In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import sqlalchemy as sal
import plotly.express as px
```

In [2]:
```
data = pd.read_csv("Attrition data.csv")
```

In [3]:
```
data = data.dropna()
```

In [4]:
```
data
```

Out[4]:

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...	TotalWorkingYears	TrainingTimesLastYear	YearsAtC
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...	1.0	6	
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...	6.0	3	
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...	5.0	2	
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...	13.0	5	
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...	9.0	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4404	4405	29	No	Travel_Rarely	Sales	4	3	Other	1	Female	...	6.0	2	
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...	10.0	5	
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...	10.0	2	
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...	5.0	4	
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...	10.0	2	

4300 rows × 29 columns

In [5]:
```
data.isna().sum()
```

Out[5]:
```
EmployeeID                0
Age                       0
Attrition                 0
BusinessTravel            0
Department                0
DistanceFromHome          0
Education                 0
EducationField            0
EmployeeCount             0
Gender                    0
JobLevel                  0
JobRole                   0
MaritalStatus             0
MonthlyIncome             0
NumCompaniesWorked        0
Over18                    0
PercentSalaryHike         0
StandardHours             0
StockOptionLevel          0
TotalWorkingYears         0
TrainingTimesLastYear     0
YearsAtCompany            0
YearsSinceLastPromotion   0
YearsWithCurrManager      0
EnvironmentSatisfaction   0
JobSatisfaction           0
WorkLifeBalance           0
JobInvolvement            0
PerformanceRating         0
dtype: int64
```

In [6]:
```
data
```

Out[6]:

