|  |  |
| --- | --- |
| **GRADUATION THESIS** | |
| **DATA ANALYSIS** | |
| **Exploring Salary Determinants and Predictions** | |
| **Student: Tran Phuc Toan**  **Ho Chi Minh City, September 2024** | |
|  | |

Contents

[**I.** **INTRODUCTION** 4](#_Toc176388859)

[**1.** **Research Overview** 4](#_Toc176388860)

[**2.** **Research objective** 4](#_Toc176388861)

[**3.** **Research question** 4](#_Toc176388862)

[**4.** **Research scope** 4](#_Toc176388863)

[**5.** **Research object** 5](#_Toc176388864)

[**6.** **Scientific contribution** 5](#_Toc176388865)

[**7.** **Technologies** 5](#_Toc176388866)

[**II.** **DATA AND METHODOLOGY** 6](#_Toc176388867)

[**1.** **Proposed Research model** 6](#_Toc176388868)

[**2.** **Research proxy variables** 7](#_Toc176388869)

[**a.** **Dependant variables** 7](#_Toc176388870)

[**b.** **Independent variables** 8](#_Toc176388871)

[**3.** **Estimation Method** 10](#_Toc176388872)

[**a.** **Exploratory Data Analysis (EDA)** 10](#_Toc176388873)

[**b.** **Correlation Analysis** 10](#_Toc176388874)

[**c.** **Linear Regression Models** 11](#_Toc176388875)

[**III.** **RESEARCH RESULTS AND DISCUSSION** 12](#_Toc176388876)

[**1.** **Descriptive statistics of variables** 12](#_Toc176388877)

[**2.** **Overview findings** 14](#_Toc176388878)

[**a.** **Relationship between Salary and Genders** 14](#_Toc176388879)

[**b.** **Relationship bewteen Salary and Education** 14](#_Toc176388880)

[**c.** **Highest-paying Jobs** 15](#_Toc176388881)

[**d.** **Gen Z’s ideal occupation** 16](#_Toc176388882)

[**3.** **Correlation Analysis** 16](#_Toc176388883)

[**a.** **Age** 17](#_Toc176388884)

[**b.** **Gender:** 18](#_Toc176388885)

[**c.** **Education Level:** 18](#_Toc176388886)

[**d.** **Years of Experience:** 19](#_Toc176388887)

[**e.** **Level:** 19](#_Toc176388888)

[**f.** **Field** 19](#_Toc176388889)

[**g.** **Salary** 20](#_Toc176388890)

[**4.** **Multiple Linear Regression Model** 22](#_Toc176388891)

[**a.** **Model Summary** 22](#_Toc176388892)

[b. **Coefficients and Their Interpretations** 23](#_Toc176388893)

[**IV.** **CONCLUSION** 25](#_Toc176388894)

[1. **Factors Influencing Salary Levels** 25](#_Toc176388895)

[2. **Proposed Salary Prediction Model** 25](#_Toc176388896)

[3. **Limitations of the research** 26](#_Toc176388897)

[4. **Suggestions for Salary Improvement** 26](#_Toc176388898)

1. **INTRODUCTION**
2. **Research Overview**

A person's pay is determined by a number of things. The purpose of this study is to investigate and evaluate the factors that influence Vietnamese salaries and to create a predictive model using relevant information. The age, gender, years of experience, education level, job level, and job field are among the independent variables taken into account in the study. Salary is the dependent variable. Utilizing statistical methodologies and data modeling strategies, the research will evaluate the influence of every component on payment and suggest a useful model for wage prediction. The results of this study can offer individuals, employers, and legislators useful information to help them make well-informed decisions on compensation.

1. **Research objective**

This study specifies three main objectives, including:

(1) Examining the factors that influence salary levels.

(2) Proposing a model for predicting salary.

(3) Suggesting methods for individuals to improve their salary.

1. **Research question**

What factors influence salary, and how do they impact it?

1. **Research scope**

Data of the Vietnamese labor force in 2023 gathered from a variety of sources, including as polls, websites that promote jobs, and other publicly accessible information. There were 6698 data points gathered in all. The dataset included five variables: age, experience, job role, and education level and salary.

1. **Research object**

This research examines the factors affecting Vietnamese labor’s salary

1. **Scientific contribution**

Scientifically, research has contributed three fundamental meanings.

(1) This study confirmed the elements that have effects on Vietnamese income

(2) This was one of the most recent studies on Vietnamese Salary.

(3) This study can be used as a starting point for further research.

