Web data extraction Requirement document

Suggested Code Policy:

1. Copy the Project folder structure from bellow github link.

<https://github.com/hackathonScrapping/BANK.git>

1. Use Python(Mandatory) with BeautifulSoup, Scrapy and Selenium (Feel free to use any other Python libraries)
2. Store Scrapped data with provided file name and CSV format.

Work 1:

**Scarped Data Format (File Name - mybanktracker\_Data\_Deposit\_<MM\_DD\_YYYY>.csv)**

|  |  |
| --- | --- |
| **Column** | **Value** |
| Date | Today <YYYY-MM-DD> |
| Bank\_Native\_Country | US |
| State | New York |
| Bank\_Name | Scrape data |
| Bank\_Local\_Currency | USD |
| Bank\_Type | Bank |
| Bank\_Product | Deposits |
| Bank\_Product\_Type | Scrape data |
| Bank\_Product\_Code |  |
| Bank\_Product\_Name | Scrape data |
| Balance | Scrape data |
| Bank\_Offer\_Feature | Scrape data (Default Offline unless mentioned) |
| Term\_in\_Months | Scrape data (applicable only for CD) |
| Interest\_Type | Fixed |
| Interest | Scrape data (if available) |
| APY | Scrape data |
| Source | Scrape data (web\_url): www.mybanktracker.com |

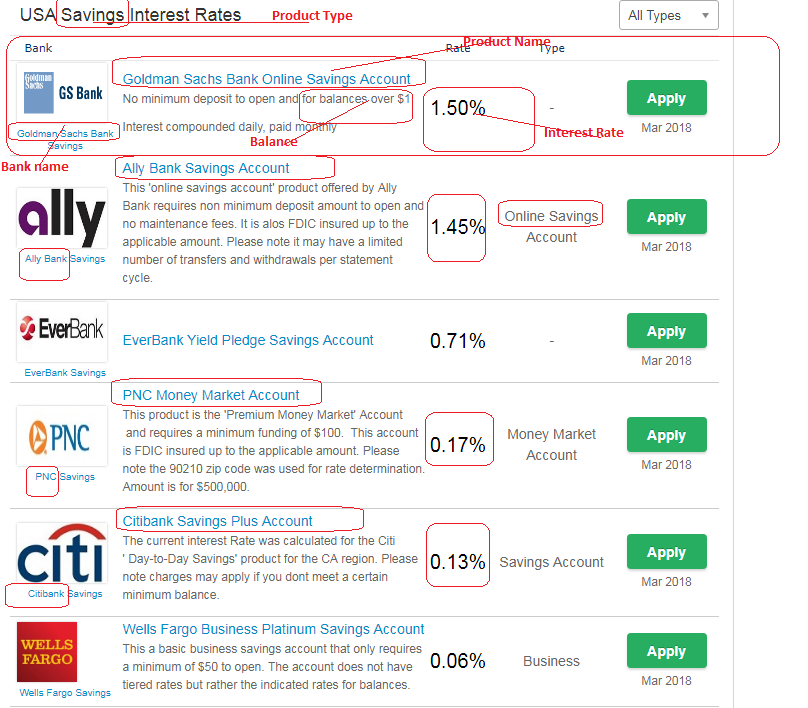
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  |  |  |  |  | | | |  |
|  |  | |  |  |  |  |  |  | | | |  |
|  | |  | | | | | | | | | |  |
|  |  | |  |  |  | | | |  |  |  |  |

**Steps**

1. Click on below bank aggregator site

URL: <https://us.deposits.org/savings-accounts/>

1. Scrape deposit account related data as per details provided below



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | |  | |  | |  | |  | |  |  |  |  |  |
|  |  | | |  | |  | |  | |  | |  | | | | | |

**Scrape all records except business account**

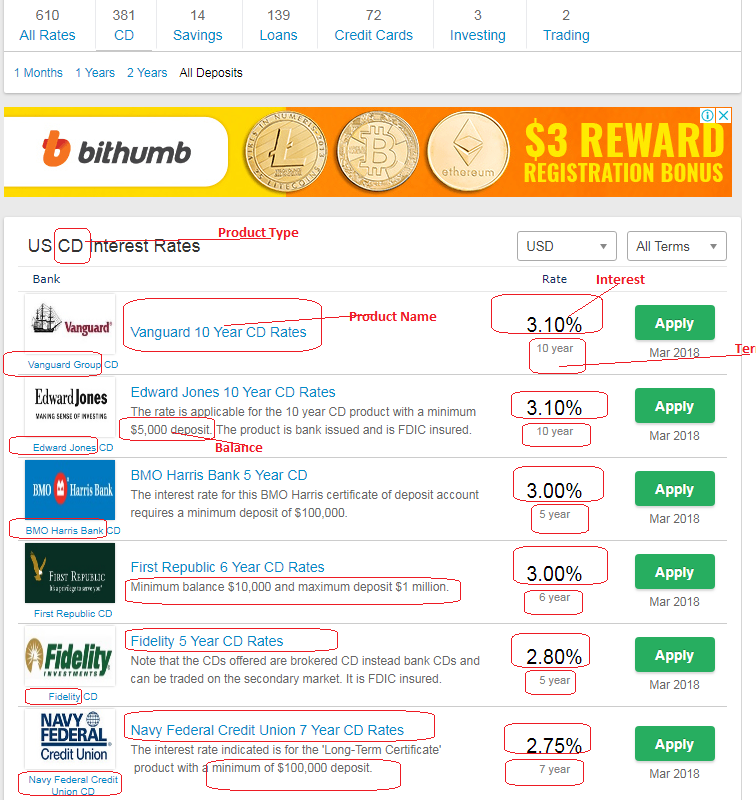
**Checking Account**

**NOT AVAILABLE**

**Scrape all records**

**CD**

[**https://us.deposits.org/deposits/**](https://us.deposits.org/deposits/)

****

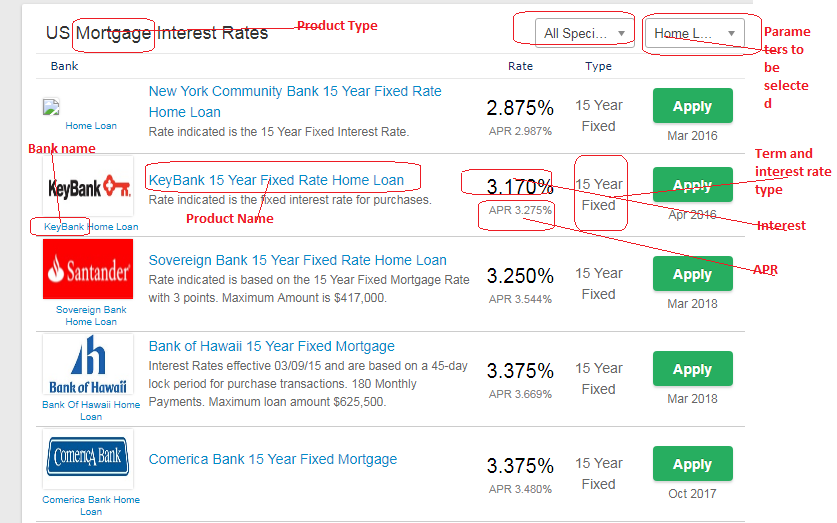
**Remove all records except for 6, 12 and 36 months from the file**

Work 2: Mortgage Rates

|  |  |
| --- | --- |
| **Column** | **Value** |
| Date | Today <YYYY-MM-DD> |
| Bank\_Native\_Country | US |
| State | New York |
| Bank\_Name | Scrape data |
| Ticker | Ticker file |
| Bank\_Local\_Currency | USD |
| Bank\_Type | Bank |
| Bank\_Product | Mortgages |
| Bank\_Product\_Type | Mortgages |
| Bank\_Product\_Code |  |
| Bank\_Product\_Name | Scrape data |
| Min\_Loan\_Amount | Scrape data |
| Bank\_Offer\_Feature | Scrape data, default Offline |
| Term\_in\_Year | Scrape data |
| Interest\_Type | Scrape data |
| Interest | Scrape data |
| APR | Scrape data |
| Mortgage\_Loan\_Amt | Scrape data |
| Mortgage\_Down\_Payment | Scrape data, Default 20% |
| Credit\_Score | 720+ |
| Mortgage\_Category | New Purchase |
| Mortgage\_Reason | Primary Residence |
| Mortgage\_Pymt\_Mode | Principal + Interest |
| Source | us.deposits.org |

**URL:** <https://us.deposits.org/loans/>

Select All Specific types, Jumbo, VA, 7/1 arm, 5/1 arm, fha



[**https://us.deposits.org/home-loans/wells-fargo/**](https://us.deposits.org/home-loans/wells-fargo/) **(ignore this)**