

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [3]: df = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv')
```

```
In [4]: df.head(5)
```

```
Out[4]:
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sci
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sci
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sci
4	27	No	Travel_Rarely	591	Research & Development	2	1	Ma

5 rows × 35 columns

```
In [5]: df.tail(5)
```

```
Out[5]:
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Educa
1465	36	No	Travel_Frequently	884	Research & Development	23	2	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life
1468	49	No	Travel_Frequently	1023	Sales	2	3	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	

5 rows × 35 columns

```
In [6]: df.shape
```

```
Out[6]: (1470, 35)
```

```
In [7]: df.describe()
```

Out[7]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000

8 rows × 26 columns



In [8]: `df.info()`

```

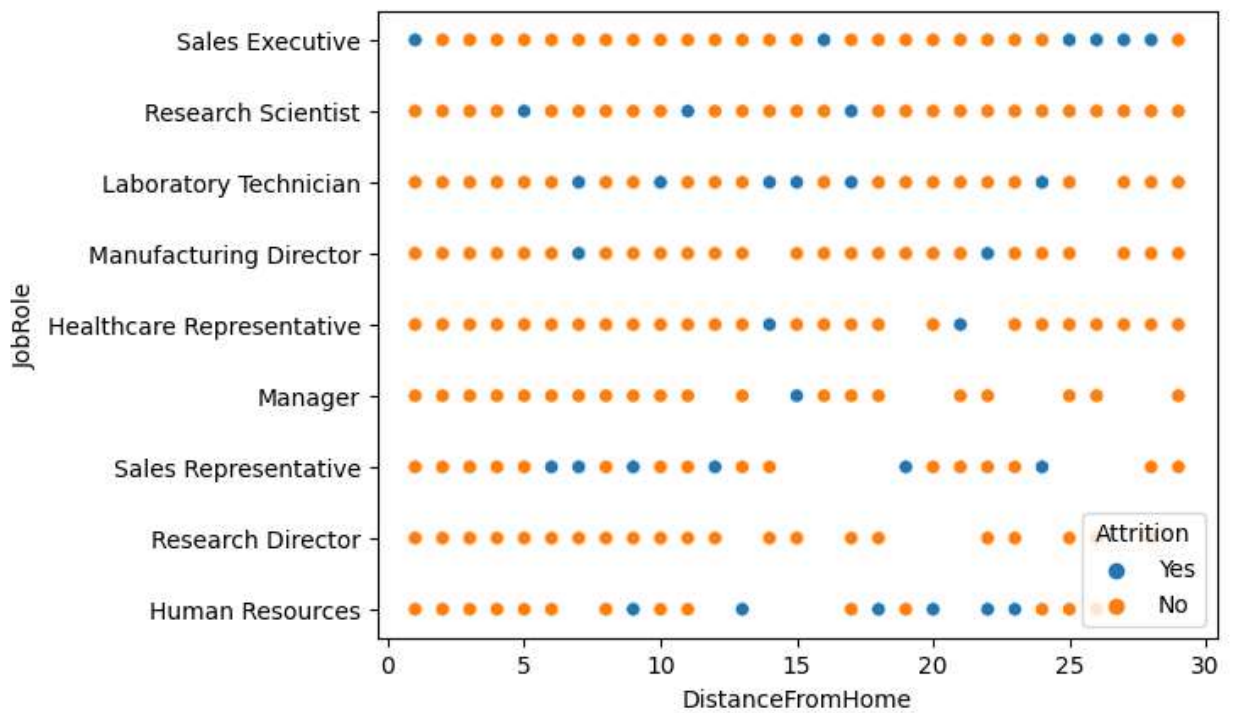
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Age                                   1470 non-null   int64
1   Attrition                           1470 non-null   object
2   BusinessTravel                       1470 non-null   object
