**Visual Data Analysis.**

**Data:** Human Resources Analytics (*This dataset is simulated.*)

(<https://www.kaggle.com/ludobenistant/hr-analytics/downloads/HR_comma_sep.csv>)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **satisfaction\_level** | **last\_evaluation** | **number\_project** | **average\_montly\_hours** | **time\_spend\_company** | **Work\_accident** | **left** | **promotion\_last\_5years** | **sales** | **salary** |
| 0.38 | 0.53 | 2 | 157 | 3 | 0 | 1 | 0 | sales | low |
| 0.8 | 0.86 | 5 | 262 | 6 | 0 | 1 | 0 | sales | medium |
| 0.11 | 0.88 | 7 | 272 | 4 | 0 | 1 | 0 | sales | medium |
| 0.72 | 0.87 | 5 | 223 | 5 | 0 | 1 | 0 | sales | low |
| 0.37 | 0.52 | 2 | 159 | 3 | 0 | 1 | 0 | sales | low |

Fields in the dataset include answers on such questions:

* How would you rate your level of satisfaction on the current position?
* How much time has passed since the last promotion?
* How many projects you have completed on the current position?
* How many hours do you work per week in average?
* How many months have you been working in the current company?
* Have you had an accident at work?
* Have you been promoted in the last 5 years?
* Choose your department.
* Choose your salary level.

**Data types:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Column name** | **Type** |  | **Column name** | **Type** |
| satisfaction\_level | float64 |  | Work\_accident | int64 |
| last\_evaluation | float64 |  | left | int64 |
| number\_project | int64 |  | promotion\_last\_5years | int64 |
| average\_montly\_hours | int64 |  | sales | object |
| time\_spend\_company | int64 |  | salary | object |

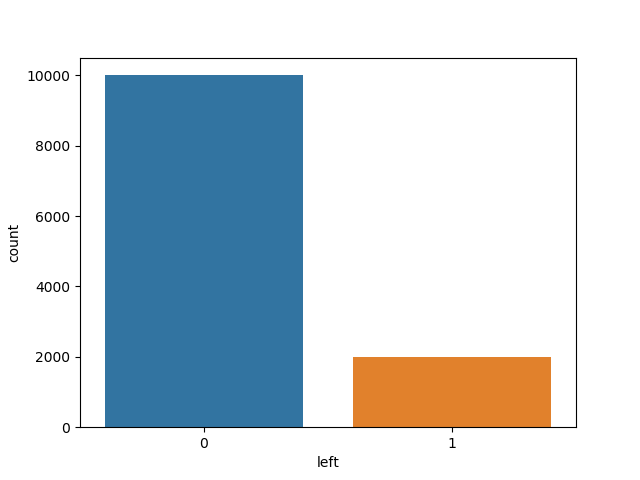
The data column "left", is target variable. It is binary: 1 indicates that the employee left the company, and 0 indicates that the employee remained in the company.

**The exact numerical statistics of the distribution**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **satisfaction\_level** | **last\_evaluation** | **number\_project** | **average\_montly\_hours** | **time\_spend\_company** |
| **count** | 11991.000000 | 11991.000000 | 11991.000000 | 11991.000000 | 11991.000000 |
| **mean** | 0.629658 | 0.716683 | 3.802852 | 200.473522 | 3.364857 |
| **std** | 0.241070 | 0.168343 | 1.163238 | 48.727813 | 1.330240 |
| **min** | 0.090000 | 0.360000 | 2.000000 | 96.000000 | 2.000000 |
| **25%** | 0.480000 | 0.570000 | 3.000000 | 157.000000 | 3.000000 |
| **50%** | 0.660000 | 0.720000 | 4.000000 | 200.000000 | 3.000000 |
| **75%** | 0.820000 | 0.860000 | 5.000000 | 243.000000 | 4.000000 |
| **max** | 1.000000 | 1.000000 | 7.000000 | 310.000000 | 10.000000 |

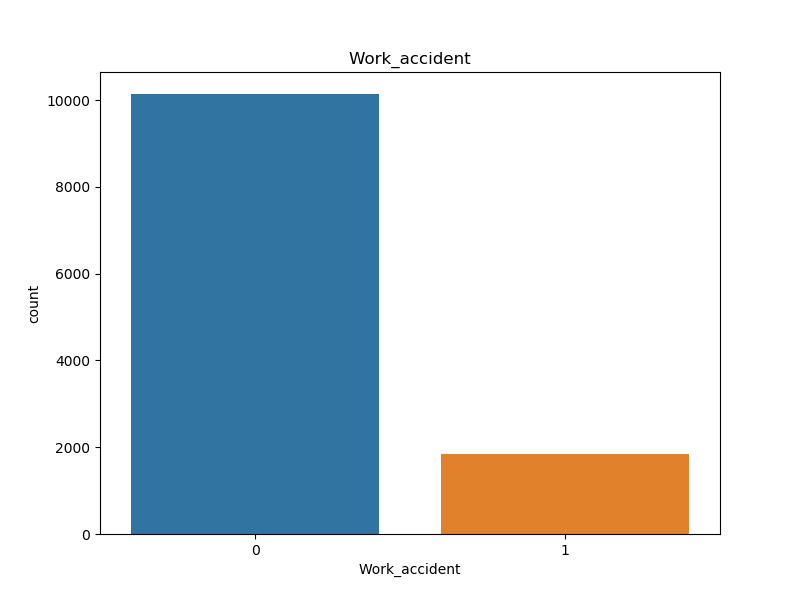
1. **Univariate visualization**

*Distribution of all features:*

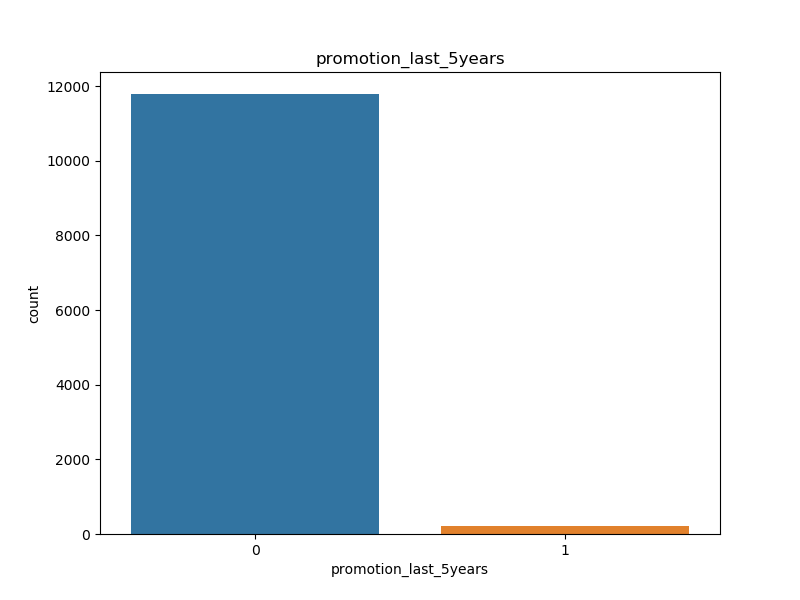


The chart above vividly illustrates the imbalance in our target variable.

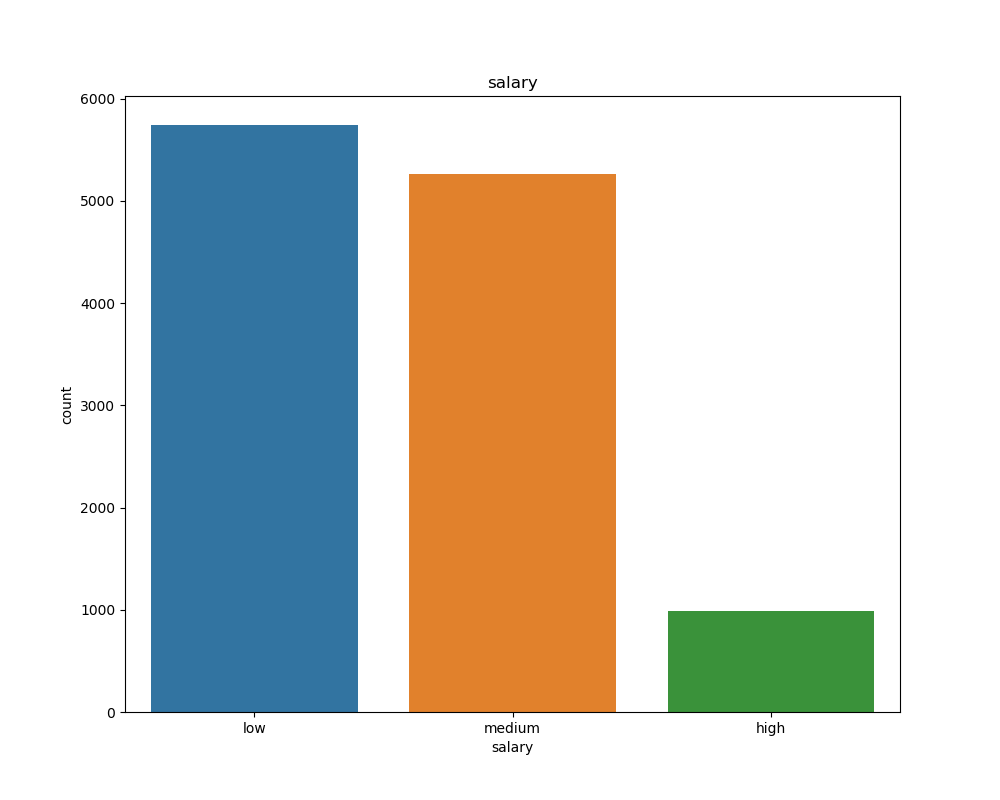
Most of the employees, which took part on the survey, remained in the company.



