Analysis

**Data Pre-processing**

The raw data was pre-processed step by step to generate clean data. During the course of pre-processing, initially, some of the predictor variables (like 'NumCompaniesWorked', 'Attrition', 'EmpJobRole', 'YearsWithCurrManager', 'TotalWorkExperienceInYears', 'ExperienceYearsAtThisCompany', 'Age', 'EmpHourlyRate', 'TrainingTimesLastYear', 'BusinessTravelFrequency') were dropped due to their insignificance discovered in the model in the later stages.

Then, the cleaned data was used to identify categorical variables such as, "Gender", "MaritalStatus", "OverTime", "EducationBackground", "EmpDepartment". Then these categorical variables were one hot coded, to convert categorical variables to numerical form.

**Data Modeling**

Then, the Linear Regression Machine Learning algorithm was chosen to create the model. Linear Regression algorithm was chosen because of its reliability.