

ENEM 2021

Esta é uma breve análise sobre os dados de inscritos no **ENEM 2021** de acordo com os dados encontrados neste dataset (<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>).

In [1]:

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

In [2]:

```
plt.style.use('seaborn-colorblind')
```

In [3]:

```
enem = pd.read_csv(r"C:\Users\user-08\Downloads\microdados_enem_2021\DADOS\MICRODADOS_ENEM_2021.csv")
```

Primeiro foram selecionadas algumas informações principais dos candidatos, descartando informações prescindíveis.

In [4]:

```
enem.head()
```

Out[4]:

	TP_FAIXA_ETARIA	TP_SEXO	TP_COR_RACA	TP_ESCOLA	SG_UF_RESIDENCIA	
0	20 anos	Feminino	Branca	Não Respondeu	MG	1.650,01 2.20
1	Entre 31 e 35 anos	Masculino	Branca	Não Respondeu	AL	Até R\$ 1.10
2	Entre 36 e 40 anos	Feminino	Branca	Não Respondeu	SP	1.100,01 1.65
3	18 anos	Masculino	Parda	Pública	CE	Até R\$ 1.10
4	17 anos	Feminino	Parda	Pública	CE	1.100,01 1.65

```
enem.columns
```

In [5]:

```
enem[['TP_FAIXA_ETARIA', 'TP_SEXO', 'TP_COR_RACA', 'TP_ESCOLA', 'SG_UF_RESIDENCIA']]
```

Out[5]:

	TP_FAIXA_ETARIA	TP_SEXO	TP_COR_RACA	TP_ESCOLA	SG_UF_RESIDENCIA
0	20 anos	Feminino	Branca	Não Respondeu	MG
1	Entre 31 e 35 anos	Masculino	Branca	Não Respondeu	AL
2	Entre 36 e 40 anos	Feminino	Branca	Não Respondeu	SP
3	18 anos	Masculino	Parda	Pública	CE
4	17 anos	Feminino	Parda	Pública	CE
...
1048570	23 anos	Feminino	Branca	Não Respondeu	MS
1048571	Entre 26 e 30 anos	Feminino	Parda	Não Respondeu	SP
1048572	18 anos	Feminino	Parda	Pública	SP
1048573	18 anos	Feminino	Branca	Não Respondeu	AL
1048574	17 anos	Feminino	Parda	Não Respondeu	PA

1048575 rows × 5 columns

Após termos os números, como a quantidade de candidatos era pujante, mostrar tais informações através de gráficos facilitaria no entendimento.

In [6]:

```
enem['TP_SEXO'].value_counts()/enem.shape[0]*100
```

Out[6]:

```
Feminino    67.503517
Masculino   32.496388
Name: TP_SEX0, dtype: Float64
```

In [7]:

```
enem['TP_COR_RACA'].value_counts()/enem.shape[0]*100
```

Out[7]:

```
Parda        44.7617
Branca       38.28596
Preta        12.197315
Não declarado  2.147772
Amarela       2.092173
Indígena      0.514985
Name: TP_COR_RACA, dtype: Float64
```

Neste gráfico é possível ver as diferentes etnias que se inscreveram para o Enem 2021, e a discrepância que existe entre elas.

De acordo com os canais jornalísticos, houve uma queda no número de inscritos de participantes indígenas, pretos e pardos. O que representa um retrocesso em relação à inclusão e diversidade de alunos em universidades públicas, já que o Enem é a forma de ingresso nas mesmas.

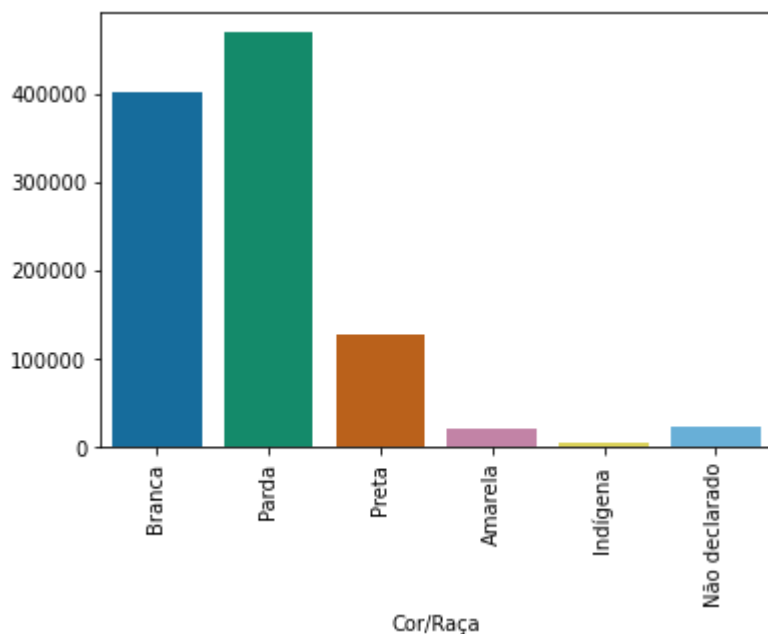
Fonte: <https://g1.globo.com/educacao/enem/2021/noticia/2021/08/27/enem-2021-cai-negros-pardos-indigenas-inscritos.ghtml> (<https://g1.globo.com/educacao/enem/2021/noticia/2021/08/27/enem-2021-cai-negros-pardos-indigenas-inscritos.ghtml>)

In [8]:

```
sns.countplot(enem['TP_COR_RACA'])
plt.xlabel('Cor/Raça')
plt.xticks(rotation=90)
plt.xlabel("Cor/Raça")
plt.ylabel("")
plt.show()
```

C:\Users\user-08\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



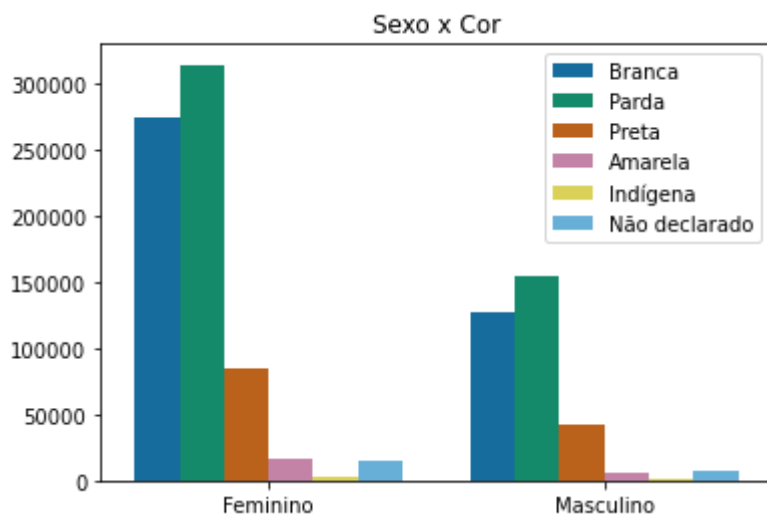
Numa breve análise podemos ver que o número de mulheres que se inscreveram é maior que dos homens.

In [9]:

```
sns.countplot(enem['TP_SEX0'], hue=enem['TP_COR_RACA'], dodge=True)
plt.title('Sexo x Cor')
plt.xlabel('')
plt.ylabel('')
plt.legend(loc='best')
plt.show()
```

C:\Users\user-08\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



In [10]:

```
enem['TP_SEX0'].value_counts()/enem.shape[0]*100
```

Out[10]:

```
Feminino      67.503517
Masculino     32.496388
Name: TP_SEX0, dtype: Float64
```

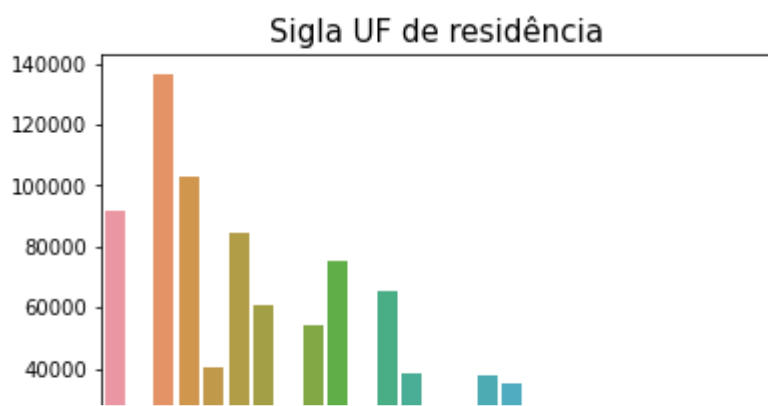
Através desse gráfico podemos visualizar as quantidades de estados que mais houve inscritos no Enem 2021, como São Paulo, Ceará e Minas Gerais. Assim como os que menos houve candidatos inscritos, como Roraima, Acre e Amapá.

In [11]:

```
sns.countplot(enem['SG_UF_RESIDENCIA'], orient='h')
plt.title('Sigla UF de residência', size=15)
plt.xticks(rotation=90)
plt.xlabel("")
plt.ylabel("")
plt.show()
```

C:\Users\user-08\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



In [12]:

```
enem['SG_UF_RESIDENCIA'].value_counts()/enem.shape[0]*100
```

Out[12]:

```
SP      12.986577
CE       9.836683
MG       8.734425
BA       8.027847
RJ       7.178409
PE       6.221205
GO       5.816561
PA       5.140977
RS       3.831295
MA       3.687004
PR       3.62387
PB       3.35007
PI       2.41709
RN       2.310946
SC       2.001669
ES       1.930715
DF       1.867296
SE       1.758386
AL       1.745893
AM       1.733686
MT       1.607801
MS       1.098252
TO       0.868321
RO       0.85106
AC       0.609208
AP       0.547314
RR       0.217343
```

Name: SG_UF_RESIDENCIA, dtype: Float64

In [13]:

```
enem.describe()
```

Out[13]:

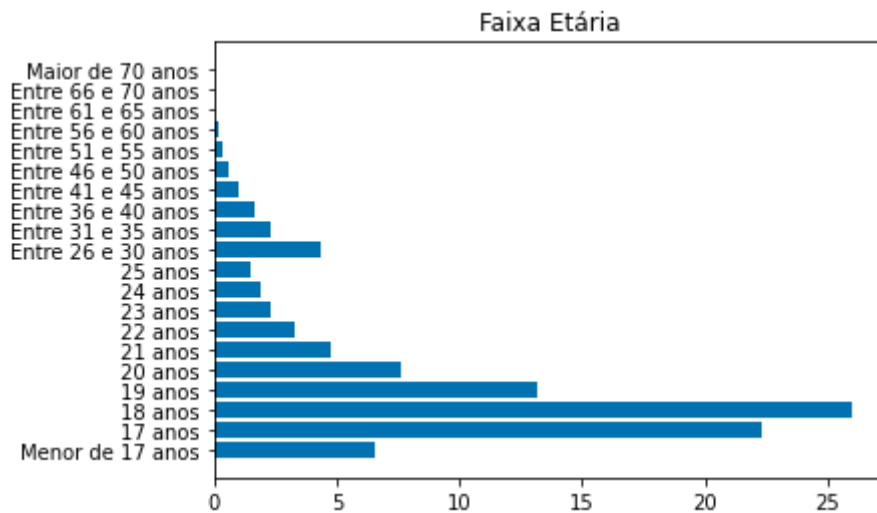
	TP_FAIXA_ETARIA	TP_SEXO	TP_COR_RACA	TP_ESCOLA	SG_UF_RESIDENCIA	Q0
count	1048574	1048574	1048574	1048574	1048574	1048574
unique	20	2	6	3	27	1
top	18 anos	Feminino	Parda	Não Respondeu	SP	Até 1.100
freq	272709	707825	469360	586061	136174	3346



Foi feita uma análise entre a faixa etaria dos inscritos, onde mostra que a maioria ainda está entre 17 e 19 anos, época onde muitos ainda tentam uma bolsa nas universidades logo que saem do ensino médio.

In [14]:

```
y=['Menor de 17 anos',  
'17 anos',  
'18 anos',  
'19 anos',  
'20 anos',  
'21 anos',  
'22 anos',  
'23 anos',  
'24 anos',  
'25 anos',  
'Entre 26 e 30 anos',  
'Entre 31 e 35 anos',  
'Entre 36 e 40 anos',  
'Entre 41 e 45 anos',  
'Entre 46 e 50 anos',  
'Entre 51 e 55 anos',  
'Entre 56 e 60 anos',  
'Entre 61 e 65 anos',  
'Entre 66 e 70 anos',  
'Maior de 70 anos',]  
x=[6.595713,  
22.297308,  
26.007582,  
13.163198,  
7.644088,  
4.793076,  
3.280357,  
2.316668,  
1.906778,  
1.471426,  
4.368882,  
2.29378,  
1.65558,  
1.018716,  
0.603009,  
0.327778,  
0.164414,  
0.061894,  
0.020695,  
0.008965,]  
plt.barh(y, x) ,  
  
plt.ylabel("")  
plt.xlabel("")  
plt.title("Faixa Etária")  
plt.show()
```



In [15]:

```
enem['TP_FAIXA_ETARIA'].value_counts()/enem.shape[0]*100
```

Out[15]:

```
18 anos          26.007582
17 anos          22.297308
19 anos          13.163198
20 anos           7.644088
Menor de 17 anos  6.595713
21 anos           4.793076
Entre 26 e 30 anos 4.368882
22 anos           3.280357
23 anos           2.316668
Entre 31 e 35 anos 2.29378
24 anos           1.906778
Entre 36 e 40 anos 1.65558
25 anos           1.471426
Entre 41 e 45 anos 1.018716
Entre 46 e 50 anos 0.603009
Entre 51 e 55 anos 0.327778
Entre 56 e 60 anos 0.164414
Entre 61 e 65 anos 0.061894
Entre 66 e 70 anos 0.020695
Maior de 70 anos  0.008965
Name: TP_FAIXA_ETARIA, dtype: Float64
```

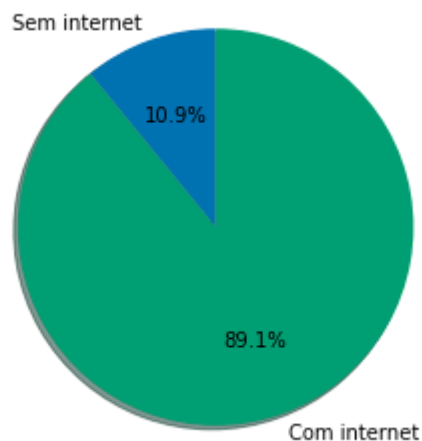
In [16]:

```
from scipy import stats
from statsmodels.stats import diagnostic
```

No gráfico abaixo é mostrado a porcentagem de candidatos sem acesso a internet na própria residência.

In [17]:

```
labels = 'Sem internet', 'Com internet'  
sizes = [10.934459, 89.065446]  
fig1, ax1 = plt.subplots()  
ax1.pie(sizes, labels=labels, autopct='%1.1f%%',  
        shadow=True, startangle=90)  
ax1.axis('equal')  
plt.show()
```



Alunos de escola pública ainda são a maioria entre os candidatos, se não contarmos com a grande porcentagem de "não respondeu".

In [18]:

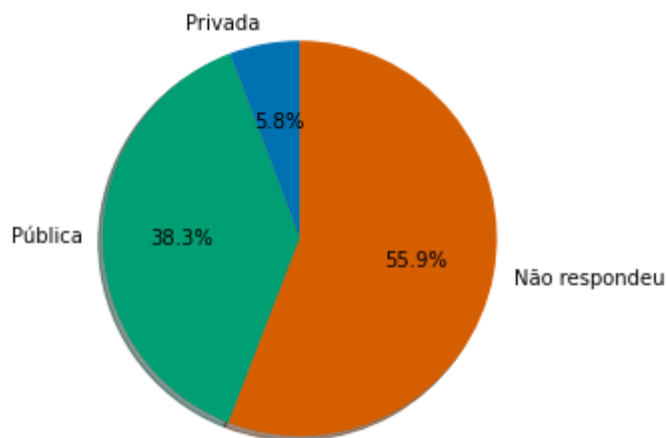
```
enem['Q025'].value_counts()/enem.shape[0]*100
```

Out[18]:

```
Sim      89.065446  
Não      10.934363  
Name: Q025, dtype: Float64
```

In [19]:

```
labels = 'Privada', 'Pública', 'Não respondeu'  
sizes = [5.778032, 38.330687, 55.891281]  
fig1, ax1 = plt.subplots()  
ax1.pie(sizes, labels=labels, autopct='%1.1f%%',  
        shadow=True, startangle=90)  
ax1.axis('equal')  
plt.show()
```



In [20]:

```
enem['TP_ESCOLA'].value_counts()/enem.shape[0]*100
```

Out[20]:

```
Não Respondeu    55.891186  
Pública          38.330687  
Privada          5.778032  
Name: TP_ESCOLA, dtype: Float64
```

Logo abaixo, conseguimos ter uma ideia sobre a renda fixa - por pessoa - na residência entre candidatos. Nota-se um grande número de "nenhuma renda" fixa.

In [21]:

```
enem['Q006'].value_counts()/enem.shape[0]*100
```

Out[21]:

Até R\$ 1.100,00	31.911213
De R\$ 1.100,01 até R\$ 1.650,00.	17.595213
De R\$ 1.650,01 até R\$ 2.200,00.	12.297642
Nenhuma Renda	7.034213
De R\$ 2.750,01 até R\$ 3.300,00.	6.256968
De R\$ 2.200,01 até R\$ 2.750,00.	5.729776
De R\$ 3.300,01 até R\$ 4.400,00.	4.957395
De R\$ 4.400,01 até R\$ 5.500,00.	3.685192
De R\$ 5.500,01 até R\$ 6.600,00.	2.24848
De R\$ 6.600,01 até R\$ 7.700,00.	1.356889
De R\$ 9.900,01 até R\$ 11.000,00.	1.307489
Acima de R\$ 22.000,00.	1.066686
De R\$ 7.700,01 até R\$ 8.800,00.	1.048566
De R\$ 13.200,01 até R\$ 16.500,00.	0.894547
De R\$ 8.800,01 até R\$ 9.900,00.	0.891496
De R\$ 16.500,01 até R\$ 22.000,00.	0.890447
De R\$ 11.000,01 até R\$ 13.200,00.	0.827599

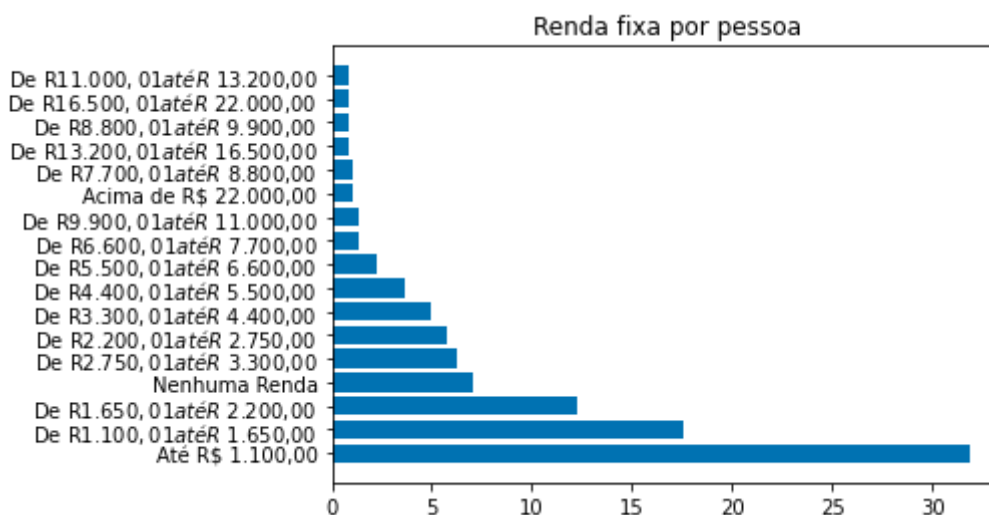
Name: Q006, dtype: Float64

In [22]:

```

y=['Até R$ 1.100,00',
'De R$ 1.100,01 até R$ 1.650,00',
'De R$ 1.650,01 até R$ 2.200,00',
'Nenhuma Renda',
'De R$ 2.750,01 até R$ 3.300,00',
'De R$ 2.200,01 até R$ 2.750,00',
'De R$ 3.300,01 até R$ 4.400,00',
'De R$ 4.400,01 até R$ 5.500,00',
'De R$ 5.500,01 até R$ 6.600,00',
'De R$ 6.600,01 até R$ 7.700,00',
'De R$ 9.900,01 até R$ 11.000,00',
'Acima de R$ 22.000,00',
'De R$ 7.700,01 até R$ 8.800,00',
'De R$ 13.200,01 até R$ 16.500,00',
'De R$ 8.800,01 até R$ 9.900,00',
'De R$ 16.500,01 até R$ 22.000,00',
'De R$ 11.000,01 até R$ 13.200,00',]
x=[31.911213,
17.595213,
12.297642,
7.034308,
6.256968,
5.729776,
4.957395,
3.685192,
2.24848,
1.356889,
1.307489,
1.066686,
1.048566,
0.894547,
0.891496,
0.890447,
0.827599,]
plt.barh(y, x)
plt.ylabel("")
plt.xlabel("")
plt.title("Renda fixa por pessoa")
plt.show()

```



Com essas informações podemos concluir que o número de inscritos do sexo feminino é maior que o masculino, independente da Cor/Raça. Concluímos também que jovens de 17 a 19 anos, época onde muitos ainda tentam uma bolsa nas universidades logo que saem do ensino médio. Além disso, podemos ver que quanto maior a renda, menor o número de inscrição no Enem.