

# EMPLOYEE RETENTION

By Janny Velazquez

# OBJECTIVE

The objective of this project is to create, analyze, and understand insights about employee retention

# RESEARCH QUESTIONS

- What specific factors have a significant effect on employee retention?
- What are some of the characteristics of employees that decide to leave a job?
- How can we help predict predict the length of time that employyes will stay at a job?

# **DATASET**

**The dataset used for this project was obtained from Kaggle and its publicly available.**

**<https://www.kaggle.com/datasets/gummulasrikanth/hr-employee-retention>**

**The dataset includes information from 15,000 employees, and has a total of 10 variables.  
The dataset did not have any missing values.**

# VARIABLES

**satisfaction\_level  
last\_evaluation  
number\_project  
average\_montly\_hours  
time\_spend\_company  
work\_accident  
left  
promotion\_last\_5years  
salary\_num  
department\_num**

# LOGISTIC REGRESSION

As our first statistical analysis method, logistic regression, which is tailored for predicting binary outcomes, such as whether an employee will leave the company or not.

# FULL MODEL

left~satisfaction\_level + last\_evaluation+ number\_project +  
average\_monthly\_hours + + time\_spend\_company +  
Work\_accident + promotion\_last\_5years + salary\_num +  
+ department\_num

AIC: 3840.8

Number of Fisher Scoring iterations: 6

Accuracy: 0.7932874

# REDUCED

left ~ satisfaction\_level + average\_monthly\_hours +  
promotion\_last\_5years + salary\_num + department\_num

AIC: 4075.2

Number of Fisher Scoring iterations: 5

0.7783952

# FULL MODEL W/ INTERACTIONS

```
left~satisfaction_level + last_evaluation+ number_project +  
average_monthly_hours + + time_spend_company +  
Work_accident + promotion_last_5years + salary_num +  
+ department_num
```

AIC: 1831.4

Number of Fisher Scoring iterations: 20

Accuracy: 0.9466548



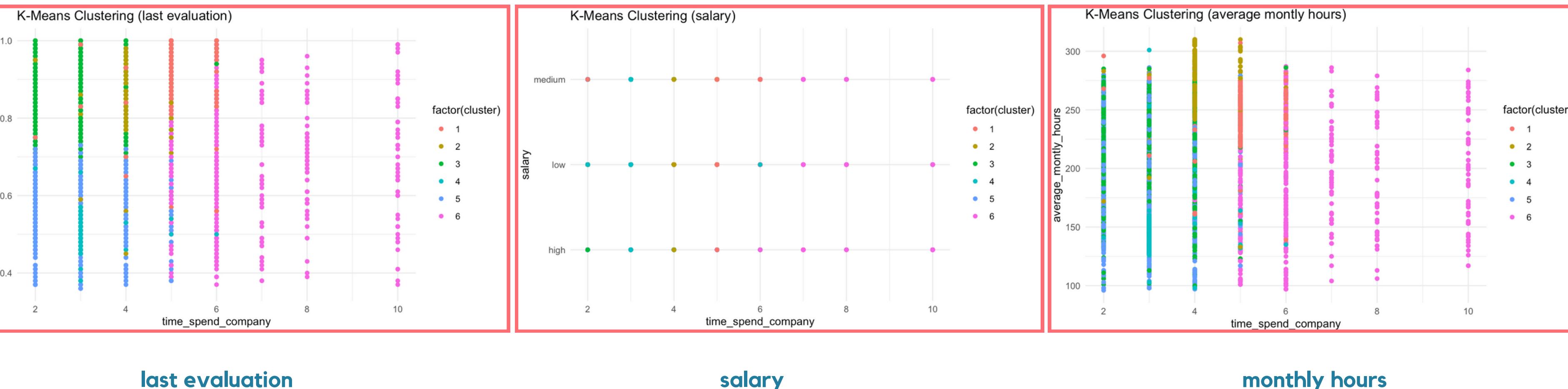
After performing logistic regression analysis on our dataset, it is evident that the full model with interactions consistently outperforms the reduced model, demonstrating a remarkable accuracy of 0.947 in predicting the likelihood of employee retention. These findings highlight the significance of considering variable interactions in understanding employee retention.

# K-MEAN CLUSTERING

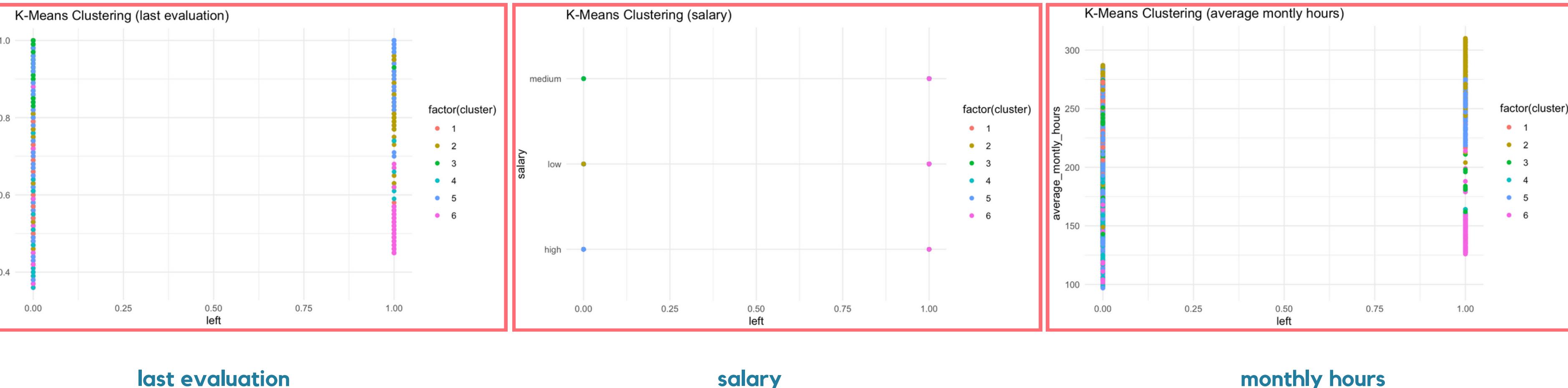
The second method used for this project is k-means clustering. Applied to employee retention, k-means clustering allows us to uncover natural segments and behaviors among employees. This unsupervised approach provides a unique perspective, revealing hidden structures that can enrich our understanding of employee dynamics.

