**Student Name** : **Tanzeel Ur Rehman**

**Student Number : 1604233**

**Course Title : Higher Diploma in Science in Data Analytics**

**Module Title : B8IT105**

**Table of Contents**

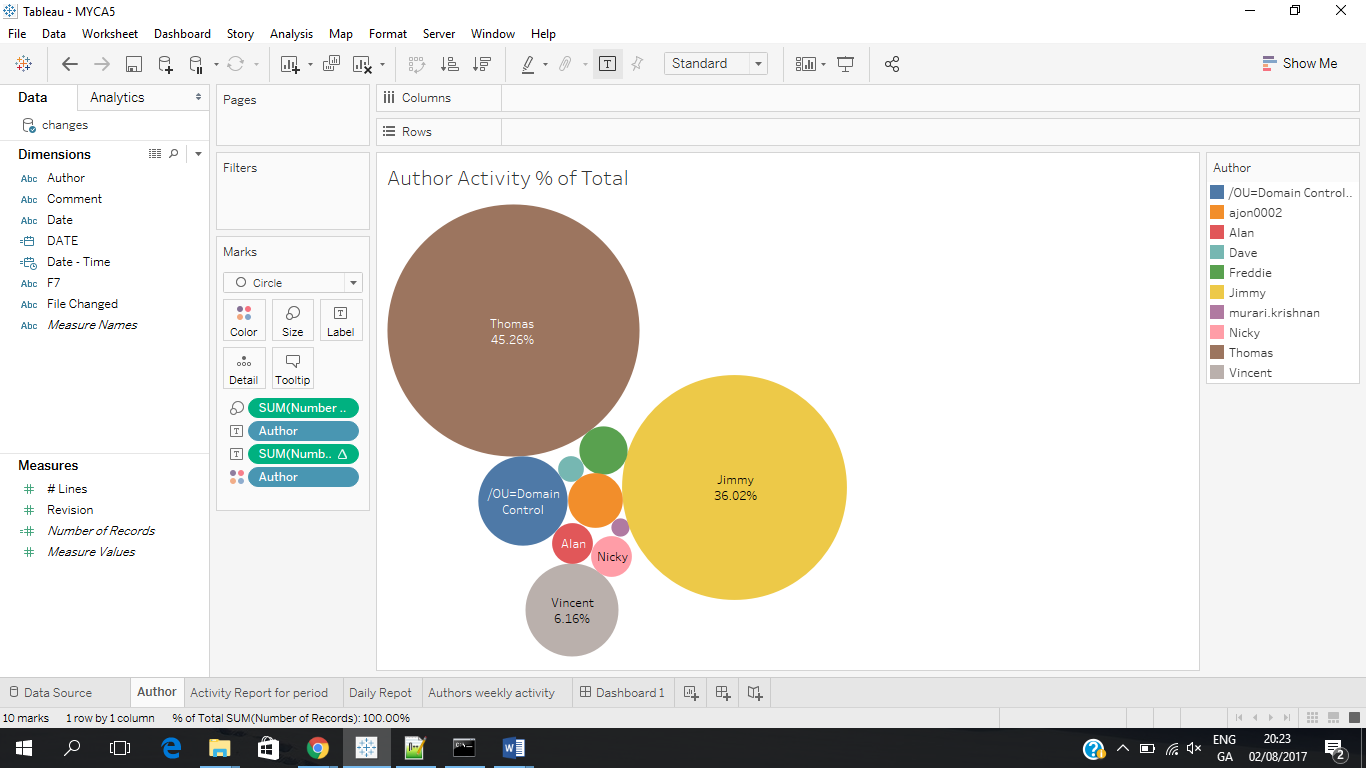
|  |  |
| --- | --- |
| **Title** | **Page** |
| **Report and analysis** | **2** |
| **Continues Area Chart** | **3** |
| **Area Chart** | **4** |
| **Side to Side Bar Graph** | **5** |
| **PERFORMANCE DASHBOARD** | **6** |

**Introduction:** Assignment is based on transforming a large dataset in text format - over 5000 lines of text.

**Methodologies used**: A log file named (changes\_python.log) is converted into changes.csv file using python codes after scrubbing and cleaning dataset we created an excel CSV file. Which is extracted into tableau visualization tools to analyse the dataset.