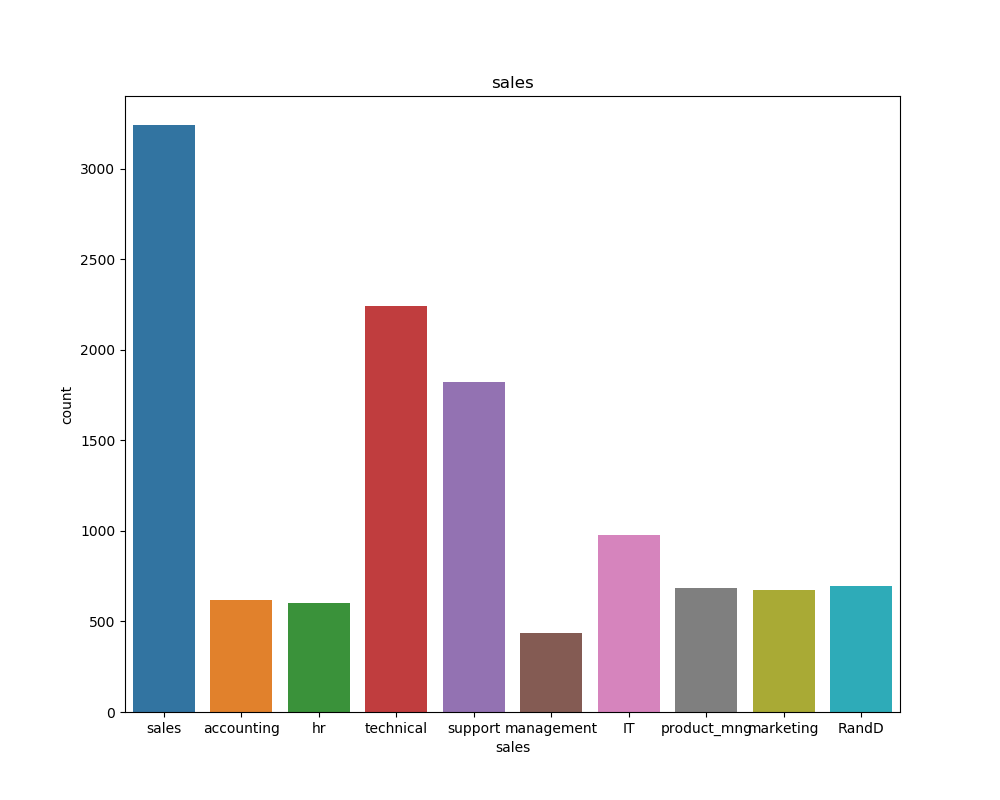
Usually, employees did not have accident at work.



Most employees have not been promoted in the past 5 years.

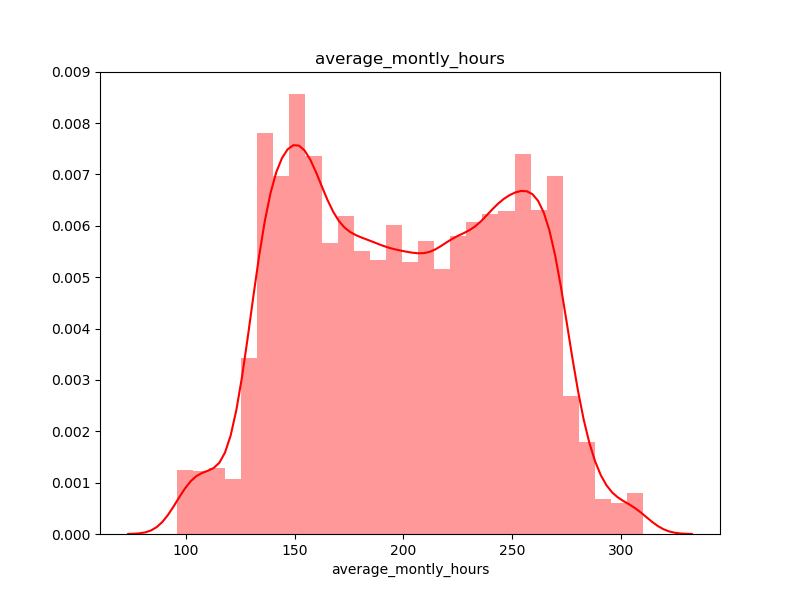


Not many employees have a high salary. Usually, it is small or medium.



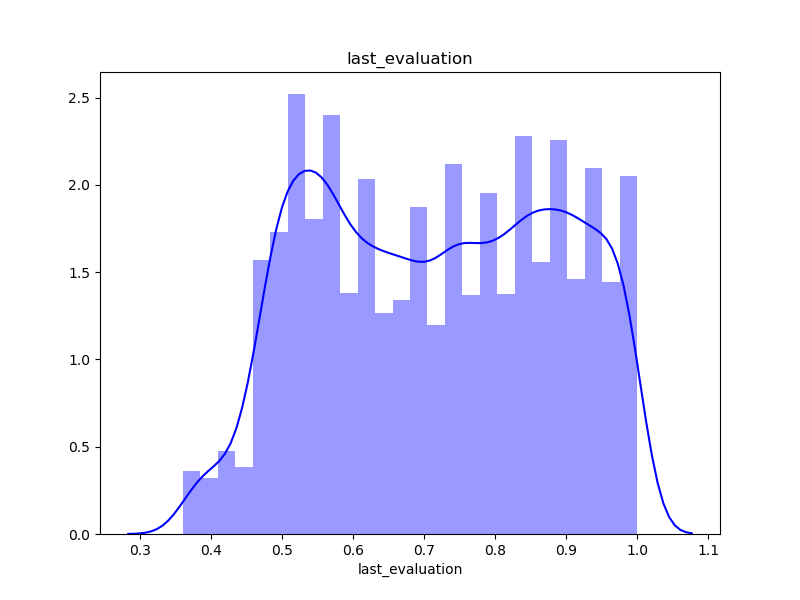
Most of employees work in "sales", "technical" and "support" departments.

Least employees work in “management” department.



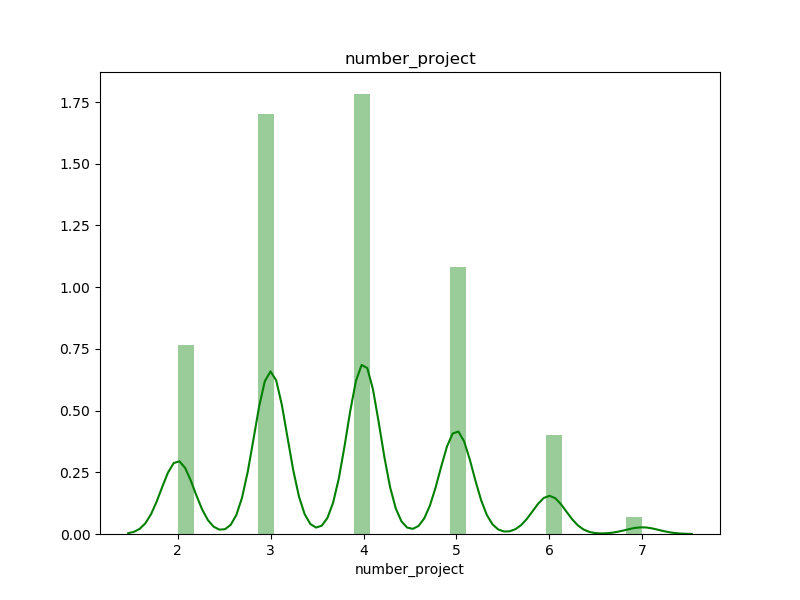
The graph above has two peaks: about 150 and 250 working hours. And the average working time per month for most employees is within these peaks.

( **!** The height of the histogram bars here is normed and shows the density rather than the number of examples in each bin.)



