Brief Note:

The attached script, `generate\_excel.py`, is designed to create an Excel file with bold column names and values. Upon execution, the script generates an Excel file named `Output Data Structure.xlsx ` that contains the specified values along with bold column names.

Included in this submission:

a) `generate\_excel.py` - The Python script responsible for creating the Excel file.

b) `Output Data Structure.xlsx` - The Excel file generated by executing the script.

c) ‘Instruction.docx’ - A word file containing brief about the assignment

The script uses the openpyxl library to achieve the desired formatting and population of data.

## **Instructions for Running the Text Analysis Script**

This script is designed to extract text articles from URLs provided in an Excel file, perform text analysis, and compute various variables as outlined in the assignment. The script is written in Python and utilizes libraries such as BeautifulSoup, TextBlob, and others for web scraping and text analysis.

### **Prerequisites**

* **Python**: Make sure you have Python 3.x installed on your machine.
* **Required Python Libraries**: The script uses various Python libraries, including BeautifulSoup, TextBlob, syllables, and nltk. To install these libraries, you can use the following commands:

pip install beautifulsoup4 textblob syllables nltk

* **Input Excel File**: Place the input Excel file named **Input.xlsx** in the **Resources** folder. This Excel file should contain at least two columns: **URL\_ID** and **URL**. Each row should correspond to a URL you want to extract and analyze.
* The script will start processing each URL from the input Excel file. It will extract article text, perform analysis, and save the results in the `Output Data Structure.xlsx’ file present in the Output folder.

Please feel free to run the script and examine the generated ` Output Data Structure.xlsx ` file.

Thank you!

Subham Dutta

rupaidutta66@gmail.com