```
=== readme-prelim01.htm ===

Please Verify you can do the filtering/analysis with with the excel file created so far.
And maybe provide some example results.

You should be able to open 1629104-excel.xlsx in excel directly, without opening the csv at all.

The csv file would eventually grow to over 122x10^6 cassettes, and then be unwieldly or impossible to do in excel.
At that point, we'll have to do the equivalent analysis with additional command-line scripts.


Note: For tricks to get excel to open csv with millions of rows
see https://www.masterdataanalysis.com/ms-excel/analyzing-50-million-records-excel/


This folder contains the output for 1629104 (~ 1.6 x 10^6)
cassettes, in 3 file formats: csv, xlsx (excel) and gnumeric.
That is, the exact same date should be in all 3 1629104 files.

-rw-rw-r-- 1 aeh-group aeh-group 158151200 Apr 25 11:34 1629104.csv
-rw-rw-r-- 1 aeh-group aeh-group  89673482 Apr 25 11:46 1629104.gnumeric
-rw-rw-r-- 1 aeh-group aeh-group  84948352 Apr 25 11:54 1629104-excel.xlsx
-rw-rw-r-- 1 aeh-group aeh-group      8860 Apr 25 12:08 readme-prelim01.htm  (this file)

From the output of htop (below), so far it looks like the bottleneck is cpu, not ram. It may also be i/o.
We'll try adding the onecsv functionality to mcff2.c to try to reduce cpu usage, context-swaps, and generally streamline things.

  1  [||||||||||||||||||||||||||||||||||||100.0%]   4  [||||||||||||||||||||||||||||||||||||100.0%]
  2  [||||||||||||||||||||||||||||||||||||100.0%]   5  [||||||||||||||||||||||||||||||||||||100.0%]
  3  [||||||||||||||||||||||||||||||||||||100.0%]   6  [||||||||||||||||||||||||||||||||||||100.0%]
  Mem[||||||||||||||||||||||||||||9.24G/15.7G]   Tasks: 515, 1178 thr; 6 running
  Swp[|                           21.5M/36.1G]   Load average: 130.05 125.20 123.53
                                                 Uptime: 19 days, 19:06:58

  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
25515 aeh-group  20   0 24052  5944  3244 R 83.3  0.0  0:00.17 htop --sort-key=PERCENT_CPU
15718 aeh-group  20   0 37320 13632  5564 R 16.7  0.1  0:00.10 python3 ./onecsv9.py - -
18850 aeh-group  20   0 36236 12740  5564 R 16.7  0.1  0:00.11 python3 ./onecsv9.py - -
21236 aeh-group  20   0 35648 12208  5752 R 16.7  0.1  0:00.10 python3 ./onecsv9.py - -
25212 aeh-group  20   0 15520  9960  2040 R 16.7  0.1  0:00.05 mcff -v -n  001637389>GGAAACAAaaggaug
25451 aeh-group  20   0 14992  9084  1668 R 16.7  0.1  0:00.02 mcff -v -n  001637548>GGAAUACCaaggaug
25453 aeh-group  20   0     0     0     0 R 16.7  0.0  0:00.02 mcff -v -n  001637599>GGAAUUUCaaggaug
25463 aeh-group  20   0 15520  9960  2040 R 16.7  0.1  0:00.07 mcff -v -n  001637589>GGAAUUACaaggaug
25505 aeh-group  20   0 14860  9080  1668 R 16.7  0.1  0:00.01 mcff -v -n  001637593>GGAAUUCGaaggaug
25518 aeh-group  20   0 14860  9104  1684 R 16.7  0.1  0:00.01 mcff -v -n  001637612>GGACAAGUaaggaug
```