**dataScript**

**v1.4**

Contents

[Installation 1](#_Toc526233952)

[Requirement 1](#_Toc526233953)

[Set-up 1](#_Toc526233954)

[Usage procedure 1](#_Toc526233955)

[Notice 1](#_Toc526233956)

[Issues 1](#_Toc526233957)

[Contact me 1](#_Toc526233958)

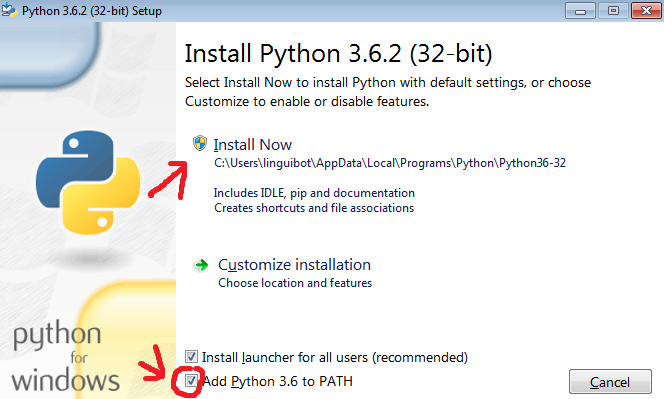
# Installation

**For windows user:**

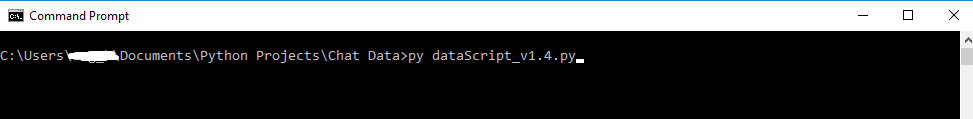
1. The 1st way to do it is to unzip the dataScript\_v1.4(Windows Only).rar file and create a shortcut of dataScript\_v1.4.exe application to a desirable location to run.

**For Mac, GNU/Linux, or Windows user:**

1. Install Python 3.6.2 from the official website (<https://www.python.org/downloads/release/python-362/>) (figure 1). Remember to check “Add Python 3.6 to PATH” before clicking “Install Now”.
2. You can now create shortcut of dataScript\_v1.4.pyc and run it.
3. Alternatively, you can run dataScript\_v1.4.py from the cmd or terminal by navigating to the folder that contains the script and typing “py dataScript.py” or “python dataScript.py” and press *Enter* (figure 2).
4. If you are unsure of what version of python your computer is running on, you can type in “py –version” or “python –version” and press *Enter*.



(Figure 1)

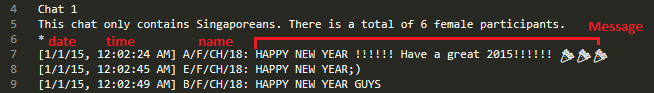


(Figure 2)

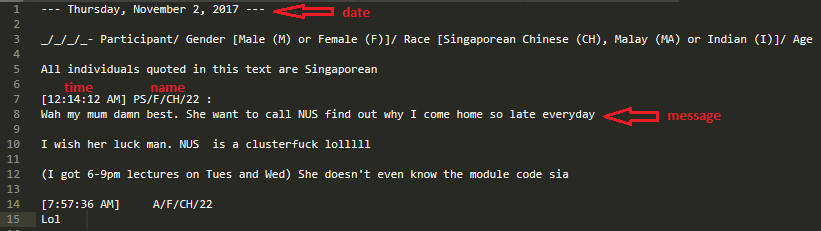
# Requirement

In this section we will be talking about the important requirement of the data (text file) that can be processed by the dataScript.

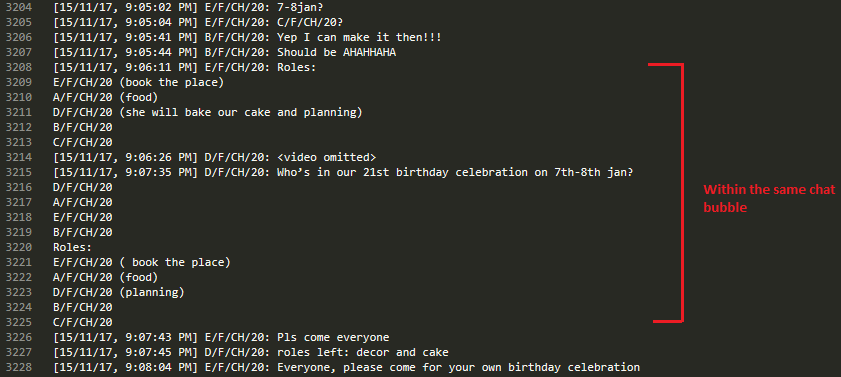
1. There should be a clear difference between name, date, and start of the chat (figure 3). Figure 4 shows a bad example this. The name, date, and start of the message have to be in the **same** **row**. Having different line is fine as long as the start of the message is in the same row as the information of the chat bubble (figure 5).



(Figure 3)

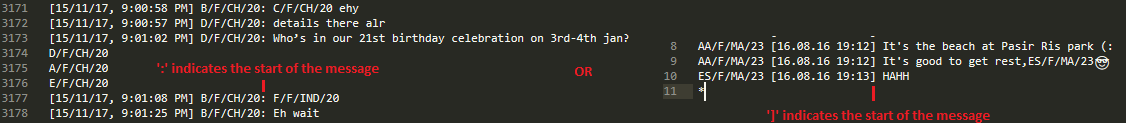


(Figure 4)

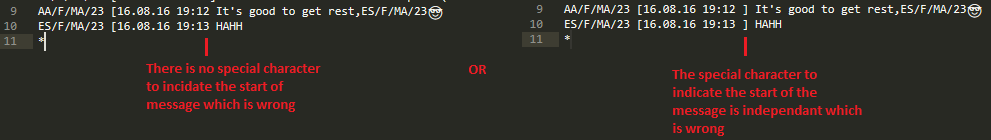


(Figure 5)

1. There should be a punctuation or a special character that indicates the start of the chat and is has to be in part of the last word before start of the message (figure 6). Figure 7 shows a bad example.



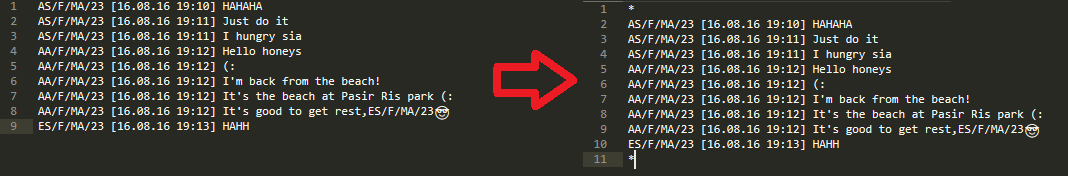
(Figure 6)



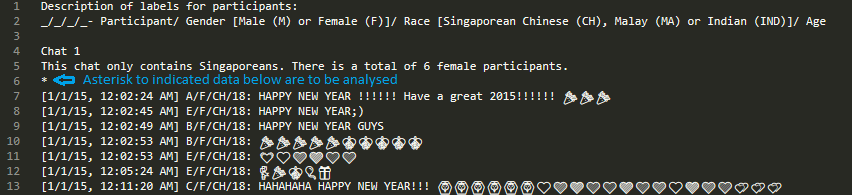
(Figure 7)

# Set-up

1. Set up a folder to contain the script and data (text file) for easy access.
2. Copy the data (chat) into a new text file or use the existing file that contain the data.
3. Include asterisk only onto the start and end of the chat for the script to know where the start and end is for data extraction and analysis (figure 8). You can add descriptions before or after the start of the data to be analysed as long as it is not inside the asterisks (figure 9).
4. You are now ready to run the script. Double click on the script to start.



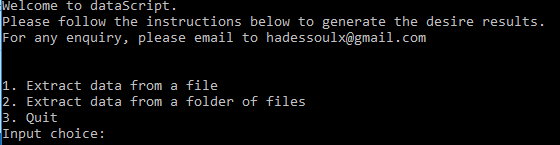
(Figure 8)



(Figure 9)

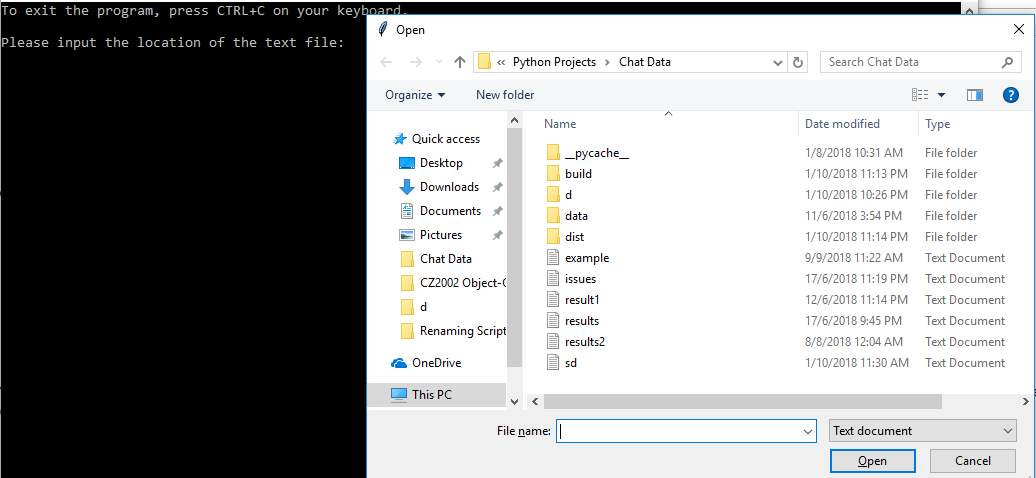
# Usage procedure

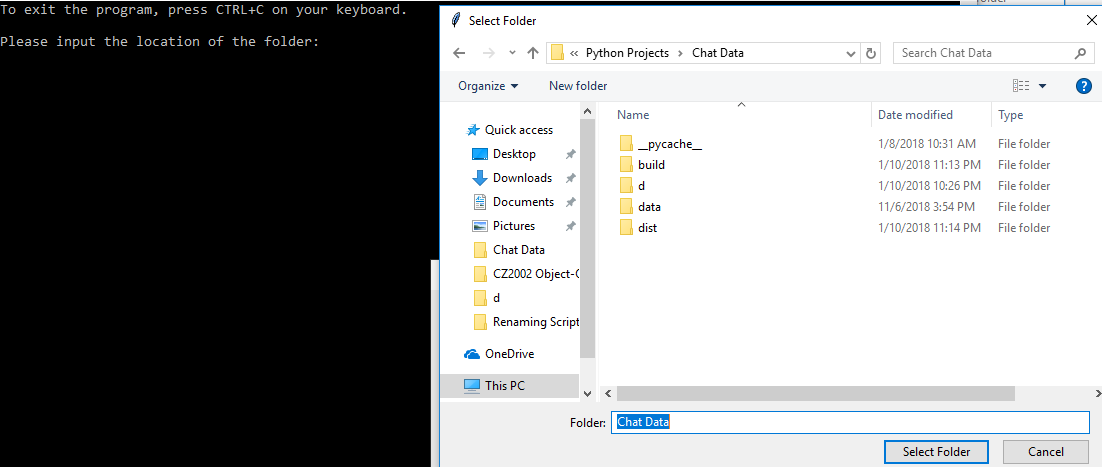
1. At the start of the program, it should display the main page (figure 10). Input **1** and press *enter* to extract data from a file. Input **2** and press *enter* to select a folder to extract data from. To quit, input **3** and press *enter*.



(Figure 10)

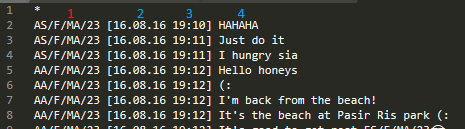
1. After selecting the 1st or 2nd option, you will be prompted to select the location of the text file or folder to be analyse (figure 11).



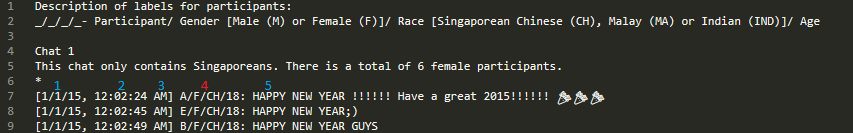


(Figure 11)

1. You will be prompted to input the position of the name in the data. In figure 12, the name is in position **1**. In figure 11, the name is in position **4**. As long as the character(s) is/are separated from other character(s) through spacing, it will be counted as a word just like **AM** which is in position **3** (figure 13).

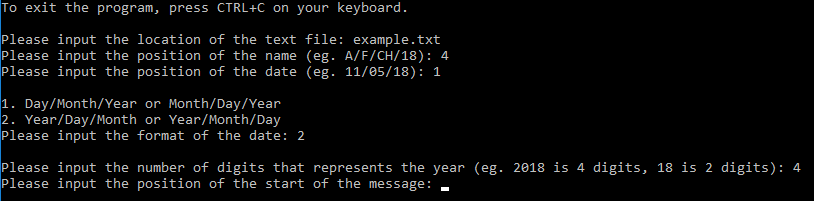


(Figure 12)



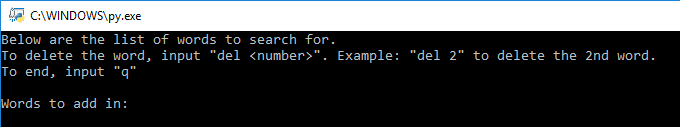
(Figure 13)

1. You will then be prompted to input the position of the date. In figure 12, the date will be in position **2**. In figure 13, the position will be in position **1**.
2. Next, you will be prompted to input the date format. Ignore the punctuations or separators that separating the day, month, and year. The results will be the same. Choose the option which the year is at the start or the end like the date that is in the text file (data) that you want to extract. Example if the date is 2016/05/12, choose option 2 (figure 14).
3. You will then be prompted to number of digits of the year. If your data is showing 2018/05/12, the digit will be **4** due to 2018. If it’s 18/05/12, it’s **2** due to 18.
4. Next, you will be promoted to input the position of the start of the chat (figure 14). In figure 12, the start of the message is in position **4**. In figure 13, the start of the message is in position **5**.

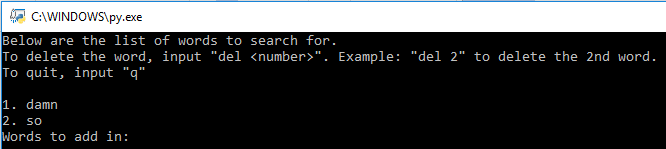


(Figure 14)

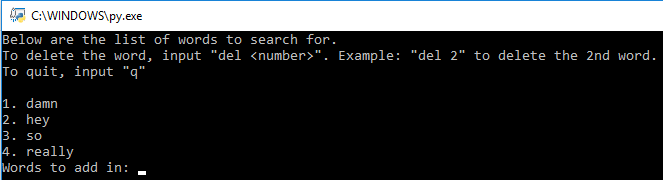
1. Next, you will be prompted to input the words you want to track the number of times it is produced each year (figure 15). You can key in a list of words that you would like to get its data. For example, if we want to count “damn” and “so” produced each year, we will key in **damn**, press *enter*, key in **so**, press *enter*. The result is shown in figure 16. If there is any input you would like to delete, input **del <number>**. In figure 17, we wants to delete the 2nd word, we will input **del 2** and press *enter*. To exit, input **q**, and press *enter*.



(Figure 15)

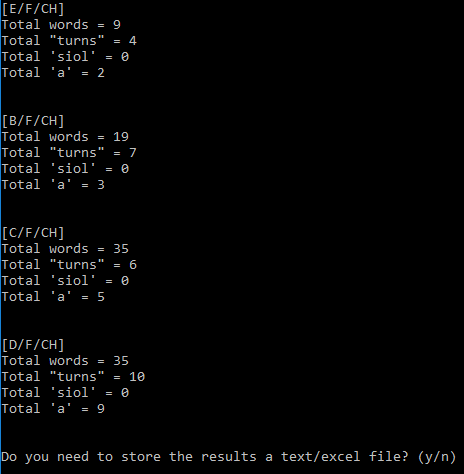


(Figure 16)



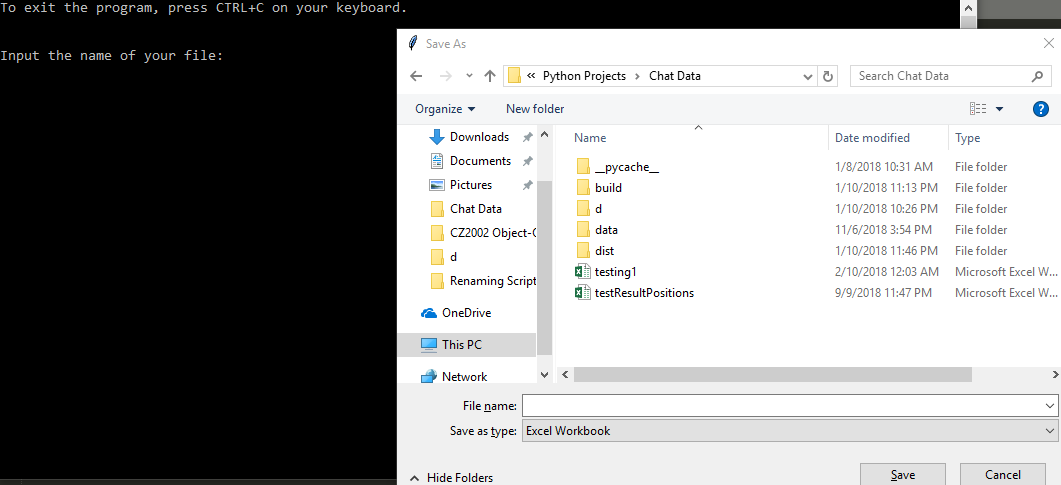
(Figure 17)

1. The results will then be displayed in the console if you had selected option 1 (figure 10) (figure 18). You will be prompted if you would like to save the results into a text/excel file. For this example, we will input **y** which stands for yes.



(Figure 18)

1. Create a name for a text/excel file which the results will be stores in that file (figure 19). You can select the file type to be stored. There are .txt, .xls (Excel 2003 or older), and .xlsx (Excel 2010 or newer). Do not worry about overriding existing ***excel files*** as new content will be appended to the ***excel file*** instead of overriding it (Take note, this does not apply to text files).

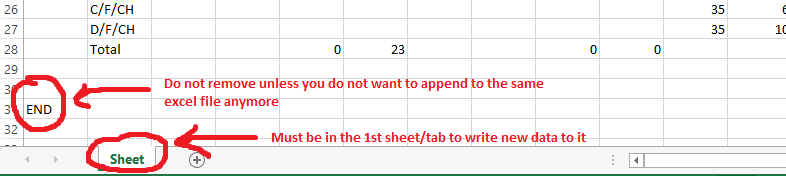


(Figure 19)

1. You will now be brought back to the main menu. To extract and analysis the data of another file, select option **1** or **2** again. Else, select option **3** to exit the program.
2. At any point of time in the program, you can press *CTRL+C* or click on the close button to exit the script.

# Notice

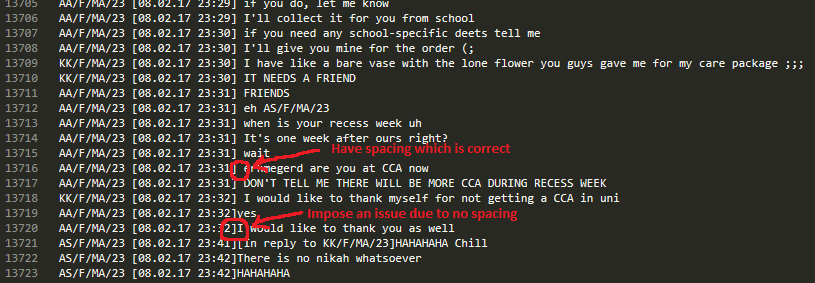
1. All punctuations except for **/** and **',** will be removed when counting the number of words. **AA/F/MA/23** will be counted as a word. **It's** will be counted as a word. **It's good to get rest,ES/F/MA/23** will be counted as 6 words because , will be replaced with a space by the script when counting the number of words. **Hi !!! =)** will be counted as one word.
2. All emoji will not be counted as words.
3. Do not remove “END” word from the excel file as it is used by the script to identify which position to continue writing to the file unless you have no intend to append to the same excel file (Figure 20).
4. To append to the same sheet in the excel file, it must be the 1st sheet/tab in the excel file (Figure 20).



(Figure 20)

# Issues

1. If the chat message uses **’** instead of **'**, there will be issue. E.g "Who's" is treated as one word. However, "Who’s" is treated as two word (who s).
2. In data file, *2018-MAL\_F2\_R.txt*, there is no spacing in certain section of chat 3. This will impose an error as there should be a space between the information of the chat bubble and the start of the start bubble (figure 21).
3. Words that differs from the original word will not be detected. Example “hiiiiiiiiiiiiii” will not be detected as “hi”. Do not worry about case-sensitive as I have already taken that into consideration.



(Figure 21)

# Contact me

If there are any issues or additional features that you would like me to add in, please contact me at [hadessoulx@gmail.com](mailto:hadessoulx@gmail.com), send me a screenshot of the issue if there is, inform me the name of the script/software. Thank you.

**The End**