Very Good Evening Everyone!

I hope all of you are doing absolutely great 🙌 and are ready for another exciting learning session! 🚀

📊 Today's agenda: We'll be diving into a real-world HR Analytics Case Study 💼 — it's going to be super insightful, practical, and interesting 🔍 📈.

💡 I hope all of you are as excited as I am to explore and learn together! 🤩

💬 Please make sure to be interactive in the chat box — share your thoughts, ideas, and questions 🧠 💭. ❓ If you have any doubts, don't hesitate to ask — I'm here to help you throughout the session 🤝 ✨

⏳ Let's wait for a few more minutes ⏱️ so that everyone can join, and then we'll start the session officially 🎓 🔥.

Get ready for an engaging and insightful learning experience! 🚀 📊 💼

```
# 📊 HR Analytics Dataset Column Descriptions

# satisfactoryLevel          -> Employee's job satisfaction score (0 to 1 scale)
# lastEvaluation             -> Most recent performance evaluation score (0 to 1 scale)
# numberOfProjects           -> Total number of projects the employee has worked on
# avgMonthlyHours            -> Average monthly working hours of the employee
# timeSpent.company          -> Number of years the employee has been in the company
# workAccident               -> Whether the employee had a work accident (1 = Yes, 0 = No)
# left                       -> Whether the employee left the company (1 = Yes, 0 = No)
# promotionInLast5years      -> Whether the employee got a promotion in the last 5 years (1 = Yes, 0 = No)
# dept                       -> Department the employee belongs to (e.g., sales, IT, HR, etc.)
# salary                     -> Salary level of the employee (low, medium, high)
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("/content/people.csv")
df
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotionInLast
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

```
#head
#tail
#info
#describe
#null values - how
#duplicate values - how
#outliers - in which cols ?

#8:42pm
```

```
df.head()
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotionInLast5yea
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.tail()
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotionInLast
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	

```
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
```

```
df.describe()
```

	satisfactoryLevel	lastEvaluation	numberOfProjects	avgMonthlyHours	timeSpent.company	workAccident	left	promotic
count	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	
mean	0.612834	0.716102	3.803054	201.050337	3.498233	0.144610	0.238083	
std	0.248631	0.171169	1.232592	49.943099	1.460136	0.351719	0.425924	
min	0.090000	0.360000	2.000000	96.000000	2.000000	0.000000	0.000000	
25%	0.440000	0.560000	3.000000	156.000000	3.000000	0.000000	0.000000	
50%	0.640000	0.720000	4.000000	200.000000	3.000000	0.000000	0.000000	
75%	0.820000	0.870000	5.000000	245.000000	4.000000	0.000000	0.000000	
max	1.000000	1.000000	7.000000	310.000000	10.000000	1.000000	1.000000	

```
#null values
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.isnull().sum().sum()
```

```
np.int64(0)
```

```
#duplicate
df.duplicated().sum()
```

```
np.int64(3008)
```

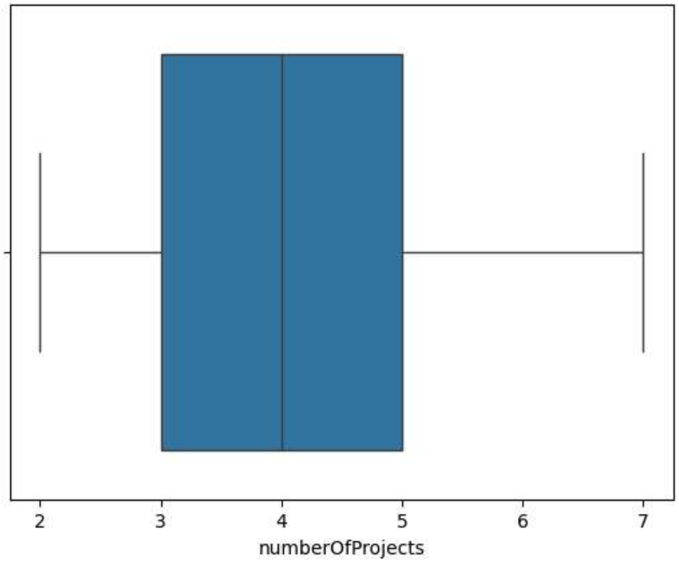
```
#remove the duplicate values

df = df.drop_duplicates()
```

```
df.duplicated().sum()
```

```
np.int64(0)
```

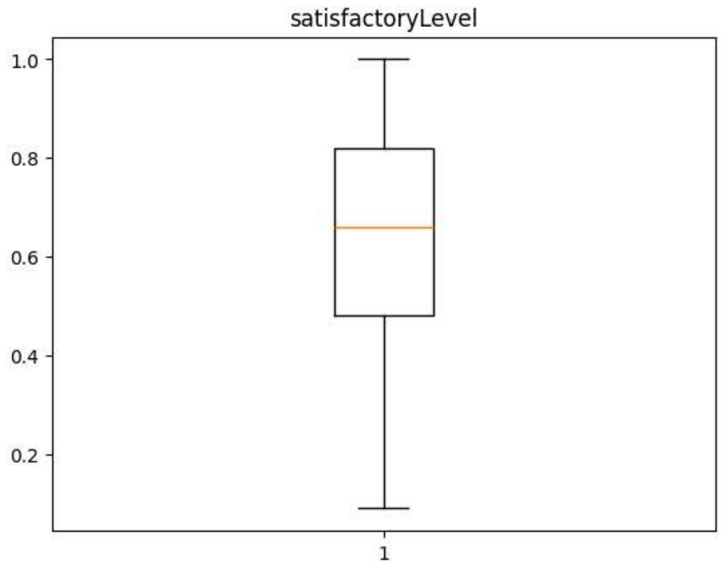
```
sns.boxplot(data = df, x ='numberOfProjects')
plt.show()
```

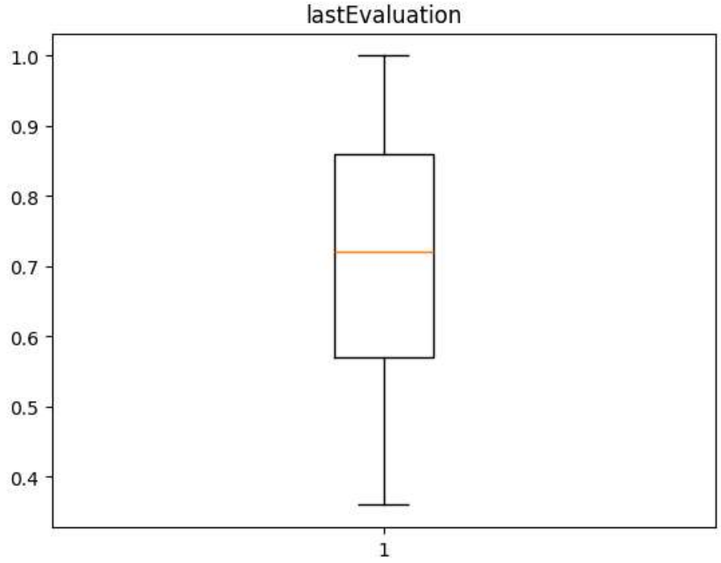


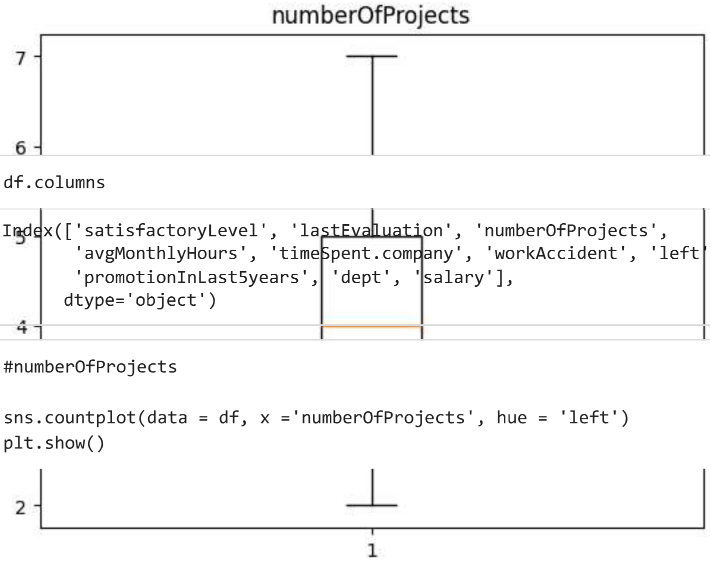
```
df.columns
```

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

```
for col in df.columns:
  if df[col].dtype !='object':
    plt.boxplot(df[col])
    plt.title(col)
    plt.show()
```





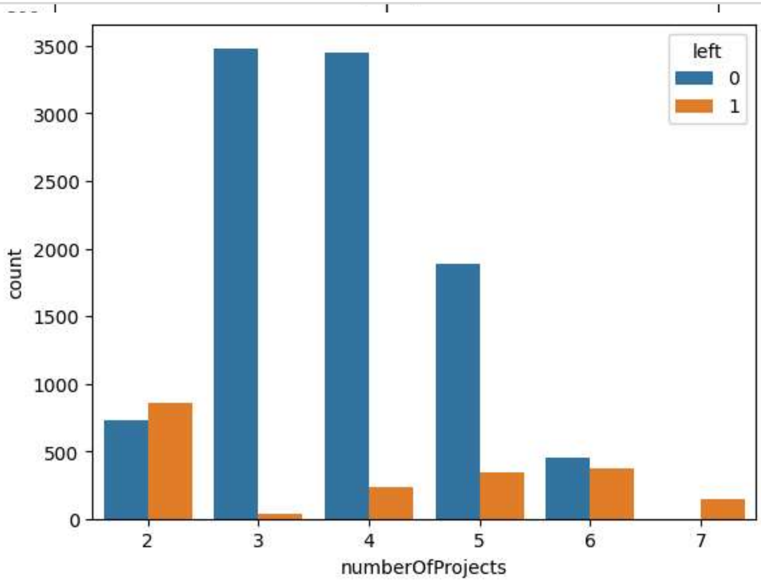


```
df.columns
```

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

```
#numberOfProjects

sns.countplot(data = df, x ='numberOfProjects', hue = 'left')
plt.show()
```



