CS Salary

Harrison Grogan

5/1/2021

ABSTRACT

My project is about Computer Science Salaries and what the difference is into how much you are paid. This is important because it is never a bad idea to educate yourself on how certain aspects of your line of work are decided. I saw someone answer someone else's question about whether age has anything to do with people losing their jobs in the Computer Science field. The person went on to explain that a lot of Computer Science workers lose their job around 50, not due to ageism, but due to people just slacking off and not keeping up with their programming skills. My research question is: Is there any link to how someone is paid when it comes to their age? I got my data from a website called Data USA. I analyzed my data through R and I found that age has little to do with whether or not someone loses their job, but as the data showed, as people aged, they got more experience and skilled, allowing them to get a pay increase. So, what I saw is that age has something to do with salary, just not job loss. Of course, a person who is older may lose their job, but it's probably from slacking off or lack of "keeping up with tech".

Introduction

My name is Harrison Grogan and I did my R project over how age effects salary in the Computer Science field of work.

Literature Review

This needs to be addressed because you'll want to know how to keep yourself in good standing and to be able to know how to keep up. As someone named Thomas B Walsh described to someone on Quora; "A lot of people lose their jobs around age 50, but it wasn't about ageism. The work is very demanding. Some people tend to begin to feel entitled and slack off." (Walsh B Thomas, Quora) He said that he personally burned out when he was 56. I read an article by Howard Williams, the article talked about how most Baby Boomers don't apply out of fear of being too old. Howard also mentioned how 61 percent of developers over 45 are concerned that their age is limiting their career options. Most developers are between the ages of 22-29. Basically, a developer reaches their mid-40s, they're likely to face career worries. Howard goes on to talk about how the tech industry favors younger developers. One of the biggest reasonings is because older developers, while having more experience, tend to cost more then the younger ones. Personally, I agree with Thomas, the reason someone is fired probably won't be because of their age, but, as Thomas said, because of a feeling of entitlement and/or slacking off.

Data

The application that I used to download the data is called ParseHub. ParseHub is a really good application when it comes to data mining. It's easy to use and very fast. The way to use ParseHub is by putting the link to the page/website you want to data mine from and then you select the data, but you don't select the data one-by-one, you select two pieces of data and it auto fills the rest for you, then when you're ready you can download it into a csv file and transfer it to R.

Methodology

After I read my csv file into R, I then proceed to check the contents of the file and noticed that there were a few columns there that I did not need, so I began by taking the columns out using the subset function with -c(. . .). Then I proceeded to analyze the data using a scatterplot graph. I used the lm() function to create a linear model and look over the data there. I then used the slope-intercept formula as a way to predict the salary using years of experience. Example: 605.4*12 + 61027 = 68291.8, with 12 being years of experience.

Results

As mentioned earlier in Abstact, the results that I found weren't too surprising from what I read in the Quora answer from Thomas B Walsh. The data showed that as years go by people would see a pay increase steadily. This does not include promotions, company, and the size/cost of the city/town that a person lives in. The scatterplot is nice to see a kind of visual representation of the data, but not as telling as the slope-intercept formula and linear model, which with the curve that the scatterplot shows it almost, in a way, works against the linear model. So, the take away is to focus more on the linear model than the scatterplot graph.

Implications

As mentioned in the Results section; the slope-intercept formula, obtained from the linear model, isn't very accurate because it is only count age and not promotions, company, and/or size/cost of the city or town a person lives in, but it gives people a good idea of what their salary may be years from now.

Conclusion

What I have achieved is showing that age does have a factor in the salary someone will obtain after working at a company for maybe, 5, 10, 15+ years. Also, keep your skills in programming up and knowledge of tech fresh.

References

Walsh B T. (2015). Is it true that computer science people only have good jobs till the age of 35-40? Quora. https://www.quora.com/Is-it-true-that-computer-science-people-only-have-good-jobs-till-the-age-of-35-40

N/A. (2019). Computer Science. Data USA. https://datausa.io/profile/cip/computer-science-110701#demographics

Williams H. (March 29, 2019). Ageism in tech: the not-so-invisible age limit developers face. https://bdtechtalks.com/2019/03/29/ageism-in-tech-age-limit-software-developers-face/https://bdtechtalks.com/2019/03/29/ageism-in-tech-age-limit-software-developers-face/

This helps us read in and clean our data

```r
salaryAge <- read_csv("C:/Users/groga/OneDrive/Desktop/Workforce_Age.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   ‘ID Age‘ = col_double(),
##   Age = col_double(),
##   ‘ID Year‘ = col_double(),
##   Year = col_double(),
##   ‘ID Workforce Status‘ = col_logical(),
##   ‘Workforce Status‘ = col_logical(),
##   ‘Total Population‘ = col_double(),
##   ‘Total Population MOE Appx‘ = col_double(),
##   ‘Average Wage‘ = col_double(),
##   ‘Average Wage Appx MOE‘ = col_double(),
##   ‘Record Count‘ = col_double(),
##   CIP2 = col_character(),
##   ‘ID CIP2‘ = col_double(),
##   share = col_double()
## )
```

```
salaryAge
```

```
## # A tibble: 324 x 14
##    ‘ID Age‘   Age ‘ID Year‘  Year ‘ID Workforce Status‘ ‘Workforce Status‘
##       <dbl> <dbl>     <dbl> <dbl> <lgl>                 <lgl>
##  1       20    20      2019  2019 TRUE                  TRUE
##  2       21    21      2019  2019 TRUE                  TRUE
##  3       22    22      2019  2019 TRUE                  TRUE
##  4       23    23      2019  2019 TRUE                  TRUE
##  5       24    24      2019  2019 TRUE                  TRUE
##  6       25    25      2019  2019 TRUE                  TRUE
##  7       26    26      2019  2019 TRUE                  TRUE
##  8       27    27      2019  2019 TRUE                  TRUE
##  9       28    28      2019  2019 TRUE                  TRUE
## 10       29    29      2019  2019 TRUE                  TRUE
## # ... with 314 more rows, and 8 more variables: Total Population <dbl>,
## #   Total Population MOE Appx <dbl>, Average Wage <dbl>,
## #   Average Wage Appx MOE <dbl>, Record Count <dbl>, CIP2 <chr>, ID CIP2 <dbl>,
## #   share <dbl>
```

```
salaryAge_df = subset(salaryAge, select = -c(`ID Age`, `ID Year`, `Workforce Status`, `Record Count`, CI
salaryAge_df
```

```
## # A tibble: 324 x 3
##      Age ‘Total Population‘ ‘Average Wage‘
##    <dbl>              <dbl>          <dbl>
##  1    20               1718         58053.
##  2    21               5682         32310.
##  3    22              23899         31881.
##  4    23              45486         48923.
##  5    24              40960         61719.
##  6    25              64725         68733.
##  7    26              63287         70502.
##  8    27              60986         74096.
##  9    28              57810         76344.
## 10    29              67395         81090.
## # ... with 314 more rows
```

This gives us a summary of our data

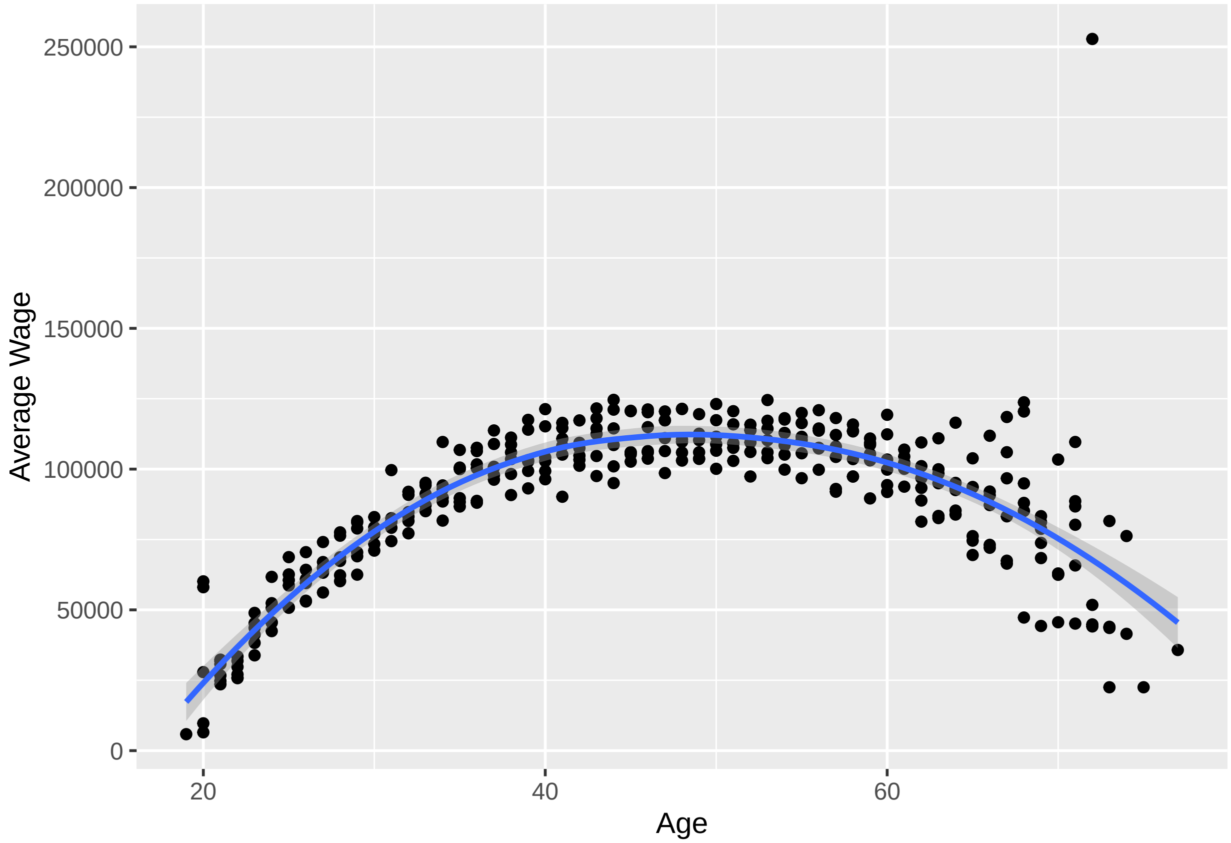
```
lapply(salaryAge_df, FUN=summary)
```

```
## $Age
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   19.00   33.00   46.50   46.53   60.00   77.00 
## 
## $`Total Population`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##     293   11758   42003   35346   52577   77218 
## 
## $`Average Wage`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    5844   74019   96793   89194  108644  252793
```

The scatterplot for the data with a smooth line for ease of read

```
ggplot(salaryAge_df, aes(x = Age, y = `Average Wage`)) + geom_point() + stat_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



##This here is the linear model that we will base my slope-intercept equation off of

```
lin <- lm(`Average Wage` ~ Age ,salaryAge_df)
summary(lin)
```

```
##
## Call:
## lm(formula = ‘Average Wage‘ ~ Age, data = salaryAge_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -83915 -14645   6079  17615 148180
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 61027.64    4592.63   13.29  < 2e-16 ***
## Age           605.36      93.56    6.47 3.64e-10 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 26350 on 322 degrees of freedom
## Multiple R-squared:  0.1151, Adjusted R-squared:  0.1123
## F-statistic: 41.87 on 1 and 322 DF,  p-value: 3.638e-10
```

This has 12 included into the equation as “years of experience”

```
predicted_salary = 605.4*12 + 61027
predicted_salary
```

```
## [1] 68291.8
```