

HR Analytics Case Study

SUBMISSION

Problem Statement and Analysis approach

- **Problem Statement**

A large company wants to identify the factors driving attrition of employees which they can focus on to curb the attrition of employees. Current Attrition rate is 16%

- **Analysis approach**

Understand the driving factors behind the attrition of employees, i.e. the variables which are strong indicators of attrition using EDA and to build the model to predict the attrition rate using Logistic Regression.

The Analysis Approach (Continued.....)

- Data Gathering (Different datasets provided are merged for analysis. The required variable Working Hours is derived using In and Out time provided.
- Univariate and Bivariate Analysis is performed under Exploratory Data Analysis to understand the factors/ indicators driving attrition of employees.
- Derived Variables- Numerical variables like Age, Total Working Years, Years at company, Years with Current Manager, Years since last Promotion have been identified following a strong attrition pattern when converted into bins. Hence, converted to Categorical Variables for analysis and Modelling as well.
- Dummy Variables created for building the model
- Attrition Variable converted to the factor

The Analysis Approach

- Training (70%) and Test (30%) Data set is created
- Model Building on train data
- Model Validation on test data

Final Model-Iteration5

	Estimate	Std. Error	z value
(Intercept)	-0.69852	0.49057	-1.424
Mean_Hours_Worked	0.03135	0.01003	3.126
BusinessTravel.xTravel_Frequently	1.55556	0.26196	5.938
BusinessTravel.xTravel_Rarely	0.82469	0.24557	3.358
Department.xResearch...Development	-1.05237	0.23526	-4.473
Department.xSales	-1.00053	0.24509	-4.082
JobLevel.x5	-0.58357	0.29459	-1.981
JobRole.xManager	-0.63571	0.24843	-2.559
JobRole.xManufacturing.Director	-0.48611	0.20912	-2.325
JobRole.xResearch.Director	0.60023	0.22237	2.699
MaritalStatus.xSingle	0.94508	0.11582	8.160
NumCompaniesWorked.x1	0.63298	0.15187	4.168
NumCompaniesWorked.x4	0.82126	0.23005	3.570
NumCompaniesWorked.x5	1.51858	0.25482	5.960
NumCompaniesWorked.x6	0.98129	0.26307	3.730
NumCompaniesWorked.x7	1.39675	0.24275	5.754
NumCompaniesWorked.x9	1.44970	0.28006	5.176
TrainingTimesLastYear.x6	-1.03632	0.35928	-2.884
EnvironmentSatisfaction.xLow	0.87270	0.13059	6.683
JobSatisfaction.xLow	0.59861	0.13694	4.371
JobSatisfaction.xVery.High	-0.60285	0.14070	-4.285
WorkLifeBalance.xBest	-0.96710	0.25609	-3.776
WorkLifeBalance.xBetter	-1.36392	0.20938	-6.514
WorkLifeBalance.xGood	-1.09640	0.22688	-4.832
AgeGroup	-0.60139	0.14515	-4.143
TotalWorkingYearsGroup.x12.28	-0.86116	0.22050	-3.906
TotalWorkingYearsGroup.x7.12	-0.51209	0.16846	-3.040
YearsAtCompanyGroup	-0.54873	0.16425	-3.341
YearsSinceLastPromotionGroup.xLess.than.2.Years.since.last.promotion	-0.70743	0.16912	-4.183
YearsSinceLastPromotionGroup.xNo.Promotion	-0.49739	0.14841	-3.351
YearsWithCurrManagerGroup.xLess.than.2.years	0.57966	0.21925	2.644
YearsWithCurrManagerGroup.xLess.than.6.Years	-0.48911	0.19652	-2.489

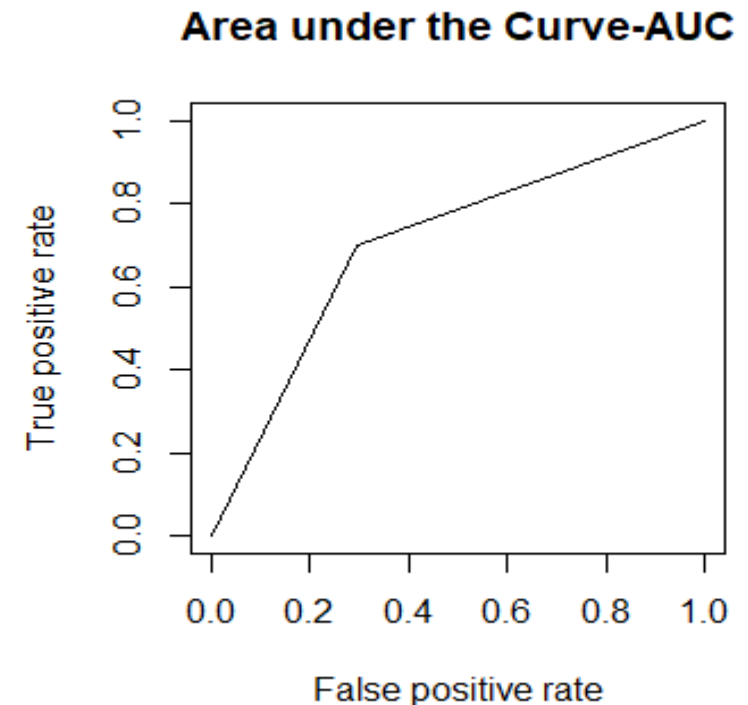
Variables in Final Model with P-Values Pr(> Z)	
(Intercept)	0.154476
Mean_Hours_Worked	0.001769 **
BusinessTravel.xTravel_Frequently	2.88e-09 ***
BusinessTravel.xTravel_Rarely	0.000784 ***
Department.xResearch...Development	7.71e-06 ***
Department.xSales	4.46e-05 ***
JobLevel.x5	0.047595 *
JobRole.xManager	0.010500 *
JobRole.xManufacturing.Director	0.020096 *
JobRole.xResearch.Director	0.006951 **
MaritalStatus.xSingle	3.36e-16 ***
NumCompaniesWorked.x1	3.08e-05 ***
NumCompaniesWorked.x4	0.000357 ***
NumCompaniesWorked.x5	2.53e-09 ***
NumCompaniesWorked.x6	0.000191 ***
NumCompaniesWorked.x7	8.73e-09 ***
NumCompaniesWorked.x9	2.26e-07 ***
TrainingTimesLastYear.x6	0.003921 **
EnvironmentSatisfaction.xLow	2.34e-11 ***
JobSatisfaction.xLow	1.23e-05 ***
JobSatisfaction.xVery.High	1.83e-05 ***
WorkLifeBalance.xBest	0.000159 ***
WorkLifeBalance.xBetter	7.31e-11 ***
WorkLifeBalance.xGood	1.35e-06 ***
AgeGroup	3.42e-05 ***
TotalWorkingYearsGroup.x12.28	9.40e-05 ***
TotalWorkingYearsGroup.x7.12	0.002367 **
YearsAtCompanyGroup	0.000835 ***
YearsSinceLastPromotionGroup.xLess.than.2.Years.since.last.promotion	2.88e-05 ***
YearsSinceLastPromotionGroup.xNo.Promotion	0.000804 ***
YearsWithCurrManagerGroup.xLess.than.2.years	0.008198 **
YearsWithCurrManagerGroup.xLess.than.6.Years	0.012816 *

Model Evaluation-Confusion Matrix & Area Under Curve

The cut off probability used to predict attrition on test data is 14.56%.
Based on this cut off, the Confusion Matrix Results are below:

