**Context**

The goal of the project is to provide a wage offer list and a total weekly salary budget for the new employees acquired through the recent acquisition of another company. These objectives will be achieved using linear regression. The data available has eleven attributes. There are attributes for an employee’s ID number, weekly wage in dollars, hours worked per week, IQ, education, experience, tenure, age, marital status, urban residence status, and the number of siblings.

**Data Preparation**

The first data preparation step checked for any missing values in both datasets. There were no missing values, so no further data preparation steps are necessary for handling missing values. Next, the ranges of each attribute are checked. The ranges are checked because the training data set cannot be relied upon to predict wages for employees whose values fall outside the scope of the training set’s values. Therefore, observations outside the training set’s range are filtered out. Lastly, the training data set is checked for outliers. Outliers were detected, and the outliers were replaced with the mean of their respective attribute. Only the outliers in training data are handled because that is the data that trains the linear regression model. So, outliers are handled to improve the model’s reliability by reducing the skew of the data. Outliers are not addressed in the scoring data set because those observations are the new employees who need a predicted wage; hence none of those observations should be removed.

**Building the Model**

The first step in building the model is to remove Employee ID because it is not predictive of wage; instead, it is an identification attribute. The initial model ran a linear regression using wage as the dependent variable and every other attribute as the independent variable. The initial results revealed that age and the number of siblings do not predict wage. So, those attributes are removed, and the model is rerun. The second regression revealed that hours are not predictive of wage, so hours are removed, and the model reran. The third regression confirmed that an employee’s IQ is predictive to the 99% confidence level, an employee’s education is predictive to the 99.9% confidence level, an employee’s experience is predictive to the 99% confidence level, an employee’s marital status is predictive to the 99% confidence level, and an employee’s urban residential status is predictive to the 99% confidence level. The model is checked for collinearity using a variance inflation factor and a correlation matrix. The VIF and correlation matrix showed small to moderate correlation in the model but no severe correlation that will reduce the model's reliability.

**Applying the Model**

With the model complete, it is applied to the data set of the new employees. After the model is applied to the data set, a wage can be predicted for each new employee. Therefore, the first objective is complete. Lastly, the sum of the predicted wages is calculated. Since the wage is weekly wages in dollars, the sum of the predicted wages represents the total number of dollars spent in wages per week. The supporting code and visuals can be found on the following pages.

**Supporting Code**

**Graphical user interface, text, application, chat or text message

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Supporting Visuals**

Table

Description automatically generated

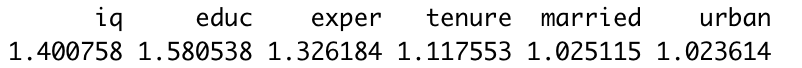
\*Results of final linear model

-------------------------------------------------------------------------------------------------------------------------------



\*Correlation matrix

-------------------------------------------------------------------------------------------------------------------------------



\*Variance inflation factors

-------------------------------------------------------------------------------------------------------------------------------

Table

Description automatically generated with medium confidence

\*First 15 observations of wage list

-------------------------------------------------------------------------------------------------------------------------------

A picture containing shape

Description automatically generated

\*Weekly salary budget