Predicting Attrition Risk using Workterra Data

Faizan Javed & Phuong Hoang

CareerBuilder Hackathon
November, 2017

What is post-hire Human Capital Management (HCM)?

Payroll

Benefits administration (health insurance/wellness, retirement accounts, vacation/sick leave, ..)

Time & attendance

Learning Management

Talent Management

May seem like a mature market ..but is more of an early mainstream market since focus so far has been on automating established workflows

Talent Management (aka employee engagement)

Workforce planning
(aligning needs of an organization with its workforce)

Optimizing performance

Retraining

Identifying leaders

Reduce/prevent voluntary employee attrition

Overview

Objectives

- Determine factors that leads to attrition.
- Predict attrition risk.

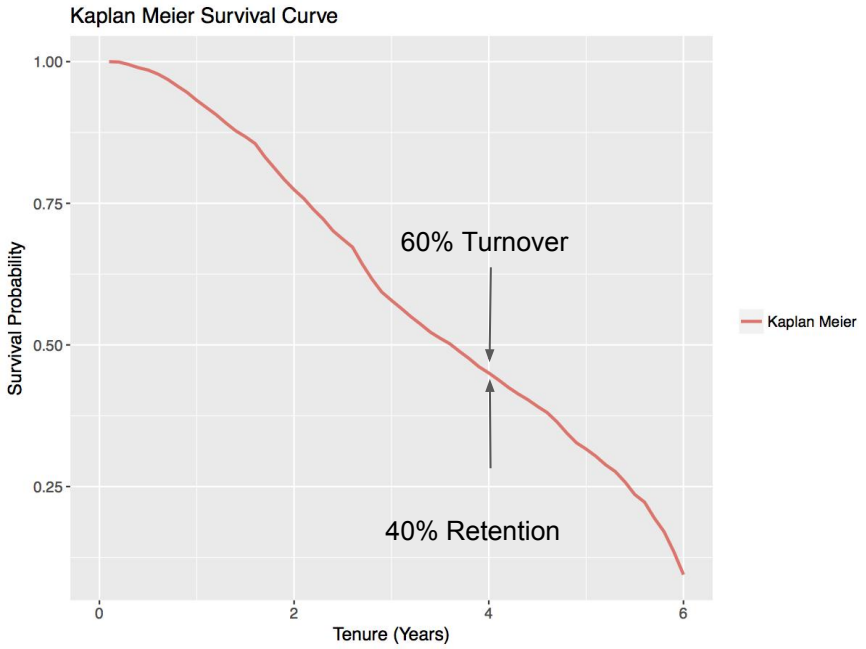
Data Source

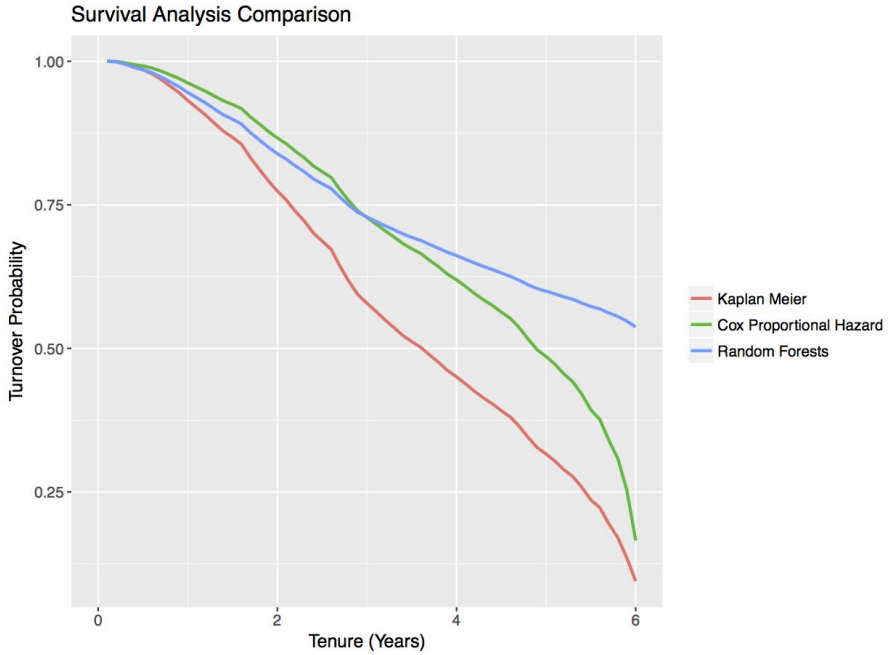
- Workterra anonymous dataset (65K data points, 18 companies)
- Available features: age, location, job title, tenure, client ID, employment status

Methodology

- Survival analysis methods
  - Kaplan Meier
  - Cox Proportional Hazard
- Machine learning
  - Random Forests

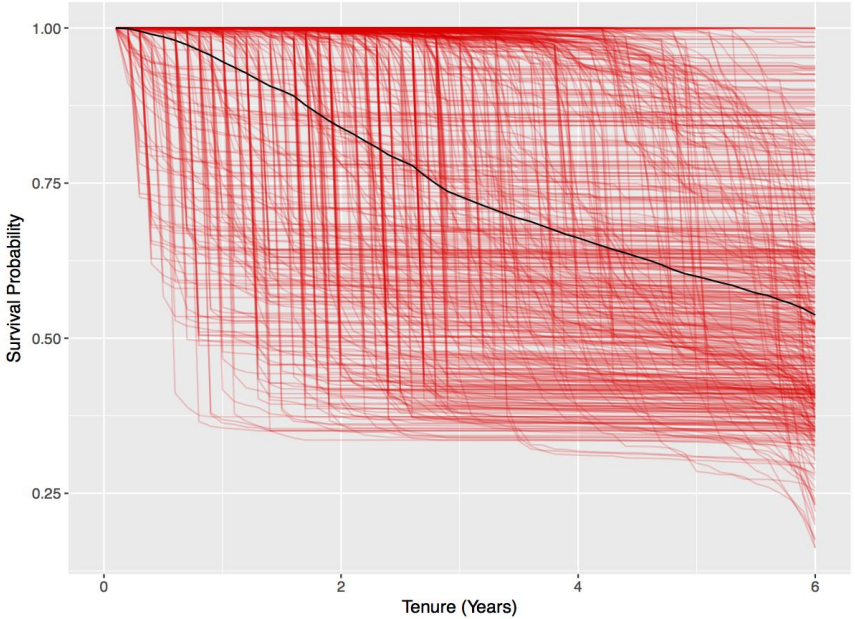
Survival Analysis





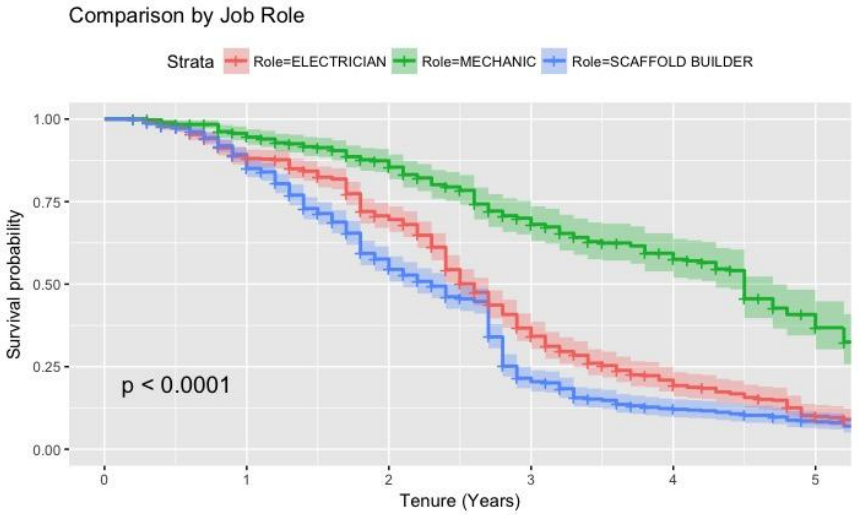
Results

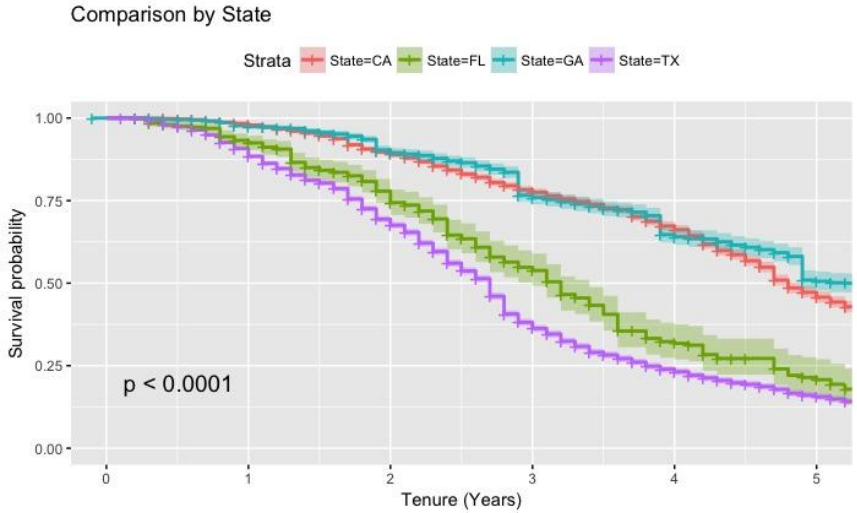
- Our model is **83% accurate** on Workterra dataset.
- Everyone has a different survival curve
- Attrition Risk = 1 - Survival Probability
- **Location**: Employees from California and Georgia are less likely to turnover than those from Texas and Florida.
- **Job Role and Job Level** are important factors, correlate with other variables such as overtime, salary hike, education, etc.

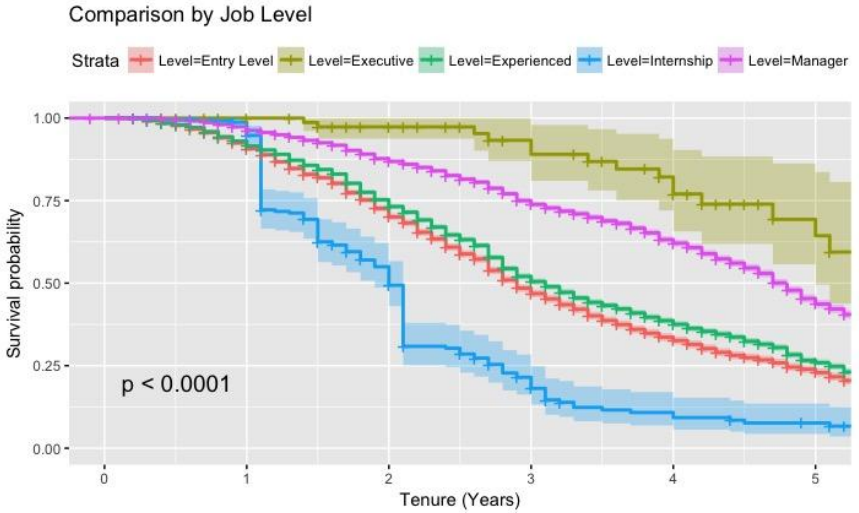


Results

- **Location**: Employees from California and Georgia are less likely to turnover than those from Texas and Florida.
- **Job Role and Job Level** are important factors, correlate with other variables such as overtime, salary hike, education, etc.







Future Works

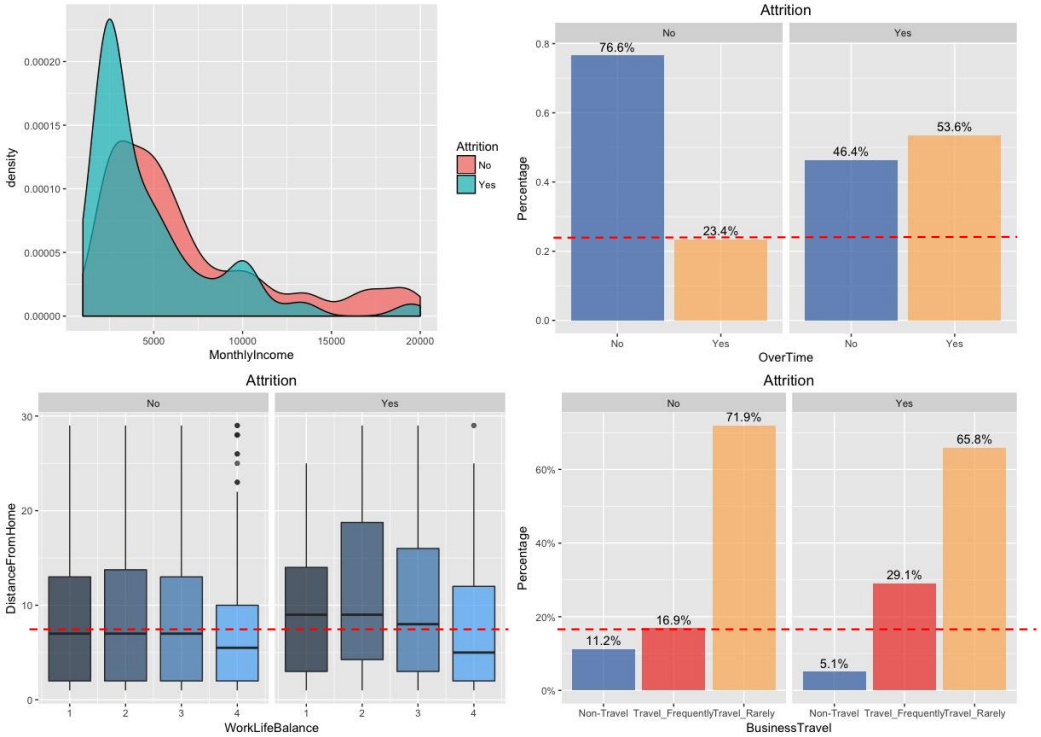
What else do we need from Workterra?

- Effort-reward factors are important.
  - Salary comparison ratio
  - Overtime
  - Distance from home/ Work life balance
  - Business Travel

  Example: Identify people who are working overtime and having relatively low salary

- Performance review data:
  - Years since last promotion
  - Employment/job/relationship satisfaction
  - Performance ratings

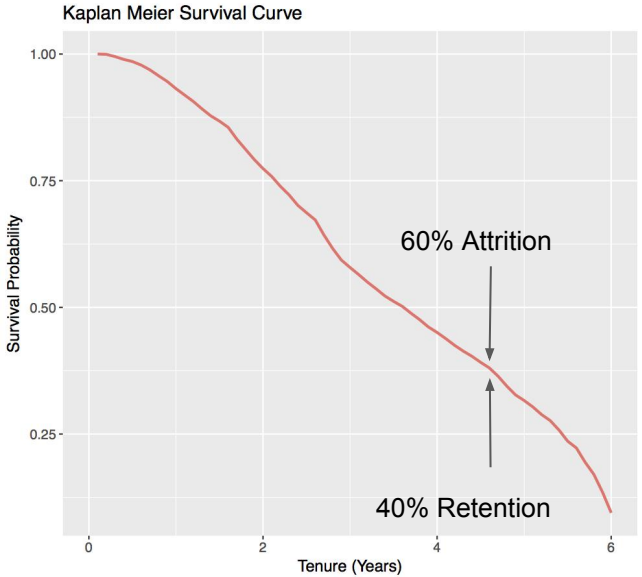
We need quarterly data snapshots to keep track of profile changes, discovering new patterns that lead to attrition, and continuously refine our model to improve performance **accuracy** and **stability**.

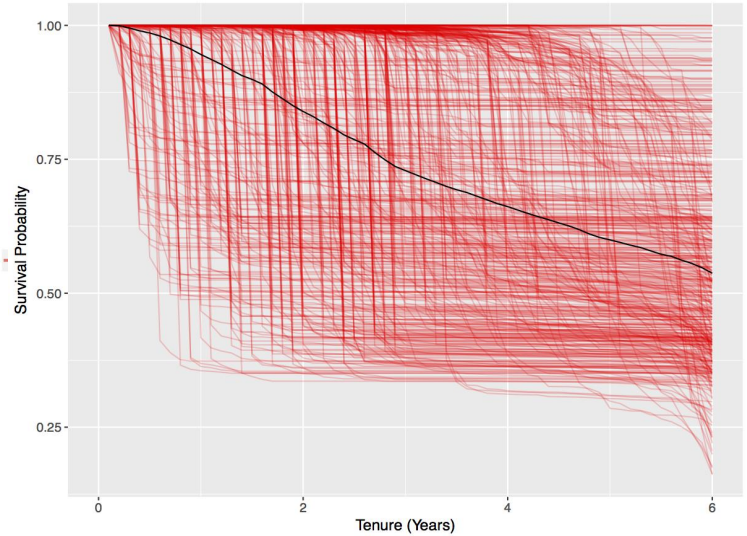


Experiments with IBM Watson Analytics fictional dataset

End

Survival Analysis





- Everyone has an attrition curve
- Attrition risk = 1 - Survival Probability

- Different people, different attrition curve

Results

- Our model is **83% accurate** on Workterra dataset.
- **Location**: Employees from California and Georgia are less likely to turnover than those from Texas and Florida.
- **Job Role and Job Level** are important factors, correlate with other variables such as overtime, salary hike, education, etc.

Results

- Our model is **83%** **accurate** on Workterra dataset.
- **Location**: Employees from California and Georgia are less likely to turnover than those from Texas and Florida.
- **Job Role and Job Level** are important factors, correlate with other variables such as overtime, salary hike, education, etc.