1. **Technologies**

The initial step of the research involves cleaning the data using Google Sheets to make sure the dataset is accurate and consistent. Following cleaning, correlation analysis and linear regression are used to analyze the data. The dependent variable (salary) and the independent variables (age, gender, education level, years of experience, job level, and job field) can be related to one another with the use of correlation analysis. Next, a salary prediction model is created by quantifying these correlations through the use of Python's Pandas package and linear regression. Finally, in order for readers to understand easily, Power BI is used to visualize the findings.

1. **DATA AND METHODOLOGY**

This chapter provides an overview of the research process, including problem identification, to propose a specific research model. Besides, the variables in the proposed model are also explained and explicitly analyzed. Methods of research and analyzing data are also presented at the end of this chapter.

1. **Proposed Research model**

This research expects to fulfill the needs to estimate oneself’s reasonable income based on their knowledge and experience. Accordingly, the author propose a hypothesis and a research model that can evaluate the impact of the mentioned variables has on the salary of the Vietnamese workforce, and that effect bares the control of internal factors. Therefore, the following hypothesis and model have been suggested:

H1: Age, Gender, Education Level, Years of Experience, Job Level, Job Field all have impact on income

*E\_Salary = 𝛽0 + 𝛽1 x Age + 𝛽2 x Gender + 𝛽3 x Education Level + 𝛽4 x Job Level + 𝛽5 x Job Field*

AGE

GENDER

EDUCATION

EXPERIENCE

JOB LEVEL

JOB FIELD

SALARY

1. **Research proxy variables**

The preceding section of this chapter has introduced the research hypothesis and model, the main variables used to estimate results, and their general descriptions. In this part, the following table will show the characteristics of each variable group, including concise definitions and the source of data. Besides, the next part of this chapter will analyze the measurement of dependent and explanatory variables.

Table 1 Variables description

| **Variables** | **Definition** |
| --- | --- |
| **Dependent variable** | |
| *E\_Salary* | Expected Salary (yearly) |
| **Explanatory variables** | |
| *age* | Labor’s age |
| *gender* | Labors’ gender |
| *Education level* | Labor’s educational level |
| *Years of Experience* | Labor’s years of experience |
| *Job Level* | Level of the occupation |
| *Job Field* | Field of the occupation |

Source: Author

* 1. **Dependant variables**

The primary dependent variable in this study is the predicted YEARLY pay, which indicates the amount of money people expect to make depending on a range of variables like their experience, education, and type of work. Numerous explanatory factors, including as age, gender, years of experience, education level, job level, and job field, have an impact on it. The relationship between these elements and an individual's expected earning capacity is measured by the predicted salary, and constructing precise salary prediction models requires an understanding of this relationship.

* 1. **Independent variables**

The study considers several independent variables that are hypothesized to influence the expected salary:

1. **Age:** This variable represents the age of an individual. It may impact salary as it often correlates with experience and career progression.
2. **Gender:** This categorical variable denotes whether the individual is male or female. Gender may influence salary due to potential disparities in pay across different genders. In the data cleaning process, the author has digitized this category, transforming “male” and “female” into “1” and “0” accordingly.
3. **Education Level:** This variable reflects the highest level of education an individual has completed (e.g., high school, bachelor's degree, master's degree). Higher education levels are often associated with higher earning potential. To understand more vividly about this variable, the data has been transformed as follow:

|  |  |
| --- | --- |
| **Education** | **Education level** |
| High school | 1 |
| Bachelor's | 2 |
| Master's | 3 |
| PhD | 4 |

1. **Years of Experience:** This variable measures the total number of years an individual has worked in their field. More experience typically correlates with higher salaries as individuals develop skills and expertise.
2. **Job Level:** This variable represents the hierarchical level of the individual's current position (e.g., entry-level, mid-level, senior-level). Higher job levels are usually associated with higher salaries due to increased responsibilities and leadership roles. After seeking help from a colleague working in HR department, the author has transformed the data as follow:

|  |  |
| --- | --- |
| **Job level** | **Level No.** |
| Junior | 1 |
| Mid-level | 2 |
| Senior | 3 |
| Junior Specialist | 4 |
| Specialist | 5 |
| Senior Specialist | 6 |
| Junior Manager | 7 |
| Manager | 8 |
| Senior Manager | 9 |
| Chief Officer | 10 |

1. **Job Field:** This categorical variable denotes the specific industry or field in which the individual works (e.g., technology, finance, healthcare). Different job fields often have varying salary ranges due to industry-specific factors. We have the following digitalized field table, this table was randomized as the author didn’t consider any specific field to be superior to the others:

|  |  |
| --- | --- |
| **Field** | **Field No.** |
| IT | 1 |
| Data | 2 |
| Other | 3 |
| Sales | 4 |
| Marketing | 5 |
| Product | 6 |
| Scientist | 7 |
| Human Resources | 8 |
| Project Management | 9 |
| Customer Service | 10 |
| Operations | 11 |
| Finance | 12 |
| Consultant | 13 |
| Design | 14 |
| Creative | 15 |
| Admin | 16 |
| Research | 17 |
| Business Development | 18 |
| Driver | 19 |
| Reception | 20 |

