Centre for Data Science
Institute of Technical Education & Research, SOA, Deemed to be University

Data Science Workshop-1
(CSE 2195)

ASSIGNMENT-7: Data Loading, Storage, and File Formats

1. Write a function in python to create a text file named as 'states.txt' and read its contents line by line and display the same on screen using- 'with' statement and without it.

2. Create a 'sales' Data Frame consisting of 5 rows and the following column headings- product_id, product_name, sold_units, and profit. Sort the frame by 'product_id'. Export it to CSV file using panda's Data Frame delimited by '|' with column headings- id, name, units, and gain. Read and display the first 3 rows of the same file.

3. Write a Python program to convert Python dictionary object (sort by key) to JSON data. Print the object members using indent level 4.

4. Write a program to create a JSON file with the following columns- name, branch, roll_number, and cgpa. Read the JSON file into a Data Frame and insert a new column 'grade' according to 'cgpa'-

    - If icgpa <= 2 then grade='D'
    - If cgpa > 2 & cgpa <= 5 then grade='C'
    - If cgpa > 5 & cgpa <= 8 then grade='B'
    - If cgpa > 8 then grade= 'A'

    Save the updated Data Frame into JSON, HTML, XML, pickle, and Excel file formats. Read the contents of saved files into a Data Frame object and display it.

5. Write a function to search and display details of student whose grade is 'A' from the binary file student.pkl having structure [name, branch, roll_number, cgpa, and grade]

6. A binary file "Book.pkl" has structure [BookNo, Book_Name, Author, Price].

    - Write a user defined function createFile() to input data for a record and add to Book.pkl.
    - Write a function countRec(Author) in Python which accepts the Author name as parameter and count and return number of books by the given Author are stored in the binary file "Book.pkl"