

Atividade - Analise simples de dados

Bibliotecas e Leitura da base:

```
In [129... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [130... df = pd.read_csv("tabela1.csv", sep=';', index_col = [0])
df
```

Out[130]:

	Turma	Sexo	Idade	Alt	Peso	Filhos	Fuma	Toler	Exerc	Cine	OpCine	TV	OpTV
Id													
1	A	F	17	1.60	60.5	2	NAO	P	0	1	B	16	R
2	A	F	18	1.69	55.0	1	NAO	M	0	1	B	7	R
3	A	M	18	1.85	72.8	2	NAO	P	5	2	M	15	R
4	A	M	25	1.85	80.9	2	NAO	P	5	2	B	20	R
5	A	F	19	1.58	55.0	1	NAO	M	2	2	B	5	R
6	A	M	19	1.76	60.0	3	NAO	M	2	1	B	2	R
7	A	F	20	1.60	58.0	1	NAO	P	3	1	B	7	R
8	A	F	18	1.64	47.0	1	SIM	I	2	2	M	10	R
9	A	F	18	1.62	57.8	3	NAO	M	3	3	M	12	R
10	A	F	17	1.64	58.0	2	NAO	M	2	2	M	10	R
11	A	F	18	1.72	70.0	1	SIM	I	10	2	B	8	N
12	A	F	18	1.66	54.0	3	NAO	M	0	2	B	0	R
13	A	F	21	1.70	58.0	2	NAO	M	6	1	M	30	R
14	A	M	19	1.78	68.5	1	SIM	I	5	1	M	2	N
15	A	F	18	1.65	63.5	1	NAO	I	4	1	B	10	R
16	A	F	19	1.63	47.4	3	NAO	P	0	1	B	18	R
17	A	F	17	1.82	66.0	1	NAO	P	3	1	B	10	N
18	A	M	18	1.80	85.2	2	NAO	P	3	4	B	10	R
19	A	NI	20	1.60	54.5	1	NAO	P	3	2	B	5	R
20	A	F	18	1.68	52.5	3	NAO	M	7	2	B	14	M
21	A	F	21	1.70	60.0	2	NAO	P	8	2	B	5	R
22	A	F	18	1.65	58.5	1	NAO	M	0	3	B	5	R
23	A	F	18	1.57	49.2	1	SIM	I	5	4	B	10	R
24	A	F	20	1.55	48.0	1	SIM	I	0	1	M	28	R
25	A	F	20	1.69	51.6	2	NAO	P	8	5	M	4	N
26	A	F	19	1.54	57.0	2	NAO	I	6	2	B	5	R
27	B	F	23	1.62	63.0	2	NAO	M	8	2	M	5	R
28	B	F	18	1.62	52.0	1	NAO	P	1	1	M	10	R
29	B	F	18	1.57	49.0	2	NAO	P	3	1	B	12	R
30	B	F	25	1.65	59.0	4	NAO	M	1	2	M	2	R
31	B	F	18	1.61	52.0	1	NAO	P	2	2	M	6	N
32	B	M	17	1.71	73.0	1	NAO	P	1	1	B	20	R
33	B	F	17	1.65	56.0	3	NAO	M	2	1	B	14	R
34	B	F	17	1.67	58.0	1	NAO	M	4	2	B	10	R

	Turma	Sexo	Idade	Alt	Peso	Filhos	Fuma	Toler	Exerc	Cine	OpCine	TV	OpTV
Id													
35	B	M	18	1.73	87.0	1	NAO	M	7	1	B	25	B
36	B	F	18	1.60	47.0	1	NAO	P	5	1	M	14	R
37	B	M	17	1.70	95.0	1	NAO	P	10	2	M	12	N
38	B	M	21	1.85	84.0	1	SIM	I	6	4	B	10	R
39	B	F	18	1.70	60.0	1	NAO	P	5	2	B	12	R
40	B	NI	18	1.73	73.0	1	NAO	M	4	1	B	2	R
41	B	F	17	1.70	55.0	1	NAO	I	5	4	B	10	B
42	B	F	23	1.45	44.0	2	NAO	M	2	2	B	25	R
43	B	M	24	1.76	75.0	2	NAO	I	7	0	M	14	N
44	B	F	18	1.68	55.0	1	NAO	P	5	1	B	8	R
45	B	F	18	1.55	49.0	1	NAO	M	0	1	M	10	R
46	B	F	19	1.70	50.0	7	NAO	M	0	1	B	8	R
47	B	F	19	1.55	54.5	2	NAO	M	4	3	B	3	R
48	B	F	18	1.60	50.0	1	NAO	P	2	1	B	5	R
49	B	M	17	1.80	71.0	1	NAO	P	7	0	M	14	R
50	B	M	18	1.83	86.0	1	NAO	P	7	0	M	20	B

