**Project One: Part 1**

Kayla Sacks

Southern New Hampshire University

DAT-430 Leverage Data for Org Results

Professor Aubrey Condor

September 29, 2024

## Introduction

The goal of this project is to first find the conditions that worsen attrition for the company then to visualize the patterns in the data to highlight these conditions. In this part of the project, I will express what information needs to be gathered to find what conditions worsen attrition. I will also explain the data source and how it will be handled and how reports will be monitored. Finally, I will also explain the creation of different visuals to help understand the data and why they are important.

## Defining the Questions

As stated, the first goal of this project is to answer the question of “what conditions worsen attrition?” The dependent variable will be attrition and the features that affect it will be found based on correlation analysis and regression models.

## Sourcing and Preparing the Data

The data that is being used is from 15 different files which the required information is saved in. These files will be merged to make it easier to work with the data, which means that the preprocessing method of aggregation is going to be used. Since I noticed that the column for ‘training’ was different between files, I renames all ‘training’ columns to be the same before merging to ensure consistency.

Once the data is merged, it will have to be cleaned as there is missing and mismatched data such as Gender being listed as M/F in some files and Male/Female in others. There are also redundant variables that will be removed such as the ‘Over18’ column since there is an ‘Age’ column. The number of columns regarding rate or income will also be reduced as they also become redundant. Rows that are missing important data such as attrition will be removed. Categorical data such as ‘Gender’ and ‘BusinessTravel’, among others, will be changed to numeric to be easier to handle.

## Monitoring Reports

Keeping the reports up to date will be essential as changes to the report can provide us with new or clearer insights into the problem we are trying to solve. It is also possible that the cause for the high attrition can change over time, so we want to be sure to watch for any change once the current cause is found and resolved. To help us with this task, we can have the report automatically updated when new information is entered. This will allow us to spot changes in patterns or new patterns that become prevalent.

## Creating Visual Reports

Visualization has become essential when providing analysis reports as they make the data easier to read and understand for those that may not have technical knowledge. Problems need to be addressed and resolved as quickly as possible and to do that a basic understanding of the problem and its cause is necessary. Dynamic visualization can help to reduce the time and effort that would usually be required to understand the data. One example of this is a chart that shows how attrition increases, or decreased, based on a variable such as job satisfaction or level of training.

It is essential that the visualizations used show the variables connected to the problem being solved, which is this project will be centered around attrition and what conditions increase attrition. It is also important to remove any potential bias that may show in the visuals due to things like incorrect scales, colors, or shapes that might not express the data correctly. Using to many visuals can also cause data to be misleading as it can overwhelm the audience (Yung, 2024). To conclude, it is important for the visuals to show only the data that is needed to solve the problem while using the least amount of visuals to clearly express the data.

## Citation

Yung, Z. (2024, April 19). *Misleading Data Visualization - What to Avoid*. Coupler.io Blog. <https://blog.coupler.io/misleading-data-visualization-examples/>