analysis

June 12, 2021

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

[3]:
```
# Load Data
df = pd.read_csv('attrition_data.csv')

# Convert outcome to binary
df['Attrition'] = np.where(df['Attrition']=='Yes', 1, 0)
```

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

[4]:

	Attrition	Age	BusinessTravel	Department	\
0	1	41	Travel_Rarely	Sales	
1	0	49	Travel_Frequently	Research & Development	
2	1	37	Travel_Rarely	Research & Development	
3	0	33	Travel_Frequently	Research & Development	
4	0	27	Travel_Rarely	Research & Development	

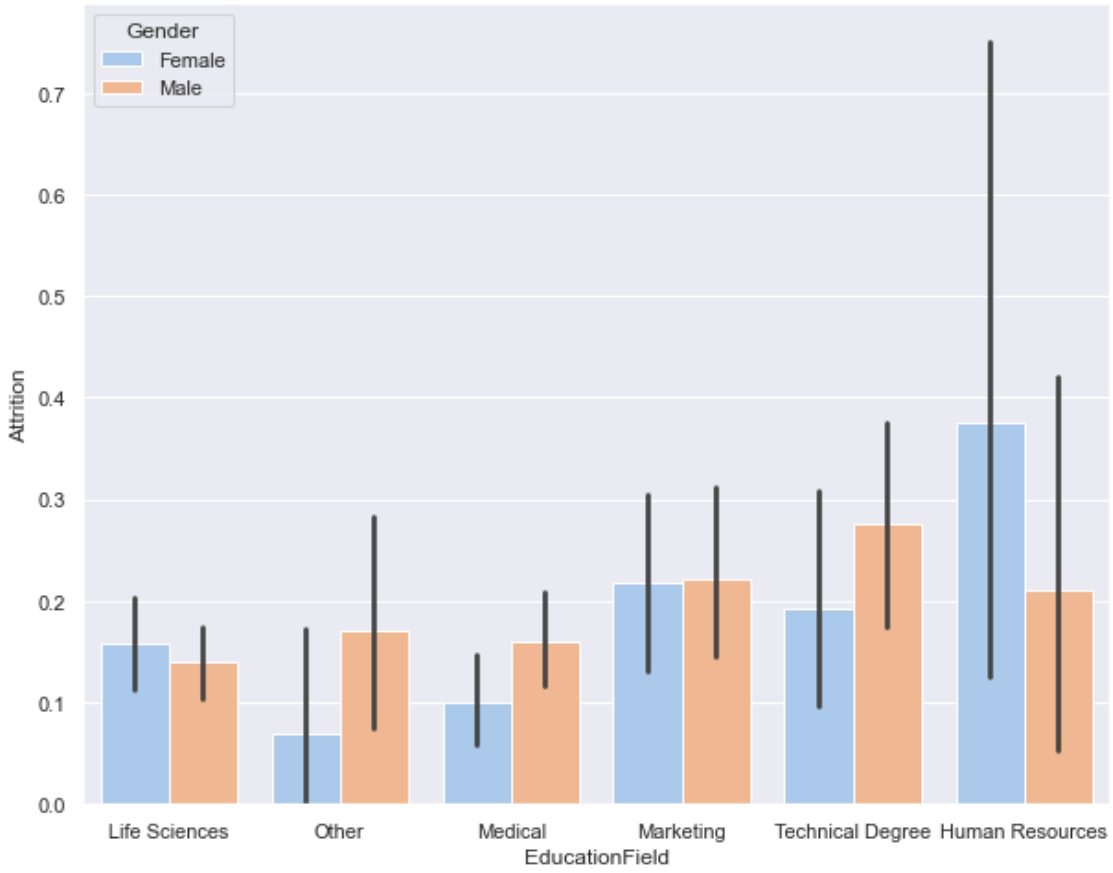
	DistanceFromHome	Education	EducationField	EnvironmentSatisfaction	\
0	1	2	Life Sciences	2	
1	8	1	Life Sciences	3	
2	2	2	Other	4	
3	3	4	Life Sciences	4	
4	2	1	Medical	1	

