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:

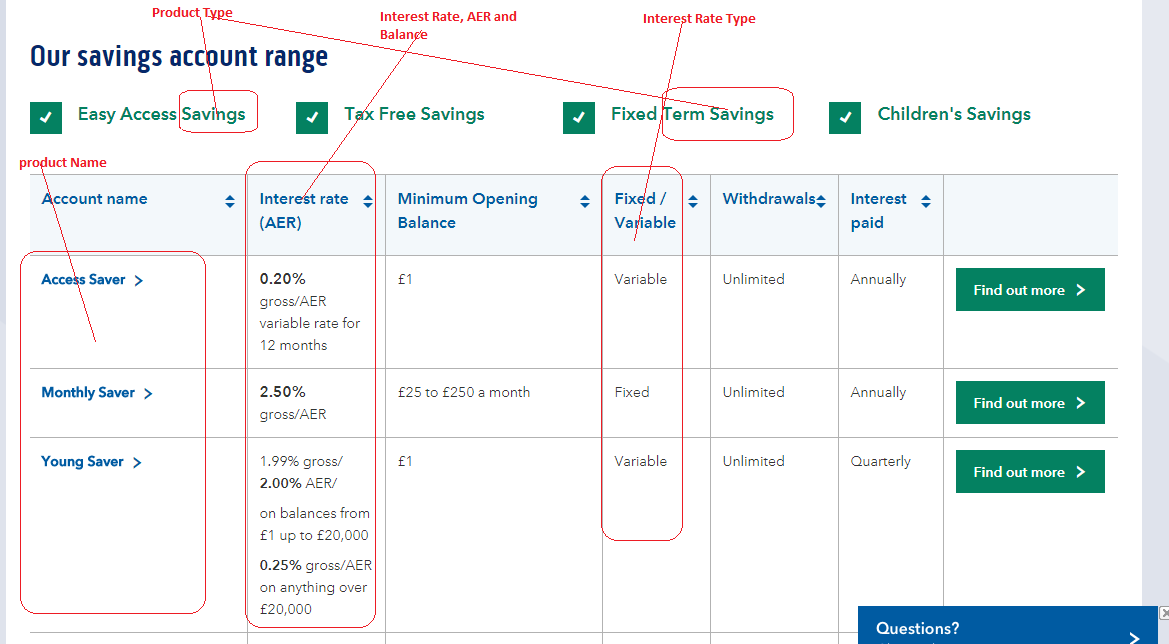
**Scarped Data Format (File Name - Santander\_Data\_Deposit\_<YYYY\_MM\_DD>.csv)**

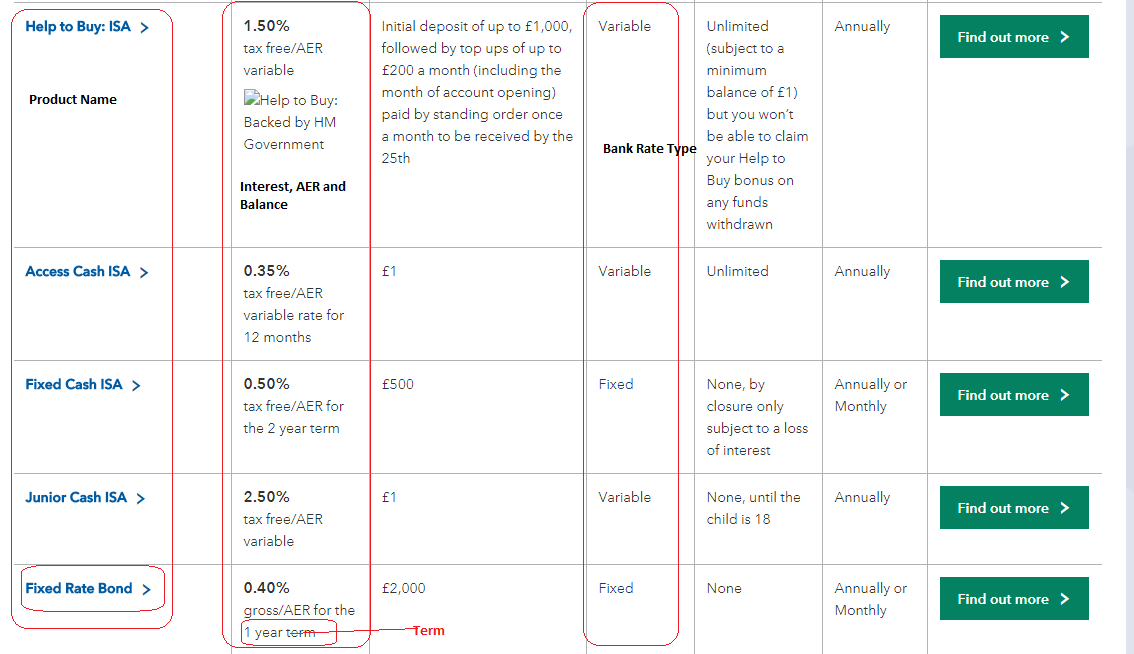
|  |  |
| --- | --- |
| **Column NAMES for Deposits** | **Default Values** |
| Date | Date(YYYY-MM-DD) |
| Bank\_Native\_Country | UK |
| State | London |
| Bank\_Name | Bank of Scotland |
| Bank\_Local\_Currency | GBP |
| Bank\_Type | Bank |
| Bank\_Product | Deposits |
| Bank\_Product\_Type | SCRAPED 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 |
| Interest | SCRAPED DATA |
| AER | SCRAPED DATA |
| Bank\_Product\_Code | LEAVE THIS COLUMN BLANK |

**PDF:** [**https://www.bankofscotland.co.uk/assets/pdf/savings/pdf/interest-rates.pdf**](https://www.bankofscotland.co.uk/assets/pdf/savings/pdf/interest-rates.pdf)

**OR**

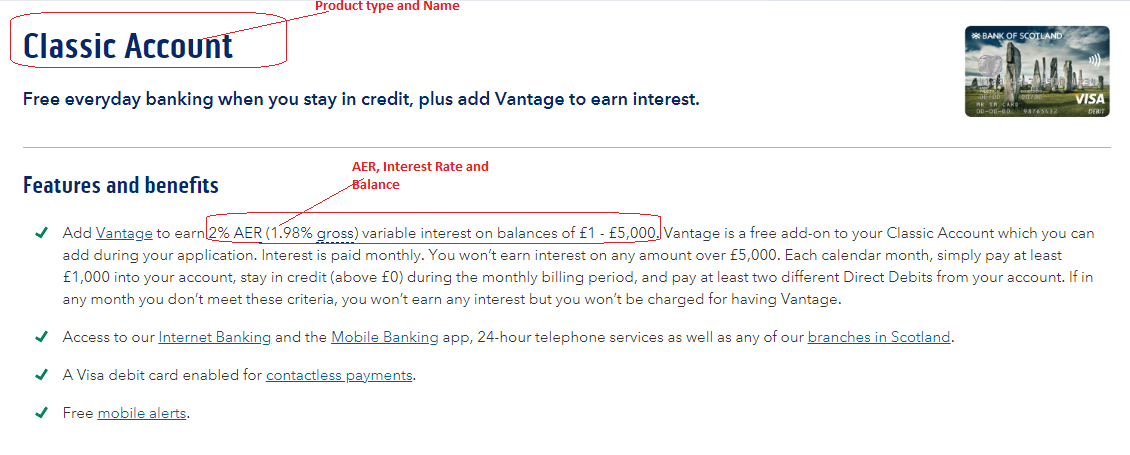
<https://www.bankofscotland.co.uk/savings/>

****

****

**Current Account:**

[**https://www.bankofscotland.co.uk/bankaccounts/classic/**](https://www.bankofscotland.co.uk/bankaccounts/classic/)

****

Work 2: Mortgage Rates

Part of Halifax bank mortgage (NO NEED TO WORK)