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

Work 1:

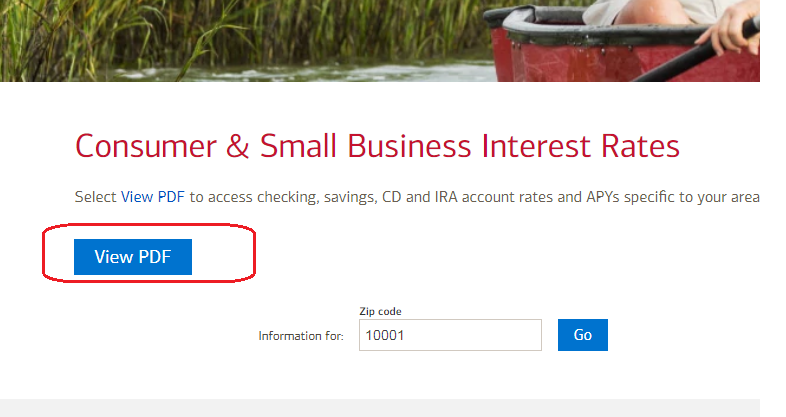
**Scarped Data Format (File Name - BOA\_Data\_Deposit\_<MM\_DD\_YYYY>.csv)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Bank Name | TAB Name | Product Name | Minimum Opening Balance | Minimum Balance | Maximum Balance | Interest Rate | APY |
|  |  |  |  |  |  |  | N/A |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

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

**Steps**

1. Bank URL: <https://www.bankofamerica.com/deposits/bank-account-interest-rates/> (PDF)



1. Scrape deposit rates from the PDF. **(PDF to be done separately) – for CD’s below details are provided for web scraping**

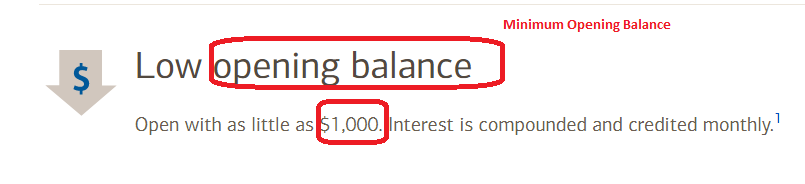
**Savings Rates – As per PDF**

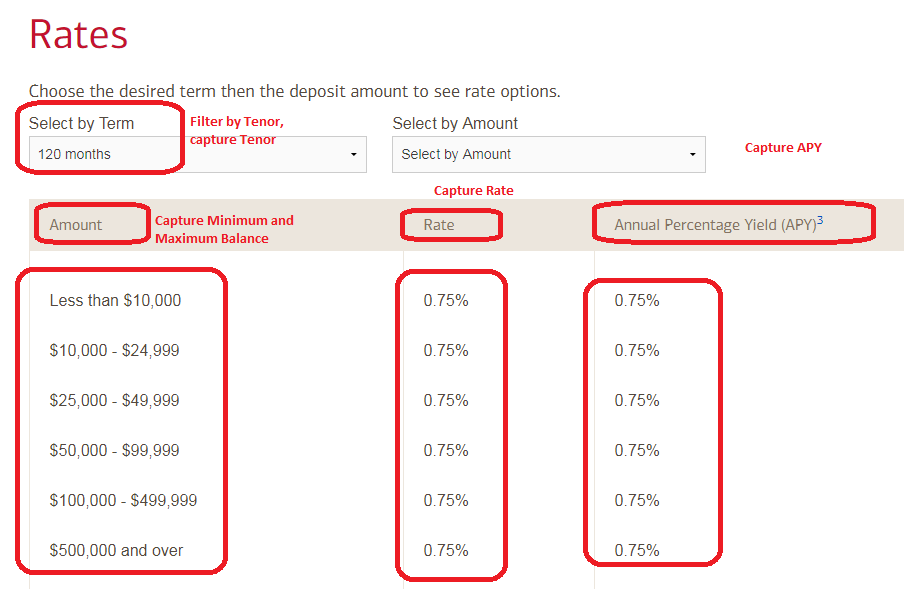
**Checking Rates – As per PDF**

Work 1: Capturing Deposit/CD Rates

**For Standard Term CD Account**

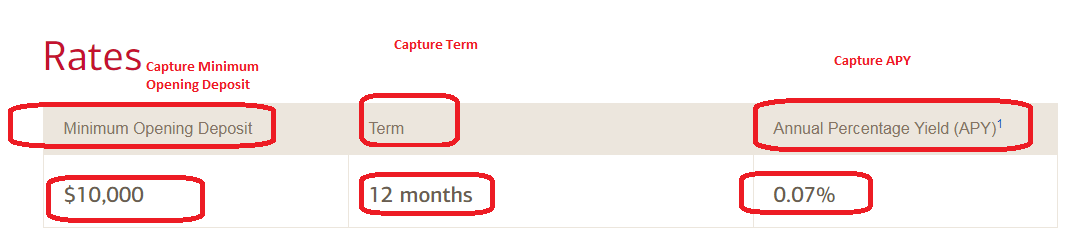
* **Click:** [**https://www.bankofamerica.com/deposits/bank-cds/standard-term-cd/**](https://www.bankofamerica.com/deposits/bank-cds/standard-term-cd/)
* **Select 3 Terms ( 6 months, 1 year and 3 year)**
* **Capture data as per below detail**

****

****

**For Featured CD Account**

https://www.bankofamerica.com/deposits/bank-cds/featured-cd/



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Work 2: Capturing Mortgage data  **Capture data for both Refinance rates and Buying Mortgage rates, additional data needs to be captured from Details**  **Scarped Data Format (File Name - BOA\_Data\_Mortgage\_<MM\_DD\_YYYY>.csv)**   |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Date | Bank Name | Property Value | Loan amount | State | City | Credit Score | Loan to Value Ratio | Loan Type | Product Name | Interest Rate | APR | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |   [**https://www.bankofamerica.com/mortgage/home-mortgage/**](https://www.bankofamerica.com/mortgage/home-mortgage/)    [**https://www.bankofamerica.com/mortgage/refinance/**](https://www.bankofamerica.com/mortgage/refinance/) |  |

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