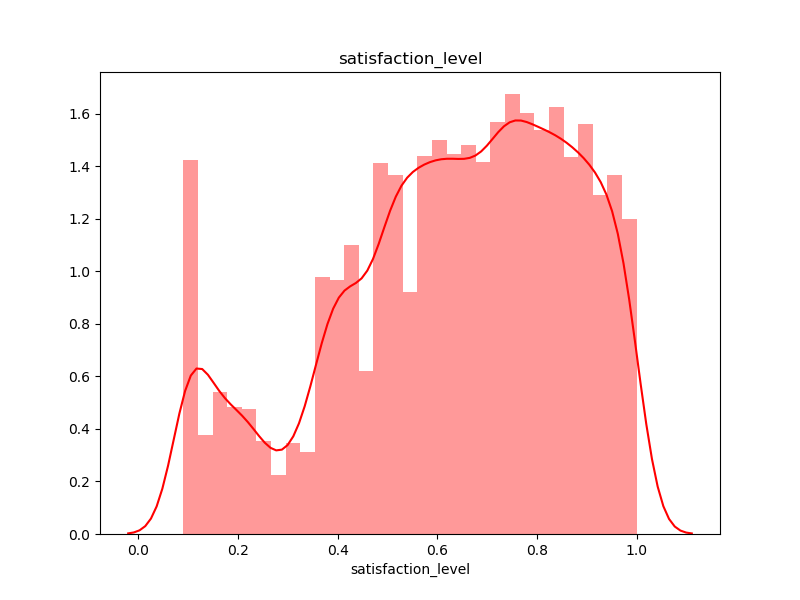
On average, from 6 months to 1 year has passed since the last promotion in most employees.

( **!** The height of the histogram bars here is normed and shows the density rather than the number of examples in each bin.)



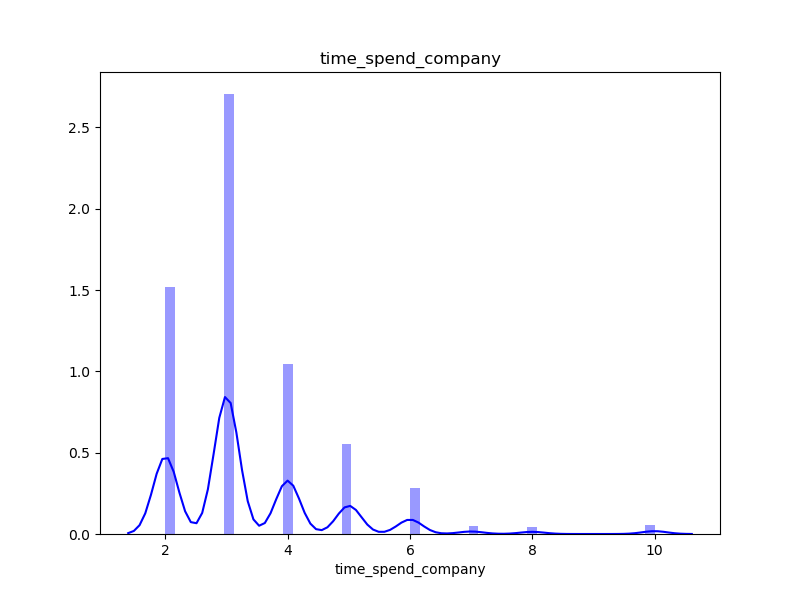
All in all, one employee worked on 2-3 projects.

( **!** The height of the histogram bars here is normed and shows the density rather than the number of examples in each bin.)



In general, level of satisfaction on the current position is greater equal 40%. It means that most of people like their work.

( **!** The height of the histogram bars here is normed and shows the density rather than the number of examples in each bin.)

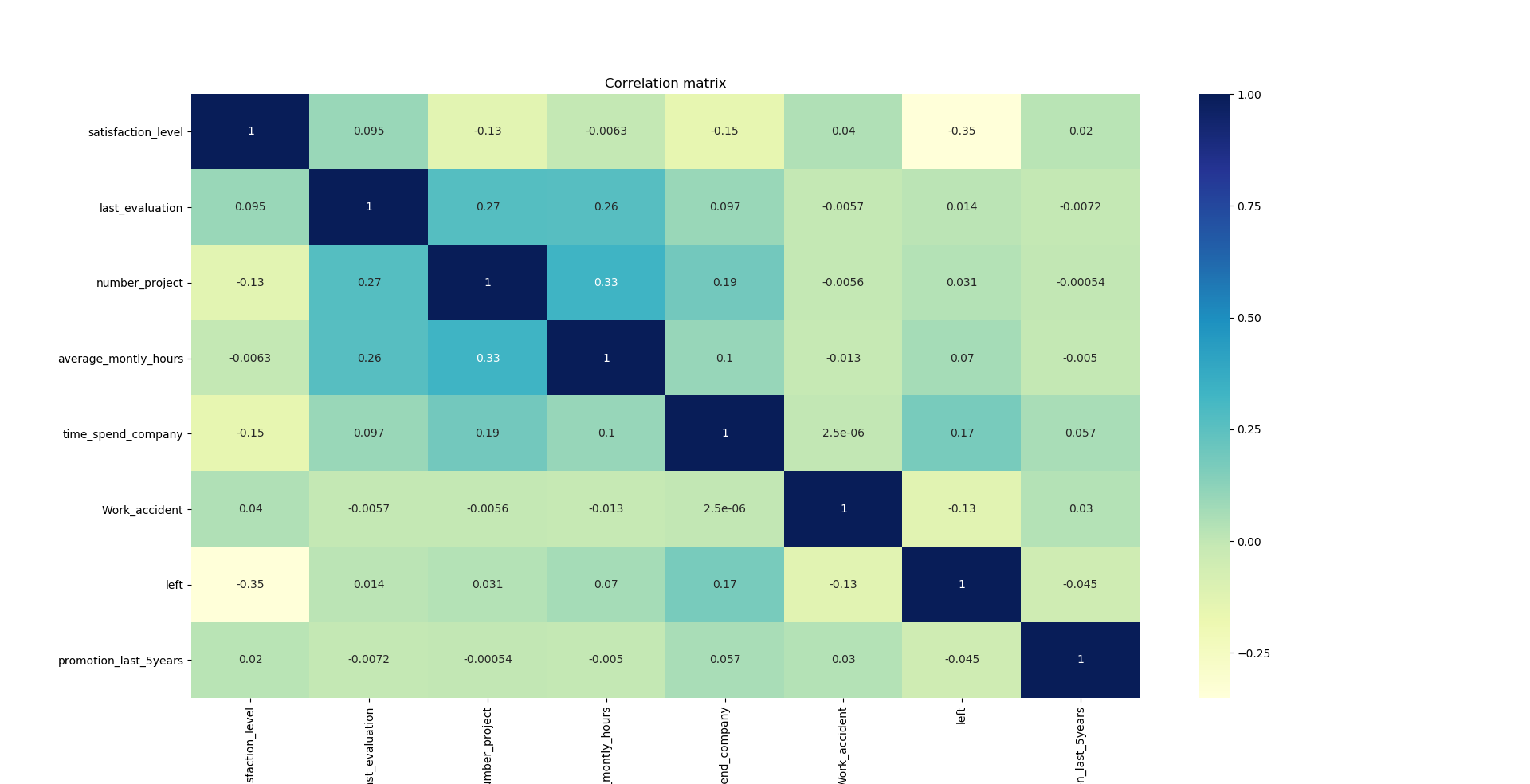


Mostly employees work in the company for 2-4 years. Only a small part of them works more than 6 years.

( **!** The height of the histogram bars here is normed and shows the density rather than the number of examples in each bin.)

1. **Multivariate visualization**

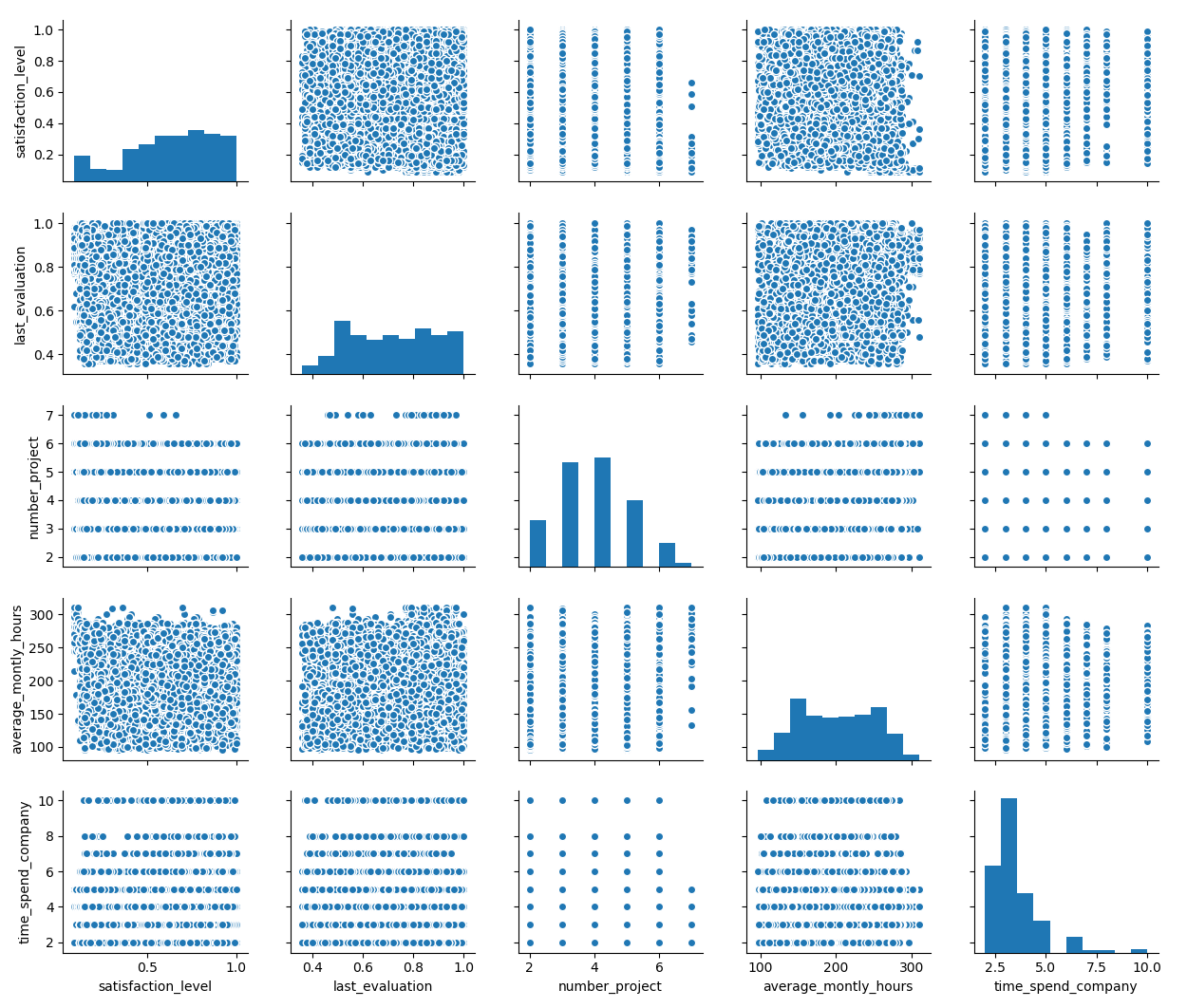
* *Correlations among the numerical features in dataset.*



