An Analysis of Salary of People and its correlation to Age They First Started to Code

*B\_Group 30*

*Rajat Shrestha, Pranav Khatri, Sanjog Chaudhary,*

*Darpan Pokhrel, Aashish Upadhyaya*

Friday, 14 Jan 2022

**Abstract:**

Context: Tech is the highest paid profession all over the world. Every day younger developers get into the market who are more exposed to the digital elements, this research is to find if starting to code earlier will affect their earning potential.

Question: This study asks, if there is any co-relation between the age the individual first started to code and their total yearly income?

Method: We analyzed the reported total annual compensation of people (in USD) of thousands of respondents of stack overflow developer survey 2020.

Results: We expected to see the developers who started to code early to be paid higher than the other developers who started to code late in their life and our analysis supports this claim.

Conclusion: The results suggest that there is a solid correlation between the age the individual first started to code and their total annual compensation since it effects the experience level of the individual.

# Introduction

The cash rewards paid to the employees in exchange of their services is called compensation (Cambridge University Press, 1995). It is a common to measure compensation on basis of annual total income (MaRS, 2022).

Coding is an act of writing computer instructions (Cambridge University Press, 1995).

This study aims to understand if there is any relation to the age they started to code for their compensation. Specifically, we ask the following research question: Is a correlation between total compensation and age first started to code?

The null hypothesis is:

**H0**: There is no correlation between total compensation and age first started to code.

The alternative hypothesis is:

**Halt**: There is correlation between total compensation and age first started to code.

The data set we used is “Stack Overflow Developer Survey 2020”. It is the largest public survey participated by developers all around the world (Stack Overflow, 2022). In 2020 65000 participants took part on this survey and we choose this year’s data as there was a column which gave the data about the annual total compensation of each participant in USD and a column which indicated at what age they first started to code.

The URL of the dataset is: “<https://data.world/technology/stack-overflow-developer-survey/workspace/file?filename=2020-survey_results_public.csv>”

Results: We plotted the points of each surveyor according to their age first started to code and their total compensation and found that the trend line was slightly slanted towards higher compensation of the individuals claiming to start to code early on their life

In the rest of this paper, we present a visual view of the data, followed by a statistical analysis, and end the report with a discussion of the implications of the results.

A picture containing qr code

Description automatically generated

Figure 1: The official stack overflow Developer survey overview (Stack Overflow, 2022).

Visualization

The dataset used for this study comprises of 64,461 rows (entry of each respondent) and 62 columns (different data points of each respondent). The age of the developers since they started to code is limited from 0 to 90. The maximum compensation threshold is set to 200000 USD.

Scatter chart

Description automatically generated

Figure 2: Age First Started to code vs Total compensation graph with trend

Figure 2 plots all the entries by everyone as a single point in a graph with horizontal distance from the left suggesting the age the individual first started to code and the vertical distance from the bottom depicting their total compensation. The Red line highlights the Trend to describe more clearly what is happening in the graph. It appears that the trend line is higher at the left and gradually declines at the right suggesting that the people who started to code early has relatively more total compensation from those who started coding later in their life.

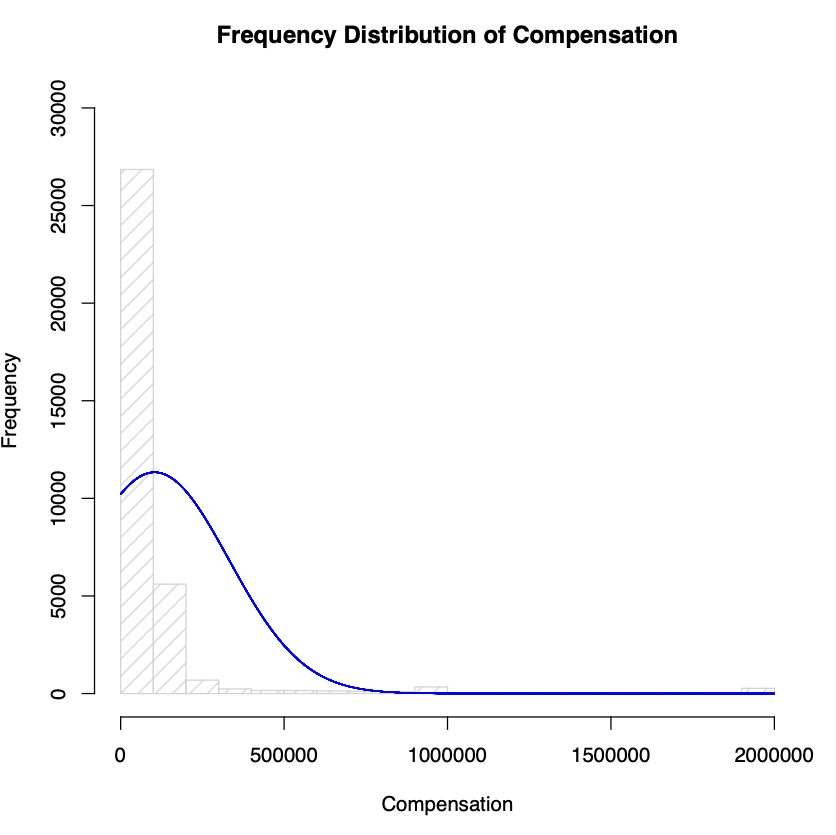


Figure 3: Frequency Distribution of the Total compensation

Figure 3 depicts a histogram showing the frequency distribution of total compensation earned by all the Respondents. This histogram shows that the total compensation is skewed toward the lower end, this is unusual as normal distribution would follow a rather distinct “bell curve” (Herrnstein & Murray, 1994).

# Analysis

*This section should describe:*

1. *The test statistic used to answer the question, and why it is appropriate for the data. You should say how your chosen test statistic is suitable for your data’s apparent distribution.*
2. *The value of the test statistic.*
3. *The p-value reported by R, and whether this value suggests the null hypothesis can be rejected in favor of the alternative hypothesis.*

# Conclusions

*What do the results of your analysis mean?*

*If you reject the null hypothesis, what does this tell you? Is the relationship causal or coincidental?*

*If you do not reject the null hypothesis, does this mean the answer to your question is “no,” or does it suggest that you don’t have enough data?*

# References

Cambridge University Press, 1995. *Cambridge international dictionary of English.* Cambridge: Cambridge University Press.

Dalgraad, P., 2008. *Introductory Statistics with R.* 2nd ed. s.l.:Springer.

MaRS, 2022. *Employee compensation: Salary, wages, incentives and commissions.* [Online]   
Available at: https://learn.marsdd.com/article/employee-compensation-salary-wages-incentives-and-commissions/

Stack Overflow, 2022. *Stack Overflow Annual Developer Survey.* [Online]   
Available at: https://insights.stackoverflow.com/survey/