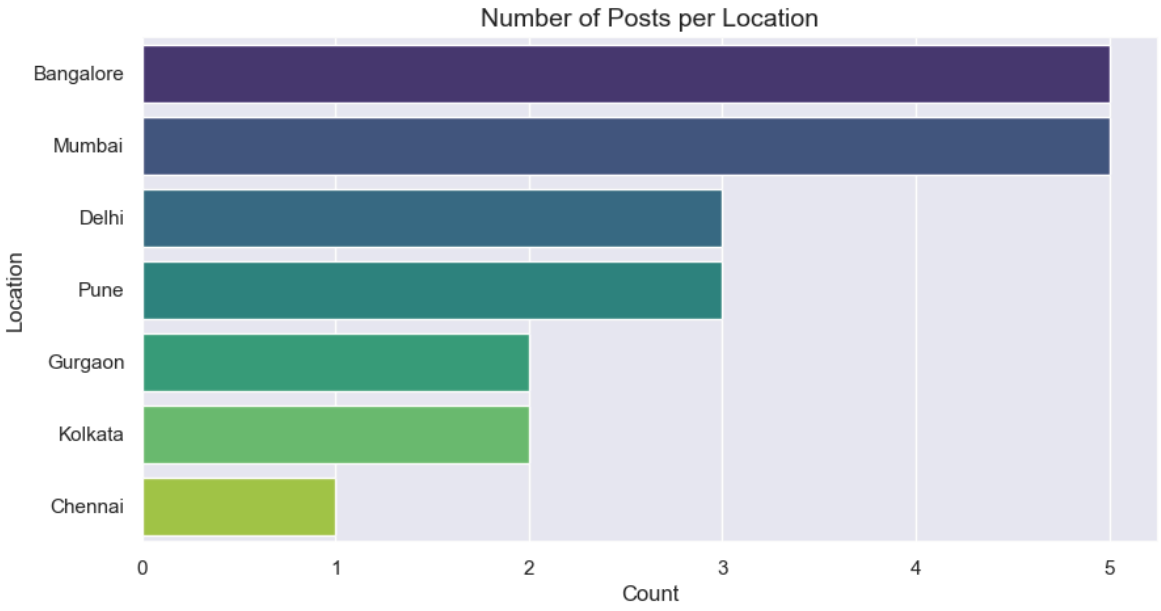
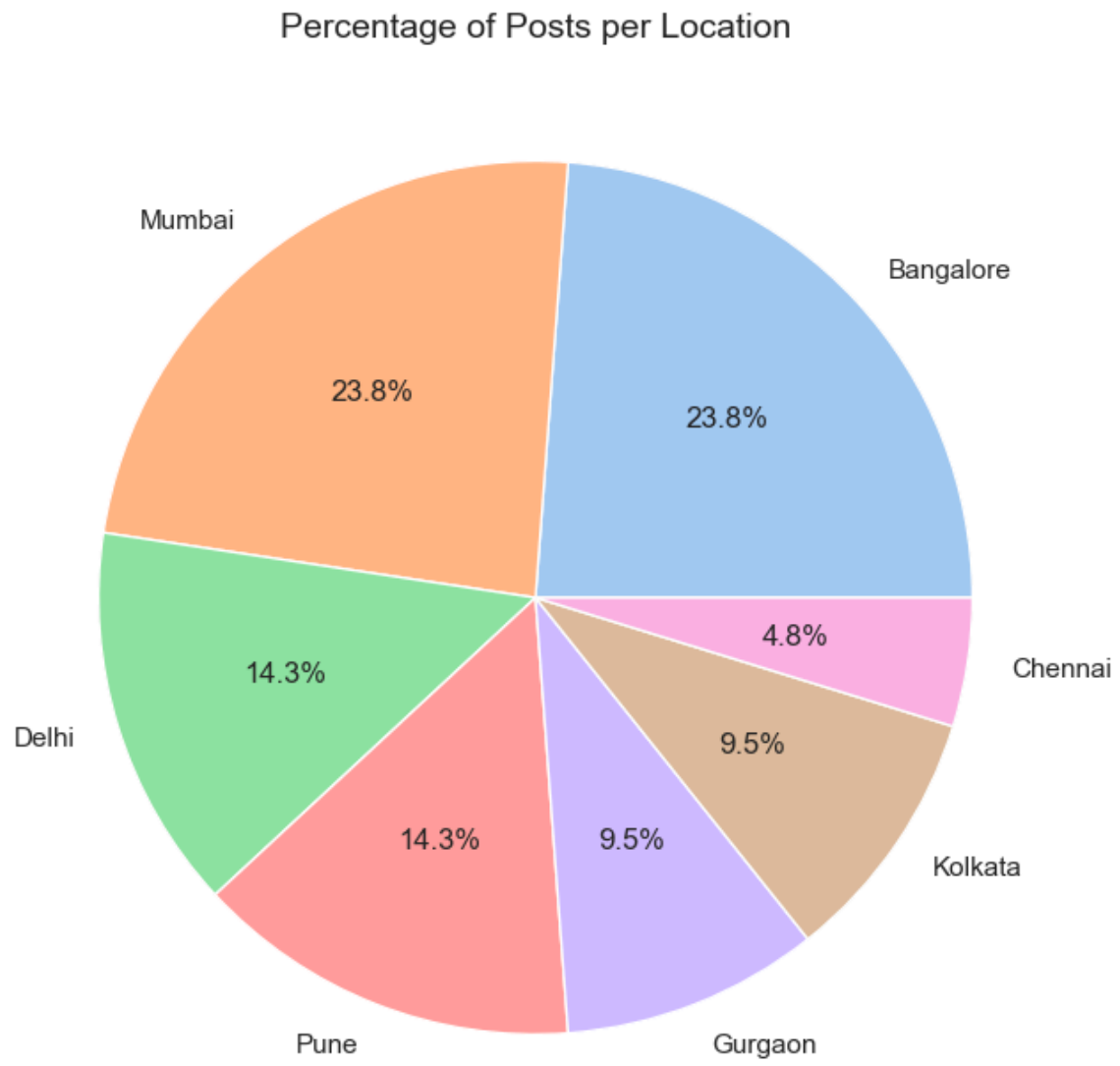
In [14]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
import json

