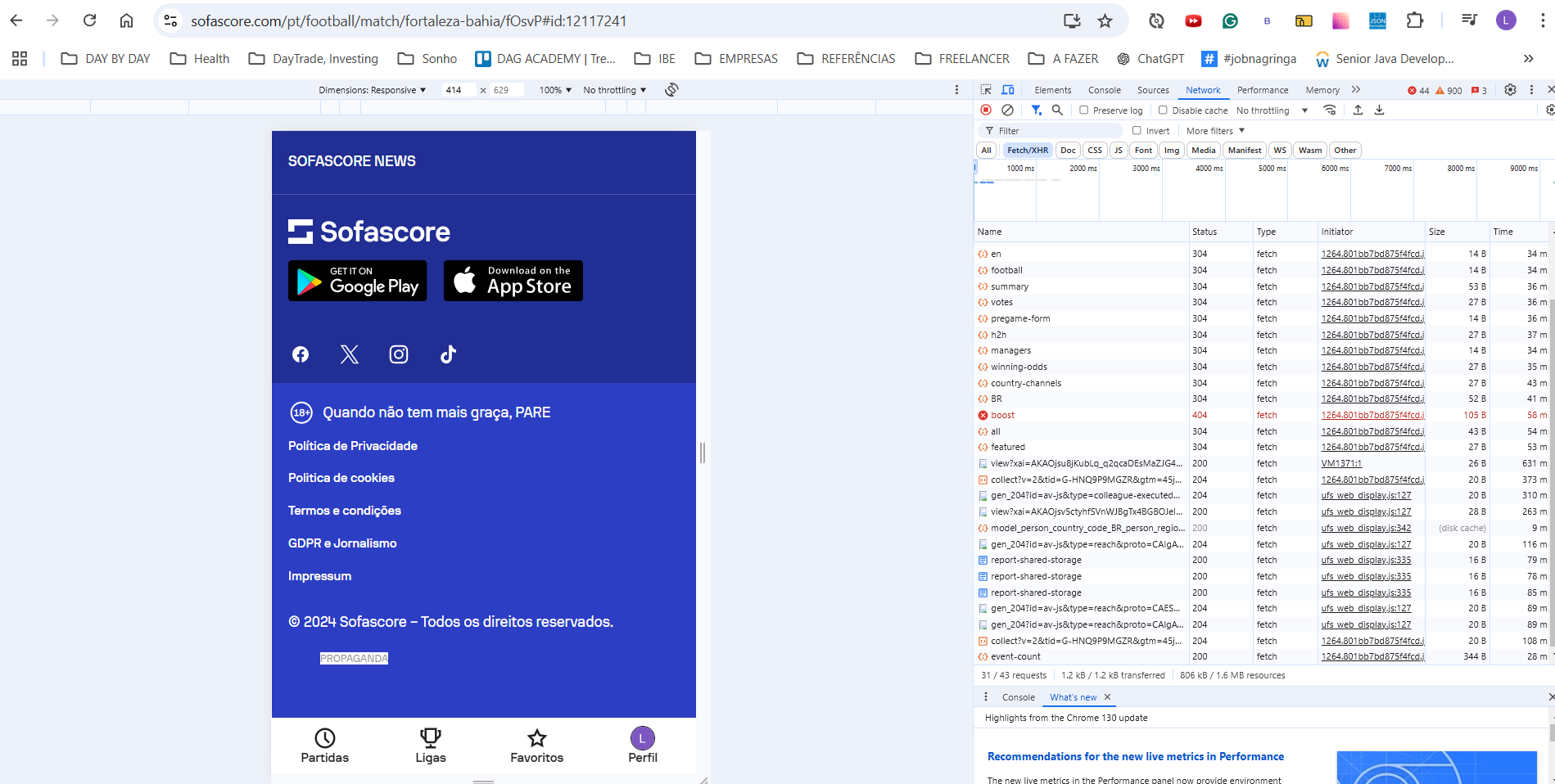
**Como Capturar Dados de Futebol, NBA e Mais com Python e Web Scraping – Tutorial Completo!**

