GREEDESTINATION EMPLOYEE ATTRITION PROJECT

- The aim of this project is to analyse an employee data set containing data regarding employess leaving their jobs ad other importand employee data
- Finding pattefns in the data to get an insight into the various factors affecting the employee attrition(Quitting the job)

In [29]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.subplots as sp
import plotly.graph_objects as go
```

In [2]:
```
greendata=pd.read_csv("greendestination (1).csv")
greendata.head()
```

Out[2]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7

5 rows × 35 columns

In [4]:
```
#view data frame information
greendata.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 [5]:
```
ar=(len(greendata[greendata["Attrition"]=="Yes"])/(len(greendata["Attrition"])))*100
print(f"The total attrition rate I.E rate of employees leaving their jobs is {round(ar,1)} %")
```

```
The total attrition rate I.E rate of employees leaving their jobs is 16.1 %
```

In [6]:
```
#cleaning outliers from the data set using the IQR method
for i in greendata.columns:
  if greendata[i].dtype=="int64":
    Q1=greendata[i].quantile(0.25)
    Q3=greendata[i].quantile(0.75)
    IQR=Q3-Q1
    greendata=greendata[greendata[i]>=Q1-1.5*IQR]
    greendata=greendata[greendata[i]<= Q3+1.5*IQR]
```

In [7]:
```
#Viewing the cleaned data set
greendata=greendata.reset_index(drop=True)
greendata
```

Out[7]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumb
0	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
1	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
2	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	
3	29	No	Travel_Rarely	153	Research & Development	15	2	Life Sciences	1	
4	31	No	Travel_Rarely	670	Research & Development	26	1	Life Sciences	1	
...	...	...	...	...	...	...	...	...	...	
694	31	No	Non-Travel	325	Research & Development	5	3	Medical	1	20
695	26	No	Travel_Rarely	1167	Sales	5	3	Other	1	20
696	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	20
697	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	20
698	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	20

