```python
# ------------------------------------------------------------
# SECTION B - Q1: CSV Data Exploration & Visualization
# Using Wine-dataset.csv provided
# ------------------------------------------------------------

# 1. Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt

# 2. Load the CSV file into a Pandas DataFrame
# Make sure 'Wine-dataset.csv' is in the same folder as this script
df = pd.read_csv("Wine.csv")


# 3. Display first 10 rows to understand the structure of the dataset
print("----- FIRST 10 ROWS -----")
print(df.head(10))
```

```
----- FIRST 10 ROWS -----
   class  Alcohol  Malic acid   Ash  Alcalinity of ash  Magnesium  \
0      1    14.23        1.71  2.43               15.6        127
1      1    13.20        1.78  2.14               11.2        100
2      1    13.16        2.36  2.67               18.6        101
3      1    14.37        1.95  2.50               16.8        113
4      1    13.24        2.59  2.87               21.0        118
5      1    14.20        1.76  2.45               15.2        112
6      1    14.39        1.87  2.45               14.6         96
7      1    14.06        2.15  2.61               17.6        121
8      1    14.83        1.64  2.17               14.0         97
9      1    13.86        1.35  2.27               16.0         98

   Total phenols  Flavanoids  Nonflavanoid phenols  Proanthocyanins  \
0           2.80        3.06                  0.28             2.29
1           2.65        2.76                  0.26             1.28
2           2.80        3.24                  0.30             2.81
3           3.85        3.49                  0.24             2.18
4           2.80        2.69                  0.39             1.82
5           3.27        3.39                  0.34             1.97
6           2.50        2.52                  0.30             1.98
7           2.60        2.51                  0.31             1.25
8           2.80        2.98                  0.29             1.98
9           2.98        3.15                  0.22             1.85

   Color intensity   Hue  OD280/OD315 of diluted wines  Proline
0             5.64  1.04                          3.92     1065
1             4.38  1.05                          3.40     1050
2             5.68  1.03                          3.17     1185
3             7.80  0.86                          3.45     1480
4             4.32  1.04                          2.93      735
5             6.75  1.05                          2.85     1450
```

```
6              5.25  1.02                            3.58      1290
7              5.05  1.06                            3.58      1295
8              5.20  1.08                            2.85      1045
9              7.22  1.01                            3.55      1045
```

```
# 4. Basic dataset information
print("\n----- BASIC INFO -----")
print(df.info())  # Column names, non-null counts, data types
print("\nShape of dataset (rows, columns):", df.shape)
```

```
----- BASIC INFO -----
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 14 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   class                         178 non-null    int64
 1   Alcohol                       178 non-null    float64
 2   Malic acid                    178 non-null    float64
 3   Ash                           178 non-null    float64
 4   Alcalinity of ash             178 non-null    float64
 5   Magnesium                     178 non-null    int64
 6   Total phenols                 178 non-null    float64
 7   Flavanoids                    178 non-null    float64
 8   Nonflavanoid phenols          178 non-null    float64
 9   Proanthocyanins               178 non-null    float64
 10  Color intensity               178 non-null    float64
 11  Hue                           178 non-null    float64
 12  OD280/OD315 of diluted wines  178 non-null    float64
 13  Proline                       178 non-null    int64
dtypes: float64(11), int64(3)
memory usage: 19.6 KB
None

Shape of dataset (rows, columns): (178, 14)
```

```
# 5. Descriptive statistics for numerical columns
print("\n----- SUMMARY STATISTICS -----")
print(df.describe())
```

```
----- SUMMARY STATISTICS -----
            class     Alcohol  Malic acid         Ash  Alcalinity of
ash  \
count  178.000000  178.000000  178.000000  178.000000
178.000000
mean     1.938202   13.000618    2.336348    2.366517
19.494944
std      0.775035    0.811827    1.117146    0.274344
```

```
3.339564
min      1.000000   11.030000    0.740000    1.360000
10.600000
25%      1.000000   12.362500    1.602500    2.210000
17.200000
50%      2.000000   13.050000    1.865000    2.360000
19.500000
75%      3.000000   13.677500    3.082500    2.557500
21.500000
max      3.000000   14.830000    5.800000    3.230000
30.000000

        Magnesium  Total phenols  Flavanoids  Nonflavanoid phenols  \
count  178.000000     178.000000  178.000000            178.000000
mean    99.741573       2.295112    2.029270              0.361854
std     14.282484       0.625851    0.998859              0.124453
min     70.000000       0.980000    0.340000              0.130000
25%     88.000000       1.742500    1.205000              0.270000
50%     98.000000       2.355000    2.135000              0.340000
75%    107.000000       2.800000    2.875000              0.437500
max    162.000000       3.880000    5.080000              0.660000

       Proanthocyanins  Color intensity         Hue  \
count       178.000000       178.000000  178.000000
mean          1.590899         5.058090    0.957449
std           0.572359         2.318286    0.228572
min           0.410000         1.280000    0.480000
25%           1.250000         3.220000    0.782500
50%           1.555000         4.690000    0.965000
75%           1.950000         6.200000    1.120000
max           3.580000        13.000000    1.710000

       OD280/OD315 of diluted wines     Proline
count                    178.000000   178.000000
mean                       2.611685   746.893258
std                        0.709990   314.907474
min                        1.270000   278.000000
25%                        1.937500   500.500000
50%                        2.780000   673.500000
75%                        3.170000   985.000000
max                        4.000000  1680.000000

# 6. Check for missing values
print("\n----- MISSING VALUES PER COLUMN -----")
print(df.isnull().sum())


----- MISSING VALUES PER COLUMN -----
class                           0
Alcohol                         0
```

```
Malic acid                      0
Ash                             0
Alcalinity of ash               0
Magnesium                       0
Total phenols                   0
Flavanoids                      0
Nonflavanoid phenols            0
Proanthocyanins                 0
Color intensity                 0
Hue                             0
OD280/OD315 of diluted wines    0
Proline                         0
dtype: int64
```

```
# 7. Handle missing values (if any)
# Here we fill missing numeric values with the column mean
df = df.fillna(df.mean(numeric_only=True))

# 8. Filter: Select wines with Alcohol content > 14
filtered_df = df[df['Alcohol'] > 14]

# 9. Sort: Sort filtered wines by Alcohol in descending order
sorted_df = filtered_df.sort_values(by='Alcohol', ascending=False)
print("\n----- FILTERED & SORTED DATA (Alcohol > 14) -----")
print(sorted_df.head())
```

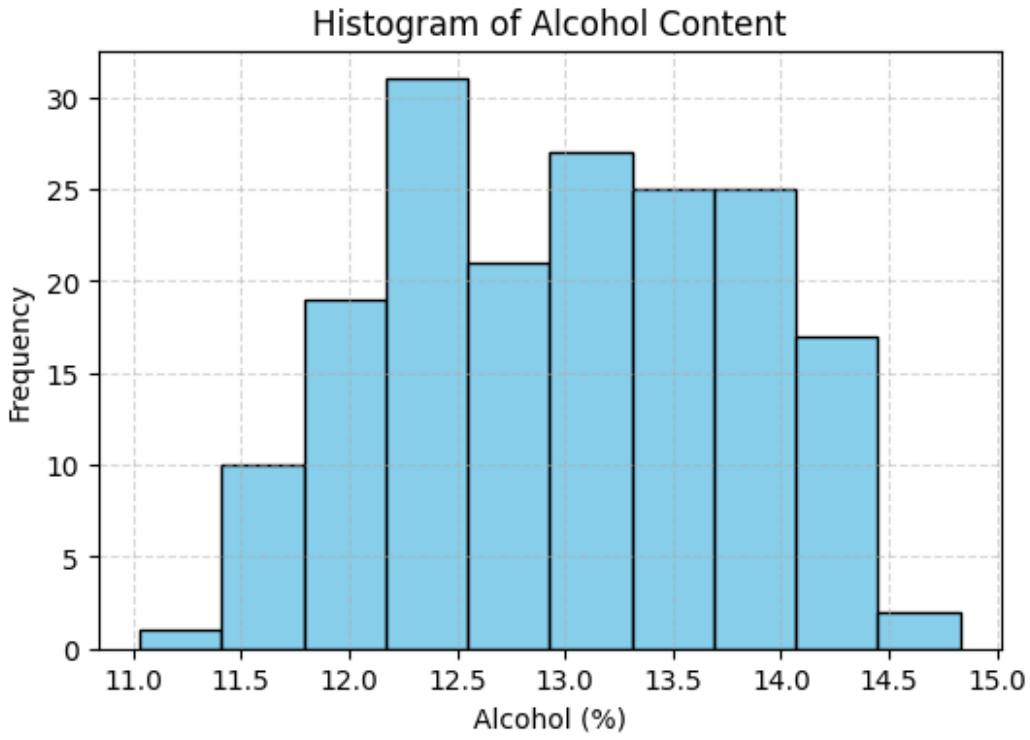
```
----- FILTERED & SORTED DATA (Alcohol > 14) -----
    class  Alcohol  Malic acid   Ash  Alcalinity of ash  Magnesium  \
8       1    14.83        1.64  2.17               14.0         97
13      1    14.75        1.73  2.39               11.4         91
6       1    14.39        1.87  2.45               14.6         96
46      1    14.38        3.59  2.28               16.0        102
14      1    14.38        1.87  2.38               12.0        102

    Total phenols  Flavanoids  Nonflavanoid phenols
Proanthocyanins  \
8            2.80        2.98                  0.29             1.98

13           3.10        3.69                  0.43             2.81

6            2.50        2.52                  0.30             1.98

46           3.25        3.17                  0.27             2.19

14           3.30        3.64                  0.29             2.96


    Color intensity   Hue  OD280/OD315 of diluted wines  Proline
8              5.20  1.08                          2.85      1045
13             5.40  1.25                          2.73      1150
```

```
6               5.25  1.02                              3.58       1290
46              4.90  1.04                              3.44       1065
14              7.50  1.20                              3.00       1547
```

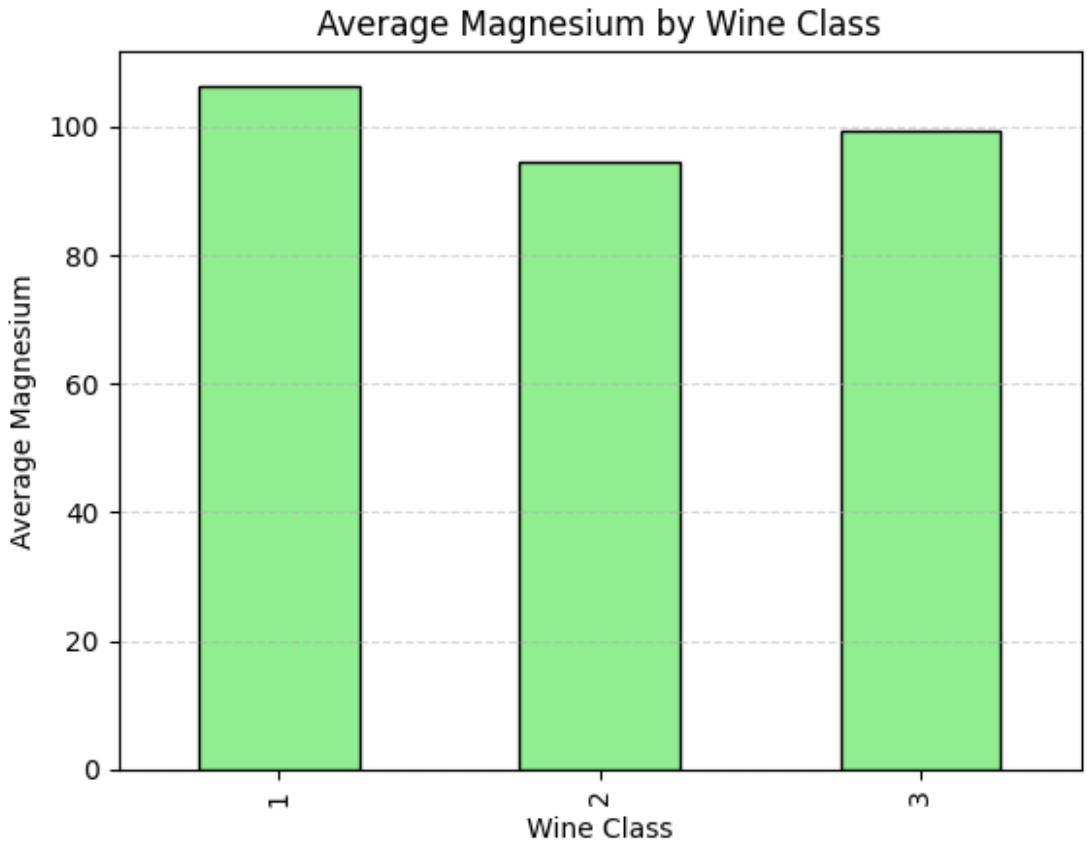
```
# 10. Group By: Calculate average 'Magnesium' level by wine class
group_mean_magnesium = df.groupby('class')['Magnesium'].mean()
print("\n----- AVERAGE MAGNESIUM BY WINE CLASS -----")
print(group_mean_magnesium)
```

```
----- AVERAGE MAGNESIUM BY WINE CLASS -----
class
1    106.338983
2     94.549296
3     99.312500
Name: Magnesium, dtype: float64
```

```
# 11. Visualization 1: Histogram of Alcohol content
plt.figure(figsize=(6,4))
plt.hist(df['Alcohol'], bins=10, color='skyblue', edgecolor='black')
plt.title('Histogram of Alcohol Content')
plt.xlabel('Alcohol (%)')
plt.ylabel('Frequency')
plt.grid(True, linestyle='--', alpha=0.5)
plt.show()
```



```python
# 12. Visualization 2: Bar chart of average Magnesium by wine class
group_mean_magnesium.plot(kind='bar', color='lightgreen',
edgecolor='black')
plt.title('Average Magnesium by Wine Class')
plt.xlabel('Wine Class')
plt.ylabel('Average Magnesium')
plt.grid(True, axis='y', linestyle='--', alpha=0.5)
plt.show()
```




```python
# 13. Save the cleaned dataset to a new CSV file
df.to_csv('Wine-dataset-cleaned.csv', index=False)
print("\nCleaned dataset saved as 'WineCleaned.csv'")
```

```
Cleaned dataset saved as 'WineCleaned.csv'
```

N