

CSCI 183: Project Proposal

Project Title: Salary Dynamics - Exploring the Impact of Experience on Earnings

2 Team Members

Project Idea:

In this project, we aim to use a dataset obtained from Kaggle that provides information on the years of experience and corresponding salaries for thirty employees, as of now (i.e. data set may grow if there is more data available). The project is inspired by our current research on the profound effects of human capital on economic prosperity. We are also motivated to understand the priority of employee development and career mobility, which both continue to be derived from professional experiences.

The primary objective is to develop a linear regression model to predict salaries (dependent variable) based on years of experience (independent variable). Given that we are using a simple linear regression model, we will most likely be able to minimize overfitting by deleting any possible outliers and reducing the complexities of the model. However, overfitting will, again, most likely not occur because our simple linear regression model includes one independent variable.

Process:

Below is a rough outline of the steps we will be taking in order to address the problem:

- **Importing Library:** In this initial step of our data analysis process, we lay the groundwork by importing essential libraries and loading our dataset. We will import NumPy, Pandas, and Matplotlib libraries.
- **Visualization:** Visualizations can aid in evaluating the performance of the linear regression model. We will be concentrating our learning and observations on two types of plots: heat map and scatter plot.
- **Data Preprocessing/Train + Test:** This phase is in case we need to conduct any type of feature scaling in order to scale numerical features to a similar range, which would

ultimately avoid biases in the model. Moreover, this step will also be necessary to split the dataset into two subsets: a training set and a testing set.

- **Generating a Final Linear Regression Model:** First, we will create an instance of the linear regression model and, then, use the `fit()` function to train using training data. During this process, we believe the model should be able to learn the coefficients (weights) that best fit the data, minimizing the error. Once the model is trained, we will use `predict()` to make predictions on testing data.
- **Final Model Evaluation:** After making predictions, we will evaluate the performance of the model using mean squared error (MSE) metric. A lower value of MSE indicates better predictive accuracy, as it means that the model's predictions are closer to the actual values. We also do not wish to have a 100% accuracy because achieving this accuracy may indicate overfitting, where the model has learned the training data too well and fails to generalize to unseen data.

Software Needed:

- Kaggle
 - We retrieved the data from Kaggle in a .csv file.
- Google Sheets
 - We imported the data into a google sheet to store it.
- Jupyter Notebooks
 - We implemented our project code on a shared notebook using Google Colab.
- Python
 - We used Python and its libraries to write our code and implement our linear regression model.
- MATLAB
 - We visualized our data with graphs by using this tool.

Scholarly Papers:

CNBC: We believe that years of college experience also come under the umbrella of “experience” these years contribute to overall education and skill development. This experience includes not only academic learning but also practical experiences gained through internships, research projects, and extracurricular activities, which can be valuable assets in the workforce.

According to the CNBC article, “those with a bachelor’s degree earned a median of \$61,600 a year in 2021, compared with a median of just \$39,700 among those with only a high school diploma.” The difference between these two salaries is significant and, therefore, years of experience (i.e. from college to professional career setting) impact salary heavily.

Dice/Indeed: We desired a better understanding of the impact (in percentage or ratio) of experience on salary because we recognize that experience is not the sole reason for salary increase. According to a recent Indeed article, employees who showcase higher levels of dedication, reliability, and have higher performance reviews are more likely to get a raise. How about the individuals who carry many years of experience? According to a Dice article, there was a “6.6 percent average salary increase [in 2022] for those with 3-5 years of technical experience” in the tech industry. Therefore, although the employee’s technical knowledge and work efficiency is all important, it has become more common for managers and organizations to demand a high level of experience to ensure salary raise.

McKinsey: We read this article in order to get a better understanding of how work experience is a crucial component of human capital. The article discusses how work experience impacts earnings, citing that in countries like the US, Germany, and the UK, work experience contributes a substantial percentage to average lifetime earnings. Specifically, the article states, “skills acquired or deployed through work experience contribute an average of 46 percent of this value over a typical working life.” Overall, this theoretical framework and empirical evidence leads us to further contemplate our project idea.