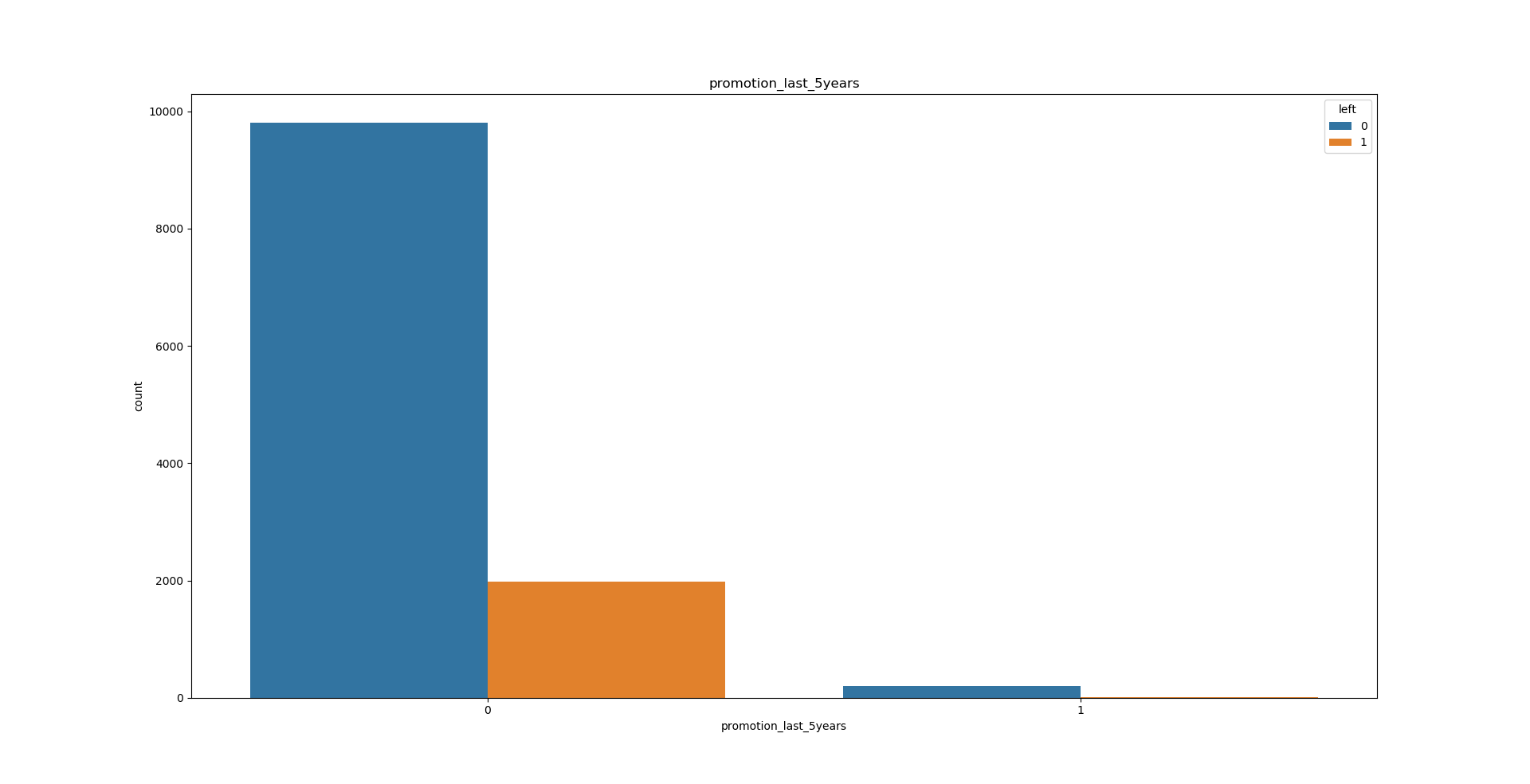
There is not strong correlation between data, so they are linearly independent.

* *Scatterplot matrix.*

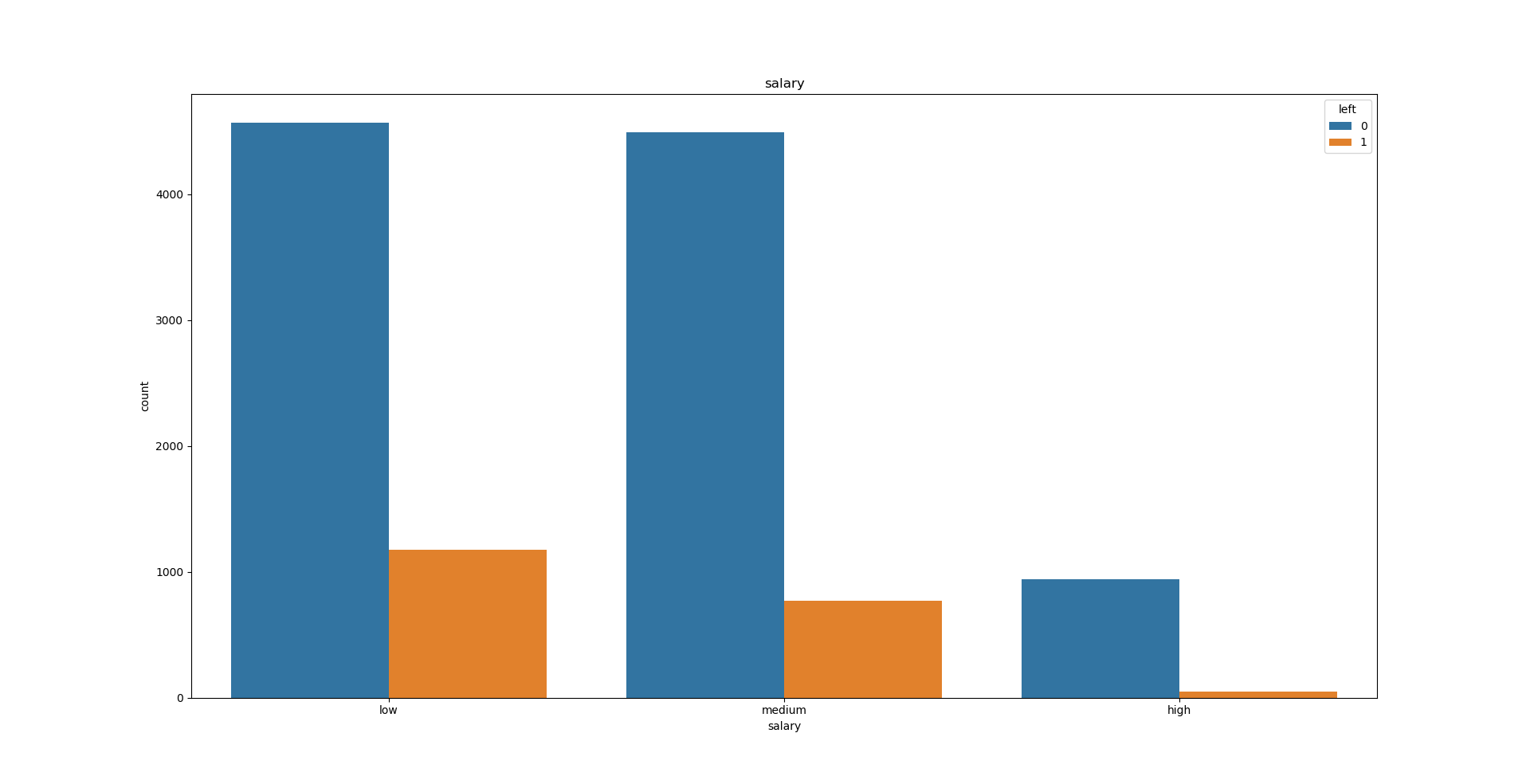
Scatterplot matrix such that the its diagonal contains the distributions of the corresponding variables, and the scatter plots for each pair of variables fill the rest of the matrix.



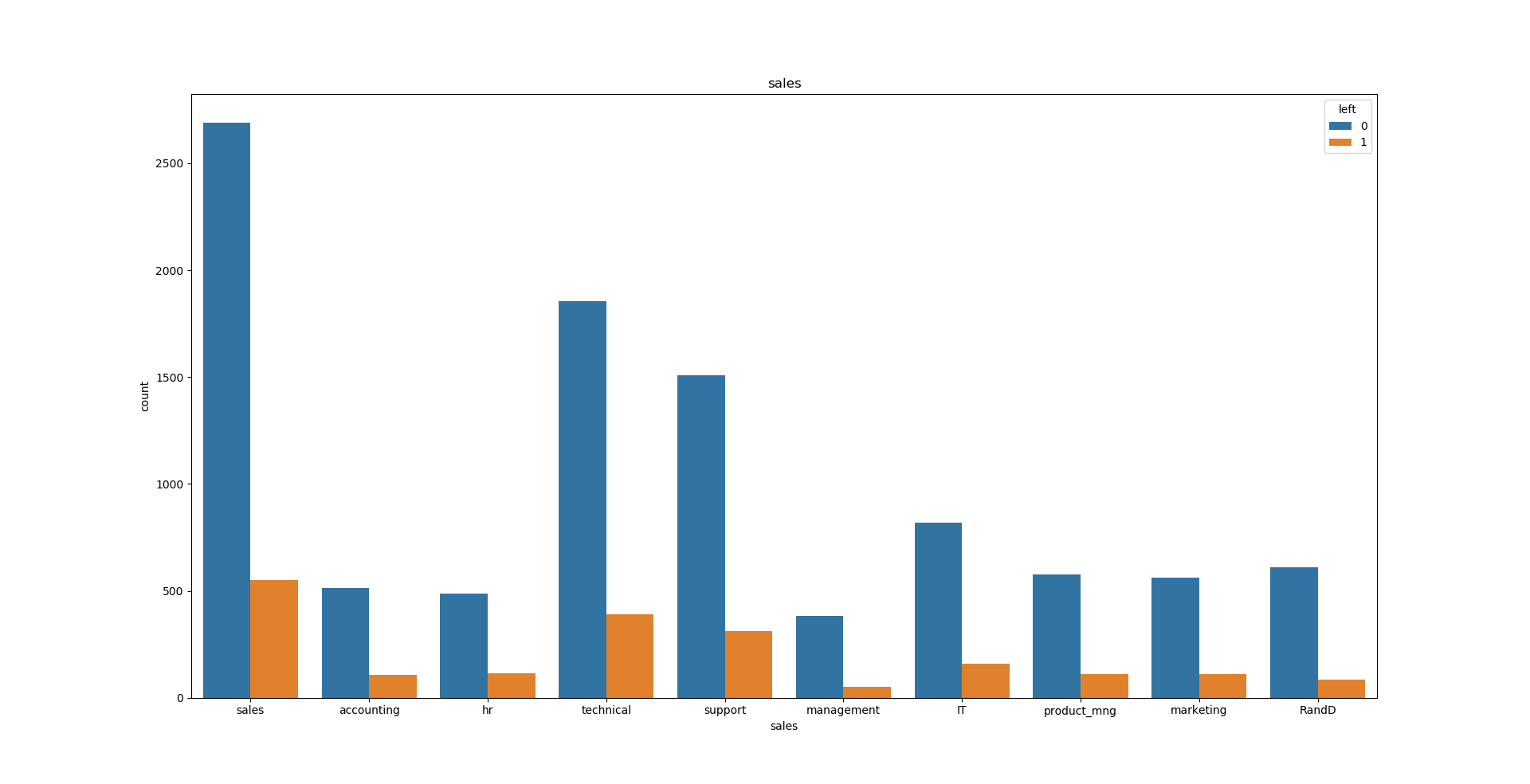
* *How the input variables are related to the target variable "left".*



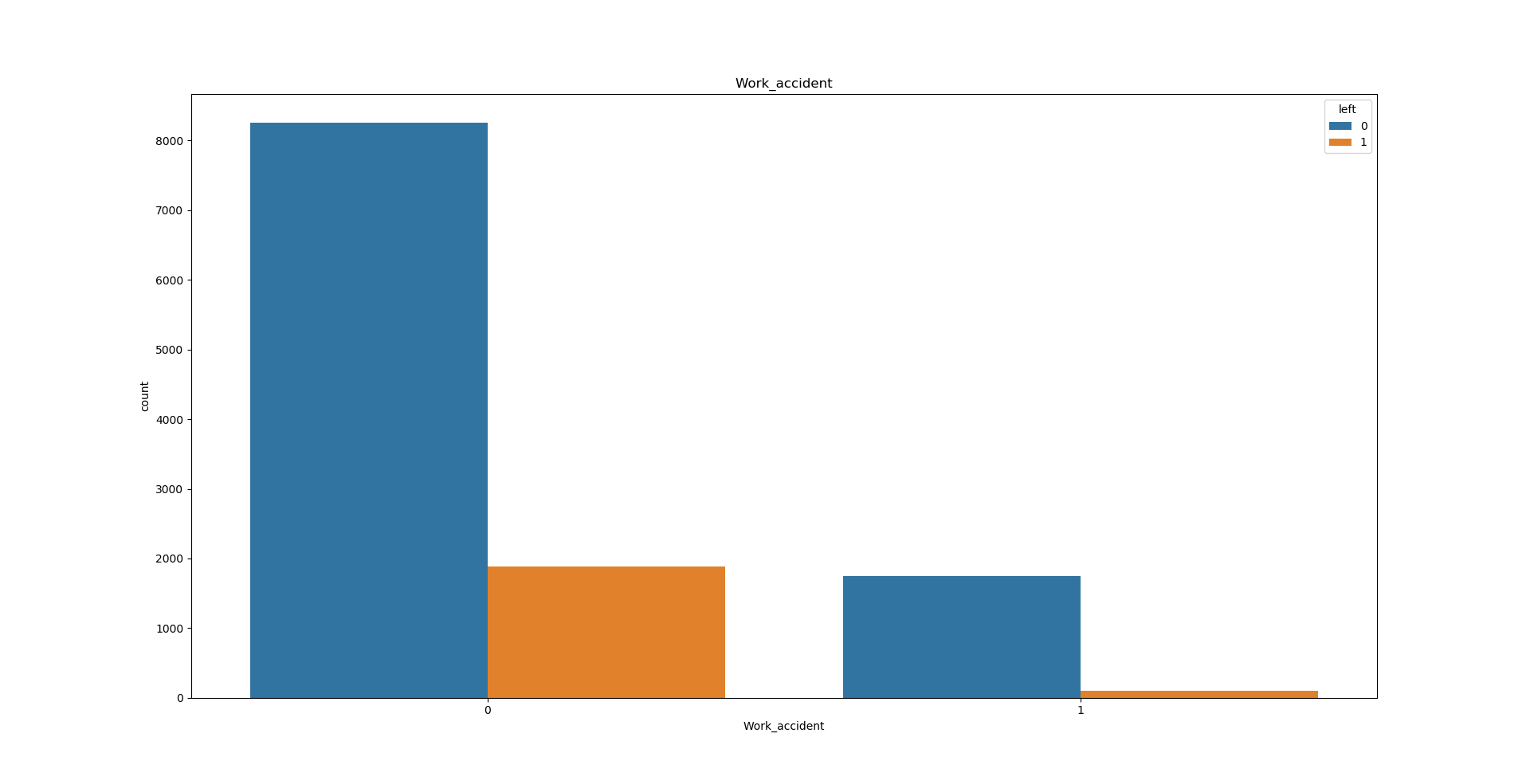
Only employees which have not been promoted in the past 5 years, left the company.



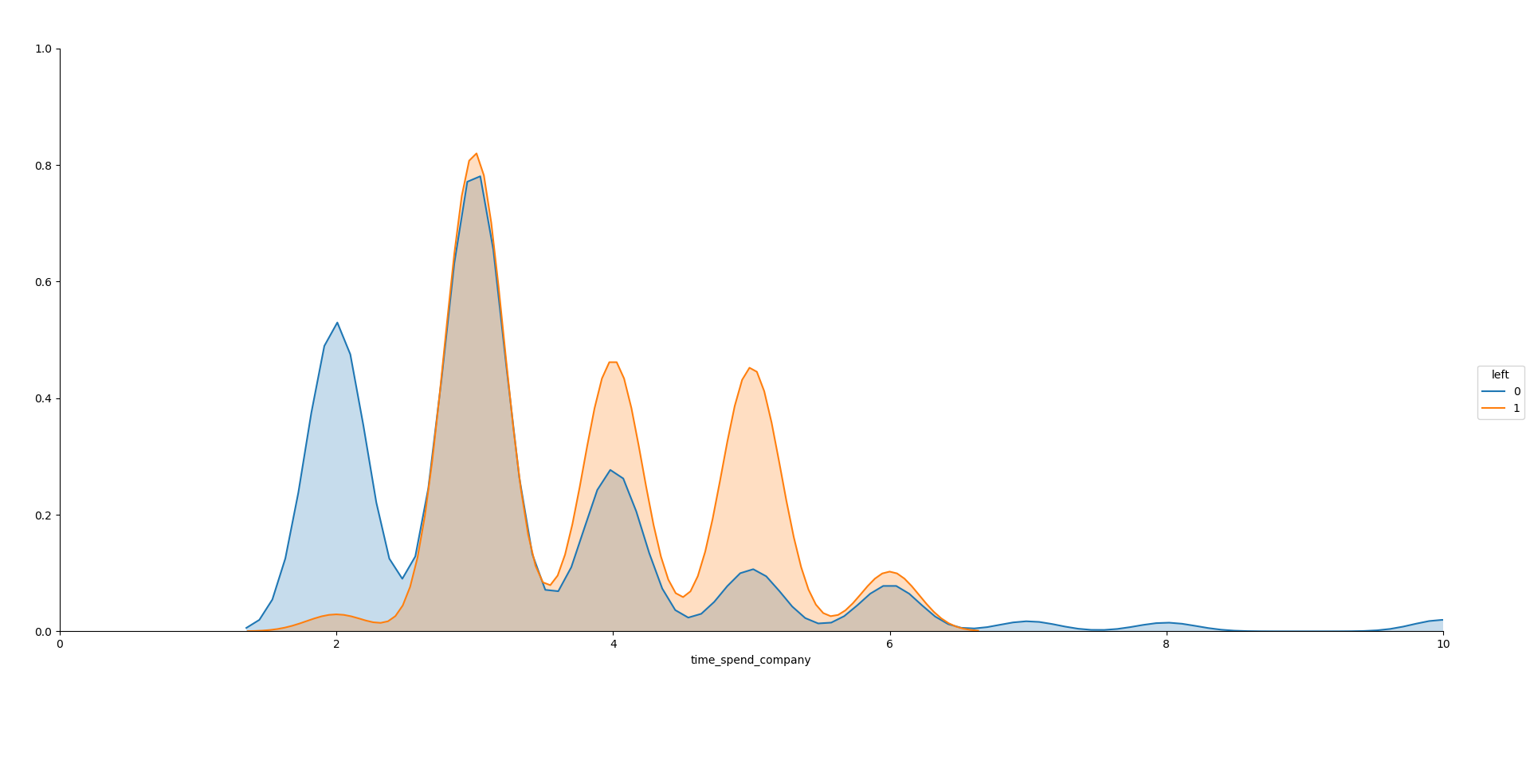
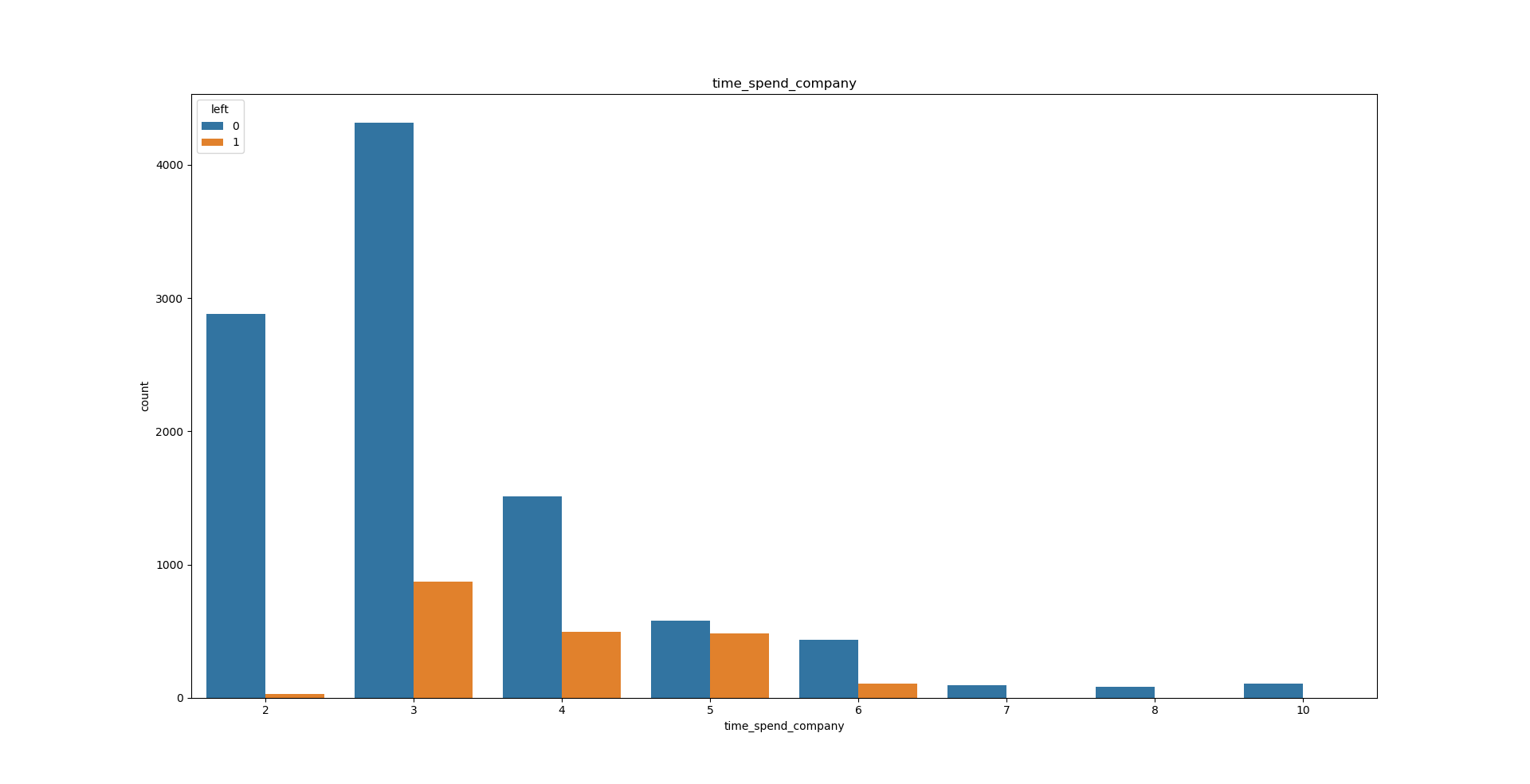
When employees have low and average wages, they usually leave the company.

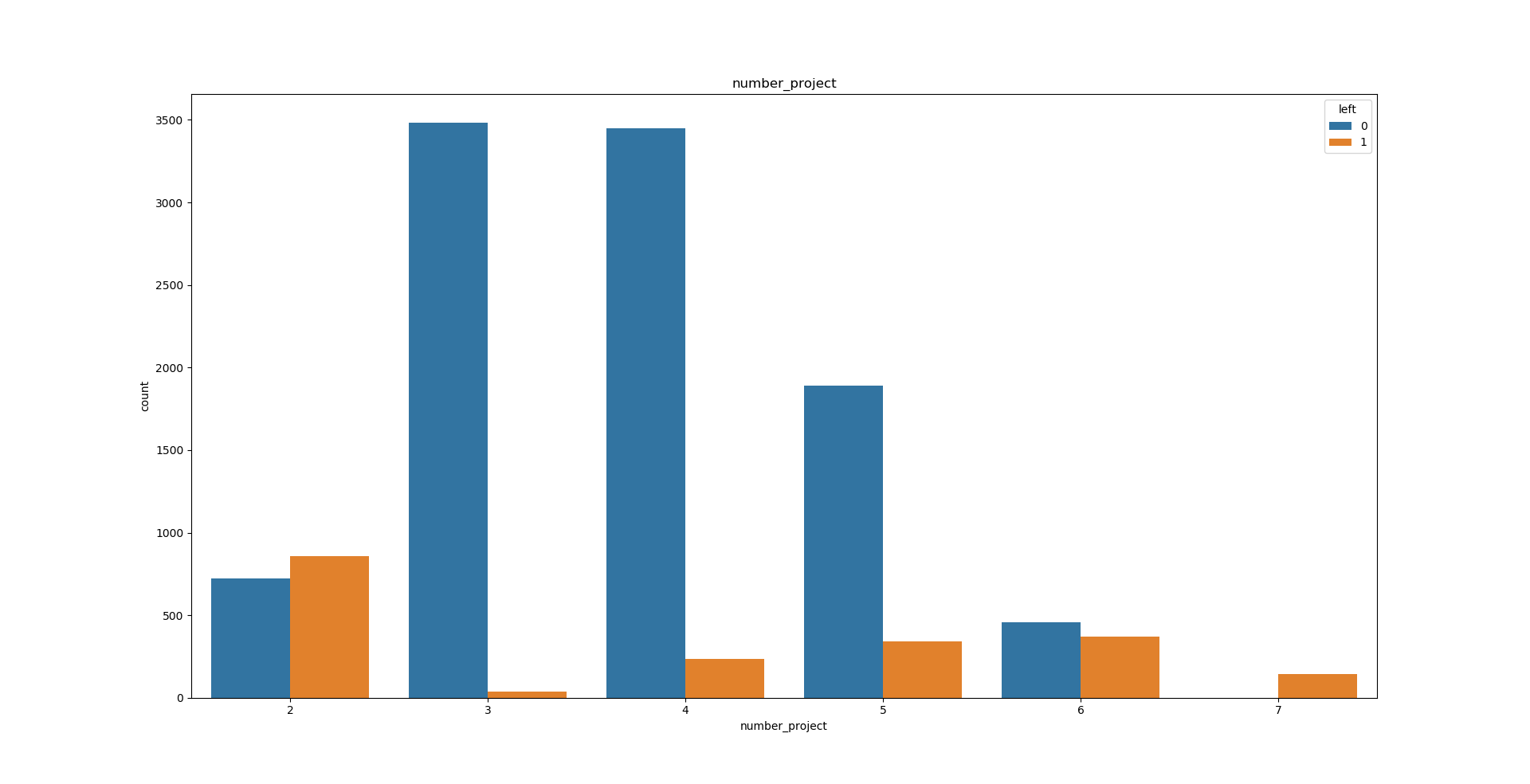


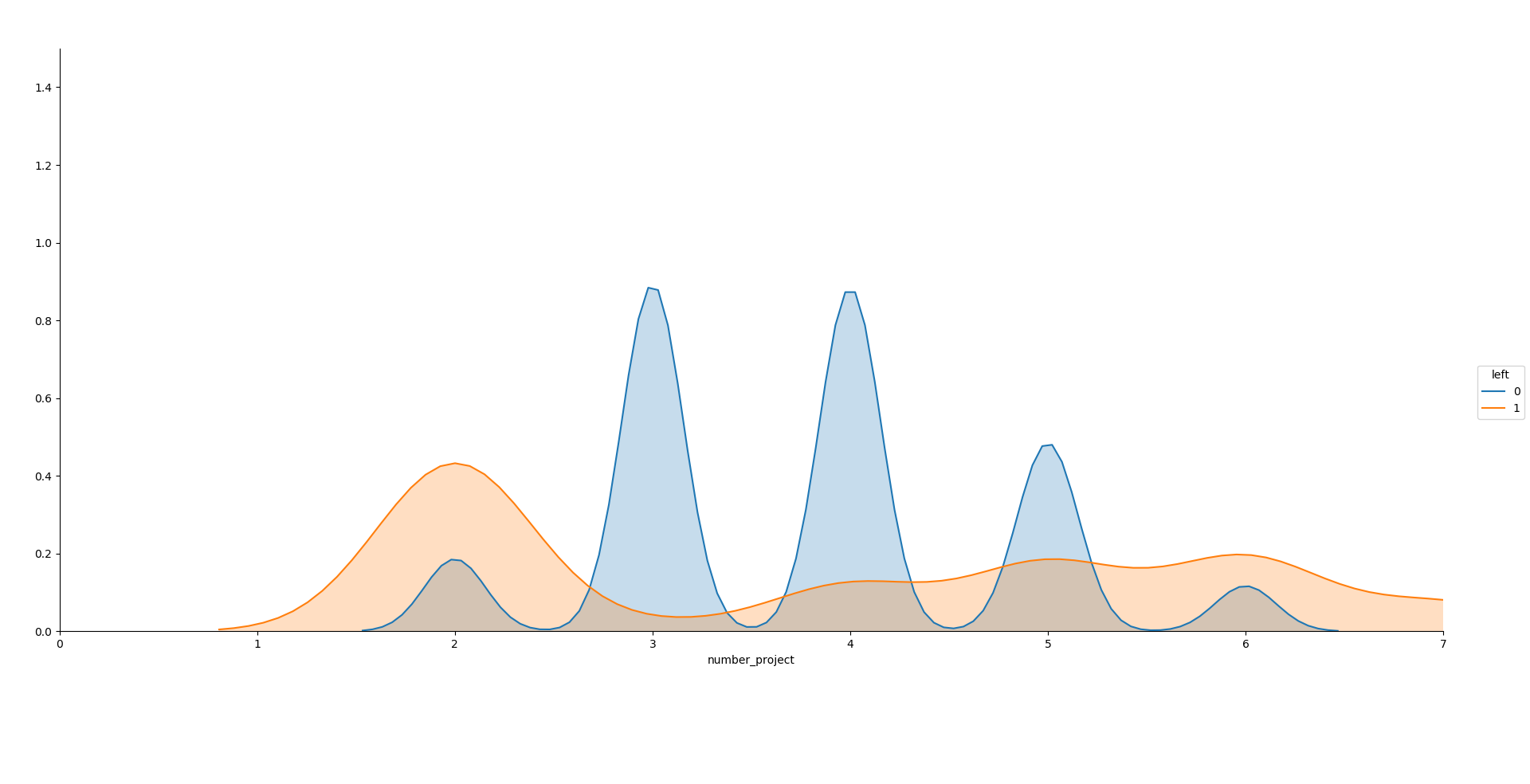
On average 1/6 of employees leave their department.

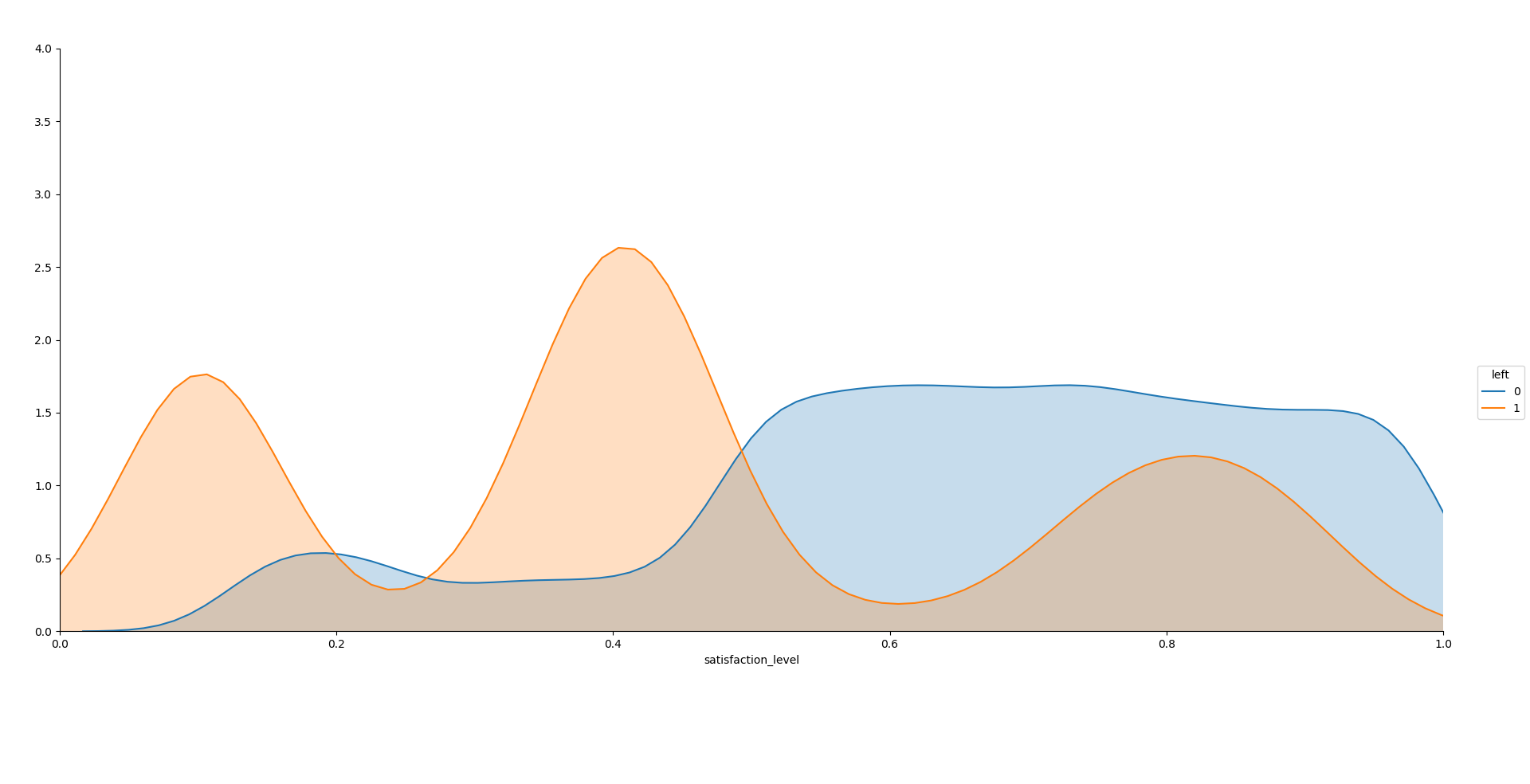


Workers are released in both cases (had accident at work or not).

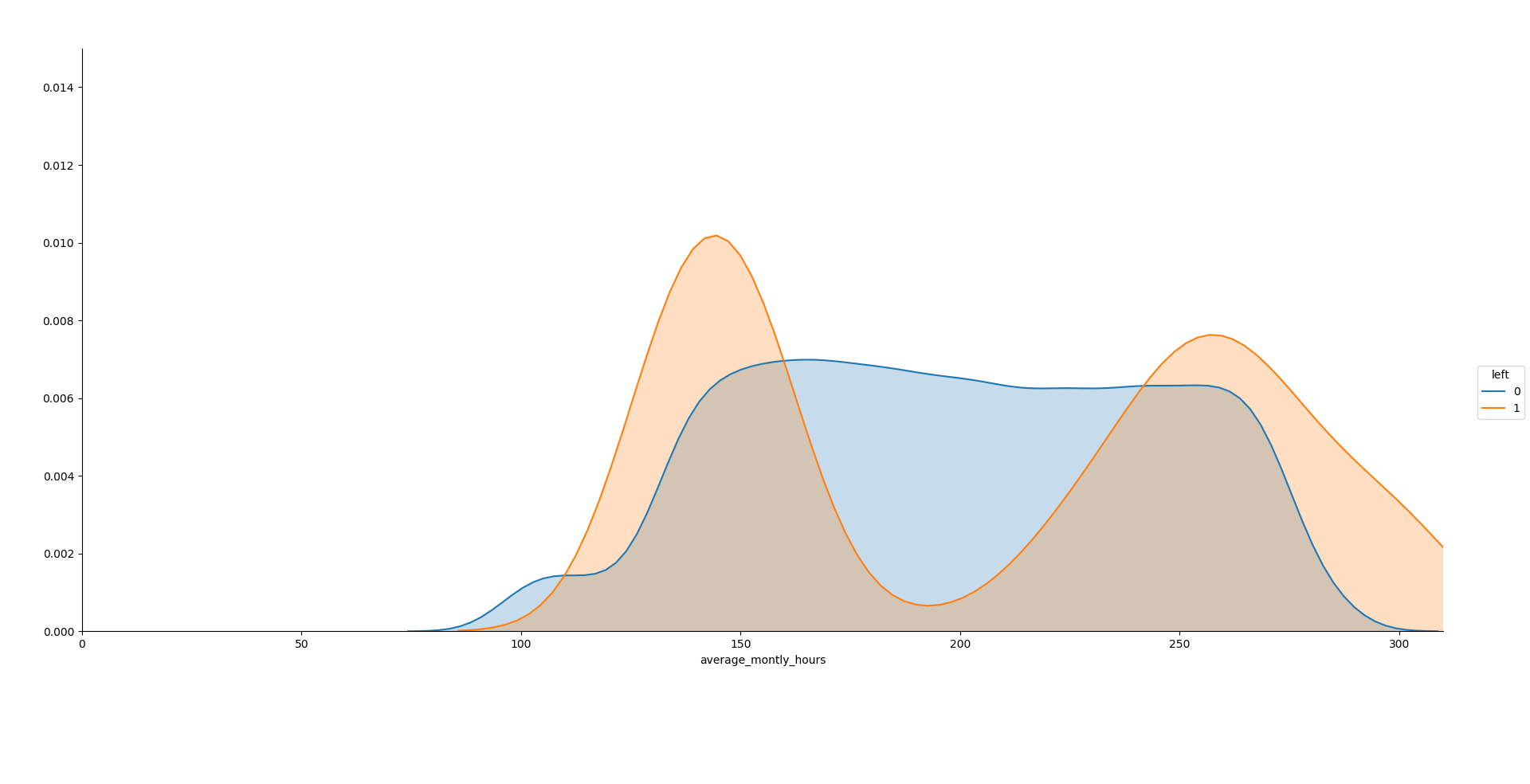
Here, we see that the probability that a person leaves, will grow while working in the company. But if they work 7 years or more then the probability that they will leave the company is very small.



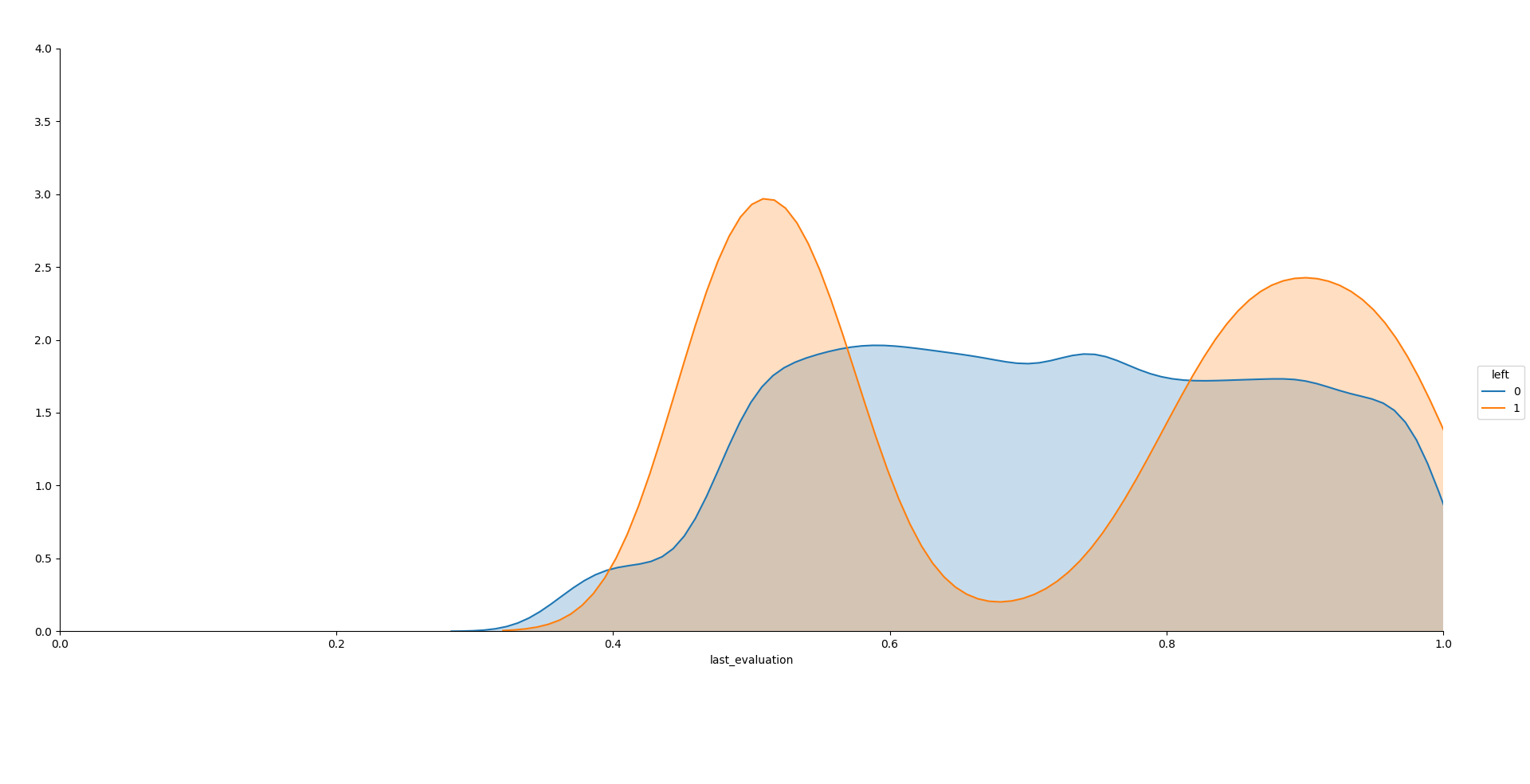
Clear regularities are not visible. But after completing 4 projects in a current position, the possibility that the employee will leave the company is increasing.



From this distribution, we see that people often leave the company when the level of satisfaction is small. At the same time, some people leave when the level is high.



Employees who work around 140 or 260+ hours a month are usually dismissed.



From the distribution of last\_evaluation, we can see that employees are going to another place if they are not promoted for a long time.

1. **Summary**

* If the level of satisfaction of the current position is lower than the average (<0.5), the employee will leave the company.
* When people have a lot of working hours per month, they can leave the company.
* Employees who work 7 years or more are practically not released.
* Many people with low wage can go away.
* A lot of people who left the company are from the following departments: HR, Accounting, Technical.

*Important factors are:*

* + low salary;
  + a certain department;
  + low level of satisfaction;
  + significant amounts of working hours per month;
  + large number of projects.