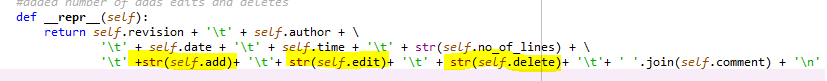
Student No 10388106

Changes Made to existing files

I have changed the how the change\_log.csv is saved, instead of saving as comma delimited it now tab delimited see method “def \_\_repr\_\_(self):” It just makes the loading the changes.csv easier into the data frame as I don't have to worry about comma or doubles and having right number of these, in terms of there being extra in the comments field.

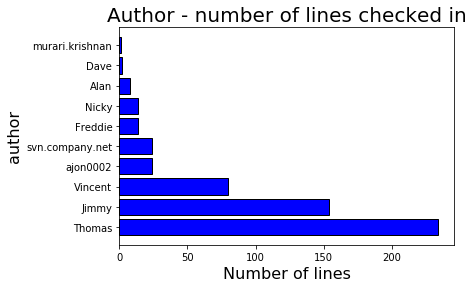
I have also changed the "def get\_commits(data):" method process\_changes\_with\_object.py. I used a replace a piece of coder in the method "details[1].strip().replace('/OU=Domain Control Validated/CN=svn.company.net', 'svn.company.net')," as I wanted to shorten the author name so I could analysis the data a bit easier from a viewing point of view

I also changed the file output as below to show the number of files that have been added, edited, and deleted during the commit. So I added the fields add, delete and edit to the init so we could write out the information from “commit.changed\_path” which had a value at the start of it then files changes in the rest of the field.



Interesting Facts

1: Commits by Users



|  |  |
| --- | --- |
| Author | NoOfLines |
| Thomas | 234 |
| Jimmy | 154 |
| Vincent | 80 |
| ajon0002 | 24 |
| svn.company.net | 24 |
| Freddie | 14 |
| Nicky | 14 |
| Alan | 8 |
| Dave | 2 |
| murari.krishnan | 1 |

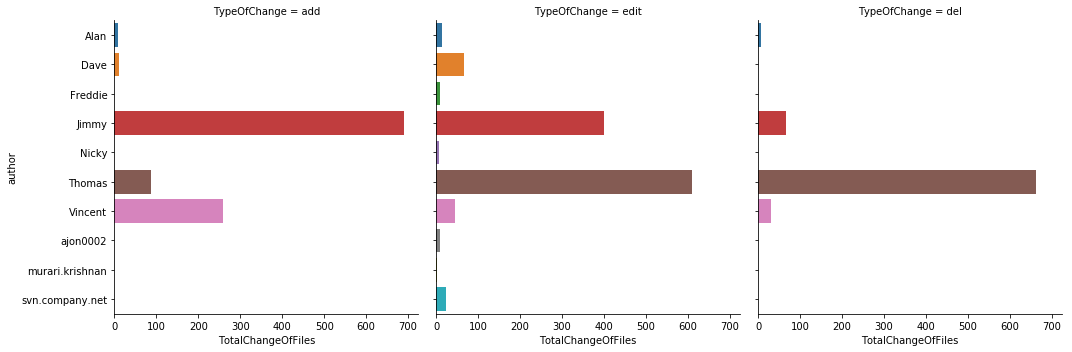
There are 10 developers/authors listed. The svn.company.net user systems to be a system author, it seems to be used for creating automatic releases, we can see an example of this by looking at the comments field eg “gradle-release] prepare for next development iteration”

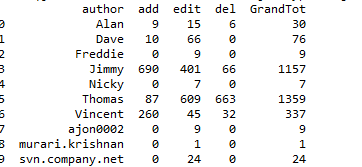
All the other users are only using their first name so maybe they are all employed for quite a while, the user who has supplied there full name maybe they are a new starter where the other are mid level or junior developer/authors.

As we have seen each commit can affect a number of files as shown by add, del edit fields, we will look at this next from the dataset.

The Information above is interesting I suppose because it shows there to be three main developers, Thomas and Jimmy they seem to be the most senior, Thomas is checking in well over two hundred lines of code while Jimmy is quite close to this with over 150 lines of code. The next level down from the senior pairing is Vincent on 80. There is quite big drop off of lines of code being checked in after that to the early twenties and early teens.

