

```
In [5]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('attorney_case_dataset.csv', na_values=[], keep_default_na=False)

print(df.head())
print(df.info())
print(df.describe())

print("\nMissing values:\n", df.isnull().sum())

# 1) Distribution of Case Outcomes
sns.countplot(x='Case_Outcome', data=df)
plt.title("Distribution of Case Outcomes")
plt.xlabel("Case Outcome")
plt.ylabel("Count")
plt.show()

# 2) Attorney Experience vs Case Outcome
sns.boxplot(x='Case_Outcome', y='Experience_Years', data=df)
plt.title("Attorney Experience vs Case Outcome")
plt.xlabel("Case Outcome")
plt.ylabel("Years of Experience")
plt.show()

# 3) Average Win/Loss Ratio by Education Level
sns.barplot(x='Education_Level', y='Win_Loss_Ratio', data=df)
plt.title("Average Win/Loss Ratio by Education Level")
plt.xlabel("Education Level")
plt.ylabel("Avg. Win/Loss Ratio")
plt.xticks(rotation=45)
plt.show()

# 4) Weighted Score Distribution by Case Outcome
sns.violinplot(x='Case_Outcome', y='Weighted_Score', data=df)
plt.title("Weighted Score Distribution by Case Outcome")
plt.xlabel("Case Outcome")
plt.ylabel("Weighted Score")
plt.show()

# 5) Correlation Matrix of Numeric Features
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix of Numeric Features")
plt.show()
```

	Experience_Years	Win_Loss_Ratio	Education_Level	Weighted_Score	\
0	16	0.72	JD	0.62	
1	0	0.49	None	0.23	
2	11	0.59	LLM	0.44	
3	13	0.37	JD	0.47	
4	1	0.38	JD	0.27	

Case_Outcome	
0	1
1	0
2	0
3	0
4	0

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Experience_Years 200 non-null   int64
1   Win_Loss_Ratio   200 non-null   float64
2   Education_Level   200 non-null   object
3   Weighted_Score   200 non-null   float64
4   Case_Outcome     200 non-null   int64
dtypes: float64(2), int64(2), object(1)
memory usage: 7.9+ KB
None
```

	Experience_Years	Win_Loss_Ratio	Weighted_Score	Case_Outcome
count	200.000000	200.000000	200.000000	200.000000
mean	14.905000	0.545000	0.529000	0.355000
std	9.665042	0.203154	0.175660	0.479714
min	0.000000	0.200000	0.150000	0.000000
25%	6.000000	0.377500	0.380000	0.000000
50%	14.500000	0.530000	0.520000	0.000000
75%	24.000000	0.730000	0.672500	1.000000
max	30.000000	0.900000	0.890000	1.000000

Missing values:

Experience_Years	0
Win_Loss_Ratio	0
Education_Level	0
Weighted_Score	0
Case_Outcome	0

dtype: int64



