

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

```
data = pd.read_csv("hsb2f.csv", sep=";")df =
data.loc[:, data.columns != 'id']
```

```
sns.set(style="whitegrid")
```

Clique duas vezes (ou prima Enter) para editar.

```
df.head(10)
```

	genero	raca	clasocial	tipescola	programa	ler	escrever	matematica	ciencias	estsociais
0	masculino	branca	baixa	pública	básico	57	52	41	47	57
1	feminino	branca	média	pública	técnico	68	59	53	63	61
2	masculino	branca	alta	pública	básico	44	33	54	58	31
3	masculino	branca	alta	pública	técnico	63	44	47	53	56
4	masculino	branca	média	pública	acadêmico	47	52	57	53	61
5	masculino	branca	média	pública	acadêmico	44	52	51	63	61
6	masculino	afro-americana	média	pública	básico	50	59	42	53	61
7	masculino	hispânica	média	pública	acadêmico	34	46	45	39	36
8	masculino	branca	média	pública	básico	63	57	54	58	51
9	masculino	afro-americana	média	pública	acadêmico	57	55	52	50	51

Passos seguintes

Gerar código com df

Ver gráficos recomendados

New interactive sheet

```
df.info()
```

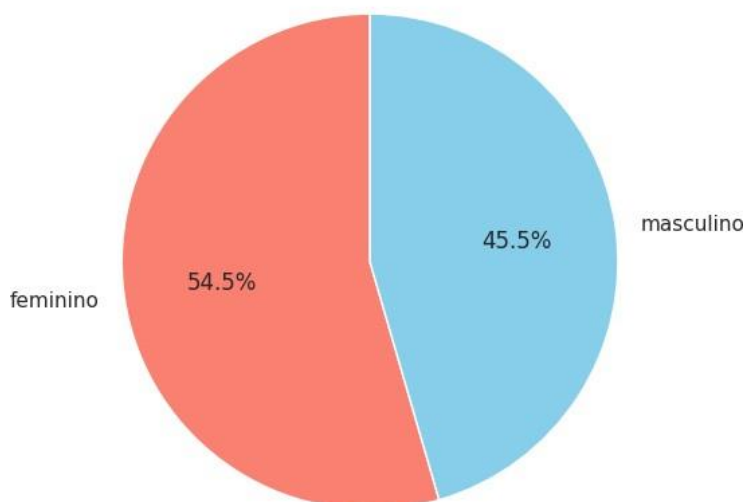
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 10 columns):
#   Column      Non-Null Count  Dtype
---  -
0   genero      200 non-null   object
1   raca        200 non-null   object
2   clasocial   200 non-null   object
3   tipescola   200 non-null   object
4   programa    200 non-null   object
5   ler         200 non-null   int64
6   escrever    200 non-null   int64
7   matematica  200 non-null   int64
8   ciencias    200 non-null   int64
9   estsociais  200 non-null   int64
dtypes: int64(5), object(5)
memory usage: 15.8+ KB
```

▢ Distribuição do gênero

```
plt.figure(figsize=(6, 6))
df['genero'].value_counts().plot.pie(autopct='%1.1f%%', startangle=90, colors=['salmon', 'skyblue'])
plt.title('Distribuição do gênero')
plt.ylabel('')
plt.show()
```



Distribuição do gênero



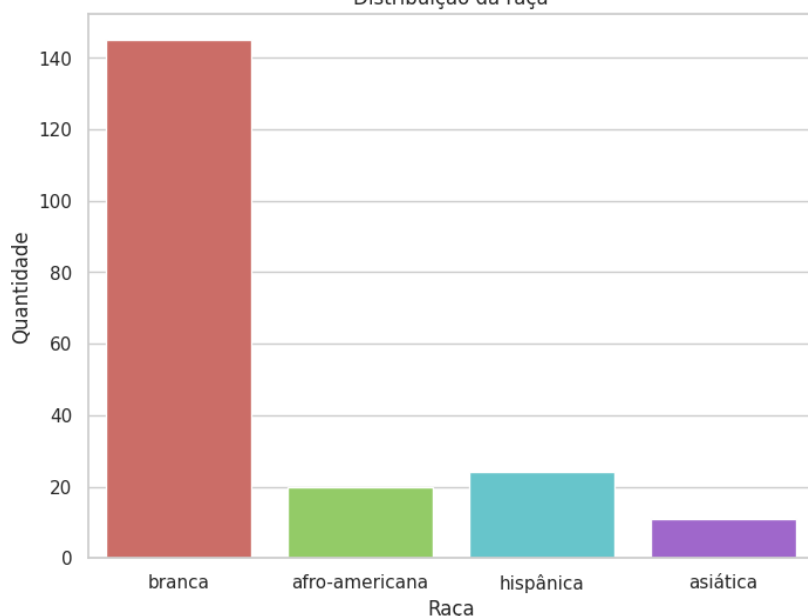
O gráfico mostra a distribuição de gênero observado nas amostras da pesquisa, realizada de forma aleatoriamente na High School and Beyond (2024), representado por um gráfico de pizza. A porcentagem de indivíduos do gênero feminino é de 54,5%, enquanto a porcentagem de indivíduos do gênero masculino é de 45,5%. Isso indica que há uma leve predominância de mulheres em relação aos homens no grupo analisado, com uma diferença de 9% entre os dois gêneros.

□ Distribuição da raça

