

```
import pandas as pd
import numpy as np
import seaborn as sb
import matplotlib.pyplot as plt
import seaborn as sb
from google.colab import drive

drive.mount('/content/drive')

path="/content/drive/MyDrive/DVA/Employee-Attrition.csv"
df=pd.read_csv(path)
df.head(5)
```

↗

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	.
0	41	Yes	Travel_Rarely	1102	Sales		1	2	Life Sciences	1	1
1	49	No	Travel_Frequently	279	Research & Development		8	1	Life Sciences	1	2
2	37	Yes	Travel_Rarely	1373	Research & Development		2	2	Other	1	4
3	33	No	Travel_Frequently	1392	Research & Development		3	4	Life Sciences	1	5
4	27	No	Travel_Rarely	591	Research & Development		2	1	Medical	1	7

5 rows × 35 columns

```
df.shape
```

↗

(1470, 35)

```
#checking duplicates
print(f"NO. of duplicate data {df.duplicated().sum()}")

#checking for null values
print(f"NO. of null values {df.isnull().sum().sum()}")

#checking data types
print(f"\n Data types {df.dtypes}")
```

↗

NO. of duplicate data 0  
NO. of null values 0

Data types

Age	int64
Attrition	object
BusinessTravel	object
DailyRate	int64
Department	object
DistanceFromHome	int64
Education	int64
EducationField	object
EmployeeCount	int64
EmployeeNumber	int64
EnvironmentSatisfaction	int64
Gender	object
HourlyRate	int64
JobInvolvement	int64
JobLevel	int64
JobRole	object
JobSatisfaction	int64
MaritalStatus	object
MonthlyIncome	int64
MonthlyRate	int64
NumCompaniesWorked	int64
Over18	object
OverTime	object
PercentSalaryHike	int64
PerformanceRating	int64
RelationshipSatisfaction	int64
StandardHours	int64
StockOptionLevel	int64
TotalWorkingYears	int64

```
TrainingTimesLastYear      int64
WorkLifeBalance            int64
YearsAtCompany             int64
YearsInCurrentRole         int64
YearsSinceLastPromotion    int64
YearsWithCurrManager       int64
dtype: object
```

```
#describe data
summary=df.describe()
summary.T
```

	count	mean	std	min	25%	50%	75%	max
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0	43.00	60.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0	4.00	4.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	2.0	3.00	5.0
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4919.0	8379.00	19999.0
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14235.5	20461.50	26999.0
NumCompaniesWorked	1470.0	2.693197	2.498009	0.0	1.00	2.0	4.00	9.0
PercentSalaryHike	1470.0	15.209524	3.659938	11.0	12.00	14.0	18.00	25.0
PerformanceRating	1470.0	3.153741	0.360824	3.0	3.00	3.0	3.00	4.0
RelationshipSatisfaction	1470.0	2.712245	1.081209	1.0	2.00	3.0	4.00	4.0
StandardHours	1470.0	80.000000	0.000000	80.0	80.00	80.0	80.00	80.0
StockOptionLevel	1470.0	0.793878	0.852077	0.0	0.00	1.0	1.00	3.0
TotalWorkingYears	1470.0	11.279592	7.780782	0.0	6.00	10.0	15.00	40.0
TrainingTimesLastYear	1470.0	2.799320	1.289271	0.0	2.00	3.0	3.00	6.0
WorkLifeBalance	1470.0	2.761224	0.706476	1.0	2.00	3.0	3.00	4.0
YearsAtCompany	1470.0	7.008163	6.126525	0.0	3.00	5.0	9.00	40.0
YearsInCurrentRole	1470.0	4.229252	3.623137	0.0	2.00	3.0	7.00	18.0
YearsSinceLastPromotion	1470.0	2.187755	3.222430	0.0	0.00	1.0	3.00	15.0
YearsWithCurrManager	1470.0	4.123129	3.568136	0.0	2.00	3.0	7.00	17.0

```
df['Attrition'].value_counts(normalize=True)
```