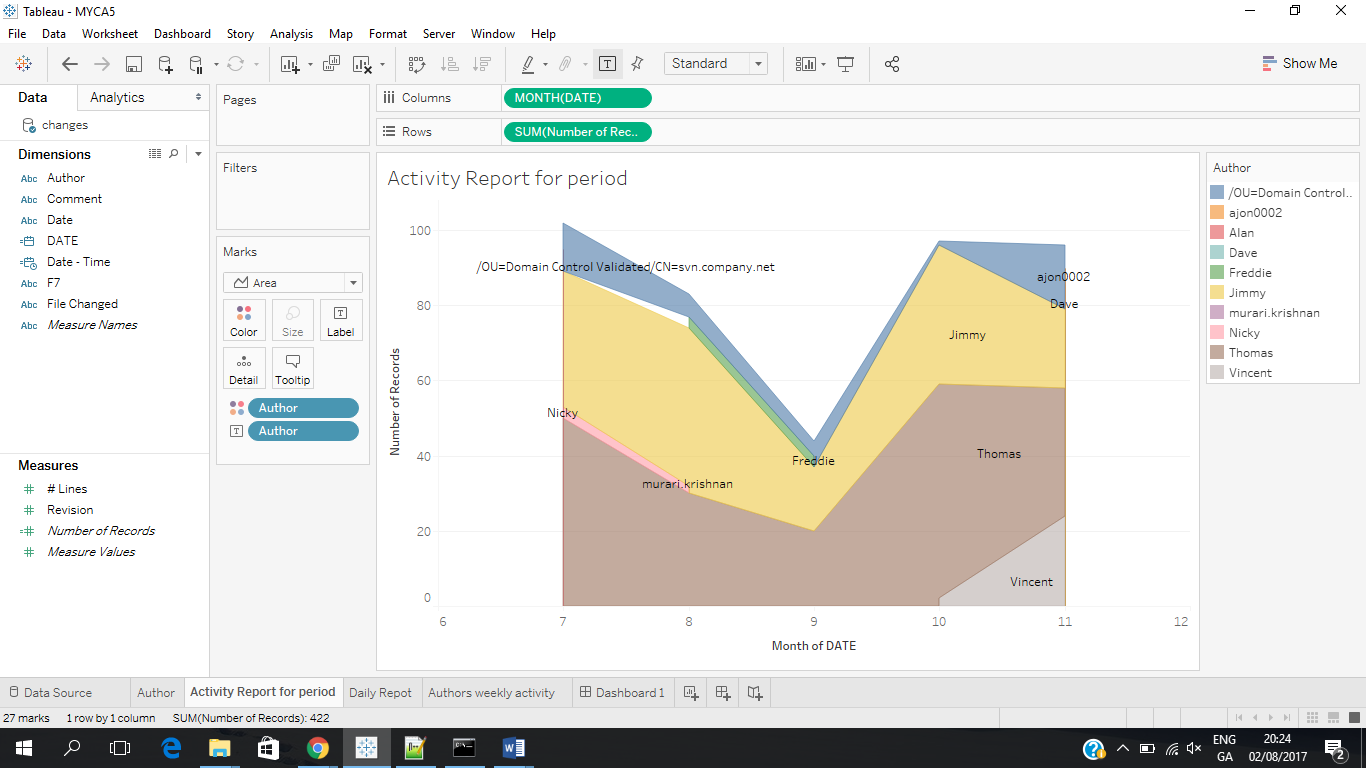
**Reporting and analysis**

**Packed Bubble Chart**



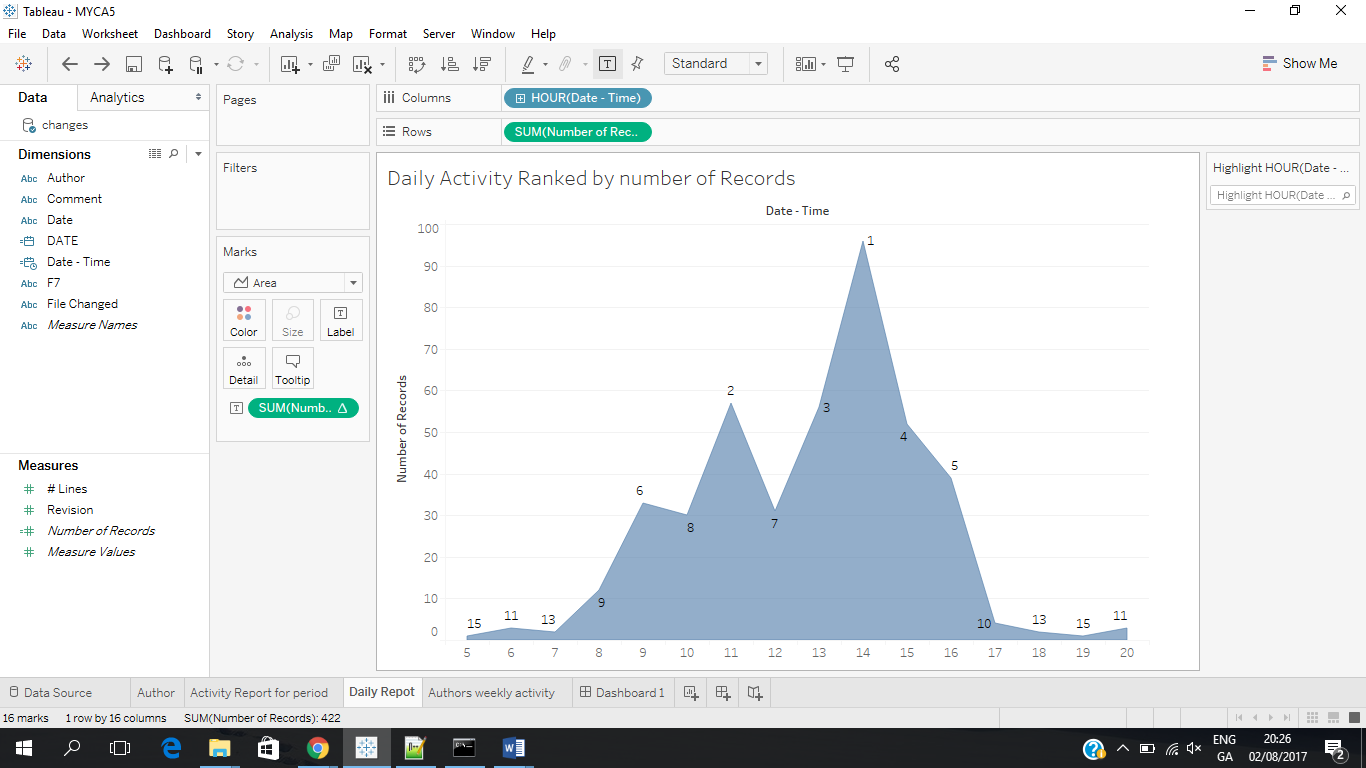
* Packed bubble chart is a visual presentation of our dataset.
* We have nine authors working on system and making changes.
* OU=domain Control is presented as Author but it’s log generated by system itself
* Author Thomas represent largest % of activity in given dataset
* Total number of record is used as measure to calculate % of authors activity
* Size of bubble is displaying volume of tasks performed by authors
* Different Colour of bubble is assigned to Authors

