

Disclaimer: Please note that the data is from Kaggle and is fictional. This project was completed as part of the MSDS 498 Capstone Project course within the Northwestern University - Data Science Program. All data, dashboards, and insights used throughout this project are completely simulated and not in any way connected to or a reflection of The Walt Disney Company. Please do not duplicate or distribute outside of the context of this course.

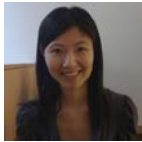





# RETENTION RISK MODEL & EMPLOYEE TURNOVER STRATEGY

JUNE 9, 2019  
ANNUAL HR SUMMIT - DAY 3

AGNES CHENG  
BRENT YOUNG  
MARY TAYLOR  
MICHELLE EURE

# Meet Our People Insights Team

Roles & Responsibilities				
	Agnes Cheng Toronto, CAN Time Zone: EST	Brent Young Los Angeles, CA Time Zone: PST	Mary Taylor Raleigh, NC Time Zone: EST	Michelle Eure Chicago, IL Time Zone: CST
PROJECT MANAGER		●	●	
PROGRAMMERS	●	●	●	
TECHNICAL WRITER		●		●
VISUALIZATION	●	●		●
DASHBOARD/MOBILE				●

● Primary Role
 ● Secondary Role



*“Whatever we accomplish is due to the combined effort.”*

WALT DISNEY

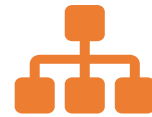
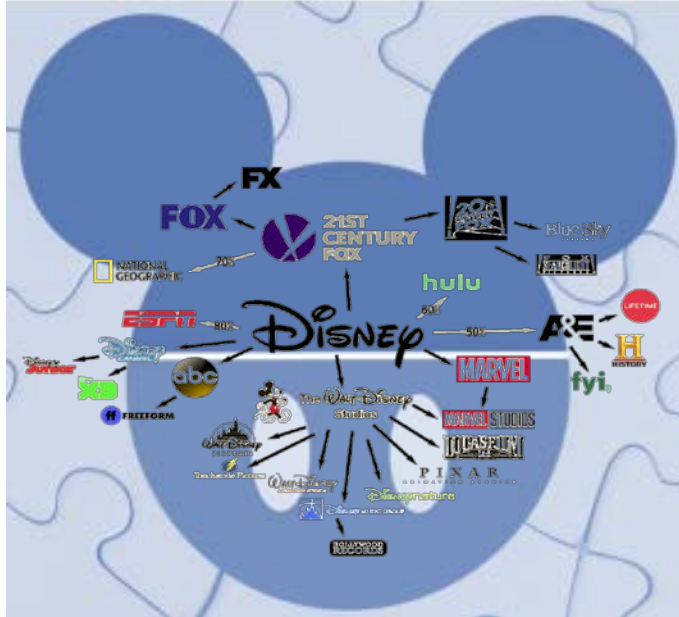
# Agenda

1. Business Landscape & Problem Statement
2. Objectives & Business Value
3. Description of the Data
4. Initial Observations & Cluster Analysis
5. Transformations
6. Analysis of the Data
7. Observations & Learnings
8. Recommendations
9. Dashboard & Mobile User Interface
10. Final Words



# Business Landscape & Problem Statement

## Business Landscape



- Significant organizational changes (Disney/21CF integration)



- Turnover has increased



- Strong competition for top talent



- Desire to use data driven approach

## Outcomes of High Voluntary Turnover



- High employee replacement cost



- Employee dissatisfaction due increase workload from work redistribution



- Decrease productivity & project delays







- Loss of institutional knowledge



# Objectives & Business Value

## Objectives

- 1) Build *retention risk model* 
- 2) Determine *voluntary resignation factors* 
- 3) Employee segmentation *cluster analysis* 
- 4) Obtain insight from model results to recommend *HR programs, policies & initiatives* to retain high risk employees 

- 5) Create an interactive *retention risk dashboard & mobile interface* 



## Business Value

- Retain top talent by proactive identification of high flight-risk employees
- Understand predictors of voluntary turnover so it can be addressed via policy changes, etc.
- Decrease voluntary turnover
- Increase cost avoidance
- Shift from reactive to proactive mindset

### Cost Avoidance

Estimated Attrition cost per departing employee: \$75,000

Average salary of \$50, 000 \* 150% cost of turnover

Estimated Cost Avoidance per year with a 10% reduction: \$3 million

\$75,000 \* 400 voluntary exits per year\*10% reduction



Note: This is not actual data.

