Employee Churn Analysis

Group G1

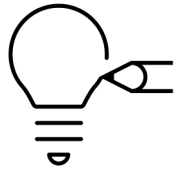
Albert, Gizem, İhsan, İpek, Osman, Safiye

CLARUSWAY
WAY TO REINVENT YOURSELF

AGENDA

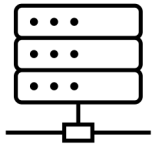
BUISNESS PROBLEMS

DATA UNDERSTANDING

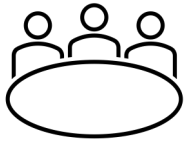


DATA PROCESSING

DATA VISUALIZATION



MODEL BUILDING



FUTURE ENHANCEMENT

BUSINESS PROBLEMS

People are expected to give their all – labor, passion, and time – to their jobs. But if their jobs don't give back enough, they will leave.

As have 4.5 million burned-out American employees who quit their jobs since November 2021 due to low satisfaction.

Ref:https://datasciencedojo.com/blog/churn-rate-prediction/#

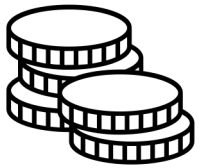
WHAT IS EMPLOYEER CHURN ANALYSIS?

Employee churn analysis refers to the process of examining and evaluating the factors that contribute to employees leaving an organization voluntarily or involuntarily. Also known as employee turnover or attrition, employee churn is a critical metric for businesses to understand, as high turnover rates can be costly and disruptive.

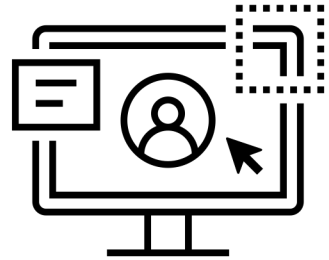
The analysis involves studying patterns, trends, and reasons behind employees leaving the company. It aims to identify the root causes of turnover, assess the impact on the organization, and develop strategies to reduce or manage employee churn.

WHY DO WE CARE ?

Cost



Talent Retention



ATTRIBUTES

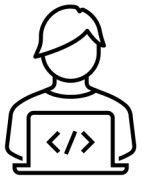
Target Variable
Left

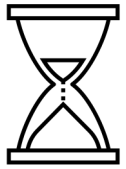
Numerical


Satisfaction Level


Last Evaluation


Number of Project

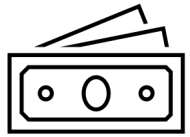

Average Montly Hours


Time Spend Company

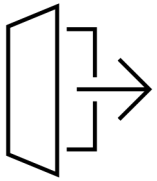
Non-Numerical


Work Accident


Promotion Last 5 Year


Salary


Department

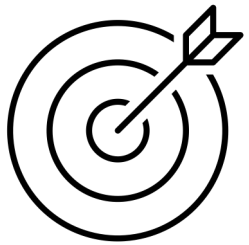

Left

DATA UNDERSTANDING

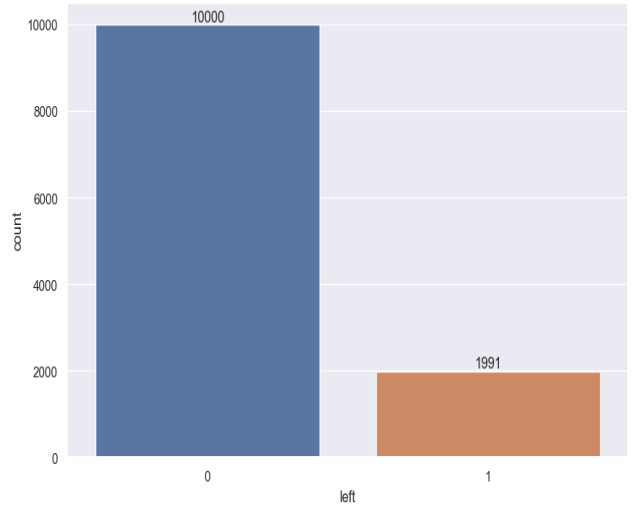
- ✓ Total number of observations is 14999 with 9 independent attributes.
- ✓ The given data has mix of numerical and categorical attributes.

Target Variable

Left



It is discrete and has high class imbalance.



ATRIBUTE	0	1
LEFT	11428	3571
LEFT (DISTINCT)	10000	1991

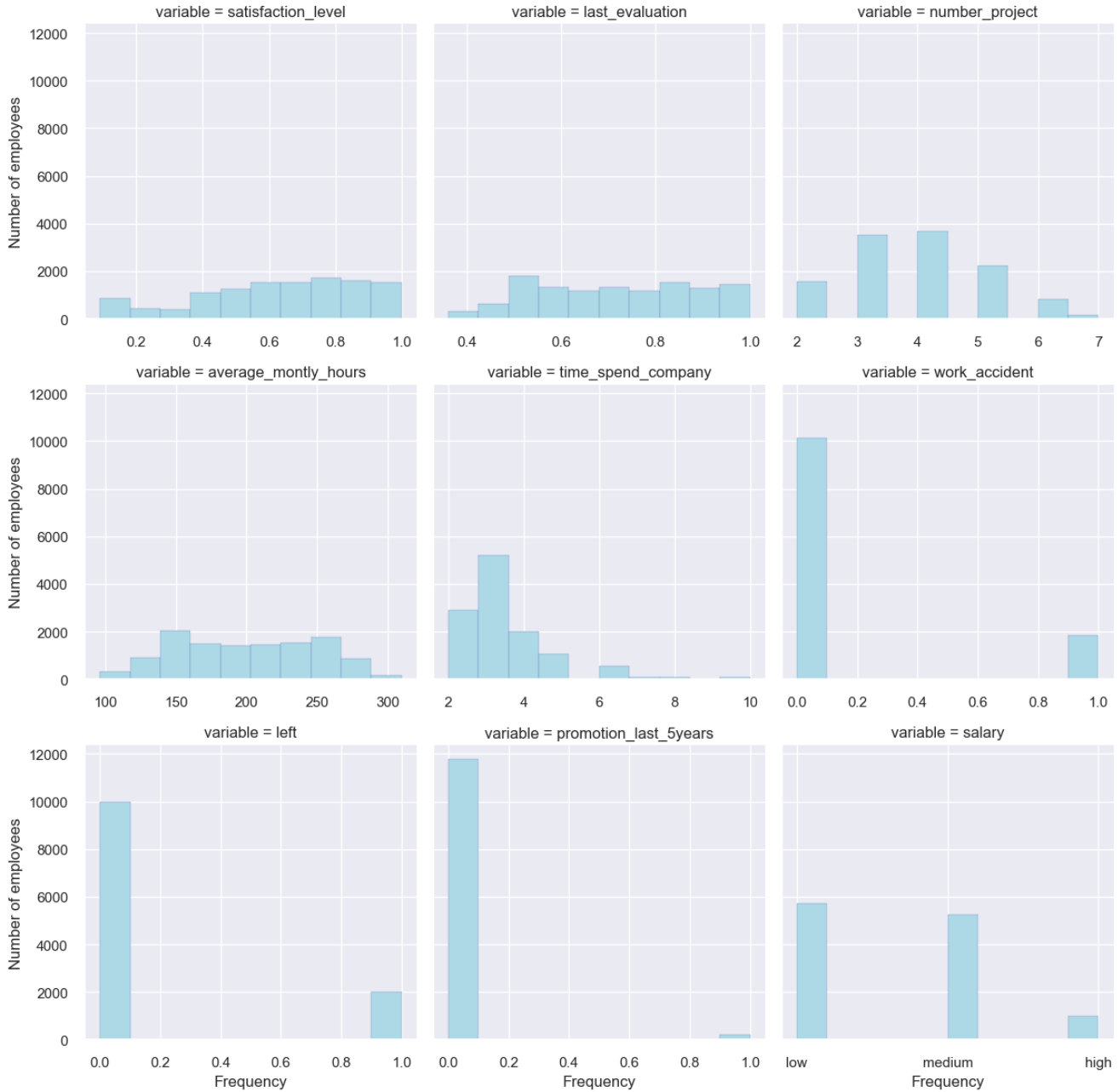
ATTRIBUTE RANGE LEVELS

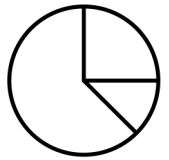
There are eight continuous variables and two categorical variables in the data set that offers information about 14999 employees. Continuous variables are those with numerical values, and categorical variables group things into category headers, like “Departments” that can have values similar to sales, marketing, consumer, operations, and so on.

ATTRIBUTE	RANGE/LEVEL
Satisfaction Level	**0.09 – 1.00**
Last Evaluation	**0.36 to 1.00**
Number of Project	**2 - 7**
Average Monthly Hours	**96- 310**
Time Spend Company	**2- 10**
Department	**10 levels**
Salary	**3 levels**
Left, Work Accident, Promotion Last % Years	**0,1**

DATA PROCESSING

Task	Implementation
✓ Dublicate records	✓ Using Distinct function
✓ No Missing Values	✓ ---
✓ Subsetting and Categorical Conversion	✓ Using as factor () function
✓ Standardization	✓ Using Range
✓ Handling Class Imbalance	✓ Using smote





These graph provides a convenient way to visualize the distribution of values across different variables in Data Frame, making it easier to compare and analyze patterns. The use of facets allows to see these distributions side by side for better insights.

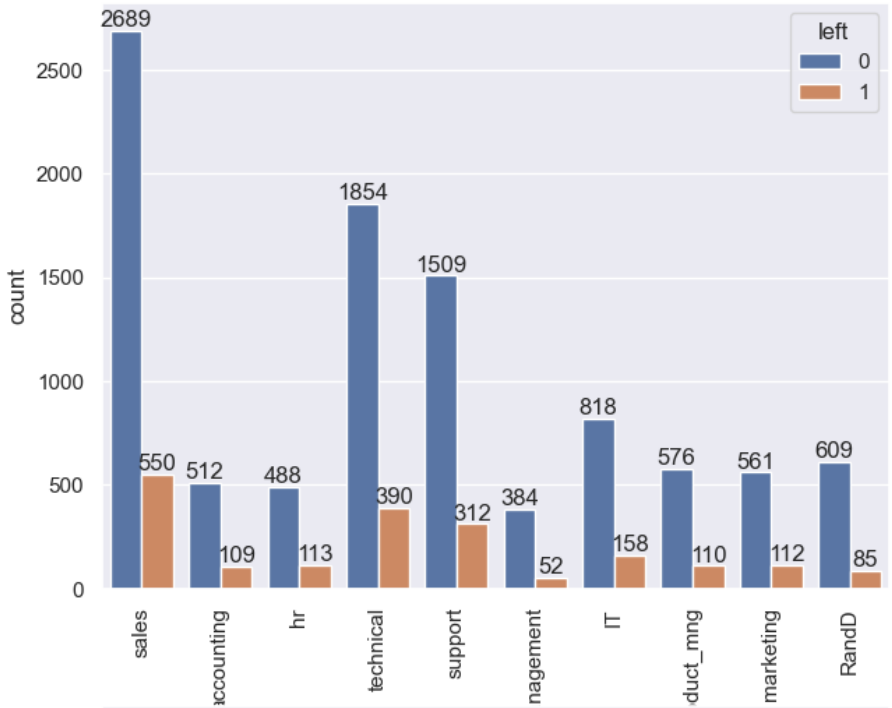
Satisfaction level: Most employees are highly satisfied.

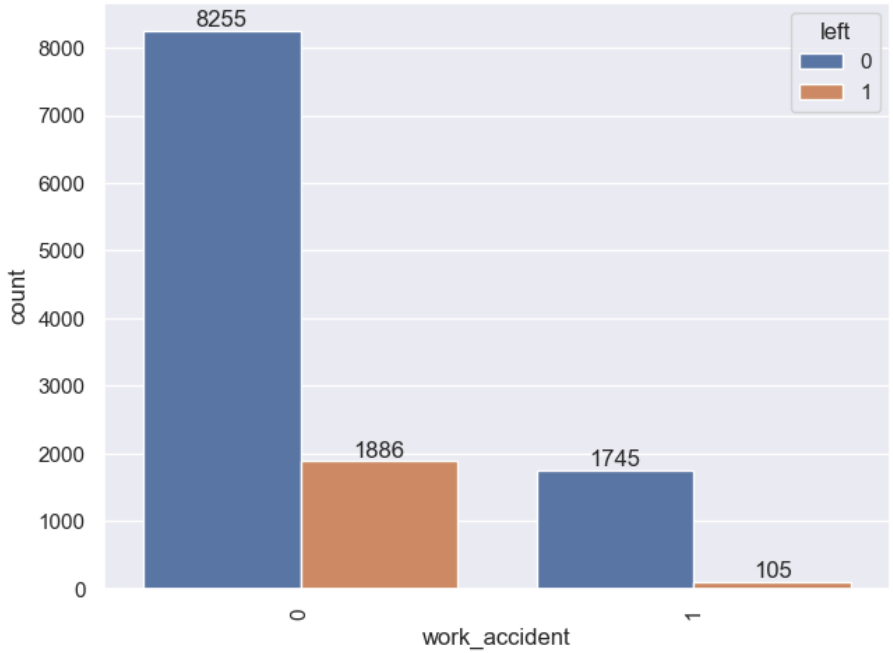
Last evaluation: Most employees are good performers with 75% of the data set being evaluated between 56%-87%.

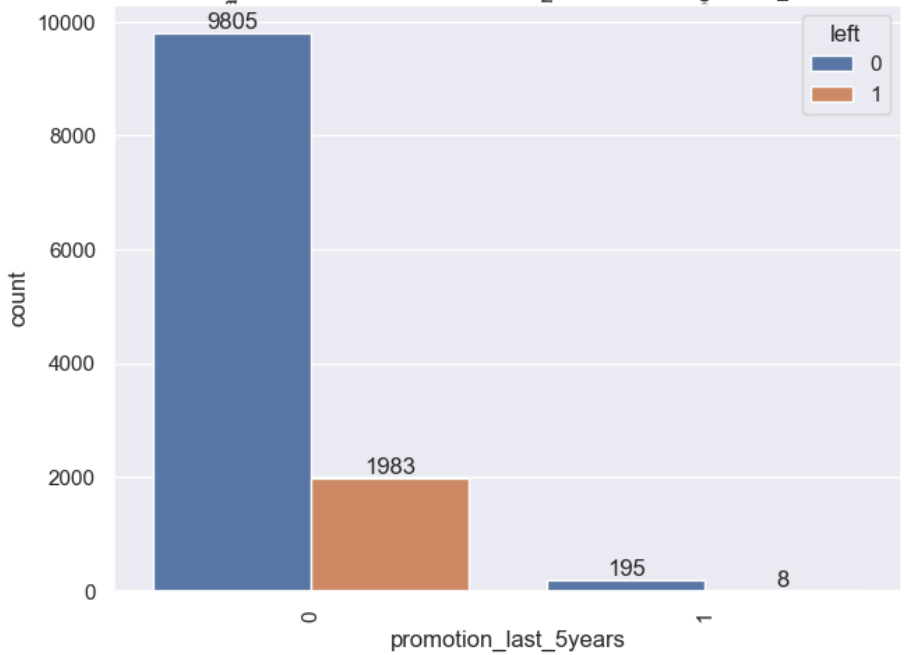
Number of projects: most employees do a reasonable number of projects.

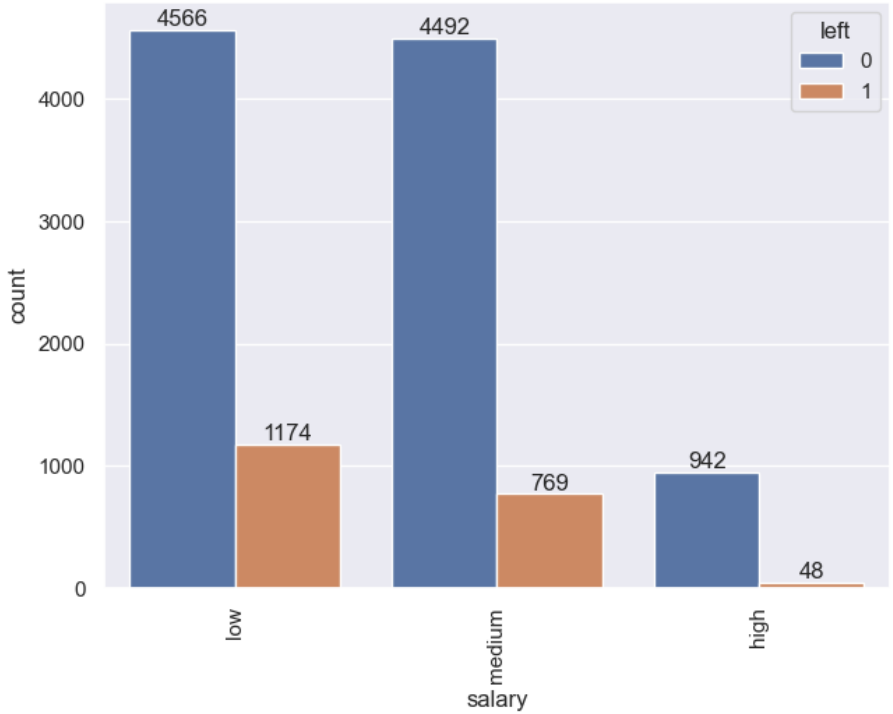
Average monthly hours: Most employees spend, fairly, a higher number of hours at work.

Time spent in the company: Fewer employees stay beyond 4 years.







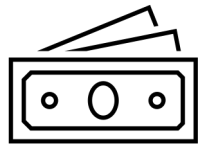


The rates of departures in departments are proportionally close to each other.

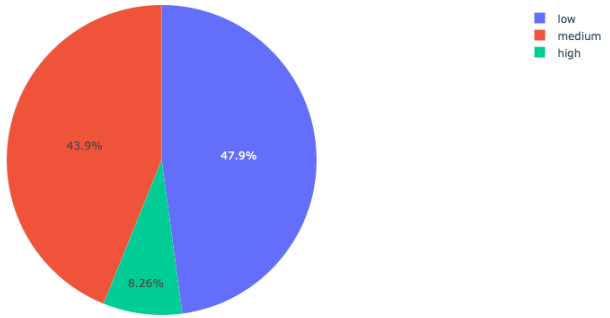
It has been observed that individuals who have not experienced work accidents tend to leave their positions more than those who have not.

The proportion of individuals who did not receive promotions and left is relatively higher.

It can be stated that the salary category with the highest departure rate is 'low.'

