



Project Title	IBM HR Analytics Employee Attrition & Performance
Tools	Python, ML, SQL, Excel
Domain	Data Analyst & Data scientist
Project Difficulties level	intermediate

Dataset : Dataset is available in the given link. You can download it at your convenience.

[Click here to download data set](#)

About Dataset

Uncover the factors that lead to employee attrition and explore important questions such as ‘show me a breakdown of distance from home by job role and attrition’ or ‘compare average monthly income by education and attrition’. This is a fictional data set created by IBM data scientists.

Education

1 'Below College'

2 'College'

3 'Bachelor'

4 'Master'

5 'Doctor'

EnvironmentSatisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

JobInvolvement

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

JobSatisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

PerformanceRating

1 'Low'

2 'Good'

3 'Excellent'

4 'Outstanding'

RelationshipSatisfaction

1 'Low'

2 'Medium'

3 'High'

4 'Very High'

WorkLifeBalance

1 'Bad'

2 'Good'

3 'Better'

4 'Best'

Example: You can get the basic idea how you can create a project from here

what steps you should have to follow

Example: You can get the basic idea how you can create a project from here

Sample code and output

HR Attrition Analysis¶

In the business world, companies often face the challenge of retaining talented employees. One of the most pressing issues is the increasing rate of employee turnover, commonly known as HR attrition. Turnover can have a significant impact on a company's productivity, stability, and long-term sustainability. High attrition rates

can lead to increased recruitment and training costs, disrupt team dynamics, and result in the loss of valuable institutional knowledge. Therefore, understanding the factors contributing to attrition and implementing effective retention strategies is crucial for maintaining a competitive edge and ensuring

Objectives of the Analysis

1. Understand Current Turnover Rates: Gain a comprehensive understanding of the current employee turnover rate and analyze the demographic distribution of attrition by age, gender, education, department, and job role.
2. Identify Key Factors Influencing Turnover: Examine the main factors contributing to employee attrition, including job satisfaction indicators (job involvement and work-life balance), salary factors (monthly income and salary hikes), and benefit factors (stock option levels), to uncover patterns and correlations that drive higher attrition rates.

Data Cleaning

In [1]:

```
# import data manipulation package
import pandas as pd
import numpy as np

# import data visualization package
import matplotlib.pyplot as plt
import seaborn as sns

# importing the warnings library
```

```
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
# set pandas options
pd.set_option('display.max_columns', 35)

# load dataset
df =
pd.read_csv('/kaggle/input/ibm-hr-analytics-attrition-dataset/W
A_Fn-UseC_-HR-Employee-Attrition.csv')
df.head()
```

Out[2]:

		A	B	D	D	D	E	E	E	E	E		H	J	J	J	M	M	M	N		C	P	P	R	S	S	T	T	W	Y		
		t	u	a	e	i	d	d	m	m	n		G	o	o	o	a	o	o	u		C	e	e	e	t	t	o	r	o	e		
A	t	s	i	p	s	u	u	p	p	v		e	u	b	b	b	r	n	n	m		v	e	r	r	l	a	o	t	a	r	a	
g	r	i	l	a	t	c	c	l	l	i	r		n	r	l	e	R	S	i	t	t	C		c	f	a	n	c	a	i	k	r	
e	i	n	y	r	a	a	a	o	o	o		d	l	n	v	e	o	a	t	h	h	o		1	n	r	o	a	O	W	i	A	
	t	e	R	t	n	t	t	y	y	n		e	y	v	e	i	t	a	l	l	m		8	t	m	n	r	p	o	n	f	t	
	i	s	a	n	c	i	i	e	e	m		R	o				i	l	y	y	p			S	a	s	d	t	r	g	e	C	
	o	s	t	e	e	o	o	e	e	e		a	l			s	S	I	R	a		e	S	a	s	d	t	r	g	e	C		

[illegible]

[illegible]

