The dataset has 11 columns, so we need to include the 'Id' column when assigning the column names.

```
[9]: import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

# Load the Glass Identification Dataset from UCI
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/glass/glass.data"
glass_data = pd.read_csv(url, header=None)
glass_data.columns = ['Id', 'Refractive Index', 'Sodium', 'Magnesium', 'Aluminum', 'Silicon', 'Potassium', 'Calcium', 'Barium', 'Iron', 'Class']

# Drop the 'Id' column
glass_data = glass_data.drop('Id', axis=1)

#check the first few rows
print(glass_data.head())
```

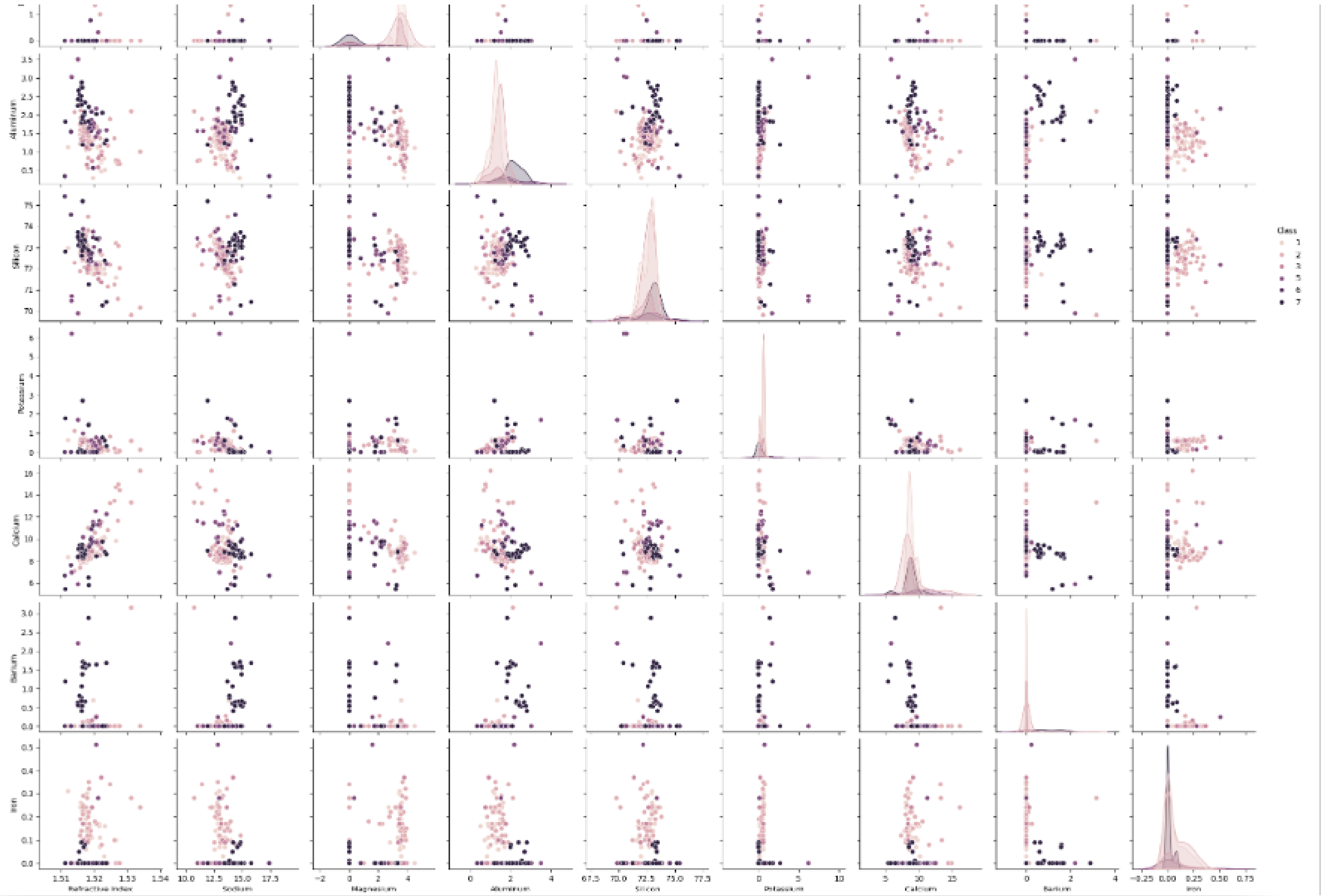
```
   Refractive Index  Sodium  Magnesium  Aluminum  Silicon  Potassium  Calcium  \
0           1.52101   13.64       4.49      1.10    71.78       0.06     8.75
1           1.51761   13.89       3.60      1.36    72.73       0.48     7.83
2           1.51618   13.53       3.55      1.54    72.99       0.39     7.78
3           1.51766   13.21       3.69      1.29    72.61       0.57     8.22
4           1.51742   13.27       3.62      1.24    73.08       0.55     8.07

   Barium  Iron  Class
0     0.0   0.0      1
1     0.0   0.0      1
```

Listing 6 : Complete any 8 calculations and plottings using seab orn package

6.1 Pairplot Scatter

[13]:
```
# part 3
# 6.1
# Pairplot scatter
sns.pairplot(glass_data, hue='Class')
plt.title('Pairplot of Glass Identification Attributes by Class')
plt.show()
```

