**Project One**

Lesley Potts-Langdon

Southern New Hampshire University

DAT 300: Data Valid: Getting Right Data

Professor Alex Pettit

March 29, 2020

Determine whether preliminary data represents/identifies the organizational challenge:

When looking at our preliminary data, there are some organizational challenges. The Kansas City data is stored in a different format than the Warehouse data. We have options for how to approach this, but much depends on the quantity of data. If we are transferring a small to moderate amount of data and only need to complete the task once the approach would be different than if there is a large amount of data and/or there will be an ongoing need.

If Kansas City is going to transfer over to the system currently used by Warehouse and the data can be moved in small bits, we could utilize an Excel spreadsheet as a midpoint for the data. Data can be exported from the as400 to Excel (Interform400, n.d.) and it can be imported to the SQL data base from the same spreadsheet (Microsoft, 2019). If there will be an ongoing need or we will be transferring very large quantities, it could be beneficial to write an application to read from as400, convert data as needed, and write to SQL.

Examine the data set to determine whether the data is usable or able to be analyzed:

The data set we have will need some conversions before it is able to be analyzed. The data from Kansas City is pulling into the Excel spread sheet in scientific format. How this is addressed will be determined by when the conversion from an integer to scientific format occurs.

In looking at the Excel spreadsheet, we can see that the number in the formula bar matches the sales numbers from the list we were provided. In this case, we simply need to change the formatting of the cells to allow the number to remain in its original formatting. If the conversion was occurring in the originating database, we would need to either modify the way the sales numbers are stored in the as400 database or manually adjust as needed after the data in imported into the SQL database.

Assess the preliminary data set for completeness and accuracy:

As it is right now, the Kansas City data set is not complete or accurate. The accuracy of the data that we do have would be remedied by fixing the formatting. The data is incomplete. We have three months with errors, one month with no data at all and we have one less month of data than we do for the other cities in the Warehouse data. If this were all the data we were dealing with, it would be easy enough to manually fill in the needed data but, as this is likely to be a bigger issue we will need to evaluate why the data is not transferring correctly and find a remedy.

Determine which data should be retained or analyzed to address the organizational challenge:

We will utilize all the data in this limited data set to address the challenge of merging the incompatible databases. It is to our benefit to solve the problem utilizing a smaller subset of data because there are fewer factors to work with at first, but it will also be a scalable solution that we will be able to apply to the entire database.

Identify limitations for data usage in regard to regulatory, ethical, and legal considerations:

As we are looking at this smaller data set, there should be no regulatory, ethical or legal issues with this project. The only thing we need to verify would be if there are any restrictions on sharing data from the time prior to the acquisition of the Kansas City company. There could be other considerations when we pursue a complete merge of all data.

References

Interform400. (n.d.). *InterExcel400.* Interform. <https://interform400.com/en/interexcel400>

Microsoft. (2019). *Import data from Excel to SQL Server or Azure SQL Database.* Microsoft. <https://docs.microsoft.com/en-us/sql/relational-databases/import-export/import-data-from-excel-to-sql?view=sql-server-ver15>