**Tópico 2 – Gráficos de densidade com seaborn**

**OBJETIVOS**

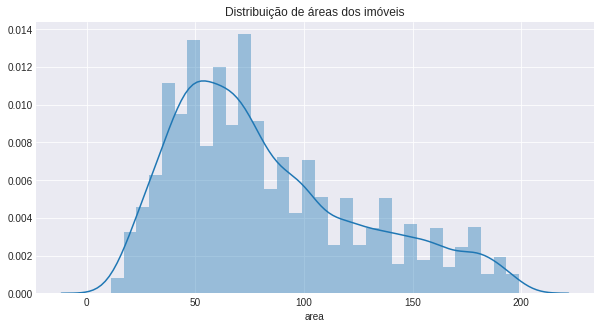
* Compreender como funciona o gráfico de densidade;
* Criar e modificar gráficos de densidade com a biblioteca seaborn.

O gráfico de densidade possui a mesma função do histograma, mas representa a distribuição como uma curva contínua ao invés de bins. A altura da curva no eixo vertical também não é uma contagem de elementos com aquele valor, como no histograma, e sim a estimativa da probabilidade de um elemento possuir aquele valor. Ou seja, é uma interpretação probabilística dos dados. Como são representados por uma curva, gráficos de densidade geram visualizações mais suaves que histogramas e podem facilitar a compreensão da distribuição.

Neste tópico, você deverá aprender a interpretar, criar e modificar gráficos de densidade, utilizando a biblioteca seaborn. Sigamos!

* Criar gráficos de densidade com a biblioteca seaborn é bastante direto, basta utilizar a função distplot(). Essa função é bastante útil, pois funciona de forma dois em um, mostrando, ao mesmo tempo, tanto a curva de probabilidade como os bins do histograma. Confira, no exemplo a seguir, como utilizar essa função para visualizar a distribuição de áreas dos dados imobiliários:
* 1 plt.figure(figsize=(10, 5))
* 2 plt.title('Distribuição de áreas dos imóveis')
* 3 sns.distplot(df['area'])
* 4 plt.show()

Download do código sem numeração no link a seguir: [Download código](https://leadfortaleza.com.br/ead2pcd/conteudo/tmp/myopenolat_1_102915070706483/aula/code/codet2p2s11.txt)

Figura 3 – Gráfico de densidade com histograma

Note que a única diferença, entre este exemplo e os exemplos anteriores, foi a chamada da função distplot() do seaborn, passando como argumento a coluna com as áreas de forma muito similar ao feito com a função hist() do matplotlib. A figura resultante desse código mostra uma curva contínua em azul escura, que corresponde à curva de probabilidade de ocorrência dos dados, e um histograma em azul claro ao fundo. Observe que os valores da curva mostrados no eixo vertical, como foi explicado, não são quantidades, e sim probabilidades, indo de 0 (0% de chance de ocorrência) a 1 (100% de probabilidade de ocorrência). Logo, pela curva de probabilidade, as regiões mais prováveis estão entre 50 e 70 m². Você percebe como o resultado ficou melhor de compreender?

Vamos seguir e conhecer mais sobre o gráfico de densidade, mas, antes disso, confira uma informação importante sobre a curva de probabilidade.

#### Ícone Saiba Mais

Essa curva de probabilidade é estimada através de um método conhecido como **Estimativa de Densidade de** **Kernel** (em inglês, a sigla KDE é utilizada).

A curva de probabilidade é estimada a partir dos dados reais, que são representados pelo histograma. A função distplot() mostra as duas representações em uma mesma imagem. Porém, você pode querer uma figura mais limpa, mostrando apenas a distribuição de forma contínua e suave. Se você quiser apenas o gráfico de densidade, pode-se usar a função kdeplot(). Confira a seguir:

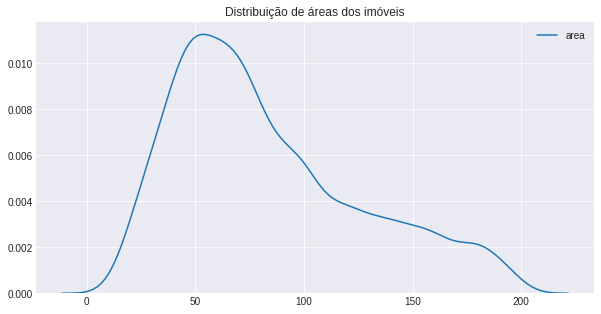
1 plt.figure(figsize=(10, 5))

2 plt.title('Distribuição de áreas dos imóveis')

3 sns.kdeplot(df['area'], shade=True)

4 plt.show()

Download do código sem numeração no link a seguir: [Download código](https://leadfortaleza.com.br/ead2pcd/conteudo/tmp/myopenolat_1_102915070706483/aula/code/codet2p3s12.txt)

Figura 4 – Gráfico de densidade sem histograma

Na linha 3, é chamada a função kdeplot(), passando como argumentos os dados e o valor True para o parâmetro shade, indicando que a área da curva será preenchida, o que causa um efeito visual que auxilia a percepção. Se não for usado esse parâmetro, só a curva será visível. Para uma lista completa dos parâmetros das funções distplot() e kdeplot(), consulte a documentação oficial disponível nos links: <https://seaborn.pydata.org/generated/seaborn.distplot.html#seaborn.distplot> e [https://seaborn.pydata.org/generated/seaborn.](https://seaborn.pydata.org/generated/seaborn.kdeplot.html#seaborn.kdeplot)[kde](https://leadfortaleza.com.br/ead/glossary/KDE)plot.html#seaborn.[kde](https://leadfortaleza.com.br/ead/glossary/KDE)plot.

Perceba que uma vantagem do gráfico de densidade é que ele fornece uma visualização da distribuição dos dados, que não depende da quantidade de bins, como é o caso do histograma.

Caro estudante, você chegou ao fim desta aula. Agora, com certeza, você está equipado com as ferramentas necessárias para visualizar efetivamente a distribuição de um [atributo](https://leadfortaleza.com.br/ead/glossary/Atributo_Visualizacao_de_dados) quantitativo. Esse é o primeiro passo para compreender esse tipo de dado, e é uma ferramenta de análise muito reveladora.

Como sempre, é importante exercitar o que foi aprendido e experimentar as opções disponíveis. Por isso, tente visualizar os outros atributos desse conjunto de dados, utilizando essas técnicas, variando os parâmetros mostrados. Procure também consultar a documentação das bibliotecas, para conhecer e experimentar as opções de estilos.

Espero que você tenha aprendido bastante! Até a próxima.

### Referências

* Matplotlib. <https://matplotlib.org/index.html>. Acessado pela última vez em Maio de 2020.
* Seaborn. <https://seaborn.pydata.org/tutorial.html>. Acessado pela última vez em Maio de 2020.
* WILKE, Claus O. **Fundamentals of Data Visualization**. O’Reilly, 2019.