```
#install all libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

pd.read_csv('/content/employe.csv')
df=pd.read_csv('/content/employe.csv')
df.head()
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotionInLast5year
0	0.38	0.53	2	157	3	0	1	
1	0.80	0.86	5	262	6	0	1	
2	0.11	0.88	7	272	4	0	1	
3	0.37	0.52	2	159	3	0	1	
4	0.41	0.50	2	153	3	0	1	

```
df.shape
```

```
(14999, 10)
```

Dataset consists of 14999 rows so it contains information of 14999 employees

```
df.isnull().sum()
```

	0
satisfactoryLevel	0
lastEvaluation	0
numberOfProjects	0
avgMonthlyHours	0
timeSpent.company	0
workAccident	0
left	0
promotionInLast5years	0
dept	0
salary	0

dtype: int64

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14999 entries, 0 to 14998
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   satisfactoryLevel      14999 non-null  float64
 1   lastEvaluation         14999 non-null  float64
 2   numberOfProjects       14999 non-null  int64
 3   avgMonthlyHours        14999 non-null  int64
 4   timeSpent.company      14999 non-null  int64
 5   workAccident           14999 non-null  int64
 6   left                   14999 non-null  int64
 7   promotionInLast5years  14999 non-null  int64
 8   dept                   14999 non-null  object
 9   salary                 14999 non-null  object
dtypes: float64(2), int64(6), object(2)
memory usage: 1.1+ MB
```

```
a=df.dept.unique()
print(a)

df['salary'].replace(['low','medium','high'],[0,1,2] ,inplace=True)
#df['dept'].replace(['sales', 'accounting', 'hr', 'technical', 'support', 'IT', 'product_mng','marketing', 'management', 'RandD'],[1,2,3
df
```

```
['sales' 'accounting' 'hr' 'technical' 'support' 'IT' 'product_mng'
 'marketing' 'management' 'RandD']
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotionInLast5
0	0.38	0.53	2	157	3	0	1	
1	0.80	0.86	5	262	6	0	1	
2	0.11	0.88	7	272	4	0	1	
3	0.37	0.52	2	159	3	0	1	
4	0.41	0.50	2	153	3	0	1	
...	...	...	...	...	...	...	...	
14994	0.11	0.85	7	275	4	0	1	
14995	0.99	0.83	4	274	2	0	0	
14996	0.72	0.72	4	175	4	0	0	
14997	0.24	0.91	5	177	5	0	0	
14998	0.77	0.83	6	271	3	0	0	

14999 rows × 10 columns

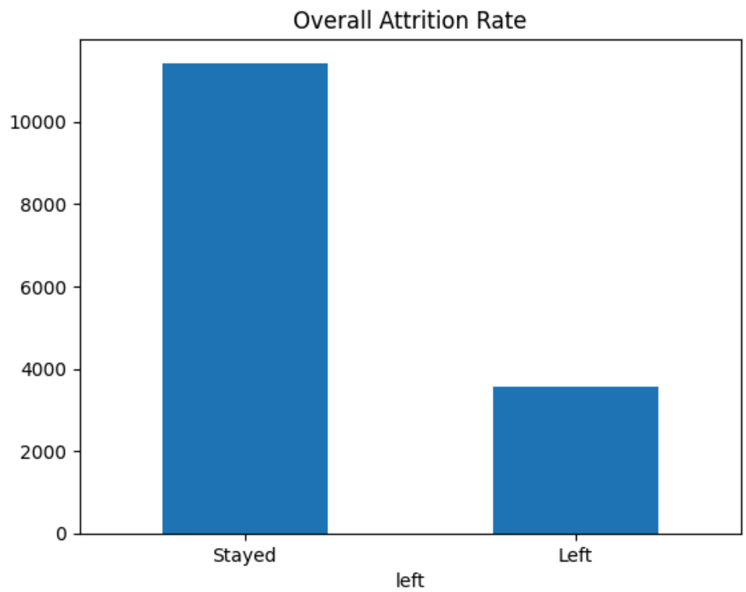
Start coding or generate with AI.

