

✓ Avaliação 1º Bimestre

✓ Parte 1 – Questões Objetivas e Teóricas

Questões estão disponíveis no AVA (5 pontos)

✓ 🔐 Parte 2 – Prática Google Colab

Informe seu nome e RA

```
usuario = input("Digite seu nome: ")
print("Nome do aluno:", usuario)
RA = input("Digite seu RA: ")
print("RA do aluno:", RA)

Digitando: Digite seu nome: Rafael de Lima Rodrigues
Nome do aluno: Rafael de Lima Rodrigues
Digite seu RA: 2024102194
RA do aluno: 2024102194
```

1. (0,5 pontos) Carregue o dataset JSON Netflix e exiba as 5 primeiras linhas.

- JSON está disponível em:
<https://raw.githubusercontent.com/prust/wikipedia-movie-data/master/movies.json>

```
import pandas as pd
url = 'https://raw.githubusercontent.com/prust/wikipedia-movie-data/master/movies.json'
df = pd.read_json(url)
df.head()
```

		title	year	cast	genres	href	extract	thumbnail	thumbnail_width	thumbnail_height	
0		After Dark in Central Park	1900	[]	[]	None	NaN	NaN	NaN	NaN	NaN
1		Boarding School Girls' Pajama Parade	1900	[]	[]	None	NaN	NaN	NaN	NaN	NaN
2		Buffalo Bill's Wild West Parad	1900	[]	[]	None	NaN	NaN	NaN	NaN	NaN

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

2. (1 ponto) Para cada coluna, apresentar a contagem de não nulos e tipo do objeto .

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36273 entries, 0 to 36272
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   title            36273 non-null   object 
 1   year             36273 non-null   int64  
 2   cast              36273 non-null   object 
 3   genres            36273 non-null   object 
 4   href              34540 non-null   object 
 5   extract           34532 non-null   object 
 6   thumbnail         30368 non-null   object 
 7   thumbnail_width   30368 non-null   float64 
 8   thumbnail_height  30368 non-null   float64 
dtypes: float64(2), int64(1), object(6)
memory usage: 2.5+ MB
```

3. (1 ponto) Apresente uma tabela com a quantidade de títulos por gênero.

```
from collections import Counter
```

```
genre_counter = Counter()
for genres in df['genres']:
    genre_counter[genres] += 1
```

```
for genre in genres:
    genre_counter.update(genres)