```
plt.figure(figsize=(8, 6))
sns.countplot(data=data, x='raca', palette='hls', hue='raca')
plt.title('Distribuição da raça')
plt.xlabel('Raça')
plt.ylabel('Quantidade')
plt.show()
```



Distribuição da raça



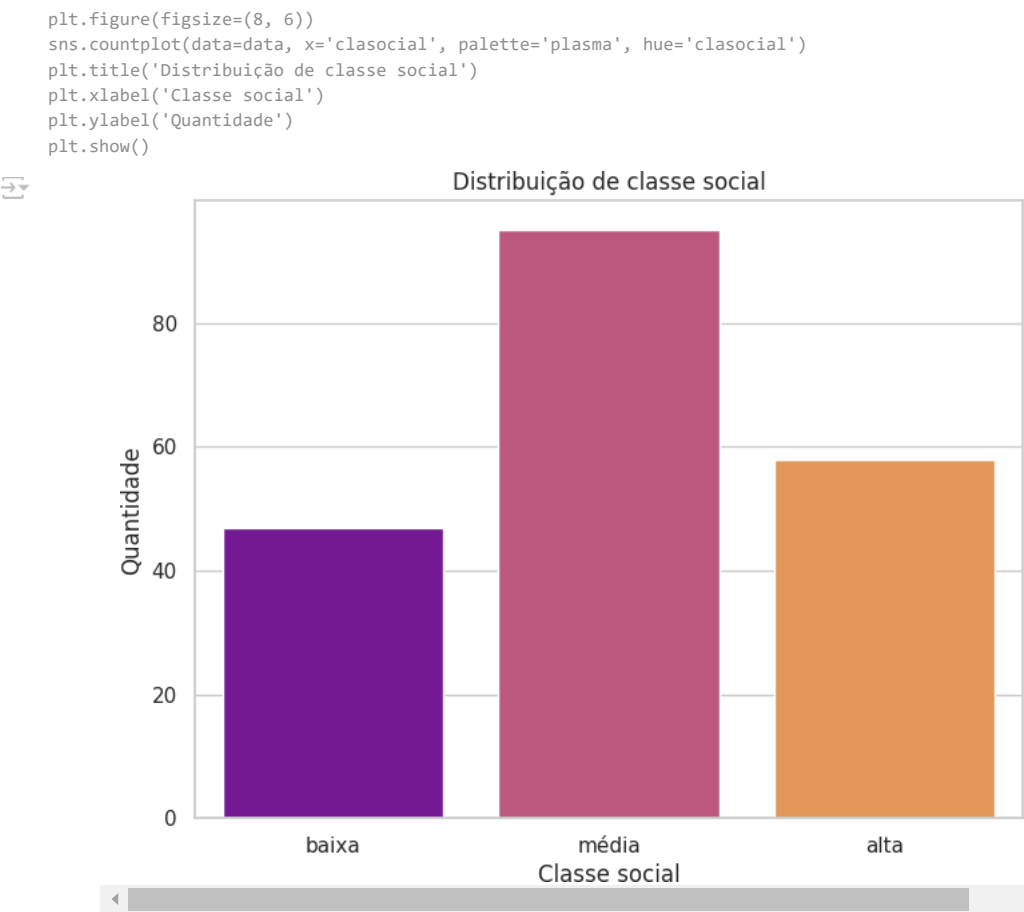
O gráfico de barras mostra a distribuição de uma variável rotulada "raça", entre as diversidades étnicas raciais encontrada nos dados observados de forma aleatoriamente na High School and Beyond (2024). As categorias apresentadas são "branca", "afro-americana", "hispânica" e "asiática".

Observações do gráfico:

- 1. **Categoria Dominante** : A categoria "branca" tem a maior quantidade, com um valor acima de 140. Isso sugere que os indivíduos nesta categoria constituem a maioria absoluta do conjunto de dados.
- 2. **Categorias Minoritárias**: As outras categorias raciais ("afro-americana", "hispânica" e "asiática") possuem contagens significativamente mais baixas. Os grupos "afro-americana" e "hispânica" possuem quantidades quase semelhantes, enquanto a categoria "asiática" tem a contagem mais baixa.
- 3. **Disparidade**: Há uma clara disparidade na representação entre a categoria "branca" e as outras categorias, indicando um desequilíbrio na composição racial deste conjunto de dados.

Essa distribuição pode ser relevante para análises demográficas, e a representação significativa de um grupo pode impactar interpretações em estudos onde a diversidade racial é um fator.

▢ Distribuição da classe social



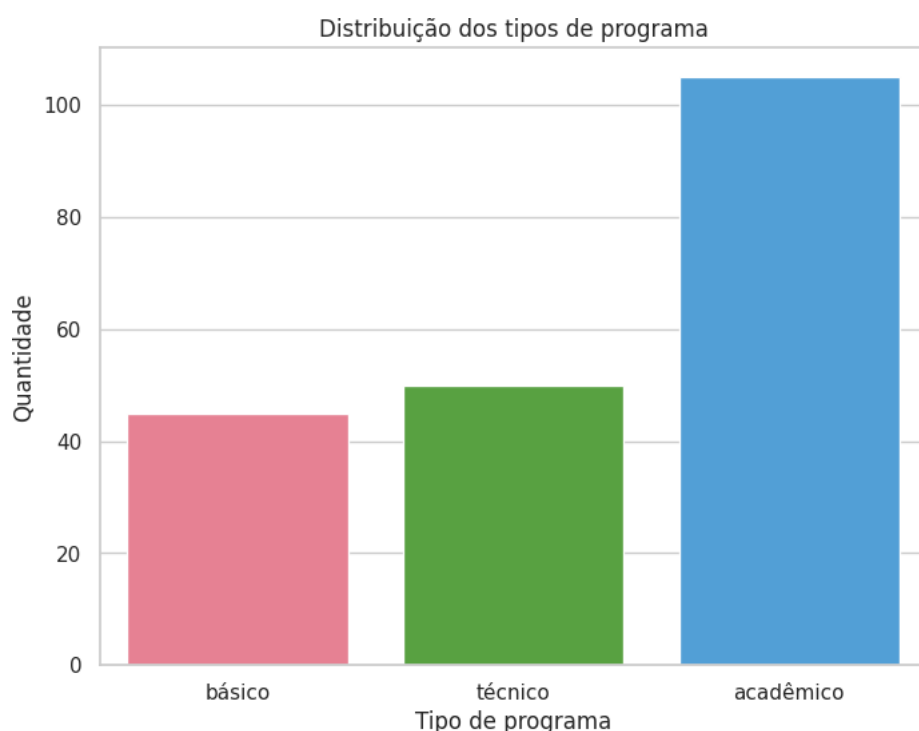
O gráfico apresentado representa a distribuição da classe social em um determinado grupo de pessoas. Cada barra representa uma classe social (baixa, média e alta), e a altura da barra indica a quantidade de pessoas que se encaixam nessa classificação.

- **Maioria da classe média:** A barra correspondente à classe média é a mais alta, indicando que a maior parte das pessoas no grupo analisado se classifica como classe média.
- **Menor quantidade na classe baixa:** A barra da classe baixa é a menor, sugerindo que há uma menor quantidade de pessoas nessa categoria.
- **Classe alta em segundo lugar:** A classe alta possui uma quantidade intermediária de pessoas, ficando em segundo lugar em relação à classe média.

Este gráfico nos permite visualizar a distribuição da classe social de forma clara e concisa. Ele indica que, no grupo analisado, a classe média é a predominante, seguida pela classe alta e, por fim, pela classe baixa.

□ Distribuição dos tipos de programa

```
plt.figure(figsize=(8, 6))
sns.countplot(data=data, x='programa', palette='husl', hue='programa')
plt.title('Distribuição dos tipos de programa')
plt.xlabel('Tipo de programa')
plt.ylabel('Quantidade')
plt.show()
```



O gráfico de barras apresentado demonstra a quantidade de cada tipo de programa (básico, técnico e acadêmico). Cada barra representa um tipo de programa, e a altura da barra indica a quantidade total desse tipo.

- **Maioria dos programas:** É evidente que a maior parte dos programas é do tipo acadêmico. A barra correspondente a esse tipo é significativamente mais alta que as demais, indicando um número consideravelmente maior de programas nessa categoria.
- **Menor quantidade:** Os programas básicos apresentam a menor quantidade entre os três tipos. A barra correspondente é a mais baixa do gráfico.
- **Quantidade intermediária:** Os programas técnicos ocupam uma posição intermediária, com uma quantidade maior que os básicos, mas menor que os acadêmicos.

Sendo assim, podemos concluir que:

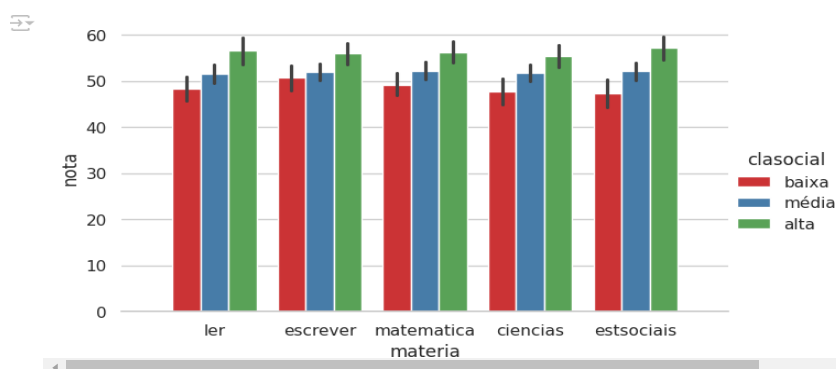
- **Foco em programas acadêmicos:** A instituição ou área analisada possui um forte foco em programas acadêmicos, com uma oferta muito maior desse tipo de programa em comparação aos demais.
- **Diversidade de opções:** A instituição oferece uma variedade de programas, abrangendo desde os básicos até os acadêmicos.
- **Desigualdade na distribuição:** A distribuição dos programas não é uniforme, com uma concentração significativa nos programas acadêmicos.

□ Classe social X Desempenho

```
data_long = data.melt(id_vars=["genero", "clasocial"],
                      value_vars=["ler", "escrever", "matematica", "ciencias", "estsociais"],
                      var_name="materia", value_name="nota")

sns.set_style('whitegrid')g = sns.catplot(x="materia", y="nota", hue="clasocial", data=data_long, kind="bar", height=4,
aspect=1.5, palette="Set1"for ax in g.axes.flat:
    sns.despine(ax=ax, left=True)
    ax.margins(x=0.1)

plt.subplots_adjust(wspace=0, bottom=0.18, left=0.06)
```



O gráfico de barras apresentado compara o desempenho de alunos em diferentes matérias (leitura, escrita, matemática, ciências e estudos sociais), divididos por classe social (baixa, média e alta). Cada barra representa a nota média de um grupo específico de alunos em uma determinada matéria, e as cores diferenciam as classes sociais.

- **Desempenho geral:** De forma geral, o gráfico sugere que o desempenho dos alunos varia de acordo com a matéria e a classe social.
- **Influência da classe social:** As barras verdes (classe alta) tendem a ser mais altas, indicando um desempenho geralmente superior em comparação aos alunos de classe média (azul) e baixa (vermelha). Essa diferença é mais evidente em algumas matérias do que em outras.
- **Desempenho por matéria:** O desempenho varia significativamente entre as diferentes matérias. Algumas matérias, como ciências e estudos sociais, apresentam uma diferença mais acentuada entre as classes sociais, enquanto em outras, como leitura e escrita, essa diferença parece ser menor.
- **A classe social influencia o desempenho escolar:** Alunos de classe social mais alta tendem a apresentar melhores resultados nas diferentes matérias.
- **As diferenças de desempenho variam entre as matérias:** Algumas matérias são mais influenciadas pela classe social do que outras, de acordo com o gráfico.
- **Fatores além da classe social:** É importante ressaltar que a classe social é apenas um dos fatores que podem influenciar o desempenho escolar. Outros fatores, como a qualidade do ensino, recursos disponíveis em casa, expectativas familiares, etc., também podem desempenhar um papel importante.

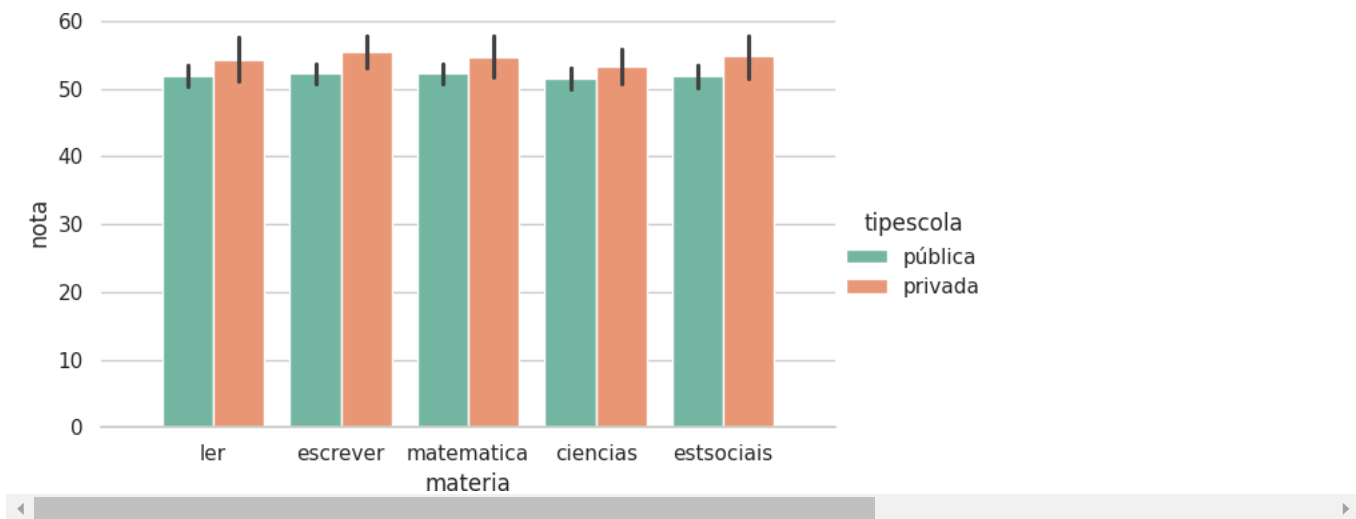
□ Tipo escola X Desempenho