**Continues Area Chart**



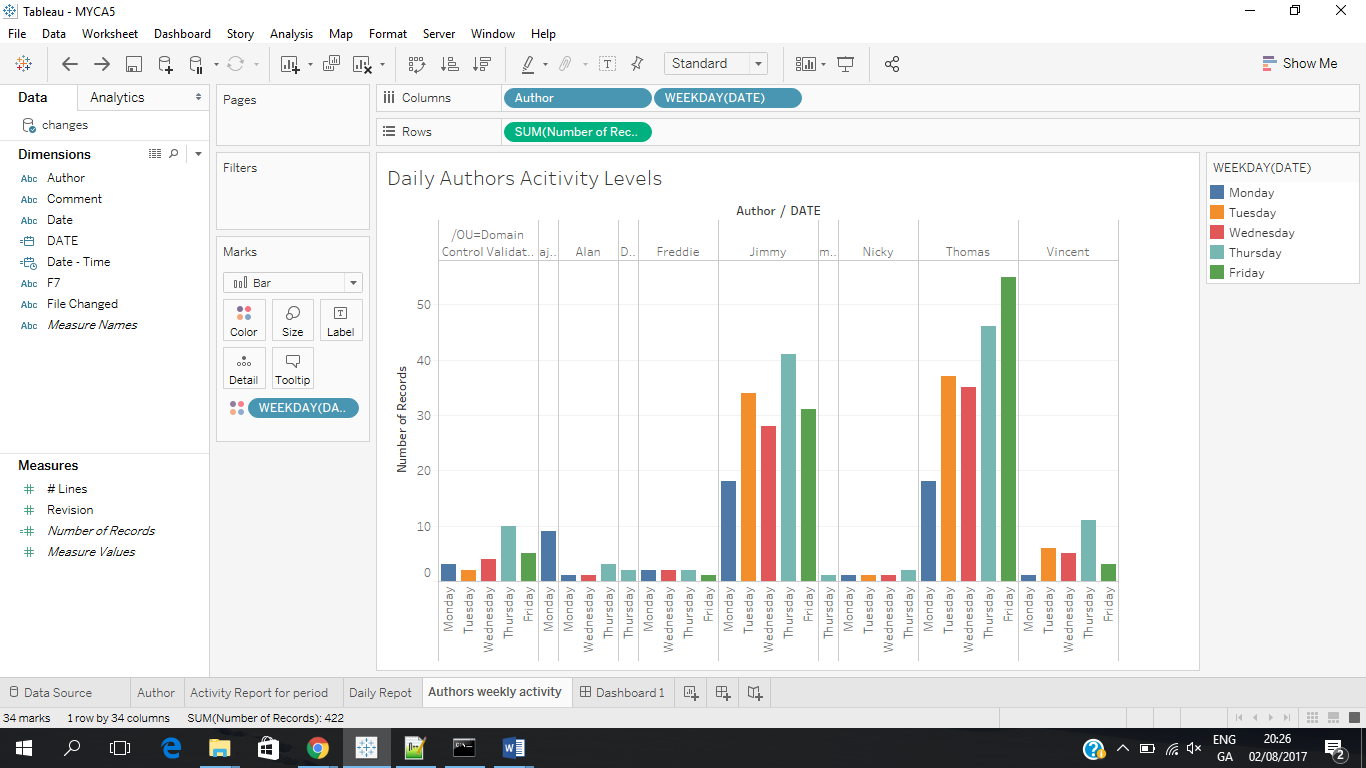
* Number for month is displayed on x-axis and number of Records on Y-axis
* Colour and size of the area is showing us level of monthly Authors activity on the system
* We can see Thomas and jimmy being active as main users from July to November
* Author Nicky only made changes in July and august
* Autor Vincent being active from October to November
* Other Authors seems to be just occasional user for making changes to the system
* Level of overall activity is decreasing from July to September and increasing from start of October and being constant from October to November
* Lowest level of activity is recorded by end of September