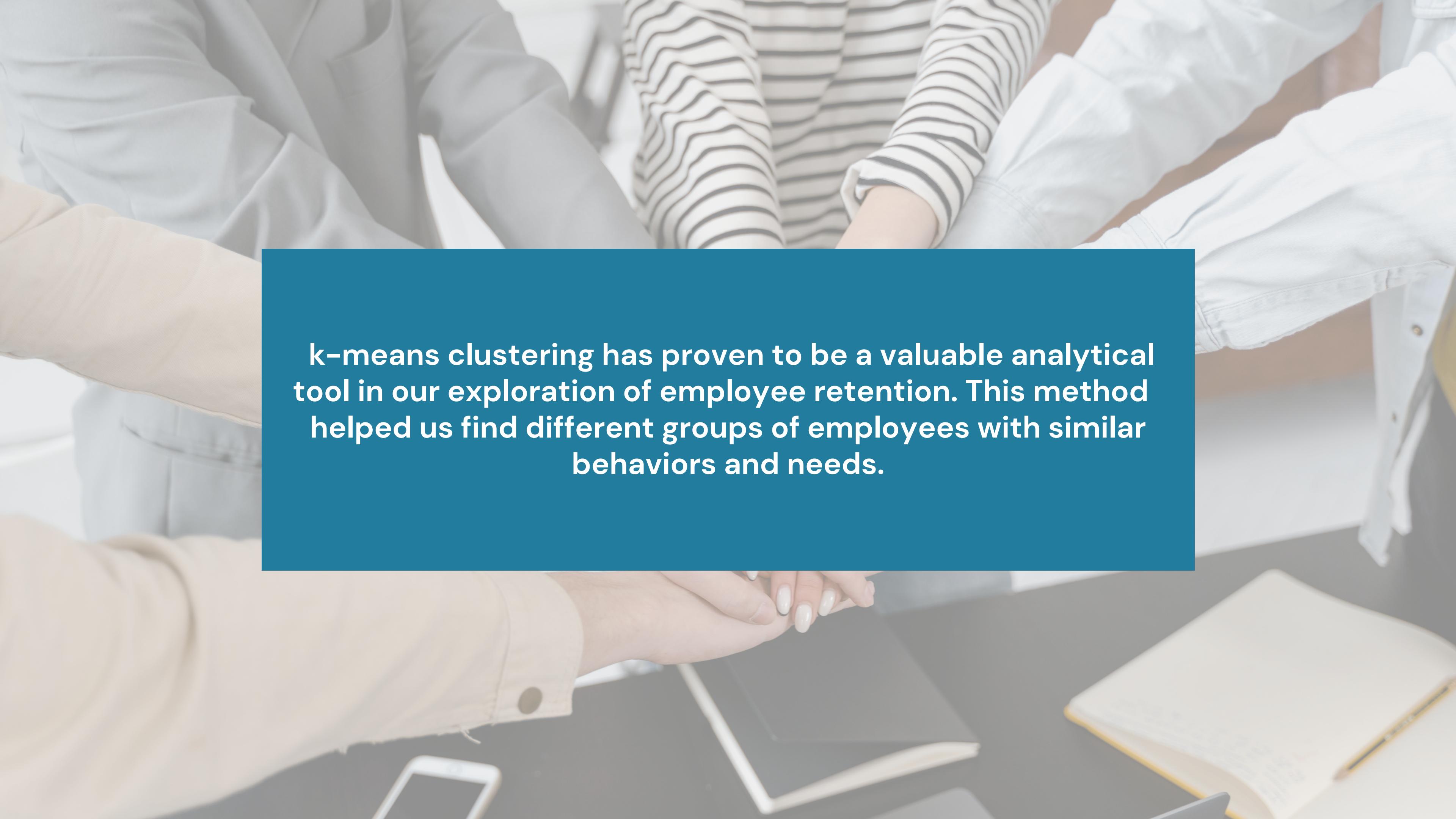
level of clustering is 6 based on elbow method

# CLUSTERING BASED ON TIME SPENT AT COMPANY



# CLUSTERING BASED ON WHETHER THE EMPLOYEE LEFT OR NOT



A photograph showing a group of people from the waist up, sitting around a table. They appear to be in a meeting or discussion. One person in the foreground is holding a tablet. A blue rectangular box is overlaid on the image, containing the text.

**k-means clustering has proven to be a valuable analytical tool in our exploration of employee retention. This method helped us find different groups of employees with similar behaviors and needs.**

# SURVIVAL ANALYSIS

As our 4 method we use Survival Analysis.  
Unlike traditional analyses, survival analysis  
doesn't just tell us 'who' will leave, but 'when',  
showing patterns in the time-to-event data.

Event:

