



Phân loại khách hàng tiềm năng

Môn học: Khoa học dữ liệu

Thực hiện bởi: Dương Hoàng Vũ

Lớp: 57kmt

GV hướng dẫn: TS. Nguyễn Văn Huy

Mục tiêu đề tài

Phân nhóm khách hàng từ dữ liệu giao dịch tại trung tâm thương mại.

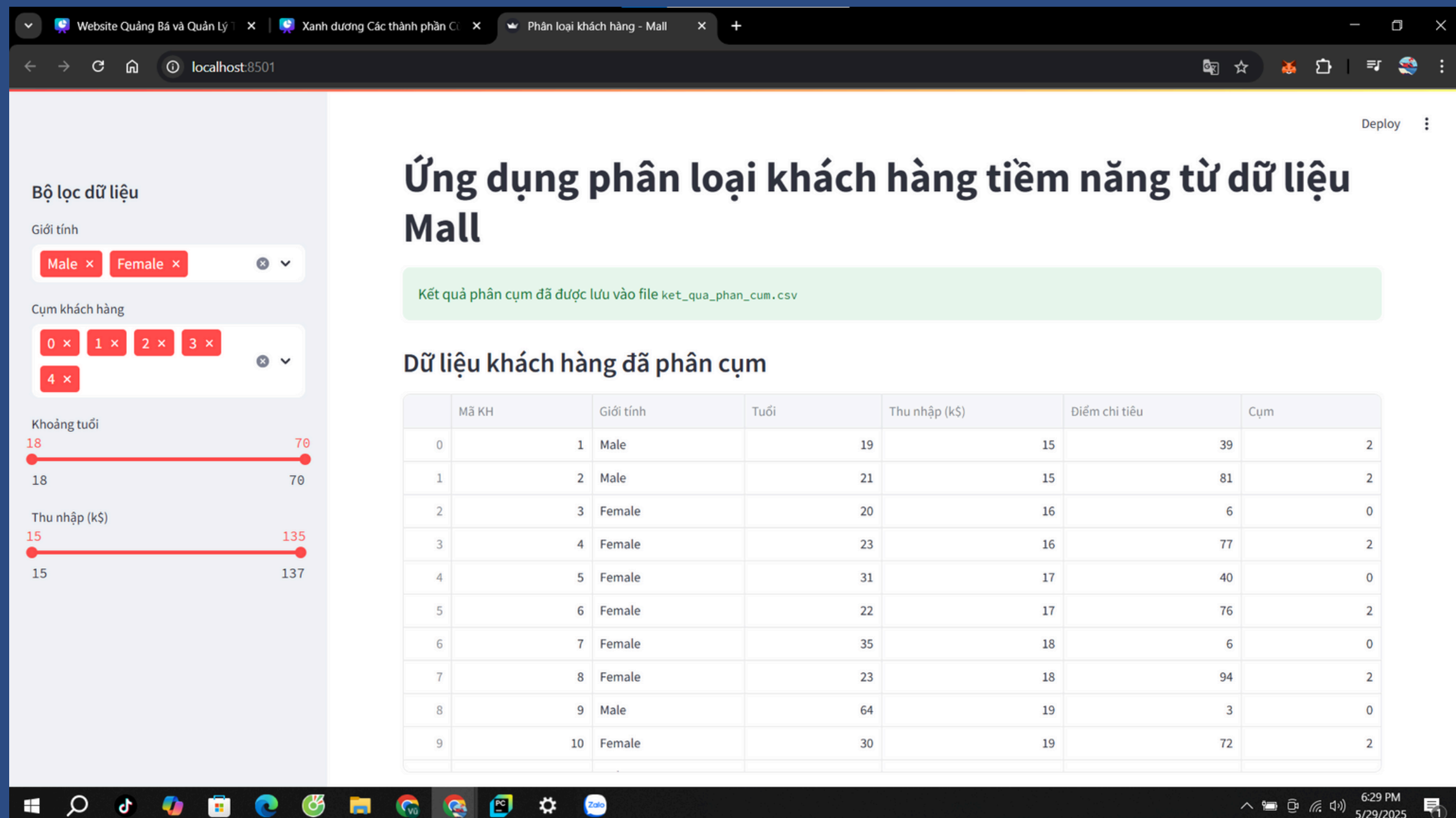
Xác định các nhóm khách hàng tiềm năng phục vụ chiến lược marketing.

