**B8IT105 Programming for Data Analytics**

**Rob Farrell – 10340158**

**Lecturer: Darren Redmond**

Table of Contents

[Table of Contents 2](#_Toc481909702)

[1. CA4 - Perform Analysis on a 5000 line dataset 3](#_Toc481909703)

[Tableau Dashboard 3](#_Toc481909704)

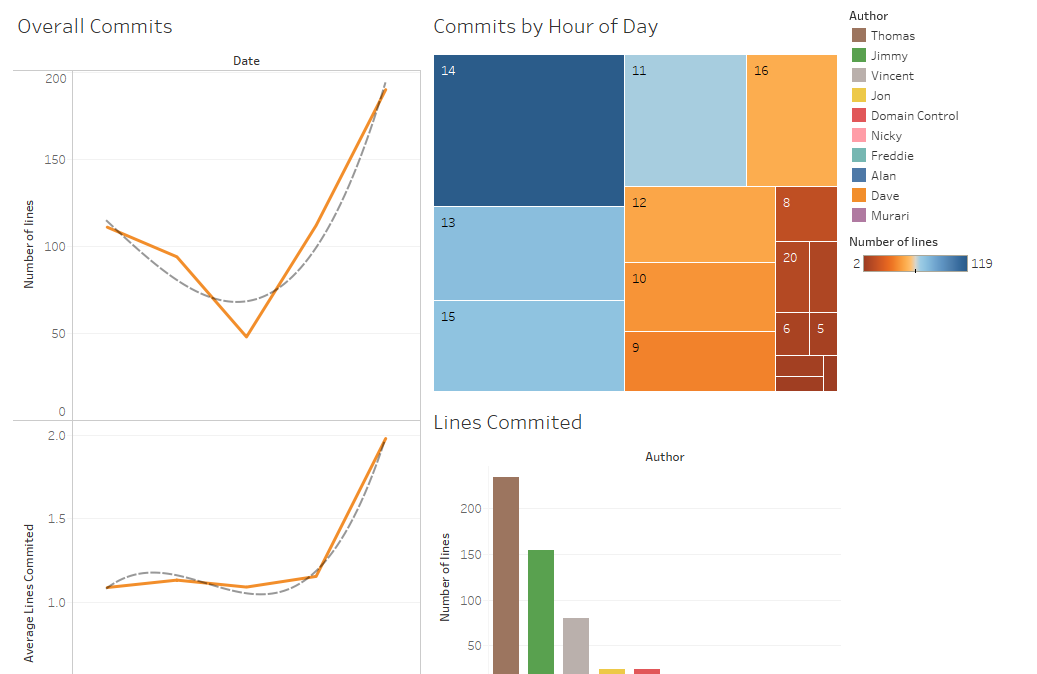
[Thomas is the Most Productive Developer 3](#_Toc481909705)

[Nobody Burns the Midnight Oil 5](#_Toc481909706)

1. [CA4 - Perform Analysis on a 5000 line dataset](https://elearning.dbs.ie/mod/assign/view.php?id=427240)

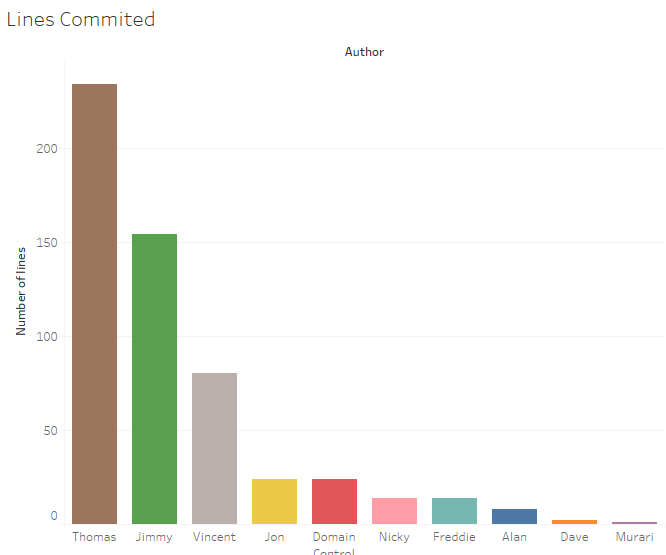
Assignment 4 is based on transforming a large dataset in text format - over 5000 lines of text. You will need to scrub (clean) the data and place it into the relevant holder/container objects. Once in these objects you will see that there are 422 different sets of commit objects. So your task will be to analyse these 422 objects that are in a list and come up with 3 interesting statistical pieces of information for this dataset with supporting evidence of "interestingness'. You code for calculating the analysis should be documented and tested. Test should be in a separate file runnable from the command line. Your statistical analytics conclusions should be in a word document explaining in approximately 500 words the information that you have gleamed from the dataset. You will be required to submit your code via github along with all documentation and tests. The deadline is the 7th May 2017 on moodle @ 23:55.

