**PROJECT – EXCEL SPREADSHEET CONASOLIDATION AND DATA EXTRACTION**

**CONFIDENTIAL: CONFIDENTIALITY AGREEMENT MUST BE SIGNED**

**DATE:** 12/30/2024

**GOAL:** Extract relevant text messages from numerous excel files using search terms and put into a single excel file

**DEADLINE:** As soon as possible (2-3 days)

1. **Structure of existing data:**

We have text messages from two cell phones in excel and CSV format. The the file structure is as follows:

* Stevens Iphone
  + Iphone
    - Messages
      * 2024-04-04[[1]](#footnote-1) 18 16 21[[2]](#footnote-2) – William Shakespeare[[3]](#footnote-3)
      * 2022-01-01 08 59 59 – Joseph & John
      * …
      * …
      * …
      * …
      * 2023-10-10 12 14 10 – Jack & Jill
* Johns Iphone
  + Iphone
    - Messages
      * …
      * …
      * …
      * …

Each of the sub-folders has a CSV and Excel file named as follows:

Messages – 2024-09-27 01 16 21[[4]](#footnote-4) – Jack & Jill.csv

Messages – 2024-09-27 01 16 21 – Jack & Jill.xls

1. **Sample Files**

I have attached the following SAMPLE files:

1. Messages - 2024-09-27 01 16 21 - Joseph & John (Input 1).xls

*This is an example of a single csv/xls file containing a single text thread.*

1. Messages - 2024-09-27 01 16 21 - Jack & Jill (Input 2).xls

*This is an example of a single csv/xls file containing a single text thread.*

1. Combined Spreadsheet (Pre-Search)

*This is an example of the first output – a spreadsheet containing all text threads*

1. PREFERRED OUTPUT

*This is the final output that we want.*

1. **Description of Assignment**

Consolidate all of the text threads (rows) contained in each file into a single excel spreadsheet. Then run a series of searches for specific text found in the **text column. (**In the attached example files the search term was “**memor!”).** We want to keep the 4 texts ***before***each search term hit and the 4 texts ***after*** each search term hit to give the text message some context. The final result will be a single spreadsheet for each

1. This is the date the text thread started [↑](#footnote-ref-1)
2. This is the time the text thread started [↑](#footnote-ref-2)
3. This is the person or person(s) who were in the thread (or group thread) (besides Steven) [↑](#footnote-ref-3)
4. This is the date on which the data was extracted. [↑](#footnote-ref-4)