Dublin Business School

Programming for Big Data – B8IT 105

CA4

Peter Hutt

10360322

The approach used for this application was to use the individual record ‘type’ and transfer into map / reduce functionality provided by Python applications.

The ‘mapper.py’ application took each record, examined it to see if it was a header record and, if so, flagged that the next record would be a detail record needed. This detail record was then processed for the details needed and then piped to the ‘reducer.py’ application.

This application used the available details to find detail, for each commit, of the day of the week, the period of the day and the week of the year. This was added to each record and passed to the output file ‘output.csv’.

Visualisations were then done from this file using Power BI.

All files are included in the GitHub repository.

This dashboard is available at

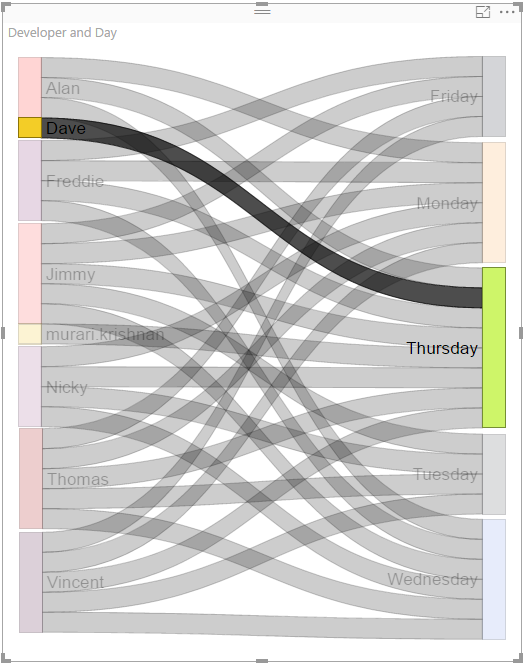
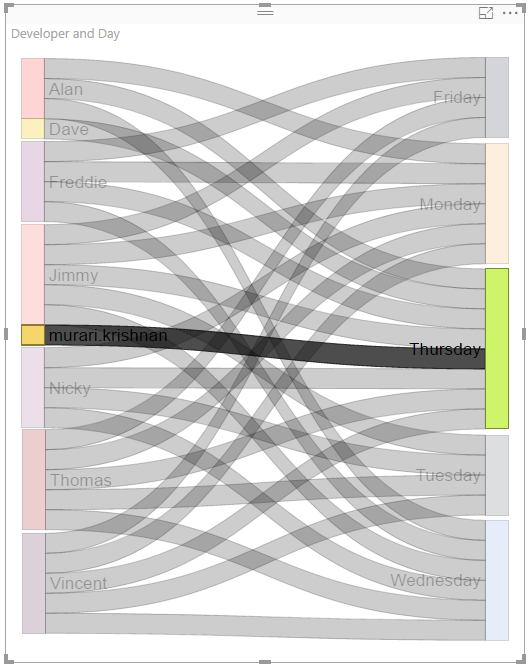
<https://app.powerbi.com/groups/me/dashboards/400eaf30-e47c-4caf-b9e6-093db4ba225b>

Using the PowerBI visualisations some analysis could be done of the data and could be used by management for staff review or process improvement.

**Posting Day for Developers**

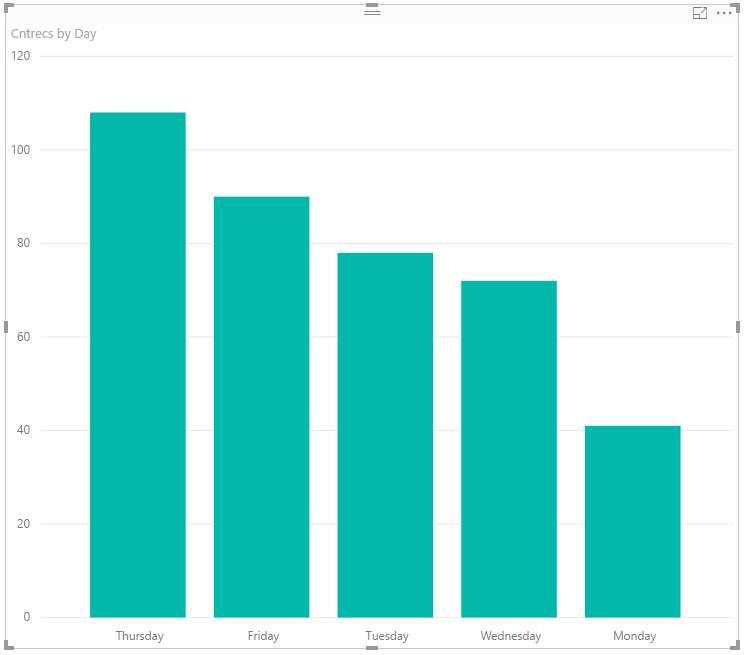
From our Sankey Chart, we find that two developers only post on Thursdays.

These developers, Dave and Murari, should be reviewed, either they work only on Thursdays or they are saving their commits too long and bringing in unnecessary risk.

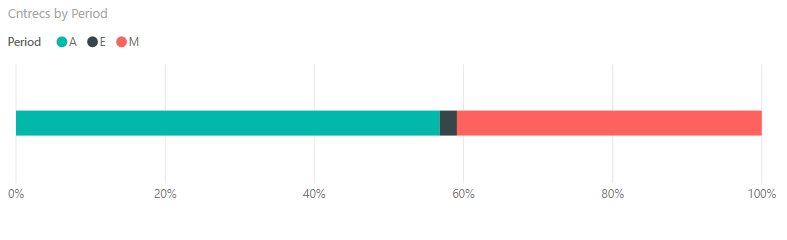
**Posting Day for Team**

Possibly due to the developers above, Thursday is the most common day for committing updates, as can be seen in the bar chart. This may conform with testing expectations or may be due to other causes but generally, in a fully performant team, the daily submissions should be relatively equal.



**Posting by Time of Day**

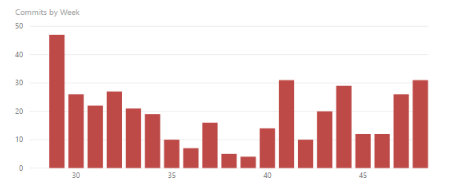
It would generally be expected that most commits would be done towards the end of the day, and this is supported by the graphic showing the Morning / Afternoon / Evening submissions.



It may be of concern, however, that the morning commits are above 40% as this may point to a commits first thing in the morning rather than last thing in the evening, which should be an absolut minimum for all the developer team.

**Posting Week**

The Commits by Week Chart shows, for the period covered, that the largest number of commits were done in week 29.

****

As week 29 is mid-July this may coincide with summer breaks, or possibly an end-of-project tidy up and documentation but it may also indicate that focus on the process was weak over the summer period which may be an issue.

**Posts by Developer**

This chart shows the number of commits by developer.



It can be seen that the most prolific are Thomas and Jimmy. This could be the nature of their work, e.g. they are fixing many bugs or updating documentation, or it could indicate their own work process of frequent saves and unit tests. It could, however, be an indicator of the number of issues found in their own work which may need to be addressed.