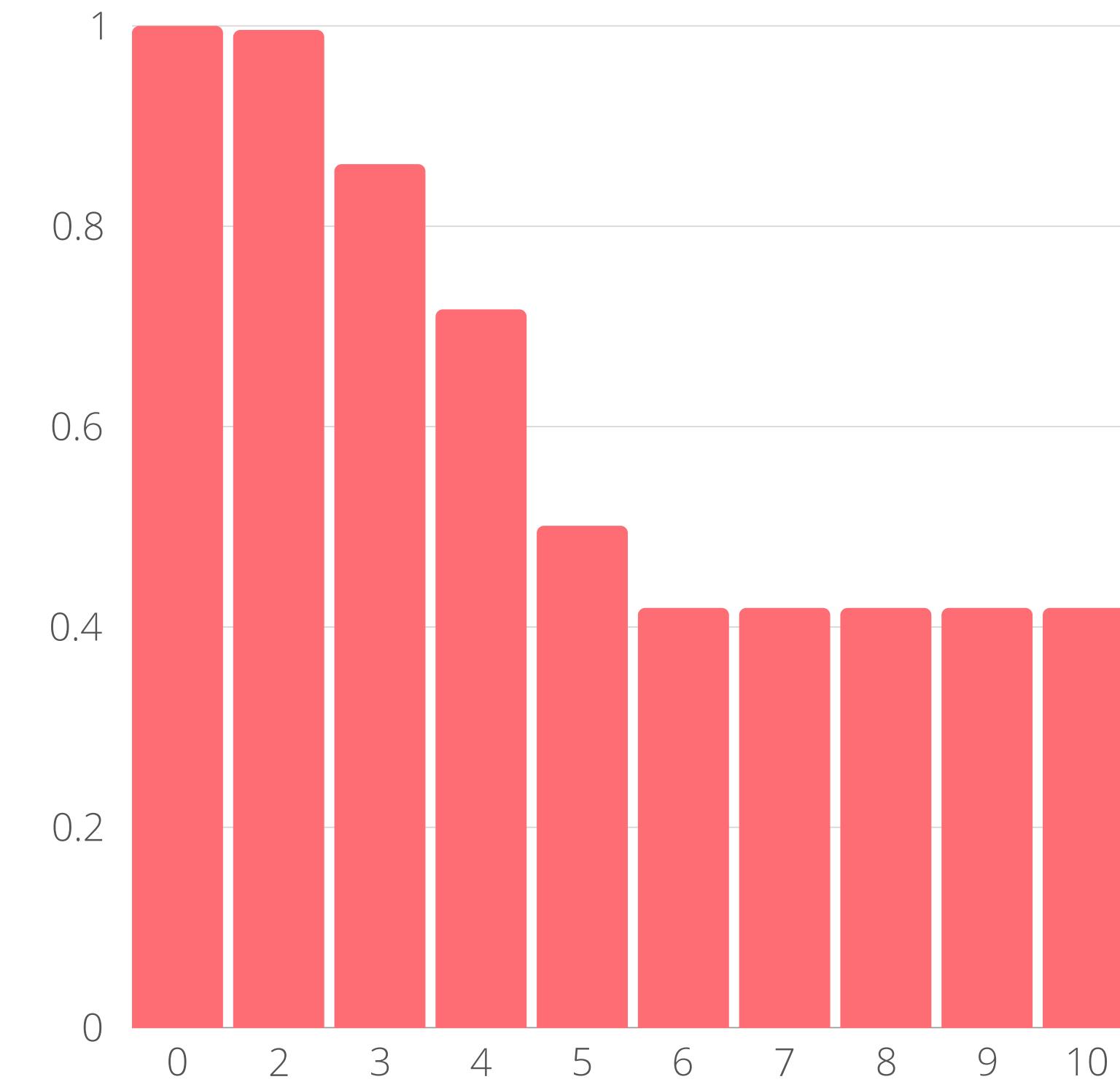
The moment when an employee leaves the job. (Variable name: "left")

Survival Time:

The duration until the event occurs. (Variable name:  
"time\_spend\_company")

