

## 4

### Exploratory Data Analysis and visualization of Social Media Data for business

# No need to install these in Colab unless seaborn/matplotlib are missing

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
import os
```

# Set consistent plot style

```
sns.set_style("whitegrid")
```

# Simulated upload section – if you're uploading manually, use:

```
# from google.colab import files
```

```
# uploaded = files.upload()
```

# For testing, create a small dummy dataset

```
data = {
    'platform': ['Instagram', 'Twitter', 'Facebook', 'Instagram', 'Twitter'],
    'likes': [120, 340, np.nan, 540, 220],
    'followers': [1500, 2300, 1800, 2000, np.nan],
    'comments': [12, 34, 15, 45, 23],
    'engagement_rate': [0.08, 0.12, 0.09, 0.15, 0.11]
}
social_media_data = pd.DataFrame(data)
```

# Handle missing values

```
print("Missing values:\n", social_media_data.isnull().sum())
```

```
social_media_data.fillna(0, inplace=True)
```

# Save cleaned data

```
social_media_data.to_csv("cleaned_social_media_data.csv", index=False)
```

```
files.download("cleaned_social_media_data.csv")
```

```
print("Summary statistics:\n")
```

```
print(social_media_data.describe())
```

```
plt.figure(figsize=(12,6))
```

```
sns.histplot(social_media_data['likes'], kde=True, bins=30)
```

```
plt.title('Distribution of Likes')
```

```
plt.xlabel('Number of Likes')
```

```
plt.ylabel('Frequency')
```

```
plt.savefig("likes_distribution.png")
```

```
plt.show()
```

```
files.download("likes_distribution.png")
```

```
plt.figure(figsize=(12,6))
```

```
sns.boxplot(x='platform', y='followers', data=social_media_data)
```

```
plt.title('Followers by Platform')
```

```
plt.xticks(rotation=45)
```

```
plt.savefig("followers_by_platform.png")
```

```
plt.show()
files.download("followers_by_platform.png")

numerical_data = social_media_data.select_dtypes(include=np.number)
corr_matrix = numerical_data.corr()

plt.figure(figsize=(10,6))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.savefig("correlation_matrix.png")
plt.show()
files.download("correlation_matrix.png")

plt.figure(figsize=(12,6))
sns.scatterplot(x='comments', y='engagement_rate', data=social_media_data)
plt.title('Comments vs Engagement Rate')
plt.savefig("comments_vs_engagement.png")
plt.show()
files.download("comments_vs_engagement.png")

plt.figure(figsize=(12,6))
sns.boxplot(x='platform', y='likes', data=social_media_data)
plt.title('Likes by Platform')
plt.xticks(rotation=45)
plt.savefig("likes_by_platform.png")
plt.show()
files.download("likes_by_platform.png")

sns.pairplot(data=social_media_data)
plt.savefig("pairplot.png")
plt.show()
files.download("pairplot.png")
```