BÀI TẬP

Phân tích và trực quan hóa dữ liệu

- **Bộ dữ liệu giả lập (Dữ liệu truyền thông xã hội)**

Bộ dữ liệu có cấu trúc tương tự như sau:
Link tải DL:
https://drive.google.com/file/d/1Jm8vpJNkKLmkpxmTGIfYGavlTNuQDG6x/view?usp=sharing

Post_ID	User_ID	Age	Gender	Post_Content	Likes	Shares	Comments	Post_Date
1	754	25	Male	Enjoying the weather	62	14	4	2024-09-24 3:43:00
2	858	52	Male	Watching a movie	108	2	0	2024-09-03 6:14:00
3	617	56	Male	Celebrating a birthday	50	45	20	2024-09-23 17:26:00
4	325	46	Female	Great day at the park	194	10	22	2024-09-14 10:17:00
5	259	31	Female	New blog post up!	23	24	3	2024-09-12 11:38:00
6	370	20	Female	Celebrating a birthday	31	24	2	2024-09-18 9:53:00
7	743	57	Female	Watching a movie	49	45	2	2024-09-02 21:14:00
8	891	36	Male	Coffee and chill	25	24	8	2024-09-15 20:53:00
9	473	28	Female	Out with friends	53	42	8	2024-09-23 21:41:00
10	173	56	Male	Celebrating a birthday	186	15	5	2024-09-15 12:17:00
11	1047	58	Male	Busy with work	196	49	1	2024-09-08 1:51:00
12	423	43	Female	Working hard today	54	36	28	2024-09-23 10:13:00
13	771	49	Female	Learning something n	36	16	4	2024-09-08 23:35:00
14	651	34	Female	Watching a movie	102	23	7	2024-09-05 16:31:00
15	193	21	Male	Weekend vibes	160	10	25	2024-09-22 13:38:00
16	165	42	Female	Exploring the city	119	33	8	2024-09-18 0:43:00

Bộ dữ liệu bao gồm 1000 điểm dữ liệu với các thuộc tính:

- Post_ID: Mã định danh của bài đăng
- User_ID: Mã định danh của người dùng
- Age: Độ tuổi của người dùng
- Gender: Giới tính của người dùng (Male/Female)
- Post_Content: Nội dung của bài đăng (giả lập)
- Likes: Số lượt thích
- Shares: Số lượt chia sẻ
- Comments: Số lượt bình luận
- Post_Date: Thời gian đăng bài

- **Câu hỏi phân tích:**
  - **Câu hỏi 1:Liệu số lượt thích của một bài đăng có phụ thuộc vào độ tuổi của người đăng bài không?**
    - Yêu cầu sinh viên phân tích mối quan hệ giữa độ tuổi của người dùng (Age) và số lượt thích (Likes).
    - Sinh viên phải chia độ tuổi thành các nhóm (ví dụ: dưới 20, từ 20-30, 30-40, trên 40) và tính trung bình số lượt thích cho mỗi nhóm tuổi.
  - **Câu hỏi 2:Bài đăng được đăng vào các khung giờ nào trong ngày có xu hướng nhận được nhiều/ít lượt tương tác nhất?**
    - Phân tích sự tương tác (tổng số lượt thích, chia sẻ và bình luận) theo các khung giờ trong ngày (ví dụ: sáng, trưa, chiều, tối). Yêu cầu sinh viên nhóm các thời gian Post_Date theo giờ đăng (sáng: 6h-12h, trưa:

12h-18h, tối: 18h-24h, đêm: 0h-6h) và tính toán tổng số tương tác (Likes, Shares, Comments) cho mỗi khung giờ.

- **Yêu cầu:**

Trực quan hóa dữ liệu:

- **Biểu đồ hộp (box plot)** hoặc **biểu đồ cột (bar chart)** để trực quan hóa mối quan hệ giữa độ tuổi và số lượt thích.
- **Biểu đồ đường (line chart)** hoặc **biểu đồ cột (bar chart)** để trực quan hóa số lượng tương tác (Likes + Shares + Comments) theo các khung giờ trong ngày.