4-2 Project One: Part One

DAT-430

Ashley Littles

Southern New Hampshire University

My organization has grown very rapidly which in turn has caused concern with our leadership team as they notice that their employees are lacking direct engagement and job satisfaction because of this new growth. I will be analyzing my leadership team's HR attrition data and creating metrics to help them draw conclusions related to why the employees are choosing to leave their jobs.

For this situation we are trying to answer if there is a correlation between the lack of direct engagement and job satisfaction with our organization's rapid growth. In order to answer this question, I will need to figure out exactly what variables are needed. Based on the question, we know that the leadership team is specifically concerned with the variable HR attrition. So, we will also need to find the surrounding variables in relation to HR attrition. Doing this will help set up the appropriate parameters in order to answer the above question.

We have a total of 15 files with data that have 23 columns and 7040 rows that have to be merged and preprocessed. A preprocess approach that will be used is data cleaning. Looking over the available data we can see that there are missing and duplicate data that need to be taken care of through cleaning the data set. For missing data, we will use a global constant such as N/A. With categorical data we will create new columns to assign to each value. Lastly, to handle feature scaling we will use normalization or standardization for the range of independent variables.

Over time the data will update, and we will have to monitor these reports as the new data loads. With the new data load the reports could display different or new reasons for why employees are choosing to leave the company other than the first initial reason. With these new changes the report will either have more or less data that must be analyzed.

Creating reports visually can allow for analysts and stakeholders to be able to better understand how the results came together. To create dynamic visualizations that are easy to understand as well as influence decision making you can start by choosing a chart with only the necessary data needed. For example, most people can easily understand a line or bar chart. In the visualizations, it is important to show how the variables relate to each other and their outcomes to help decision makers come to accurate conclusions.

References

Azevedo, N. (Sep 29, 2023) *“Data Preprocessing Techniques: 6 Steps to Clean Data in Machine Learning”* scalablepath <https://www.scalablepath.com/data-science/data-preprocessing-phase>

Vashisht, R. (Jan 06, 2021) *“When to perform a Feature Scaling?”* Atoti <https://www.atoti.io/articles/when-to-perform-a-feature-scaling/#:~:text=What%20is%20Feature%20Scaling%3F,during%20the%20data%20preprocessing%20step>.