	proportion
Attrition	
No	0.838776
Yes	0.161224

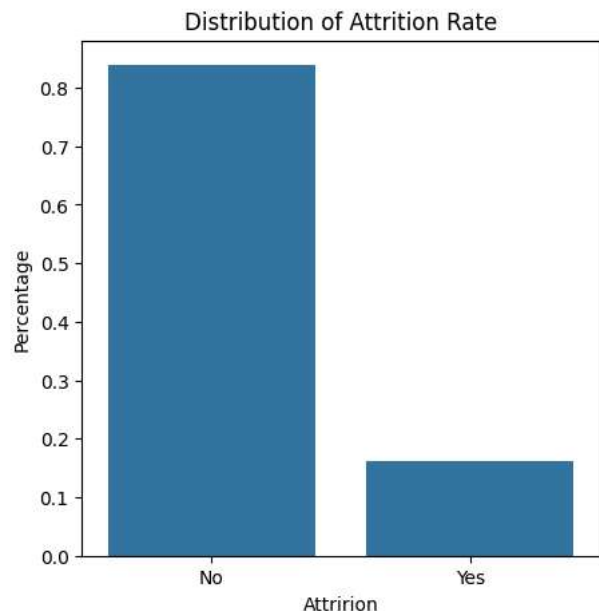
```
#need to work on annotations
```

```
attrition = df['Attrition'].value_counts(normalize=True)
print(attrition)

plt.figure(figsize=(5,5))
sb.barplot(x=attrition.index, y=attrition.values)
```

```
plt.xlabel('Attririon')
plt.ylabel('Percentage')
plt.title('Distribution of Attrition Rate')
plt.show()
```

```
Attrition
No      0.838776
Yes     0.161224
Name: proportion, dtype: float64
```



Based on the analysis, the company's attrition rate is 16.12%. This means that about 16.12% of the employees decided to leave the company during the analyzed period.

Average of Tenure Average tenure: The average tenure measures the average number of years an employee stays with the company before leaving.

It can provide insight into workforce stability and employee satisfaction within the organization.

```
avg_tenure=df['YearsAtCompany'].mean()
print(f"Average years of employee to leave the company is {avg_tenure:.2f} years")
```

```
Average years of employee to leave the company is 7.01 years
```

The average tenure of employees before they decided to leave was 7.01 years. With this average tenure, it can be concluded that many employees feel comfortable and have been with the company for a long time.

#Employee's Demographics

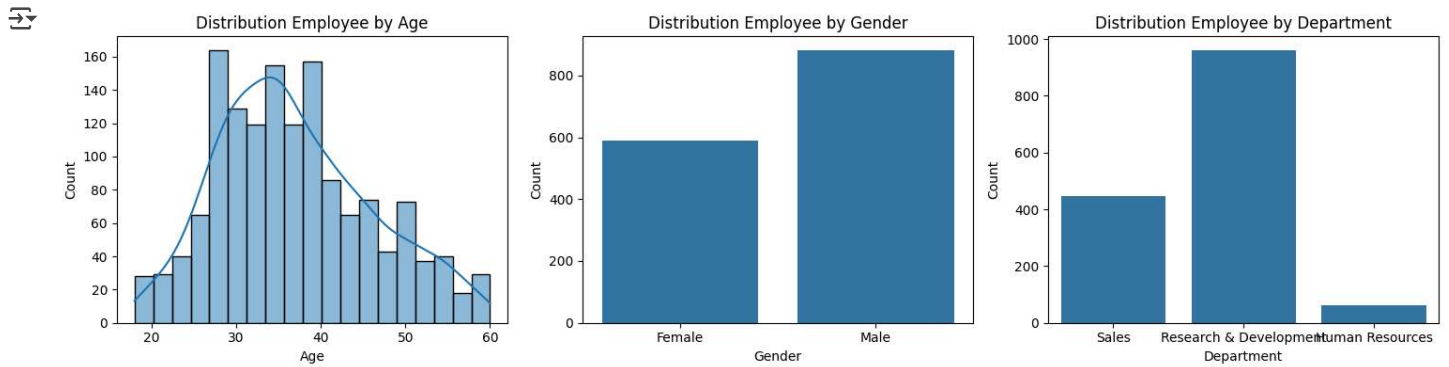
```
fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(15,4))
```

```
sb.histplot(data=df, x='Age', kde=True, ax=axes[0])
axes[0].set_title('Distribution Employee by Age')
axes[0].set_xlabel('Age')
axes[0].set_ylabel('Count')
```

```
sb.countplot(data=df, x='Gender', ax=axes[1])
axes[1].set_title('Distribution Employee by Gender')
axes[1].set_xlabel('Gender')
axes[1].set_ylabel('Count')
```

```
sb.countplot(data=df, x='Department', ax=axes[2])
axes[2].set_title('Distribution Employee by Department')
axes[2].set_xlabel('Department')
axes[2].set_ylabel('Count')
```

```
plt.tight_layout()
plt.show()
```



1. Age: Most of the company's employees are in the 30-35 age group. This indicates that the company has many employees who are at a productive and experienced age.
2. Gender: The majority of employees at this company are male. There are significantly more male employees than female employees.
3. Department: Most of the company's employees are concentrated in the research and development department. This indicates that the company is heavily focused on product or service research and development activities.

```
df_attrition = df[df['Attrition'] == 'Yes']
df_attrition.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	.
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	
14	28	Yes	Travel_Rarely	103	Research & Development	24	3	Life Sciences	1	19	
21	36	Yes	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	
24	34	Yes	Travel_Rarely	699	Research & Development	6	1	Medical	1	31	

5 rows × 35 columns

```
def calculate_attrition_rate(df, column):
    attrition_counts = df.groupby([column, 'Attrition']).size().unstack(fill_value=0)
    attrition_rate = attrition_counts['Yes']/attrition_counts.sum(axis=1) * 100
    attrition_rate_df = attrition_rate.reset_index()
    attrition_rate_df.columns = [column, 'AttritionRate']
    print(attrition_rate)
    return attrition_rate_df
```

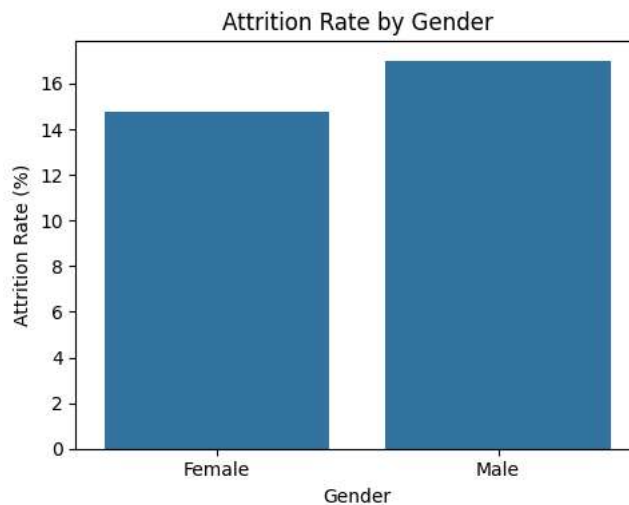
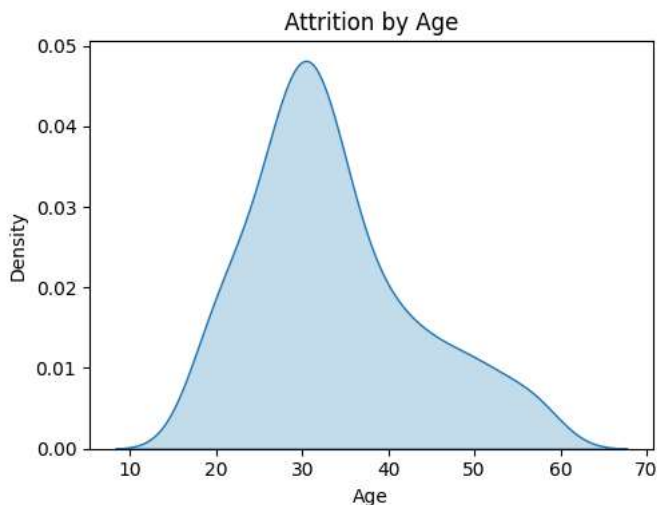
```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(10,4))
```

```
# Plot 1: KDE plot of Age with Attrition hue
sb.kdeplot(data=df_attrition, x='Age', fill=True, ax=axes[0])
axes[0].set_title('Attrition by Age')
axes[0].set_xlabel('Age')
axes[0].set_ylabel('Density')
```

```
# Plot 2: Bar plot of Gender count with Attrition hue
attrition_rate_df = calculate_attrition_rate(df, 'Gender')
sb.barplot(data=attrition_rate_df, x='Gender',
y='AttritionRate', ax=axes[1])
axes[1].set_title(f'Attrition Rate by Gender')
axes[1].set_xlabel('Gender')
axes[1].set_ylabel('Attrition Rate (%)')
```

```
plt.tight_layout()
plt.show()
```

```
Gender
Female    14.795918
Male      17.006803
dtype: float64
```



```
!git clone https://github.com/Saish0619/IBM-HR-Attrition.git
```

```
fatal: destination path 'IBM-HR-Attrition' already exists and is not an empty directory.
```

```
!cp "/content/drive/MyDrive/Colab Notebooks/IBM-HR-DA.ipynb" /content/IBM-HR-Attrition
```

```
%cd /content/IBM-HR-Attrition
```

```
/content/IBM-HR-Attrition
```

```
!git config --global user.email "saishgaonkar06@gmail.com"
```

```
!git config --global user.name "Saish0619"
```

```
!git add IBM-HR-DA.ipynb
```

```
!git commit -m "Notebook Added"
```

```
[main (root-commit) 731e3be] Notebook Added
1 file changed, 1 insertion(+)
create mode 100644 IBM-HR-DA.ipynb
```

```
!git push https://Saish0619:ghp_H3NeoqvD3PIcxPU114QaHNxwLLh39r2s01zN@github.com/Saish0619/IBM-HR-Attrition.git
```

```
Enumerating objects: 3, done.
Counting objects: 100% (3/3), done.
Delta compression using up to 2 threads
Compressing objects: 100% (2/2), done.
Writing objects: 100% (3/3), 102.83 KiB | 8.57 MiB/s, done.
Total 3 (delta 0), reused 0 (delta 0), pack-reused 0
To https://github.com/Saish0619/IBM-HR-Attrition.git
* [new branch]    main -> main
```

