Minor Project – HR Analytics

Student Name: Eekshith Sai kandukuri
Organization: EdiGlobe
Submission Date: 10th July 2025

Objective

SimpleYard, a growing organization, is facing high employee attrition, which leads to operational challenges and increased hiring costs. In this project, we use HR analytics to explore employee data and uncover key reasons behind attrition using visualizations and statistics.

```
#uploading the CSV file
from google.colab import files
uploaded = files.upload()
```

```
Choose Files employe.csv
• employe.csv(text/csv) - 566770 bytes, last modified: 7/10/2025 - 100% done
Saving employe.csv to employe.csv
```

```
# Importing required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Loading dataset
df = pd.read_csv("employe.csv")

# Display first 5 rows
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	

Next steps: Generate code with df | View recommended plots | New interactive sheet

Generate print hello world using rot13 Close

```
# Shape of dataset
print("Rows and Columns:", df.shape)

# Column names
print("\nColumn Names:\n", df.columns)

# Info about dataset
df.info()

# Check for null values
print("\nNull Values:\n", df.isnull().sum())
```

```
Rows and Columns: (14999, 10)

Column Names:
 Index(['satisfactoryLevel', 'lastEvaluation', 'numberOfProjects',
       'avgMonthlyHours', 'timeSpent.company', 'workAccident', 'left',
       'promotionInLast5years', 'dept', 'salary'],
      dtype='object')
<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
```

What can I help you build?

```
 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

Null Values:
 satisfactoryLevel        0
lastEvaluation           0
numberOfProjects         0
avgMonthlyHours          0
timeSpent.company        0
workAccident             0
left                     0
promotionInLast5years    0
dept                     0
salary                   0
dtype: int64
```

Q1: What is the current workforce size, and how many employees have already left the organization?

Generate | randomly select 5 items from a list | Close

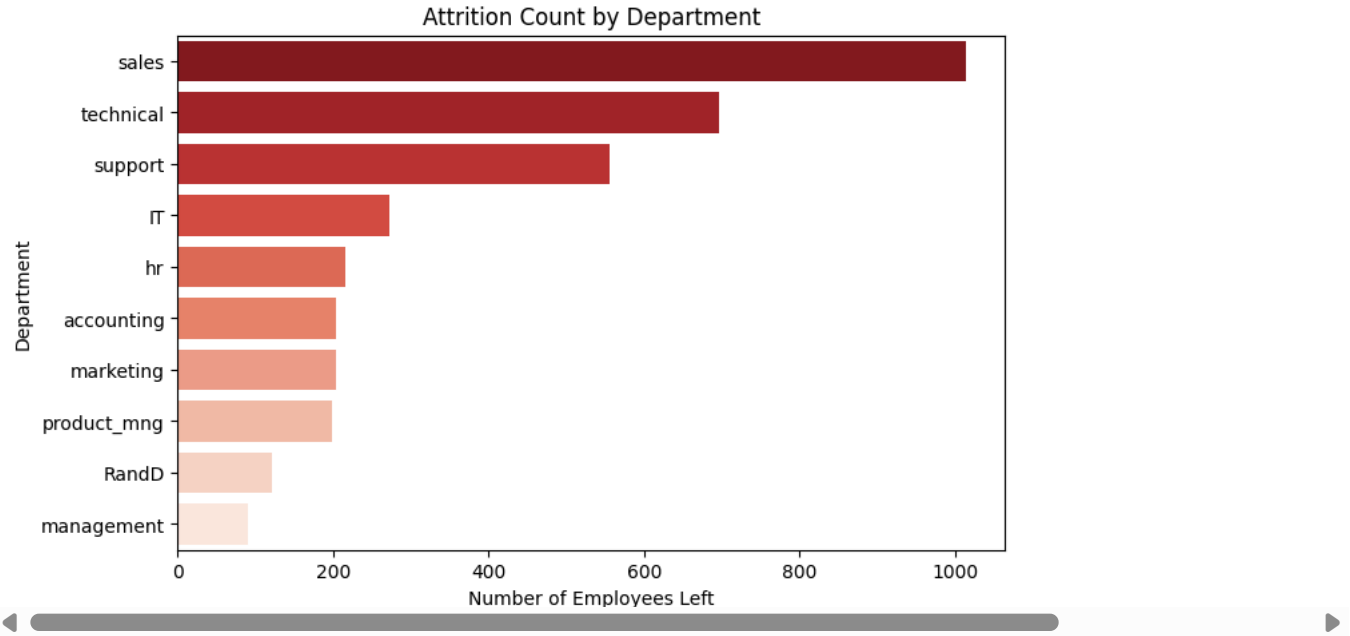
```
# Total employees and those who left
total_employees = df.shape[0]
left_employees = df[df['left'] == 1].shape[0]

print(f"Total Employees: {total_employees}")
print(f"Employees who left: {left_employees}")
```

```
Total Employees: 14999
Employees who left: 3571
```

Q2: Which departments are experiencing the highest rates of attrition?

```
# Employees who left, grouped by department
dept_attrition = df[df['left'] == 1]['dept'].value_counts()

# Plot
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(8,5))
sns.barplot(x=dept_attrition.values, y=dept_attrition.index, palette='Reds_r')
plt.title("Attrition Count by Department")
plt.xlabel("Number of Employees Left")
plt.ylabel("Department")
plt.show()
```

```
/tmp/ipython-input-9-3536703683.py:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `le

  sns.barplot(x=dept_attrition.values, y=dept_attrition.index, palette='Reds_r')
```



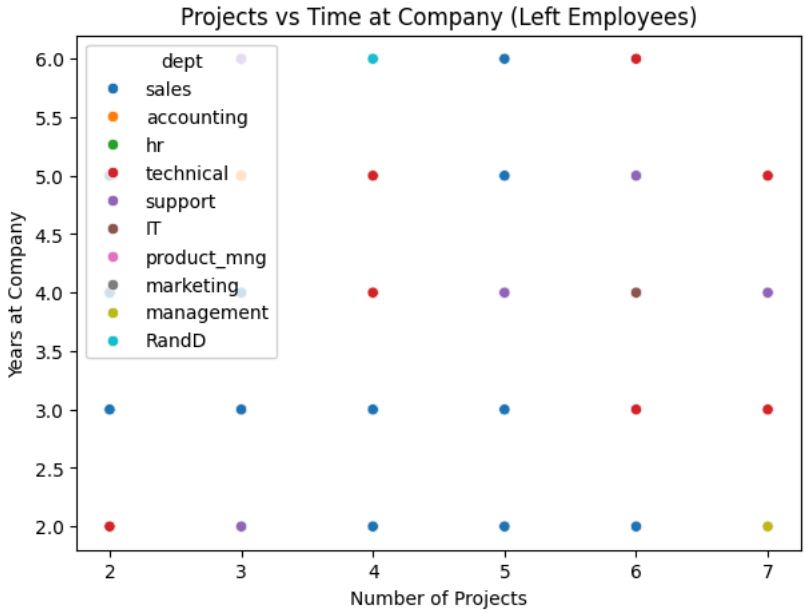
Q3: Are employees working on fewer than 3 projects more likely to leave the company?

```
# Employees with fewer than 3 projects
less_than_3 = df[df['numberOfProjects'] < 3]

# Attrition rate
attrition_rate = (less_than_3['left'] == 1).mean() * 100
print(f"Attrition rate for employees with <3 projects: {attrition_rate:.2f}%")
```

```
Attrition rate for employees with <3 projects: 65.62%
```

Q4: How does the number of projects correlate with time spent at the company, particularly for those who have left?

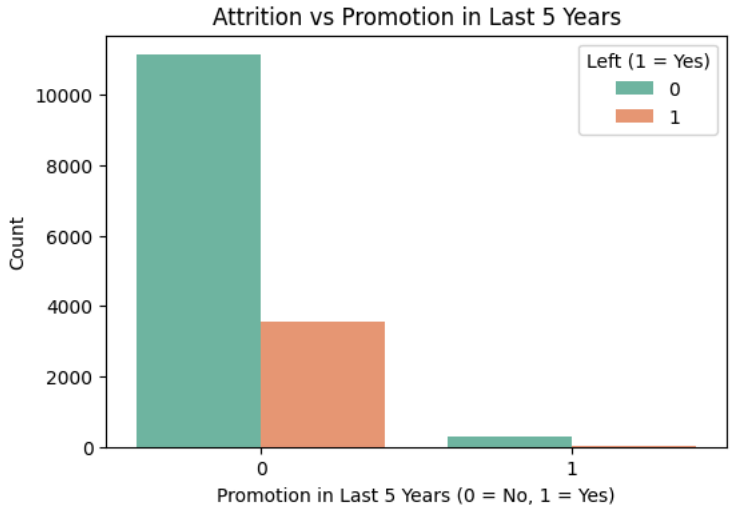
```
# Filter those who left
left_df = df[df['left'] == 1]

# Plot
plt.figure(figsize=(7,5))
sns.scatterplot(x='numberOfProjects', y='timeSpent.company', data=left_df, hue='dept')
plt.title("Projects vs Time at Company (Left Employees)")
plt.xlabel("Number of Projects")
plt.ylabel("Years at Company")
plt.show()
```



Q5: Could compensation levels be influencing an employee's decision to leave?

```
# Salary levels vs attrition count
plt.figure(figsize=(8,5))
sns.countplot(x='salary', hue='left', data=df, palette='coolwarm')
plt.title("Attrition by Salary Level")
plt.xlabel("Salary Level")
plt.ylabel("Count")
plt.legend(title='Left (1 = Yes)')
plt.show()
```


```
# Promotion vs Attrition
plt.figure(figsize=(6,4))
sns.countplot(x='promotionInLast5years', hue='left', data=df, palette='Set2')
plt.title("Attrition vs Promotion in Last 5 Years")
plt.xlabel("Promotion in Last 5 Years (0 = No, 1 = Yes)")
plt.ylabel("Count")
plt.legend(title='Left (1 = Yes)')
plt.show()
```



Conclusion & Key Insights