

Histograma

Transcrição

[0:00] Legal, moçada. A gente já aprendeu a criar tabelas de distribuições de frequências, agora vou mostrar para vocês uma forma de representar isso graficamente com o uso de histogramas.

[0:09] Todo mundo já deve ter ouvido falar dos histogramas.

[0:12] Eu vou introduzir agora uma biblioteca nova no nosso projeto, que é a Seaborn.

[0:16] Então, a gente precisa fazer um `import seaborn as sns`, o `sns` é o padrão, tudo mundo usa como um apelido para biblioteca, um nome menor.

[0:26] A gente vai encontrar isso na própria documentação. Aqui está o link da documentação. Deixa eu rodar aqui o `import`.

[0:31] E aqui está a documentação. É uma biblioteca de visualização de dados estatísticos, especializada nisso, baseada no Matplotlib.

[0:39] O nosso objetivo não é entrar muito a fundo nessa biblioteca.

[0:42] Aqui na Alura a gente tem cursos que entram a fundo nessa biblioteca. Então, se você tiver algum interesse é só você procurar os cursos.

[0:49] Aqui eu vou mostrar para vocês como criar o histograma, também o box plot, mas a gente vai ficar basicamente nisso, não vamos nos aprofundar muito.

[0:58] Eu vou mostrar o que eu vou usar, você vai entender e tendo interesse você procura a biblioteca e o curso também.

[1:03] Então, vamos lá. Primeiro, eu estou usando aqui como nome da variável, eu vou criar um objeto do Seaborn, que é o gráfico, e vou atribuir ele a essa variável que eu estou chamando de `ax`.

[1:12] Estou chamando de `ax` porque na documentação você vai encontrar muito esse tipo de nome, então fica mais fácil você fazer o link para conseguir entender melhor o que eu estou fazendo aqui, comparando com a documentação.

[1:25] Então, vamos lá, `ax` é igual, eu vou chamar a biblioteca `sns` ponto, para criar um histograma eu tenho a função `distplot`.

[1:35] Abre e fecha e passo para ela o dado que eu quero plotar. Então, no caso aqui eu vou plotar altura, já deixei configurado aqui para altura.

[1:44] Eu vou, inicialmente, suprimir aqui, é KDE, eu vou retirar o KDE, que é uma função de densidade que ele traça junto com o histograma.

[1:57] No próximo eu deixo o KDE, só para você ver como é que fica a representação.

[2:04] Criei aqui esse cara, eu vou pegar agora o `ax` e vou começar a configurar alguns parâmetros nele.

[2:10] Eu passo `ax` ponto `figure` ponto `set size` e vou configurar o tamanho que eu quero que apareça o gráfico aqui no meu notebook.

[2:21] Eu estou dando 12 de largura com seis de altura. Você configura o que ficar melhor aí no seu, na sua tela.

[2:27] Aqui também eu estou configurando um título, estou chamando de Distribuição de Frequências, variável Altura, o tamanho da fonte 18.

[2:35] Aqui eu configuro qual vai ser o label do eixo x, o eixo debaixo, falando que vai ser metros, uma fonte 14 e aqui eu peço para mostrar o gráfico.

[2:44] Então vai lá, ele vai mostrar o meu histograma. Está aqui, é um gráfico de barras, que estão juntinhas, onde a área dessa barra é proporcional à frequência da classe, que a gente criou as tabelas, lembra? Basicamente é isso.

[2:59] Eu vou plotar aqui embaixo, tem a mesma configuração, só que está com os títulos um pouco diferentes.

[3:07] Eu vou deixar o KDE. Como o padrão é true, então é só tirar.

[3:13] Distribuição de Frequências, Altura, vou botar aqui KDE, que é a função de densidade. Aqui novamente a gente pode deixar só metros, eu botei Altura ali, botar M, mas são metros.

[3:31] Ele ficou um pouco maior aqui, eu deixei uma configuração maior, vamos botar o 12 que é igual o de cima, e ele traçou a função de densidade.

[3:40] Para representar bem esse gráfico aqui a gente pode, depois, fazer umas comparações, eu vou mostrar para vocês essa coisa.

