**Employee Churn Prediction: An Exploratory Data Analysis Report**

Helmi Kaittikkattil Abraham(K00302088),

Srujana Kurudi Adishesha Gupta(K00302089)

1. **Introduction**

Employee turnover in an organisation can result from a variety of circumstances. Designing retention tactics will be made easier by identifying such elements. Because it takes more time and money to identify, interview, and hire new staff than to keep the ones you already have. A higher rate of staff retention will benefit the business.   
In a company's HR department, this dataset is used to analyse employee turnover and the factors that influence it. It includes 14,999 employee records and 10 attributes that have datas related to worker satisfaction, the number of tasks they must complete, time spent with the organisation, their history of promotions over the last five years, departments, and the associated pay. And it will be used to determine whether or not an employee is likely to leave.

For data examination, the mix of binary, category, and numeric data types were used from different attributes. Various analytics tasks are supported by the dataset, including determining the causes of employee attrition, evaluating the relationships between workload and happiness, and looking at retention patterns particular to a department.

1. **Individual Attributes**

Each of the ten attributes in the provided dataset offers a distinct perspective on the churn behaviour of employees.

* **Satisfaction\_level**: It is the score for employee satisfaction which goes from 0 to 1, with 1 denoting complete satisfaction. While the majority of values fall somewhere in the middle, some employees exhibit extremely high or low levels of contentment or discontent.
* **Number\_project**: This attribute, which ranges from 2 to 7, indicates the number of projects that each person is working on. It aids in determining workload and possible burnout**.**
* **Average\_monthly\_hours**: This is the average number of hours worked per month, ranging from 96 to 310. Higher monthly hours may be associated with burnout, and the distribution seems unbalanced.
* **Time\_spend\_company**: This represents the number of years one working for the company. It ranges from 2 to 10 years, indicates employee longevity; most employees have less years, which may indicate high churn or recent recruits.
* **Work\_accident**: This binary attribute indicates if the employee was involved in an accident at work. Most have not been in any accidents, indicating general safety with a few odd incidences.
* **Promotion\_last\_5\_years**: This binary field indicates if the worker has received a promotion during the previous five years., which ranges from 0 to 1. The majority of employees have not been promoted which acts as a major cause for churn.
* **Department**: This is a categorical property that lists departments such as support, technical, and sales and more. Comparing responsibilities enables one to determine which departments have the highest turnover rates. This feature helps to analyse churn trends within specific departments.
* **Salary**: This nominal attribute contains salary data as low, medium, and high levels. Fewer workers are top earners, with the majority falling into the low- and medium-income ranges.
* **Left**: This special binary attribute shows whether an employee stayed with the company (0) or left (1). Analysing the elements influencing employee churn requires this quality.

1. **Groups of Attributes**

After importing the dataset, the initial analysis indicates balanced data with minor outliers in some attributes (like monthly hours). The dataset appears well-prepared for further exploration however some duplicates were there and ‘Remove Duplicates’ operator resolved this issue. This reduced the unique values into 11991.

Exploring the relationships between attributes reveals several important trends:

* Satisfaction Level (Satisfaction\_level Vs left): There is a negative correlation between satisfaction and turnover; employees with low satisfaction levels are more likely to leave, indicating that improving satisfaction could reduce attrition. This satisfaction level has a direct relation with the time spend in the company, as the employee spend more time, it shows tendency to produce a churn. Whereas more satisfied workers maintain a minimum working hour. (Figure 1)
* Workload and Satisfaction (Number\_project Vs Time\_spend\_company): High project loads and monthly hours are associated with lower satisfaction levels, suggesting possible burnout among overworked employees. Employees with higher number of projects who works for a minimum time tends to leave the company, however those employees who maintains a minimum number of projects and time remains in the firm. (Figure 2)
* Promotion and working hours (Promotion\_last\_5\_years Vs Time\_spend\_company): Longer tenure is positively correlated with promotions, but very few employees have received promotions, which may indicate limited growth opportunities. And if promotion is there, then working hours doesn’t matter. (Figure 3)
* Salary and churn rate (Salary Vs Left): Certain departments, such as management, have more high-salary employees, while others like support and sales have predominantly low-salary employees. This could affect turnover rates in different departments. Anyhow, low level salary exhibit churn rate more compared to high level (Figure 4)
* Department and Retention (Department vs. left): Higher churn rate was experienced in sales department whereas lower numbers are in management level. Departments like accounting, HR, IT, marketing exhibit similar rate of churn. Employees in technical as well as support department tends to remain in the company for long (Figure 5)
* Department and working hours (Department Vs Time\_spend\_company): Almost all departments show similar pattern other than management department.
* Promotion and Department (Department Vs Promotion): Department with less promotion shows more turnover rates which is IT and product department. Whereas employees tend to remain in the company if they get promoted.