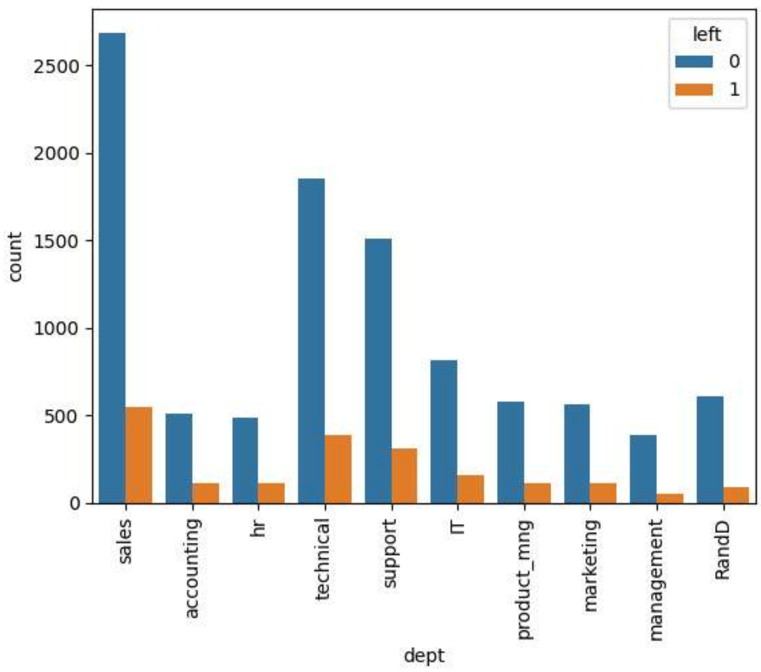
Analysis

- people who are woking in 2,5,6,7 projects are leaving the most

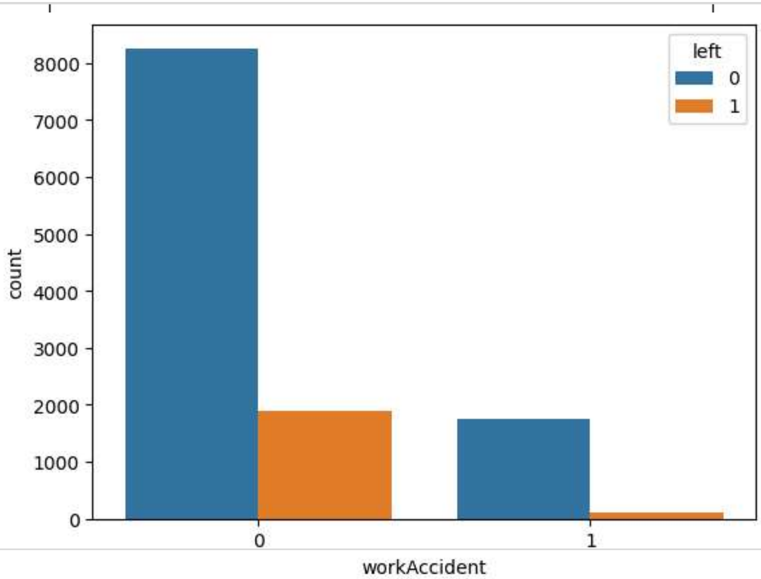
suggestion to hr

distribute ythe projects in such a way a people shouldnt get less proects and fo few it should not be burden peope who are wokeing only in 2 projects are levaung becoiz theyare not getting oppurtunity to showcase the skills and people who are working in 5,6,&7 are leavung the most due to multiple projects stress so givre projects in a balanced way and people who are woking in multiple projects give them salary hikes an dsome bonus

```
sns.countplot(data = df, x ='dept', hue = 'left')
plt.xticks(rotation = 90)
plt.show()
```



```
sns.countplot(data = df, x ='workAccident', hue = 'left')
plt.show()
```



work accident is not the reson for the employess to leave

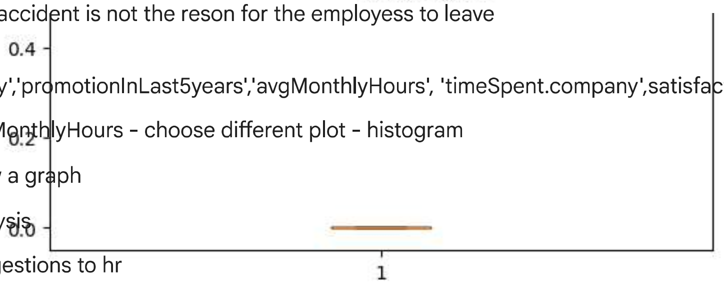
'salary','promotionInLast5years','avgMonthlyHours', 'timeSpent.company',satisfactoryLevel

-avgMonthlyHours - choose different plot - histogram

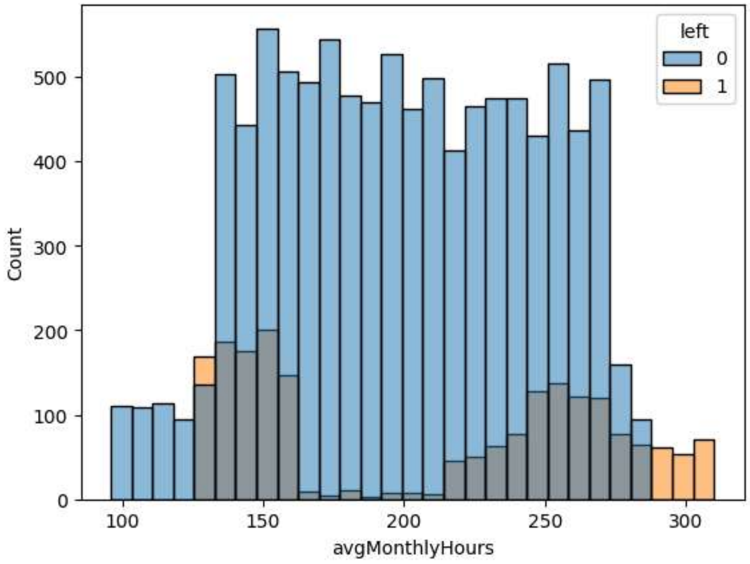
-draw a graph

-analysis

-suggestions to hr



```
sns.histplot(data = df, x ='avgMonthlyHours', hue = 'left')
plt.show()
```


left
Count
avgMonthlyHours