|  |  |
| --- | --- |
| **Course Title** | **Higher Diploma in Science in Data Analytics** |
| **Unit Title** | **Programming For Bid Data** |
| **Unit Code** | **B8IT105** |
| **Assignment Title** | **CA4 Perform analysis on a 5000 line dataset** |
| **Word Count** | **669 excluding tables, charts and TOC** |
| **Date** | **6th May 2017** |
| **Student Name** | **John Butler** |
| **Student Number** | **10353776** |

Contents

[Introduction 3](#_Toc481863704)

[General analysis 4](#_Toc481863705)

[Analysis By Path Type 7](#_Toc481863706)

[Analysis by workstream 9](#_Toc481863707)

[Summary 11](#_Toc481863708)

# Introduction

This report provides detailed analysis of a project undertaken by a 10 person team over a 20 week period from July to November 2015. GITHUB was used for version control and over the course of the project 422 commits were recorded.

The changes section specifies whether each change relates to a file addition, modification or deletion. It also provides details on the type of work performed by the author, including:

* Build-Config: work done on build configuration of various phone models
* Resolution: with pixel density ranging from 600dp, to mdpi, hdpi,xhdpi and xxhdpi
* Java:
* Gradle: os [build automation](https://en.wikipedia.org/wiki/Build_automation) system, builds upon the concepts of [Apache Ant](https://en.wikipedia.org/wiki/Apache_Ant) Mavenwork
* XML
* Other: other ad-hoc work which does not lend itself to appropriate grouping.

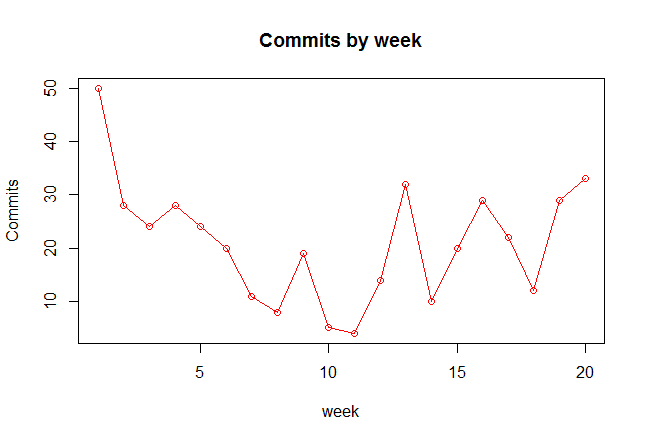
Three categories of analysis of commits are provided, being:

* General analysis :
* Analysis by path type:
* Analysis by workstream:

# General analysis

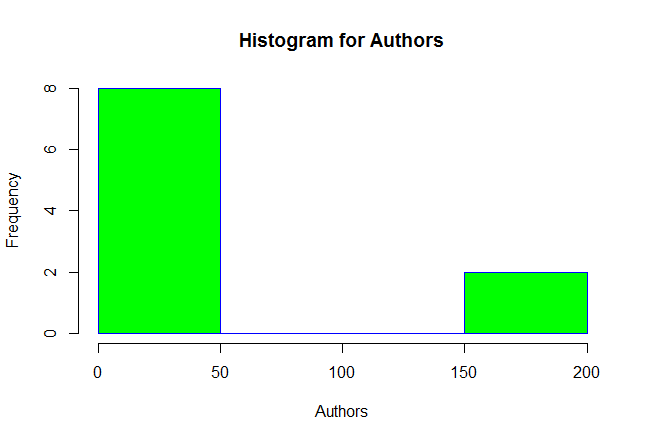
The table below is based on the python generated csv files and chart below (created using R studio analysis of python csv files) show analysis of the 422 commits by week. (see excel workbook “CommitsByWeekAndAuthor.xlsx”).

The project started and ended strongly. However, there are an unusually low number of commits in the middle of the project (9 commits in fortnight starting on week 37 – fortnight beginning 7/09/2015.) Perhaps members of the team were on holidays during that period.



|  |  |  |
| --- | --- | --- |
| **Project commits by week** | | |
|  |  |  |
| **Week** | **Commits** | **CumComms** |
| 28 | 50 | 50 |
| 29 | 28 | 78 |
| 30 | 24 | 102 |
| 31 | 28 | 130 |
| 32 | 24 | 154 |
| 33 | 20 | 174 |
| 34 | 11 | 185 |
| 35 | 8 | 193 |
| 36 | 19 | 212 |
| 37 | 5 | 217 |
| 38 | 4 | 221 |
| 39 | 14 | 235 |
| 40 | 32 | 267 |
| 41 | 10 | 277 |
| 42 | 20 | 297 |
| 43 | 29 | 326 |
| 44 | 22 | 348 |
| 45 | 12 | 360 |
| 46 | 29 | 389 |
| 47 | 33 | 422 |
| Total | 422 |  |

