**2017**



Svetlana Nicolenco

10265376

*https://lh3.googleusercontent.com/yxZJpqHgtcB78CgPzgCgXr1L8jwLDUu9G1SPzRFu72--8k76VqiOdQX4uqu9sFVPkBLHF_JP0WfTFueclpxX1ypMYzAvjSPT7fGiw6E_jzoLHpiAFmvwPI4T-5wRfsSJgsHZIu5sng*

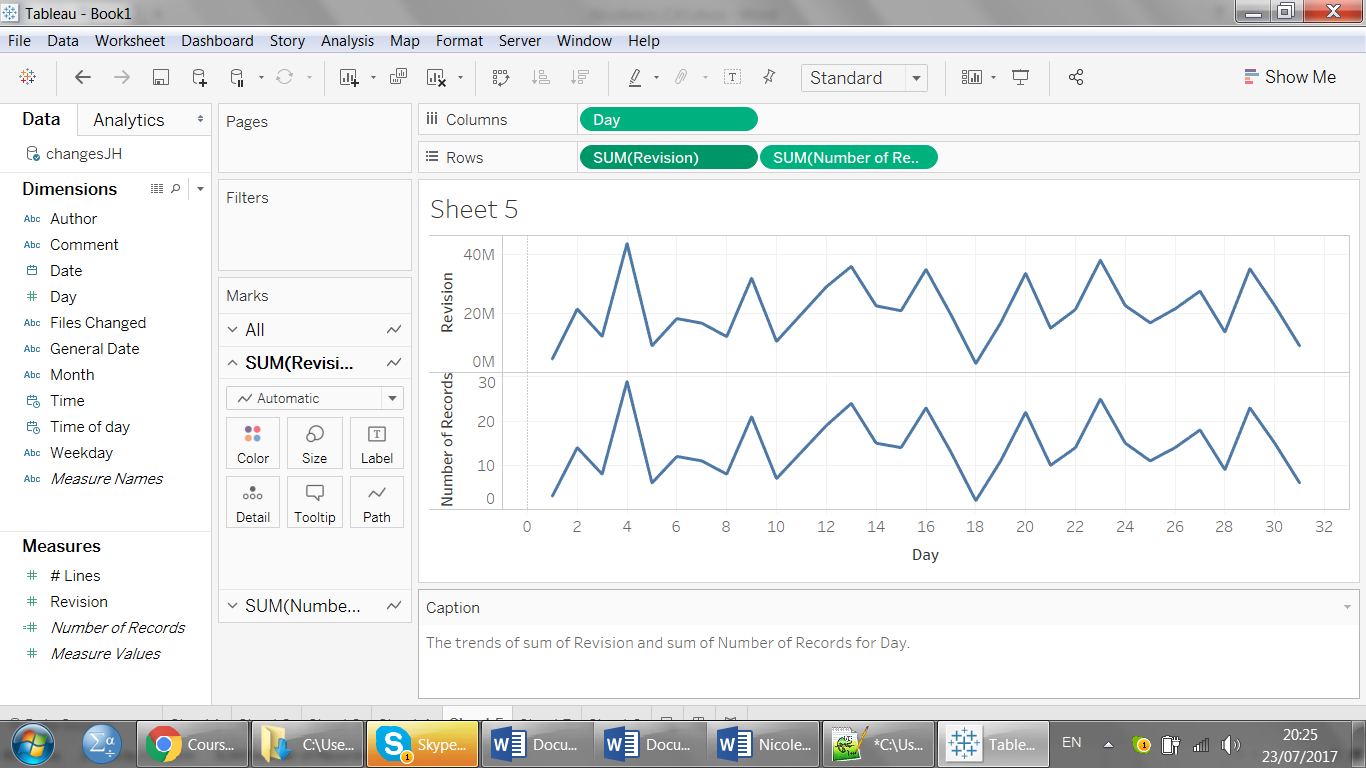
Programming for   
Big Data - ca5

B8IT105

The purpose of this report is to present results from the log file. Dataset was cleaned and prepared in Python. Traditionally CRISP-DM methodology is a good practice to work on data analysis. However, limited time designated for this project, thus only Exploratory Data Analysis is done in Tableau.