3   DailyRate                            1470 non-null   int64
4   Department                           1470 non-null   object
5   DistanceFromHome                     1470 non-null   int64
6   Education                            1470 non-null   int64
7   EducationField                       1470 non-null   object
8   EmployeeCount                        1470 non-null   int64
9   EmployeeNumber                       1470 non-null   int64
10  EnvironmentSatisfaction               1470 non-null   int64
11  Gender                               1470 non-null   object
12  HourlyRate                           1470 non-null   int64
13  JobInvolvement                       1470 non-null   int64
14  JobLevel                             1470 non-null   int64
15  JobRole                              1470 non-null   object
16  JobSatisfaction                       1470 non-null   int64
17  MaritalStatus                        1470 non-null   object
18  MonthlyIncome                        1470 non-null   int64
19  MonthlyRate                          1470 non-null   int64
20  NumCompaniesWorked                  1470 non-null   int64
21  Over18                              1470 non-null   object
22  OverTime                             1470 non-null   object
23  PercentSalaryHike                   1470 non-null   int64
24  PerformanceRating                   1470 non-null   int64
25  RelationshipSatisfaction              1470 non-null   int64
26  StandardHours                       1470 non-null   int64
27  StockOptionLevel                    1470 non-null   int64
28  TotalWorkingYears                   1470 non-null   int64
29  TrainingTimesLastYear                1470 non-null   int64
30  WorkLifeBalance                      1470 non-null   int64
31  YearsAtCompany                       1470 non-null   int64
32  YearsInCurrentRole                   1470 non-null   int64
33  YearsSinceLastPromotion              1470 non-null   int64
34  YearsWithCurrManager                 1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB

```

In [9]: *#Show a breakdown of distance from home by job role and attrition.*

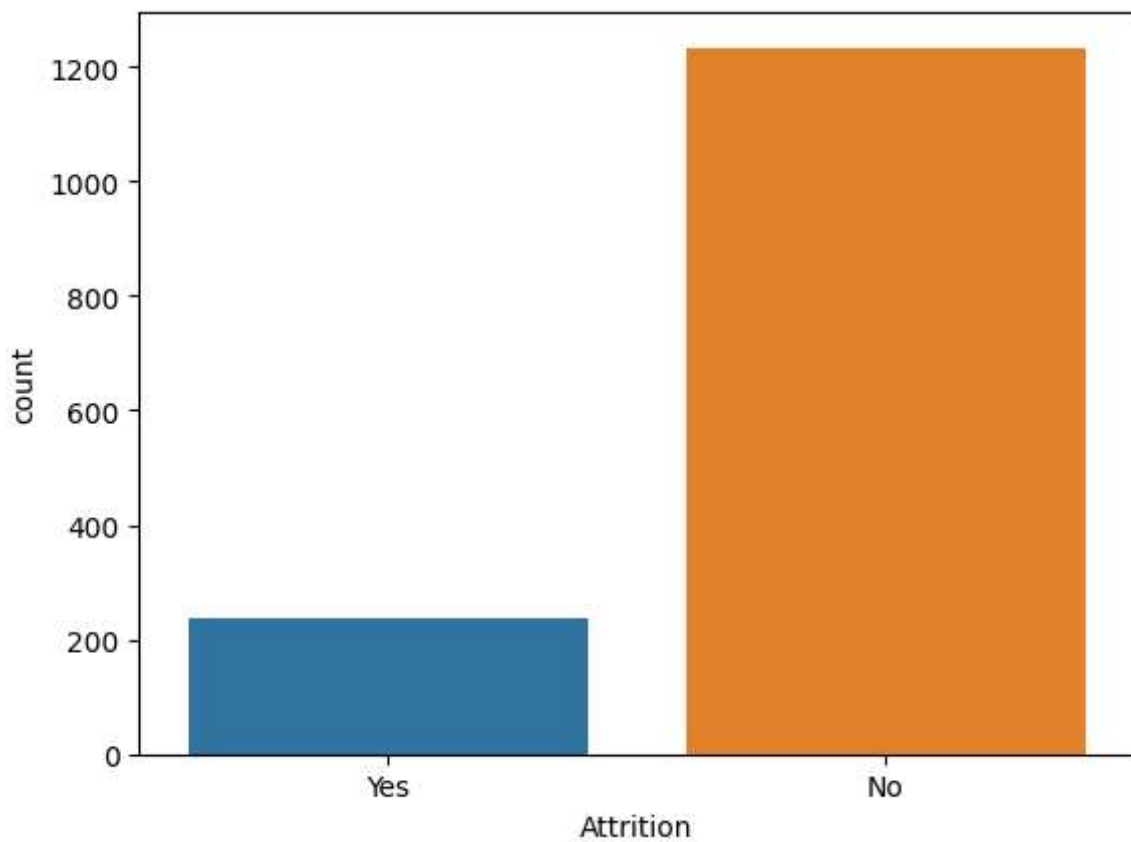
```
sns.scatterplot(data=df, x='DistanceFromHome', y='JobRole', hue='Attrition')
```

Out[9]: <AxesSubplot: xlabel='DistanceFromHome', ylabel='JobRole'>



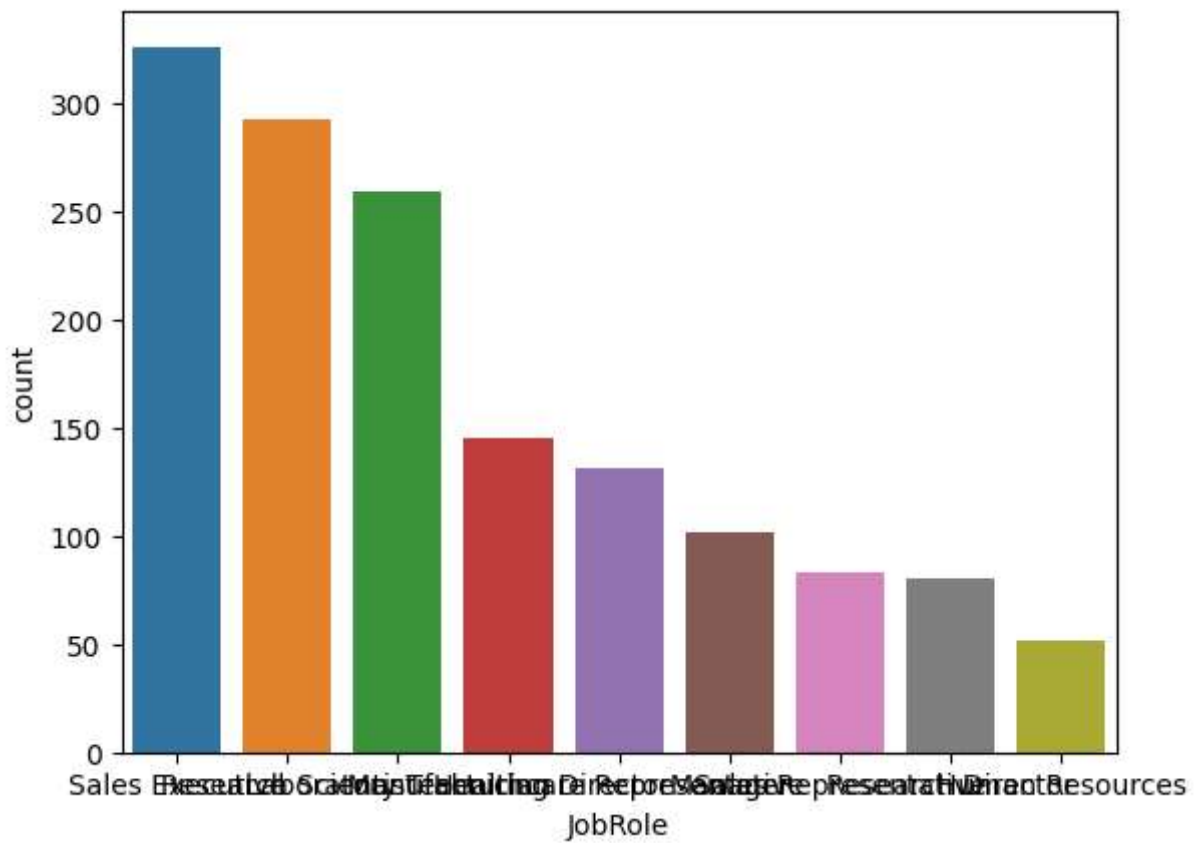
```
In [10]: sns.countplot(data=df, x='Attrition')
```