Tableau Dashboard

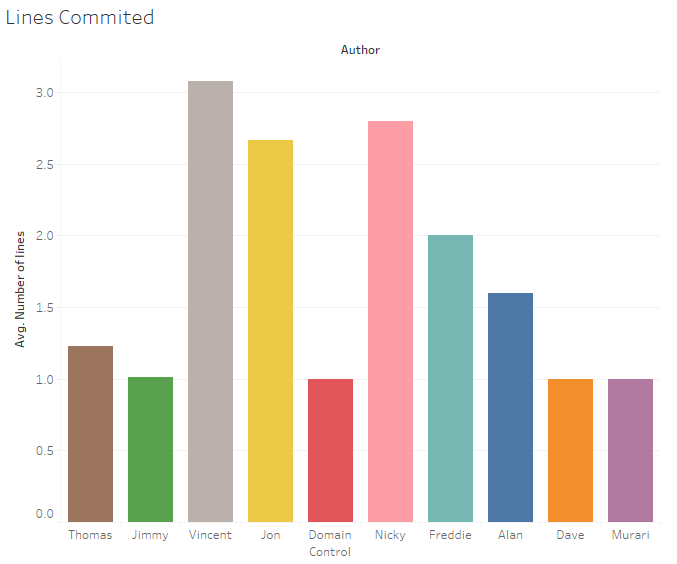


The results that presented themselves were pretty mundane. To get the results I wanted firstly in Python I tidied up some of the names, had to change the dictionary index on “number\_of\_lines” and then to get the dates and times I wanted I had to do some custom splits within Tableau.

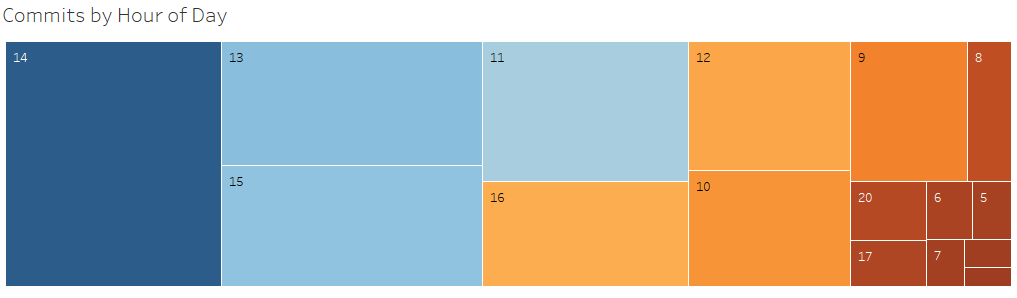
Thomas is the Most Productive Developer



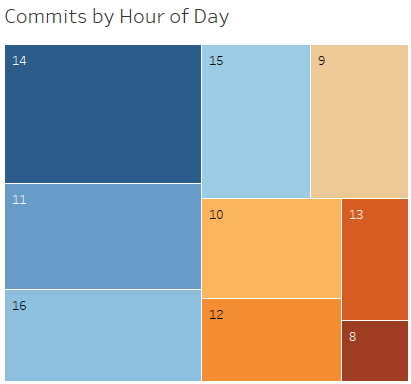
This is in an overall sense, although the data shows on average during each commit Vincent has been committing more lines of code but he obviously just does it less often. Thomas is making smaller but more regular contributions to the code.



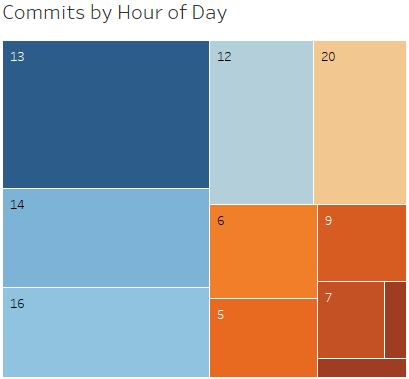
Nobody Burns the Midnight Oil



Most code is committed at 2PM in the middle of the day. Vincent seems to start the earliest but takes breaks throughout the day. Thomas appears a little more organised and consistent. Let’s compare the two tables.



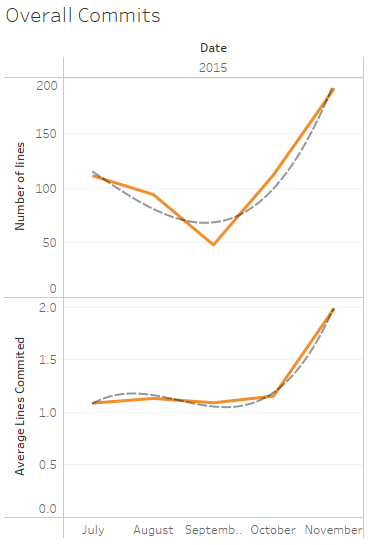
Thomas

* 47 lines committed around 2PM
* Standard day of 8AM to after 4PM
* Relatively consistent blocks of submissions

Vincent

* Submissions more erratic
* Less consistency
* Works from as early as 5AM and as late as after 8PM

Something Happened in November

As can be seen from the graphs there is a polynomial trend distribution for both total line committals and average line committals.

As you can see both curves trend positively towards November, one might infer that this is indicative of summer holidays or a previous project coming to an end around September. Either way it looks like going into November that the developers shift gears.