```
data_long = data.melt(id_vars=["genero", "tipescola"],
                      value_vars=["ler", "escrever", "matematica", "ciencias", "estsociais"],
                      var_name="matéria", value_name="nota")

sns.set_style('whitegrid')
g = sns.catplot(x="matéria", y="nota", hue="tipescola", data=data_long, kind="bar", height=4, aspect=1.5, palette="Set2")
for ax in g.axes.flat:
    sns.despine(ax=ax, left=True)
    ax.margins(x=0.1)

plt.subplots_adjust(wspace=0, bottom=0.18, left=0.06)
plt.show()
```





O gráfico nos dá uma visão geral do desempenho dos alunos em escolas públicas e privadas. Ele mostra que, em geral, as notas são bastante semelhantes nas duas redes de ensino, com pequenas variações em algumas disciplinas. No entanto, para entender completamente o que esses números significam, precisamos considerar o contexto em que os dados foram coletados e outras informações relevantes. É como se estivéssemos comparando duas turmas de uma mesma série: As notas podem ser muito parecidas, mas cada aluno tem sua própria história e suas próprias dificuldades.

Com base nas informações fornecidas, podemos concluir que a pesquisa realizada na High School and Beyond (2024) revela várias tendências importantes:

1. Distribuição de Gênero: Há uma leve predominância de mulheres (54,5%) em relação aos homens (45,5%) na amostra analisada.
2. Diversidade Étnica: A categoria "branca" é a mais representada, com uma quantidade significativamente maior do que as categorias "afro-americana", "hispanica" e "asiática". Isso indica uma disparidade na composição racial do grupo.
3. Classe Social: A maioria das pessoas pertence à classe média, seguida pela classe alta e, por fim, pela classe baixa. Isso mostra uma predominância da classe média no grupo analisado.
4. Tipos de Programas: A maioria dos programas oferecidos é do tipo acadêmico, com uma quantidade menor de programas técnicos e básicos. Isso sugere um foco maior em programas acadêmicos na instituição analisada.
5. Desempenho Escolar por Classe Social: O desempenho dos alunos varia de acordo com a matéria e a classe social, com alunos de classe alta geralmente apresentando melhores resultados. No entanto, essa diferença é mais acentuada em algumas matérias do que em outras.

6. Desempenho em Escolas Públicas e Privadas: As notas dos alunos são bastante semelhantes nas duas redes de ensino, com pequenas variações em algumas disciplinas. Isso indica que, apesar das diferenças contextuais, o desempenho acadêmico é comparável entre escolas públicas e privadas.

Essas conclusões destacam a importância de considerar fatores como gênero, diversidade étnica, classe social e tipo de programa ao analisar dados educacionais, pois eles podem influenciar significativamente os resultados e interpretações dos estudos.