<https://www.sofascore.com/pt/>

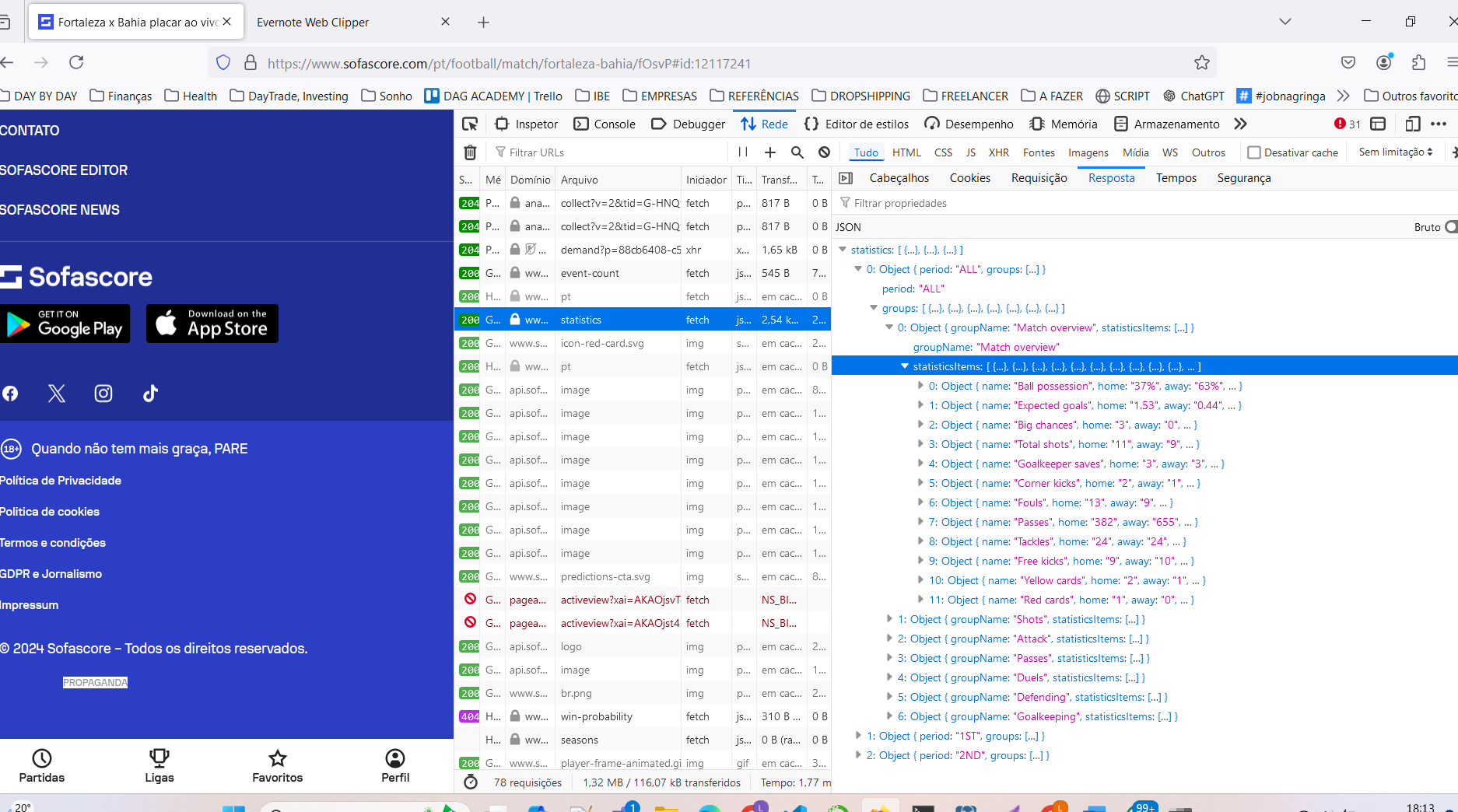
<https://www.sofascore.com/pt/torneio/futebol/brazil/brasileirao-serie-a/325#id:58766>

