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

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Bank Name | Product Name | Account Balance | Minimum to Open | Maximum Balance | Rate % | APY % | Bonus Interest Rate | Bonus APY |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

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

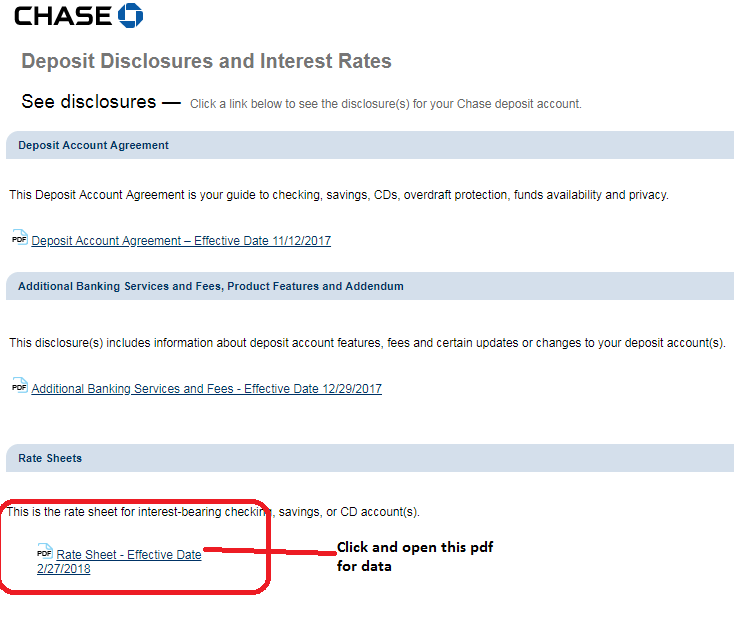
**Steps**

1. Click on below bank site
2. Scrape the below highlighted data points

**Savings & Checking Accounts**

**URL**: <https://apply.chase.com/oao/ViewDisclosures.aspx?type=PER> - (PDF)

**PDF to be worked separately**



Work 2: Capturing Mortgage data

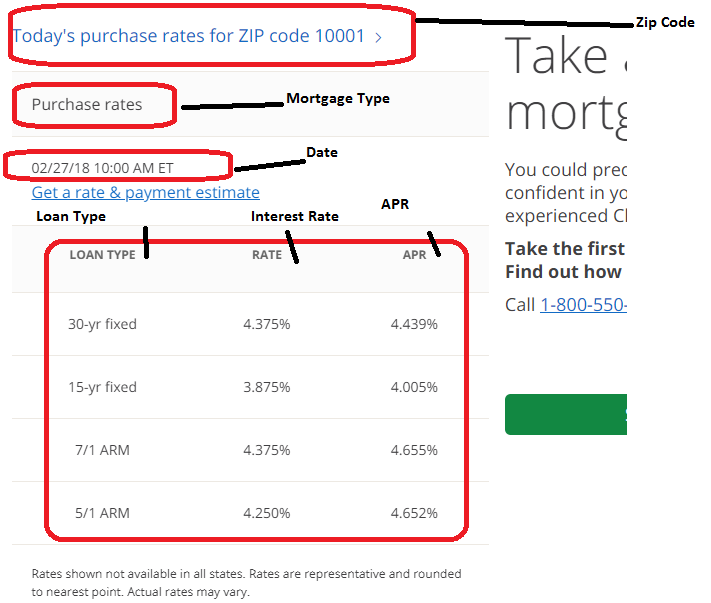
**Capture data for both Mortgage and Refinance along with any other additional data as per below**

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

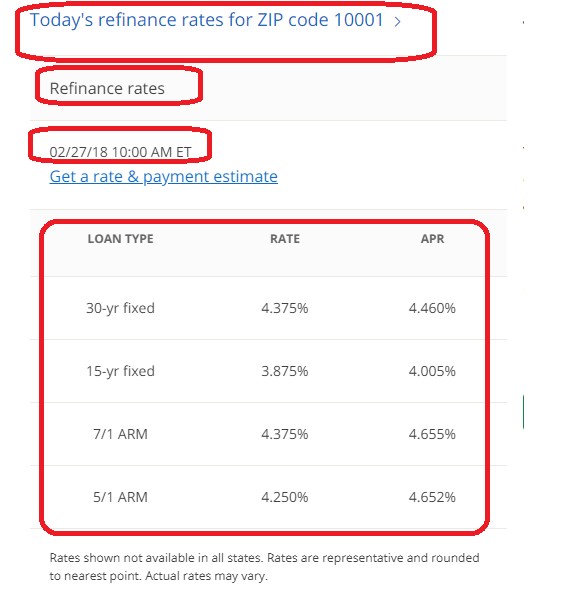
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | Bank Name | Interest Type | Repayment Method | Loan amount | State | Loan Term | Credit Score | Loan to Value Ratio (LTV) | Loan Type | Product Name | Interest Rate | APR |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

Capture the following:

**URL:** <https://www.chase.com/personal/mortgage/mortgage-purchase#today>

****

**URL:** <https://www.chase.com/personal/mortgage/mortgage-refinance#rate>

d