

Clique duas vezes (ou pressione "Enter") para editar

```
#Iniciando a utilização da biblioteca pandas
```

```
import pandas as pd
```

```
import seaborn as sns
```

```
orig_url='https://drive.google.com/file/d/1pkOP40FvztNt0pnWgKMG5iTXzRt4TW_N/view?usp=sharing'
```

```
import requests
```

```
from io import StringIO
```

```
file_id = orig_url.split('/')[-2]
```

```
dwn_url='https://drive.google.com/uc?export=download&id=' + file_id
```

```
url = requests.get(dwn_url).text
```

```
csv_raw = StringIO(url)
```

```
dfs = pd.read_csv(csv_raw)
```

```
# Quantas reclamações foram registradas em 2016, contidas nesta base?
```

```
dfs['i»¿AnoCalendario'].value_counts().head(1)
```

```
↳ 2016      203486  
   Name: i»¿AnoCalendario, dtype: int64
```

```
# Qual região houve mais reclamação? E qual houve menos?
```

```
testeMaior = dfs['Regiao'].value_counts().head(1)
```

```
testeMenor = dfs['Regiao'].value_counts().tail(1)
```

```
print(testeMaior)
```

```
print(testeMenor)
```

```
Nordeste      66411
```

```
Name: Regiao, dtype: int64
Norte    19202
Name: Regiao, dtype: int64
```

Qual a quantidade de reclamações realizadas por homens? E por mulheres?

```
dfs['SexoConsumidor'].value_counts().head(2)
```

```
F    107229
M     94446
Name: SexoConsumidor, dtype: int64
```

Em qual região os homens fizeram mais reclamações que as mulheres?

```
dfs_temp_ = dfs[['Regiao', 'SexoConsumidor']]
```

```
dfs_2 = dfs_temp_[dfs['SexoConsumidor']=='M'].value_counts()
```

```
dfs_2[1:2]
```

```
Regiao    SexoConsumidor
Nordeste  M                29065
dtype: int64
```

Quais assuntos apresentaram mais reclamações em 2016?

```
dfs[['CodigoAssunto', 'DescricaoAssunto', 'i»¿AnoCalendario']].value_counts().head()
```

CodigoAssunto	DescricaoAssunto	i»¿AnoCalendario	
101	Telefone (Convencional, Celular, Interfone, Etc.)	2016	21272
187	Telefonia Celular	2016	15102
54	Cartão de Crédito	2016	11774
186	Telefonia Fixa (Plano de Expansão / Compra e Venda / Locação)	2016	11056
53	Banco comercial	2016	9274

```
dtype: int64
```

Analisando o nome fantasia, determine qual empresa recebeu mais reclamações em 2016.

```
dfs[['strNomeFantasia', 'i»¿AnoCalendario']].value_counts().head(1)
```

```
strNomeFantasia  i»¿AnoCalendario
OI                2016                2732
dtype: int64
```

Qual empresa possui um maior percentual de resolução de reclamações?

```
dfs[['Atendida', 'strNomeFantasia']].value_counts().head(1)
```

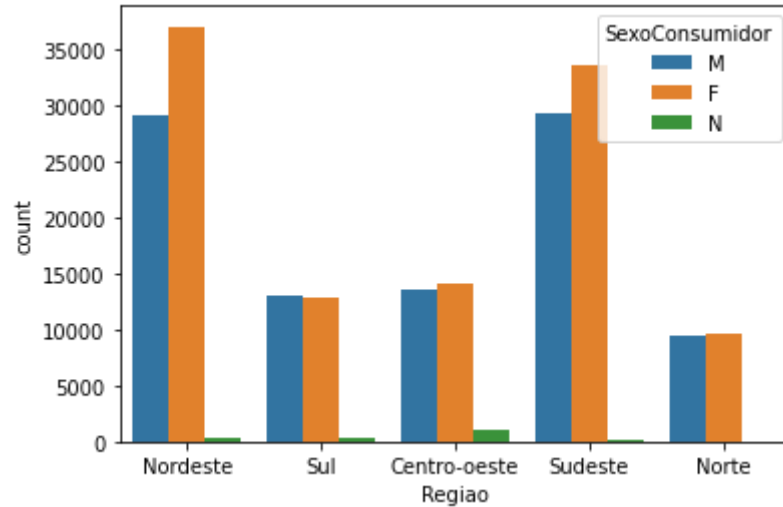
```
Atendida  strNomeFantasia
S         OI                2098
dtype: int64
```

Monte um gráfico mostrando a quantidade de reclamações por região, separando as mesmas por sexo.

```
dfs_temp = dfs[['Regiao', 'SexoConsumidor']]
```

```
sns.countplot(x='Regiao', hue='SexoConsumidor', data=dfs_temp)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fe5433dc510>

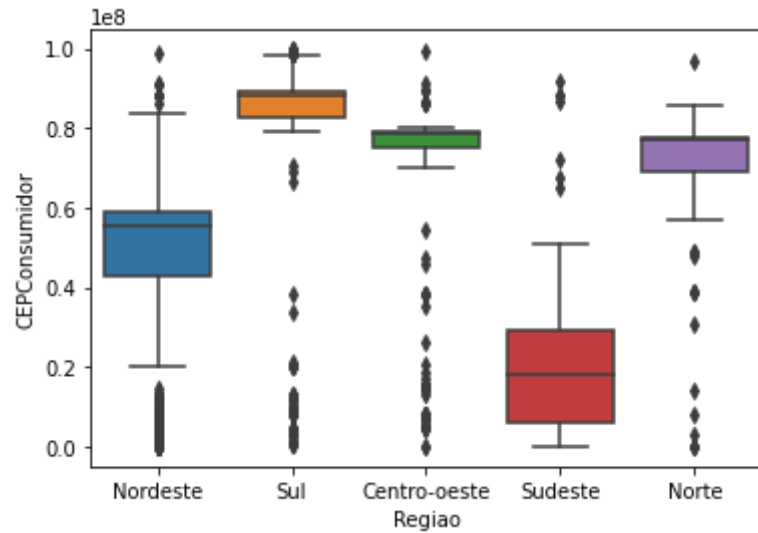


Elabore um gráfico de linhas com a quantidade de reclamações por mês.

Este gráfico lembra alguma distribuição estatística?

```
# Elabore um gráfico boxplot mostrando a quantidade de reclamações por região.  
sns.boxplot(x=dfs['Regiao'],y=dfs['CEPConsumidor'], data=dfs)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe548391290>
```



```
# É possível afirmar que existe correlação entre o número de reclamações  
# E o número de habitantes por Estado? Elabore um gráfico de dispersão e  
# calcule o índice de correlação destes dois fatores.
```

