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 - CAPONE\_Data\_Deposit\_<MM\_DD\_YYYY>.csv)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Bank Name | Product Name | Bank Offer | Minimum Balance | Minimum Balance | Maximum Balance | Interest Rate | APY |
|  |  |  | Online |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

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

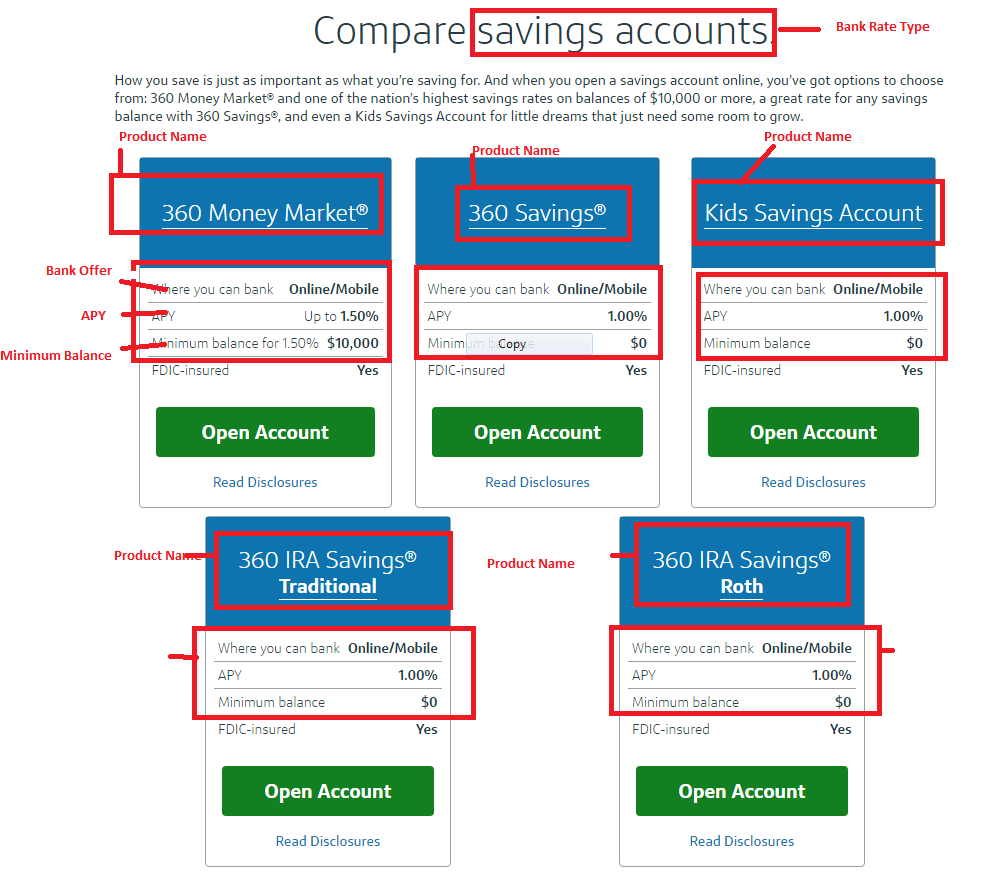
**Steps**

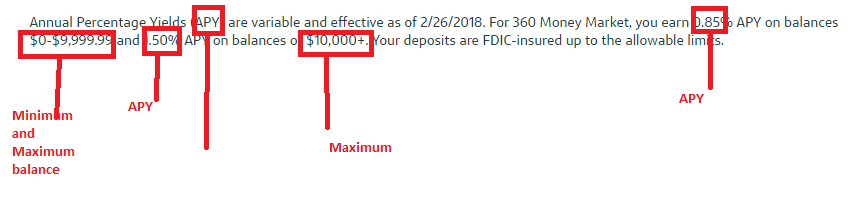
1. Click on below bank site

URL: <https://www.capitalone.com/bank/savings-accounts/#id_comparesavingsaccounts>

Scrape data for each of the 5 products (360 Money market, 360 Savings, Kids saving Account, 360 IRA Savings Traditional, 360 IRA savings Roth)

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



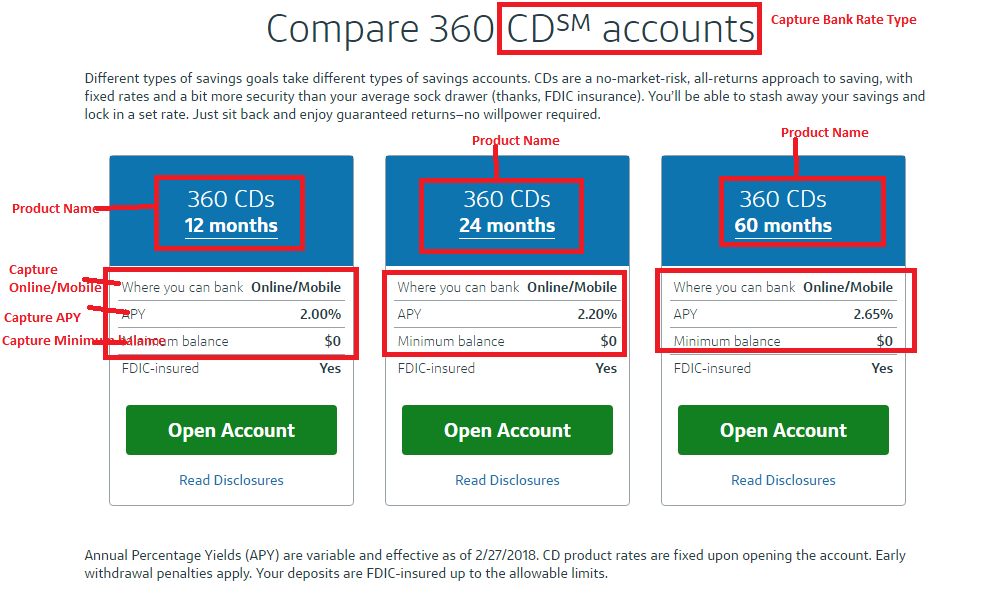


As per snapshot below, scrape data

1. Capture: Bank Rate type (headline, appearance 1)
2. Product name, Offer type, APY, Minimum Balance
3. Capture additional details from text for 360 money market as shown above

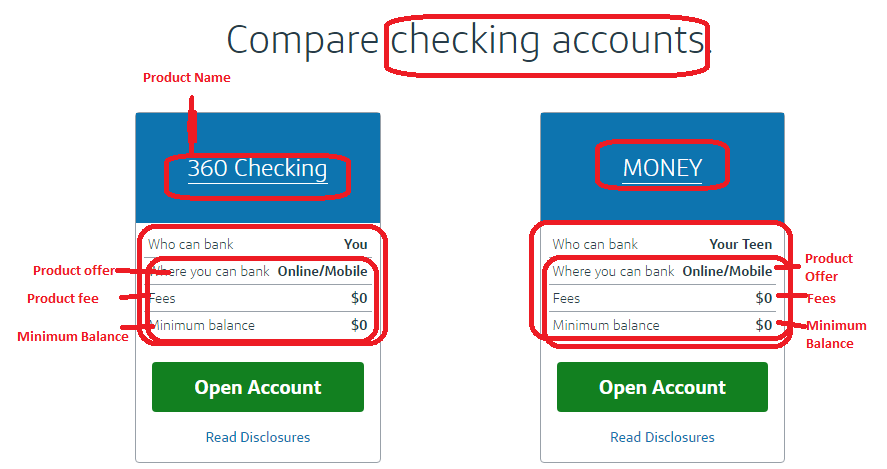
Click

<https://www.capitalone.com/bank/savings-accounts/#id_comparecdaccounts>





<https://www.capitalone.com/bank/open-an-account/#id_checkingaccounts>



|  |  |
| --- | --- |
| Work 2: Capturing Mortgage data  **No data available – Do not capture** |  |

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