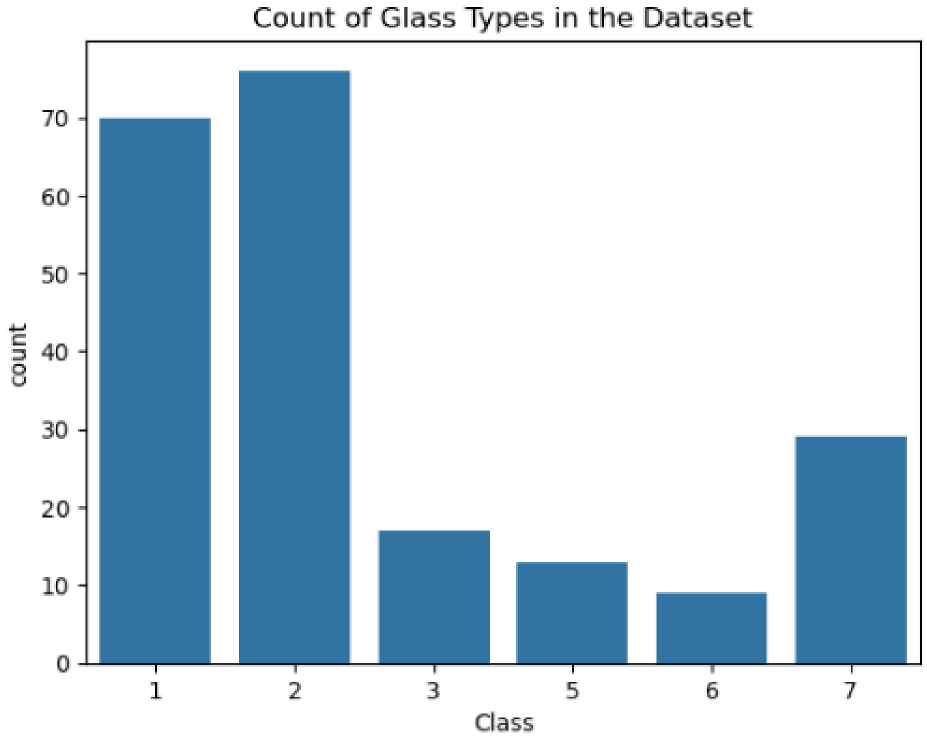
Analysis:
The pairplot scatter shows that while some classes (like **Class 7**) form more distinct clusters, most classes overlap across several attribute combinations. This indicates that some glass types share similar chemical properties, making them harder to distinguish visually.



Aluminum
Silicon
Potassium
Calcium
Barium
Iron
Refractive Index
Sodium
Magnesium
Class
1
2
3
5
6
7

6.2 Count plot of the class variable

[15]:
```
# 6.2 Count plot of the class variable
sns.countplot(x='Class', data=glass_data)
plt.title('Count of Glass Types in the Dataset')
plt.show()
```



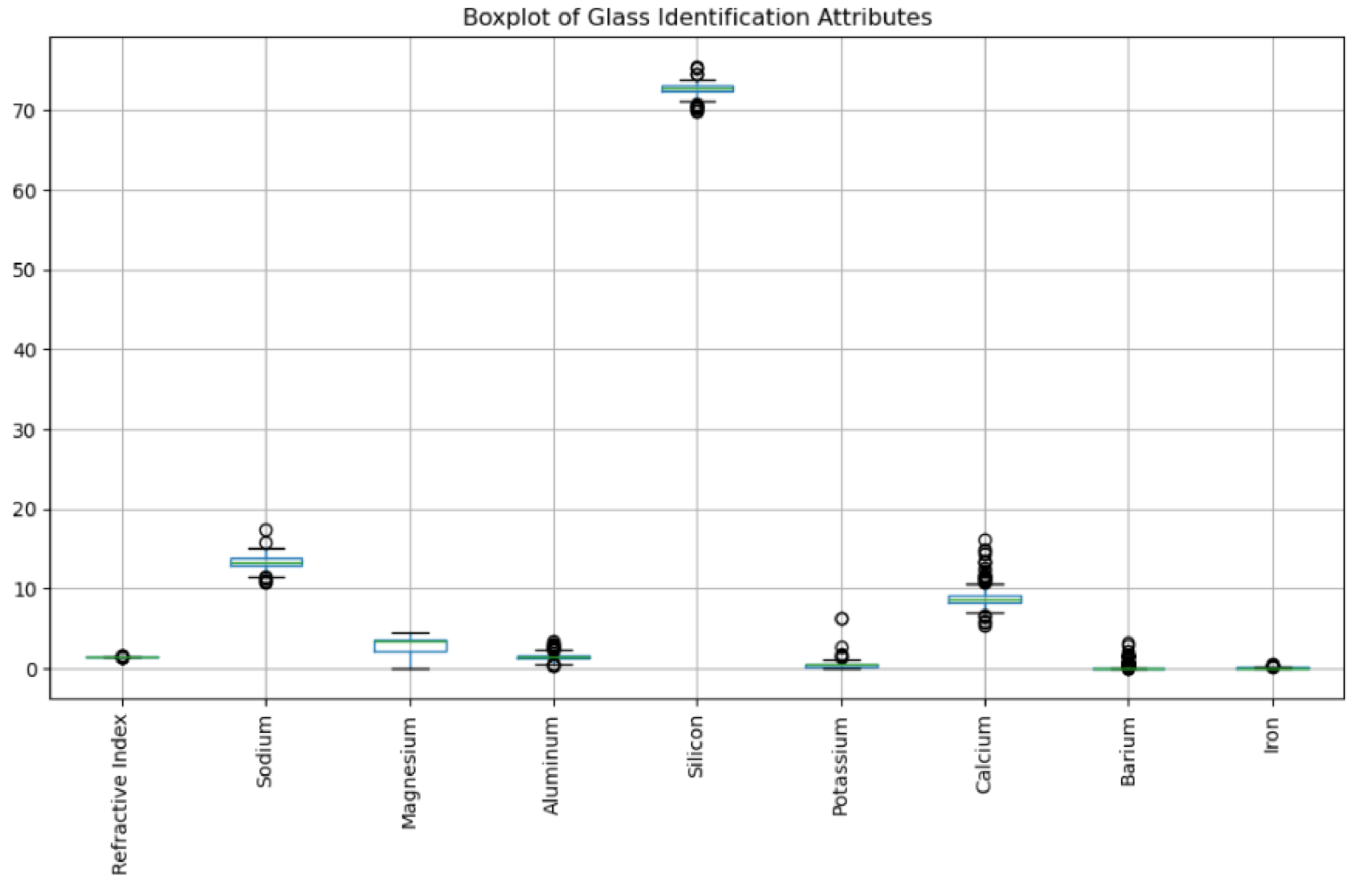
Analysis:

