

# Analyzing the Changing Trends in the Computer Science Job Market Over Time (COMP3125 Individual Project)

Malachi Elisnord

*Data Science*

**Abstract**— This study explores the evolving trends in the computer science job market, specifically focusing on salary changes, the disparity between software engineer salaries and fresh graduate earnings, geographic salary differences, and the connection between employment status and median salaries. By analyzing datasets from 2012 to 2023, this project aims to provide a comprehensive understanding of the factors influencing the software engineering profession, offering insights into career trajectory, employment opportunities, and salary growth in the field.

**Keywords**—software engineer salaries, employment trends, geographic disparities, computer science graduates, job market analysis.

## I. INTRODUCTION (*HEADING 1*)

The computer science job market has undergone significant changes over the past decade, with shifts in salaries, job availability, and employment conditions. This report examines the changing trends in the software engineering field from 2012 to 2023, focusing on several key questions. These include how software engineer salaries have evolved, how they compare to the earnings of recent graduates, and how geographic and employment factors influence salary trends. The topic is particularly important as it provides insights into the economic factors shaping the technology industry, highlighting disparities in opportunities and compensation for software developers across different regions and stages of their careers.

Understanding salary trends and employment dynamics is vital for both industry professionals and new graduates. As software development continues to be one of the most in-demand fields globally, knowing the trends can guide career decisions, especially in a time when remote work and the cost of living vary dramatically by location. Current research on this subject points to an increasing demand for software engineers, but salary growth has been uneven, often reflecting factors like geographic location, experience, and educational background [5].

This analysis uses datasets from the U.S. Bureau of Labor Statistics and other reputable sources such as Kaggle and Zenodo to investigate these trends. The findings will provide a clearer understanding of how the software engineering job market has evolved and help both new and experienced developers navigate the landscape.

## II. DATASETS

### A. Source of dataset (*Heading 2*)

This study utilizes several datasets from credible sources that provide insights into the software engineering

job market, salaries, and employment trends. The following datasets are used:

#### 1. *Software\_Engineer\_Salaries\_Employment\_2012-2023.csv*:

**Source:** This dataset was created by me using data from the U.S. Bureau of Labor Statistics (BLS)[5], which is a reliable and authoritative source for labor data. The BLS provides comprehensive employment and salary data across various sectors, including software engineering.

**Date Generated:** The dataset spans from 2012 to 2023, incorporating annual data points for the analysis of salary trends in the software engineering field.

**Dataset Creation:** The data was extracted directly from the U.S. Bureau of Labor Statistics, which compiles national employment statistics for various industries and roles.

#### 2. *SoftwareDeveloperIncomeExpensesperUSACity.csv*:

**Source:** This dataset was obtained from Kaggle [3], with further contributions from Zenodo [1]. It offers valuable insights into software developer salaries adjusted for cost of living across different cities in the U.S.

**Date Generated:** The data was compiled in 2022, reflecting the latest trends available at the time.

**Dataset Creation:** The dataset includes salary data adjusted and unadjusted for cost of living, along with city-level economic parameters such as cost of living, median home prices, and local purchasing power.

#### 3. *All-ages.csv*:

**Source:** The data for this dataset was also sourced from Kaggle [3] and Data.World [4]. This dataset contains information on the employment status, salary, and educational background of graduates across various majors.

**Date Generated:** The dataset's most recent version is based on data from 2014.

**Dataset Creation:** The dataset includes information about major codes, employment rates, median salaries, and full-time employment status for graduates across various academic backgrounds.

#### 4. *Grad-students.csv*:

**Source:** Like the All-ages dataset, this data is sourced from Kaggle [3] and Data.World [4].

**Date Generated:** The dataset is also based on data from 2014.

**Dataset Creation:** It includes detailed data on the number of graduates, unemployment rates, median salaries, and the share of graduates employed in each field.

### B. Character of the datasets

The datasets are primarily in CSV format and cover a range of parameters related to salaries, employment, and geographic conditions. Below is a summary of the key characteristics of each dataset.

Dataset	Size	Columns	Key Parameters
Software_Engineer_Salaries_Employment_2012-2023.csv	12 rows	(Annual data) Year, Mean Salary, Employed Individuals	Year, Mean Salary, Employment Data
SoftwareDeveloperIncomeExpensesperUSACity.csv	78 rows	60+ rows (City-level data)	City, Adjusted Salary, Unadjusted Salary, Cost of Living, Job Count
All-ages.csv	173 rows	Major_code, Major, Students, Employed, Unemployed, Median Salary	Major, Employment Data, Unemployment Rate, Median Salary
Grad-students.csv	174 rows	Major_code, Major, Grad-Total, Employed, Unemployed, Grad_Premium	Graduate Employment, Salary, Unemployment Rate

### Data Cleaning and Conversion:

- **Software\_Engineer\_Salaries\_Employment\_2012-2023:** The dataset required no significant cleaning but was organized to ensure consistency in the format for analysis.

### Combination of Datasets:

- Two datasets were combined for comparative analysis of salary data. For reference, the all-ages.csv dataset was merged with the grad-students.csv dataset to compare employment outcomes across different stages of a career.

## III. METHODOLOGY

This project employs regression analysis, comparative data analysis, and visualization to answer the research questions. Each method was chosen to uncover specific patterns, correlations, or trends in the datasets.

### A. Regression Analysis

Regression models were used to understand relationships between employment status and median salaries. Specifically, a linear regression model was applied to investigate correlations such as median salary versus total employed individuals across workers, graduates, and non-graduates. The assumptions for linear regression are that the relationship between the variables is straight-line and that the data points are spread evenly without patterns.

#### Advantages:

- Easy to use and interpret.

- Shows clear relationships between variables.

#### Disadvantages:

- Assumes the relationships are simple and linear, which might not capture more complex trends.
- Datasets dating back to 2014, so may not show an accurate representation of current trends.

To implement regression, the Python library Scikit-learn was used. Key tools included LinearRegression for the model and matplotlib for visualizations. I cleaned the data by removing unnecessary columns and adjusting data as needed to improve accuracy.

### B. Comparative Analysis

I compared datasets like grad-students.csv [2] and SoftwareDeveloperIncomeExpensesperUSACity.csv [3] to explore:

- Differences in salaries between new graduates and experienced developers.
- Geographic differences in salaries, both adjusted for cost of living and unadjusted.

#### Advantages:

- Makes it easy to compare important factors side by side.
- Highlights differences in the data.

#### Disadvantages:

- Sometimes trends might seem too simple, especially with limited data.

Since there wasn't much variety in the data for comparison, I focused on key metrics and used pandas functions to combine datasets. I then visualized the results using matplotlib.

### Visualization

To make the data easier to understand, I used graphs like bar charts, line graphs, and scatter plots to show:

- Salary trends from 2012 to 2023 (Line chart).
- Relationships between salaries and job availability by city (Bar Graph).
- Comparison between Software Developer salaries versus graduate median salaries (Box Plot).
- Correlation between computer science graduates, non-graduates, and general workforce (Linear Regression).

### Extra Work and Adjustments

- **Merging Datasets:** I combined the grad-students.csv and SoftwareDeveloperIncomeExpensesperUSACity.csv files to analyze the data together.
- **Filtering Data:** I filtered out unnecessary data to focus on the most relevant information for analysis.

These steps helped streamline the data and focus on the key factors for generating insights into the computer science job market.

## IV. RESULTS

### A. Software Engineer Salary Trends (2012-2023)

The line chart shows a consistent increase in mean software engineer salaries from 2012 to 2023, with minor fluctuations. This upward trend suggests steady growth in the

demand for software engineers, likely driven by technological advancements and industry needs.

- **Pattern:** Salaries steadily rise, with occasional fluctuations, indicating stability despite external factors.
- **Prediction:** Based on the consistent growth observed, it's likely that salaries will continue to increase, though some fluctuations may occur due to market or economic changes.
- **Interpretation:** The data reflects growing demand for software engineers, and while future salary increases seem probable, external factors could cause short-term deviations.

### B. Comparison of Software Developer Salaries vs Graduate Median Salaries

The box plot compares the adjusted and unadjusted software developer salaries with the median salary for software engineering graduates from 2014, which is approximately \$92,667.

