Experiment 2:

Topic: Select the social media platforms of your choice (Twitter, Facebook, LinkedIn, Youtube,Web blogs etc), connect to and capture social media data for business.

Group no. 51

Data source: Reddit API wrapper

Dataset link: https://github.com/Hiten-Dusseja/SMA_Repo/blob/main/exp2_gh/legal_advice_india_all.json

Code:

```
import praw
import json
from datetime import datetime
import re


reddit = praw.Reddit(
    client_id="YkjnPbh7fE-NVhEMap86pw",
    client_secret="kg0xkGmXuQmQZpo48cW00OkQ4nSxvQ",
    user_agent="LegalAdviceIndiaBot by YOUR_USERNAME"
)


subreddit = reddit.subreddit("LegalAdviceIndia")


def extract_location(text):
    # Example list of Indian cities; expand this list as needed
    locations = ["Delhi", "Mumbai", "Bangalore", "Chennai", "Hyderabad", "Kolkata", "Pune", "Noida", "Gurgaon"]
    for location in locations:
        if location.lower() in text.lower():
            return location
    return "Unknown"
```

```
data = []
limit = 2000  # Number of posts to fetch
batch_size = 100  # PRAW allows a maximum of 100 posts per request
count = 0
after = None  # Used to paginate backward

print(f"Fetching up to {limit} posts from r/LegalAdviceIndia...\n")

while count < limit:
    # Fetch a batch of posts
    posts = subreddit.new(limit=batch_size, params={"after": after})

    batch_data = []
    for post in posts:
        count += 1
        post_data = {
            "title": post.title,
            "author": str(post.author),
            "url": post.url,
            "score": post.score,
            "created_utc": post.created_utc,
            "created_date": datetime.utcfromtimestamp(post.created_utc).strftime('%Y-%m-%d %H:%M:%S'),
            "num_comments": post.num_comments,
            "selftext": post.selftext,
            "id": post.id,
            "subreddit": str(post.subreddit),
            "location": extract_location(post.selftext + " " + post.title),  # Combine title and selftext
        }
        batch_data.append(post_data)
```

```
        # Stop if we reach the limit
        if count >= limit:
            break

    data.extend(batch_data)

    if batch_data:
        after = batch_data[-1]["id"]
    else:
        break

    print(f"Fetched {len(batch_data)} posts. Total so far: {count}.\n")

file_name = "legal_advice_india_all.json"
with open(file_name, "w", encoding="utf-8") as json_file:
    json.dump(data, json_file, ensure_ascii=False, indent=4)

print(f"Data saved to '{file_name}' with {len(data)} entries.")
```