import pandas as pd

import matplotlib.pyplot as plt

from bs4 import BeautifulSoup

from urllib.request import urlopen

# url site B3:"http://www.b3.com.br/pt\_br/market-data-e-indices/indices/indices-amplos/

# indice-ibovespa-ibovespa-composicao-da-carteira.htm"

# no entanto a referencia com os dados ainda é o site bmfbovespa a seguir

url = urlopen("http://bvmf.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=IBOV&idioma=pt-br")

# guarda os dados do site

soup = BeautifulSoup(url.read(), 'html.parser')

# descrição e data da tabela, encontra o id inspecionando o html da página

span = soup.find("span", id="ctl00\_contentPlaceHolderConteudo\_lblTitulo")

# pega os dados no corpo principal e separa os itens da tabela (começam com tr)

tr = soup.tbody.find\_all('tr')

# pega as linhas da tabela

dados = []

for t in tr:

dados.append(t.text.split('\n'))

# converte em dataFrame

df = pd.DataFrame(dados)

# monta o cabecalho no dataframe

# primeiro apaga as colunas desnecessárias

df.drop([0, 1, 3, 5, 7, 9, 11, 12], axis=1, inplace=True)

# segundo dá nome às colunas

cabecalho = ('Codigo', 'Nome', 'Tipo', 'Numero acoes', 'Participacao')

df.columns = cabecalho

# algumas ações tem outros codigos além de somente ON e PN, abaixo ficamos só com as 2 primeiras letras ON, PN ou UN

df['Tipo'] = list(map(lambda x: x[:2], df['Tipo']))

# salva a tabela em arquivo csv

df.to\_csv("C:/Users/Documents/Ibovespa.csv",index=False,sep=';')

# exibe o cabecalho e as linhas do dataframe

print('\n'+span.text+'\n')

linhas = len(df)+1

print(df.head(linhas))

# exibe a linha 4 e suas colunas separadamente só de exemplo

# i = 4

# for j in range(len(df.columns)):

# print(str(j)+") "+cabecalho[j]+": "+df.iloc[i, j])

# pega as 2 primeiras letras da linha 4 coluna 2 do dataframe

# print(df.iloc[4,2][:2])

# mostra gráfico de barras horizontal do número de ações por tipo: ON, PN, ou UNit

tipos = df.groupby('Tipo').count().iloc[:, 0].sort\_values()

tipos.plot(kind='barh', title='IBOV')

plt.ylabel('Número de ações', fontsize=10)

plt.xlabel('Tipos de ações', fontsize=10)

plt.title('Ações por tipo no IBOVESPA ('+str(len(df))+' no total)')

# coloca os valores no final da barra

for indice, valor in enumerate(tipos):

plt.text(valor, indice, str(valor))

# exibe o gráfico

plt.show()