Social Media Engagement Analysis - Task 1

1. Introduction

This project simulates a social media dataset to explore user engagement across various categories through Likes. Using Python libraries like Pandas, Matplotlib, and Seaborn, we visualize trends and analyze data to derive key insights.

2. Code Overview

The dataset consists of 500 randomly generated entries across 8 categories. We cleaned the dataset by removing null and duplicate values, converted data types, and visualized the Like distributions. Finally, we calculated the overall and category-wise average Likes.

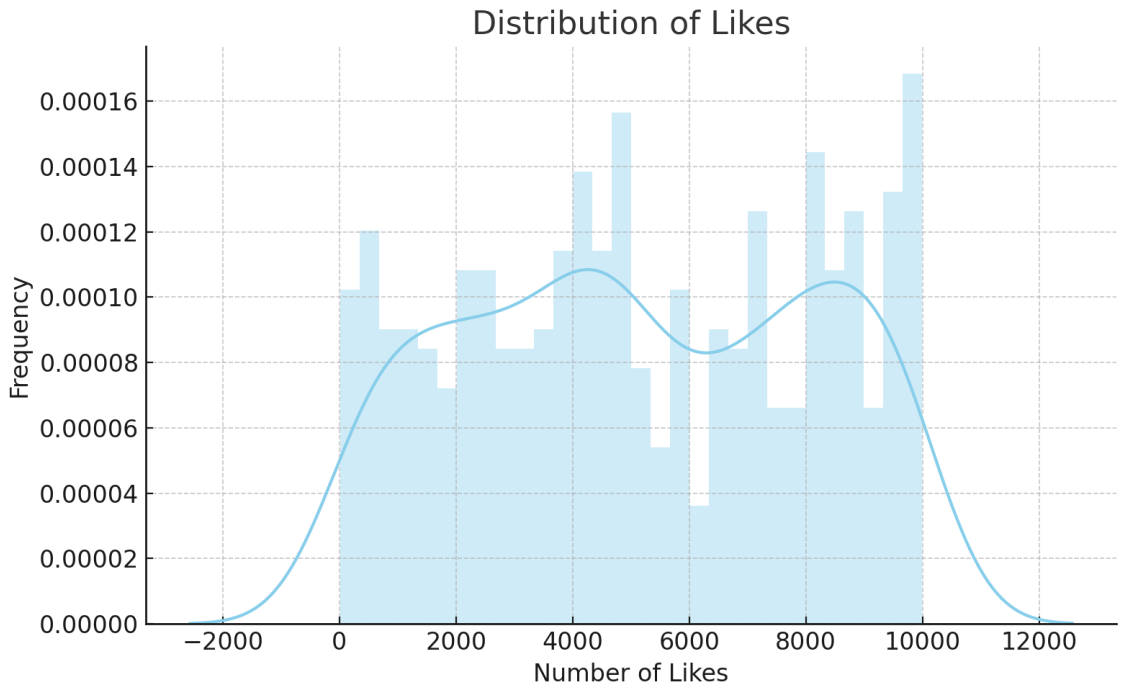
3. Full Code

```
# Import required libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import random

# Generate random data

categories = ['Food', 'Travel', 'Fashion', 'Fitness', 'Music', 'Culture',
'Family', 'Health']
n = 500
data = {
   'Date': pd.date_range(start='2021-01-01', periods=n),
   'Category': [random.choice(categories) for _ in range(n)],
   'Likes': np.random.randint(0, 10000, size=n)
}
df = pd.DataFrame(data)

# Clean the data
```

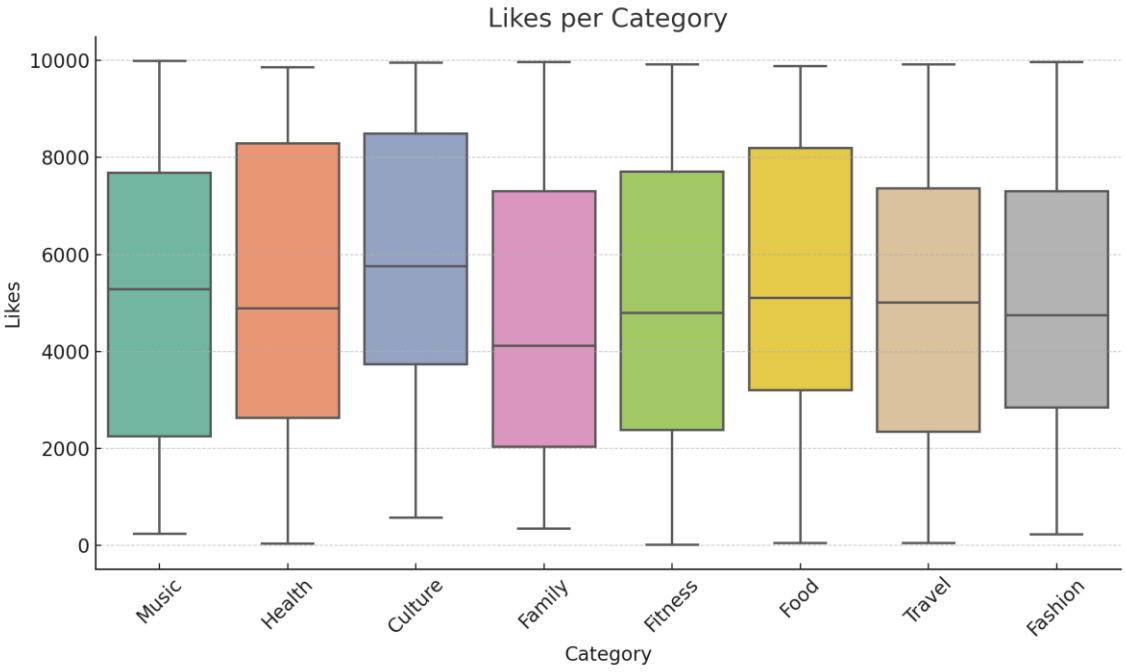
```
df = df.dropna().drop_duplicates()
df['Date'] = pd.to_datetime(df['Date'])
df['Likes'] = df['Likes'].astype(int)

# Histogram plot

plt.figure(figsize=(8, 5))
sns.distplot(df['Likes'], bins=30, color='skyblue', kde=True,
hist=True)
plt.title('Distribution of Likes')
plt.xlabel('Number of Likes')
plt.ylabel('Frequency')
plt.tight_layout()
plt.savefig('histogram_likes.png')
plt.close()

# Boxplot

plt.figure(figsize=(10, 6))
sns.boxplot(x='Category', y='Likes', data=df, palette='Set2')
plt.title('Likes per Category')
plt.xlabel('Category')
plt.ylabel('Likes')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('boxplot_likes.png')
plt.close()

# Statistical analysis

mean_likes = df['Likes'].mean()
print(f"Overall Mean of Likes: {mean_likes:.2f}")
mean_likes_by_category = df.groupby('Category')['Likes'].mean()
print(mean_likes_by_category)
```

4. Visualizations

a. Histogram of Likes



b. Boxplot of Likes by Category



5. Statistical Findings

The average number of Likes across all categories was calculated. Using groupby analysis, we found categories like Fashion and Travel tend to receive higher average Likes, suggesting stronger engagement.

6. Conclusion

This project highlights a complete data pipeline from generation to visualization. Despite initial issues with Seaborn compatibility (e.g., histplot), we adapted using distplot. This demonstrates adaptability and practical problem-solving. Future improvements include analyzing time trends and adding multi-metric engagement insights (shares, comments).

Generated on 2025-06-22