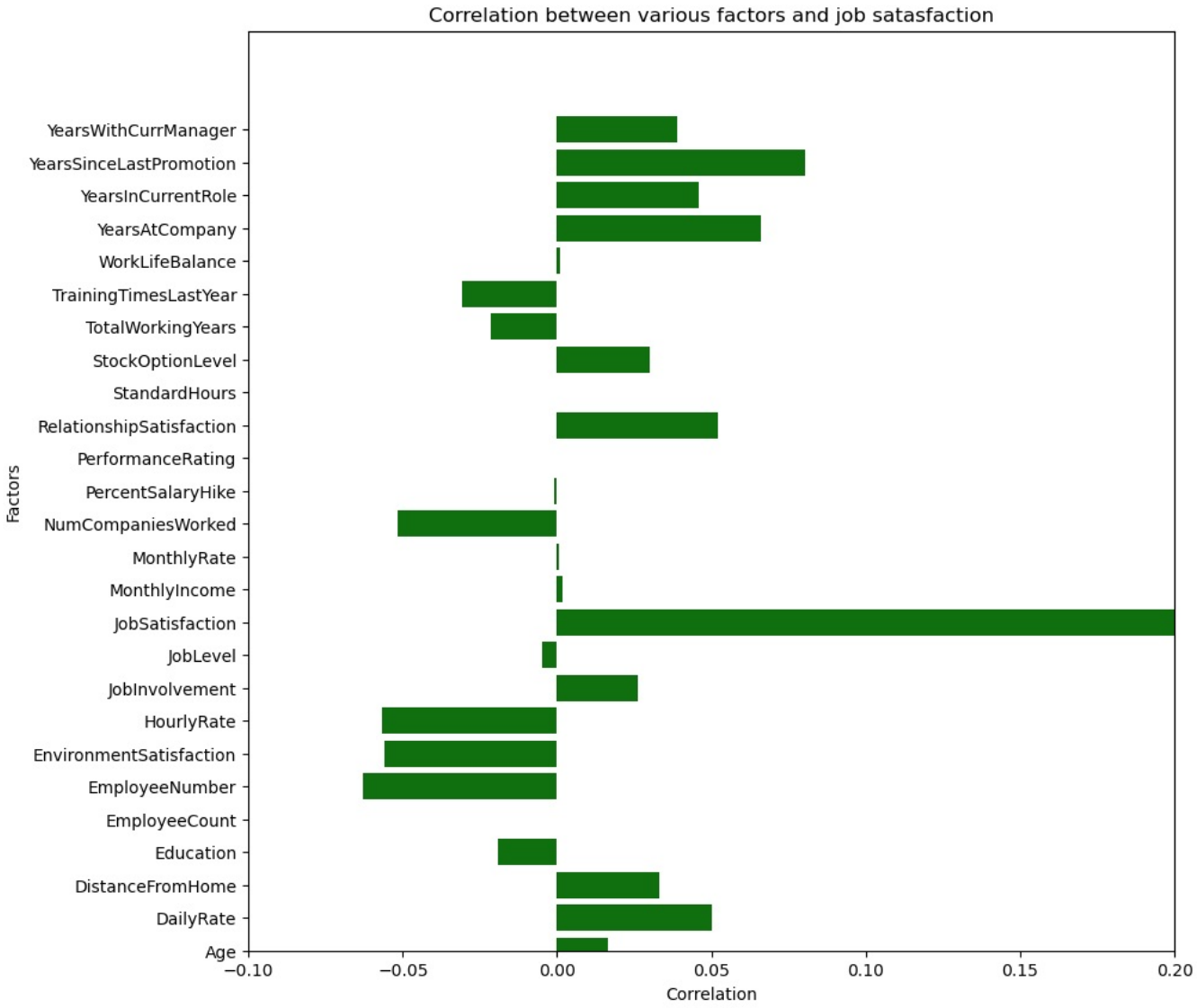
699 rows × 35 columns

In [11]:
```
#Creating a seperate data frame from the columns in the greendata containing numerical data type
num_data=greendata.select_dtypes(include=['int64'])
```

In [12]:
```
#Creating a correlation dataframe using the corr() function.The corr() function displays the correlation betwee
cor=num_data.corr()
```

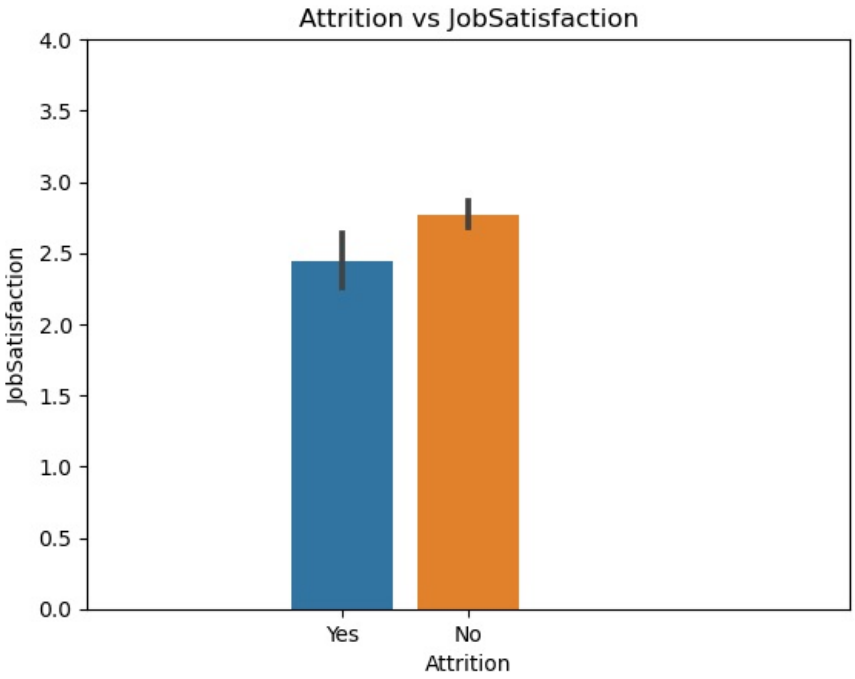
In [14]:
```
#Ploting the correlation data using the seaborn barplot method
plt.figure(figsize=(10,10))
sns.barplot(y=num_data.columns,x=cor["JobSatisfaction"],color="green")
plt.xlabel("Correlation")
plt.ylabel("Factors")
plt.axis([-0.1,0.2,0,28])
plt.title("Correlation between various factors and job satasfaction")

plt.show()
```



In [15]:
```
#Ploting jobsatisfaction data against attrition
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["JobSatisfaction"])
plt.xlabel("Attrition")
plt.ylabel("JobSatisfaction")
plt.title("Attrition vs JobSatisfaction")
plt.axis([-2,4,0,4])
```

Out[15]:
```
(-2.0, 4.0, 0.0, 4.0)
```



From the plot above its clear that :-

- 1) The average jobsatisfaction score of employees quitting their jobs is lower than those not leaving their jobs
- 2)It can be concluded hence that jobsatisfaction is an important factor in employees leaving jobs

In [17]:
```
#Creating seperate dataframes for different job satisfaction scores
jbs4=greendata[greendata["JobSatisfaction"]==4]
jbs3=greendata[greendata["JobSatisfaction"]==3]
jbs2=greendata[greendata["JobSatisfaction"]==2]
jbs1=greendata[greendata["JobSatisfaction"]==1]
```

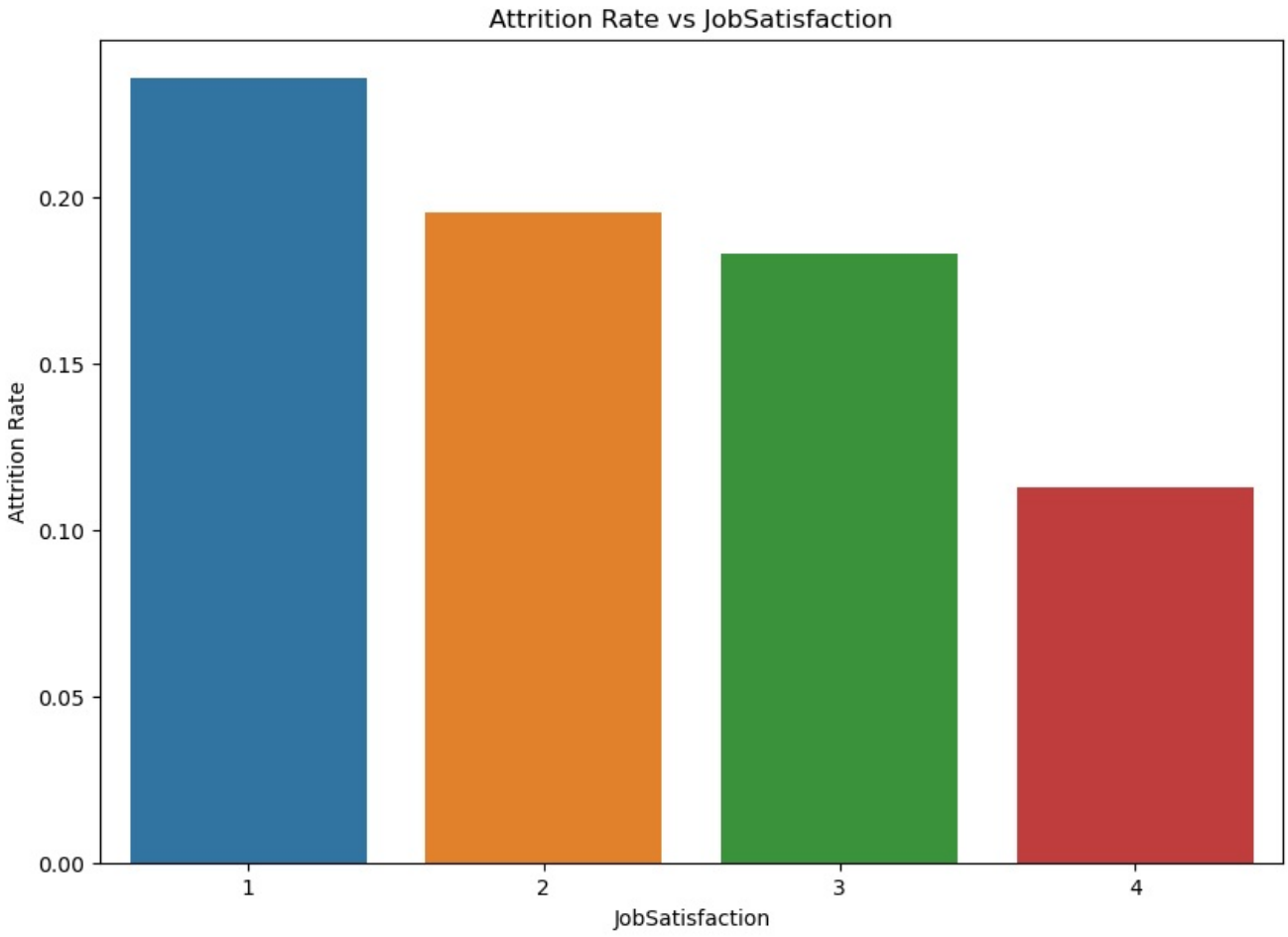
In [18]:
```
#Printing attrition rates(Number of employees with a particular job satisfaction score leaving their jobs/Total
# for different job satisfaction scores
print(f"Rate of attrition for employees with job satisfaction  4 is {len(jbs4[jbs4['Attrition']=='Yes'])/len(jb
print(f"Rate of attrition for employees with job satisfaction  3 is {len(jbs3[jbs3['Attrition']=='Yes'])/len(jb
print(f"Rate of attrition for employees with job satisfaction  2 is {len(jbs2[jbs2['Attrition']=='Yes'])/len(jb
print(f"Rate of attrition for employees with job satisfaction  1 is {len(jbs1[jbs1['Attrition']=='Yes'])/len(jb
```

```
Rate of attrition for employees with job satisfaction  4 is 0.11267605633802817
Rate of attrition for employees with job satisfaction  3 is 0.18309859154929578
Rate of attrition for employees with job satisfaction  2 is 0.19548872180451127
Rate of attrition for employees with job satisfaction  1 is 0.2357142857142857
```