The histogram (from R Studio) and table below show an uneven distribution of the commits submitted by authors. Of the 10 authors only 2 provided between 150 and 200 commits over the length of the project while the remaining 8 authors only provided 0 to 50 commits



|  |  |  |
| --- | --- | --- |
| **Author** | **Commits** | **%** |
| Vincent | 26 | 6% |
| Thomas | 191 | 45% |
| Jimmy | 152 | 36% |
| /OU=Domain Control Validated/CN=svn.company.net | 24 | 6% |
| Freddie | 7 | 2% |
| Dave | 2 | 0% |
| ajon0002 | 9 | 2% |
| murari.krishnan | 1 | 0% |
| Nicky | 5 | 1% |
| Alan | 5 | 1% |
| Total | 422 | 100% |

# Analysis By Path Type

This analysis is based mainly on python generated summary reports. See workbook “7-8-9-AddsDeletesModsByAuthorv1.xlsx”

Thomas has generated more commit paths (45% of total) than Jimmy (38% of total). However, 49% of Thomas’s path changes relate to deletes and only 6% relate to adds while 60% of Jimmy’s path changes relate to adds and only 6% relate to deletes. Depending on the relative weightings of the path types this could indicate that Jimmy adds more value to the project than Thomas even though Thomas committed more paths in absolute terms.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Commits by path type in "changes" section** | | | | | |
|  |  |  |  |  |  |
| **Values** | | | | | |
| **Author** | **Adds** | **Delete** | **Mod** | **Total** | **%Total** |
| murari.krishnan | 0 | 0 | 1 | 1 | 0% |
| ajon0002 | 0 | 0 | 9 | 9 | 0% |
| Vincent | 260 | 32 | 45 | 337 | 11% |
| Thomas | 87 | 663 | 609 | 1,359 | 45% |
| Nicky | 0 | 0 | 7 | 7 | 0% |
| Jimmy | 690 | 66 | 401 | 1,157 | 38% |
| Freddie | 0 | 0 | 9 | 9 | 0% |
| Dave | 10 | 0 | 66 | 76 | 3% |
| Alan | 9 | 6 | 15 | 30 | 1% |
| /OU=Domain Control Validated/CN=svn.company.net | 0 | 0 | 24 | 24 | 1% |
| Total | 1,056 | 767 | 1,186 | 3,009 | 100% |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Percentage** | | | | |  |
| **Author** | **Adds** | **Delete** | **Mod** | **Total** |  |
| murari.krishnan | 0% | 0% | 100% | 100% |  |
| ajon0002 | 0% | 0% | 100% | 100% |  |
| Vincent | 77% | 9% | 13% | 100% |  |
| Thomas | 6% | 49% | 45% | 100% |  |
| Nicky | 0% | 0% | 100% | 100% |  |
| Jimmy | 60% | 6% | 35% | 100% |  |
| Freddie | 0% | 0% | 100% | 100% |  |
| Dave | 13% | 0% | 87% | 100% |  |
| Alan | 30% | 20% | 50% | 100% |  |
| /OU=Domain Control Validated/CN=svn.company.net | 0% | 0% | 100% | 100% |  |

The table below uses Python generated detailed report to show path Adds per author per commits. It shows an unequal distribution of Adds over commits. For example, Vincent was responsible for 260 Adds of which 259 were part of 1 commit. Jimmy generated 690 adds over 13 commits but one commit alone accounts for 262 (38%) of these.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sum of Adds** | **Column Labels** |  |  |  |  |  |
| **Row Labels** | **Alan** | **Dave** | **Jimmy** | **Thomas** | **Vincent** | **Grand Total** |
| 1 |  |  | 15 | 19 | 1 | 35 |
| 2 |  |  | 10 | 8 |  | 18 |
| 3 | 3 |  | 9 | 3 |  | 15 |
| 4 |  |  | 4 | 8 |  | 12 |
| 5 |  |  | 5 | 10 |  | 15 |
| 6 | 6 |  | 6 | 12 |  | 24 |
| 7 |  |  |  | 7 |  | 7 |
| 8 |  |  |  | 8 |  | 8 |
| 9 |  |  | 9 |  |  | 9 |
| 10 |  | 10 |  |  |  | 10 |
| 12 |  |  |  | 12 |  | 12 |
| 13 |  |  | 13 |  |  | 13 |
| 16 |  |  | 16 |  |  | 16 |
| 66 |  |  | 66 |  |  | 66 |
| 91 |  |  | 91 |  |  | 91 |
| 92 |  |  | 184 |  |  | 184 |
| 259 |  |  |  |  | 259 | 259 |
| 262 |  |  | 262 |  |  | 262 |
| **Grand Total** | **9** | **10** | **690** | **87** | **260** | **1056** |

# Analysis by workstream

This analysis is based mainly on python generated summary reports. See workbook “10-15-WorkstreamsByAuthorv1.xlsx” and table below.