# Description of the Data

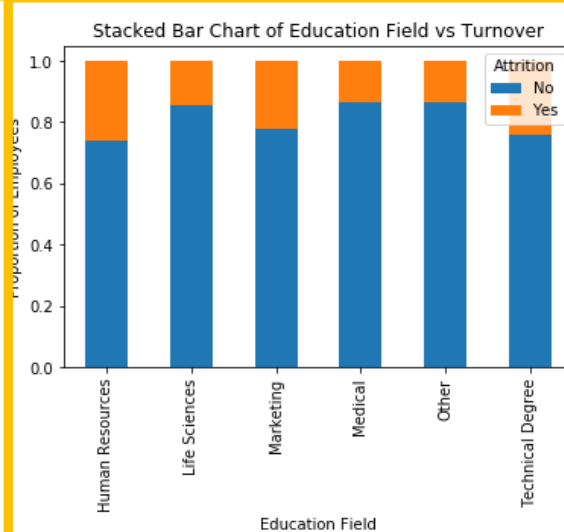
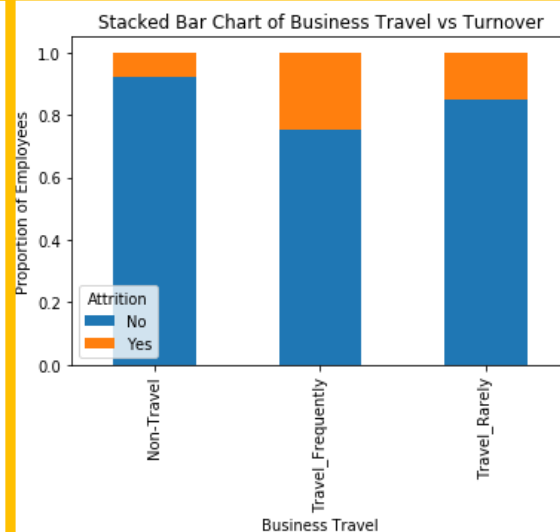
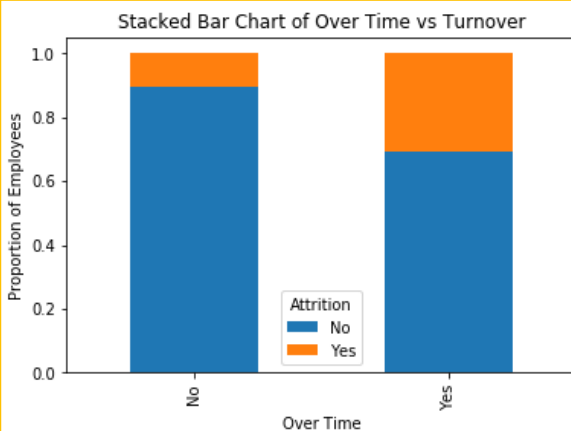
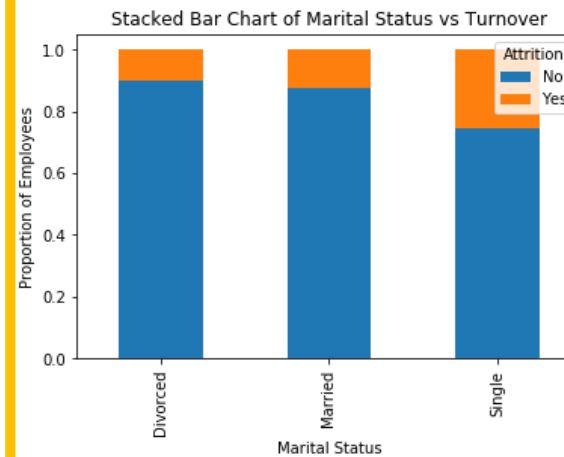
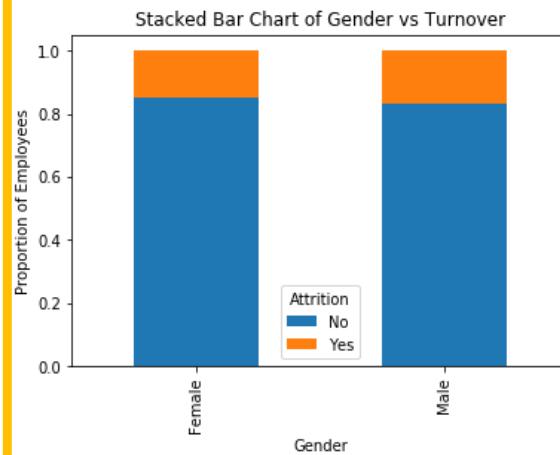
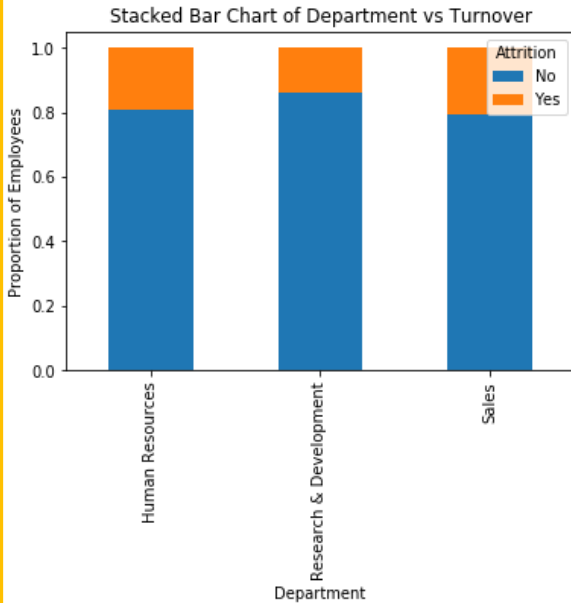
Category	Description	Additional Information
File Format	Comma delimited (CSV) file	Header rows
Number of Records	1470	Imbalanced data set
Number of Fields	35	Demographic Employee job/performance Employee history Employee survey
Response Variable	Whether the employee left the company (0 = No, 1 = Yes)	Yes = 237 No = 1233
Variable Types	Mix of numeric, categorical, and binary.	
Number of Missing Records	0	No imputation was conducted
Time Frame	2015 - 2016	