# Load JSON file
with open("preprocessed_legal_advice.json", "r") as f:
    data = json.load(f)

# Convert JSON to DataFrame
df = pd.DataFrame(data)

# Convert UNIX timestamp to datetime
df["created_date"] = pd.to_datetime(df["created_date"], unit="ms")

# Set Seaborn style
sns.set(style="darkgrid")

#  🔹 1. Bar Chart - Number of Posts per Location
plt.figure(figsize=(10, 5))
sns.countplot(y=df["location"], order=df["location"].value_counts().index, palet
plt.title("Number of Posts per Location", fontsize=14)
plt.xlabel("Count")
plt.ylabel("Location")
plt.show()

#  🔹 2. Pie Chart - Percentage of Posts per Location


#  🔹 3. Scatter Plot - Score vs. Number of Comments


#  🔹 4. Heatmap - Correlation Matrix


#  🔹 5. Word Cloud - Most Frequent Words in Selftext
```

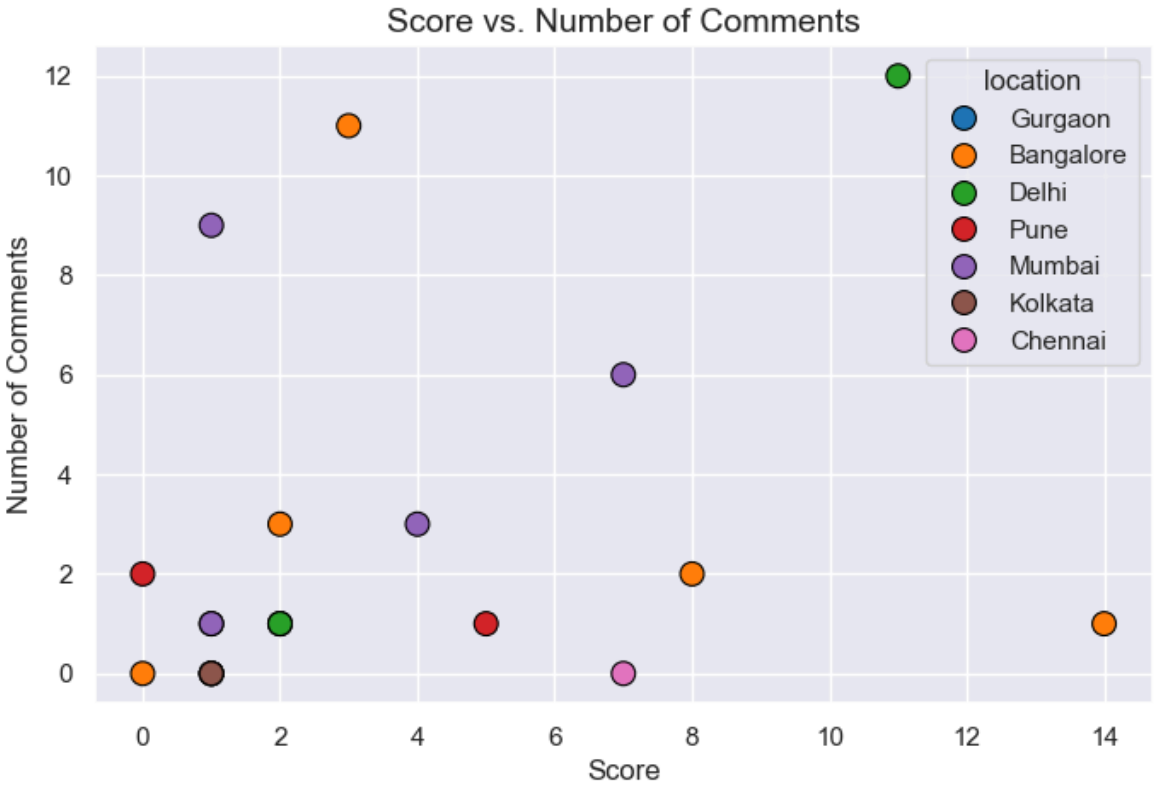
```
C:\Users\Komal\AppData\Local\Temp\ipykernel_8692\2463527399.py:22: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v
0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effe
ct.