The count plot reveals a noticeable imbalance between the glass types, particularly with Classes 1 and 2 being overrepresented, while Classes 3, 5, 6, and 7 are underrepresented.

[ ]:

6.3 **Box and Whisker Plot for Each Attribute**

[17]:
```python
# Boxplot for each attribute
# visualize the spread and outliers
plt.figure(figsize=(12, 6))
glass_data.drop('Class', axis=1).boxplot()
plt.title('Boxplot of Glass Identification Attributes')
plt.xticks(rotation=90)
plt.show()
```

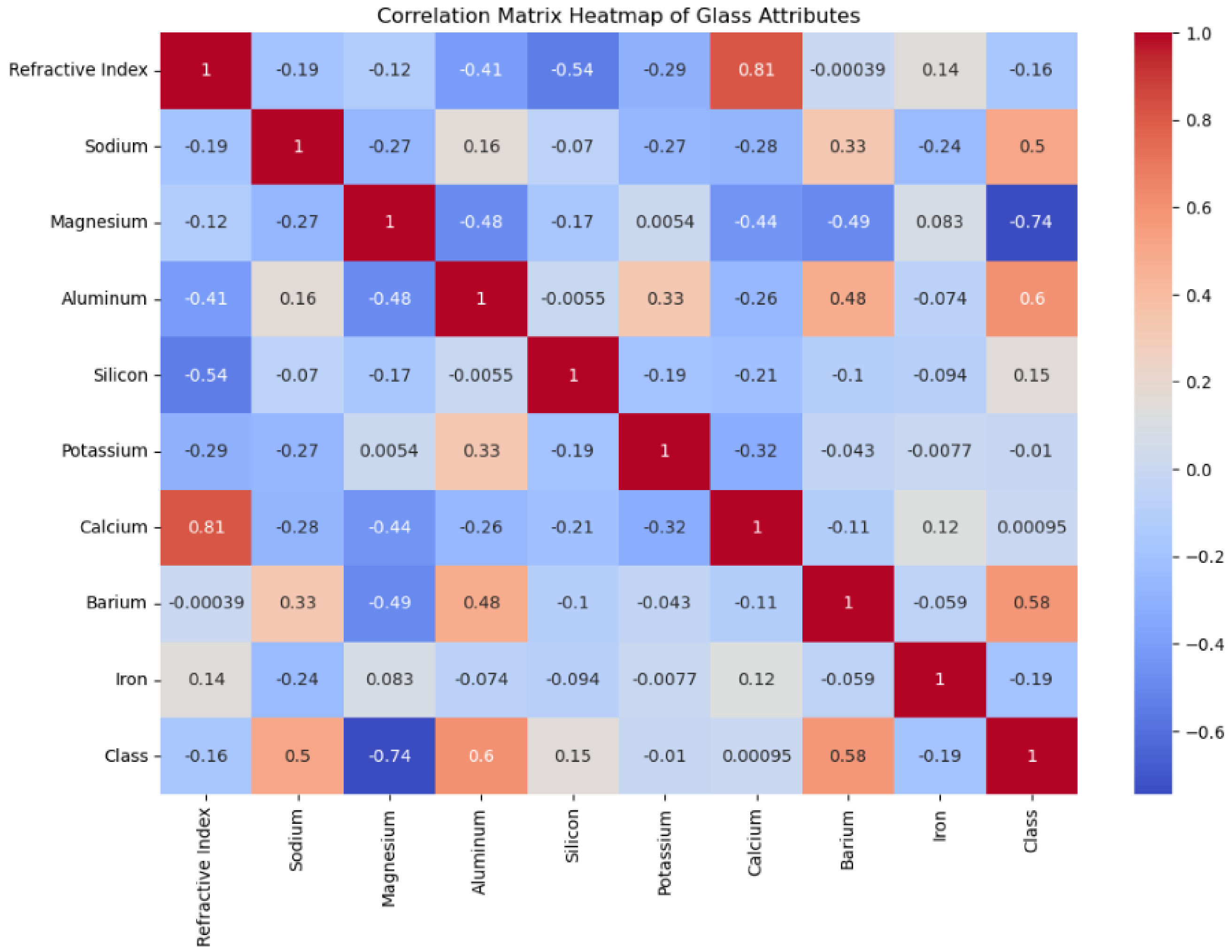


Analysis:

The plots show us that while some features like **Sodium** and **Calcium** vary a lot between glass samples, others like **Iron** stay relatively constant. We also see some outliers in features like **Potassium** and **Magnesium** that may need to be addressed.

