Mercado Financeiro: Como coletar dados de BTC diários com Python, Selenium e Pandas

[](https://medium.com/@joaolggross?source=post_page-----59ab0b3b39f0----------------------)

[João Gross](https://medium.com/@joaolggross?source=post_page-----59ab0b3b39f0----------------------)

[Jul 30](https://medium.com/@joaolggross/mercado-financeiro-como-coletar-dados-de-btc-di%C3%A1rios-com-python-selenium-e-pandas-59ab0b3b39f0?source=post_page-----59ab0b3b39f0----------------------) · 5 min read

Uma forma simples e rápida de obter informações da Web

Image for post



Photo by [Carlos Muza](https://unsplash.com/@kmuza?utm_source=medium&utm_medium=referral) on [Unsplash](https://unsplash.com/?utm_source=medium&utm_medium=referral)

Atualmente qualquer investidor consegue encontrar na Internet uma vasta gama de informações disponibilizadas gratuitamente. De posse das URLs dos diferentes websites o investidor consegue verificar informações fundamentalistas sobre ações e FIIs, ver os rendimentos históricos e futuros desses ativos, consultar opções (derivativos de ações) disponíveis para negociação, taxas de aluguel de ações, dentre outras tantas.

Porém a análise dessas informações é restrita ao modo como o website as disponibiliza. Ter essas informações em uma estrutura de dados e poder analisá-las como se quer pode ser bastante interessante.

Nestes artigo explico como realizar a coleta diária de BTC (aluguel de ações) do BTG Pactual. A coleta é feita com um script Python que utiliza o pacote Selenium com o webdriver do Chrome para a coleta de dados, e o pacote Pandas para o armazenamento e manipulação das informações.

**Configurando o ambiente de desenvolvimento**

Antes de começarmos, é importante que você tenha instalado no seu computador o Python 3, os pacotes Selenium e Pandas, bem como o Chromedriver.

Abaixo compartilho três tutoriais detalhados que explicam passo a passo como deixar o seu ambiente de desenvolvimento pronto para que possamos executar os scripts. Os tutoriais são focados no Windows 10, mas o procedimento para Linux e Mac é bastante semelhante.

**[Como instalar o Python 3 no Windows 10](https://www.linkedin.com/pulse/como-instalar-o-python-3-windows-10-jo%C3%A3o-gross/" \t "_blank)**

[O procedimento de instalação do Python 3 no Windows 10 é um pouco mais complicado do que instalar no Linux, porém com…](https://www.linkedin.com/pulse/como-instalar-o-python-3-windows-10-jo%C3%A3o-gross/" \t "_blank)

[www.linkedin.com](https://www.linkedin.com/pulse/como-instalar-o-python-3-windows-10-jo%C3%A3o-gross/" \t "_blank)

**[Como instalar o Chromedriver no Windows 10](https://www.linkedin.com/pulse/como-instalar-o-chromedriver-windows-10-jo%C3%A3o-gross/" \t "_blank)**

[O Chromedriver é um recursos de software utilizado por pacotes de programação com o Selenium para acessar URLs em um…](https://www.linkedin.com/pulse/como-instalar-o-chromedriver-windows-10-jo%C3%A3o-gross/" \t "_blank)

[www.linkedin.com](https://www.linkedin.com/pulse/como-instalar-o-chromedriver-windows-10-jo%C3%A3o-gross/" \t "_blank)

**[Como instalar o Selenium webdriver no Windows 10](https://www.linkedin.com/pulse/como-instalar-o-selenium-webdriver-windows-10-jo%C3%A3o-gross/" \t "_blank)**

[Este tutorial funciona para qualquer sistema operacional que possua Python e pip instalados, seja Windows, Linux ou…](https://www.linkedin.com/pulse/como-instalar-o-selenium-webdriver-windows-10-jo%C3%A3o-gross/" \t "_blank)

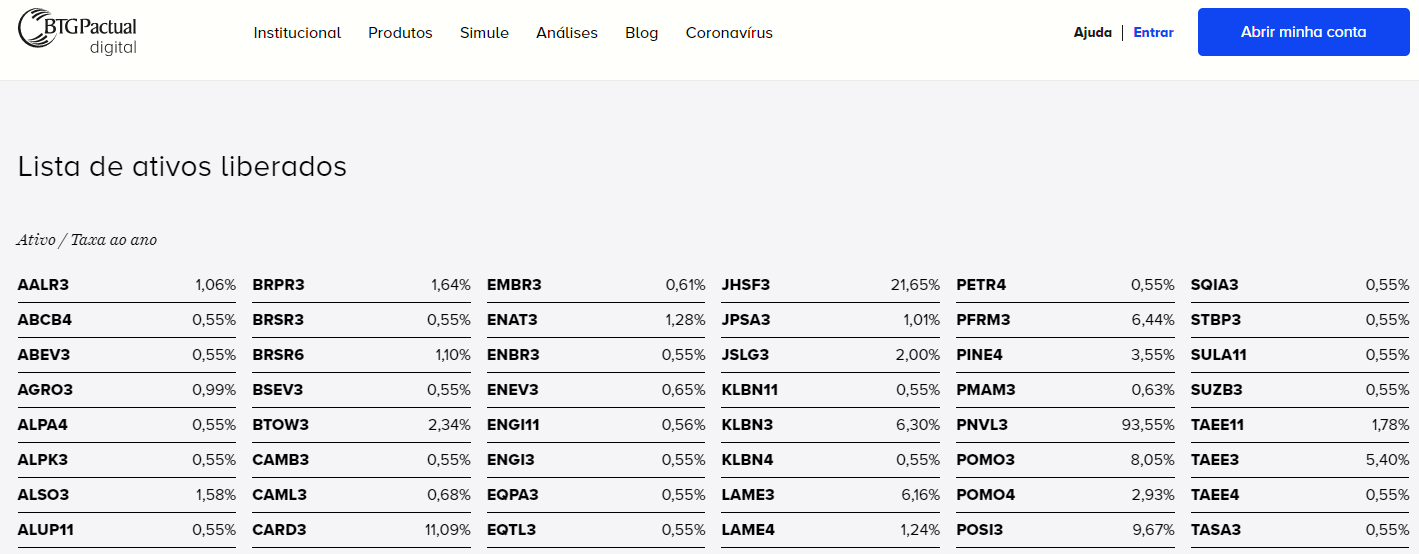
[www.linkedin.com](https://www.linkedin.com/pulse/como-instalar-o-selenium-webdriver-windows-10-jo%C3%A3o-gross/" \t "_blank)

**Análise do website e identificação de tags HTML**

Queremos coletar dados da página de Venda Descoberta do BTG Pacutal. Os dados estão presentes na URL <https://www.btgpactualdigital.com/renda-variavel/venda-descoberta>.

Ao acessar a página temos acesso às taxas de aluguel de ações para operação de venda a descoberto. As informações que vemos no website tem a seguinte aparência:

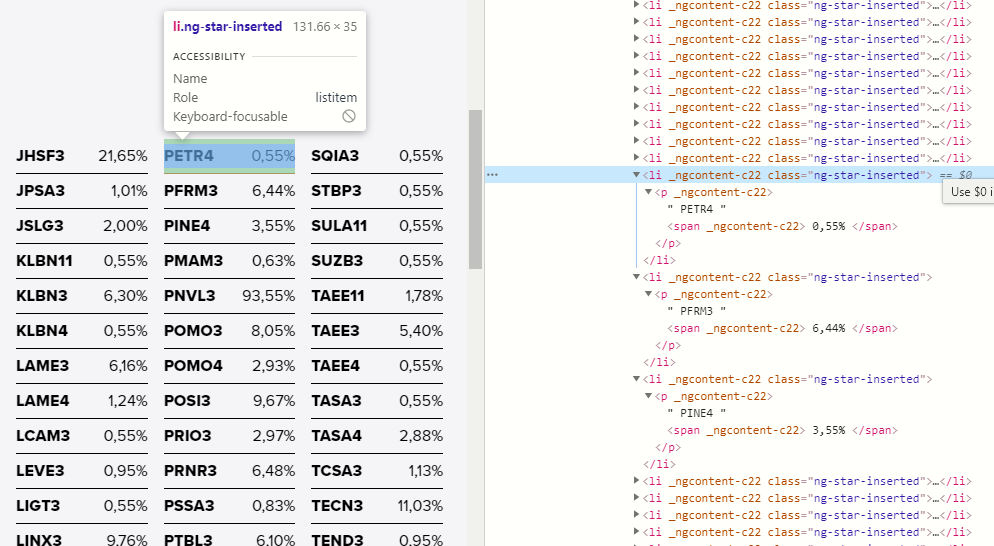
Image for post



Taxas a.a. em % para a venda a descoberto de ações

Já ao inspecionar os pares código de ação e taxa % a.a. observamos que eles são compostos de tags HTML <li> (lista de item), cada um aninhado com um tag <p> (parágrafo).

Image for post



Descrição HTML de cada elemento

O caminho desses elementos desde a origem do documento pode ser obtido ao consultar o XPath, ou seja, podemos ter acesso ao caminho percorrido pelas tags até encontrar o elemento de interesse. Esse caminho, para os elementos pesquisados, é

/html/body/app-root/div/app-variable-sale/section/app-variable-assets-list/section/div/div[2]/div/ul/li[x]

onde x corresponde à posição do elemento na lista, sendo 1 (ação AALR3) para o primeiro elemento e 244 (ação YDUQ3) para o último elemento. Devemos guardar esse caminho, pois ele será muito importante adiante.

**Realizando a coleta de dados de BTC**

Para poder acessar o website e realizar a coleta, primeiramente criamos um driver com o Chromedriver e acessamos a URL desejada com o comando driver.get(). O XPath do primeiro elemento já está definido.

O pacote Selenium disponibiliza diversos métodos para a busca de tags no corpo do HTML da página. O método find\_elements\_by\_xpath() retorna todos os elementos que possuam o XPath indicado. Utilizamos a string xpath sem os últimos 3 caracteres, de modo a encontrar todos os elementos <li> ao invés de apenas o primeiro.

É bastante simples, porém o trecho de **código acima não funciona**. Por que? A questão é que esta página é gerada dinamicamente com o AngularJS e não uma página estática. Sabemos disso, pois no xpath temos a tag app-root, que não pertence ao HTML padrão. O problema com páginas dinâmicas geradas por código JavaScript é que logo que acessamos a página com o driver e partimos para a busca do elemento de interesse, a página ainda não foi completamente gerada e o trecho de código acima não consegue encontrar nenhum componente <li> no caminho especificado.

Uma forma de contornar esse problema é forçando o script a aguardar até que os elementos estejam presentes na página. A partir daí temos certeza que já foram gerados e estão disponíveis para coleta.

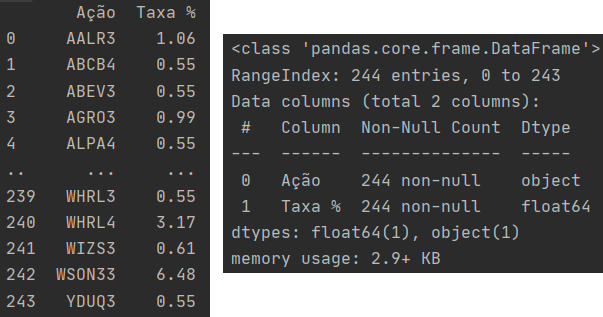
O script fica durante 20 segundos procurando o XPath especificado. Assim que a página é gerada e o elemento encontrado, a coleta via find\_elements\_by\_xpath() consegue encontrar todos os elementos de interesse.

**Criação do DataFrame com os códigos das ações e taxas de BTC**

Com a coleta de dados concluída podemos acessar os valores e inseri-los em um DataFrame para posterior manipulação, provido pelo pacote Pandas.

O DataFrame é criado com as colunas “Ação” e “Taxa %”. O acesso ao conteúdo de cada elemento Selenium é realizado com .text. Os dados aparecem em formato de string e são separados pelo caractere de nova linha '\n'. Cada parte da string é armazenada em uma coluna do DataFrame. Por fim é realizada uma limpeza nos dados da coluna “Taxa %” de modo que as informações possam ser transformadas de string para float. O conteúdo final do DataFrame fica da seguinte forma:

Image for post

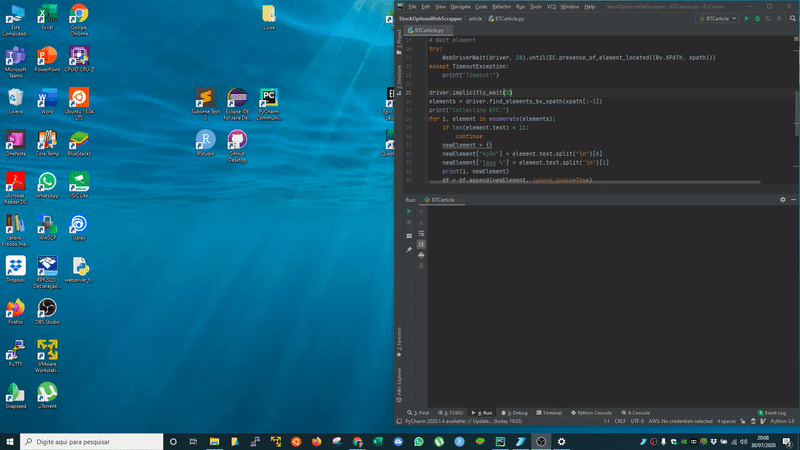


Conteúdo do DataFrame | Informações sobre o DataFrame

**Coleta diária das taxas de BTC**

É possível que o script seja automatizado e executado diariamente para a coleta de dados desejada. Mas para que o artigo não fique muito extenso vou apenas mostrar o script em execução e os resultados obtidos. Esse procedimento pode ser realizado no momento que o investidor desejar as informações.

Image for post



Execução do script

**Considerações Finais**

Neste artigo foi apresentado um script que acessa o website do BTG Pactual e coleta as taxas de aluguel de ações para todas as ações disponíveis. O script executa rapidamente e pode ser ativado toda vez que o investidor queira coletar as taxas de BTC do momento. Foram utilizados os pacotes Selenium e Pandas, o primeiro para criar um driver de acesso e manipulação do website, enquanto o segundo permitiu a criação de uma estrutura de dados chamada DataFrame para armazenar as informações.

Com as informações coletadas é possível realizar diversas análises, como, por exemplo, encontrar a distribuição por percentual de taxa de BTC, encontrar a menor taxa ou a maior, dentre outras. Em artigo oportuno haverá uma continuação deste artigo, no qual vou apresentar diferentes formas de visualização dos dados para auxiliar análises realizadas pelo investidor.

Deixe seu comentário, dúvida ou sugestão.

Forte abraço!