

```
In [18]: from platform import python_version
print('Versao da Linguagem ', python_version())
```

Versao da Linguagem 3.7.4

```
In [20]: import numpy as np
import pandas as pd
```

```
In [21]: import matplotlib
```

```
In [22]: import matplotlib.pyplot as plt
```

```
In [23]: import seaborn as sns
```

```
In [25]: dadosRH = pd.read_csv('dadosRH.csv')
```

```
In [27]: dadosRH.head()
```

Out[27]:

	id_funcionario	departemento	regiao	educacao	genero	canal_recrutamento	numero_treiname
0	65438	Marketing	Sudeste	Mestrado ou Doutorado	F	Outsourcing	
1	65141	Logistica	Nordeste	Superior	M	Outro	
2	7513	Marketing	Sudeste	Superior	M	Outsourcing	
3	2542	Marketing	Norte	Superior	M	Outro	
4	48945	Tecnologia	Nordeste	Superior	M	Outro	

```
In [28]: dadosRH.shape
```

Out[28]: (54808, 11)

```
In [29]: dadosRH.isnull().sum()
```

```
Out[29]: id_funcionario      0
departamento      0
regiao             0
educacao           2409
genero             0
canal_recrutamento 0
numero_treinamentos 0
idade             0
aval_ano_anterior  4124
tempo_servico      0
promovido          0
dtype: int64
```

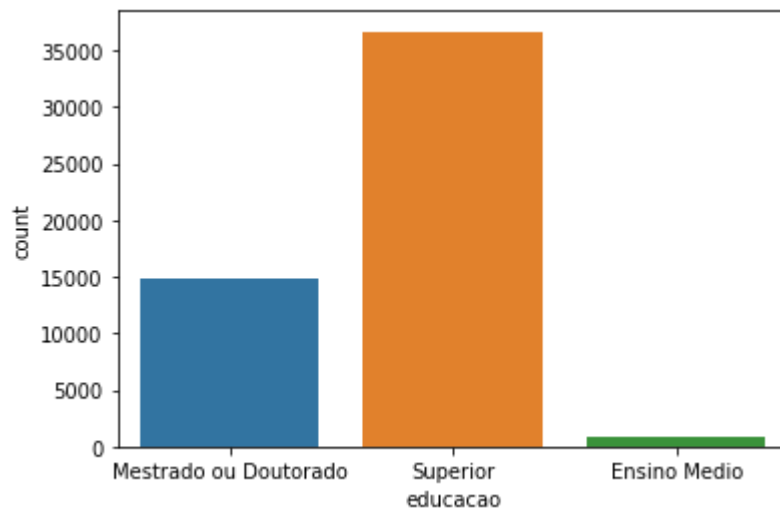
```
In [34]: dadosRH.groupby(['educacao']).count()
```

```
Out[34]:
```

	id_funcionario	departamento	regiao	genero	canal_recrutamento	numero_treinamentos
educacao						
Ensino Medio	805	805	805	805	805	805
Mestrado ou Doutorado	14925	14925	14925	14925	14925	14925
Superior	36669	36669	36669	36669	36669	36669

```
In [36]: sns.countplot(dadosRH['educacao'])
```

```
Out[36]: <matplotlib.axes._subplots.AxesSubplot at 0x2117dd5d8c8>
```



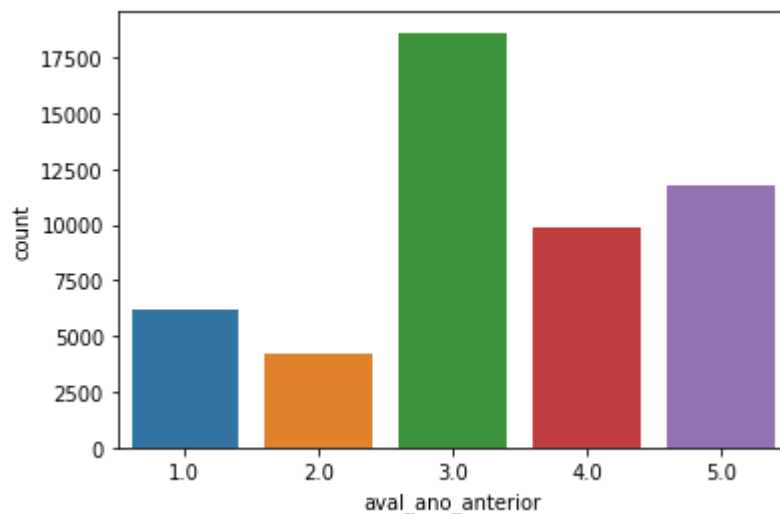
```
In [37]: dadosRH.groupby(['aval_ano_anterior']).count()
```

Out[37]:

	id_funcionario	departemento	regiao	educacao	genero	canal_recrutamento	nun
aval_ano_anterior							
1.0	6223	6223	6223	5887	6223		6223
2.0	4225	4225	4225	4035	4225		4225
3.0	18618	18618	18618	17867	18618		18618
4.0	9877	9877	9877	9509	9877		9877
5.0	11741	11741	11741	11362	11741		11741

```
In [38]: sns.countplot(dadosRH['aval_ano_anterior'])
```

Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x2117ddd9508>



```
In [42]: dadosRH['educacao'].fillna(dadosRH['educacao'].mode()[0], inplace = True)
```

```
In [44]: dadosRH['aval_ano_anterior'].fillna(dadosRH['aval_ano_anterior'].median(), inplace = True)
```

```
In [45]: dadosRH.isnull().sum()
```

```
Out[45]: id_funcionario      0
departamento      0
regiao             0
educacao           0
genero             0
canal_recrutamento 0
numero_treinamentos 0
idade              0
aval_ano_anterior  0
tempo_servico      0
promovido          0
dtype: int64
```

```
In [46]: dadosRH.shape
```

```
Out[46]: (54808, 11)
```

```
In [48]: dadosRH.isnull().sum()
```

```
Out[48]: id_funcionario      0
departamento      0
regiao             0
educacao           0
genero             0
canal_recrutamento 0
numero_treinamentos 0
idade              0
aval_ano_anterior  0
tempo_servico      0
promovido          0
dtype: int64
```

```
In [49]: dadosRH.groupby(['educacao']).count()
```

```
Out[49]:
```

	id_funcionario	departamento	regiao	genero	canal_recrutamento	numero_treinamentos
educacao						
Ensino Medio	805	805	805	805	805	805
Mestrado ou Doutorado	14925	14925	14925	14925	14925	14925
Superior	39078	39078	39078	39078	39078	39078

```
In [50]: dadosRH.groupby(['aval_ano_anterior']).count()
```

Out[50]:

	id_funcionario	departemento	regiao	educacao	genero	canal_recrutamento	nun
aval_ano_anterior							
1.0	6223	6223	6223	6223	6223		6223
2.0	4225	4225	4225	4225	4225		4225
3.0	22742	22742	22742	22742	22742		22742
4.0	9877	9877	9877	9877	9877		9877
5.0	11741	11741	11741	11741	11741		11741

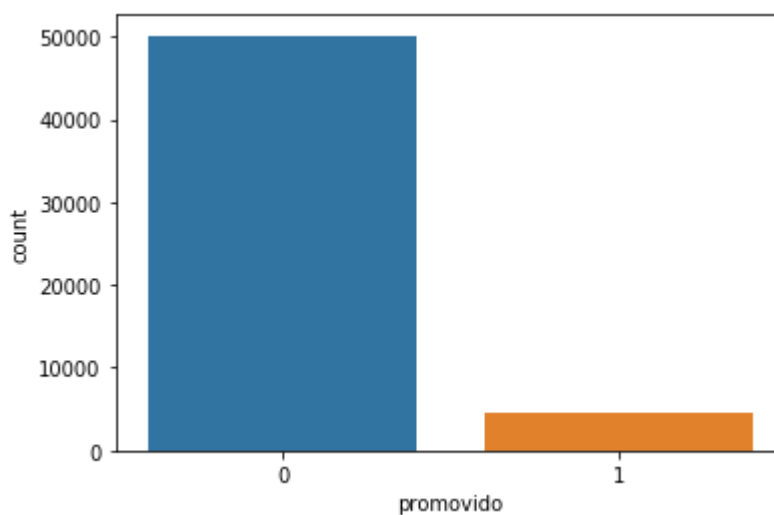
```
In [51]: dadosRH.groupby(['promovido']).count()
```

Out[51]:

	id_funcionario	departemento	regiao	educacao	genero	canal_recrutamento	numero_tr
promovido							
0	50140	50140	50140	50140	50140		50140
1	4668	4668	4668	4668	4668		4668

```
In [52]: sns.countplot(dadosRH['promovido'])
```

Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0x2117debfd08>



```
In [54]: df_classe_majoritaria = dadosRH[dadosRH.promovido==0]
df_classe_minoritaria = dadosRH[dadosRH.promovido==1]
```

```
In [61]: df_classe_majoritaria.shape
```

```
Out[61]: (50140, 11)
```

```
In [62]: df_classe_minoritaria.shape
```

```
Out[62]: (4668, 11)
```

```
In [85]: from sklearn.utils import resample
df_classe_minoritaria_unsampled = resample (df_classe_minoritaria,replace = True,
```

```
In [87]: dadosRH_balanceados = pd.concat ([df_classe_majoritaria, df_classe_minoritaria_u
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-87-6c9e5a38ca23> in <module>
----> 1 dadosRH_balanceados = pd.concat ([df_classe_majoritaria, df_classe_minoritaria_upsampled])
```

```
NameError: name 'df_classe_minoritaria_upsampled' is not defined
```

```
In [88]: dadosRH_balanceados.promovido.value_counts()
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-88-70c2d9a55c67> in <module>
----> 1 dadosRH_balanceados.promovido.value_counts()
```

```
NameError: name 'dadosRH_balanceados' is not defined
```

```
In [78]: dadosRH_balanceados.info()
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-78-49fe9403c544> in <module>
----> 1 dadosRH_balanceados.info()
```

```
NameError: name 'dadosRH_balanceados' is not defined
```

```
In [79]: sns.countplot(dadosRH_balanceados['promovido'])
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-79-84d70fddb6ac> in <module>
----> 1 sns.countplot(dadosRH_balanceados['promovido'])
```

```
NameError: name 'dadosRH_balanceados' is not defined
```

In []: