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 the day of the week of the commit and the period of the day for the commit, added this to the 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

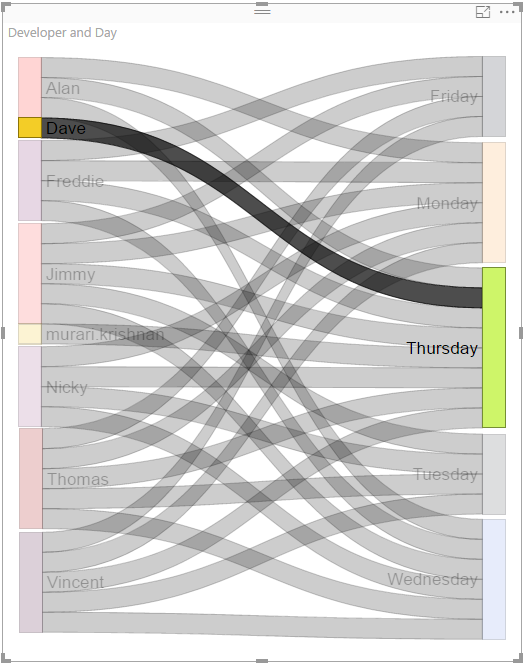
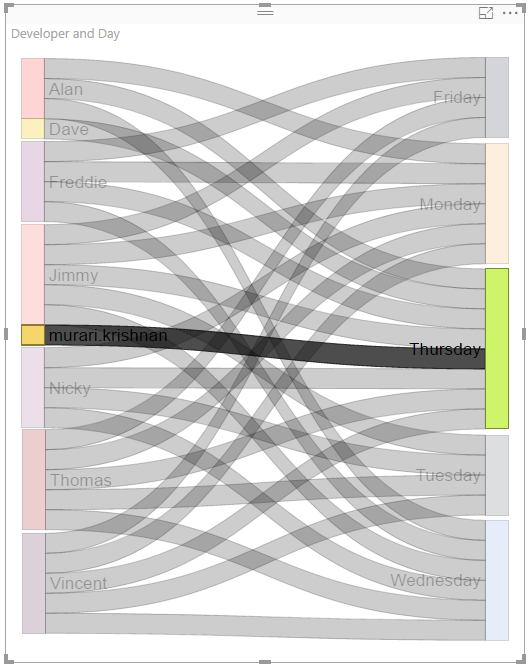
<https://app.powerbi.com/groups/me/dashboards/400eaf30-e47c-4caf-b9e6-093db4ba225b>

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**

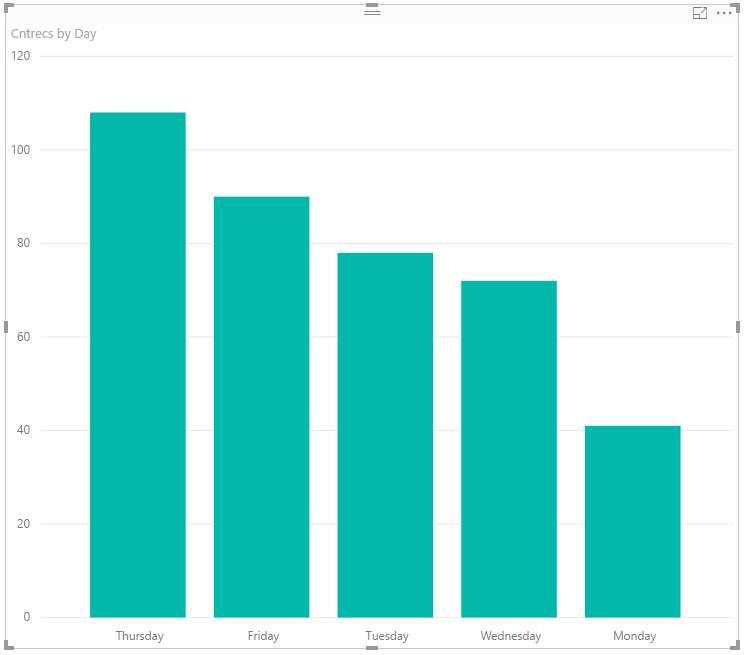
From our Sankey Chart, we find that two developers only post on Thursdays.

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.

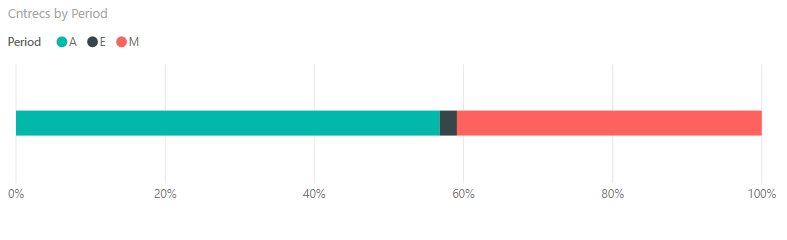
**Posting Day for Team**

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.