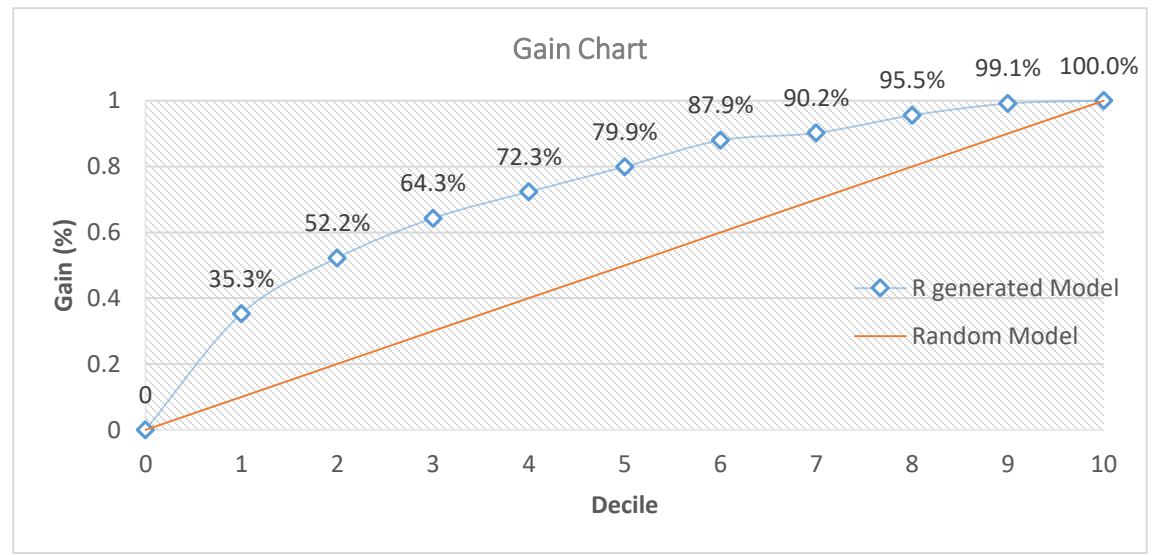
Matrix	Values	Comment
Accuracy	70.60%	It represents the accuracy of the model against Actuals
Sensitivity	70.08%	True Positive Rate
Specificity	70.70%	True Negative Rate

- The model is able to predict 70% of predictions accurately. Model is also able to predict 70% of true positives and true negatives correctly.
- Area under the curve representing the trade off between True positives and false positives is 70.09%



Gains Chart

Gain Chart				
Decile	Observations	Attrition	Cum- Churn	Gain(%Cum-Churn)
1	133	79	79	35.3%
2	132	38	117	52.2%
3	132	27	144	64.3%
4	133	18	162	72.3%
5	132	17	179	79.9%
6	132	18	197	87.9%
7	133	5	202	90.2%
8	132	12	214	95.5%
9	132	8	222	99.1%
10	132	2	224	100.0%
Total	1323	224		



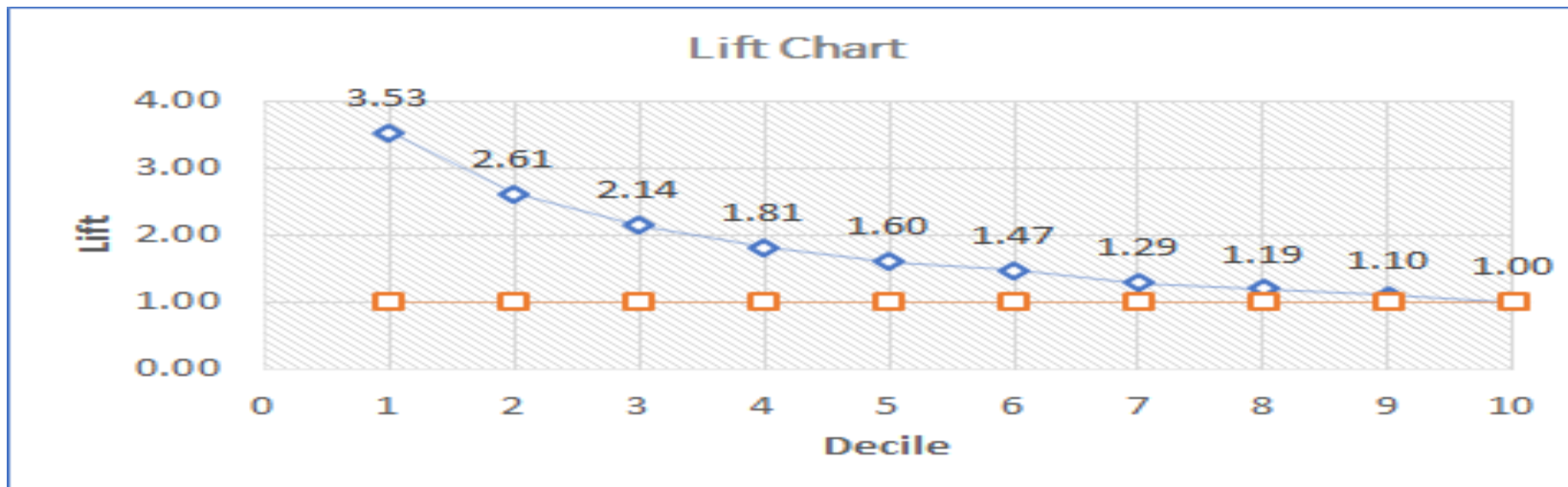
Looking at the gains chart, Model is able to predict 72.3% of Attrition in first top 4 deciles vs the random approach where only 40% attrition is predicted in top 4 deciles.



Lift Curve

Lift Chart						
Decile	Observations	Churn	Cum- Attrition	Gain(%Cum-Churn)	Gain (Random Model)	Lift
1	133	79	79	35.3%	10%	3.53
2	132	38	117	52.2%	20%	2.61
3	132	27	144	64.3%	30%	2.14
4	133	18	162	72.3%	40%	1.81
5	132	17	179	79.9%	50%	1.60
6	132	18	197	87.9%	60%	1.47
7	133	5	202	90.2%	70%	1.29
8	132	12	214	95.5%	80%	1.19
9	132	8	222	99.1%	90%	1.10
10	132	2	224	100.0%	100%	1.00
Total	1323	224				

The model is performing 3.53 times better than the random approach in 1st decile and 1.81 times better than the random approach in 4th decile.

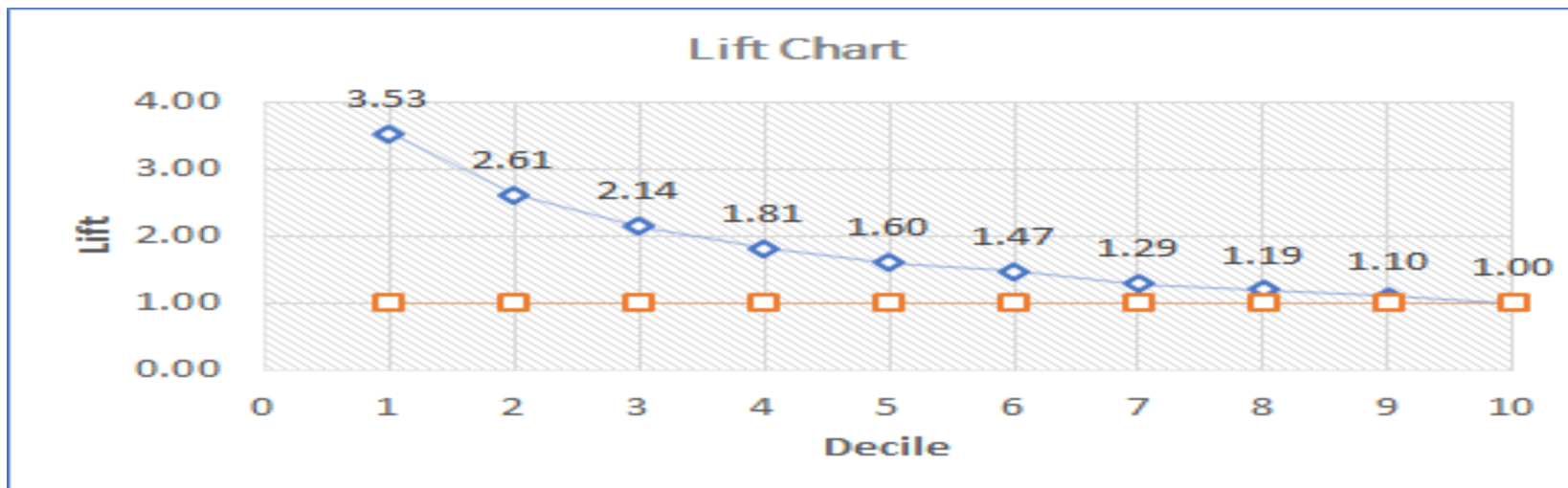




Lift Curve

Lift Chart						
Decile	Observations	Churn	Cum- Attrition	Gain(%Cum-Churn)	Gain (Random Model)	Lift
1	133	79	79	35.3%	10%	3.53
2	132	38	117	52.2%	20%	2.61
3	132	27	144	64.3%	30%	2.14
4	133	18	162	72.3%	40%	1.81
5	132	17	179	79.9%	50%	1.60
6	132	18	197	87.9%	60%	1.47
7	133	5	202	90.2%	70%	1.29
8	132	12	214	95.5%	80%	1.19
9	132	8	222	99.1%	90%	1.10
10	132	2	224	100.0%	100%	1.00
Total	1323	224				

The model is performing 3.53 times better than the random approach in 1st decile and 1.81 times better than the random approach in 4th decile.



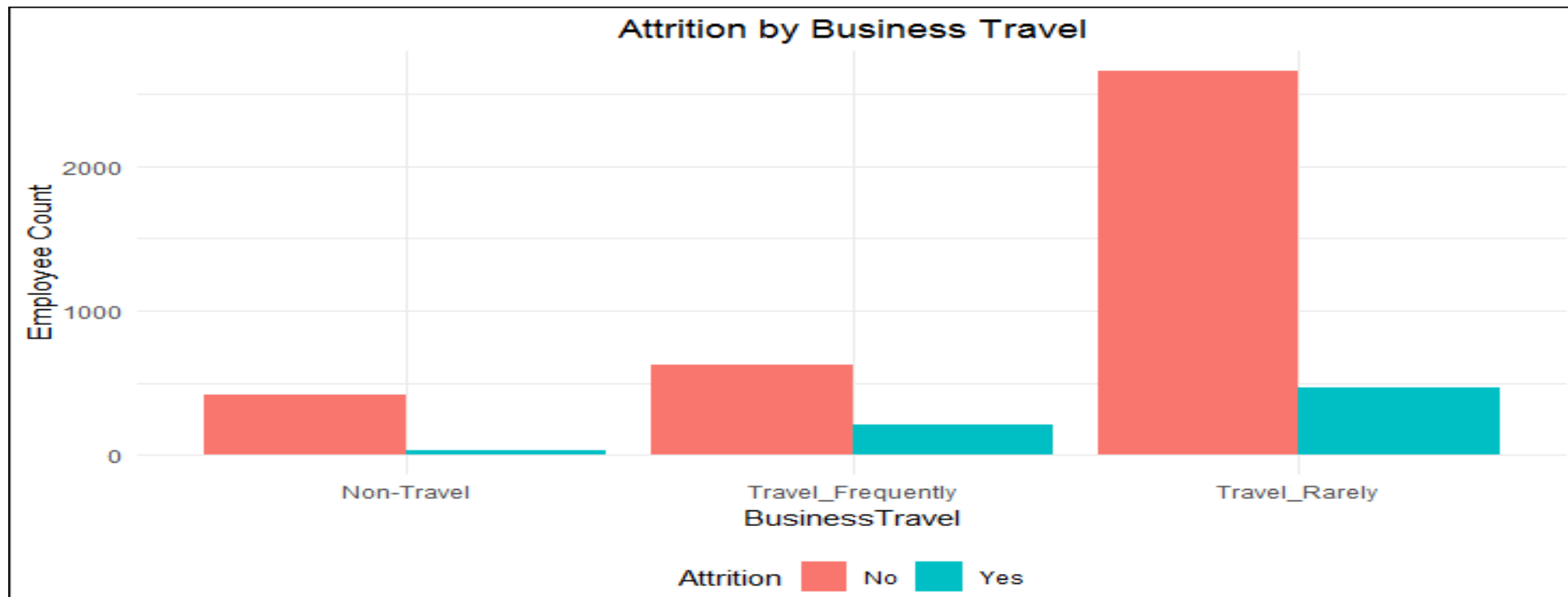
Mean_Hours_Worked :- As the table below suggests with the Increase in working hours, there is increase in attrition rate

	working_Hours	min	max	total	count	perc_Attrition
	<int>	<dbl>	<dbl>	<dbl>	<int>	<dbl>
1	1	6	14	221	1470	15
2	2	14	20	236	1470	16
3	3	20	30	254	1470	17

Business Travel :- As you can infer from the below table Employees who travel rarely are higher than who travel frequently. There is significant pattern of attrition seen for Employees travelling frequently and rarely.

```
> table(HRAnalytics$BusinessTravel)

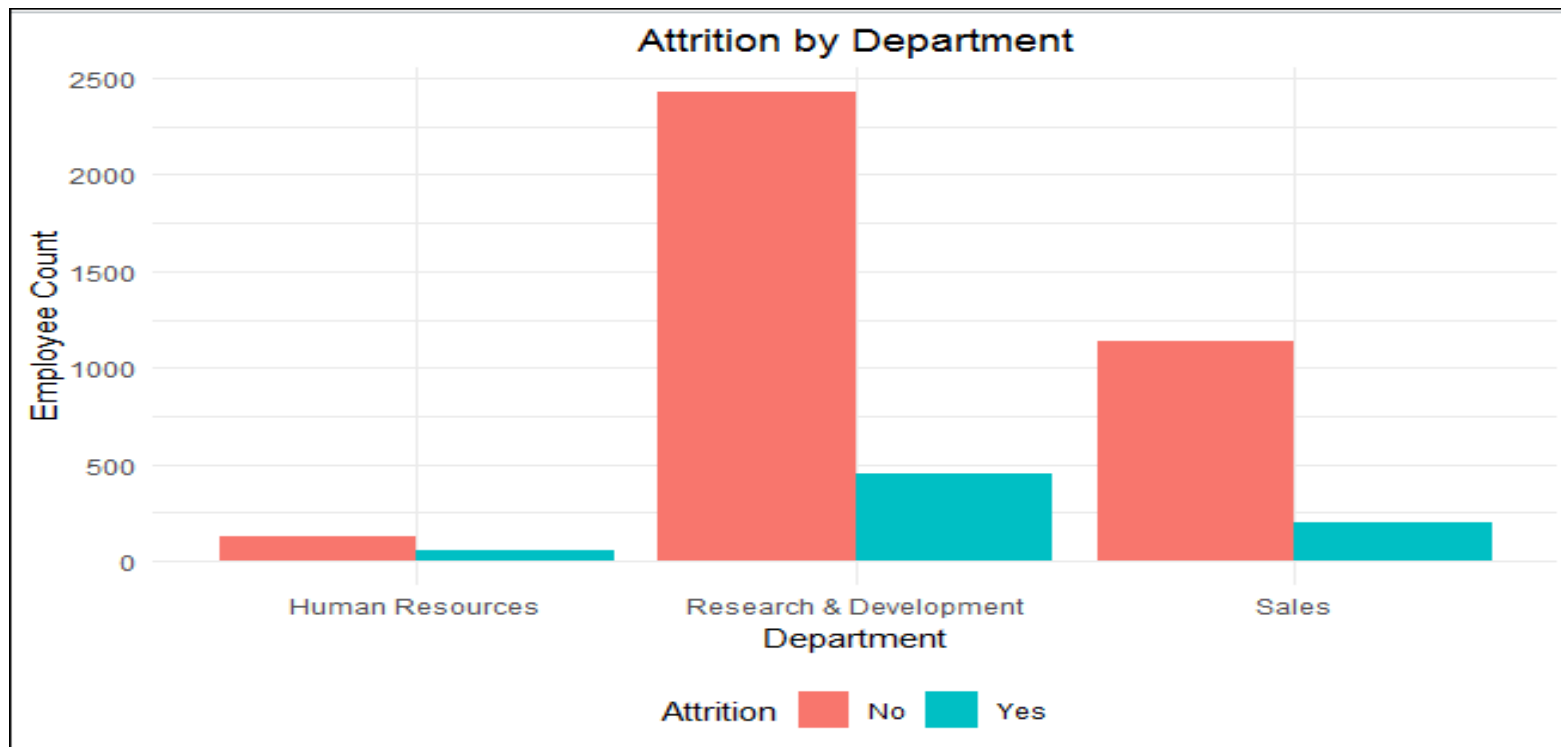
Non-Travel  Travel_Frequently  Travel_Rarely
      450             831             3129
```



Department :- R&D and Sales department has significantly higher attrition.

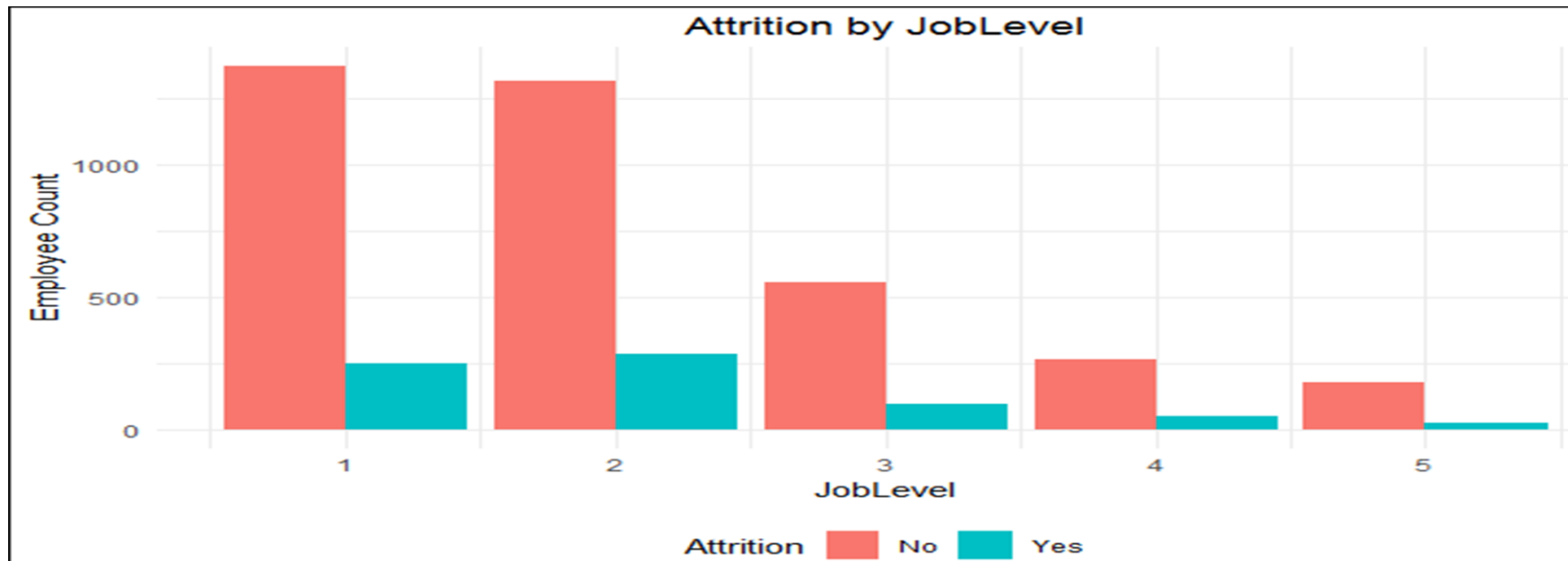
```
> table(HRAnalytics$Department)
```

Human Resources	Research & Development	Sales
189	2883	1338



JobLevel :- Job Level 5 also has significantly higher attrition when compared with the volume at that level.
 Job level 1 & 2 has around 70% of employees i.e. 3000+ employees.
 Considering the volume at job level 1 & 2, we can focus on job level 1 & 2 to control the attrition.

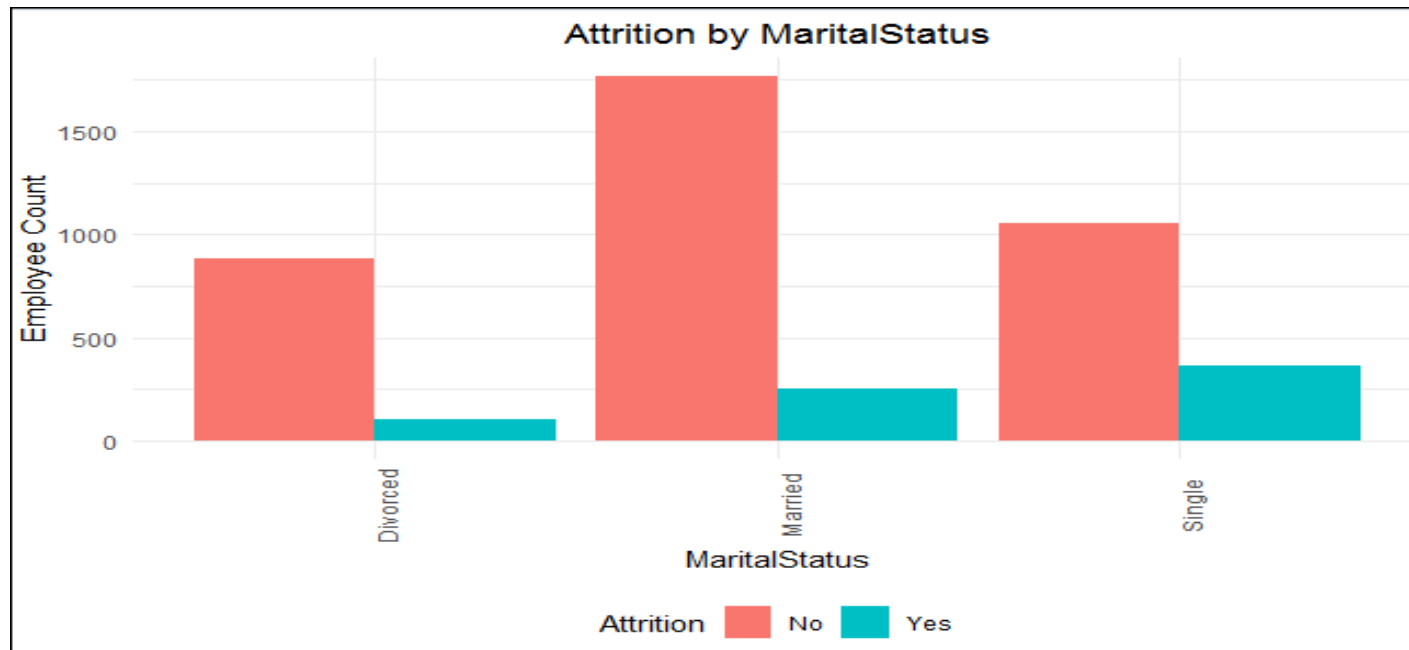
```
> table(HRAnalytics$JobLevel)
 1      2      3      4      5 
1629 1602  654  318  207
```



MaritalStatus :- Single status has higher attrition rate @ 26% , Divorced and Married have attrition rates of 10 and 12% respectively.

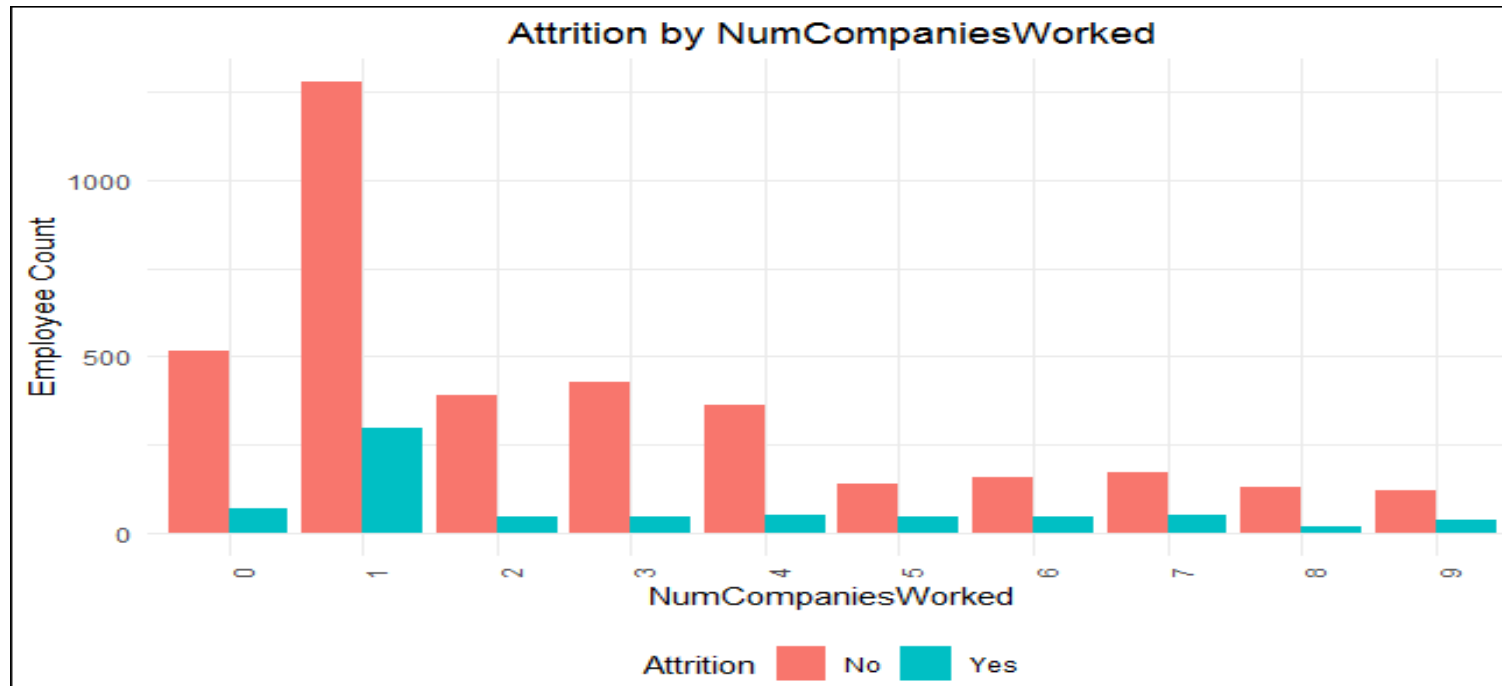
```
> table(HRAnalytics$MaritalStatus)
```

Divorced	Married	Single
981	2019	1410



NumCompaniesWorked :- 35% of employees have worked with 1 company where attrition rate is 19%. Looking at the volume of this category, focus on this group of employees would result in solving attrition to a large extend.

```
> prop.table(table(HRAnalytics$NumCompaniesWorked))#35% of employees worked for 1 company.
 0      1      2      3      4      5      6      7      8
0.13287982 0.35759637 0.09931973 0.10748299 0.09410431 0.04240363 0.04716553 0.05034014 0.03333333
 9
0.03537415
```



TrainingTimesLastYear :-

- Maximum Employees i.e. 70% of employees have attained 2-3 trainings last year. Attrition Rate is 17-18% here.
- Employees who have attained zero training last year have highest attrition rate i.e. 19%

```
> table(HRAnalytics$TrainingTimesLastYear)

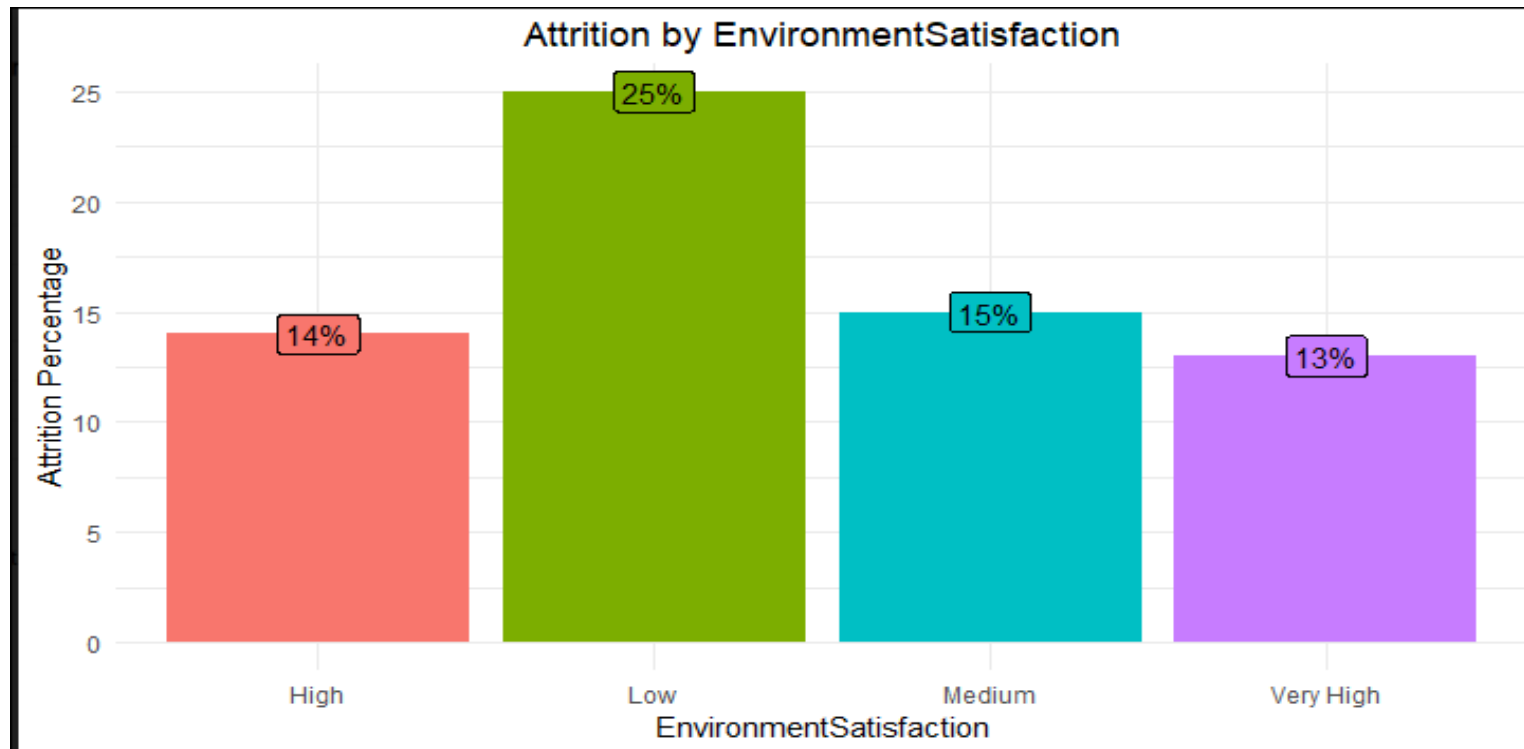
 0      1      2      3      4      5      6 
162    213   1641   1473    369    357    195
```



EnvironmentSatisfaction :- Low Environment Satisfaction leads to high Attrition rate (25%)

```
> table(HRAnalytics$EnvironmentSatisfaction)
```

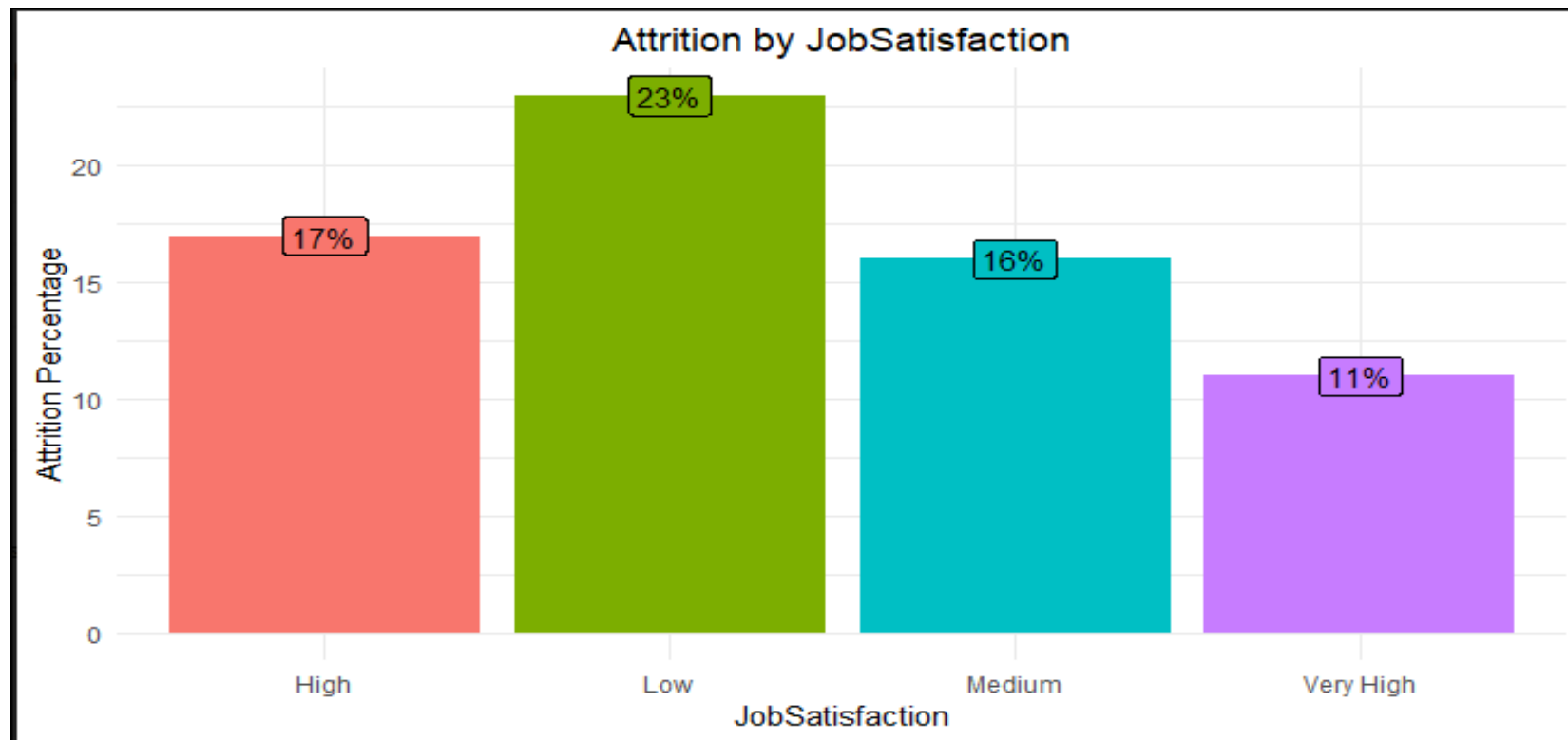
High	Low	Medium	Very High
1375	845	856	1334



JobSatisfaction :- Low Job Satisfaction leads to higher attrition rate

```
> table(HRAnalytics$JobSatisfaction)
```

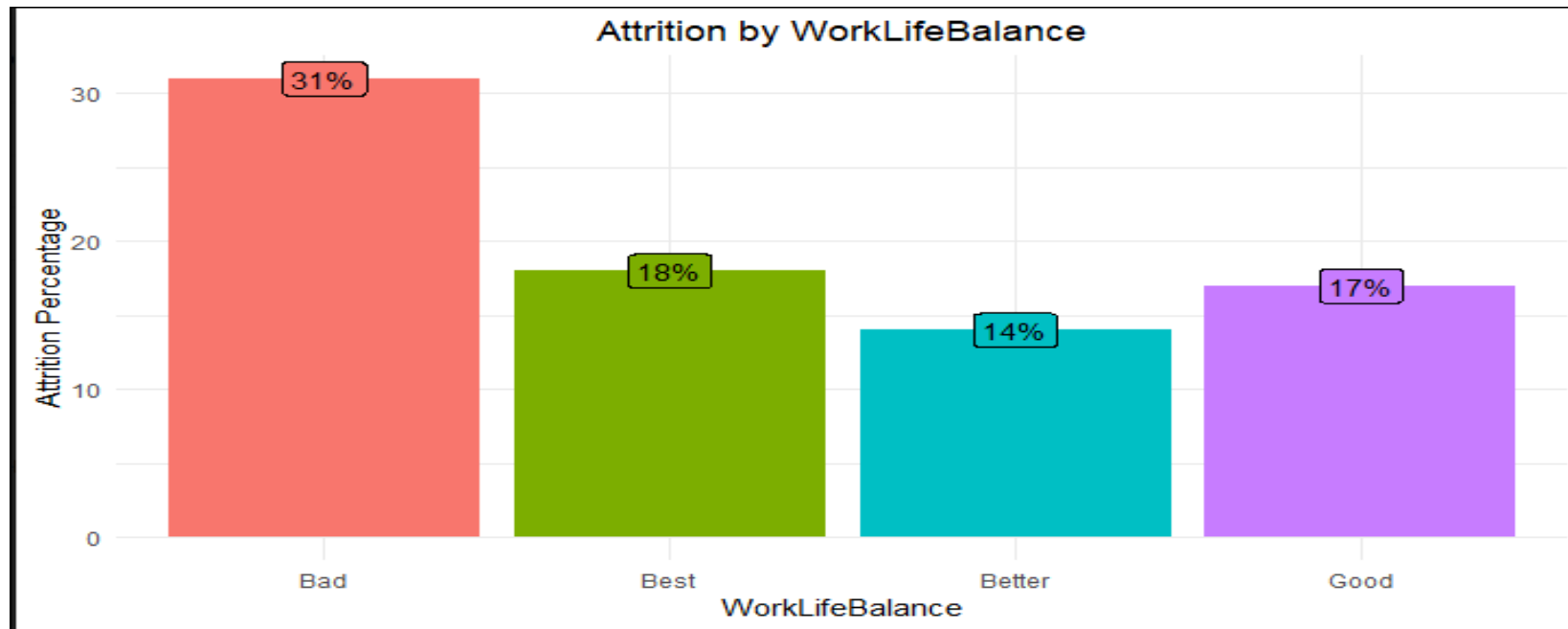
High	Low	Medium	Very High
1323	860	840	1387



WorkLifeBalance :- Attrition rate is higher with Bad work life balance. However, the volume of Employees with Bad work life balance is very less i.e. 5% of total employees with bad worklife balance.

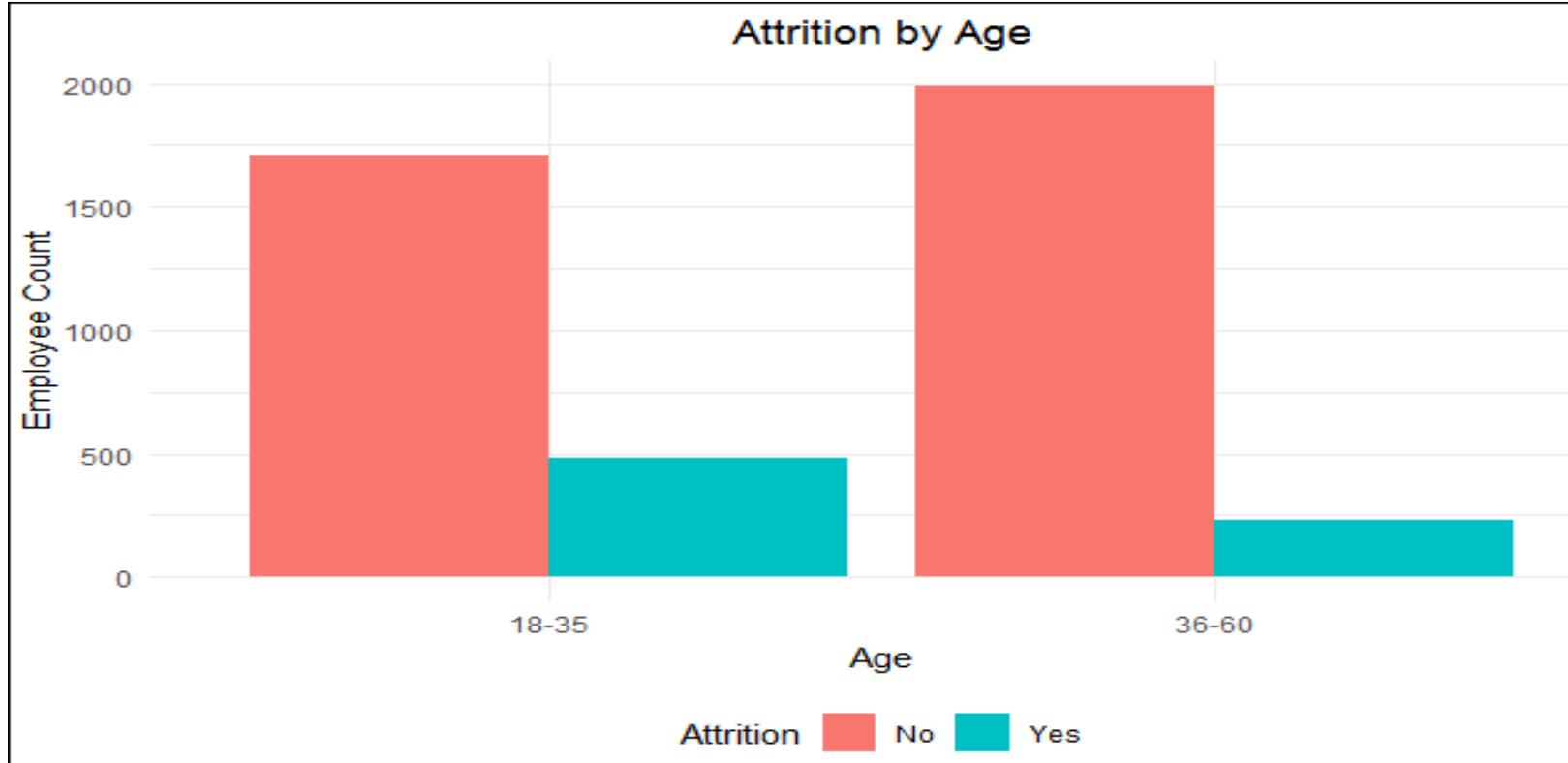
```
> table(HRAnalytics$workLifeBalance)
```

Bad	Best	Better	Good
239	454	2698	1019

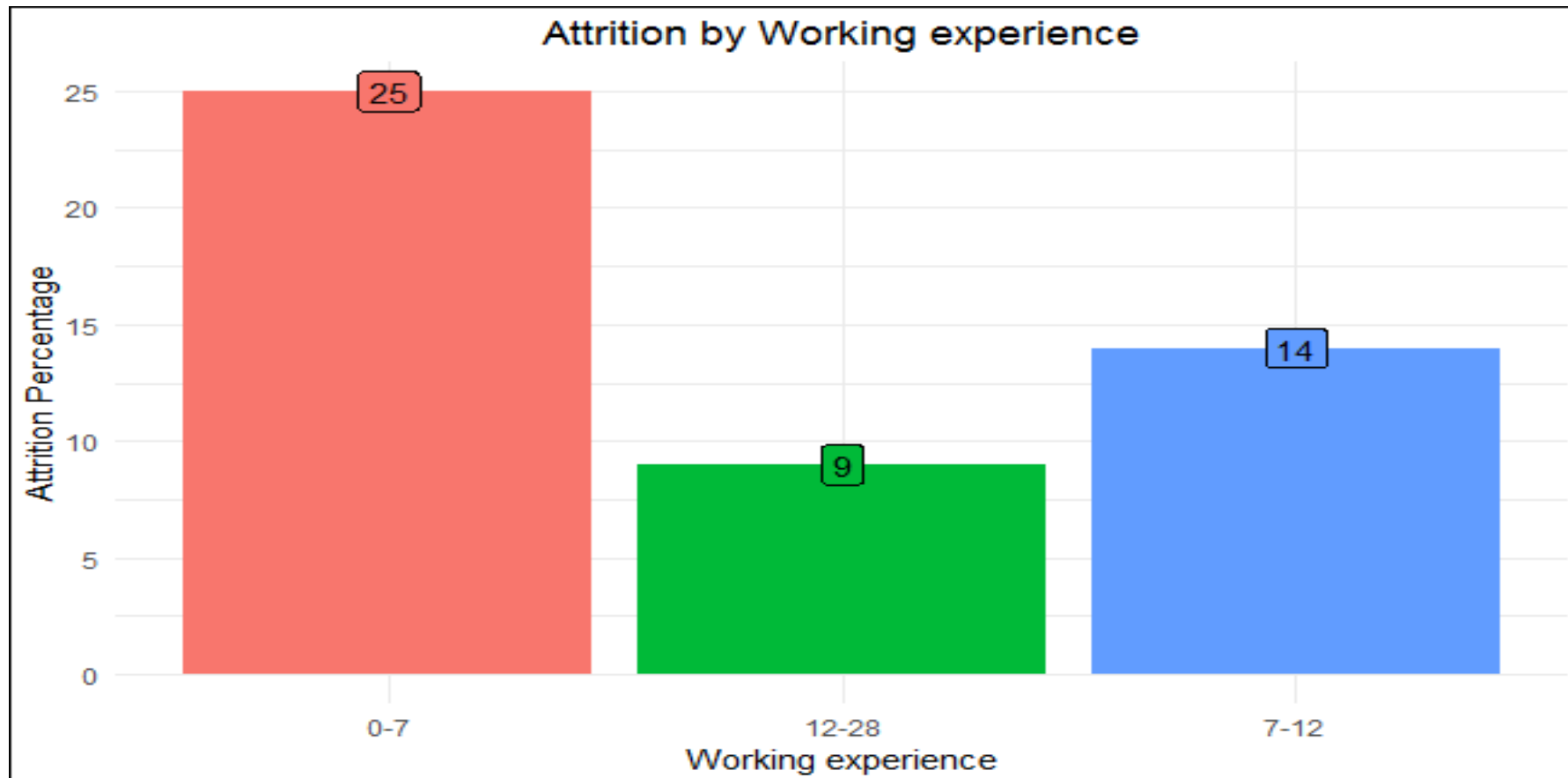


Age :- Attrition Rate is 2 times in age group 18-35 compared to age group 36-60.

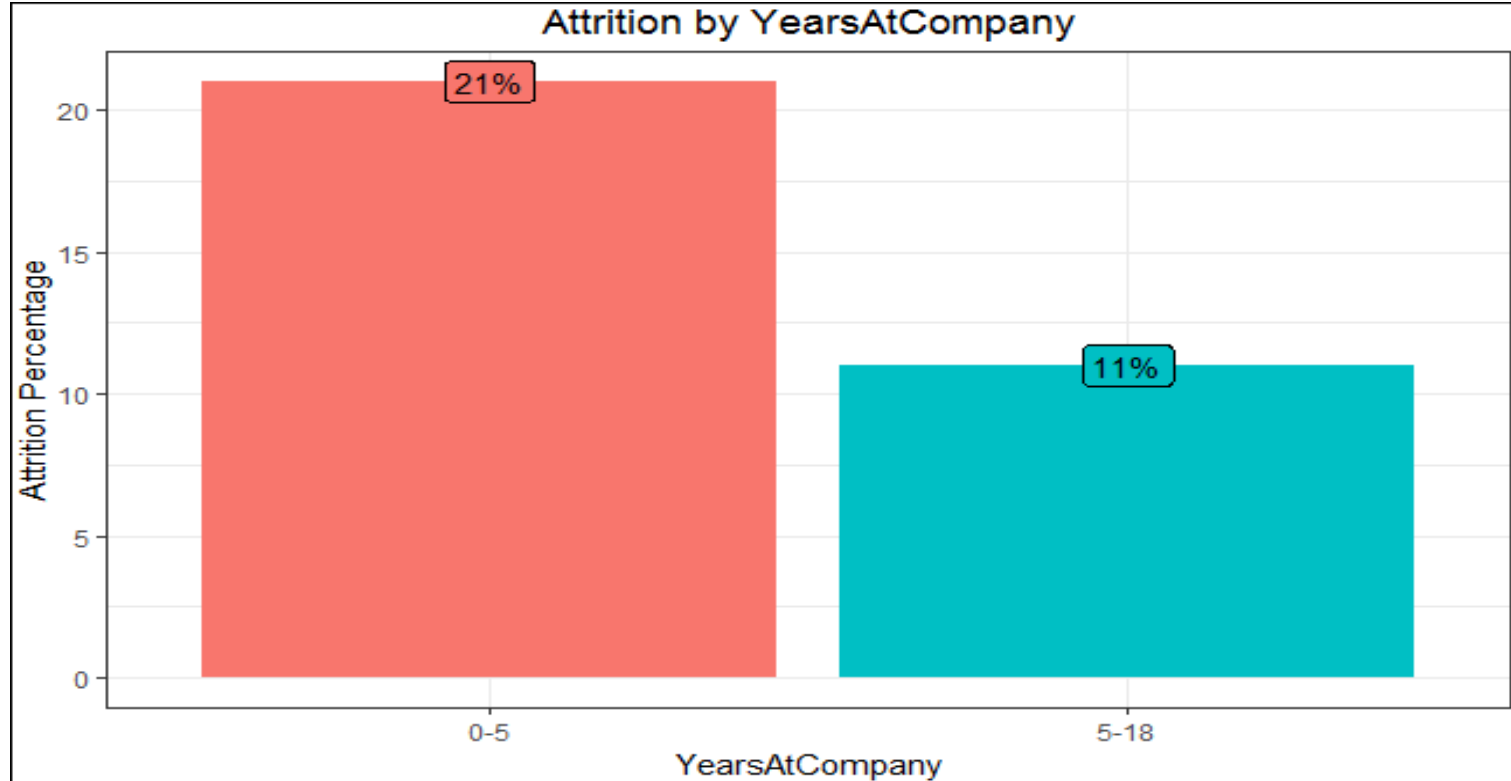
```
> summary(HRAnalytics$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00  30.00   36.00   36.92  43.00   60.00
```



