Analysis of Tech Industry Salaries and Experience Level

By Ebenezer Omari

1. Descriptive Introduction

The tech industry is known for its competitive salaries and rapid career progression. This report analyzes the relationship between salaries and experience levels in the tech sector, aiming to provide insights for both job seekers and employers. The main objectives are to:

1. Explore the distribution of salaries across different experience levels
2. Quantify the relationship between years of experience and salary
3. Identify any notable trends or patterns in salary progression

2. Data Preprocessing

To prepare the data for analysis, the following steps were taken:

1. Checked for duplicates
2. Checked any rows with missing values
3. Checked the structure of the dataset
4. Converted experience_level, employment_type, company_size to a factor
5. Split the data into training (70%) and testing (30%) sets for our decision tree model

3. Exploratory Analysis

Salary Distribution

summary(salary_data$salary_in_usd)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2859	62726	101570	112298	150000	600000

The summary of salary in usd gives a strong indication to the distribution of salary in usd in the dataset.

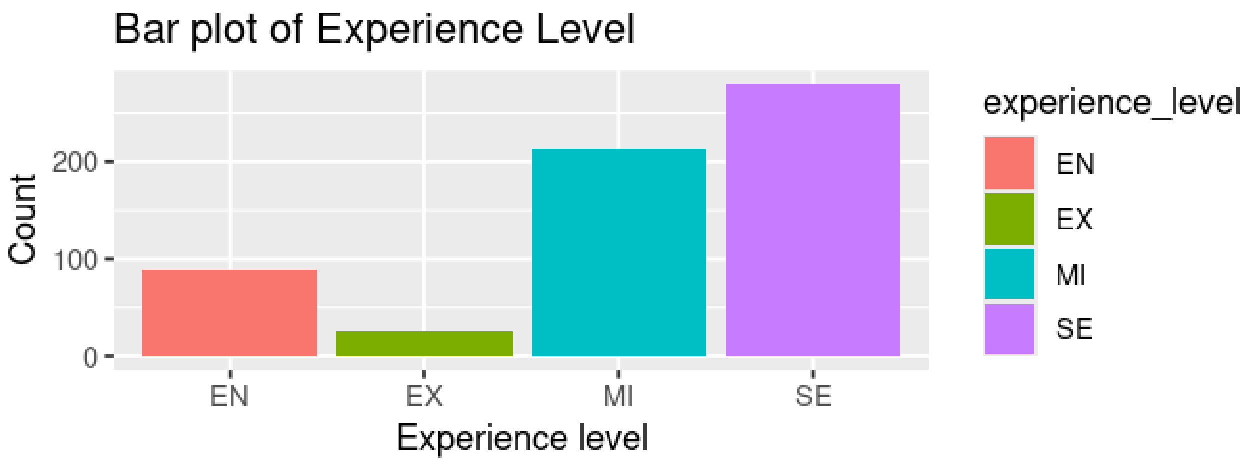
Range: The range of salaries is quite wide. The Salary ranges from 2,859.00 USD to 600,000.00 USD. The average Salary is 112,298.00 USD

Quartile: 1st Quartile: 25% of the salaries is lower than 62,726.00 USD(1st Quartile)

3rd Quartile: 75% of the salaries is higher than 150,000.00 USD(3rd Quartile)

Skewness: The salary distribution shows a right-skewed pattern, with a median salary of $101,570 and a mean of $112,298.

Data Visualization on experienced level using Bar Plot



Analysis

This Bar plot depicts the distribution of experience levels among a group of individuals. The x-axis represents different experience levels, while the y-axis shows the count of individuals in each category.

Here's a breakdown of the information presented:

EN (Entry): Represented by a pink bar, with about 100 individuals.

EX (Expert): Shown in green, this is the smallest group with about 50 individuals.

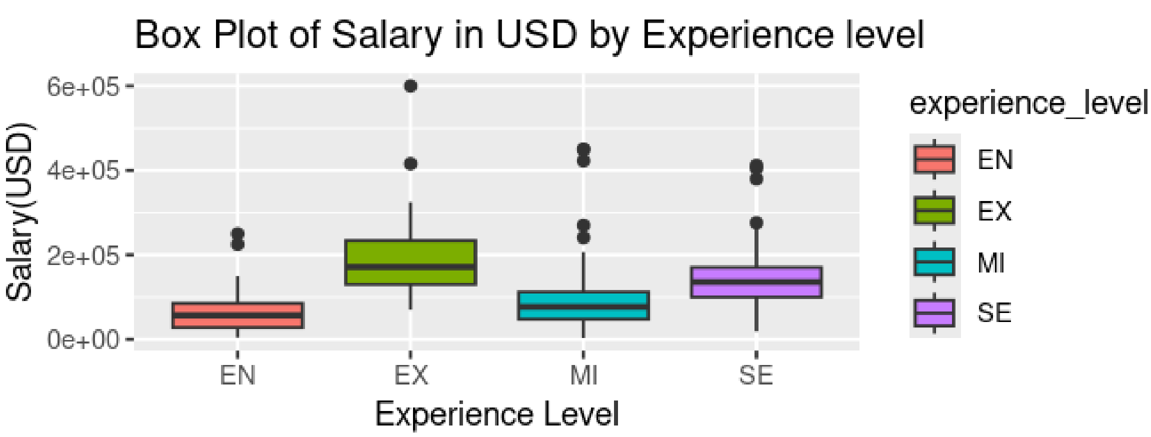
MI (Mid-level): Depicted by a teal bar, this is the second-largest group with about 200 individuals.

SE (Senior): Represented by a purple bar, this is the largest group with over 200 individuals.

The plot effectively illustrates the relative proportions of each experience level

It appears that mid-level (MI) and senior (SE) experience levels are the most common, while expert (EX) level is the least represented in this dataset.

Data Visualization on salary in USD by experience level using Box Plot



Analysis

This image shows a box plot comparing salaries (in USD) across different experience levels

Entry level (EN) has the lowest median salary and the smallest range.

Expert level (EX) shows the highest median salary and the largest range, indicating more variability in pay.

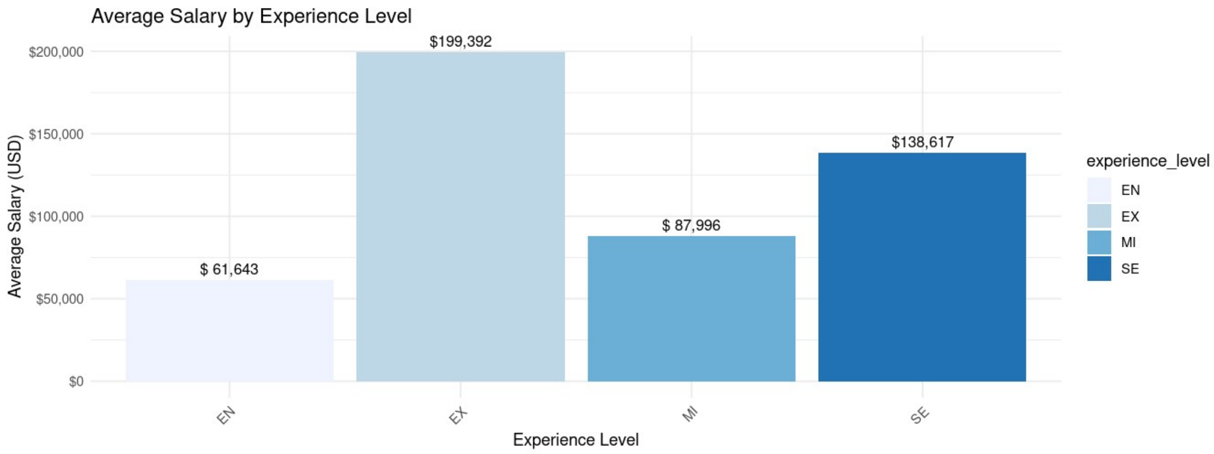
Mid-level (MI) and Senior level (SE) fall between EN and EX, with SE having a higher median than MI.

All levels show some outliers above their main distributions.

There's a clear progression in salary as experience level increases.

The range of salaries tends to increase with experience level, suggesting more pay variability in higher positions.

Data Visualization of a bar chart depicting the average salary by experience level



<u>Analysis</u>

This image shows a bar chart depicting the average salary by experience level. Here's a breakdown of the information presented:

Experience Levels:

- EN: Entry level
- EX: Expert level
- MI: Mid-level
- SE: Senior level

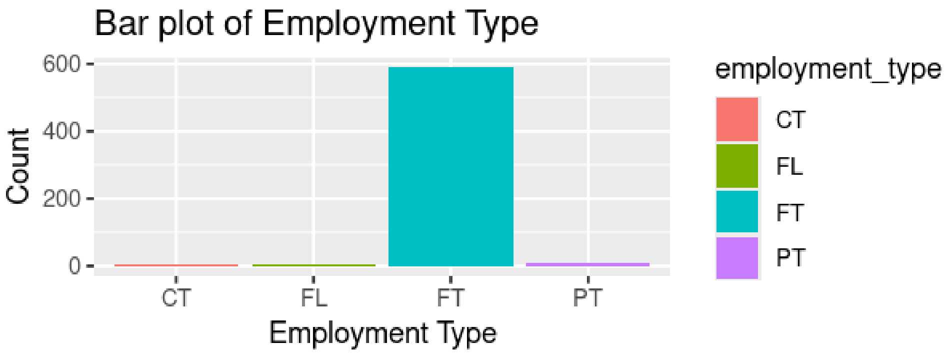
- Average Salaries:

- EN (Entry): $61,643
- EX (Expert): $199,392
- MI (Mid-level): $87,996
- SE (Senior): $138,617

The observations made were:

- The Expert level (EX) has the highest average salary at $199,392.
- Entry level (EN) has the lowest average salary at $61,643.
- There's a significant jump from Entry to Mid-level and then again to Senior level.
- Interestingly, the Expert level average salary is higher than the Senior level.

Data Visualization on Bar Plot of Employment Types



<u>Analysis</u>

This bar plot (also known as count plot) shows the distribution of different employment types in a data set.

The employment type CT: Likely stands for Contract

FL: Likely stands for Freelance

FT: Likely stands for Full-Time

PT: Likely stands for Part-Time

From the distribution, FT (Full-Time) is overwhelmingly the most common employment type, with almost 600 counts.

The other employment types (CT, FL and PT) have very low counts, barely visible on the graph compared to FT. Looking at the visual representation, Each employment type is represented by a different color bar. The y-axis shows the count, ranging from 0 to 600. The x-axis lists the different employment types. A Key observation is that, There's a stark contrast between full-time employment and all other types, suggesting that in this data set, full-time positions are significantly more prevalent than contract, freelance, or part-time positions. This visualization effectively illustrates the dominance of full-time employment in the data set, with other employment types being comparatively rare.

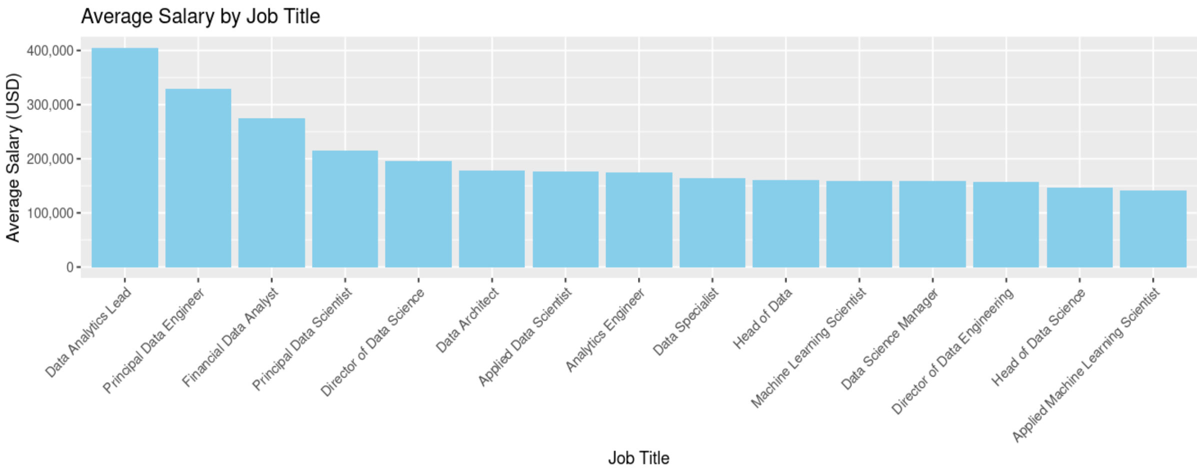
Data Visualization on Box Plot of Salary by Employment Type

<u>Analysis</u>

This box plot compares salary distributions in USD across different employment types. The y-axis represents salary in USD, ranging from 0 to 6e+05 (600,000). Each box represents the interquartile range (IQR) for that employment type. From the observations, Full-Time (FT) jobs have the highest median salary and the widest range. Contract (CT) positions show the second-highest median and a wide range. Freelance (FL) and Part-Time (PT) positions have lower median salaries. FT has several high outliers, indicating some very high-paying full-time positions. Looking at the comparisons: FT and CT show higher overall salaries compared to FL and PT. PT has the narrowest salary range, suggesting less variability. FL shows a compressed salary range, but with some high outliers.

This visualization effectively illustrates how salary distributions vary across different employment types,with full-time positions generally offering higher salaries.

Data Visualization on Top 10 Job Titles by Average Salary



<u>Analysis</u>

This image shows a bar plot of average salaries for various job titles. The x-axis represents different job titles, while the y-axis shows the average salary in dollars. The y-axis ranges from 0 to $300,000 with increments of $100,000. Job titles are arranged in descending order of average salary from left to right. The highest-paying job title appears to be "Data Analytics Lead" Other high-paying roles include "Principal Data Architect", "Finance Data Analyst"and "Principal Data Scientist". The lowest-paying job title among those shown is "Applied Machine Learning Scientist" Leadership positions (e.g., "Lead", "Principal") tend to be higher on the salary scale. This visualization effectively displays the salary hierarchy among various data-related job titles, highlighting the financial value placed on different skills and positions within the industry.

4. Confirmatory Analysis

Chi-Square Test

H0 = The proportion of full-time employees does not differ significantly across small, medium

and large companies.

H1 = The proportion of full-time employees differs significantly across small, mediumand large companies.

Pearson's Chi-squared test

data: salary_data$employment_type and salary_data$company_size

X-squared = 6.9349, df = 6, p-value = 0.3269

Analysis

Based on the Chi-squared test, the p-value = 0.3629 is greater than the significant level (0.05), This suggests that we fail to reject the null hypothesis. There is not enough evidence to conclude that the proportion of full-time employees differs significantly across small, mediumand large companies.

Logistic Regression

H0 = There is no significant difference in salary (in USD) among different employment types (e.g., full-time, part-time, contract).

H1 = There is a significant difference in salary (in USD) among different employment types.

summary(logistic)

Call:

glm(formula = employment_type ~ salary_in_usd, family = "binomial",

    data = salary_data)

Deviance Residuals:

```
    Min       1Q   Median       3Q      Max
-3.3525   0.0956   0.1105   0.1336   0.7012
```

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept)   5.859e+00 7.719e-01   7.591 3.17e-14 ***

salary_in_usd -7.636e-06 3.492e-06 -2.187   0.0287 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 57.950 on 606 degrees of freedom

Residual deviance: 54.628 on 605 degrees of freedom

AIC: 58.628

Number of Fisher Scoring iterations: 8

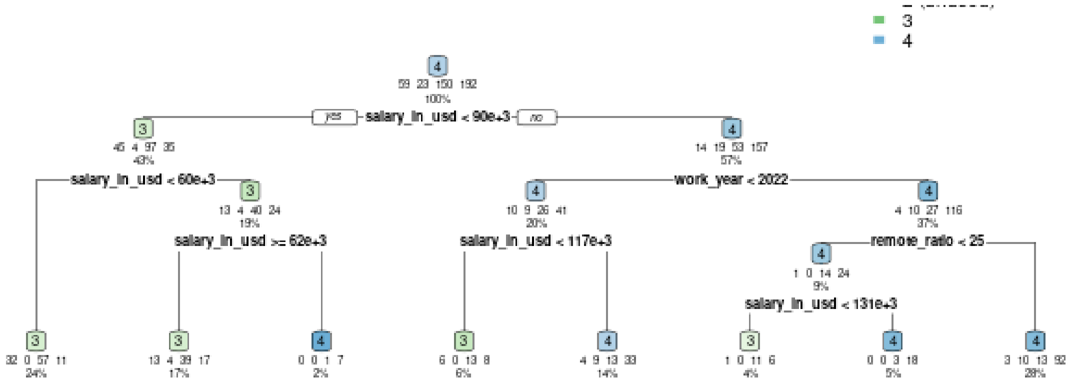
Analysis

From the analysis conducted, the p-value(3.17e-14) is extremely small, indicating strong evidence against the null hypothesis.

The p-value (3.17e-14) is much smaller than the common significance level of 0.05 (or even 0.01).

Therefore we reject the null hypothesis and conclude there is a significant difference in salary (in USD) among different employment types.

5. Decision Tree Visualization



A decision tree model was developed to further investigate the relationship between various factors and experience level in the tech industry. This model aimed to predict experience levels based on salary, company size, employment type, work year and remote work ratio.

Model Performance:

The decision tree model achieved an accuracy of 59% on the test dataset. This result indicates that the model correctly predicts the experience level for nearly 6 out of 10 cases,

Interpretation:

The 59% accuracy suggests that:

1. There is a discernible pattern in how factors like salary, company size and work arrangements relate to experience levels.
2. These factors provide meaningful information for distinguishing between different experience levels, but they don't tell the whole story.

Model Insights

The decision tree provided valuable insights showing influential variables in the model

	feature importance
salary_in_usd	salary_in_usd 52.065904
work_year	work_year 12.471510
company_size	company_size 6.926430

The top three important variable were found to be the most influential in predicting experiencing levels. This suggests these factors are strongly associated with career progression in the tech industry.

The moderate accuracy indicates that the relationship between these factors and experience levels is not straightforward.

While providing useful insights, the model's performance suggests room for improvement:

1. Additional Features: Incorporating more relevant features, such as specific technical skills or educational background could potentially improve the model's predictive power.
2. Larger Dataset: A larger, more diverse dataset could help the model learn more significance relationships between the variables and experience levels.

In conclusion, the decision tree model, with its 59% accuracy, serves as a valuable starting point for understanding the factors influencing experience levels in the tech industry.

6. Insights and Conclusions

Key findings from our analysis include:

1. There's a strong positive correlation between years of experience and salary in the tech industry.
2. The variability in salaries increases significantly at senior and expert levels.
3. Different job roles show distinct patterns of salary progression
4. Full-time employment is the most common type in the tech industry.
5. There is a strong relationship between salary level and employment type.
6. Higher salaries are generally associated with full-time positions.