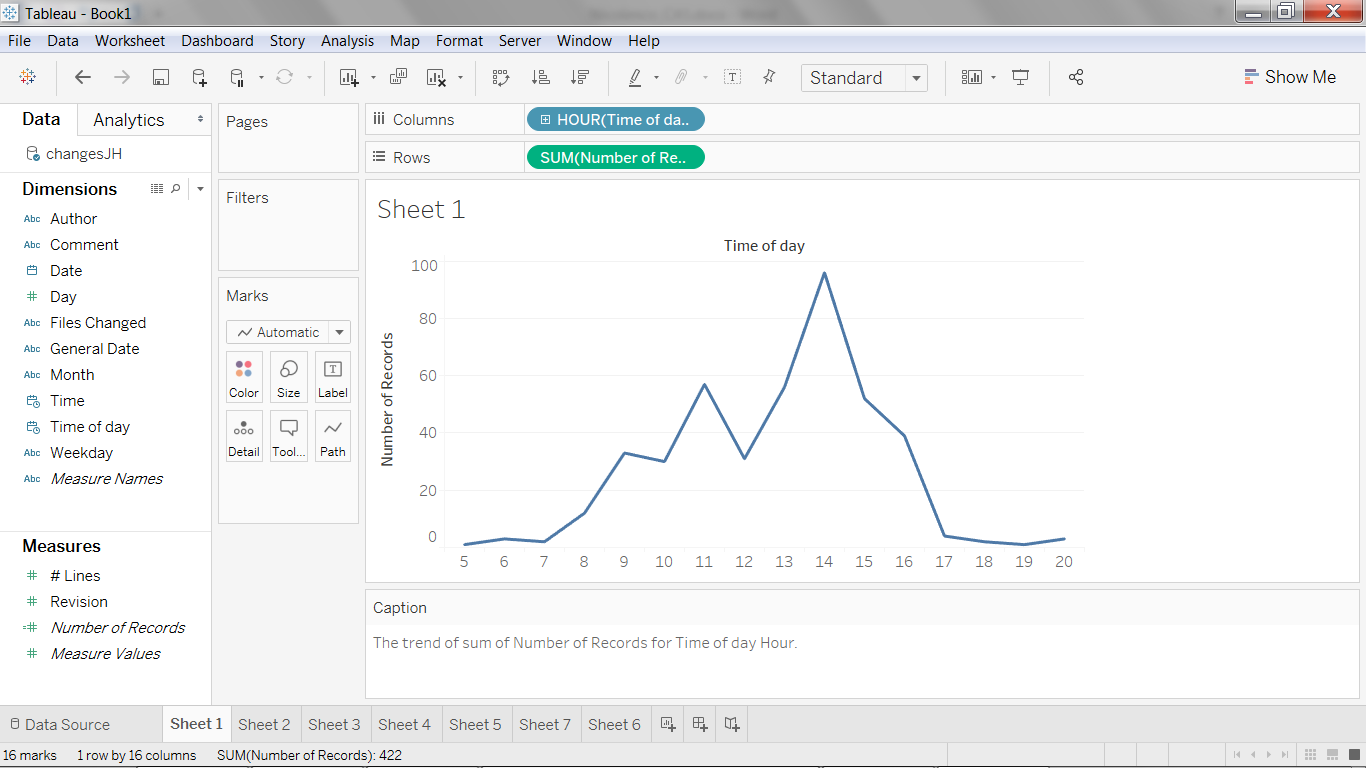
Since there is a limited information on this dataset, the meaning of Revision column data is unclear. Looking at the Revision and Records data in comparison (Fig.1) shows that it probably reflects the same type of information. For further investigation Records has to be used as a Sum Number of Records.

Fig 1: trends of sum of Revision and sum of Number of Records for Day



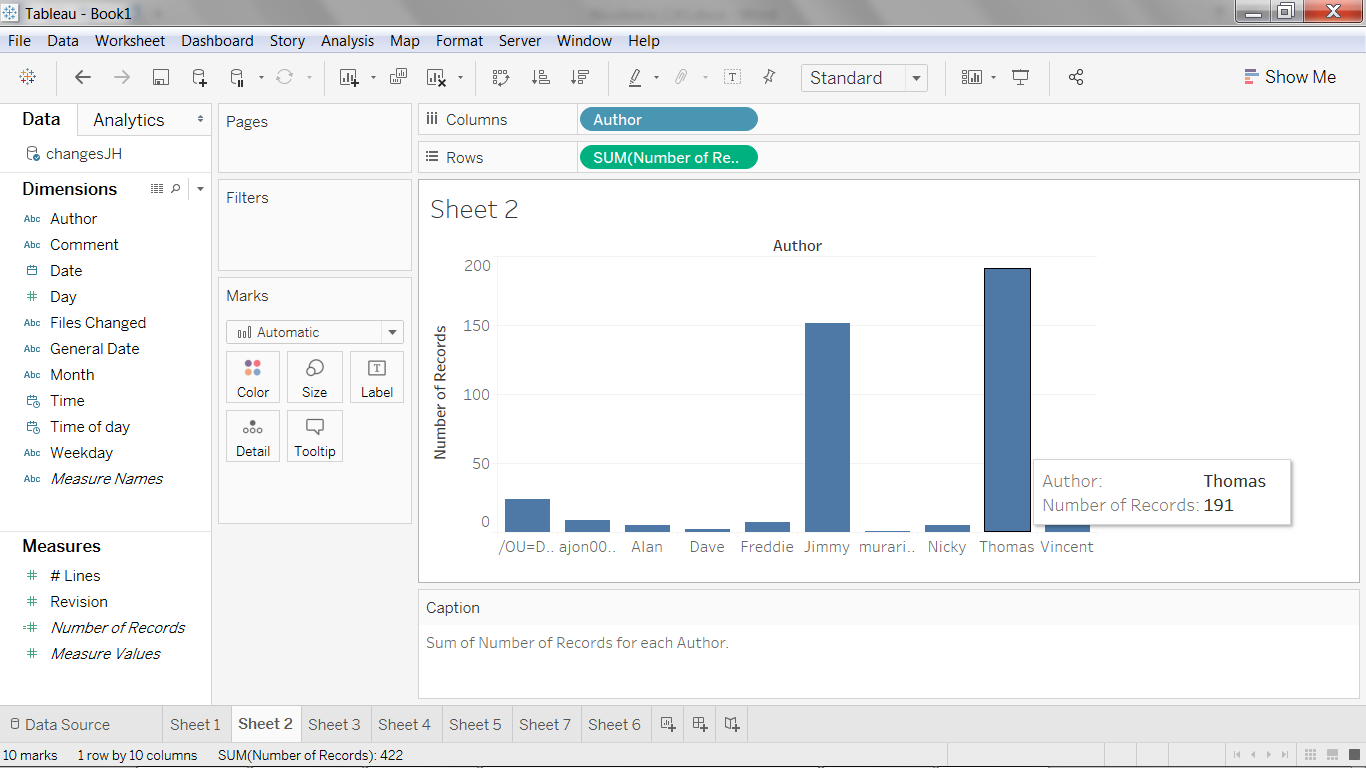
Looking at the trend of Hourly records shows that the most records was done after lunchtime. Productivity is better when people are not hungry ☺

Fig. 2: the trend of sum of Number of Records for Time of day Hour



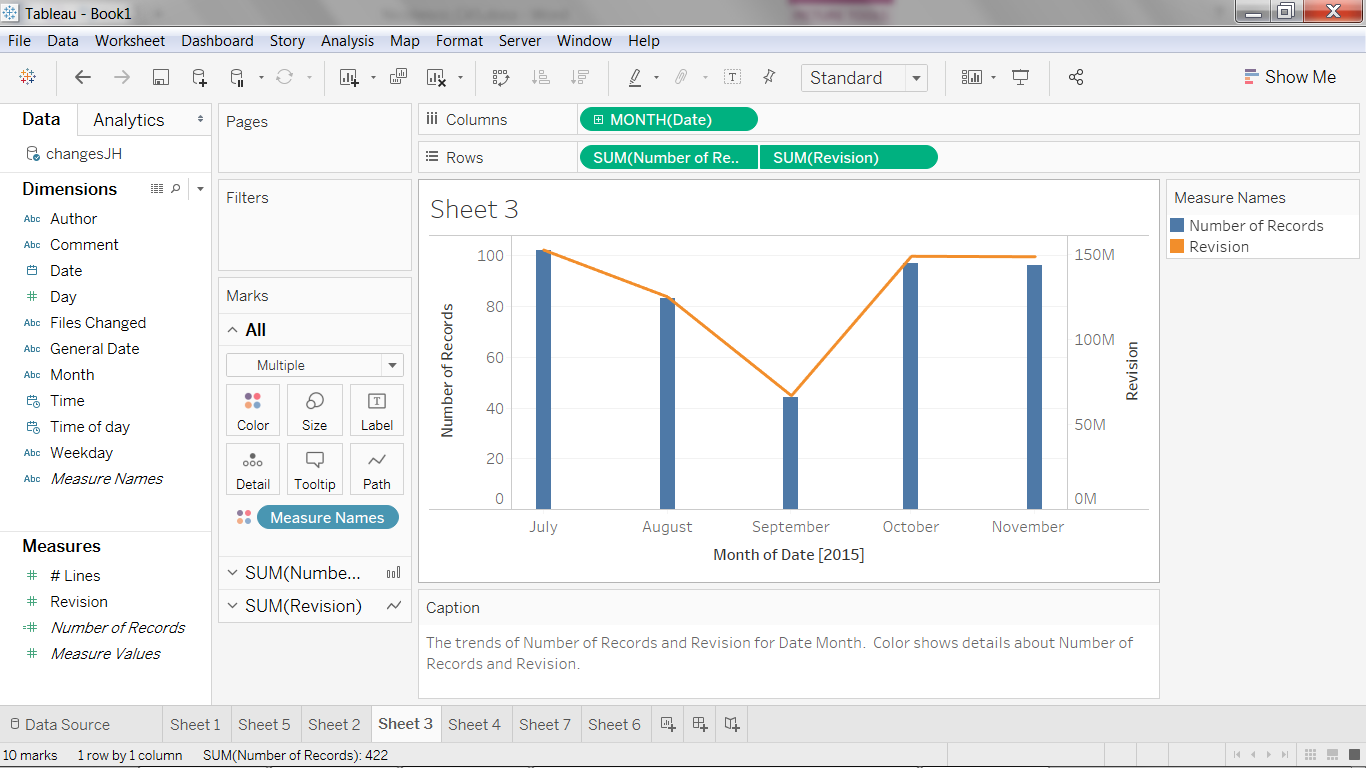
The most number of records belongs to two authors – Thomas (191) and Jimmy (152)

