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

|  |  |
| --- | --- |
| **Column** | **Value** |
| Date | Today <YYYY-MM-DD> |
| Bank\_Native\_Country | UK |
| State | London |
| Bank\_Name | Scrape data |
| Bank\_Local\_Currency | GBP |
| 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 |
| Term\_in\_Months | Scrape data |
| Interest\_Type | Fixed |
| Interest | Scrape data |
| AER | Scrape data |
| Source | moneysupermarket.com |

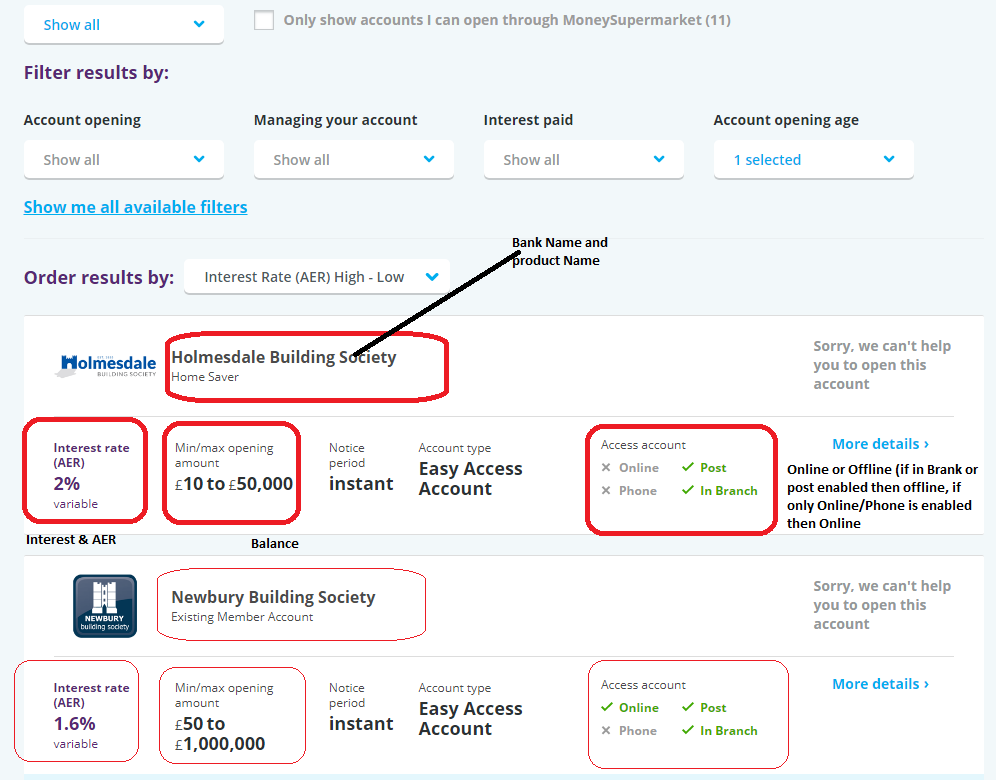
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | |  | **SAVINGS** |  |  |  |  | | | |  |
|  |  | |  |  |  |  |  |  | | | |  |
|  | |  | | | | | | | | | |  |
|  |  | |  |  |  | | | |  |  |  |  |

**Steps**

1. Click on below bank aggregator site

URL: <https://www.moneysupermarket.com/savings/results/>

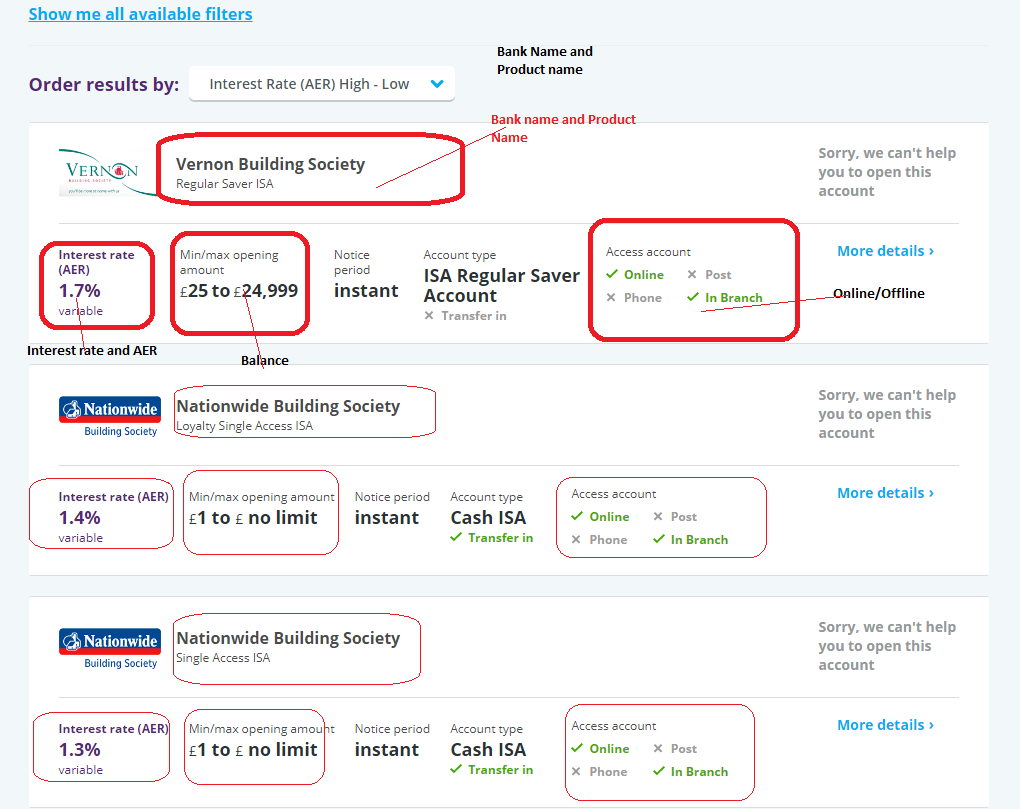
1. Review Instant Access Savings Account
2. Scrape deposit account related data as per details provided below



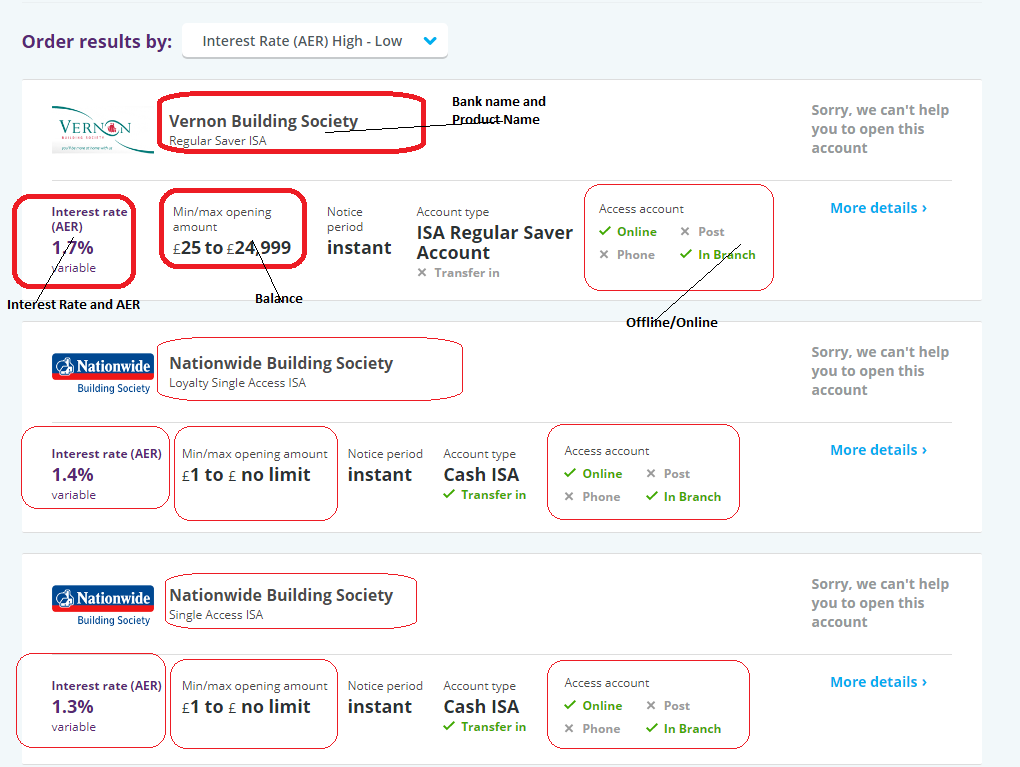
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | |  | |  | |  | |  | |  |  |  |  |  |
|  |  | | |  | |  | |  | |  | |  | | | | | |

**Scrape all records and select Cash ISA and Fixed Rate Savings Bond**

[**https://www.moneysupermarket.com/savings/results/?goal=SAV\_CASHISA**](https://www.moneysupermarket.com/savings/results/?goal=SAV_CASHISA)

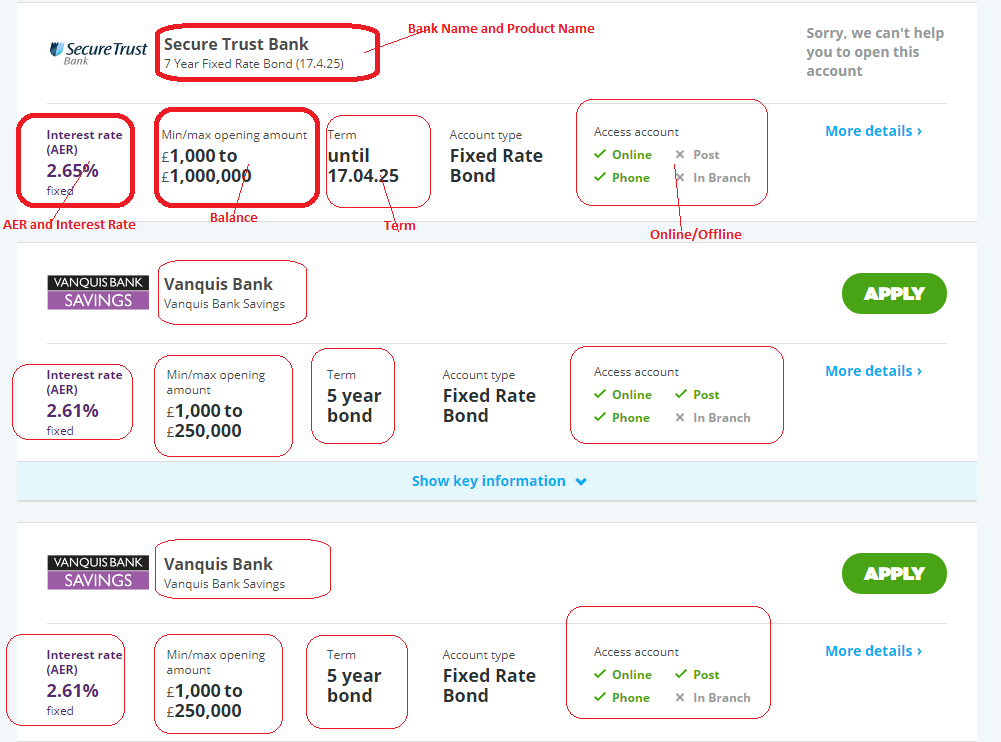
****

[**https://www.moneysupermarket.com/savings/results/?goal=SAV\_CASHISA**](https://www.moneysupermarket.com/savings/results/?goal=SAV_CASHISA)

****

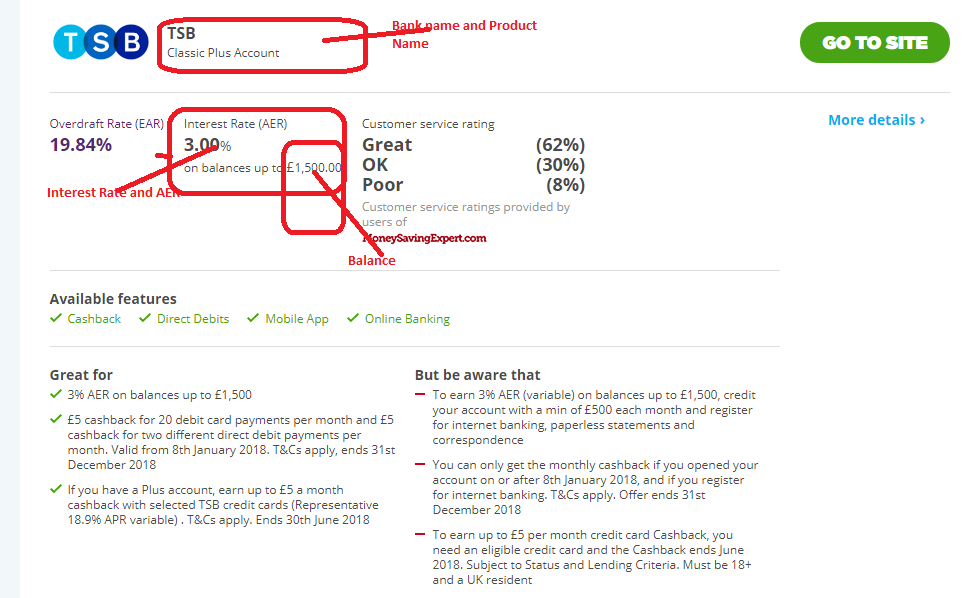
**Term Deposits**

[**https://www.moneysupermarket.com/savings/results/?goal=SAV\_FIXEDRATEBONDS**](https://www.moneysupermarket.com/savings/results/?goal=SAV_FIXEDRATEBONDS)

