Student Number: 10392589

Link to files on GitHub: <https://github.com/kieranregan/CA4>

**Summary:**

* There are 422 records in the cleaned file “changes.csv.”
* Pandas was throwing errors when importing the changes.csv file as some of the comments contained commas.
* I specified Pandas to pick the first 6 columns when importing the data and this solved the problem.

**Statistic 1 – Number of Commits**

* Out of the 422 records, 24 were associated with an automated system called SVN and did not have a user name.
* The majority of the commits (81.27%) were done by Thomas and Jimmy:

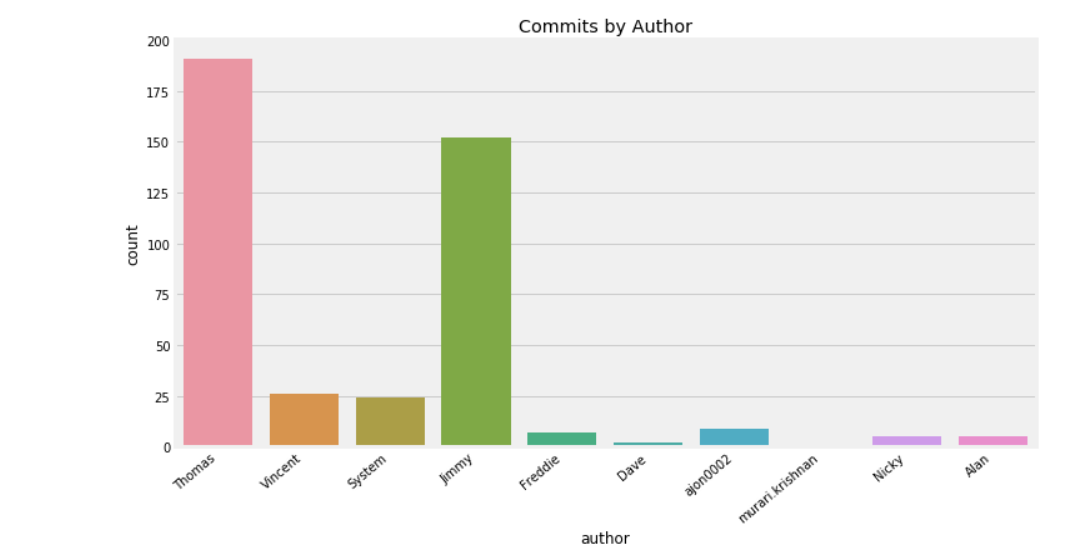


Figure 1 - Volume of commits made by author

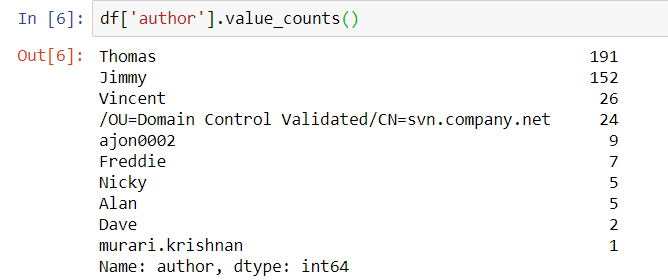


Figure 2 - Count of commits by author

**Statistic 2:**

The data in the changes.csv file contained a date each commit was made

When the changes.csv was loaded into a Pandas Data frame I could see that the date was of type object and not type date

I used the Pandas to\_datetime function to convert it to the correct format.

When the date was in the correct format, I could map the dates to their corresponding day of the week to get more interesting statistics on the commits:

