HR Analytics Case Study

SUBMISSION

Group Name:

1. Anshika Misra
2. Madhulika Joshi
3. Sharanya Vijay
4. Piyu Chatterjee

Business Objectives

A large company named **XYZ**, employs 4410 employees. However, every year, around 16% of its employees leave the company and need to be replaced with the talent pool available in the job market. This level of **attrition** is bad for the company, because of the following reasons -

- The former employees' projects get delayed, which makes it difficult to meet **timelines**, resulting in a reputation loss among consumers and partners
- A sizeable department has to be maintained, for the purposes of **recruiting** new talent
- More often than not, the new employees have to be **trained** for the job and/or given time to acclimatize themselves to the company

Hence, the management wants to understand what factors they should focus on, in order to curb attrition. They want to know what changes they should make to their workplace, in order to get most of their employees to stay. Also, they want to know which of these variables is most important and needs to be addressed right away.

The results thus obtained will be used by the management to understand what changes they should make to their workplace, in order to get most of their employees to stay.

Steps involved

1. Data Import
2. Data Cleaning
   - Check the duplicates in employee records
   - Find the number of unique members for each column
   - Find and remove the column with only NA values
   - Find and remove the columns that have only a single value in the entire column
   - Find no of columns which have only NA data for both in_time and out_time and remove them from the data frame
   - Impute the NA's in the data wherever possible and remove the ones that cannot be imputed.
3. Data Conversion
   - Convert to factors – Attrition, BusinessTravel, Department, EducationField, Gender, JobRole, MaritalStatus, Education, JobLevel, StockOptionLevel, JobSatisfaction, WorkLifeBalance, JobInvolvement, PerformanceRating, WorkingCondition
   - Convert to numeric – EmployeeID, Age, DistanceFromHome, MonthlyIncome, NumCompaniesWorked, PercentSalaryHike, TotalWorkingYears, TrainingTimesLastYear, YearsAtCompany, YearsSinceLastPromotion, YearsWithCurrManager, Leaves, TotalWorkingHrs, AvgTime
   - Convert to date – All the dates in in_time and out_time files
4. Add Derived Columns
   - Find the number of leaves that an employee has taken and add it to the employee data frame
   - Find the total in-office time per employee
   - Find the total working hours per year by aggregating the difference
   - Find the average working hours per employee using the formula (TotalWorkingHours/(total no of days - Leaves))
   - Divide the employees into Overworked/Underworked/Normal categories based on average no of hours worked

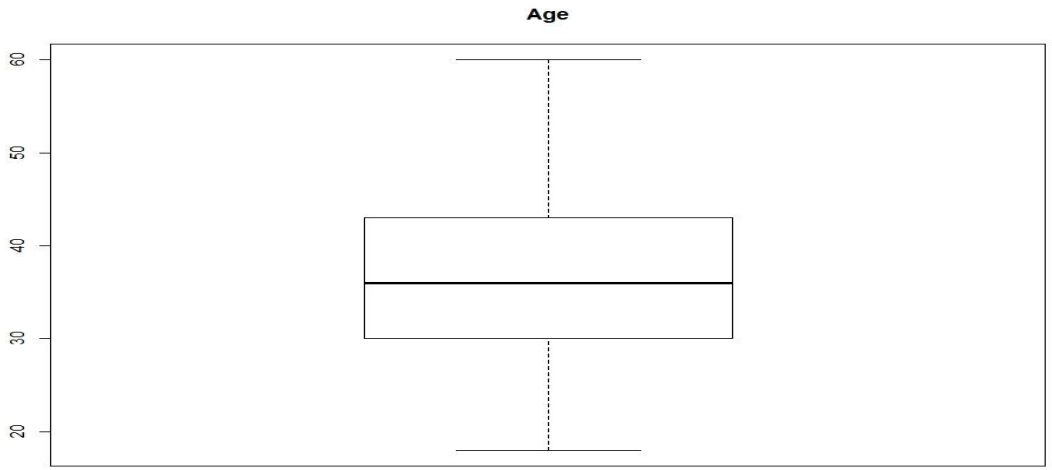
Steps involved – contd.

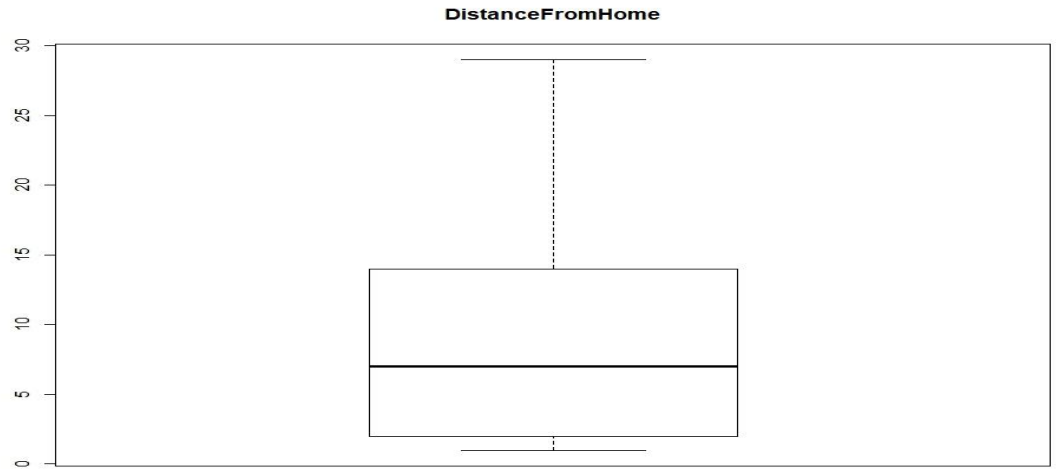
5. Outlier Treatment
   - MonthlyIncome , YearsSinceLastPromotion capped at 0.90
   - YearsAtCompany capped at 0.92
   - NumCompaniesWorked, TotalWorkingYears, TrainingTimesLastYear, capped at 0.95
   - YearsWithCurrManager, TotalWorkingHrs capped at 0.97
   - TrainingTimesLastYear floored at 0.05
6. Perform Univariate analysis on Continuous and Discrete Variables
   - Characteristics of Continuous Variables
   - Plots for Discrete Variables:
7. Perform Bivariate analysis
8. Derive Correlations
9. Normalizing continuous features
10. Convert factors with 2 levels to numerical variables
11. Creating dummy variables for categorical variables having more than 2 factors.
12. Split the data into Test and Train
13. Build Logistic Regression Models
14. Evaluate the model and find the optimal cutoff
15. Derive KS –statistic and Lift & Gain
16. Draw Conclusions

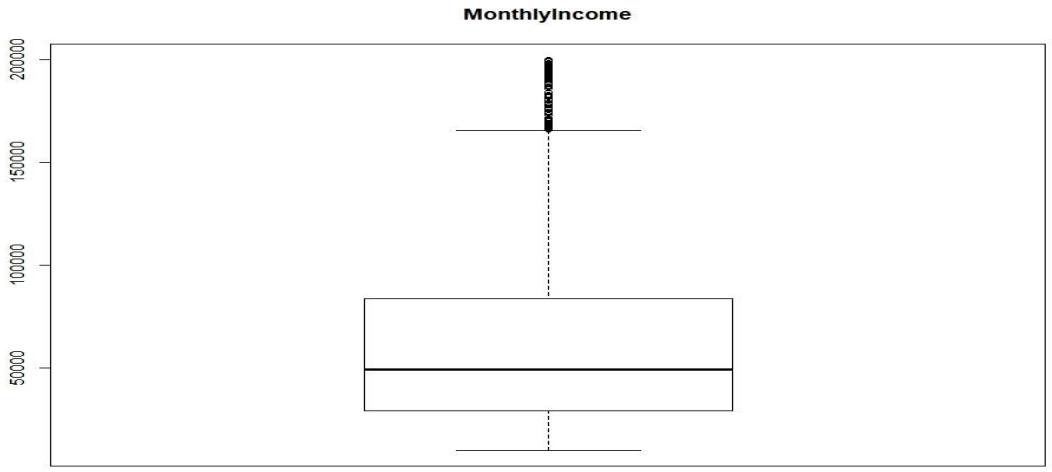
Continuous Variables – Characteristics

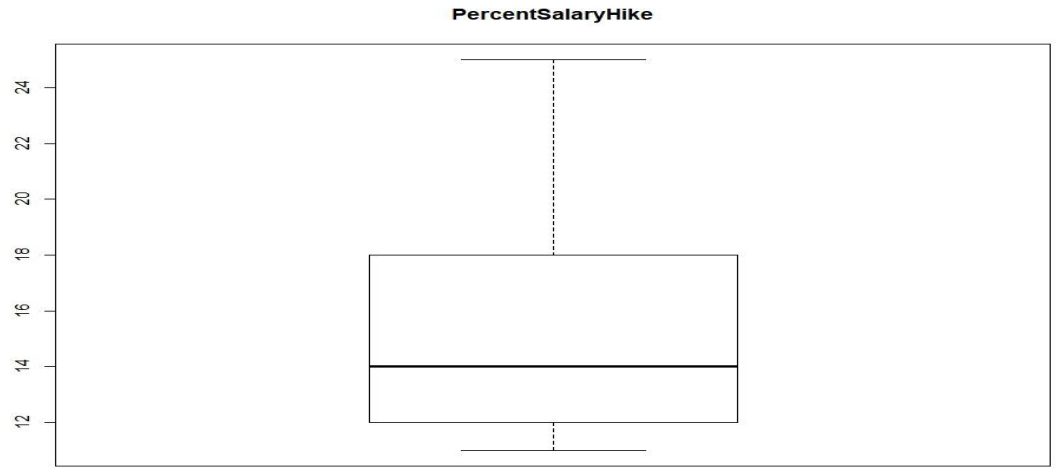
Variables	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
EmployeeID	1	1107	2208	2207	3308	4410
DistanceFromHome	1	2	7	9.2	14	29
MonthlyIncome	10090	29110	49300	65033	83760	199990
NumCompaniesWorked	0	1	2	2.69	4	9
PercentSalaryHike	11	12	14	15.2	18	25
TotalWorkingYears	0	6	10	11.3	15	40
TrainingTimesLastYear	0	2	3	2.8	3	6
YearsAtCompany	0	3	5	7.01	9	40
YearsSinceLastPromotion	0	0	1	2.19	3	15
YearsWithCurrManager	0	2	3	4.13	7	17
Leaves	1	8	13	12.7	17	24
TotalWorkingHrs	1349	1563	1746	1821	1967	2723
AvgTime	5.95	6.67	7.41	7.7	8.36	11.03

Outlier Analysis



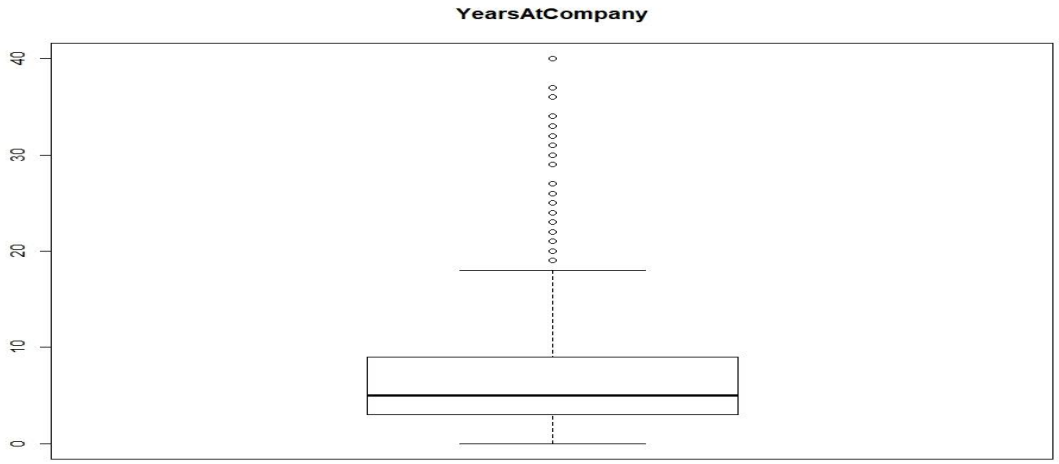


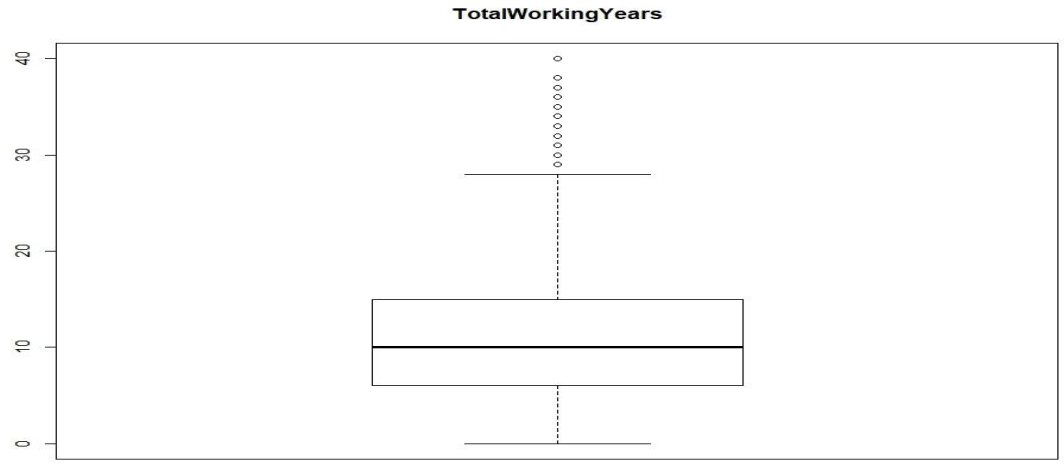


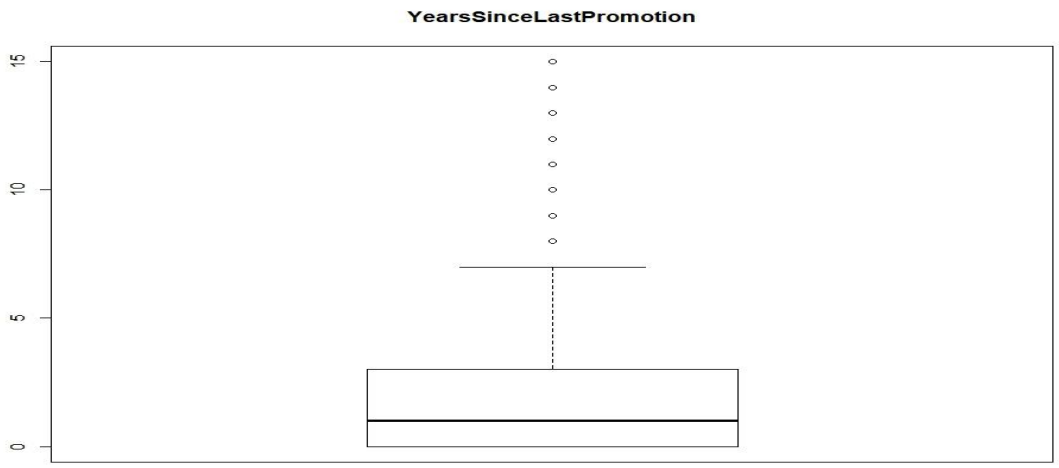


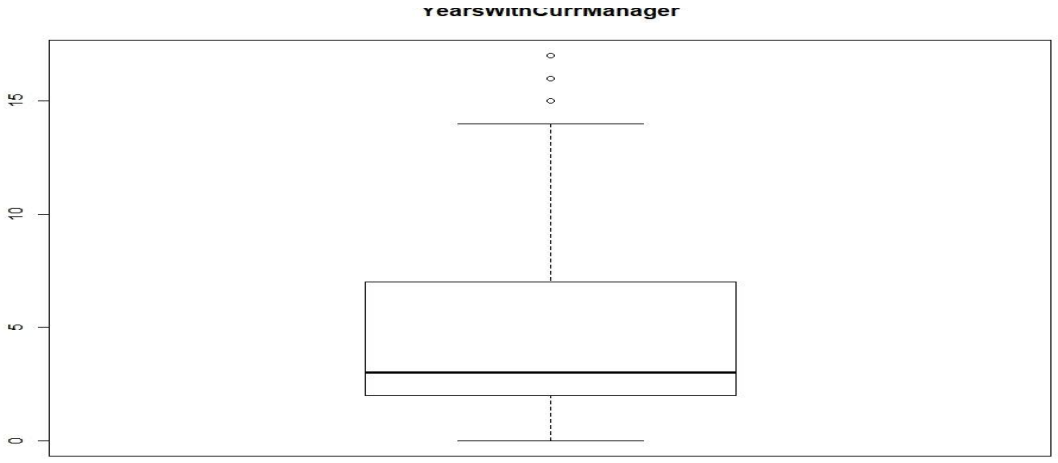
Monthly Income has Outliers. Age, Distance From Home and Percent Salary Hike do not have Outliers.

Outlier Analysis



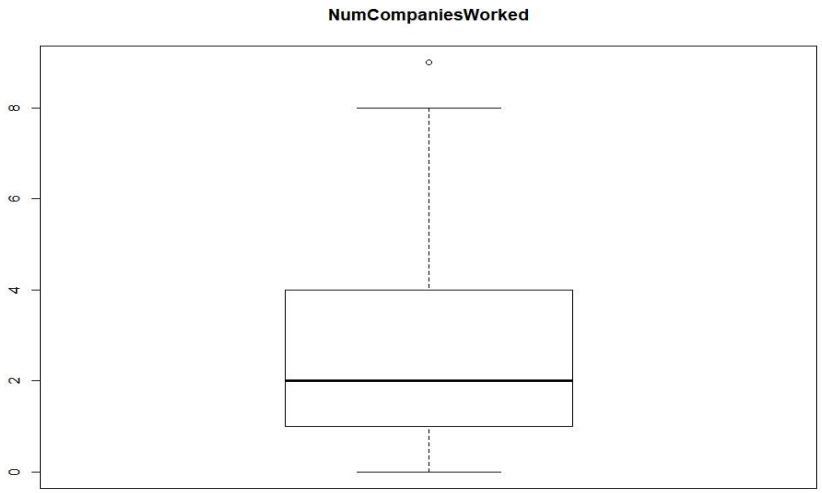


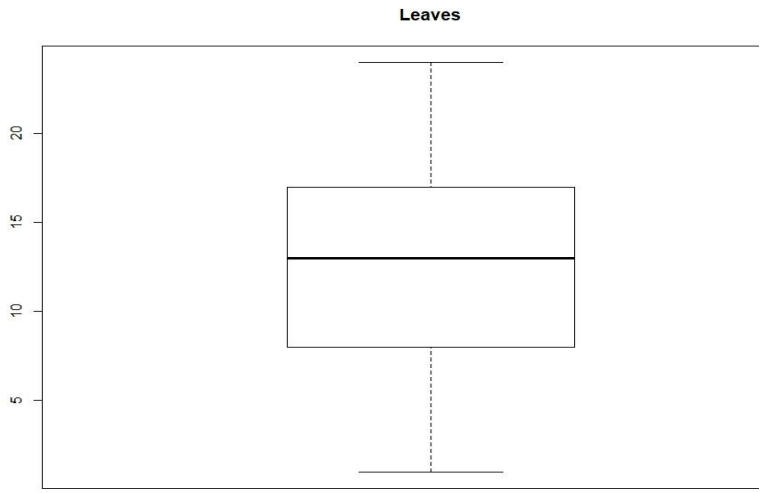


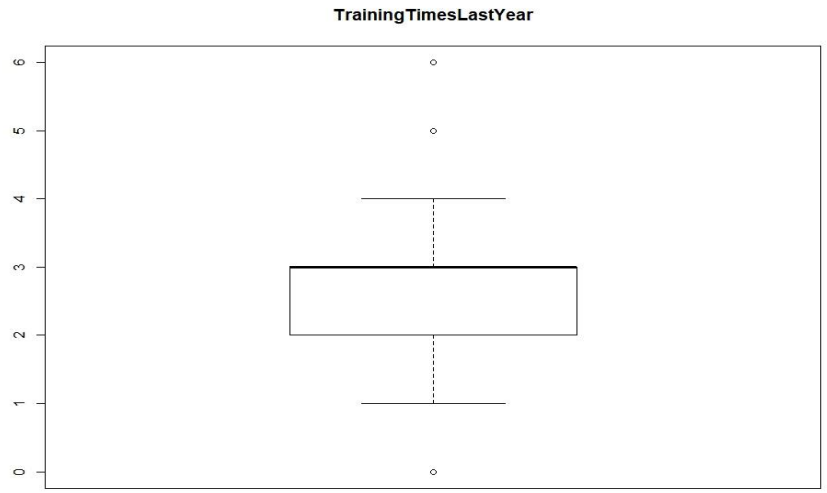


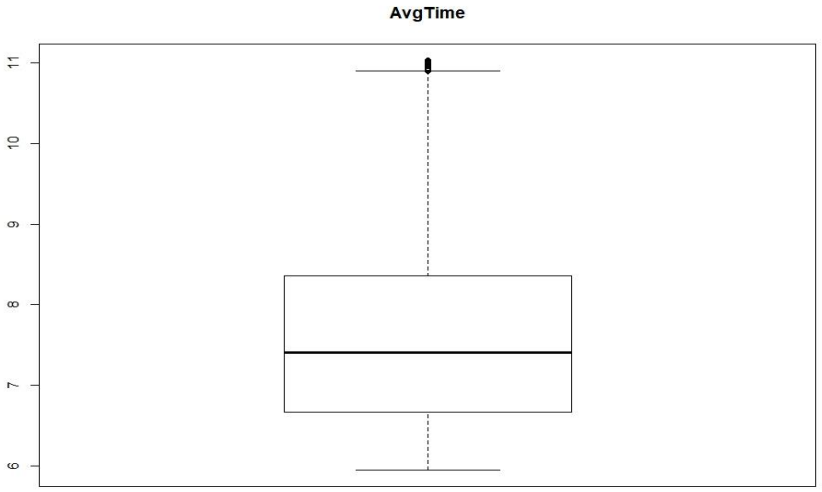
Years at Company, Total Working Years, Years Since Last Promotion, Years with Current Manager have Outliers.

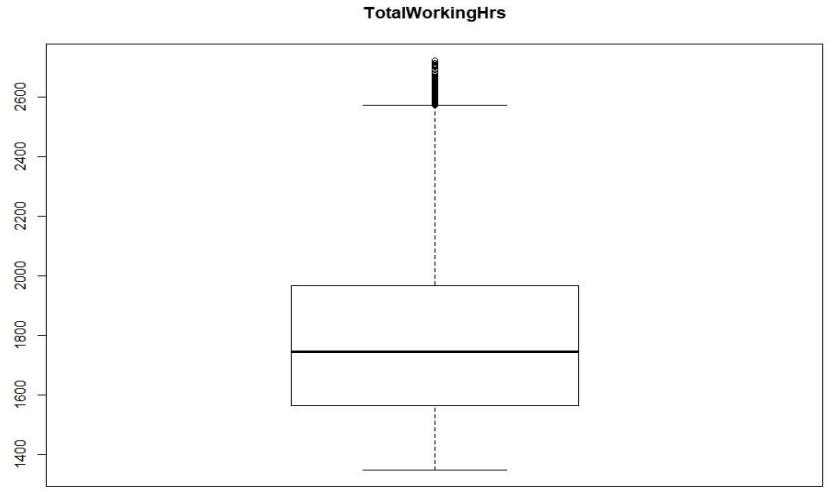
Outlier Analysis









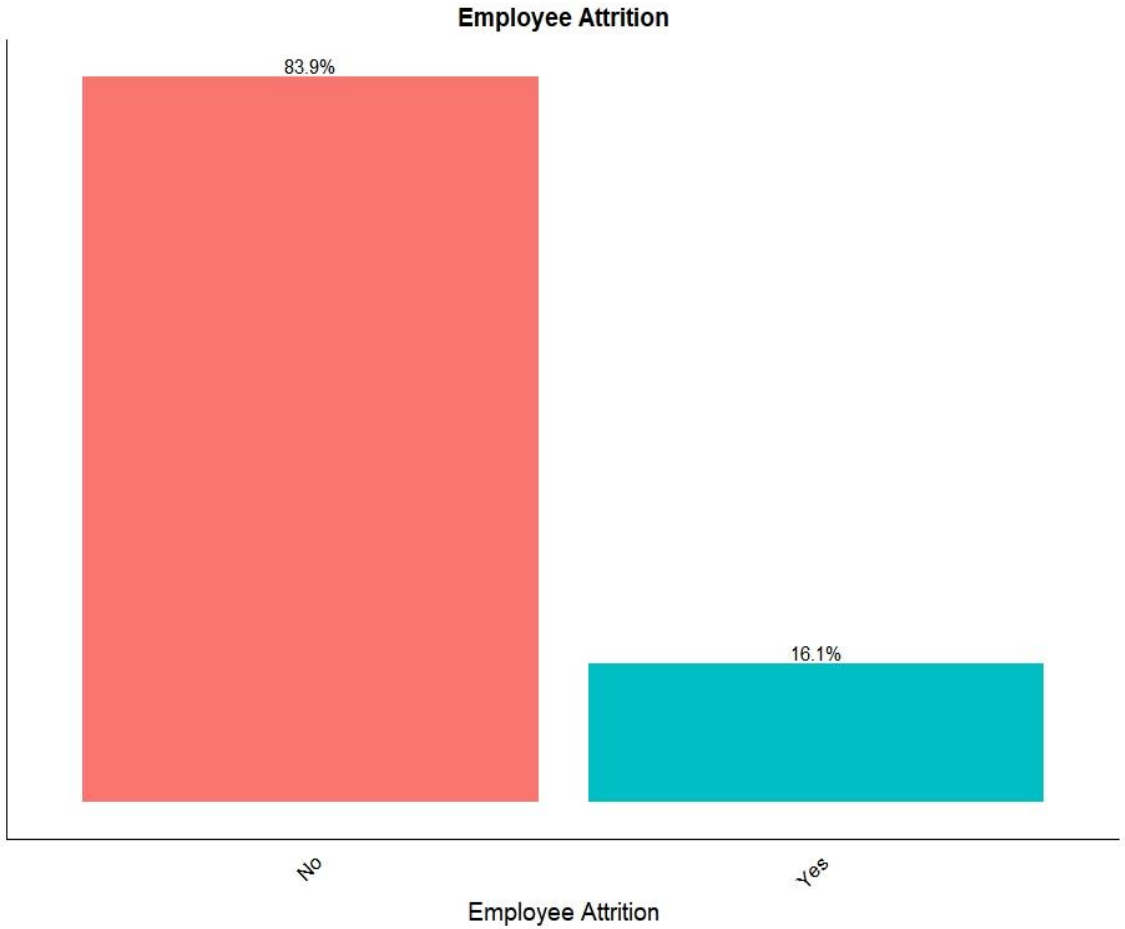


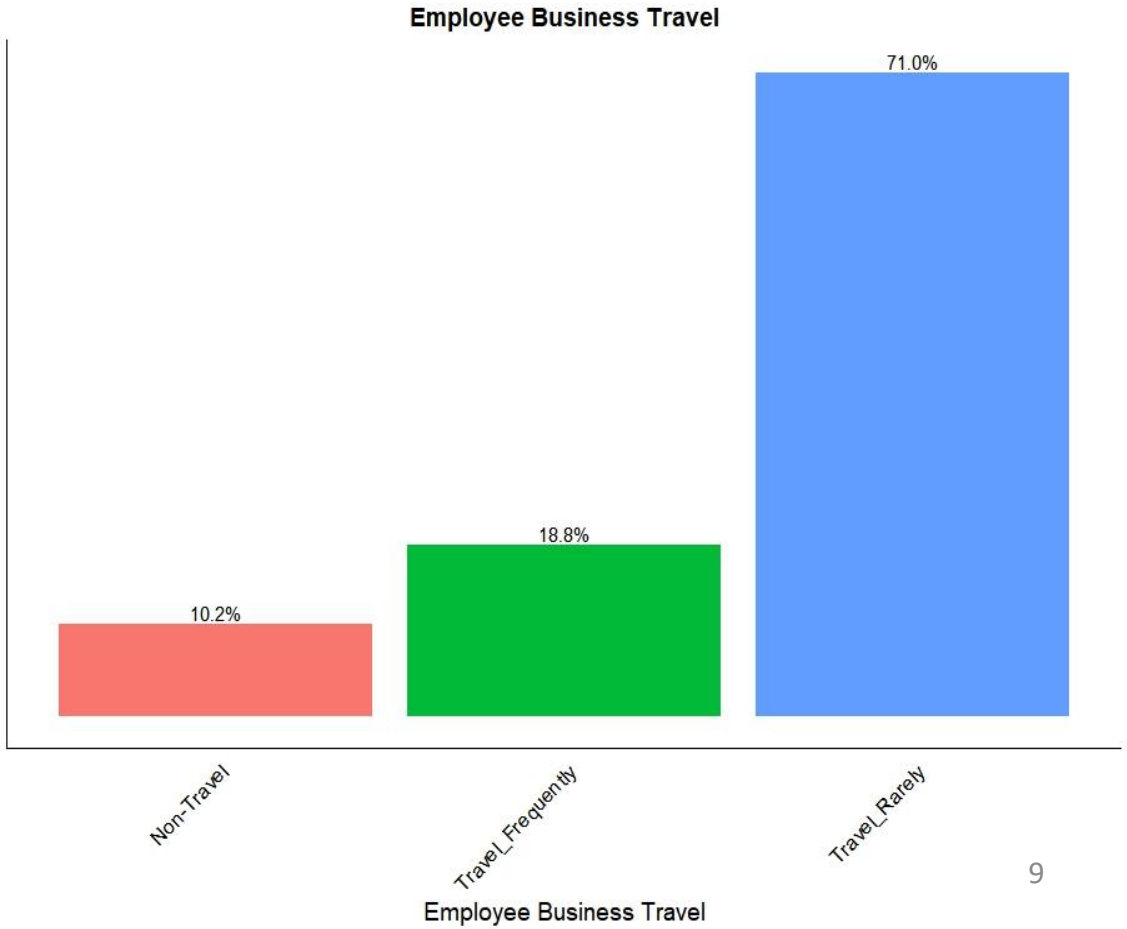
Num Companies Worked, Training Times Last Year, Avg Time, Total Working Hours have outliers. Leaves has no Outliers

Univariate Analysis – Attrition and Business Travel

Attrition	Count	Percentage
No	3688	83.93%
Yes	706	16.07%

BusinessTravel	Count	Percentage
Non-Travel	448	10.20%
Travel_Frequently	827	18.82%
Travel_Rarely	3119	70.98%

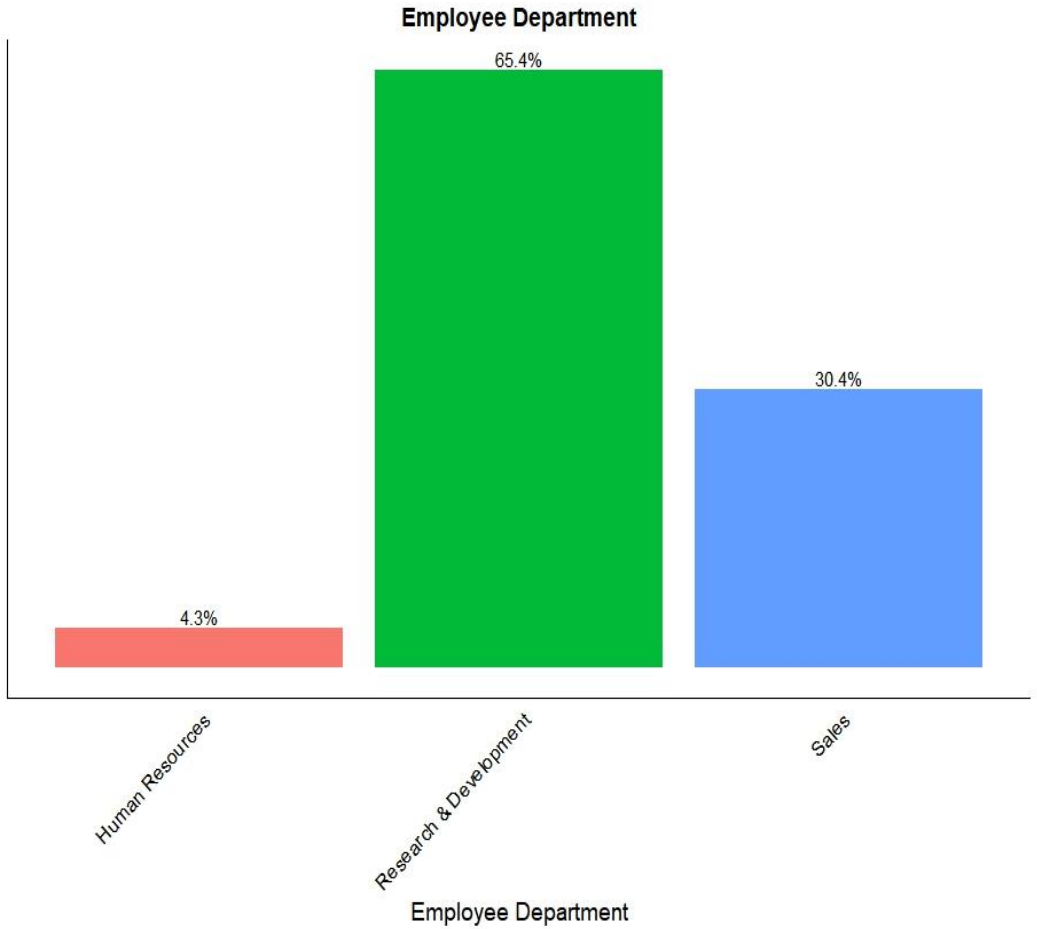


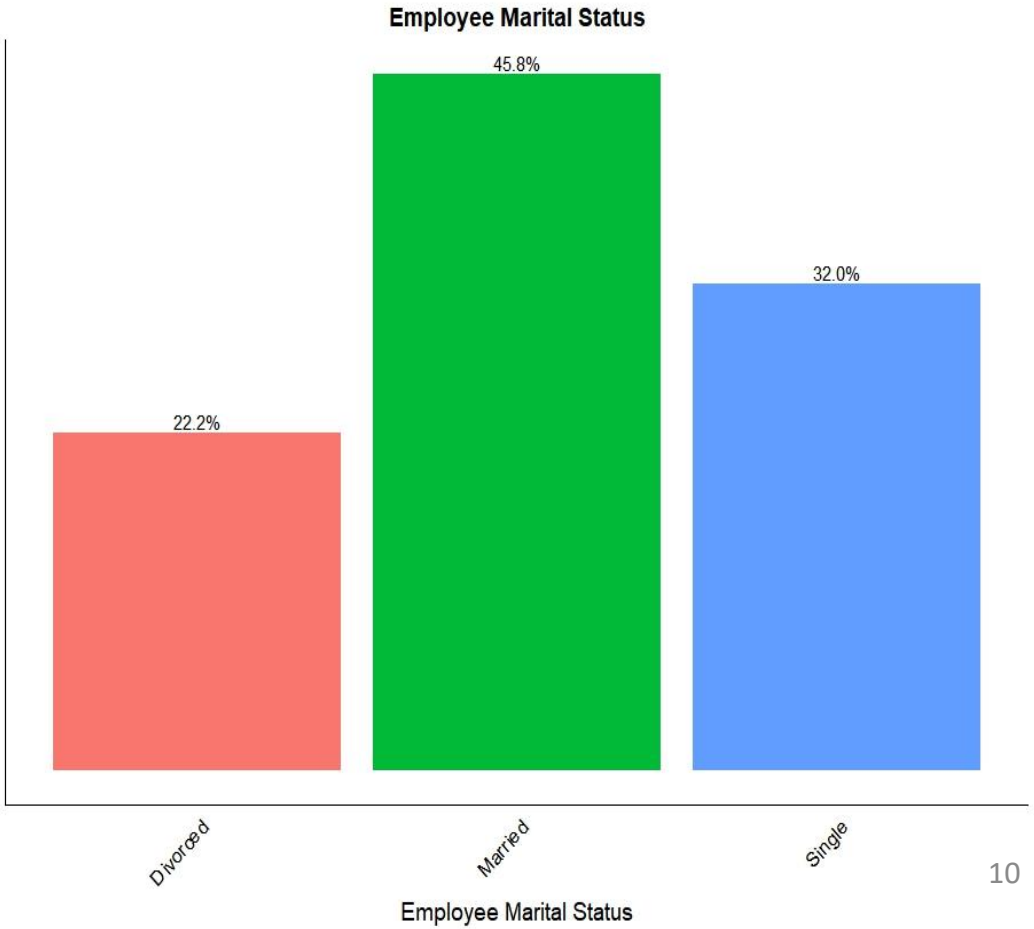


Univariate Analysis – Department and Marital Status

Department	Count	Percentage
Human Resources	187	4.26%
Research & Development	2873	65.38%
Sales	1334	30.36%

MaritalStatus	Count	Percentage
Divorced	976	22.21%
Married	2012	45.79%
Single	1406	32.00%

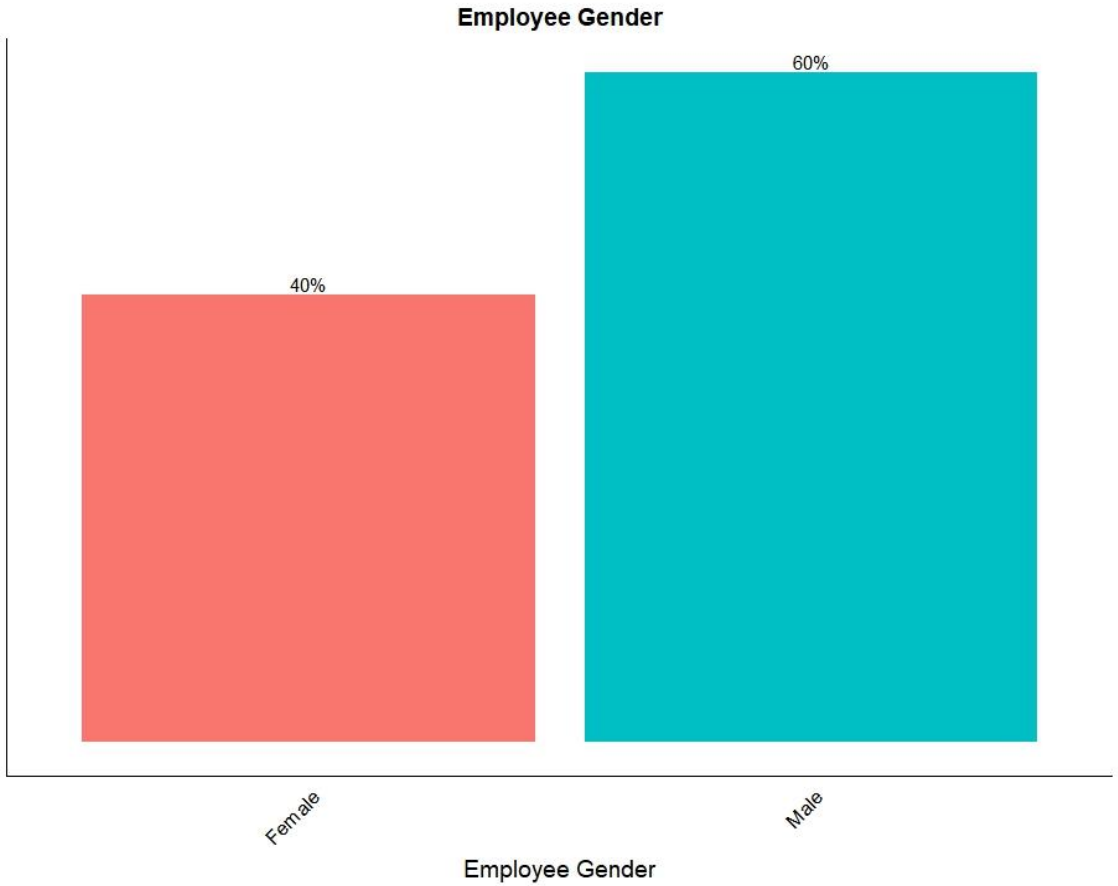


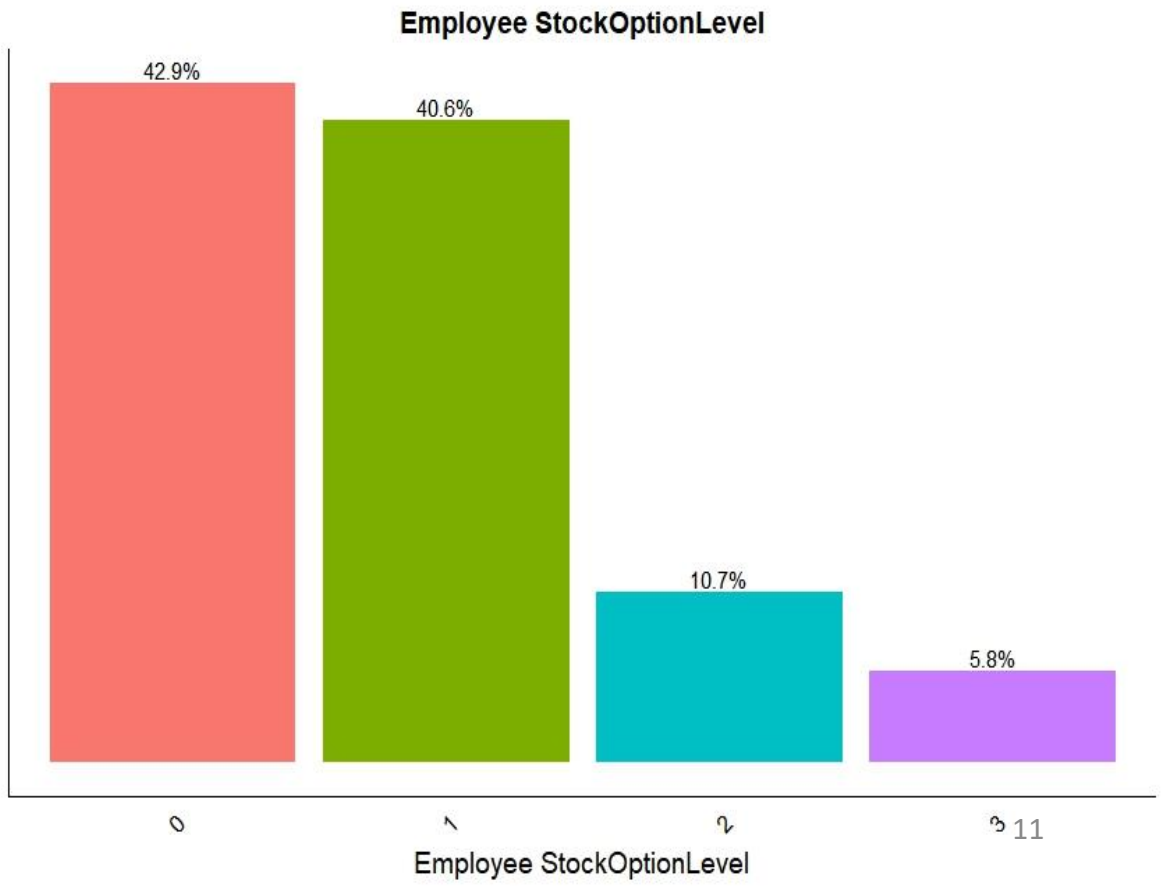


Univariate Analysis – Gender and Stock Option Level

Gender	Count	Percentage
Female	1759	40.03%
Male	2635	59.97%

StockOptionLevel	Count	Percentage
0	1886	42.92%
1	1782	40.56%
2	472	10.74%
3	254	5.78%

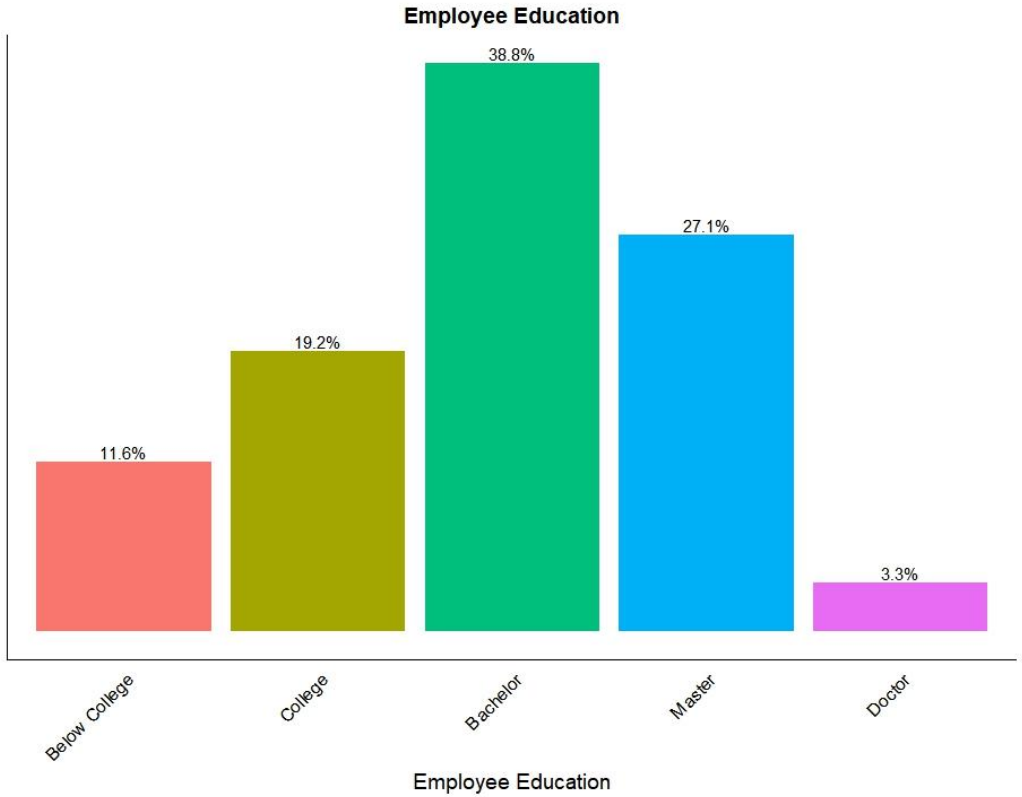


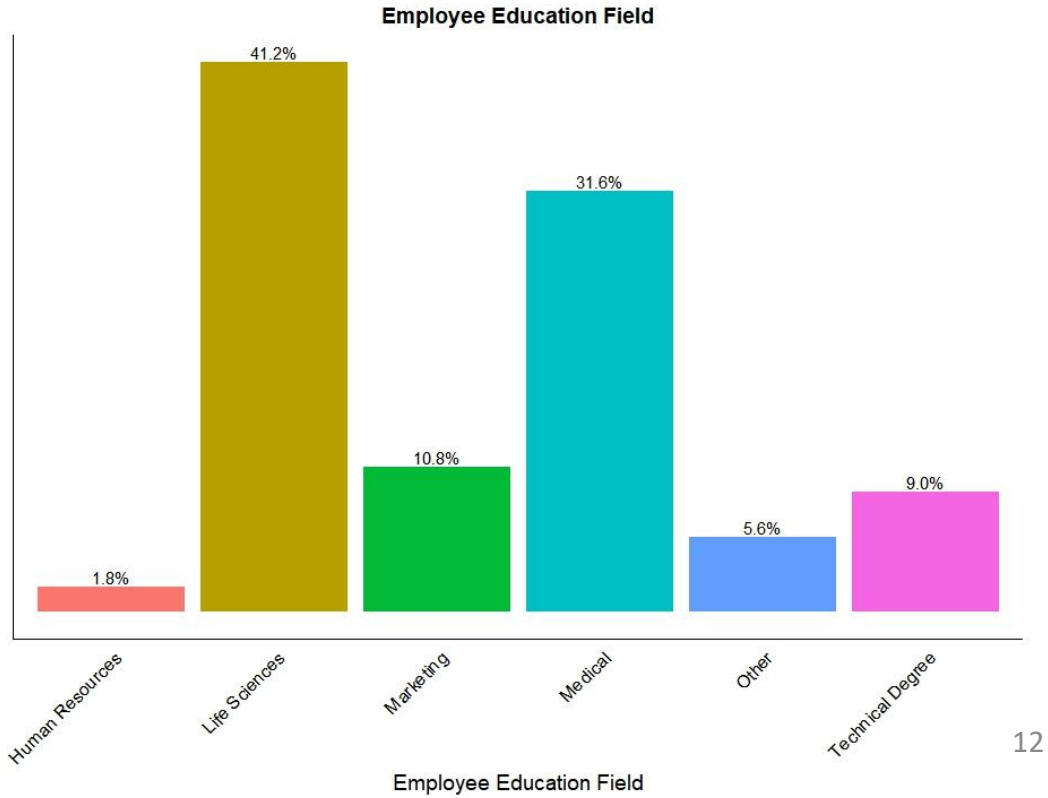


Univariate Analysis – Education and Education Field

Education	Count	Percentage
Below College	509	11.58%
College	843	19.19%
Bachelor	1707	38.85%
Master	1191	27.11%
Doctor	144	3.28%

EducationField	Count	Percentage
Human Resources	80	1.82%
Life Sciences	1811	41.22%
Marketing	476	10.83%
Medical	1387	31.57%
Other	245	5.58%
Technical Degree	395	8.99%

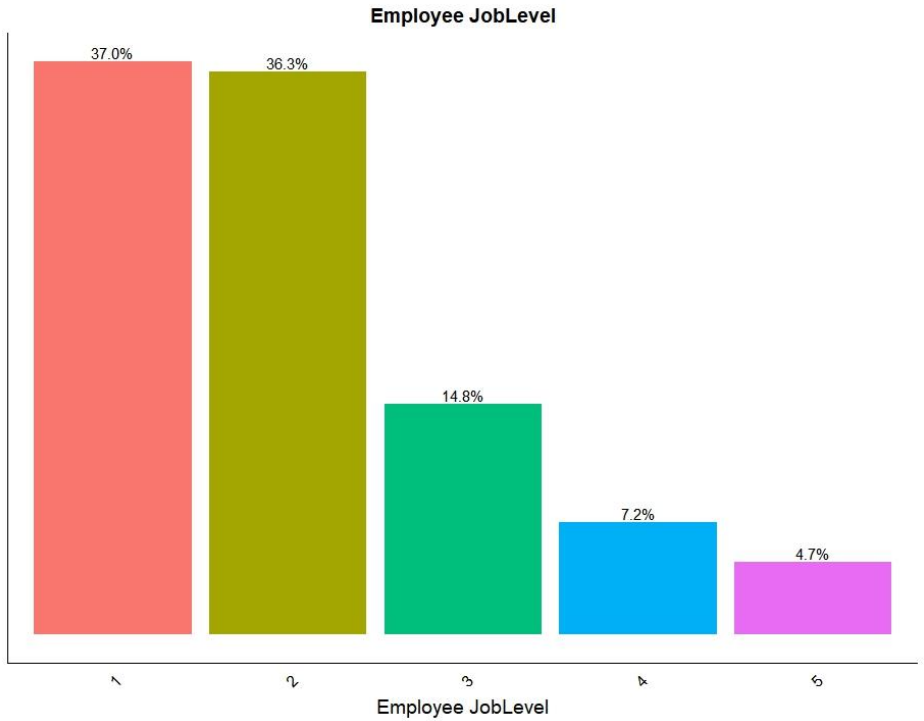


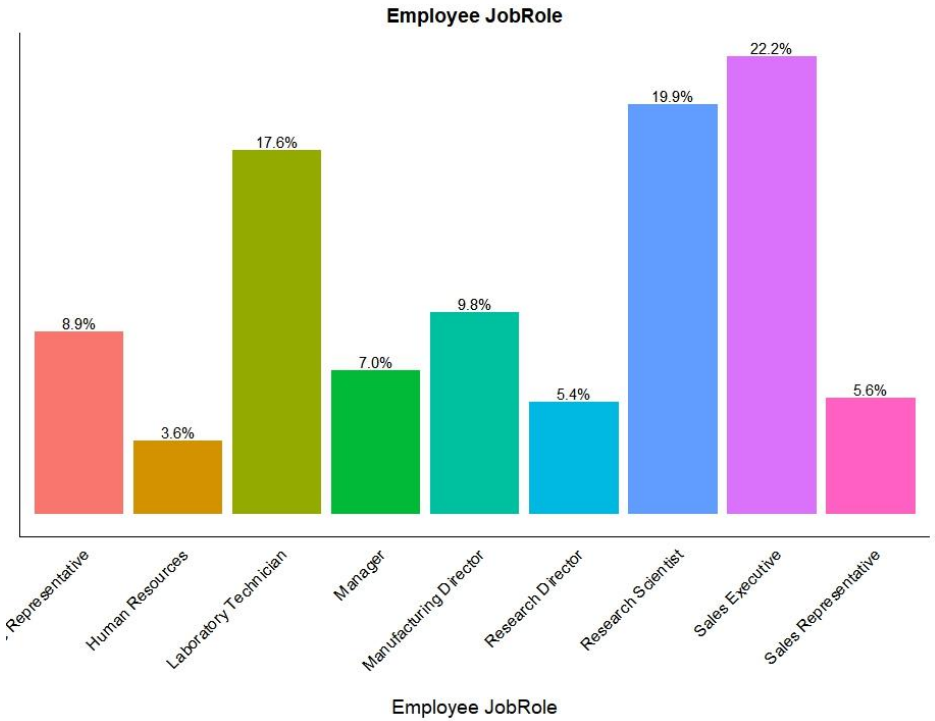


Univariate Analysis – Job Level and Job Role

JobLevel	Count	Percentage
1	1624	36.96%
2	1595	36.30%
3	652	14.84%
4	318	7.24%
5	205	4.67%

JobRole	Count	Percentage
Sales Executive	977	22.23%
Research Scientist	874	19.89%
Laboratory Technician	775	17.64%
Manufacturing Director	431	9.81%
Healthcare Representative	390	8.88%
Manager	306	6.96%
(Other)	641	14.59%

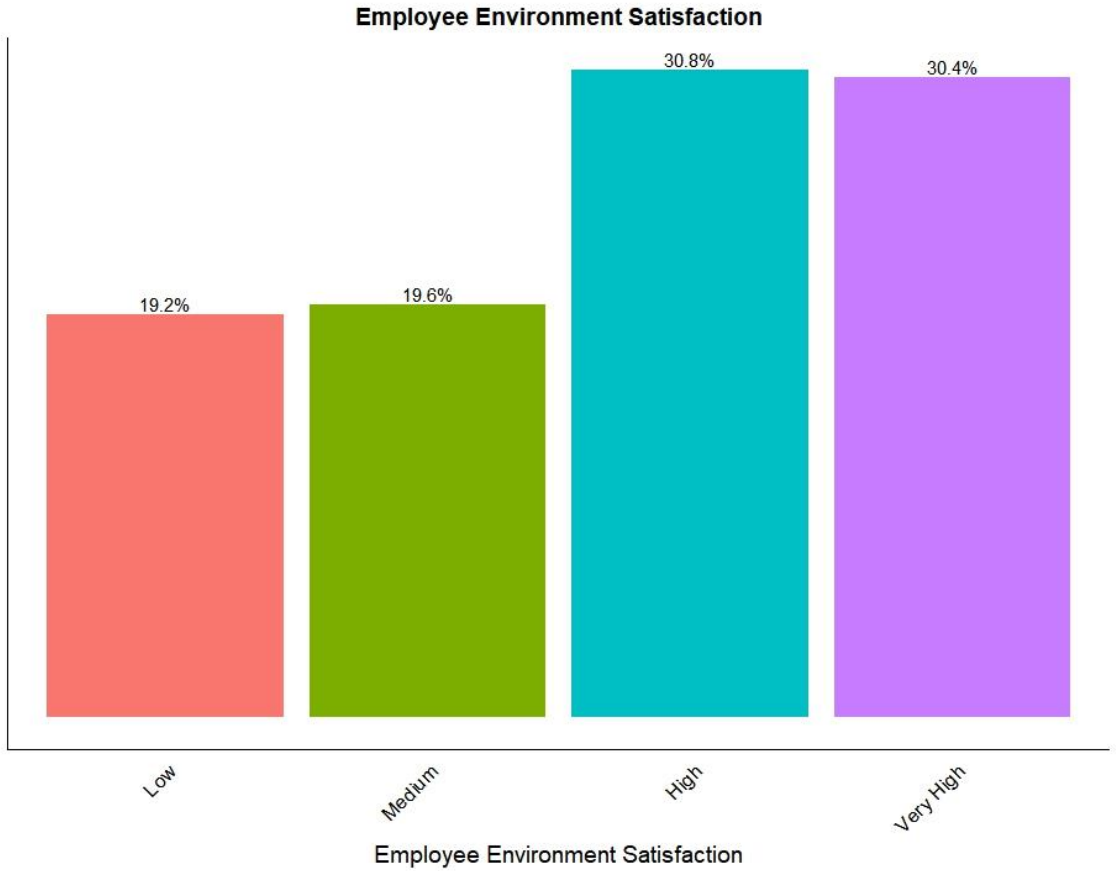


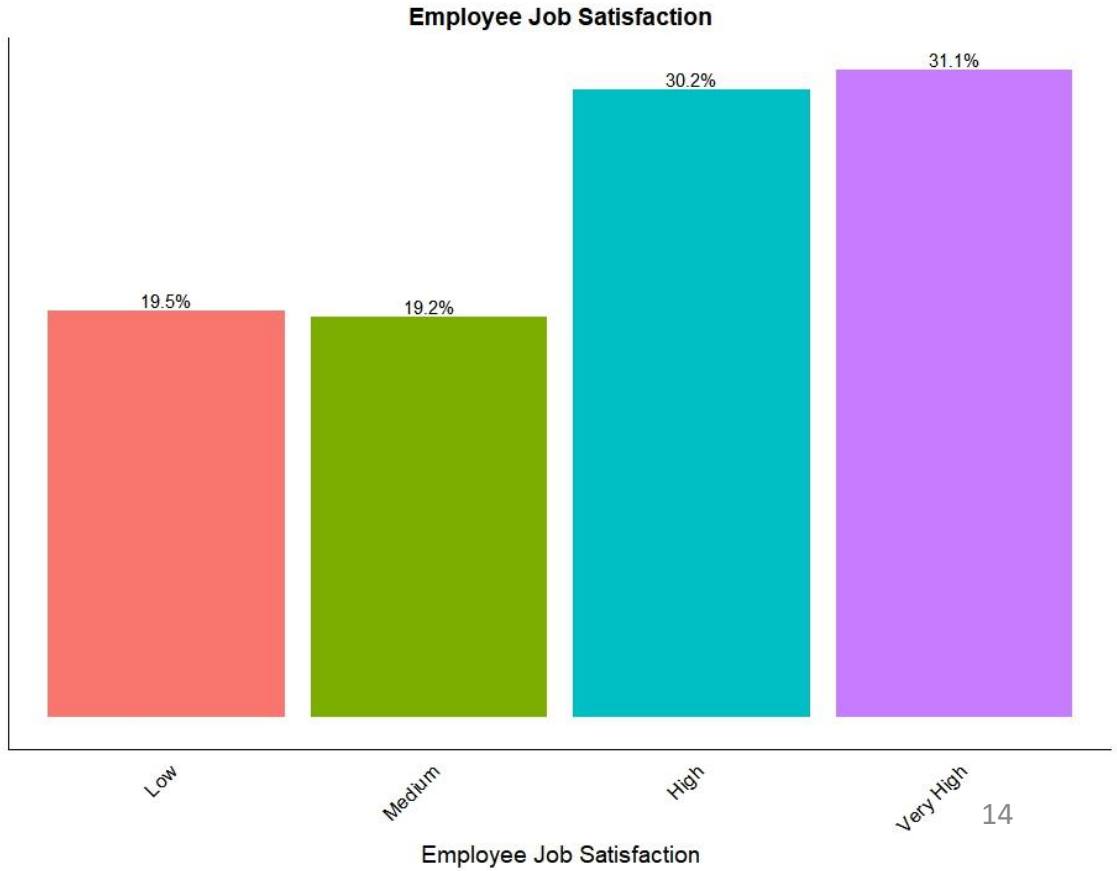


Univariate Analysis –Environment & Job Satisfaction

EnvironmentSatisfaction	Count	Percentage
Low	843	19.19%
Medium	862	19.62%
High	1352	30.77%
Very High	1337	30.43%

JobSatisfaction	Count	Percentage
Low	857	19.50%
Medium	845	19.23%
High	1326	30.18%
Very High	1366	31.09%

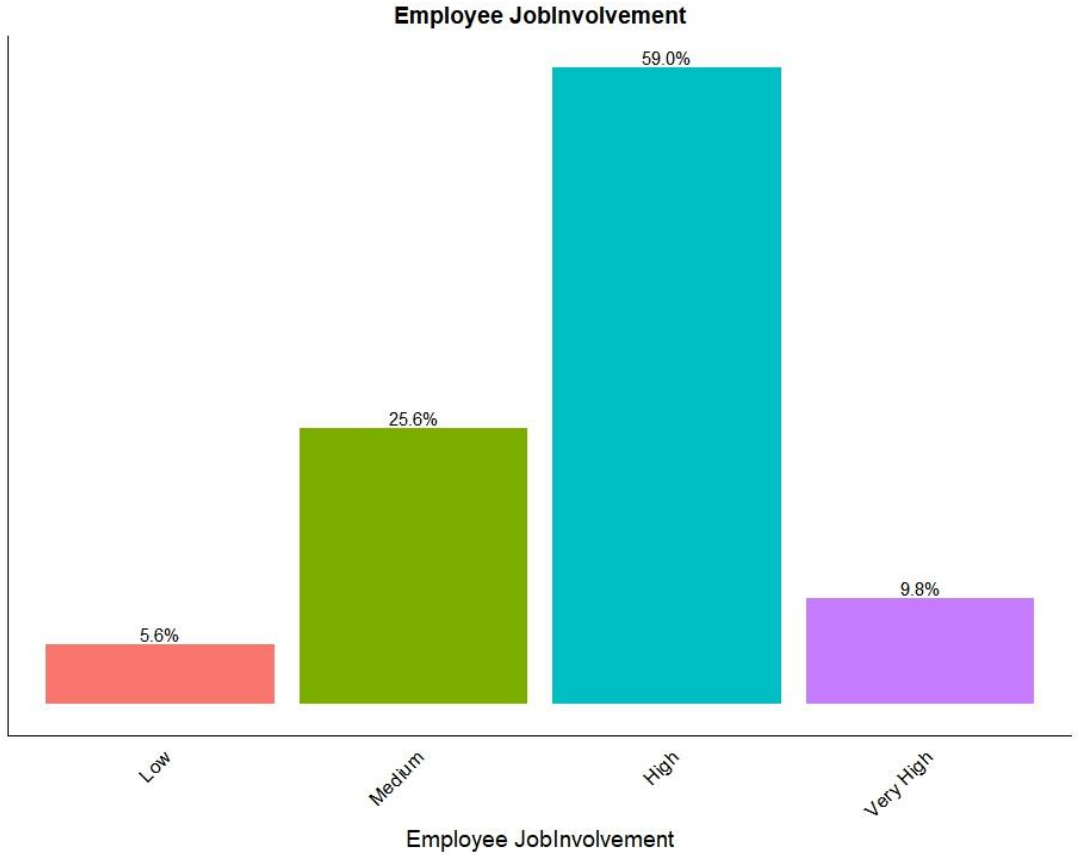


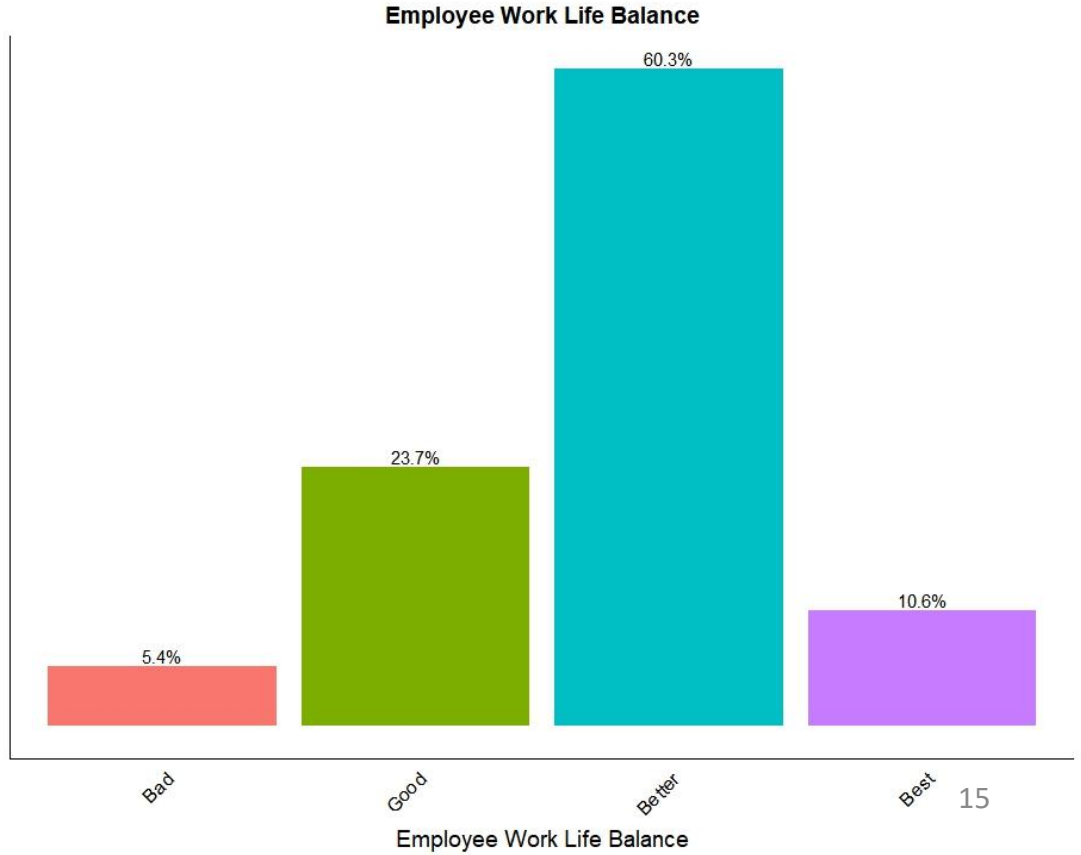


Univariate Analysis – Job Involvement & Work Life Bal

JobInvolvement	Count	Percentage
Low	245	5.58%
Medium	1124	25.58%
High	2593	59.01%
Very High	432	9.83%

WorkLifeBalance	Count	Percentage
Bad	238	5.42%
Good	1042	23.71%
Better	2649	60.29%
Best	465	10.58%

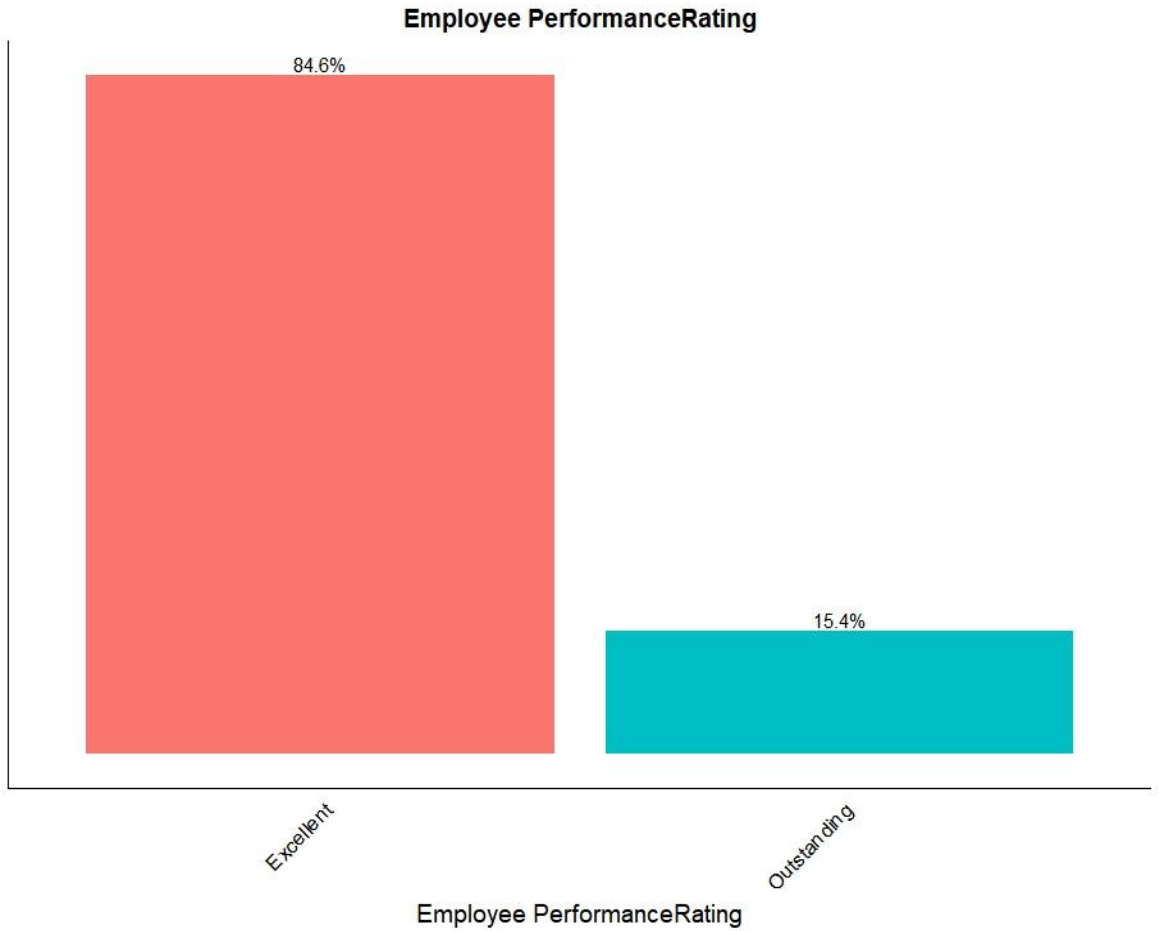


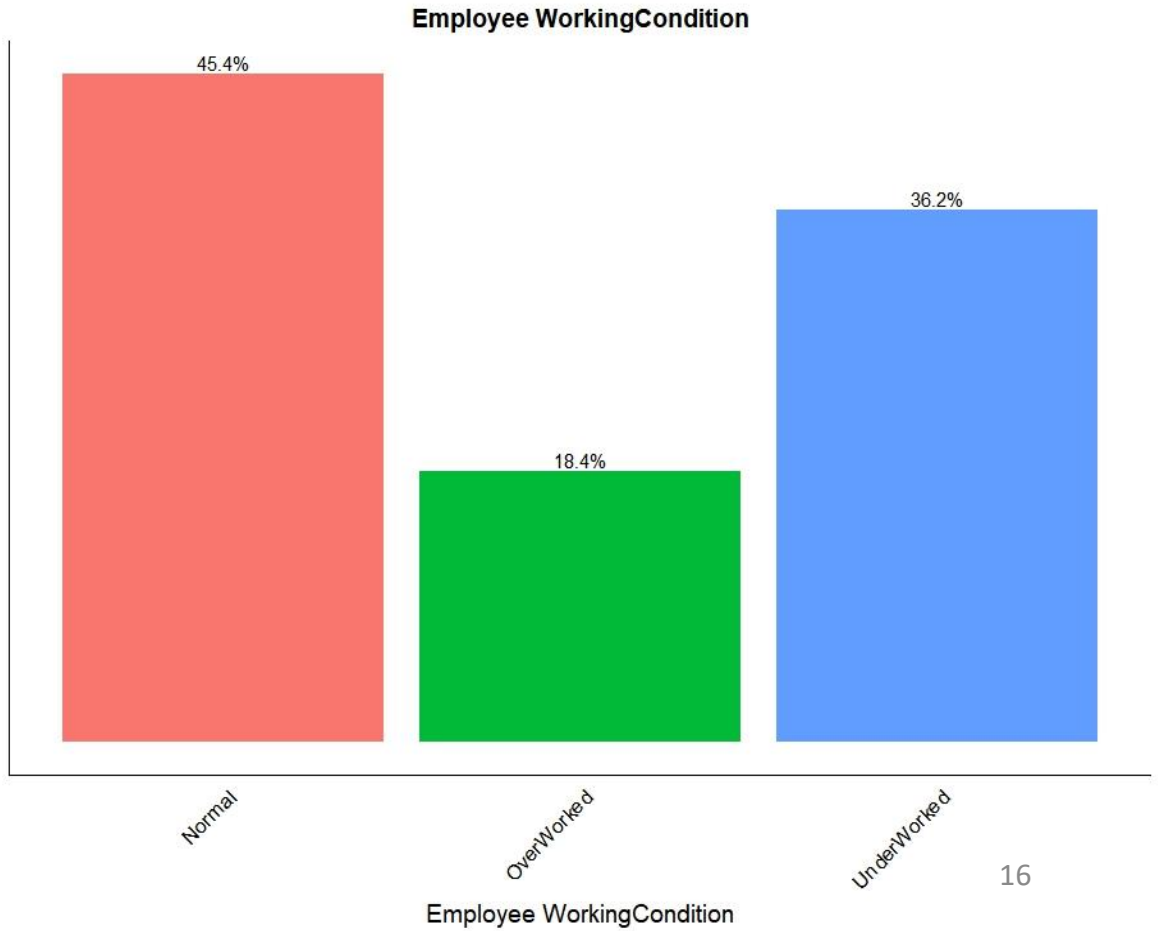


Univariate Analysis – Perf Rating and Working Cond

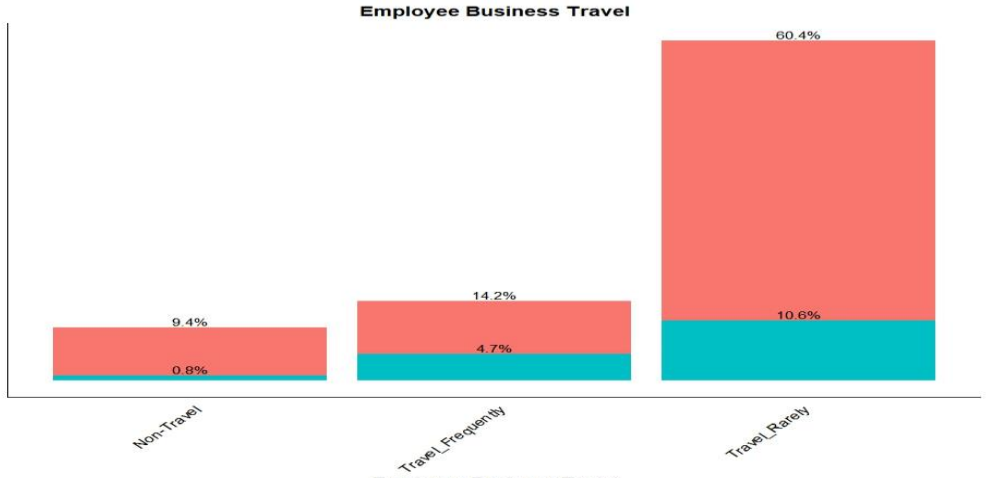
PerformanceRating	Count	Percentage
Excellent	3718	84.62%
Outstanding	676	15.38%

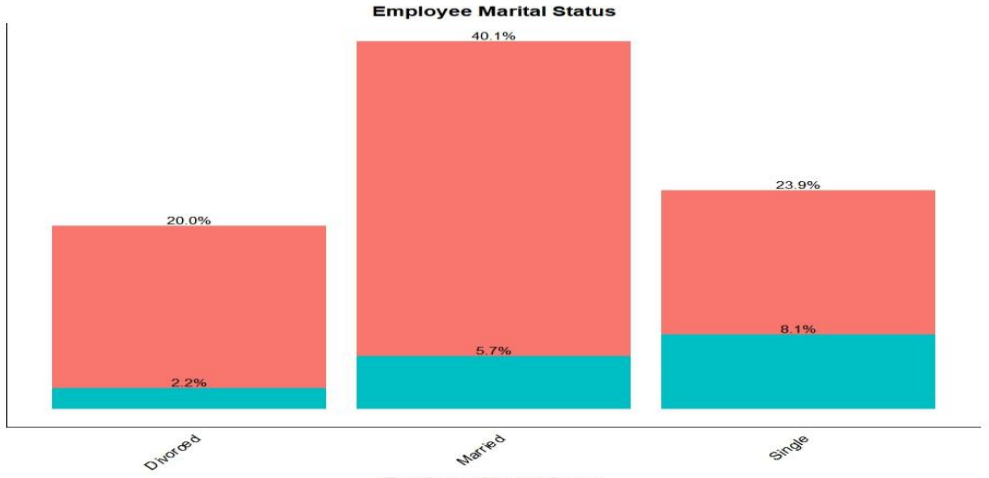
WorkingCondition	Count	Percentage
Normal	1996	45.43%
OverWorked	809	18.41%
UnderWorked	1589	36.16%

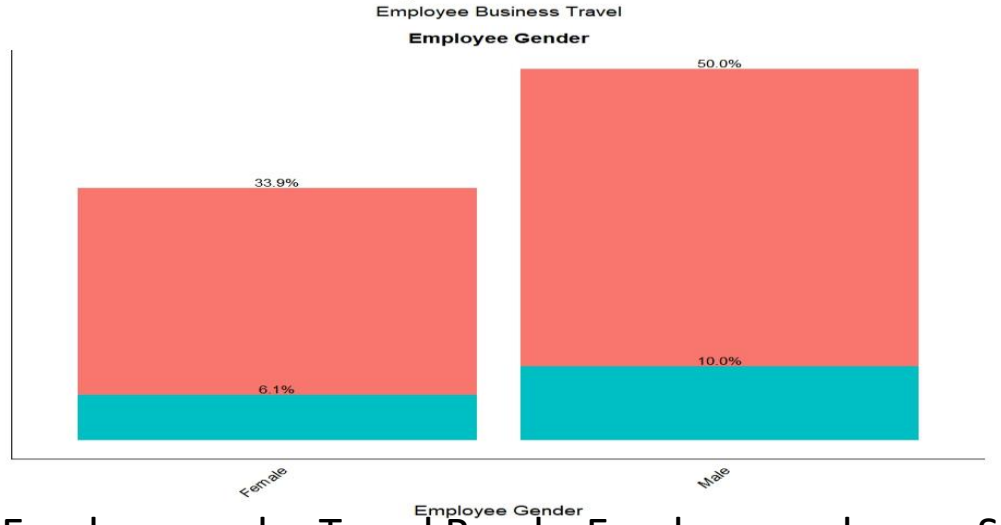


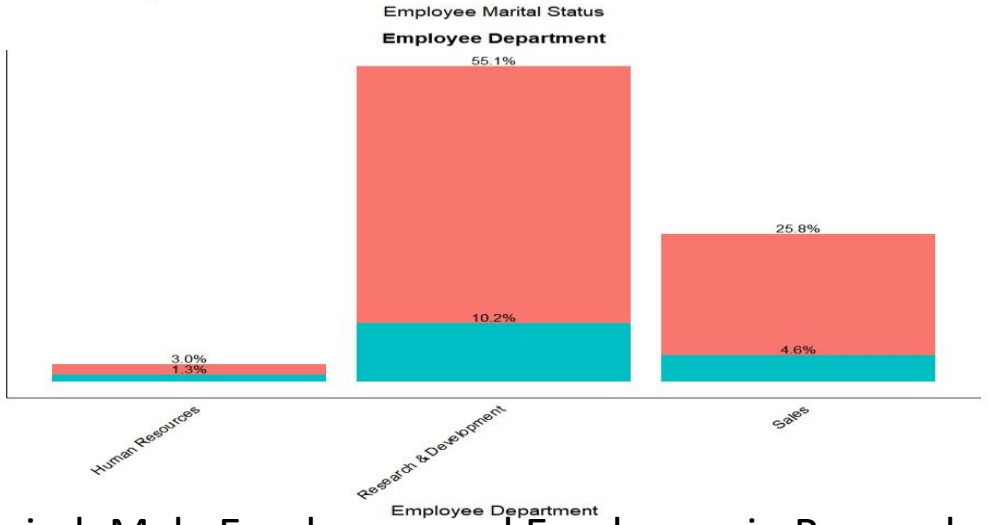


Bivariate Analysis – Against Attrition



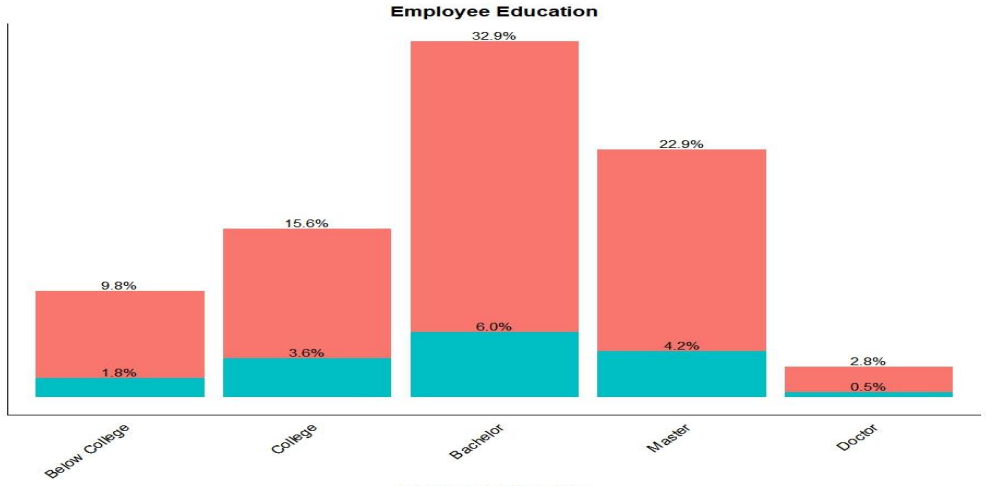


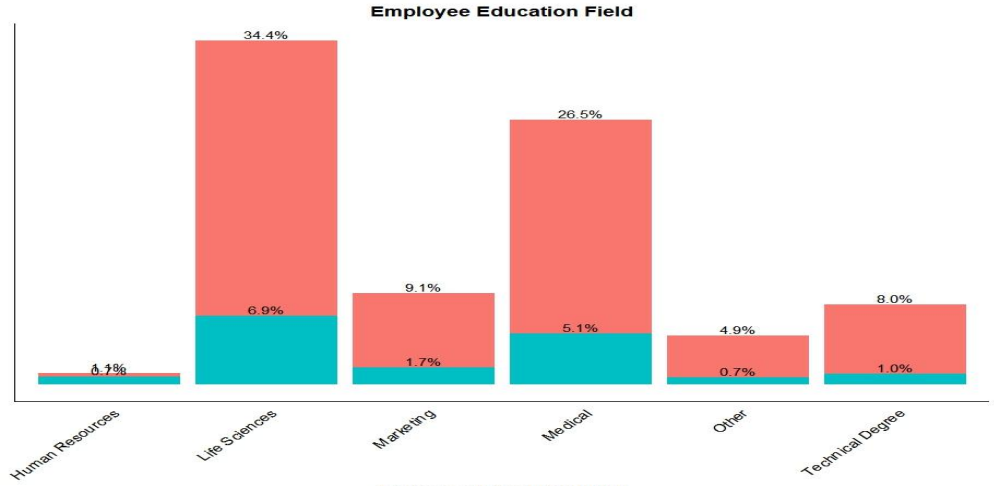


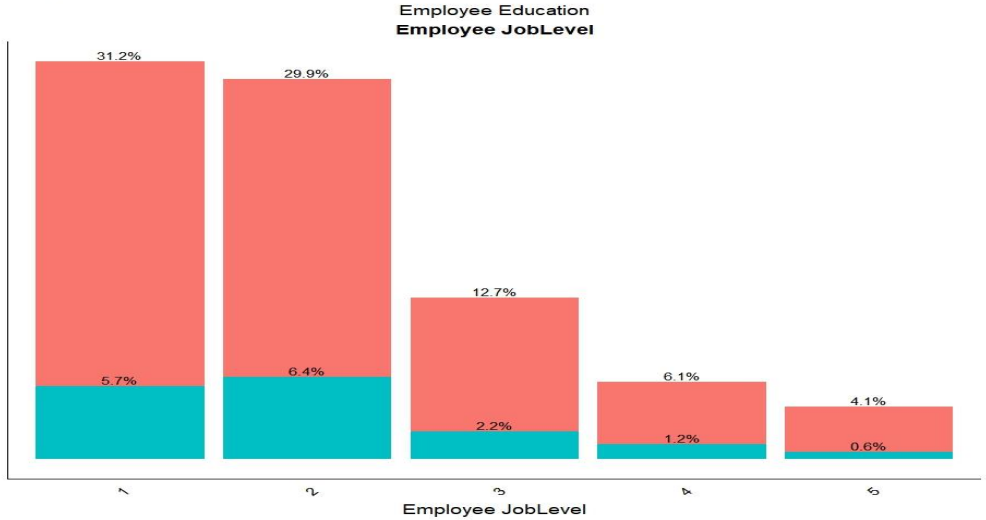


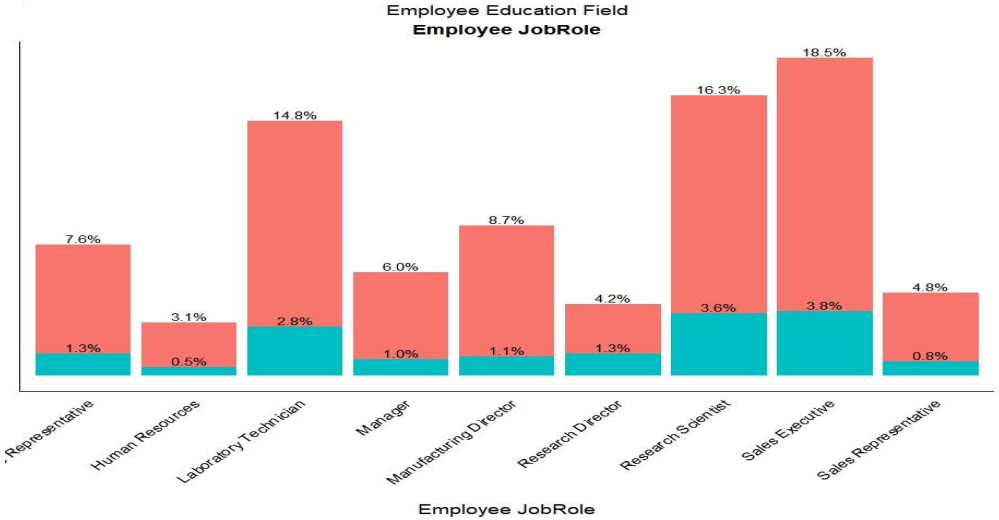
Employees who Travel Rarely, Employees who are Single/Married, Male Employees and Employees in Research and Development Department have a higher probability (>5%) of quitting the organization. (The % values shown in the graph are % of Total)

Bivariate Analysis – Against Attrition



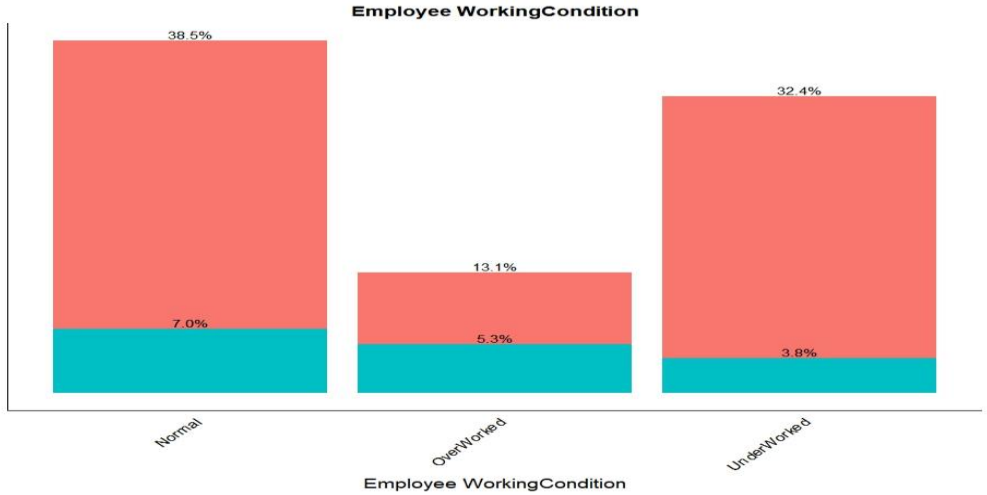


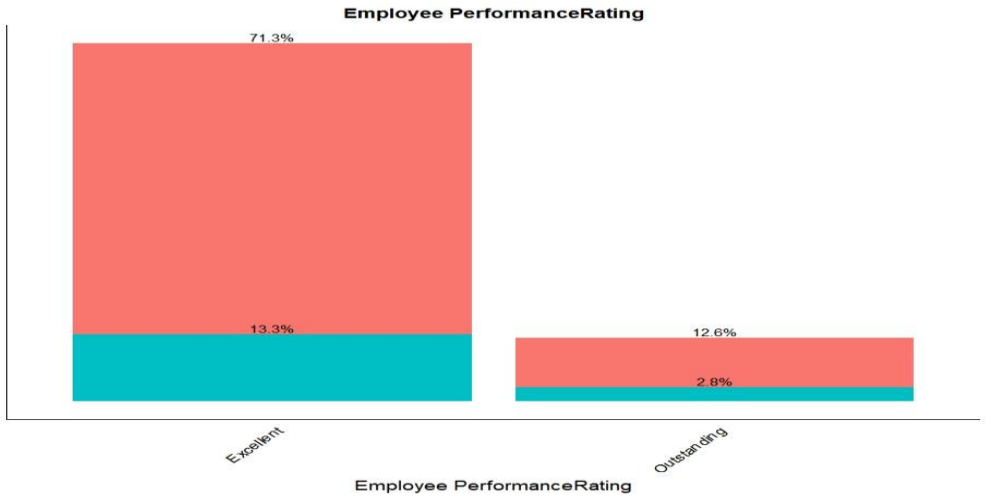


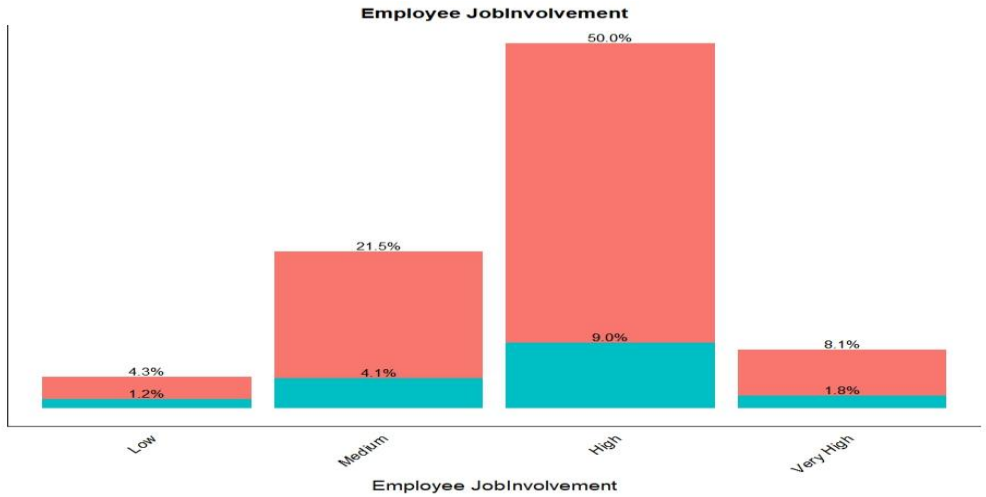


Employees who are Bachelors, Employees who have studied Life Sciences/ Medical, Employees in Job Level 1 and 2 have a higher probability (>5%) of quitting the organization. No Job Roles have greater than 5% Attrition. (The % values shown in the graph are % of Total)

Bivariate Analysis – Against Attrition

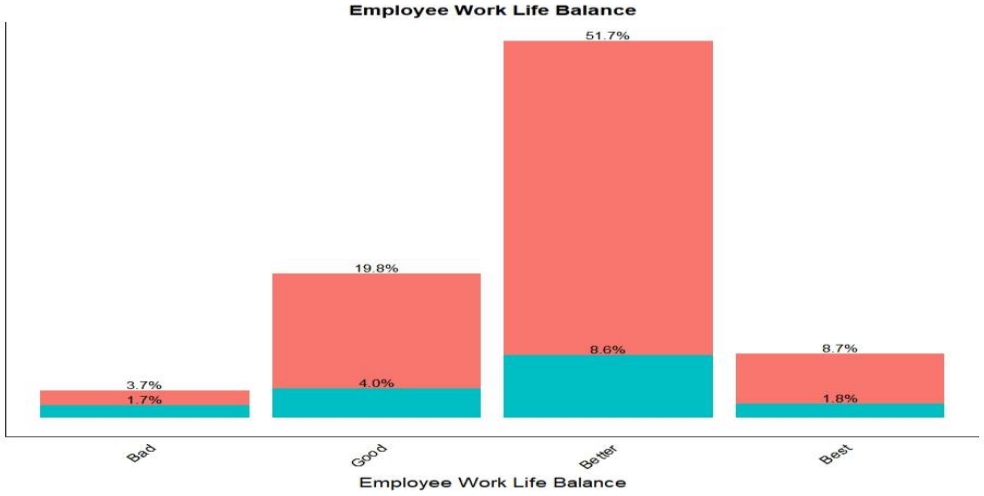


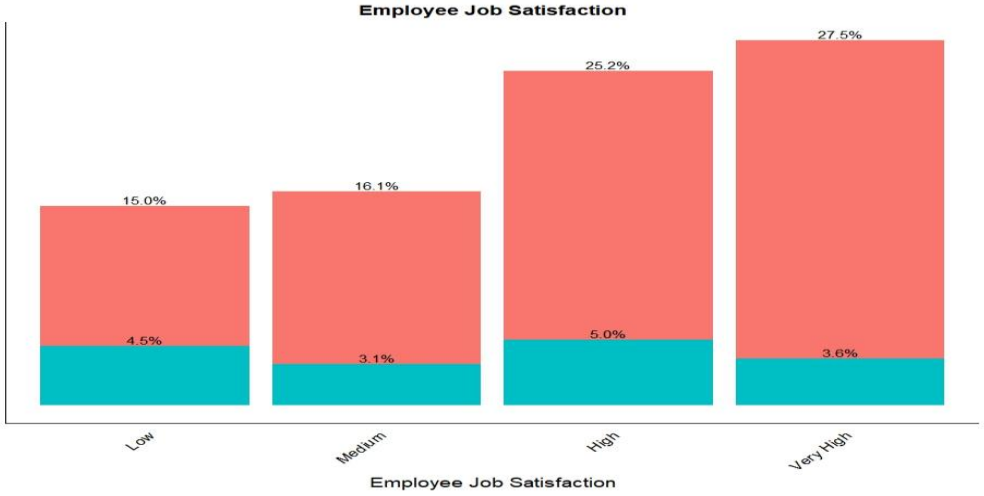


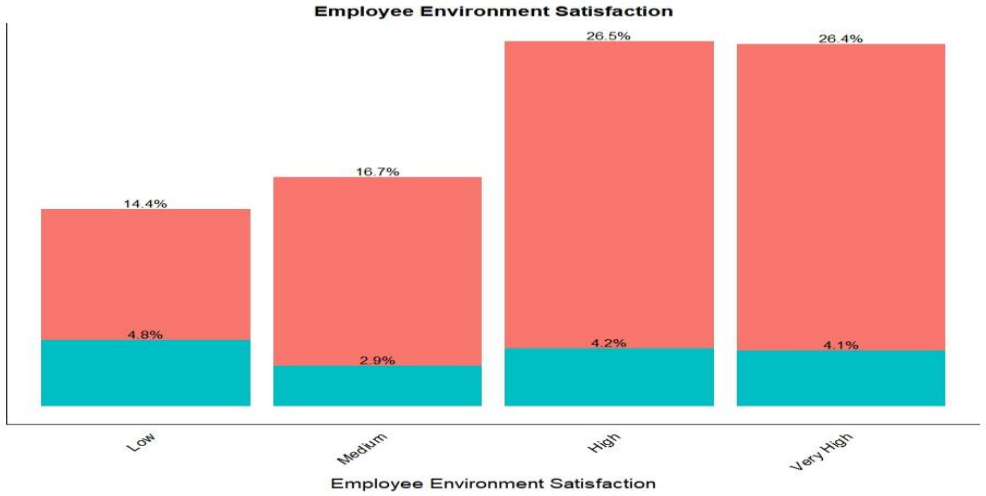


Employees with Stock Option Level 1 and 2, Employees who work between 7-9 hours or more than 9 hours, Employees with Excellent Performance Rating, and Employees with High Job Involvement have a higher probability (>5%) of quitting the organization. (The % values shown in the graph are % of Total)

Bivariate Analysis – Against Attrition

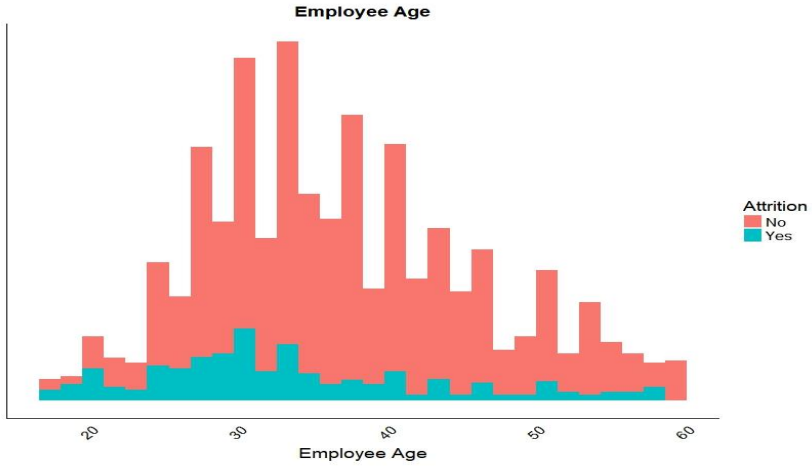


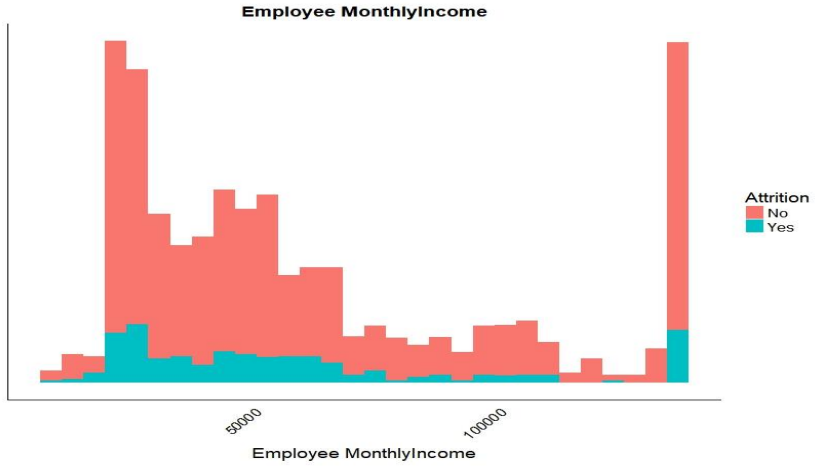


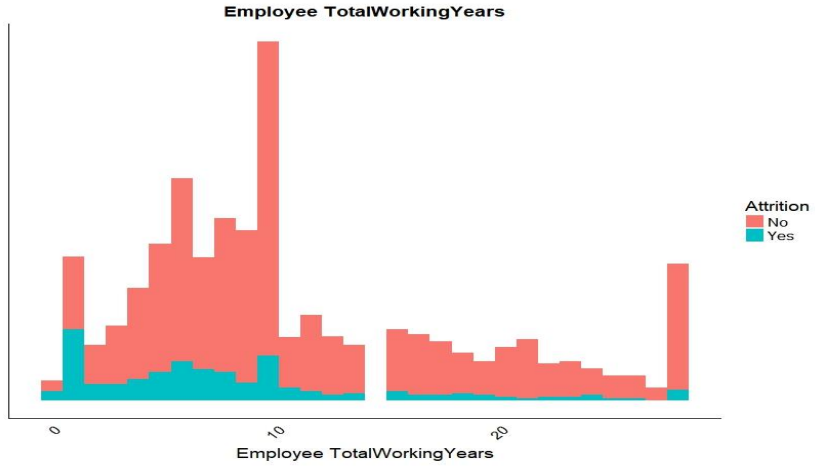


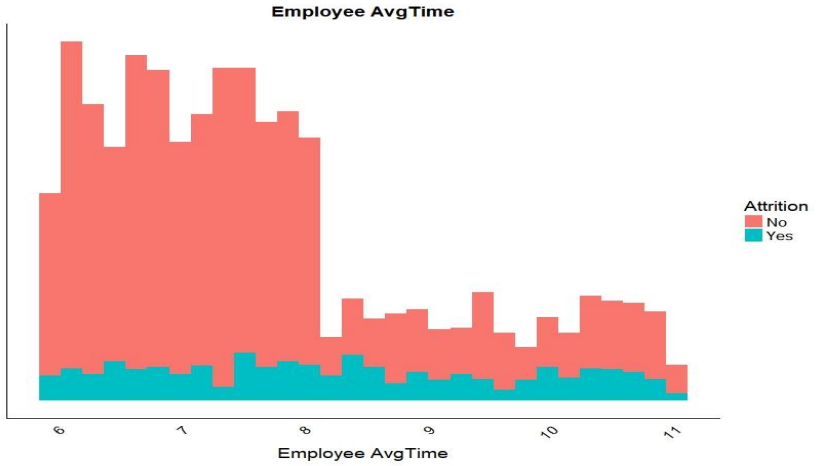
Employees with Better Work Life balance, and Employees with High Job Satisfaction have a higher probability (>5%) of quitting the organization. (Looks Counter intuitive) (The % values shown in the graph are % of Total)

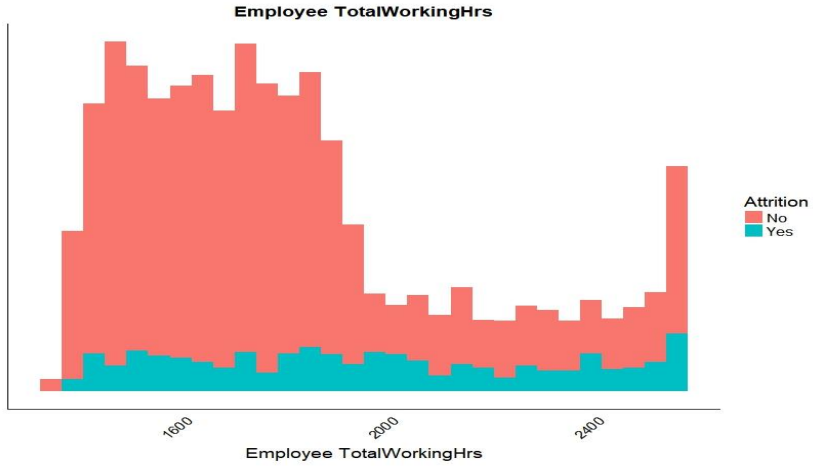
Bivariate Analysis – Against Attrition





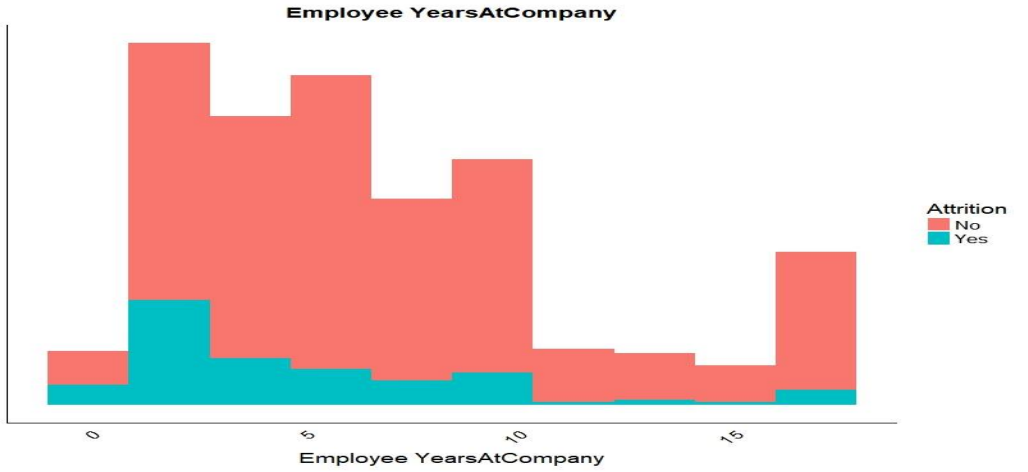


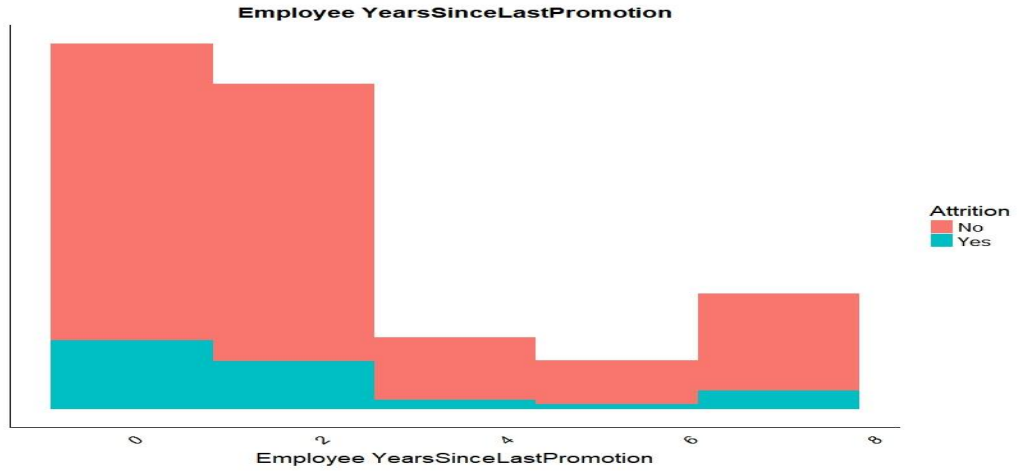


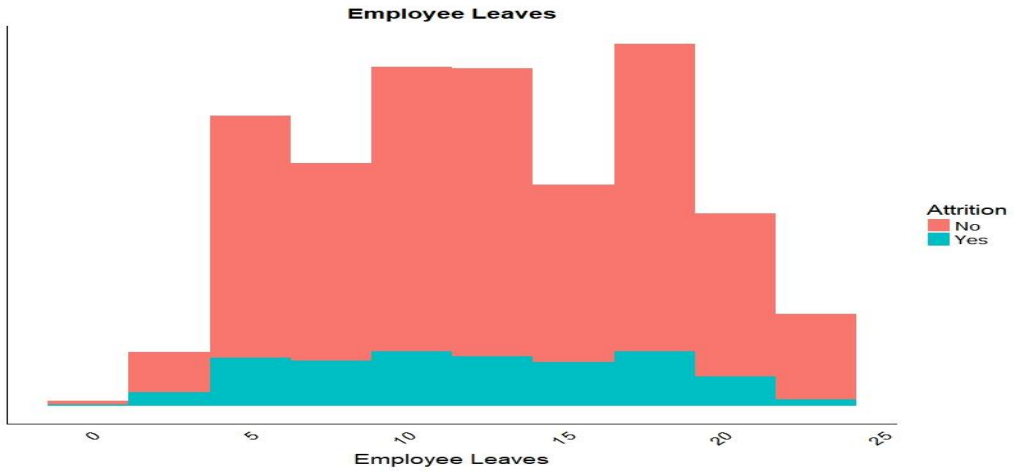


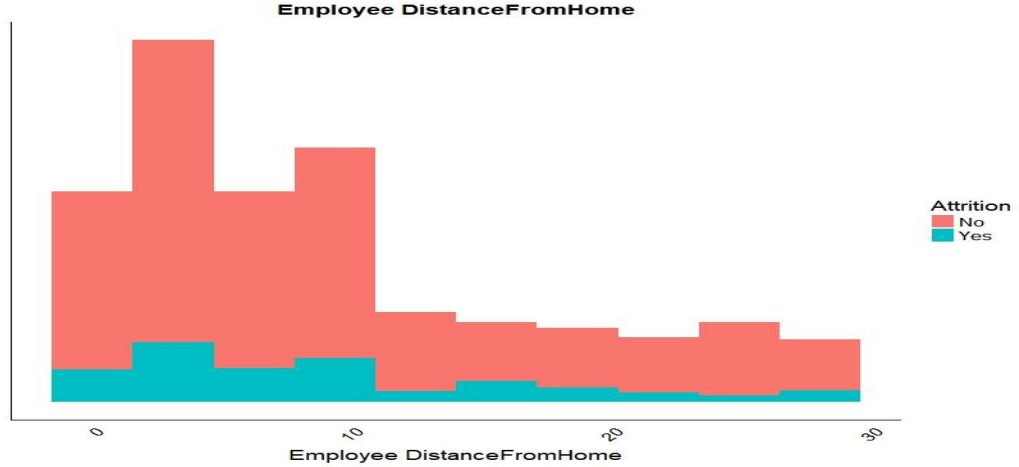
Young Employees, Employees with less Monthly income, Employees with less number of working years have relatively higher Attrition.

Bivariate Analysis – Against Attrition



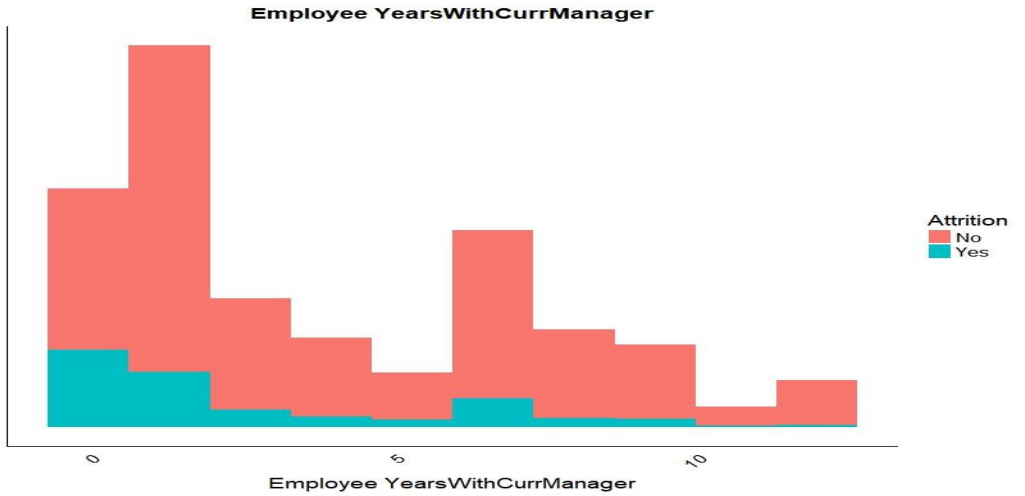


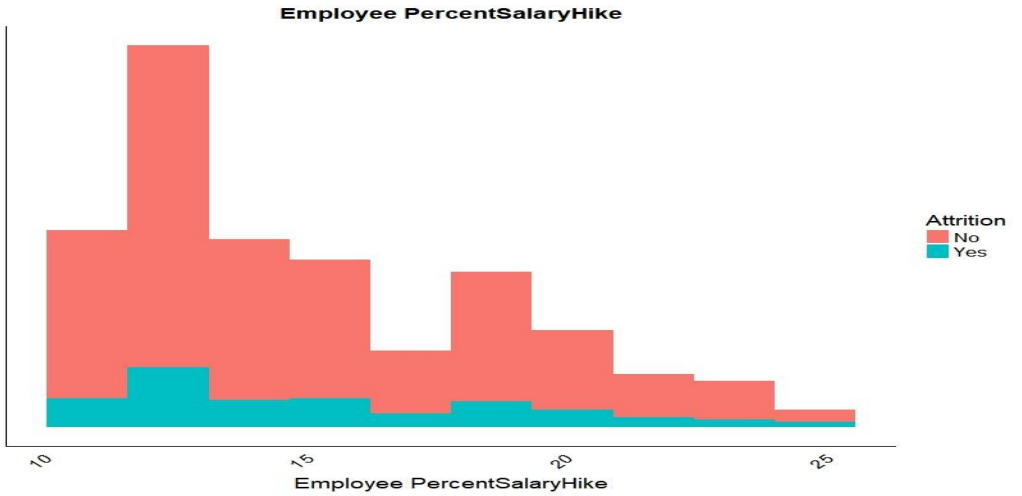


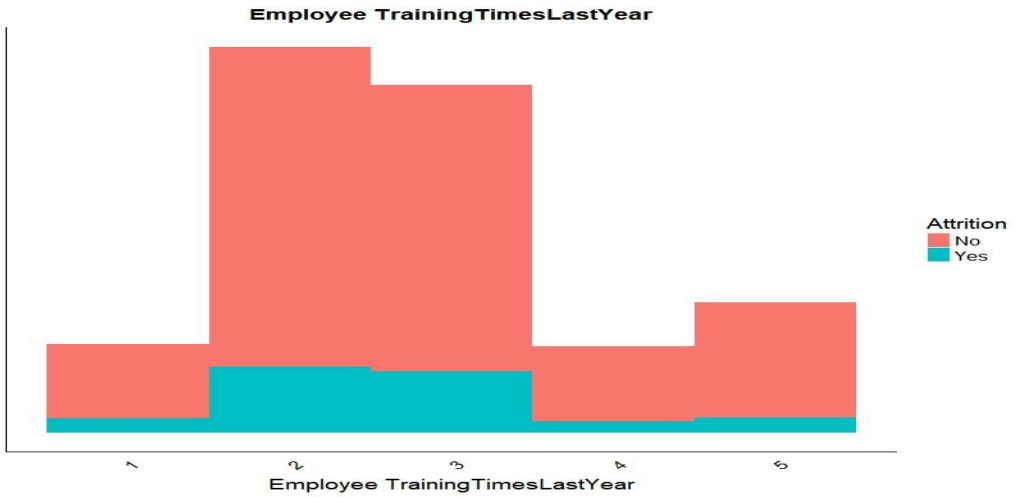


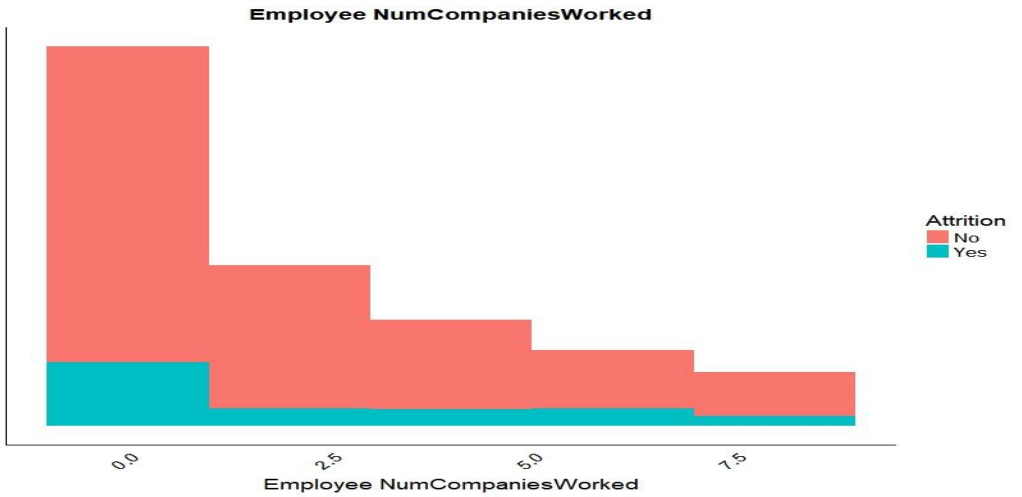
Employees who have worked in the company for fewer years, Employees who had their promotion recently and employees who live closer have a more likelihood of leaving.

Bivariate Analysis – Against Attrition



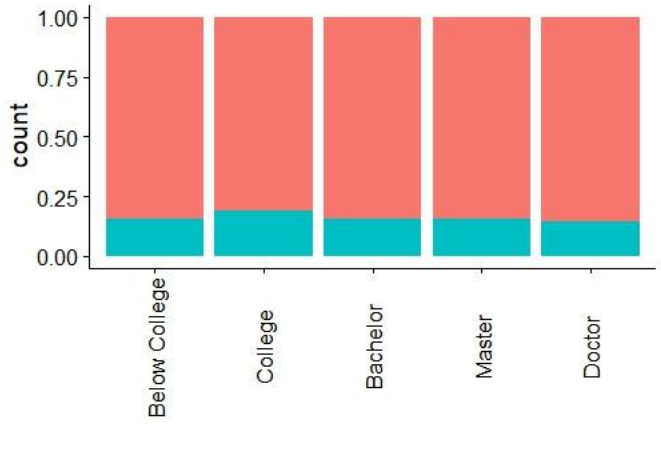


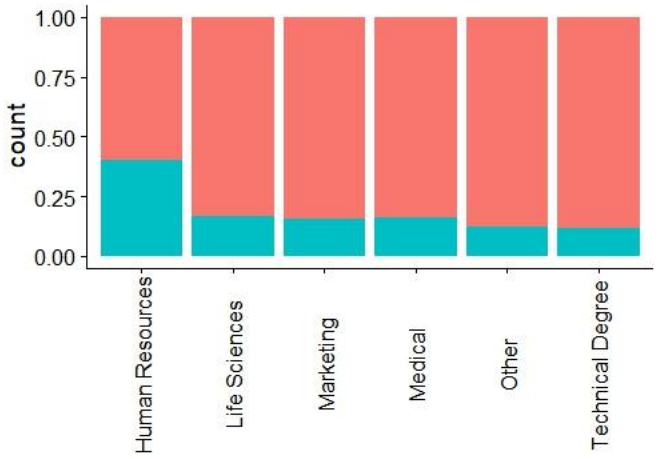


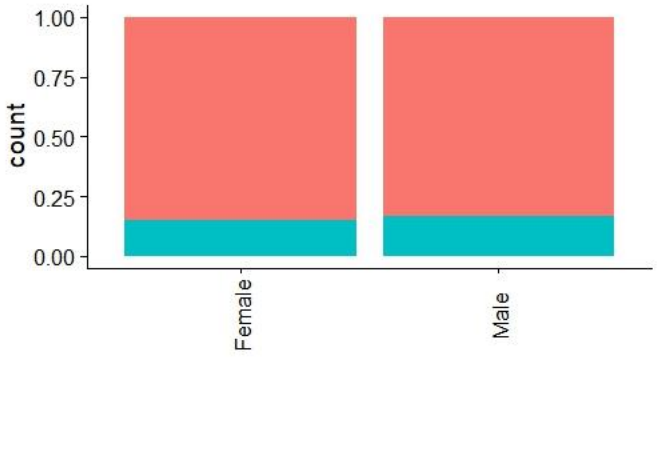


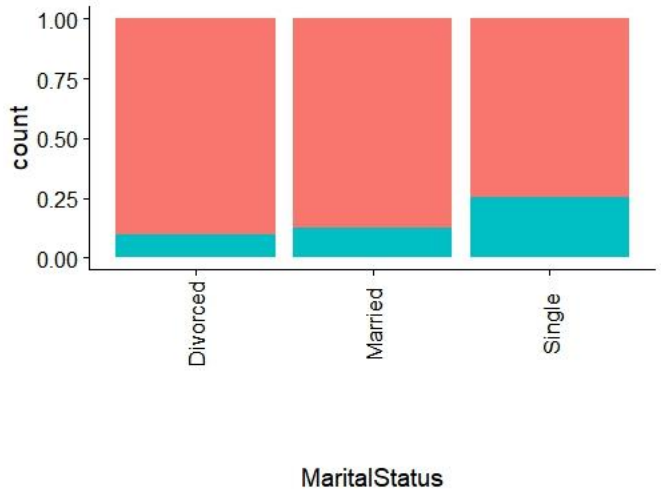
Employees who have new managers, Employees with lower Salary Hike, Employees with average number of trainings and employees who have worked in fewer companies are more likely to leave.

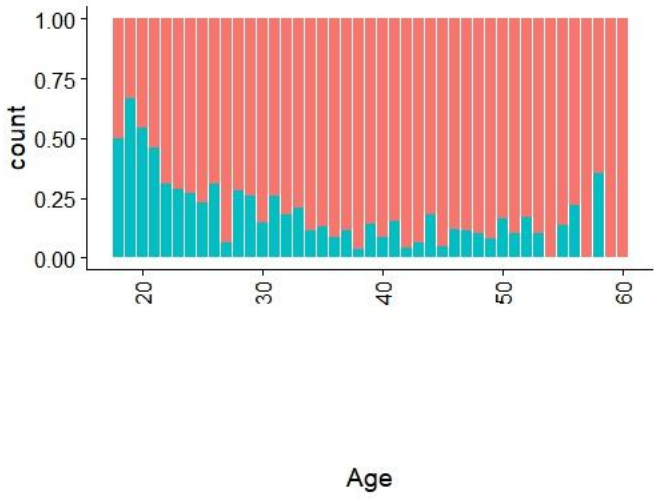
Bivariate Analysis – Personal Attributes Vs Attrition





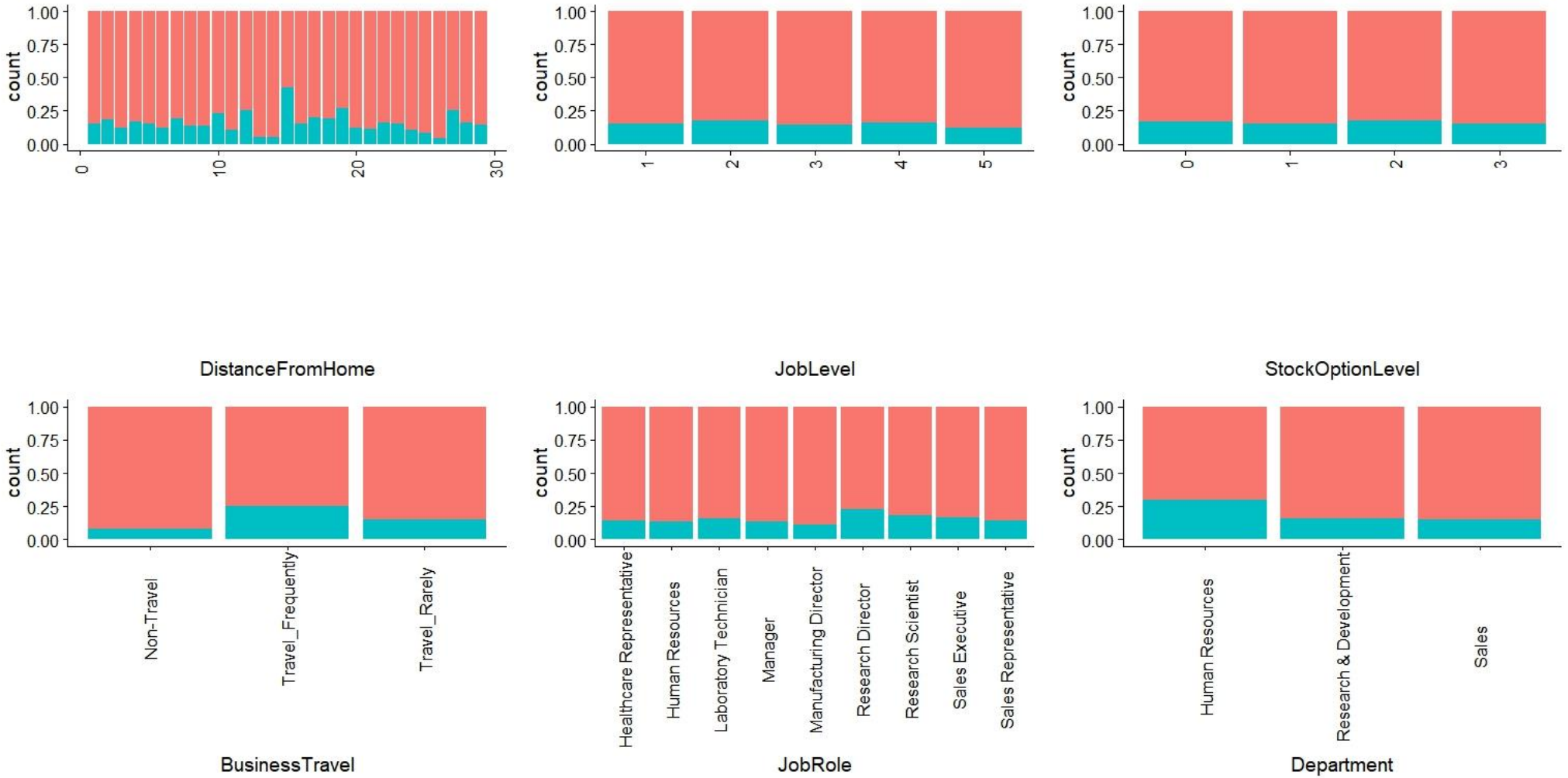






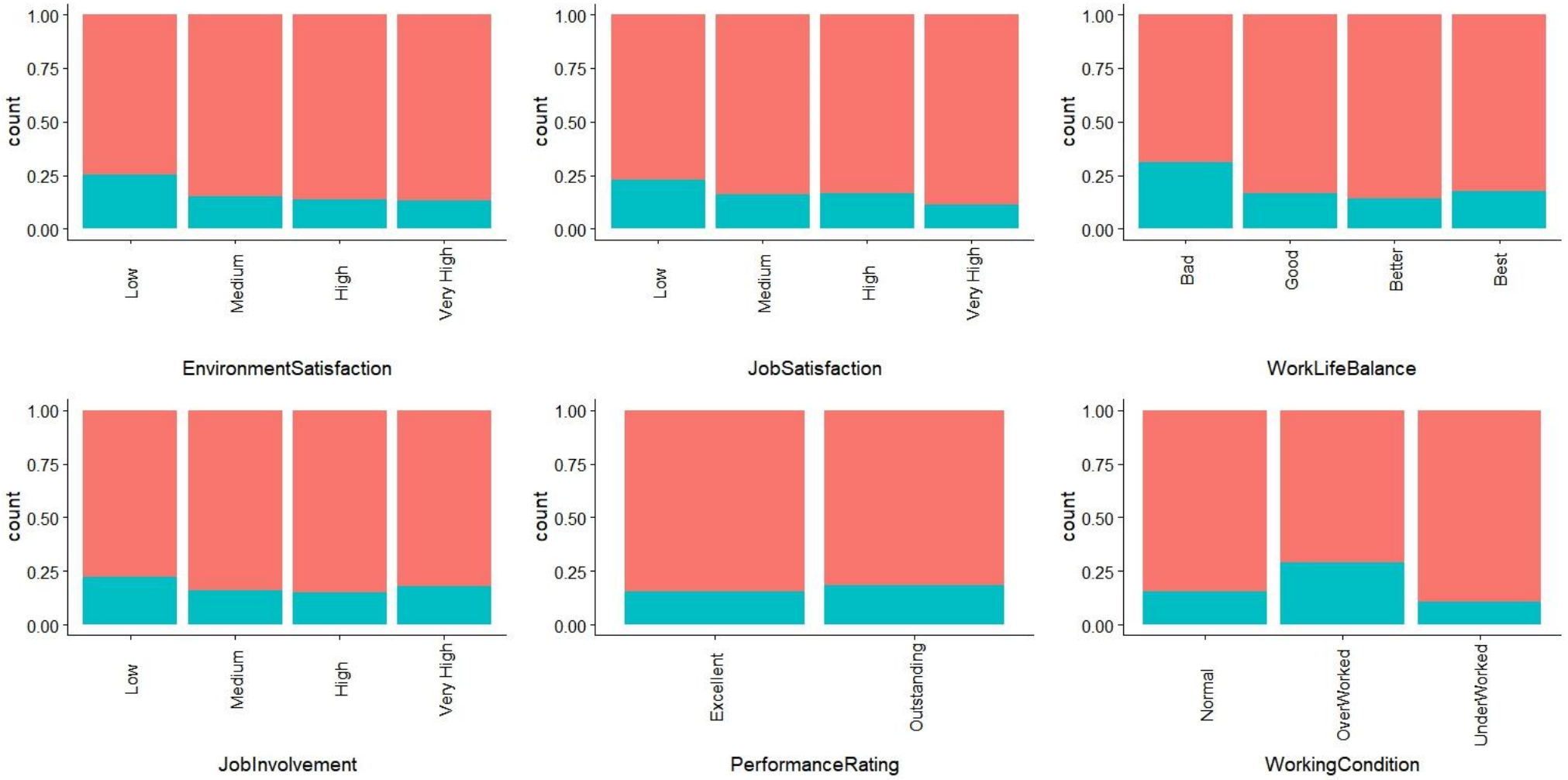
As a Percentage of total, Employees who have studied Human Resource, Single and Young Employees are more likely to leave.

Bivariate Analysis – Job Attributes Vs Attrition



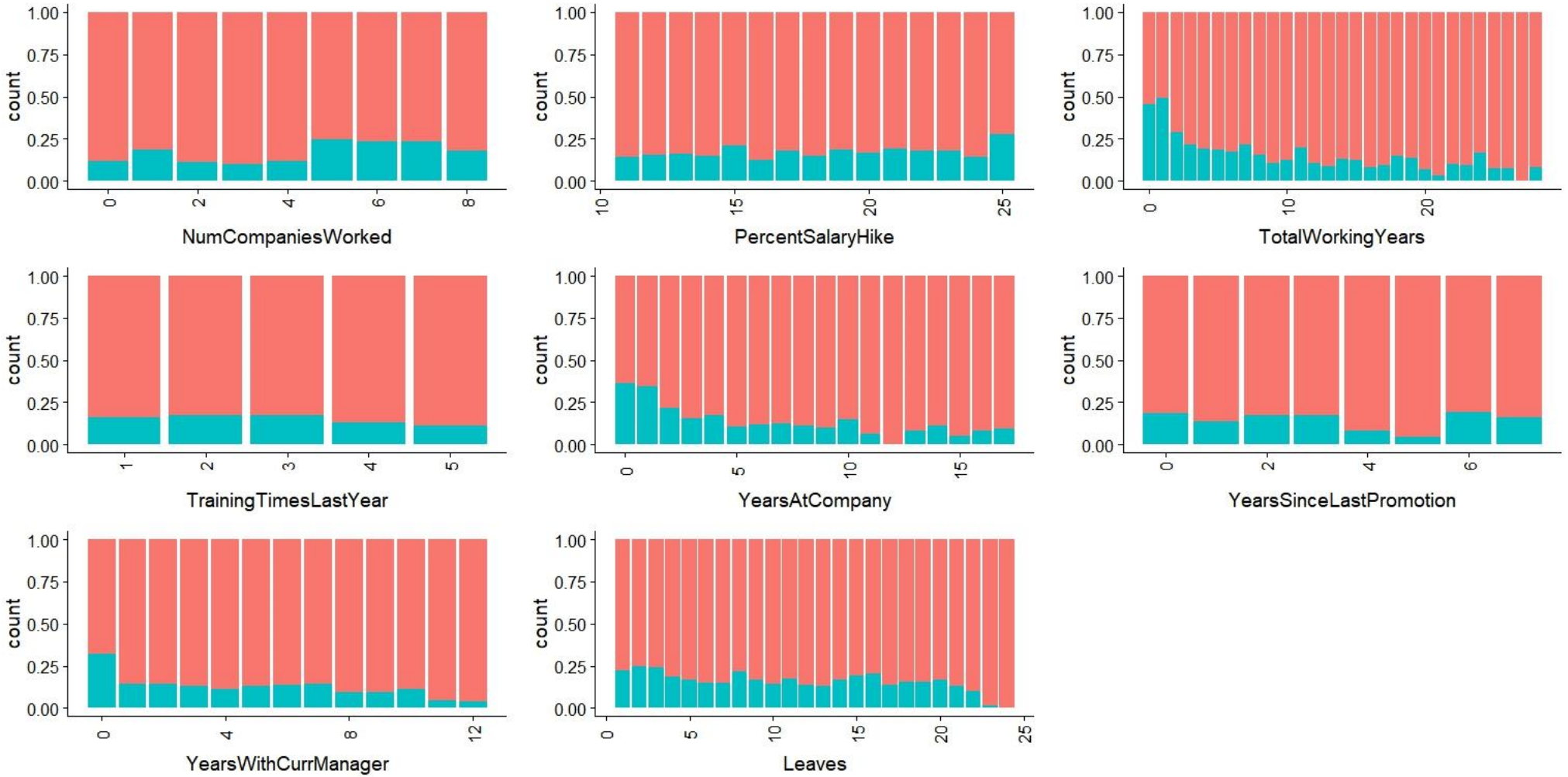
As a Percentage of total, Employees who work further from home, Employees who travel frequently, in Human Resource and Research Directors are more likely to leave.

Bivariate Analysis – Performance Attributes Vs Attrition



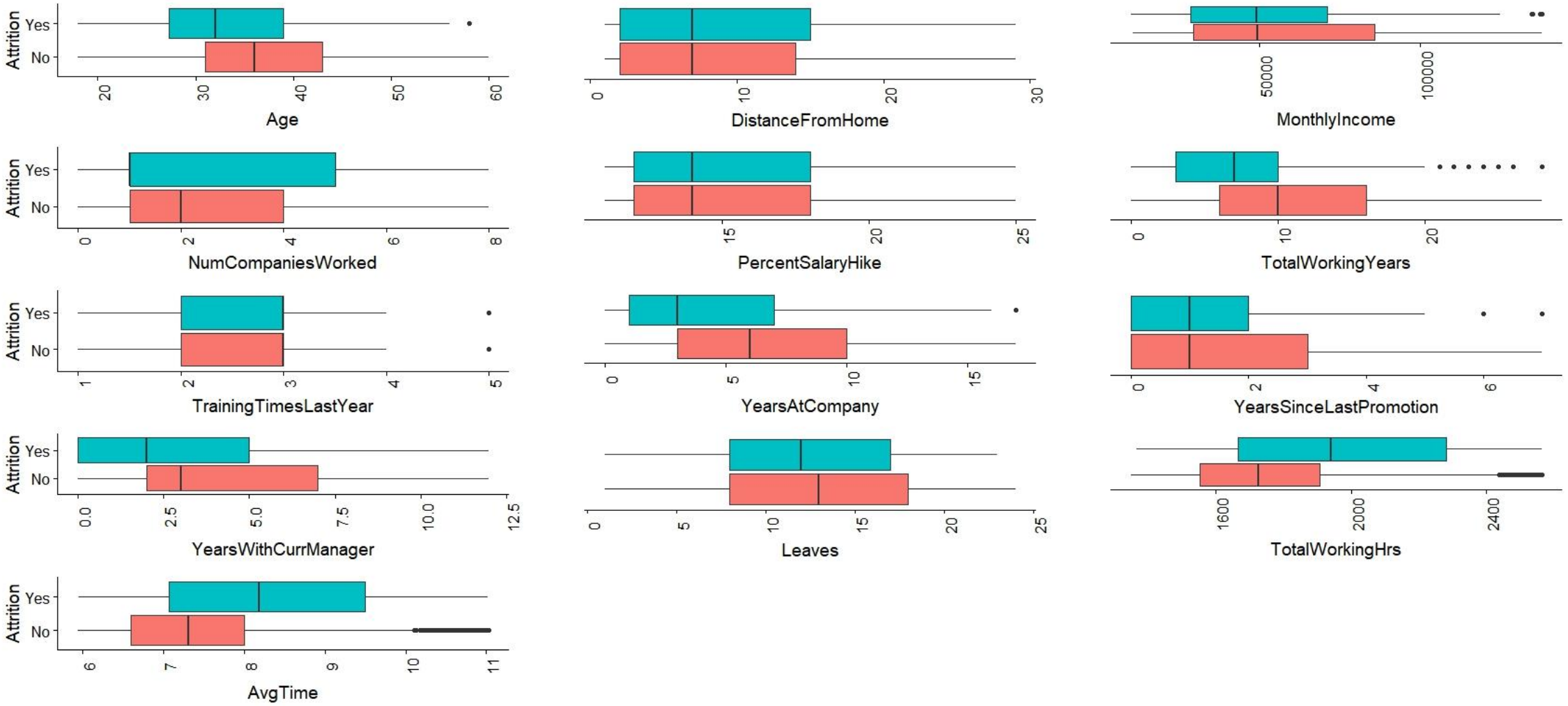
As a Percentage of total, Employees who have low Performance Attributes and Overworked employees are more likely to leave.

Bivariate Analysis – Other Attributes Vs Attrition



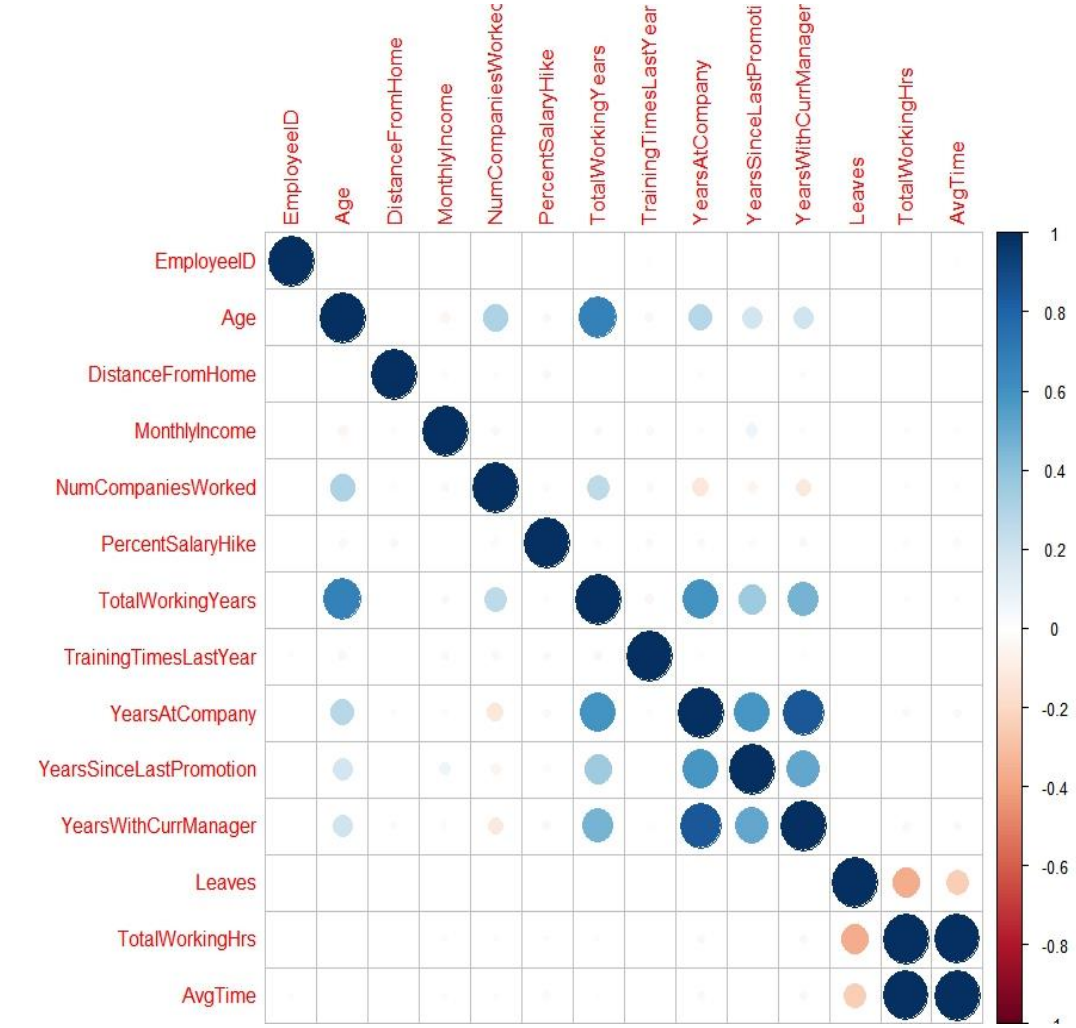
As a Percentage of total, Employees who worked in more companies, employees with less years of working experience, employees who are at the company and manager for fewer years are more likely to leave.

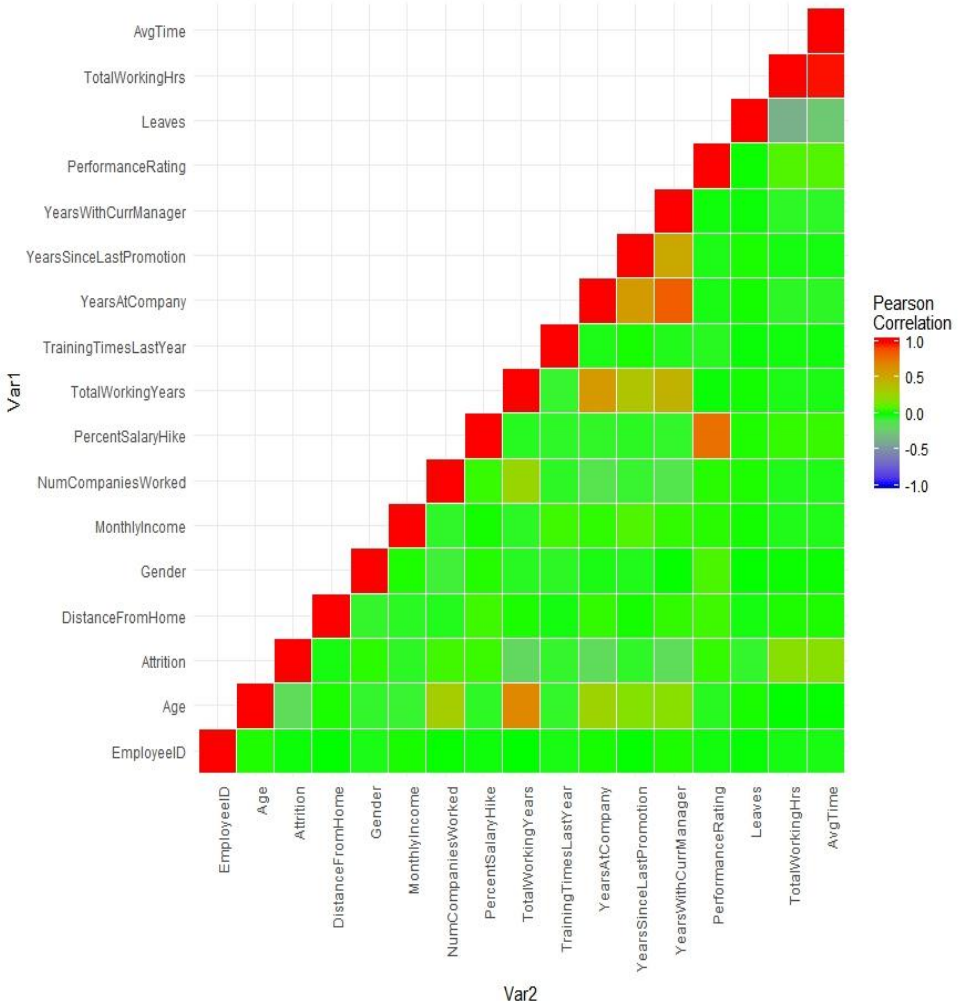
Bivariate Analysis – Box Plot of Variables wrt Attrition



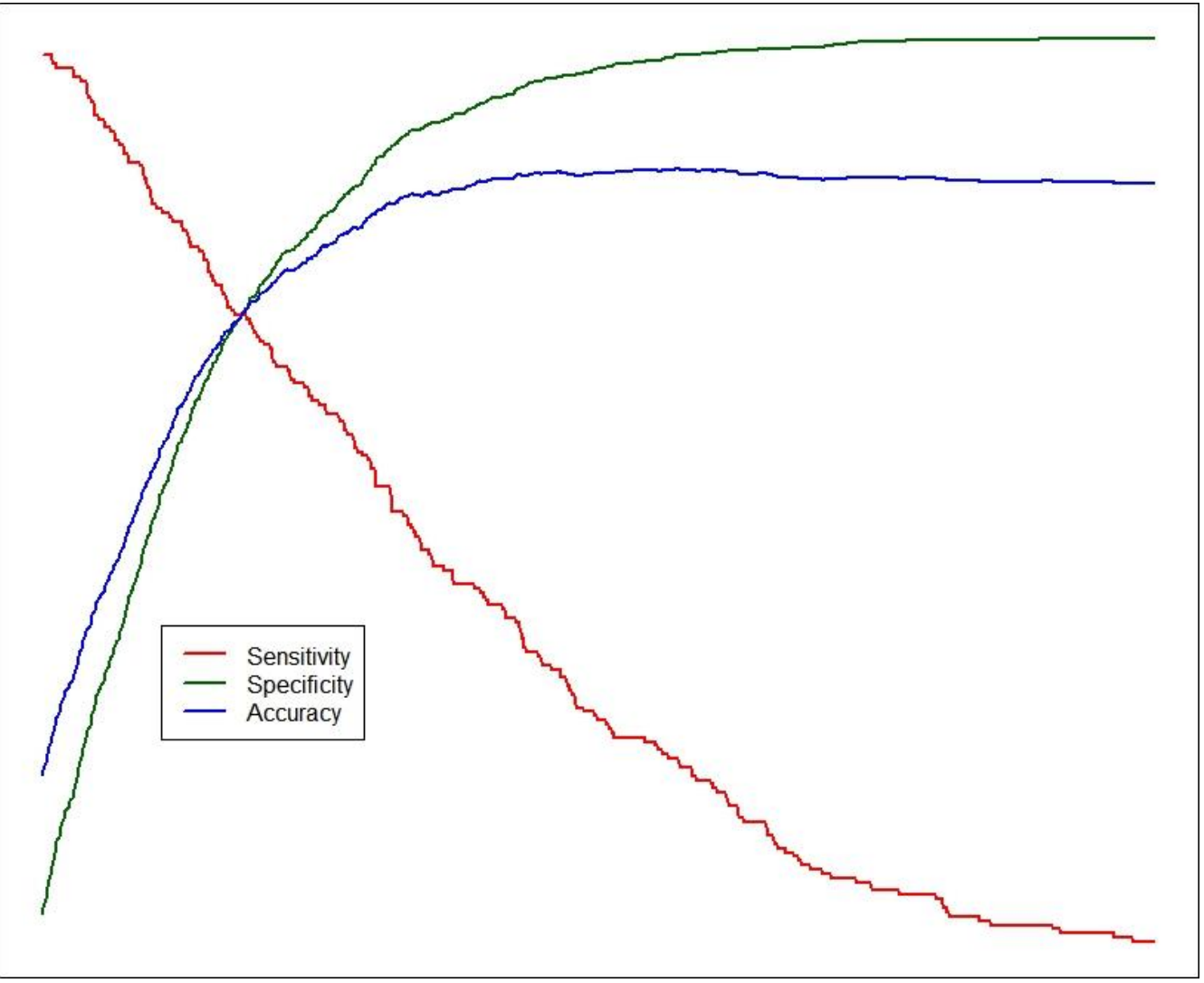
Variables that impact attrition are: Age, NumCompaniesWorked, TotalWorkingYears, YearsAtCompany, YearsWithCurrManager, Leaves, TotalWorkingHrs, AvgTime

Correlation Matrix – For all numeric columns





Model Evaluation – Deciding the Optimum Cutoff

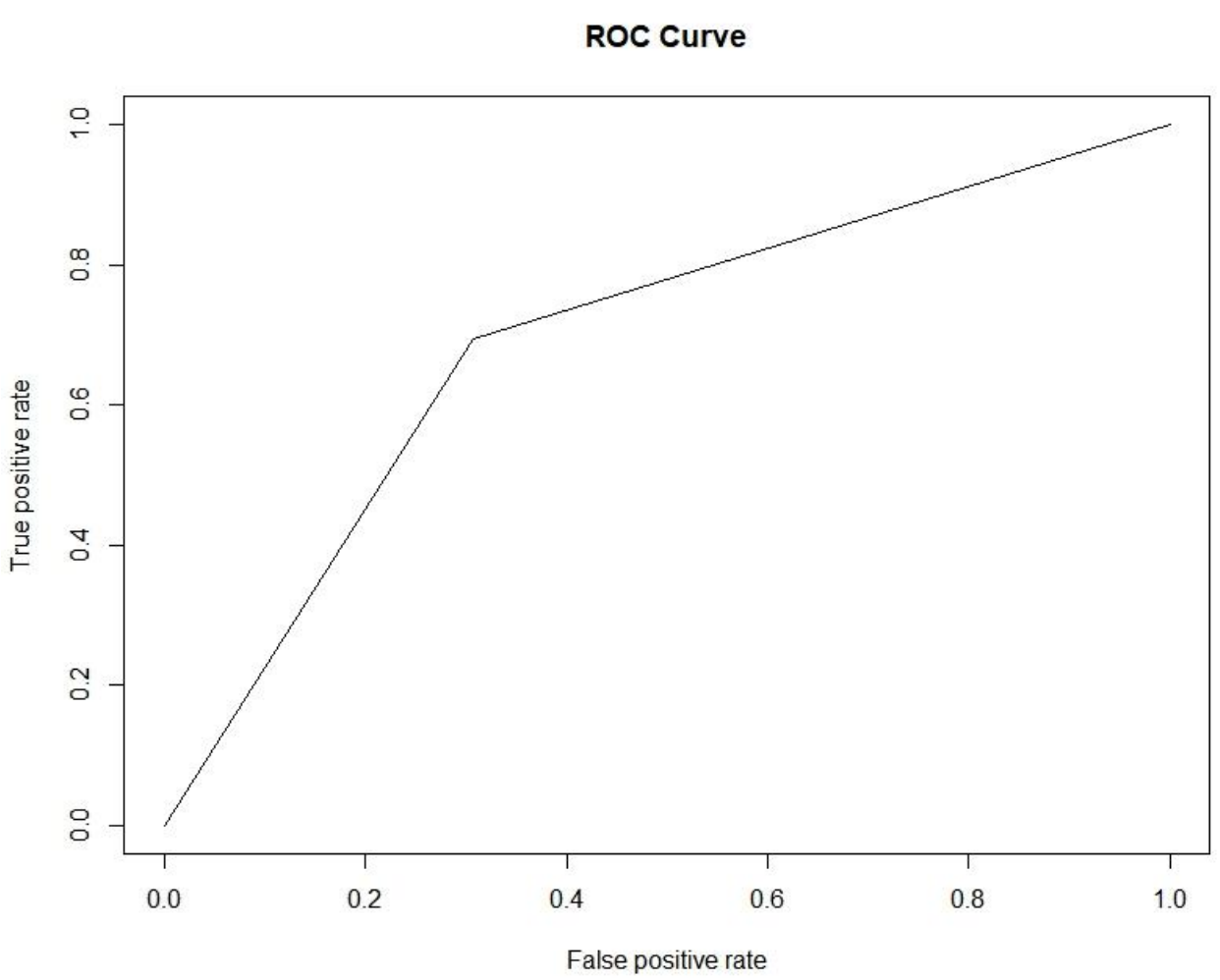


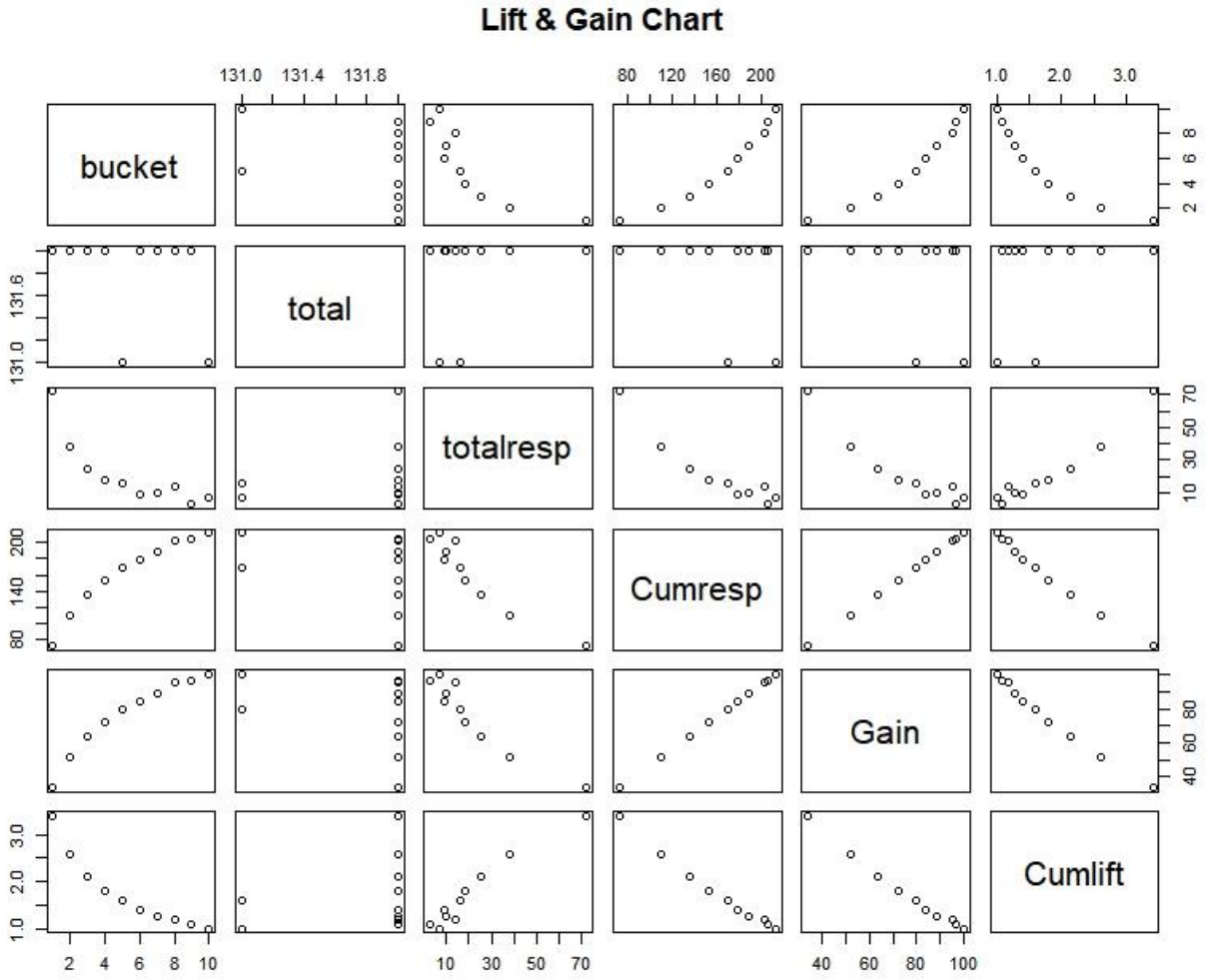
	NA Imputation							NA Removal
Characteristics	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26	Model 27	Model 25
Cutoff	0.1555	0.1639	0.1658	0.1647	0.1605	0.1672	0.1672	0.1616
Accuracy	0.6882	0.6897	0.6912	0.692	0.6927	0.6904	0.6859	0.7303
Sensitivity	0.6887	0.6887	0.6934	0.6887	0.6934	0.6887	0.6887	0.7416
Specificity	0.6881	0.6899	0.6908	0.6926	0.6926	0.6908	0.6854	0.7281
AUC	0.6884	0.6893	0.6934	0.6906	0.693	0.6897	0.687	0.7349
KS Statistics		0.3786	0.3869	0.3813	0.386	0.3795	0.374	0.4697

All the 8 models were evaluated to find the Cutoff, Accuracy, Sensitivity, Specificity, AUC and KS Statistics.
Based on the Characteristics, Model 25 is chosen to be the optimal model.

MODEL 25	Reference	
Prediction	No	Yes
No	766	65
Yes	340	147

Model Evaluation – KS Stat and Gain and Lift Chart





Conclusions

- As per the chosen model, the variables significantly impacting the Attrition are:
  - NumCompaniesWorked – Employees who have worked in more companies have higher probability of Attrition. This points out to the trend if a person has been switching jobs very frequently, he/she may leave the organization earlier(hopping trend). To retain such employees, we may look for the needs (salary, work culture, benefits etc.) of this specific group and try to retain good performers from this group by offering them higher salaries/perks.
  - TotalWorkingYears – Employees who have worked for fewer years have higher probability of Attrition. Younger crowd seem to leave the organization in much higher numbers.
  - YearsSinceLastPromotion – Employees who have had a recent promotion / did not have a promotion for long have higher probability of Attrition. The higher the person has to wait for a promotion, higher is the dissatisfaction amongst the employee, making him lookout for a change. If the person has been a good performer, we can have negotiations and offer timely promotions to them.
  - TotalWorkingHrs – Employees who are overworked have higher probability of Attrition. Continuously overworked employees tend to be under a lot of stress and this cause them to look out for a change.
  - MaritalStatusSingle – Employees who are single have higher probability of Attrition.
  - BusinessTravelTravel_Frequently - Employees who travel frequently have higher probability of Attrition.
  - EnvironmentSatisfactionHigh and Very.High – Employees who have High and Very High Environmental Satisfaction have lower probability of Attrition. These two values ensure that an employee tend to stay with the organization.
  - JobSatisfactionVery.High – Employees who have High Job Satisfaction have lower probability of Attrition.
- Using our predictive model, XYZ organization will be able to predict with accuracy which employees have a higher probability of resigning. This helps not only to take preventive measures but also into making better hiring decisions.

Recommendations

It is understandable that the Employees who are leaving the company are young, efficient, talented and ambitious who should be engaged more to stay with the Company

We should take following measure to retain these group of employees.

- We can have a tie up with couple of prestigious universities and have the employees pursue Higher Education. To ensure that they remain with the company and that the company recovers the ROI, they may be asked to sign a bond of reasonable duration.
- They should be given chance to work on good projects based on latest technologies, where they can gain skills and knowledge further.
- Their work should be appreciated in the team meetings at various levels e.g. from team meetings to All Hands meet. They should be recognized for work. They can even we awarded a token Prize money and Certificates in these meetings
- There should be a website for employee's to give anonymous feedback and genuine feedback should be acted upon.
- Employees who travel frequently are having high attrition, as employees with family/commitments seem to find it difficult. To handle this, the employees should travel on rotation basis and have travels planned in advance so that there is minimum disruption and to have extra travel allowance as perks.
- Continuously overworked employees also tend to be under a lot of stress and this cause them to look out for a change. A regular report needs to be taken where if an employee is found to over-work regularly, we need to understand the issue beneath, whether its performance or the manager has not set the expectations correctly. Having regular meetings will ensure that employees feels heard and get a sense of belonging.
- Overall to retain the employees, look for the needs (salary, work culture, benefits etc..) to retain good performers.

Thank You