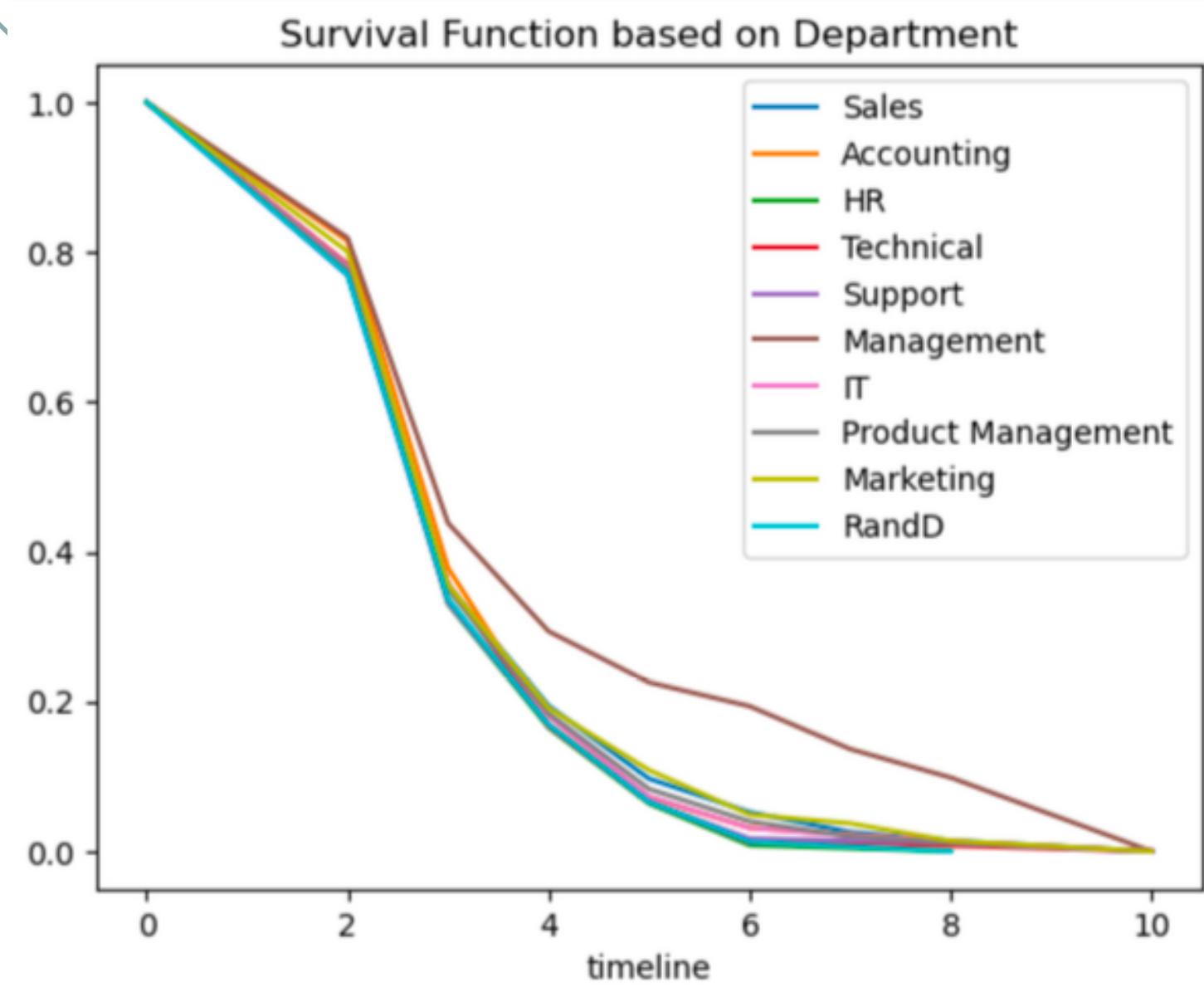
# PROBABILITY OF STAYING AT THE COMPANY BY YEAR

timeline	probability
0.0	1.000000
2.0	0.996466
3.0	0.862022
4.0	0.717594
5.0	0.500623
6.0	0.419008
7.0	0.419008
8.0	0.419008
10.0	0.419008



We can see that probability of an employee staying for longer than 2 years at the company is 99% however probability of surviving longer than 5 years drops to 50%, and probability of staying from 6 to 10 years remains 42%.

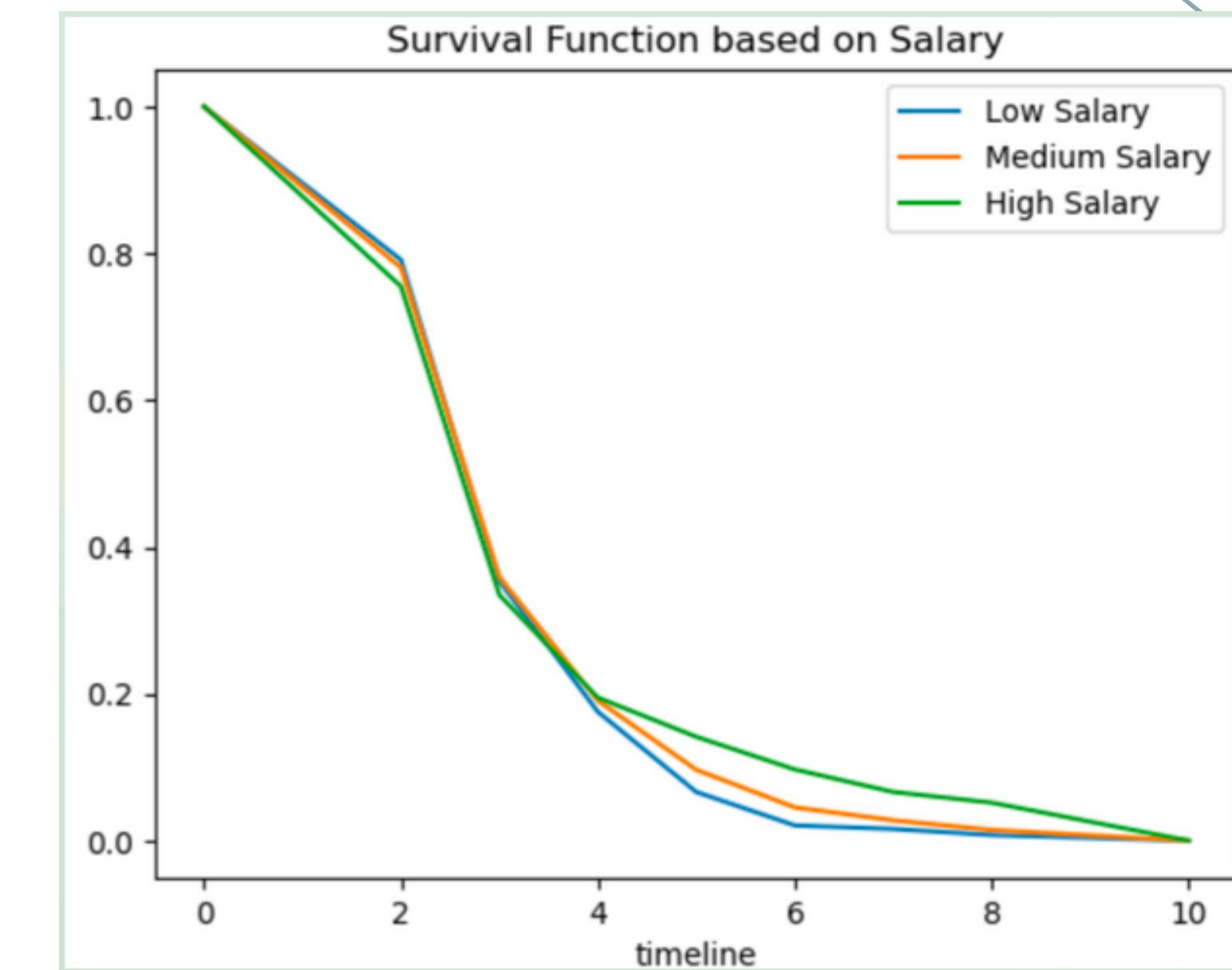
# SURVIVAL FUNCTION BASED ON DEPARTMENTS



Here can notice that employees who work in management have a significantly higher survival probability than the rest of the fields

# SURVIVAL FUNCTION BASED ON SALARY LEVELS

We can notice that employees with higher salary have a slightly higher survival probability, than the employees with low and medium salaries



A collage of various office scenes. In the top left, a woman with glasses and a white shirt is looking down at a laptop. In the bottom left, hands are shown working on a laptop keyboard. In the center, a person's hands are on a laptop keyboard. To the right, a person is working at a desk with a laptop, a red phone, and a clipboard. The background is a grid pattern.

This project looked at different ways to understand and improve employee retention. We used various tools to find out what factors influence employees staying or leaving, identified different employee groups, and even explored when people might leave.

**THANK YOU!**