1. **Introduction:**

Getting a good salary is very important for any employee, for this there are many factors plays an important role, in that employee experience and degree of study is more important, here we have a data of employees having salary, and other variables like years employed, total experience, department etc.

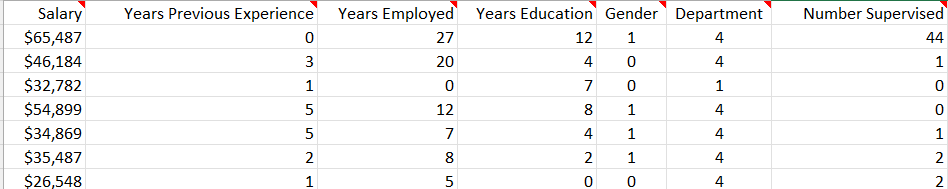
We performed descriptive analysis on these variables how salary and years employed variability happened, what is the central tendency of both variables? Any outliers we can analyse? All these questions answer in by descriptive statistics.

Later we performed regression analysis weather these variables are correlate each other. If it is how much percentage? And what is the variability in one variable explained by another variable all these we analysed using regression.

Finally, we draw some important conclusions, on our analysis is really years employed impacted salary variable. Or any other variable impacted? How can we treat other variables in future studies what insights we can draw from descriptive analysis and regression, at last we conclude this report by our understanding of analysis?

**2.0. Descriptive analysis:**

Here we have a data regarding of employee’s professional summary, having 46 entries with 8 columns such as salary, years previous experience, years employed, years education etc. The data does not contain any missing values, and it has all are numerical data type.



**Figure 1. Sample employee’s data**

Graphical user interface, text, application

Description automatically generated **Figure 2. Data types of employee’s data**

**2.1 Salary:**

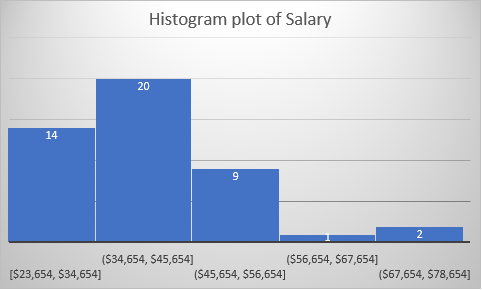
Salary is a numerical variable in employee’s data which provides the salary of employee, using descriptive statistics in excel will get the clear idea about salary data of employee’s how the central tendency and variance happens.

Graphical user interface, application, table, Excel

Description automatically generated

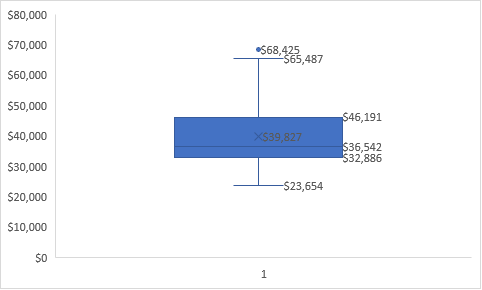
**Figure 3. Descriptive statistics of Salary**

Figure (3) shows the descriptive statistics of salary, the mean salary is 39827.39, minimum salary is 23654, maximum salary is 69246, and the median salary is 36541 that means out of 46 employees 50% percentile employees are having salary less than 36541.5.



**Figure 4. Histogram of Salary data**

Figure (4) shows the histogram plot of salary data, the data has skewed slightly in positive side, there are 3 members are having salary more than 56k pounds, 20 members having salary between 34k pounds to 45k pounds.

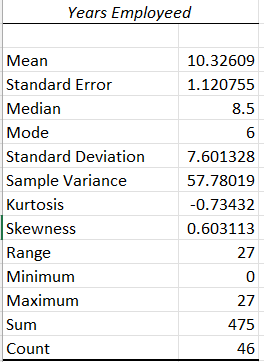


**Figure 5. Box plot of Salary**

Figure (5) shows the box plot of salary data, in entire data it considers as only one outlier point at 68,425 dollars, 75% percentile value at 46,191 dollars, 25% percentile value at 32,886 dollars.

**2.2 Years Employed:**

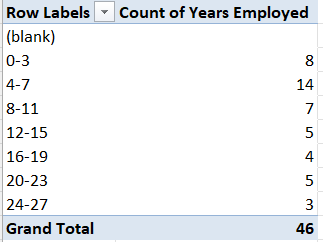
Years employed is a numerical variable provides the details about number of years working in the same organization, descriptive analysis gives clear idea about how data floating according all the employees experience, this will provide employee attrition rate also.



**Figure 6. Descriptive statistics of Years Employed**

2

Graphical user interface, table

Description automatically generated

**Figure 7. Pivot table of years of employed**

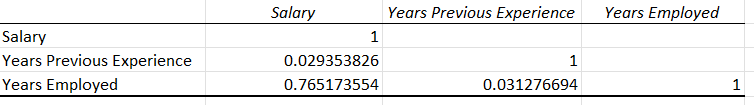
Figure (7) shows the pivot table of years of employed variable, below 7 years employees are total 22 members, 14 members are having 4 to 7 years of service which is 30% of total employees, between 8 and 11 years 7 members are there which is 15% of total employees, 8 employees having more than 20 years of service which is 17% of total employees.

**2.3 Multivariate analysis between Salary and Years of employed:**

We are much interested about how salary and years of employed varies together,

**Figure 8. Scatter plot between Salary and Years of employed**

Figure (8) shows scatter plot between Salary and Years of employed, there is some positive relation between these two variables, as salary increases years of employed also increase vice versa.



**Figure 9. Correlation between Salary and Years of employed**

Figure (9) shows the correlation between Salary and Years employed it is a justification of Figure (8), the value in 0.76 which means these two variables are highly positively correlated.

**3.0. Regression Analysis:**

Regression analysis is a way of mathematically sorting out which of those variables does indeed have an impact. It answers the questions: Which factors matter most? Which can we ignore? How do those factors interact with each other? And, perhaps most importantly, how certain are we about all of these factors?

In regression analysis, those factors are called variables. You have your dependent variable — the main factor that you’re trying to understand or predict. In Redman’s example above, the dependent variable is monthly sales. And then you have your independent variables — the factors you suspect have an impact on your dependent variable.

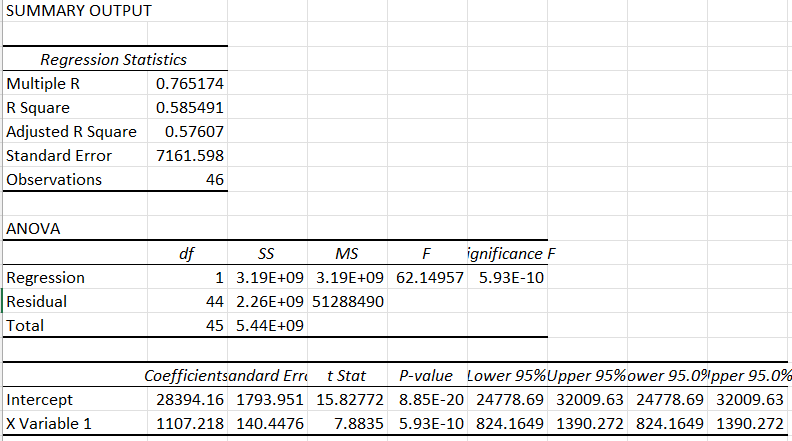
**3.1 Regression analysis between Salary and Years employed:**

We are much interested to see the regression analysis between salary and years employed, we already seen both variables are having Pearson correlation (r) is 0.765, so there could be explainable between these variables.

**Figure 10. Regression line between Salary and Years employed**

Figure (10) shows the regression line between Salary and Years employed the equation shown by

With R-square value 0.5855, the same we can see by regression summary



**Figure 11. Regression summary**

Figure (11) shows the regression summary between Salary and Years employed, salary consider as dependent variable, Years employed consider as independent variable so the equation becomes

**3.1.1 Interpretation:**

Here the intercept value is 28394.16 dollars which means if the years employed variable value zero, the salary will remain same amount of 28395.16, x-coefficient value is 1107.218 which means if you are increasing years employed by one unit then salary will increase by 1107.218 times, which indicates a positive relationship.

**3.1.2. R-squared:**

R-Squared (R² or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (the goodness of fit).

Here R-squared 0.58, which means 58% percentage of variance in the salary explained by years employed.

**3.1.3. Hypothesis testing:**

Hypothesis testing is used to confirm if our beta coefficients are significant in a linear regression model. Every time we run the linear regression model, we test if the line is significant or not by checking if the coefficient is significant.

We have two hypotheses:

* Null Hypothesis (Ho)
* Alternate Hypothesis (Ha)

Ha: There is relation between Salary and Years employed

Ho: There is no relation between Salary and Years employed

**3.1.4. P-value:**

So here P-value is 5.93E-10 which is lower than 0.05 so we reject the null hypothesis, which concludes there is a relation between salary and years employed.

**4.0 Managerial Interpretations and Implications:**

Here we performed analysis of employee salary data, to get a good salary there many factors place a role. Company, total experience, how many years stays in a particular organization, college and degree of study.

We studied how salary variable varying across the data in an organization, there are only few employees getting highest salary in the organization, many employees’ getting salary below 50,000 dollars, then we checked how salary variable impacted by years employed in an organization, which shows those are staying more years they are getting highest salary. Which is obvious true if you are stay more time in a company, you will get good hike, the maximum years stay is up to 27 years, an average stay year around 10 years, this indicates there is less attrition rate.

We performed regression analysis between salary and years employed which shows a string relation ship with R-square 0.58, R-Squared (R² or the coefficient of determination) is a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable, and Pearson correlation 0.75, Correlation is a statistic that measures the degree to which two variables in relation each other, by using Pearson correlation factor (r ) we have some values to measure the relation. R values between [-1, +1] where -1 means highly negative correlated, +1 means highly positive correlated

P-Value is defined as the most important step to accept or reject a null hypothesis. Since it tests the null hypothesis that its coefficient turns out to be zero i.e., for a lower value of the p-value (<0.05) the null hypothesis can be rejected otherwise null hypothesis will hold.

**5.0 Conclusion:**

In this report we showed what factors are impacted for an employee salary, we performed descriptive and regression analysis draw some good insights like we got a significant variable year employed has strong relation with salary. In this report we drew bar, histogram plots also added pivot tables to understand the descriptive statistics, in regression we provided regression summary with regression equation interpretation R-square, with all statistical values like p, t, and F-values.

There are many other variables also there like year of experience, department in future we need to perform analysis on these variables to show a better result on variability in salary of employee.