Fig. 3: sum of Number of Records for each Author



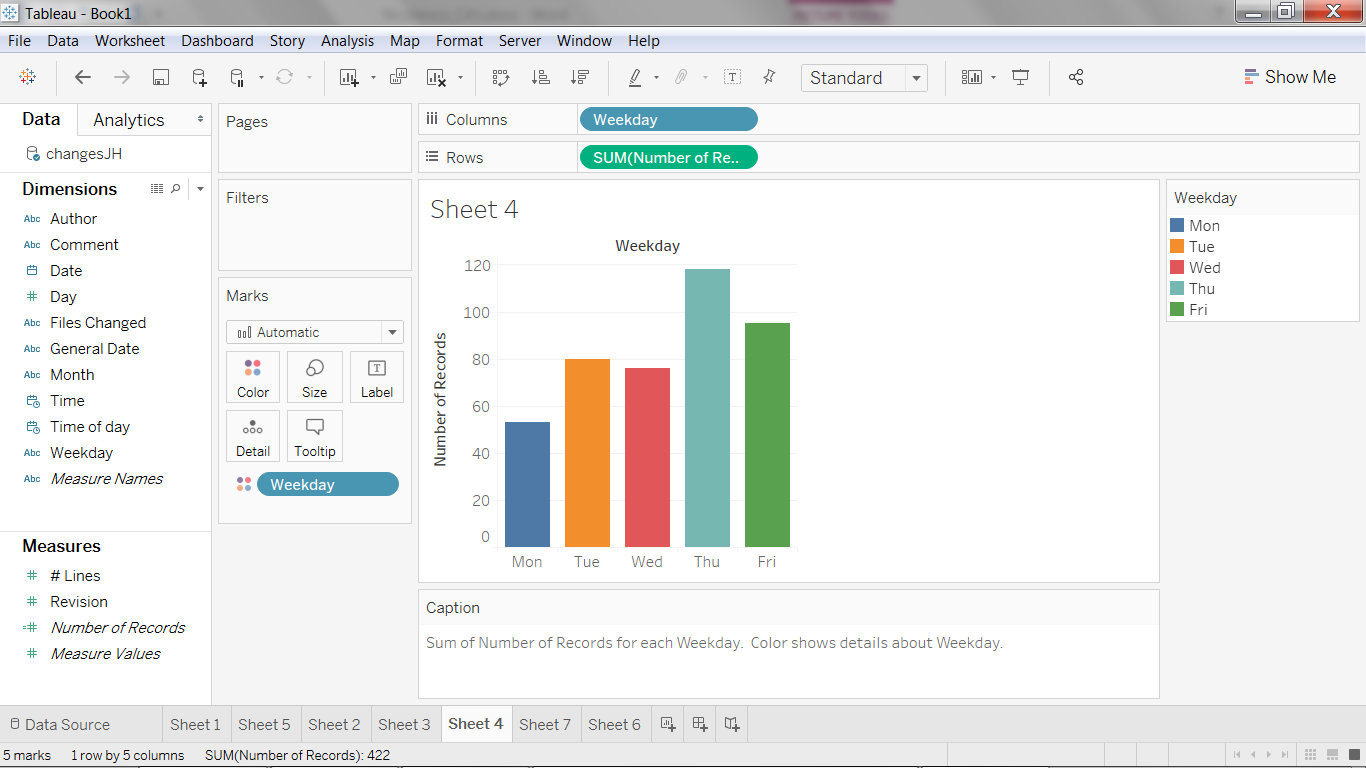
Least number of Records has been processed in September. Back to School is not only for children, but for parents too…

Fig.4: the trends of Number of records for Months



Monday saying reflects the situation at its best: “It takes twice as long for me to not get anything done on Monday as it does the rest of the week.”

Fig.5: Sum of number of Records for each Weekday



Observing data in this log file display that data is not synthetic and follow the natural “human trend of life”.