[3:46] Esse aqui, como ele foi criado a partir de uma distribuição normal, ele tem formato normal, forma de sino, bem simétrico.

[3:59] Essa variável foi feita de propósito, ela foi criada por mim. Por isso eu quis mostrar ela aqui, porque ela fica mais bonitinha no histograma e também a função de densidade.

[4:10] Uma outra forma de fazer isso, já não usando o Seaborn, a gente pode usar o Pandas, só que a gente tem menos recursos que o Seaborn.

[4:18] Mas vamos lá, vamos mostrar aqui. A mesma coisa, dados ponto altura ponto hist, você consegue traçar um histograma já desse jeito aqui. Ficou meio tosco porque aqui ele não está usando, como default acho que ele vem 10 barras, então vou bins, você aumenta o número de classes, você cria mais classes para fazer o seu histograma.

[4:43] Eu posso também configurar aqui no Pandas um figsize também. Lá a gente botou 12, seis, então a gente pode por 12, seis aqui também, ele vai aumentar a nossa figura.

[4:56] Ficou bem parecido lá, parece até que ele está bem igual. Deve ter 50 barras aqui também.

[5:02] Ele faz isso aqui, só que com menos recurso. Ali, entrando no ajuste você vai ver que consegue fazer muito mais, tem muito mais recurso a biblioteca do que usando o Pandas. O Pandas é um quebra galho nesse caso gráfico aqui.

[5:16] Um outro cara que a gente criou, eu vou criar mais um aqui embaixo, para trazer ele aqui. Deixa eu ver se eu acho ele aqui. Lá no começo a gente criou aquela tabelinha. Aqui, as classes A, B, C, D e E, e eu chamei esse cara aqui de dist freq. quantitativas personalizadas. É isso que eu não estava lembrando.

[5:38] Então, vamos lá. Copiei esse cara aqui. Vamos tentar fazer uma representação gráfica dele também, porque é uma classe que a gente criou, é uma tabela de frequência que a gente criou personalizada.

[5:49] Se você usar esses recursos aqui, você não vai conseguir criar esse cara.

[5:53] Então, vou mostrar para você, como criar um gráfico de barras também, que vai ficar bem parecido com aquele histograma. Não é um histograma, mas você consegue fazer uma representação naquele mesmo formato.

[6:04] Aqui, como eu vou querer a variável frequência, vou copiar ela aqui, vou passar para cá. Eu quero só a coluna de frequência. Podia fazer com percentual, vai ficar o mesmo tamanho, ela tem o mesmo comportamento, mesma distribuição.

[6:20] Ponto plot ponto bar. Eu quero plot, um gráfico de barra. Abre aqui.

[6:26] Para ele ficar juntinho, na forma que é um histograma, eu vou passar esse parâmetro, que é a largura de um, ele vai ficar com as barras juntinhas, uma do lado da outra.

[6:38] Eu posso configurar também a cor, vou botar isso aqui só de macete para vocês, para vocês conseguirem brincar, vou deixar azul.

[6:45] Posso também, para ficar bem próximo desse formato aqui, talvez, um alpha. Alpha é uma transparência. E o figsize a gente pode copiar daqui de cima, esse aqui, igualzinho esse aqui.

[7:00] Vai ficar uma coisa parecida com esse aqui de cima, por isso que eu botei um Alpha.

[7:07] Tem vírgula demais aqui, vamos tirar uma. Vamos rodar e ver o que ele roda para a gente.

[7:12] Olha lá. Ele fez uma representação gráfica de barra, ficou até bem parecida aqui na cor, de um histograma. Seria um histograma da nossa distribuição de frequência que a gente criou aqui.

[7:24] Uma forma de criar, a gente já criou a tabela, você começa a criar um gráfico de barras, com os resultados daquela tabela.

[7:29] Ou então, se você não criou essa questão de frequência, aí você não quer, você consegue criar utilizando esses métodos aqui, ele já cria para a gente as classes e cria o gráfico automaticamente.

[7:41] Era isso que eu queria mostrar para vocês.

[7:42] No próximo a gente vai mudar sessão, agora a gente vai falar de medidas de tendência central. Tá bom?

[7:48] Vejo você no próximo vídeo.