INTRODUCTION:

The aim of this analysis was to investigate the relationship between an employee’s age and their salary in a given company. The study focused on analyzing the available data on employee age and salary to determine whether there is a significant linear relationship between the two variables. The study used regression analysis to examine the strength and direction of the relationship between employee age and salary, and to provide insights into how age impacts salary in the company.

The findings of this study are important for understanding the factors that contribute to employee salaries and for making informed decisions regarding compensation policies. By identifying the relationship between employee age and salary, the study can help the company optimize its compensation strategies and better align employee salaries with their experience and performance. The results can also inform discussions around diversity and inclusion, as the study can shed light on potential disparities in compensation based on age.

Overall, the regression analysis provided valuable insights into the factors influencing employee salaries in the company and can help drive more informed decision-making.

DATA AND METHODS:

The data used in this study was obtained from an online database and included information on 1000 employees in a given company. The dataset included variables such as age, salary, job title, department, business unit, and gender. The data was cleaned to remove any duplicates or missing values and ensure consistency and accuracy.

The study employed regression analysis to examine the relationship between employee age and salary. Specifically, a simple linear regression model was used, with age as the independent variable and salary as the dependent variable. The model was fitted to the data using least squares method to estimate the regression coefficients and predict the expected salary for each age.

Assumptions:

Several assumptions were made during the analysis, including:

1. Linearity. The relationship between age and salary was assumed to be linear, meaning that changes in salary were assumed to be proportional to changes in age.
2. Independence. It was assumed that the observations in the dataset were independent of each other, meaning that one observation did not influence the other.
3. Normality. The errors in the regression model were assumed to be normally distributed, meaning that the residuals followed a normal distribution.
4. Homoscedasticity. It was assumed that the variance of the residuals was constant across all levels of age, meaning that the spread of the residuals was the same for all ages.

These assumptions were tested and verified using various diagnostic tools, including residual plots and tests for normality and homoscedasticity. Overall, the assumptions were made during the analysis were necessary for ensuring the accuracy and validity of the results.

RESULTS:

The regression model indicated a significant relationship between the age of an employee and their salary, with a p-value of 7.88e-15, indication a very low probability of observing such a relationship by chance. The regression coefficient for age was 112944, indicating that on average, for each additional year in age, the employee’s salary increases by $112944.

The R-squared value of 0.0003 suggests that 0.03% of the variation in the dependent variable (salary) can be explained by the independent variable (age). This suggests that age may not be a significant predictor of salary in this dataset.

Additionally, the analysis revealed a few outliers, indicating some extreme values in the dataset. These outliers may have had a significant impact on the regression model’s results. Therefore, it may be necessary to investigate these outliers further to determine whether they should be excluded or included in the analysis.

However, it is important to keep in mind the limitations of the analysis, including potential confounding variables that were not accounted for, and the possibility of measurement error or other data quality issues. It may be necessary to gather additional data or use alternative analysis methods to fully explore the relationship between age and salary in this population.

INTERPRETATION:

Based on the results obtained from the regression analysis, it was found that there is a statistically significant relationship between an employee’s age and their salary (p-value < 0.05). The regression equation that was obtained (y = -81.487 + 112944) shows a positive coefficient for age, indicating that as an employee’s age increases, their salary tends to increase as well.

However, the R-squared value of 0.0003 indicates that only a very small proportion of the variation in salaries can be explained by age alone. This means that other factors not included in the analysis could be influencing salaries as well.

It is important to note that these results are bases on certain assumptions such as linearity, normality, and independence of data. Violation of these assumptions can affect the validity of the results. Therefore, caution should be exercised when interpreting the results and making conclusions.

CONCLUSION:

Bases on the analysis conducted on the relationship between employee age and the salary, the following conclusion can be drawn:

The regression analysis revealed a statistically significant relationship between employee age and salary (p-value < 0.05), indicating that age has an impact on employee salaries within the given dataset. The regression coefficient for age represents the average change in salary associated with a one-unit increase in age.

However, it is important to note that the magnitude of the relationship between age and salary is relatively small, as indicated by the low R-squared value. This suggests that age alone does not explain a substantial portion of the variation in employee salaries. Other factors, such as education level, job experience, or performance, might also influence salary levels and should be considered in future analyses.

Furthermore, it is essential to acknowledge the limitations of the analysis. The dataset included a random selection of 1000 employees from a larger population of the company. While efforts were made to ensure the representativeness of the sample, the findings might not generalize to the entire employee population. Additionally, other unmeasured variables, such as job performance metrics or market conditions, could have an influence on salaries.

Based on these results, it is recommended to conduct further research and analysis, considering additional variables that may impact employee salaries. This could provide a more comprehensive understanding of the determinants of salaries within the company. Additionally, exploring interactions between age and other variables might uncover more nuanced insights into the relationship between age and salary.

In conclusion, while age was found to have a statistically significant relationship with employee salary, it should be considered alongside other factors when making decisions regarding compensation policies, employee development, and career progression within the company

APPENDICES:

1. Scatter plot of the data points: This graph shows the relationship between the employee's age and their salary.
2. Residual plot: This graph shows the residuals (the difference between the actual values and predicted values) plotted against the independent variable (age). This graph helps to check the assumption of homoscedasticity.
3. Histogram of residuals: This graph shows the distribution of the residuals. This graph helps to check the assumption of normality.
4. QQ plot of residuals: This graph helps to check the assumption of normality.

5.Box plot of salary: This graph shows the distribution of the salary.