bl\_ms\_search.zip will contain:

bl\_ms\_data\_analysis.py

full-list-digitised-mss-jan-2022.csv

bl\_manuscript\_search\_instructions.docx

out.csv (example of search)

you will need python3 and pandas installed

download the excel file from https://blogs.bl.uk/digitisedmanuscripts/ and save as a cvs file

to run the search, open a terminal (for example terminal on mac, mobaXterm on windows)

for a *quick start* type

*data\_analysis.py 'file.csv 'search\_string' 'exclude\_string'*

file.csv is the file you want to search

search\_string is a single word you want to search for

exclude\_string is a single word to exclude a manuscript OR type 'none'

for *interactive process* with more options type

*data\_analysis.py 'file.csv'*

the program will ask for a list of single words to include and another list of single words to exclude

for example:

*data\_analysis.py full-list-digitised-mss-jan-2022.csv hours leaves*

will search the list for all manuscripts with hours in the comment column and after finding all those with "hours" will exclude all those that have the term leaves (ie those with only individual pages)

the output will look like this

[tricia@] $ data\_analysis.py full-list-digitised-mss-jan-2022.csv hours leaves

input file is full-list-digitised-mss-jan-2022.csv

search string is ['hours']

exclude string is ['leaves']

starting dimensions (rows, columns)=(4808, 3)

including hours

after search (rows, columns)=(67, 3)

excluding leaves

after exclusions (rows, columns)=(66, 3)

new file is called out.csv

for example:

*data\_analysis.py full-list-digitised-mss-jan-2022.csv*

will search the list for all manuscripts with hours in the comment column and after finding all those with "hours" will exclude all those that have the term leaves leaf binding bindings

the output will look like this, out.csv provided as an example.

[tricia@]/ $ data\_analysis.py full-list-digitised-mss-jan-2022.csv

type terms to search, (no strange characters) hours

hours

number of terms is 1

type terms to exclude, (no strange characters) leaves leaf binding bindings

leaves leaf binding bindings

number of terms is 4

starting dimensions (rows, columns)=(4808, 3)

including hours

after search (rows, columns)=(67, 3)

excluding leaves

excluding leaf

excluding binding

excluding bindings

after exclusions (rows, columns)=(51, 3)

new file is called out.csv