Kelsey Barton
DSC530-T304
Final Assignment
March 2, 2024

Attrition in the Workplace

The data scientist was presented with an opportunity to explore an area of interest while incorporating a dataset. While presented with many opportunities, the scientist recently experienced attrition within the workplace and chose to explore the causation of attrition. Certain questions that the scientist used to aid in discovering the perfect dataset included: What elements lead to an employee leaving an organization? What elements keep an employee interested in an organization? What are the main causes for the longest employed employees within the organization? Does salary play into attrition? Is age a factor in attrition? Does work from home affect attrition?

The dataset chosen provided a large amount of useful information and had many variables incorporated within it. The main variables explored within the experiment were age, attrition, daily rate, travel distance from home, and education. The experiment showed that many of the employees were between the ages of 25 and 45. The ages ranged from 18 to 60 with most of the employees between that middle position. Many of the employees were paid similar daily rates, but there were some instances where certain employees were paid drastically more than others. This study showed that many of the employees did not travel far from their homes to their jobs. The scientist believes that many of the employees work remotely while about half of the employees travel distances of 2 miles up to 30 miles to get to work. The education levels differed within the organization, but it appeared that most employees did not have an extensively long educational history. The attrition results showed that with over 1,400 employees researched, only around 200 of them have left the organization. The majority of the employees remain employed at the organization. The outliers within the experiment involved employees below 20 and above 60 years of age. Other outliers were extremely high or low daily rates, very minimal distances to work, and unusual education levels. The data had to be cleaned, and if the data entries were legitimate, they were incorporated into the analysis. The data entry errors were adjusted out of the dataset. While the scientist was presented with a large amount of data within this experiment, the most prominent variable that leads to an employee leaving an organization was inconclusive.

The scientist could have explored more relationships between variables, specifically the relationships of each variable to the attrition factor within the study. While the study provides a lot of information, the scientist did not come to conclusion on the exact cause of attrition. The dataset provided the appropriate number of variables for the analysis, but the scientist needed to create better relationships between the variables. The scientist believed that salary played a large role in attrition. This does not prove true within the dataset. The scientist faced challenges in her coding skill set. As she is a young data scientist, she was unable to appropriately connect the variables through the code to find a direct cause for her research question.

Citations

Downey, A. B. (n.d.). Think Stats. Retrieved from https://platform.virdocs.com/read/629508/198/#/4/4/2,/1:0,/1:0

Pavansubhash. (2017). IBM HR Analytics Employee Attrition & Performance [Data file and code book]. Retrieved from https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset