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

df = pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
df.head()
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2         Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   ...  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  ...                         1            80                 0
1  ...                         4            80                 1
2  ...                         2            80                 0
3  ...                         3            80                 0
4  ...                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                  8                      0               1
6
1                 10                      3               3
10
2                  7                      3               3
0
3                  8                      3               3
```

```
8
4                  6                        3               3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

```
df.shape
```

```
(1470, 35)
```

```
df.describe()
```

```
                Age    DailyRate  DistanceFromHome    Education
EmployeeCount  \
count  1470.000000  1470.000000       1470.000000  1470.000000
1470.0
mean     36.923810   802.485714          9.192517     2.912925
1.0
std       9.135373   403.509100          8.106864     1.024165
0.0
min      18.000000   102.000000          1.000000     1.000000
1.0
25%      30.000000   465.000000          2.000000     2.000000
1.0
50%      36.000000   802.000000          7.000000     3.000000
1.0
75%      43.000000  1157.000000         14.000000     4.000000
1.0
max      60.000000  1499.000000         29.000000     5.000000
1.0

       EmployeeNumber  EnvironmentSatisfaction   HourlyRate
JobInvolvement  \
count     1470.000000              1470.000000  1470.000000
1470.000000
mean      1024.865306                 2.721769    65.891156
2.729932
std        602.024335                 1.093082    20.329428
0.711561
min          1.000000                 1.000000    30.000000
1.000000
25%        491.250000                 2.000000    48.000000
2.000000
```

```
50%       1020.500000                    3.000000    66.000000
3.000000
75%       1555.750000                    4.000000    83.750000
3.000000
max       2068.000000                    4.000000   100.000000
4.000000

          JobLevel  ...  RelationshipSatisfaction  StandardHours  \
count  1470.000000  ...               1470.000000         1470.0
mean      2.063946  ...                  2.712245           80.0
std       1.106940  ...                  1.081209            0.0
min       1.000000  ...                  1.000000           80.0
25%       1.000000  ...                  2.000000           80.0
50%       2.000000  ...                  3.000000           80.0
75%       3.000000  ...                  4.000000           80.0
max       5.000000  ...                  4.000000           80.0

       StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
count       1470.000000        1470.000000            1470.000000
mean           0.793878          11.279592               2.799320
std            0.852077           7.780782               1.289271
min            0.000000           0.000000               0.000000
25%            0.000000           6.000000               2.000000
50%            1.000000          10.000000               3.000000
75%            1.000000          15.000000               3.000000
max            3.000000          40.000000               6.000000

       WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
count      1470.000000     1470.000000         1470.000000
mean          2.761224        7.008163            4.229252
std           0.706476        6.126525            3.623137
min           1.000000        0.000000            0.000000
25%           2.000000        3.000000            2.000000
50%           3.000000        5.000000            3.000000
75%           3.000000        9.000000            7.000000
max           4.000000       40.000000           18.000000

       YearsSinceLastPromotion  YearsWithCurrManager
count              1470.000000           1470.000000
mean                  2.187755              4.123129
std                   3.222430              3.568136
min                   0.000000              0.000000
25%                   0.000000              2.000000
50%                   1.000000              3.000000
75%                   3.000000              7.000000
max                  15.000000             17.000000

[8 rows x 26 columns]
```

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

```
Age                         0
Attrition                   0
BusinessTravel              0
DailyRate                   0
Department                  0
DistanceFromHome            0
Education                   0
EducationField              0
EmployeeCount               0
EmployeeNumber              0
EnvironmentSatisfaction     0
Gender                      0
HourlyRate                  0
JobInvolvement              0
JobLevel                    0
JobRole                     0
JobSatisfaction             0
MaritalStatus               0
MonthlyIncome               0
MonthlyRate                 0
NumCompaniesWorked          0
Over18                      0
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

```
attrition_count = pd.DataFrame(df['Attrition'].value_counts())
attrition_count
```

```
     Attrition
No        1233
Yes        237
```

```
plt.pie(attrition_count['Attrition'] , labels = ['No' , 'Yes'] ,
explode = (0.2,0))
```

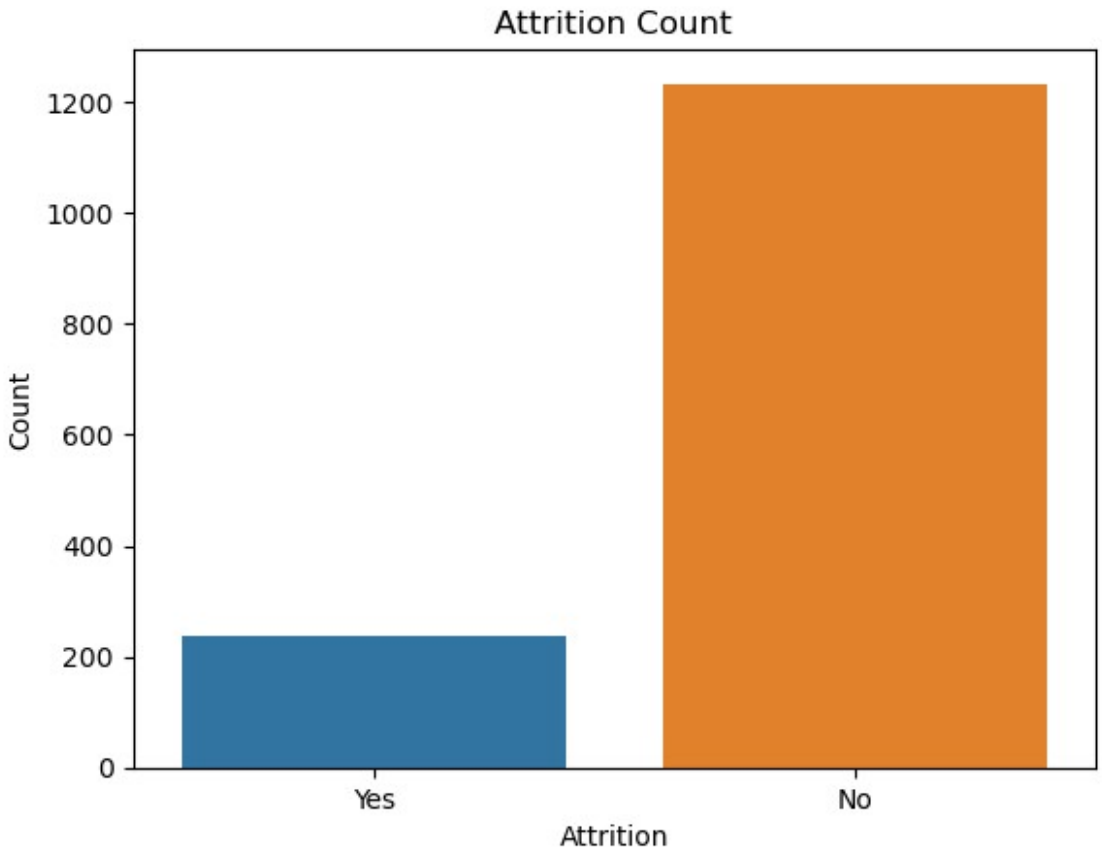
```
([<matplotlib.patches.Wedge at 0x232d4de5090>,
  <matplotlib.patches.Wedge at 0x232d4dd16c0>],
```

```
[Text(-1.136781068348268, 0.6306574368426737, 'No'),
 Text(0.961891673217765, -0.5336332157899547, 'Yes')])
```


```
sns.countplot(data=df, x='Attrition')

plt.title('Attrition Count')
plt.xlabel('Attrition')
plt.ylabel('Count')

plt.show()
```




```
df.drop(['EmployeeCount' , 'EmployeeNumber'] , axis = 1)
```

```
      Age Attrition     BusinessTravel  DailyRate
Department  \
0      41       Yes      Travel_Rarely       1102
Sales
1      49        No  Travel_Frequently        279  Research &
Development
2      37       Yes      Travel_Rarely       1373  Research &
Development
3      33        No  Travel_Frequently       1392  Research &
Development
4      27        No      Travel_Rarely        591  Research &
Development
...   ...       ...                ...         ...
...
1465   36        No  Travel_Frequently        884  Research &
Development
1466   39        No      Travel_Rarely        613  Research &
Development
1467   27        No      Travel_Rarely        155  Research &
Development
1468   49        No  Travel_Frequently       1023
```

```
Sales
1469   34        No      Travel_Rarely        628  Research &
Development

      DistanceFromHome  Education EducationField
EnvironmentSatisfaction  \
0                    1          2  Life Sciences
2
1                    8          1  Life Sciences
3
2                    2          2          Other
4
3                    3          4  Life Sciences
4
4                    2          1        Medical
1
...                ...        ...            ...
...
1465                23          2        Medical
3
1466                 6          1        Medical
4
1467                 4          3  Life Sciences
2
1468                 2          3        Medical
4
1469                 8          3        Medical
2

      Gender  ...  RelationshipSatisfaction  StandardHours
StockOptionLevel  \
0     Female  ...                         1             80
0
1       Male  ...                         4             80
1
2       Male  ...                         2             80
0
3     Female  ...                         3             80
0
4       Male  ...                         4             80
1
...      ...  ...                       ...            ...
...
1465    Male  ...                         3             80
1
1466    Male  ...                         1             80
1
1467    Male  ...                         2             80
1
```

```
1468    Male  ...                          4             80
0
1469    Male  ...                          1             80
0

     TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
0                    8                      0               1
6
1                   10                      3               3
10
2                    7                      3               3
0
3                    8                      3               3
8
4                    6                      3               3
2
...                ...                    ...             ...
...
1465                17                      3               3
5
1466                 9                      5               3
7
1467                 6                      0               3
6
1468                17                      3               2
9
1469                 6                      3               4
4

      YearsInCurrentRole  YearsSinceLastPromotion YearsWithCurrManager

0                      4                        0                     5

1                      7                        1                     7

2                      0                        0                     0

3                      7                        3                     0

4                      2                        2                     2

...                  ...                      ...                   ...

1465                   2                        0                     3

1466                   7                        1                     7

1467                   2                        0                     3
```

```
1468                    6                           0                    8

1469                    3                           1                    2


[1470 rows x 33 columns]
```

```
attrition_dummies = pd.get_dummies(df['Attrition'])
attrition_dummies.head()
```

```
   No  Yes
0   0    1
1   1    0
2   0    1
3   1    0
4   1    0
```

```
df = pd.concat([df, attrition_dummies] , axis = 1)
df.head()
```

```
   Age Attrition     BusinessTravel  DailyRate              Department
\
0   41       Yes      Travel_Rarely       1102                   Sales

1   49        No  Travel_Frequently        279  Research & Development

2   37       Yes      Travel_Rarely       1373  Research & Development

3   33        No  Travel_Frequently       1392  Research & Development

4   27        No      Travel_Rarely        591  Research & Development


   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2          Other              1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   ...  StockOptionLevel TotalWorkingYears  TrainingTimesLastYear  \
0  ...                 0                 8                      0
1  ...                 1                10                      3
2  ...                 0                 7                      3
3  ...                 0                 8                      3
```

```
4  ...                  1                  6                       3

   WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0                1               6                  4
1                3              10                  7
2                3               0                  0
3                3               8                  7
4                3               2                  2

   YearsSinceLastPromotion YearsWithCurrManager  No  Yes
0                        0                    5   0    1
1                        1                    7   1    0
2                        0                    0   0    1
3                        3                    0   1    0
4                        2                    2   1    0

[5 rows x 37 columns]
```

```
df = df.drop(['Attrition' , 'No'] , axis = 1)
df.head()
```

```
   Age     BusinessTravel  DailyRate              Department  \
0   41      Travel_Rarely       1102                   Sales
1   49  Travel_Frequently        279  Research & Development
2   37      Travel_Rarely       1373  Research & Development
3   33  Travel_Frequently       1392  Research & Development
4   27      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount
EmployeeNumber  \
0                 1          2  Life Sciences              1
1
1                 8          1  Life Sciences              1
2
2                 2          2         Other               1
4
3                 3          4  Life Sciences              1
5
4                 2          1        Medical              1
7