6.4 Heatmap Analysis

[19]:
```python
# Correlation matrix
correlation_matrix = glass_data.corr()
# plot a heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix Heatmap of Glass Attributes')
plt.show()
```



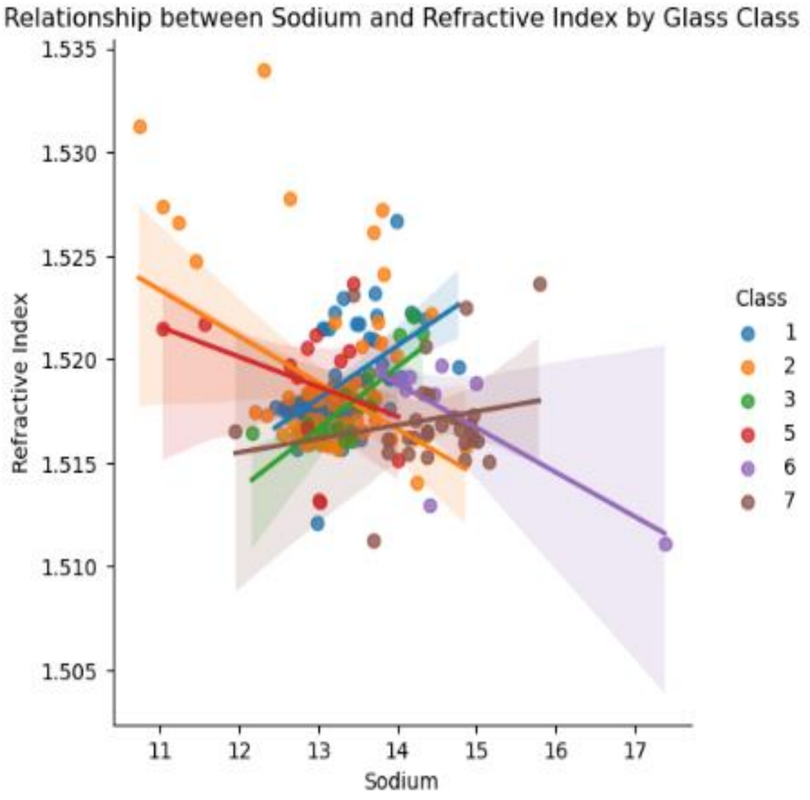
Analysis:

- The heatmap reveals that there are strong positive correlations between **Sodium** and **Calcium**, and between **Magnesium** and **Sodium.** There is also a negative correlation between **Calcium** and **Aluminum.** These relationships suggest that certain glass elements tend to increase or decrease together, which may have implications for glass classification. However, some features like **Iron** seem to act independently, with little to no correlation with other features.

6.5 **Relationship Plots**

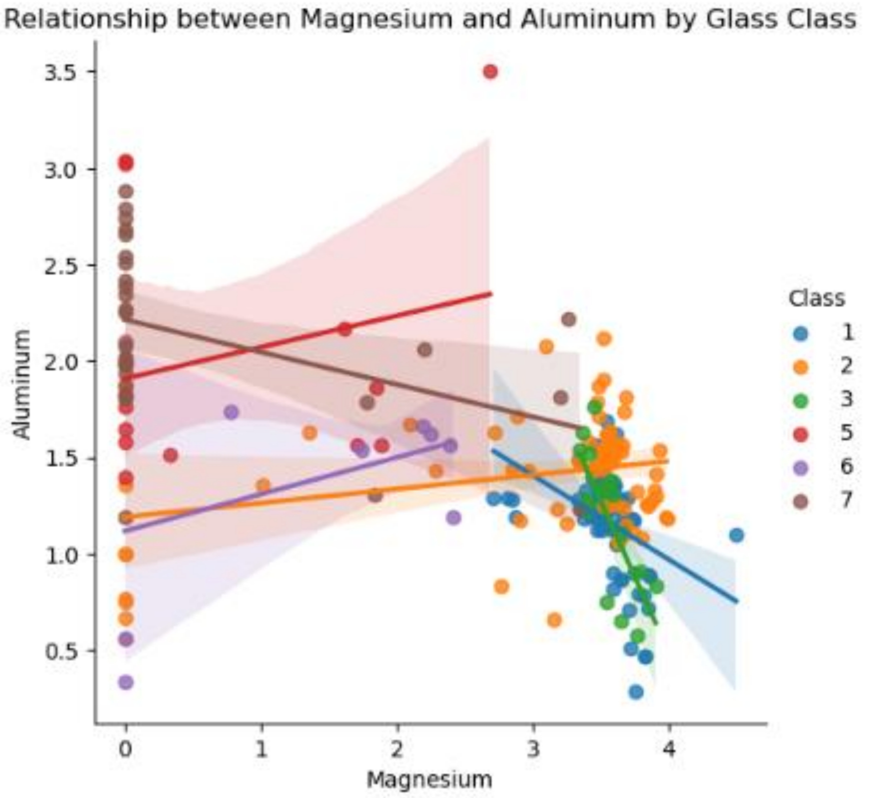
[33]:

```
# Example of a relationship plot between Sodium and Refractive Index
sns.lmplot(x='Sodium', y='Refractive Index', hue='Class', data=glass_data)
plt.title('Relationship between Sodium and Refractive Index by Glass Class')
plt.show()
```

