

Introduction

Employee retention strategies are integral to the success and well-being of a company. There are often many reasons why employees leave an organization, and in this project, I will explore some of the key drivers of employee attrition. Employee attrition measures how many workers have left an organization and is a common metric companies use to assess their performance. While turnover rates vary from industry to industry, the Bureau of Labor Statistics reported that among voluntary separations the overall turnover rate was 25% in 2020.

In this notebook, I will explore IBM's dataset on HR Analytics. The data consists of nearly 1,500 current and former employees with information related to their job satisfaction, work-life balance, tenure, experience, salary, and demographic data. Below is a brief overview and summary statistics of the data.

```
In [1]: #Import the libraries

In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
import plotly
import plotly.express as px

In [3]: #Import the data

In [4]: hr_data = pd.read_csv(r'C:\Users\ASUS\Downloads\WA_Fn-UseC-HR-Employee-Atrition.csv')

In [5]: hr_data.head()
```

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	YearsInCurrentRole
0	41	Yes	Travel_Rarely	1102	Sales	1	Life Sciences	1	1	...	1	80	0	8	8
1	49	No	Travel_Frequently	279	Research & Development	2	Life Sciences	1	2	...	4	80	1	10	10
2	37	Yes	Travel_Rarely	1373	Research & Development	2	Other	1	4	...	2	80	0	7	7
3	32	No	Travel_Frequently	1392	Research & Development	3	Life Sciences	1	5	...	3	80	0	8	8
4	27	No	Travel_Rarely	591	Research & Development	2	Medical	1	7	...	4	80	1	6	6

5 rows x 35 columns

EDA

```
In [6]: # Data Cleaning
# Exploratory Data Analysis (EDA)

In [7]: hr_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1476 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Age                 1476 non-null   int64
 1   Attrition           1476 non-null   object
 2   BusinessTravel      1476 non-null   object
 3   DailyRate           1476 non-null   int64
 4   Department          1476 non-null   object
 5   DistanceFromHome    1476 non-null   int64
 6   Education            1476 non-null   object
 7   EducationField       1476 non-null   object
 8   EmployeeCount       1476 non-null   int64
 9   EmployeeNumber      1476 non-null   int64
10   EnvironmentSatisfaction 1476 non-null   int64
11   Gender              1476 non-null   object
12   HourlyRate          1476 non-null   int64
13   JobInvolvement      1476 non-null   int64
14   JobLevel            1476 non-null   int64
15   JobRole             1476 non-null   object
16   JobSatisfaction      1476 non-null   int64
17   MaritalStatus       1476 non-null   object
18   MonthlyIncome       1476 non-null   int64
19   MonthlyRate         1476 non-null   int64
20   NumCompaniesWorked  1476 non-null   int64
21   Over18              1476 non-null   object
22   OverTime            1476 non-null   object
23   PercentSalaryHike    1476 non-null   int64
24   PerformanceRating    1476 non-null   int64
25   RelationshipSatisfaction 1476 non-null   int64
26   StandardHours       1476 non-null   int64
27   StockOptionLevel    1476 non-null   int64
28   TotalWorkingYears   1476 non-null   int64
29   TrainingTimesLastYear 1476 non-null   int64
30   WorkLifeBalance     1476 non-null   int64
31   WorkLifeBalance     1476 non-null   int64
32   YearsInCurrentRole  1476 non-null   int64
33   YearsSinceLastPromotion 1476 non-null   int64
34   YearsWithCurrManager 1476 non-null   int64
dtypes: int64(26), object(9)
memory usage: 480.1+ KB
```

```
In [28]: hr_data.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
WorkLifeBalance     0
YearsInCurrentRole  0
YearsSinceLastPromotion 0
YearsWithCurrManager 0
dtype: int64
```

```
In [219]: print("There are {} missing values in the data.".format(hr_data.isnull().sum().sum()))

There are 0 missing values in the data.

There are no missing values in dataset so we can proceed with visualization
```

Visualizing metrics using barplots

```
In [256]: # Barplot no 1 (Attrition of employees for EducationField v/s MonthlyIncome gender wise.)

sns.set(style='darkgrid', font_scale=1.3)
sns.catplot(x="EducationField", y="MonthlyIncome", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.3, palette="bar", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be854d73d>
```

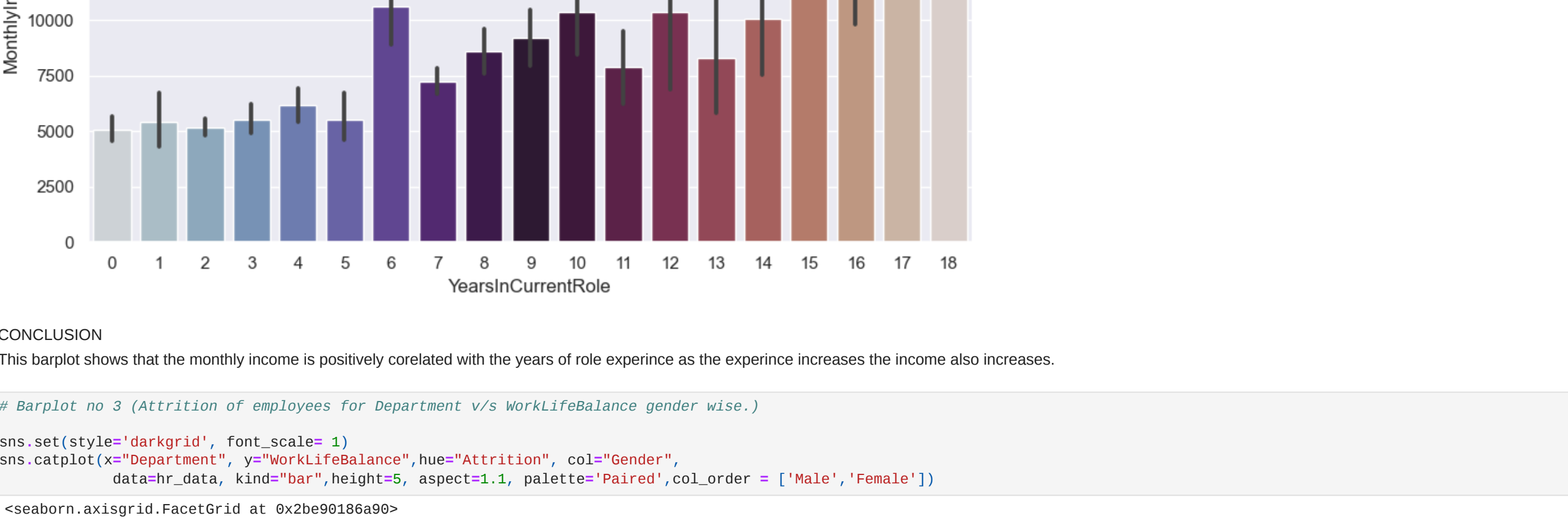


CONCLUSION
In barplot no 1 mostly attrition of male employees are from marketing field with monthly income less than 6000 and for female employees it is also from marketing field with monthly income around 7000

```
In [319]: # Barplot no 2 (Attrition of employees for YearsInCurrentRole v/s MonthlyIncome.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="YearsInCurrentRole", y="MonthlyIncome",
            data=hr_data, kind="bar", height=5, aspect=1.6, palette="twilight")

<seaborn.axisgrid.FacetGrid at 0x2bea235d4c>
```

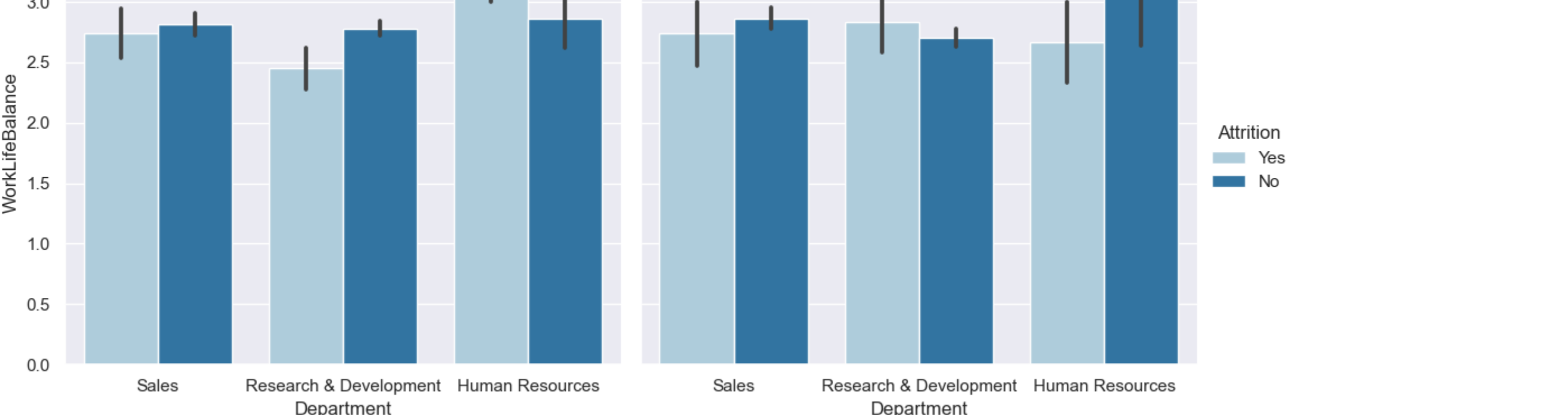


CONCLUSION
This barplot shows that the monthly income is positively correlated with the years of role experience as the experience increases the income also increases.

```
In [256]: # Barplot no 3 (Attrition of employees for Department v/s WorkLifeBalance gender wise.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="Department", y="WorkLifeBalance", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.1, palette="Paired", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be90186a8>
```

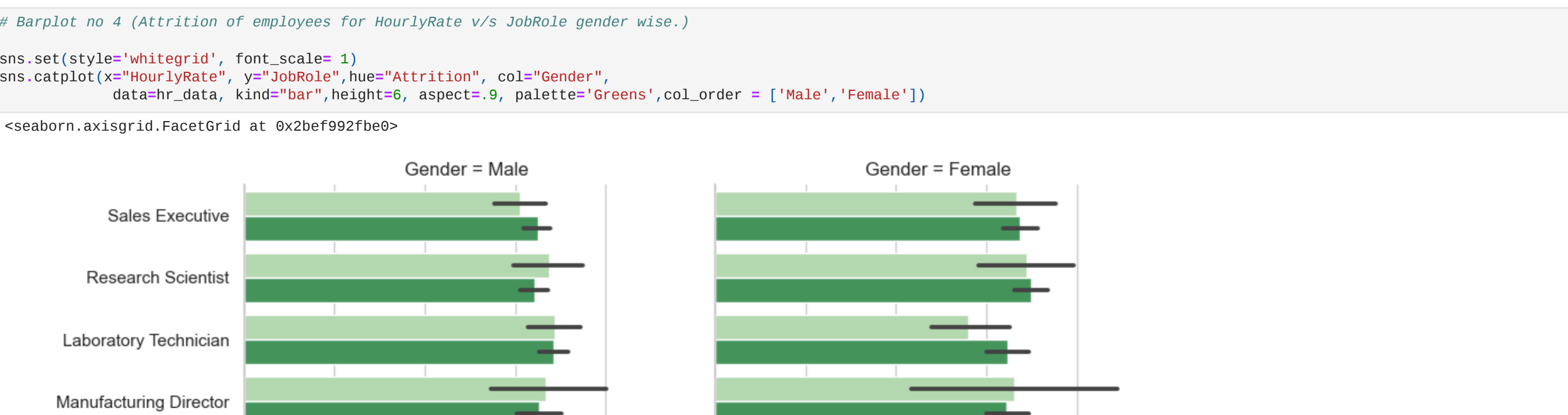


CONCLUSION
In barplot no 3 attrition of male employees are from Human resources department whose work life balance are more than 3 and for female employees it is from Research and development department whose work life balance are between 2.5 and 3

```
In [284]: # Barplot no 4 (Attrition of employees for HourlyRate v/s JobRole gender wise.)

sns.set(style='whitegrid', font_scale=1)
sns.catplot(x="HourlyRate", y="JobRole", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.2, palette="greens", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9927b2b>
```