  sns.countplot(y=df["location"], order=df["location"].value_counts().index, pale
tte="viridis")
```



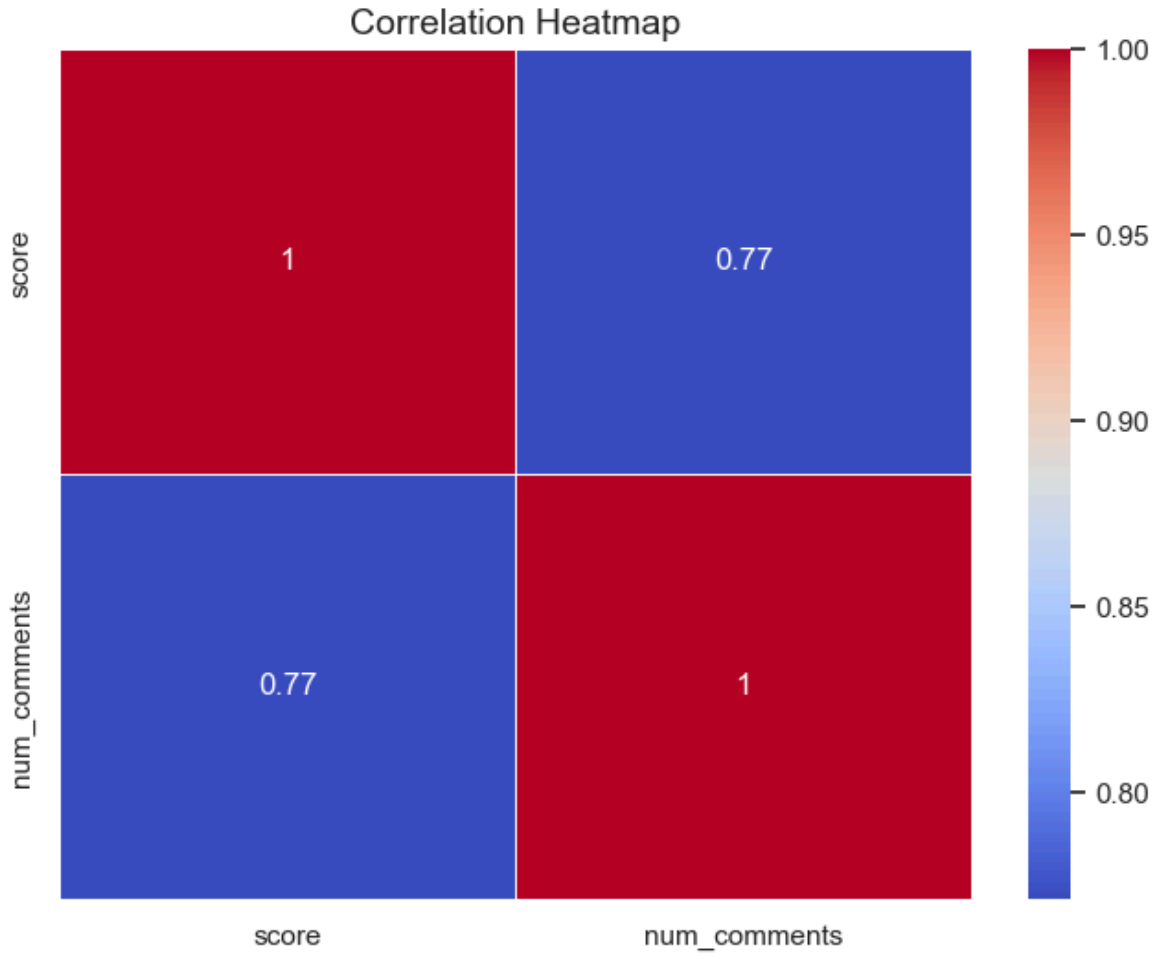
In [15]:
```
plt.figure(figsize=(8, 8))
df["location"].value_counts().plot.pie(autopct="%1.1f%%", colors=sns.color_palet
plt.title("Percentage of Posts per Location", fontsize=14)
plt.ylabel("")  # Hide y-label
plt.show()
```



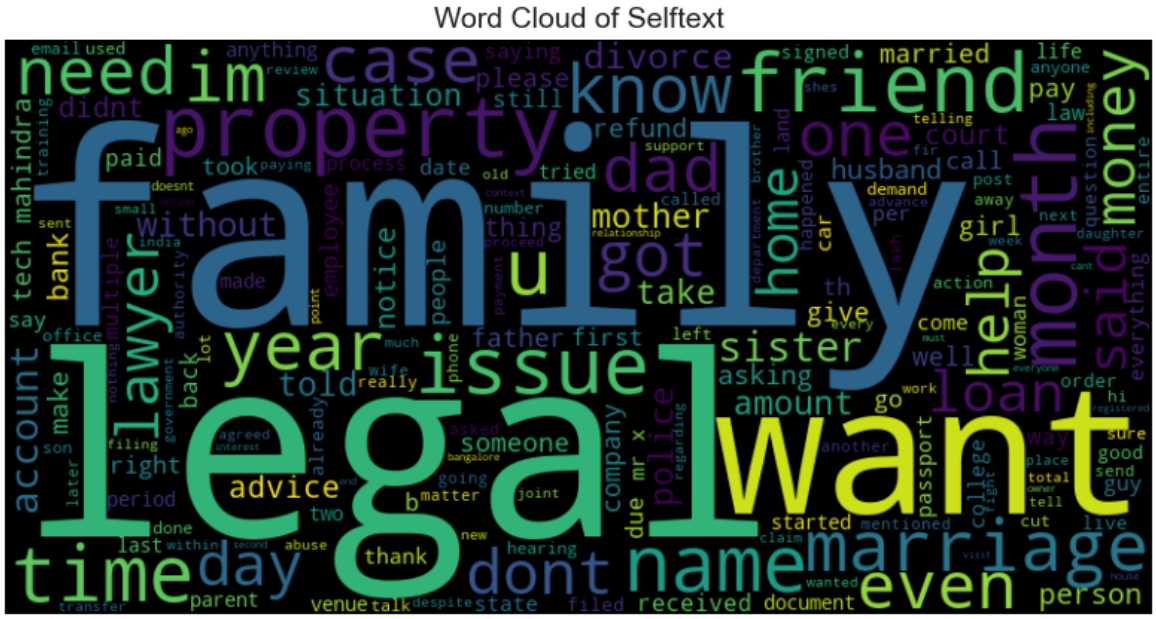
In [16]:
```
plt.figure(figsize=(8, 5))
sns.scatterplot(x=df["score"], y=df["num_comments"], hue=df["location"], palette
plt.title("Score vs. Number of Comments", fontsize=14)
plt.xlabel("Score")
plt.ylabel("Number of Comments")
plt.show()
```



In [10]:
```