****

**Checking Account**

[**https://www.moneysupermarket.com/current-accounts/results/?goal=cur\_all**](https://www.moneysupermarket.com/current-accounts/results/?goal=cur_all)

****

**Scrape all records**

Work 2: Mortgage Rates

**File output name: Comparethemarket\_Data\_Mortgages\_<MM\_DD\_YYYY>.csv)**

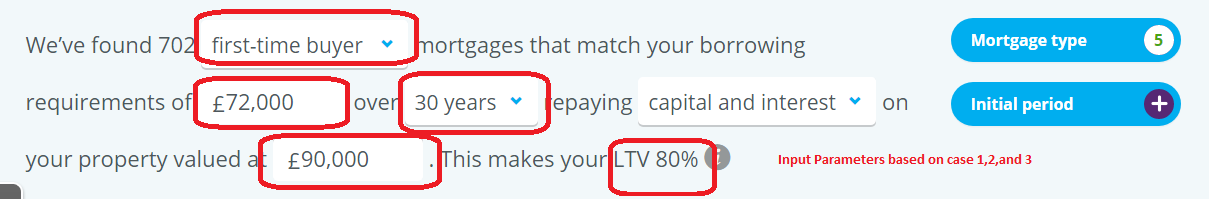
|  |  |
| --- | --- |
| **Column** | **Value** |
| Date | Today <YYYY-MM-DD> |
| Bank\_Native\_Country | UK |
| State | London |
| Bank\_Name | Scrape data |
| Ticker | Ticker file |
| Bank\_Local\_Currency | GBP |
| Bank\_Type | Bank |
| Bank\_Product | Mortgages |
| Bank\_Product\_Type | Mortgages |
| Bank\_Product\_Code |  |
| Bank\_Product\_Name | Scrape data |
| Min\_Loan\_Amount | Scrape data |
| Bank\_Offer\_Feature | Scrape data, default Offline |
| Term (Y) | Scrape data |
| Interest\_Type | Scrape data |
| Interest | Scrape data |
| APRC | Scrape data |
| Mortgage\_Loan\_Amt | Scrape data |
| Mortgage\_Down\_Payment | Scrape data (20%) |
| Mortgage\_Category | New Purchase |
| Mortgage\_Reason | Primary Residence |
| Mortgage\_Pymt\_Mode | Principal + Interest |
| Source | moneysupermarket.com |

[**https://money.comparethemarket.com/mortgages/first-time-buyer/?AFFCLIE=CM01**](https://money.comparethemarket.com/mortgages/first-time-buyer/?AFFCLIE=CM01)