Analise 01 - Fumantes por genero

In [131...

```
quantidade_machos = len(df.loc[df['Sexo'] == 'M'])
quantidade_machos_fumantes = len(df.loc[(df['Sexo'] == 'M') & (df['Fuma'] == 'SI
quantidade_femeas = len(df.loc[df['Sexo'] == 'F'])
quantidade_femeas_fumantes = len(df.loc[(df['Sexo'] == 'F') & (df['Fuma'] == 'SI

print(f"Quantidade de homens: {quantidade_machos}")
print(f"Quantidade de homens fumantes: {quantidade_machos_fumantes}")
print(f"Porcentagem de homens fumantes: {(quantidade_machos_fumantes/quantidade_
print(f"Quantidade de mulheres: {quantidade_femeas}")
print(f"Quantidade de mulheres fumantes: {quantidade_femeas_fumantes}")
print(f"Porcentagem de mulheres fumantes: {(quantidade_femeas_fumantes/quantidad
```

Quantidade de homens: 12
Quantidade de homens fumantes: 2
Porcentagem de homens fumantes: 16.67%

Quantidade de mulheres: 36
Quantidade de mulheres fumantes: 4
Porcentagem de mulheres fumantes: 11.11%

Essa analise demosntra que embora existam mais mulheres que fumam, elas fazem parte de uma menor porcentagem do que os homens.

Podendo implicar que é mais fácil um homem fumar do que uma mulher.

Análise 02 - Verificar se a disposição do grupo influencia na decisão de fumar

```
In [132... df_machos = df.loc[df['Sexo'] == 'M']
macho_count = df_machos.value_counts("Toler", normalize=True)

df_femeas = df.loc[df['Sexo'] == 'F']
femea_count = df_femeas.value_counts("Toler", normalize=True)

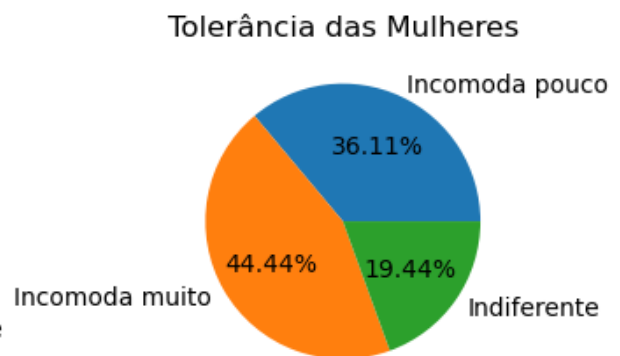
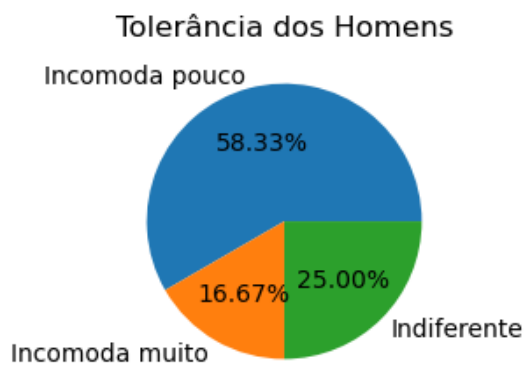
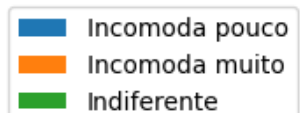
indexes = ["Incomoda pouco", "Incomoda muito", "Indiferente"]
macho_tol = [macho_count['P'], macho_count['M'], macho_count['I']]
femea_tol = [femea_count['P'], femea_count['M'], femea_count['I']]

fig, axs = plt.subplots(1, 2, constrained_layout = True)
fig.suptitle("Tolerância ao fumo por gênero", fontsize=14)
axs[0].set_title("Tolerância dos Homens")
axs[0].pie(macho_tol, labels=indexes, autopct='%1.2f%%')

handles, labels = axs[0].get_legend_handles_labels()
fig.legend(handles, labels)

axs[1].set_title("Tolerância das Mulheres")
axs[1].pie(femea_tol, labels=indexes, autopct='%1.2f%%')
plt.show()
```

Tolerância ao fumo por gênero



A partir do gráfico acima, podemos chegar à conclusão de que, em geral, os homens tendem a se importar menos com fumantes ao seu redor, enquanto as mulheres demonstram maior preocupação.

Dado que o ambiente influencia nas ações do indivíduo, as mulheres tendem a não fumar, já que a tolerância ao seu redor não o permitiria.

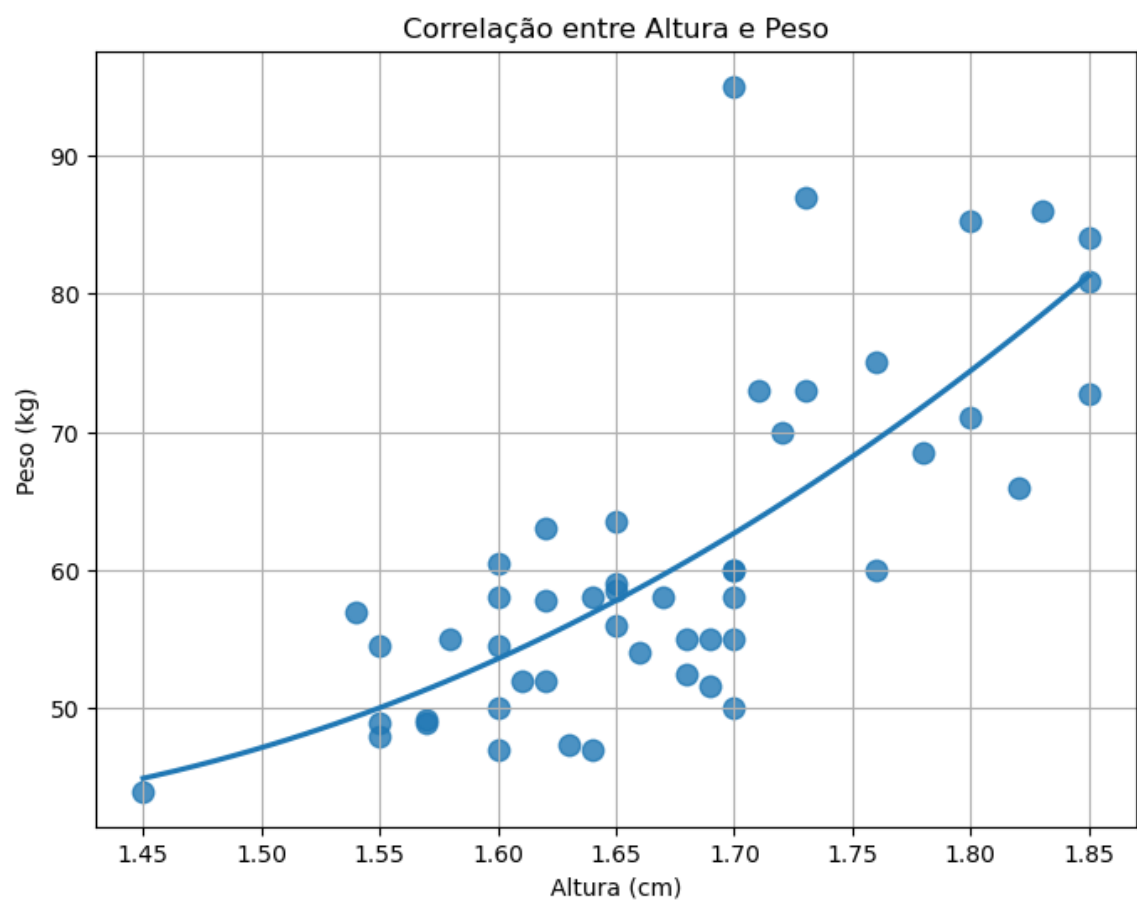
Por outro lado, os homens enfrentarão uma pressão menor e terão uma chance maior de se tornarem fumantes.

