

Insights into the software development Industry: What factors affect the yearly compensation in the software development industry?

STUDENT NAME
CLASS
PROF NAME

Abstract

It would not be wrong to say that the software industry has grown significantly. It is said that around 26.9 million people worldwide are software developers (Evans Data Corporation: 2021). So why are so many people joining this industry? Is it because it is highly paid, or are most jobs lucrative? Which specific programming languages are in demand, or does it depend on the type of employment status? In this work, these are our research questions, and we will evaluate each of them to see more insights into how they contribute to the trend in yearly compensations.

Firstly, we made two subsets, one for the variables of interest and the second one specifically to take out professional developers. Then, after data wrangling, we dive into exploratory data analysis using different plots for better understanding. Our results show that these factors positively impact the yearly compensation in the software development industry.

Introduction:

The software development industry is a growing industry. Most college students who wish to end up in the software industry usually opt for STEM majors, including Computer Science, Computer/Software Engineering, Information Technology, etc. It is not merely just the degree that matters but also there has been a positive trend in the usage of programming languages over the years. It is an essential factor for professionals in the field or those considering entering it. There is a considerable variation in the yearly compensations among employees and one of the significant factors contributing. This was an exciting area for our group to explore; hence we decided to work on it. Our research was inspired by a Stack Overflow survey conducted in 2021. The survey had more than 83439 respondents. By analyzing

the survey data, we were able to successfully create a detailed picture of how education and programming languages relate to salary by formulating specific research questions. We decided upon one general and four specific research questions: Our general question was what is the average salary in the software development industry, and does it depend on years of experience, number of technical skills, education level, or type of software development? Our more specific questions were the type of degree needed, the most common programming language, types of software development jobs that are more lucrative, and lastly, which languages are the highest paid.

Statistical Analysis (with Data Visualization)

Firstly, we wanted to see which salary group our respondents are present in. To do that, we made a histogram (Appendix: Figure1) to see the yearly salary distribution, which clearly showed that we could see a decline in the number of respondents with increasing salaries. However, we did notice some outliers, including a 30million per year income, so to avoid skewness, we created a limit for a yearly salary below \$100,000.

After that, we specifically wanted to see how compensations vary on coding experience, respondents' number, and employment status (full-time, part-time, and freelance).

We made a scatter plot (Figure 2) where we compared yearly compensations with

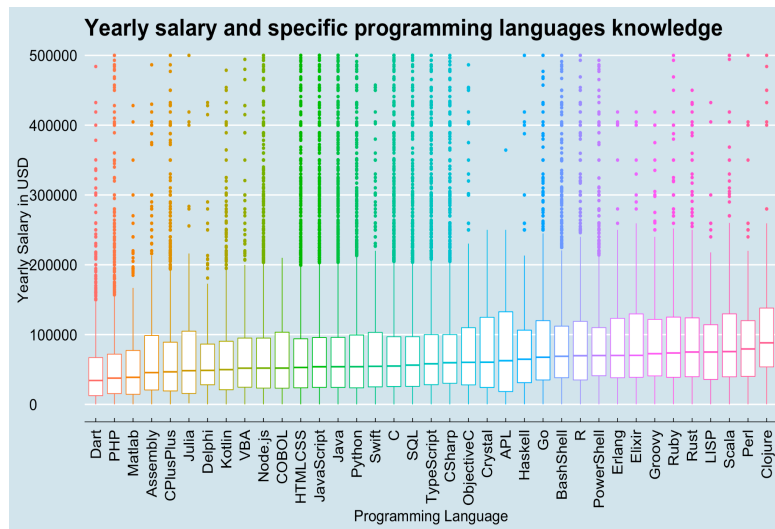


Figure 2. Compensation with employment statuses and coding experience

years of coding and employment statuses.

Then we wanted to see the relationship between education and yearly salary. We decided to make a boxplot (Appendix: Figure 3) for each education level with yearly compensation on the y-axis. Most respondents had bachelor's and master's degrees, but doctoral degrees had the highest median income.

We had a long list of languages that our respondents used and decided to draw a boxplot (Figure



4). It was clear from the plot that Clojure had the highest median salary approaching \$100k/year, followed by Perl and Scala. The most popular languages ranked in the middle of the scale at around \$50k/year.

Figure 4. Programming languages vs yearly compensation

Our last research question was to find which roles in the industry are most lucrative. Using boxplot, we used variable

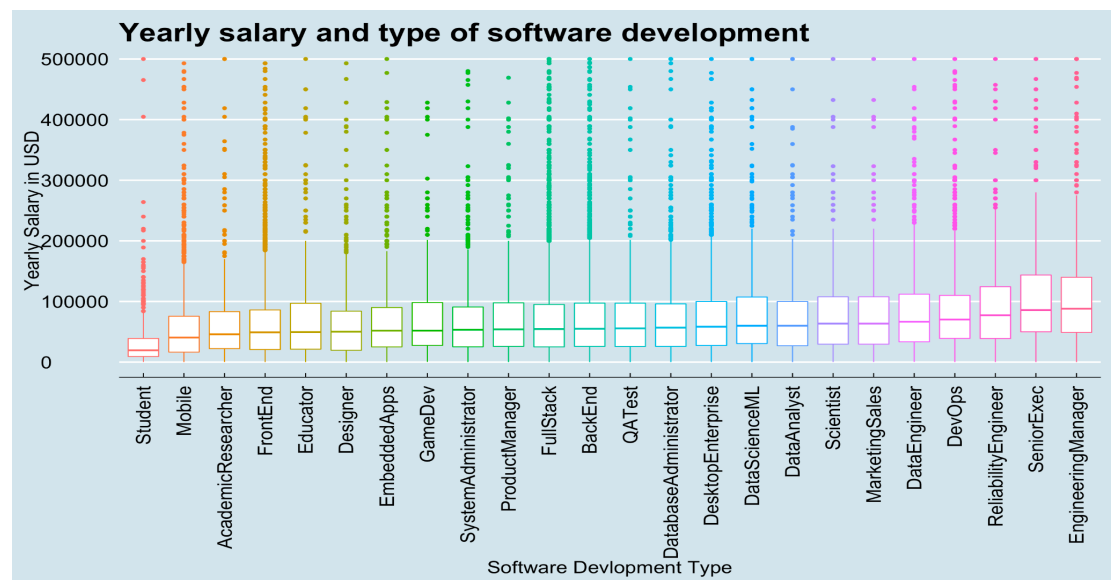


Figure 5(above). Boxplot for Yearly Salary and Software Development type

software development types to see how compensation changes with positions (Figure 5 above).

We found that senior positions, including Senior Executive and Engineering Manager, are more lucrative, followed by the data roles.

After doing all this work, we wanted to look back into our general question. From these steps, we know that, to some extent, these research factors make a difference in the yearly compensation, but does it also depend on the representation of a geographical region. For example, some countries in our dataset might have more professional developers than others. We wanted to add this variable and other variables to our regression model to see this. First regression we wanted to do with the subset data for US respondents only and one with top five countries.

The first regression with US respondents only:

$$\begin{aligned} \text{CoveredCompYearly} = & 78478.6 + 3540.9(\text{YearsCodePro}) - 40456.8(\text{EdLevelSecondary school}) - 37005.8(\text{EdLevelSome} \\ & \text{university}) - 29112.3 (\text{EdLevelAssociate}) - 31357.6 (\text{EdLevelBachelor}) - 27552.2 (\text{EdLevelMaster}) - 7720.3 (\text{EdLevelDoctoral}) - \\ & 40204.4 (\text{EdLevelProfessional degree}) - 53270.2 (\text{EdLevelOther}) - 44481.6 (\text{EmploymentPart-time}) - \\ & 8693.1(\text{EmploymentFreelance}) - 1805.8 (\text{LanguageNo}) \end{aligned}$$

R-squared= 0.1716

The second regression with top 5 countries:

$$\begin{aligned} \text{CoveredCompYearly} = & 37746.8 + 2226.92 (\text{YearsCodePro}) - 31711.57 (\text{EdLevelSecondary school}) - 36324.58 (\text{EdLevelSome} \\ & \text{university}) - 44295.08 (\text{EdLevelAssociate}) - 26015.66 (\text{EdLevelBachelor}) - 18196.43 (\text{EdLevelMaster}) - 8685.51 \\ & (\text{EdLevelDoctoral}) - 23883.97 (\text{EdLevelProfessional degree}) - 32928.66 (\text{EdLevelOther}) - 27581.26 (\text{EmploymentPart-time}) + \\ & 6912.84 (\text{EmploymentFreelance}) + 404.92 (\text{LanguageNo}) + 41062.85(\text{CountryGermany}) - 732.97 (\text{CountryIndia}) - \\ & 48106.98(\text{CountryUnited Kingdom of Great Britain and Northern Ireland}) - 96225.71 (\text{CountryUnited States of America}) \end{aligned}$$

R-squared= 0.4638

We noticed that by adding country as a predictor, our model improved, which could be seen from the increase in our R-squared value, representing the percentage of variations explained by the model. For example, from 17.16% to 46.38% was a significant change.

Conclusion and Discussion

After all the exploratory data analysis, we can confidently say that we can get a better model by adding individual countries. This is because the proportion of respondents representing each country varies a lot in this dataset. We also conclude a few points, including the average salary in the software development industry is \$75,000 with a median of \$54,000. In terms of degree, we found out that 50% of our survey respondents have bachelor's degrees, and the higher we get educated, the higher our yearly compensation. Because of the popularity of programming languages, we found out that the most common programming languages are C, Java, JavaScript, HTML/CSS, and SQL. Also, we found out that there is a variation in median salaries. Some languages have higher median salaries while some than others. For example, Clojure, Perl, and Scala have much higher median salaries than other languages in our survey. In terms of jobs, we observed that the senior roles are usually the most lucrative, typically requiring 14 years of experience. In our survey, these roles included Engineering Manager and Senior Executive.

REFERENCES

Data resources:

Stack overflow developer survey 2021. (n.d.). Stack Overflow. <https://insights.stackoverflow.com/survey/2021>

Article Resource

Software developer statistics 2021: How many software engineers are in the US and in the world? [Updated]. (2020, February 9). Daxx Software Development Teams. <https://www.daxx.com/blog/development-trends/number-software-developers-world>

APPENDIX

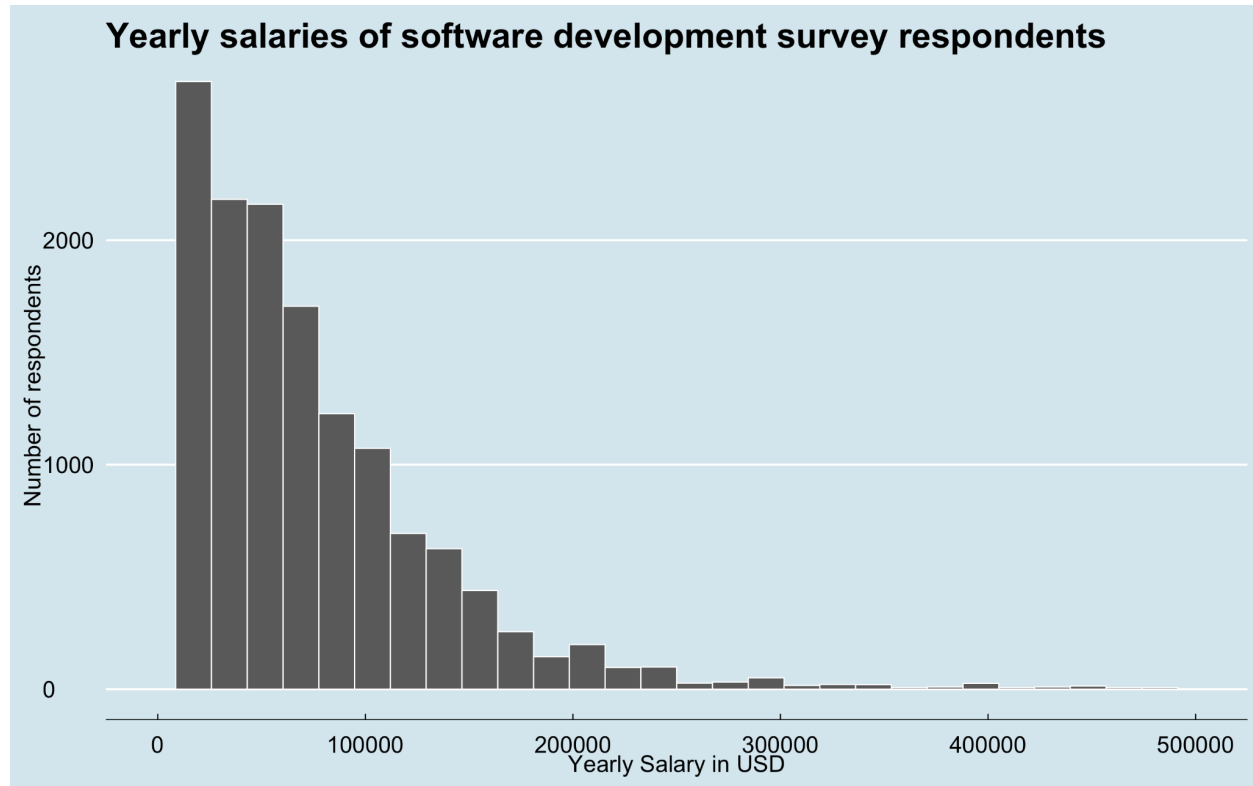


Figure 1. Histogram of yearly salary distribution. We can notice that with increasing salaries we see a decline in number of respondents.

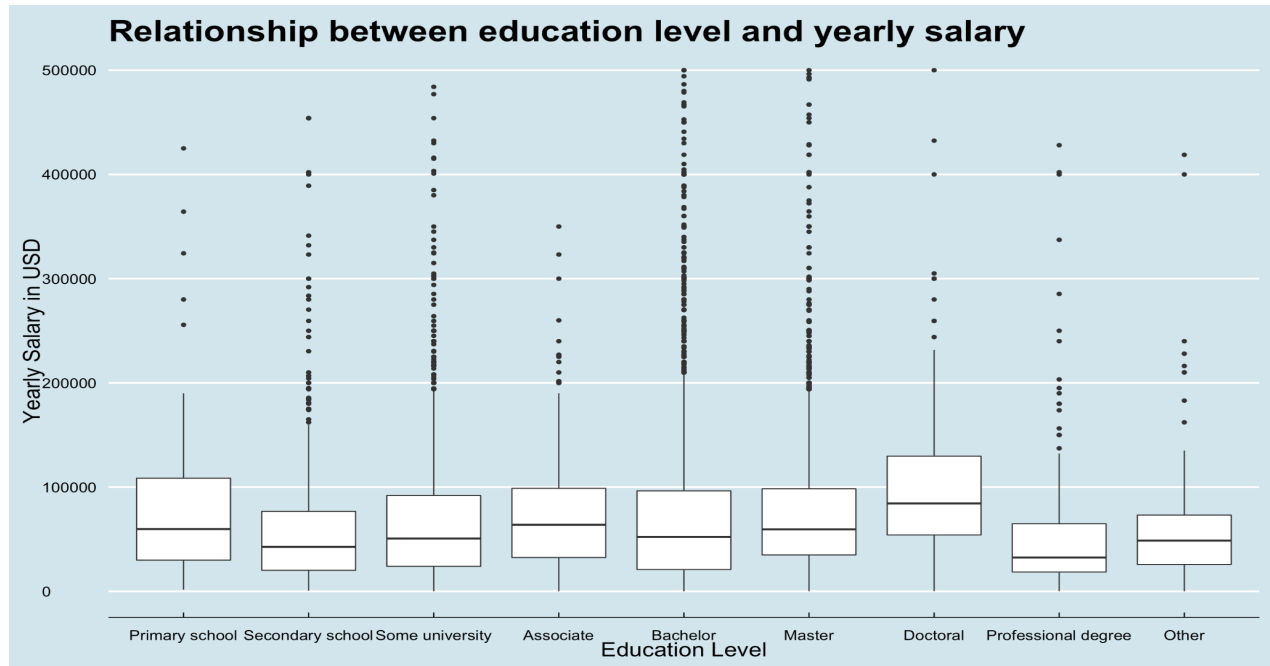


Figure 3. Boxplot for Education level and yearly salary