   EnvironmentSatisfaction  ... StandardHours  StockOptionLevel  \
0                        2  ...            80                 0
1                        3  ...            80                 1
2                        4  ...            80                 0
3                        4  ...            80                 0
4                        1  ...            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance
YearsAtCompany  \
```

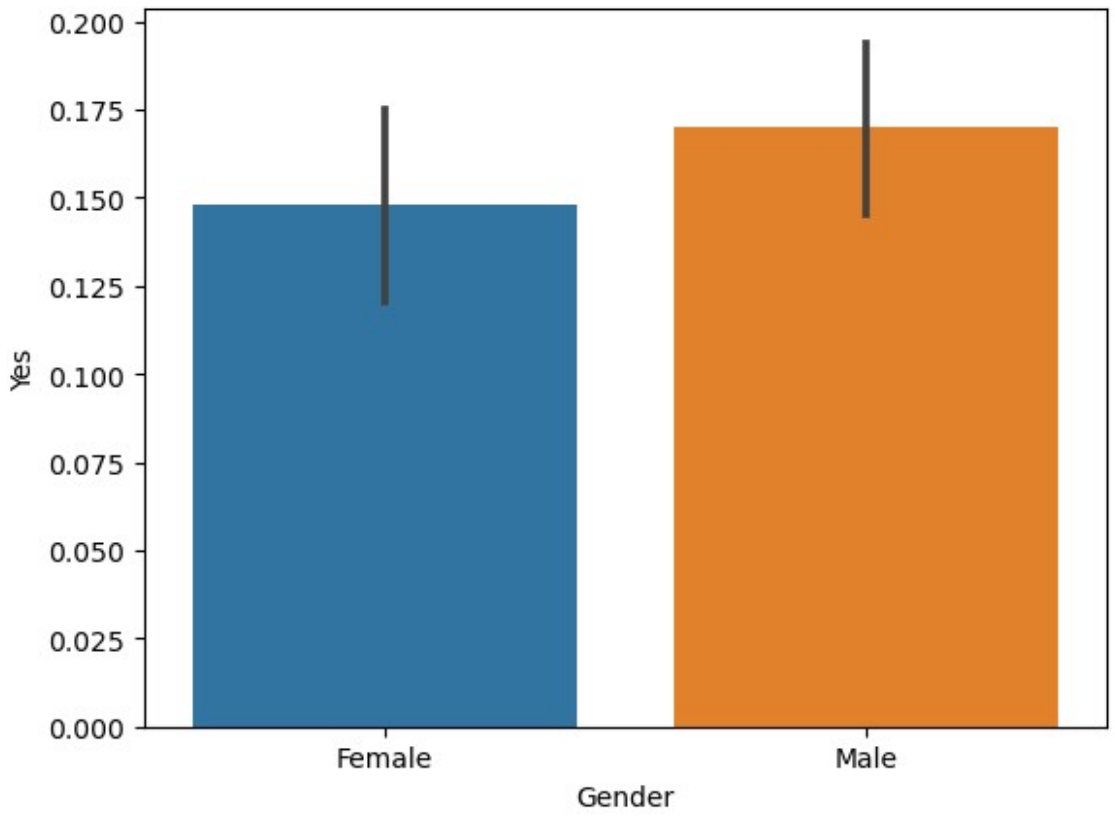
```
0                    8                         0                 1
6
1                   10                         3                 3
10
2                    7                         3                 3
0
3                    8                         3                 3
8
4                    6                         3                 3
2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
Yes
0                    4                        0                    5
1
1                    7                        1                    7
0
2                    0                        0                    0
1
3                    7                        3                    0
0
4                    2                        2                    2
0

