SALARY TRENDS
&
INSIGHTS ANALYSIS

Using R

Team Members :
- 2022UCM2305 (Prabhmeet)
- 2022UCM2372 (Tisha)
- 2022UCM2374 (Adithyan)
- 2022UCM2684 (Chinmay)

INTRODUCTION

- For this project, we worked with a **salary dataset** consisting of around 6,700 entries.
- Throughout the analysis, we made use of R's powerful tools for:
  - data manipulation,
  - statistical analysis, and
  - visualization

  to produce clear, informative charts and graphs.

INTRODUCTION

The analysis is divided into four main sections:

1. **Understanding demographics of Dataset.**

2. **Exploring relationship between salary and key factors**
   l
   - Education
   - Years of Experience
   - Job Titles
   - Gender

3. **Effects of YoE on salary by gender, education level & job title.**

4. **Gender Specific trends in Education Level & Job Titles.**

ANALYSIS PROCESS OVERVIEW

- DATA READING & PREPROCESSING
- UNDERSTANDING THE DEMOGRAPHICS
- PLOTS for Salary vs A Single Factor
- SALARY vs YoE
- LOOKING INTO GENDERS

DATA READING & PREPROCESSING

- Taking a look at the data.
- We loaded the necessary libraries (readr, ggplot2)
- Read the CSV File
- Check by printing the head & summary of dataset.

Head of Dataset

A tibble: 6 × 6

Age	Gender	Education Level	Job Title	Years of Experience	Salary
<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>
32	Male	Bachelor's	Software Engineer	5	90000
28	Female	Master's	Data Analyst	3	65000
45	Male	PhD	Senior Manager	15	150000
36	Female	Bachelor's	Sales Associate	7	60000
52	Male	Master's	Director	20	200000
29	Male	Bachelor's	Marketing Analyst	2	55000

Summary of Dataset

```
      Age            Gender          Education Level     Job Title
 Min.   :21.00   Length:6704        Length:6704        Length:6704
 1st Qu.:28.00   Class :character   Class :character   Class :character
 Median :32.00   Mode  :character   Mode  :character   Mode  :character
 Mean   :33.62
 3rd Qu.:38.00
 Max.   :62.00
 NA's   :2
 Years of Experience     Salary
 Min.   : 0.000      Min.   :   350
 1st Qu.: 3.000      1st Qu.: 70000
 Median : 7.000      Median :115000
 Mean   : 8.095      Mean   :115327
 3rd Qu.:12.000      3rd Qu.:160000
 Max.   :34.000      Max.   :250000
 NA's   :3           NA's   :5
```

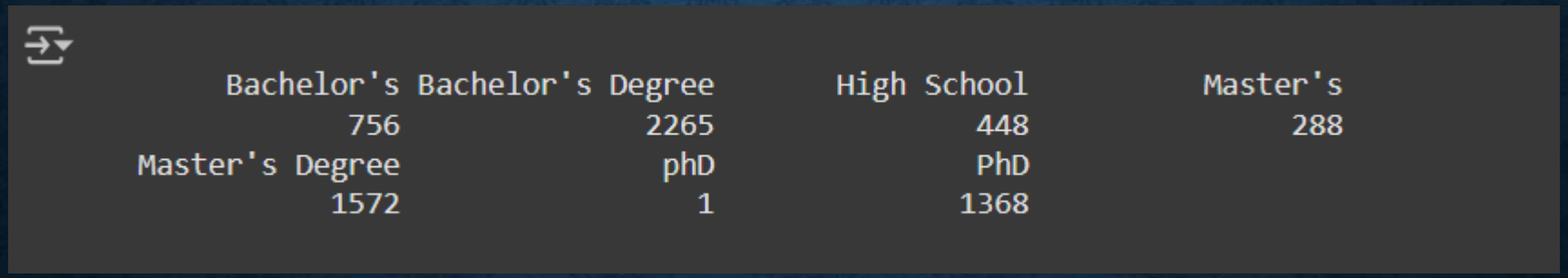
DATA CLEANING

- Out of 6704 entries, only 6 had NULL values.
- So, we removed those rows.

```
Missing Rows
# A tibble: 6 x 6
    Age Gender `Education Level` `Job Title`        `Years of Experience` Salary
  <dbl> <chr>  <chr>             <chr>                              <dbl>  <dbl>
1    NA NA     NA                NA                                    NA     NA
2    NA NA     NA                NA                                    NA     NA
3    27 Male   NA                Developer                              7 100000
4    31 Male   Master's Degree   Full Stack Engine…                     8     NA
5    26 Female Bachelor's Degree Social M                              NA     NA
6    36 Male   Bachelor's Degree Sales Director                         6     NA
```

INCONSISTENT VALUES

- Different values (e.g. PhD and phD) were used to refer to the same degree in the 'Education Level' column.
- We made them consistent.



Bachelor's	Bachelor's Degree	High School	Master's
756	2265	448	288
Master's Degree	phD	PhD	
1572	1	1368	

INCONSISTENT VALUES

- Final values, after making them consistent.

```
Bachelor High School      Master         PhD
    3021         448        1860        1369
```

DUPLICATE ENTRIES

- Identified and removed duplicate rows using duplicated() and distinct() functions.
- Grouped the data by key variables and calculated the frequency of each unique combination.
- Sorted and identified the top 10 most common combinations.

DUPLICATE ENTRIES

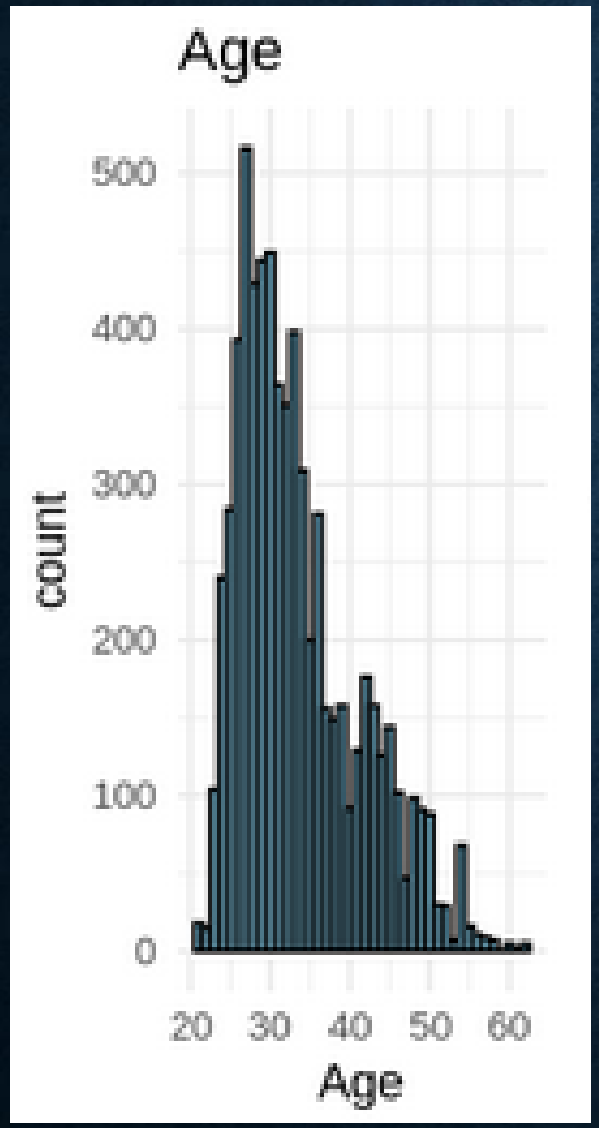
```
# A tibble: 10 × 7
     Age Gender `Education Level` `Job Title`  `Years of Experience` Salary  Freq
   <dbl> <chr>  <chr>             <chr>                        <dbl>  <dbl> <int>
 1    24 Female High School       Receptioni…                      0  25000    45
 2    27 Male   Bachelor          Software E…                      3  80000    45
 3    32 Male   Bachelor          Product Ma…                      7 120000    45
 4    32 Male   Bachelor          Software E…                      8 190000    39
 5    33 Female Master            Product Ma…                     11 198000    38
 6    27 Male   Bachelor          Software E…                      4 140000    37
 7    29 Female Master            Data Scien…                      6 180000    37
 8    25 Male   Bachelor          Product Ma…                      1  60000    36
 9    26 Male   Bachelor          Data Analy…                      3 130000    35
10    29 Male   Bachelor          Marketing …                      4  70000    33
```

It shows that there are duplicates, but since it is possible that 2 people have the same Age, education levels, gender, etc, we didn't remove them.

WHO PARTICIPATED IN SURVEY ?

- OBJECTIVE: To understand the demographic and participant characteristics of Survey Participants.

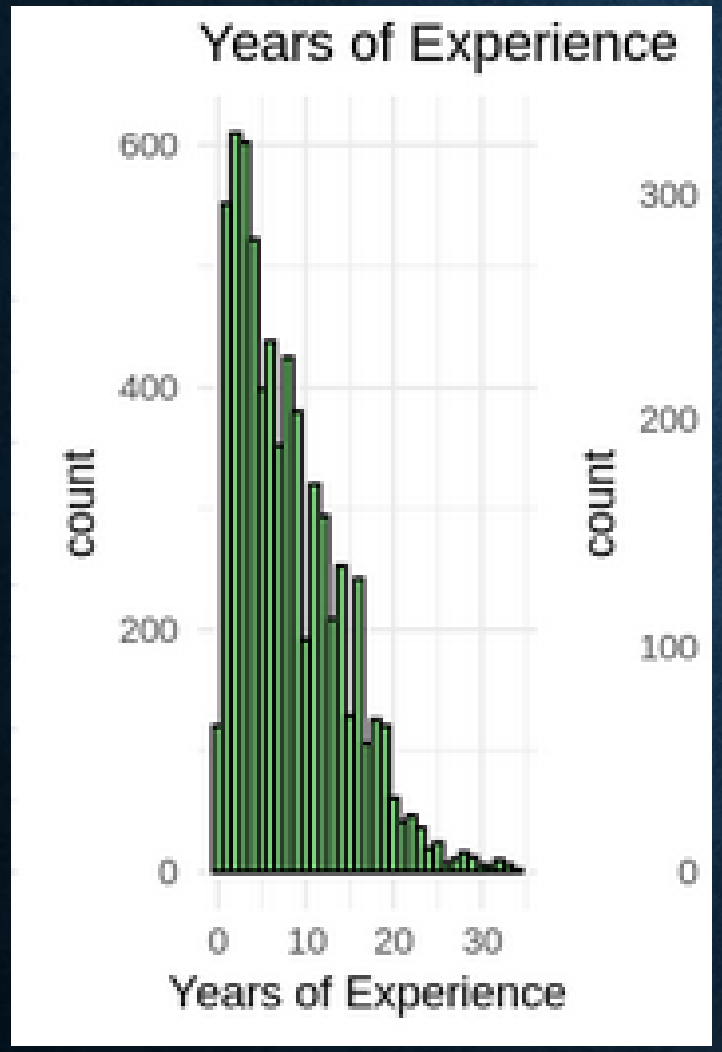
- Ensured **Age** and **Years of Experience** columns had Integers only.
- Created **histograms** for **Age**, **Years of Experience**, and **Salary.**
- Created **bar plots** for **Gender** and **Education Level.**
- Visualized the **top 20 most frequent job titles.**



This showed that:

Most of the participants are Young

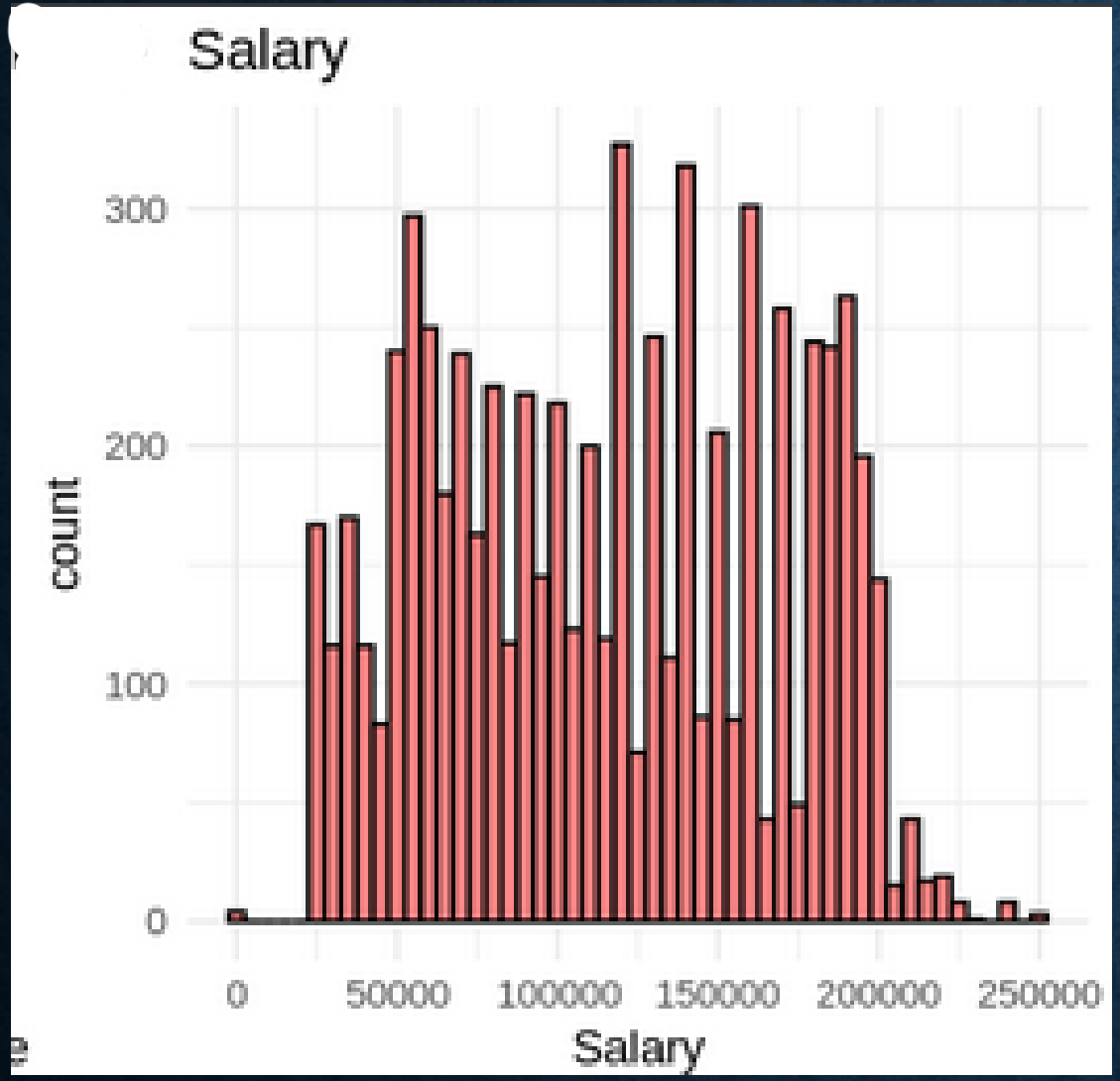
(aged 25 to 35)



Around 2000 people

had less than 5 years of

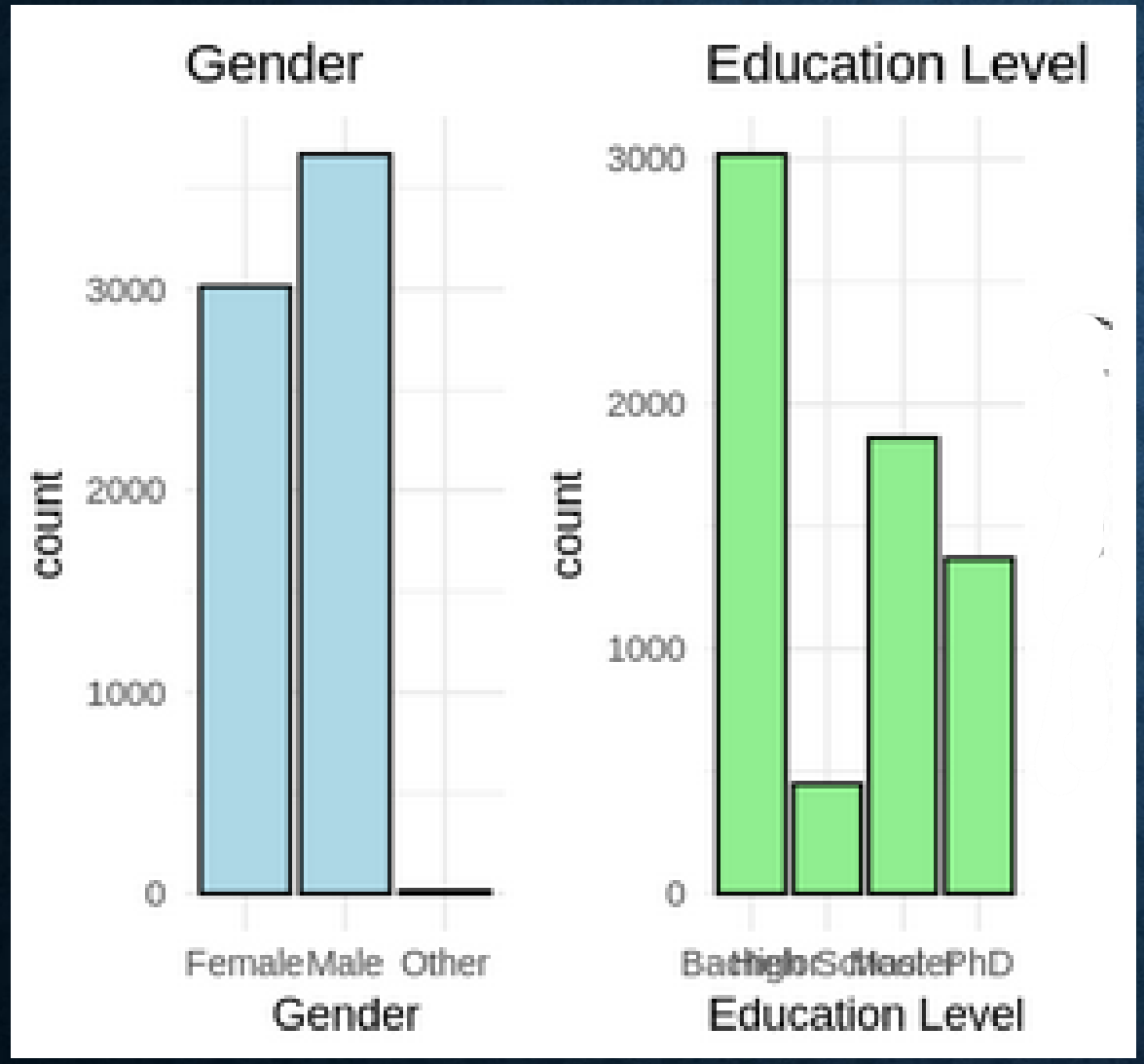
experience



Most Salary offered is

between

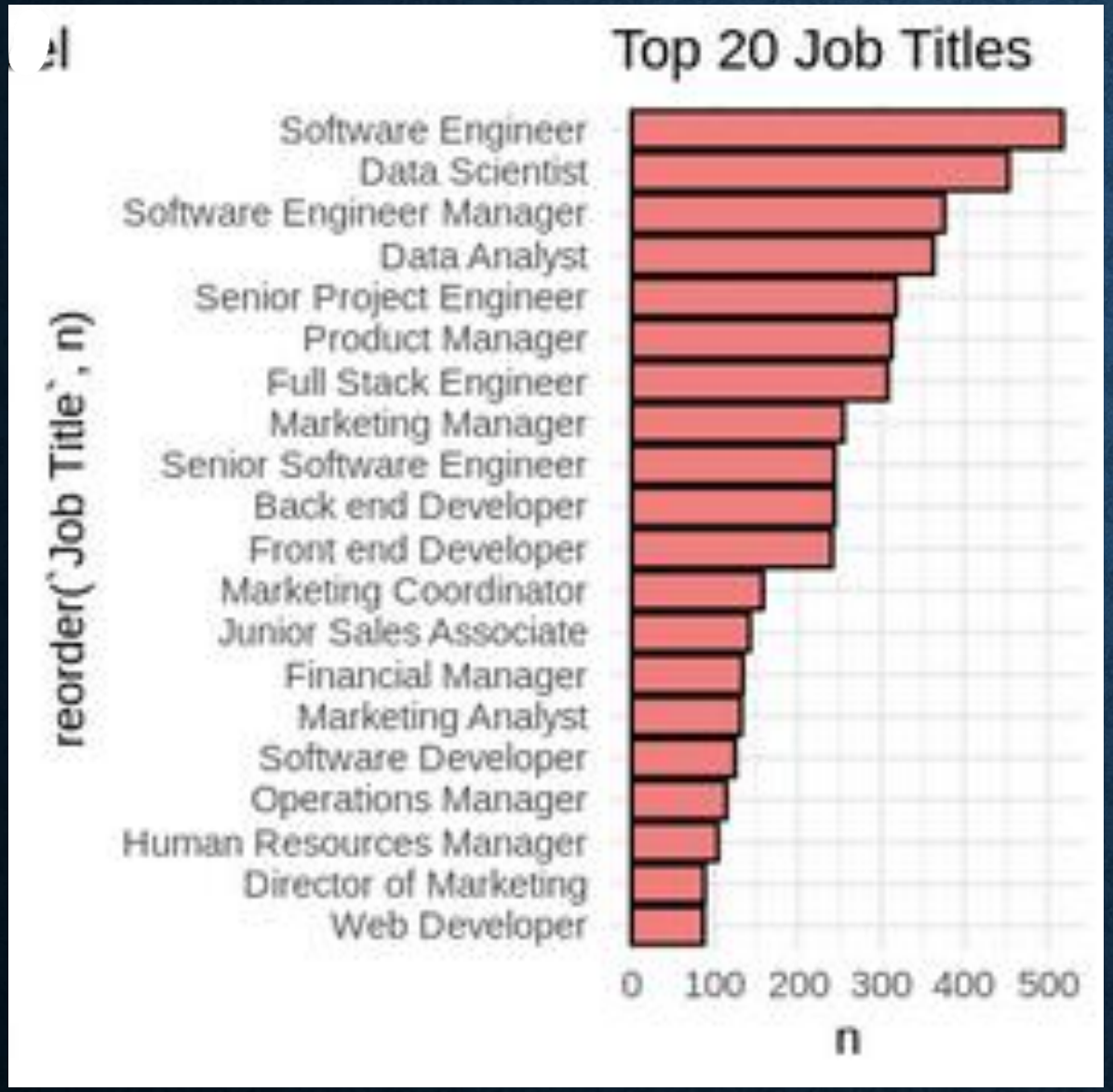
50,000 and 2,00,000



Number of Male participants were more than number of Female participants.

People with Bachelor's Degree are roughly 45%,

And 30% have Master's Degree.



The Survey focused more

on IT industry

Top 3 popular Job Titles :

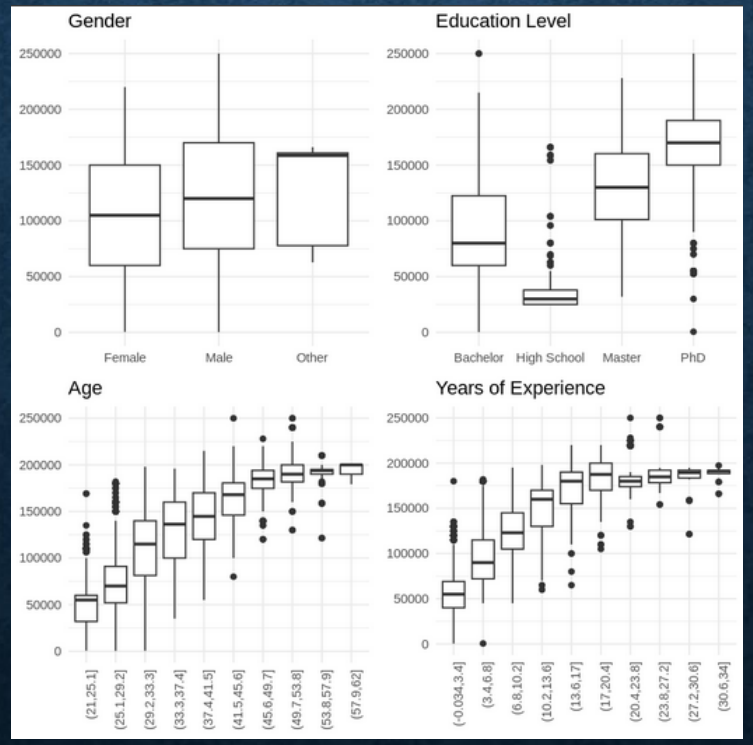
Software Engineer

Data Scientist

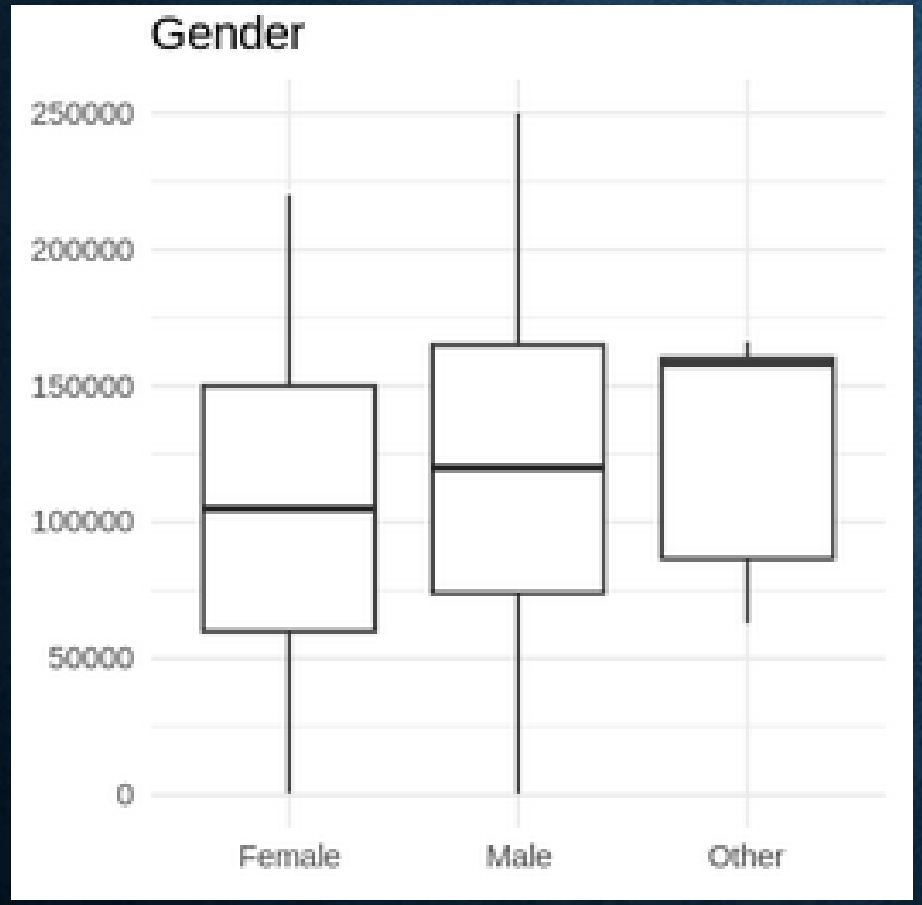
Software Engineer Manager

RELATIONSHIP BETWEEN SALARY & OTHER FACTORS

- Factors : Gender, Education Level, Age, Years of Experience



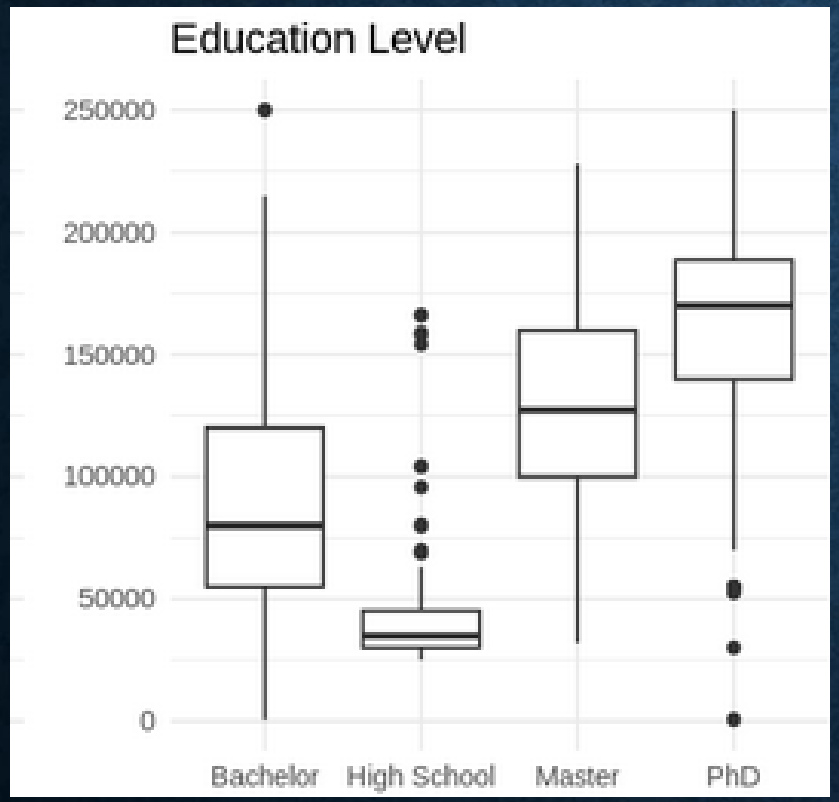
Gender vs Salary



Visualized with Box Plots to compare Salary distributions across genders.

Salary of Male participants were slightly higher than Females

Education Level vs Salary

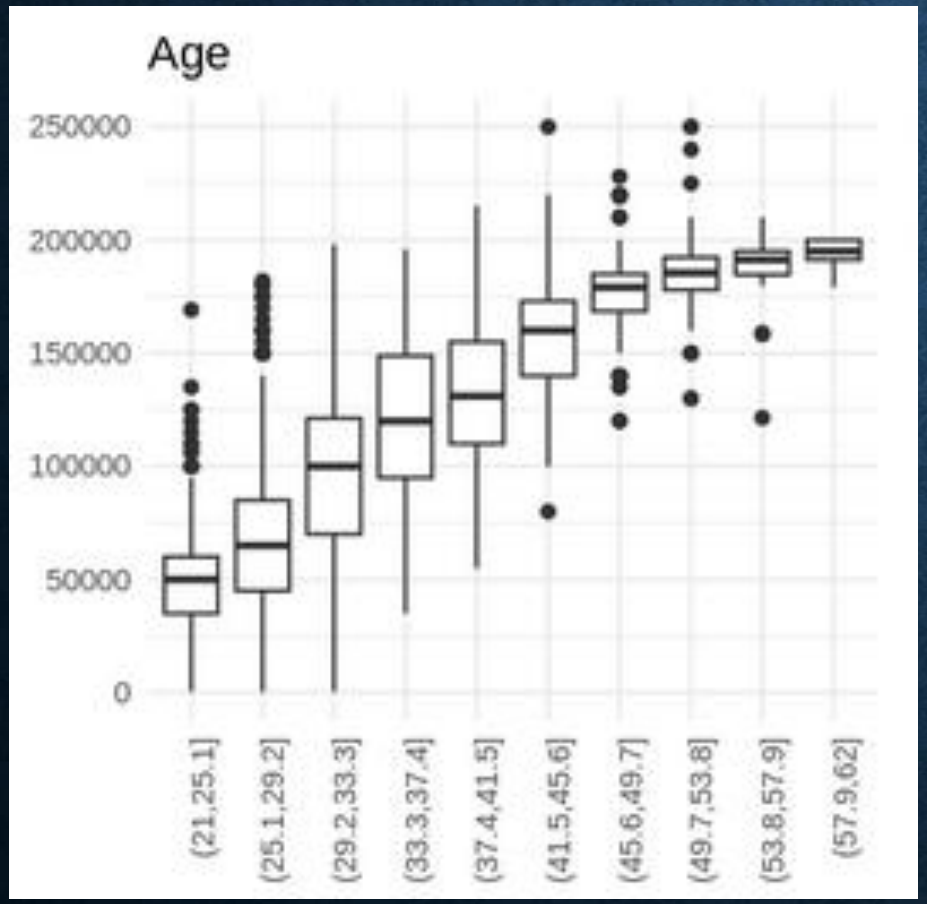


Used a **boxplot** to examine salary variations by education level.

The salary of participants with a higher degree is higher.

(Phd > Master > Bachelor > High School)

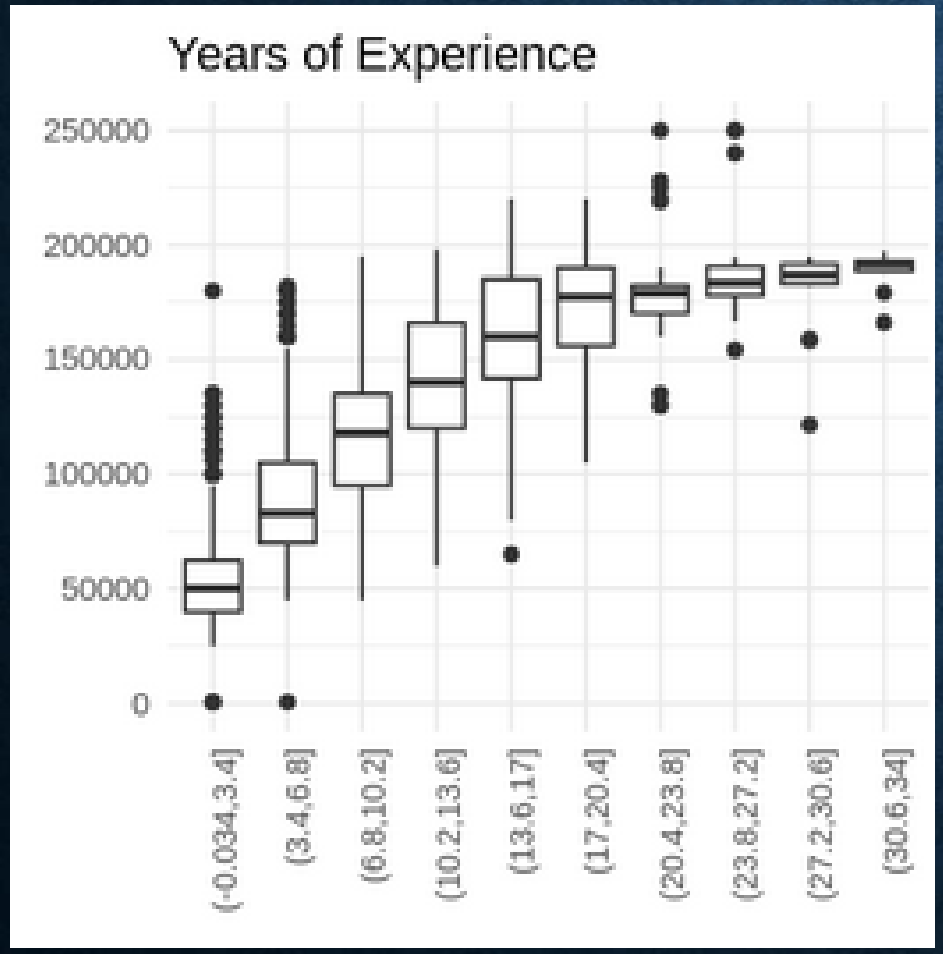
Age vs Salary



Age was binned into groups and visualized with a **boxplot** to explore salary differences across age groups.

Salary roughly Increases as Age increases

Years of Experience vs Salary



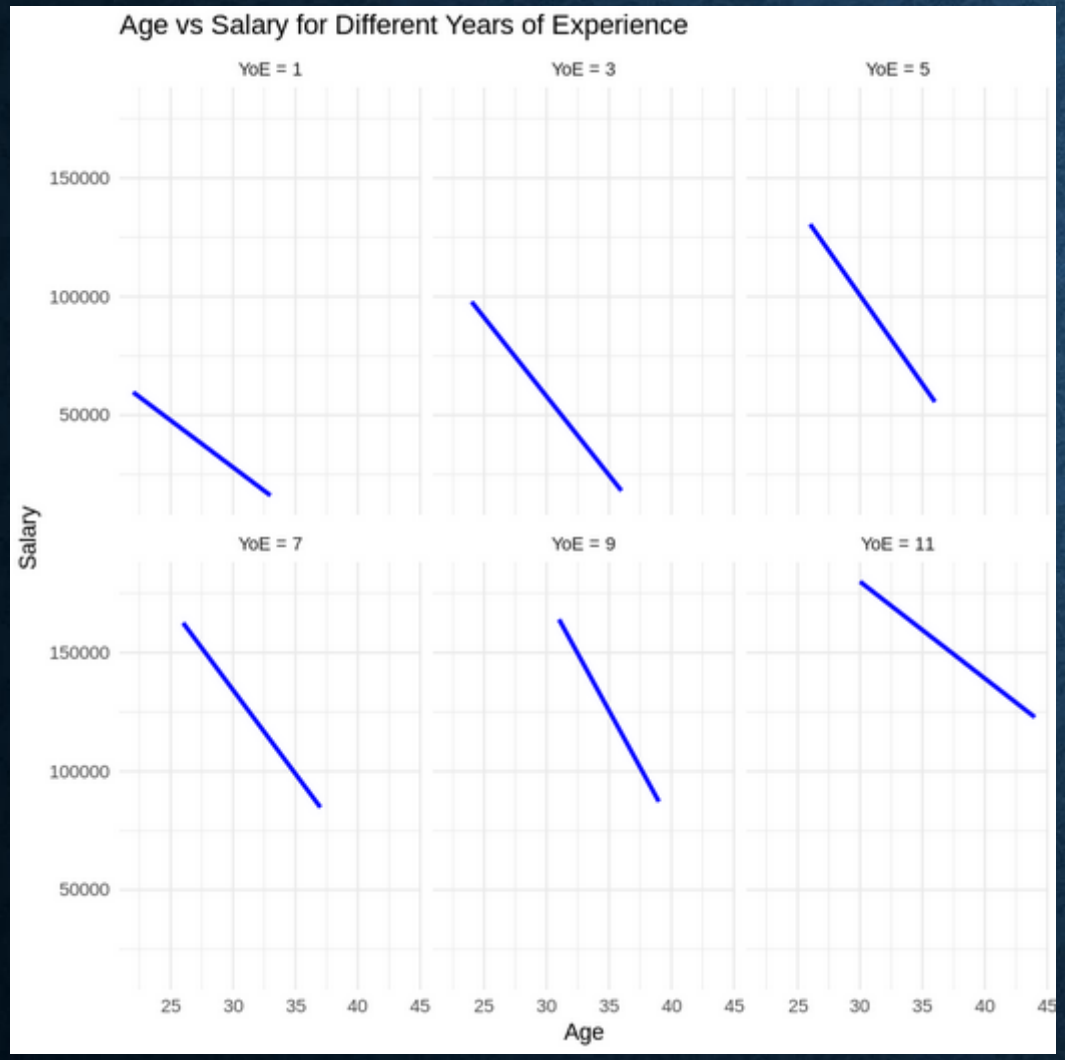
Years of Experience were binned and visualized with a **boxplot** to analyze salary trends based on experience.

The salary roughly increases as YoE increases.

AGE VS SALARY FOR DIFFERENT YEARS OF EXPERIENCE

- We saw that there was a Positive Correlation between YoE & Salary.
- There is also a Positive Correlation between YoE & Age.
- As people age, they typically accumulate more Experience.

- So, by Transitivity (Age ->YoE -> Salary), we might expect that Age and Salary are positively correlated as well.
- But our analysis showed otherwise. Age & Salary are Negatively Correlated.



Method:

The data was filtered for different for specific YoE values (1, 3, 5, 6, 9, 11)

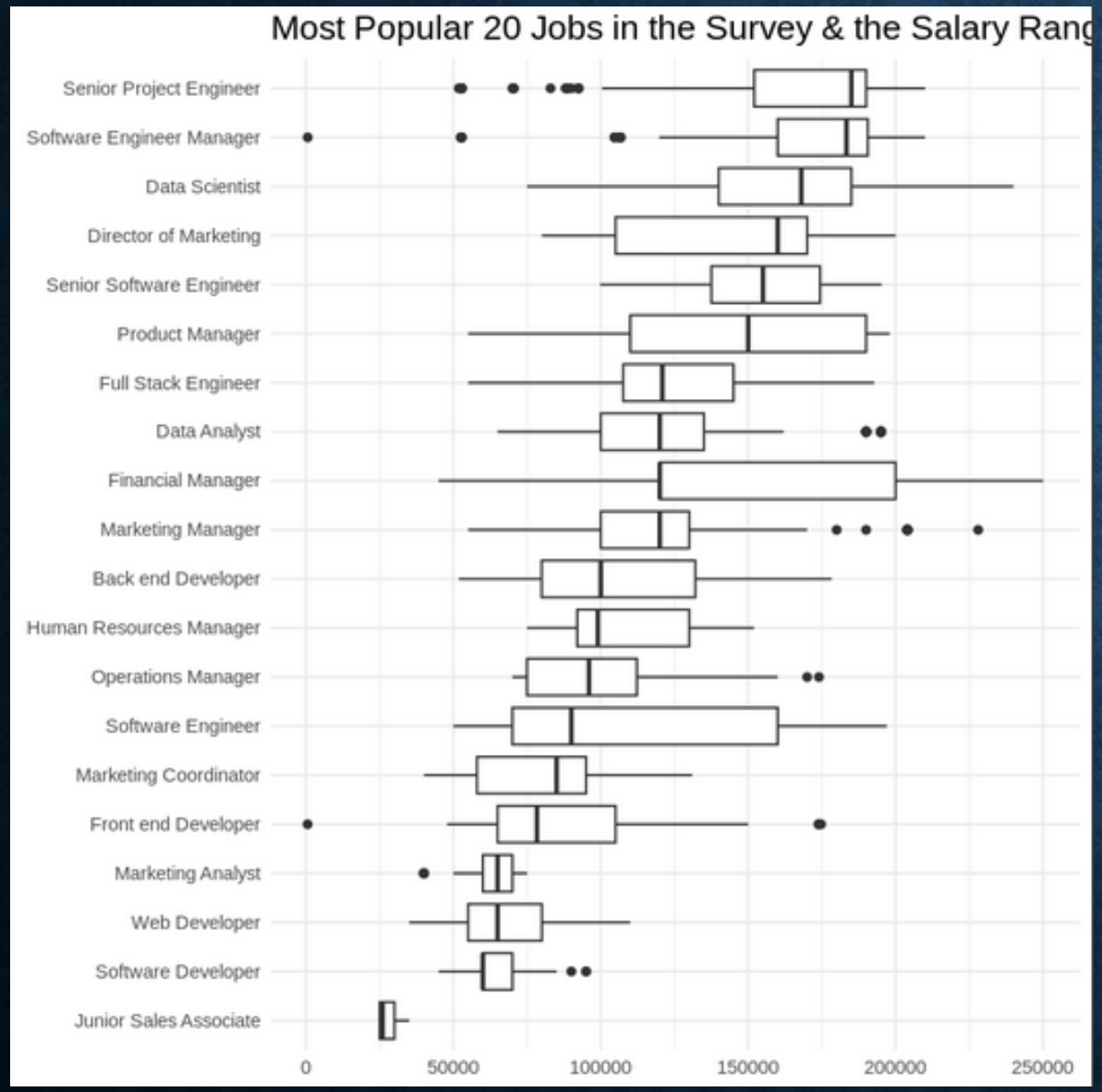
Linear Regression Model (**geom_smooth(method = "lm"**) was used to visualize the relationship between Age & Salary

Older Individuals with same YoE as Younger Individuals may earn less.

Younger workers are seen to have more potential for growth. Thus, their Higher Salary.

MOST POPULAR JOBS AND THEIR SALARY RANGES

- To Analyze the salary distribution for 20 most popular Jobs in the dataset.
- Top 20 Jobs were identified based on Frequency using **top_n()**
- **Data was** filtered to include only these jobs.
- To show Salary Ranges, boxplot was created for each, with Jobs ordered by Median Salary.



This Plot shows the Most Popular Jobs and the Salary Ranges of them in decreasing order of the Median Salaries.

Top 5 Highest Paying Jobs are:

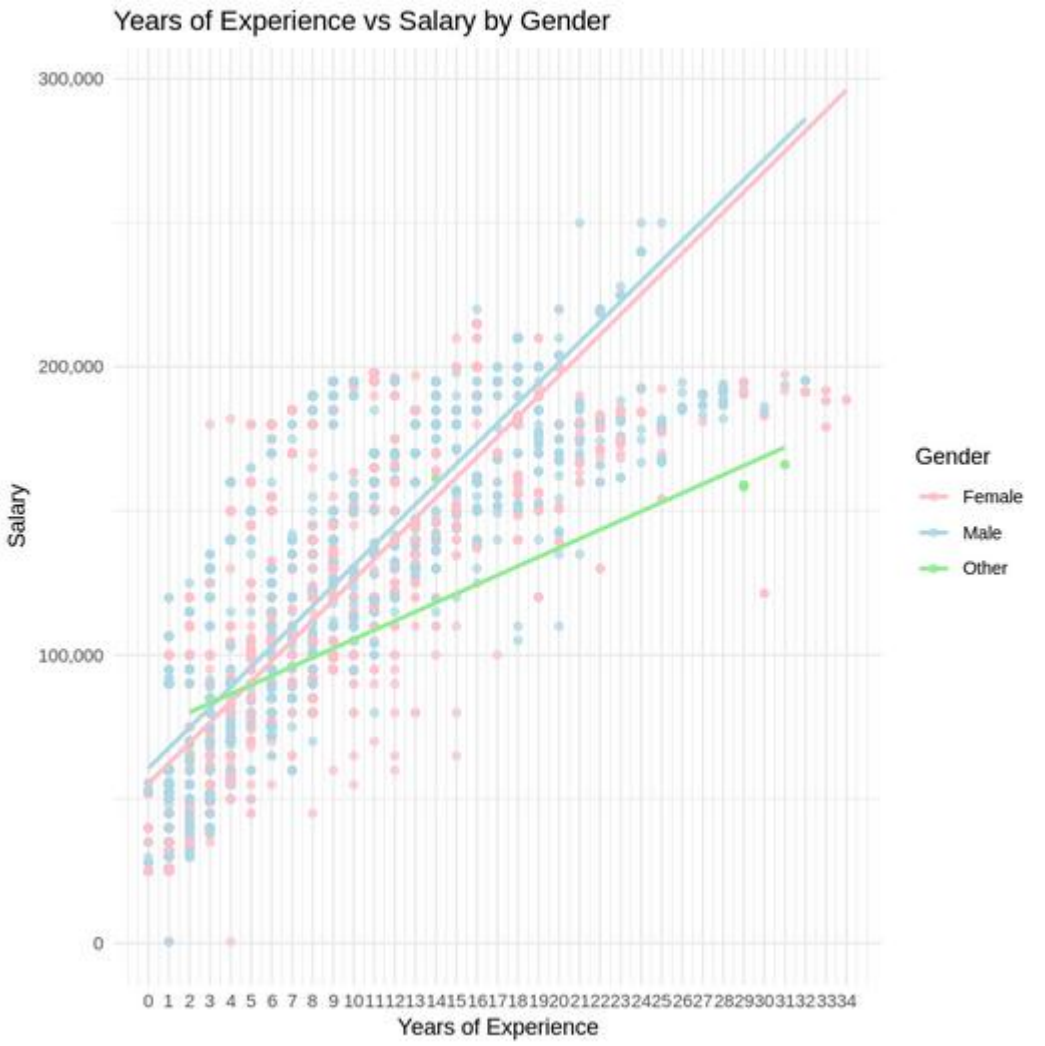
Senior Project Engineer
Software Engineer Manager
Data Scientist
Director of Marketing
Senior Software Engineer

Median Salaries of top 5 Jobs are all above Rs. 1, 50, 000

SALARY VS YEARS OF EXPERIENCE

- In the previous analysis, we found that salary and YoE have positive correlation - the higher YoE, the higher salary.
- Now checking the potential Growth rate w.r.t.:
  - Gender
  - Education Level
  - Job Title

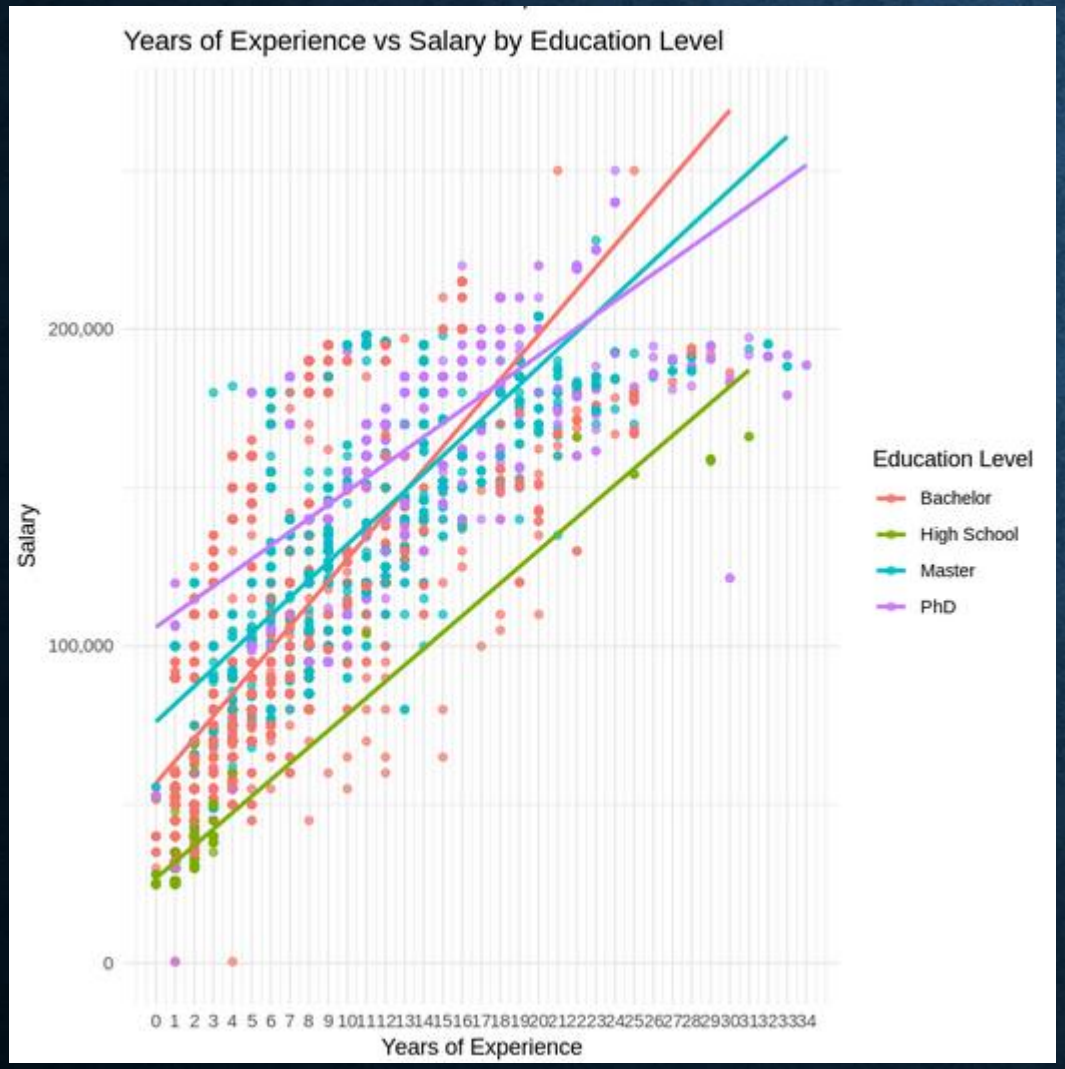
YoE vs Salary by Gender



Displays data points and Linear Regression lines for Males and Females.

Growth Rates of Salary for Males & Females look the same.

YoE vs Salary by Education Level



Similar Plot grouped by different Education Levels, showing the trend across various degrees.

Growth Rates of Salary of Different education levels are not that different.

PhD has the lowest Rate.

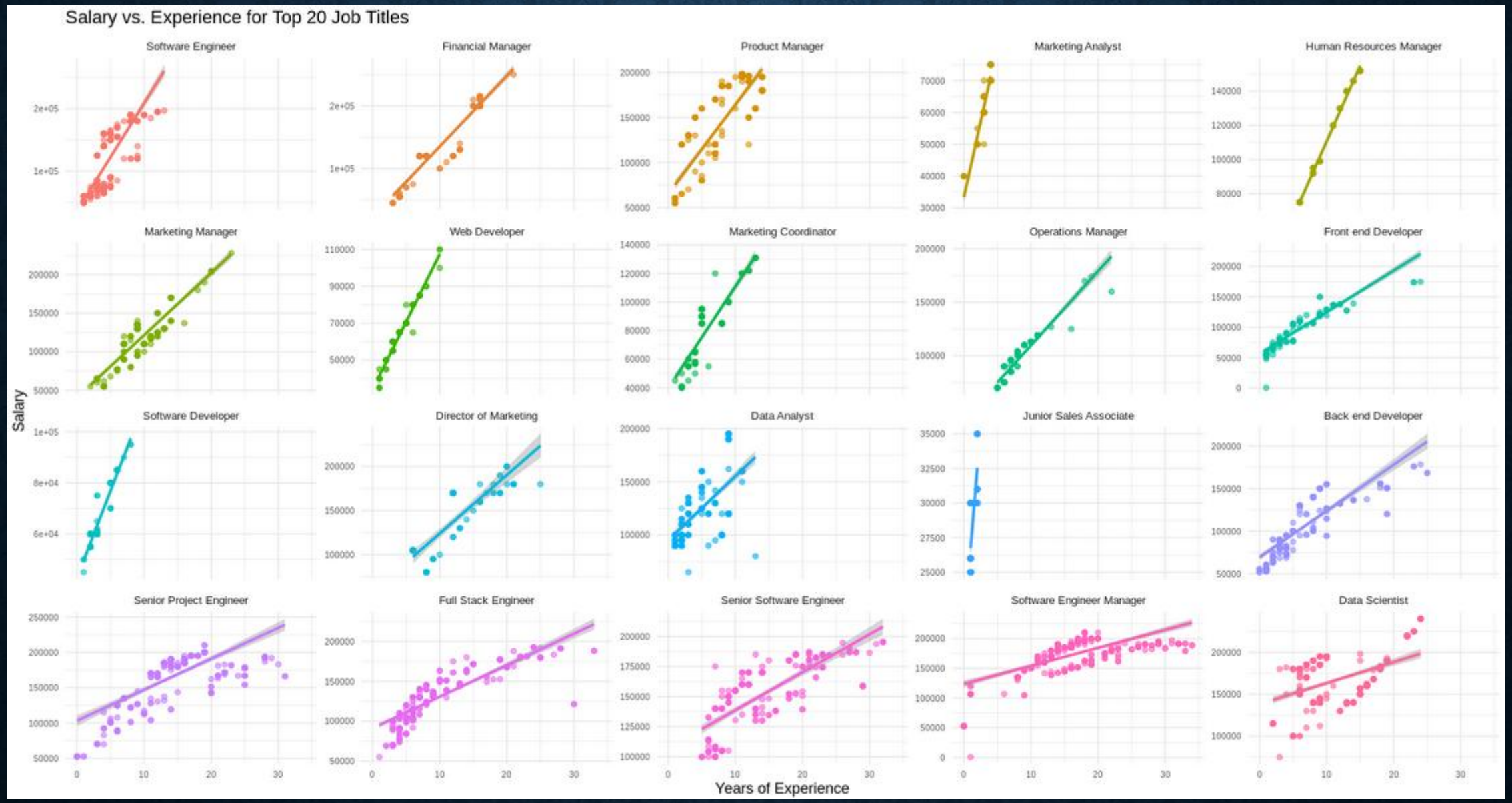
YOE VS SALARY FOR DIFFERENT JOB TITLES

- For each of top 20 Job Titles:

- Built **linear regression models** to calculate slopes (rate of salary growth with experience) and intercepts.

- Displayed individual scatter plots with regression lines and confidence intervals.

- Job titles are ranked by the slope of their regression line, indicating how quickly salaries grow with experience.


Salary vs. Experience for Top 20 Job Titles
Software Engineer
Financial Manager
Product Manager
Marketing Analyst
Human Resources Manager
Marketing Manager
Web Developer
Marketing Coordinator
Operations Manager
Front end Developer
Software Developer
Director of Marketing
Data Analyst
Junior Sales Associate
Back end Developer
Senior Project Engineer
Full Stack Engineer
Senior Software Engineer
Software Engineer Manager
Data Scientist
Salary
Years of Experience

The top 3/20 Jobs with highest growth rates are:

Software Engineer
Financial Manager
Product Manager

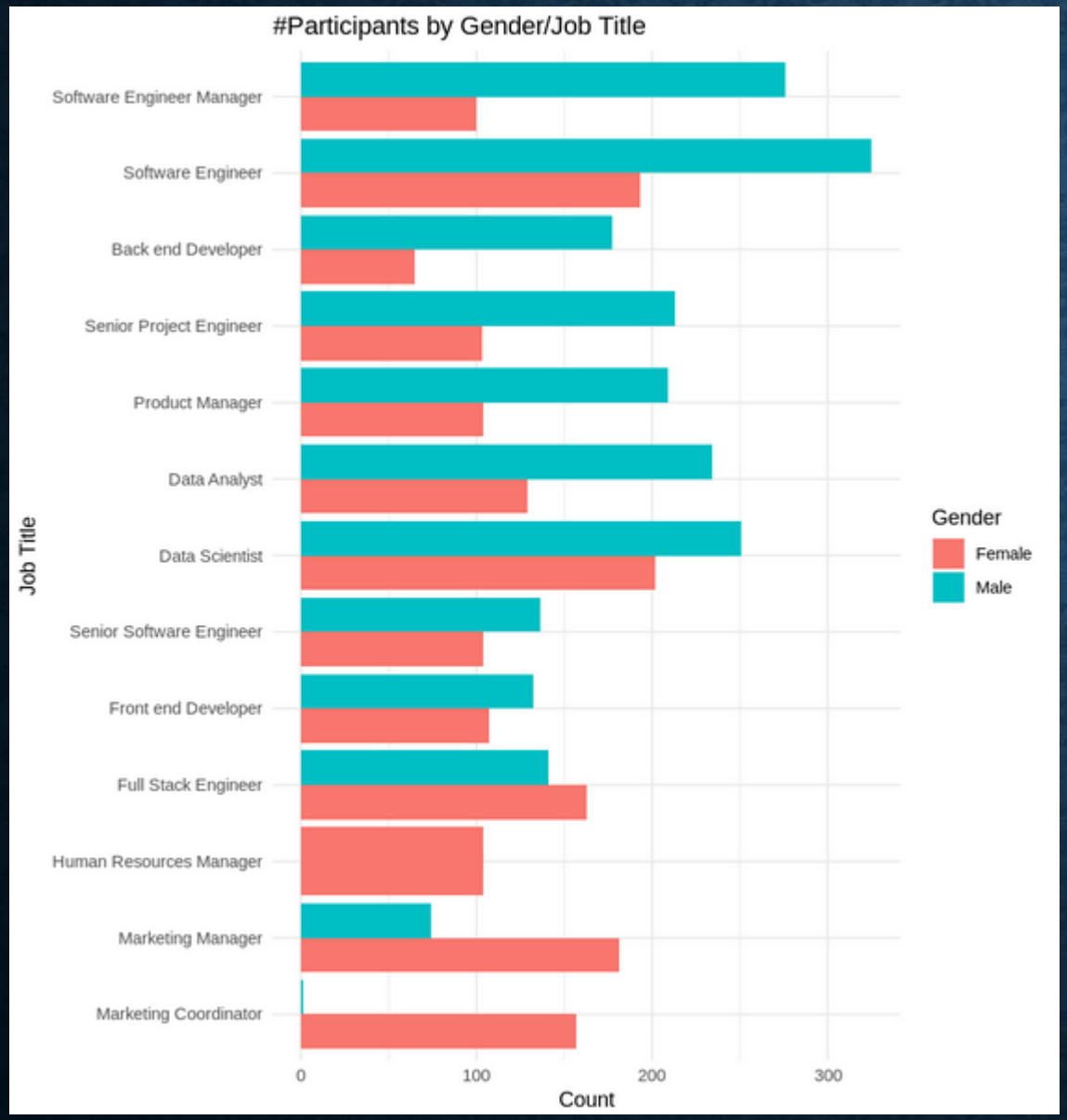
The Bottom 3/20 Jobs with Lowest growth rates are:

Senior Software Engineer
Software Engineer Manager
Data Scientist

GENDER DISTRIBUTION AND SALARY ANALYSIS

- Identified the most common Job Titles among Males and Females.
- Plotted a Bar graph to show the number of Male & Female participants for each Job title.
- Job titles are ordered by difference in participation.

- Used a Box plot to show salary distribution for each popular job title, categorized by gender.
- Job titles are ordered by Median Salary.

