**Umuzi assignmentc3: Predicting salary differences**

**Multiple Linear Regression**

In the next series of challenges, we will predict employee salaries from different employee characteristics (or features).

Import the data salary.csv to a Jupyter Notebook. A description of the variables is given in the metadata. Answer the following questions:

Use multiple linear regression to predict salary from all the variables in the dataset.

1. Create scatterplots, histograms, and a descriptive statistics table of the variables of interest.
2. Recode the variable Field into three dummy variables, using HR as the reference category.
3. Produce a correlation matrix comparing the relationship of salary to the predictor variables. Is there any multicollinearity or other problems that may be a problem in the multiple regression?
4. Run the multiple linear regression and interpret the correlation coefficients. What are the most important features when predicting employee salary?
5. Plot the standardised residuals versus the standardised predicted values. Are there any problems with the regression?
6. Get and interpret the partial coefficients and tolerance values for the regression model. Are there any problems with this model?