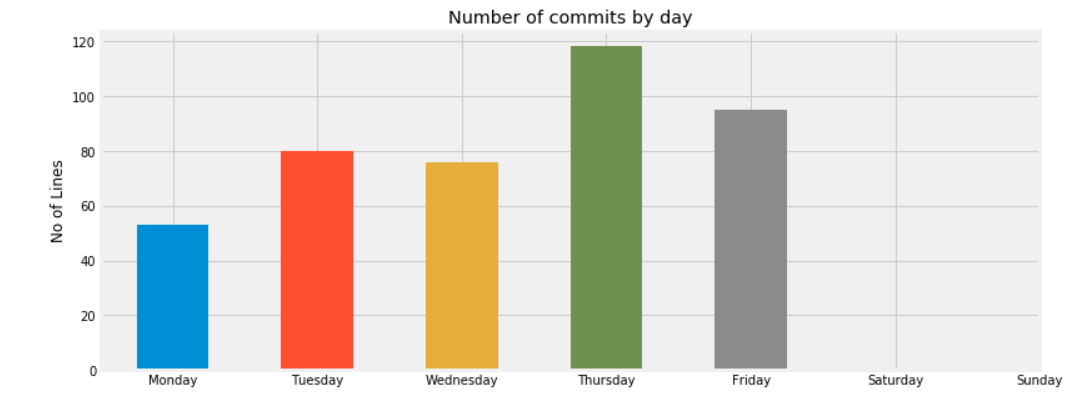


Figure 3 - Number of Commits by day

The above chart shows the number of commits by day.

As some what expected there were no commits on a Saturday or Sunday.

Monday is a quiet day for commits, with Thursdays being the most popular followed by Fridays.

I also did analysis on the number of commits by week:

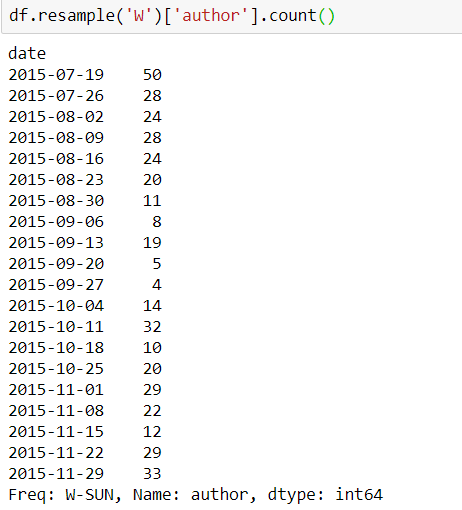


Figure 4 - Number of commits by week

As you can see from the results there was a quiet period from 2015-08-30 to 2015-09-27

**Statistic 3.**

Like the dates described above the data also contained the time a commit was made at.

The time was classed as an object when I created a data frame from the changes.csv file.

I again converted this to time format using pandas to\_date time and just took the hour the change was made for my analysis.

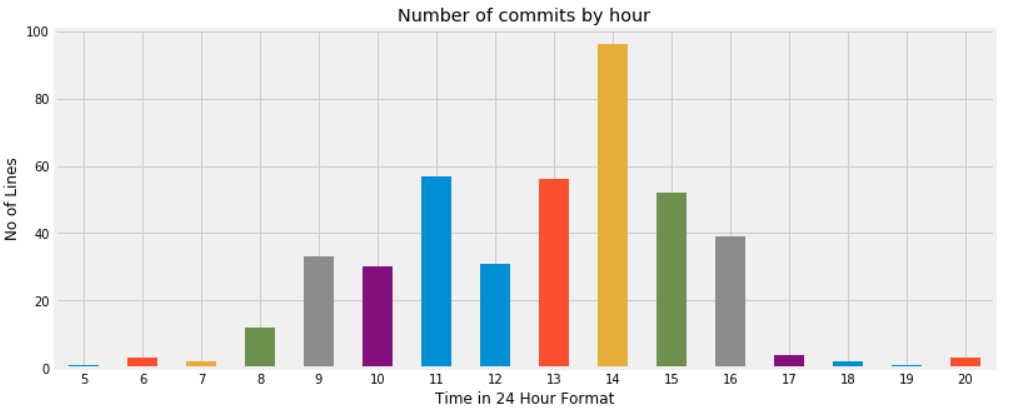


Figure 5 - Time commits are made

The above chart shows that the most common time for commits is 2pm.

On the whole it looks like the majority of the commits are being made usually within standard business hours between 8am – 4pm

I also grouped the number of commits by authors and weekdays to show what days the authors made their most commits.