# Initial Observations

## Stacked Box Plots – Categorical Variables

## Observations



Highest turnover in HR & Sales

- Attrition slightly higher for males than females



- Single

- Overtime



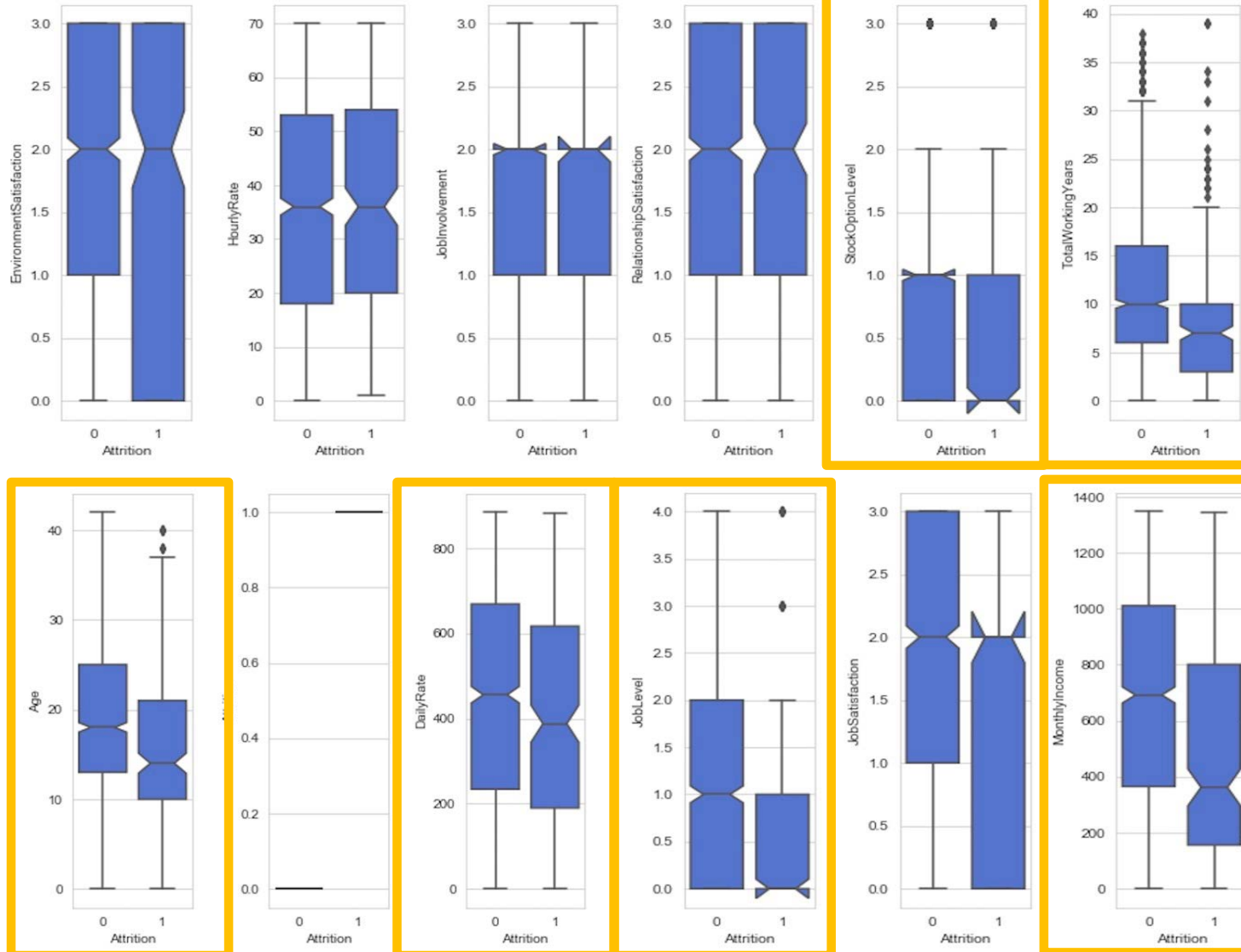
- Travel frequently

- HR, marketing and technical degrees



# Initial Observations

## Notched Box Plots – Continuous Variables



## Statistically Different Medians



- Stock Options Level

- Total Working Years



- Age

- Daily Rate



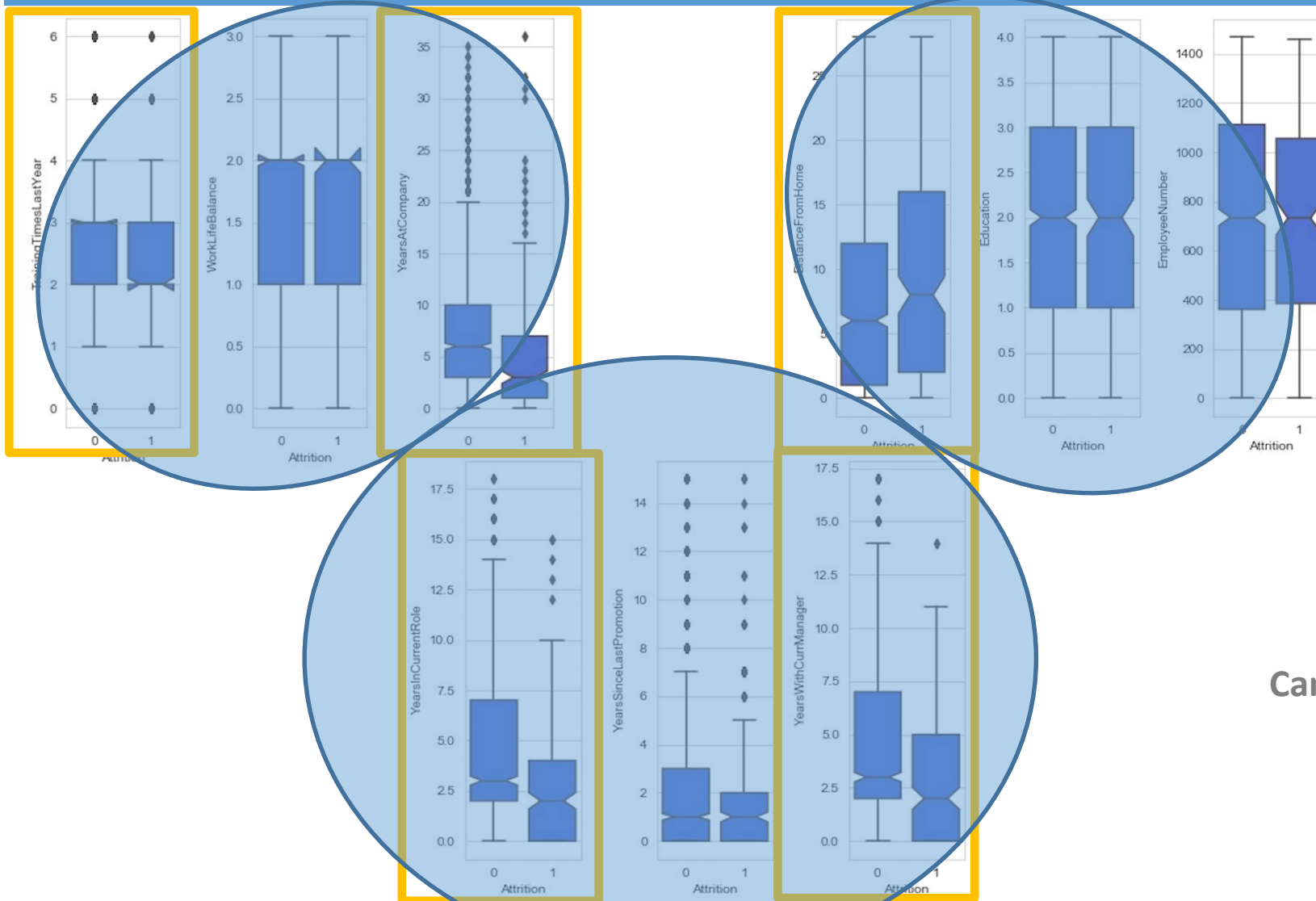
- Job Level

- Monthly Income



# Initial Observations

## Notched Box Plots- Continuous Variables



## Statistically Different Medians



- Training Times Last Year

- Years at Company



- Distance From Home

- Years in Current Role



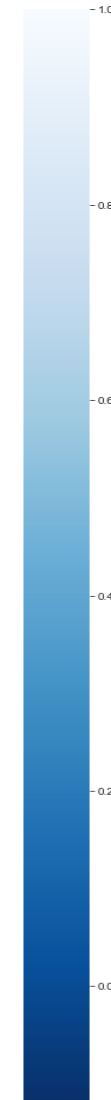
- Years with Current Manager

Can you spot the hidden Mickey?



