Survival Analysis
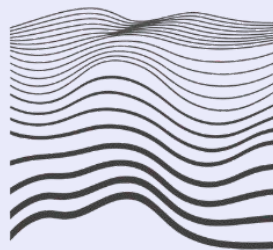
project by:

LINDSAY MCFARLANE

www.lindsaymsba.com

831.601.7684 | lindsay.alexandra14@gmail.com | linkedin: lindsay-mcfarlane

1 Summary

2 Full Deck

Summary

Research question:
"What factors impact employee churn?"

Methodology

We only know the outcome/tenure for *churned* employees.

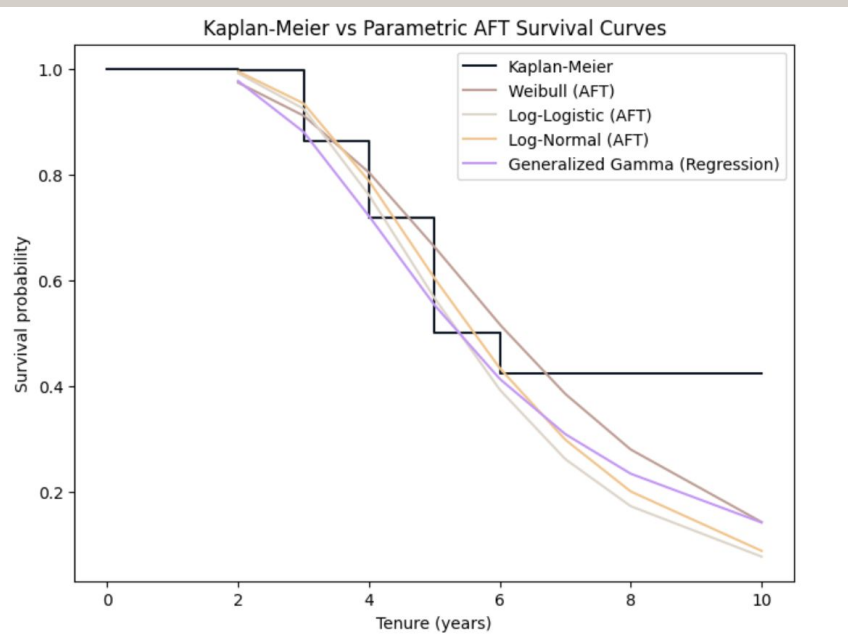**We cannot see the total time to churn for those who have not churned yet.**

If we *only* look at churned customers, *and/or use regression* the data will be biased.

We therefore use Survival Analysis, which accounts for censored data

Best Model

Log-Logistic AFT ("accelerated failure time") was the best model fit and was tuned. It is only reliable for covariate rankings, not time-to-event predictions
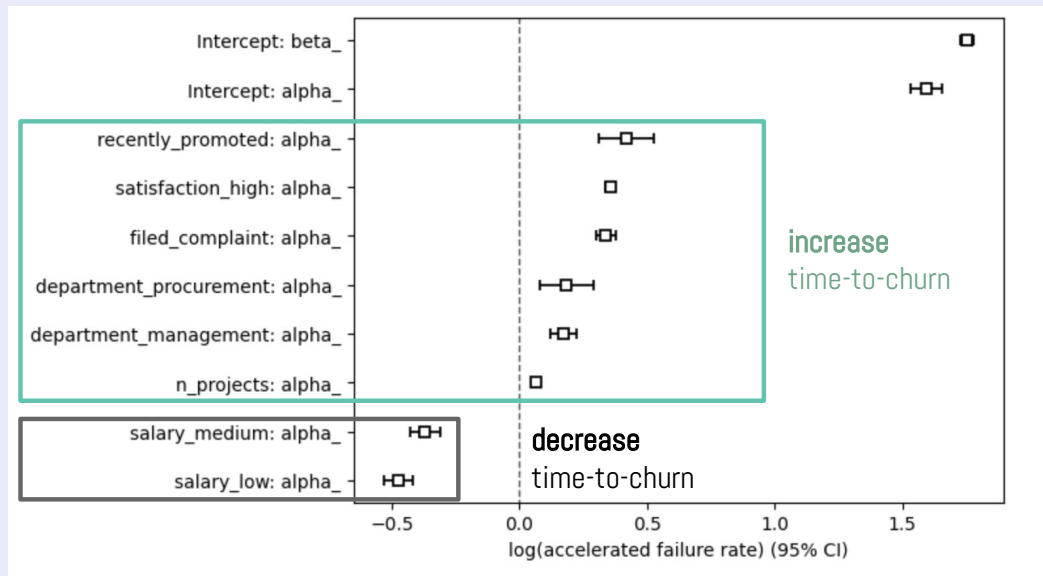

Kaplan-Meier vs Parametric AFT Survival Curves

Legend:
— Kaplan-Meier
— Weibull (AFT)
— Log-Logistic (AFT)
— Log-Normal (AFT)
— Generalized Gamma (Regression)

X-axis: Tenure (years)
Y-axis: Survival probability

Results

Major factors affecting time-to-churn:

1. Being recently promoted *(increases)* 👍
2. Having high satisfaction *(increases)* 👍
3. Filing a complaint *(increases)* !
4. Having a low salary *(decreases)* 👍
5. Being in Management Dept *(increases)*
6. Being in Procurement Dept *(increases)*
7. Number of projects *(increases)* !

NOT important:

- Monthly hours !
- Being in any other dept

increase
time-to-churn

decrease
time-to-churn

log(accelerated failure rate) (95% CI)

👍 *makes sense*

! *interesting*

Recommendation

I recommend that leaders prioritize the following:

1. **Promotions** when appropriate
2. **Surveys to understand satisfaction drivers**
3. **Investigate cases where complaints were filed**
   to understand pattern of it improving retention
4. **Industry-competitive salaries & raises**
5. **Survey questions/analysis**
   specifically for mgmt. & procurement
6. **Check-ins to offer more projects**
   to interested employees

#1
#2
#3

notebook link:

Technical Setup

Data

Source: Kaggle "employee churn" synthetic dataset

Type: Structured

Observations: 14,249 (originally)

Features: 9 (work data points)

Target: "tenure (years)"

Setup

Language: Python

Packages: lifelines, numpy, pandas, matplotlib, seaborn

Compute: Python3 CPU in Google Colab

Evaluation Metrics

- Model Selection: AIC & C-index
- Covariate impact: exp(coef) rankings

Click for Full Deck