

Desvendando Histórias com Dados: Um Guia Visual do Seaborn

Imagine que você é um detetive de dados. Seu trabalho é interrogar um conjunto de dados até que ele confesse seus segredos. O Seaborn é o seu kit de ferramentas de interrogatório visual, onde cada gráfico é uma lente de aumento diferente, projetada para revelar um tipo específico de verdade.

Nossos Dados de Estudo: O Restaurante

Para nossa investigação, vamos usar os dados de um restaurante, contidos no famoso dataset "tips" (gorjetas). Ele nos diz o valor total da conta, a gorjeta, o sexo do cliente, o dia da semana, etc.

Python

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Carregando nossos dados
df_gorjetas = sns.load_dataset('tips')
```

Pergunta 1: Como os dias da semana se comparam?

A Ferramenta: O Raio-X dos Dados (`sns.boxplot`)

Quando sua pergunta envolve **comparar a distribuição** de um número (como o valor da conta) entre diferentes grupos (como os dias da semana), você precisa de um raio-x. O boxplot é essa ferramenta. Ele mostra o "esqueleto" dos dados para cada grupo.

Lendo o Gráfico:

Um boxplot é rico em detalhes. Ele nos mostra:

- **A linha no meio da caixa (Mediana):** O valor "típico". 50% dos valores são maiores e 50% são menores que essa linha. É o coração do grupo.
- **A caixa em si:** Onde se concentra a maioria (os 50% centrais) dos seus dados. Uma caixa alta significa muita variação; uma caixa curta significa que os valores são muito parecidos.
- **Os "bigodes" (linhas):** A faixa de valores "normal" ou esperada.
- **Os pontos fora:** Valores atípicos (outliers), que são muito diferentes da maioria.

Python

```
# O Código
sns.boxplot(data=df_gorjetas, x='day', y='total_bill')
plt.title('Raio-X do Valor da Conta por Dia')
```

```
plt.show()
```

A História que ele Conta:

Ao olhar para o raio-x, vemos imediatamente que os fins de semana (Sábado e Domingo) têm o "coração" (mediana) mais alto – as pessoas gastam mais. Suas caixas também são mais altas, mostrando maior variedade de gastos, e eles possuem os "outliers", aquelas contas excepcionalmente altas.

Pergunta 2: Qual o "ponto de encontro" dos clientes?

A Ferramenta: O Horizonte dos Dados (`sns.histplot`)

Quando sua pergunta é sobre a **frequência e distribuição** de uma única variável numérica, você quer ver onde os dados se aglomeram. O histograma desenha o "horizonte" dos seus dados, onde cada barra é um prédio, e a altura do prédio nos diz quantas contas caem naquela faixa de valor.

Lendo o Gráfico:

Onde os "prédios" são mais altos, é onde a maioria dos seus dados vive. A forma geral do horizonte (se é simétrico, inclinado, etc.) revela a natureza da sua distribuição.

Python

```
# O Código
sns.histplot(data=df_gorjetas, x='total_bill', kde=True, bins=20)
plt.title('Onde se Concentram os Valores das Contas?')
plt.show()
```

A História que ele Conta:

O horizonte nos mostra que o "bairro" mais populoso é o de contas entre \$10 e \$20. A cidade se espalha para a direita, com prédios cada vez menores, indicando que contas muito altas são raras. A linha suave (kde) confirma essa tendência.

Pergunta 3: Uma coisa leva à outra?

A Ferramenta: O Detetive de Relações (`sns.lmplot`)

Quando você suspeita que **duas variáveis numéricas** podem estar conectadas (uma influencia a outra), você precisa de um gráfico de dispersão. Ele coloca cada um dos seus dados como um ponto em um mapa para ver se eles formam um padrão.

Lendo o Gráfico:

Cada ponto é uma observação. O mais importante é o padrão geral: os pontos sobem juntos? Descendo juntos? Ou estão espalhados como estrelas aleatórias no céu? O Implot já nos ajuda desenhando uma linha de tendência.

Python

```
# O Código
sns.lmplot(data=df_gorjetas, x='total_bill', y='tip', height=6)
plt.title('Contas Maiores Geram Gorjetas Maiores?')
plt.show()
```

A História que ele Conta:

A constelação de pontos forma um claro caminho ascendente da esquerda para a direita. A linha de tendência confirma: sim, existe uma relação positiva. Contas mais altas estão claramente associadas a gorjetas mais altas.

Pergunta 4: Qual é a conexão mais forte?

A Ferramenta: O Mapa de Calor (`sns.heatmap`)

Quando você tem várias variáveis numéricas e quer ter uma visão geral e rápida de **todas as relações** entre elas de uma só vez, o mapa de calor é a ferramenta perfeita. Ele transforma uma tabela de correlações em um mapa colorido.

Lendo o Gráfico:

É como um jogo de "quente ou frio".

- **Cores quentes (vermelhas):** Indicam que as variáveis "andam juntas" (correlação positiva forte).
- **Cores frias (azuis):** Indicam que as variáveis andam em direções opostas (correlação negativa forte).
- **Cores neutras:** Indicam pouca ou nenhuma relação.

Python

```
# O Código
numeric_df = df_gorjetas.select_dtypes(include=['number'])
correlation_matrix = numeric_df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('O Mapa das Relações Numéricas')
plt.show()
```

A História que ele Conta:

O mapa nos mostra instantaneamente o "ponto mais quente": a célula que cruza `total_bill` e `tip`, com um valor de 0.68. Isso nos diz que, de todas as relações possíveis, a mais forte é a do valor da conta com a gorjeta.