**Area Chart**



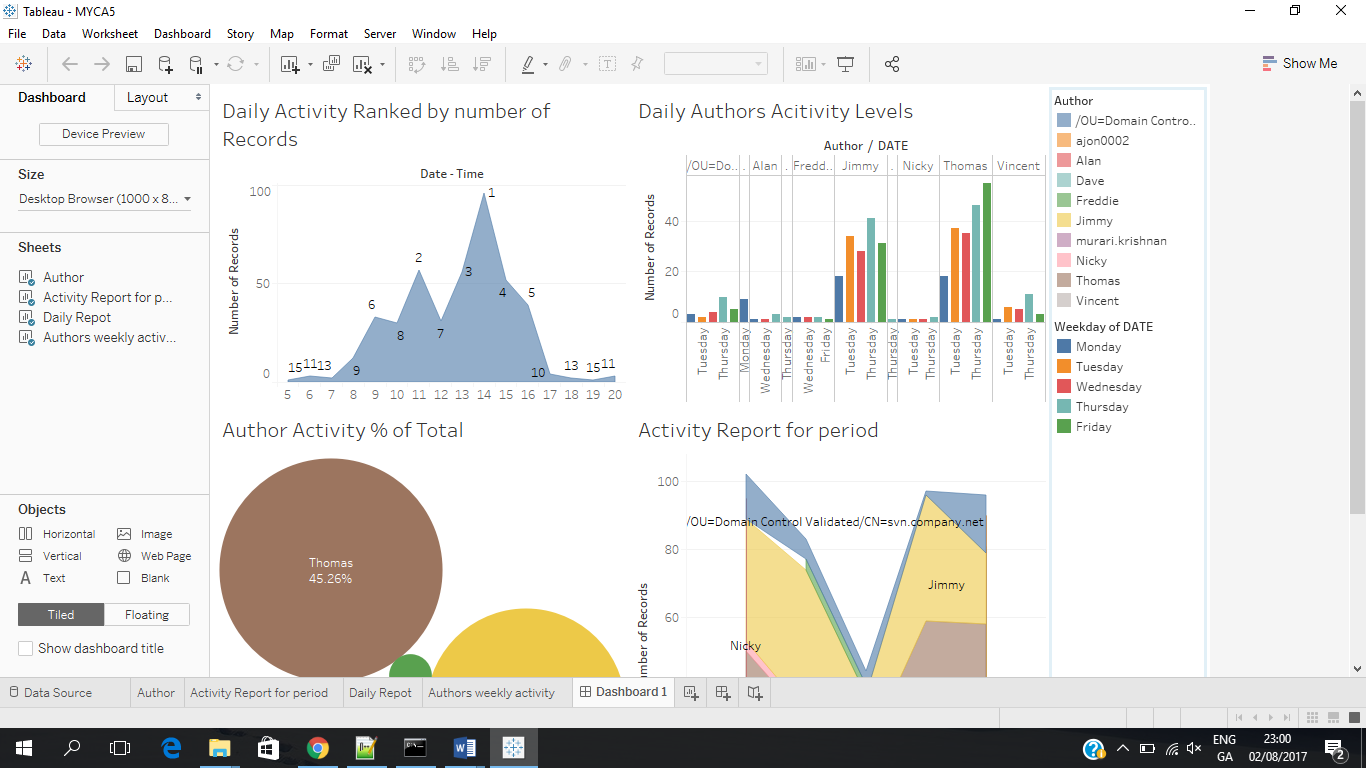
* Daily activity is analysed using area chart
* On x-axis time of the day is displayed and on y-axis number of records
* We can see all activity are recorded between 5:00 am to 20:00
* Hours are ranked from 1 to 15 by level of activity using numbers of records by Authors
* We can see from visual at 14:00 hour of the day user’s activity reaching at highest point
* Activity’s out of normal working hours can be seen at its lowest point

**Side to Side Bar Graph**



* Bar graph is displaying Authors daily levels of activity
* Working week is Monday to Friday
* We can see Thursday seem to be busiest day for all authors combined
* OU=domain Control Validation is highest on Thursday implying its effected by overall system activity. Meaning increase on system usage will increase logs of domain Control validation a system generated log
* Author Thomas is the only user who has most records on Friday compare to other days of the week

**PERFORMANCE DASHBOARD**



* I have created a dashboard for quick visualization for all KPI, s