	EmployeeID	Age	Attrition	BusinessTravel	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Gender	...	TotalWorkingYears	TrainingTimesLastYear	YearsAtC
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...	1.0	6	
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...	6.0	3	
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...	5.0	2	
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...	13.0	5	
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...	9.0	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4404	4405	29	No	Travel_Rarely	Sales	4	3	Other	1	Female	...	6.0	2	
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...	10.0	5	
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...	10.0	2	
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...	5.0	4	
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...	10.0	2	

4300 rows × 29 columns

In [7]:
```
data.columns = data.columns.str.lower()
```

In [8]:
```
data.columns
```

Out[8]:
```
Index(['employeeid', 'age', 'attrition', 'businesstravel', 'department',
       'distancefromhome', 'education', 'educationfield', 'employeecount',
       'gender', 'joblevel', 'jobrole', 'maritalstatus', 'monthlyincome',
       'numcompaniesworked', 'over18', 'percentsalaryhike', 'standardhours',
       'stockoptionlevel', 'totalworkingyears', 'trainingtimeslastyear',
       'yearsatcompany', 'yearssincelastpromotion', 'yearswithcurrmanager',
       'environmentsatisfaction', 'jobsatisfaction', 'worklifebalance',
       'jobinvolvement', 'performancerating'],
      dtype='object')
```

In [9]:
```
data_columns = ['employee_id', 'age', 'attrition', 'business_travel', 'department',
       'distance_from_home', 'education', 'education_field', 'employee_count',
       'gender', 'job_level', 'job_role', 'marital_status', 'monthly_income',
       'num_companies_worked', 'over_18', 'percent_salary_hike', 'standard_hours',
       'stock_option_level', 'total_working_years', 'training_times_lastyear',
       'years_at_company', 'years_since_last_promotion', 'years_with_curr_manager',
       'environment_satisfaction', 'job_satisfaction', 'work_life_balance',
       'job_involvement', 'performance_rating']
```

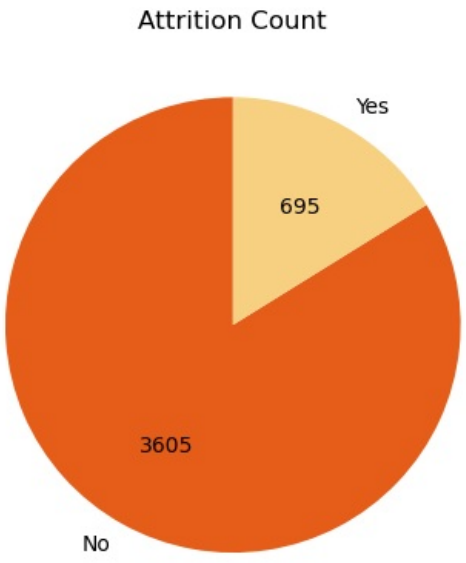
In [10]:
```
data.columns = data_columns
```

In [11]:
```
data.shape
```

Out[11]: (4300, 29)

In [29]:
```
labels = data.attrition.value_counts().index
values = data.attrition.value_counts()
colors=["#E65C19","#F8D082"]

plt.pie(values,labels=labels,colors=colors,autopct=lambda p: "{:.0f}".format(p*sum(values)/100),startangle=90)
plt.title("Attrition Count")
plt.show()
```

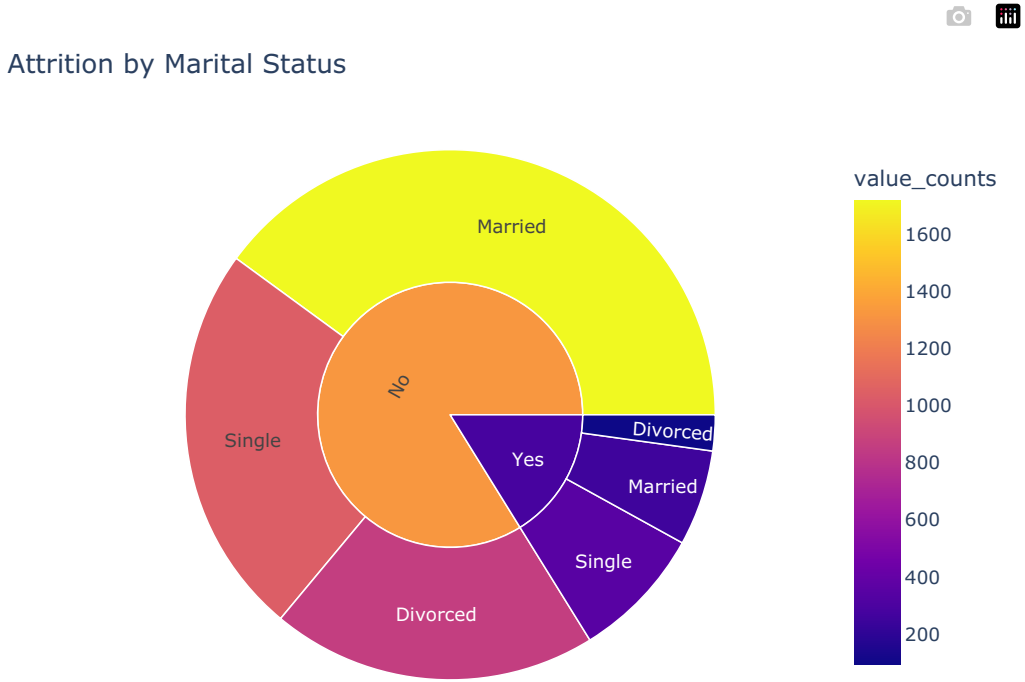


In [13]:
```
data.marital_status.value_counts()
```

Out[13]:
```
marital_status
Married     1969
Single      1382
Divorced     949
Name: count, dtype: int64
```

In [14]:
```
maritial_data = data.marital_status.value_counts()
maritial_label = data.marital_status.value_counts().index
value_c = data.groupby(["attrition","marital_status"]).size().reset_index(name="value_counts")

fig =px.sunburst(
    value_c,
    path=['attrition','marital_status'],
    values="value_counts",
    color="value_counts",
    title="Attrition by Marital Status"
)
fig.show()
```

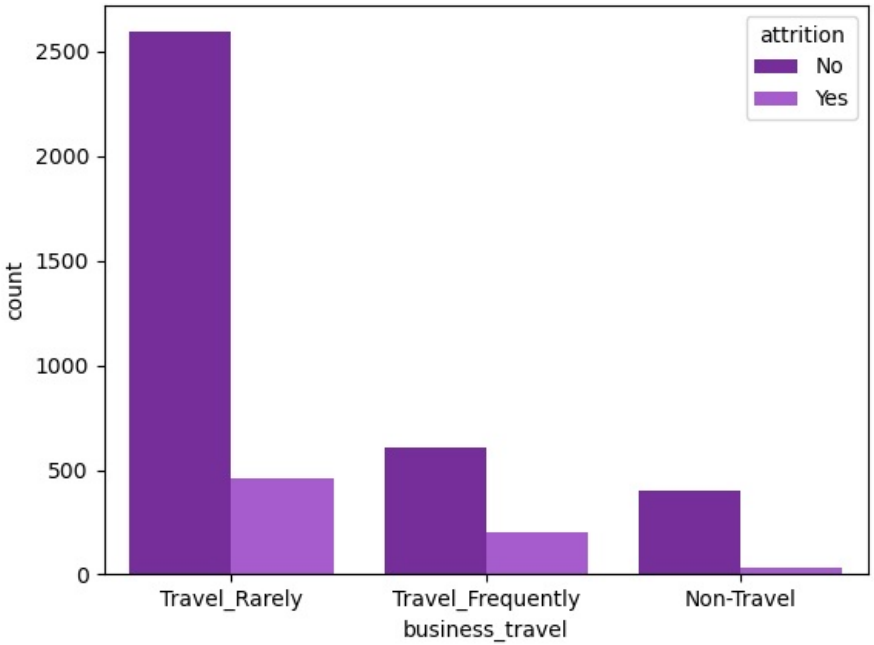


In [15]:
```
data.business_travel.value_counts()
```

Out[15]:
```
business_travel
Travel_Rarely        3051
Travel_Frequently     809
Non-Travel            440
Name: count, dtype: int64
```

In [25]:
```
color = ["#7A1CAC","#AD49E1"]
sns.countplot(data=data,x="business_travel",hue="attrition", palette=color)
```

Out[25]:
```
<Axes: xlabel='business_travel', ylabel='count'>
```



In [17]:
```
data.education_field.value_counts()
```

Out[17]:
```
education_field
Life Sciences       1766
Medical             1364
Marketing            469
Technical Degree     384
Other                237
Human Resources       80
Name: count, dtype: int64
```

In [18]:
```
value_c = data.groupby(["attrition","education_field"]).size().reset_index(name="value_counts")
fig=px.sunburst(
    value_c,
    path=['attrition','education_field'],
    values="value_counts",
    color="value_counts",
```

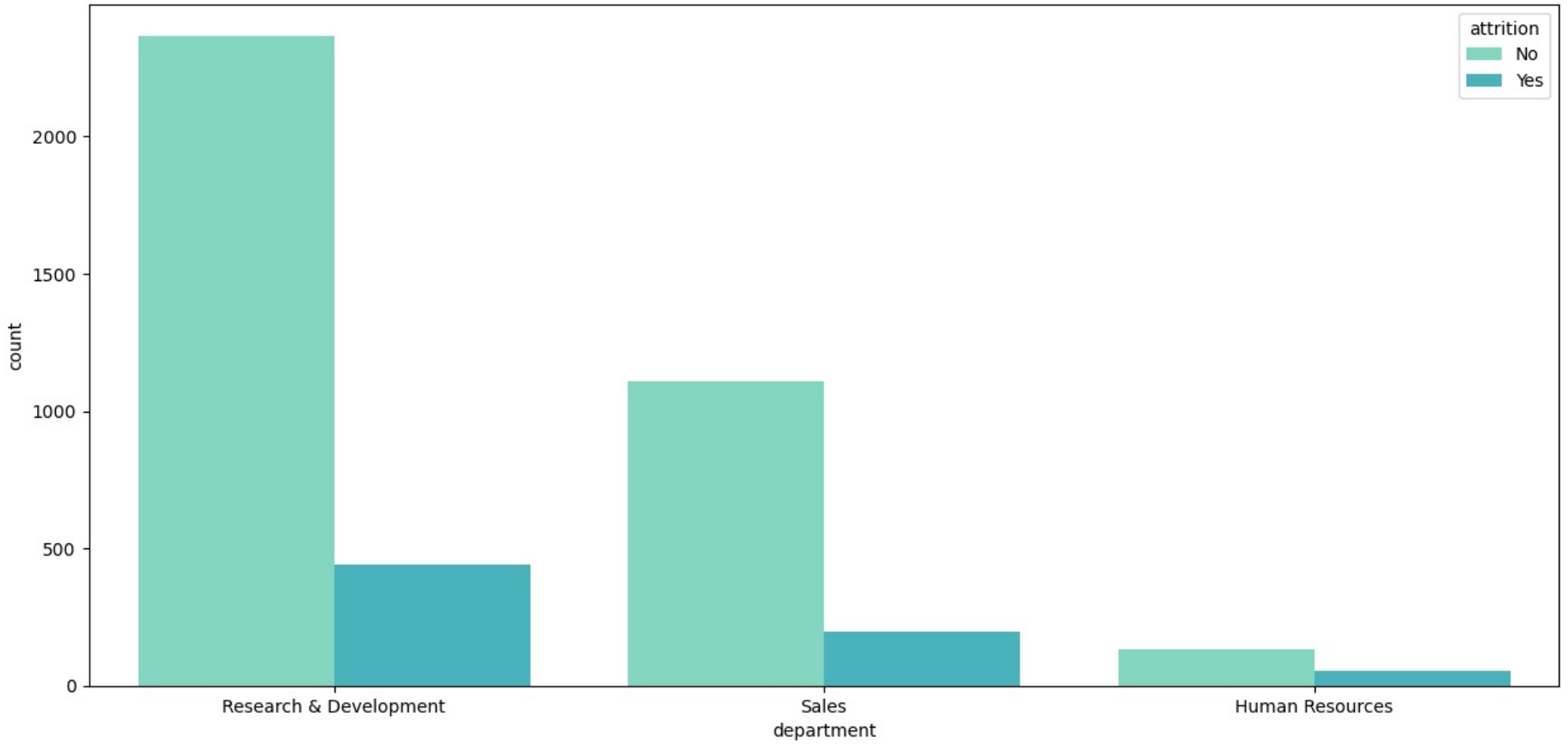
```
    title="Attrition by Education Status"
)
fig.show()
```

In [19]:
```
data.department.value_counts()
```

Out[19]:
```
department
Research & Development    2807
Sales                     1307
Human Resources            186
Name: count, dtype: int64
```

In [28]:
```
plt.figure(figsize=(15,7))
b_order=data.department.value_counts().index
color=['#77E4C8',"#36C2CE"]
sns.countplot(data=data,x="department",order=b_order,hue="attrition",palette=color)
```

Out[28]:
```
<Axes: xlabel='department', ylabel='count'>
```


attrition
No
Yes
2000
1500
1000
500
0
count
Research & Development
Sales
Human Resources
department

In [21]:
```
data
```

Out[21]:

	employee_id	age	attrition	business_travel	department	distance_from_home	education	education_field	employee_count	gender	...	total_working_years	training_times_lastyear	years_
0	1	51	No	Travel_Rarely	Sales	6	2	Life Sciences	1	Female	...	1.0	6	
1	2	31	Yes	Travel_Frequently	Research & Development	10	1	Life Sciences	1	Female	...	6.0	3	
2	3	32	No	Travel_Frequently	Research & Development	17	4	Other	1	Male	...	5.0	2	
3	4	38	No	Non-Travel	Research & Development	2	5	Life Sciences	1	Male	...	13.0	5	
4	5	32	No	Travel_Rarely	Research & Development	10	1	Medical	1	Male	...	9.0	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
4404	4405	29	No	Travel_Rarely	Sales	4	3	Other	1	Female	...	6.0	2	
4405	4406	42	No	Travel_Rarely	Research & Development	5	4	Medical	1	Female	...	10.0	5	
4406	4407	29	No	Travel_Rarely	Research & Development	2	4	Medical	1	Male	...	10.0	2	
4407	4408	25	No	Travel_Rarely	Research & Development	25	2	Life Sciences	1	Male	...	5.0	4	
4408	4409	42	No	Travel_Rarely	Sales	18	2	Medical	1	Male	...	10.0	2	

4300 rows × 29 columns

In [22]:
```
engine = sal.create_engine("mysql+mysqlconnector://root:12345678@localhost/Unified_Mentor_Attition")
connection = engine.connect()
```

In [24]:
```
data.to_sql("attriton_data",con=connection,index = False,if_exists= "append")
```

Out[24]: 4300

In [ ]: