**Assignment No. 1**

**Name: Om Sudhir Waghmare**

**PRN No: 122B1F134**

**Problem Statement: Reading and Writing Different Types of Datasets**

**Objective:**  
The objective of this assignment is to familiarize ourselves with reading and writing different types of datasets such as .txt, .csv, and .xml from both web sources and local disk storage. This task will involve loading datasets into memory, processing the data using Python, and saving the processed datasets to a specified location on the disk.

**Prerequisites:**

1. **Python environment** with the necessary libraries:
   * **pandas** for reading/writing .csv and .txt files.
   * **xml.etree.ElementTree** for handling .xml files.
   * **requests** for fetching datasets from the web.
2. **Internet connection** (for downloading datasets from the web).
3. A **text editor** for writing and running Python scripts.
4. **Basic knowledge** of file handling in Python (e.g., opening, reading, writing files).

**Theory:**

Different types of datasets require different approaches for reading, processing, and writing. Below are common formats we encounter:

1. **Text Files (.txt):**
   * Simple, unstructured text files.
   * Each line of the file is often a new record.
   * Reading .txt involves parsing the text, usually line by line.
2. **Comma-Separated Values (.csv):**
   * A structured format where data is separated by commas.
   * Pandas makes it easy to load and manipulate .csv files.
3. **Extensible Markup Language (.xml):**
   * A hierarchical data structure.
   * We can use the xml.etree.ElementTree library to parse and process XML files.

**Algorithm for Reading and Writing Datasets:**

**Step 1: Read datasets from the web/local disk**

**For .csv file**

*import pandas as pd*

*# Reading from a local disk*

*df\_csv = pd.read\_csv('file.csv')*

*# Reading from a web source*

*df\_csv = pd.read\_csv('http://example.com/file.csv')*

**Step 2: Process the data**

* You can manipulate the dataset based on the requirements. For example, in pandas, you can filter data, remove missing values, or modify the structure.

**Step 3: Write datasets to a specified disk location**

* **Writing .csv files:**

*df\_csv.to\_csv('output.csv', index=False)*

**References:**

* Pandas Documentation
* Requests Documentation
* [xml.etree.ElementTree Documentation](https://docs.python.org/3/library/xml.etree.elementtree.html)

**Conclusion:**

By completing this task, we have become familiar with reading and writing various dataset formats including .txt, .csv, and .xml files from both local disk storage and web sources. These skills are essential for handling real-world data and preparing it for further analysis or storage.