```
Out[10]: <AxesSubplot: xlabel='Attrition', ylabel='count'>
```



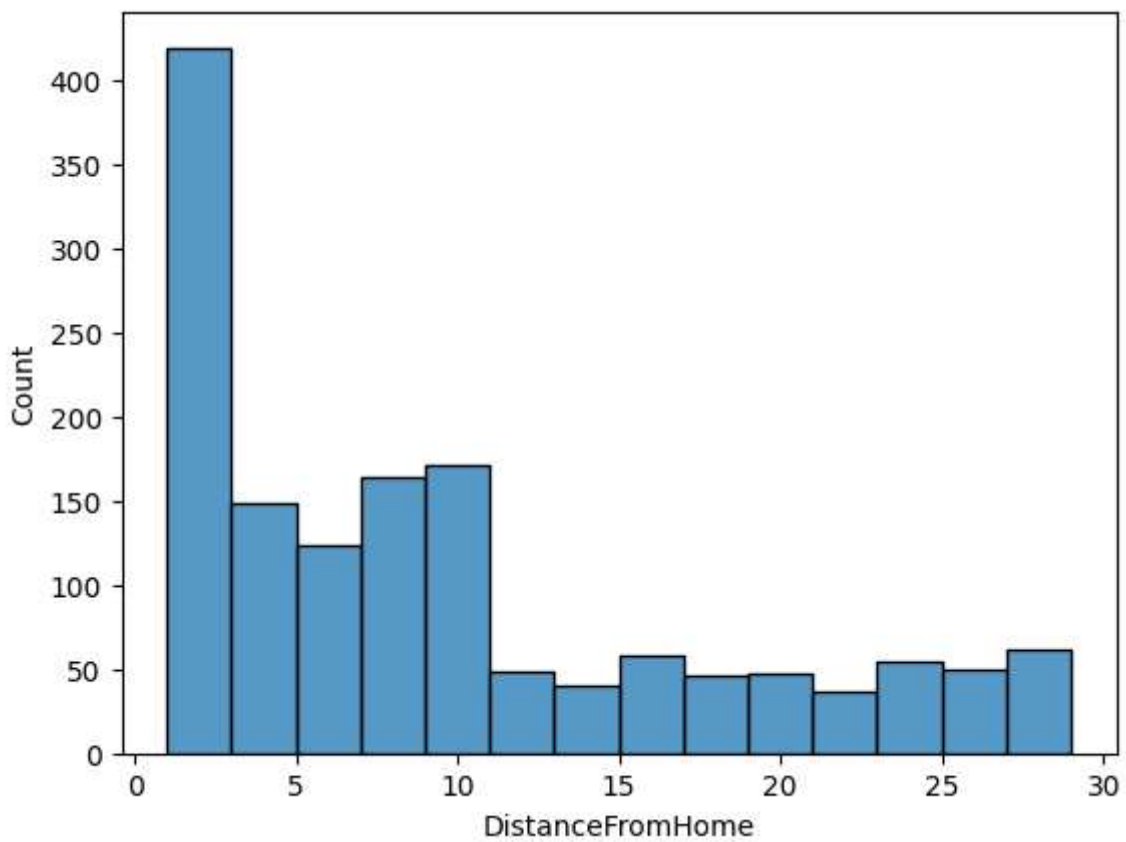
```
In [11]: sns.countplot(data=df, x='JobRole')
```