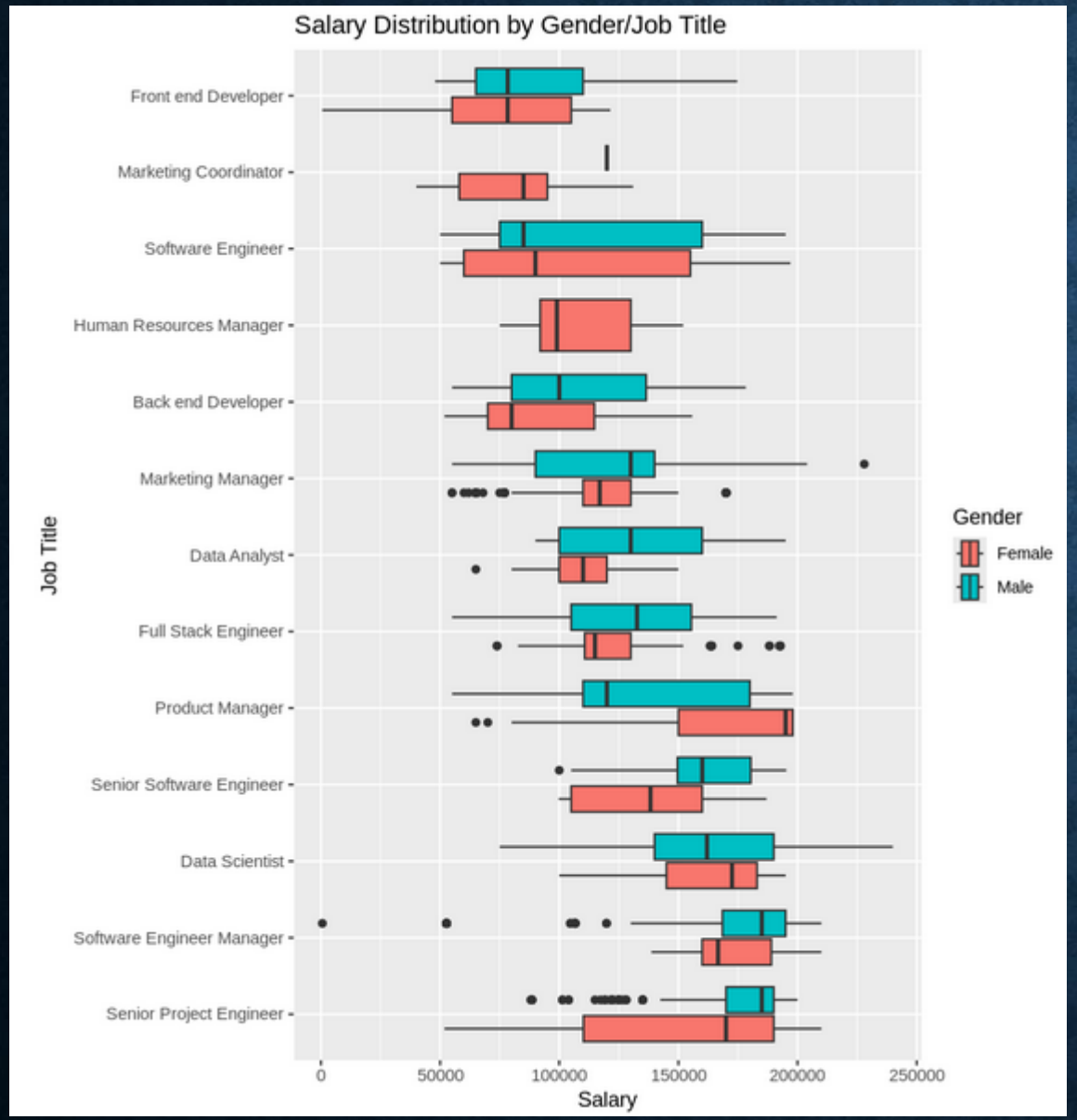


Female participants are higher in Jobs like:

Full Stack Engineer (Tech)
Human Resource Manager (Non-Tech)
Marketing Manager (Non-Tech)
Marketing Coordinator (Non-Tech)

Male Representation more in Tech Fields

Female representation more in Non-Tech Fields



This shows that Salary Ranges for Males and Females overlap for most roles.

Generally,

Tech Roles : higher salaries

Non-Tech Roles: Lower salaries

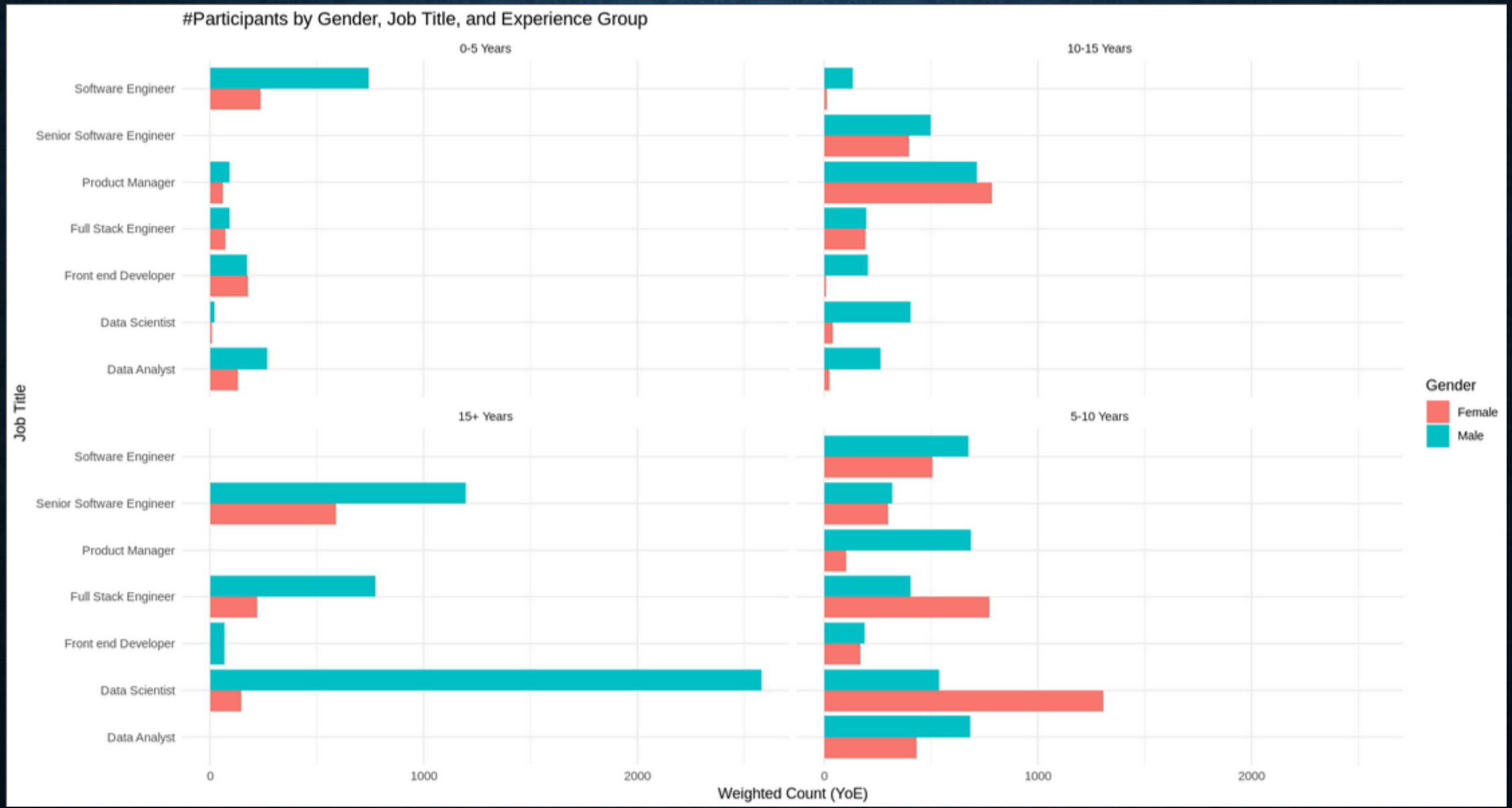
We also found some Pay Gaps.

Men earned more for same Job Titles.

UNCOVERING THE UNDERLYING CAUSES OF THESE PAY GAPS

- We considered Years of Experience, as it could have been a factor.

- Instead of counting each participant equally, the count was adjusted based on the number of YoE they had.
- Data was categorized into Experience Levels: 0-5 Years, 5-10 Years, 10-15 Years and 15+ Years.
- Analysed for the Jobs which were popular among both Males & Females.


#Participants by Gender, Job Title, and Experience Group
0-5 Years
10-15 Years
15+ Years
5-10 Years
Software Engineer
Senior Software Engineer
Product Manager
Full Stack Engineer
Front end Developer
Data Scientist
Data Analyst
Job Title
0
1000
2000
Weighted Count (YoE)
Gender
Female
Male

Findings

Among the survey Participants, Males in general had more YoE, except for 5-10 YoE group.

So, this could have been the reason for Pay Gaps.

However, we cannot be entirely sure yet.

To draw more accurate conclusions, we need to consider additional factors and gather more data.

This will be part of our **Future Work**.

THANK YOU !!