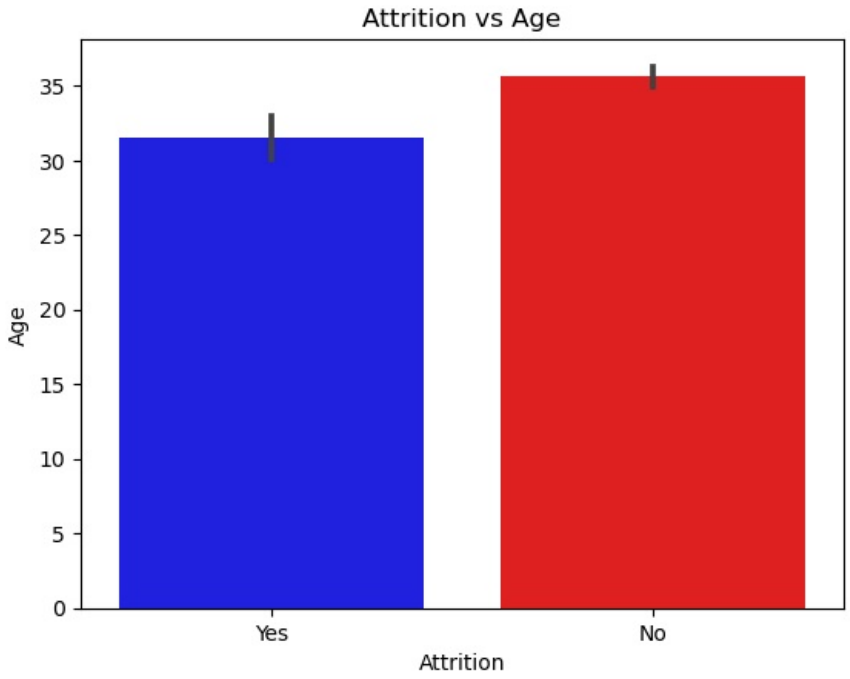
It can be established from the above statements that an employee with a jobsatisfaction score of 1 is twice more likely to leave his job than an employee with a job satisfaction score of 4 Lesser the job satisfaction higher are the chances of an employee leaving their job

Below is a barplot demonstrating the same

In [71]:
```
plt.figure(figsize=(10,7))
sns.barplot(x=[1,2,3,4],y=[len(jbs1[jbs1['Attrition']=='Yes'])/len(jbs1),len(jbs2[jbs2['Attrition']=='Yes'])/le
plt.xlabel("JobSatisfaction")
plt.ylabel("Attrition Rate")
plt.title("Attrition Rate vs JobSatisfaction")
plt.show()
```



In [90]:
```
#A barplot showing the realation between attrition and age of employees
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["Age"],palette=('blue','red'))
plt.xlabel("Attrition")
plt.ylabel("Age")
plt.title("Attrition vs Age")
plt.show()
```



In [24]:
```
#It can be clearly noticed that the average age of an employee leaving their job is lesser than those not leavi

a1=greendata[greendata["Attrition"]=="Yes"]["Age"].mean()
a2=(greendata[greendata["Attrition"]=="No"]["Age"].mean())
print(f"Average age of  an employee leaving their job is {a1} ")
print(f"Average age of an employee not leaving their job is {a2} ")
```

```
Average age of  an employee leaving their job is 31.516393442622952
Average age of an employee not leaving their job is 35.61005199306759
```

In [95]:
```
x=greendata["Attrition"]
x
greendata.head()
```

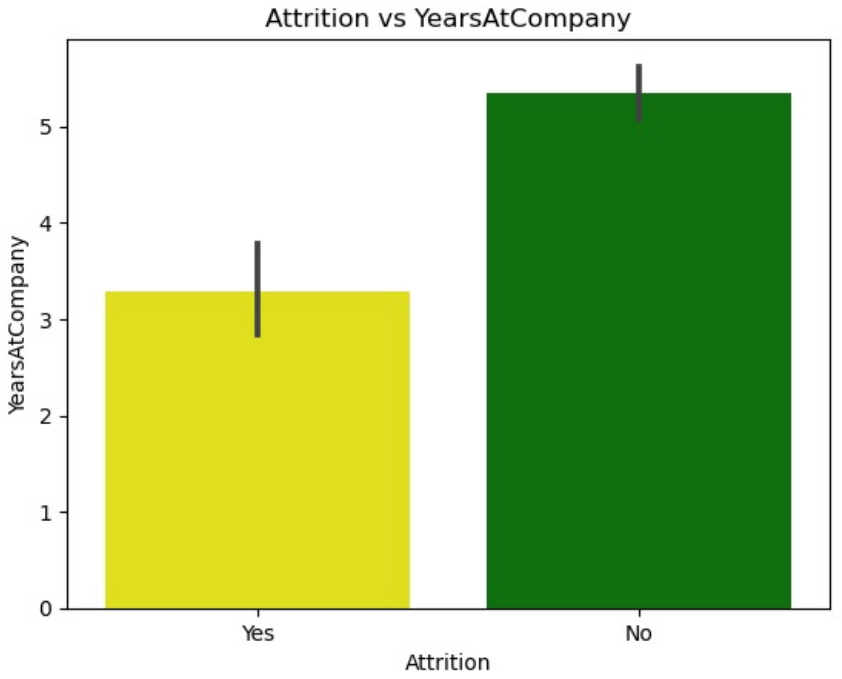
Out[95]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4
1	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5
2	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8
3	29	No	Travel_Rarely	153	Research & Development	15	2	Life Sciences	1	15
4	31	No	Travel_Rarely	670	Research & Development	26	1	Life Sciences	1	16

5 rows × 35 columns

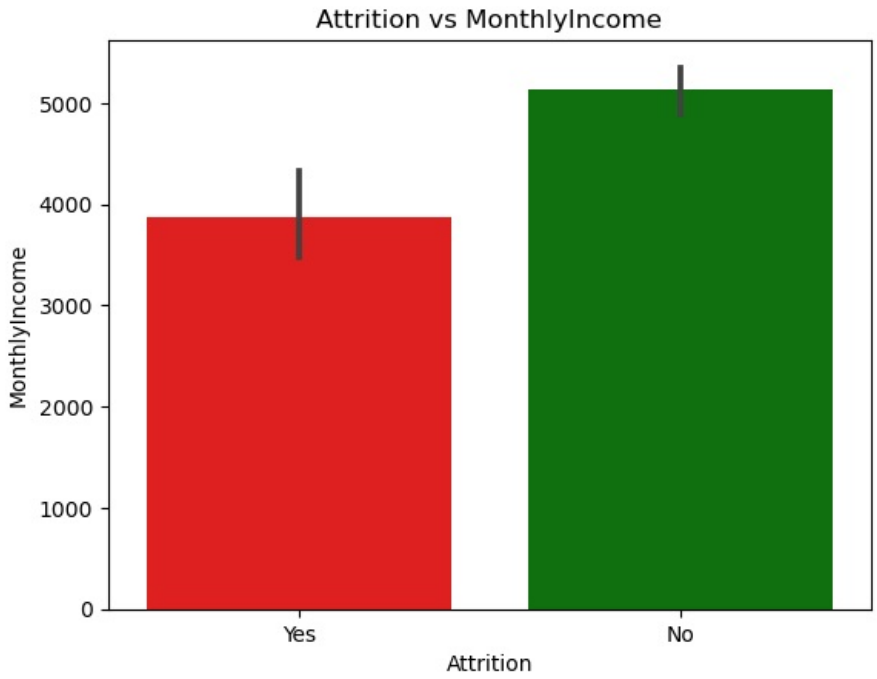
Its clear from the above information that a younger employee is more likely to quit their job than an older employee

In [104…
```
#using Barplot technique to plot average years at company of an employee for both attrition categories
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["YearsAtCompany"],palette=["yellow","green"])
plt.xlabel("Attrition")
plt.ylabel("YearsAtCompany")
plt.title("Attrition vs YearsAtCompany")
plt.show()
```



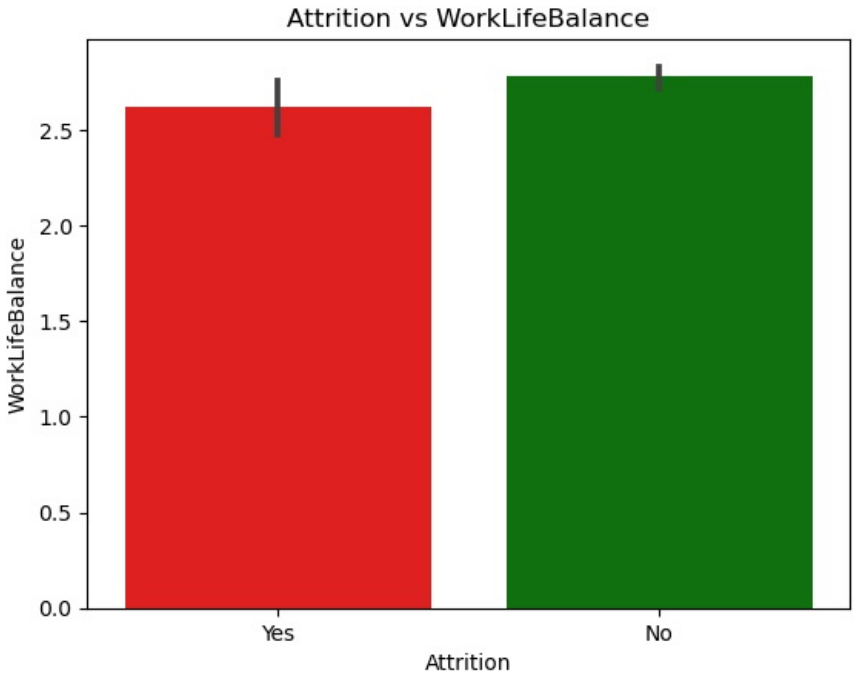
In [38]:
```
#The average years an employee leaving the company has spent working for the company is significantly lower tha
# Hence it can be concluded from this that the longer an employee has worked for the company the lesser are his
```

In [39]:
```
#barplot method used to show relationship between monthly income of the employees and their monthly incomes
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["MonthlyIncome"],palette=["red","green"])
plt.xlabel("Attrition")
plt.ylabel("MonthlyIncome")
plt.title("Attrition vs MonthlyIncome")
plt.show()
```



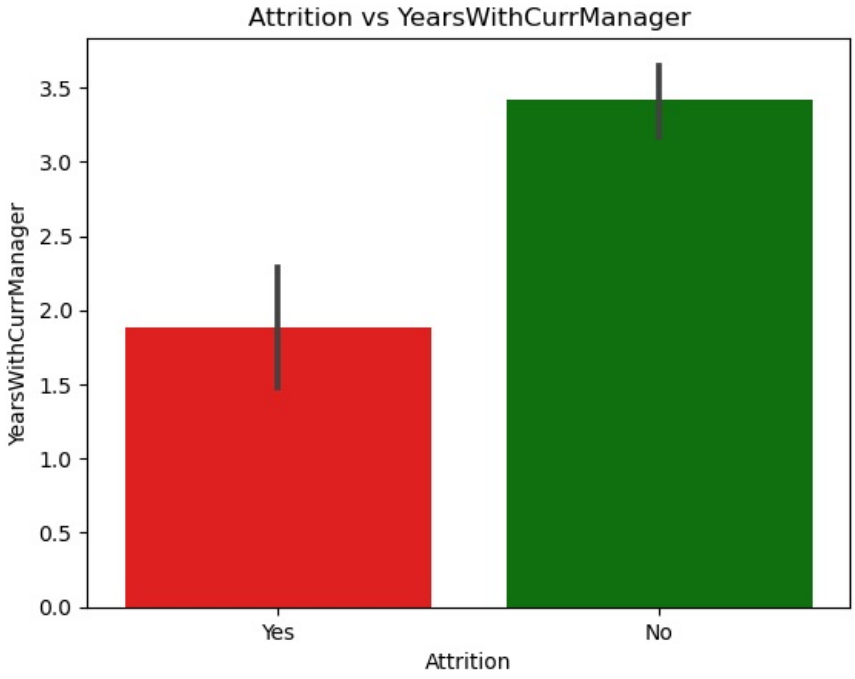
- From the above plot it can be concluded.
- The chances of an employee leaving his job increase as their monthly income decreases.

