**B8IT105 - CA5**

**John Fraser (1036455)**

**23 July, 2017**

*Overall notes – I used the “process\_changes.py” program from Moodle, adjusting slightly and re-naming as “CA5\_process\_changes\_JFraser.py” to create the file “CA5\_changes\_JF1.csv”. However, this contained 425 records, rather than 422 (excluding header line), because of an overflow problem. Also, some of the “lines” (see below) were placed in an incorrect column. In the end, this did not affect my analysis below, but to demonstrate this, I converted the .csv file to excel, adjusted for the above issues and then reconverted to .csv: “CA5\_changes\_JF2.csv”. The column headed “Files Changes” was pushed rightwards by one column in the latter. This latter data was then loaded by “CA5\_analysis\_JFraser.py” which produced the analysis outlined below.*

*The only test that I carried out was a line near the end of the analysis program which summed the number of operations attributed to each author - see “check\_count\_authors” – and produced a result of 422, the actual total.*

**CA5 – Statistical Analytics Conclusion:**

The 3 main pieces of information of potential interest which I identified from an initial perusal of the data were as follows:

1. Number and identity of the individual authors/operators.
2. Number of operations attributed to each author/operator and number of “lines” in excess of a chosen number – also split by operator. (This could be seen as two pieces of information!!)
3. Number of operations carried out on different days of the week.

The output (see sample screenshot from command prompt below) identified 10 separate operators who carried out 422 operations. However, an examination of the names shows one is called “OU/=DomainControl….”, presumably this relates to a number of tasks carried out by anonymous operators who did not record their identity. Of these, the main operators appear to be Thomas and Jimmy, who did 191 and 152 operations respectively, or 343 of the 422 operations recorded (81%). The full list, as established by the program “CA5\_analysis\_JFraser.py”, were Thomas (191), Jimmy (152), Alan (5), “OU/=DomainControl….” (24), Nicky (5), Dave (2), Freddie (7), Murari.Krishnan (1), Vincent (26), ajon0002 (9).

The second measure, number of lines, I took as a crude measure of size or difficulty of the job. I set up the program so that the user running it could enter a number and then the number of operations with “size” greater than this counted. Re-running a number of times, entering different critical numbers, established that there were 49 records/operations in excess of “length” 10, 23 of these were in excess of 20, 12 in excess of 30, 6 in excess of 100 and 3 (these all between 250 and 300) in excess of 200.

The program was also set up to establish how many “excess” operations were attributed to each author/operator (see the “cross\_count” variable in the program). Examining percentages more closely, Thomas and Jimmy accounted for 21 of the 23 that were in excess of 20. So 91% of the excess numbers, compared with 81% of the overall. For in excess of 30, the calculation was 10/12 = 83% (see screenshot below). Similar calculations for other excess figures show that Thomas and Jimmy tended to do more of the “large” operations, lending further possible evidence to the theory that they are the senior operators.

For days of the week, the output show the following numbers:

Monday 53

Tuesday 80

Wednesday 76

Thursday 118

Friday 95

From this, we can see that Thursday is the busiest day of the week by some distance. For example, it

is more than twice as busy as Monday.