	Gender	JobInvolvement	…	OverTime	PercentSalaryHike	PerformanceRating	\
0	Female	3	…	Yes	11	3	
1	Male	2	…	No	23	4	
2	Male	2	…	Yes	15	3	
3	Female	3	…	Yes	11	3	
4	Male	3	…	No	12	3	

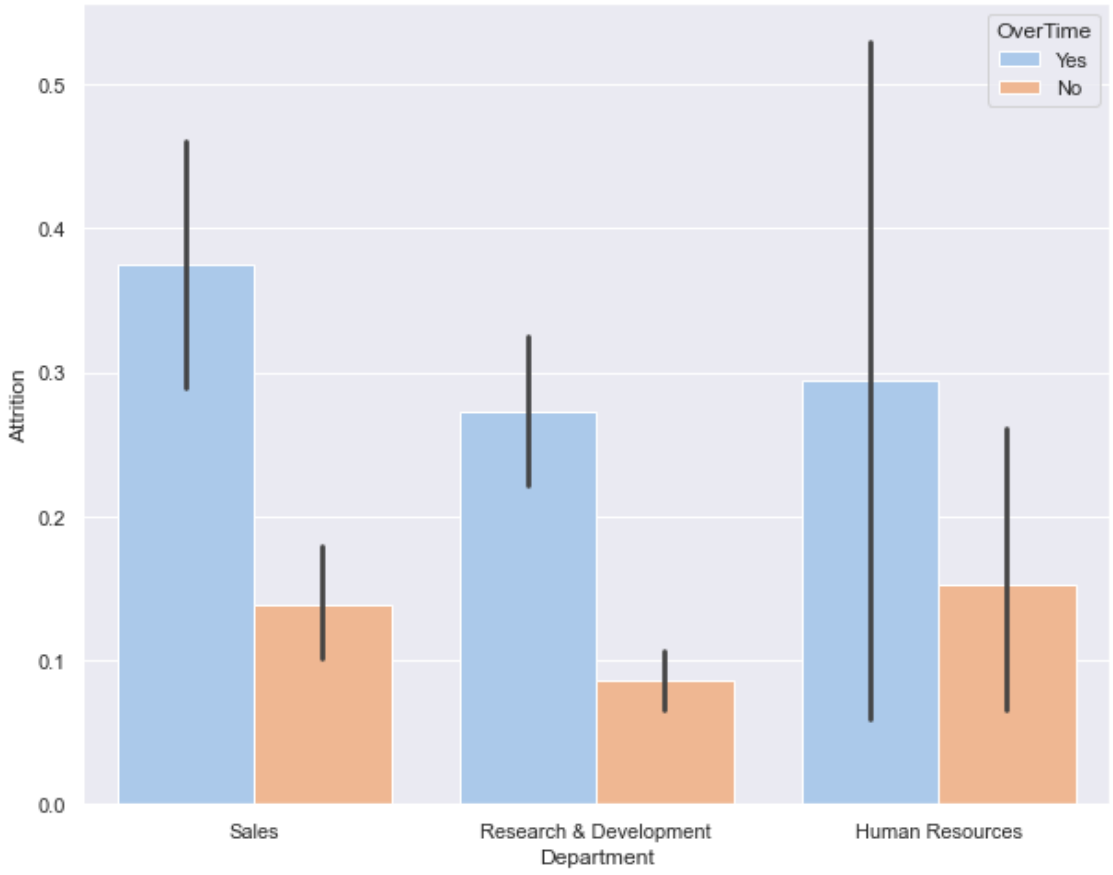
	RelationshipSatisfaction	TotalWorkingYears	WorkLifeBalance	\
0	1	8	1	
1	4	10	3	

```
2                         2                  7                 3
3                         3                  8                 3
4                         4                  6                 3

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

   YearsWithCurrManager
0                     5
1                     7
2                     0
3                     0
4                     2

[5 rows x 25 columns]
```