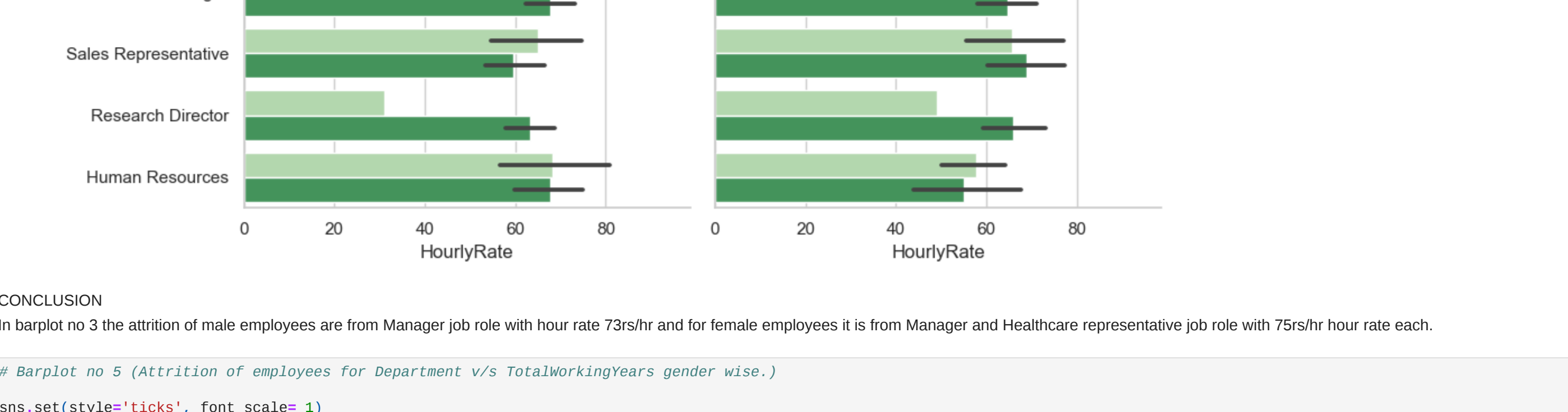


CONCLUSION
In barplot no 3 the attrition of male employees are from Manager job role with hour rate 73\$shr and for female employees it is from Manager and Healthcare representative job role with 75\$shr hour rate each.

```
In [256]: # Barplot no 5 (Attrition of employees for Department v/s TotalWorkingYears gender wise.)

sns.set(style='ticks', font_scale=1)
sns.catplot(x="Department", y="TotalWorkingYears", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.2, palette="icefire", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9d3c49e>
```

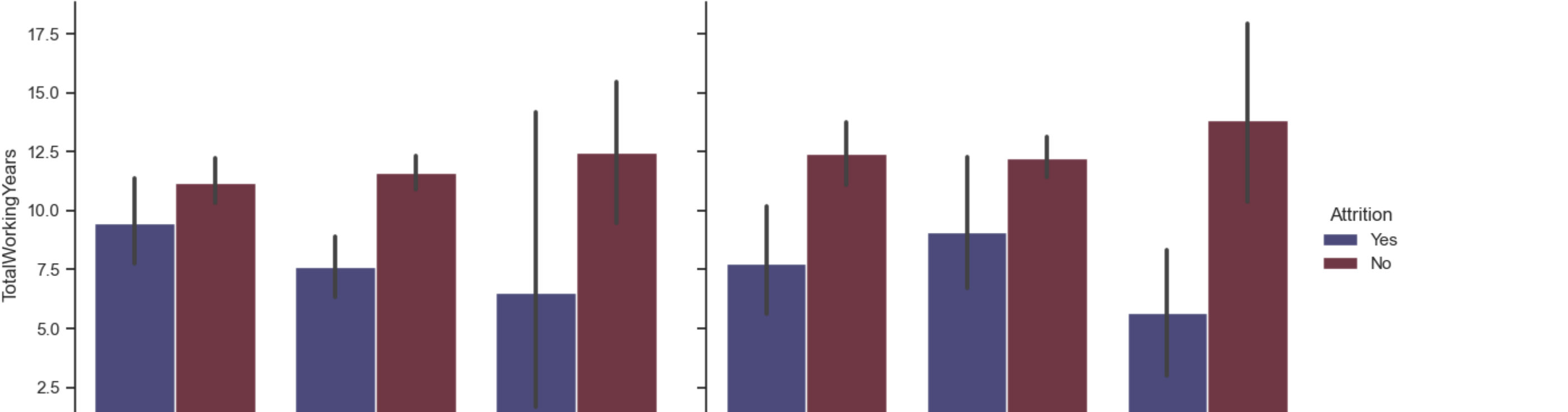


CONCLUSION
In this barplot the attrition of male employees are mostly from sales department with total working years of 9 years and other side for female employees it is from Research and development department with total working years 8.5 years.

```
In [292]: # Barplot no 6 (Attrition of employees for MaritalStatus v/s JobSatisfaction gender wise.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="MaritalStatus", y="JobSatisfaction", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.4, palette="Reds", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2beaf19b73>
```