2: File Changes by Authors





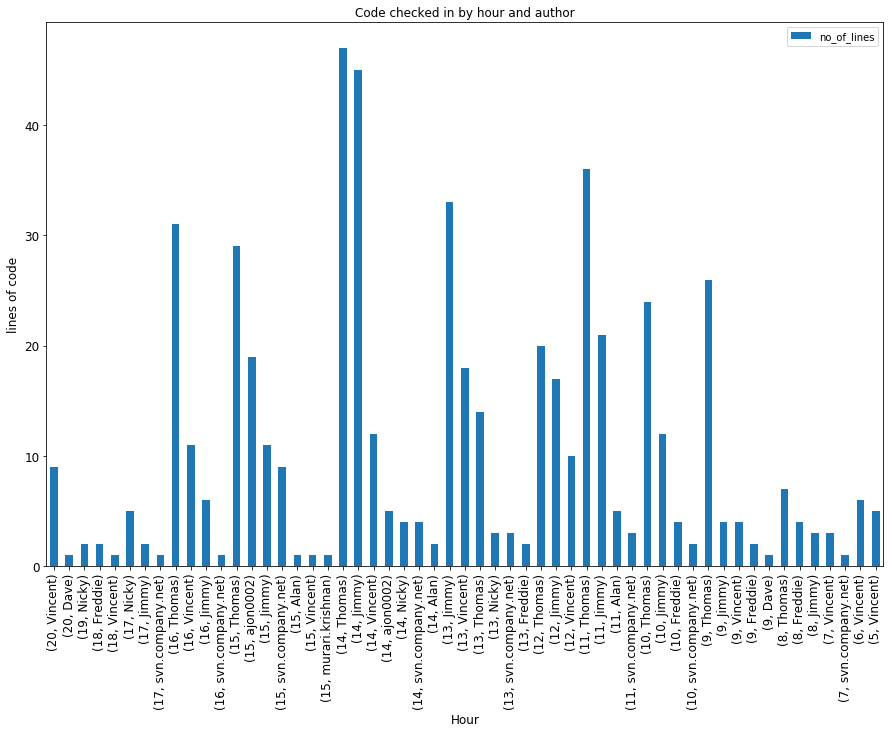
So a commit can be made up by many file changes of either add or delete or an edit.

We have seen that Thomas is the leader, but is he really?

This graph shows a different story, we can see in fact Jimmy is making the most significant code changes in terms of editing and adding whereas Thomas does a lot of deletes.

This data is interesting in that we can now see the difference between Thomas and Jimmy is mainly down to the number of deletes per file. But now if we discount the deletes as a field then we might say that Jimmy is now the biggest contributor, but I suppose this all comes down to whether or not we feel deletes are significantly important or not.

3: Time of the day by Authors



But let us just go back to the plain commits. I will now look at Thomas check in pattern in terms of the hours he does this. He seems to be a consistent worker he commits from 9am through till up until 1pm, it seems it then he must take a break probably lunch from 1pm till 2pm as his commits total drop back and he is the 3rd highest between 1 and 2 pm. Then between 2 and 5pm he is back at the front leader board for all commits with his highest commits coming between 2 and 3pm.

It seems Jimmy is consistently second on the leader board and seems to start an hour later than Thomas at 10am. Jimmy’s second highest level of commits is when Thomas appears to be on his lunch.

Another interesting fact, as a side note I wrote a test to see if we had any entries where the date was a Saturday or a Sunday, this was to see if anyone was working at the weekend, I then wrote the code for the test, there were no dates that were a Saturday or a Sunday which I was a bit surprised by. I was so surprised that I changed a couple of dates in the changes\_log.csv file so they are now dates for a Saturday and a Sunday, these date are being return by the method. .But maybe I shouldn’t have been surprised as there have been very little late night commits so appears the team do very little overtime if any. Certainly from the figures it seems it is all the other the developers who work later. Thomas has no check in’s after the hour of 6pm.