[illegible]


```
df.isnull().sum() / len(df) * 100
```

```
Out[5]:
```

Age	0.0
Attrition	0.0
BusinessTravel	0.0
DailyRate	0.0
Department	0.0
DistanceFromHome	0.0
Education	0.0
EducationField	0.0
EmployeeCount	0.0
EmployeeNumber	0.0
EnvironmentSatisfaction	0.0
Gender	0.0
HourlyRate	0.0
JobInvolvement	0.0
JobLevel	0.0
JobRole	0.0
JobSatisfaction	0.0
MaritalStatus	0.0
MonthlyIncome	0.0
MonthlyRate	0.0

NumCompaniesWorked	0.0
Over18	0.0
OverTime	0.0
PercentSalaryHike	0.0
PerformanceRating	0.0
RelationshipSatisfaction	0.0
StandardHours	0.0
StockOptionLevel	0.0
TotalWorkingYears	0.0
TrainingTimesLastYear	0.0
WorkLifeBalance	0.0
YearsAtCompany	0.0
YearsInCurrentRole	0.0
YearsSinceLastPromotion	0.0
YearsWithCurrManager	0.0

dtype: float64

There are any missing values in the dataset.

In [6]:

check data types

df.dtypes

Out[6]:

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

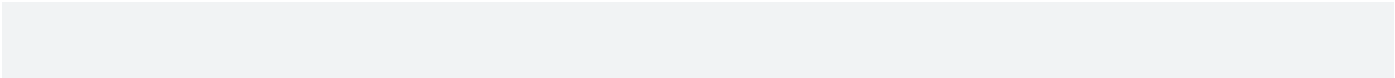
RelationshipSatisfaction	int64
StandardHours	int64
StockOptionLevel	int64
TotalWorkingYears	int64
TrainingTimesLastYear	int64
WorkLifeBalance	int64
YearsAtCompany	int64
YearsInCurrentRole	int64
YearsSinceLastPromotion	int64
YearsWithCurrManager	int64

dtype: object

All columns have appropriate data types, ensuring that the data is correctly formatted for analysis.

In [7]:

```
# check data describe
df.describe()
```



Out[7]:

	A	D	D	E	E	E	E	H	J	J	J	M	M	N	P	P	R	S	S	T	Tr	W	Y	Y	Y	Y
	g	a	i	d	m	m	n	o	o	o	o	o	o	u	e	e	el	t	t	o	ai	o	e	e	e	e
	e	i	s	u	p	p	vi	u	b	b	b	n	n	m	r	rf	at	a	o	t	ni	r	a	a	ar	ar
		l	t	c	l	l	ro	r	l	L	S	t	t	C	c	o	io	n	c	a	n	k	r	r	s	s

m e a n															s t d														
3	6	.	9	2	3	8	1	0							9	.	1	3	5	3	7	3							
8	0	.	4	9	8	5	7	1	4						4	0	.	1	0	6	1	0							
9	.	1	9	2	5	5	1	7							8	.	1	0	6	8	6	4							
2	.	9	1	2	9	2	2	5							1	.	0	2	4	1	6	5							
	1	.	0												0	.	0												
1	0	.	8	6	5	3	0	6							6	0	.	0	2	4	3	3							
2.	7	2	1	7	6	9									1.	0	9	3	0	8	2								
6	5	.	8	9	1	1	5	6							2	0	.	3	2	9	4	2							
2	.	7	2	9	9	3	4	6							0	.	7	1	6	5	6	1							
2	.	0	6	3	9	9	4	6							1	.	1	0	2	9	4	4							
6	5	.	9	3	1	2	9	3							4	7	.	0	5	6	7	8							
1	4	3	1	3	0	1	3	4	0	1					7	1	.	7	8	0	0	4							
2.	6	9	3	1	9	7									2.	4	9	8	0	0	9								
1	5	.	2	0	9	5	2	4							3	.	6	5	9	9	3	8							
3	.	1	5	3	7	4	1								0	.	3	6	0	8	2	4							
2.	7	1	2	2	4	5									1.	0	8	1	2	0	9								
	8	0	.	0											0	.	0												
0	.	7	9	3	8	7	8								0	.	8	5	2	0	7	7							
1	1	.	2	7	9	5	9	2							7	.	7	8	0	7	8	2							
2.	7	9	9	3	2	0									1.	2	8	9	2	7	1								
2	.	7	6	1	2	2	4								0	.	7	0	6	4	7	6							
7	.	0	0	8	1	6	3								6	.	1	2	6	5	2	5							
4	.	2	2	9	2	5	2								3	.	6	3	1	3	7								
2.	1	8	7	7	5	5									3.	2	2	2	4	3	0								
4.	1	2	3	1	2	9									3.	5	6	8	1	3	6								