Essa conclusão está em consonância com a Análise 1 e pode revelar uma das implicações do resultado obtido.

Analise 03 - Correlação entre Peso e Altura

In [133...

```
plt.figure(figsize=(8, 6))
sns.regplot(x='Alt', y='Peso', data=df, ci=None, scatter_kws={"s": 80}, order=2)
plt.title('Correlação entre Altura e Peso')
plt.xlabel('Altura (cm)')
plt.ylabel('Peso (kg)')
plt.grid(True)
plt.show()
```



Observando o gráfico, fica evidente que, em geral, à medida que a altura de um indivíduo aumenta, seu peso também tende a aumentar. No entanto, é importante notar que o gráfico inclui uma linha que representa a tendência dos dados, calculado a partir de uma regressão.

Essa linha visa indicar que, apesar da tendência geral de aumento, os dados apresentam diversos desvios em relação à média. Mesmo que haja uma inclinação crescente, há variações consideráveis em torno dessa tendência normativa.

Analise 04 - Correlação entre Peso e Altura e Horas de Atividade Física

In [134...

```
imc = df['Peso'] / (df['Alt'] ** 2)
imc
```

```
Out[134]: Id
1      23.632812
2      19.257029
3      21.271001
4      23.637692
5      22.031726
6      19.369835
7      22.656250
8      17.474717
9      22.024082
10     21.564545
11     23.661439
12     19.596458
13     20.069204
14     21.619745
15     23.324151
16     17.840340
17     19.925130
18     26.296296
19     21.289062
20     18.601190
21     20.761246
22     21.487603
23     19.960242
24     19.979188
25     18.066594
26     24.034407
27     24.005487
28     19.814053
29     19.879103
30     21.671258
31     20.060954
32     24.964946
33     20.569330
34     20.796730
35     29.068796
36     18.359375
37     32.871972
38     24.543462
39     20.761246
40     24.391059
41     19.031142
42     20.927467
43     24.212293
44     19.486961
45     20.395421
46     17.301038
47     22.684703
48     19.531250
49     21.913580
50     25.680074
dtype: float64
```

```
In [135... fig, axs = plt.subplots(1, 2, constrained_layout = True, figsize=(8,6))
fig.suptitle("Correlação entre Índice de Massa Corporal (IMC) e horas de exercíc

CORES = {
    'C': (0.8, 0.8, 0.8, 0.3),
    'Y': (1, 1, 0, 0.3),
    'G': (0, 1, 0, 0.3),
```

```

'M': (1, 0, 1, 0.3),
}

axs[0].plot(imc, df["Exerc"], "o")
axs[0].axvspan(10, 18.5, color=CORES['C'], label='Abaixo do peso')
axs[0].axvspan(18.6, 24.9, color=CORES['Y'], label='Peso ideal')
axs[0].axvspan(25.0, 29.9, color=CORES['G'], label='Levemente acima do peso')
axs[0].axvspan(30.0, 34.9, color=CORES['M'], alpha=0.3, label='Obesidade grau I')
axs[0].set_xlim([15, 35])

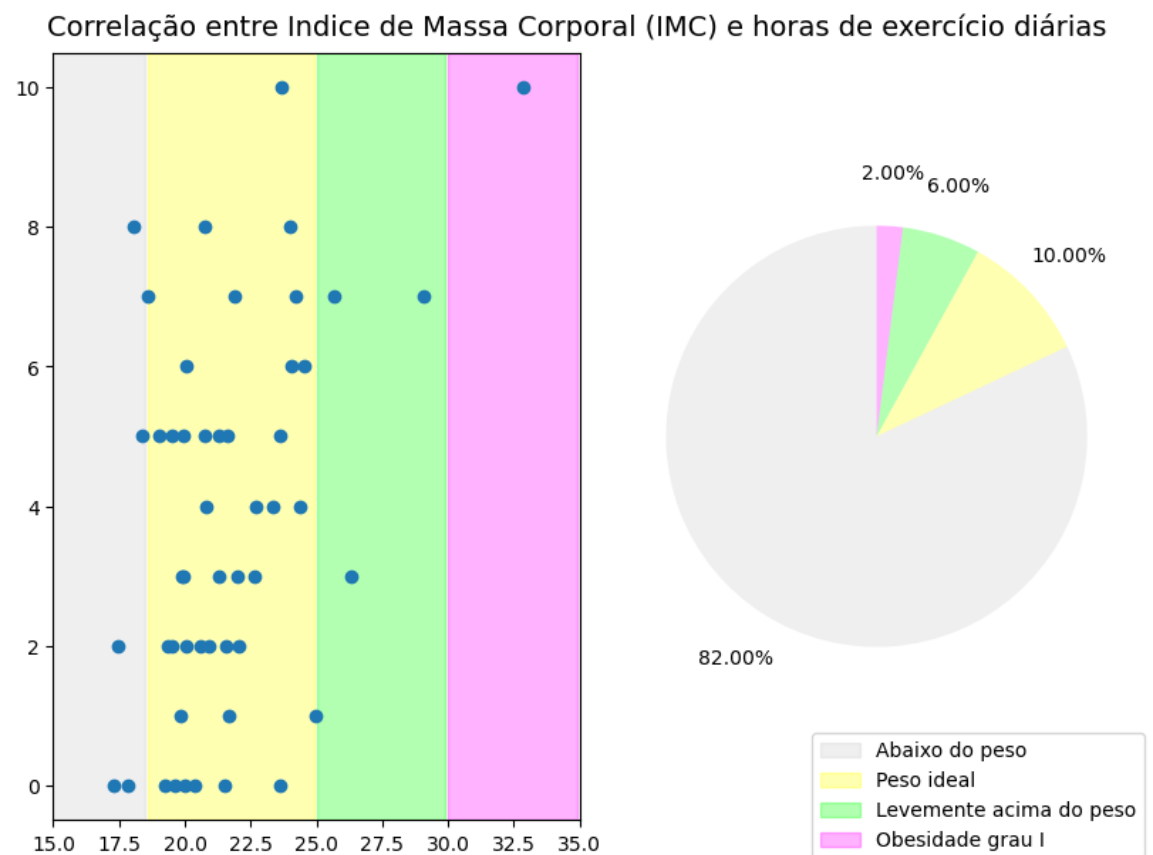
lim = pd.cut(imc, bins=[0, 18.5, 25, 30, 35, 40, 45, 50]).value_counts()
lim = lim[lim!=0]

axs[1].pie(lim, autopct='%1.2f%%', startangle=90, pctdistance=1.25, colors=[CORES['C'], CORES['Y'], CORES['G'], CORES['M']])

handles, labels = axs[0].get_legend_handles_labels()
fig.legend(handles, labels, loc='lower right')

plt.show()

```



Em primeiro lugar, esta análise ilustra como é possível agrupar duas ou mais variáveis através de cálculos, criando uma variável composta que, em determinadas situações, pode ser mais significativa.

Além disso, a análise revela que uma quantidade reduzida de horas de exercício diário não necessariamente resulta em uma qualidade inferior do Índice de Massa Corporal (IMC), e a grande maioria está dentro dos padrões considerados aceitáveis.

Análise 05 - Tipo de mídia preferida pelas turmas

In [136...

```
dfA = df[df['Turma'] == 'A']
dfB = df[df['Turma'] == 'B']

turmas = ("Turma A", "Turma B")
preferencias = {
    'Cinema Bom': (dfA['OpCine'].value_counts().get('B', 0), dfB['OpCine'].value
    'Cinema Muito Bom': (dfA['OpCine'].value_counts().get('M', 0), dfB['OpCine']
    'TV ruim': (dfA['OpTV'].value_counts().get('R', 0), dfB['OpTV'].value_counts
    'TV media': (dfA['OpTV'].value_counts().get('M', 0), dfB['OpTV'].value_count
    'TV boa': (dfA['OpTV'].value_counts().get('B', 0), dfB['OpTV'].value_counts(
    'TV ns': (dfA['OpTV'].value_counts().get('N', 0), dfB['OpTV'].value_counts()
}
print(preferencias)

x = np.arange(len(turmas)) # the label locations
width = 0.1 # the width of the bars
multiplier = 0

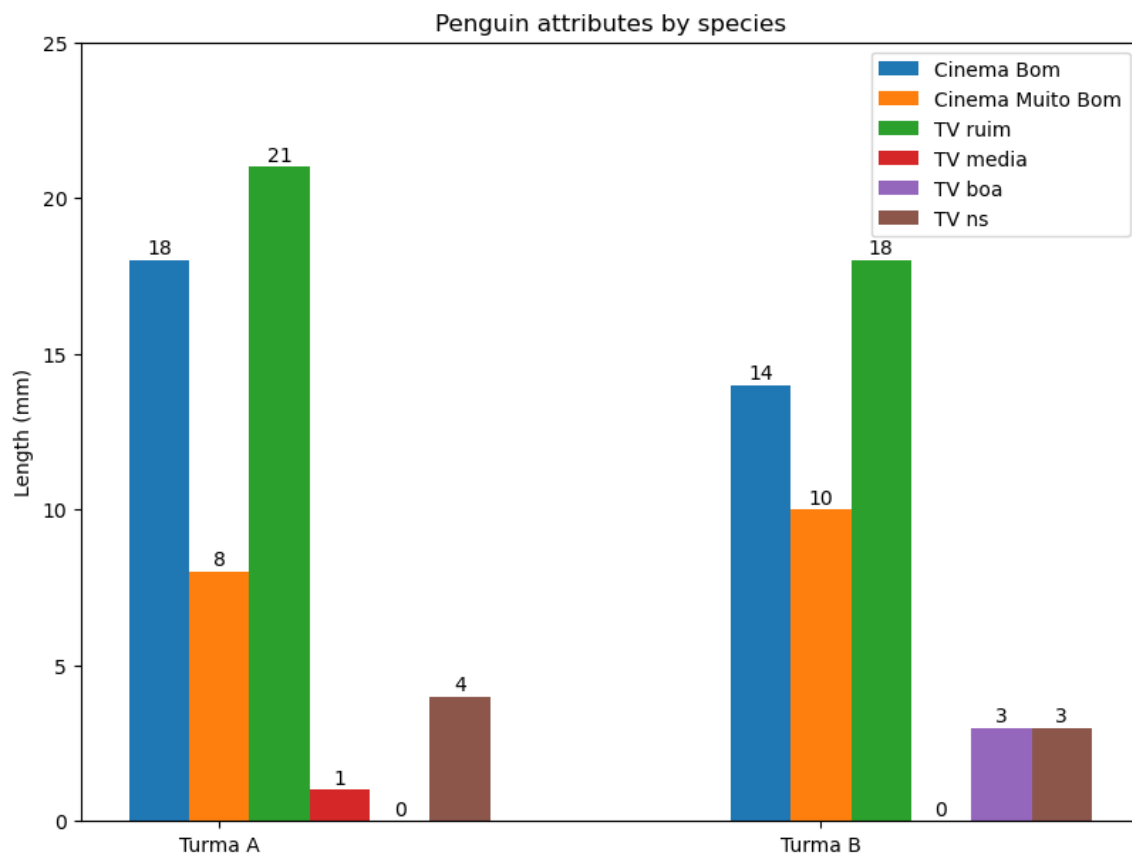
fig, ax = plt.subplots(constrained_layout = True, figsize=(8,6))

for attribute, measurement in preferencias.items():
    offset = width * multiplier
    rects = ax.bar(x + offset, measurement, width, label=attribute)
    ax.bar_label(rects, padding=1)
    multiplier += 1

# Add some text for labels, title and custom x-axis tick labels, etc.
ax.set_ylabel('Length (mm)')
ax.set_title('Penguin attributes by species')
ax.set_xticks(x + width, turmas)
ax.legend(loc='upper right')
ax.set_ylim(0, 25)

plt.show()
```

```
{'Cinema Bom': (18, 14), 'Cinema Muito Bom': (8, 10), 'TV ruim': (21, 18), 'TV
media': (1, 0), 'TV boa': (0, 3), 'TV ns': (4, 3)}
```



Este gráfico evidencia como os alunos percebem as mídias (TV e cinema).

Em geral, ambas as turmas indicam uma preferência negativa em relação à TV. Na turma A, a grande maioria demonstra apreço por assistir a filmes, sugerindo que essa atividade é recomendável para eles. Por outro lado, na turma B, as opiniões são bastante divergentes. Portanto, seria aconselhável buscar uma atividade alternativa, que não envolva nem TV nem cinema, para atender a um número maior de alunos.