

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df= pd.read_csv("Downloads/Employee.csv")
df
```

	Education	JoiningYear	City	PaymentTier	Age	Gender
0	Bachelors	2017	Bangalore	3	34	Male
1	Bachelors	2013	Pune	1	28	Female
2	Bachelors	2014	New Delhi	3	38	Female
3	Masters	2016	Bangalore	3	27	Male
4	Masters	2017	Pune	3	24	Male
...	...	...	...	...	...	...
4648	Bachelors	2013	Bangalore	3	26	Female
4649	Masters	2013	Pune	2	37	Male
4650	Masters	2018	New Delhi	3	27	Male
4651	Bachelors	2012	Bangalore	3	30	Male
4652	Bachelors	2015	Bangalore	3	33	Male

	ExperienceInCurrentDomain	LeaveOrNot
0	0	0
1	3	1
2	2	0
3	5	1
4	2	1
...	...	...
4648	4	0
4649	2	1
4650	5	1
4651	2	0
4652	4	0

[4653 rows x 9 columns]

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

```
Education      0
JoiningYear    0
```

```
City 0
PaymentTier 0
Age 0
Gender 0
EverBenched 0
ExperienceInCurrentDomain 0
LeaveOrNot 0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 4653 entries, 0 to 4652
```

```
Data columns (total 9 columns):
```

#	Column	Non-Null Count	Dtype
0	Education	4653 non-null	object
1	JoiningYear	4653 non-null	int64
2	City	4653 non-null	object
3	PaymentTier	4653 non-null	int64
4	Age	4653 non-null	int64
5	Gender	4653 non-null	object
6	EverBenched	4653 non-null	object
7	ExperienceInCurrentDomain	4653 non-null	int64
8	LeaveOrNot	4653 non-null	int64

```
dtypes: int64(5), object(4)
```

```
memory usage: 327.3+ KB
```

```
df.describe()
```

	JoiningYear	PaymentTier	Age
count	4653.000000	4653.000000	4653.000000
mean	2015.062970	2.698259	29.393295
std	1.863377	0.561435	4.826087
min	2012.000000	1.000000	22.000000
25%	2013.000000	3.000000	26.000000
50%	2015.000000	3.000000	28.000000
75%	2017.000000	3.000000	32.000000
max	2018.000000	3.000000	41.000000

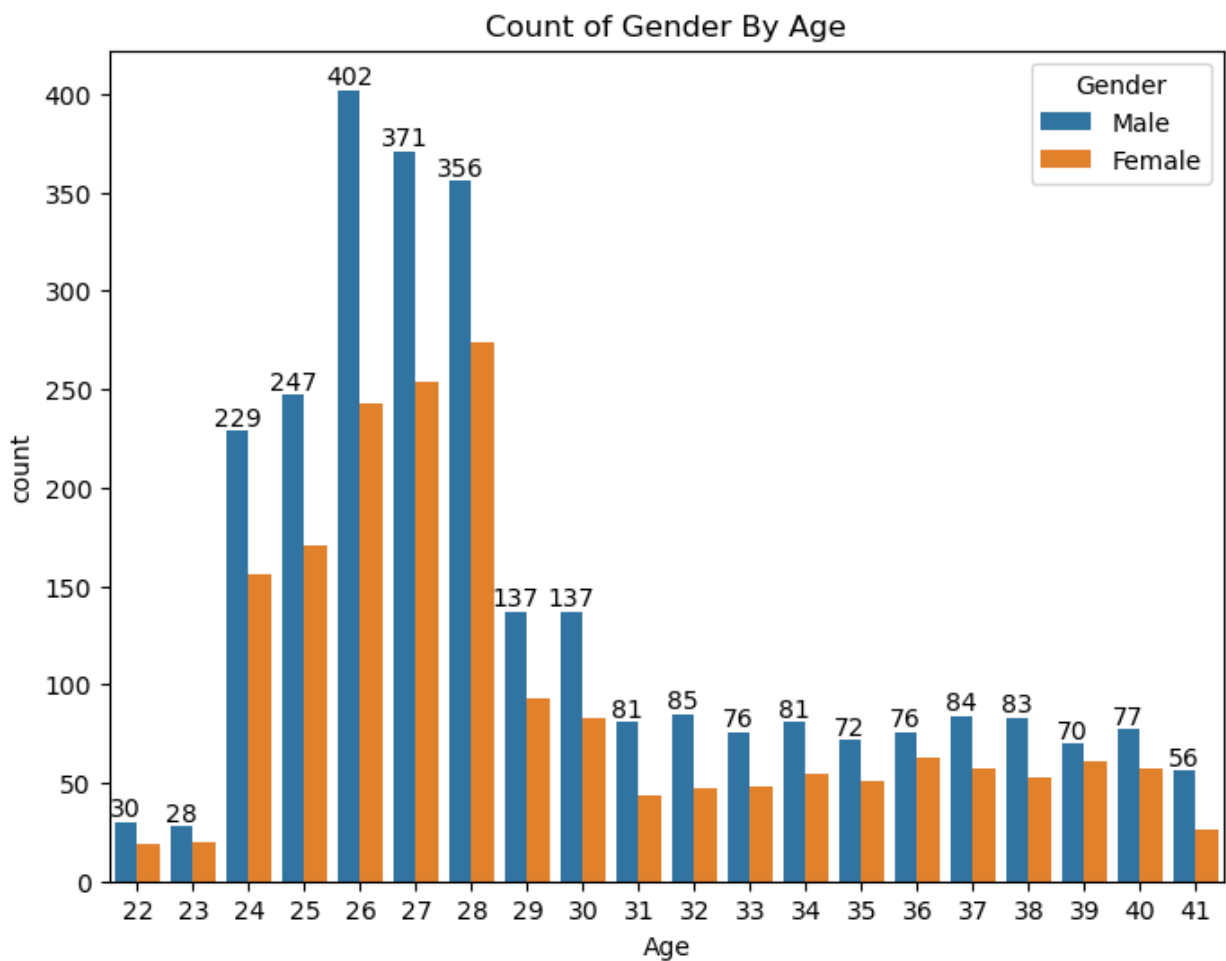
```
LeaveOrNot
```

```
count    4653.000000
mean      0.343864
std       0.475047
min       0.000000
25%      0.000000
50%      0.000000
75%      1.000000
max       1.000000
```

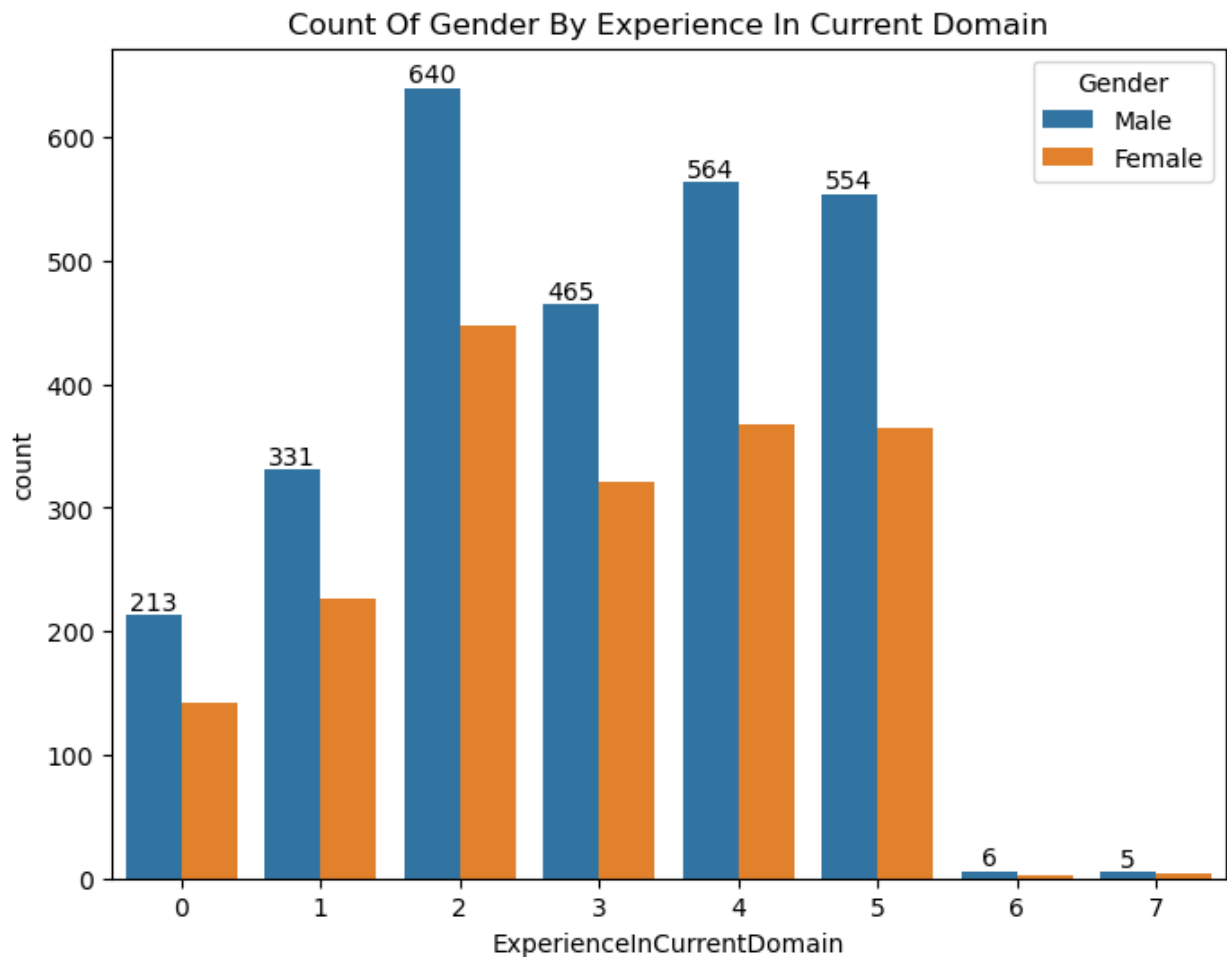
```
df.duplicated().sum()
```

```
1889
```

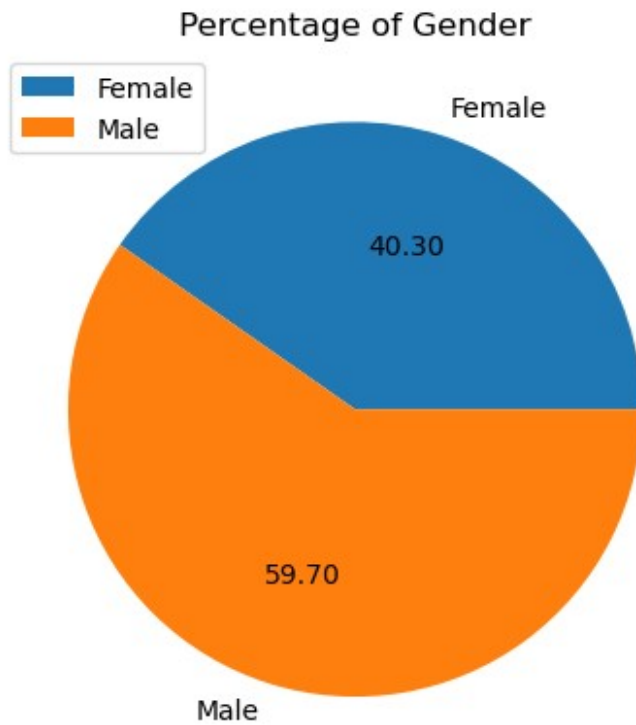
```
plt.figure(figsize=(8,6))
ax=sns.countplot(data=df,x="Age",hue="Gender",)
ax.bar_label(ax.containers[0])
plt.title("Count of Gender By Age")
plt.show()
```



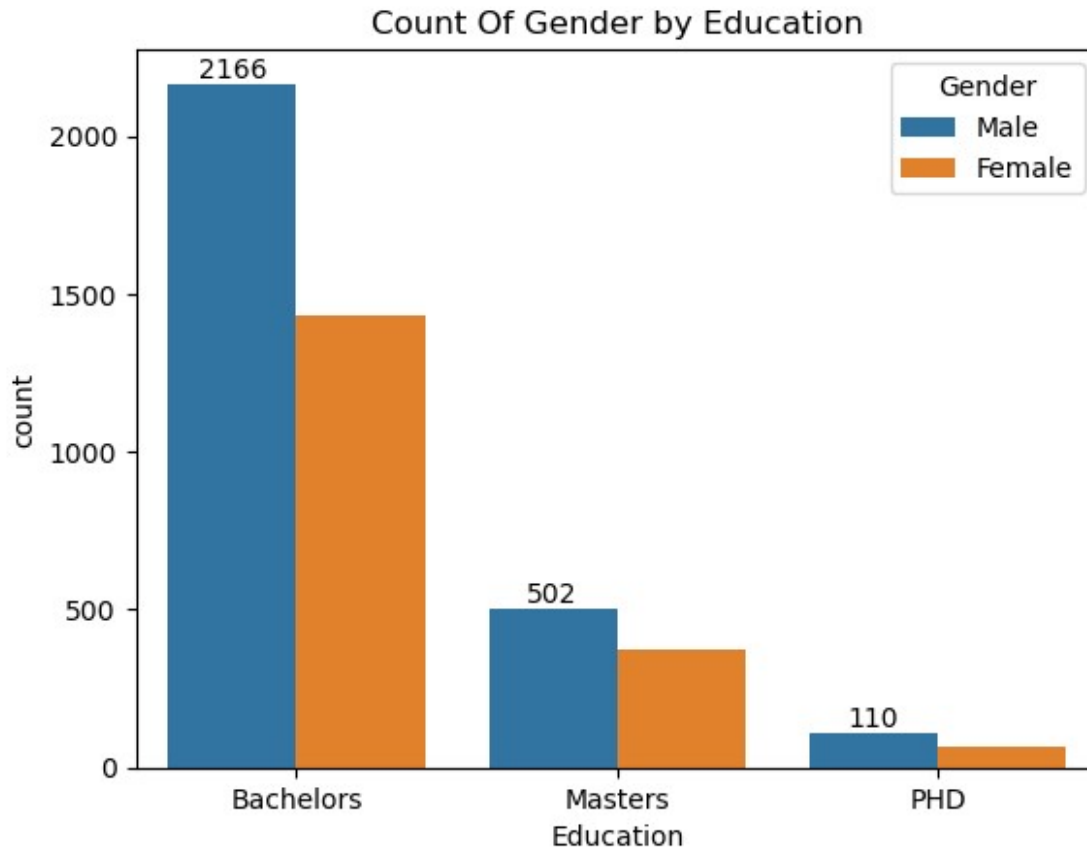
```
plt.figure(figsize=(8,6))
ax=sns.countplot(data=df,x="ExperienceInCurrentDomain",hue="Gender",)
plt.title("Count Of Gender By Experience In Current Domain")
ax.bar_label(ax.containers[0])
plt.show()
```



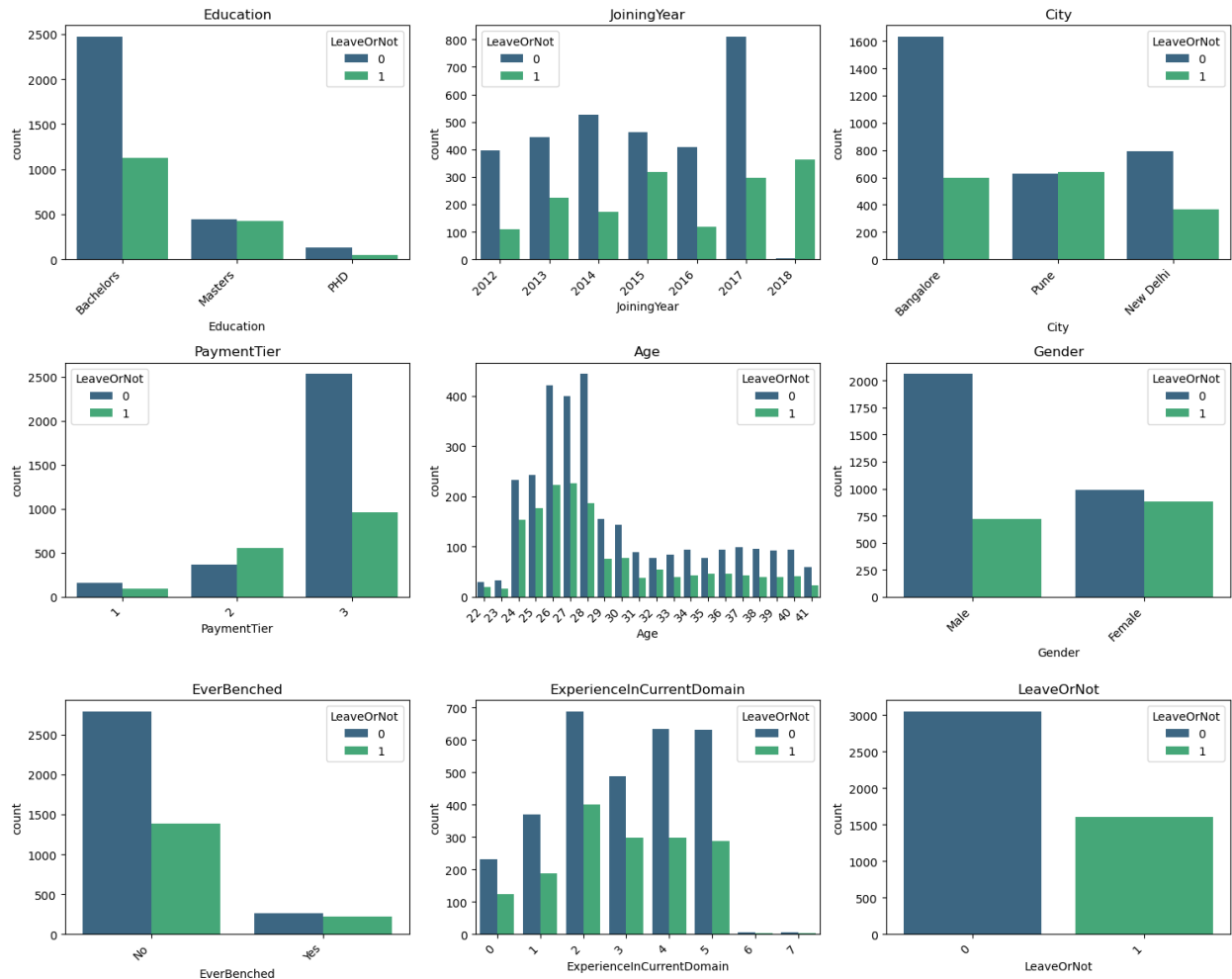
```
gp= df.groupby("Gender").agg({"Gender": "count"})
plt.pie(gp["Gender"], labels=gp.index, autopct="%1.2f")
plt.title("Percentage of Gender")
plt.legend()
plt.show()
```



```
ax=sns.countplot(data=df,x="Education",hue="Gender")
ax=ax.bar_label(ax.containers[0])
plt.title("Count Of Gender by Education")
plt.show()
```



```
cols_to_plot = [  
    'Education', 'JoiningYear', 'City', 'PaymentTier',  
    'Age', 'Gender', 'EverBenched',  
    'ExperienceInCurrentDomain', 'LeaveOrNot'  
]  
  
# Define subplot layout: 3 rows x 3 columns  
n_cols = 3  
n_rows = (len(cols_to_plot) + n_cols - 1) // n_cols  
  
plt.figure(figsize=(15, 12)) # Adjust size as needed  
  
for i, col in enumerate(cols_to_plot, 1):  
    plt.subplot(n_rows, n_cols, i)  
    sns.countplot(data=df, x=col, hue="LeaveOrNot", palette='viridis')  
    plt.title(col)  
    plt.xticks(rotation=45, ha='right') # Rotate x labels for  
    readability  
  
plt.tight_layout()  
plt.show()
```



Most employees hold a Bachelor's degree, work in Payment Tier 3, are based in Bangalore, and have not been benchd. Attrition is notably higher among those with lower education, less domain experience, and lower payment tiers.

```
plt.figure(figsize=(5,5))
ax=sns.countplot(data=df,x="PaymentTier",hue="Gender",)
ax.bar_label(ax.containers[0])
plt.title("Count of Gender By Payment Tier")
plt.show()
```