[illegible]

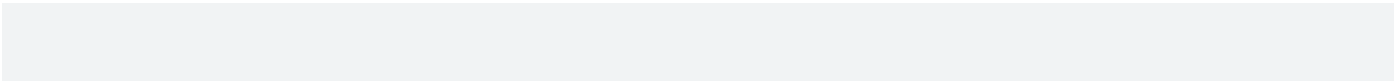
a	0	4	9	.	.	0	0	0	.	.	.	9	6	0	5	.	0	0	.	0	0	.	0	8	5.	7.
x	.	9	.	0	0	6	0	0	0	0	0	9	9	0	.	0	0	.	0	.	0	0	.	.	0	0
	0	9	0	0		8	0	.	0	0	0	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	.	0	0		.	0	0	0	0	0	9	9	0	0	0	0		0	0	0	0	0	0	0	0
	0	0	0	0		0	0	0	0	0	0	.	.	0	0	0	0		0	0	0	0	0	0	0	0
	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0
	0	0	0	0		0		0	0	0	0	0	0		0	0		0	0		0	0	0	0	0	0
	0	0	0			0		0	0	0	0	0	0		0	0		0	0		0	0	0	0	0	0
	0	0	0			0		0				0	0		0			0			0	0	0	0	0	0
		0				0		0				0	0		0			0			0	0		0		
		0				0		0				0	0					0			0	0		0		
						0						0	0								0	0				
						0						0	0													
						0						0	0													

Based on this summary, there are no apparent outliers in the dataset, as the values fall within expected ranges.

Exploratory Data Analysis

In [8]:

```
df.head()
```



Out[8]:

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

		i t i o n	n R a v e l	y R a m e n t	r a n c i e n t	a a t t i e e n t	o o y y e e n t	o o n n e e n t	e r y R o a l t v e m e n t	l y v l e e l	o a t i l s f a c t i o n	a t a l y S l a n c t i o n	h h o l m p R a n c i e s W o r k e d	1 8 i n t e n s i v e	T e n r m s a n c i p l a r y H i k e	o r m s h c i p R a t i s f a c t i o n	t i o n d a r p t H i o o u n s e v e l	k O W i n g T i m e s L a s t Y e a r	l W i n g T i m e s L a s t Y e a r	n i f e B o m b a n c y	L i f e C o m p a n y	S A t t e n t i o n						
0	4 1	Y e s	T r a v e l - R a	1 1 0 2	S a l e s	1	L i f e S c i e n	1	1	2	F e m a l e	9 4	S a l e s E x e c	4	S i n g l e	5 9 9 3	1 9 4 8	Y e s	1 1	3	1	8 0	0	0	8	0	1	6

[illegible]

[illegible]

[illegible]

dataset. Let's visualize it!

