HR Analytics

CASE STUDY SUBMISSION

Group Name:

1. Abhishek Bansal
2. Puneet Tripathi
3. Melvin Vallully
4. Srishti Khanna

Abstract

- **XYZ**, employs over 4000 employees
- Their yearly attrition is around 15%, resulting in
  - Delayed Projects
  - Need for more backup resources
  - More number of trainings and onboarding sessions for New employees
- All the above factors is impacting their operational cost and brand value

Problem solving methodology

Data Preparation -

- Test for duplicates
- Fix NA values
- Fix outliers
- Create Segmented variables – like age group, overtime etc.

EDA -

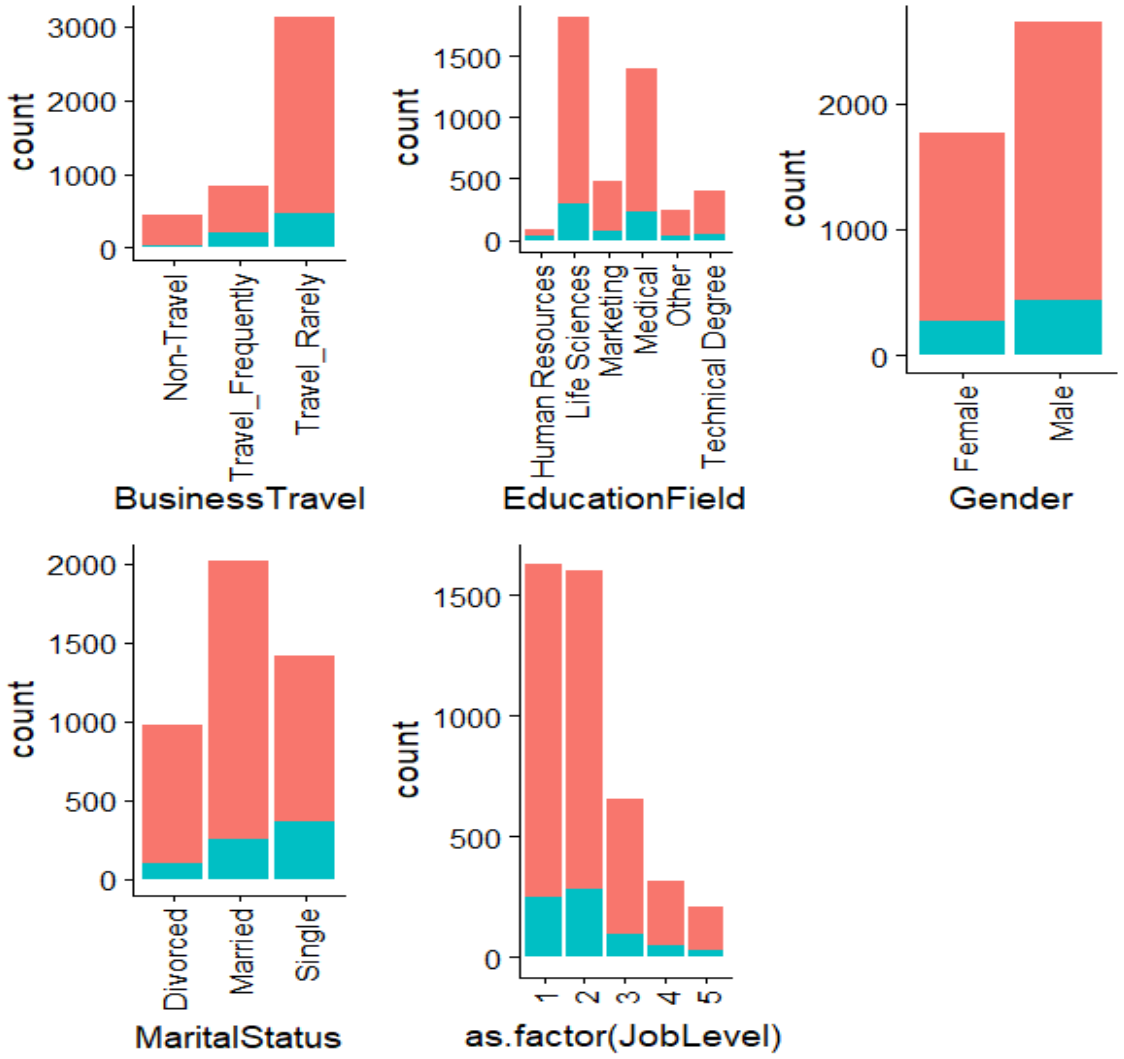
- Perform univariate and multivariate analysis

Modeling -

- Create train/test data by spliting the data in 70:30 ratio
- Run stepAIC on train data, run iteration to remove columns with High VIF and high p-value
- Take final model and use it to predict attrition in test model
- Evaluate model on sensitivity, specificity and accuracy of model
- Create Gain/Lift charts to validate the model strength
- Find KS-statistics to determine the discriminating power of model

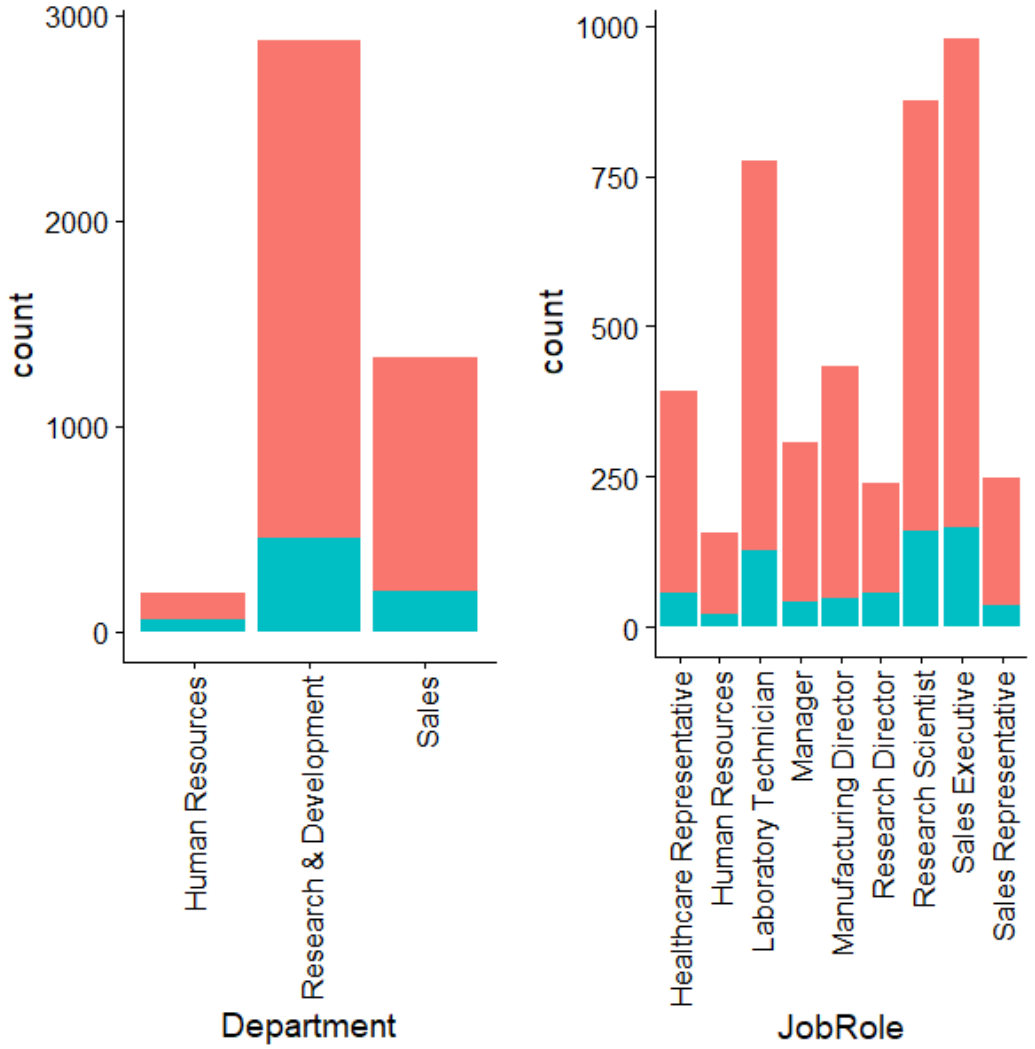
Variables across attrition

- BusinessTravel has a clear impact on attrition
- Other variables too are showing significant spread for attrition
- Numberwise More Males leave job but percentagewise women leave more
- Singles tend to quit jobs more than married and divorced



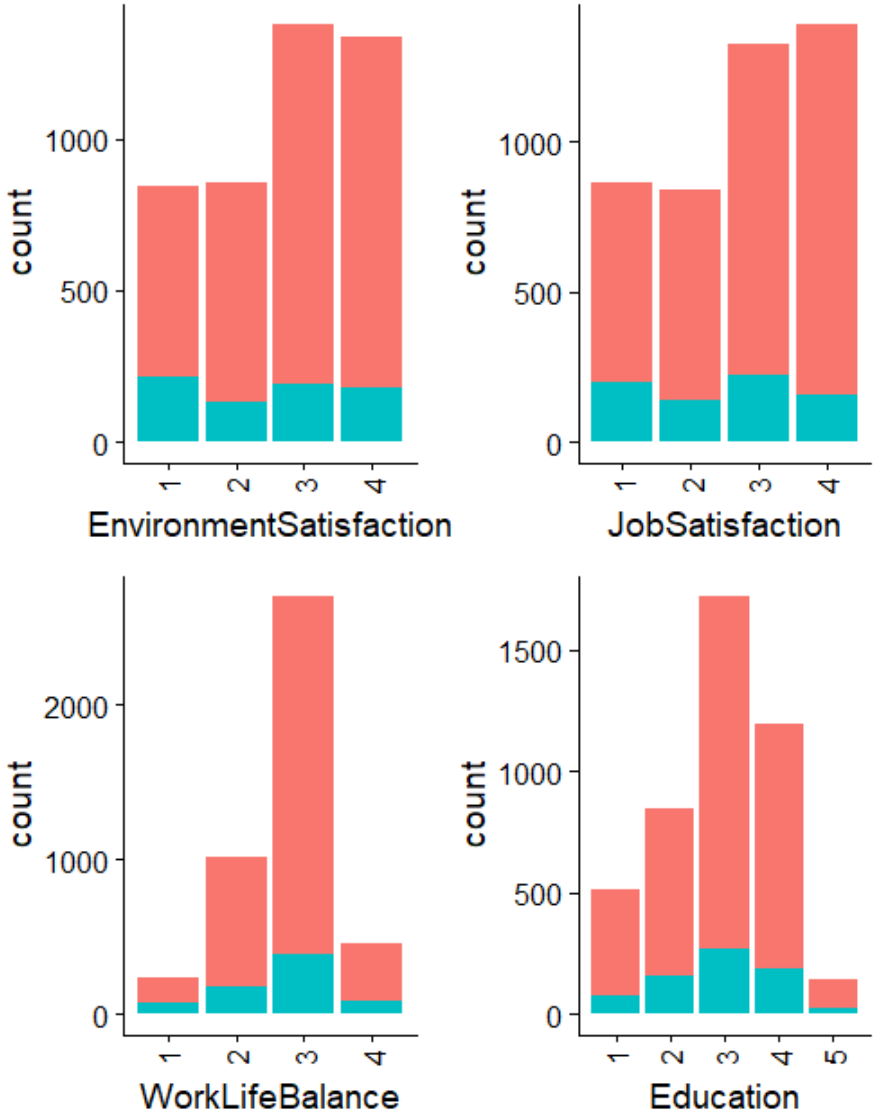
Variables across attrition

- People working in Research be it Research Scientist or Lab Technician quits job more than others
- Sales executives are also attrition prone more than others