```
Out[11]: <AxesSubplot: xlabel='JobRole', ylabel='count'>
```



```
In [12]: sns.histplot(data=df, x='DistanceFromHome')
```

```
Out[12]: <AxesSubplot: xlabel='DistanceFromHome', ylabel='Count'>
```



```
In [13]: # Create a pivot table to show the average distance from home by job role and attrition
pivot_table = df.pivot_table(values='DistanceFromHome', index='JobRole', columns='Attrition')
print(pivot_table)
```

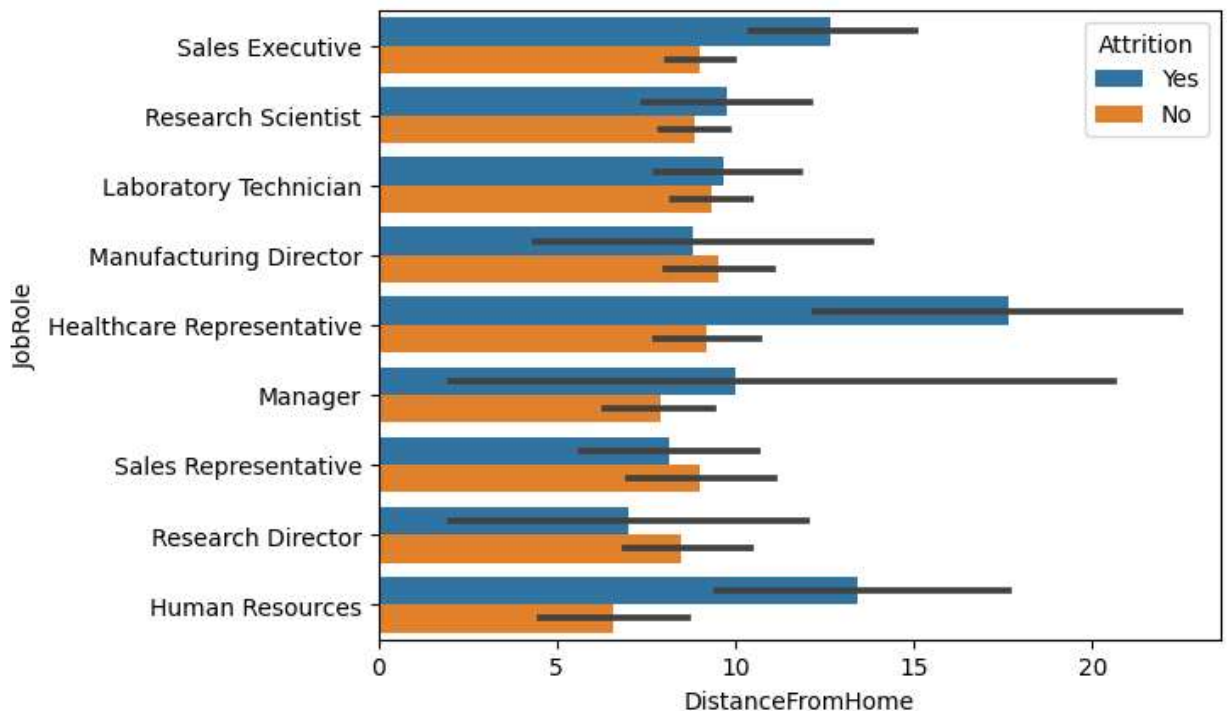
Attrition	No	Yes
JobRole		
Healthcare Representative	9.204918	17.666667
Human Resources	6.600000	13.416667
Laboratory Technician	9.329949	9.661290
Manager	7.927835	10.000000
Manufacturing Director	9.533333	8.800000
Research Director	8.474359	7.000000
Research Scientist	8.869388	9.765957
Sales Executive	9.026022	12.649123
Sales Representative	9.000000	8.151515

```
In [16]: # Create a pivot table to show the average monthly income by education and attrition
pivot_table = df.pivot_table(values='MonthlyIncome', index='Education', columns='Attrition')
print(pivot_table)
```

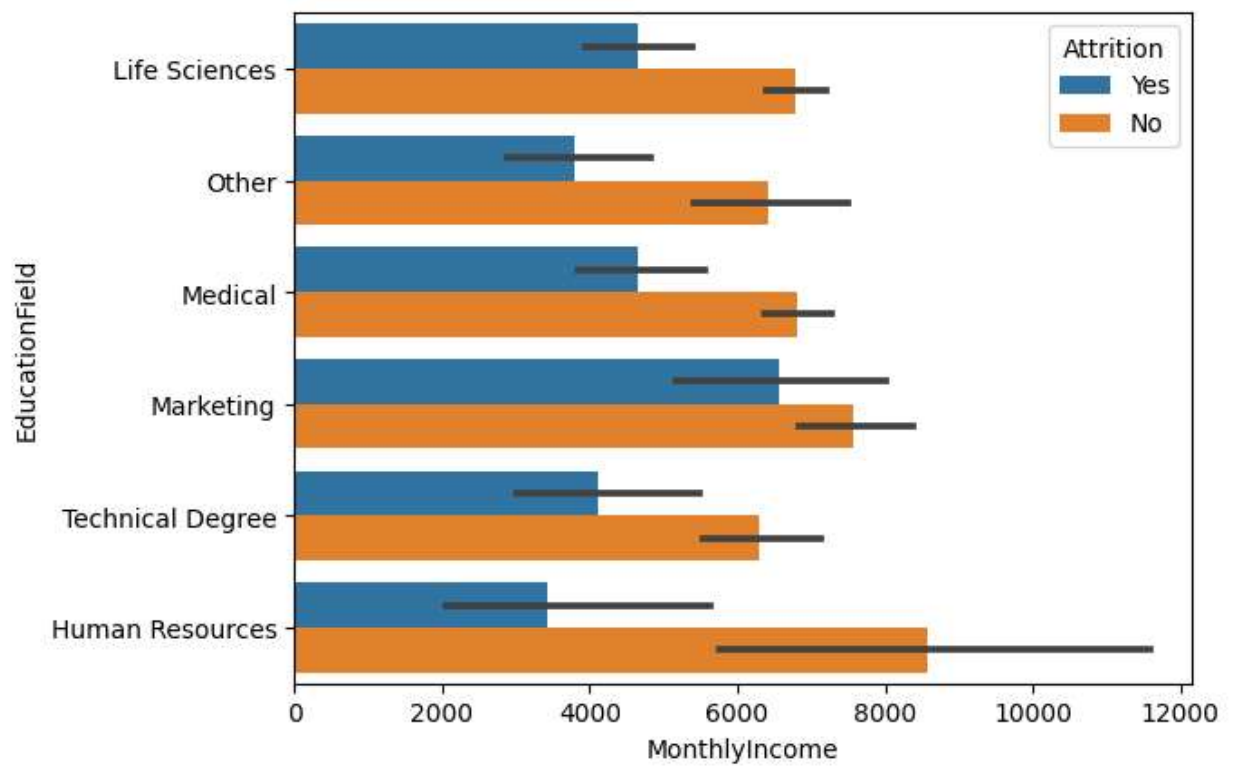
Attrition	No	Yes
Education		
1	5926.129496	4360.161290
2	6586.058824	4282.545455
3	6882.919662	4770.242424
4	7087.814706	5335.155172
5	8559.906977	5850.200000

```
In [14]: sns.barplot(x='DistanceFromHome', y='JobRole', hue='Attrition', data=df)
```

```
Out[14]: <AxesSubplot: xlabel='DistanceFromHome', ylabel='JobRole'>
```



```
In [15]: sns.barplot(x='MonthlyIncome', y='EducationField', hue='Attrition', data=df)
plt.show()
```



In []: