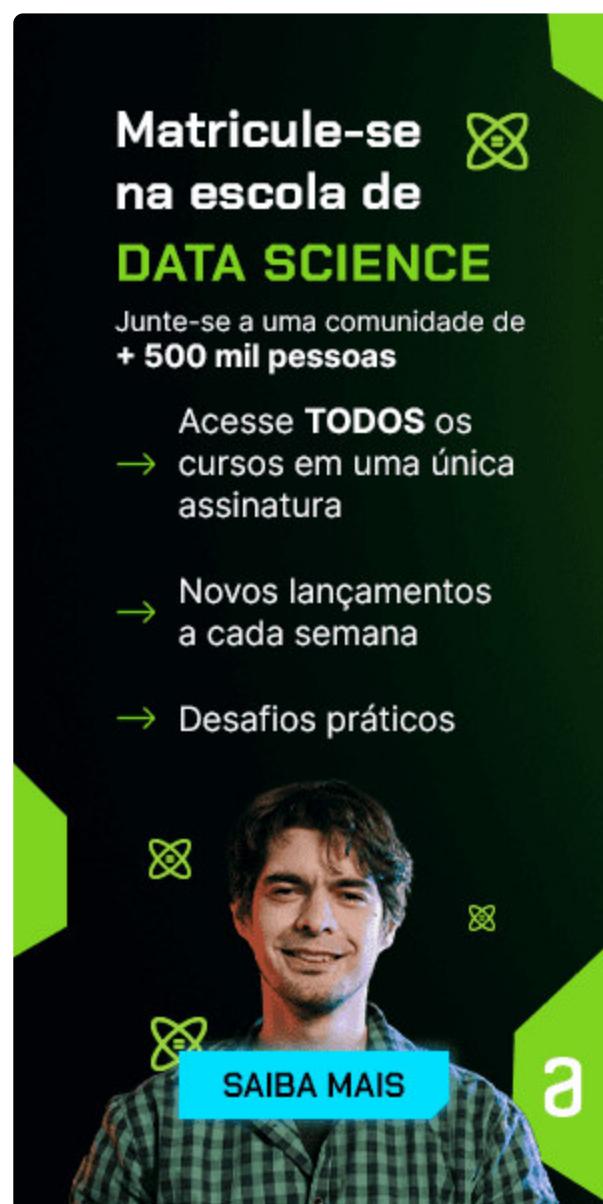
30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Para saber mais sobre o tema, você pode conferir o artigo de [histograma] (<https://www.alura.com.br/artigos/o-que-e-um-histograma>). E também checar a [documentação](#).



30%

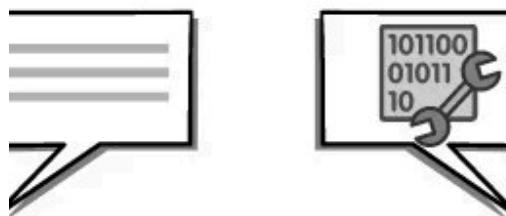
[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Daniel é instrutor na escola de Dados e professor de Matemática, Física, Química e Inglês. Tem verdadeira paixão em aprender coisas e assuntos novos, e transmitir seus conhecimentos.

## Leia também



[Aplicações de NLP no mercado de trabalho](#)

[Otimização de hiperparâmetros](#)

[Machine  
Learning  
Spa](#)



[Veja outros artigos sobre  
Data Science](#)

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Receba conteúdos, dicas, notícias, inovações e tendências sobre o mercado tech diretamente na sua caixa de entrada.

[bins.br@gmail.com](mailto:bins.br@gmail.com)

**ENVIAR**

## Nossas redes e apps



## Institucional

Sobre nós

Trabalhe na Alura

Para Empresas

Para Sua Escola

Política de Privacidade

## A Alura

Como Funciona

Formações

Plataforma

Depoimentos

Instrutores(as)

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

## Conteúdos

Alura Cases

Imersões

Artigos

Podcasts

Artigos de educação  
corporativa

Imersão Dados com Python II

## Fale Conosco

Email e telefone

Perguntas frequentes

## Novidades e Lançamentos

bins.br@gmail.com

[ENVIAR](#)

## TRILHAS POR CARREIRA

### Carreiras de IA

Engenharia de IA

Especialista em IA

30%

[Compre com desconto](#)

TÁ ACABANDO!

02 : 13 : 38 : 54  
DIAS HORAS MIN SEG

Platform Engineering

SRE (Site Reliability Engineering)

#### Carreiras de UX & UI

UI Design

UX Design

#### Carreiras de Mobile & Front-End

Desenvolvimento Mobile com Flutter

Desenvolvimento Front-End React

#### Carreiras de Back-End

Desenvolvimento Back-End PHP

Desenvolvimento Back-End .NET

Desenvolvimento Back-End Python

Desenvolvimento Back-End Java

Desenvolvimento Back-End Node.js

#### Carreiras de Negócios

Liderança

Recursos Humanos (RH)

Social Media Marketing

Growth Marketing

## CURSOS UNIVERSITÁRIOS FIAP

Graduação | Pós-graduação | MBA