```
Start coding or generate with AI.
```

Task

Perform statistical analysis for comparing two business strategies (e.g., t-test) using the dataset "/content/WineQT.csv".

Identify the groups

Subtask:

Determine how the data is structured to represent the two business strategies you want to compare.

Reasoning: Examine the columns and unique values of the DataFrame to identify a suitable column for grouping, then select a column and check its unique values to confirm it can split the data into two groups.

```
print(df.columns)
print(df['quality'].unique())
```

```
Index(['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
       'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
       'pH', 'sulphates', 'alcohol', 'quality', 'Id'],
      dtype='object')
[5 6 7 4 8 3]
```

Check assumptions

Subtask:

Verify if the data meets the assumptions for a t-test (e.g., normality, equal variances).

Reasoning: Create two groups based on the 'quality' column and perform normality and equal variance tests on the 'alcohol' feature for these groups to verify the assumptions for a t-test.

```
from scipy.stats import shapiro, levene

# Create two groups based on 'quality'
high_quality = df[df['quality'] >= 6]['alcohol']
low_quality = df[df['quality'] < 6]['alcohol']

# Check for normality using Shapiro-Wilk test
shapiro_high, p_high_shapiro = shapiro(high_quality)
shapiro_low, p_low_shapiro = shapiro(low_quality)
```

```
# Perform Levene's test for equal variances
levene_test, p_levene = levene(high_quality, low_quality)

# Print the results
print(f"Shapiro-Wilk test for high quality group (alcohol): Statistic={shapiro_high:.4f}, p
print(f"Shapiro-Wilk test for low quality group (alcohol): Statistic={shapiro_low:.4f}, p-v
print(f"Levene's test for equal variances (alcohol): Statistic={levene_test:.4f}, p-value={
```

```
Shapiro-Wilk test for high quality group (alcohol): Statistic=0.9775, p-value=0.0000
Shapiro-Wilk test for low quality group (alcohol): Statistic=0.8514, p-value=0.0000
Levene's test for equal variances (alcohol): Statistic=98.0234, p-value=0.0000
```

Perform the t-test

Subtask:

Apply the appropriate t-test (independent samples t-test or paired samples t-test) based on the data and the nature of the comparison.

Reasoning: Import the ttest_ind function and perform an independent samples t-test with equal_var=False using the high_quality and low_quality series, storing the results in separate variables and printing them.

```
from scipy.stats import ttest_ind

# Perform independent samples t-test with equal_var=False (Welch's t-test)
t_statistic, p_value = ttest_ind(high_quality, low_quality, equal_var=False)

# Print the results
print(f"Welch's t-test (alcohol): t-statistic={t_statistic:.4f}, p-value={p_value:.4f}")
```

```
Welch's t-test (alcohol): t-statistic=17.0671, p-value=0.0000
```

Interpret the results

Subtask:

Explain the t-test output, including the p-value and confidence interval, in the context of the business problem.

Reasoning: Explain the t-test output including the p-value and t-statistic, interpreting them in the context of the business problem as outlined in the instructions.

```
# Interpretation of the t-test results

# 1. Examine the printed output (already done in the previous step).
```

```
#    t-statistic = 17.0671
#    p-value = 0.0000 (effectively < 0.0001)

# 2. Interpret the p-value
alpha = 0.05
print(f"Significance level (alpha): {alpha}")
print(f"P-value: {p_value:.4f}")

if p_value < alpha:
    print("Since the p-value is less than the significance level (alpha), we reject the nul
    print("This means there is a statistically significant difference in the mean alcohol c
else:
    print("Since the p-value is greater than or equal to the significance level (alpha), we
    print("This means there is no statistically significant difference in the mean alcohol

# 3. Discuss the t-statistic
print(f"\nT-statistic: {t_statistic:.4f}")
print("The t-statistic measures the difference between the two group means in units of the
print("A larger absolute t-statistic indicates a greater difference between the group means
print(f"In this case, the t-statistic of {t_statistic:.4f} is large, indicating that the di

# 4. Conclude based on the p-value
print("\nConclusion based on p-value:")
if p_value < alpha:
    print("There is strong evidence to conclude that the mean alcohol content of high-quali
else:
    print("There is not enough evidence to conclude that the mean alcohol content of high-q


# 5. Relate to the business problem/strategies
print("\nRelating to business strategies:")
print("If 'high quality' and 'low quality' represent outcomes of two different business str
print("For example, if a business strategy aims to produce higher quality wine (as defined
print("This finding could inform decisions related to optimizing production processes, unde
```

```
Significance level (alpha): 0.05
P-value: 0.0000
Since the p-value is less than the significance level (alpha), we reject the null hypothesis
This means there is a statistically significant difference in the mean alcohol content betwe

T-statistic: 17.0671
The t-statistic measures the difference between the two group means in units of the standard
A larger absolute t-statistic indicates a greater difference between the group means relativ
In this case, the t-statistic of 17.0671 is large, indicating that the difference in mean al

Conclusion based on p-value:
There is strong evidence to conclude that the mean alcohol content of high-quality wines is

Relating to business strategies:
If 'high quality' and 'low quality' represent outcomes of two different business strategies
For example, if a business strategy aims to produce higher quality wine (as defined in this
This finding could inform decisions related to optimizing production processes, understandin
```

Present findings

Subtask:

Summarize the statistical findings and their implications for the business strategies.

Reasoning: Summarize the statistical findings and their implications for the business strategies based on the previous steps.

```
print("Summary of Statistical Findings and Business Implications\n")

# 1. Purpose of the analysis
print("1. Purpose of the Analysis:")
print("   The analysis compared a specific characteristic, alcohol content, between two gro

# 2. Key statistical findings from the t-test
print("2. Key Statistical Findings:")
print(f"   - T-statistic: {t_statistic:.4f}")
print(f"   - P-value: {p_value:.4f} (which is effectively less than 0.0001)")
print(f"   - Significance Level (alpha): {alpha}")
print("   - Conclusion: Since the p-value (0.0000) is much less than the significance level
print("   - Interpretation: There is a statistically significant difference in the mean alc

# 3. Meaning of the statistically significant difference in the context of business strateg
print("3. Meaning in the Context of Business Strategies:")
print("   The statistically significant difference indicates that the business strategy ass
print("   Specifically, examining the mean alcohol content (though not explicitly printed i

# 4. Potential business implications
print("4. Potential Business Implications:")
print("   - Production Decisions: Businesses might explore how different production techniq
print("   - Marketing and Product Positioning: Alcohol content could be a factor used in ma
print("   - Understanding Customer Preferences: The finding suggests alcohol content might
print("   - Process Optimization: Businesses aiming for higher quality might need to optimi

# 5. Limitations of this analysis
print("5. Limitations of the Analysis:")
print("   - Correlation vs. Causation: This analysis shows an association between quality a
print("   - Single Characteristic: Only alcohol content was examined in this comparison. Wi
print("   - Definition of Quality: The definition of 'high' and 'low' quality is based on a
print("   - Assumptions: While Welch's t-test was used to account for unequal variances, th
```

```
Summary of Statistical Findings and Business Implications

1. Purpose of the Analysis:
   The analysis compared a specific characteristic, alcohol content, between two groups of w

2. Key Statistical Findings:
   - T-statistic: 17.0671
   - P-value: 0.0000 (which is effectively less than 0.0001)
   - Significance Level (alpha): 0.05
   - Conclusion: Since the p-value (0.0000) is much less than the significance level (0.05),
   - Interpretation: There is a statistically significant difference in the mean alcohol con

3. Meaning in the Context of Business Strategies:
   The statistically significant difference indicates that the business strategy associated
   Specifically, examining the mean alcohol content (though not explicitly printed in this s
```

```
4. Potential Business Implications:
   - Production Decisions: Businesses might explore how different production techniques infl
   - Marketing and Product Positioning: Alcohol content could be a factor used in marketing
   - Understanding Customer Preferences: The finding suggests alcohol content might be a cha
   - Process Optimization: Businesses aiming for higher quality might need to optimize their

5. Limitations of the Analysis:
   - Correlation vs. Causation: This analysis shows an association between quality and alcoh
   - Single Characteristic: Only alcohol content was examined in this comparison. Wine quali
   - Definition of Quality: The definition of 'high' and 'low' quality is based on a specifi
   - Assumptions: While Welch's t-test was used to account for unequal variances, the assump
```

Summary:

Data Analysis Key Findings

- The Shapiro-Wilk test indicated that the alcohol content in both high-quality (p-value = 0.0000) and low-quality (p-value = 0.0000) wine groups is not normally distributed, violating a key assumption for a standard independent samples t-test.
- Levene's test showed that the variances of alcohol content between the high-quality and low-quality groups are significantly different (p-value = 0.0000), violating the equal variance assumption of a standard independent samples t-test.
- Welch's t-test, which does not assume equal variances, was performed.
- The Welch's t-test yielded a t-statistic of 17.0671 and a p-value of 0.0000.
- Based on a significance level of $\alpha = 0.05$, the p-value (0.0000) is less than α, leading to the rejection of the null hypothesis.
- There is a statistically significant difference in the mean alcohol content between high-quality and low-quality wines.

Insights or Next Steps

- The significant difference in alcohol content suggests that business strategies impacting wine quality may also have a noticeable influence on the final alcohol level. Further investigation into *how* different strategies affect alcohol content could be beneficial.
- Since the normality assumption was violated, consider non-parametric tests (like the Mann-Whitney U test) for a more robust comparison of the two groups, especially if the sample size were small, although t-tests are often robust to this violation with large samples.