Employee Churn: Survival Analysis

Background and Goals:

1. Analyzing employee life cycle could have significant impact on talent operation strategy regarding talent attraction, activation and attrition. Figuring out what's the most hazardous point for employee is absolutely important for in-time intervention.
2. Predicting employees' attrition could be beneficial in multiple ways: 1) estimating headcounts 2) Reducing attrition rate by taking action in advance.
3. In this analysis, we not only identify important factors that affect employee tenure and further dive in the degree of impact of these factors, quantifying the influence.

Methods

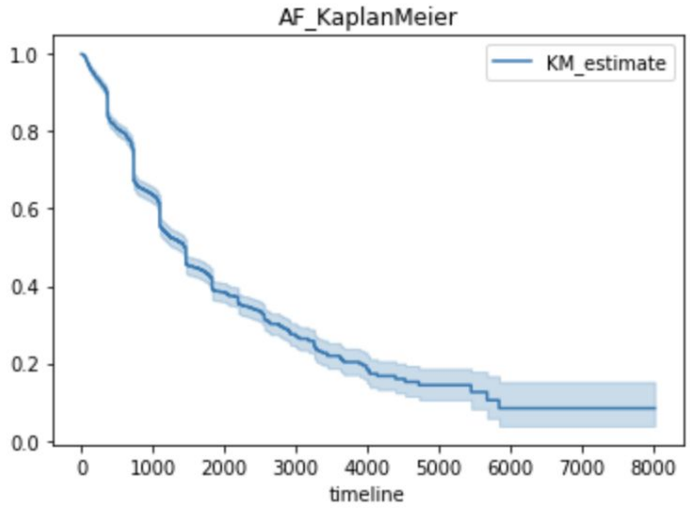
1. We constructed both parametric and non-parametric models using lifelines package on Python 3.0.
2. There are **1800+** complete cases imported to the models. The data was split into training and testing subsets for validation purpose
3. We only look at one position, the pillar role within the organization.
4. Data used from multiple resources, includes **Annual Salary, Ethnicity, Gender, Site Region, Race, Seniority Level, Tenure and Attrition Status**
5. The models helps identify the survival(Retention) and hazard(Dropout) probability

Method to look at the attrition of this role: Survival Analysis

Non-parametric model: Kaplan Meier model

This line graph roughly estimates the probability of surviving longer than a certain time point.

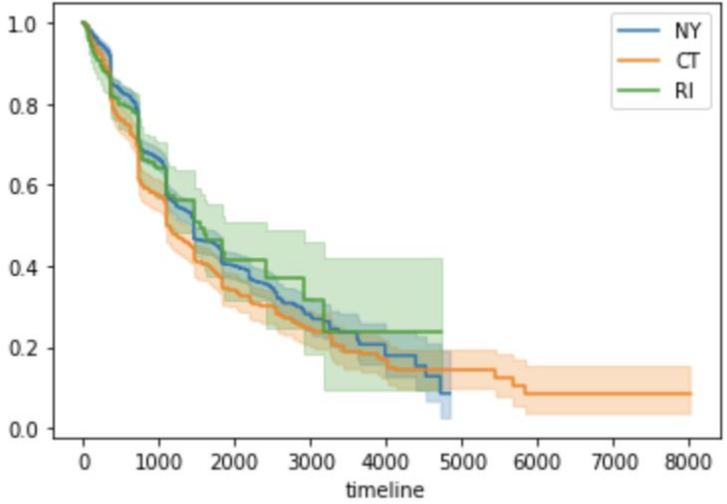
E.g The chance for this position to stay longer than 3.9 years is 50%. However, the median survival time here should not be 100% trusted since over 50% of the data are censored. Just a reference.



Method to look at the attrition of this role: Survival Analysis

Non-parametric model: Kaplan Meier model

Based on previous research outcomes, we noticed that geographical location will have little noticeable impact on attrition. The survival graph by location again indicates the difference due to geography is not statistically significant, proved by the intersection of confidence interval could prove that.

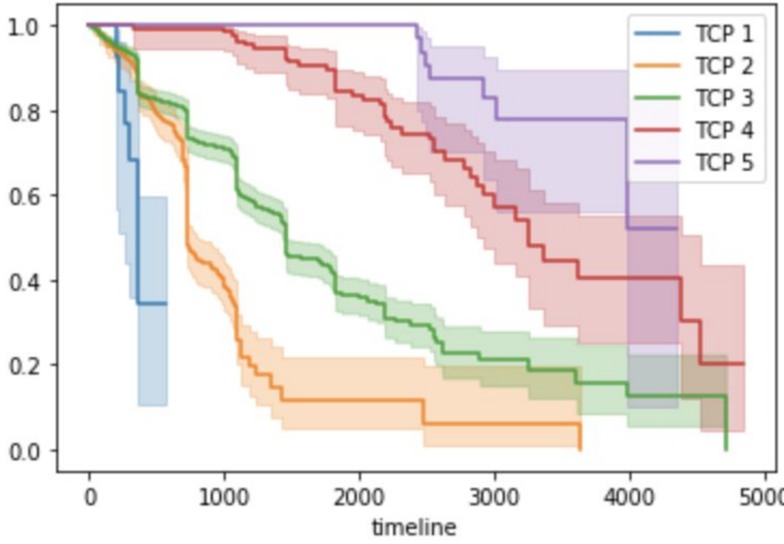


Method to look at the attrition of this role: Survival Analysis

Non-parametric model: Kaplan Meier model

Let's now look at how seniority distinguishes each other in terms of the survival possibility. Roughly speaking, the higher level the role is, later the survival function start to decline. This makes sense, because you have to stay long enough to get promoted to certain levels.

Of course, we have external hire for certain higher level positions. Nevertheless, due to the special characteristics of the role in nature. Level 4 and level 5 are most often developed internally, leaving small space for outsiders. Therefore, this small amount of outsiders will not have an apparent impact on the survival trend.



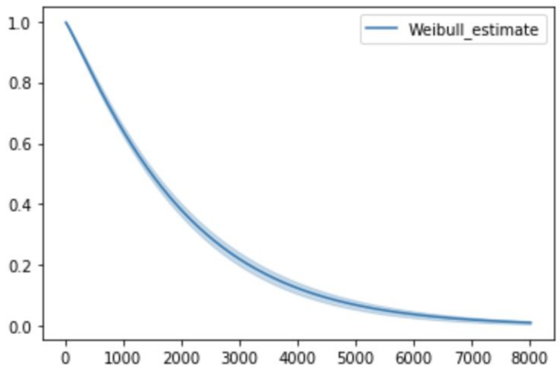
*note:TCP here indicates seniority, the larger the number is, the higher level

Method to look at the attrition of this role: Survival Analysis

Parametric model to quantify impact of covariate: **Weibull model**

The model index rho is ~1.1, indicating that the event rate(dropout) increase over time. But, the increase speed is closer to constant, the change in the possibility of dropping out should not be drastic.

This can be one of the metrics for evaluating any of our new retention practice.



Method to look at the attrition of this role: Survival Analysis

Parametric model to quantify impact of covariate: **Weibull model**

Weibulll is able to include continuous covariates to evaluate its/their impacts on the survival function. The following is the outcome of the Weibull model:

param	covariate	coef	exp(coef)	p
lambda_	ANNUAL SALARY	0.000028	1.000028	3.522389e-56
	CT	-1.690473	0.184432	~~9.999995e-01~~
	Level	0.399794	1.491517	3.472677e-31
	NY	-1.999146	0.135451	~~9.999994e-01~~
	RI	-1.439487	0.237049	~~9.999996e-01~~
	Intercept	6.676487	793.526556	9.999981e-01
rho_	Intercept	0.407519	1.503084	9.230174e-74

param	covariate	coef	exp(coef)	p
lambda_	ANNUAL SALARY	0.000028	1.000028	3.522389e-56
	CT	-1.690473	0.184432	~~9.999995e-01~~
	Level	0.399794	1.491517	3.472677e-31
	NY	-1.999146	0.135451	~~9.999994e-01~~
	RI	-1.439487	0.237049	~~9.999996e-01~~
	Intercept	6.676487	793.526556	9.999981e-01
rho_	Intercept	0.407519	1.503084	9.230174e-74

The p-value indicates statistical significance. **Annual Salary and Seniority level are statistically significant**. The geographical location are not statistically significant in impacting the survival function. Before we dive deeper into other property, we will rebuild the model with only significant features. The interaction between the two factors(annual salary and seniority level) is not statistically significant either and thus will not be put into the model again.

param	covariate	coef	exp(coef)	se(coef)
lambda_	Intercept	5.232678	187.293770	0.083534
	Level	0.496990	1.643766	0.034318
	WAGE	0.019305	1.019492	0.001526
rho_	Intercept	0.388904	1.475363	0.022315

After fitting we found that with one-unit increase in Level, the average tenure duration change by a factor of 1.64, meaning staying for a longer time, 64% longer. As for wage, the average tenure duration change by a factor of 1.019, namely, 1.9% longer if wage increases by 1 unit increase(1k).

Method to look at the attrition of this role: Survival Analysis

Parametric model to quantify impact of covariate: **Cox Proportional Hazards Model**

Hazard function describe the probability that event happens at some time given the subject survive up to that time. Before we run the Cox model we have to run tests to check on the assumptions. The tests outcome show it meets the requirements. The following is the model summary, after removing non-significant factors:

covariate	coef	exp(coef)	p
Level	-0.720446	0.486535	5.483017e-40
WAGE	-0.028770	0.971639	3.752341e-36

covariate	coef	exp(coef)	p
Level	-0.720446	0.486535	5.483017e-40
WAGE	-0.028770	0.971639	3.752341e-36

If one unit in level increases, hazard(dropout) decrease by 51%. This indicates the importance of career advancement program when it comes to retention. In some special years like 2020-2021, when we are facing potential pitfall of retention of this position, extra effort should be devoted to expedite advancement process in order to have a better retention result. As for wage, one-unit(1k) increase will bring down 2.8% hazard rate. Compensation team could use this metric reevaluate current payment program.These are compared with the baseline hazards.

Prediction with the model

What we can do with the prediction: **we could forecast the business employee count based on expected survival rates.**

Predicting median remaining times for current employees

E.g Number 1051 employee is most likely to leave the soonest, with an median estimation of 129.0 days left at the organization

Predicting survival function conditional on existing durations

E.g The chance for employee 1051 to survive longer than 124-133 days after his current situation is estimated to be 50 %

```
n1051[n1051==50]
```

```
124.0    50.0
127.0    50.0
128.0    50.0
129.0    50.0
130.0    50.0
131.0    50.0
133.0    50.0
```

Results in short:

1. Among all the selected features, tenure and annual salary play significant roles in the models
2. The longer this type of employee stays, the higher level they could be and vice versa.
3. The average tenure duration would be 1.9% longer if wage increases by 1 unit increase(1k). Meanwhile, one-unit(1k) increase in wage will bring down 2.8% hazard(dropout) rate.
4. The model is more likely for estimation of headcount instead of flagging employees on the brink of the dropout (For real-life practice's sake)
5. The impact of **annual salary** on the probability to stay longer is small. Instead, seniority does more work in sustaining employees.(Important insight for advancement program)