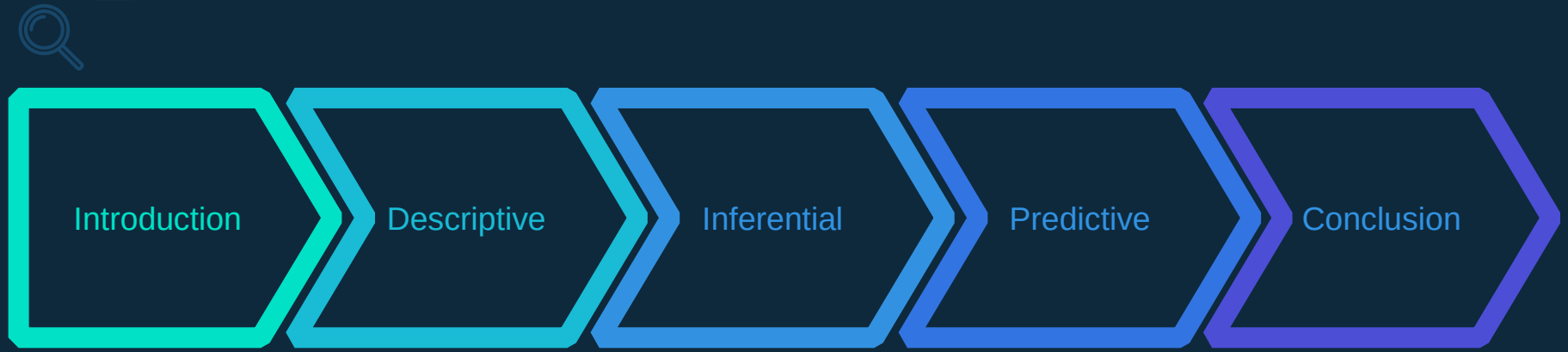
HR Analytics

Khoo Kian Sim, Tong Zi Heng, Wu Tong, Yeo Yi Xuan, Zhang Jieyuan

Agenda


Introduction
Descriptive
Inferential
Predictive
Conclusion
Yi Xuan
Zi Heng
Kian Sim
Wu Tong
Jieyuan

1

Introduction

Analytics in HR

Objective

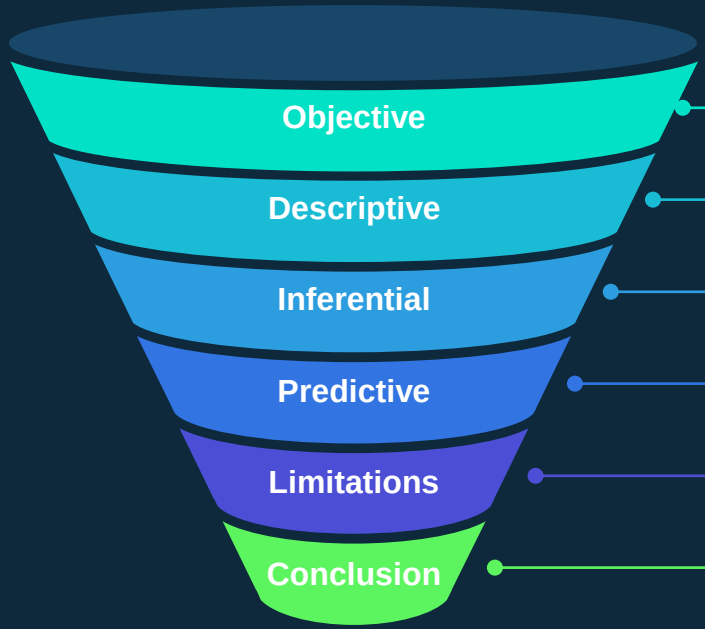
Utilize analytics to assist HR in reducing attrition and improving performance

slido

What do you think is the most important factor affecting employee attrition?

ⓘ Start presenting to display the poll results on this slide.

Overall Concept



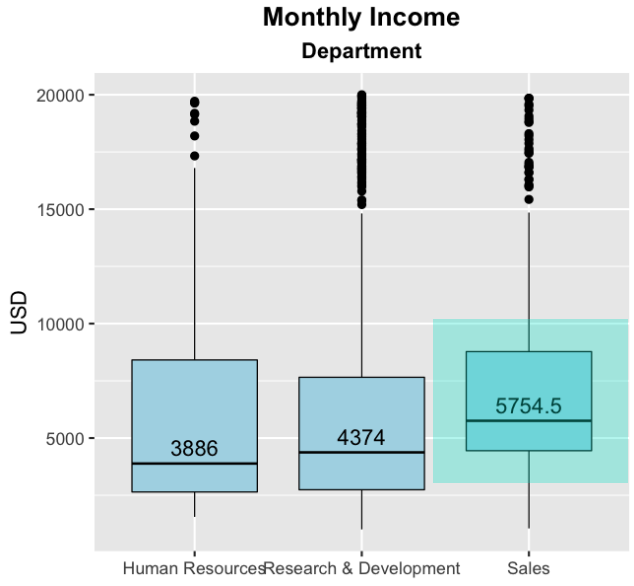
- Reduce attrition and improve performance
- Explore employee profile
- Comparison of means and test of association
- Logistic regression
- Potential limitations
- Recommendations for further work

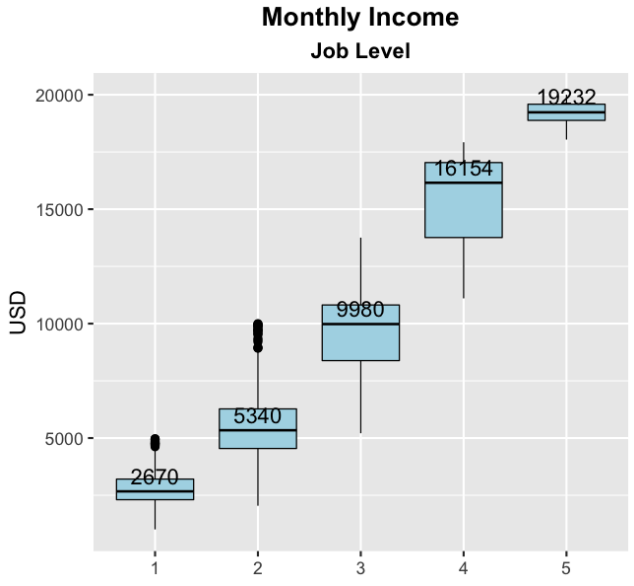
2

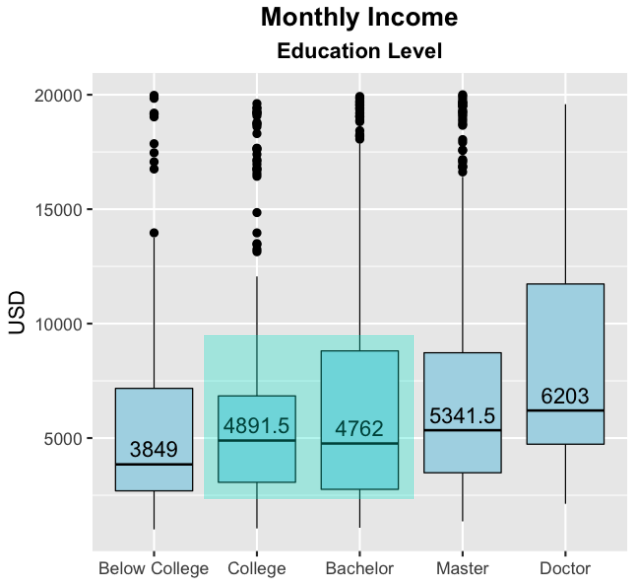
Descriptive Statistics

Key findings from diving into employee monthly income

Monthly Income







1. Sales has Highest Median Monthly Income

The company may have prioritised paying a higher salary to its sales staff

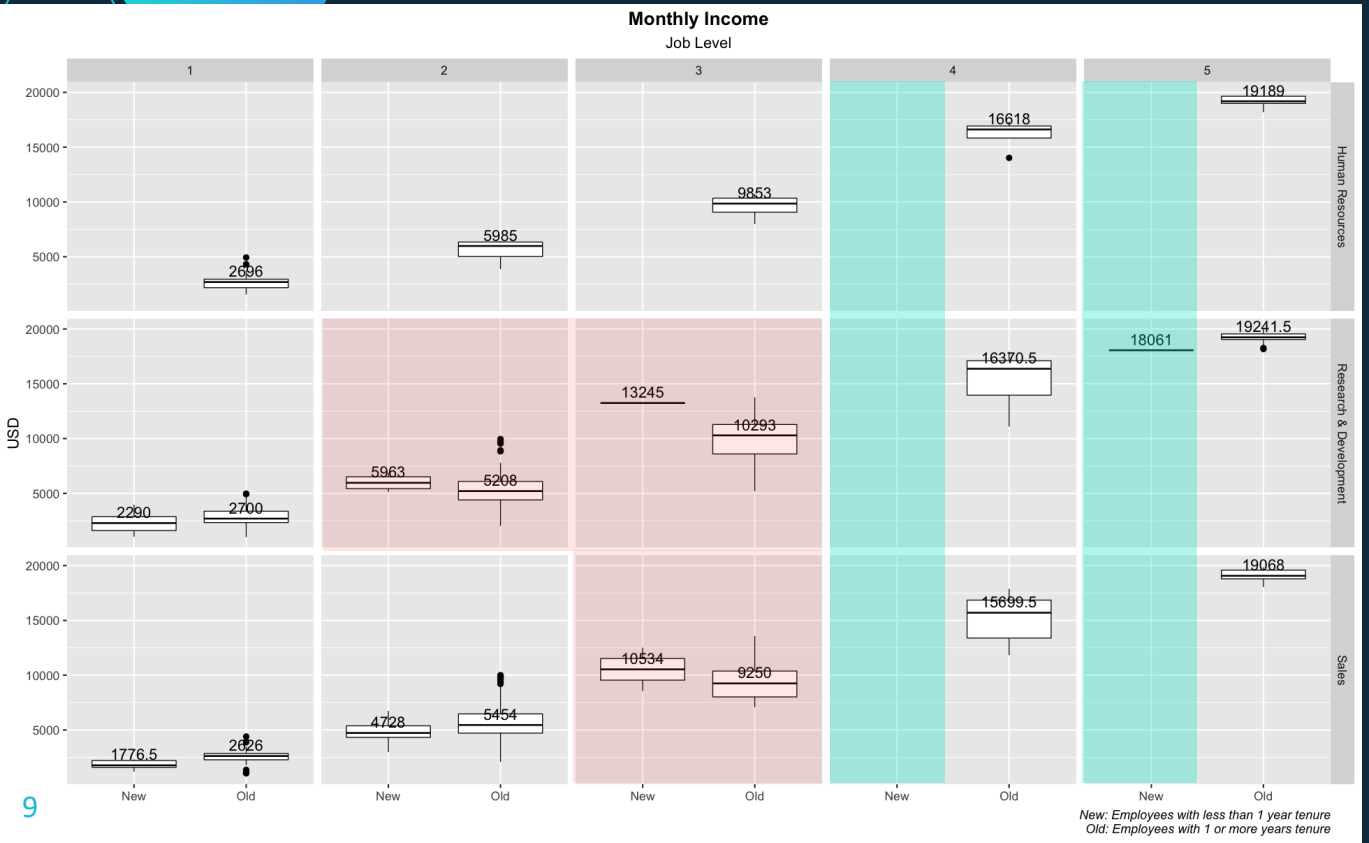
2. Income Increases with Job Level

In line with the norm, median income increases as employees progress in job level

3. College has slightly Higher Median Monthly Income than Bachelor

The company may be valuing both education level equally as both have around the same average years of working experience

Monthly Income



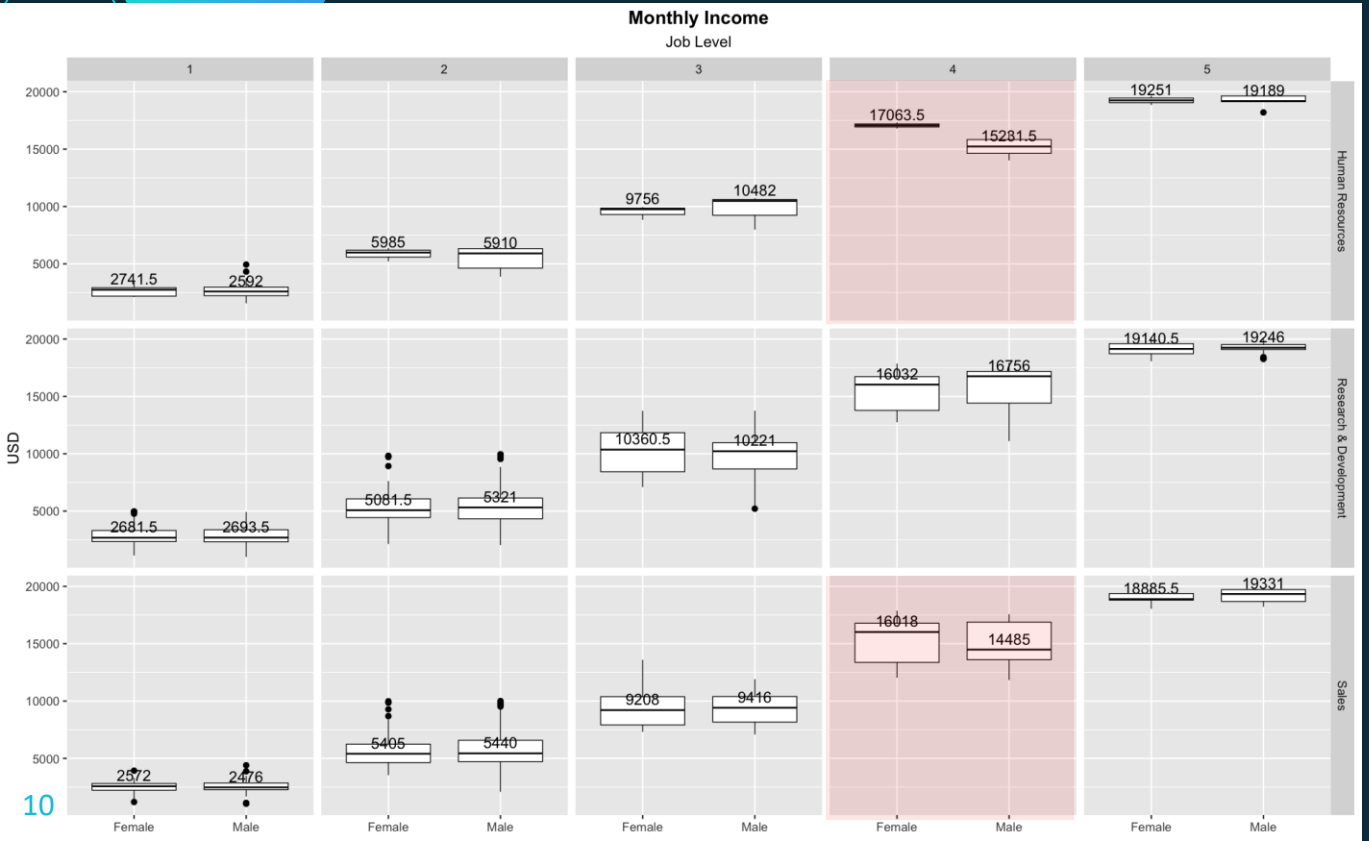
1. New Employees with Job Level 2 and 3 in R&D and Job Level 3 in Sales have Higher Median Monthly Income

The company may want to take note and manage it well to prevent dissatisfaction amongst employees with longer tenure in the company

2. Only 1 New Employee with Job Level 4 and 5

The company may prefer to promote employees from within

Monthly Income



1. Median Income of Male and Female are largely comparable

The company might have placed importance in ensuring gender pay equality

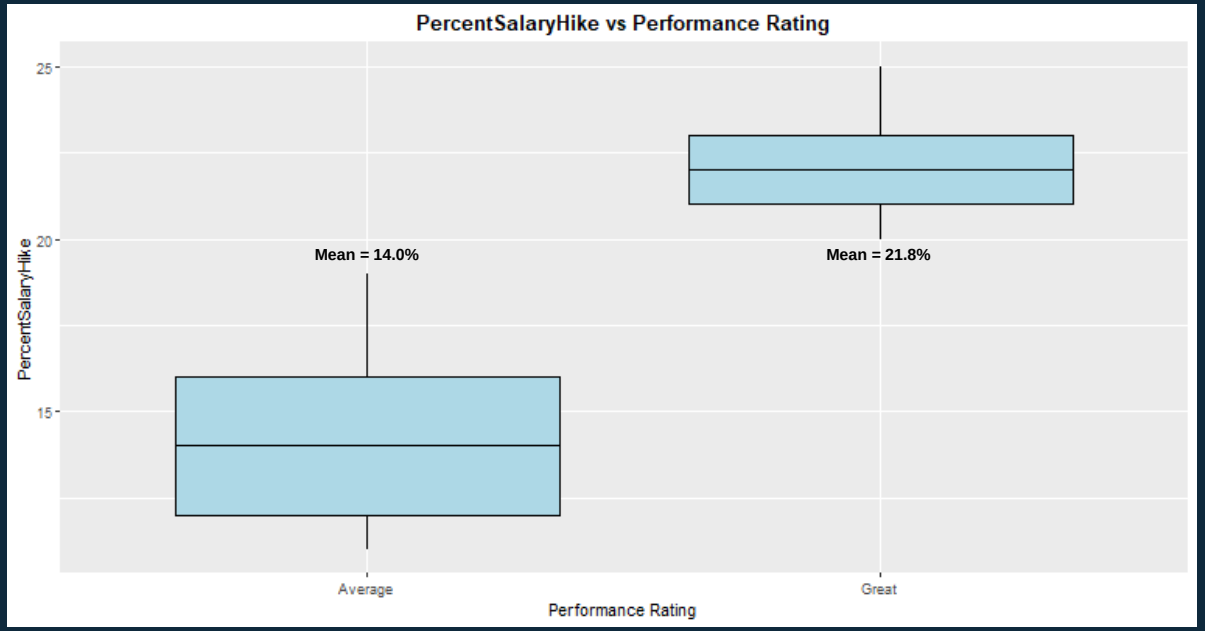
However, there are exceptions for employees with Job Level 4 in HR department and Sales department

3

Inferential Statistics 

Key findings from comparison of means and test of association for employee performance and attrition

Employee Performance



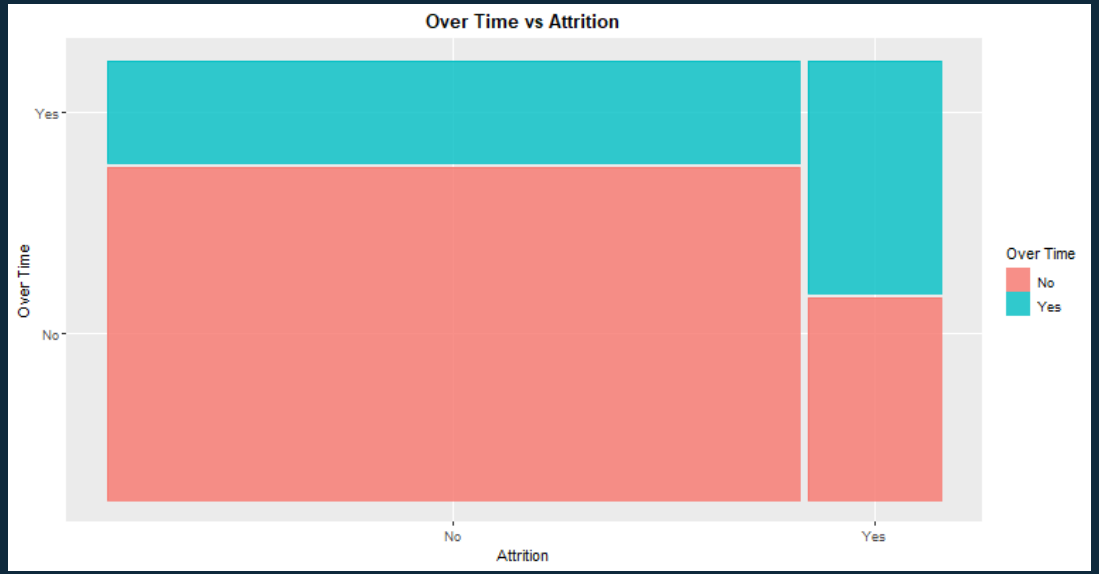
*Difference of mean percentage salary hike between employee with average performance ratings and great performance rating significant at 95% confidence level

Attrition (Comparison of Means)

Factors	Attrited mean	Non-attrited mean	Difference	p-value
Age	33.6	37.6	-4.0	6.5E-10
Daily rate	750	813	-62	0.03
Distance from home	10.6	8.9	1.7	2.7E-03
Monthly income	4,787	6,833	-2,046	5.5E-10
Total working years	8.2	11.9	-3.6	2.9E-11
Number of trainings received last year	2.6	2.8	-0.2	0.02
Number of years at company	5.1	7.4	-2.2	2.0E-07
Numbers of years at current role	2.9	4.5	-1.6	4.6E-10
Number of years with current manager	2.9	4.4	-1.5	1.4E-09

*Difference in means for the factors above were significant at 95% confidence level

Attrition (Test of Association)



Association with attrition were significant at 95% confidence level for the following factors: frequency of business travel, department, education field, environment satisfaction, job involvement, job level, job role, job satisfaction, martial status, overtime, stock option level, work life balance

4

Predictive model

Logistic regression

Predictive Model - Logistic Regression

```
glm_fit=glm(AttritionBin~ Age+BusinessTravel+DailyRate+Department+DistanceFromHome
            +Education+EducationField+EnvironmentSatisfaction+Gender+HourlyRate+JobInvolvement+JobLevel
            +JobRole+JobSatisfaction+MaritalStatus+MonthlyIncome+MonthlyRate+NumCompaniesWorked+OverTime
            +PercentSalaryHike+PerformanceRating+RelationshipSatisfaction+StockOptionLevel
            +TotalWorkingYears+TrainingTimesLastYear+WorkLifeBalance+YearsAtCompany+YearsInCurrentRole
            +YearsSinceLastPromotion+YearsWithCurrManager, family = binomial, data = train
            , contrasts = list(Education = 'contr.treatment', EnvironmentSatisfaction = 'contr.treatment',
                               JobInvolvement ='contr.treatment',JobLevel = 'contr.treatment',
                               JobSatisfaction = 'contr.treatment', RelationshipSatisfaction = 'contr.treatment',
                               StockOptionLevel = 'contr.treatment',WorkLifeBalance = 'contr.treatment'))
# Contrast level for ordered variable change to contr.treament from the default polynomial orthogonal contrasts
# as we are unable to assume the variable are equally spaced which is required for polynomial contrasts

summary(glm_fit)
```

Initial AIC = 751
Initial accuracy ≈ 88%
Number of factors = 30

Predictive Model - Logistic Regression

```
library(MASS)

step<-stepAIC(glm_fit,direction="both")
summary(step)
```

```
Call:
glm(formula = AttritionBin ~ BusinessTravel + DailyRate + DistanceFromHome +
    EducationField + EnvironmentSatisfaction + Gender + JobInvolvement +
    JobLevel + JobRole + JobSatisfaction + NumCompaniesWorked +
    OverTime + RelationshipSatisfaction + StockOptionLevel +
    TotalWorkingYears + TrainingTimesLastYear + WorkLifeBalance +
    YearsAtCompany + YearsInCurrentRole + YearsSinceLastPromotion +
    YearsWithCurrManager, family = binomial, data = train, contrasts = list(Education = "contr.treatment",
    EnvironmentSatisfaction = "contr.treatment", JobInvolvement = "contr.treatment",
    JobLevel = "contr.treatment", JobSatisfaction = "contr.treatment",
    RelationshipSatisfaction = "contr.treatment", StockOptionLevel = "contr.treatment",
    WorkLifeBalance = "contr.treatment"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6630  -0.4423  -0.2067  -0.0603   3.3491
```

Initial AIC = 751
Initial accuracy ≈ 88%
Number of factors = 30

AIC after stepwise selection = 730
Accuracy after stepwise selection ≈ 89 %
Number of factors after stepwise selection = 21

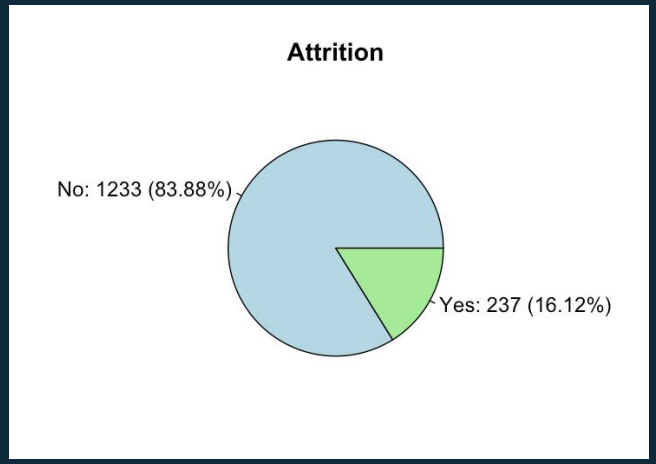
Predictive Model - Logistic Regression

Coefficients:	Estimate	Std. Error	Pr(>|z|)	**Significance**
Business Travel Travel_Frequently	2.2767641	1.5064567	0.028855	***
Over Time Yes	2.0978475	0.2306034	< 2e-16	***
Years Since Last Promotion	0.1826927	0.0507045	0.000314	***
Distance from home	0.0507604	0.0128267	0.0000758	***
⋮	⋮	⋮	⋮	⋮

Predictive Model - Limitations

.

Imbalanced dataset



Imbalanced prediction result

```
glm_pred  No Yes
      No  244  25
      Yes   7  18
> mean(glm_pred==test$Attrition)
[1] 0.8911565
```

Accuracy for predicting "No" : 244/(244+7) ≈ 97%
Accuracy for predicting "Yes" : 18/(25+18) ≈ 42%

5 Conclusion

Conclusion & Recommendations

STATISTICAL CONCEPT

- Descriptive Statistics
- Inferential Statistics:
  - Comparison of mean (using z-test)
  - Test of association (using Chi-Sq)
- Predictive Analytics:
  - Logistic Regression

SOME KEY FACTORS ASSOCIATED WITH ATTRITION

- Business Travel Frequency
- Distance from Home
- Job/Environment/Relationship satisfaction
- Overtime
- Daily Rate / Income
- Years with Company/Current Manager

FURTHER WORK

- Analysis for Performance:
  - Lack of differentiation in performance rating in current data set
  - Modification to performance appraisal metrics to enable analytics in the future

Thank you