Overall Thomas has a greater count of workstream tasks than Jimmy. Approximately 45% of Thomas tasks relate to screen/icon resolution (pixel) while only 35% of Jimmy’s tasks relate to this.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis of workstreams by author** | | | | | | | |
| **Values By Author** | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| **Auth** | **bld-config** | **pixel** | **gradle** | **java** | **xml** | **oth** | **Total** |
| **/OU=Domai** | 0 | 0 | 0 | 0 | 0 | 48 | 48 |
| **ajon0002** | 0 | 0 | 0 | 6 | 3 | 9 | 18 |
| **Alan** | 2 | 12 | 0 | 5 | 11 | 5 | 35 |
| **Dave** | 22 | 0 | 3 | 33 | 12 | 9 | 79 |
| **Freddie** | 0 | 0 | 4 | 5 | 0 | 7 | 16 |
| **Jimmy** | 98 | 456 | 44 | 257 | 208 | 247 | 1310 |
| **murari.krish** | 0 | 0 | 0 | 1 | 0 | 1 | 2 |
| **Nicky** | 0 | 0 | 5 | 2 | 0 | 5 | 12 |
| **Thomas** | 78 | 693 | 42 | 251 | 255 | 231 | 1550 |
| **Vincent** | 5 | 280 | 2 | 11 | 38 | 27 | 363 |
|  |  |  |  |  |  |  |  |
| **Percentage** |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| **Auth** | **bld-config** | **pixel** | **gradle** | **java** | **xml** | **oth** | **Total** |
| **/OU=Doma** | 0% | 0% | 0% | 0% | 0% | 100% | 100% |
| **ajon0002** | 0% | 0% | 0% | 33% | 17% | 50% | 100% |
| **Alan** | 6% | 34% | 0% | 14% | 31% | 14% | 100% |
| **Dave** | 28% | 0% | 4% | 42% | 15% | 11% | 100% |
| **Freddie** | 0% | 0% | 25% | 31% | 0% | 44% | 100% |
| **Jimmy** | 7% | 35% | 3% | 20% | 16% | 19% | 100% |
| **murari.krish** | 0% | 0% | 0% | 50% | 0% | 50% | 100% |
| **Nicky** | 0% | 0% | 42% | 17% | 0% | 42% | 100% |
| **Thomas** | 5% | 45% | 3% | 16% | 16% | 15% | 100% |
| **Vincent** | 1% | 77% | 1% | 3% | 10% | 7% | 100% |

The table below uses Python generated detailed report to show path pixel resolution related tasks per author per commit. It shows a very unequal distribution by the key authors. For example, Vincent generated 280 pixel related tasks over 3 commits, one commit alone generated 277 of these. Thomas generated 693 tasks over 15 commits , one commit alone generated represents 297 of these (almost 43%) and Jimmy generated 456 pixel related tasks over 11 commits and one commit alone generated 228 of these (50%) of these. It is interesting to note that these high volume commits appear towards the end of the project – perhaps indicating a push to complete project within deadlines.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sum of tpixel** | **Author** |  |  |  |  |
| **Row Labels** | **Alan** | **Jimmy** | **Thomas** | **Vincent** | **Grand Total** |
| 1 |  | 2 | 8 |  | 10 |
| 2 |  | 2 | 8 |  | 10 |
| 3 |  | 6 | 6 | 3 | 15 |
| 4 |  |  | 12 |  | 12 |
| 5 |  | 5 |  |  | 5 |
| 6 |  | 12 | 30 |  | 42 |
| 7 |  | 14 | 14 |  | 28 |
| 8 |  |  | 8 |  | 8 |
| 9 |  |  | 18 |  | 18 |
| 12 | 12 |  | 48 |  | 60 |
| 13 |  | 13 |  |  | 13 |
| 14 |  |  | 14 |  | 14 |
| 16 |  | 16 | 32 |  | 48 |
| 22 |  |  | 44 |  | 44 |
| 26 |  | 26 |  |  | 26 |
| 44 |  | 132 |  |  | 132 |
| 52 |  |  | 52 |  | 52 |
| 102 |  |  | 102 |  | 102 |
| 228 |  | 228 |  |  | 228 |
| 277 |  |  |  | 277 | 277 |
| 297 |  |  | 297 |  | 297 |
| **Grand Total** | **12** | **456** | **693** | **280** | **1441** |

# Summary

Utlising Python and R Studio is an effective way to analyse large volumes of data and present results in a meaningful manner. Key findings were as follows:

* 20% of authors (Jimmy and Thomas) account for 81% of commits.
* Large drop in commits during middle of project.
* Adds accounted for 60% of Jimmy’s path type changes but only 6% of Thomas’s path type changes.
* Key workstream related to pixel resolution – 45% of Thomas’s workstreams and 35% of Jimmy’s workstreams related to this.
* Unequal distribution of pixel resolution workstreams over commits with large values appearing around end of project – perhaps to meet looming deadline.