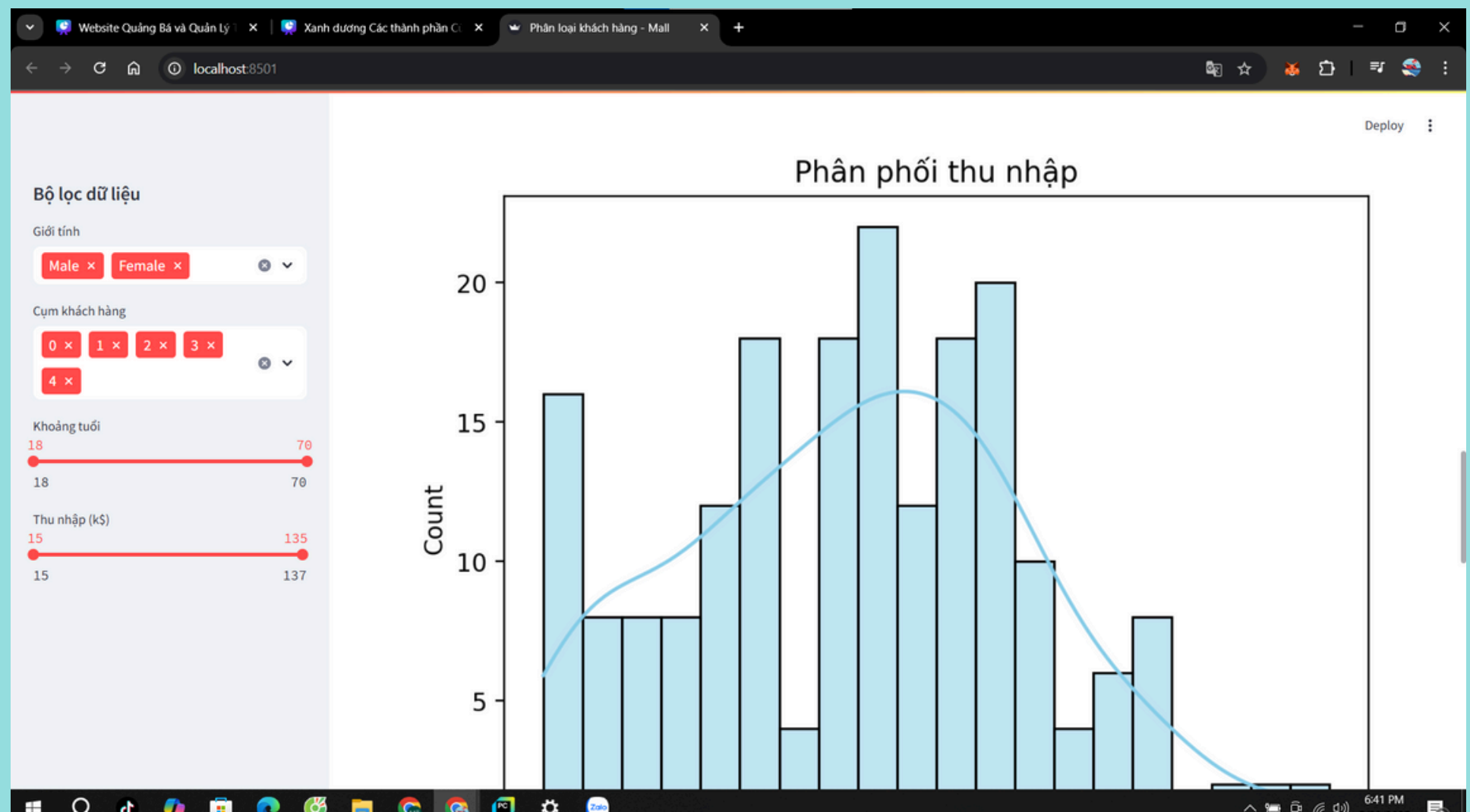
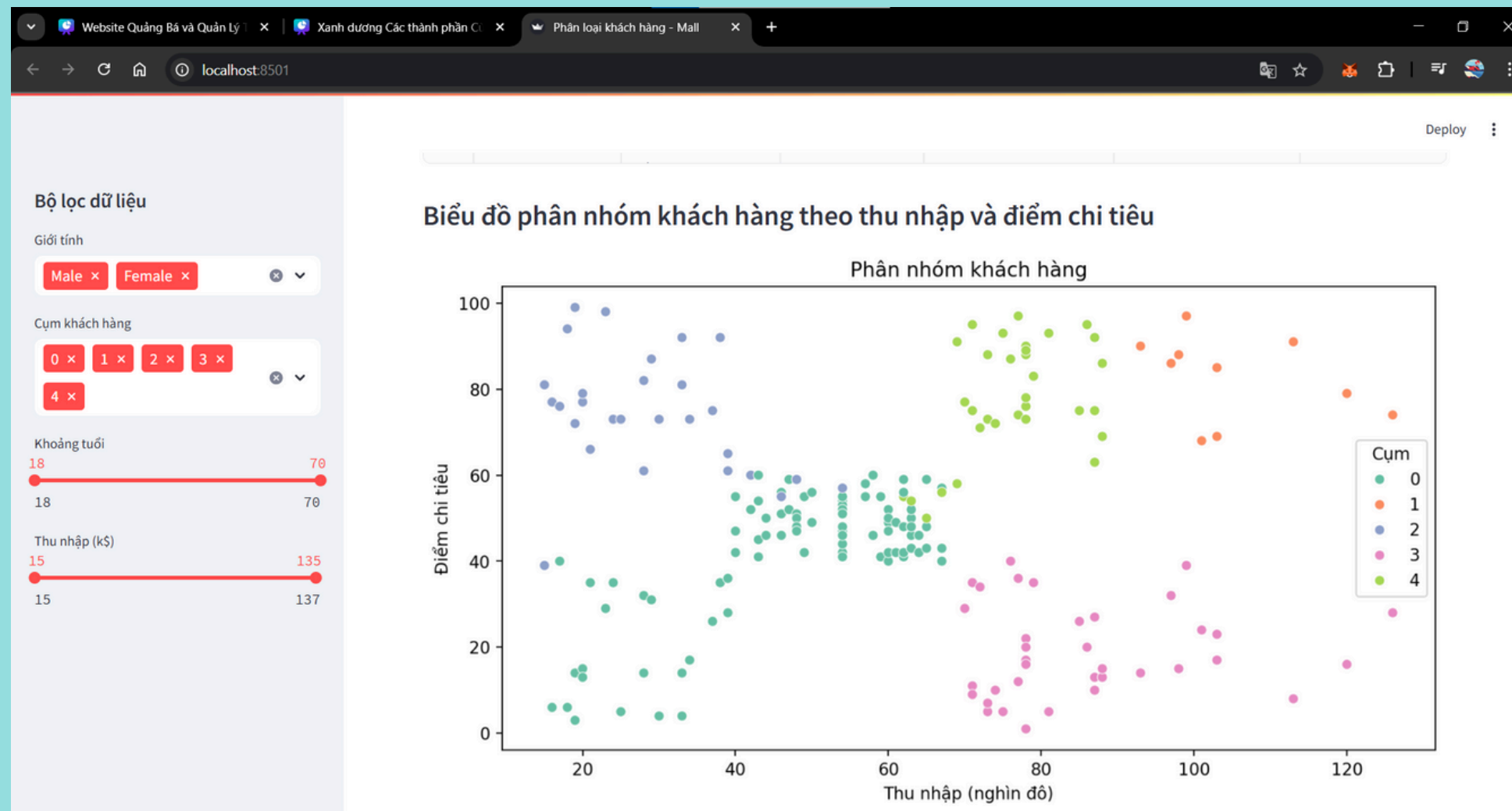
Xây dựng ứng dụng web trực quan hóa và phân tích.