<https://www.sofascore.com/pt/time/futebol/fortaleza/2020>

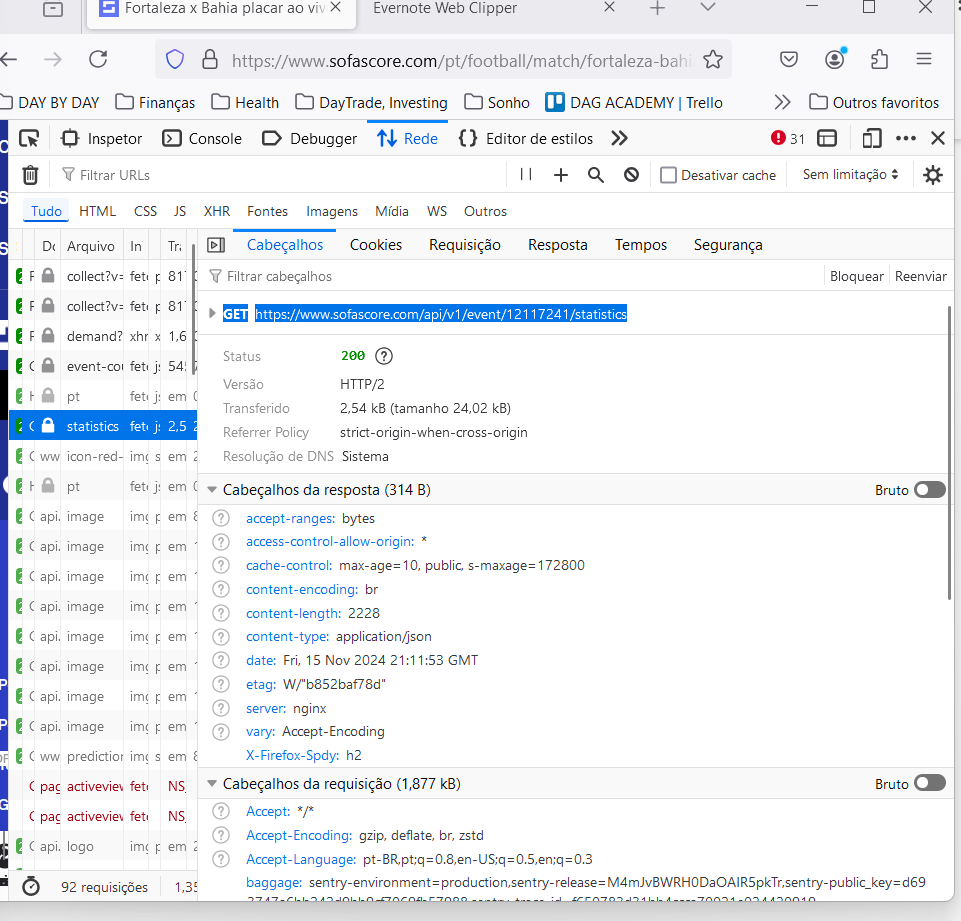
network shr



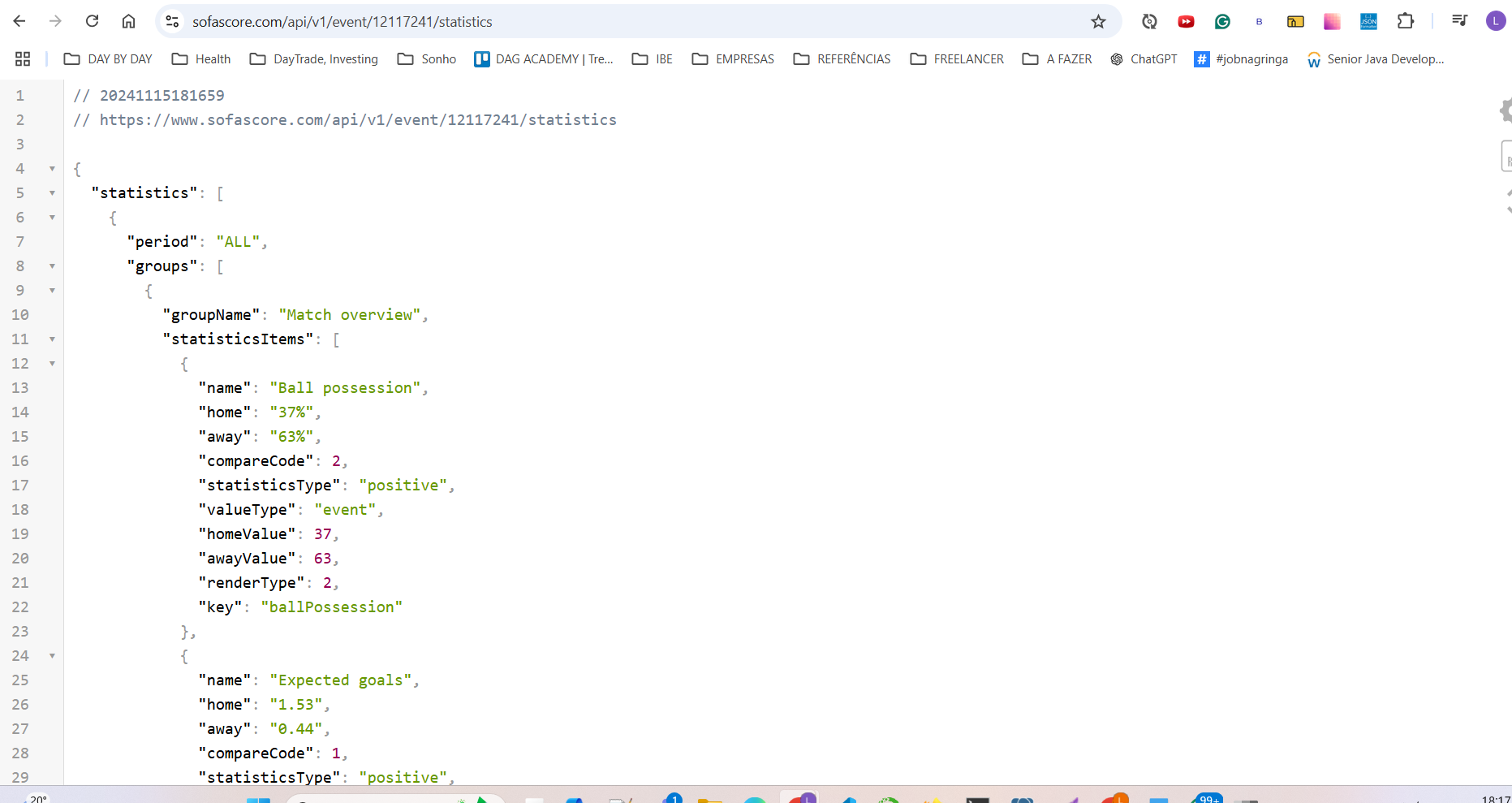
Abrindo no firefoz-> inspecionar -> rede-> arquivos -> statitics -> statisticsItems



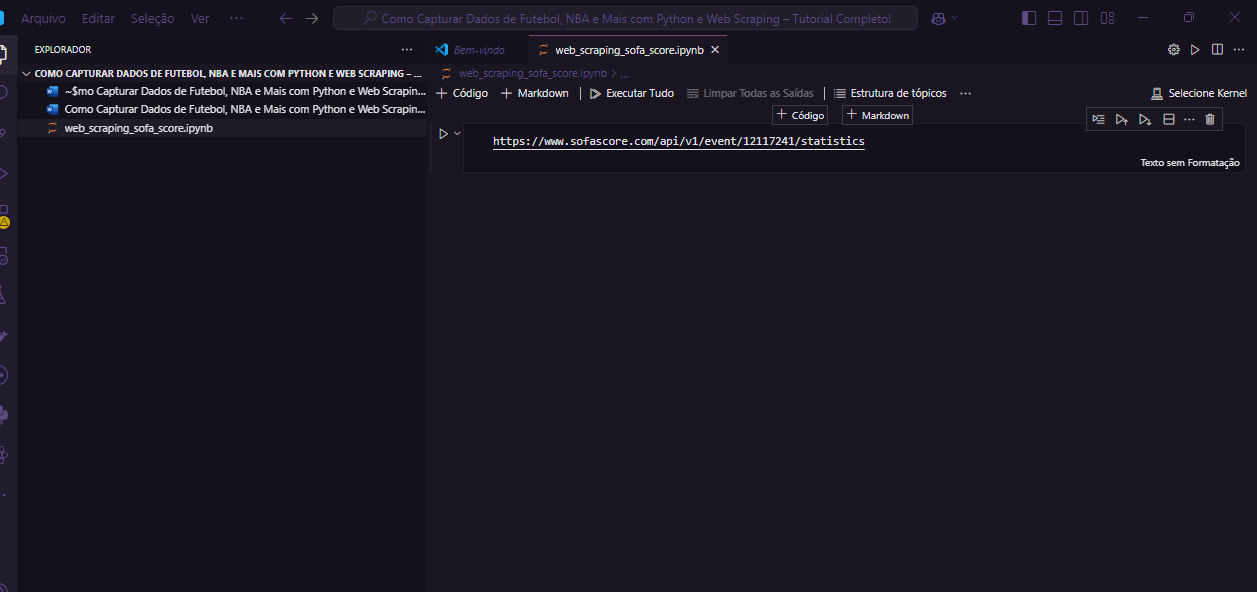
A achamda da url podemos ver emcabeçalho



|  |  |
| --- | --- |
| GET | https://www.sofascore.com/api/v1/event/12117241/statistics |

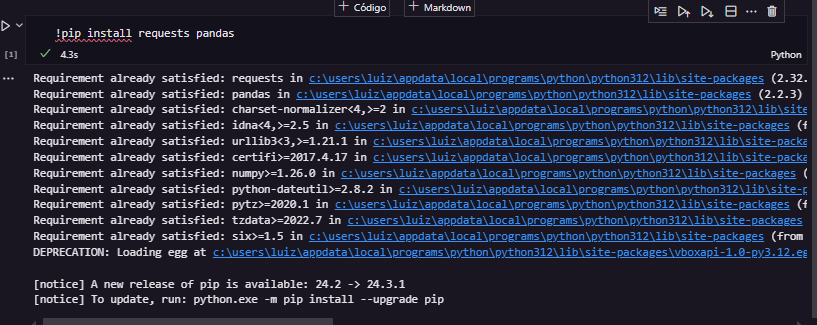


Vamos criar um notebook ou arquivo .py



Vamos instalar o requests e pandas

**!pip install requests pandas**



Vamos importar o requests e pandas

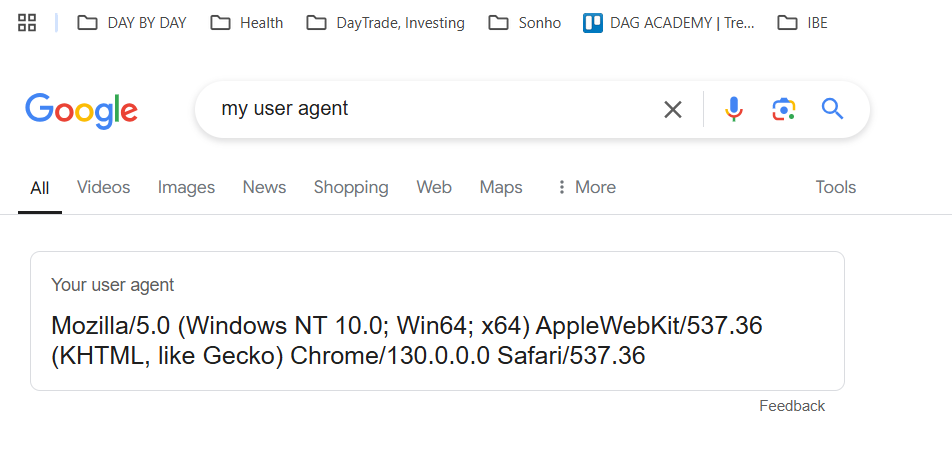
**import requests**

**import pandas as pd**

vamos colocar a url numa variável

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

vamos descrobri o nosso user agente



Vamos cria rum header com esse user agetn

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

vamos fazer uma solicitação get para a api

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

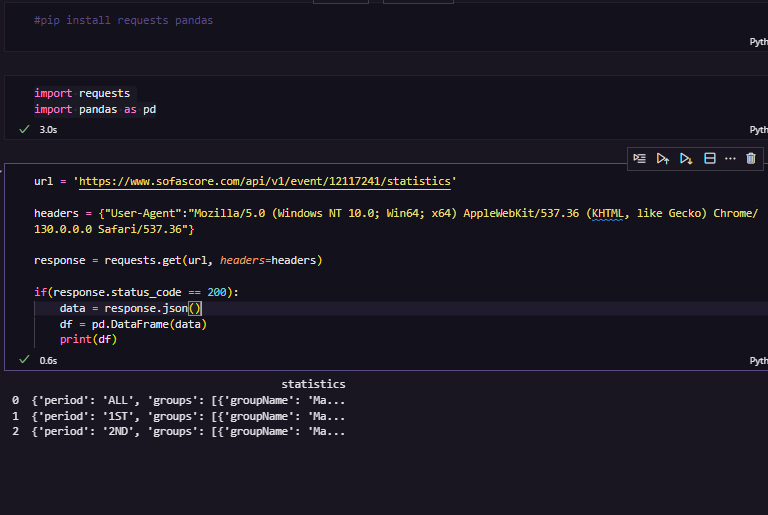
**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**df = pd.DataFrame(data)**

**print(df)**

****

vamos percorrer o data para pegar as info

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**print(match\_stat)**

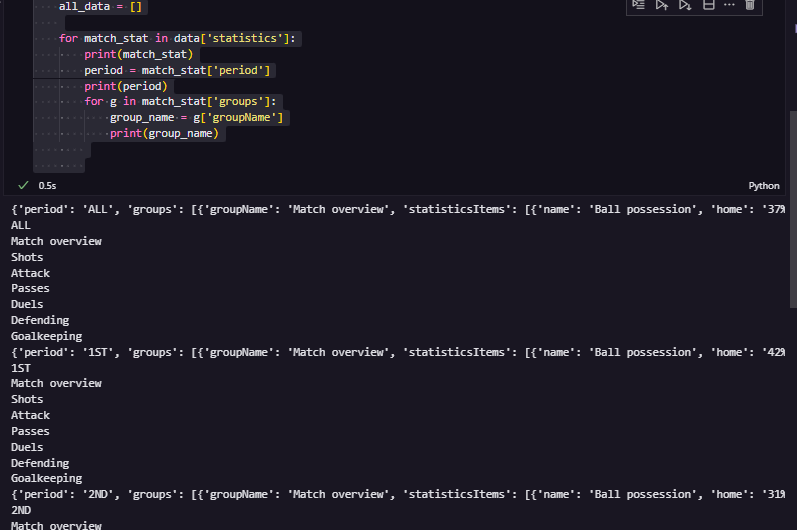
**period = match\_stat['period']**

**print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**print(group\_name)**

****

Vamos pegar os statisticsItems de cada grupo

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**print(match\_stat)**

**period = match\_stat['period']**

**print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**print(group\_name)**

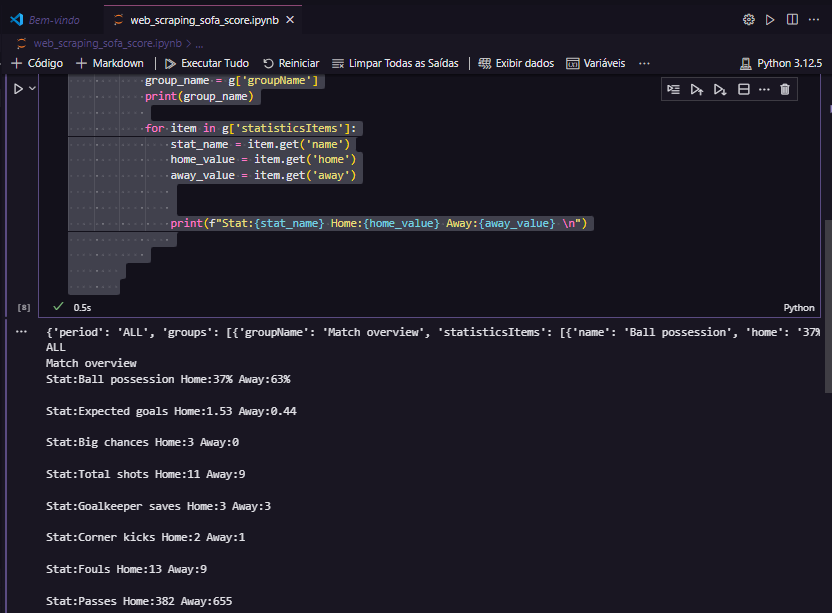
**for item in g['statisticsItems']:**

**stat\_name = item.get('name')**

**home\_value = item.get('home')**

**away\_value = item.get('away')**

**print(f"Stat:{stat\_name} Home:{home\_value} Away:{away\_value} \n")**

****

vamos ppular po all\_data com um dicionário com as informações que colhemos

Vamos popular um array vom todas info relevantes

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**#print(match\_stat)**

**period = match\_stat['period']**

**#print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**# print(group\_name)**

**for item in g['statisticsItems']:**

**stat\_name = item.get('name')**

**home\_value = item.get('home')**

**away\_value = item.get('away')**

**# print(f"Stat:{stat\_name} Home:{home\_value} Away:{away\_value} \n")**

**all\_data.append(**

**{**

**'Period': period,**

**'Group': group\_name,**

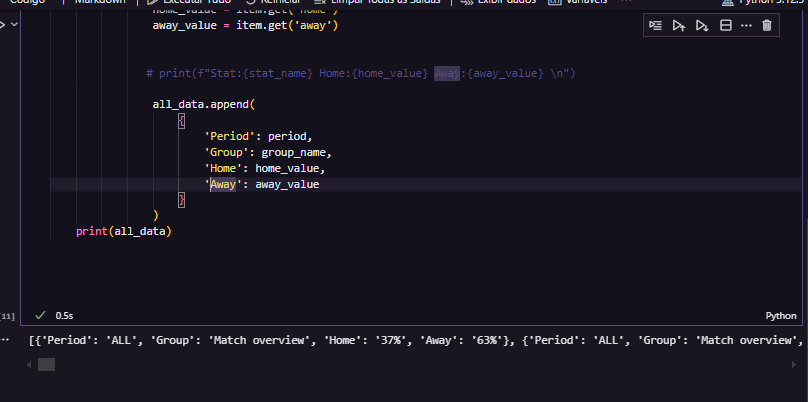
**'Home': home\_value,**

**'Away': away\_value**

**}**

**)**

**print(all\_data)**

****

Vamos construit um data frame com esses dados

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**#print(match\_stat)**

**period = match\_stat['period']**

**#print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**# print(group\_name)**

**for item in g['statisticsItems']:**

**stat\_name = item.get('name')**

**home\_value = item.get('home')**

**away\_value = item.get('away')**

**# print(f"Stat:{stat\_name} Home:{home\_value} Away:{away\_value} \n")**

**all\_data.append(**

**{**

**'Period': period,**

**'Group': group\_name,**

**'Statistic': stat\_name,**

**'Home': home\_value,**

**'Away': away\_value**

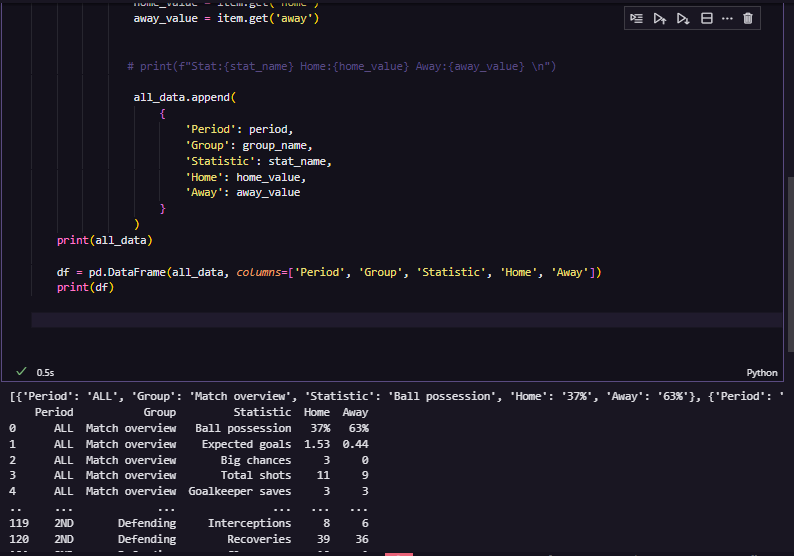
**}**

**)**

**print(all\_data)**

**df = pd.DataFrame(all\_data, *columns*=['Period', 'Group', 'Statistic', 'Home', 'Away'])**

**print(df)**

****

Podemos criar o df usando um dicionário sem ter que passar quais saõ as colunas

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**#print(match\_stat)**

**period = match\_stat['period']**

**#print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**# print(group\_name)**

**for item in g['statisticsItems']:**

**stat\_name = item.get('name')**

**home\_value = item.get('home')**

**away\_value = item.get('away')**

**# print(f"Stat:{stat\_name} Home:{home\_value} Away:{away\_value} \n")**

**all\_data.append(**

**{**

**'Period': period,**

**'Group': group\_name,**

**'Statistic': stat\_name,**

**'Home': home\_value,**

**'Away': away\_value**

**}**

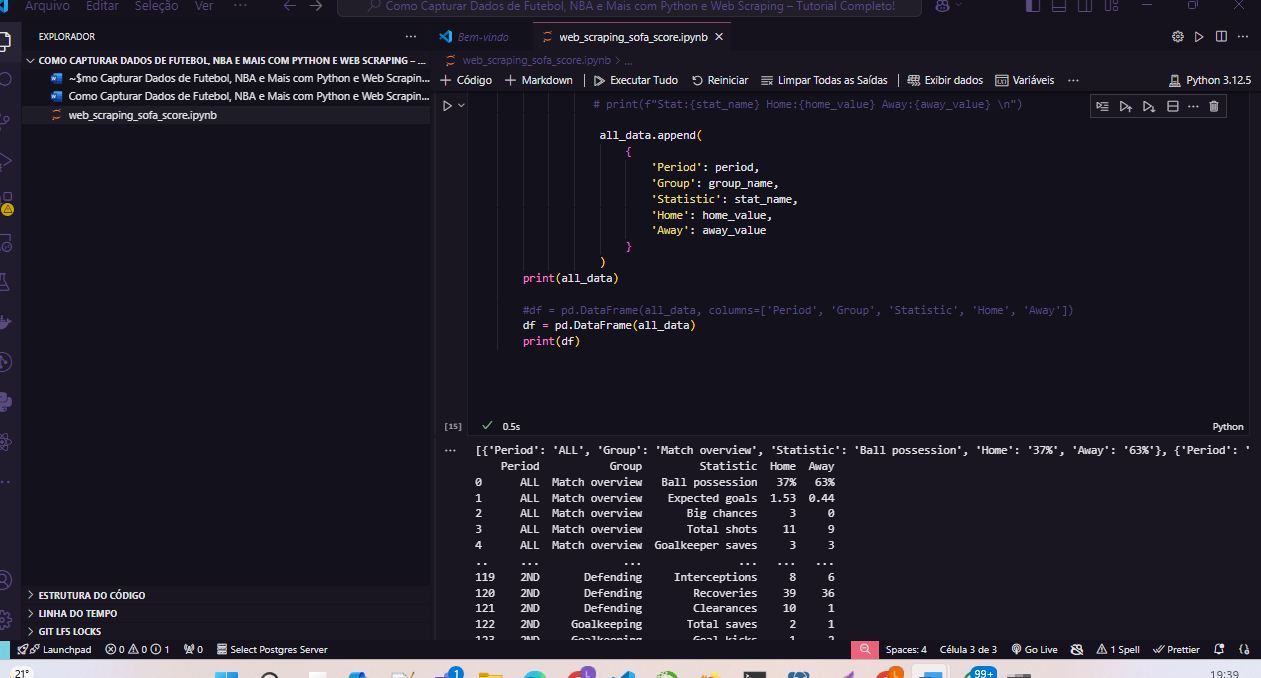
**)**

**print(all\_data)**

**#df = pd.DataFrame(all\_data, columns=['Period', 'Group', 'Statistic', 'Home', 'Away'])**

**df = pd.DataFrame(all\_data)**

**print(df)**

****

Vamos transformar esse dados para excel

**url = 'https://www.sofascore.com/api/v1/event/12117241/statistics'**

**headers = {"User-Agent":"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36"}**

**response = requests.get(url, *headers*=headers)**

**if(response.status\_code == 200):**

**data = response.json()**

**all\_data = []**

**for match\_stat in data['statistics']:**

**#print(match\_stat)**

**period = match\_stat['period']**

**#print(period)**

**for g in match\_stat['groups']:**

**group\_name = g['groupName']**

**# print(group\_name)**

**for item in g['statisticsItems']:**

**stat\_name = item.get('name')**

**home\_value = item.get('home')**

**away\_value = item.get('away')**

**# print(f"Stat:{stat\_name} Home:{home\_value} Away:{away\_value} \n")**

**all\_data.append(**

**{**

**'Period': period,**

**'Group': group\_name,**

**'Statistic': stat\_name,**

**'Home': home\_value,**

**'Away': away\_value**

**}**

**)**

**print(all\_data)**

**#df = pd.DataFrame(all\_data, columns=['Period', 'Group', 'Statistic', 'Home', 'Away'])**

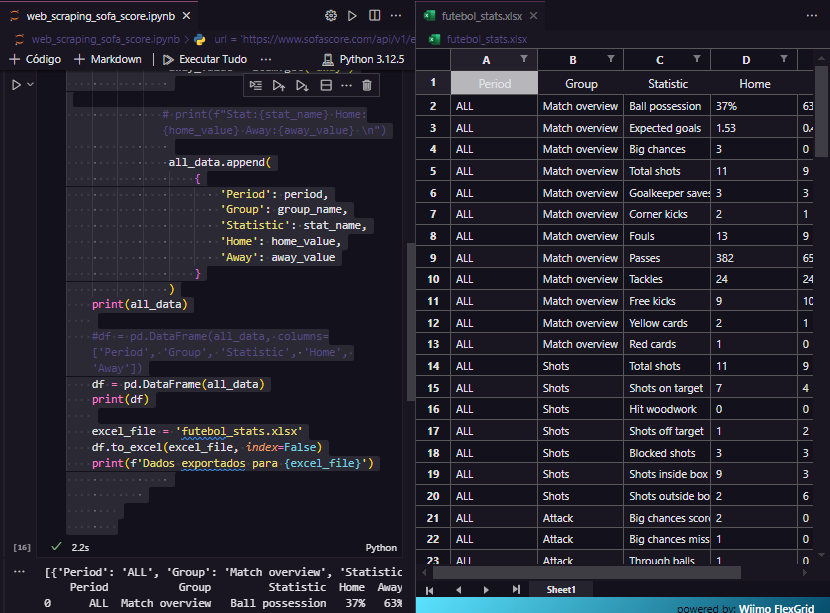
**df = pd.DataFrame(all\_data)**

**print(df)**

**excel\_file = 'futebol\_stats.xlsx'**

**df.to\_excel(excel\_file, *index*=False)**

**print(f'Dados exportados para {excel\_file}')**

****