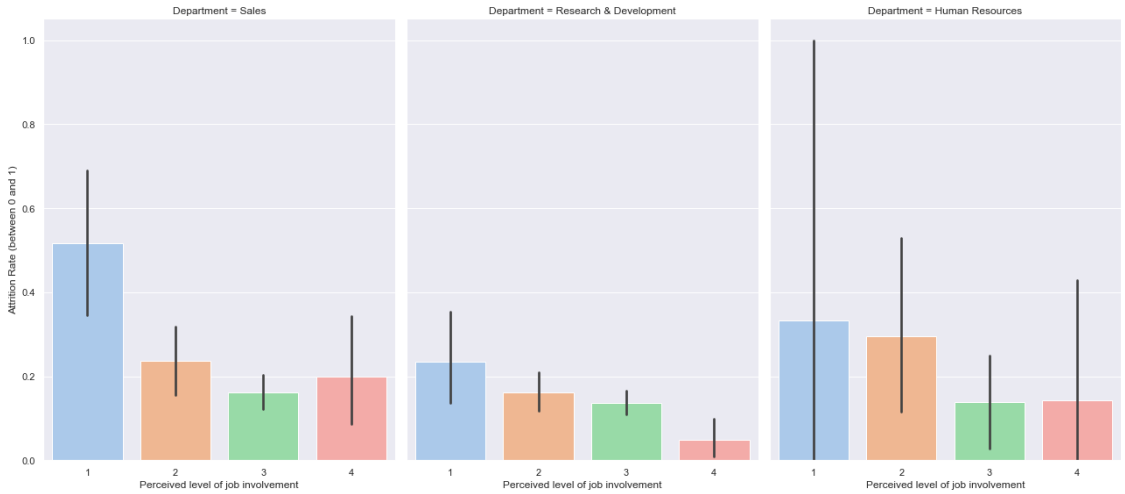
[5]:
```
# Gender and Education
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x='EducationField', y='Attrition', data=df, hue='Gender',␣
 ↪palette='pastel')
plt.show()
```



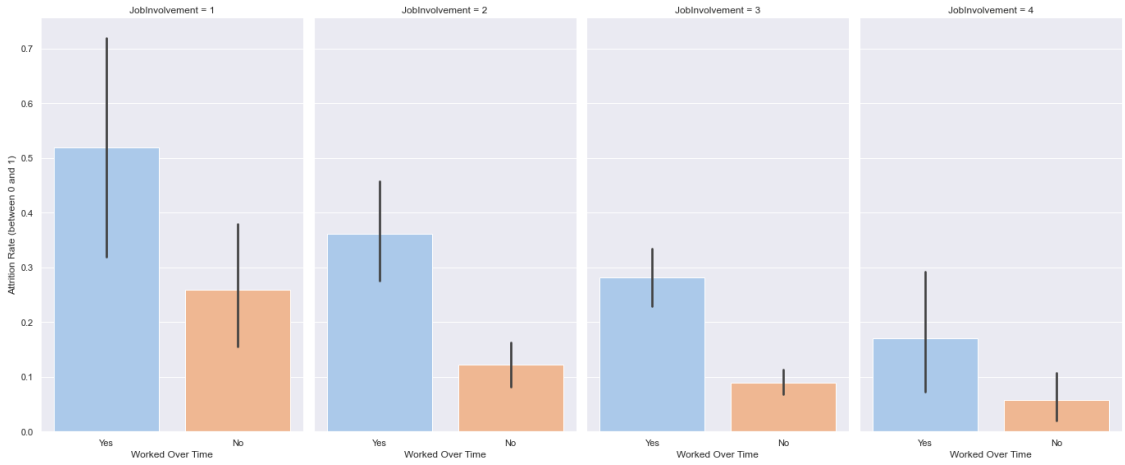
[6]:
```
# Department and Overtime
sns.set(rc={'figure.figsize':(10,8)})
sns.barplot(x='Department', y='Attrition', data=df, hue='OverTime',␣
 ↪palette='pastel')
plt.show()
```



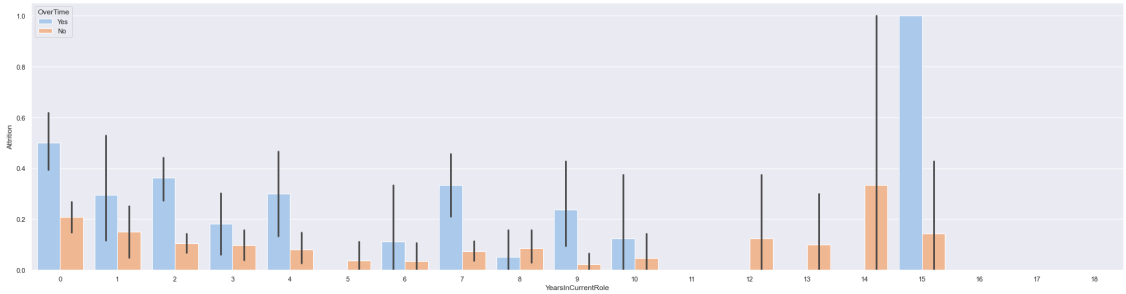
[7]:
```
# Environment and Job Involvement
s = sns.FacetGrid(df, col="Department", height=8, aspect=.75)
s.map(sns.barplot, 'JobInvolvement', 'Attrition', palette='pastel',␣
 ↪order=[1,2,3,4])
s.set_axis_labels("Perceived level of job involvement", "Attrition Rate␣
 ↪(between 0 and 1)")
plt.show()
```



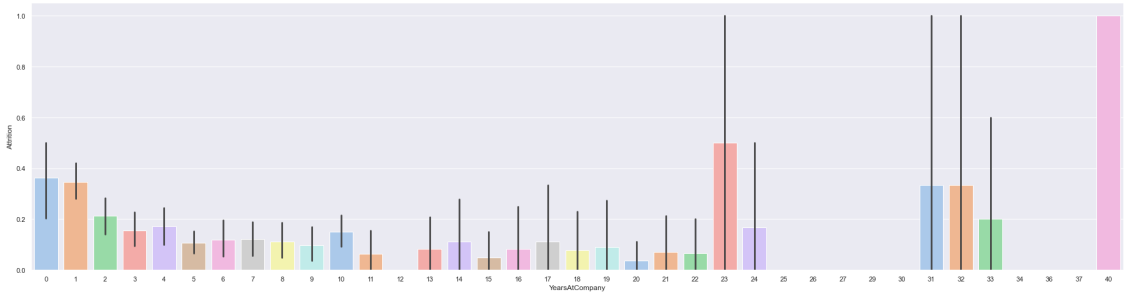
[8]:
```
# Overtime and Job Involvement
s = sns.FacetGrid(df, col="JobInvolvement", height=8, aspect=.6)
s.map(sns.barplot, 'OverTime', 'Attrition', palette='pastel', order=["Yes",␣
 ↪"No"])
s.set_axis_labels("Worked Over Time", "Attrition Rate (between 0 and 1)")
plt.show()
```



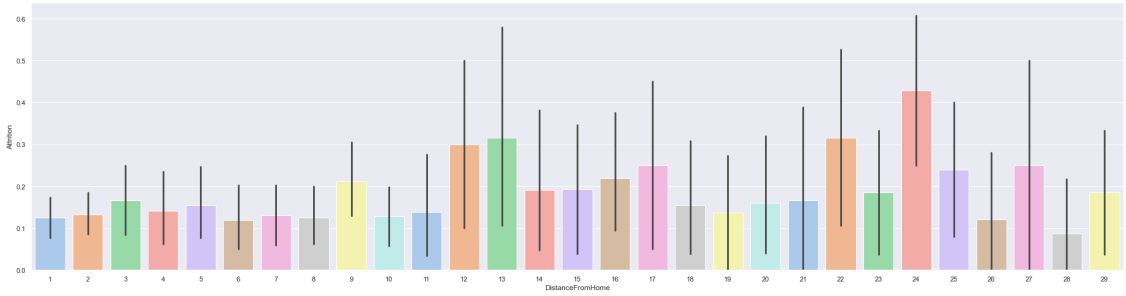
[9]:
```
# Years in current role
sns.set(rc={'figure.figsize':(32,8)})
sns.barplot(x='YearsInCurrentRole', y='Attrition', data=df, hue='OverTime',␣
 ↪palette='pastel')
plt.show()
```



[10]:
```
# Years at the same company
sns.set(rc={'figure.figsize':(32,8)})
sns.barplot(x='YearsAtCompany', y='Attrition', data=df, palette='pastel')
plt.show()
```

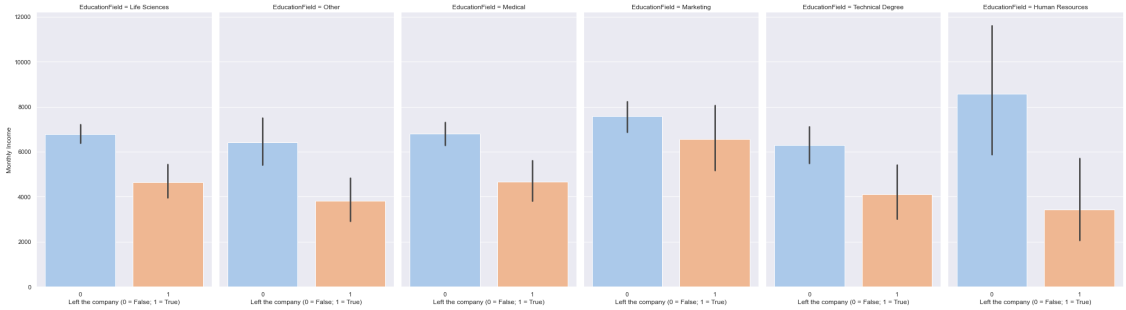


[11]:
```
# Distance From Home
sns.set(rc={'figure.figsize':(32,8)})
sns.barplot(x='DistanceFromHome', y='Attrition', data=df, palette='pastel')
plt.show()
```



[12]:
```
# Education, Attrition and Monthly Income
s = sns.FacetGrid(df, col="EducationField", height=8, aspect=.6)
s.map(sns.barplot, 'Attrition', 'MonthlyIncome', palette='pastel', order=[0, 1])
```

```
s.set_axis_labels("Left the company (0 = False; 1 = True)", "Monthly Income")
plt.show()
```



[13]:
```
# Conclusion

'''
The turnover of employees seems to be driven by three main factors - monthly␣
 ↪income, perceived level of job involvement and working over time. On␣
 ↪average, employees that left the company
earned markedly lower wages than those who chose to remain. Similarly, the␣
 ↪attrition rate was highest among employees who were the most dissatisfied␣
 ↪with the level of involvement in their
job. Therefore, if employees feel left out or insufficiently engaged, then␣
 ↪their incentive to leave is heightened. Lastly, another key differentiator␣
 ↪is the possibility of working over time.
Specifically, employees who did work over time demonstrated a significantly␣
 ↪higher likelihood of leaving the company. Taken together with the other two␣
 ↪factors discussed above, it appears that
the risk of burnout and discontentment is much more pronounced when the␣
 ↪employees have to work beyond their regular remit. All of the above problems␣
 ↪were most visible within the Sales and Human
Resources departments.
'''
```