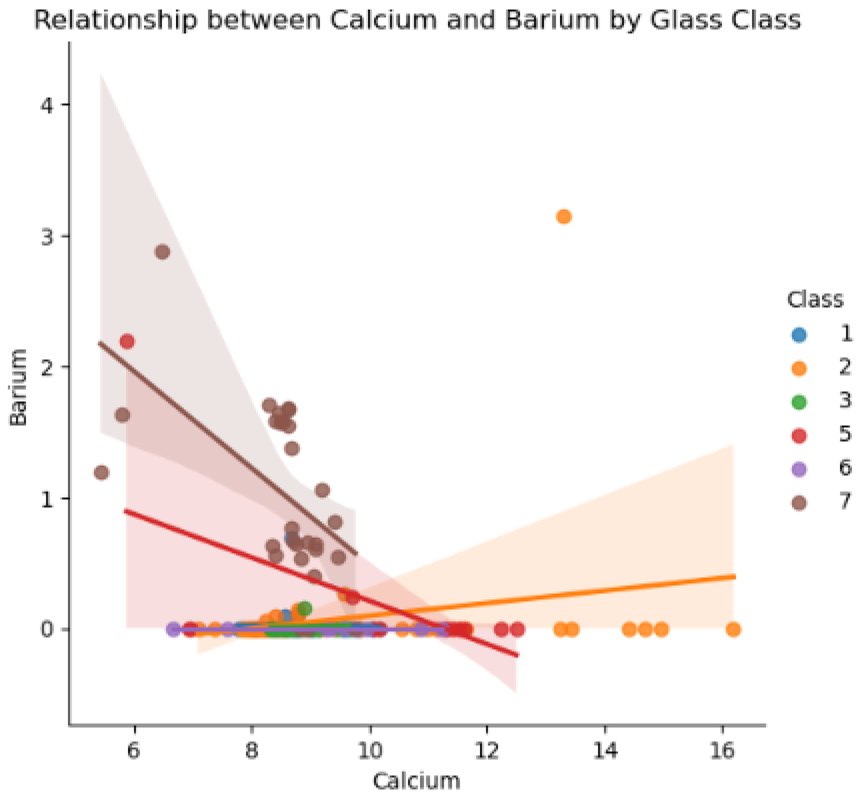


[35]:

```
# Relationship plot for Magnesium vs Aluminum
sns.lmplot(x='Magnesium', y='Aluminum', hue='Class', data=glass_data)
plt.title('Relationship between Magnesium and Aluminum by Glass Class')
plt.show()
```



```
37]: # Relationship plot for Calcium vs Barium
sns.lmplot(x='Calcium', y='Barium', hue='Class', data=glass_data)
plt.title('Relationship between Calcium and Barium by Glass Class')
plt.show()
```



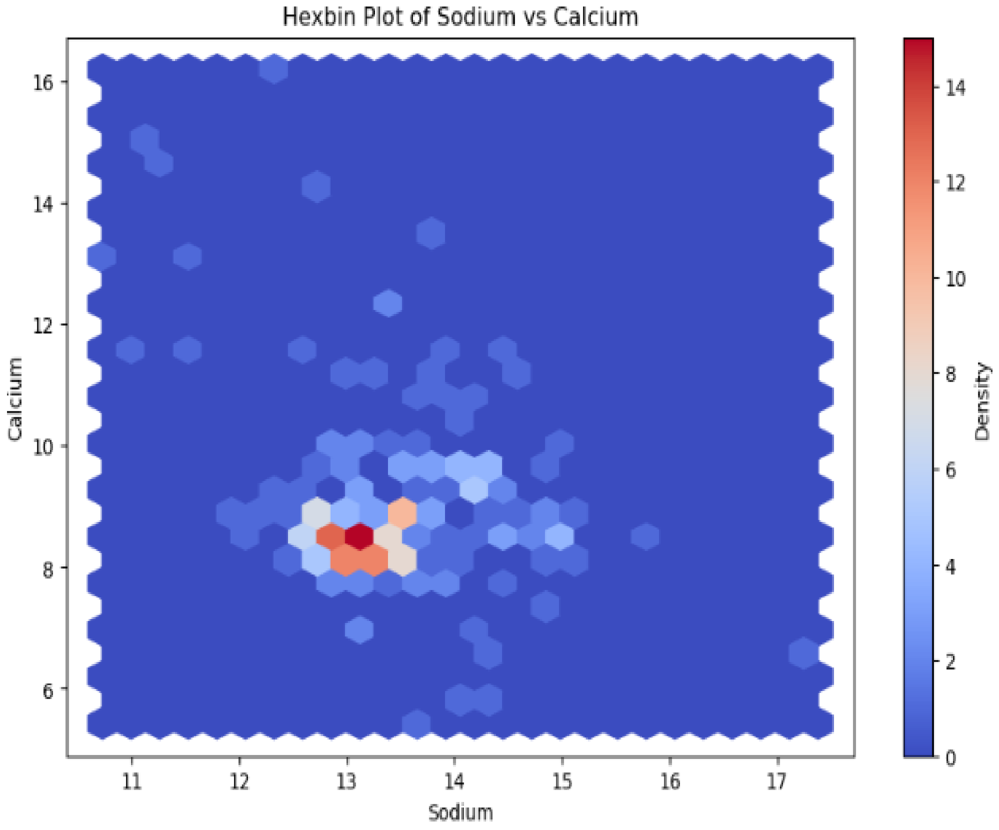
Analysis:

- **Magnesium vs. Aluminum**
- **Calcium vs. Barium**

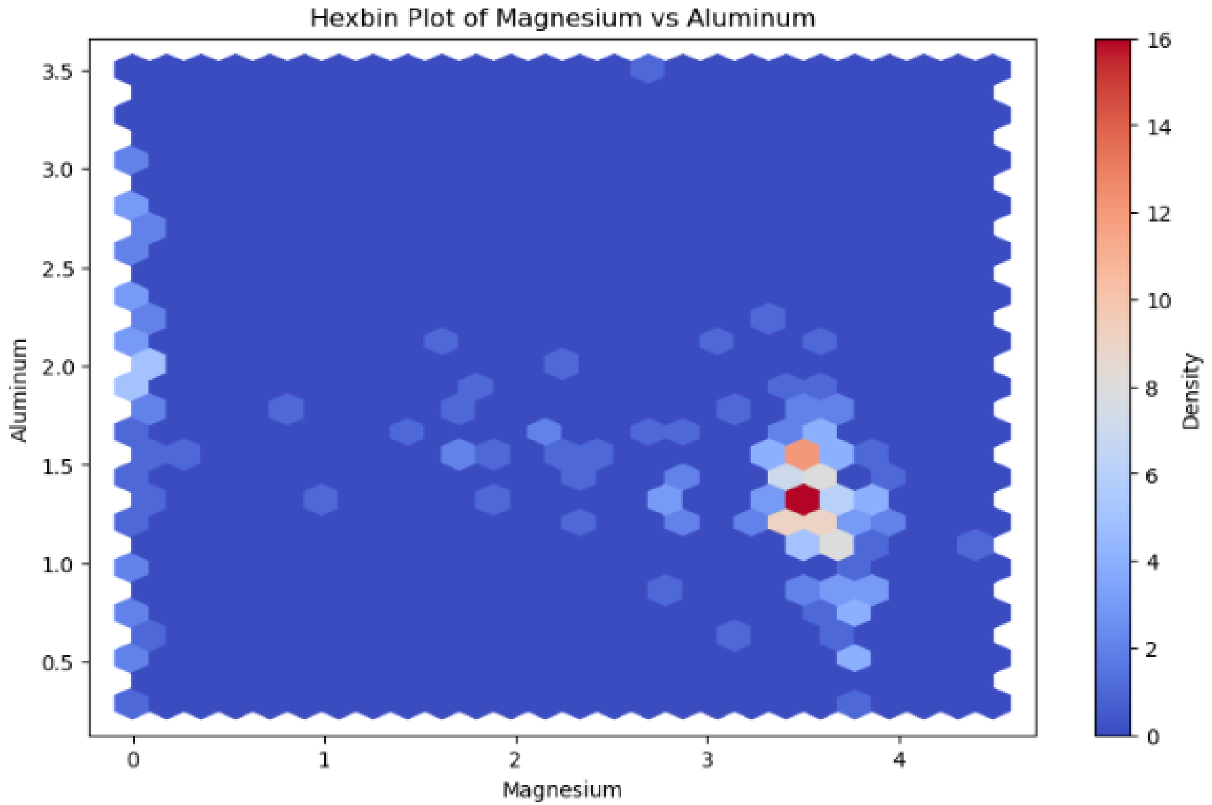
The relationships between features vary depending on the glass type. **Magnesium** and **Aluminum** show a strong, negative correlation for some glass types, indicating a trade-off in their use, while **Calcium** and **Barium** show a weaker positive correlation. The variability in relationships across glass types highlights how different elements influence the composition of glass differently, depending on its intended use (e.g., windows vs. containers).

6.6 Hexbin Plot Analysis

```python
# Hexbin plot for Sodium vs Calcium
plt.figure(figsize=(10, 6))
plt.hexbin(glass_data['Sodium'], glass_data['Calcium'], gridsize=25, cmap='coolwarm')
plt.title('Hexbin Plot of Sodium vs Calcium')
plt.xlabel('Sodium')
plt.ylabel('Calcium')
plt.colorbar(label='Density')
plt.show()
```




```python
# Hexbin plot for Magnesium vs Aluminum
plt.figure(figsize=(10, 6))
plt.hexbin(glass_data['Magnesium'], glass_data['Aluminum'], gridsize=25, cmap='coolwarm')
plt.title('Hexbin Plot of Magnesium vs Aluminum')
plt.xlabel('Magnesium')
plt.ylabel('Aluminum')
plt.colorbar(label='Density')
plt.show()
```

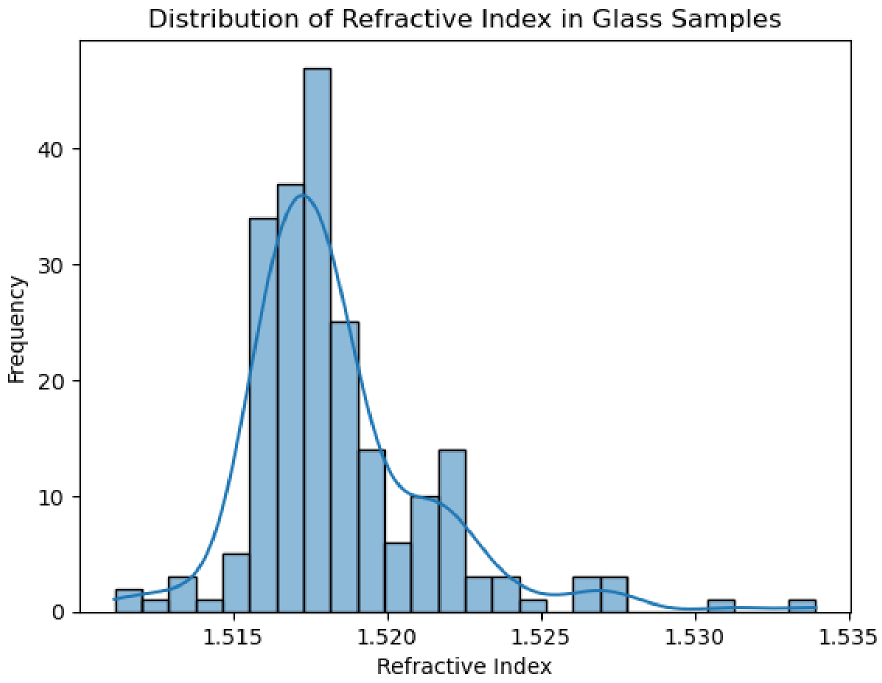


Analysis:

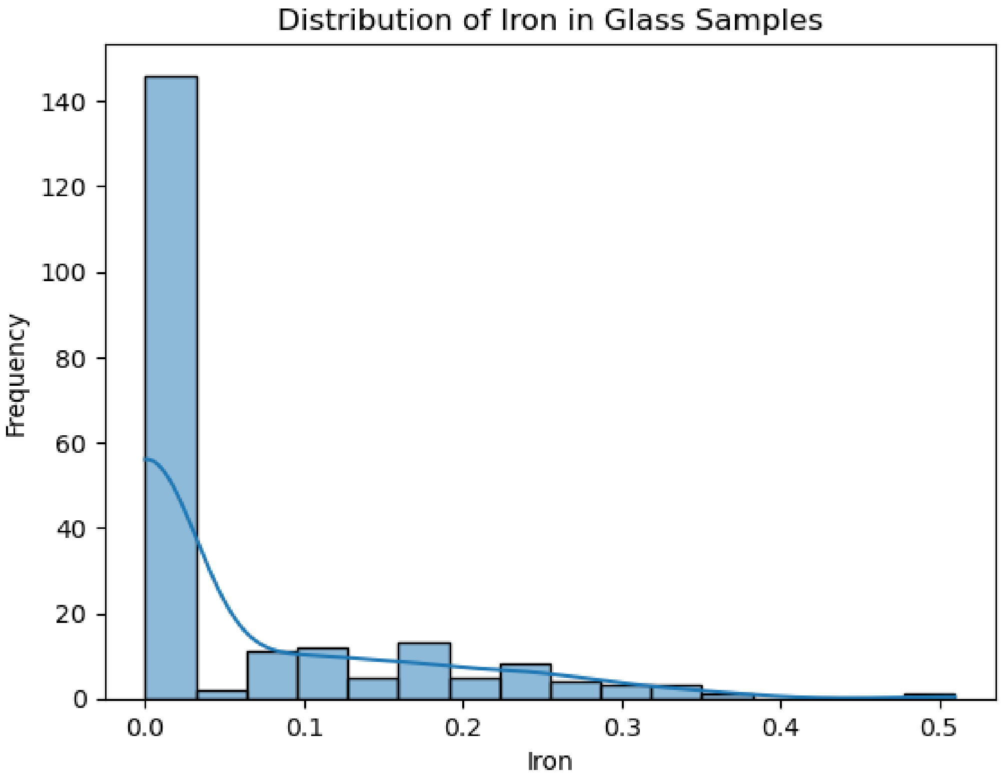
- **Sodium vs. Calcium**
- **Magnesium vs. Aluminum**
- The hexbin plots for **Sodium vs. Calcium** and **Magnesium vs. Aluminum** reveal that **moderate levels** of these elements are the most common in the glass samples, especially for **building window glass**. However, glass samples with extreme levels of these elements are rare and may represent unique or specialized types of glass. Understanding these density patterns is essential for making informed decisions in modeling, as it highlights where most data points lie and where outliers may present challenges.

6.7 Distribution Plot

```
# Distribution plot for Refractive Index
sns.histplot(glass_data['Refractive Index'], kde=True)
plt.title('Distribution of Refractive Index in Glass Samples')
plt.xlabel('Refractive Index')
plt.ylabel('Frequency')
plt.show()
```




```
# Distribution plot for Iron
sns.histplot(glass_data['Iron'], kde=True)
plt.title('Distribution of Iron in Glass Samples')
plt.xlabel('Iron')
plt.ylabel('Frequency')
plt.show()
```

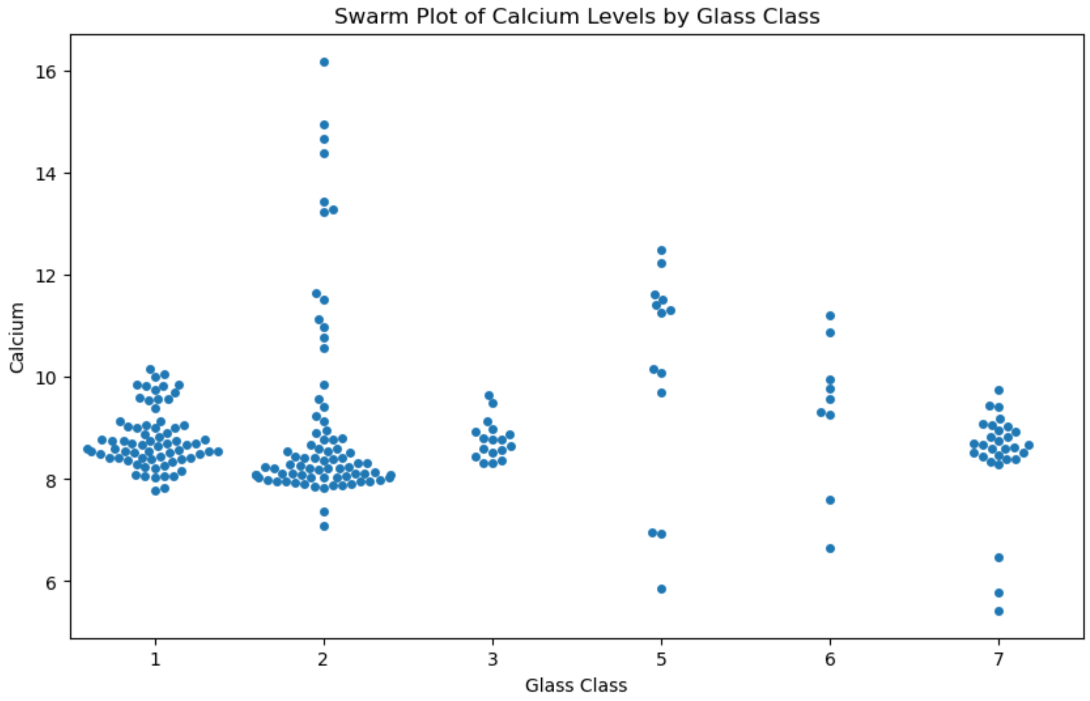


Analysis:

- The distribution plots for **Refractive Index** and **Iron** show contrasting behaviors. **Refractive Index** is normally distributed and consistent across glass samples, making it a stable and useful feature for classification. On the other hand, **Iron** is skewed and has several outliers, indicating that it could play a more significant role in identifying specialized glass types but may require transformation to reduce skewness.

6.8 Swarm Plot

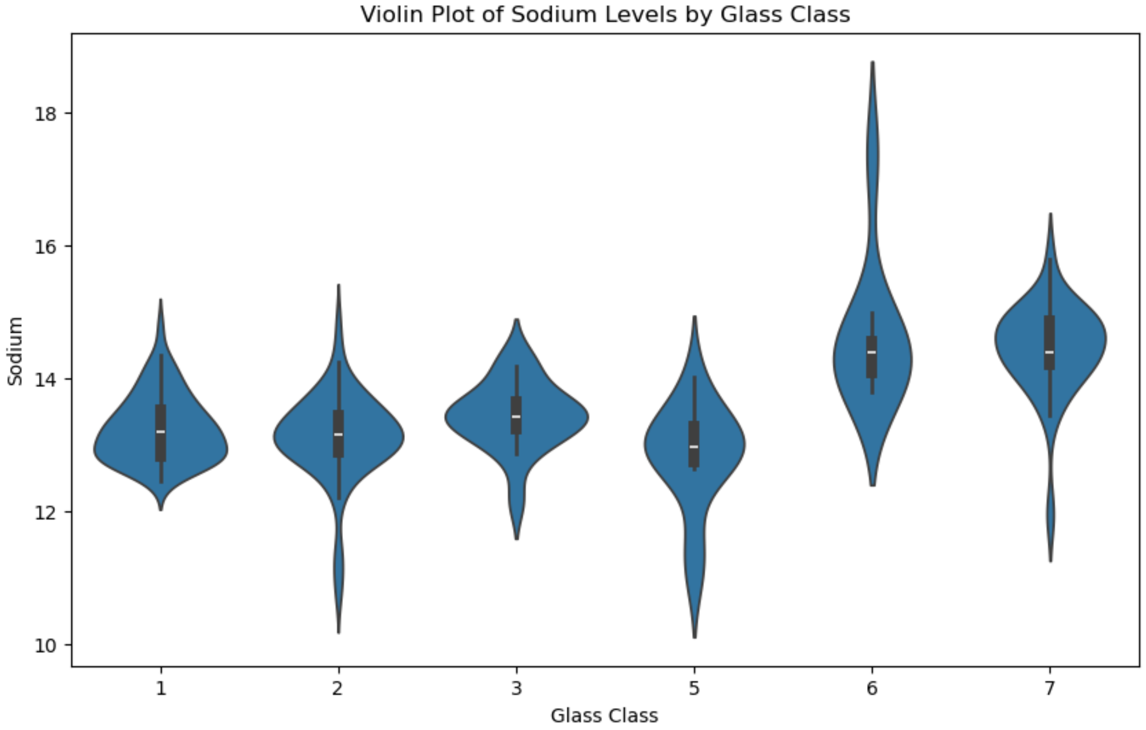
```
# Swarm plot for Calcium levels by glass class
plt.figure(figsize=(10, 6))
sns.swarmplot(x='Class', y='Calcium', data=glass_data)
plt.title('Swarm Plot of Calcium Levels by Glass Class')
plt.xlabel('Glass Class')
plt.ylabel('Calcium')
plt.show()
```



Analysis:
The swarm plot shows that **Calcium levels** are relatively consistent in **building window glass (Classes 1 and 2)** but vary widely in other classes like **headlamp glass (Class 7)** and **container glass (Class 5).** These outliers and variability suggest that Calcium plays different roles in the production of various glass types.

6.9 Violin Plot

```python
# Violin plot for Sodium levels by class
plt.figure(figsize=(10, 6))
sns.violinplot(x='Class', y='Sodium', data=glass_data)
plt.title('Violin Plot of Sodium Levels by Glass Class')
plt.xlabel('Glass Class')
plt.ylabel('Sodium')
plt.show()
```



Analysis:
The violin plot shows that **Sodium levels** are tightly clustered in **building window glass (Classes 1 and 2)**, which implies a strong consistency in their chemical composition. **Vehicle window glass (Class 3)** and **headlamp glass (Class 7)** have more spread, indicating a greater range of Sodium content.