

VISUALIZAÇÃO DE DADOS EM PYTHON

Vanessa Cadan Scheffer

0

Ver anotações

ANÁLISE E APRESENTAÇÃO DE DADOS

As bibliotecas pandas, matplotlib e seaborn podem ser utilizadas para o carregamento dos dados e a geração dos gráficos.



Fonte: Shutterstock.

Deseja ouvir este material?

Áudio disponível no material digital.

DESAFIO

Como desenvolvedor em uma empresa de consultoria de software, você foi alocado em um projeto para uma empresa de telecomunicações. Essa empresa tem interesse em habilitar um novo serviço, mas antes precisa entender qual a

disponibilidade dos satélites autorizados a operar no Brasil. Para a primeira sprint (período de 15 dias de trabalho), você foi encarregado de apresentar, uma análise preliminar da situação dos satélites.

Nessa primeira entrega, você deve apresentar a comparação da quantidade de satélites que são brasileiros, dos que são estrangeiros. Dentre os satélites brasileiros, você deve discriminar a quantidade de cada operadora comercial, bem como a quantidade de satélites operando em cada banda. As mesmas análises devem ser feitas para os satélites que pertencem a outros países.

o
Ver anotações

Onde esses dados podem ser encontrados? Qual a melhor forma de apresentar os resultados, basta levar os números? Qual biblioteca pode ser usada para resolver o desafio?

RESOLUÇÃO

Um dos grandes desafios nessa primeira entrega é encontrar uma fonte confiável de dados.

No endereço <https://www.dados.gov.br/dataset>, existe uma categoria específica para esse tipo de informação: Agência Nacional de Telecomunicações - Anatel. Dentro dessa categoria encontramos um arquivo delimitado (csv) com a relação de satélites autorizados a operar no Brasil: <https://www.dados.gov.br/dataset/relacao-de-satelites-geoestacionarios-autorizados-a-operar-no-brasil>, basta clicar no recurso e fazer download para a pasta do projeto.

Agora que identificamos uma fonte confiável podemos usar as bibliotecas pandas, matplotlib e seaborn para carregar os dados e gerar gráficos que contemplem as informações solicitadas.

```
In [19]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Vamos carregar os dados em um DataFrame pandas, chamado df_satelites. Os dados possuem como delimitador ";", logo é preciso informar ao método read_csv esse parâmetro. Também é preciso garantir que linhas duplicadas sejam removidas (linha 2) e para ter os índices variando de 0 até N, vamos resetar (linha 3).

Ver anotações

Como resultado temos um DF com 68 linhas e 7 colunas.

In [20]:

```
df_satelites = pd.read_csv('satelites_operando_comercialmente.csv',
sep=';')
df_satelites.drop_duplicates(inplace=True)
df_satelites.reset_index(drop=True, inplace=True)