**4.** **Interesting Findings**

* More satisfied employees work for less hours. Whereas less satisfied employees who works for a long time shows the tendency to quit.
* If number of projects are in medium range and the employees get more time to work on it, then they stay and there is no tendency to leave. However, if number of projects are high and the time spend in the company is less, then turnover rate is more.
* If the employees are honoured with promotion, then working hours is inversely proportional to churn rate, but if they are not getting any promotion, time spend in company get reduced and ends in churn.
* Highest salaried employees show high retention pattern whereas lowest salaried tends to churn.
* Higher churn rate and higher retention rate are in sales department whereas lower levels for both are found in management level.
* Management level employees have to spend more time in company compared to other departments.
* Product management team doesn’t have any promotion for the last 5 years whereas sales team gave the most. And also churn rate is high in IT department due to less promotions.

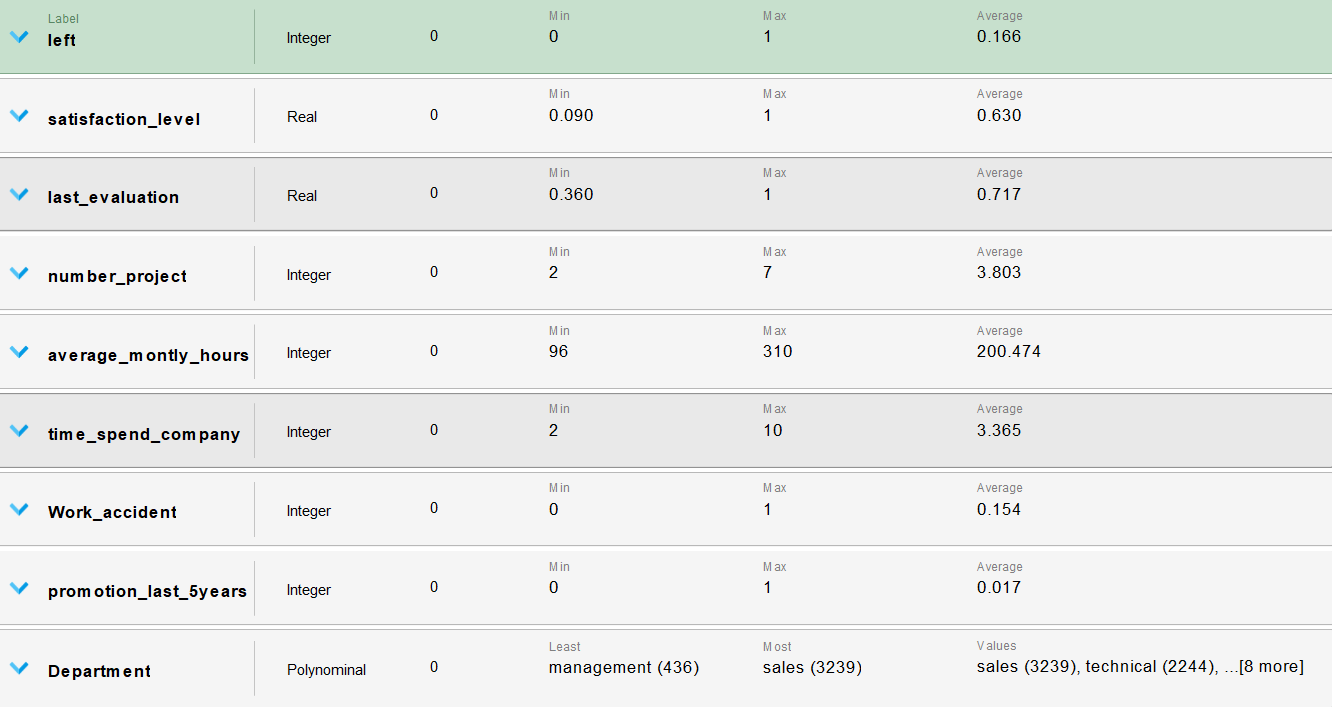
The HR division can investigate staff retention tactics with the help of these insights. There seem to be two primary things that may be done: increasing pleasure and lowering overwork.