```
df['left'].value_counts()
#3571 employees left the job
```

	count
left	
0	11428
1	3571

dtype: int64

```
df['left'].value_counts().plot(kind='bar')
plt.title('Overall Attrition Rate')
plt.xticks([0, 1], ['Stayed', 'Left'], rotation=0)
plt.show()
```

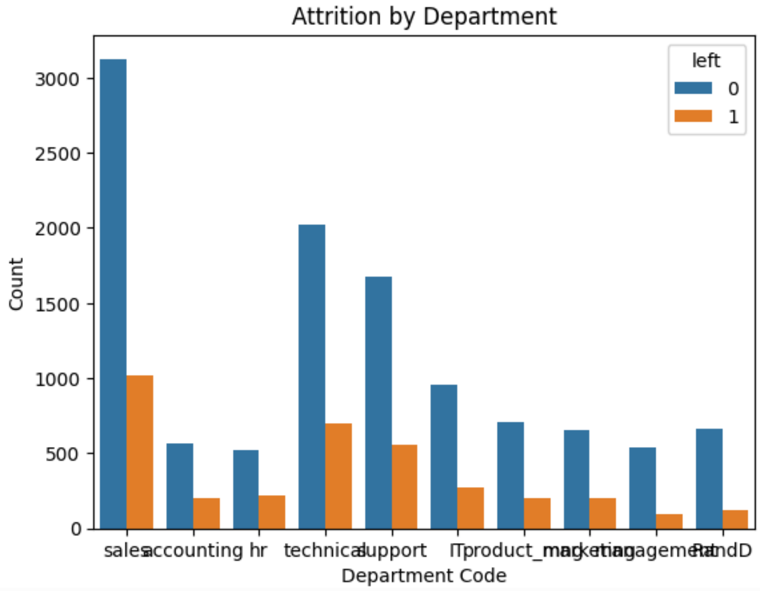


```
##INSIGHTS
Out of 14999 employees around 11428 employees stayed in company and around 3571 employees left the job(23.80%)
```

INSIGHTS

Out of 14999 employees around 11428 employees stayed in company and around 3571 employees left the job(23.80%)

```
sns.countplot(x='dept', hue='left', data=df)
plt.title("Attrition by Department")
plt.xlabel("Department Code")
plt.ylabel("Count")
plt.show()
```



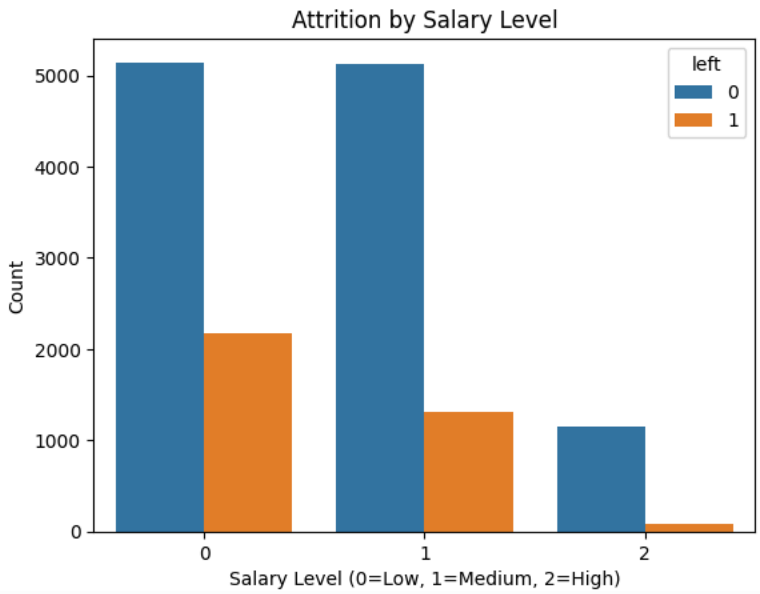
INSIGHTS

1.The attrition rate is the highest for sales department and its around 1000.

2.Attrition rate for technical department is 750

3.Attrition for suppoort department is 600 4.attrition for other departments is less than 250

```
sns.countplot(x='salary', hue='left', data=df)
plt.title("Attrition by Salary Level")
plt.xlabel("Salary Level (0=Low, 1=Medium, 2=High)")
plt.ylabel("Count")
plt.show()
```



INSIGHT

1.Around 2000 emplyees left the job who had low salary

2.Around 1300 employees left the job who had medium salary

3.Employees who had high salary and left the job were negligible

```
b=df.numberOfProjects.unique()
print(b)
```

```
[2 5 7 6 4 3]
```

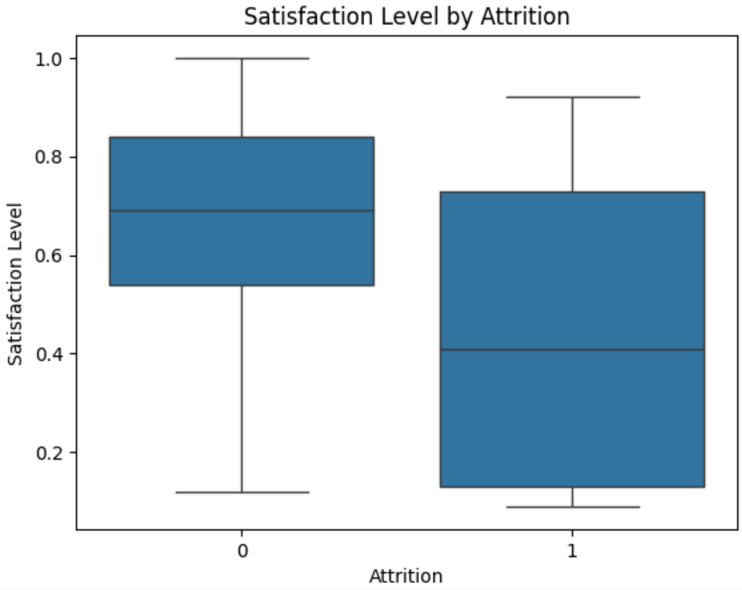
```
df['few_projects'] = df['numberOfProjects'] < 3
sns.countplot(x='few_projects', hue='left', data=df)
plt.title("Attrition for Employees with < 3 Projects")
plt.xlabel("Projects < 3")
plt.ylabel("Count")
plt.show()
```

INSIGHTS

The chart shows that employees with fewer than 3 projects (True) are more likely to leave the company than stay. In this group, the number of employees who left is greater than those who stayed. In contrast, among employees with 3 or more projects (False), most have stayed, and fewer have left.

This suggests that low project involvement may contribute to disengagement and attrition. Employees with insufficient responsibilities might feel underutilized or disconnected, prompting them to exit the organization.

```
sns.boxplot(x='left', y='satisfactoryLevel', data=df)
plt.title("Satisfaction Level by Attrition")
plt.xlabel("Attrition")
plt.ylabel("Satisfaction Level")
plt.show()
```



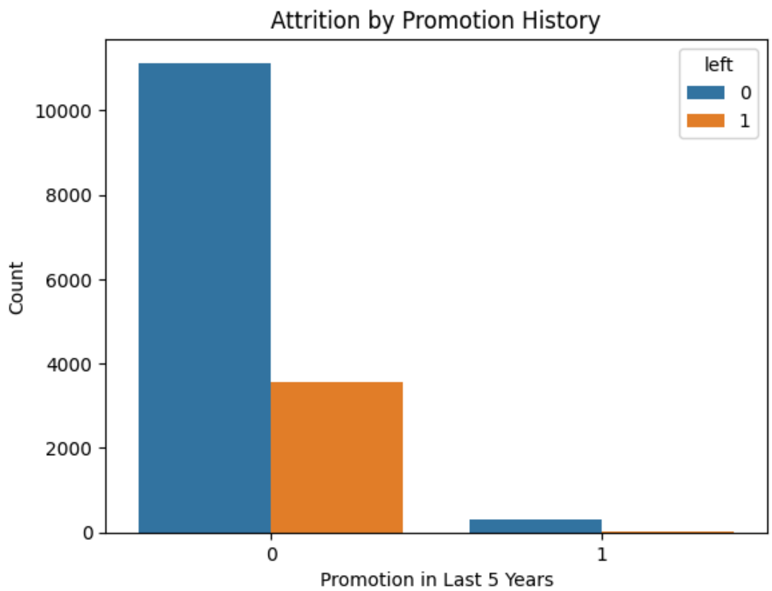
INSIGHTS

1.The boxplot clearly shows that employees who left the company tend to have significantly lower satisfaction levels compared to those who stayed.

2.The median satisfaction level among employees who left is approximately 0.4, while it is 0.7 for those who remained.

3.This suggests that satisfaction level is a strong predictor of attrition, and efforts to improve employee satisfaction could directly reduce turnover

```
sns.countplot(x='promotionInLast5years', hue='left', data=df)
plt.title("Attrition by Promotion History")
plt.xlabel("Promotion in Last 5 Years")
plt.ylabel("Count")
plt.show()
```



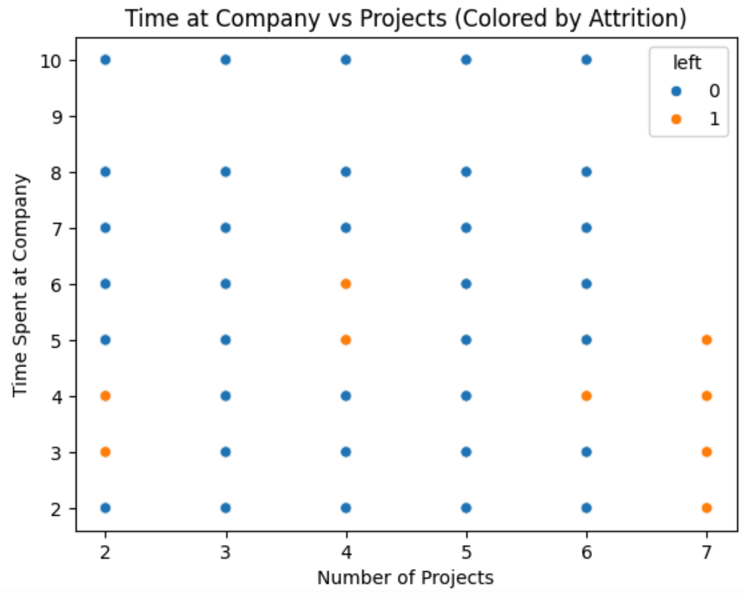
INSIGHTS

1.Employees who didnt get promotion in last 5 years and left the job are nearly 4000.

2.Employees who get promotion in last 5 years and left the job is 0.

```
sns.scatterplot(x='numberOfProjects', y='timeSpent.company', hue='left', data=df)
plt.title("Time at Company vs Projects (Colored by Attrition)")
plt.xlabel("Number of Projects")
plt.ylabel("Time Spent at Company")
```

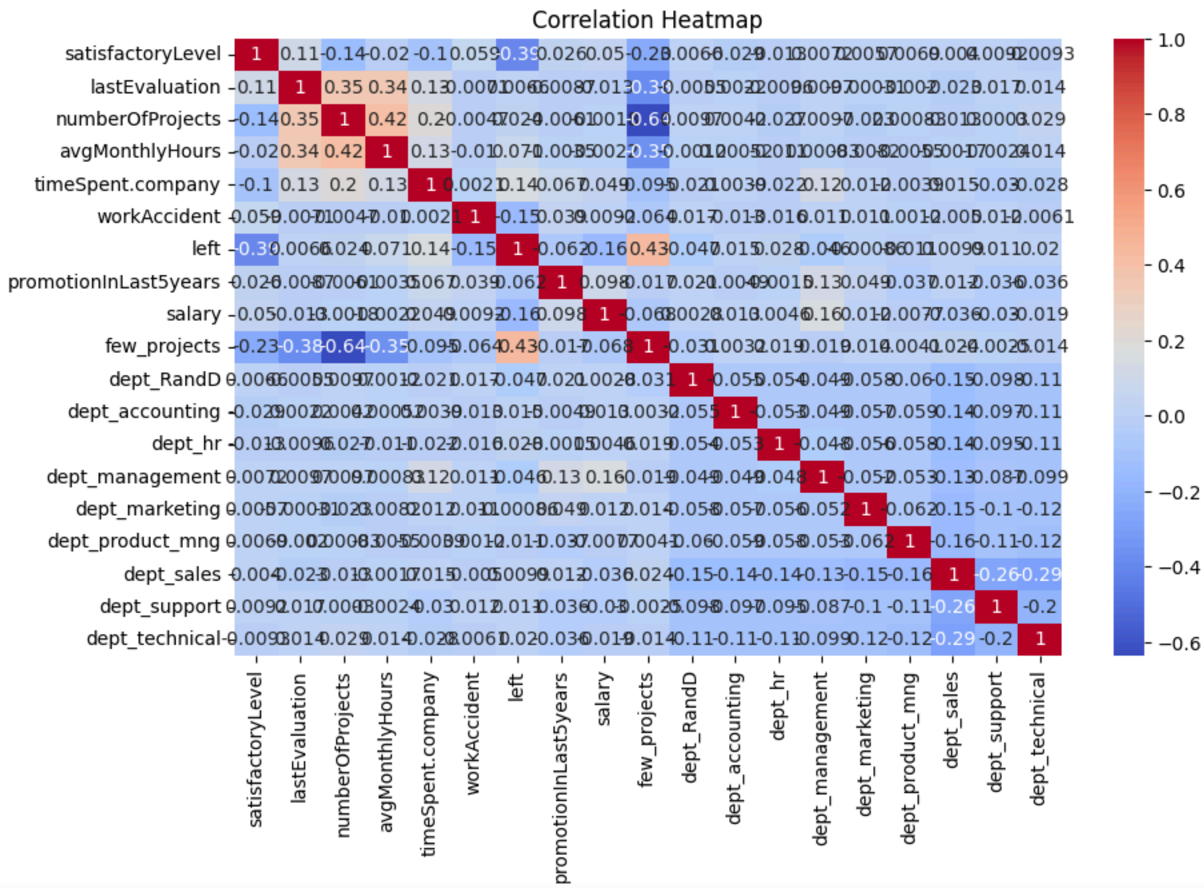
```
plt.show()
```



INSIGHT

The scatter plot shows that attrition is higher among employees who are either overloaded (working on 6–7 projects) or possibly under-engaged (2 projects). Additionally, employees who have spent 2 to 6 years at the company are more likely to leave, suggesting that mid-tenure is a critical period for retention. Long-tenure employees (7+ years) show significantly lower attrition.

```
plt.figure(figsize=(10, 6)) # Increase the figure size
df_encoded = pd.get_dummies(df, columns=['dept'], drop_first=True)
sns.heatmap(df_encoded.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```



INSIGHTS

1.satisfactoryLevel vs left: −0.39 Employees with lower satisfaction are more likely to leave. This aligns with your earlier boxplot and supports retention through satisfaction improvement.

2.few_projects vs left: +0.43

Employees with fewer than 3 projects are more likely to leave. Under-engagement may be a major churn driver.

3.numberOfProjects vs left: +0.24

Surprisingly, there's also a mild positive correlation with number of projects — indicating that both too few and too many projects might be risky

4.avgMonthlyHours vs left: +0.20

Higher working hours also correlate with higher attrition — potential burnout risk for overloaded employees.

5.salary vs left: −0.16

Lower salaries are weakly but negatively correlated with attrition — confirms earlier insights that low pay contributes to employee churn, though less strongly than satisfaction or project count.

6.promotionInLast5years vs left: −0.06 (very weak)

workAccident vs left: −0.15 Promotions and work accidents have very weak negative correlations with attrition. Their impact is minimal compared to satisfaction or project involvement.

Quantifying Attrition & Insights

Out of a total of 14999 employees, 3571 employees (~24%) have left the organization.

Through detailed analysis of various features, the following patterns were observed

1.Attrition is department-specific — Sales, Technical, and Support departments have higher exit rates compared to others.

2.Employees with low salary levels had significantly higher attrition, indicating dissatisfaction with compensation.

3.Employees working on fewer than 3 projects or more than 6 projects showed higher attrition. This suggests both under-engagement and burnout can drive exits.

4.Low Satisfaction is strongly correlated with attrition. Most employees who left had satisfaction levels below 0.45.

5.Lack of promotion in the last 5 years was a major factor — the vast majority of leavers had no recent promotions.

6.Mid-tenure employees (2–6 years) are more prone to leaving than long-tenure employees, who tend to stay.

7.The most significant predictor of attrition is low satisfaction level (r = -0.39).

8.Employees with <3projects are at high risk of leaving (r = +0.45)

9. Promotions and accidents have little linear impact on attrition.

Data-Driven Retention Strategies:

Based on the findings, the following targeted retention strategies are recommended:

1.Launch retention programs in high-attrition departments (Sales, Tech, Support) such as team engagement activities and manager check-ins.

2.Consider pairing salary increases with promotion/bonus cycles to retain talent

3.Avoid under-utilization (<3 projects) and overwork (>6 projects). Ensure fair distribution of workload.

4.Regularly assess employee satisfaction and intervene early for those below 0.5.

5.Provide clear growth paths and timely promotions to reduce stagnation-based exits.

6.Employees in the 2–6 year range should be given focused attention, including mentorship and re-engagement activities.