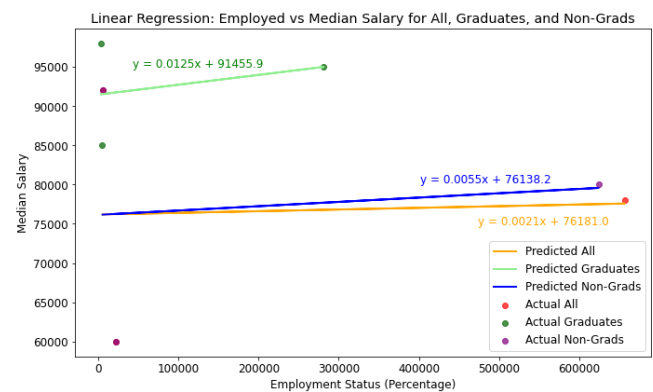
- **Salary Distribution:** Both adjusted and unadjusted software developer salaries show significant variation, with the median graduate salary falling just below the middle of both boxes. This suggests that software developer salaries are generally higher than the median graduate salary, but the range of salaries is wide.
- **Disparities:** Software developers, in both adjusted and unadjusted salaries, earn more than the median graduate salary, with adjusted salaries reflecting regional variations. However, given that the graduate data is from 2014 and the software developer data is from 2022, the comparison may not accurately reflect the current job market or that of 2014.
- **Consideration:** If we were comparing 2012 software developer salaries with 2012 graduate salaries, it's likely that graduates from 2012 would have had a relatively higher salary compared to software developers at that time. The current data suggests that, over time, developers' salaries increased more significantly compared to the starting salaries of graduates in the field.

### C. Comparison of Adjusted vs Unadjusted Software Developer Salaries Across Cities

The bar plot compares adjusted and unadjusted software developer salaries across cities, showing significant variation based on location.

- **Salary Distribution:** Cities like Columbus, OH (\$117,552), Seattle, WA (\$117,323), and Charlotte, NC (\$114,122) has higher adjusted salaries, reflecting strong demand and competitive markets. In contrast, cities such as Honolulu, HI (\$72,811), Eugene, OR (\$85,912), and Miami, FL (\$88,912) have lower salaries, likely due to weaker tech presence or higher costs of living.
- **Disparities:** Adjusted salaries reflect regional economic conditions. High salaries in top cities suggest stronger tech markets, while lower salaries in others may indicate smaller markets or cost-of-living adjustments.
- **Consideration:** Comparing 2022 data, salary trends show growth in tech hubs, though past data (e.g., from 2012) would likely show different regional dynamics based on the tech industry's evolution.

### D. Employment Status and Median Salaries for Computer Science Graduates, Non-Graduates, and the General Workforce



**Figure 1: Linear Regression Graph**

This analysis uses linear regression models to explore the relationship between employment status and median salaries across different groups: all workers, graduates, and non-graduates. The data shows the impact of employment rates (for all workers, graduates, and non-graduates) on median salaries within the computer science field.

**Model Summary:** The linear regression models were trained for each group, with employment status as the independent variable and median salary as the dependent variable. The equations for the three groups are as follows:

- Employed All vs Median Salary:

$$\text{Median Salary} = 76180.97 + 0.0021 \times \text{Employed}$$

- Intercept: 76180.97, Slope: 0.0021

This suggests that for each percentage increase in employment, the median salary for all workers increases by approximately \$0.0021. The relationship is quite weak, as shown by a very low R-squared value of 0.0024, indicating that employment status has little to no effect on median salary for the general workforce.

- Employed Graduates vs Median Salary:

$$\text{Grad Median Salary} = 91455.93 + 0.0125 \times \text{Grad Employed}$$

- Intercept: 91455.93, Slope: 0.0125

This shows that for each percentage increase in graduate employment, the median salary for graduates increases by \$0.0125. This relationship is stronger than for the general workforce, with an R-squared value of 0.0867, suggesting that employment status has a moderate effect on graduate median salary.

- Employed Non-Grads vs Median Salary:

$$\text{Non-Grad Median Salary} = 76138.22 + 0.0055 \times \text{Non-Grad Employed}$$

- Intercept: 76138.22, Slope: 0.0055

This model indicates that for each percentage increase in non-graduate employment, the median salary for non-graduates increases by \$0.0055. Similar to the general workforce model, the R-squared value of 0.0144 suggests a weak relationship between employment status and median salary for non-graduates.

Evaluation of Results:

- The R-squared values for the models show that the employment status of all workers and non-graduates has a very weak influence on median salaries, as indicated by R-squared values close to zero. This suggests that factors other than employment status (such as education, experience, or industry) may be more significant in determining salaries for these groups.
- The graduates' model shows a higher R-squared value (0.0867), indicating that employment status plays a slightly more important role in determining graduate salaries, though it still accounts for a small portion of the variance in salary.
- Outdated Data: The datasets used in this analysis are from 2012, which may not reflect current salary trends or employment dynamics. Economic factors, technological advancements, and industry changes could have significantly impacted the job market since then. To improve results, using more up-to-date datasets would provide a more accurate representation of current trends and ensure the analysis reflects the ongoing evolution of the job market.
- Limited Data Variety: The datasets used in this analysis are restricted in terms of geographic regions and industry sectors, which may not fully capture the variations in salaries across different locations or specific job roles within computer science. To improve results, expanding the datasets to include more diverse regions and industry sectors would provide a clearer picture of how salaries differ across various areas and job roles in the field.