[5 rows x 35 columns]
```

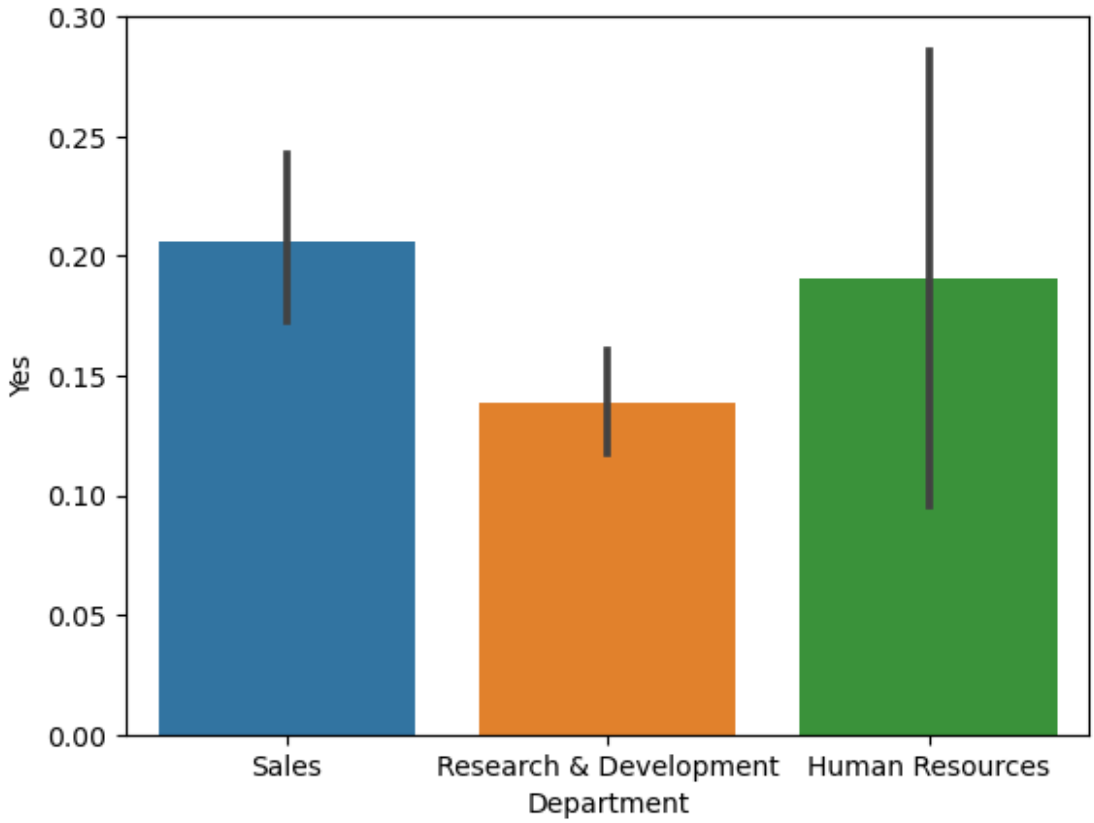
```
sns.barplot(x = 'Gender' , y = 'Yes', data = df)
```

```
<Axes: xlabel='Gender', ylabel='Yes'>
```



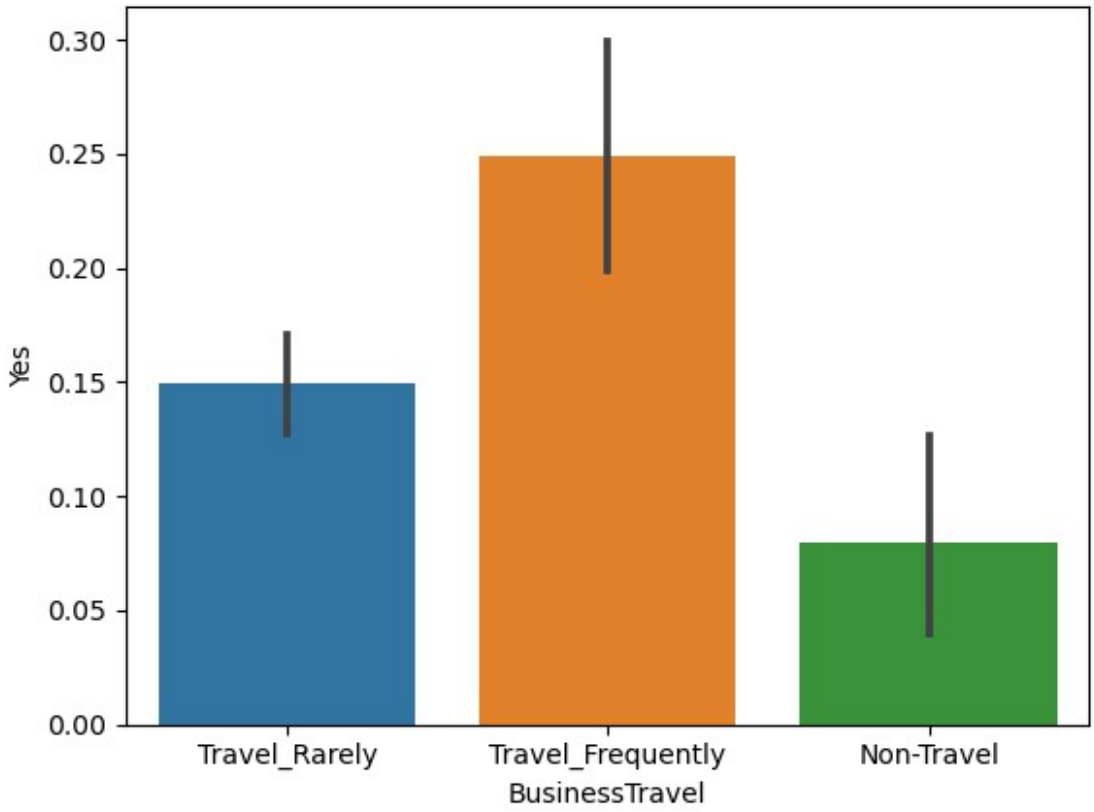

```
sns.barplot(x = 'Department', y = 'Yes', data = df)
```

```
<Axes: xlabel='Department', ylabel='Yes'>
```




```
sns.barplot(x = 'BusinessTravel', y = 'Yes', data = df)
```

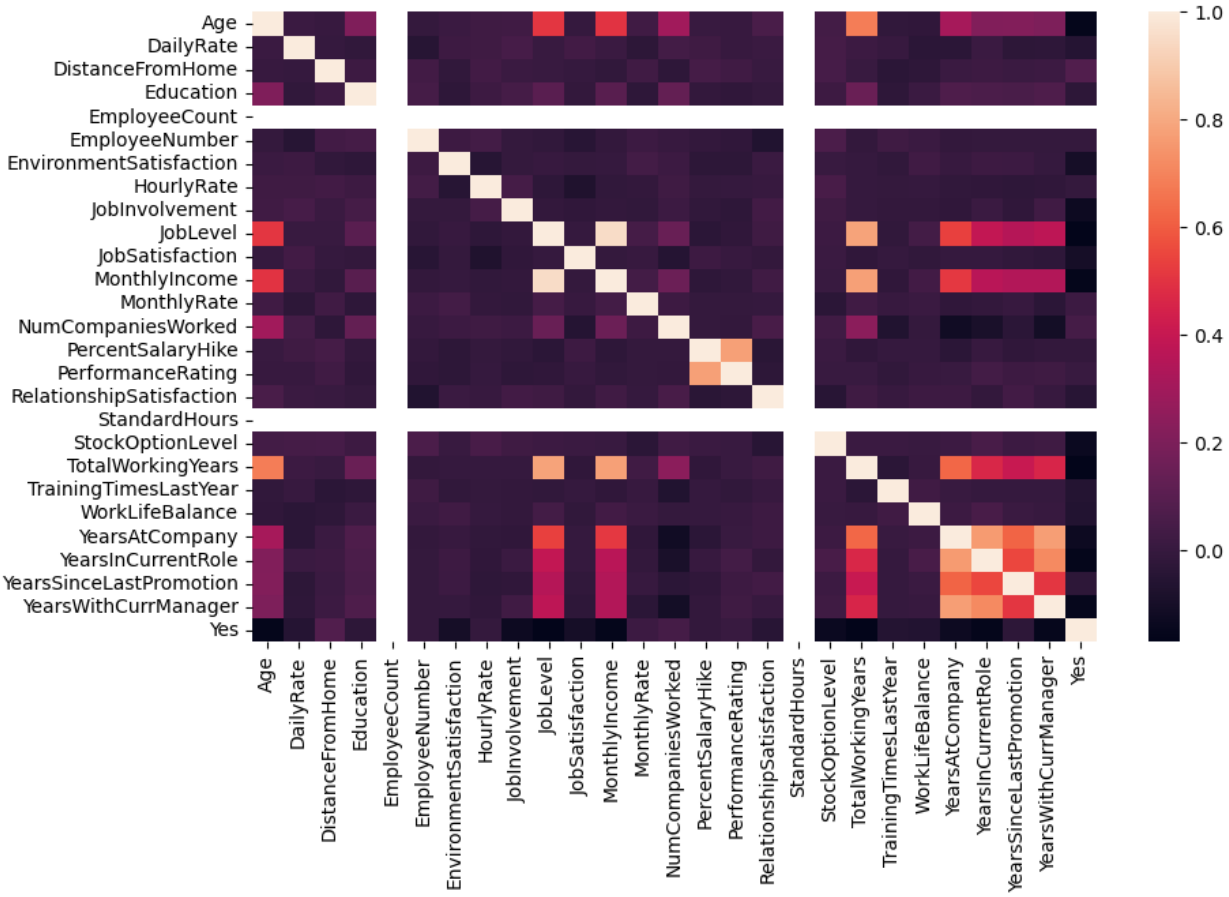
```
<Axes: xlabel='BusinessTravel', ylabel='Yes'>
```




```
plt.figure(figsize = (10,6))
sns.heatmap(df.corr())
```

```
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\3820300259.py:2:
FutureWarning: The default value of numeric_only in DataFrame.corr is
deprecated. In a future version, it will default to False. Select only
valid columns or specify the value of numeric_only to silence this
warning.
  sns.heatmap(df.corr())
```

```
<Axes: >
```




```
df = df.drop(['Age' , 'JobLevel'], axis = 1)

# Data Preprocessing

from sklearn.preprocessing import LabelEncoder
for column in df.columns:
    if df[column].dtype==np.number:
        continue
    else:
        df[column]=LabelEncoder().fit_transform(df[column])
```

```
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
```

```
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
```

```
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
```

```
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
```

```
strictly correct.
  if df[column].dtype==np.number:
C:\Users\manvi\AppData\Local\Temp\ipykernel_25812\2736618238.py:5:
DeprecationWarning: Converting `np.inexact` or `np.floating` to a
dtype is deprecated. The current result is `float64` which is not
strictly correct.
  if df[column].dtype==np.number:
```

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators = 10, criterion = 'entropy',
random_state = 0)

x  = df.drop(['Yes'], axis = 1)
y = df['Yes']

x_train, x_test , y_train, y_test = train_test_split(x,y, test_size =
0.3, random_state = 0)

x_train.head()
```

```
     BusinessTravel  DailyRate  Department  DistanceFromHome
Education  \
338               2        295           2                 4
2
363               2        159           2                 4
2
759               2        437           0                23
3
793               2        502           1                14
1
581               2        516           1                 0
2

     EducationField  EmployeeCount  EmployeeNumber
EnvironmentSatisfaction  \
338               2              0            338
3
363               2              0            363
3
759               3              0            759
1
793               1              0            793
0
581               1              0            581
3

     Gender  ...  RelationshipSatisfaction  StandardHours
StockOptionLevel  \
338       0  ...                         2             0
3
```

```
363       0  ...                           1               0
0
759       1  ...                           0               0
0
793       1  ...                           3               0
1
581       1  ...                           2               0
2

     TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  \
338                 10                      2                2
363                  1                      2                2
759                  6                      3                2
793                  4                      5                1
581                  7                      2                2

     YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
338              10                   9                        1
363               1                   0                        0
759               6                   3                        0
793               4                   2                        2
581               2                   2                        0

     YearsWithCurrManager
338                     2
363                     0
759                     4
793                     2
581                     2

[5 rows x 32 columns]
```

```
rf.fit(x_train, y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=10,
random_state=0)
```

```
rf.score(x_train, y_train)
```

```
0.9815354713313897
```

```
# Predicting for X_Test
```

```
pred = rf.predict(x_test)
```

```
from sklearn.metrics import accuracy_score
accuracy_score(y_test, pred)
```

```
0.8526077097505669
```