print(df_satelites.info())
df_satelites.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68 entries, 0 to 67
Data columns (total 7 columns):
Satelite operando      68 non-null object
Órbita                  68 non-null object
Bandas                 68 non-null object
Status do Satélite     68 non-null object
Pos. Orbital           68 non-null object
Direito                68 non-null object
Operadora Comercial    68 non-null object
dtypes: object(7)
memory usage: 3.8+ KB
None
```

Out[20]:

	Satelite operando	Órbita	Bandas	Status do Satélite	Pos. Orbital	Direito	Operadora Comercial
0	EUTELSAT 65 West A	Satélite Geoestacionário (GEO)	C (AP30B), Ku (AP30B), Ka	operação comercial	65°O	Brasileiro	EUTELSAT DO BRASIL LTDA
1	HISPASAT 74W-1	Satélite Geoestacionário (GEO)	Ku (AP30/30A)	operação comercial	74°O	Brasileiro	HISPAMAR SATELITES S.A.
2	AL YAH 3	Satélite Geoestacionário (GEO)	Ka	operação comercial	20°O	Brasileiro	YAH TELECOMUNICAÇÕES LTDA
3	SGDC	Satélite Geoestacionário (GEO)	Ka	operação comercial	75°O	Brasileiro	TELECOMUNICACOES BRASILEIRAS SA TELEBRAS
4	SES-14	Satélite Geoestacionário (GEO)	C (não planejada), Ku (não planejada), Ka	operação comercial	47,5°O	Brasileiro	SES DTH DO BRASIL LTDA