# Initial Observations

Age	1.0	0.0	-0.0	0.2	0.0	0.0	0.0	0.5	-0.0	0.5	0.0	0.3	0.0	0.1	0.0	0.7	-0.0	-0.0	0.3	0.2	0.2	0.2
DailyRate	0.0	1.0	-0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.0	-0.0	0.0	-0.0	-0.0
DistanceFromHome	-0.0	-0.0	1.0	0.0	-0.0	0.0	0.0	0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	0.0	0.0	-0.0	-0.0	0.0	0.0	0.0	0.0
Education	0.2	-0.0	0.0	1.0	-0.0	0.0	0.0	0.1	-0.0	0.1	-0.0	0.1	-0.0	-0.0	0.0	0.1	-0.0	0.0	0.1	0.1	0.1	0.1
EnvironmentSatisfaction	0.0	0.0	-0.0	-0.0	1.0	-0.0	-0.0	0.0	-0.0	-0.0	0.0	0.0	-0.0	0.0	0.0	-0.0	-0.0	0.0	0.0	0.0	0.0	-0.0
HourlyRate	0.0	0.0	0.0	0.0	-0.0	1.0	0.0	-0.0	-0.1	-0.0	-0.0	0.0	-0.0	0.0	0.1	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0
JobInvolvement	0.0	0.0	0.0	0.0	-0.0	0.0	1.0	-0.0	-0.0	-0.0	-0.0	0.0	-0.0	0.0	0.0	-0.0	-0.0	-0.0	0.0	0.0	0.0	0.0
JobLevel	0.5	0.0	0.0	0.1	0.0	-0.0	-0.0	1.0	-0.0	0.9	0.0	0.1	-0.0	0.0	0.0	0.8	-0.0	0.0	0.5	0.4	0.4	0.4
JobSatisfaction	-0.0	0.0	-0.0	-0.0	-0.0	-0.1	-0.0	-0.0	1.0	0.0	0.0	-0.1	0.0	-0.0	0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0
MonthlyIncome	0.5	0.0	0.0	0.1	-0.0	-0.0	-0.0	0.9	0.0	1.0	0.1	0.2	-0.0	0.0	0.0	0.7	-0.0	0.0	0.5	0.4	0.3	0.4
MonthlyRate	0.0	-0.0	0.0	-0.0	0.0	-0.0	-0.0	0.0	0.0	0.1	1.0	0.0	-0.0	-0.0	-0.0	0.0	0.0	-0.0	-0.0	0.0	0.0	-0.0
NumCompaniesWorked	0.3	0.0	-0.0	0.1	0.0	0.0	0.0	0.1	-0.1	0.2	0.0	1.0	-0.0	0.1	0.0	0.2	-0.1	-0.0	-0.1	-0.1	-0.0	-0.1
PercentSalaryHike	0.0	0.0	0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	1.0	-0.0	0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0
RelationshipSatisfaction	0.1	0.0	0.0	-0.0	0.0	0.0	0.0	0.0	-0.0	0.0	-0.0	0.1	-0.0	1.0	-0.0	0.0	0.0	0.0	-0.0	0.0	0.0	-0.0
StockOptionLevel	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.0	0.0	0.0	-0.0	1.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
TotalWorkingYears	0.7	0.0	0.0	0.1	-0.0	-0.0	-0.0	0.8	-0.0	0.7	0.0	0.2	-0.0	0.0	0.0	1.0	-0.0	0.0	0.6	0.5	0.4	0.5
TrainingTimesLastYear	-0.0	0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	-0.0	0.0	-0.1	-0.0	0.0	0.0	-0.0	-0.0	1.0	0.0	0.0	-0.0	-0.0	-0.0
WorkLifeBalance	-0.0	-0.0	-0.0	0.0	0.0	-0.0	-0.0	0.0	-0.0	0.0	0.0	-0.0	-0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
YearsAtCompany	0.3	-0.0	0.0	0.1	0.0	-0.0	-0.0	0.5	-0.0	0.5	-0.0	-0.1	-0.0	0.0	0.0	0.6	0.0	0.0	1.0	0.8	0.6	0.8
YearsInCurrentRole	0.2	0.0	0.0	0.1	0.0	-0.0	0.0	0.4	-0.0	0.4	-0.0	-0.1	-0.0	-0.0	0.1	0.5	-0.0	0.0	0.8	1.0	0.5	0.7
YearsSinceLastPromotion	0.2	-0.0	0.0	0.1	0.0	-0.0	-0.0	0.4	-0.0	0.3	0.0	-0.0	-0.0	0.0	0.0	0.4	-0.0	0.0	0.6	0.5	1.0	0.5
YearsWithCurrManager	0.2	-0.0	0.0	0.1	-0.0	-0.0	0.0	0.4	-0.0	0.4	-0.0	-0.1	-0.0	-0.0	0.0	0.5	-0.0	0.0	0.8	0.7	0.5	1.0



## Highly Correlated Variables

- ~~Percent Salary Hike vs. Performance Rating (0.77)~~
- ~~Years At Company vs. Years With Current Manager (0.77)~~
- ~~Years At Company vs. Years In Current Role (0.76)~~
- ~~Monthly Income vs. Job Level (0.89)~~
- ~~Monthly Income vs. Total Years Worked (0.70)~~

# Initial Observations

## Candidate Features

Feature	# of Automated Feature Selection
JobSatisfaction	5
MaritalStatus	5
OverTime	5
TotalWorkingYears	5
DistanceFromHome	4
EnvironmentSatisfaction	4
NumCompaniesWorked	4
TrainingTimesLastYear	4
YearsAtCompany	4
Age	3
DailyRate	3
JobInvolvement	3
MonthlyRate	3
RelationshipSatisfaction	3
YearsSinceLastPromotion	3
Gender	2
HourlyRate	2
WorkLifeBalance	2
Department	1
EducationField	1
JobRole	1
PerformanceRating	1
StockOptionLevel	1

## Automated Feature Selection

### Automated feature selection techniques:

- Univariate feature selection
- Recursive feature elimination (with logistic regression)
- Tree-based feature selection (feature importance)
- Decision tree
- Lasso embedded method



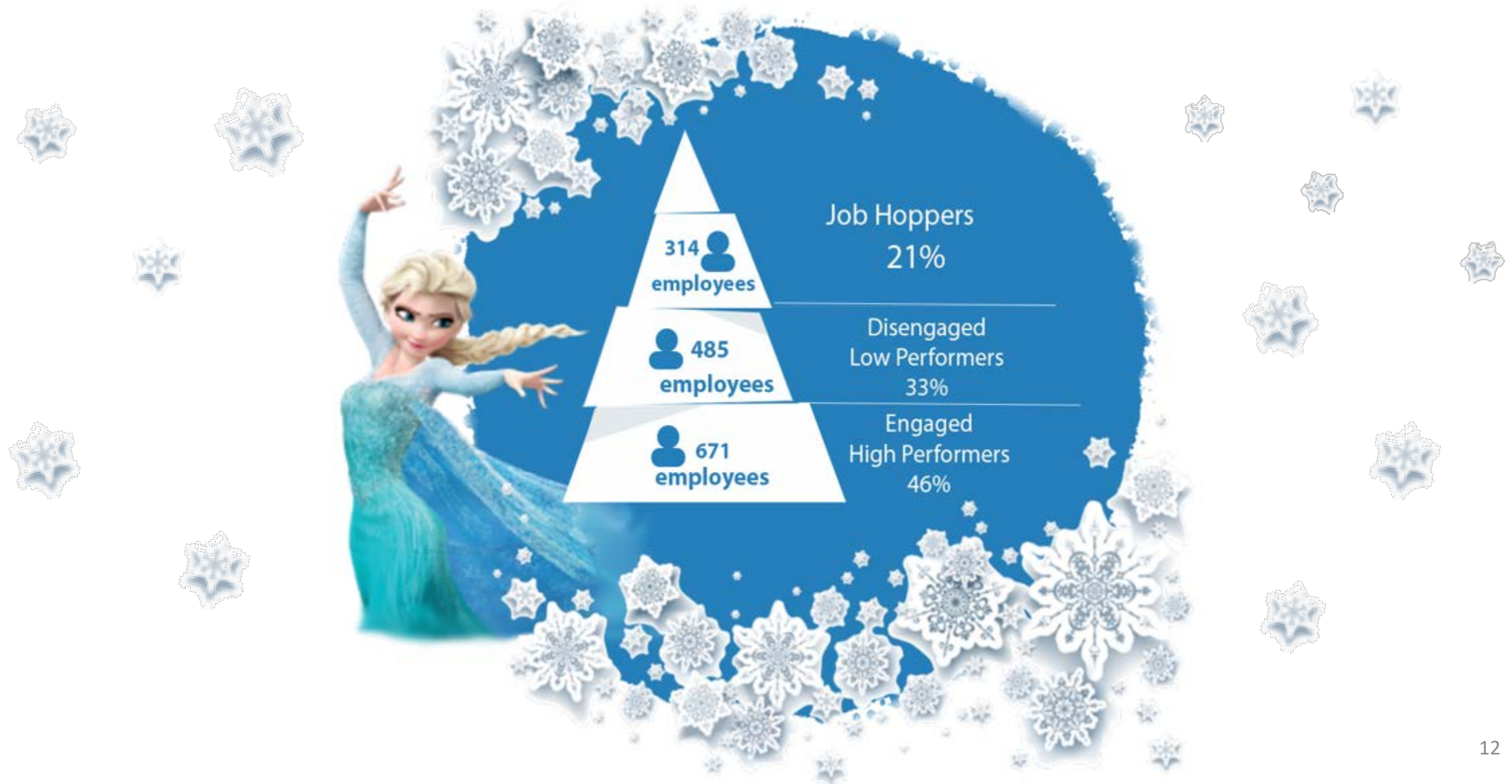
High importance features or most frequently identified features were identified as strong candidate features



Retention Risk Model

# Cluster Analysis

We also discovered three segments using cluster analysis: job hoppers, disengaged low performers, and engaged high performers.



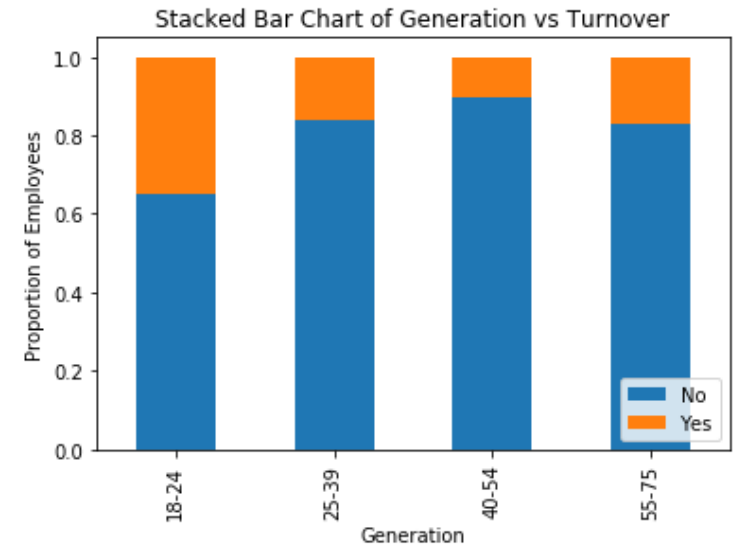
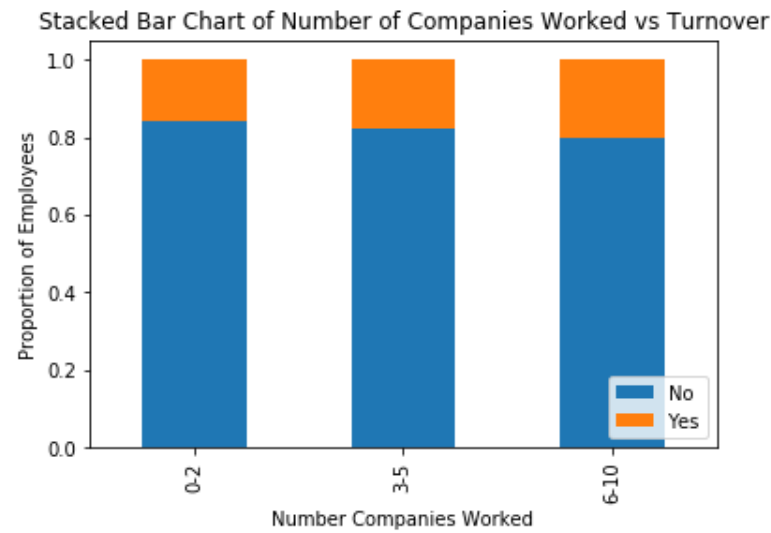
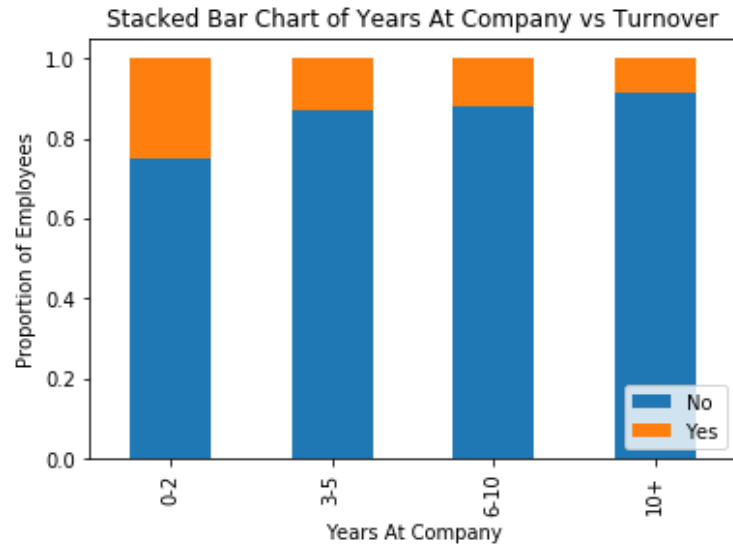
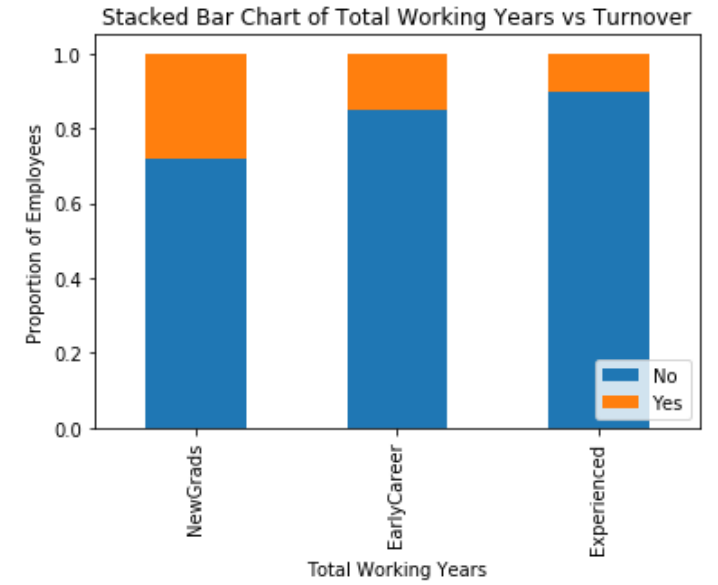
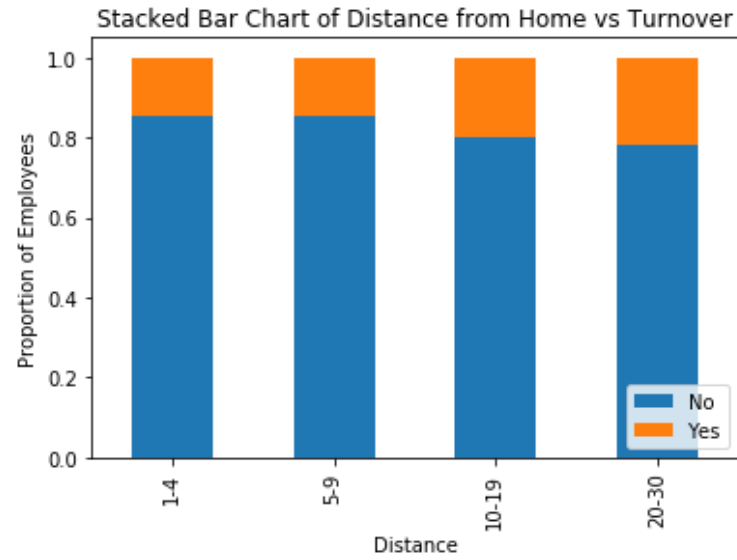
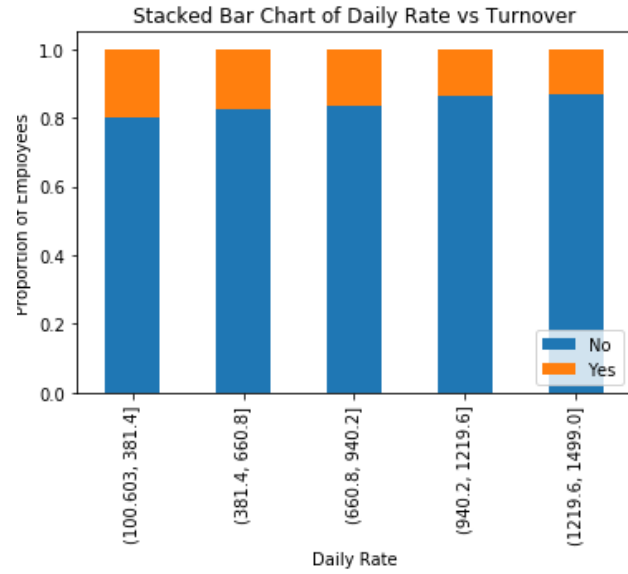


# Transformations

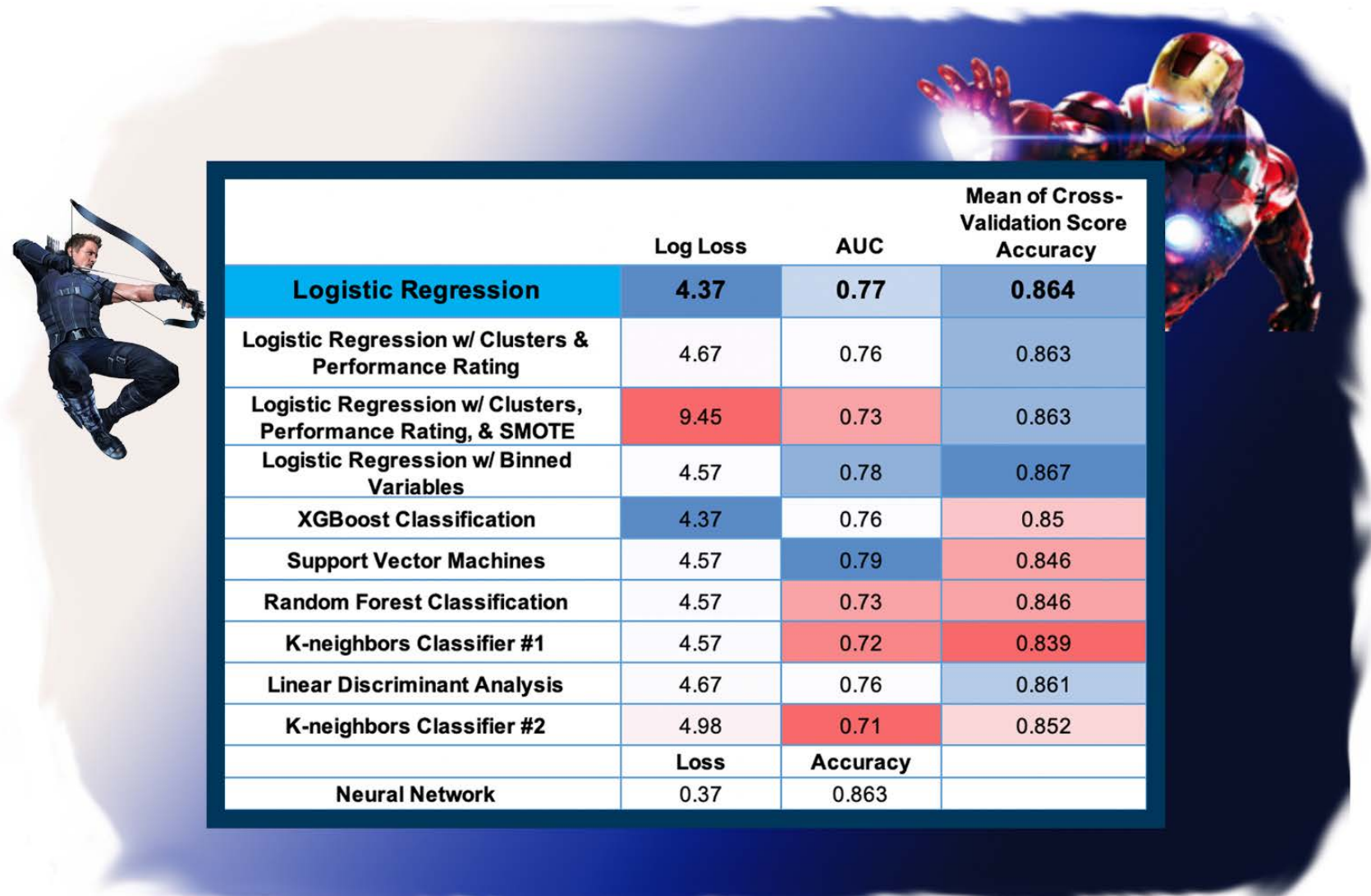


1. Majority of the variables did not include any major outliers
2. No missing values
3. Integer encoding applied to categorical variables
4. Feature Engineering
  - Binned variables
  - Cluster variable
5. Resampling Techniques & Imbalanced Data
  - SMOTE (aka: Synthetic Minority Over-sampling Technique)

# Binned Variables

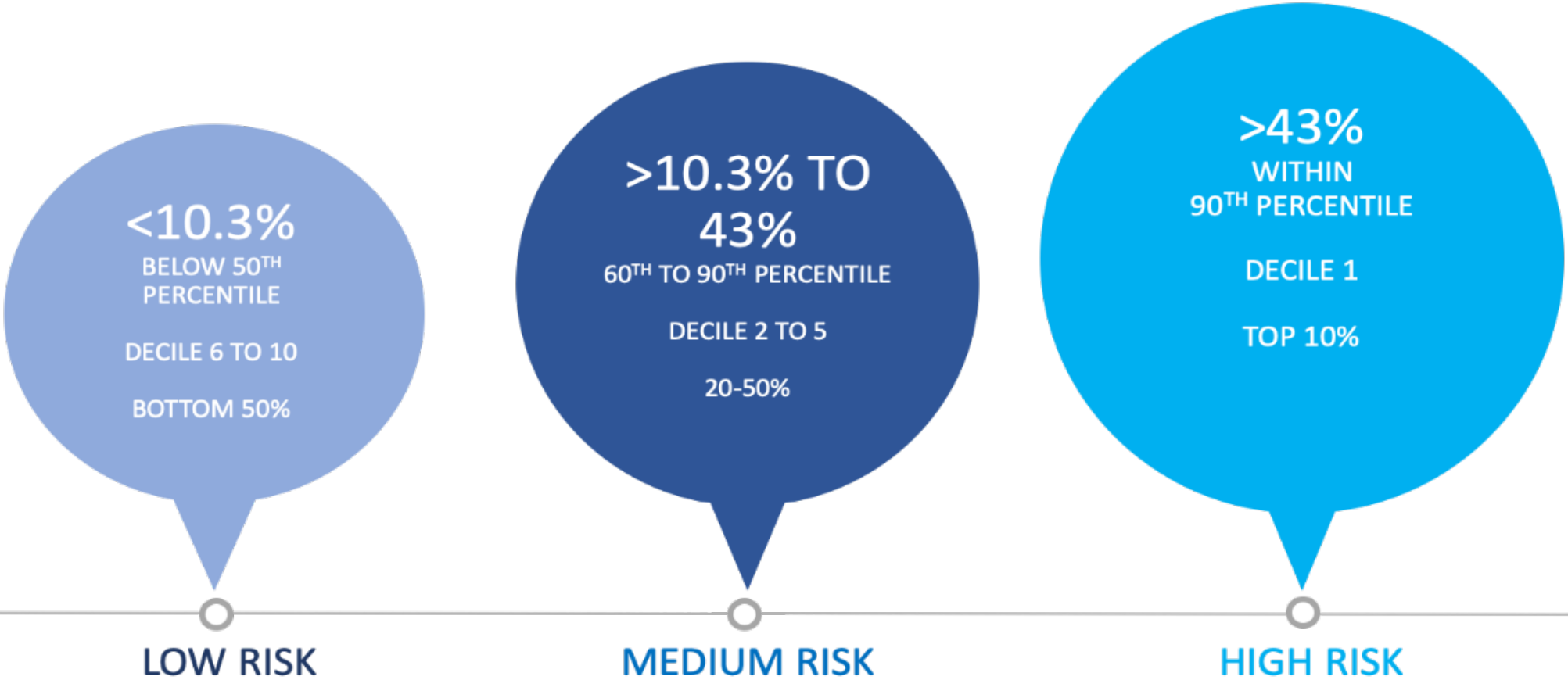


# Modeling Techniques Explored & Performance



# Retention Risk Probability Ranges

The visual below illustrates the probability ranges for low, medium, and high risk and the probability of leaving relative to the population based on our model.



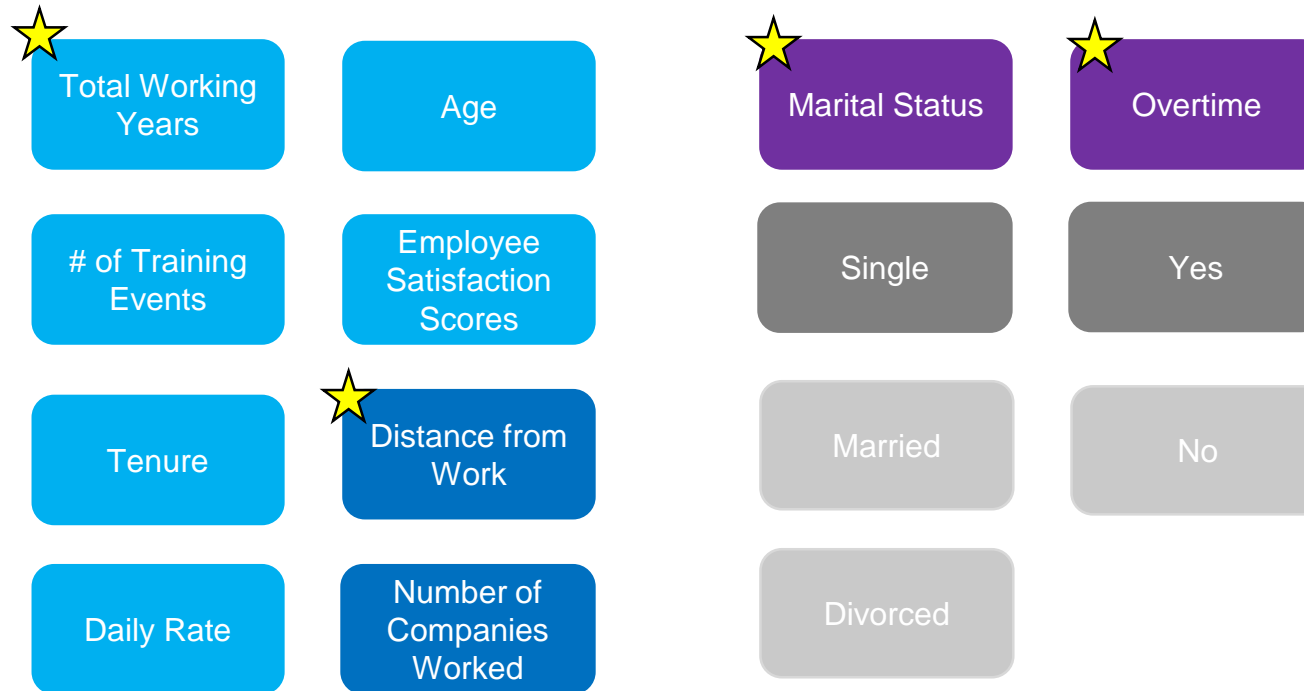
Average 16.8% Voluntary Rate Turnover



# Key Observations & Findings

Below outlines the current predictors and themes of voluntary turnover in Corporate that were included in our final model.

## Predictors that Prompt Others to Leave



- **Lower** the number, the more likely an employee is to leave the company.
- **Higher** the number, the more likely an employee is to leave the company.

- **Dark grey** represents which group is more likely to leave the company compared to the others.
- **Stars** represent most predictive variables.

# Recommendation & Implementation: Stay Survey



## Survey Content

### Stay Survey will cover:

- 19 questions (5-points scale of SD-SA including 2 open ended questions)
  - Learning
  - Pay/Compensation
  - Growth & Development Opportunities
  - Leadership
  - Work-Life Balance



# Further Recommendations

## Program



› Change anniversary rewards to 1, 3, 5, 7, 10 and every 5 years thereafter.

› Proximity based residency bonus program

› Implement mentoring programs

› Training, development, and stretch opportunities for disengaged low performers



## Policy



› Offer retention sign-on bonuses with clawback clauses

› Source local talent



› Unlimited paid time-off



› Flex hours & work from home policies



› Proactively perform salary adjustment to compete with market median



## Initiatives



› Implement a stay survey to identify targeted training and development opportunities



› Identify understaffed departments and increase staff through hiring or lateral transfers



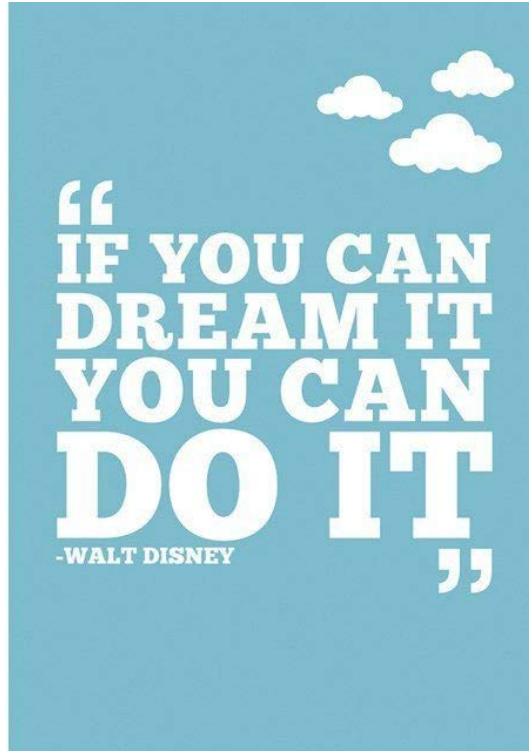
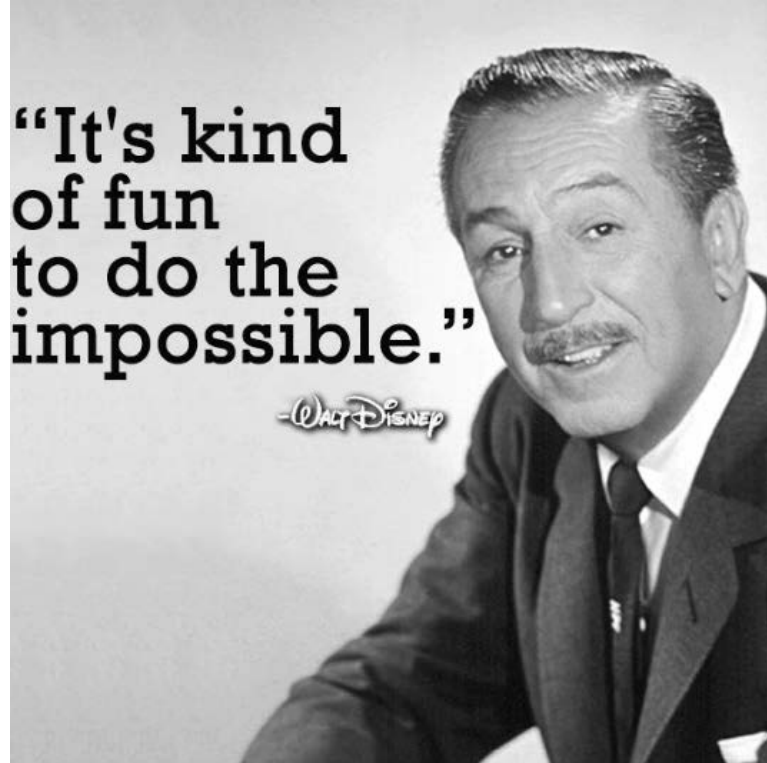
› Create career ladder/ lattice to provide opportunities for career progression



› Job hoppers/contingent workforce strategy



# Final Words







**THANK YOU FOR  
LISTENING!**