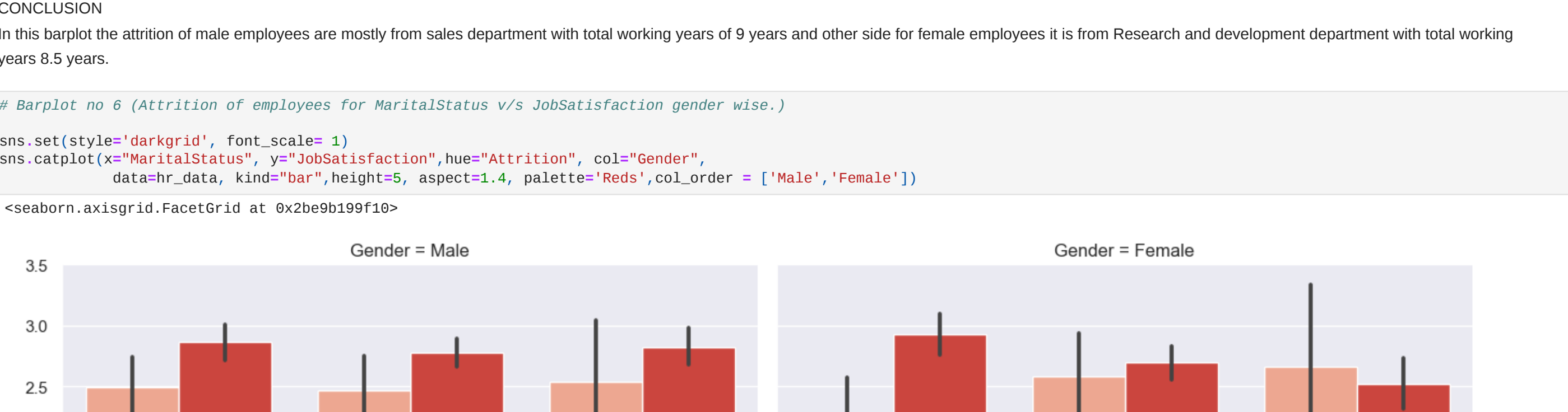


CONCLUSION
In this barplot we can see the employees which are divorced are taking attrition or thinking of it. male employees mostly with 2.6 jobsatisfaction and female employees with 2.8 jobsatisfaction.

```
In [328]: # Barplot no 7 (Attrition of employees for Department v/s MonthlyIncome.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="Department", y="MonthlyIncome", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.5, palette="twilight", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2beaf19b7c>
```



CONCLUSION
This barplot shows the highest monthly income is for sales department.

```
In [268]: # Barplot no 8 (Attrition of employees for Department v/s MonthlyIncome gender wise.)

sns.set(style='whitegrid', font_scale=1)
sns.catplot(x="Department", y="MonthlyIncome", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.4, palette="spring", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9d3c49e>
```

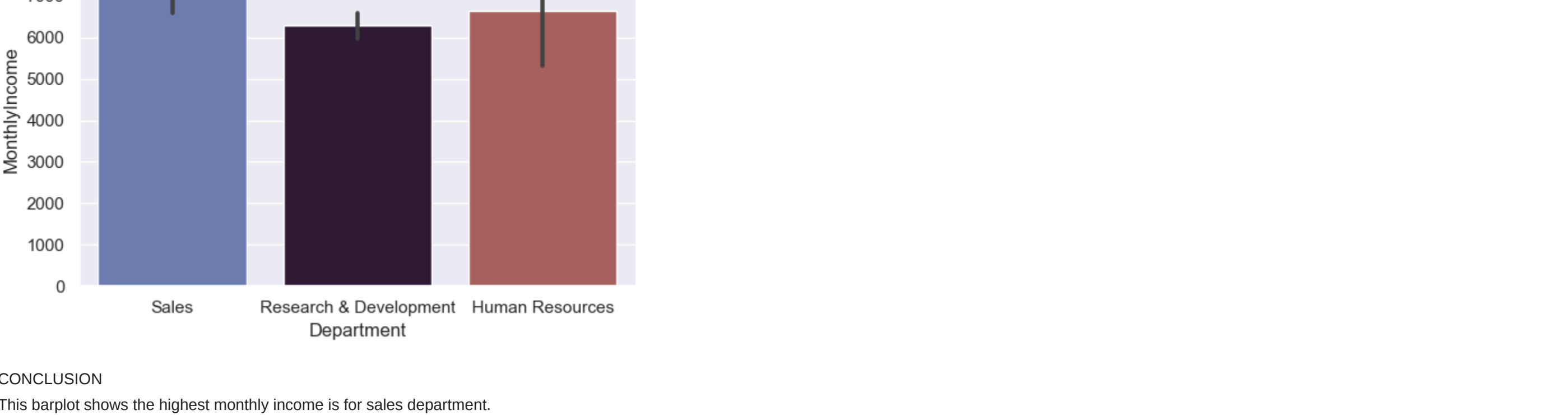


CONCLUSION
In this barplot the attrition of employees are from sales department with monthlyincome 6000 and for female employees attrition is also from sales department with monthlyincome 5700.

```
In [275]: # Barplot no 9 (Attrition of employees for JobLevel v/s JobRole gender wise.)

sns.set(style='whitegrid', font_scale=1)
sns.catplot(x="JobLevel", y="JobRole", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.3, palette="seismic", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9b0b95b>
```

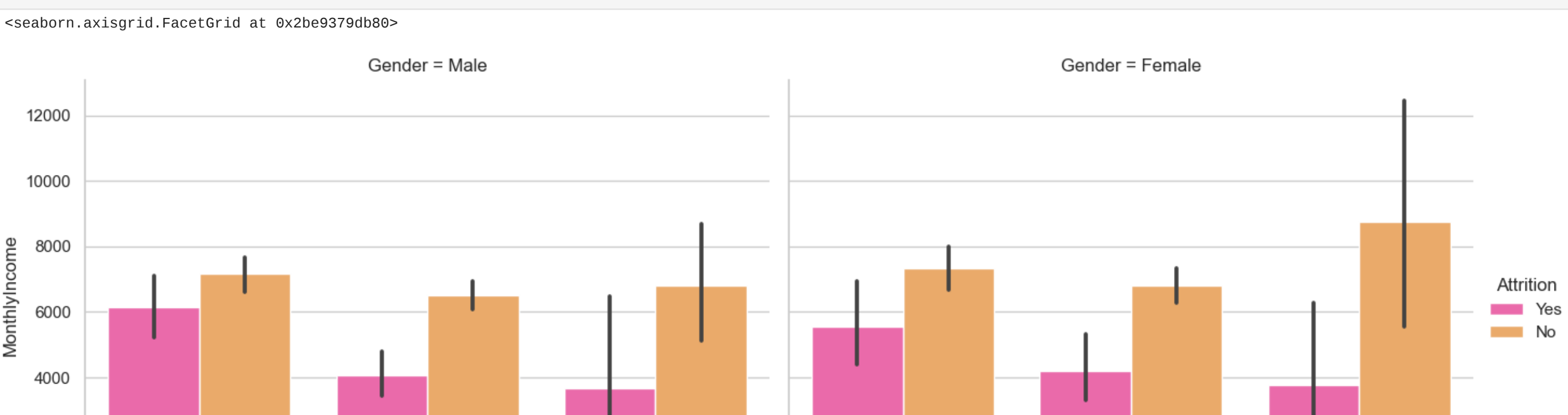


CONCLUSION
In this barplot the attrition of male employees are from research director's job role whose joblevel is 5 and also for female employees the attrition is from research director's role with same joblevel 5.

```
In [277]: # Barplot no 10 (Attrition of employees for EnvironmentSatisfaction v/s EducationField gender wise.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="EducationField", y="EnvironmentSatisfaction", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=6, aspect=1.3, palette="turbo_r", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be942c27f>
```



CONCLUSION
In this barplot the attrition of male employees are from other educational fields with environment satisfaction 3.2 and for female employees attrition is from same other educationfield with environmentsatisfaction 3.5.

```
In [288]: # Barplot no 11 (Attrition of employees for BusinessTravel v/s JobLevel gender wise.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="BusinessTravel", y="JobLevel", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.2, palette="copper", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9b0b48b>
```

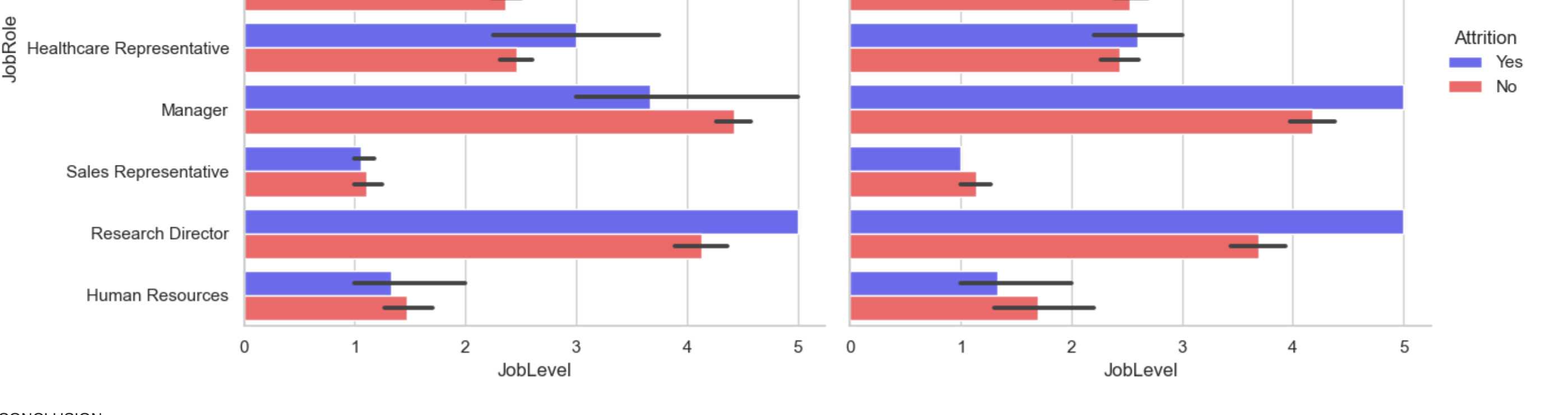


CONCLUSION
In this barplot the attrition of male employees are non travelers whose job levels are between 1.5 and 2 and female employees attrition are also non travelers whose job levels are 2.

```
In [285]: # Barplot no 12 (Attrition of employees for YearsInCurrentRole v/s JobRole gender wise.)

sns.set(style='whitegrid', font_scale=1.3)
sns.catplot(x="YearsInCurrentRole", y="JobRole", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.1, palette="binary", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9b0b63d>
```



CONCLUSION
In this barplot the attrition of male employees are research directors playing same role from more than 14 years and also same for female employees research directors playing same role from more than 14 years.

```
In [289]: # Barplot no 13 (Attrition of employees for MaritalStatus v/s MonthlyIncome.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="MaritalStatus", y="MonthlyIncome", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.3, palette="twilight", col_order = ['Male', 'Female'])

<seaborn.axisgrid.FacetGrid at 0x2be9b0b93a>
```

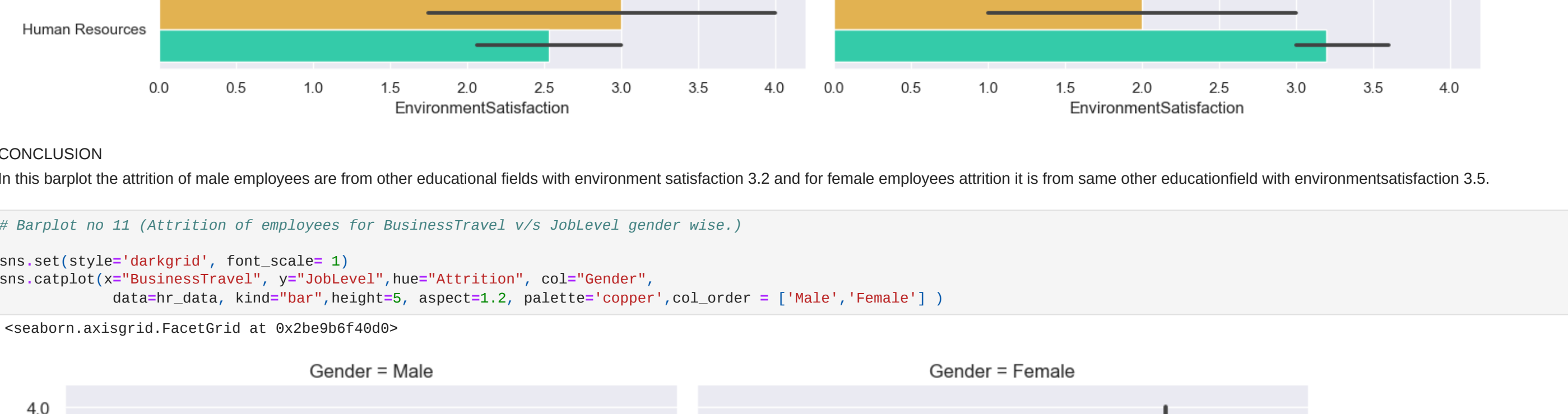


CONCLUSION
This barplot shows the attrition of male employees which are married and have monthlyincome 5500 and also married females with monthlyincome 6000 have done attrition.

```
In [333]: # Barplot no 14 (Attrition of employees for JobRole v/s MonthlyIncome.)

sns.set(style='darkgrid', font_scale=1)
sns.catplot(x="JobRole", y="MonthlyIncome", hue="Attrition", col="Gender",
            data=hr_data, kind="bar", height=8, aspect=1.2, palette="twilight")

<seaborn.axisgrid.FacetGrid at 0x2bf02b783e>
```

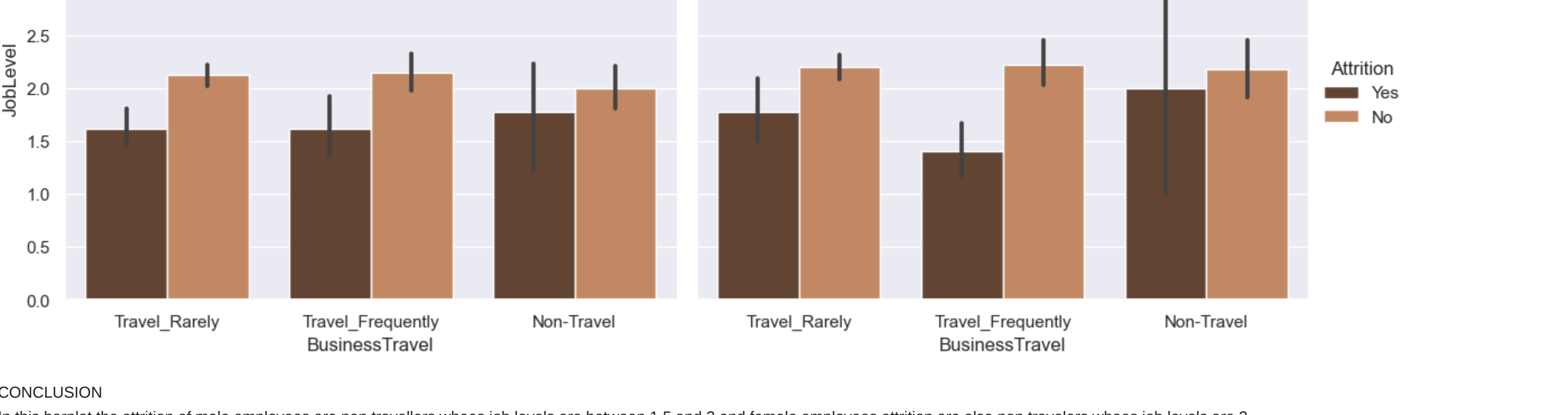


CONCLUSION
This barplot shows the highest monthly income is for Manager job role.

```
In [332]: # Barplot no 15 (Attrition of employees for Jobsatisfaction v/s MonthlyIncome.)

sns.set(style='ticks', font_scale=1)
sns.catplot(x="Jobsatisfaction", y="MonthlyIncome",
            data=hr_data, kind="bar", height=8, aspect=1.2, palette="greens")

<seaborn.axisgrid.FacetGrid at 0x2bf02b78d6>
```



CONCLUSION
It shows that employees with monthly income more than 6500 have job satisfaction 1.

```
In [ ]:
```