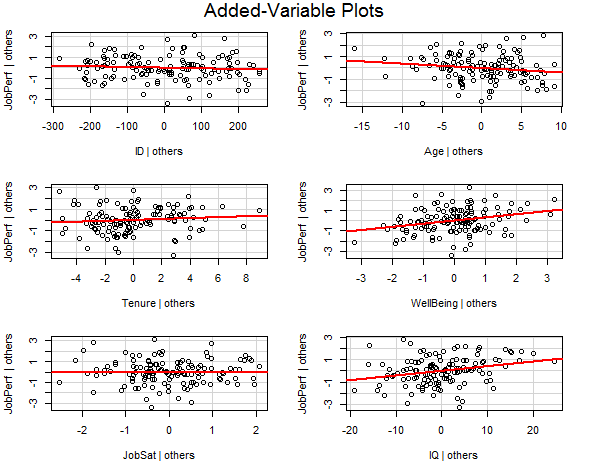
David Lowe

Job Performance EDA

1. It is known that job performance is a large factor in producing success in a company. Job satisfaction and employee wellbeing have been tied to the employee’s job performance. In an effort to see what factors affect the job performance of an employee, data was collected to rate a university’s employee age, tenure, happiness, job satisfaction, job performance, and IQ. From these variables, I hope to find the relationship between job performance and job satisfaction. If there are variables that strongly affect job performance, it can help companies to know how to get the most out of their employees.

|  |  |
| --- | --- |
| **Variable** | **# of Missing Values** |
| Age | 0 |
| Tenure | 0 |
| Wellbeing | 160 |
| Job Satisfaction | 160 |
| Job Performance | 64 |
| IQ | 0 |

1. The data is organized with each individual employee having a value for age, tenure, and IQ, all with values greater than zero; and wellbeing, job satisfaction, and job performance with values between zero and ten. Since job performance is what we are trying to understand, it will be our response variable. This may be problematic, only because the values of job performance only range from zero to ten. I need to be careful when making predictions, as they may fall outside of these bounds. However, understanding the relationship between these variables and job performance is much more important than predicting job performance based on a given set of variables. That being the case, I don’t worry too much about the predictions from this analysis. One troubling characteristic of the data is that there are many missing values in a few of the variables. Notably, 73% of the data has some type of missing value. I noticed that there is never a missing value in both job satisfaction and wellbeing, meaning that everyone answered at least one of these questions. Missing values require that I either drop data out of the study, or replace them in some way. Both methods require sacrifices in the model.



1. Besides the difficulty of having a high amount of missing data, the data seem to be linearly related to job performance, which satisfies the assumption of linearity needed for linear regression (see added-variable plots). Multiple Linear Regression (MLR) would allow me to see the relationship between the explanatory variables and job performance. MLR can provide easy interpretation of the effect the provided variables (age, tenure, etc.) have on job performance.
2. I don’t know how to replace missing data in a dataset without changing my results drastically.