1. **Estimation Method**
2. **Exploratory Data Analysis (EDA)**

In order to find underlying patterns, trends, and relationships in data, a critical phase in the data analysis process is called exploratory data analysis, or EDA. This involves summarizing and displaying data. EDA aids in determining the structure of the data, spotting abnormalities or outliers, and evaluating the data's quality. Through the use of a variety of statistical and graphical methods, including correlation matrices, scatter plots, and histograms, EDA offers insights that guide further research and modeling. Making educated conclusions and ensuring that the data is ready for more sophisticated statistical techniques or machine learning algorithms depend on this preliminary investigation.

1. **Correlation Analysis**

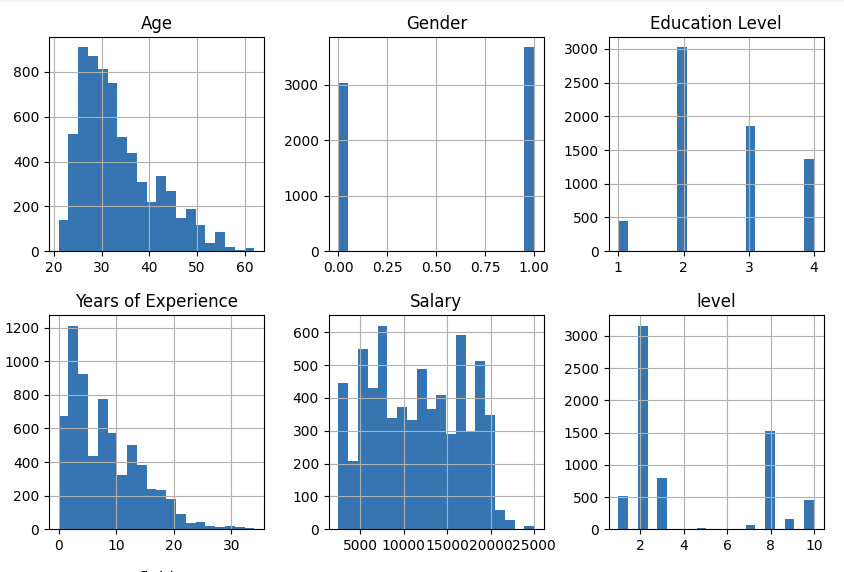
Correlation analysis is a statistical technique used to measure and describe the strength and direction of the relationship between two or more variables. Researchers can figure out how closely variables move together by computing correlation coefficients, such as Pearson's correlation coefficient. If two variables have a positive correlation, it means that when one rises, the other also tends to rise, and if they have a negative correlation, it means that when one rises, the other tends to fall. Finding correlations between variables is aided by correlation analysis, and this knowledge can be vital for deciphering data patterns, formulating predictions, and directing future study. Though two factors may be related, one does not always cause the other, it is crucial to remember that correlation does not imply causation.

1. **Linear Regression Models**

Linear regression models are a fundamental statistical method used to analyze and predict the relationship between a dependent variable and one or more independent variables. The basic concept is to fit a linear equation to the observed data such that dependent variable values can be predicted from independent variable values. This relationship is represented by the equation Y=β0+β1x+ϵ in a basic linear regression model, where x is the independent variable, Y is the dependent variable, β0 is the intercept, β1, β1 is the slope, and ϵ is the error term. This idea is expanded to incorporate numerous independent variables using multiple linear regression. Although linear regression models presume a linear relationship and may not perform well with non-linear data, they are frequently used for predicting, analyzing relationships between variables, and making data-driven decisions.

1. **RESEARCH RESULTS AND DISCUSSION**
2. **Descriptive statistics of variables**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Age** | **Gender** | **Education Level** | **Years of Experience** | **Salary** | **Level** |
| **COUNT** | 6,698 | 6,698 | 6,698 | 6,698 | 6,698 | 6,698 |
| **MEAN** | 33.62 | 0.50 | 2.6 | 6.06 | 11,535.85 | 4.18 |
| **STD** | 7.62 | 0.50 | 0.88 | 6.06 | 5,274.01 | 3.07 |
| **25%** | 28 | - | 2 | 3 | 7,000 | 2 |
| **50%** | 32 | 1 | 2 | 7 | 11,500 | 2 |
| **75%** | 38 | 1 | 3 | 12 | 16,000 | 8 |
| **MAX** | 62 | 1 | 4 | 34 | 25,000 | 10 |



The mean and standard deviation values for each variable provide a comprehensive view of the dataset's central tendencies and variability.

