



Retention Risk & Employee Turnover Strategy Executive Summary: Final Report

**Prepared for Dr. Donald Wedding, CHRO
Annual HR Summit Day 3
May 29, 2019**

People Insights Team

**Agnes Cheng
Brent Young
Mary Taylor
Michelle Eure**

***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.*

Problem Statement

The Walt Disney Company has undergone significant organizational and leadership changes over the last 5 years which has caused an increase in turnover. With the recent 20th Century Fox acquisition and competitive landscape, executives anticipate further disruption leading to the possible loss of high potential and key talent as well as increased attrition related costs.

Repercussions of high voluntary turnover:

- **High replacement costs** for the company. **150% of salary per departing employee.**
- **Projects get delayed** in a time where meeting deadlines are crucial.
- Work needs to be redistributed to the remaining employees, which causes worker fatigue, burnout, and **employee dissatisfaction.**
- New talent has to be recruited, trained and given **time to acclimatize** themselves to the company.
- Loss of **institutional knowledge.**

Business Objectives

This project leveraged data analytics to achieve four key business objectives which provide the CHRO the information, tools, and recommendations to achieve a yearly 10% reduction in voluntary turnover resulting in significant cost avoidance.

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



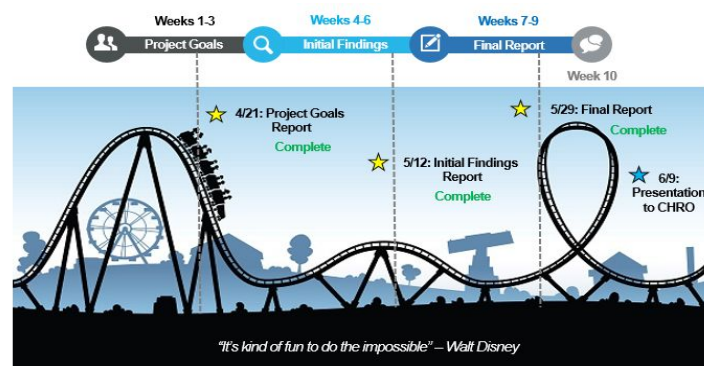
Objective	Deliverable	Success Criteria
<ul style="list-style-type: none"> • Identify high, medium, and low flight risk employees. 	<ul style="list-style-type: none"> • Exploratory data analysis of demographics, survey data, and attrition for employees. Cluster analysis. 	<ul style="list-style-type: none"> • Decrease in voluntary turnover rate and increase cost avoidance. • Increase in retaining key talent. • Accurate models.
<ul style="list-style-type: none"> • Determine which factors prompt employees to leave. 	<ul style="list-style-type: none"> • Build an explainable predictive model to uncover key insights, themes and predict the probability that someone would leave the company. 	<ul style="list-style-type: none"> • Changes to the work environment, policies, HR strategy, etc. • Decrease in voluntary turnover rate and costs. • Increase in employee satisfaction scores.

<ul style="list-style-type: none"> Create a Stay Survey and use the model results to help retain high-risk employees. 	<ul style="list-style-type: none"> Create questions, theoretical experiment process, and A/B testing approach. 	<ul style="list-style-type: none"> Shift from reactive to proactive mindset amongst leaders and HR. Increase in retaining key talent and career conversations. Increase in employee satisfaction scores.
<ul style="list-style-type: none"> Create an interactive dashboard and mobile app for Corporate HR and leaders. 	<ul style="list-style-type: none"> User-friendly dashboard and app that allows users to see low, medium, and high risk by team and individual along with relevant demographics. 	<ul style="list-style-type: none"> Usability and buy-in from CHRO, leaders, and HR.

Solution

To help Disney Corporate overcome the challenges of voluntary attrition, the People Insights' team utilized a data set that contains approximately 1,500 employee records to build several classification models to predict the probability that an employee will leave the company, determine which factors prompt employees to leave, and create a Stay Survey so that leaders/HR can use the model results to retain high-risk employees. Overall, this will help inform the development of a Corporate HR strategy to retain top talent, increase employee satisfaction/engagement survey scores, increase productivity, decrease voluntary turnover, and lower hiring/replacement costs, thus saving the company money. An interactive dashboard and mobile interface to track attrition risk were created for Corporate HR and leaders.

This project was completed over a period of ten weeks with a team of four data scientists. The workload was distributed into key streams including project management, exploratory data analysis, modeling, technical writing, visualization, dashboard and mobile application development, and presentation development. Team members were responsible for various streams. The full team provided contributions to each stream.



Key Findings

Observations and Learnings

The best performing model was Logistic Regression. We identified the factors and themes that prompt employees to leave the company so that the company can make changes to the work environment, policies, and HR strategy. Below are the characteristics that are predictive of employees who are more likely to leave. These are the variables that were included in the model in order of importance.

- Lower total working years
- Logged overtime over the course of 1 year (**7 times** more likely to leave compared to those who did not log overtime)
- Are single (**4 times** more likely to leave compared to those who are married)
- Live further away from work, especially those who live 20+ miles away from work
- Have worked at a number of companies in the recent past, especially those over 6+
- Younger in age, especially 18-24 age range
- Have low environment satisfaction, job satisfaction, job involvement, relationship satisfaction, or work-life balance survey scores
- Received a low amount of training events in the prior year
- Have lower daily rates (\$)
- Fewer years worked at the company, especially less than 6 years

Data Preparation

- The data was split into a training (70%) and test set (30%).
- Majority of the variables didn't include any major outliers. No missing values.
- The training set was used to fit the models. K-fold cross-validation, which is a resampling procedure, was also conducted on the training dataset and was used for model selection purposes. The best model was then applied to the test set for final evaluation.
- For categorical data, label encoding was applied.

Exploratory Data Analysis

- The data for employees primarily comes from the Research & Development and Sales departments, with HR having the least amount of employees.
- Most of the nine job roles are primarily Sales Executives, Research Scientists, and Laboratory Technicians.
- The data set is imbalanced with a low rate of attrition (237 out of 1470 employees left the company) so we are naturally trying to predict "rare events".
- Majority of the variables were skewed.
- A correlation matrix revealed several variables that were correlated. For instance, Percent Salary Hike vs. Performance Rating.
- Variation Selection and Resampling Techniques: Bar plots, Stacked bar plots, box plots, and automated variable selection techniques such as univariate feature selection,

recursive feature elimination (with logistic regression), tree-based feature selection (feature importance), a decision tree, and lasso embedded method on the training data set. We also experimented with SMOTE (Synthetic Minority Over-sampling Technique).

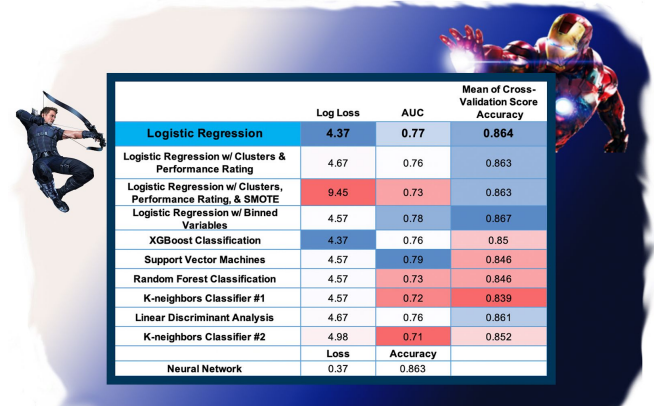
- **Feature Engineering:** New and binned variables were created to help generate additional insights, targeted recommendations, and potentially increase model performance. For example, cluster segment and ranges for Age, Total Working Years, Distance to Work, Number of Companies Worked, Years At Company, and Daily Rate.

Cluster Analysis: As part of our EDA, we also developed employee clusters using the survey and demographic data in order to better understand how employees are feeling. This allows HR to develop targeted strategies and policies to increase employee engagement and satisfaction, which can ultimately improve employee retention. Three segments were determined.



Segment	Description
Job Hoppers	This group dislikes the company's work environment, are indifferent about their satisfaction with their job, they don't stay at companies very long, have long commute times, and are high performers.
Disengaged Low Performers	This group loves the company's work environment, are somewhat dissatisfied/indifferent about their job, receive lower performance ratings, lower annual salary increases, lower income, and the least training and development amongst their peers. However, many of them have short commutes.
Engaged High Performers	This group likes the company's work environment and absolutely loves their job. They are highly committed to their role and to the company. These employees are also high performers and as a result, they are rewarded with the best performance ratings, highest salary increases, and stock options compared to their peers. They also tend to have short commutes and a good monthly income.

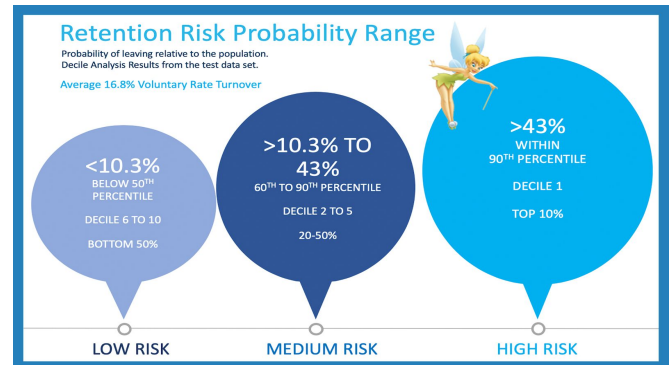
Predictive Modeling: Given that this a classification problem, a mix of modeling methods were used; from explainable modeling techniques such as logistic regression and decision trees to more complex modeling techniques such as linear discriminant analysis, XGBoost, SVM, random forest, K-Nearest



	Log Loss	AUC	Mean of Cross-Validation Score Accuracy
Logistic Regression	4.37	0.77	0.864
Logistic Regression w/ Clusters & Performance Rating	4.67	0.76	0.863
Logistic Regression w/ Clusters, Performance Rating, & SMOTE	9.45	0.73	0.863
Logistic Regression w/ Binned Variables	4.57	0.78	0.867
XGBoost Classification	4.37	0.76	0.85
Support Vector Machines	4.57	0.79	0.846
Random Forest Classification	4.57	0.73	0.846
K-neighbors Classifier #1	4.57	0.72	0.839
Linear Discriminant Analysis	4.67	0.76	0.861
K-neighbors Classifier #2	4.98	0.71	0.852
Loss	Accuracy		
Neural Network	0.37	0.863	

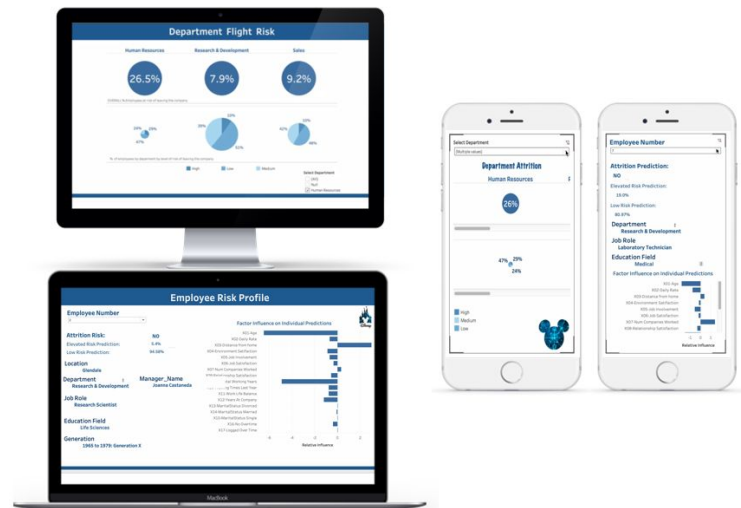
Neighbors and Neural Network. All modeling was completed in Python and the code was shared using CoLab. The models were assessed using statistical evaluation criteria such as AUC, ROC Curves, Logarithmic loss (Logloss), mean k-folds cross-validation accuracy scores, and Decile Analysis. The **Logistic Regression model yielded the best predictive performance metrics and is also the most interpretable**. As a result, the team moved forward with the Logistic regression model. *Note: We also experimented with different variations of the Logistic Regression model, but none of them outperformed the original logistic regression model.*

Decile Analysis: To determine probability cutoffs for high, medium, and low-risk employees, the People Insights team decided to use the decile analysis on the test data set for the logistic regression model. Here is a visual summary of the retention risk probability range based on the decile analysis.



Dashboard & Mobile User Interface

The People Insights team created an interactive dashboard and mobile interface that allows the CHRO and team to monitor **predicted attrition risk** in a user-friendly way (e.g., low, medium, and high risk by team and individual along with relevant demographics). Tableau was chosen because Disney has an enterprise license and the appropriate security is already integrated within the business. The results can be viewed on a desktop, laptop, tablet, or mobile phone. This dashboard is accessible through the web or the Tableau Workbook application.



Recommendations

The People Insights team have generated very thoughtful and thorough recommendations on how this project can be implemented and used to:

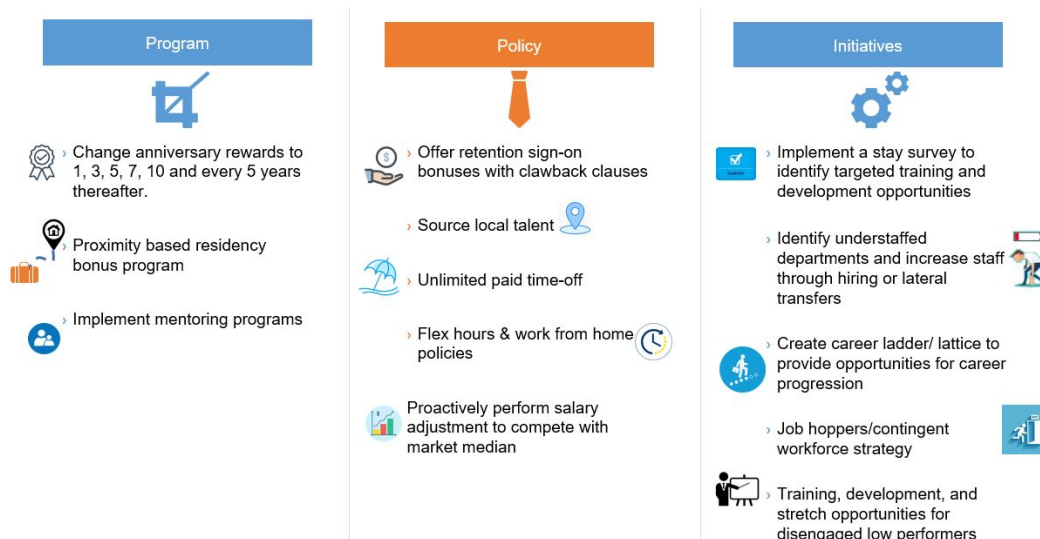
- 1) Make proactive decisions in regards to key talent using a stay survey
- 2) Make changes to employee policy, initiatives, and programs
- 3) Make further enhancements to data collection and variable creation in the future
- 4) Conduct further exploration.

Stay Survey

The overall objective and purpose of the 19 question stay survey is so that leaders/HR can use the retention risk model results to retain high-risk employees and showcase the value of the model and approach so that it can possibly be expanded to the rest of Disney Corporate and to the other business segments. This will be accomplished using a cross-sectional study to test whether the model truly helps us retain employees. The survey includes questions in regards to Learning, Pay/Compensation, Growth & Development, Leadership, and Work-Life Balance.

Policy Recommendations Based on Findings

The programs, policy changes, and/or initiatives recommended below are based upon our logistic regression model results and insights that we discovered through this project:



Additional Data Collection, Variable Creation, & Explorations Moving Forward

Our research on the use of analytics in HR also identified data that could have been useful, but was not included in the data set. This includes variables such as location data, number of dependents, etc. We also advise integrating additional employee survey related data, identify retention areas by department, leader, role, etc. and gauge the effectiveness of management in supporting both the company and the employee through their individual leadership styles.

Final Words

In the words of Walt Disney, “It’s kind of fun to do the impossible”. This 10-week project has given us the opportunity to create, innovate, and collaborate in a way that generated leading edge and forward-thinking ideas. We not only see this as an opportunity to inspire other HR and business leaders at Disney, but for the greater use of advanced modeling in human resources in all companies. It’s been an honor to work with the Disney team and contribute to the Disney of the future.