genre_df = pd.DataFrame.from_dict(genre_counter, orient='index', columns=['Quantidade'])
genre_df = genre_df.sort_values(by='Quantidade', ascending=False)
genre_df
```

	Quantidade	
Drama	14062	grid
Comedy	10498	up-down
Silent	7131	edit
Western	4436	
Romance	2857	
Crime	2612	
Action	2241	
Thriller	1998	
Musical	1785	
Adventure	1761	
Horror	1758	
War	1694	
Science Fiction	1169	
Mystery	1163	
Noir	1120	
Short	1023	
Animated	897	
Fantasy	777	
Biography	753	
Historical	567	
Documentary	527	
Sports	459	
Family	435	
Independent	319	
Supernatural	313	
Erotic	240	
Slasher	231	
Superhero	221	
Spy	214	
Teen	205	
Satire	119	
Suspense	112	
Political	110	
Martial Arts	105	
Disaster	85	
Legal	56	
Performance	42	
Sport	39	
Live Action	35	
Dance	34	
Found Footage	25	

Next steps: [Generate code with genre_df](#) [View recommended plots](#) [New interactive sheet](#)

4. (1 ponto) Crie uma coluna Década que classifica os anos de lançamento (year), agrupando em faixas:

- "70s" : Entre 1970 e 1980
- "80s" : Entre 1980 e 1990
- "90s" : Entre 1990 e 2000
- "2000s" : Entre 2000 e 2010
- "2010s" : Entre 2010 e 2020
- "2020s" : Entre 2020 e 2030

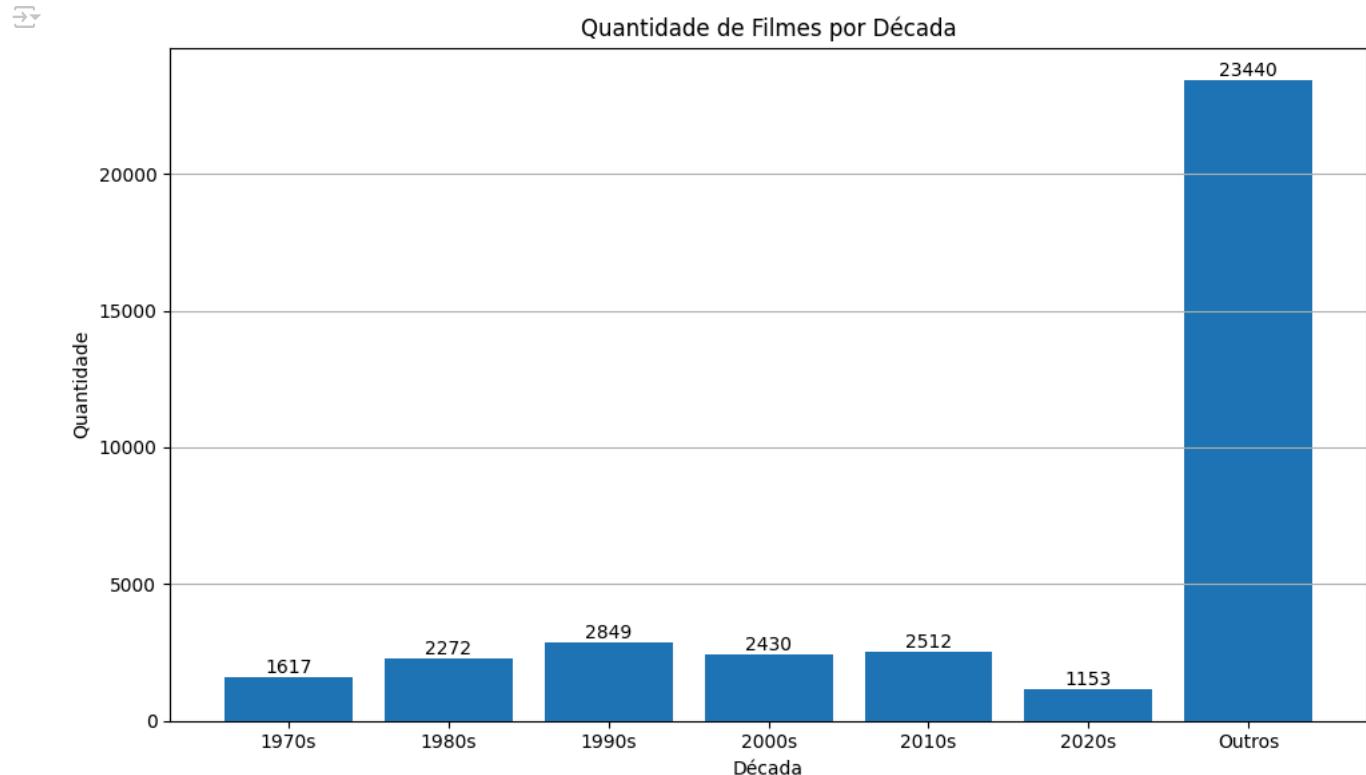
```
def get_decade(year):
    try:
        year = int(year)
    except:
        return 'Outros'
    if 1970 <= year < 1980:
        return '1970s'
    elif 1980 <= year < 1990:
        return '1980s'
    elif 1990 <= year < 2000:
        return '1990s'
    elif 2000 <= year < 2010:
        return '2000s'
    elif 2010 <= year < 2020:
        return '2010s'
    elif 2020 <= year < 2030:
        return '2020s'
    else:
        return 'Outros'

df['decade'] = df['year'].apply(get_decade)
```

5. (1 ponto) Crie um gráfico que apresente Quantidade de filmes por década

```
import matplotlib.pyplot as plt

decade_counts = df['decade'].value_counts().sort_index()
fig, ax = plt.subplots(figsize=(10, 6))
bars = ax.bar(decade_counts.index, decade_counts.values)
for bar in bars:
    yval = bar.get_height()
    ax.text(bar.get_x() + bar.get_width()/2.0, yval + 1, int(yval), ha='center', va='bottom')
ax.set_title('Quantidade de Filmes por Década')
ax.set_xlabel('Década')
ax.set_ylabel('Quantidade')
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```



6. (0,5 pontos) Explicação do Insight:

Escreva aqui sua interpretação sobre o gráfico e qual foi a década com mais produções. Quais Diretores aparecem com maior frequência?

O gráfico mostra que muitos filmes estão na categoria "Outros", o que indica registros com anos ausentes ou fora da faixa de 1970 a 2030. Entre os anos válidos, os anos 1990 se destacam com mais títulos, o que reflete o crescimento da indústria nesse período. Além disso, diretores como Woody Allen, Steven Spielberg e Clint Eastwood aparecem com frequência, mostrando sua importância e produtividade ao longo do tempo.