**Age**: The mean age of 33.62 years indicates that individuals are generally in the early stages of their careers. The standard deviation of 7.62 years shows a moderate spread around this average, suggesting a diverse age range within the dataset, with some individuals being significantly younger or older.

**Gender**: The mean value of 0.50 reflects an equal distribution between the gender categories represented. The standard deviation of 0.50 indicates that this binary variable has balanced values, meaning there is an even split between the two categories, with no significant skew towards one gender.

**Education** **Level**: The mean education level of 2.6 suggests that, on average, individuals have attained a level of education that is between bachelor's and master’s degree. The standard deviation of 0.88 indicates variability in educational backgrounds, with some individuals having lower or higher levels of education compared to the average.

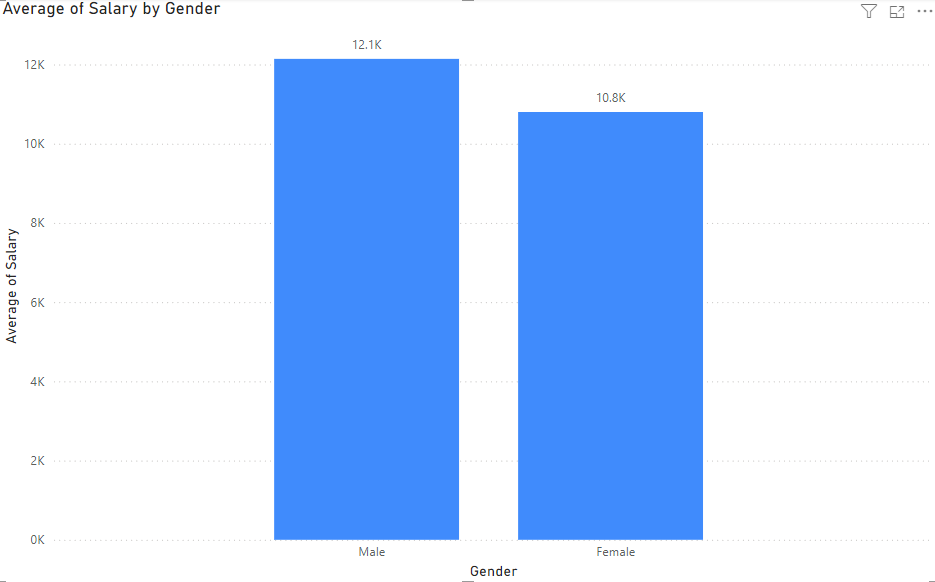
**Years of Experience**: With a mean of 6.06 years, individuals in the dataset typically have a moderate amount of experience. The standard deviation of 6.06 years reveals a wide range of experience levels, from those just starting their careers to those with extensive experience, indicating substantial variability in this aspect.

**Salary**: The average salary of $11,535.85 reflects a moderate income level. The standard deviation of $5,274.01 indicates significant variability in salaries, meaning there are notable differences in earnings across the dataset, with some individuals earning considerably more or less than the average.

**Job Level**: The mean job level of 4.18 suggests that individuals generally occupy mid-level positions. The standard deviation of 3.07 shows a broad range of job levels, from entry-level to senior positions, highlighting the diversity in roles and responsibilities within the dataset.

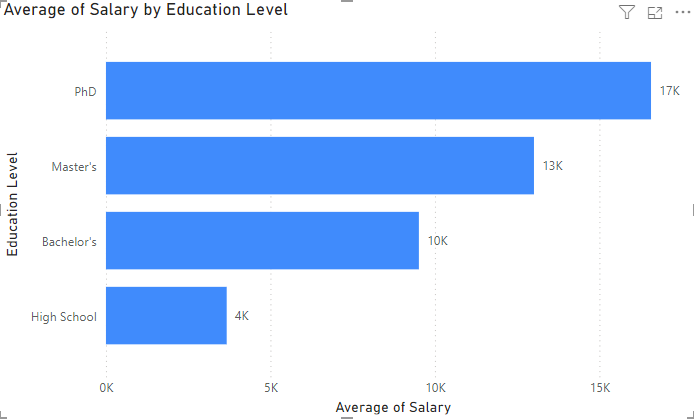
These values help to understand not only the central tendencies but also the spread and variability within the dataset, providing a clearer picture of the population being analyzed.

1. **Overview findings**
2. **Relationship between Salary and Genders**



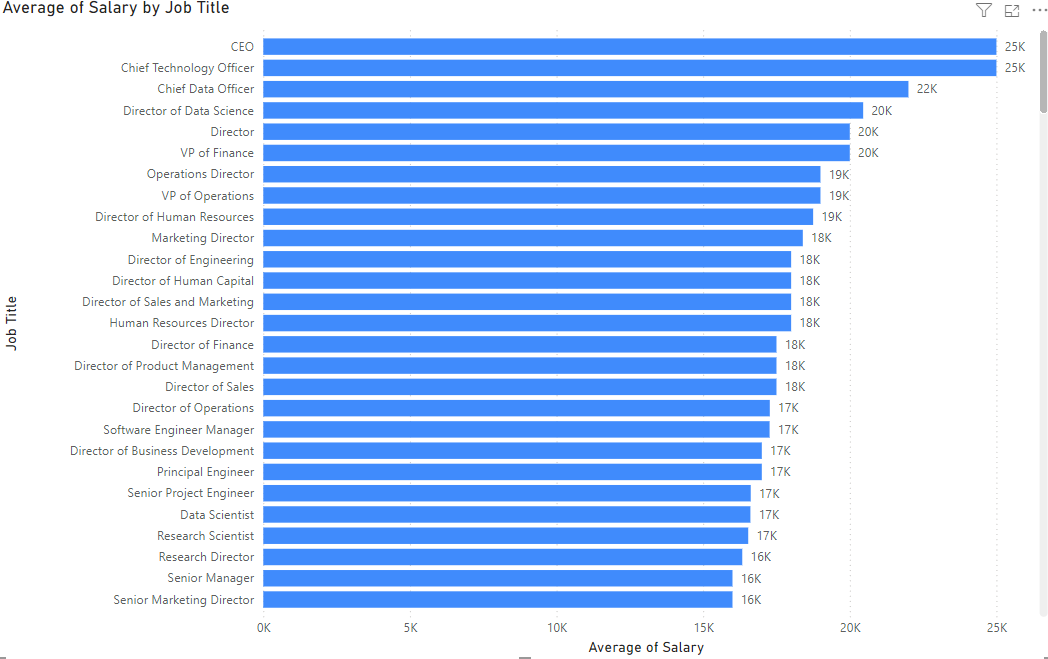
The chart shows that on average, males earn about $12,100, which is higher than the average salary for females at $10,800. This indicates a gender pay gap, where males are earning more on average than their female counterparts. The difference, though not extreme, suggests that gender may play a role in salary variations across the workforce.

1. **Relationship bewteen Salary and Education**



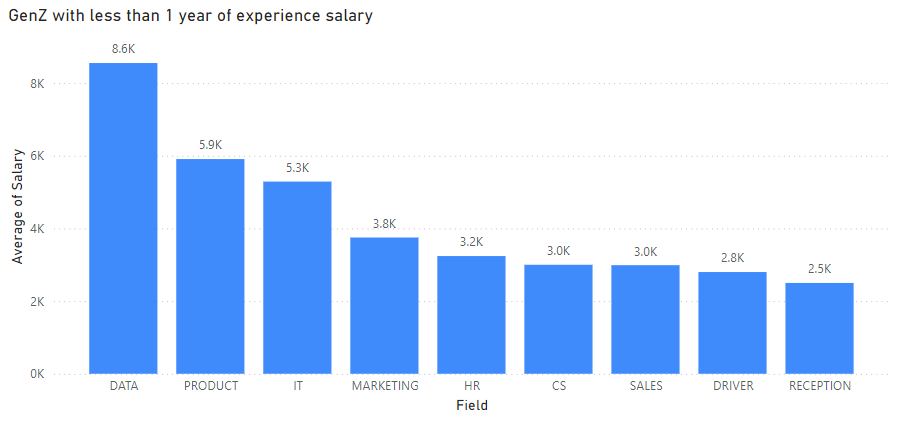
Education level has a clear impact on salary, with those holding a PhD earning the highest average salary of around $16,600. Master's degree holders earn about $13,000, while those with a Bachelor's degree average around $9,500. Individuals with only a high school education have the lowest average salary, at approximately $3,700, highlighting the strong correlation between higher education and increased earning potential.

1. **Highest-paying Jobs**



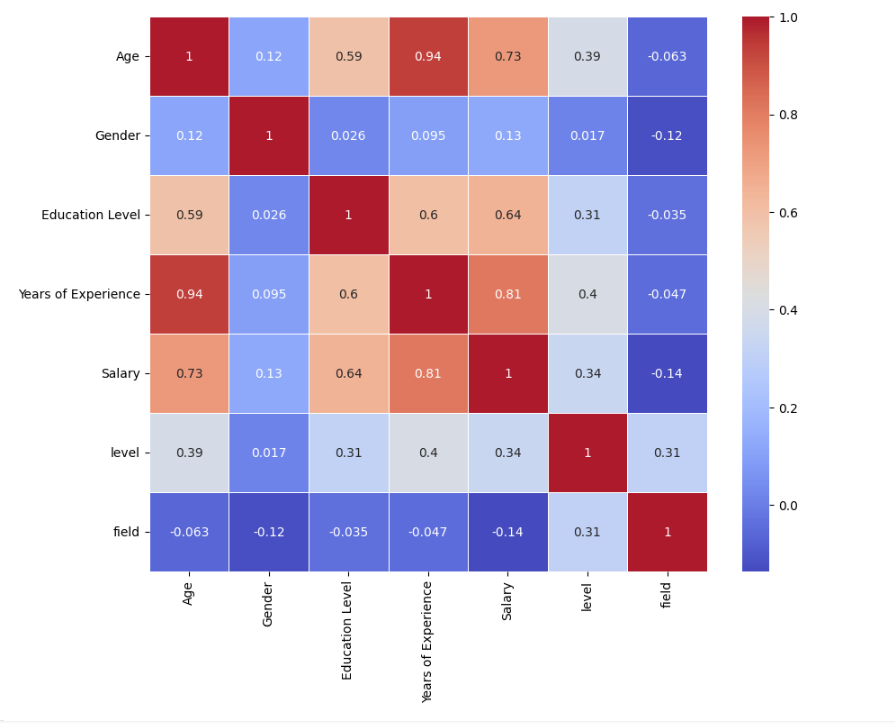
The graphic shows the average yearly salary for different job categories. CEOs and chief technology officers make the most money, with an average salary of $25,000 each. Chief Data Officers come in second, earning $22,000 on average. Director of Data Science, Director, and VP positions are among the other high-paying occupations; their typical salary ranges from $20,000 to $19,000. The pay scale is clearly arranged in a hierarchy according to the chart, with executive and director-level positions paying far more than other stated roles.

1. **Gen Z’s ideal occupation**



The bar chart illustrates the average salaries for Generation Z employees with less than one year of experience across different fields. Data roles stand out with the highest average salary of $8.6K, followed by Product and IT at $5.9K and $5.3K, respectively. Fields such as Marketing and HR offer moderate salaries around $3.8K and $3.2K, while roles in Customer Service, Sales, Driver, and Reception are at the lower end, with salaries ranging from $2.5K to $3.0K. This highlights how choosing a career in tech-related fields can significantly impact early earnings for Gen Z professionals.

1. **Correlation Analysis**



1. **Age**

* **Education Level (0.59):** A moderate positive correlation suggests that older individuals tend to have higher education levels, which may reflect longer periods of education or advanced degrees.
* **Years of Experience (0.94):** A very strong positive correlation indicates that age is highly associated with years of experience. Older individuals generally have more years of work experience.
* **Salary (0.73):** A strong positive correlation suggests that older individuals tend to earn higher salaries, likely due to greater experience and seniority.
* **Level (0.39):** A moderate positive correlation indicates that age is somewhat related to job level, with older individuals often holding higher-level positions.
* **Field (-0.06):** A very weak negative correlation implies that age has a negligible impact on the field in which individuals work.

1. **Gender:**

* **Education Level (0.03):** A very weak positive correlation suggests that gender has minimal influence on education level in this dataset.
* **Years of Experience (0.09):** A very weak positive correlation indicates a negligible relationship between gender and years of experience.
* **Salary (0.13):** A weak positive correlation implies a slight association between gender and salary, but it's not strong.
* **Level (0.02):** A very weak positive correlation shows minimal impact of gender on job level.
* **Field (-0.12):** A weak negative correlation suggests a slight association between gender and the field of work, but the impact is minimal.

1. **Education Level:**

* **Years of Experience (0.60):** A moderate positive correlation indicates that individuals with higher education levels tend to have more years of experience.
* **Salary (0.64):** A moderate to strong positive correlation suggests that higher education levels are associated with higher salaries.
* **Level (0.31):** A weak to moderate positive correlation indicates some association between education level and job level, with more education potentially leading to higher job levels.
* **Field (-0.04):** A very weak negative correlation suggests that education level has a minimal impact on the field of work.

1. **Years of Experience:**

* **Salary (0.81):** A strong positive correlation indicates that more years of experience are strongly associated with higher salaries.
* **Level (0.40):** A moderate positive correlation suggests that more experience is related to holding higher job levels.
* **Field (-0.05):** A very weak negative correlation implies that years of experience has a minimal impact on the field of work.

1. **Level:**

* **Field (0.31):** A weak to moderate positive correlation indicates that job level is somewhat related to the field of work, with certain fields potentially having higher job levels.
* **Salary (0.34):**  
  The correlation of 0.34 between Job Level and Salary is moderate. This positive correlation suggests that higher job levels are associated with higher salaries, though the relationship is not as strong as with years of experience or education because the data was conducted in a lot of enterprises, and each’s level may vary differently

1. **Field**

* The correlations with other variables are generally very weak, indicating that the field of work has minimal direct associations with age, gender, education level, years of experience, salary, and job level in this dataset.

1. **Salary**

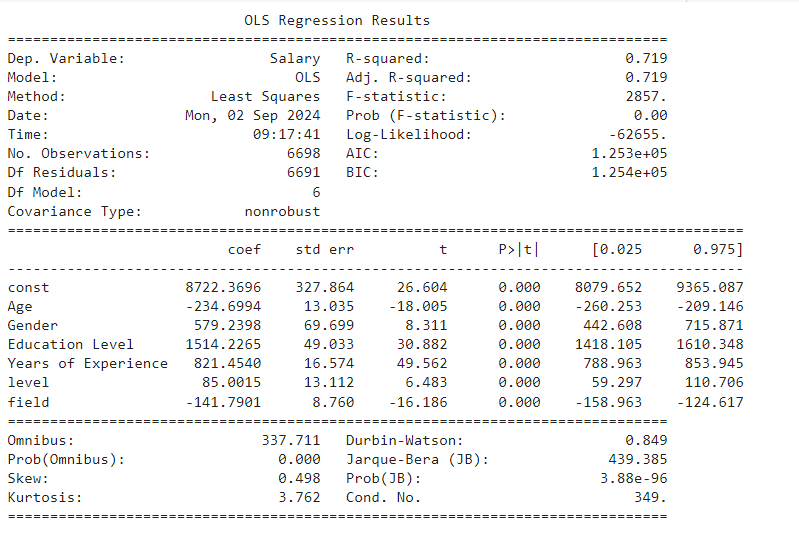
* **Age (0.73):**  
  The correlation of 0.73 between Age and Salary is quite strong and positive. This suggests that as individuals get older, their salaries tend to be higher. This relationship might be due to the accumulation of experience and seniority over time, which often leads to higher compensation.
* **Gender (0.13):**  
  The correlation of 0.13 between Gender and Salary is weak but positive. This indicates a slight association between gender and salary, suggesting that gender has a modest impact on earnings. The impact is not very strong, but it may reflect differences in salary based on gender in the dataset.
* **Education Level (0.64):**  
  A correlation of 0.64 between Education Level and Salary is moderate to strong. This positive correlation implies that individuals with higher levels of education tend to earn higher salaries. Higher education often leads to better job opportunities and higher pay.
* **Years of Experience (0.81):**  
  The correlation of 0.81 between Years of Experience and Salary is strong. This indicates a significant positive relationship, meaning that more years of experience are strongly associated with higher salaries. Experience is a key factor in determining salary, as it usually reflects skill development and seniority.
* **Level (0.34):**  
  The correlation of 0.34 between Job Level and Salary is moderate. This positive correlation suggests that higher job levels are associated with higher salaries, though the relationship is not as strong as with years of experience or education. Job levels often reflect the responsibility and seniority of a position, impacting salary.
* **Field (-0.14):**  
  The correlation of -0.14 between Field and Salary is weak and negative. This indicates a slight negative association, suggesting that certain fields might have lower average salaries compared to others. However, the effect is minor, and other factors are likely more influential in determining salary.

**Summary**

* **Strongest Positive Correlations:** **Years of Experience (0.81)** and **Education Level (0.64)** are strongly positively correlated with Salary, indicating that more experience and higher education are strongly associated with higher salaries.
* **Moderate Positive Correlations:** **Age (0.73)** and **Job Level (0.34)** show moderate positive correlations with Salary. Age has a strong positive impact, while Job Level has a moderate impact.
* **Weak Correlations:** **Gender (0.13)** and **Field (-0.14)** show weak correlations with Salary, suggesting that gender has a modest impact and field has a slight negative impact on salary.

These correlations reveal that education, experience, and age are the most significant factors affecting salary in this dataset, while job level, gender, and field have a more limited impact.

1. **Multiple Linear Regression Model**



1. **Model Summary**

* **R-squared: 0.719**  
  The model explains approximately 71.9% of the variance in the dependent variable (Salary), indicating a good fit.
* **Adj. R-squared: 0.719**  
  The adjusted R-squared value, which accounts for the number of predictors in the model, is the same as the R-squared, confirming the model's reliability in explaining the variability in salary.
* **F-statistic: 0.00**  
  This high F-statistic, with a p-value of 0.00, indicates that the overall model is statistically significant, meaning that the predictors collectively explain a significant amount of the variance in salary.

1. **Coefficients and Their Interpretations**

* **Age: -234.70**  
  For each additional year of age, the salary decreases by approximately $234.70. This negative relationship suggests that, in this model, older individuals tend to earn less, potentially due to the age-specific dynamics in the dataset.
* **Gender: 579.24**  
  Being a male is associated with an increase in salary of approximately $579.24, holding other factors constant.
* **Education Level: 1514.23**  
  Each additional unit increase in education level corresponds to an increase in salary of approximately $1514.23. This indicates a strong positive effect of higher education on salary.
* **Years of Experience: 821.45**  
  Each additional year of experience is associated with an increase in salary of approximately $821.45, highlighting the value of experience in salary determination.
* **Level: 85.00**  
  Each unit increase in job level is associated with an increase in salary of approximately $85.00, reflecting the effect of higher job levels on salary.
* **Field: -141.79**  
  Working in a particular field (assuming the variable is categorical with specific fields) is associated with a decrease in salary of approximately $141.79, suggesting that some fields may offer lower salaries compared to others.

Finally, we have the following Regression Model to predict the Expected Salary for Vietnamese Labor:

*E\_Salary = 8722.37 − 234.70\*****Age*** *+ 579.24\*****Gender*** *+ 1514.23\*****Education Level*** *+ 821.45\*****Experience*** *+ 85.00\*****Job******Level****−141.79\*****Job******Field***

For example, if you are 25 years old, male, have a Bachelor’s Degree, have 1 year of experience, want to work in mid-level tier and work in IT industry then your expected **yearly salary** would equal:

8722.37 -234.70\*25 + 579.24\*1 + 1514.23\*2 + 821.45\*1 + 85.00\*2 - 141.79\*1 = **7312.23** ($)

1. **CONCLUSION**
   1. **Factors Influencing Salary Levels**

**Age**: Our analysis reveals that as individuals age, their salary tends to decrease slightly after a certain point, which may reflect a market preference for younger employees in certain industries or roles.

**Gender**: The data indicates a gender wage gap, with males earning significantly more than females on average, highlighting ongoing disparities in the workplace.

**Education Level**: Higher education levels are strongly correlated with higher salaries, underscoring the importance of educational attainment in career advancement.

**Years of Experience**: Experience is a critical determinant of salary, with each additional year of experience leading to a substantial increase in earnings.

**Job Level**: Promotions and higher job levels contribute positively to salary, emphasizing the value of career progression.

**Job Field**: Certain job fields are associated with lower salaries, which may reflect differences in industry standards or demand for specific skills.

* 1. **Proposed Salary Prediction Model**

The proposed model accurately predicts salary based on key variables such as age, gender, education level, years of experience, job level, and job field. The model’s coefficients provide insights into how each factor influences salary, allowing for reliable salary predictions and identifying areas where interventions could be made to address disparities. Using the formula, people can predict their salary based on their own qualities. Note that gender, education level, job level and job field can be looked up at part II.2.b above:

*E\_Salary = 8722.37 − 234.70\*****Age*** *+ 579.24\*****Gender*** *+ 1514.23\*****Education Level*** *+ 821.45\*****Experience*** *+ 85.00\*****Job******Level****−141.79\*****Job******Field***

* 1. **Limitations of the research**

**Sample Size and Representation**: The dataset used may not be representative of the entire population, especially if it is drawn from a specific industry, region, or demographic group. This could limit the generalizability of the findings to other populations.

**Data Quality**: Incomplete, outdated, or inaccurate data can affect the validity of the results. Any missing or erroneous data points could skew the analysis and lead to biased conclusions. Moreover, different titles and levels from different companies may have led to the abnormal job levels’ salary

**Variable Selection:** The model includes specific variables like age, gender, education level, years of experience, job level, and job field, but it may omit other important factors such as industry-specific skills, location, economic conditions, or company size, which can also influence salary.

**Measurement Error:** Variables like education level or years of experience may not fully capture the nuances of an individual's qualifications or job performance, leading to potential measurement errors.

**Cultural and Geographic Differences:** The results may not be generalizable to different cultural or geographic contexts where salary determinants might differ due to varying societal norms or economic conditions.

* 1. **Suggestions for Salary Improvement**

**Pursue Higher Education**: Investing in further education can significantly boost earning potential, as higher education levels are strongly linked to higher salaries. Especially when it has the highest coefficient in the formula.

**Gain More Experience**: Accumulating years of relevant experience is crucial for increasing salary, making continuous skill development and career longevity important.

**Seek Promotions**: Actively pursuing higher job levels and responsibilities can lead to higher salaries, emphasizing the importance of career growth and advancement.

**Choose High-Demand Fields**: Selecting or transitioning to fields with higher salary potential can also be a strategic move for individuals looking to improve their earnings.

In conclusion, understanding the factors that influence salary, utilizing predictive models, and taking proactive steps in education, experience, and career choices are key strategies for individuals seeking to maximize their earning potential.