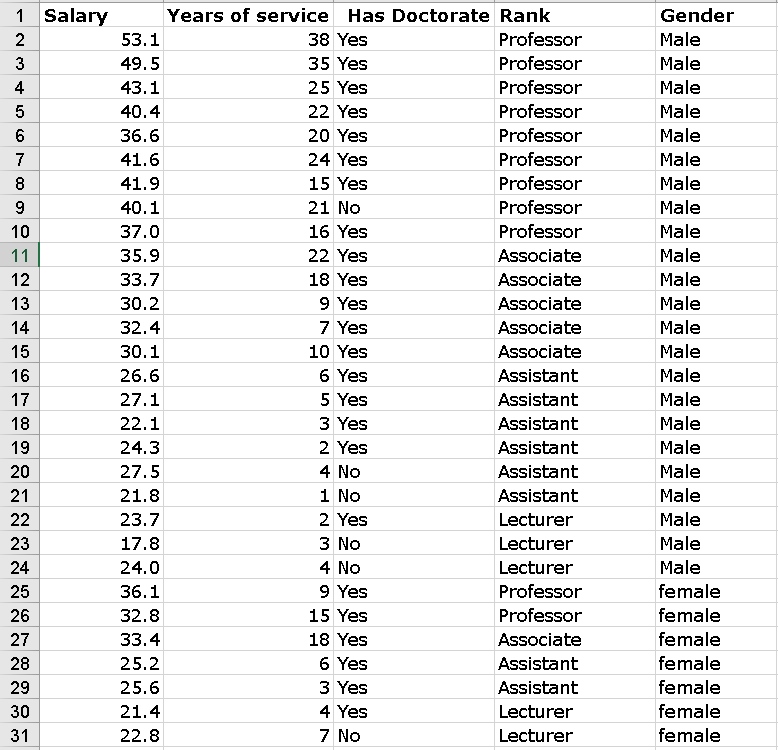
**Linear Regression: Use of Dummy Variables**

Linear Regression requires that all explanatory variables (independent variable) are continuous numeric variables. Consider the following data set for Linear Regression, where **Salary** is the dependent variable, and the rest are explanatory variables. Note that **Has Doctorate**, **Rank**, and **Gender** are categorical variables.



When a **categorical** variable like **Gender** is an explanatory variable, it can be transformed into numeric data with the help of **one** dummy variable (**indicator** variable)

**Gender** has two possible values (Male, Female). Hence one dummy variable is sufficient to represent the information. **1** if the instructor has Doctorate, **0** Otherwise.

Similarly, **Has Doctorate** alsohas two possible values (**Yes**, **No**). Hence one dummy variable is sufficient to represent the information. **1** if the instructor has Doctorate, **0** Otherwise.

The categorical variable **Rank,** however**,** has 4 possible values (Professor, Associate, Assistant, and Lecturer). Hence 3 dummy variables that take 0 or 1 values are needed to represent the **Rank** information (see the transformed dataset below). **1** in Professor Column indicates the record (row) is of a Professor. Similarly, **1** for Associate Professor or Assistant Professor. (0, 0, 0) for Professor, Associate, Assistant indicates the row is of a Lecturer.

In the transformed dataset below, all categorical variables have been replaced by dummy variables. In general, a categorical variables with **k** values can be represented by **k-1** dummy variables.

