**Programming for Big Data**

**B8IT105**

**Darren Redmond**

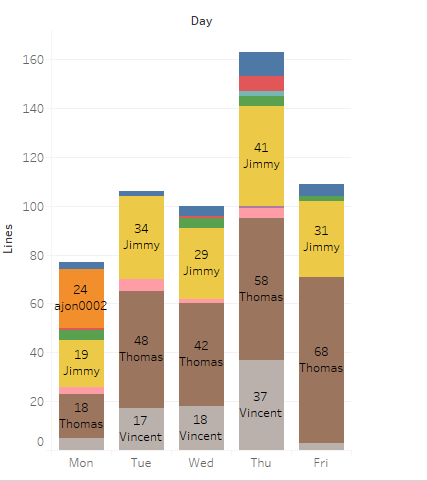
**CA 5**

**TJ Cronin**

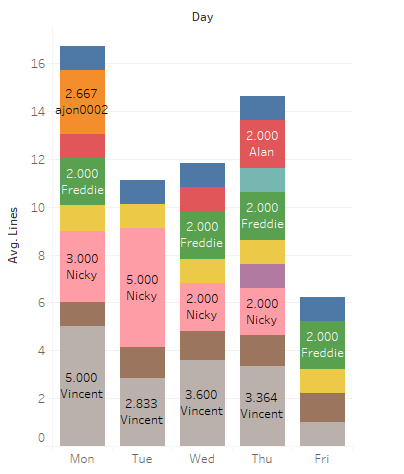
**10364180**

After cleaning the data from ‘changes\_python.log’, using ‘process-changes.py’, I was left with ‘changes.csv’. This csv file needed further rearranging however as some of the lines in ‘Files Changed’ were occupying multiple lines in the file. To rectify this, I opened notepad and pasted all the parts of the offending lines together, replaced them back into the file in the correct place and deleted the now vacant lines so that there was a total of 422 lines in the file. The ‘Date’ column needed further editing also as it included the ‘date’, ‘time’, ‘day’, ‘calendar date’, ‘month’ and ‘year’. I copied the ‘Date’ column and split the original into the above columns using the ‘Text to Column’ options in ‘Excel’. Finally, I ‘delimited’ the ‘commas’ and ‘brackets’ from the new columns and read the new file with additional columns and options for analysis into ‘Tableau’.

With the additional columns, I could now analyse the ‘Authors’ ‘Lines’ by time, day, month or calendar date. Below is a ‘bar chart’ which shows how many lines were done by day. The day is subdivided by author using different colours. There are some very interesting titbits to the graph below, starting with the fact that Thursday seems to be the most productive day by some distance. Monday is the least productive way which indicates that the office has ‘a case of the Mondays’.

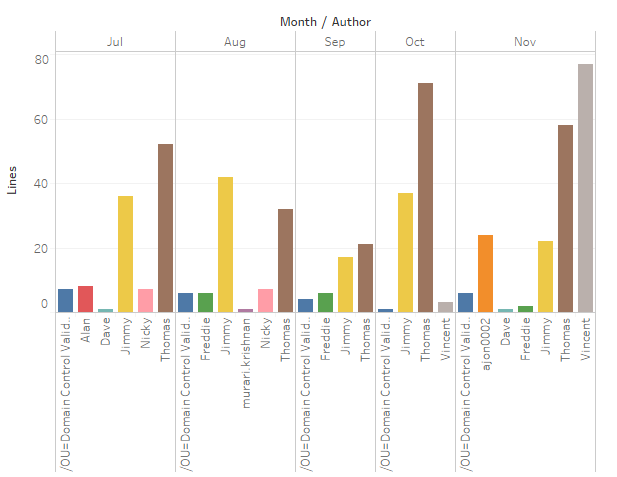


The light blue in Thursday above represents ‘Dave’. I find it interesting that ‘Dave’ only works on Thursday evidently. Similarly, the orange represents ‘ajon002’ who only works Mondays. Perhaps these fellows only work on coding one day a week. It is also clear that ‘Vincent’, ‘Thomas’ and ‘Jimmy’ perform the bulk of the work. ‘Thomas’ gets better as the week progresses, which makes me wonder what is he doing on the weekend? ‘Vincent’ peaks on Thursday before phoning it in on Friday. Only ‘Jimmy’ is somewhat consistent throughout the week, save for Monday when nobody really feels like working. ‘Alan’ represented by red, performed one line on a ‘Monday’, one line on a Wednesday and four lines on a Thursday. ‘Murari.krishnan’ did one line throughout the dataset on a Thursday. The ‘Author’ is shown below ‘Jimmy’ in purple.



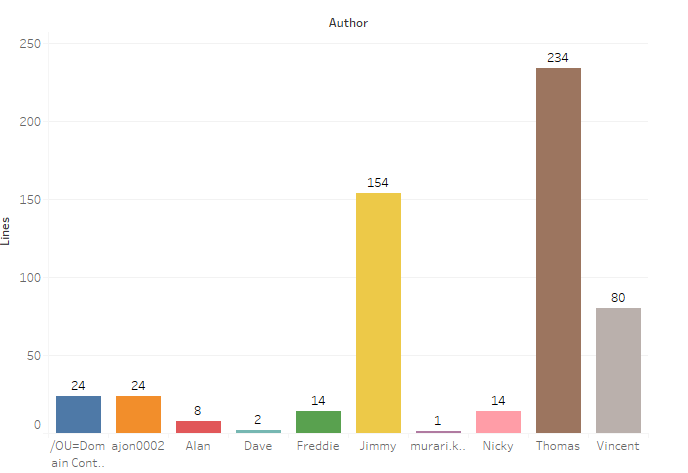
When we switch to average line per day, we get some interesting results. Monday is not a ‘lazy day’ so much as they are not asked to code then. Perhaps they do plans for the week every second Monday and write no lines. When they do code however, everyone is pretty productive. ‘Nicky’ who didn’t seem to be a major factor when counting total lines, has now ascended to the most productive employee alongside ‘Vincent’ on a per day basis when he actually does some work. Vincent, as noted earlier, perhaps has one eye on the weekend on Fridays. ‘Thomas and ‘Jimmy’, again represented by brown and yellow, do the most work in totality, but take a lot of time to do so. Jimmy’s consistency is again prevalent as he does approximately 1 line every day he works. Freddie is also consistent, save for the fact that he doesn’t do Tuesdays.

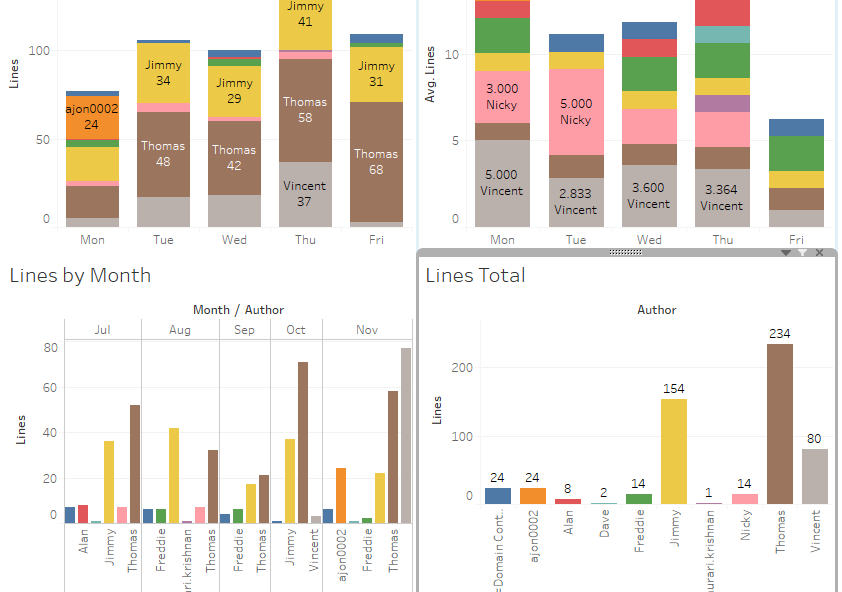
Below, we can see the ‘Lines’ by ‘Author’ split into the ‘Months’ which they were performed. Once more, the work habits of ‘Dave’ are fascinating. From the first graph, we know that he only did 2 lines on Thursdays. Now we can see that those 2 lines were 4 months apart. It appears ‘Freddie’ replaced ‘Alan’ at the beginning of August, perhaps because ‘Alan’ only worked 3 days a week, whereas ‘Freddie’ will do 4. ‘Thomas’ was the most productive and consistent along with ‘Jimmy’, save for September, when everyone was busy getting their children back to school. Thomas’ time as the most productive member of the employees may already have ended due to Vincent’s diligence. Vincent who I assume started in late October, had the most productive November to leave a great first impression on the company.



Using all three graphs, we can tell that Nicky worked a total of 5 days over 2 months, possibly as a contractor, and was never contacted again despite his proficiency. The new man Vincent is primed to be the most productive member of the team, particularly if starts trying on Fridays. Dave is only asked to write a line every other season. We can now tell that ‘Jimmy’ and ‘Thomas’ were simply not asked to do too much in September because we know they do the same amount of work every day they are asked to write some lines.

Below, we can see the total number of lines by all Authors and the dashboard containing all 4 graphs.





Having analysed the employees of the company and their work habits, I find some of the working patterns odd, particularly ‘Dave’ who I assume fills other roles in the company and codes part-time. The company should have utilised ‘Nicky’ better when they had him. ‘Alan’ was replaced by ‘Freddie’, who works 4 days as opposed to 3, but they couldn’t find someone to work all 5 days. ‘Jimmy’ and ‘Thomas’ are the only ever-present employees, which is interesting over a 5-month period and ‘Vincent’ and ‘ajon002’ are two great finds for the company. The turnover of employee in general raises even more questions.