```
#Loading Librabries
import pandas as pd
from scipy.stats import f_oneway

# Load the dataset
file_path = "Grade_Sheet[1].xlsx"  # Change this path if needed
xls = pd.ExcelFile(file_path)
grade_df = xls.parse('Sheet1')
grade_df.head()
```

	Mathematics	English	Education	Biology
0	2.59	3.64	4.00	2.78
1	3.13	3.19	3.59	3.51
2	2.97	3.15	2.80	2.65
3	2.50	3.78	2.39	3.16
4	2.53	3.03	3.47	2.94

Hypothesis for GPA Comparison

- **Null Hypothesis (H_0):** The average GPAs of all four majors (Mathematics, English, Education, Biology) are equal.
- **Alternative Hypothesis (H_1):** At least one major has a significantly different average GPA.

```
# Extract GPA data for each major
math_gpa = grade_df['Mathematics']
eng_gpa = grade_df['English']
edu_gpa = grade_df['Education']
bio_gpa = grade_df['Biology']

# Perform one-way ANOVA
anova_result = f_oneway(math_gpa, eng_gpa, edu_gpa, bio_gpa)

# Display the result
print("ANOVA Test Result:")
print(f"F-statistic: {anova_result.statistic:.4f}")
print(f"p-value: {anova_result.pvalue:.4f}")
```

```
ANOVA Test Result:
F-statistic: 3.1992
p-value: 0.0335
```

```
# Interpretation
alpha = 0.05
if anova_result.pvalue < alpha:
    print("\nConclusion: Reject the null hypothesis.")
    print("There is a significant difference in average GPAs among the
four majors.")
else:
    print("\nConclusion: Fail to reject the null hypothesis.")
    print("There is no significant difference in average GPAs among
the four majors.")
```

```
Conclusion: Reject the null hypothesis.
There is a significant difference in average GPAs among the four majors.
```

```
df=pd.read_csv('IBM_HR.csv')
```

```
print("Shape of dataset:", df.shape)
```

```
Shape of dataset: (1470, 35)
```

```
# List all attribute names
print("Attributes in the dataset:\n", df.columns.tolist())
```

```
# Check for missing values
print("\nMissing values in each column:\n",df.isnull().sum())
```

```
Attributes in the dataset:
 ['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
'YearsWithCurrManager']

Missing values in each column:
 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
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

```
# Value counts of 'Attrition'
print("Attrition counts:\n",df['Attrition'].value_counts())

# Value counts of 'JobSatisfaction'
print("\nJob Satisfaction counts:\
n",df['JobSatisfaction'].value_counts())
```

```
Attrition counts:
 Attrition
No     1233
Yes     237
Name: count, dtype: int64

Job Satisfaction counts:
 JobSatisfaction
4    459
3    442
1    289
2    280
Name: count, dtype: int64
```

```
# Cross-tabulation
contingency_table_1 = pd.crosstab(df['Attrition'],
df['JobSatisfaction'])
print("Contingency Table (Attrition vs JobSatisfaction):\n",
contingency_table_1)
```

```
Contingency Table (Attrition vs JobSatisfaction):
 JobSatisfaction    1    2    3    4
Attrition
No                223  234  369  407
Yes                66   46   73   52
```

```
from scipy.stats import chi2_contingency
```

```python
# Chi-square test
chi2, p, dof, expected = chi2_contingency(contingency_table_1)

print(f"Chi-square Statistic: {chi2:.4f}")
print(f"p-value: {p:.4f}")
print(f"Degrees of Freedom: {dof}")
print("Expected Frequencies Table:\n", expected)

# Interpret the result
alpha = 0.05
if p < alpha:
    print("\nConclusion: Reject the null hypothesis.")
    print("There is a significant relationship between Attrition and
JobSatisfaction.")
else:
    print("\nConclusion: Fail to reject the null hypothesis.")
    print("There is no significant relationship between Attrition and
JobSatisfaction.")
```

```
Chi-square Statistic: 17.5051
p-value: 0.0006
Degrees of Freedom: 3
Expected Frequencies Table:
 [[242.40612245 234.85714286 370.73877551 384.99795918]
 [ 46.59387755  45.14285714  71.26122449  74.00204082]]

Conclusion: Reject the null hypothesis.
There is a significant relationship between Attrition and
JobSatisfaction.
```

```python
# Cross-tabulation for Attrition and Education
contingency_table_2 = pd.crosstab(df['Attrition'],df['Education'])
print("Contingency Table (Attrition vs Education):\n",
contingency_table_2)

# Chi-square test
chi2_edu, p_edu, dof_edu, expected_edu =
chi2_contingency(contingency_table_2)

print(f"\nChi-square Statistic: {chi2_edu:.4f}")
print(f"p-value: {p_edu:.4f}")
print(f"Degrees of Freedom: {dof_edu}")
print("Expected Frequencies Table:\n", expected_edu)

# Interpret the result
if p_edu < alpha:
    print("\nConclusion: Reject the null hypothesis.")
    print("There is a significant relationship between Attrition and
Education.")
else:
```

```
    print("\nConclusion: Fail to reject the null hypothesis.")
    print("There is no significant relationship between Attrition and
Education.")

Contingency Table (Attrition vs Education):
 Education    1    2    3    4   5
Attrition
No         139  238  473  340  43
Yes         31   44   99   58   5

Chi-square Statistic: 3.0740
p-value: 0.5455
Degrees of Freedom: 4
Expected Frequencies Table:
 [[142.59183673 236.53469388 479.77959184 333.83265306  40.26122449]
 [ 27.40816327  45.46530612  92.22040816  64.16734694   7.73877551]]

Conclusion: Fail to reject the null hypothesis.
There is no significant relationship between Attrition and Education.
```