In [10]:

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

```
plt.figure(figsize=(8,6))
```

```
ax = sns.barplot(x=attrition.index, y=attrition)
```

```
for p in ax.patches:
```

```
    ax.annotate(f'{p.get_height() * 100:.2f}%',  
                (p.get_x() + p.get_width() / 2.,  
p.get_height()),
```

```
                ha='center', va='bottom')
```

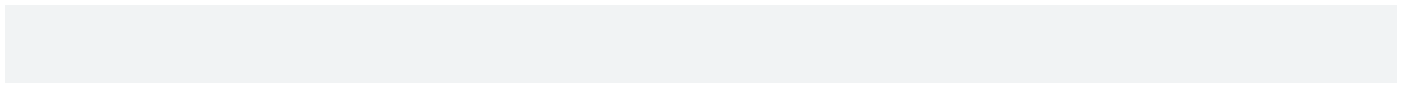
```
plt.title('Distribution of Attrition Rate')
```

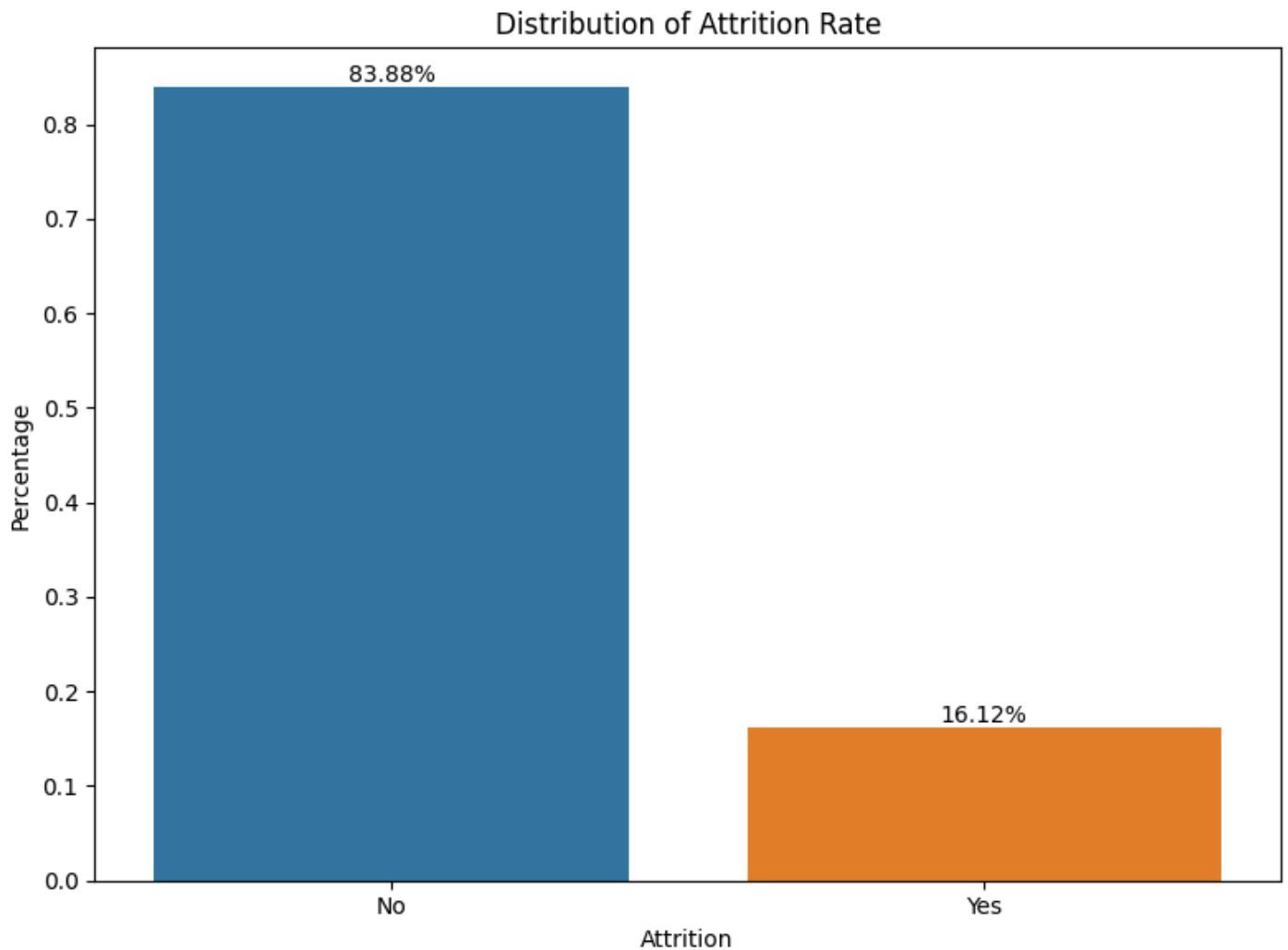
```
plt.xlabel('Attrition')
```

```
plt.ylabel('Percentage')
```

```
plt.tight_layout()
```

```
plt.show()
```





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.

In [11]:

```
avg_tenure = df['YearsAtCompany'].mean()
```

```
print(f'Average years of employee to leave the company is  
{avg_tenure} years')
```

```
Average years of employee to leave the company is  
7.0081632653061225 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

In [12]:

```
fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(15,5))
```

```
sns.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')
```

```
sns.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')
```

```
sns.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()
```

[illegible]

[illegible]

[illegible]

[illegible]


```
    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']
    return attrition_rate_df
```

In [15]:

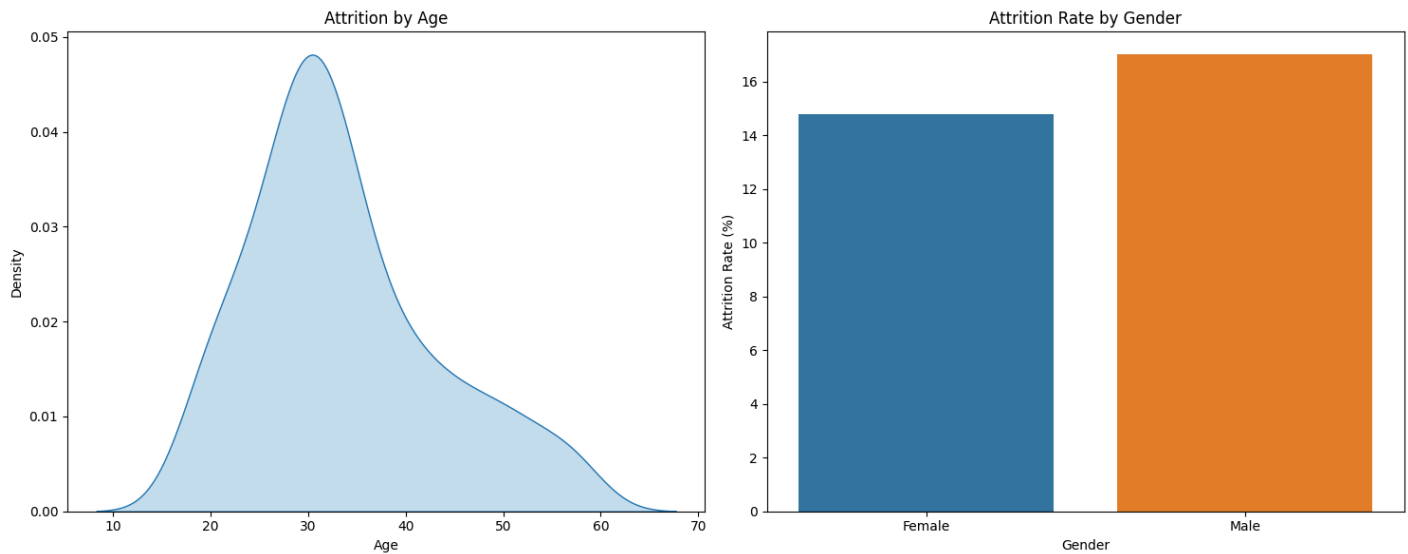
```
fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15,6))

# Plot 1: KDE plot of Age with Attrition hue
sns.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')
sns.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()
```



[Reference link](#)