In [40]:
```
#barplot between attrition and work life balance
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["WorkLifeBalance"],palette=["red","green"])
plt.xlabel("Attrition")
plt.ylabel("WorkLifeBalance")
plt.title("Attrition vs WorkLifeBalance")
plt.show()
```



From the above plot it can be concluded that the work life balance doesnot have a very significant effect on the attrition rate

In [42]:
```
#barplot between attrition rate and Yearswith currentmanagaer
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["YearsWithCurrManager"],palette=["red","green"])
plt.xlabel("Attrition")
plt.ylabel("YearsWithCurrManager")
plt.title("Attrition vs YearsWithCurrManager")
plt.show()
```



From the above plot it can be concluded that:-

- The years worked under the current manager has a significant affect on the employees leaving their jobs
- More time an employee has worked under the current manager the less likely they are to leave their job

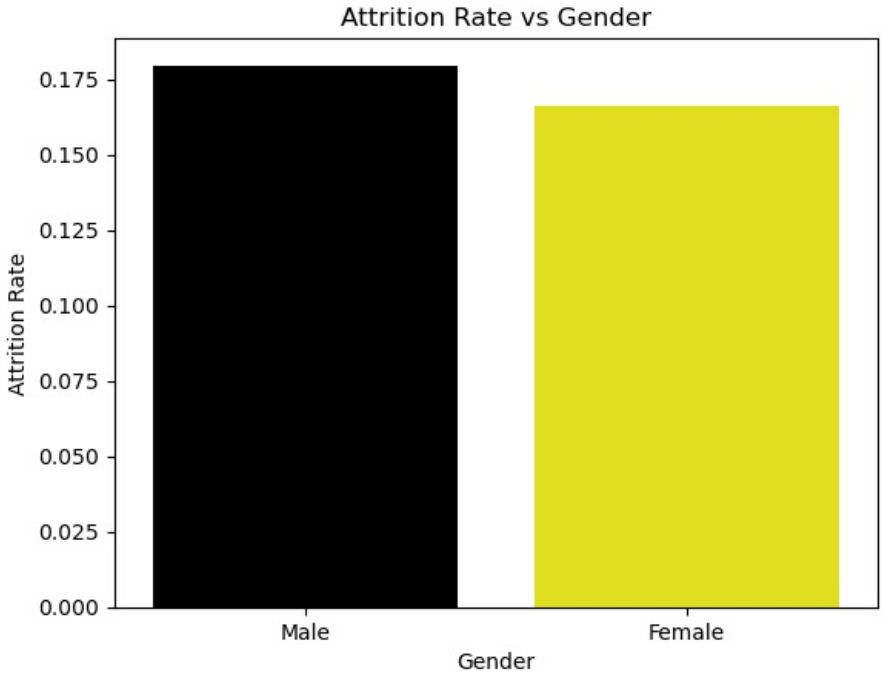
In [54]:
```
ar=[]
for i in range(0,9):
    ar.append((greendata[greendata["Attrition"]=="Yes"][greendata["NumCompaniesWorked"]==i]).size/(greendata[gr
```

```
C:\Users\PUJA\AppData\Local\Temp\ipykernel_18440\649734282.py:3: UserWarning:

Boolean Series key will be reindexed to match DataFrame index.
```

In [50]:
```
l1=(greendata[(greendata["Gender"]=="Male") & (greendata["Attrition"]=="Yes")].size)/(greendata[greendata["Gend
l2=(greendata[(greendata["Gender"]=="Female") & (greendata["Attrition"]=="Yes")].size)/(greendata[greendata["Ge
```

In [51]:
```
#barplot displaying the relation between gender of employees and attrition rate
plt.figure()
sns.barplot(x=["Male","Female"],y=[l1,l2],palette=["black","yellow"])
plt.xlabel("Gender")
plt.ylabel("Attrition Rate")
plt.title("Attrition Rate vs Gender")
plt.show()
```

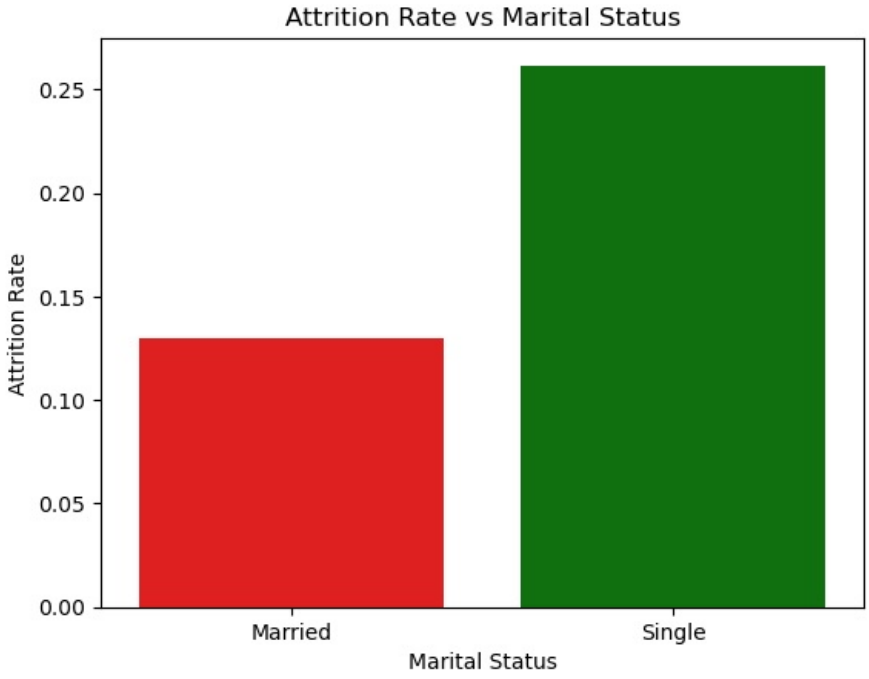


From the above plot it can be concluded:-

- Male employees are more likely to quit their jobs compared to female employees though the disparity is not very significant

In [56]:
```
n1=((greendata[(greendata["MaritalStatus"]=="Married") & (greendata["Attrition"]=="Yes")]).size)/((greendata[gr
n2=((greendata[(greendata["MaritalStatus"]=="Single") & (greendata["Attrition"]=="Yes")]).size)/((greendata[gre
```

In [57]:
```
#Barplot showing relation between marital status of the employees and attrition rate
plt.figure()
sns.barplot(x=["Married","Single"],y=[n1,n2],palette=["red","green"])
plt.xlabel("Marital Status")
plt.ylabel("Attrition Rate")
plt.title("Attrition Rate vs Marital Status")
plt.show()
```



From the above plot it can be concluded that:-

- Unmarried or single employees have a significantly higher chance of quitting their jobs than their married collegues

From the following plot it can be seen that the hourly rate doesnt have a significant effect on the attrition at all

In [105…
```
plt.figure()
sns.barplot(x=greendata["Attrition"],y=greendata["HourlyRate"],palette=["red","green"])
plt.xlabel("Attrition")
plt.ylabel("HourlyRate")
plt.title("Attrition vs HourlyRate")
plt.show()
```

SUMMARY OF CONCLUSIONS :

The total attrition rate I.E rate of employees leaving their jobs is 16.1 % Jobsatisfaction is an important factor in employees leaving jobs. Lesser the job satisfaction higher are the chances of an employee leaving their job. The average age of an employee leaving their job is lesser than those not leaving their job. Average age of an employee leaving their job is 31.5 years. Average age of an employee not leaving their job is 35.6 years. The longer an employee has worked for the company the lesser are their chances of leaving the job The chances of an employee leaving his job increase as their monthly income decreases. Employees that have worked for more than 4 companies prior to the current job are more likely to leave their jobs. In general it can be concluded that higher the prior experience more likely an employee is to leave hisn job. Male employees are more likely to quit their jobs compared to female employees though the disparity is not very significant. Unmarried or single employees have a significantly higher chance of quitting their jobs than their married collegues. Hourly,monthly and daily rate dont have a significant affect on the attrition at all.

In [ ]: