Assignment

Please provide the Python file for every question.

Q1. Parse all the JSON files in the given folder name ie. **Json_dump** and extract the following columns.

ColumnName	Look Up Value in JSON
ASIN	JSON FILE NAME
MRP	price
brandName	brandName
ProductName	itemName
browseNode	browseNode

All the result should be compiled in a single excel sheet and where columnName is not found represent as 'NA'.

Q2. Perform the following Data Cleaning and Manipulation Techniques Using Pandas.

Filename – data_dump.xlsx

- Given **purchase-date** column is in UTC time format. You need to convert it to IST then split the column into two named as **date_stamp** and **time_stamp**.
- Bucket the time_stamp into given intervals which are as follows:
    - 00:00:00 - 08:00:00 – **bucket1**
    - 00:00:00 - 16:00:00 – **bucket2**
    - 00:00:00 - 23:59:59 – **bucket3**

  Every bucket will be in separate column.

- Obtain total sales state wise and ASIN using the above data.

Note – perform necessary data cleaning methodologies.

Sample Example

	asin	ship_state	item_price	date_stamp	time_stamp	bucket1	bucket2	bucket3
0	B0B1DYTPXB	MADHYA PRADESH	399.0	2022-07-06	18:54:56	0	0	1
1	B0B1DYTPXB	HARYANA	549.0	2022-06-13	05:18:04	1	1	1
2	B0B1DYTPXB	RAJASTHAN	449.0	2022-07-13	11:32:27	0	1	1

Q3. Extract Brand Name from the Column named 'search_term' in the given file.

File Name- keyword_data.xlsx

Brand Names are provided in the sheet('**brandName'**), if the search_term doesn't have any brand mark it as '**generic_term'**. The final output should have same number of rows as the original file.

Also, include all the variations of the provided brands that you can find in the search_term,

Ex. Search_term like 'kelocks cornflakes muesli', 'kelogs','kelox' etc that comes under brand_name = 'Kellogg"s'

Note : You can use any ML Algorithm.

Final Columns: search_term, brand_name

Sample:

search_term	brand_name
Kelloggs cornflakes	kelloggs
Cornflakes under 500	generic_term