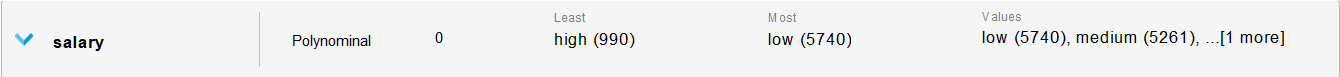
1. **Appendix**

**Reference**

<https://www.kaggle.com/datasets/liujiaqi/hr-comma-sepcsv/data>

**Table**

****

****

**Visualizations**

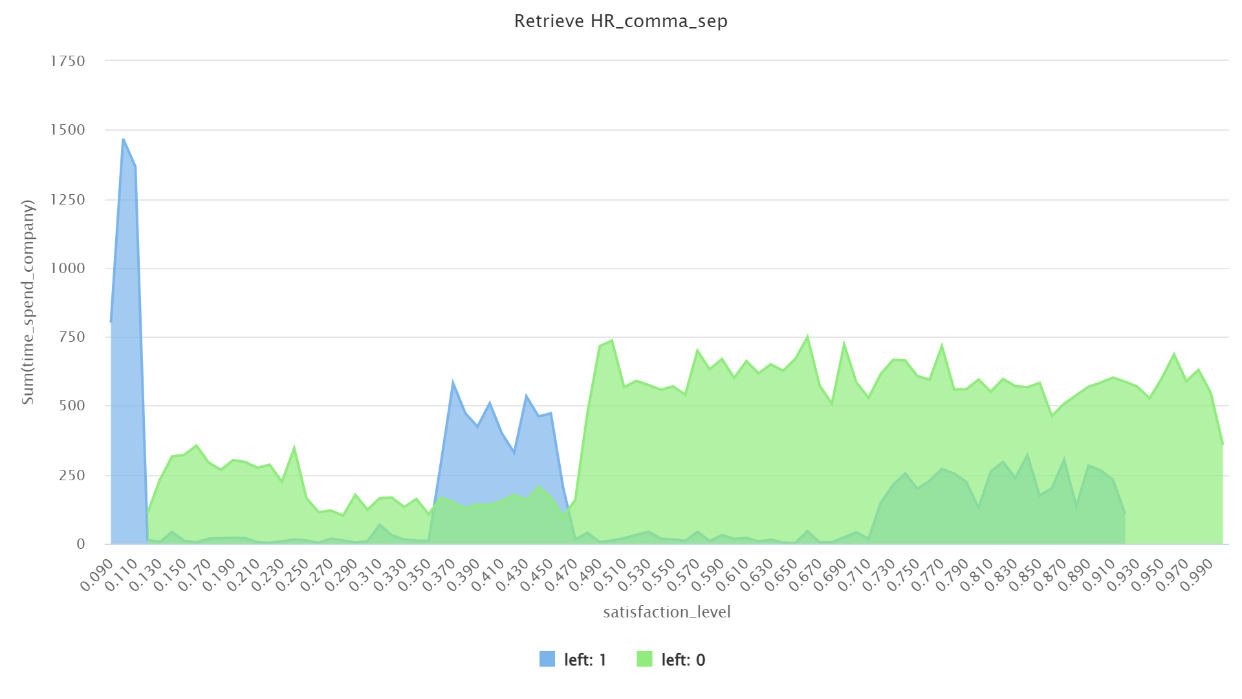
Figure 1: Satisfaction\_level Vs Time\_spend\_company

Figure 2: Time\_spend\_company Vs Number\_project

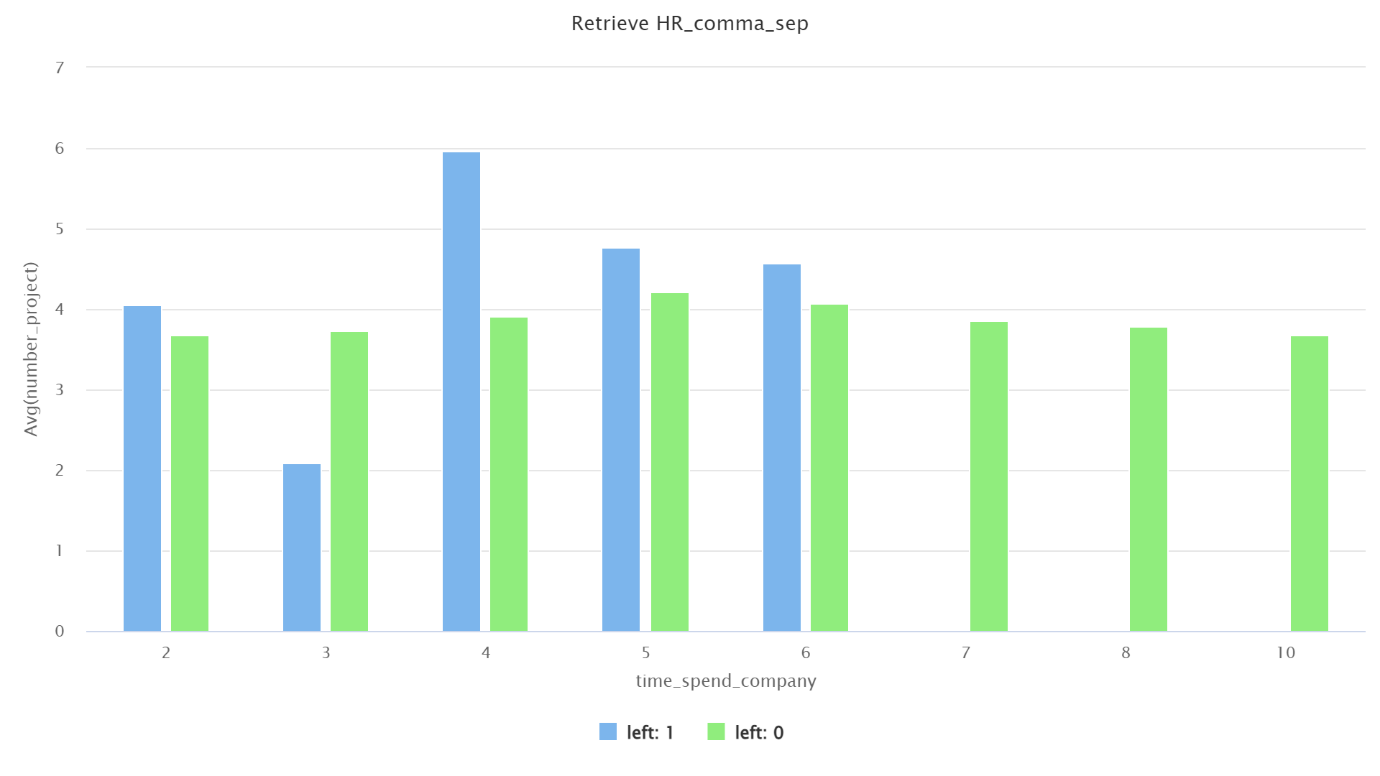


Figure 3: Promotion\_last\_5\_years Vs Time\_spend\_company

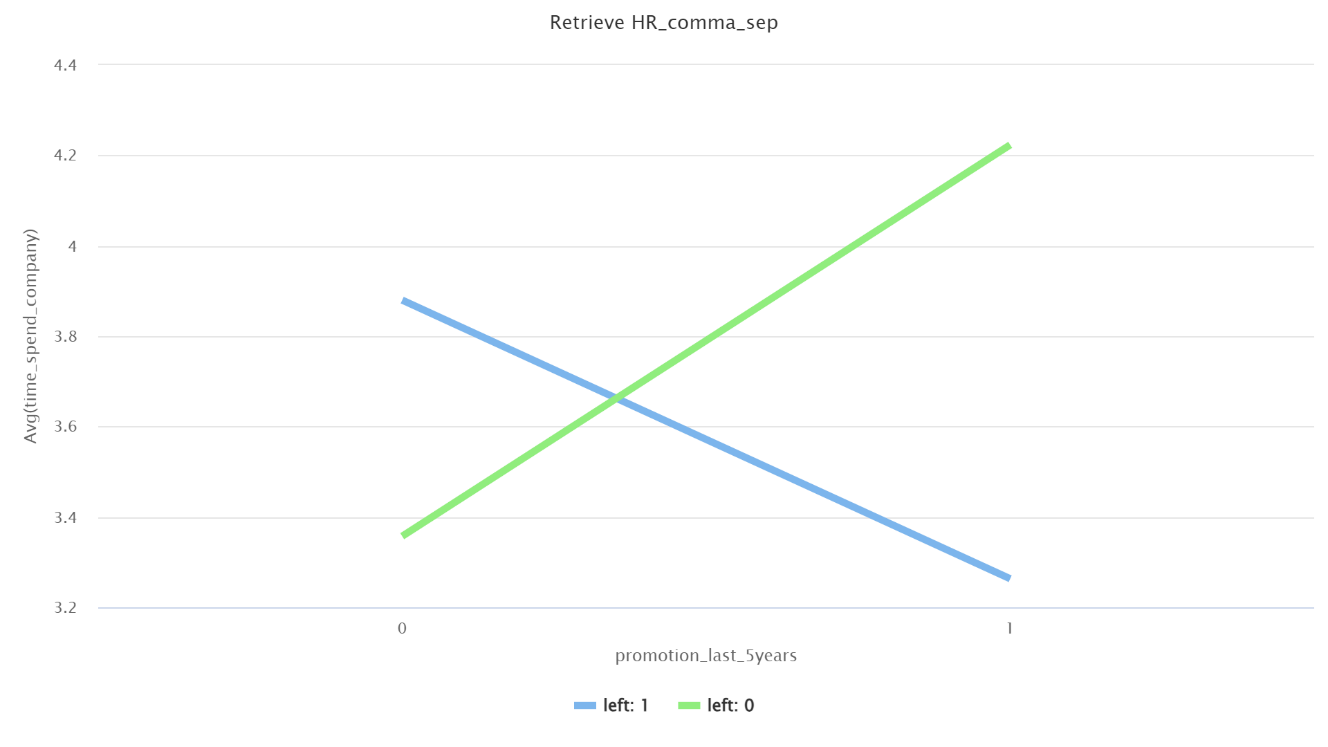


Figure 4: Salary Vs Left

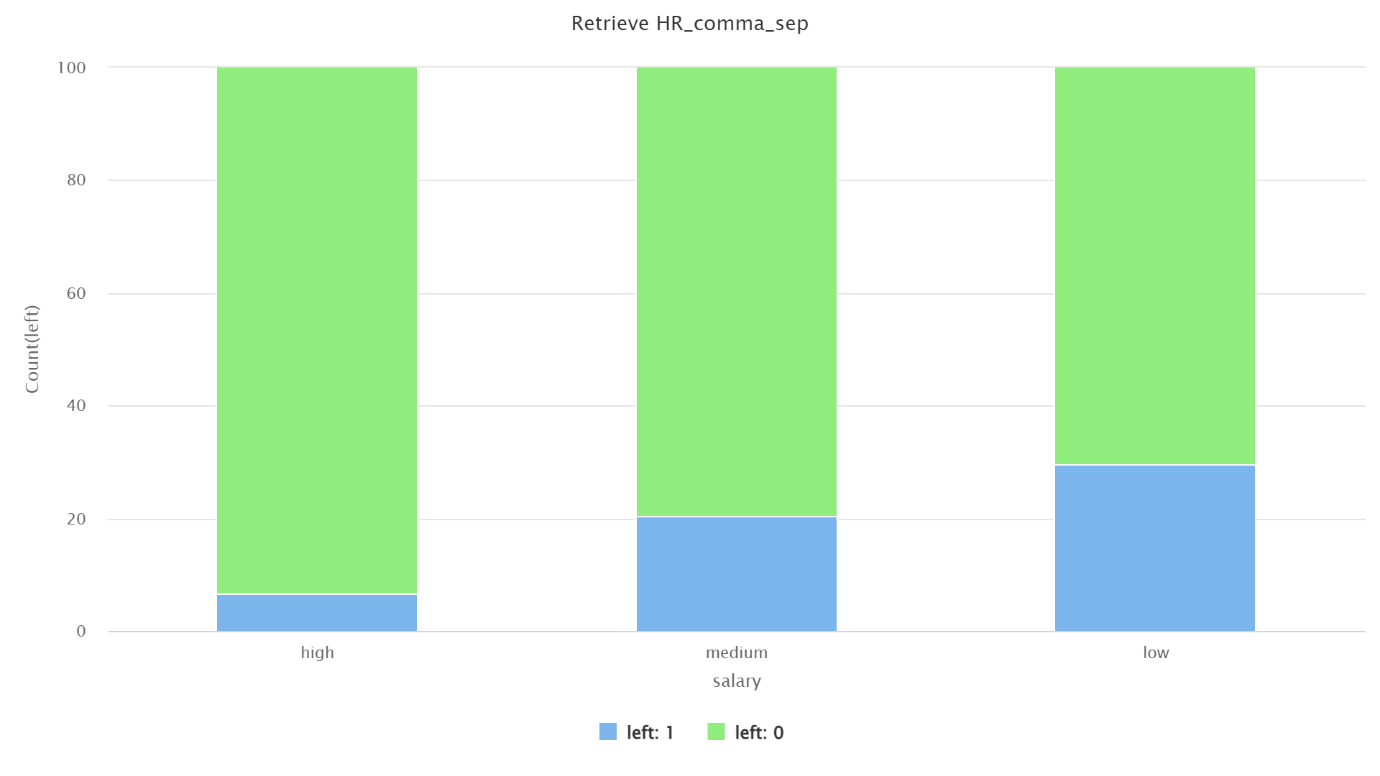


Figure 5: Department Vs left

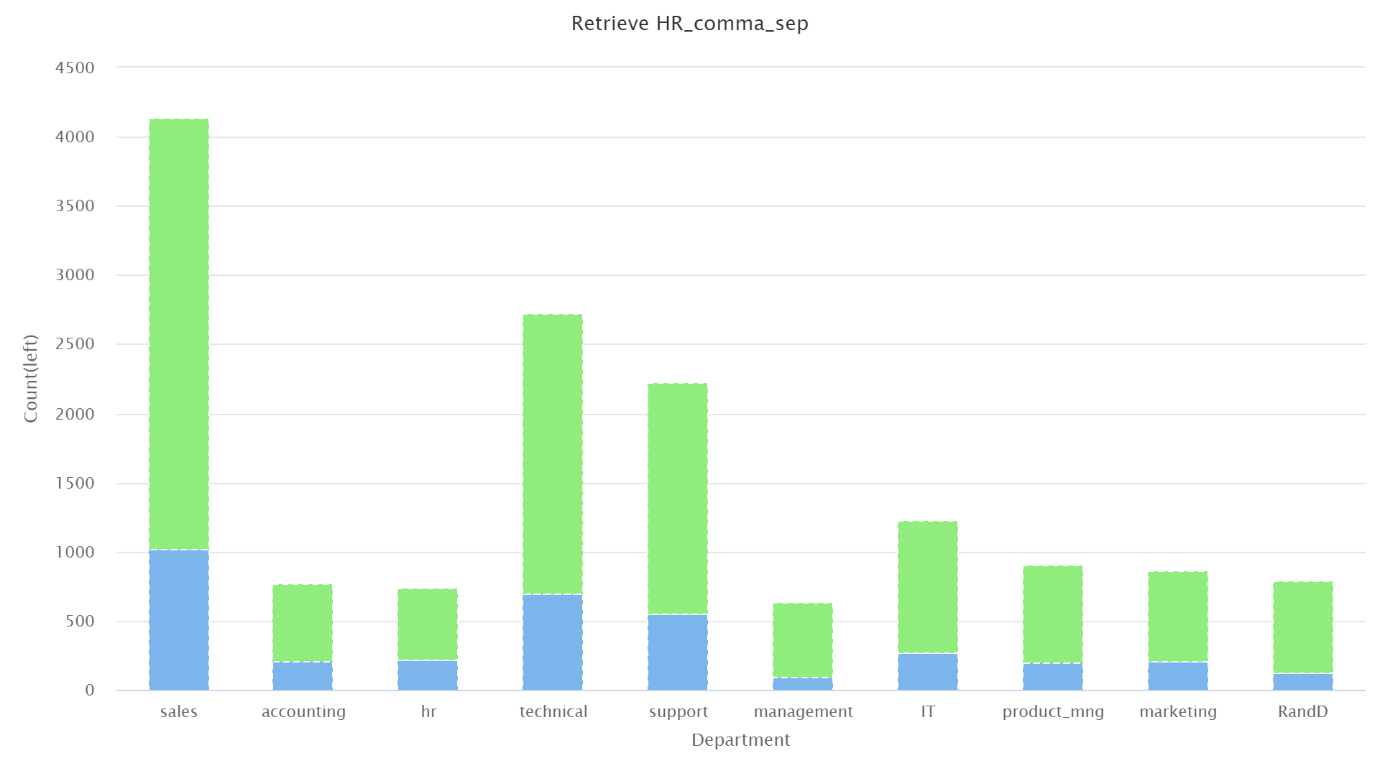