#### Unexpected Results:

- The weak relationships found in the models for all workers and non-graduates could be attributed to the fact that employment status alone is not a strong predictor of salary. Many factors, such as years of experience, skills, industry, and geographic location, likely have a more significant impact on salaries.
- The relatively stronger relationship for graduates may suggest that having a degree could make employment status more influential, but again, other factors like specific skillsets and professional networking may also play a crucial role in earning potential.

#### Data Merging and Comparison Context:

- It's important to note that the datasets used in this analysis come from 2012 for both employment status and salary data for all three groups: all workers, graduates, and non-graduates. By using consistent datasets from the same time period, the analysis allows for a more accurate comparison across groups, as external economic factors (like inflation, changes in the job market, or policy shifts) from 2012 are similarly reflected across all groups.
- The merging of these datasets ensured that we could directly compare the employment status and salary outcomes for different groups while keeping the analysis context consistent, allowing for a clearer understanding of how employment status correlated with salaries at that particular point in time.

## V. DISCUSSION

This project employed several analytical methods, including regression analysis, comparative data analysis, and visualization, to investigate salary trends in the computer science job market. While the methods provided valuable insights, there are limitations and opportunities for improvement.

#### Limitations and Future Works:

- Linear Regression Assumptions: The regression analysis assumes a simple linear relationship between variables, which may not fully capture the complexities of salary influences. For instance, employment status alone has a weak correlation with median salaries, indicating that factors like experience, skills, and geography likely play a more significant role. Future work could explore options like incorporating non-linear regression models or machine learning techniques like decision trees. This could yield more accurate results.

## VI. CONCLUSION

In conclusion, this project provides valuable insights into the evolving landscape of the computer science job market, focusing specifically on software engineering salaries and employment dynamics. Through the analysis of datasets spanning from 2012 to 2023, key trends emerged, including the steady growth of software engineer salaries, regional disparities, and the varying income levels for recent graduates versus experienced developers. Geographic factors and cost of living also played a significant role in salary variations, with tech hubs offering higher salaries compared to regions with less demand for software professionals.

The regression and comparative analysis methods used in this project effectively highlighted the impact of employment status on median salaries, emphasizing the role of degrees in salary progression. Additionally, the comparative analysis of graduate salaries versus software developer incomes showed that, on average, recent graduates with degrees in computer science tend to earn higher salaries than established software developers in the field, although regional and industry-specific variations may exist.

Overall, this study contributes to the broader understanding of the computer science job market, offering insights that are not only useful for job seekers but also for professionals and organizations planning for future hiring and compensation strategies in the tech industry. The trends observed suggest that software engineering will continue to be a highly rewarding career, though the dynamics between experience, location, and employment status will remain crucial in shaping salary outcomes.

## REFERENCES

- [1] Bruno Bonfrisco, Franco Seveso. "Software Developer Income/Expenses per USA City." Zenodo, Zenodo, 8 Dec. 2022, [zenodo.org/records/7412091#.Y9Y3ZNJBwUE](https://zenodo.org/records/7412091#.Y9Y3ZNJBwUE).
- [2] "College Majors and Their Graduates." Kaggle, [www.kaggle.com/datasets/thedevastator/uncovering-insights-to-](https://www.kaggle.com/datasets/thedevastator/uncovering-insights-to-)

college-majors-and-their?select=recent-grads.csv. Accessed 7 Dec. 2024. K. Elissa, "Title of paper if known," unpublished.

- [3] Devastator, The. "U.S. Software Developer Salaries." Kaggle, 11 Feb. 2023, [www.kaggle.com/datasets/thedevastator/u-s-software-developer-salaries/data](https://www.kaggle.com/datasets/thedevastator/u-s-software-developer-salaries/data).
- [4] "FiveThirtyEight's Datasets." Data.World, data.world/fivethirtyeight. Accessed 7 Dec. 2024.
- [5] U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, 5 Dec. 2024, [www.bls.gov/](https://www.bls.gov/).