Agora vamos criar um gráfico que faz a contagem e visualmente, faz a comparação entre a quantidade de satélites brasileiros e estrangeiros. Podemos usar o `countplot()`, passando como parâmetros o DF e a coluna 'Direito', como variável categórica a ser contada. Também podemos usar os recursos da biblioteca `matplotlib` para configurar o tamanho da figura e dos textos nos eixos.

In [21]:

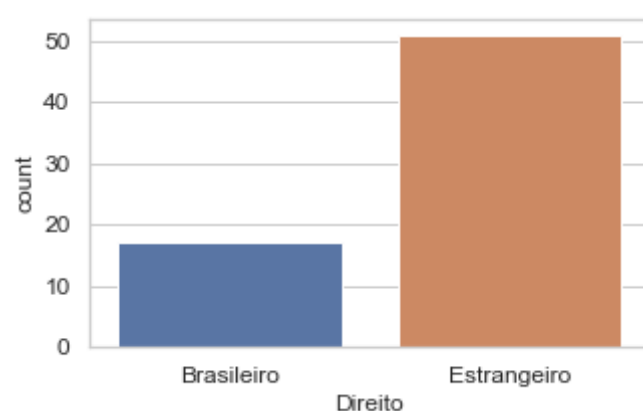
```
# Quantos satélites são brasileiros e quantos são estrangeiro?

plt.figure(figsize=(5,3))
plt.tick_params(labelsize=12)
sns.countplot(data=df_satelites, x='Direito')
```

Ver anotações

Out[21]:

<matplotlib.axes._subplots.AxesSubplot at 0x1cf35cfdbe0>



Agora vamos extrair as informações sobre os satélites brasileiros. Para facilitar, vamos criar um novo DataFrame aplicando um filtro. Veja na linha 13, que o DF `df_satelites_brasileiros`, será um filtro do DF `df_satelites` onde somente os brasileiros estarão presentes. Agora, podemos usar o `countplot` no `df_satelites_brasileiros` para contar quantos satélites cada operadora comercial no Brasil possui (linha 8). Como o nome das operadoras é longo, vamos pedir para exibir na vertical, por isso configuramos a rotação do "xticks" na linha 6. Na linha 7 configuramos o tamanho dos textos nos eixos.

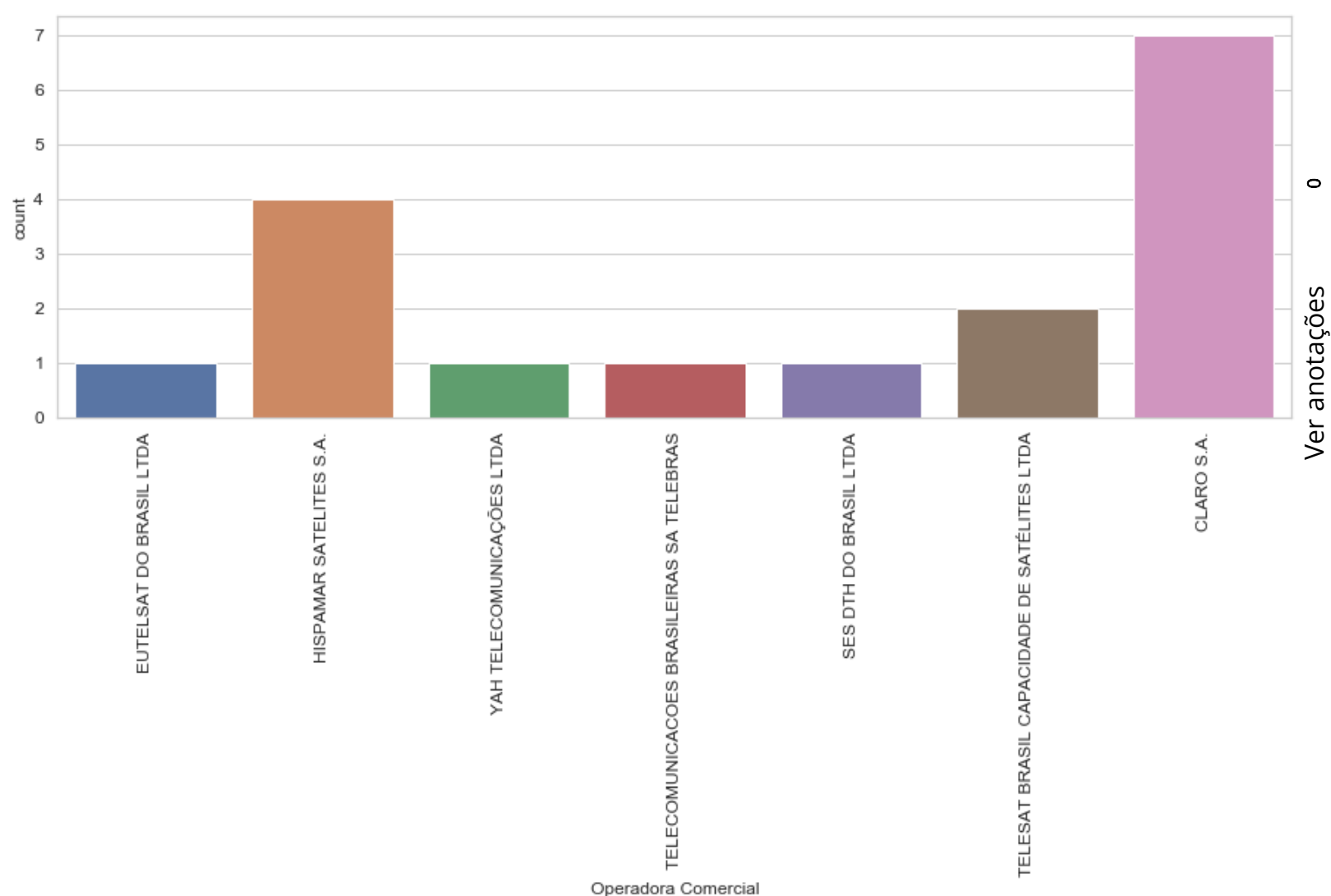
In [22]:

```
# quantos satélites cada operadora brasileira possui operando?

df_satelites_brasileiros = df_satelites.loc[df_satelites['Direito']
== 'Brasileiro']

plt.figure(figsize=(15,5))
plt.xticks(rotation=90)
plt.tick_params(labelsize=12)
sns.countplot(data=df_satelites_brasileiros, x='Operadora
Comercial')
```

Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x1cf36eb4e10>

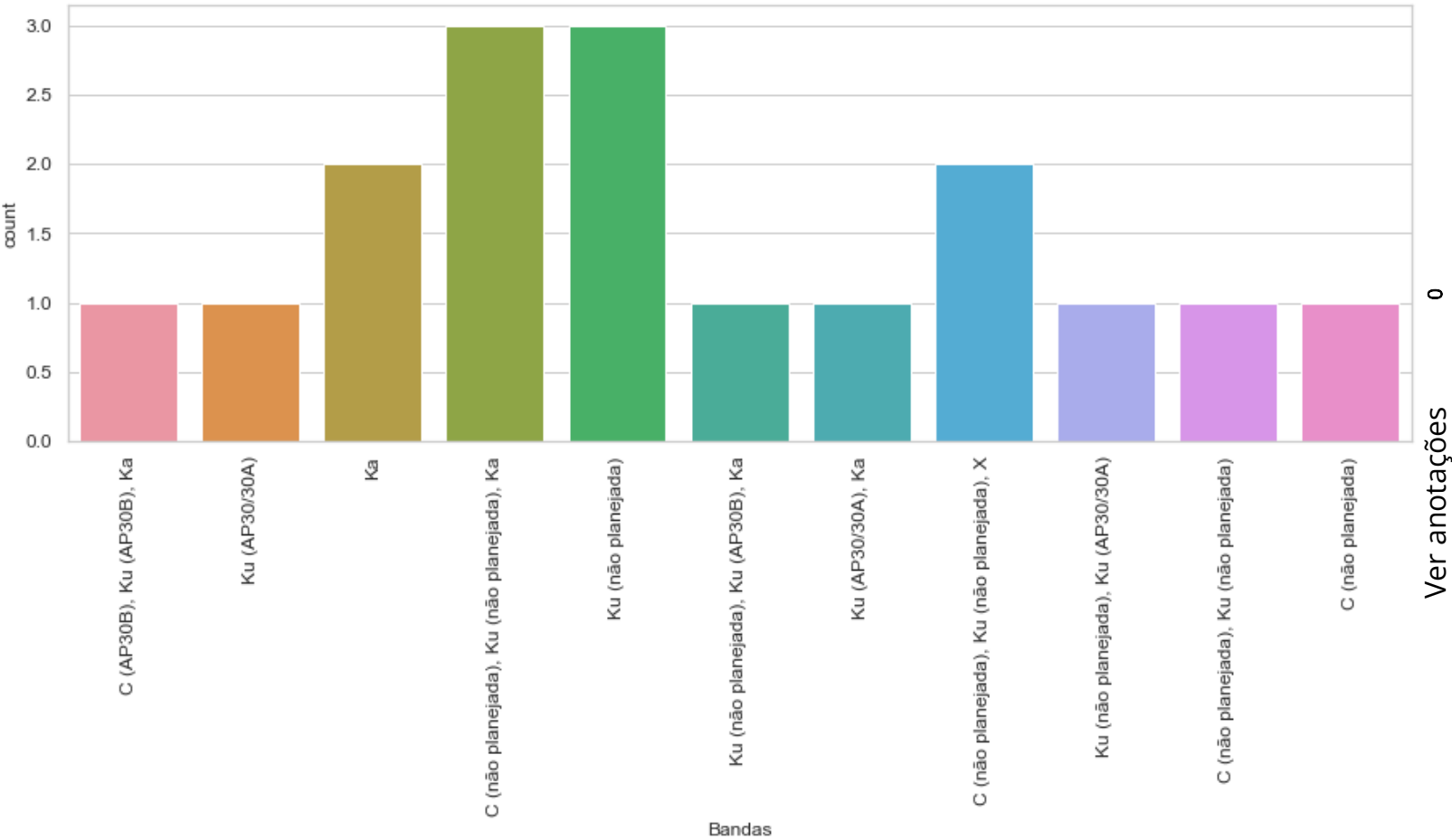


Para saber quantos satélites brasileiros estão operando em cada banda, vamos usar o countplot, passando como parâmetro o df_satelites_brasileiros e a coluna 'Bandas'. Novamente foi necessário configurar o texto nos eixos.

In [23]: *# Quantos satélites brasileiros estão operando em cada banda?*

```
plt.figure(figsize=(15,5))
plt.xticks(rotation=90)
plt.tick_params(labelsize=12)
sns.countplot(data=df_satelites_brasileiros, x='Bandas')
```

Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x1cf33aeb588>



Agora vamos repetir os mesmos processos para os satélites estrangeiros, começando pela criação de um DataFrame que contenha somente as informações sobre eles (linha 3). Esse primeiro gráfico mostra quantos satélites cada operadora estrangeira possui em operação.

In [24]:

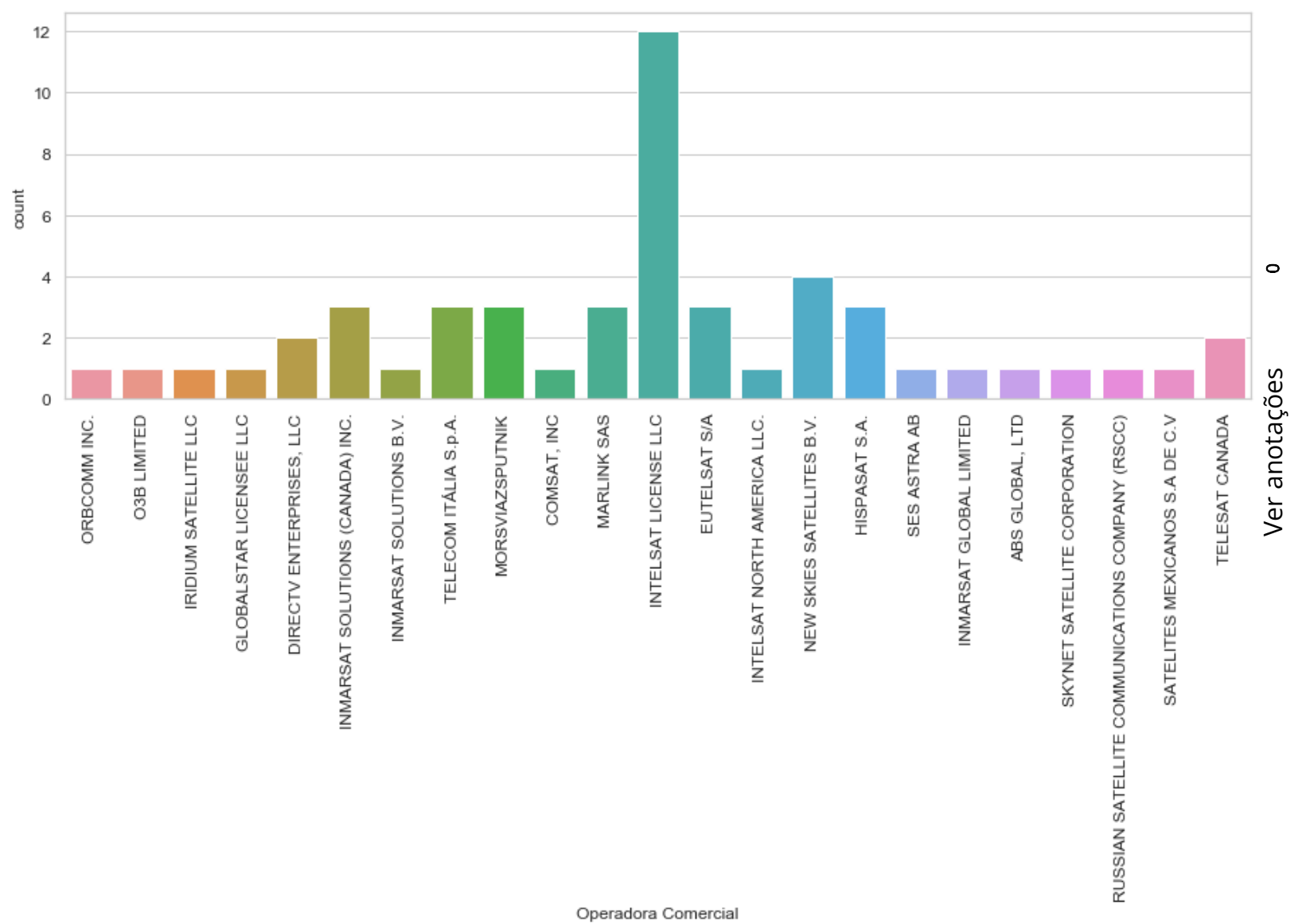
```
# Quantos satélites cada operadora estrangeira possui operando?

df_satelites_estrangeiros = df_satelites.loc[df_satelites['Direito']
== 'Estrangeiro']

plt.figure(figsize=(15,5))
plt.xticks(rotation=90)
plt.tick_params(labelsize=12)
sns.countplot(data=df_satelites_estrangeiros, x='Operadora
Comercial')
```

Out[24]:

<matplotlib.axes._subplots.AxesSubplot at 0x1cf36f5ab70>

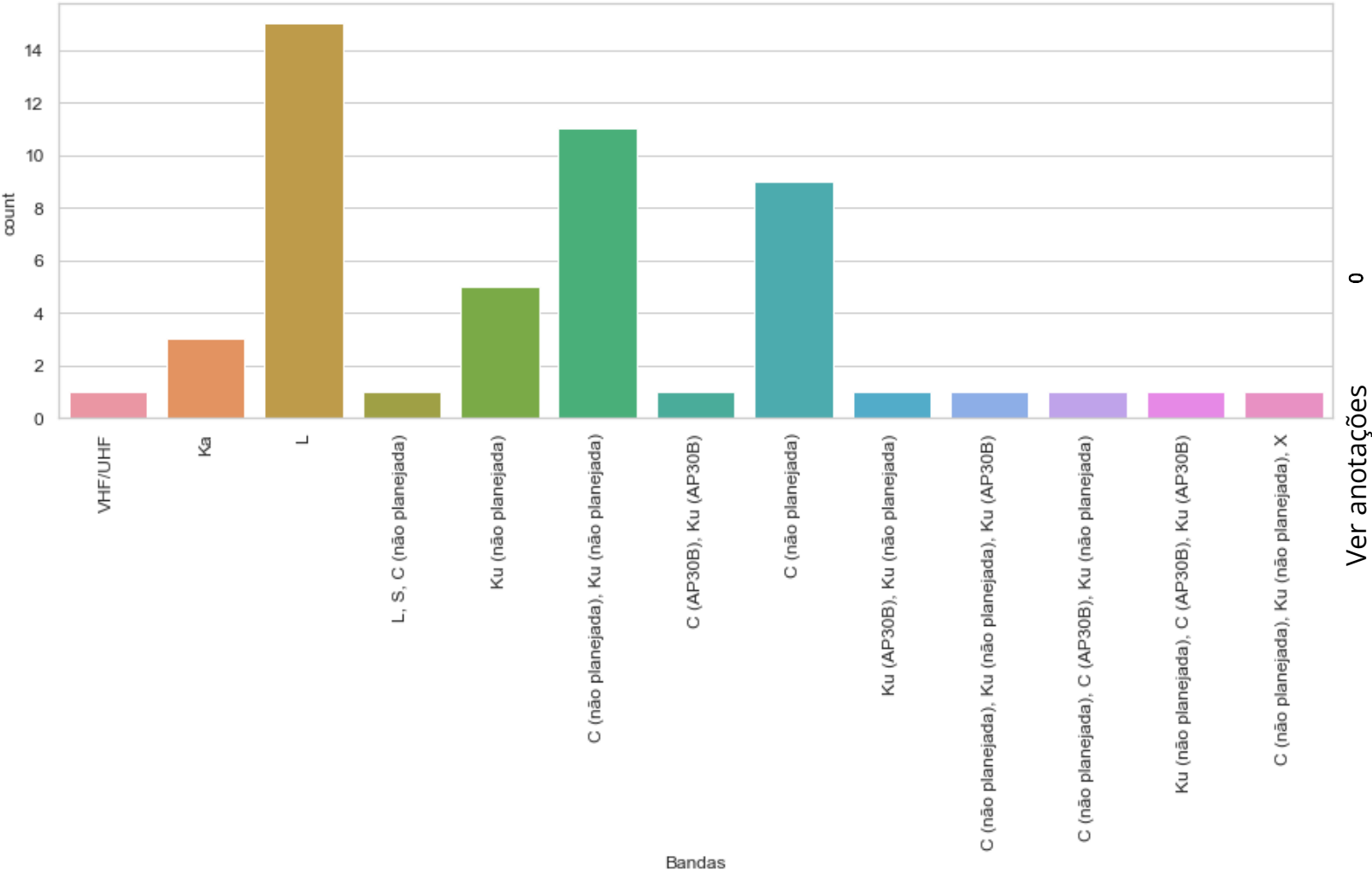


Agora vamos plotar quantos satélites estrangeiros estão operando em cada banda.

In [25]: *# Quantos satélites brasileiros estão operando em cada banda?*

```
plt.figure(figsize=(15,5))
plt.xticks(rotation=90)
plt.tick_params(labelsize=12)
sns.countplot(data=df_satelites_estrangeiros, x='Bandas')
```

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x1cf36f4da20>



Com essas informações o cliente começará a ter argumentos para a escolha de uma operadora e banda que deseja contratar. Que tal personalizar um pouco mais o tamanho das legendas, criar títulos para os gráficos?

DESAFIO DA INTERNET

Ganhar habilidade em programação exige estudo e treino (muito treino). Acesse o endereço <https://www.kaggle.com/datasets>, faça seu cadastro e escolha uma base de dados para treinar e desenvolver seu conhecimento sobre visualização de dados.