

```

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import pandas as pd

df=pd.read_csv('D:/Users/Aseel/Desktop/predicting-employee-attrition/train_data.csv')

pd.set_option('display.max_columns',None)

linkcode
df.head(1)

```

	MMM-YY	Emp_ID	Age	Gender	City	Education_Level	Salary	Dateofjoining	LastWorkingDate	Joining Designation	Designation	Total Business Value	Quarterly Rating
0	2016-01-01	1	28	Male	C23	Master	57387	2015-12-24	NaN	1	1	2381060	2

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19104 entries, 0 to 19103
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   MMM-YY                19104 non-null  object
1   Emp_ID                19104 non-null  int64
2   Age                  19104 non-null  int64
3   Gender                19104 non-null  object
4   City                  19104 non-null  object
5   Education_Level       19104 non-null  object
6   Salary                19104 non-null  int64
7   Dateofjoining         19104 non-null  object
8   LastWorkingDate       1616 non-null   object
9   Joining Designation   19104 non-null  int64
10  Designation            19104 non-null  int64
11  Total Business Value  19104 non-null  int64
12  Quarterly Rating      19104 non-null  int64
dtypes: int64(7), object(6)
memory usage: 1.9+ MB

```

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

```
MMM-YY      0
```

```

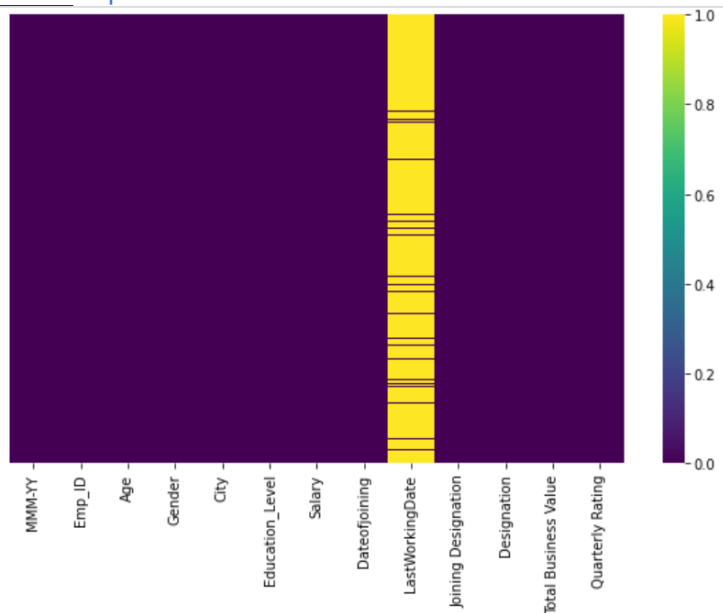
Emp_ID          0
Age             0
Gender          0
City            0
Education_Level 0
Salary          0
Dateofjoining   0
LastWorkingDate 17488
Joining Designation 0
Designation     0
Total Business Value 0
Quarterly Rating 0
dtype: int64

```

```
plt.figure(figsize=(10,6))
```

```
sns.heatmap(df.isnull(),yticklabels=False,cmap='viridis')
```

```
<AxesSubplot:>
```



```
df.columns
```

```

Index(['MMM-YY', 'Emp_ID', 'Age', 'Gender', 'City', 'Education_Level',
       'Salary', 'Dateofjoining', 'LastWorkingDate', 'Joining Designation',
       'Designation', 'Total Business Value', 'Quarterly Rating'],
      dtype='object')

```

```
df['Dateofjoining']=pd.to_datetime(df['Dateofjoining'])
```

```
df['LastWorkingDate']=pd.to_datetime(df['LastWorkingDate'])
```

```
df['Year_of_join']=df['Dateofjoining'].apply(lambda t:t.year)
```

```
df['Month_of_join']=df['Dateofjoining'].apply(lambda t:t.month)
```

```
df['Day_of_join']=df['Dateofjoining'].apply(lambda t:t.day)
```

```

df['Year_of_leave'] = df['LastWorkingDate'].apply(lambda t:t.year)

df['Month_of_leave'] = df['LastWorkingDate'].apply(lambda t:t.month)

df.drop(columns='Dateofjoining',inplace=True)

df['Attrition'] = np.nan

mypop = df.pop('Attrition')
df.insert(1,'Attrition',mypop)
mypop1 = df.pop('Year_of_join')
df.insert(8,'Year_of_join',mypop1)
mypop2 = df.pop('Month_of_join')
df.insert(9,'Month_of_join',mypop2)
mypop3 = df.pop('Day_of_join')
df.insert(10,'Day_of_join',mypop3)

df = df.astype({'Year_of_join':int,'Month_of_join':int,'Day_of_join':int})

df['Attrition']=np.where(df['LastWorkingDate'].isnull(),0,1)

df.drop(columns='LastWorkingDate',inplace=True)

linkcode
df.head(3)

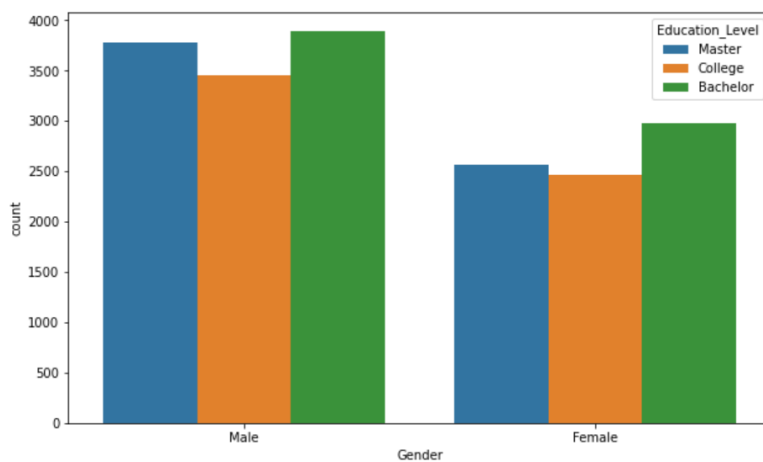
```

```

plt.figure(figsize=(10,6))

sns.countplot(data=df,x='Gender',hue='Education_Level')

```

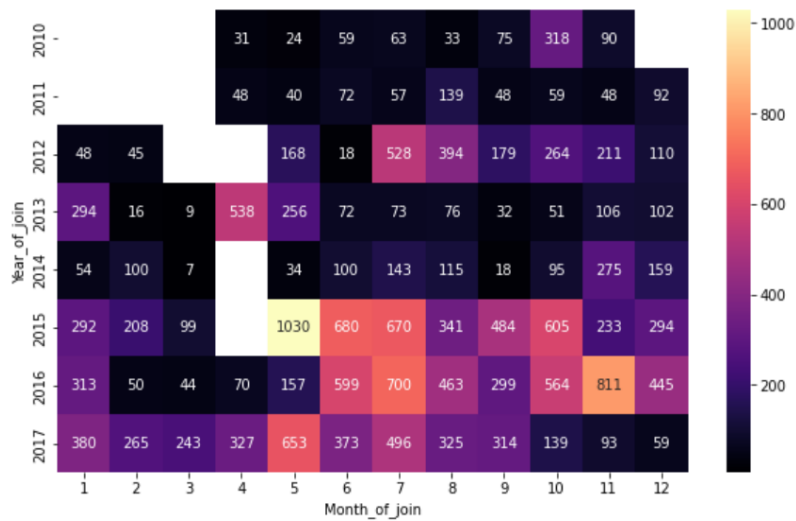


```
joiners = df.groupby(by=['Year_of_join','Month_of_join']).count()['Emp_ID'].unstack()
```

```
linkcode
```

```
plt.figure(figsize=(10,6))
```

```
sns.heatmap(joiners,annot=True,fmt='.4g',cmap='magma')
```



```
df['Attrition'].value_counts()
```

```
0    17488
```

```
1     1616
```

```
Name: Attrition, dtype: int64
```

```
plt.figure(figsize=(10,6))
```

```
fg = sns.FacetGrid(df, col="QuarterlyRating", row='Attrition')
```

```
fg.map(sns.scatterplot, "Salary", "Age")
```