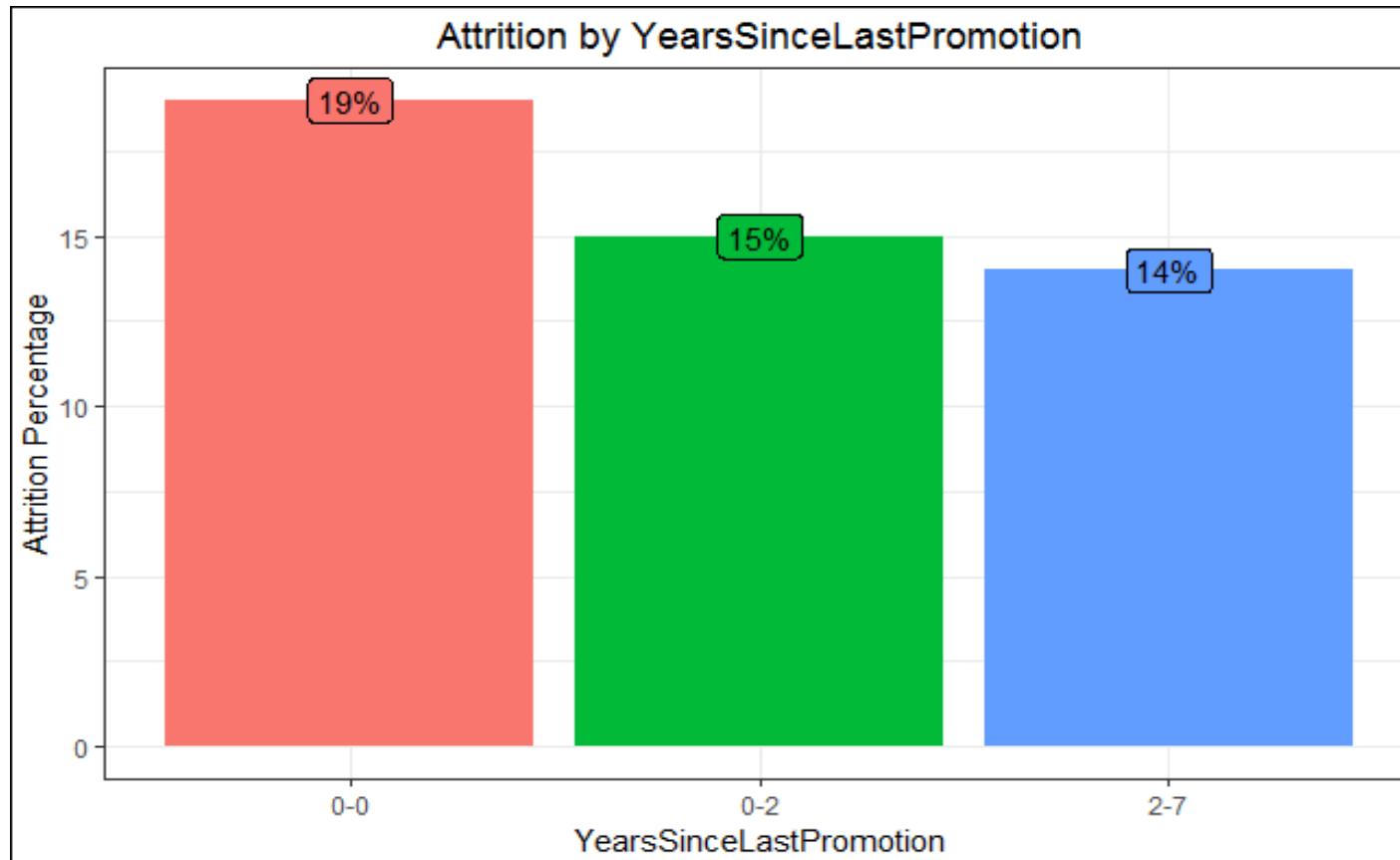
TotalWorkingYears :- Attrition is higher with employees having less total working experience. Attrition with less than 7 years working experience, attrition rate is significantly high



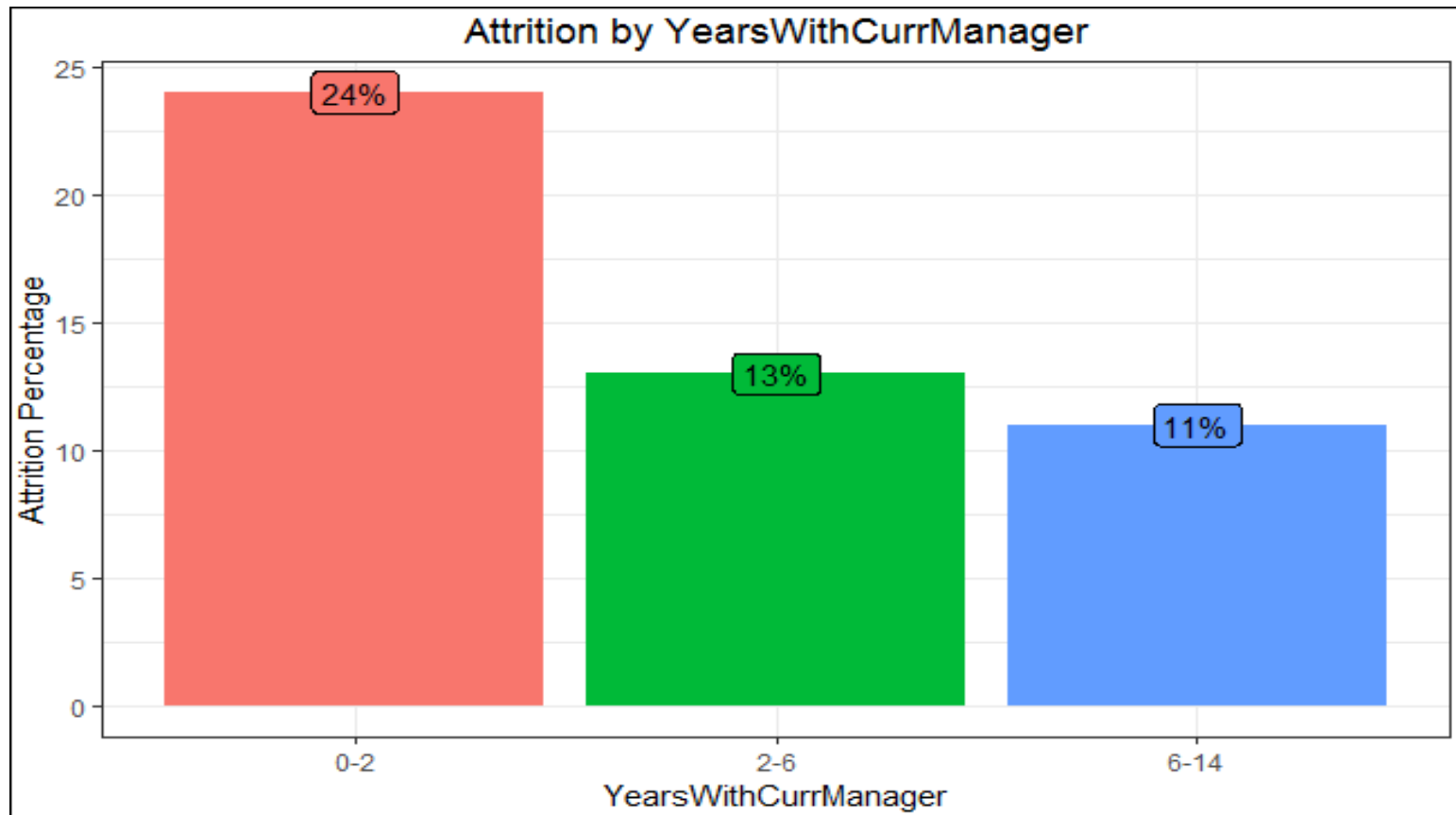
YearsAtCompany :- Employees with less than 5 years at company has relatively higher attrition rate (@ 21%) than employees above 5 years at company



YearsSinceLastPromotion :- Employees with no promotion have high attrition rates



YearsWithCurrManager :- Employees less than 2 years with current manager have higher chances of attrition.





Conclusion

1. The employees who are from R&D and Sales department attrite at high rate.
2. Employees with marital status as single tend to attrite more than married\divorced.
3. Employees who have hopped jobs less than four companies tend to attrite more frequently.
4. Employees with less work Life balance and environmental satisfaction attrite more.
5. Employees from the age group 18 to 35 have high attrition rates compared to other age groups.
6. Employees with less than seven years of exp have high attrition rate.
7. Employees whose tenure in the company less than 5 years tend to attrite more.
8. Employees with no promotions in the last two years tend to attrite more.
9. Employees who have not been provided any training in the previous year attrite more.
10. Employees with niche skills (unicorn) tend to attrite more than the normal employees