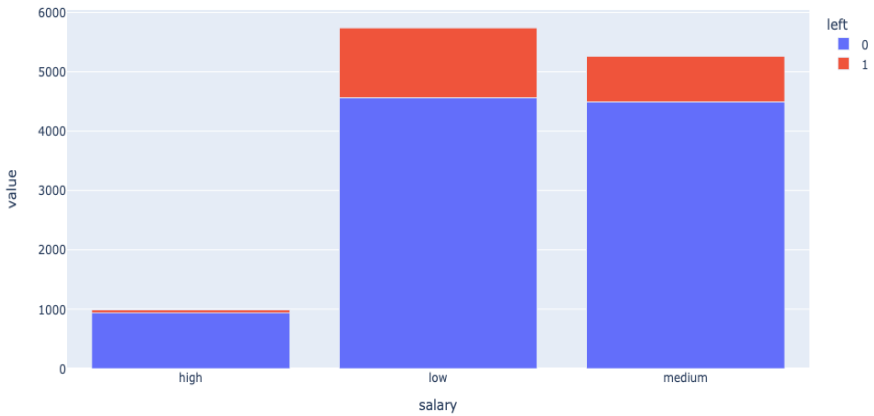


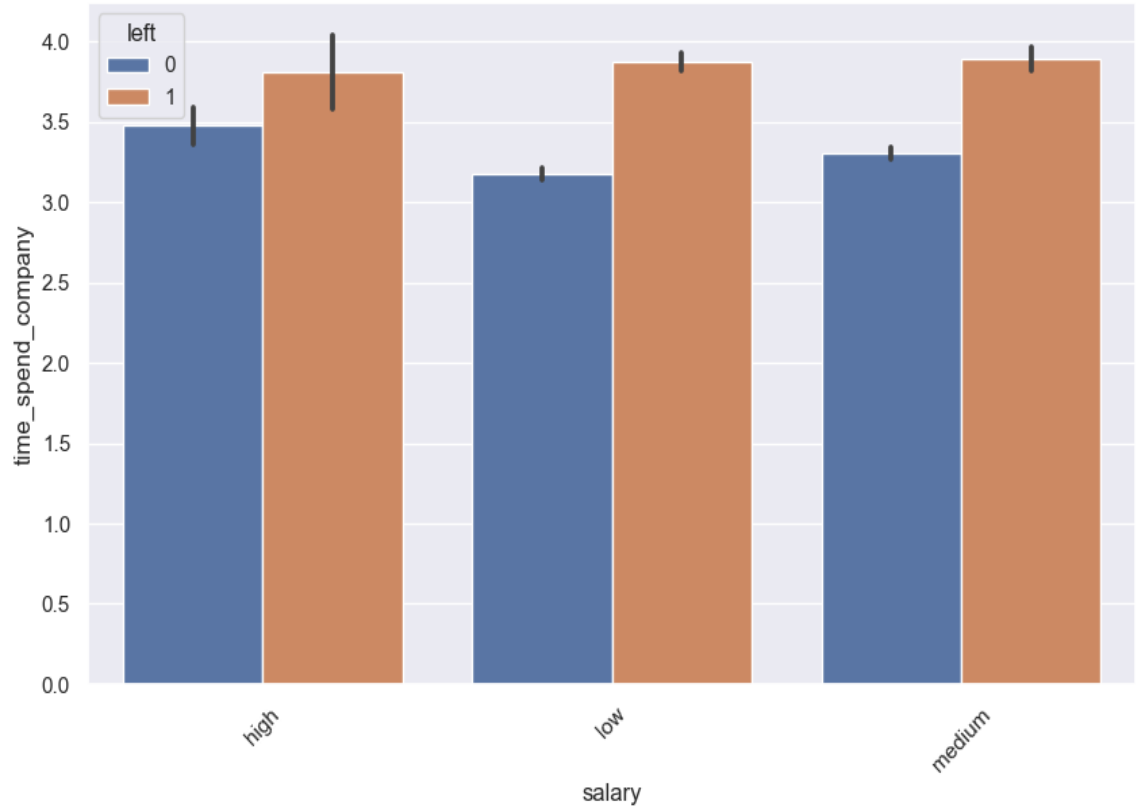
"salary" Column Distribution

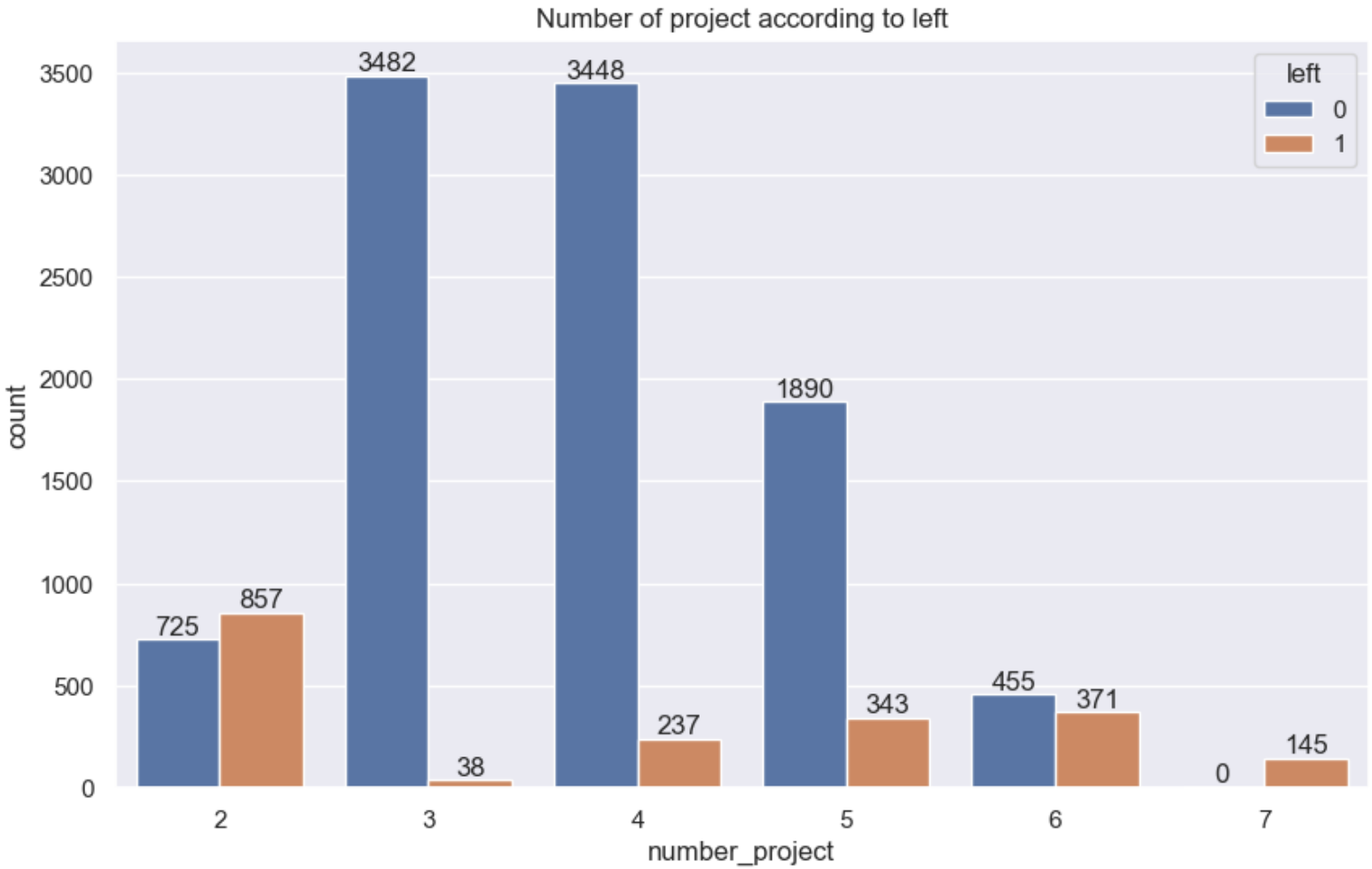


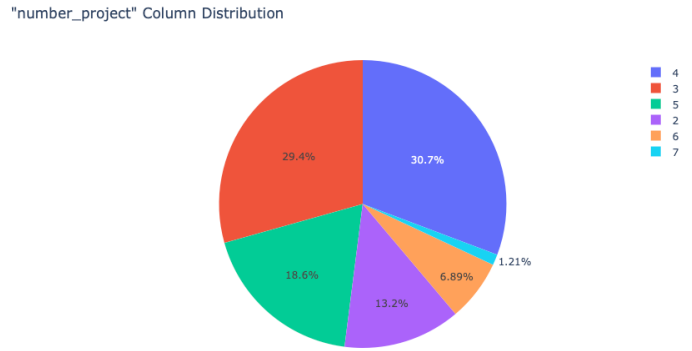
The number of individuals with high salaries in the dataset is very low.

salary and left



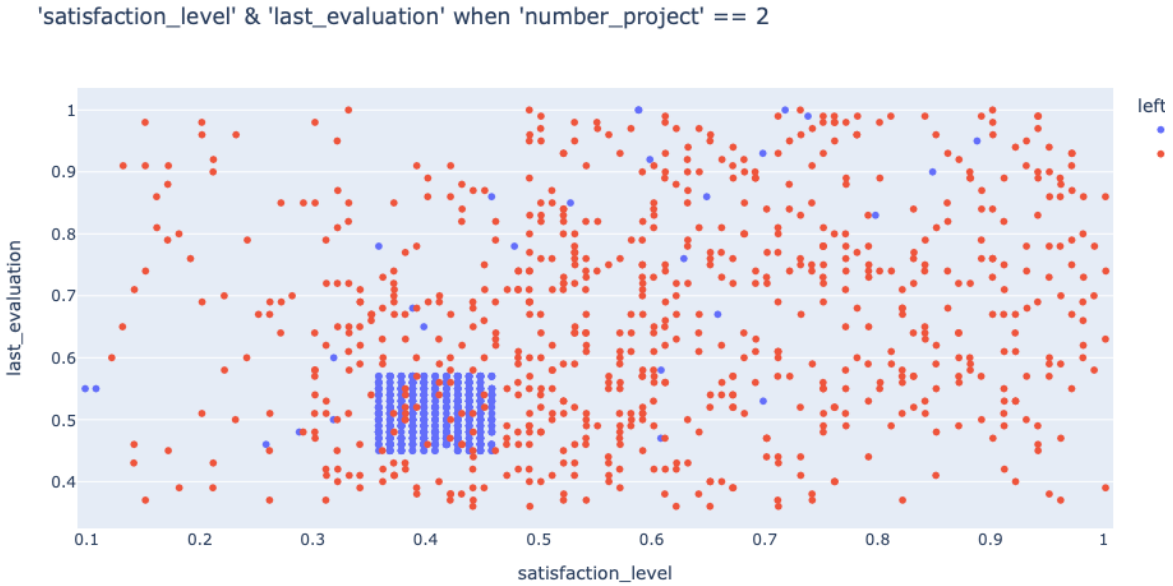


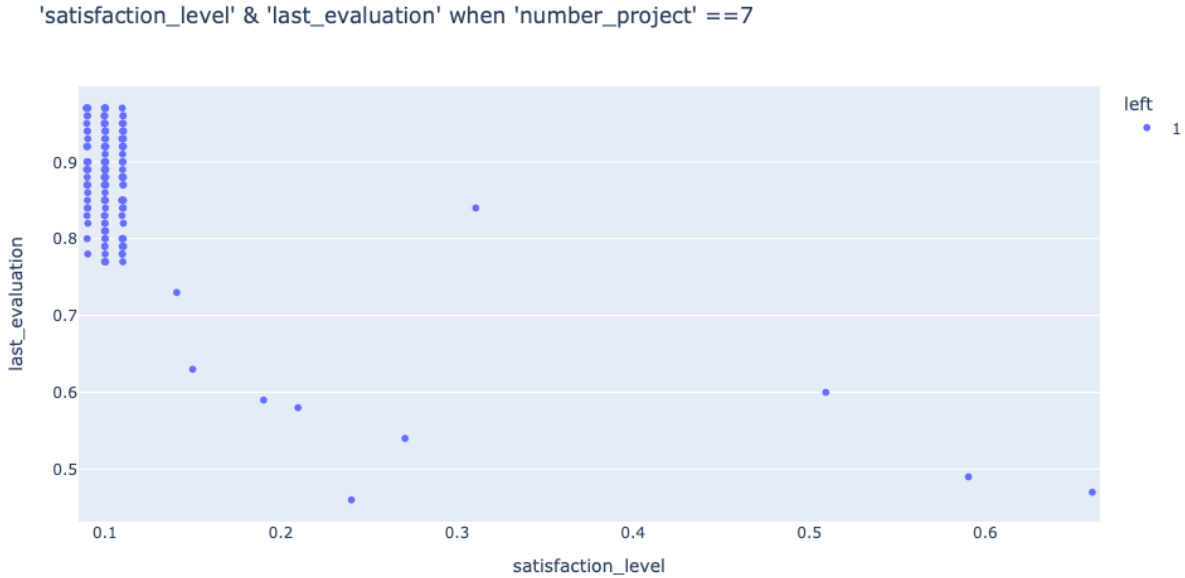




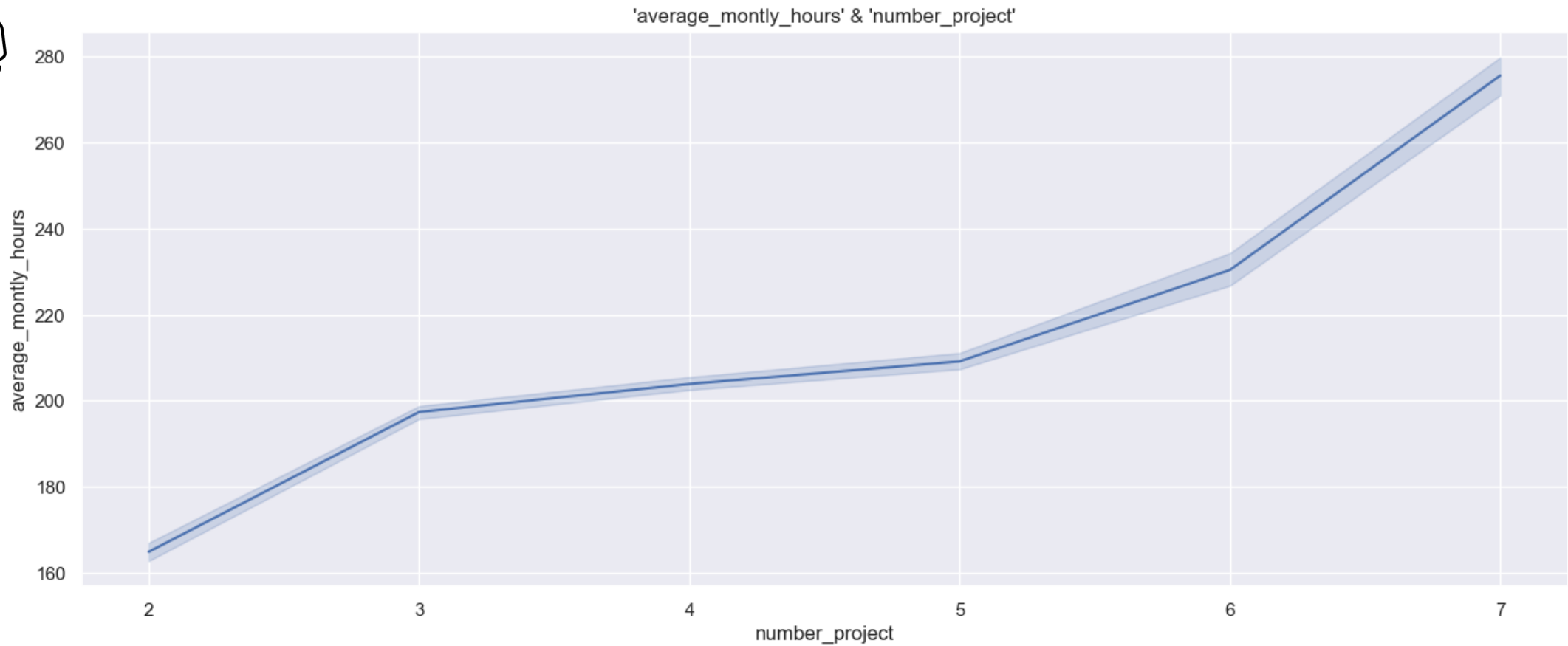
More than fifty percent of the employees are assigned to projects rated 3 or higher.

The number of leaving employees is higher among those who have only two projects during the period. This can be summed up as: "the employees with only two projects feel worthless or emptied". Because most of the employees work on three or four projects. With the 6th project, the number of resignings is getting over the number of ongoings. There are no ongoing staff members who were assigned to 7 projects. Working on more projects may cause intensive workload, regarding to this the satisfaction level may decrease with the insufficient motivators.

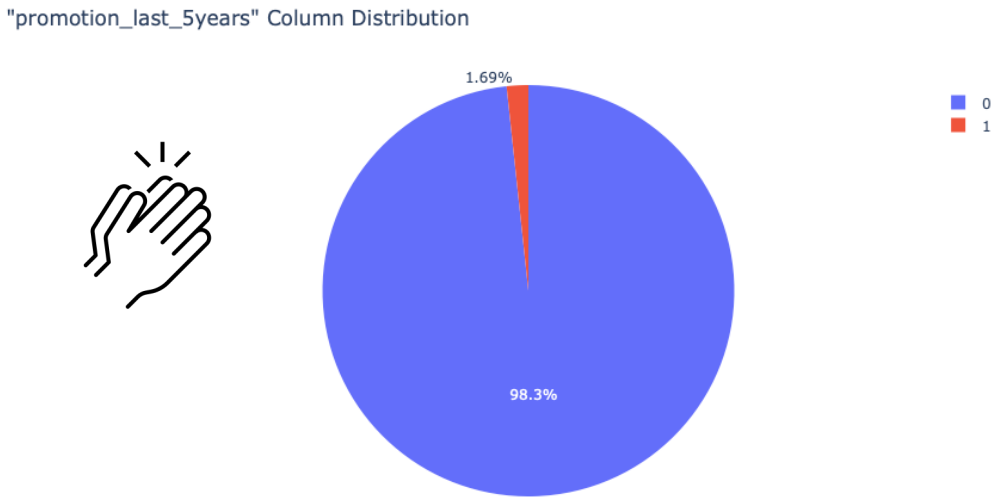




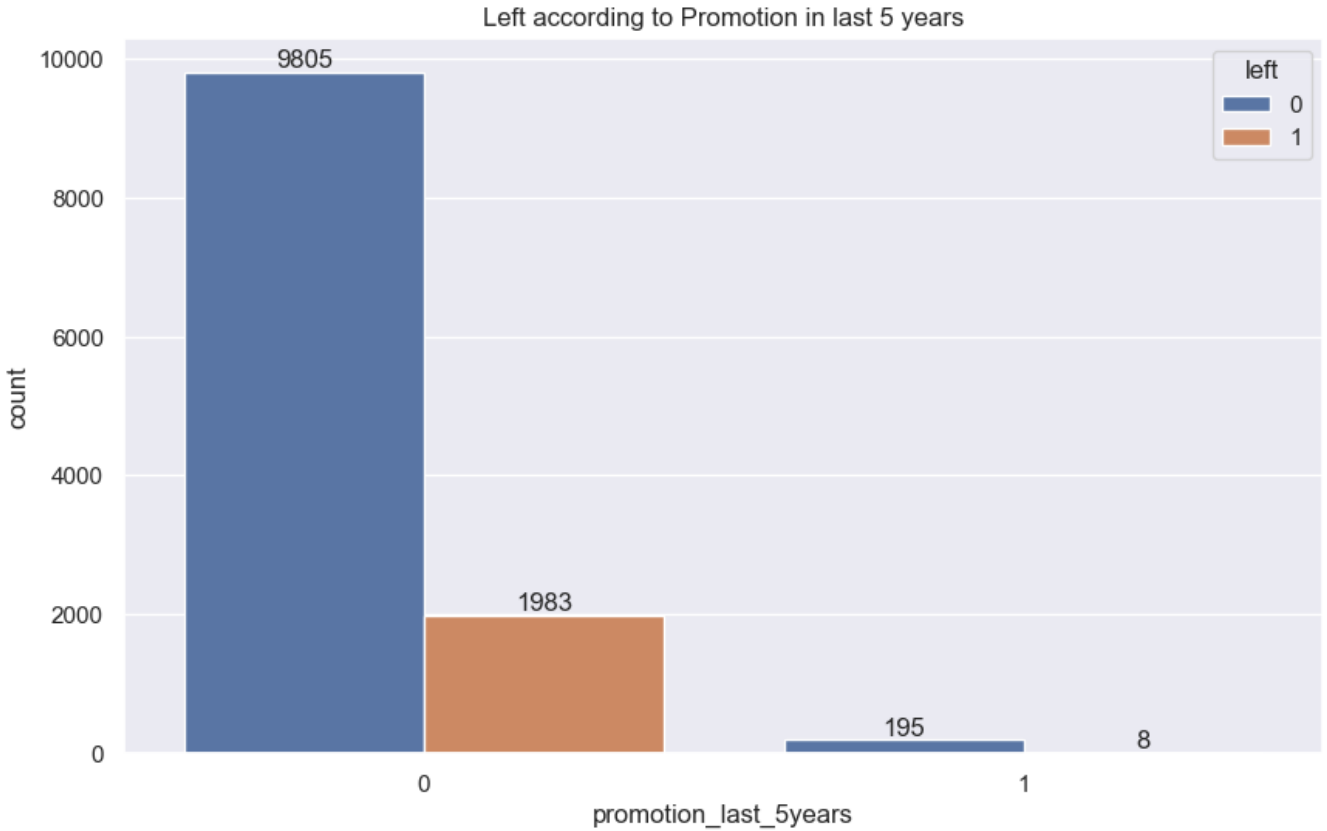
We had considered that taking on seven projects might also lead to resignations, and this trend is clearly demonstrated in the graph. Therefore, having too few or too many projects could be a reason for resigning.



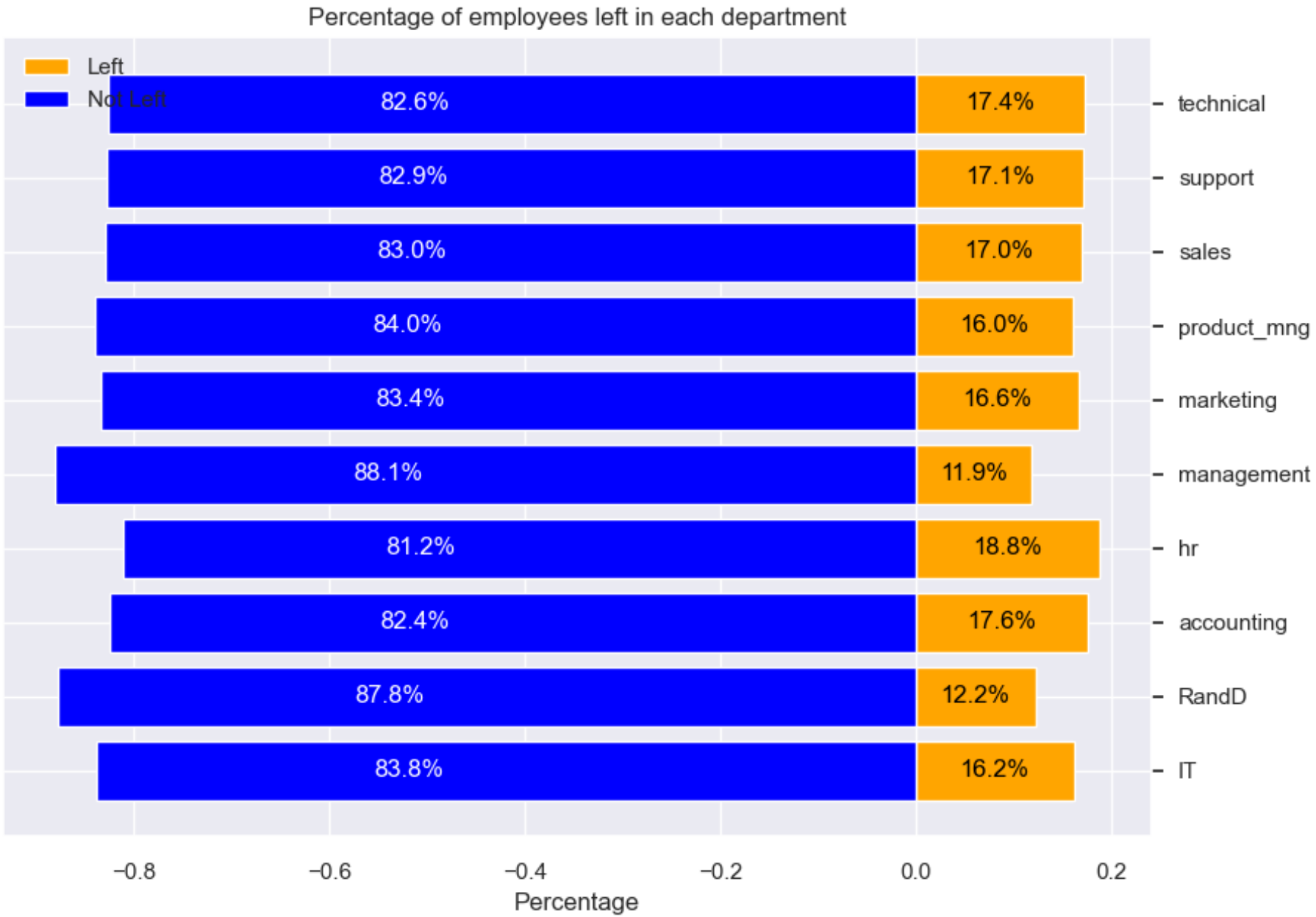
There is a dramatic increase in the monthly working hours when transitioning from two projects to three projects and six projects to seven projects.

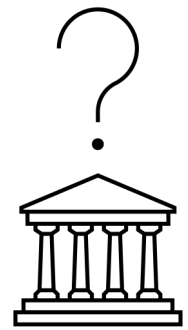


Almost nobody in the company has received a promotion in the last five years.

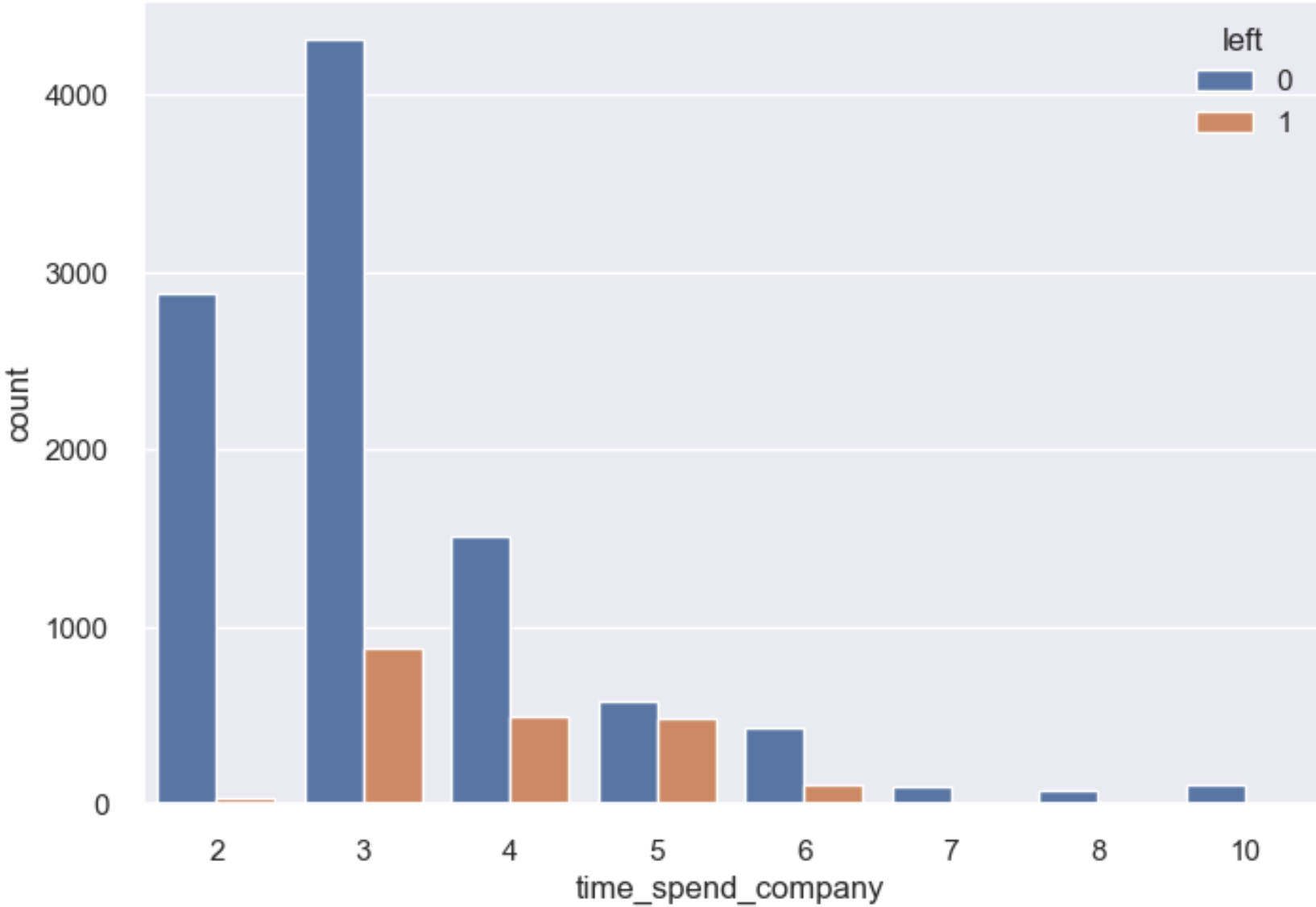


The number of individuals receiving promotions is very low. Additionally, among those who do not receive promotions, the departure rate is high.

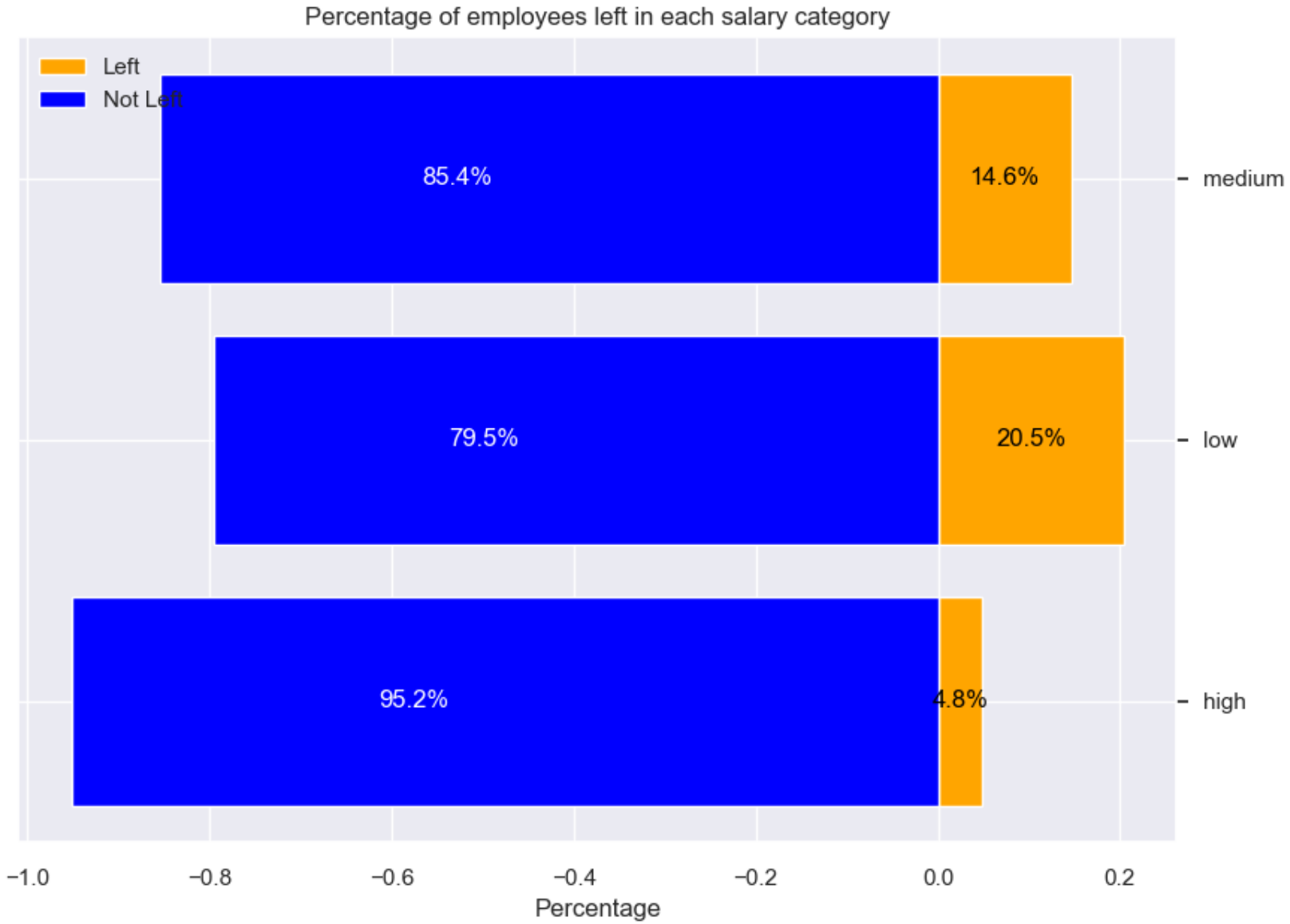


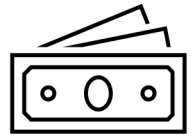


The departure rates in each department are very close to each other.



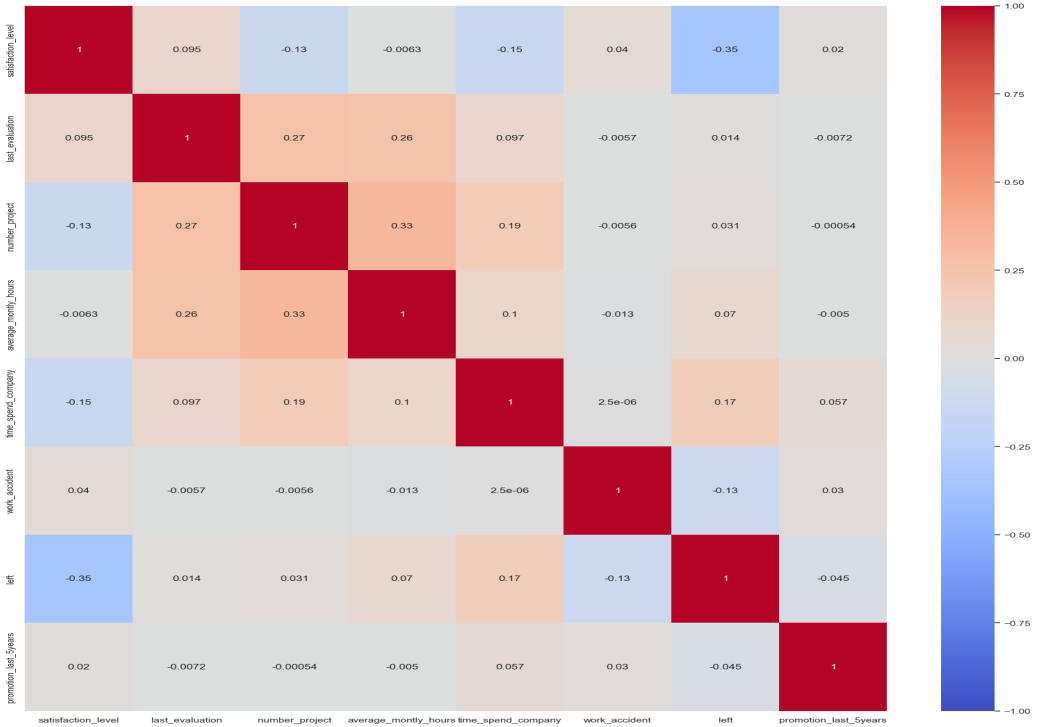
Observing the TIME SPEND COMPANY, the instances of employees leaving the company tend to increase during the 3rd year of employment, followed by a gradual decline until the 6th year.





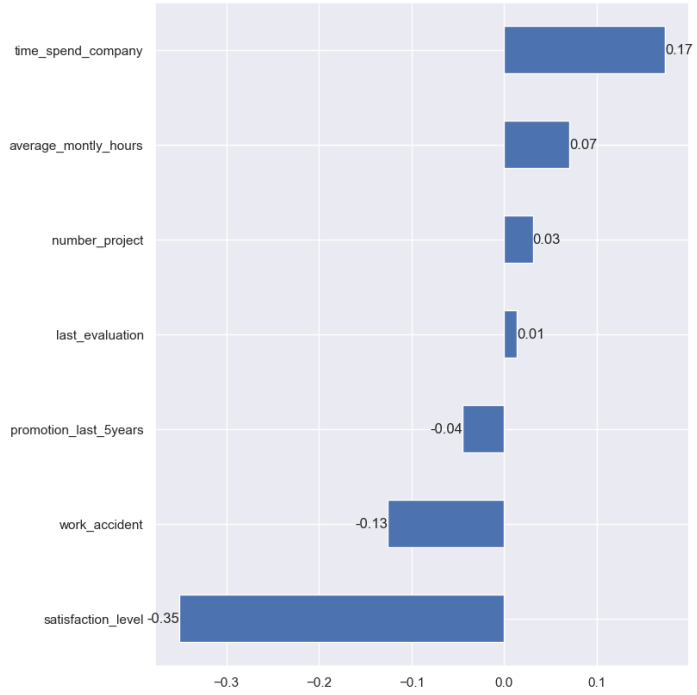
It was observed that the individuals with the highest departure rates predominantly consist of those with low salaries.

DATA VISUALIZATION



CORRELATION PLOT

There is no significant correlation among the features. Weak negative correlations are observed between 'left' and 'satisfaction_level' (-0.35), 'time_spend_company' and 'satisfaction level' (-0.15), as well as 'number_project' and 'satisfaction level' (-0.13). Additionally, weak positive correlations are observed between 'number_project' and 'average_monthly_hours' (0.33), 'last_evaluation' and 'number_project' (0.27), and 'last_evaluation' and 'average_monthly_hours' (0.26). The most prominent negative relationship is observed between 'left' and 'satisfaction_level', while the positive one is between 'number_project' and 'average_monthy_hours'.



<u>Possitive Corelation</u>

Number of Project, Last Evaluation, Average Monthly Hours, Time Spend Company

<u>Negative Corelation</u>

Satisfaction Level, Number of Project, Time Spend Company

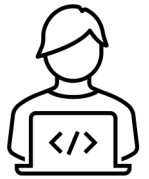
INFERENCE

REASON QUALIFIED PEOPLE LEAVING

Experienced

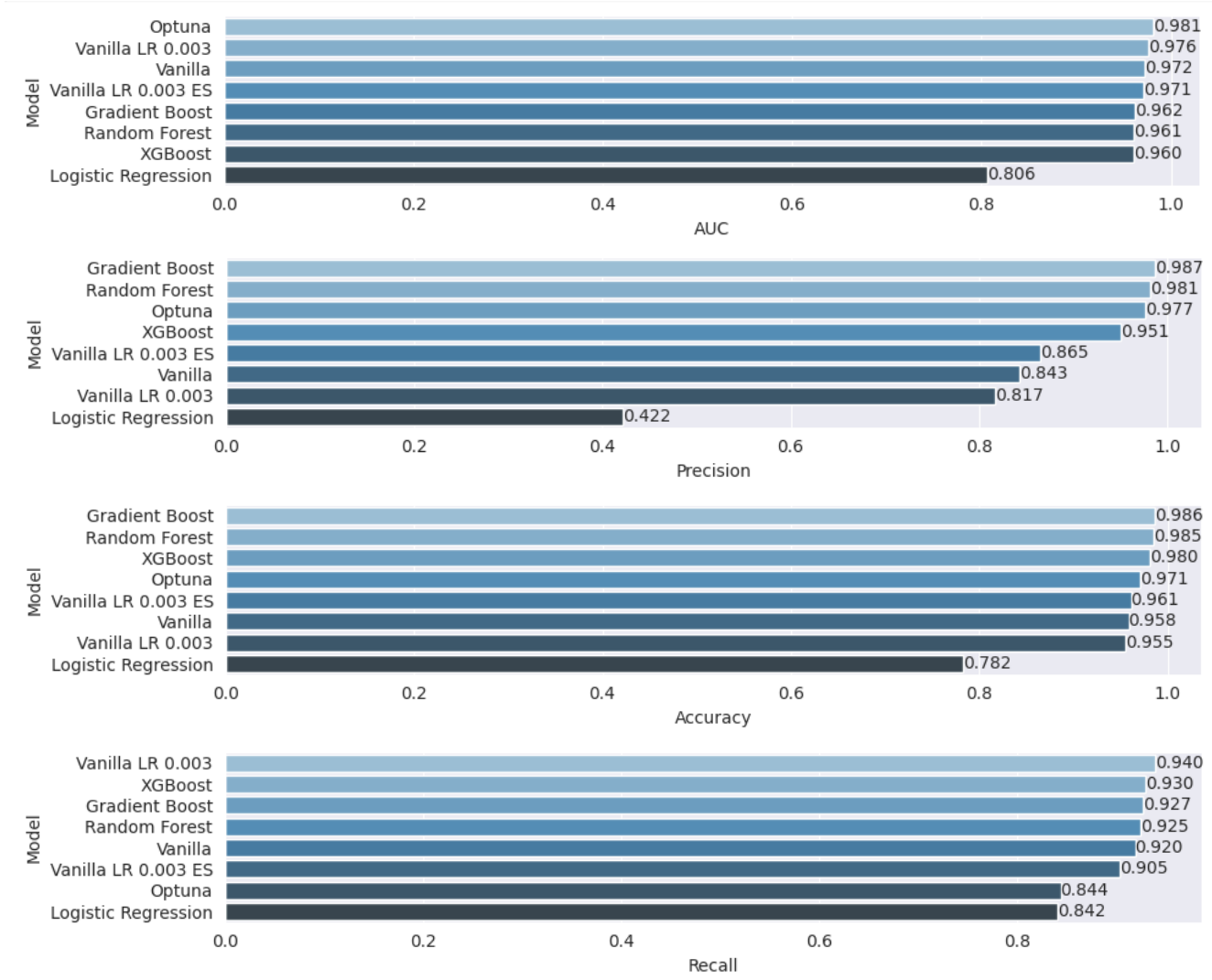
Low Satisfaction Level

Spend more Time at Work



- ✓ Experienced people may not have challenges level at work to improve their skills.
- ✓ Work to pay ratio is high : Low salary but high working hours.

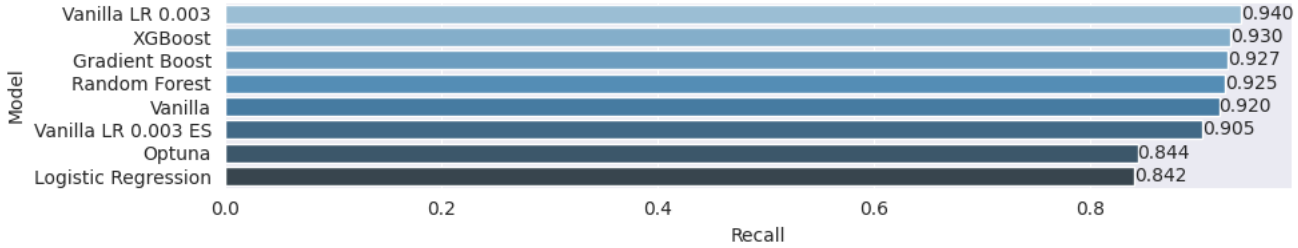
MODEL BUILDING In ML and DL



- ✓ As target variable Left is discrete hence various classification techniques has to be applied.
- ✓ Started Logistic Regression, Gradient Boast, Random Forest and last XG Boast are applied.
- ✓ After ML , DL models are built.

	Model	AUC	Accuracy	Precision	Recall	F1 Score
0	Optuna	0.981	0.971	0.977	0.844	0.906
1	Vanilla LR 0.003 ES	0.971	0.961	0.865	0.905	0.885
2	Vanilla	0.972	0.958	0.843	0.920	0.880
3	Vanilla LR 0.003	0.976	0.955	0.817	0.940	0.874
4	Logistic Regression	0.806	0.782	0.422	0.842	0.562
5	Random Forest	0.961	0.985	0.981	0.925	0.952
6	Gradient Boost	0.962	0.986	0.987	0.927	0.956
7	XGBoost	0.960	0.980	0.951	0.930	0.940

DL MODEL LAST EVALUATION



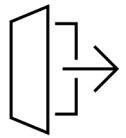
	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	work_accident	left	promotion_last_5years	departments	salary	predict
0	0.090	0.790	6	293	5	0	1	0	sales	low	1
1	0.740	0.960	4	154	4	0	0	0	support	medium	0
2	0.810	0.970	4	212	2	0	0	0	sales	low	0
3	0.370	0.540	2	149	3	0	1	0	support	low	1
4	0.100	0.770	6	272	4	0	1	0	accounting	low	1

Conclusion

Characterizing loyalty

Three conditions that affect loyalty are:

- ✓ *a high level of satisfaction (satisfaction_level >= 47%)*
- ✓ *have spent at least 4 years in the organization (time_spend_company < 5 years)*
- ✓ *good performers with an evaluation of at least 80% (last_evaluation < 81%)*



Characterizing left

Three conditions that affect 'resigned' are:

✓*low or moderate satisfaction (satisfaction_level < 47%)*

✓*have a workload of 6 or more projects (number_project >= 6 projects) and*

✓*their performance being evaluated at least 58% (last_evaluation >= 58 %)*

To Sum up :

HR analytics, the provenance of a few leading companies, a decade ago, is a solution that is being widely applied now by several growing businesses to uncover surprising sources of talent and counterintuitive insights about what drives employees to be loyal to their organization. We hope this encourages you to leverage the power of HR analytics to retain talent and save hiring costs

THANK YOU FOR YOUR ATTENTION ☺