**Assignment: Discussion #2  
Name: Kenatu Bruce  
Student ID: C0770979  
Course ID: DAT1030**

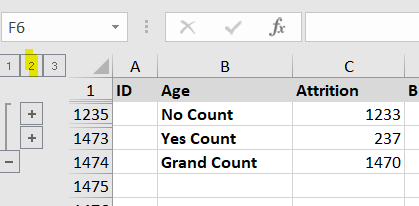
Assignment Objective:  
Using the business problem you selected, explain how the procedures you performed using SQL could have been performed with a different tool while also highlighting the key strengths and weakness of the tool you choose.

**Introduction**

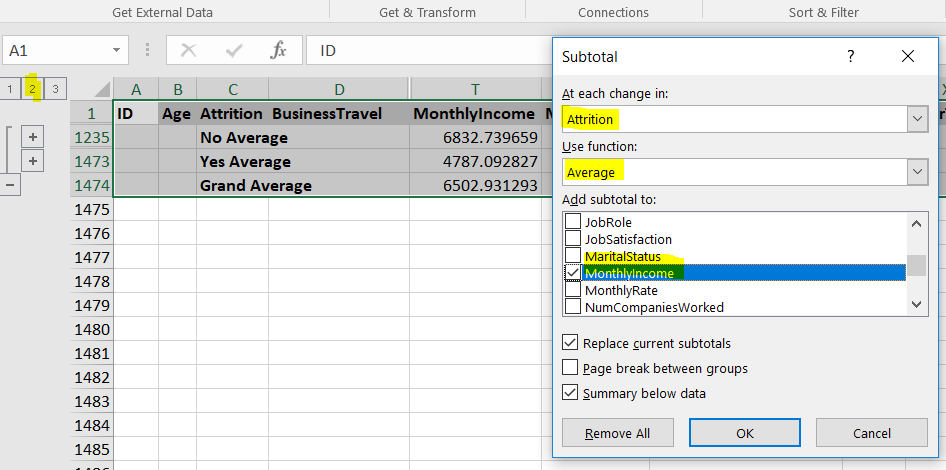
In the simulation we covered another of the seven transformative steps – aggregation | summarization. The functions looked at during the simulation were: **Average, Count, Min and Max**. These steps were relatively simple to execute using the SQL in Hive but there is another tool and platform that could have performed the same tasks much easier and more intuitively; I speak of none other than that popular work horse - **Excel Spreadsheets.** Excel is **ubiquitous;** everywhere you go you will come across some version of it.Excel has been around since the mid-1980s, and ever since its revolutionary debut, with each new release, it’s gotten more capable and more powerful.With the advent of **Office 365**, the limitations of Excel in the use of data analysis have dwindled rapidly, quieting some of the previous arguments like, **slow on large datasets**, **not able for users to collaborate** and **lack scalability and the ability to automate**.Excel thru Office 365 can now scale, it can handle millions of records at speeds comparable to relational databases of the same size, and with ***Power Query*** it handles **ETL** more elegantly, supporting the convergence of multiple data sources simultaneously i.e. (Oracle DB, MySQL, Files and CSV), something that many SQL platforms are not able to do and even has its own language **‘M’ [Mashup],** all the while remaining one of the most simple and intuitive tools for new users. *Excel is a useful tool for data aggregation. It comes with a wide variety of summary functions built in. Combine this impressive breadth with a simple user interface and you have a powerful tool for creating summaries***1.** Excel, already a tool in every data workers tool kit is keeping its top spot as the ***Swiss army knife*** of data analysis, even with the swell of Big Data; it’s positioned itself to handle the more frequent data analytic tasks for arguably more than 90% of the sector.

**Data Visualization**

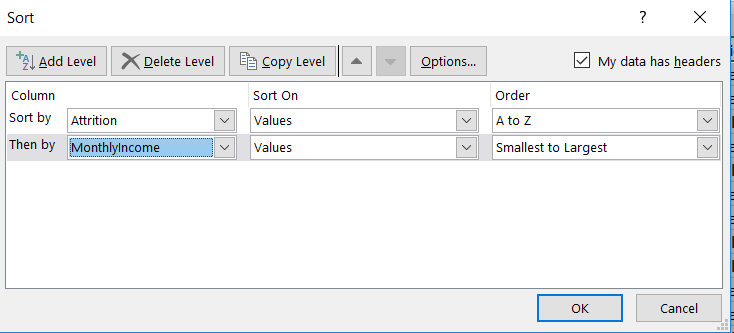
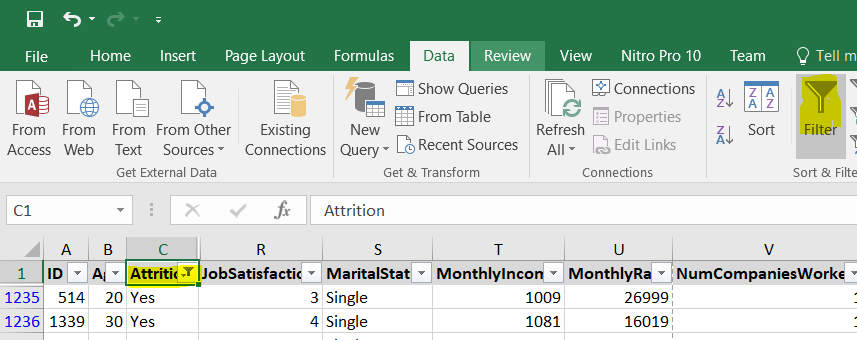
Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software *(WhatIs.com*). Data visualization is the graphic representation of data. It involves producing images that communicate relationships among the represented data to viewers of the images. (*Wikipedia*). In the simulation no meaningful data visualization that took place as it was all textual. SQL is a language for querying and manipulating data and has not such function. Excel on the other hand is simply a matter of define range > pick charts > edit appearance. In this simple similar way the functions used in the simulation could be replicated in Excel simply: **sort data** set on attrition field > select **subtotals** function under the data tab > specify **count** on change in attrition field value > subtotal on Attrition Field > view result in tab 2. To create new tables copy values to new spread sheet in work book.

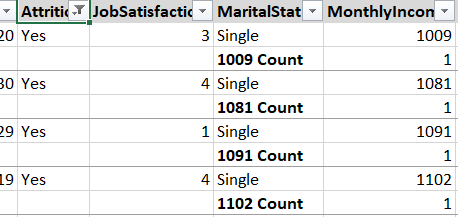


For the statistic results of Average, Min & Max monthly income, repeat the steps above selecting the appropriate subtotal functions, and the monthly income field for the subtotals. Rename columns as directed.



The third set of instruction in the simulation require that the data set be sorted on two levels, first by Attrition and then by monthly Income. Then you filter for Attrition (Yes) and then again for Non Attrition {No). The steps above can then be repeated this time aggregating on the change in value for the monthly Income field. Rename columns as directed.



**Strengths of Excel 2:**

* **Easy and effective comparisons -**With the powerful analytical tools included within Microsoft Excel you have the ability to analyze large amounts of data to discover trends and patterns that will influence decisions. Microsoft Excel’s graphing capabilities allows you to summarize your data enhancing your ability to organize and structure your data.
* **Powerful analysis of large amounts of data -**Recent upgrades to the Excel spreadsheet enhance your ability to analyze large amounts of data. With powerful filtering, sorting and search tools you are able to quickly and easily narrow down the criteria that will assist in your decisions. Combine these tools with the tables, Pivot Tables and Graphs you can find the information that you want quickly and easily even if you have hundreds of thousands of data items. While you will need the latest technology to get the best out of Microsoft Excel it is scalable and can be used at home on your low powered PC or at work on your high powered Laptop.
* **Working Together -**With the advent of the Excel Web App you can now work on spreadsheets simultaneously with other users. The ability to work together enhances your ability to streamline processes and allows for ‘brainstorming’ sessions with large sets of data – the collaboration tools allow you to get the most out of the sharing capabilities of Microsoft Excel. The added bonus is that as the Excel Worksheet is web based you can collaborate anywhere – you are no longer tied to your desk but can work on spreadsheets on the go – this is ideal for a businessman on the go.
* **Microsoft Excel Mobile & iPad Apps -**With the advent of the tablet and the smart phone it is now possible to take your worksheets to a client or a meeting without having to bring along your Laptop. The power of these mobile devices now allows you to manipulate data and update your spreadsheets and then view the spreadsheets immediately on your phone or tablet.

**Weaknesses of Excel:**

* When the datasets are very large (long), having to use a GUI to scroll through, perhaps, a few dozen columns of data, manually selecting, moving, deleting, etc. 4 can become tiresome. In Sql you can describe the table find the columns required then select them for display with a statement while filtering at the same time.
* You always have to ensure that your data set is sorted the way you want. In SQL the order is set by default and is usually ascending starting from the first column selected. There are however instances when you have to specify order in Sql to get the expected results.
* User Bias 3 - the downside of using spreadsheets is that only the information that the user chooses for analysis is included in the presentations, and therefore, other pertinent information that may influence decision making might be excluded, unintentionally. To address this many companies are using reporting tools such as Tableau and Qlik to make reporting of data more user friendly and comprehensive instead of relying solely on the spreadsheet. There is of course Power BI from Microsoft.

**Reference**

1 DAT1030 – Module 5, Section 79 – Tools for Data Aggregations

2 <https://turbofuture.com/computers/Advantages-of-Microsoft-Excel>

3 <https://smallbusiness.chron.com/advantages-disadvantages-spreadsheets-26551.html>

4 <https://www.reddit.com/r/excel/comments/bonank/sql_vs_powerquery/>