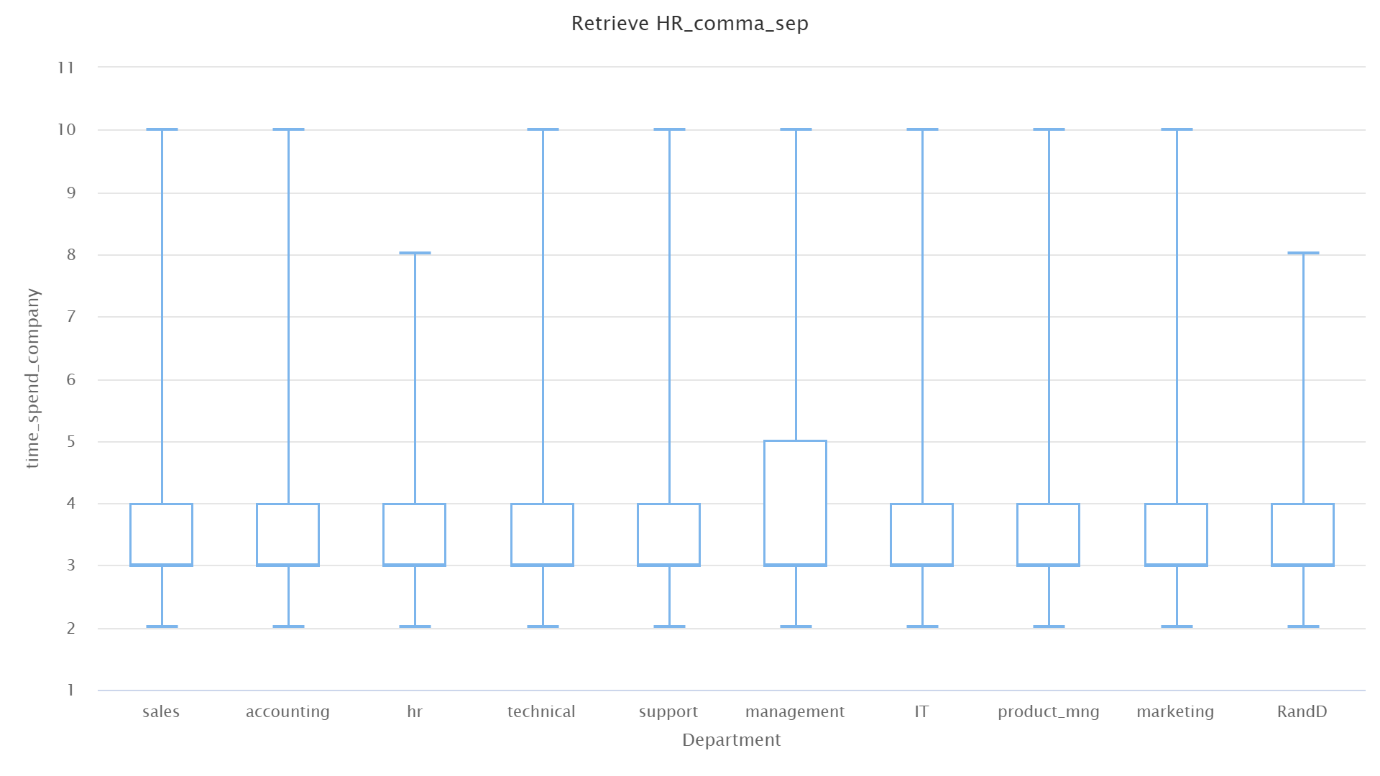
Figure 6: Department Vs Time\_spend\_company 

Figure 7: Department Vs Promotion\_last\_5years