plt.figure(figsize=(8, 6))
sns.heatmap(df[["score", "num_comments"]].corr(), annot=True, cmap="coolwarm", l
plt.title("Correlation Heatmap", fontsize=14)
plt.show()
```

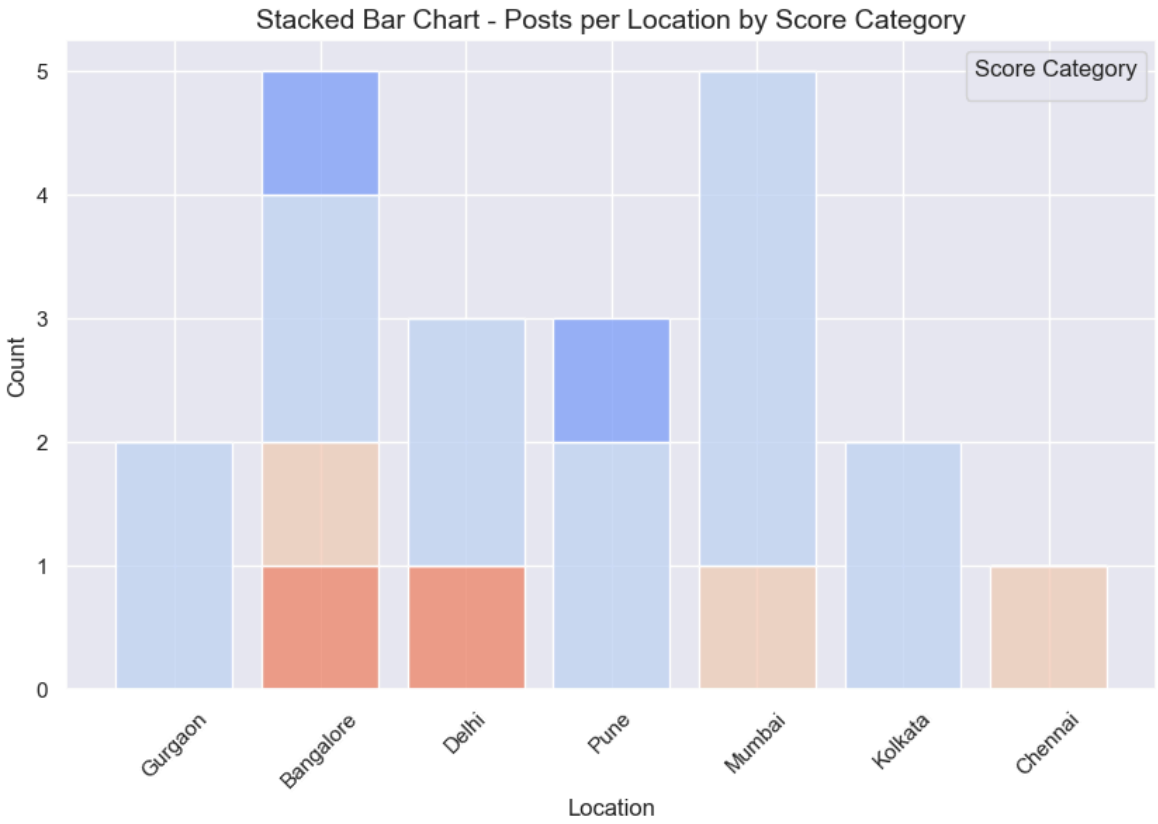


In [17]:
```
text = " ".join(df["selftext"])
wordcloud = WordCloud(width=800, height=400, background_color="black", colormap=

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.title("Word Cloud of Selftext", fontsize=14)
plt.show()
```



In [18]:
```
# Categorize scores into bins
df["score_category"] = pd.cut(df["score"], bins=[-1, 0, 5, 10, 50], labels=["Low

# Plot Stacked Bar Chart
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x="location", hue="score_category", multiple="stack", pale
plt.title("Stacked Bar Chart - Posts per Location by Score Category", fontsize=1
plt.xlabel("Location")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.legend(title="Score Category")
plt.show()
```

```
C:\Users\Komal\AppData\Local\Temp\ipykernel_8692\307793973.py:11: UserWarning: No
artists with labels found to put in legend.  Note that artists whose label start
with an underscore are ignored when legend() is called with no argument.
  plt.legend(title="Score Category")
```


Stacked Bar Chart - Posts per Location by Score Category
Score Category
Count
0
1
2
3
4
5
Gurgaon
Bangalore
Delhi
Pune
Mumbai
Kolkata
Chennai
Location

In [ ]: