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 - Depositaccounts\_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.depositaccounts.com |

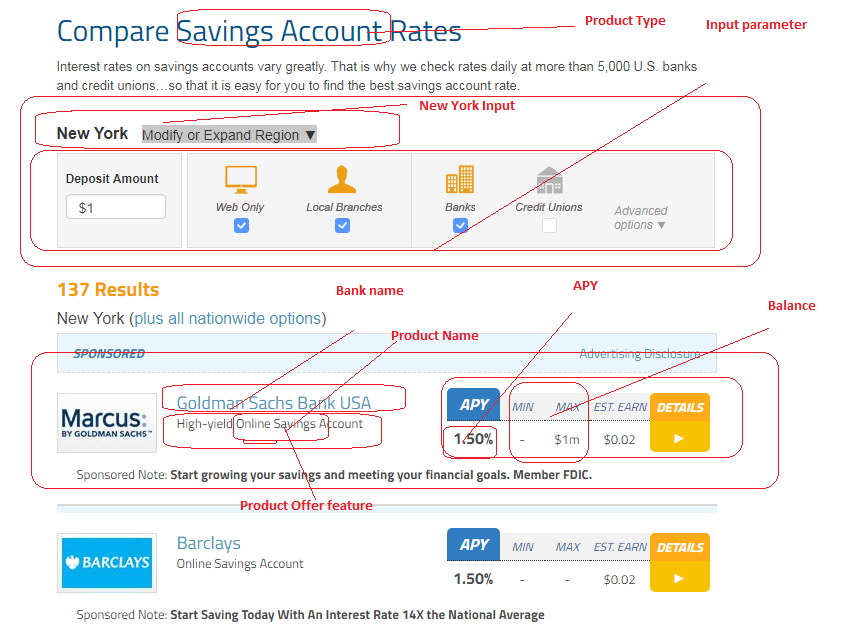
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  |  |  |  |  | | | |  |
|  |  | |  |  |  |  |  |  | | | |  |
|  | |  | | | | | | | | | |  |
|  |  | |  |  |  | | | |  |  |  |  |

**Steps**

1. Click on below bank aggregator site

URL: <https://www.depositaccounts.com/savings/>

1. Click on **Modify or Expand Region ▼** and select New York
2. Scrape deposit account related data as per details provided below

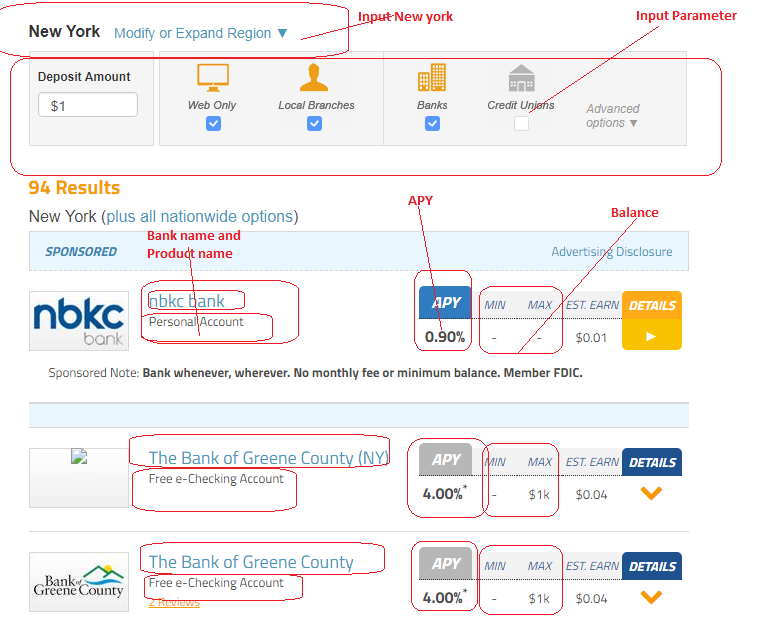


|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | |  | |  | |  | |  | |  |  |  |  |  |
|  |  | | |  | |  | |  | |  | |  | | | | | |

**Scrape all 137 records**

**Checking Account**

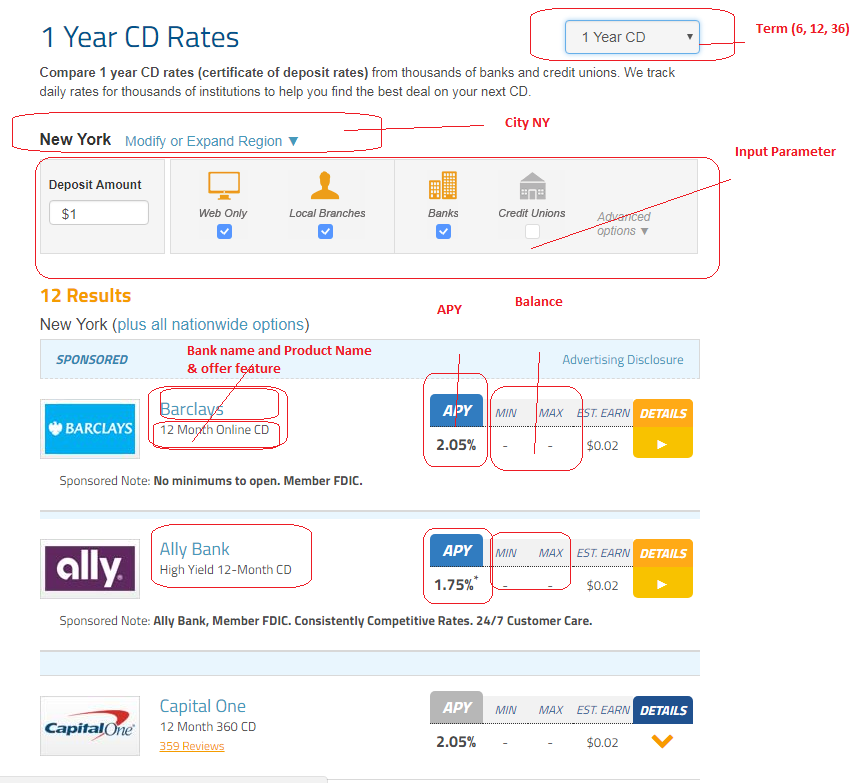
**URL:** [**https://www.depositaccounts.com/checking/**](https://www.depositaccounts.com/checking/)

****

**Scrape all 94 records**

**CD**

**Select 6, 12 and 36 months from the dropdown and scrape data**

****

Work 2: Mortgage Rates

NOT AVAILABLE