

Bases de dados e tipos de variáveis | Countplot

Programação para Advogados – 2024.2

José Luiz Nunes e Lucas Thevenard

Roteiro de Aula

- Bases de dados e tipos de variáveis
- A sintaxe do countplot
- Aula prática: vamos programar!
 - Preparação: configuração do VS Code
 - Nosso primeiro gráfico
- Exercício

Bases de dados e tipos de variáveis

- O que é uma base de dados?
 - Dados estruturados e não estruturados
 - Dados tabulares (tabela)
 - Como organizar dados tabulares?
 - Conceito de **tidy data**

Tidy Data

- **Conceito:** alinhamento entre a estrutura dos dados e sua semântica (significado).
- **Regra centrais:**
 - Linhas são observações
 - Colunas são variáveis
 - Apenas uma unidade de observação (ou de análise) por tabela

Tidy Data

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	128042583

variables

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	216766	128042583

observations

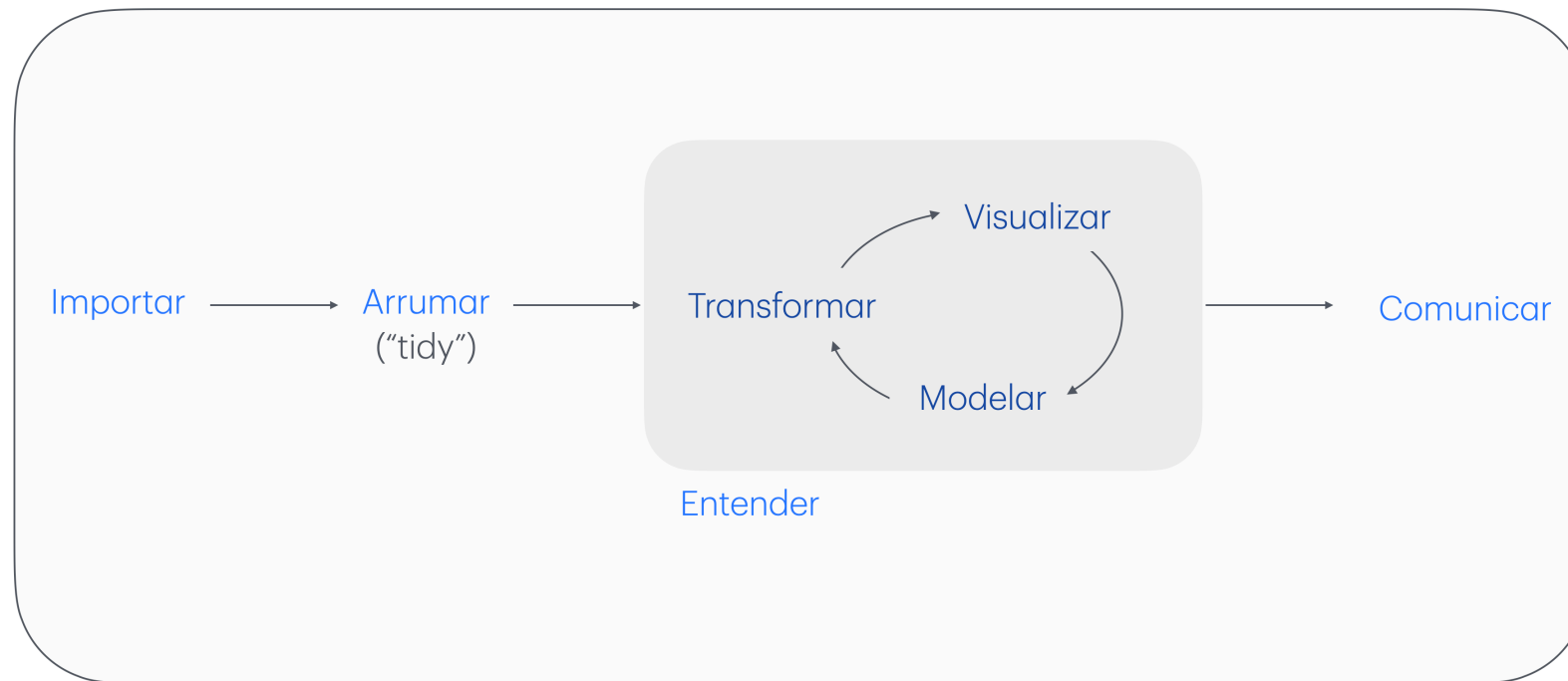
country	year	cases	population
Afghanistan	99	745	19987071
Afghanistan	00	2666	20595360
Brazil	99	37737	172006362
Brazil	00	80488	174504898
China	99	212258	1272915272
China	00	216766	128042583

values

Tidy Data

- Vamos construir um exemplo com uma tabela de alunos da sala.
 - **Unidade de observação:** Alunos
 - **Variáveis:** Nome, Matrícula, Altura, Ano de Nascimento, Signo, Cor Favorita
- Como construir uma tabela de jurisprudência no escritório?

O Ciclo da Ciência de Dados



Vamos para o código

Bibliotecas para dados e visualizações

```
# Importando as bibliotecas necessárias
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[1] ✓ 1.4s Python

Carregar os dados da aula

```
# Carregando os dados  
cartoes = pd.read_csv("https://bit.ly/dados_cartoes")  
cartoes
```

[2]



0.8s

Python

Dados de pagamentos dos cartões presidenciais

	data_pgto	ano	mes	forma_pagamento	valor	tipo_despesa	nome_fornecedor	cpf_servidor	mandato	posicao_mandato
0	2003-01-02	2003	1	Débito	1696.90	LOCACAO DE MEIOS DE TRANSPORTE	JW TRANSPORTADORA TURÍSTICA	***.010.008-**	Lula (1º Mandato)	1
1	2003-01-02	2003	1	Débito	1336.60	LOCACAO DE MEIOS DE TRANSPORTE	JW TRANSPORTADORA TURÍSTICA	***.010.008-**	Lula (1º Mandato)	1
2	2003-01-02	2003	1	Débito	1396.43	LOCACAO DE MEIOS DE TRANSPORTE	JW TRANSPORTADORA TURÍSTICA	***.868.251-**	Lula (1º Mandato)	1
3	2003-01-03	2003	1	Débito	105.00	MATERIAL P/MANUTENÇÃO DE BENS MÓVEIS	SEBASTI O ALVES DE SOUSA - ME	***.262.328-**	Lula (1º Mandato)	1
4	2003-01-03	2003	1	Débito	200.00	MATERIAL P/MANUTENÇÃO DE BENS MÓVEIS	JOÃO LEITE NETO - ME - JLN ELETRONICOS	***.262.328-**	Lula (1º Mandato)	1
...
114958	2022-12-29	2022	12	Crédito	210.80	GENEROS DE ALIMENTAÇÃO	ZAIANNA CRISTINA COMERCIO DE ALIMENTOS LTDA	***.135.521-**	Jair Bolsonaro	7
114959	2022-12-29	2022	12	Crédito	536.13	GENEROS DE ALIMENTAÇÃO	CASA DO CHOCOLATE PRODUTOS ALIMENTICIOS	***.135.521-**	Jair Bolsonaro	7
114960	2022-12-29	2022	12	Crédito	673.82	GENEROS DE ALIMENTAÇÃO	MERCADINHO LA PALMA LTDA	***.135.521-**	Jair Bolsonaro	7
114961	2022-12-29	2022	12	Crédito	2147.53	GENEROS DE ALIMENTAÇÃO	COMPANHIA BRASILEIRA DE DISTRIBUIÇÃO	***.135.521-**	Jair Bolsonaro	7
114962	2022-12-31	2022	12	Crédito	36327.36	SERV.DE APOIO ADMIN.,TECNICO E OPERACIONAL	SIGNATURE FLIGHT SUPPORT	***.725.557-**	Jair Bolsonaro	7

114963 rows x 10 columns

Dicionário dos dados

- **data_pgto**: a data em que o pagamento foi realizado, no formato YYYY-MM-DD.
- **ano**: O ano em que a transação foi realizada, como número inteiro.
- **mes**: O mês em que a transação foi realizada, como número inteiro (1 = Janeiro, 2 = Fevereiro, etc).
- **forma_pagamento**: A forma de pagamento utilizada, que pode ser "Crédito" ou "Débito".
- **valor**: O valor do pagamento realizado, em reais brasileiros (R\$), como número decimal.

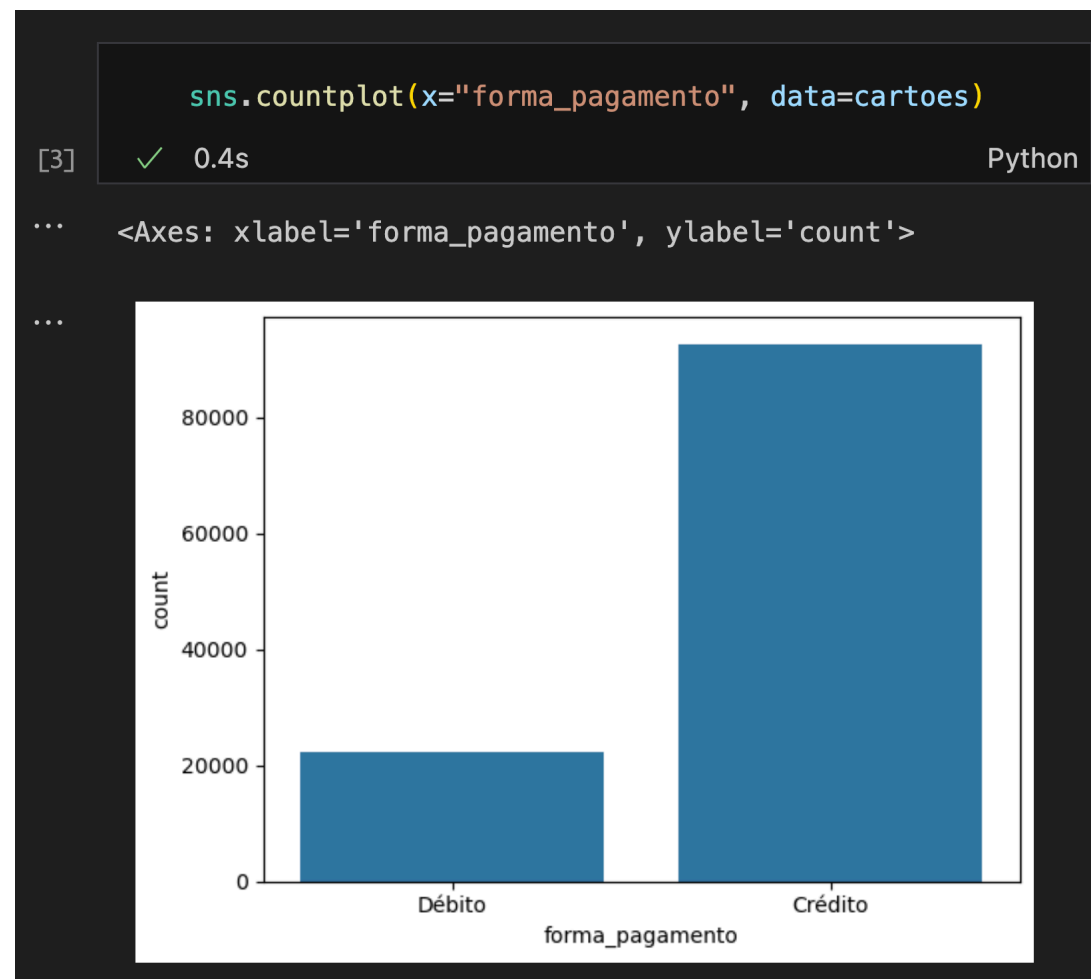
Dicionário dos dados

- **tipo_despesa**: A classificação orçamentária da despesa detalhada.
- **nome_fornecedor**: A razão social ou nome do favorecido.
- **cpf_servidor**: Os 6 últimos dígitos, antes do código verificador, do Cadastro de Pessoa Física (CPF) do agente suprido (que realizou o pagamento).
- **mandato**: Identifica o mandato no qual aconteceu o pagamento.
- **posicao_mandato**: Variável numérica que representa a ordem do mandato no qual aconteceu o pagamento.

Vamos aprender a fazer o nosso primeiro gráfico

Primeiro Gráfico

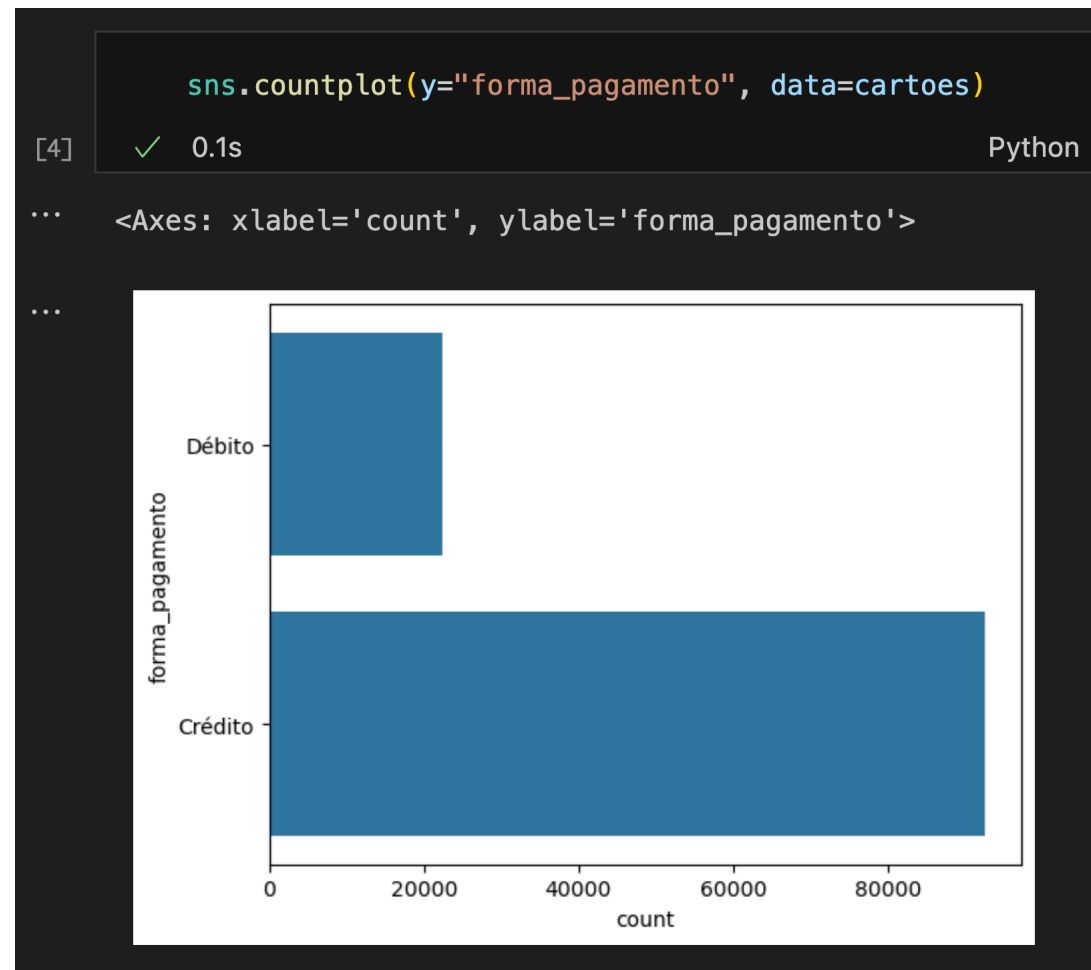
- Contagem do uso das duas formas de pagamento: Crédito vs. Débito.
- Usamos a função `sns.countplot()`
 - `data=cartoes` : definimos qual DataFrame (dados) usar.
 - `x="forma_pagamento"` : qual coluna/variável dos dados queremos plotar, em qual eixo.



Como poderíamos inverter os eixos do gráfico?

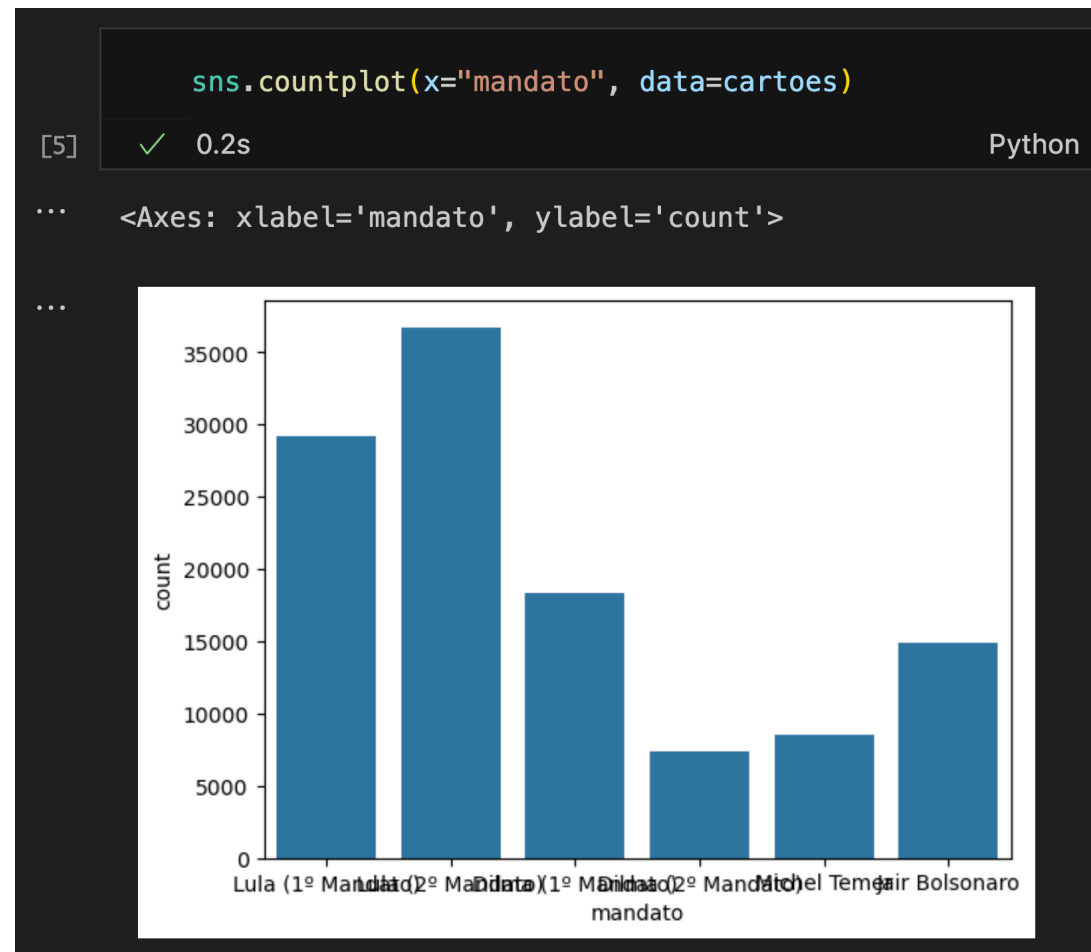
Eixo invertido

- Invertemos os eixos definindo a função com a variável de contagem no eixo y.
 - Para isso usamos o argumento `y="forma_pagamento"` na função `sns.countplot()`.
 - Em que casos faz sentido fazer isso?



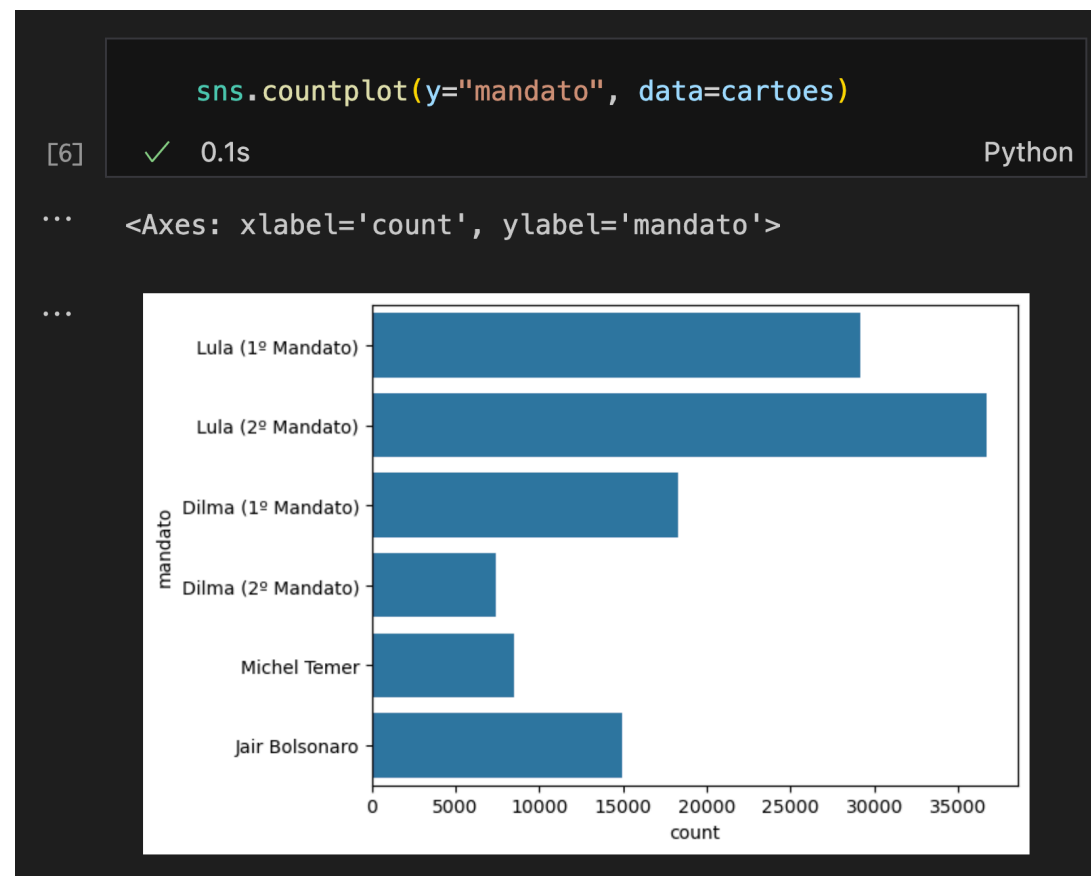
Problema dos nomes sobrepostos no eixo x

- Um problema comum é termos a sobreposição dos nomes. Isso ocorre quando temos
 - **Muitas categorias.**
 - **Nomes longos** das categorias.
- Veja ao lado o que ocorre quando tentamos plotar os mandatos no eixo x.



Invertendo os eixos eliminamos a sobreposição

- Uma forma simples de eliminar o problema das sobreposições é inverter os eixos.
- Agora responda:
 - **Em qual mandato houve o maior gasto no cartão presidencial?**



Cruzando duas variáveis

- Podemos usar o argumento `hue` para cruzar duas variáveis.
- Veja ao lado o número de transações de cada presidente por forma de pagamento.

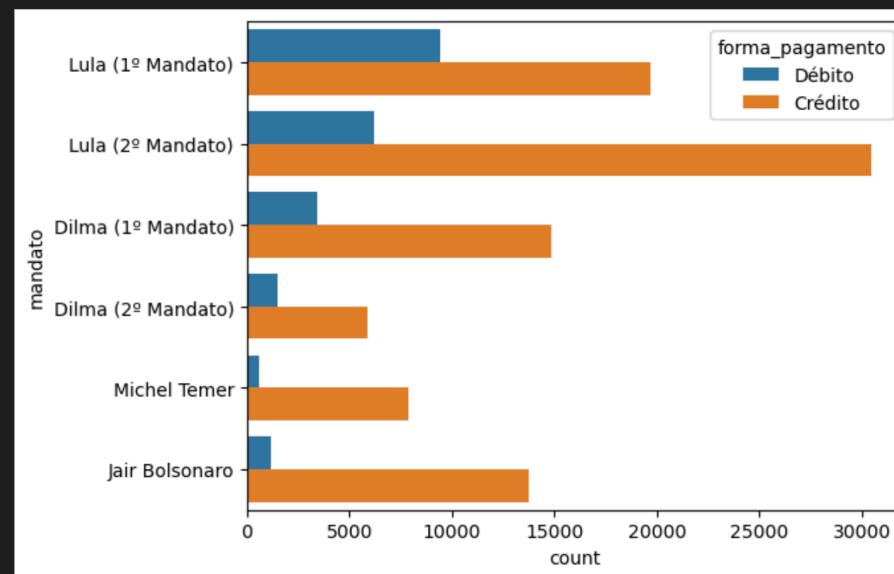
```
sns.countplot(  
    y="mandato",  
    hue="forma_pagamento",  
    data=cartoes  
)
```

[7]

✓ 0.2s

Python

<Axes: xlabel='count', ylabel='mandato'>



Ainda podemos melhorar esse gráfico...

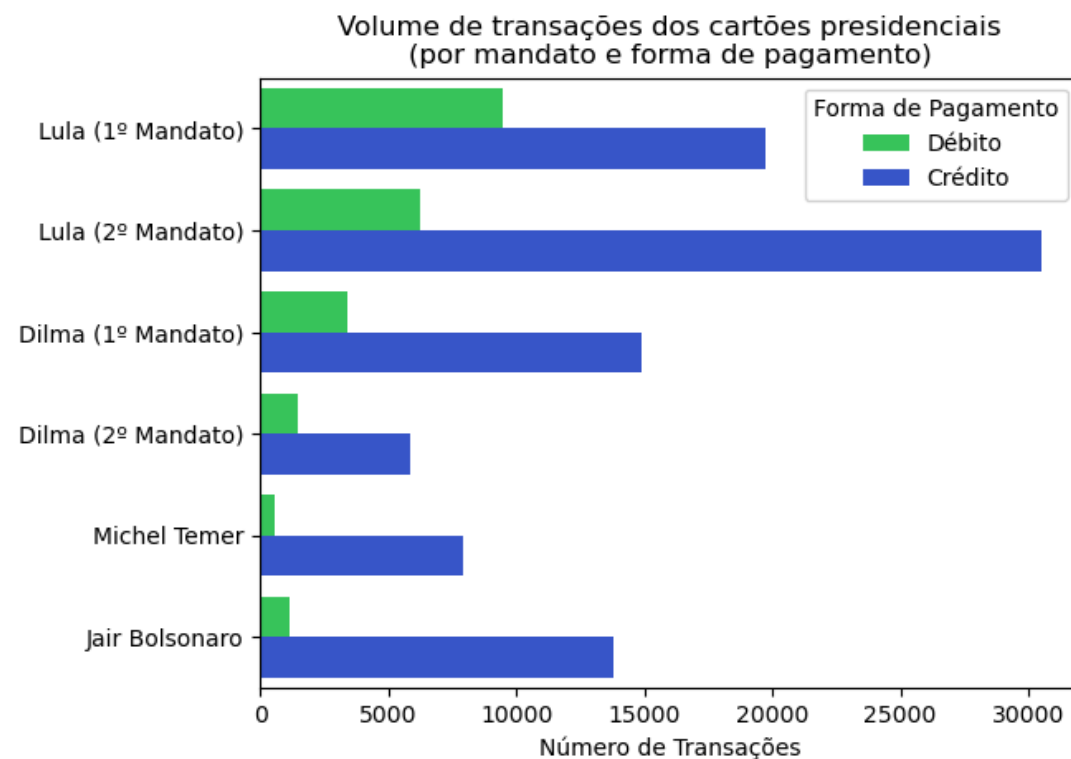
```
fig, ax = plt.subplots()

sns.countplot(
    y="mandato",
    hue="forma_pagamento",
    palette={"Crédito": "#1e4ce3",
            "Débito": "#1fdb51"},
    data=cartoes,
    ax=ax
)

ax.set_title("Volume de transações dos cartões presidenciais\n(por mandato e forma de pagamento)")
ax.set_xlabel("Número de Transações")
ax.set_ylabel("")
ax.legend(title="Forma de Pagamento")
plt.show()
```

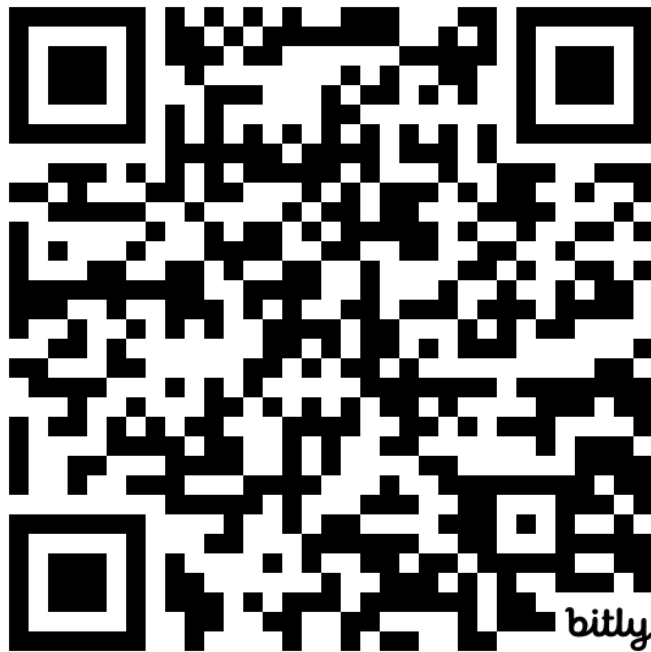
[8] ✓ 0.2s

Python



Agora mãos à obra!

Config do Visual Studio Code



- **Guia de Configuração do VS CODE**
 - Disponível no EClass
 - Link de acesso:
bit.ly/config_vscode