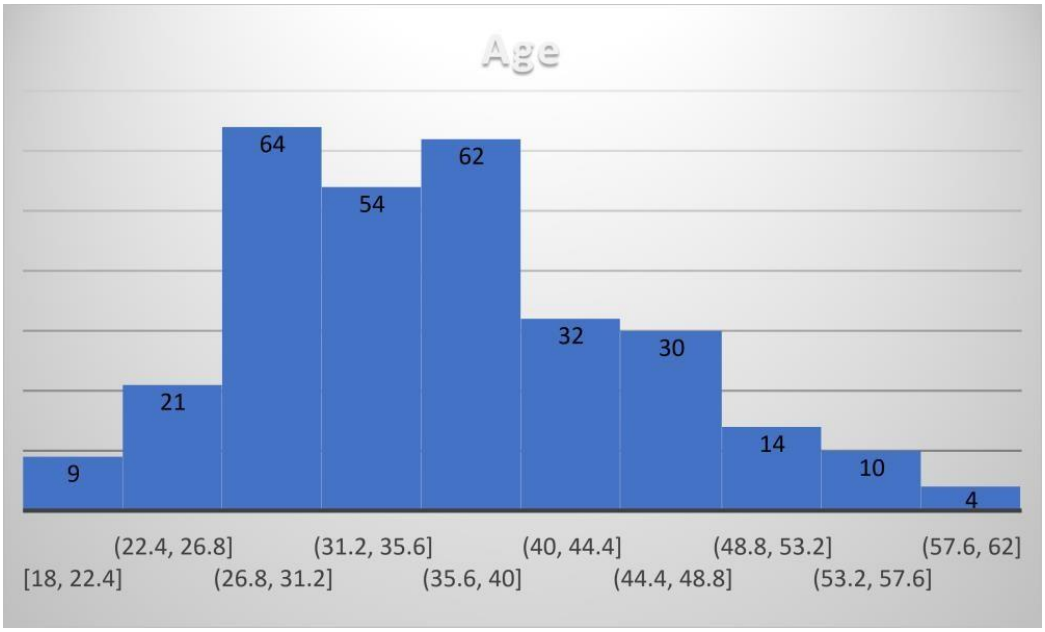
Predicting Factors affecting Employee Performance with Applications of Data Mining Classification

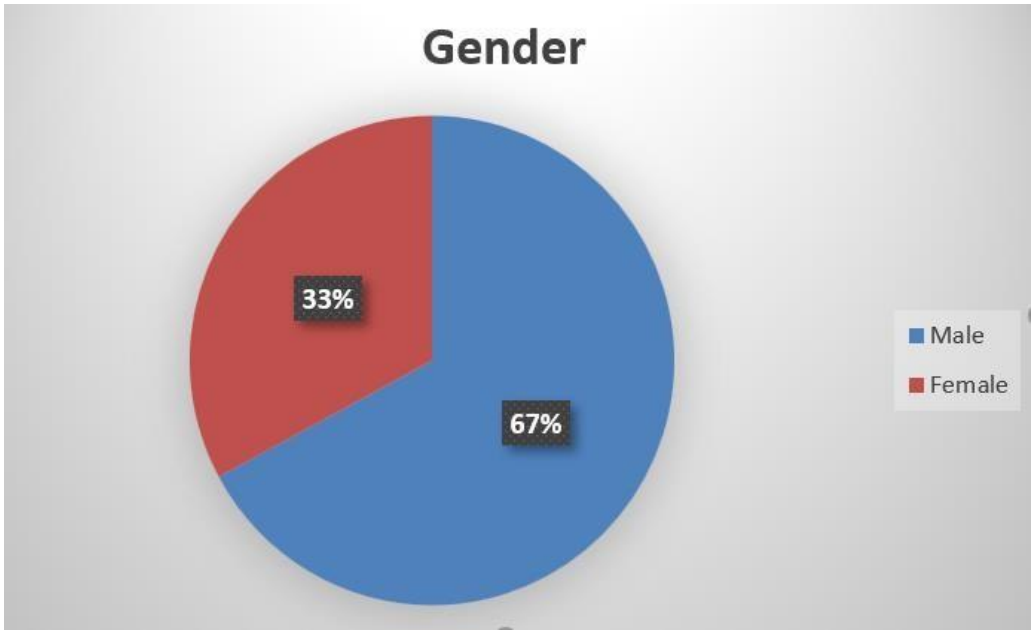
DATA ANALYSIS :



S. NO.	AGE	NO. OF RESPODENTS
1	18-22.4	9
2	22.4-26.8	21
3	26.8-31.2	64
4	31.2-35.6	54
5	35.6-40	62
6	40-44.4	32
7	44.4-48.8	30
8	48.4-53.2	14
9	53.2-57.6	10
10	57.6-62	4
	Total	300

INTERPRETATIONS

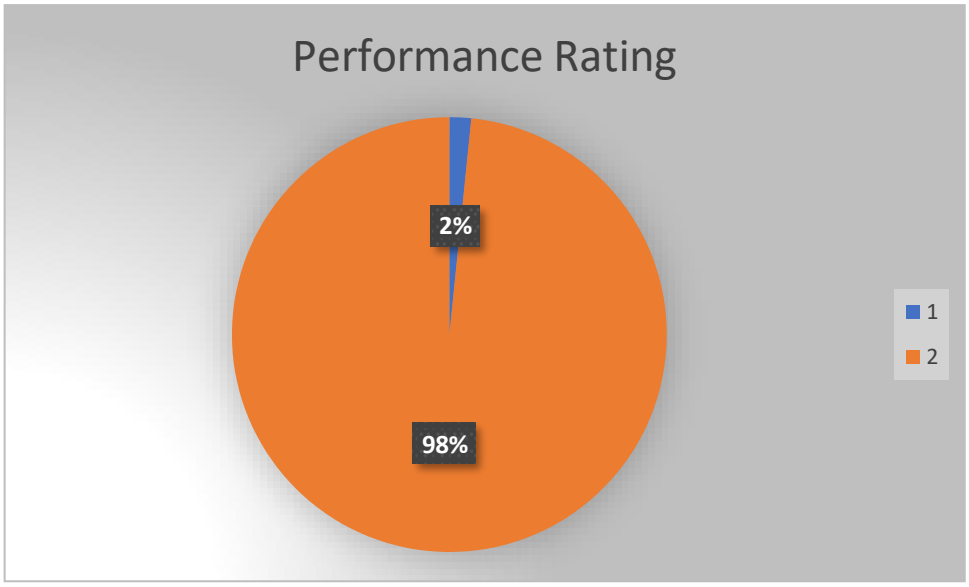
Above table shows that larger proportion of the respondents i.e, 64 were belonging to the 26.8-31.2 years range of the age followed by 62 no. of respondents are between 35.6-40 years range of age followed by 54 no. of respondents are between 31.2-35.6 years range of age followed by 32 no. of respondents are between 40-44.4 years range of age followed by 30 no. of respondents are between 44.4-48.8 years range of age followed by 21 no. of respondents are between 22.4-26.8 years range of age followed by 14 no. of respondents are between 48.4-53.2 years range of age followed by 10 no. of respondents are between 53.2-57.6 years range of age followed by 4 no. of respondents are between 57.6-62 years range of age.



S. NO.	GENDER	PERCENTAGE
1	Male	33%
2	Female	67%
	Total	100%

INTERPRETATIONS

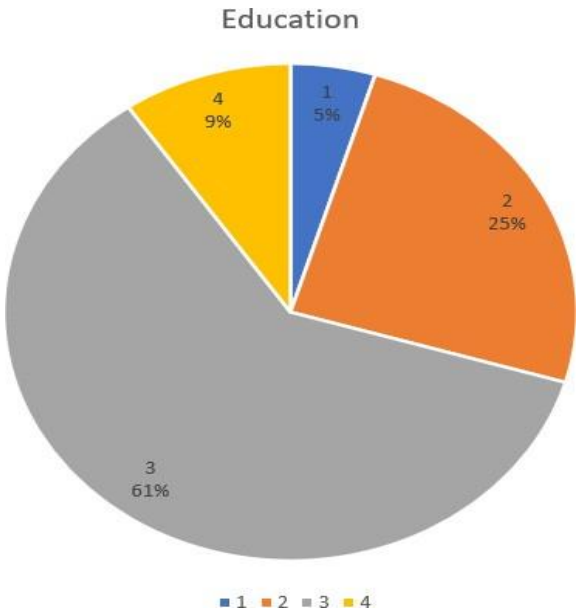
The above table shows that 33% of the employee were male in the company and 67% of the employee were female in the company so majority of the employee were female in the company.



S. NO.	PERFORMANCE RATING	PERCENTAGE
1	Good	98%
2	Bad	2%
	Total	100%

INTERPRETATIONS

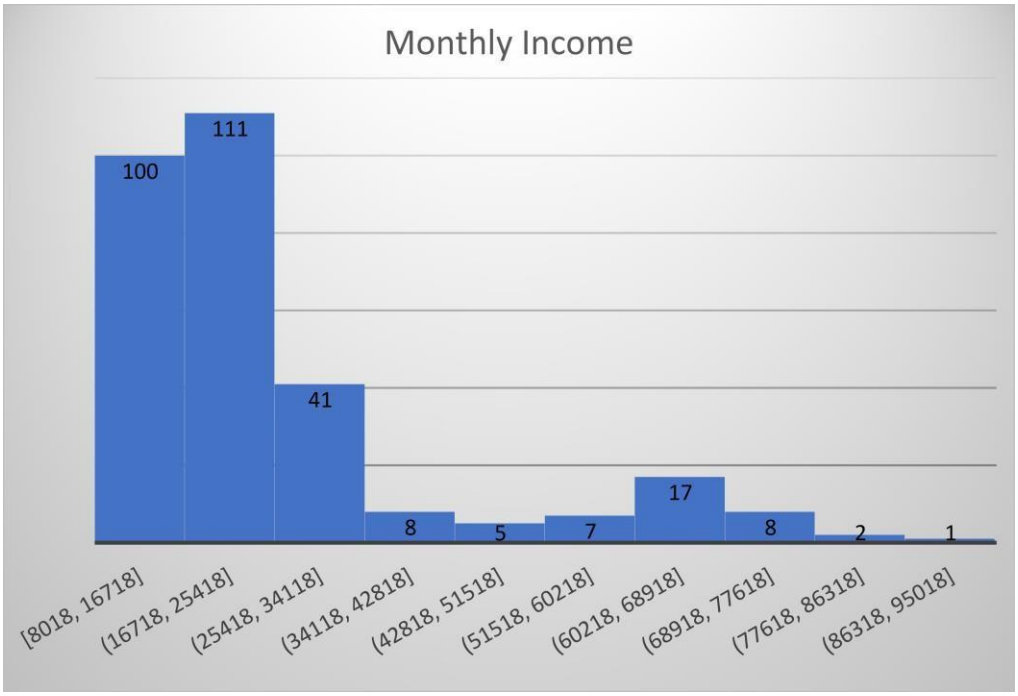
The above table shows that 98% of the respondent gave good performance rating to themselves and 2% of the respondents gave bad performance rating to themselves. So majority of the employees think that their performance is good.



S. NO.	EDUCATION	PERCENTAGE
1	Below college	5%
2	Bachelor	25%
3	Masters	61%
4	PhD	9%
	Total	100%

INTERPRETATIONS

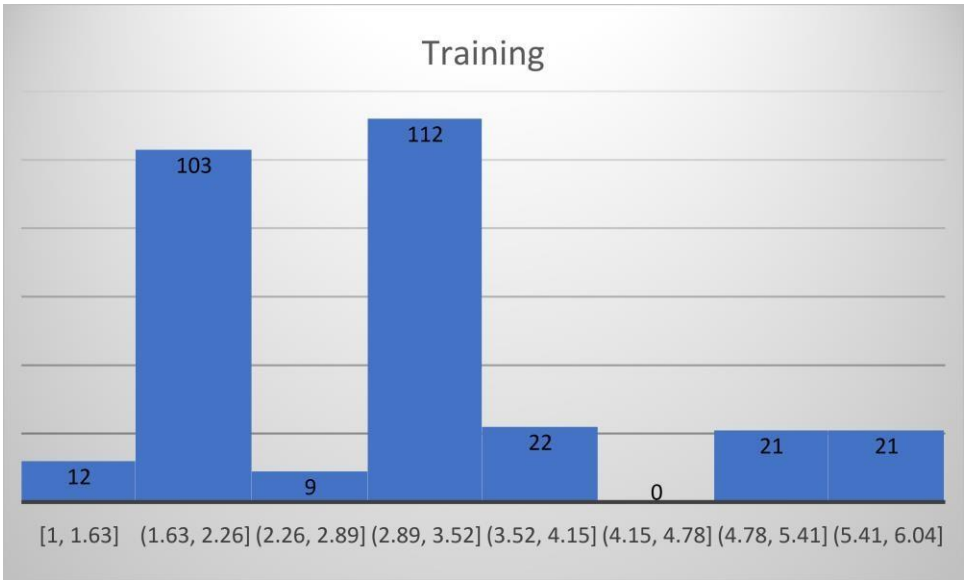
The above table shows that 5% respondents from the sample were from below college, 25% respondents from the sample were from Bachelor, 61% respondents from the sample were from Masters and 91% respondents from the sample were from PhD.



S. NO.	MONTHLY INCOME	NO. OF RESPODENTS
1	8018-16718	100
2	16718-25418	111
3	25418-34118	41
4	34118-42818	8
5	42818-51518	5
6	51518-60218	7
7	60218-68918	17
8	68919-77618	8
9	77618-86318	2
10	86318-95018	1
	Total	251

INTERPRETATIONS

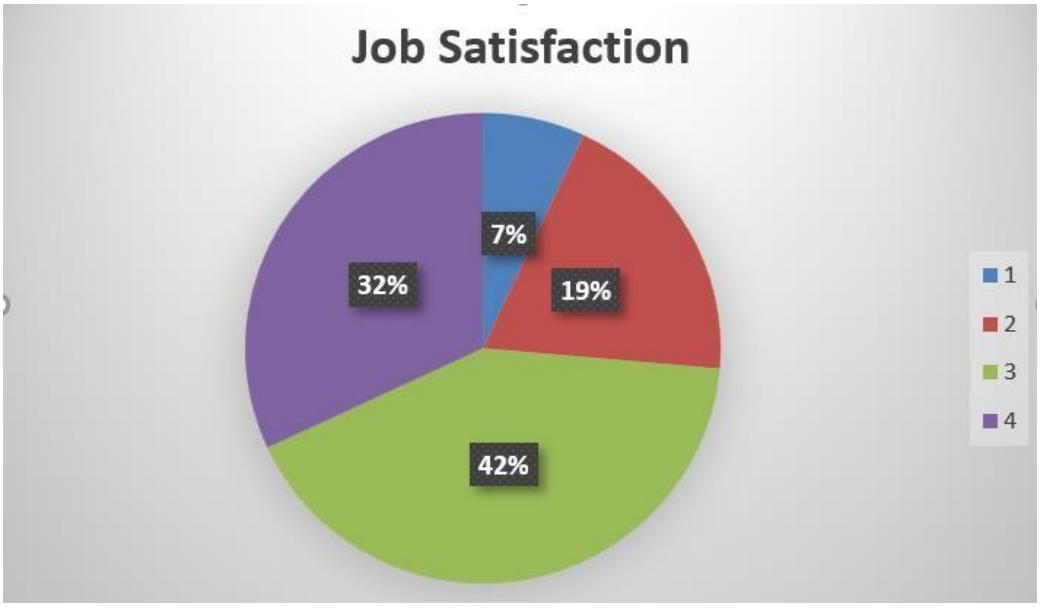
Above table shows that larger proportion i.e 111 no of respondents had 16718-25418 monthly income followed by 100 no. of respondents had 8018-16718 monthly income, followed by 41 no. of respondents had 25418-34118 monthly income, followed by 8 no. of respondents had 34118-42818 monthly income, followed by 17 no. of respondents had 60218-68918 monthly income followed by 8 no. of respondents had 34118-42818 and 68919-77618 monthly income followed by 7 no. of respondents had 51518-60218 monthly income followed by 5 no. of respondents had 42818-51518 monthly income followed by 2 no. of respondents had 77618-86318 monthly income and 1 respondent had 86318-95018 monthly income.



S. NO.	TRAINING IN HOURS	NO. OF RESPODENTS
1	1-1.63	12
2	1.63-2.26	103
3	2.26-2.89	9
4	2.89-3.52	112
5	3.52-4.15	22
6	4.15-4.78	0
7	4.78-5.41	21
8	5.41-6.04	21
	Total	300

INTERPRETATIONS

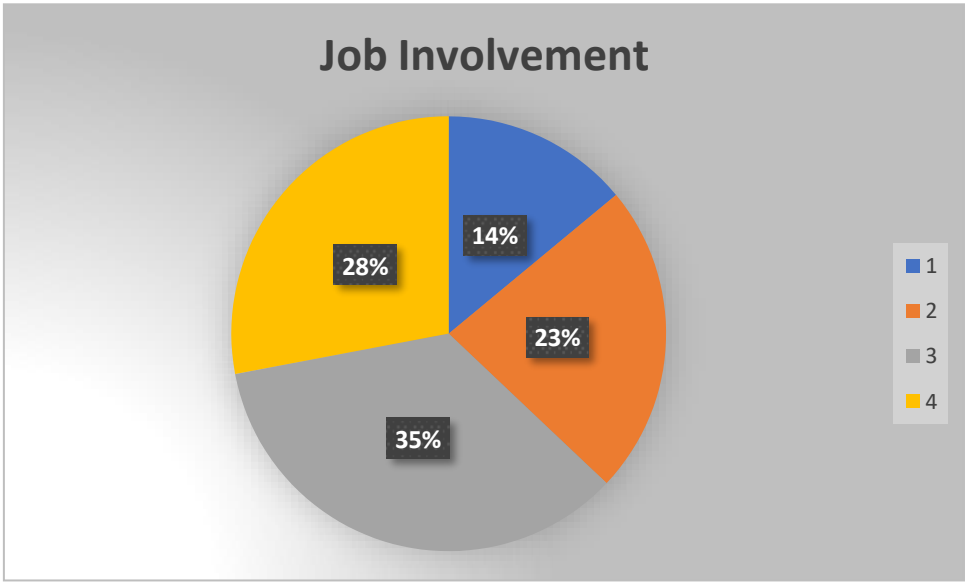
Above table shows that larger proportion i.e 103 no of respondents had 2.89-3.52 training hours followed by 103 no. of respondents had 1.63-2.26 training hours, followed by 22 no. of respondents had 3.52-4.15training hours, followed by 21 no. of respondents had 4.78-5.41 and 5.41-6.04 training hours , followed by 12 no. of respondents had 1-1.63training hours followed by 9 no. of respondents had 2.26-2.89.



S. NO.	JOB SATISFACTION	PERCENTAGE
1	Low	7%
2	Medium	19%
3	High	42%
4	Very High	32%
	Total	100%

INTERPRETATIONS

The above table clearly shows that 7% of the respondents had low job satisfaction, 19% of the respondents had medium job satisfaction, 42% of the respondents had high job satisfaction and 32% of the respondents had very high job satisfaction. So, majority of the employee were highly satisfied with their job.



S. NO.	JOB INVOLVEMENT	PERCENTAGE
1	Low	14%
2	Medium	23%
3	High	35%
4	Very High	28%
	Total	100%

INTERPRETATIONS

The above table clearly shows that 14% of the respondents had low job involvement, 23% of the respondents had medium job involvement, 35% of the respondents had high job involvement and 28% of the respondents had job involvement. So, majority of the employee were highly involved towards their job.

Logistic Regression On R

```
> View(MyData)
> rofsan=glm(PerformanceRating~Age+Training+Gender+Education+MonthlyIncome+Environme
ntSatisfaction+JobInvolvement+JobSatisfaction, data = MyData, family = "binomial")
> summary(rofsan)

Call:
glm(formula = PerformanceRating ~ Age + Training + Gender + Education +
    MonthlyIncome + EnvironmentSatisfaction + JobInvolvement +
    JobSatisfaction, family = "binomial", data = MyData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8257   0.3023   0.4724   0.6343   1.4236

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)             -3.197e+00  5.510e-01  -5.802 6.53e-09 ***
Age                      3.188e-02  1.010e-02   3.158  0.00159 **
Training                 1.574e-01  6.005e-02   2.621  0.00876 **
Gender                  -1.293e-01  1.538e-01  -0.841  0.40043
Education               -1.917e-02  7.452e-02  -0.257  0.79701
MonthlyIncome            1.034e-04  2.431e-05   4.254 2.10e-05 ***
EnvironmentSatisfaction  2.811e-01  6.741e-02   4.170 3.05e-05 ***
JobInvolvement           5.097e-01  1.002e-01   5.089 3.59e-07 ***
JobSatisfaction          2.846e-01  6.678e-02   4.262 2.02e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)
```

We have performed logistic regression between Performance and its factors of Performance to find out which factor is most prominent in getting the performance of employee. Factors like Age, Training, Monthly Income, Environment Satisfaction, Job Satisfaction, Job Involvement has p- value less than .05 means these factors have a significant impact on performance of employees. Out of these factors Job involvement and Age are most prominent in bringing out the performance of employee. Followed by Job satisfaction, Environment Satisfaction and Monthly Income.

The Logistic regression equation will be –

Performance = -3197 + 3.188(Age) +1.574(Training) + 1.034(Monthly Income) + 2.811(Environment Satisfaction) + 5.097 (Job Involvement) + 2.846 (Job Satisfaction).

If we will increase 1 unit of Job involvement by keeping others constant so it will increase the performance by 5.097 units. Followed by Age, Job satisfaction, Environment Satisfaction and Monthly Income. So, it is clear that Job involvement and Age shows most prominent effect on performance of employee.

Findings

From the above analysis we can say that the coefficient of Age, Training, Monthly Income, Environment Satisfaction, Job Involvement, Job Satisfaction are positive and the coefficient of Gender and Education are negative. At the same time, we can see that the absolute value of Job involvement coefficient is very big and the absolute value of all other's coefficient is small, which means that Job involvement may have bigger influence on the performance of employee than others and the nature of the effect is positive. When we are going to evaluate performance of employee, we can put the information of Age, Training, Monthly Income, Environment Satisfaction, Job Involvement, Job Satisfaction of the employee into this regression model, and then we can predict an estimate performance of an employee.

Conclusion

This paper is about what factors would have influenced the Employee Performance. Initially we did the preliminary analysis and chose 8 variables to predict performance. Then we used the method of Logistic regression to analyse how these factors affect the Employee Performance. After the analysis, we chose 6 variables (Age, Training, Monthly Income, Environment Satisfaction, Job Involvement, Job Satisfaction) to include in our model. We found that only Age, Training, Monthly Income, Environment Satisfaction, Job Involvement, Job Satisfaction are significant because of Sig<0.05. So we removed Gender and Education from our model. Finally, after again running the analysis we got the equation that is

Performance = -3197 + 3.188(Age) +1.574(Training) + 1.034(Monthly Income) +2.811(