- Dữ liệu sử dụng
- Nguồn dữ liệu: Mall Customer Segmentation Data

-
- Thuật toán KMeans - Tóm tắt
- Chọn số cụm k
- Khởi tạo k tâm cụm
- Gán điểm dữ liệu vào cụm gần nhất
- Cập nhật tâm cụm theo trung bình
- Lặp lại cho đến khi hội tụ
- Mục tiêu: Tối thiểu hóa khoảng cách trong cụm, giữa các cụm.

-





Giao diện ứng dụng (Streamlit)

- Nhập dữ liệu → phân cụm KMeans
- Bộ lọc: Giới tính, cụm, tuổi, thu nhập
- Hiển thị bảng dữ liệu và biểu đồ tương tác
- Cho phép tải kết quả .csv

Phương pháp sử dụng

Giải thuật phân cụm: K-Means Clustering

Số cụm: 5

Dữ liệu sử dụng cho phân cụm:

Tuổi

Thu nhập

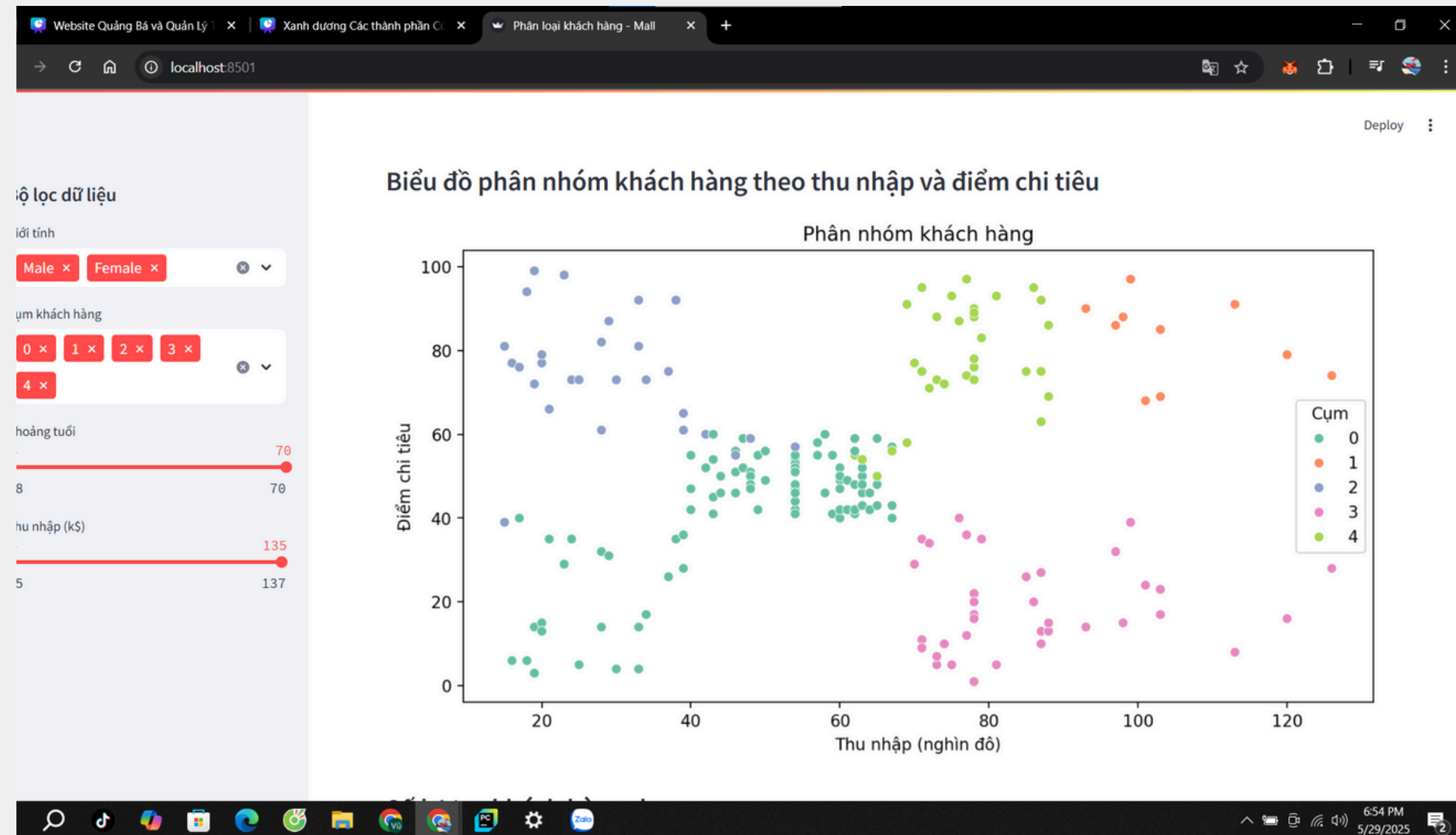
Điểm chi tiêu

Biểu đồ Scatter Plot: Phân nhóm khách hàng theo Thu nhập & Điểm chi tiêu

Ý nghĩa:

- Mô tả biểu đồ hiển thị thu nhập và điểm chi tiêu.
- Mỗi điểm là một khách hàng, mỗi màu là một nhóm phân cụm.
- Quan sát mối liên hệ giữa thu nhập và mức chi tiêu để nhận biết khách hàng tiềm năng.

- Cụm 1 (màu riêng biệt): Khách hàng thu nhập cao – chi tiêu cao → nhóm VIP.
- Cụm 3: Khách thu nhập thấp nhưng chi tiêu cao → thích mua sắm, cần chiến lược khuyến mãi.
- Cụm 0: Khách hàng thu nhập thấp – chi tiêu thấp → nhóm ít tiềm năng.
- Cụm 4: Thu nhập cao nhưng chi tiêu thấp → cần động lực thúc đẩy chi tiêu



Biểu đồ : Số lượng khách hàng theo từng cụm

Mục đích:

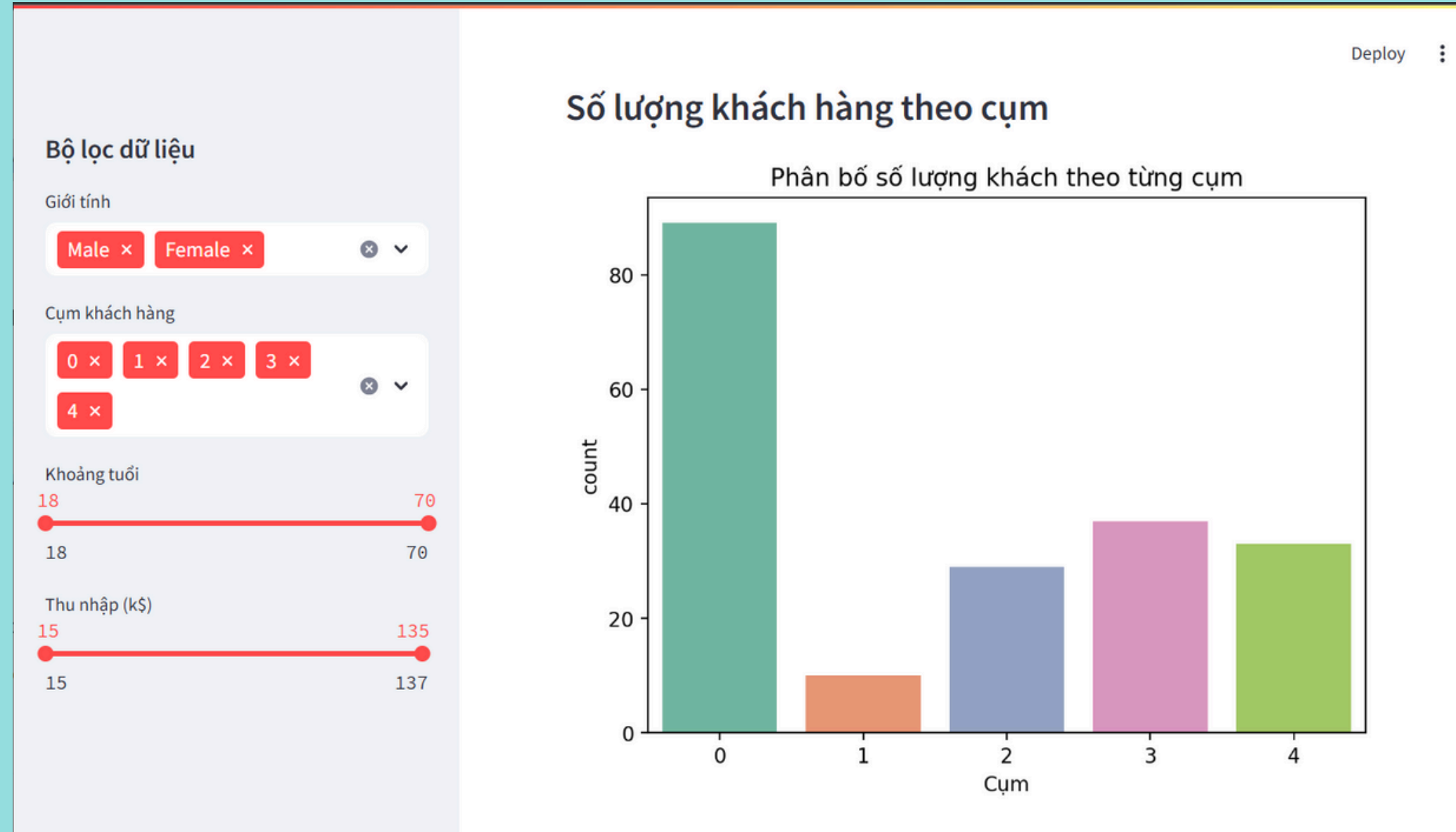
- Thể hiện số lượng khách hàng thuộc mỗi cụm.

Ý nghĩa:

- Mỗi cột biểu diễn bao nhiêu người thuộc về cụm tương ứng.
- Giúp doanh nghiệp biết cụm nào là phổ biến nhất, để tối ưu chiến lược chăm sóc hoặc quảng bá.

Nhận xét:

- Một số cụm có nhiều khách hơn → cần ưu tiên chăm sóc và giữ chân.
- Các cụm ít khách nhưng có tiềm năng cao (như cụm 1) → cần chiến lược giữ chân riêng biệt.



