Web data extraction Requirement document

Suggested Code Policy:

1. Copy the Project folder structure from bellow github link.

<https://github.com/pool23/pool_UK>

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 - Lloyds\_Data\_Deposit\_<YYYY\_MM\_DD>.csv)**

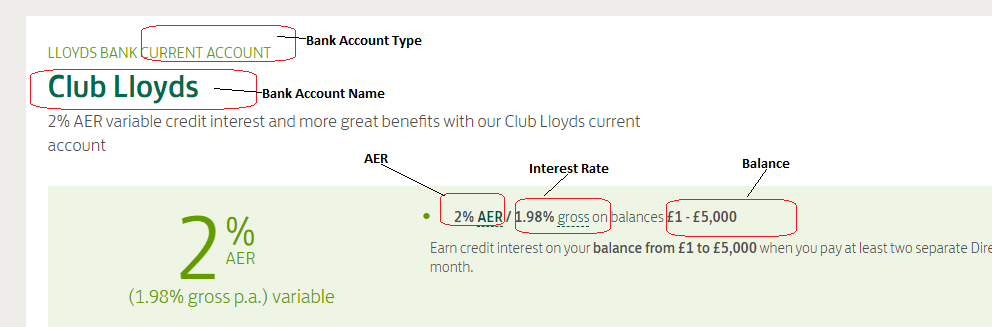
|  |  |
| --- | --- |
| **Column NAMES for Deposits** | **Default Values** |
| Date | Date(YYYY-MM-DD) |
| Bank\_Native\_Country | UK |
| State | London |
| Bank\_Name | Lloyds Bank |
| Bank\_Local\_Currency | GBP |
| Bank\_Type | Bank |
| Bank\_Product | Deposits |
| Bank\_Product\_Type | SCRAPED DATA - Format should be Savings/Current/Term Deposits |
| Bank\_Product\_Name | SCRAPED DATA |
| Balance | SCRAPED DATA |
| Bank\_Offer\_Feature | Offline by default and Online where product name has keyword Online |
| Term in Months | SCRAPED DATA |
| Interest\_Type | SCRAPED DATA - Default is Variable, if data is not available in product name |
| Interest | SCRAPED DATA |
| AER | SCRAPED DATA |
| Bank\_Product\_Code | LEAVE THIS COLUMN BLANK |

**Savings Account: PDF**

[**https://www.lloydsbank.com/assets/media/pdfs/savings/interest-rates.pdf**](https://www.lloydsbank.com/assets/media/pdfs/savings/interest-rates.pdf)

**Current Account:**

[**https://www.lloydsbank.com/current-accounts/all-accounts.asp**](https://www.lloydsbank.com/current-accounts/all-accounts.asp)

****

Work 2: Mortgage Rates

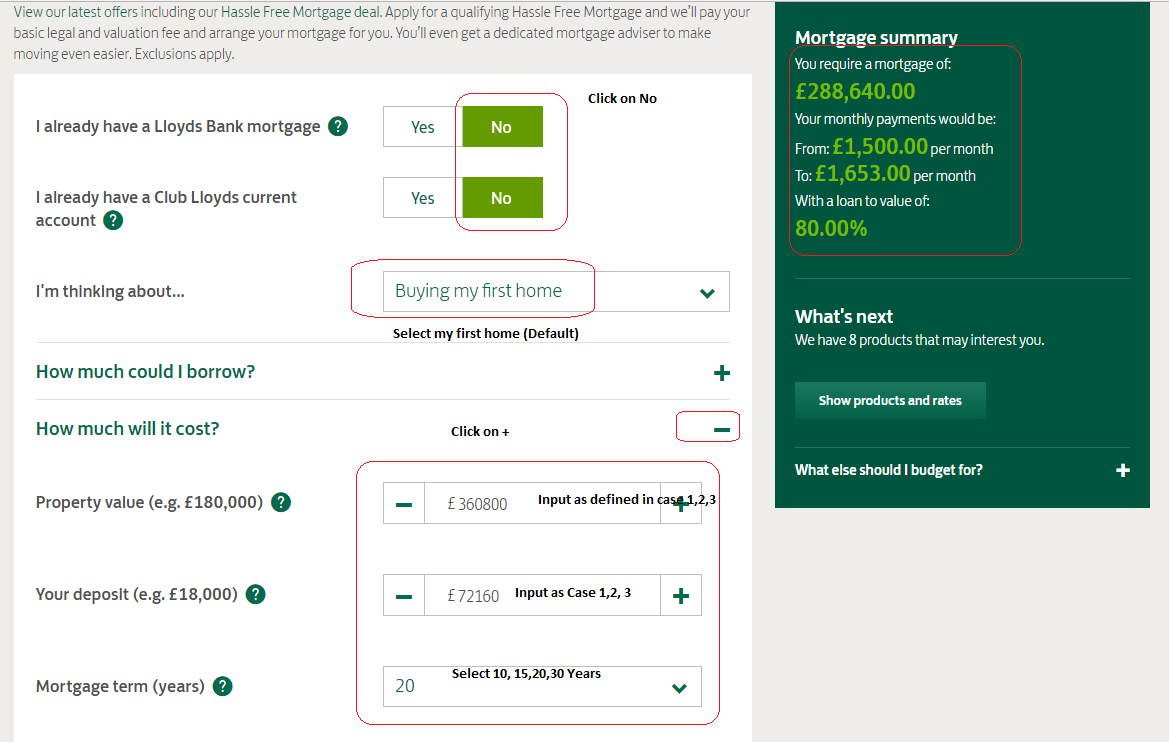
|  |  |
| --- | --- |
| **Mortgages Columns** | **Default Values** |
| Date | Date(YYYY-MM-DD) |
| Bank\_Native\_Country | UK |
| State | London |
| Bank\_Name | Lloyds Bank |
| Bank\_Local\_Currency | GBP |
| Bank\_Type | Bank |
| Bank\_Product | Mortgages |
| Bank\_Product\_Type | Mortgages |
| Bank\_Product\_Name | SCRAPED DATA |
| Min\_Loan\_Amount | SCRAPED DATA |
| Bank\_Offer\_Feature | Offline |
| Term (Y) | SCRAPED DATA |
| Interest\_Type | SCRAPED DATA (Variable, by default) |
| Interest | SCRAPED DATA |
| APRC | SCRAPED DATA |
| Mortgage\_Loan\_Amt | SCRAPED DATA |
| Mortgage\_Down\_Payment | 20% |
| Mortgage\_Category | New Purchase |
| Mortgage\_Reason | Primary Residence |
| Mortgage\_Pymt\_Mode | Principle + Interest |
| Fixed\_Rate\_Term | SCRAPED DATA <from title> |
| Bank\_Product\_Code | LEAVE THIS COLUMN BLANK |

<https://www.lloydsbank.com/mortgages/mortgage-calculator.asp>

Case 1: Property Value: 90,000; Your deposit: 18000

Case 2: Property Value: 270,000; Your deposit: 54000

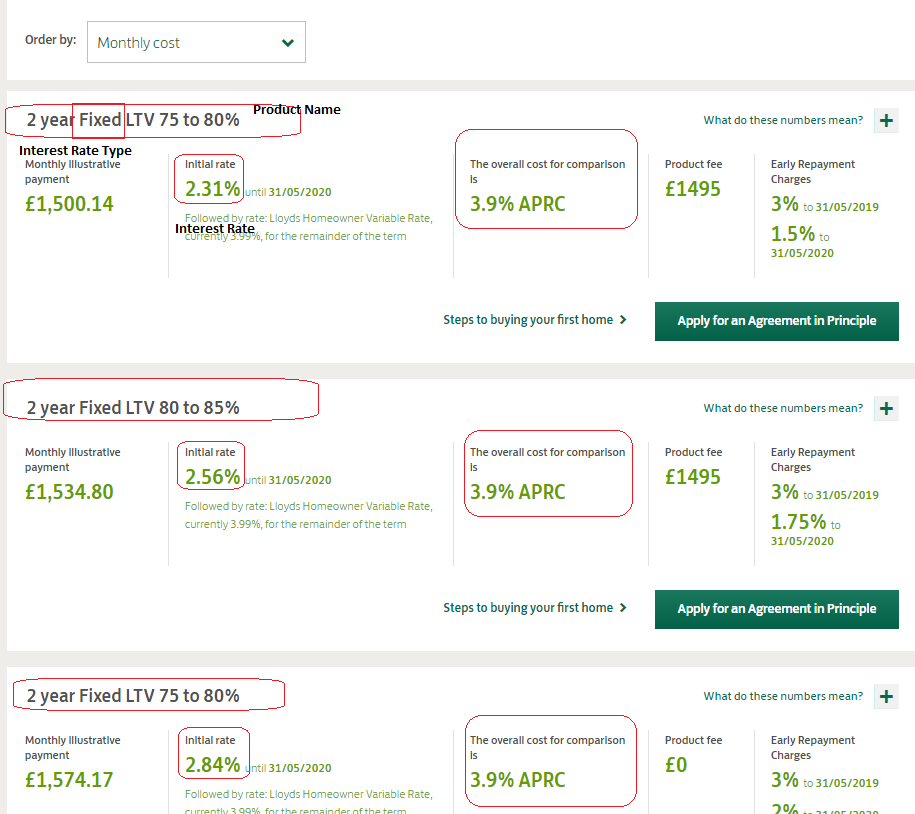
Case 2: Property Value: 450,000; Your deposit: 90,000



Click on Show products and rates

3 cases as shown above should be scraped

Output: [https://www.lloydsbank.com/mortgages/mortgage-calculator.asp#p:results](https://www.lloydsbank.com/mortgages/mortgage-calculator.asp" \l "p:results)



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

**Example for 1st Product:**

Product Name: 2 Year Fixed LTV 75-80%

Term: 20

Fixed\_Rate\_Term: 2

Mortgage\_Down\_Pymt: 20%

Interest: 2.31%

APRC: 3.9%