Zachary Tankersley

Southern New Hampshire University

DAT-430-J7783

January 29, 2023

**Module Three Assignment**

As part of a quick exploratory data analysis (EDA) of a subset of HR Training Data, a baseline will be established to help answer the following:

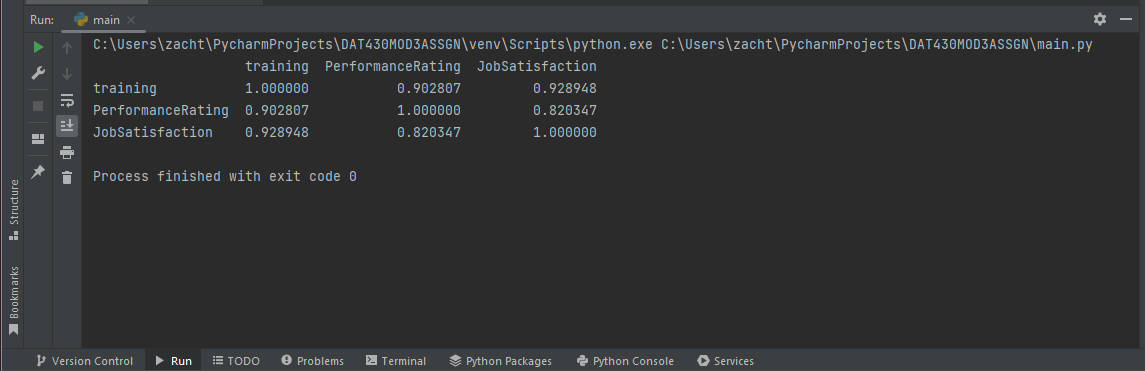
*How well are employees doing with provided training. Specifically that is, does training have an effect on performance or job satisfaction?*

As part of establishing a baseline, the data which is stored in the form of a .CSV file named “HR Training Data” will be used. This data will be imported into Python which will be used to analyze and visualize the data to create predictions for the HR department pertaining to training.

A “common sense baseline” is how a problem would be addressed if we had no knowledge of data sciences whatsoever. A common sense baseline should be simple, which therefore can be hard to beat, and often only be a slim margin (Ramakrishnan, 2018). So while python and other tools may be a great way to analyze data quickly, I want to take a quick look at the data that needs to be analyzed. I can see with a quick glance that it does indeed appear that the more training an employee gets, the higher the performance (“PerformanceRating”) and satisfaction (“JobSatisfaction”) tend to be:



While 5 is a small sample size, hopefully this crude prediction holds true, in fact, it appears that training vs performance and satisfaction are almost at a 1:1 ratio. For every unit increase in training, there is *just about* one unit increase in both performance and satisfaction. In fact, you can see the correlation matrix for the three metrics here:



Based on the correlation matrix, there appears to be a strong relationship between training and both variables.

When analyzing the data set as a whole and using an ordinary least squares regression technique, the analysis yields some interesting results. The model, according to the summary analysis function the model formula would be:

However, the significance rates shown in the OLS technique are troubling and suggest that the correlation between the variables may not be as strong as originally anticipated.

*References:*

Ramakrishnan, R. (2018, June 19). Create a Common-Sense Baseline First - Towards Data Science. Medium. https://towardsdatascience.com/first-create-a-common-sense-baseline-e66dbf8a8a47