Zachary Tankersley

Southern New Hampshire University

DAT-430-J7783

February 15, 2023

**Project One**

The primary question fueling this inquiry is: “Why are employees leaving their jobs.” To attempt to isolate as few potential root causes as possible we need to discover which factors may be affecting employee attrition rates. The prevailing ideal is that the employees are feeling negative consequences of growth via decline in job satisfaction and decline in direct engagement. To accomplish this task I must gather the HR attrition data, aggregate it into a single source and do some exploratory data analysis to gain insight. To ease this process, non-relevant parameters will be eliminated, the rest of the parameters will be inspected for “cleanliness” and cleaned if need-be, and then fed through different analysis tools and methods. Optimistically and at a first glance, it would appear that most of the numeric parameters are fine, however many of the non-numeric categorical and binary parameters need to be converted.

The data will come from a file directory stored on a virtual machine/repository and is broken into several .CSV files that will need to be imported and combined into a single larger file for manipulation and exploration in Python. Python is certainly preferable for automating the combination and import of the files; however I must admit that I do like Excel for a lot of preprocessing and cleaning tasks. Reason being you can very easily undo and revert any changes to the table(s). It is also very easy to isolate various cells and modify only that data in Excel. In this way, missing data can easily be inserted into missing cells, categorical data can easily be removed/modified/replaced with the proper data type. Excel has very robust, built in functions and features for this, packaged with a nice GUI. For feature scaling, it would depend on the scenario, but I would likely lean on Python for this as you could simply create a loop to iterate through the data and apply whatever scaling arithmetic you need. Once the data has been prepared and preprocessed, it can be moved between Python or Excel, etc. as needed in .CSV format.

For reporting I will likely also use Excel, or Tableau, as they both are fairly dynamic in their ability to intake new data and represent it in a constantly updating visual form. The report has to be able to update when new data is fed to it over time, and it needs to be easily modifiable if need be. To create visualizations that are going to be interpretable by the leadership team, the parameters being visualized need to be pared down to only what is necessary. From there, some of the basic key visualization tenets should be used: lines, colors, space, (Adobe, 2021). KPI’s should be highlighted or accentuated in some way. The message concepts should be kept simple and direct. It is important to make the visuals stand out in the *right way* in order to help stakeholders make the decisions that can be most impactful and beneficial to the organization. Attrition can be a huge detriment to the atmosphere and profitability of our organization, and any way we can decrease that should be looked at carefully.

**References:**

Adobe. (2021, January 12). *What Are The 6 Elements of Design? | Adobe XD*. Ideas. https://xd.adobe.com/ideas/process/ui-design/6-elements-design/