

【初心者向け】

# Python スクレイピング実 践入門

== MENTA教材 ==  
written by bayashi



# 目次

1. この教材の前提知識
2. この教材の対象者
3. 対応バージョン
4. スクレイピング とは
5. スクレイピング の注意事項
6. 環境構築(Anaconda・Jupyter notebook)
7. Pythonの簡単な基礎文法のおさらい
8. スクレイピング をやってみよう
9. MENTAのサイトから名前とプランタイトルを取得する
10. ページ遷移にも挑戦
11. ログインを自動化してみよう



# この教材の前提知識

- ・ProgateやドットインストールなどでHTML/CSSの基礎知識を身に附けている
- ・プログラミングの基本構文を知っている(for文、if文)  
(何となくのレベルでも構いません)



# この教材の対象者

- ・Progateをやったけど、次に何をすればいいかわからない
- ・プログラミングを使って何かしてみたい
- ・データ分析のためにweb上にあるデータを収集したい



# 対応バージョン

**python == 3.6**

**selenium 3.13.0**

**Mac**

**windows**

**Google Chrome**



# スクレイピングとは

- ・ネット上にあるwebサイトの情報を実行されたプログラムによって情報を取得し、その情報を加工したり分析したりすることです。
- ・大量のデータを収集することができる所以業務効率化やデータ分析に有効な手段です。
- ・スクレイピング の利用例  
**株価の変動を確かめる、ネットオークションの価格の変動を確かめる、  
マーケティングに使用するデータを収集する**



# スクレイピング の注意事項①

- ・スクレイピングを行う際にはサイトの規約などをしっかり読みましょう。取得したデータを再利用することを利用規約で禁じているサイトもあります。スクレイピングを禁止しているサイトだと法的に訴えられることもあるため、注意が必要です。

# 環境構築編



# 環境構築

早速ですが、Pythonを手軽に実行できるための環境である、「**Jupyter Notebook**」をインストールしましょう。手順としては

① **Anaconda**のダウンロード、インストール

② **Anaconda**に搭載されている「**Jupyter Notebook**」を起動する

というプロセスになります。

インストールのやり方は次ページにある記事を参考にしてみてください。



# Anacondaのインストールをする

- ・ダウンロードしたファイルをどこに配置するかなどは、添付先URLの記事を参考にしてみてください。
- ・なにかお困りごとや確認したいことがあればお気軽にお尋ねください。

## 【Windows用】

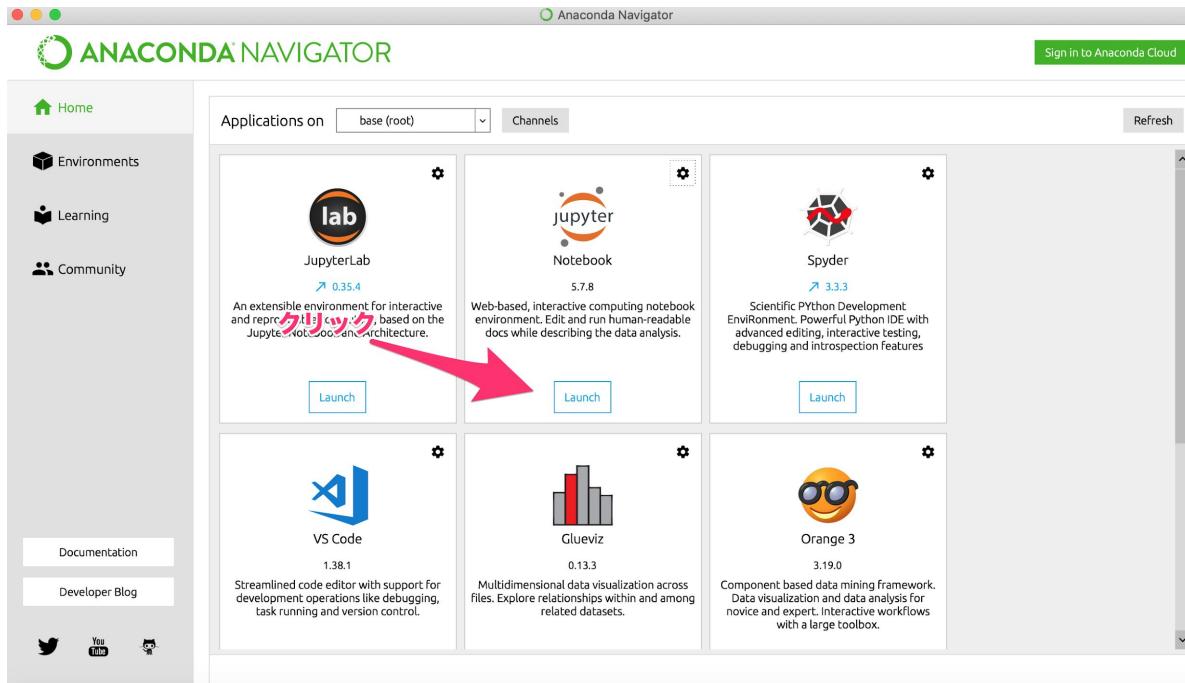
<https://did2memo.net/2017/07/09/python-anaconda-jupyter-notebook-windows/>

## 【Mac用】

<https://qiita.com/turmericN/items/37ea643439c40c556aa4>

# Jupyter Notebook使い方編

# ①Anacondaを開きJupyter notebookをクリックする





## ②Jupyter notebookが開いたら、以下の場所をクリックして「Python3」を選択する

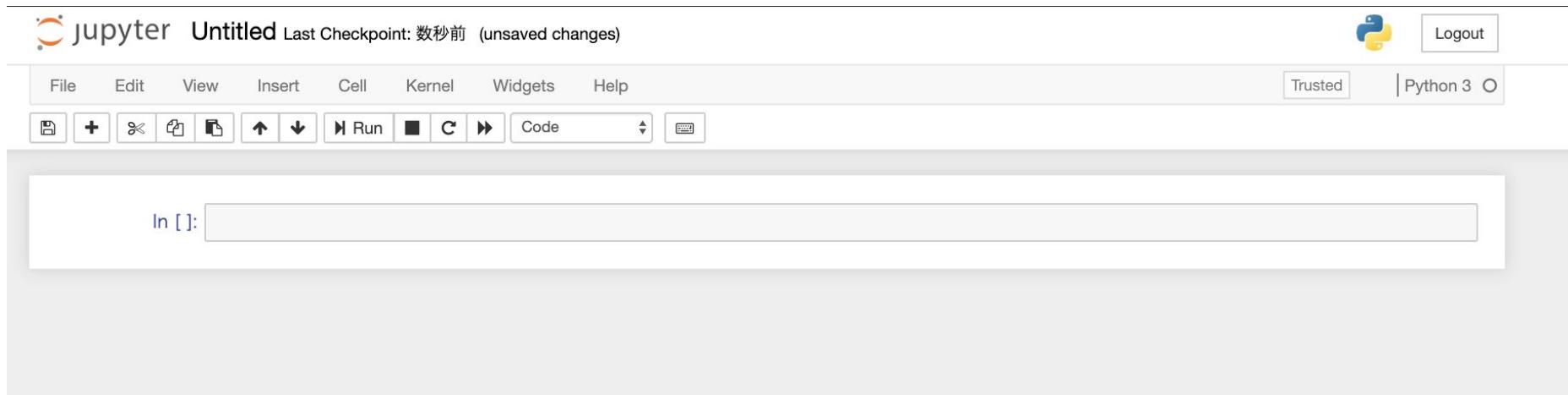


The screenshot shows the Jupyter Notebook interface. At the top, there is a logo and buttons for 'Quit' and 'Logout'. Below the logo, there are tabs for 'Files', 'Running', and 'Clusters', with 'Files' being the active tab. The main area displays a file list with the following entries:

	Name	Last Modified	File size
0	[dir_name]	9ヶ月前	
	AnacondaProjects	6ヶ月前	

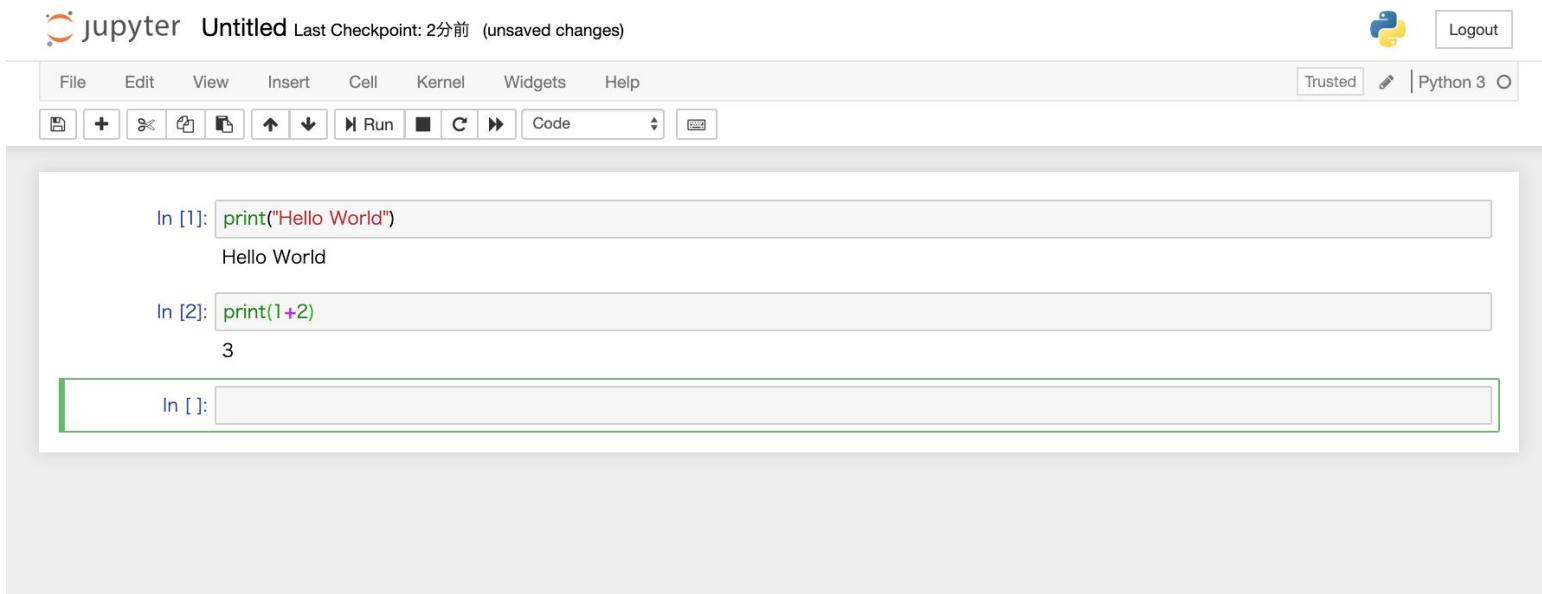
A red arrow points to the 'Last Modified' column header. The word 'クリック' (click) is written in pink over the 'Last Modified' column header of the second row.

③セルの中にPythonのプログラムを書いてみましょう。  
実行は「Shift + Enter」です。



The screenshot shows a Jupyter Notebook interface. At the top, there is a header with the Jupyter logo, the title "Untitled", and a message about the last checkpoint. On the right side of the header are "Logout" and "Python 3" buttons. Below the header is a toolbar with various icons for file operations, cell types, and kernel selection. The main area is a code cell with the prompt "In [ ]:" followed by a large empty text input field.

# 実際のプログラム出力画面(参考)



The screenshot shows a Jupyter Notebook interface with the following details:

- Header:** jupyter Untitled Last Checkpoint: 2分前 (unsaved changes)
- Toolbar:** File, Edit, View, Insert, Cell, Kernel, Widgets, Help, Trusted, Python 3
- Buttons:** New, Open, Save, Cell, Run, Kernel, Code, Help
- Code Cells:**
  - In [1]: `print("Hello World")`  
Output: Hello World
  - In [2]: `print(1+2)`  
Output: 3
  - In [ ]: (Empty cell)