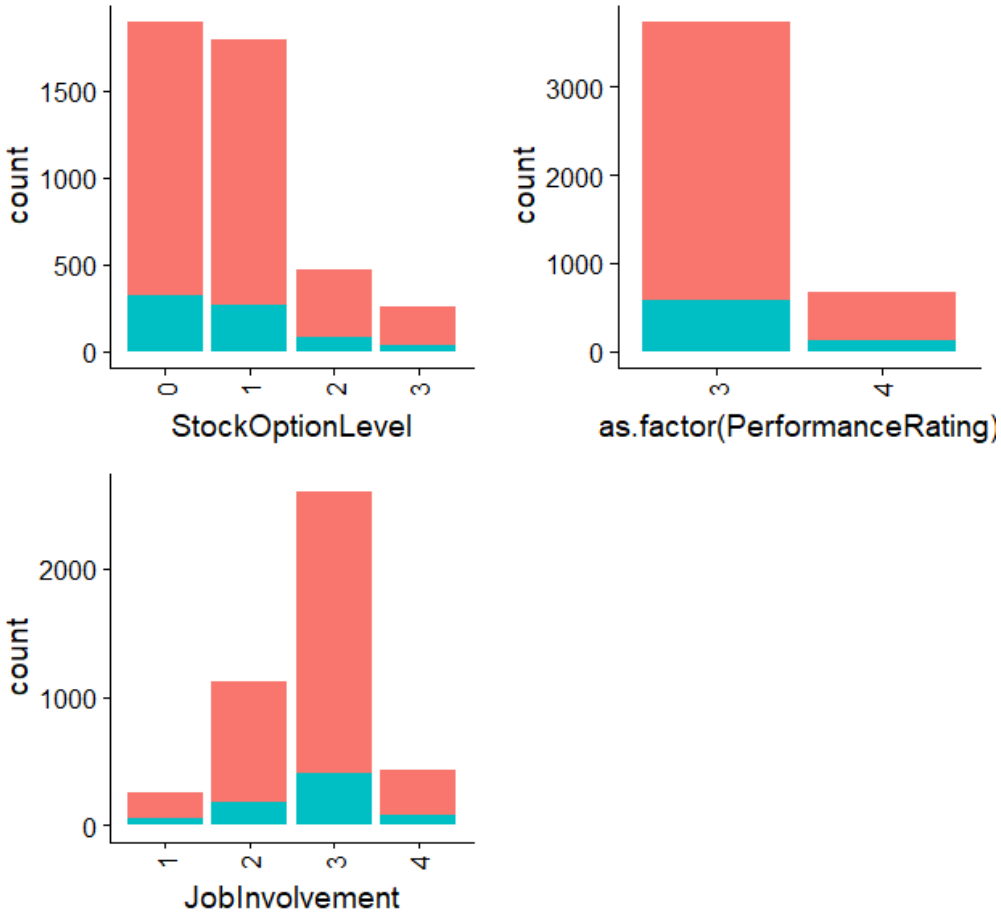
Factor (ordinal though) but have numerical value

- WorkLifeBalance & Eduation is an important Factor
- Looks like EnvironmentSatisfaction and JobSatisfaction explains variance in similar fashion



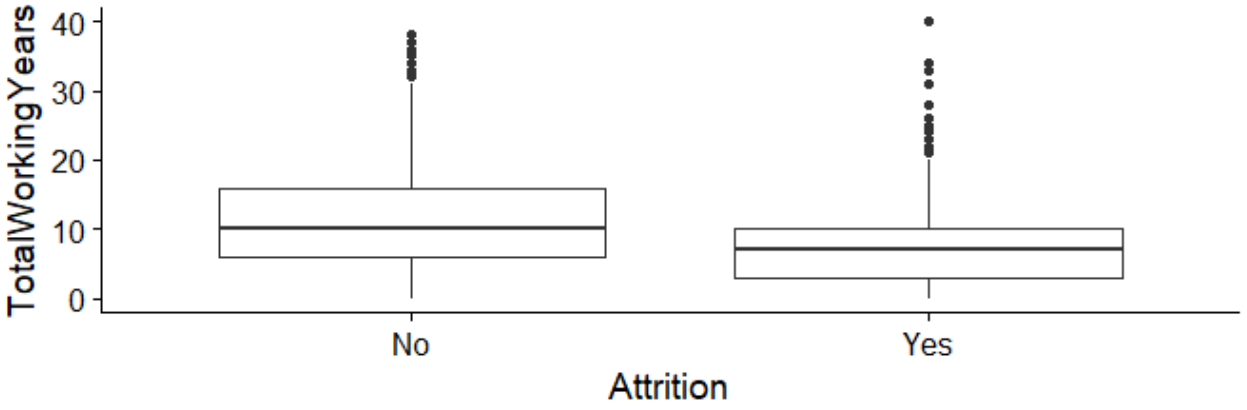
Variables categorical (ordinal) in nature

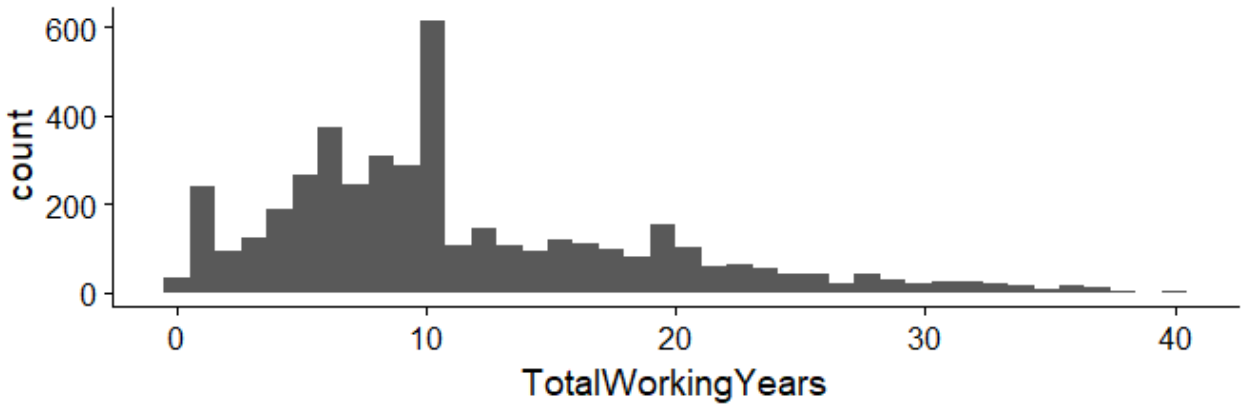
- All of these variables looks important as they have significant attrition spread
- StockOptionLevel and JobInvolvement looks like key

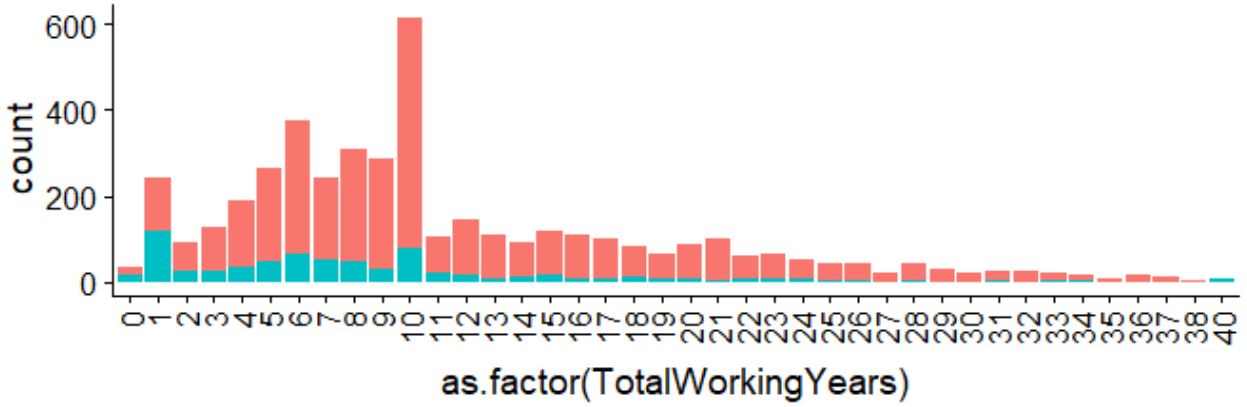


Total Work Experience

- People leaving company have median 7 years of experience and have relatively lower overall experience
- This might mean as people get more experienced they tend to stay at same company for longer time

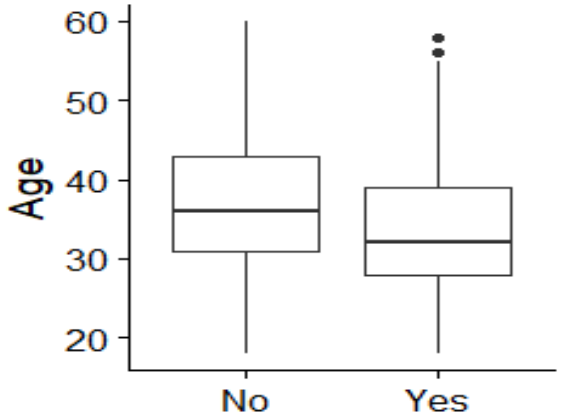


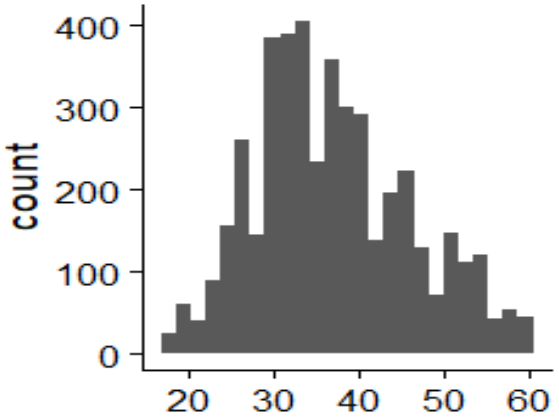


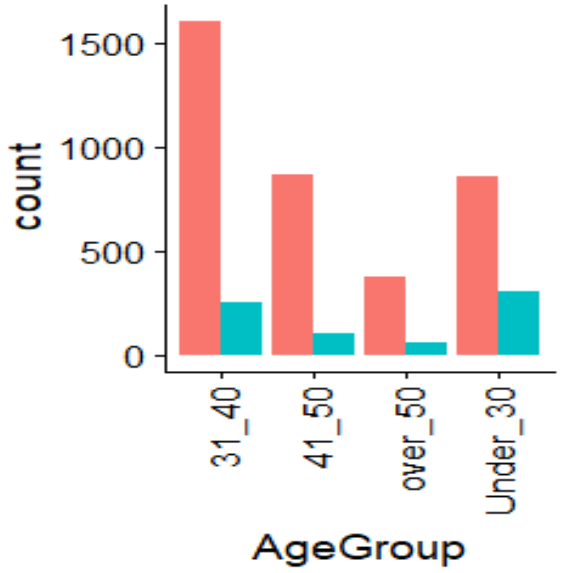


Age and AgeGroup

- Mostly work force has age between 25 to 50
- People resigning appears to be relatively younger ones
- We will use AgeGourp for modeling as it gives a creaer picture, So dropping Age

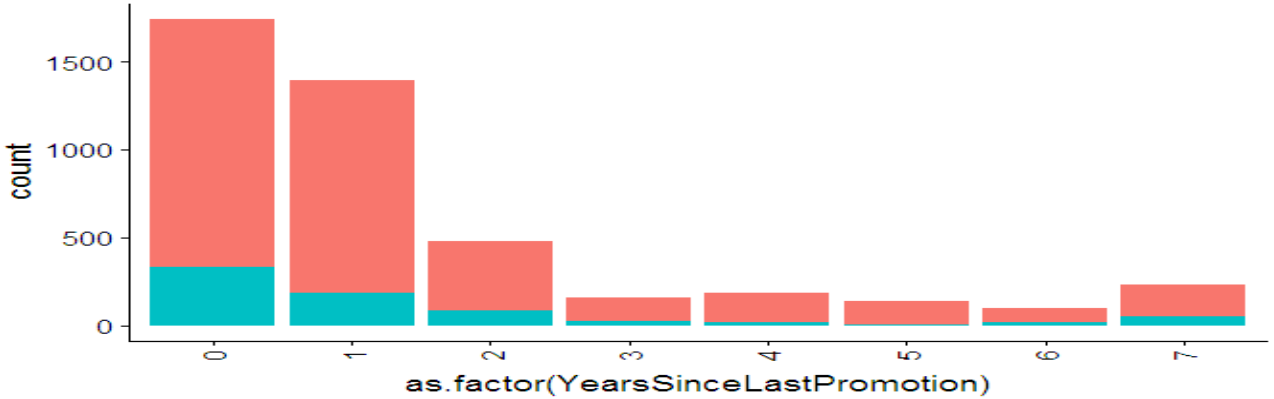


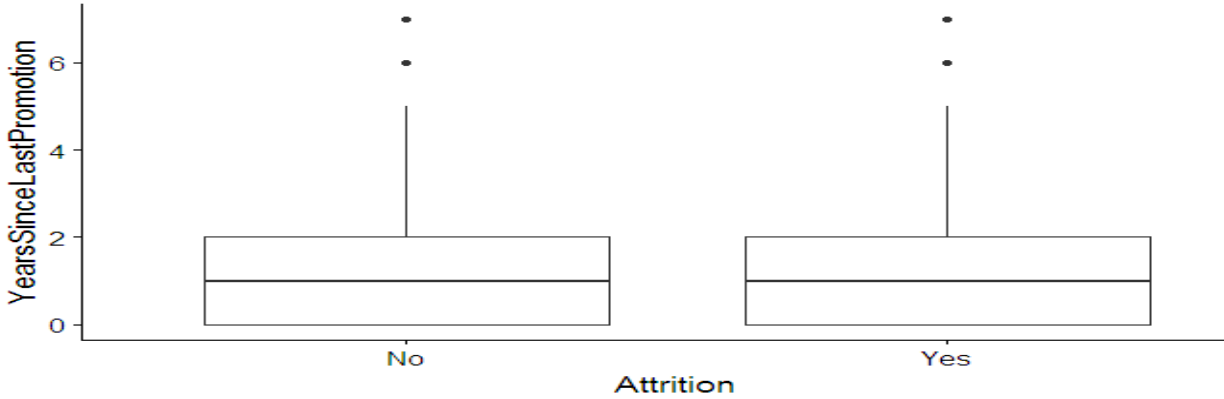


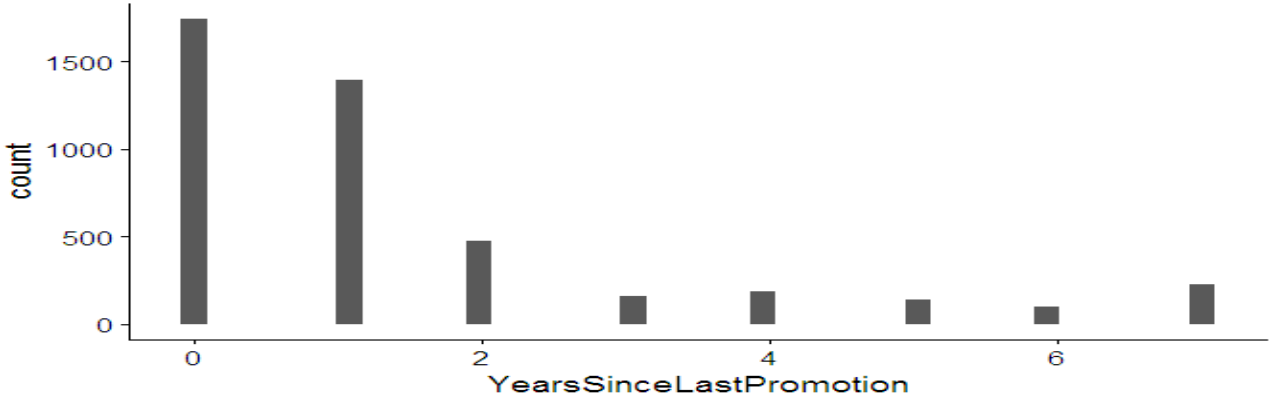


Years Since Last Promotion

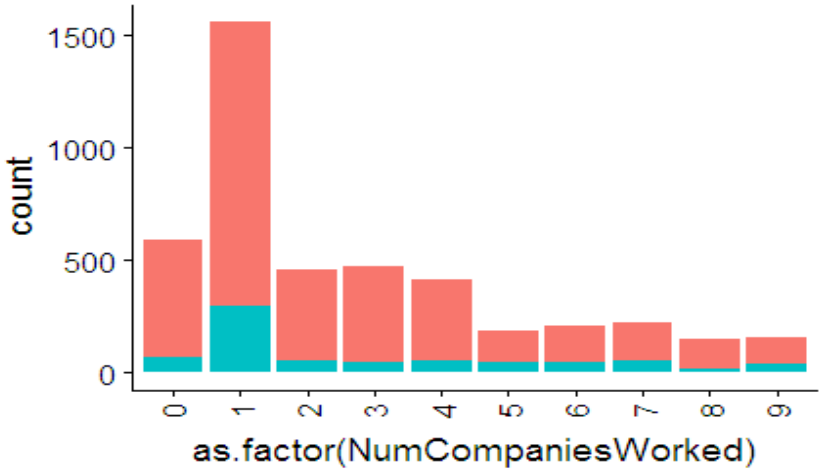
- people are trying to change job soon after getting promotion
- that might make sense as well because that means reaching higher salary/ higher grade jump in relatively lower time span



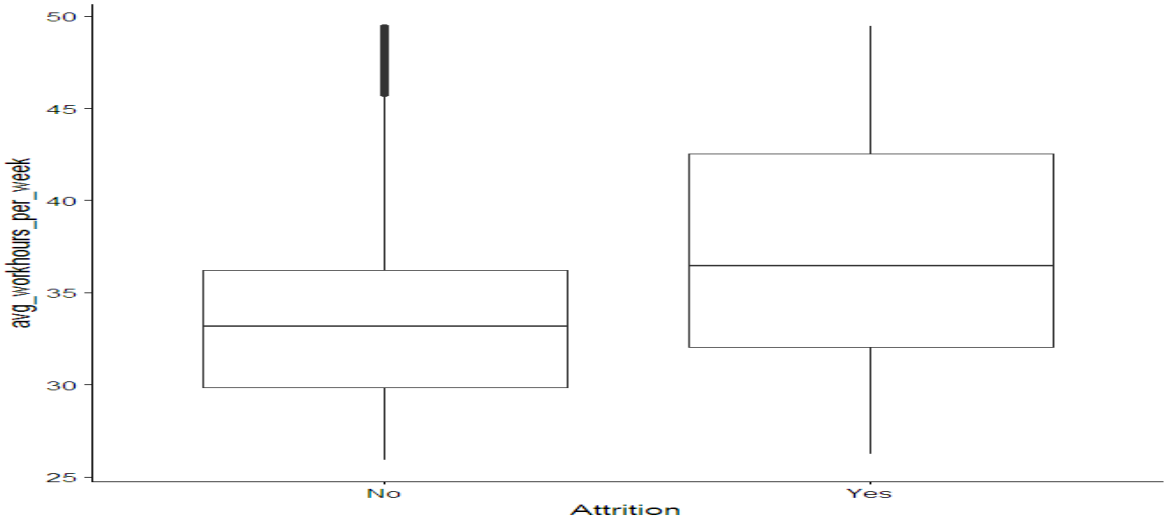


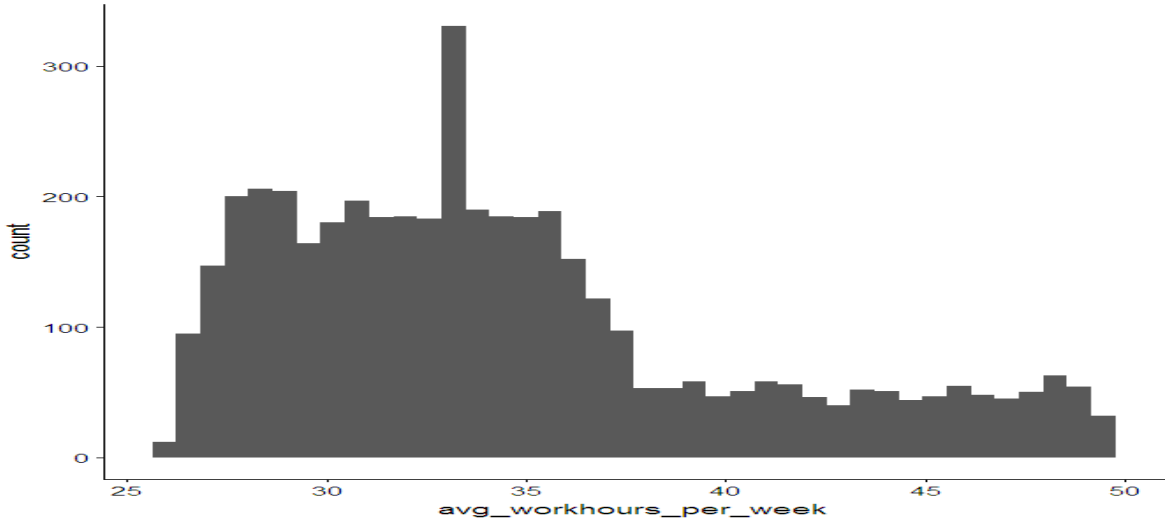


Other Variables having impact on Attrition



- Most people leaving have worked in 1 company or if it is their 1st company
- People resigning are those who are generally overworked





Model Creation

- We did a 70-30 split for Training & Testing respectively
- Post stepAIC and 29 Iterations over Train dataset, we reached a model that takes 13 input variables
- AIC – 2226.1, Null deviance: 2747.7, Residual deviance: 2198.1

Final Model Equation –

```
model_29 <- glm(formula = Attrition ~ NumCompaniesWorked + TotalWorkingYears +
                  YearsAtCompany + YearsSinceLastPromotion +
                  avg_workhours_per_week + AgeGroup.xUnder_30 +
                  EnvironmentSatisfaction.x2 + EnvironmentSatisfaction.x3 +
                  EnvironmentSatisfaction.x4 + JobSatisfaction.x4 +
                  BusinessTravel.xTravel_Frequently + JobRole.xManufacturing.Director +
                  MaritalStatus.xSingle,
                family = "binomial", data = train)
```

Model Evaluation

- We predicted the Attrition on test dataset -

```
summary(test_pred)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.002391 0.046552 0.110468 0.171465 0.238789 0.862327
```

- Then we plotted sensitivity, specificity and accuracy of model, and finalized a cutoff of 0.1775758

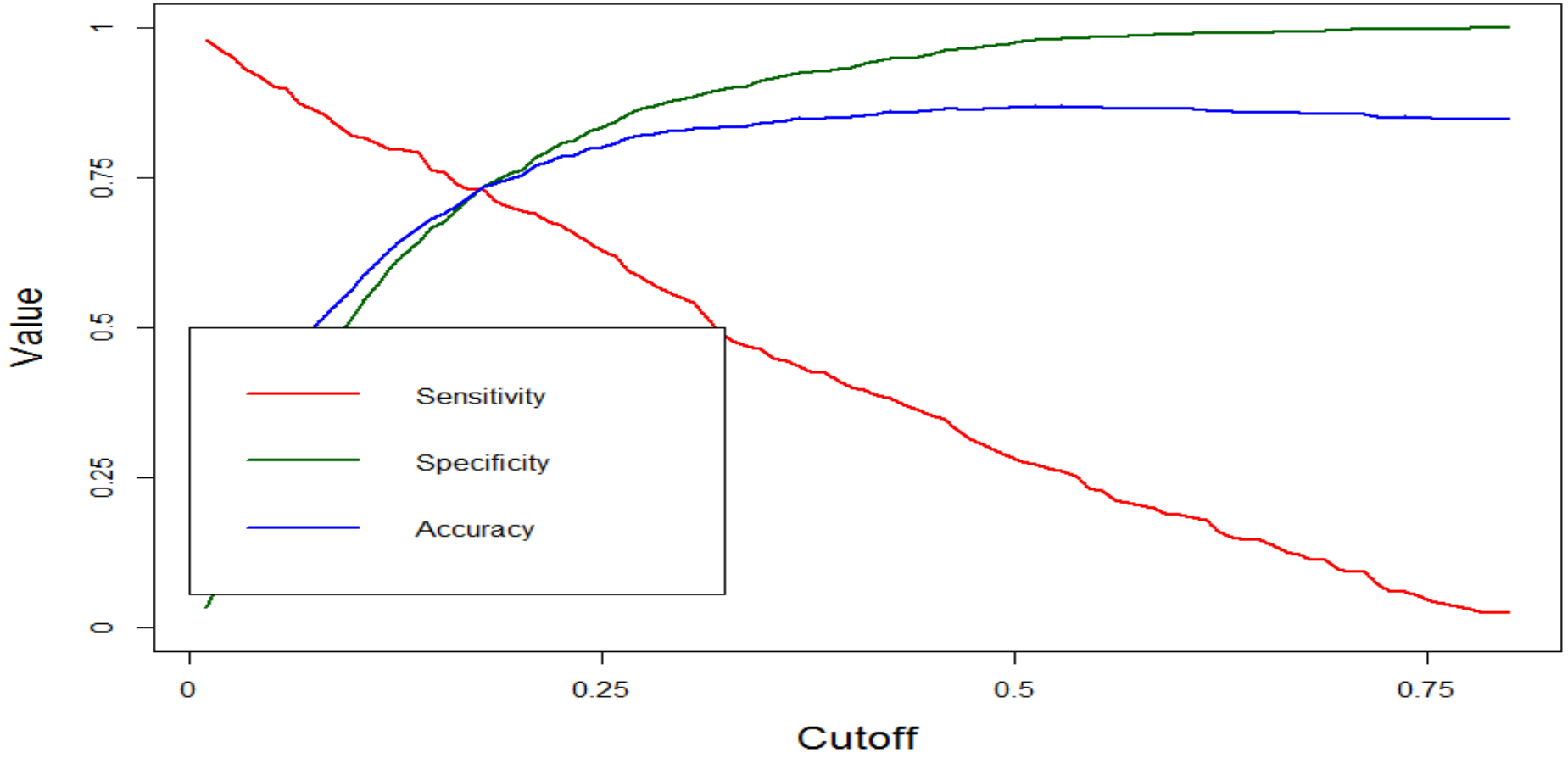
```
+-----------------------------------------+
|Accuracy   |  Sensitivity  |  Specificity |
+-----------------------------------------+
|0.733938   |  0.7294686    |  0.734767    |
+-----------------------------------------+
```

- Plot of sensitivity, specificity and accuracy on next page

- KS-Statistics of our model is 0.4642356 – which is a satisfactory discrimination power

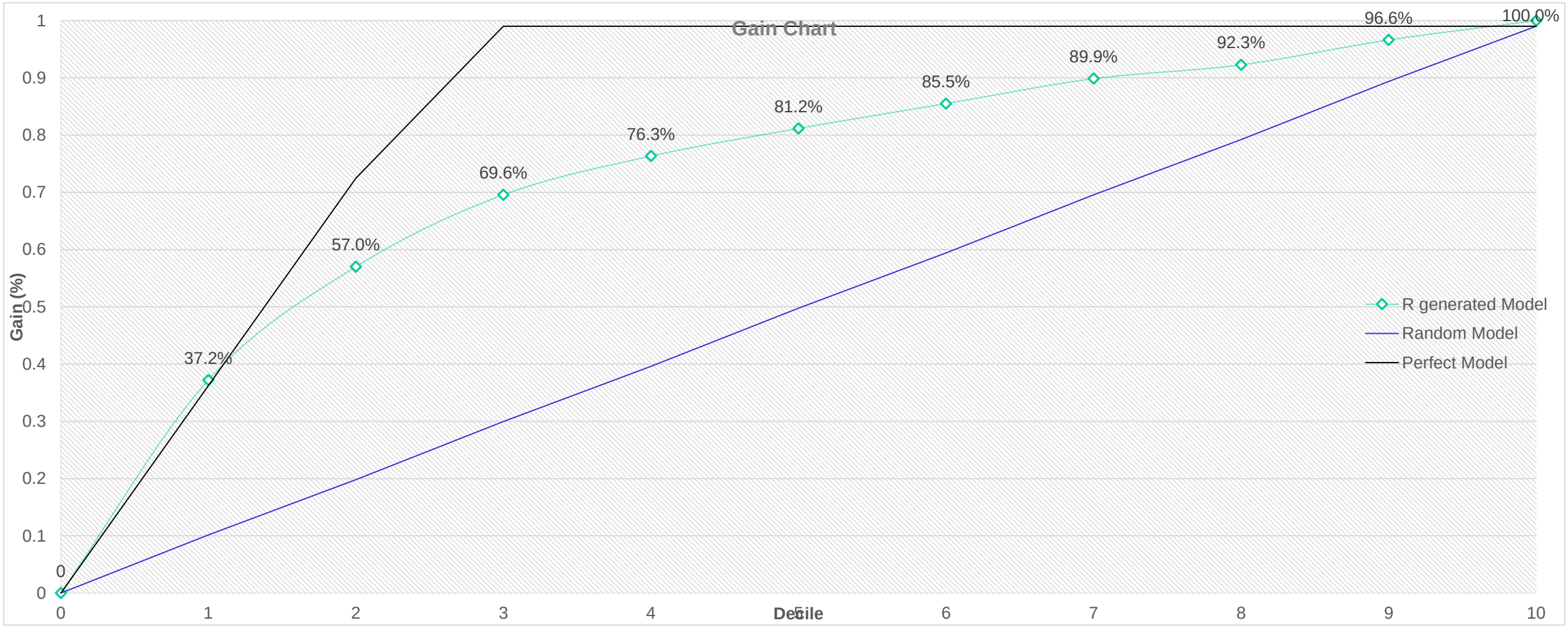
Model Evaluation – Contd.