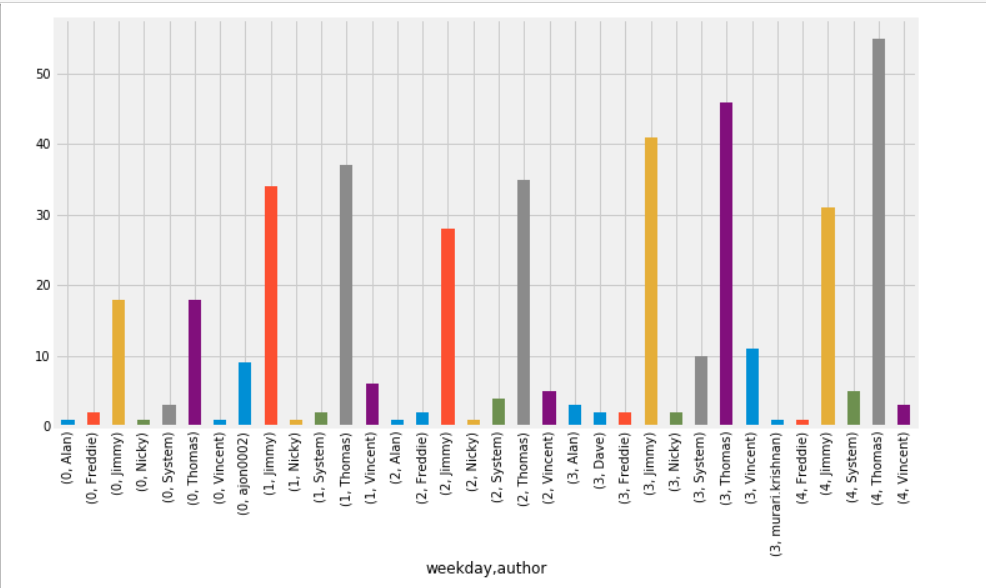


Figure 6 - Authors and Days grouped

I also did the same for count of commits by author and time. This could be useful to see who starts work early and who works late.

We can see that Vincent has the most early (before 8am) and late commits (from 6-8pm)

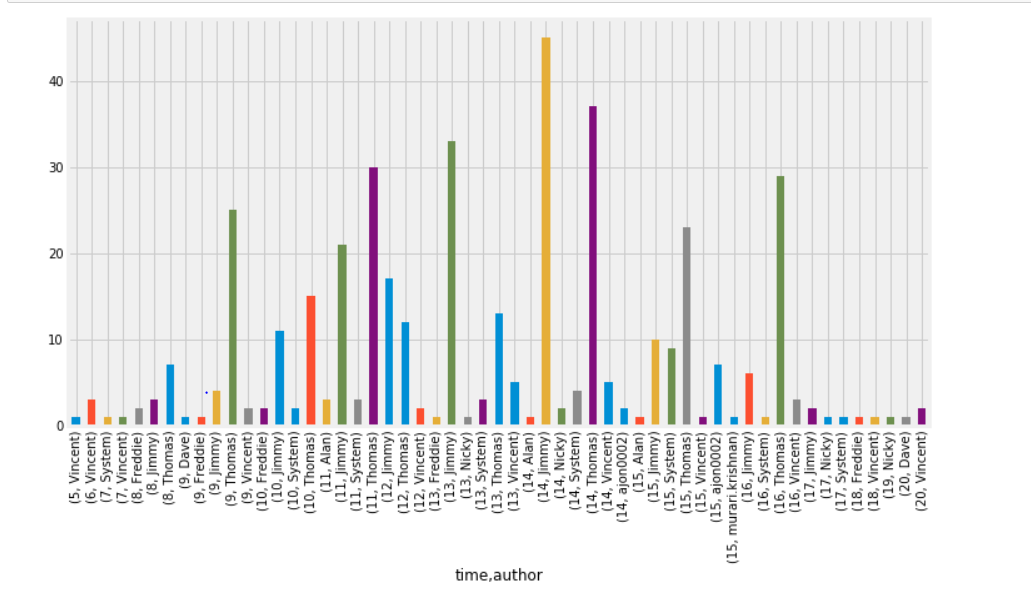


Figure 7 - Number of commits by time and author