**Analysis of the Data**

This is a word document because I have to submit one.

It is a summary of the Analysis.ipynb which should be ready instead.

All files saved on Github <https://github.com/barrysheppard/B8IT105-CA4>

**Student Name**: Barry Sheppard **Student Number**: 10387786

## 1. Commits per user

Now we have the data ready it is time to look at three interesting statistical pieces of information. This will be

1. Commits per user
2. File changes per user
3. Time series of each users commits

Revision

Author

Alan 5

Dave 2

Freddie 7

Jimmy 152

Nicky 5

Thomas 191

Vincent 26

ajon0002 9

murari.krishnan 1

svn.company.net 24

We can see the most commits are by Thomas with Jimmy coming in a close second. Lets start breaking those down to see what our programmers are actually doing. First we look at some of the file adds. Each commit can have multiple files changed at once so even if programmer only commits once they could be changing 100s of files. We can also express this as %.

Revision Percentage

Author

Alan 5 1.184834

Dave 2 0.473934

Freddie 7 1.658768

Jimmy 152 36.018957

Nicky 5 1.184834

Thomas 191 45.260664

Vincent 26 6.161137

ajon0002 9 2.132701

murari.krishnan 1 0.236967

svn.company.net 24 5.687204

**Why is this interesting?** This is interesting as it is the first indication of how much work is being done. It is boring from the perspective of impressive statistical forumlae but it is core to our understanding of what is happening. The difference between the high performers and lower is also very stark.

## 2. File changes per user

A single commit can be a change of one character or it could be the entire project from start to finish. It reallly depents on how the author is approaching things. To get into this we need to see the number of files Added, Deleted, and Modified during each commit.

A single commit can be a change of one character or it could be the entire project from start to finish. It reallly depents on how the author is approaching things. To get into this we need to see the number of files Added, Deleted, and Modified during each commit.

Added Deleted Modified All Files

Author

Alan 9 6 15 30

Dave 10 0 66 76

Freddie 0 0 9 9

Jimmy 690 66 401 1157

Nicky 0 0 7 7

Thomas 87 663 609 1359

Vincent 260 32 45 337

ajon0002 0 0 9 9

murari.krishnan 0 0 1 1

svn.company.net 0 0 24 24

Here we can see that Thomas has the majority of files changed closely by Jimmy. Vincent trails both of them and the remainder are below 100 files in total with 4 less than 10. Looking back at the commits plot we can see the svn.company.net has the same number of commits as it does files suggesting it makes one change everytime, possible to the same file.

So Jimmy ads the majority of files. Thomas deletes the majority o f files but also modifies the most. Vincent's contributions primarily come from adding.

**Why is this interesting?** While commits gave us one piece of information, files changed gives us a different one. We can now see that the difference between Jimmy and Thomas is largely down to the files deleteed by Thomas. If we had to choice between them for a single best contributor, before we might have fone with Thomas on commit volume alone, but if we take a file delete as 'lost' work then Jimmy comes out ahead.

## 3. Time series of each users commits

The raw number of commits isn't a great way to judge a users actual performance. After all, they could make 100 minor commits on one day and nothing on every other day. We saw Dave only commited twice but had 76 files! To get a better idea of what is happening in the team we see how regular posts are.

One way to visualise this is with heatmaps. I've updated some custom code in heatmap.py Depending on the author the range gets updated, so yellow doesn't always mean the same thing. The legend below each will show.

Unsurprisingly Jimmy and Thomas are the most consistent authors and have been with the project for the majojrity of the time. Interestingly, when they are visually compared they seem to work in tandem, with Jimmy having more commits on the days Thomas doesn't and vice-versa. Vincent has also been doing a pretty good job, especially if he only joined the project in November.

**Why is this interesting?** Actually getting to see the days people working does give us a good insight into what is happening. In this case, we need to look and see if Thomas and Jimmy are actually the same person trying to draw down double the salary