**OCBC GROUP – TEXT MINING HANDS-ON SESSION**

Hands-on session consists of two parts.

Part-1: Write a python script to extract text attributes from **Sample term sheet 2016.docx** file and save the result in .txt format

Part-2: Write a python script to read the attributes in text file generated in part 1 and **merge** the attributes into a template document [‘**final\_template.docx’**]

**Part-1**

**Input File name:** **Sample term sheet 2016.docx**

Required key attributes to be extracted are marked in Yellow colour in above mentioned input docx file for reference,

1. Below are the trade attributes to be extracted and store into text file in single line delimited by “|”,

* # of Months
* OCBC Reference
* Trade Date
* Initial Valuation Date
* Effective Date
* Notional Amount
* Fixed Rate

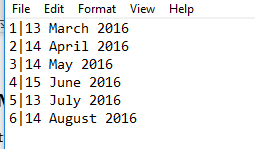
**Expected Output text file ‘key\_attributes.txt’ should be generated as seen below:**

****

1. Below are the payment attributes to be extracted and store into text file as multiple lines delimited by “|”,

* (t)
* Fixed Rate Payer Payment Date(t)

**Expected Output text ‘payment\_date.txt’ should be generated as seen below:**



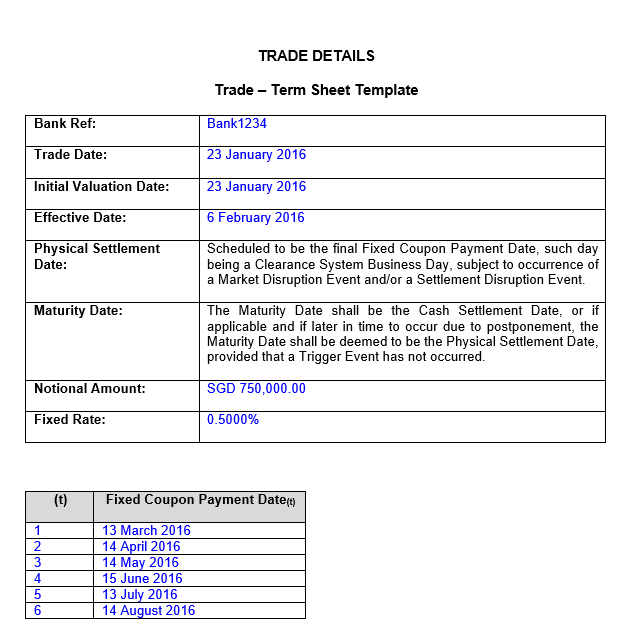
**Part-2**

You will be able to do this phase only if phase 1 is completed.

The output text files generated in part 1 are used as input file for this part 2.

Write a python script to read **‘key\_attributes.txt’** and‘**payment\_date.txt’** filesandmerge the extracted attributes into template document ‘**final\_template.docx’**

**Expected Output: Updated** ‘**final\_template.docx’ document should be seen as below:**



Please reply back with scripts for part-1 and part-2 and its corresponding output within 24 hours from the time of receiving e-mail. In case of clarifications, you can reach out to Muthu (+65-81324400) only during 7AM-7PM.