Plot of sensitivity, specificity and accuracy



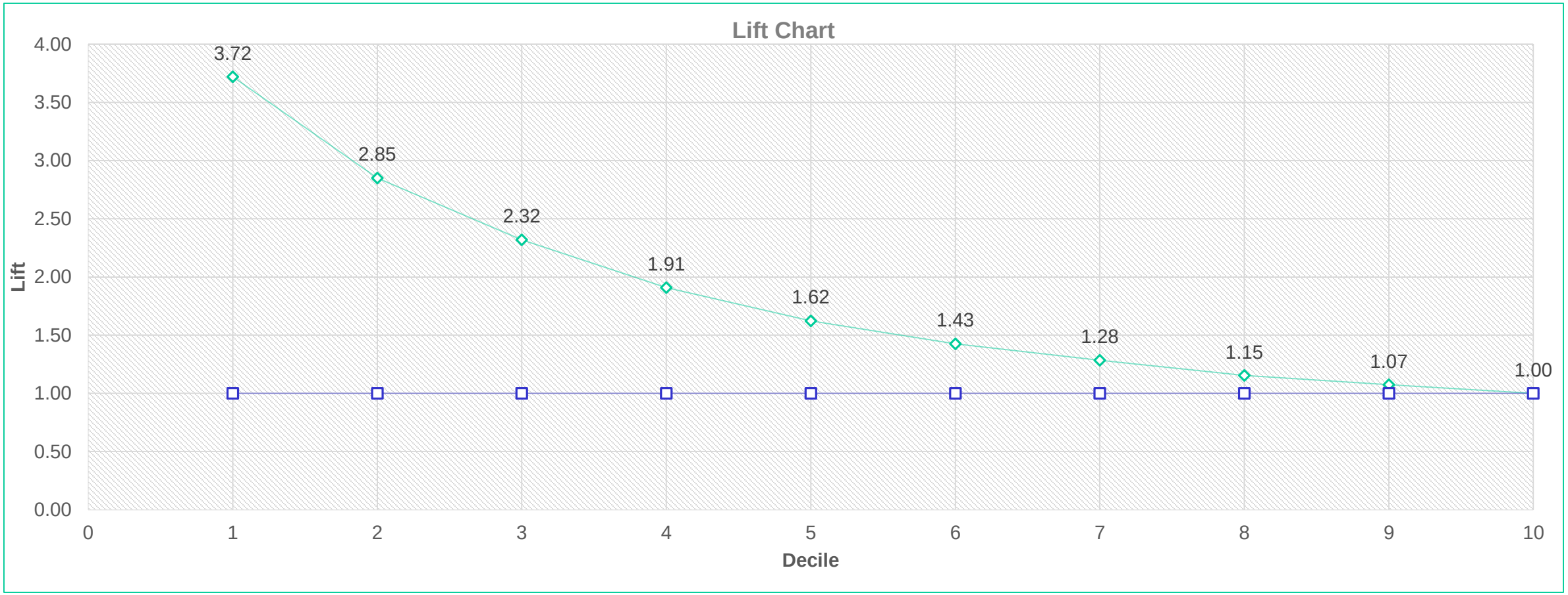
Model Evaluation – Gain Chart

From Gain Chart it is clearly visible that in 4th decile using our model we identify 76.3% of attrition.



Model Evaluation – Lift Chart

From Lift Chart it is clearly visible that in 3rd decile our model is outperforming a random model by 2.32 and in 4th by 1.91



Conclusion

Factors that affect adversely and attribute to increase in attrition are:

- **NumCompaniesWorked** – People who have worked in many companies before have higher chances of leaving.
- **YearsSinceLastPromotion** – People who have not been promoted for a long time have higher chances of leaving.
- **avg_workhours_per_week** – Has strong impact on attrition, people who are working more hours, i.e. overworked employee, are more likely to leave company.
- **AgeGroup.xUnder_30** – It tells that people with age under 30 are more likely to change job than others.
- **BusinessTravel.xTravel_Frequently** – People traveling frequently is coming out a factor for attrition, as it may be affecting their work life balance.
- **MaritalStatus.xSingle** – the employees who are single have higher chances of leaving than the ones married or divorced

Conclusion – contd.

Factors that keeps attrition in control:

- **TotalWorkingYears** – This variable has a very low p-value and has stronger impact on attrition. The employees with higher work experience tend to stay longer at company.
- **YearsAtCompany** – People who have spent more time at company tends to stay with the company.
- **EnvironmentSatisfaction.x2/ x3 / x4**- People who rated work environment as Medium/ High/ Very High are happy with the company environment and do not leave.
- **JobSatisfaction.x4 –**People only who are very highly satisfied with their job are going to stay with the organization.
- **JobRole.xManufacturing.Director** – Employees who are Director in manufacturing are satisfied with their role and tend to stay with the organization