**Changes Analysis**

Python was used to extract relevant information for this analysis from the raw data (see Simple.py and Test\_Simple.py). Microsoft Excel was used for data preparation, analysis and visualisation of the data (See Changes\_1.xls). A summary of the interesting facts of this data are:

1. The top three performers in are Thomas, Jimmy and Vincent respectively, with Thomas responsible for 44% of the changes;
2. The work performed between Monday to Friday, about 29% on Thursdays and 14% on Mondays;
3. Most of the work is performed in the morning between midnight (68%) and 12 noon and the least at night between 6pm and midnight (3%); and
4. Merge was used in most of the changes by authors (85%).

The dashboard below supports the findings.



1. Top three performers

Using Python, a smaller data set in comma separated values (csv) file based on the Python code. There are 9 authors, the top performers are Thomas, Jimmy and Vincent and least performers are Alan, Dave and Murari Krishan. See the table and chart below:

|  |  |
| --- | --- |
|  |  |
| **Row Labels** | **Sum of number\_of\_lines** |
| Thomas | 234 |
| Jimmy | 154 |
| Vincent | 80 |
| ajon0002 | 24 |
| Nicky | 14 |
| Freddie | 14 |
| Alan | 8 |
| Dave | 2 |
| murari.krishnan | 1 |
| **Grand Total** | **531** |

1. Work done during the week

From the data, it all the changes were recorded on weekdays and mostly in the morning between midnight at noon. The date and time were extracted using Python and the day of the week and classification of the time of day (morning, afternoon and night were determined in Excel. The data visualization is shown in the Dashboard above. More details are shown below.

|  |  |
| --- | --- |
|  |  |
| **Row Labels** | **Sum of number\_of\_lines** |
| Monday | 74 |
| Tuesday | 104 |
| Wednesday | 96 |
| Thursday | 153 |
| Friday | 104 |
| **Grand Total** | **531** |

|  |  |
| --- | --- |
|  |  |
| **Row Labels** | **Sum of number\_of\_lines** |
| afternoon | 349 |
| morning | 167 |
| night | 15 |
| **Grand Total** | **531** |



The Dashboard above enables further analysis to detail each authors work according to the day of the week and time of the day.

1. Work performed

This part focused on the aspect of the changes performed by the authors: M – merged, A – added, and D – delete. Most of the action taken was to merge data. Using the dashboard above, it is possible to drill down to the individual author’s actions, day of the week or time of the day.

|  |  |
| --- | --- |
|  |  |
| **Row Labels** | **Sum of number\_of\_lines** |
| A | 33 |
| D | 12 |
| M | 189 |
| **Grand Total** | **234** |

1. Project duration.

Based on the dataset provided, the project ran from July to November 2015, the months was extracted from the dates in data. the dashboard can display author’s monthly activity, change path or day of the week.

|  |  |
| --- | --- |
|  |  |
| Row Labels | Sum of number\_of\_lines |
| Jul | 104 |
| Aug | 88 |
| Sep | 44 |
| Oct | 111 |
| Nov | 184 |
| Grand Total | 531 |

Conclusion

This analysis demonstrates how Python and Excel can used together to analyse data. It starts with raw data that is refined before analysis and trends identified which are presented visually. Excel was sufficient given the size of data, However, other tools can be used on big data such as Alteryx, etc. Authors with the highest level of performance, time of the day for working and actual work performed were identified.