**PDF parsing for US CDO OC**

US CLO deals, document type: OC

|  |  |
| --- | --- |
| **Problem 1** | How to use Tabula in our Python program. |
| *Solution* | Not solved |
| **Problem 2** | How to detect which table is the required one using tabula? |
| *Solution* | In general the required table is the first table in pdf |
| **Problem 3** | Define connection rules between excel template and pdf-csv table. |
| *Solution*  Excel file names | **The ready code you can find** [**here**](https://github.com/AnoushAtayan/ManipulatingWithPDF/blob/master/My%20First%20Project.py)  Description of rules |
| Name | In the first row of the table search ‘Notes’ keyword, and all the cells which contain the keyword take as a value of name. Before that, make the following changes: throw ‘class’. ‘notes’ words and all spaces and dashes, if available. |
| Original amount | Search for ‘ principal amount ‘ in the first column and corresponding cells’ fill as original amount . It should be integer. Make sure that the value does not contain currency signs. |
| Currency | Fill in USD by default |
| Fixing deal currency | Fill in 1 by default |
| Tranche Type | Search in the first row ‘comb’ and ‘comp’ strings, if no result, fill in “Standard”, otherwise “Combo” |
| Interest deferral | Search ‘defer’ in the first column, if the corresponding row’s cell value is ‘No’ fill ‘No’ otherwise ‘Pik’ |
| Initial rating (M) | Search for ‘moody’ in the first column, take the string, throw all symbols, spaces and (sf) or sf if available and the remaining part take as a value |
| Initial rating (SNP) | Search for ‘S&P’ in the first column, take the string, throw all symbols, spaces and (sf) or sf if available and the remaining part take as a value |
| Initial ratings (F) | Search for ‘Fitch’ in the first column, take the string, throw all symbols, spaces and (sf) or sf if available and the remaining part take as a value |
| **Problem 4** | Testing for several documents |
| *Solution* | In process |