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

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

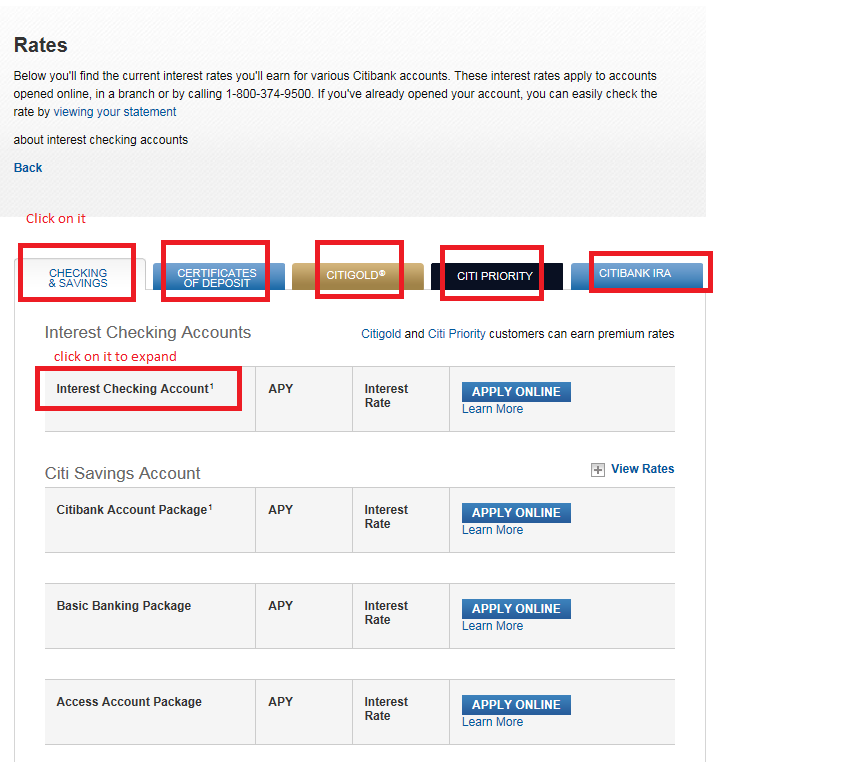
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  |  |  |  |  | | | |  |
|  |  | |  |  |  |  |  |  | | | |  |
|  | |  | | | | | | | | | |  |
|  |  | |  |  |  | | | |  |  |  |  |

**Steps**

1. Click on below bank site

URL: <https://online.citi.com/US/JRS/pands/detail.do?ID=CurrentRates&JFP_TOKEN=7JAPCVIC>

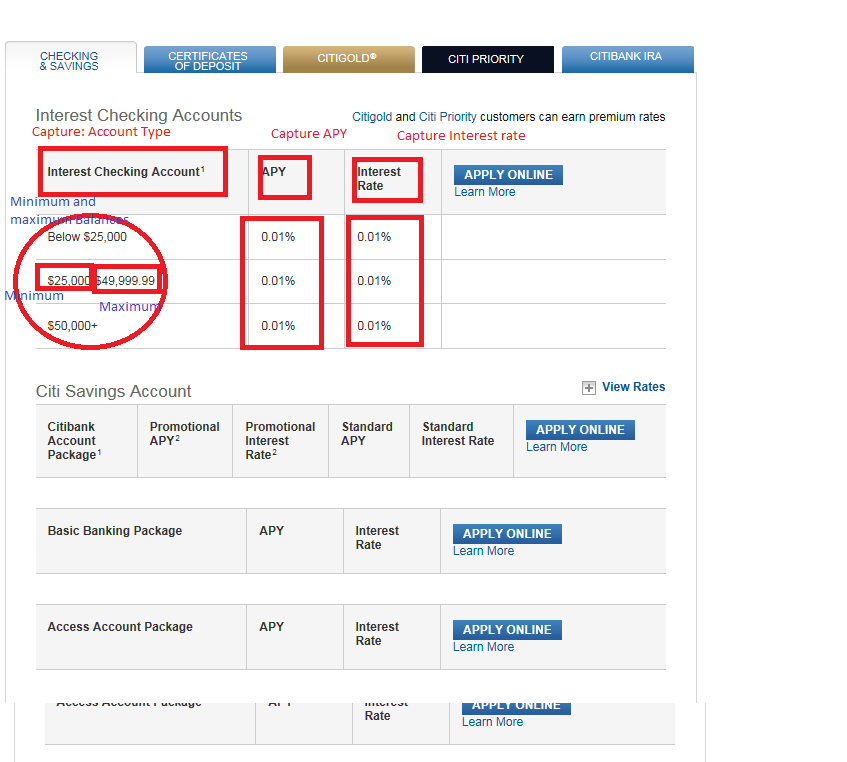
1. Click on each of the 5 Tabs (CHECKING & SAVINGS, CERTIFICATES OF DEPOSIT, CITIGOLD, CITI PRIORITY, CITIBANK IRA)
2. Scrape deposit account related data as per details provided below

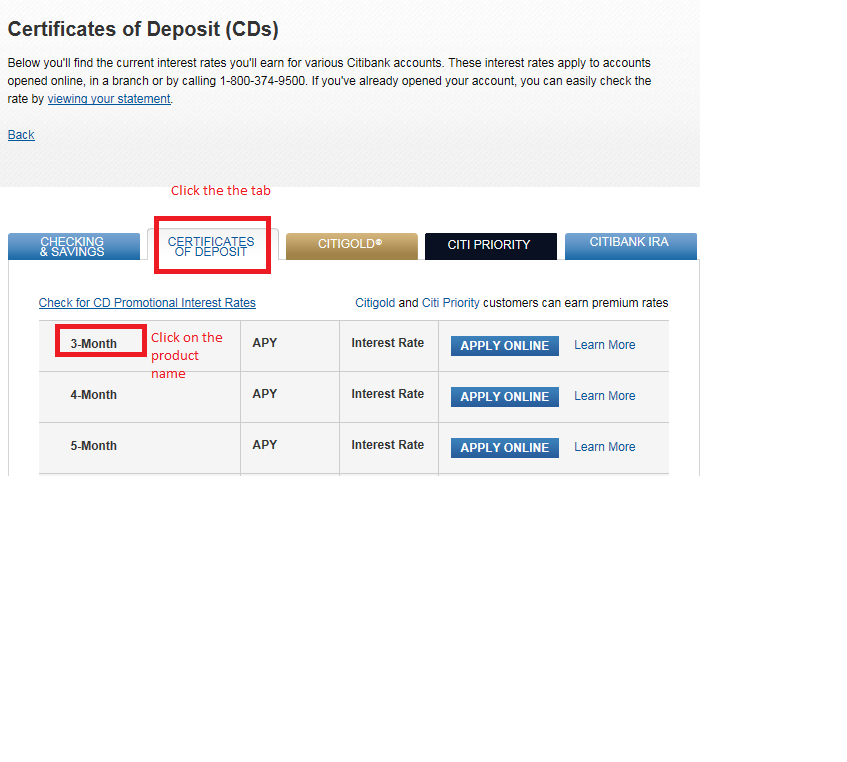


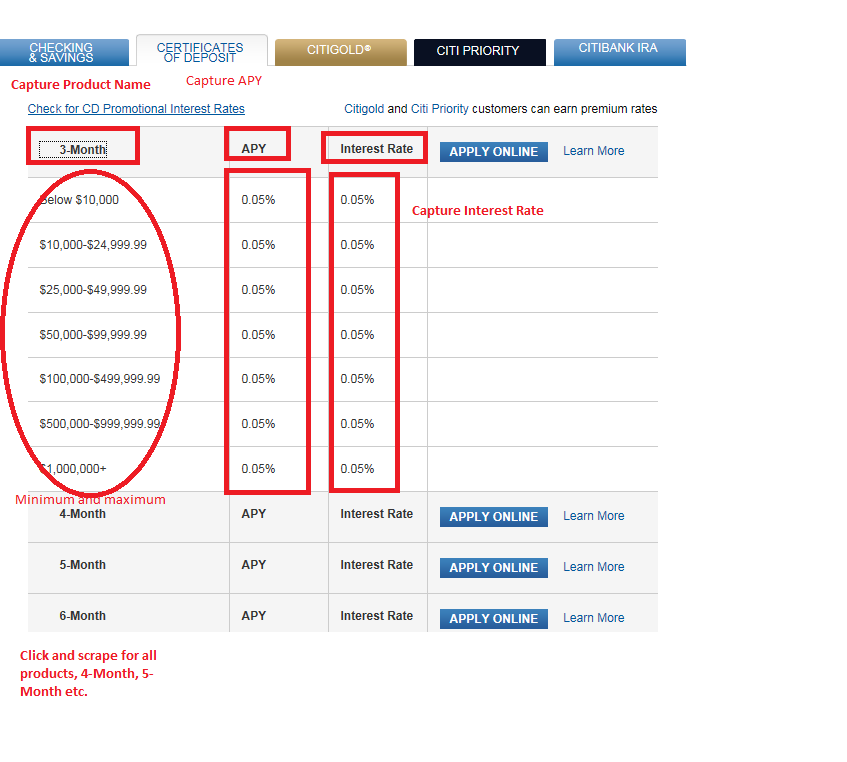
As per snapshot below, scrape data

1. Capture: Account Name Interest Checking Account
2. Capture Minimum and Maximum Balance
3. Capture APY
4. Capture Interest Rate

Repeat same for 4 other tabs (CERTIFICATE OF DEPOSIT, CITIGOLD, CITIPRIORITY, CITIBANK IRA)







|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\citi5.png**  **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\Citi6.png**  **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\Citi7.png**  **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\Citi8.png**  Work 2: Capturing Mortgage data  **Scarped Data Format (File Name - CITI\_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 | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |   **URL:** [**https://online.citi.com/US/JRS/portal/template.do?ID=mortgage\_home\_mortgage&JFP\_TOKEN=NZYB1QVK**](https://online.citi.com/US/JRS/portal/template.do?ID=mortgage_home_mortgage&JFP_TOKEN=NZYB1QVK)  **Capture data for both Refinance rates and Buying Mortgage rates, additional data needs to be captured from Details**  **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\Citi9.png**  **D:\Personal\Ramesh\BlackRock\Requirement Docs\Snapshot\citi10.png** |  |

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