Coding Challenge: Social Media Engagement Analysis

Scenario

You are working as a **Data Analyst** for a growing social media platform called **ConnectHub**. The management wants to understand how users are engaging on the platform through **posts, likes, comments, and friendships**.

You are provided with an **Excel dataset** containing user activity. Your task is to clean, analyze, and visualize the data to derive insights.

Dataset Structure (Sample Columns)

Sheet: Users

User_ID	Username	Email	Join_Date	City
1	alex_123	alex@example.com	2023-01-15	New York
2	sara_k	sara@example.com	2023-02-20	Chicago
3	raj90	raj90@example.com	2023-03-05	Mumbai

Sheet: Posts

Post_ID	User_ID	Content_Length (words)	Post_Date
101	1	45	2023-05-01
102	2	120	2023-05-02
103	3	75	2023-05-03

Sheet: Likes

Like_ID	User_ID	Post_ID	Like_Date
201	3	101	2023-05-01
202	1	102	2023-05-02
203	2	101	2023-05-03

Sheet: Comments

Comment_ID	Post_ID	User_ID	Comment_Text	Comment_Date
301	101	2	Great post!	2023-05-01
302	102	3	Thanks for sharing	2023-05-02
303	103	1	Very informative	2023-05-03

Sheet: Friendships

Friendship_ID	User_ID1	User_ID2	Since_Date
401	1	2	2023-03-10
402	2	3	2023-04-15
403	1	3	2023-04-20

Tasks

Task 1: Data Cleaning & Preparation

1. Use **Flash Fill** to extract the domain name from the Email column.
2. Format Join_Date into MMM-YYYY.
3. Remove duplicates in Users table (if any).
4. Use **Text-to-Columns** to separate city and country if provided as “City, Country”.

Task 2: Formulas & Functions

5. Calculate the **number of posts** each user has made. (Use COUNTIF).
6. Find the **average content length** of posts per user. (Use AVERAGEIF).
7. Identify the **most active user** (highest likes + comments given).
8. Use **VLOOKUP/XLOOKUP** to fetch the username of each commenter in the Comments sheet.

9. Use **DATEDIF** to calculate how long each friendship has lasted (in months).

Task 3: Pivot Tables & Charts

10. Create a **Pivot Table** showing total likes and comments per post.
11. Create a **Pivot Chart** (Column chart) for top 5 most liked posts.
12. Show **user activity by city** (number of users, posts, likes).
13. Create a **timeline filter (slicer)** for Post_Date to analyze trends.

Task 4: Conditional Formatting

14. Highlight users who joined **before Jan 2023** in light red.
15. Highlight posts that have **content length > 100 words** in green.
16. Highlight the top 3 posts with most likes.

Task 5: Dashboard

17. Build a one-page **interactive dashboard** with:

- Total users, posts, likes, comments (KPIs)
- Top 5 most active users
- Engagement trend over time (line chart)
- Posts vs Likes (bar chart)
- City-wise user distribution (pie chart)

Expected Outcomes

- Should demonstrate **Excel core skills**: Data Cleaning, Formulas, Lookup Functions, Pivot Tables, Conditional Formatting, Visualization, and Dashboarding.
- Should be able to **interpret business questions into Excel models**.
- Extra credit: use **Power Query** to automate loading & cleaning.