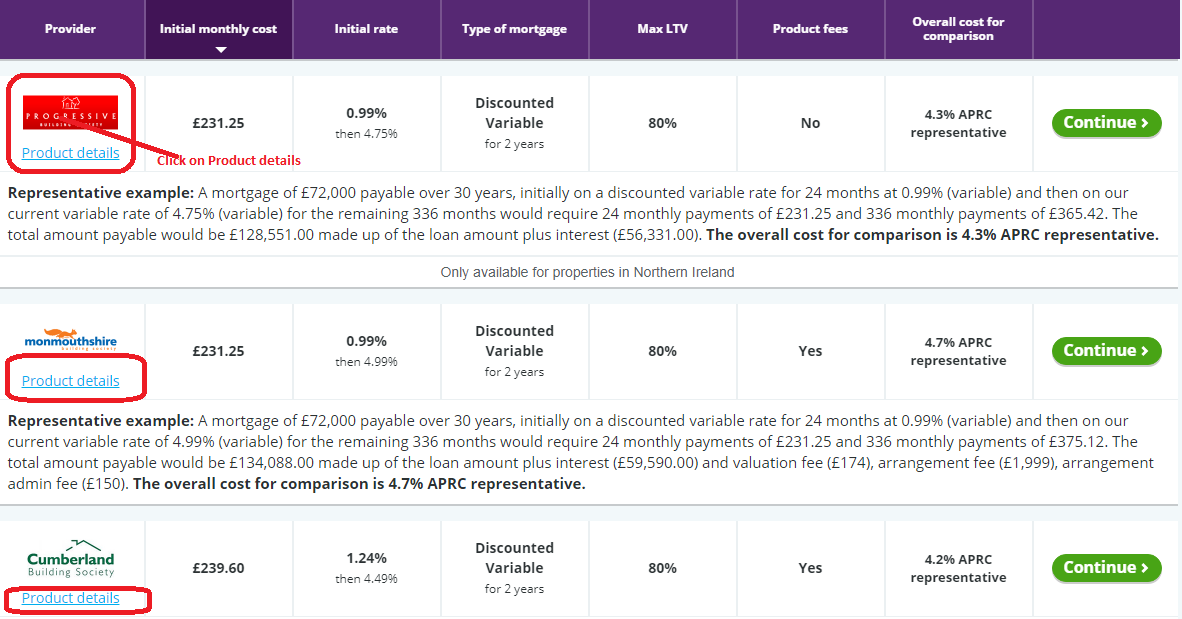
Case 1: Property Value: 90,000; Mortgage Amount: 72,000

Case 2: Property Value: 270,000; Mortgage Amount: 216,000

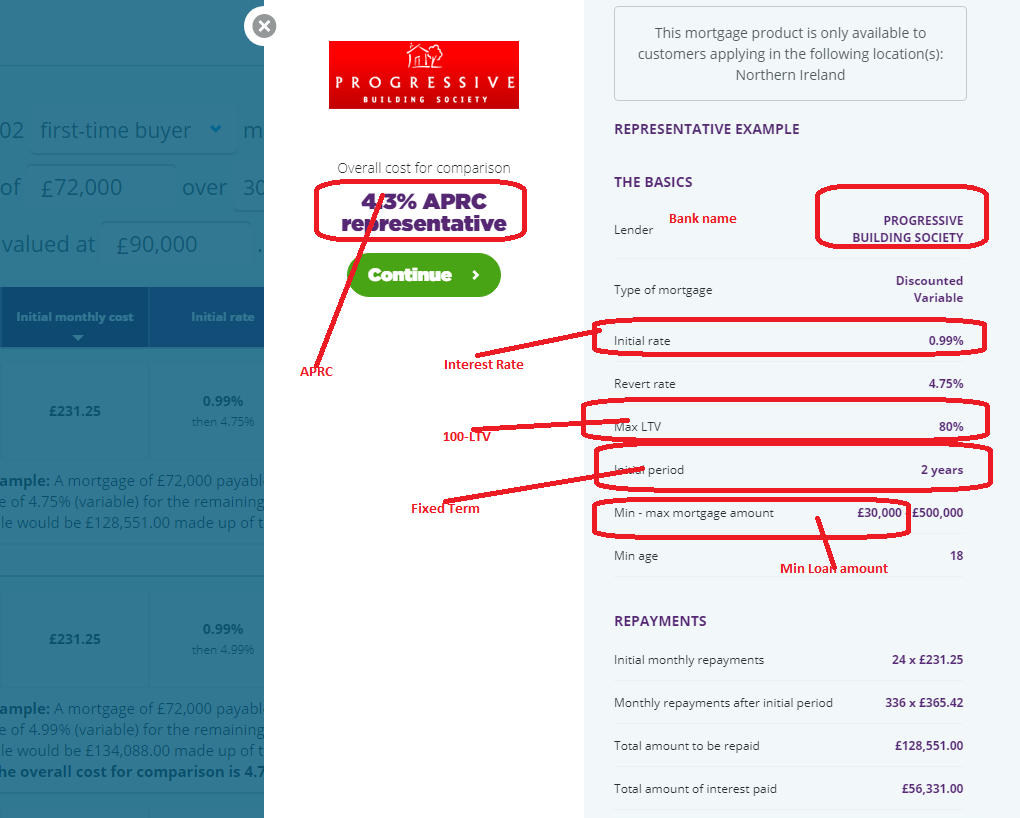
Case 2: Property Value: 450,000; Mortgage Amount: 360,000

****

**Step 2**

****

Step 3



Please update product name: loan period (2 year fixed)

Repeat for 10, 15, 25, 30 year loan term