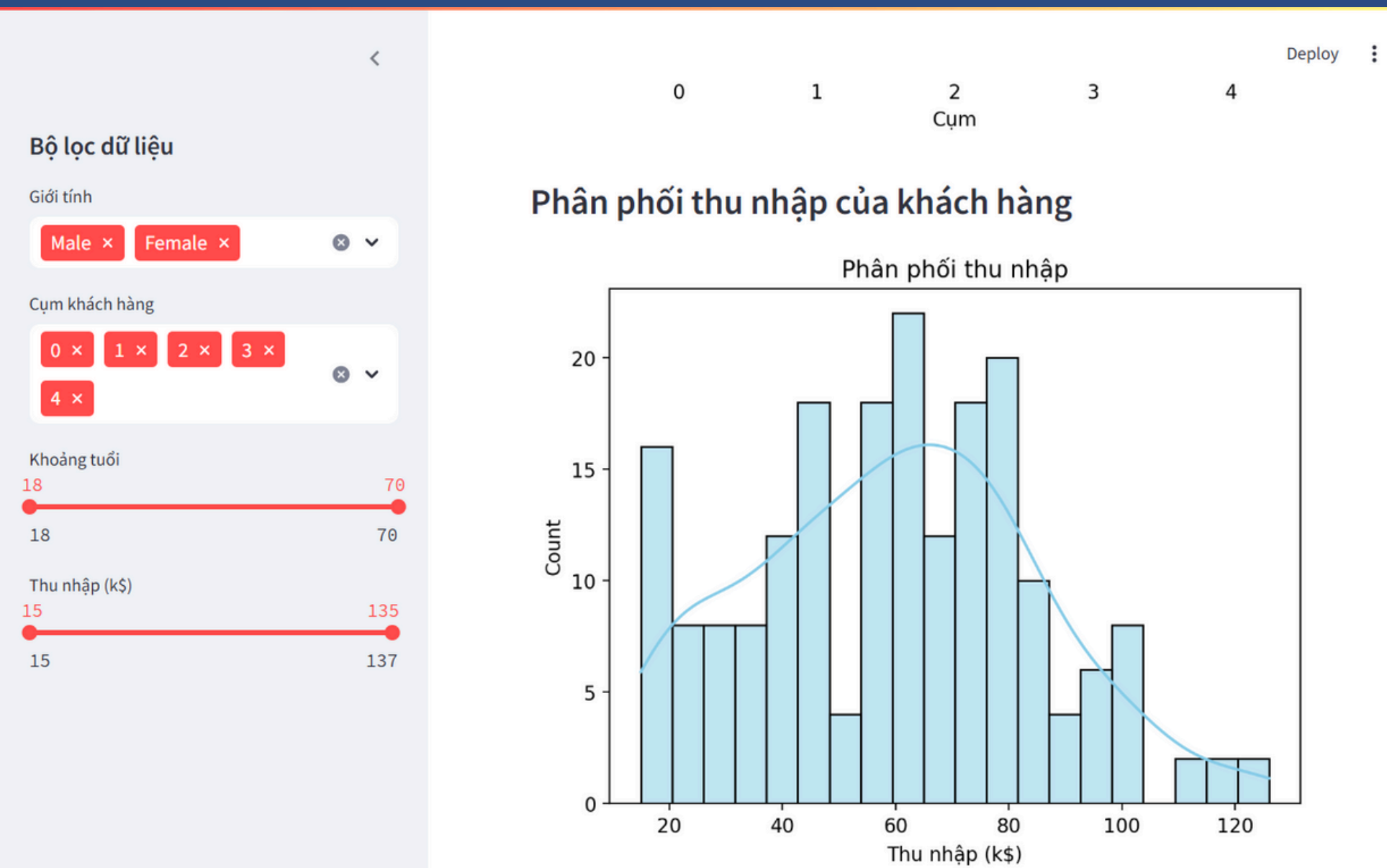
Biểu đồ Phân phối thu nhập của khách hàng

Mục đích:

- Thể hiện sự phân bố thu nhập của tất cả khách hàng.
- Ý nghĩa:
- Trục X: Mức thu nhập (nghìn đô).
- Trục Y: Số lượng khách có mức thu nhập tương ứng.
- Đường cong KDE (màu xanh): biểu diễn phân bố xác suất mượt của thu nhập.

Nhận xét:

- Thu nhập chủ yếu tập trung trong khoảng 30k – 75k.
- Có rất ít khách hàng thu nhập quá thấp hoặc quá cao.
- Giúp doanh nghiệp tập trung nhóm khách có thu nhập tầm trung trở lên.



Biểu đồ : So sánh điểm chi tiêu giữa các cụm

Nhận xét:

- Cụm 1 có mức chi tiêu trung bình cao nhất → khách VIP.
- Cụm 0 chi tiêu rất thấp → khách ít tiềm năng.
- Cụm 3 chi tiêu cao nhưng biến động lớn → cần chiến lược riêng biệt.

Mục đích:

- So sánh mức chi tiêu trung bình
- và độ phân tán của từng cụm.

Ý nghĩa:

Mỗi "hộp" đại diện cho phân
phối điểm chi tiêu của từng cụm:

- Đường giữa hộp: Giá trị trung vị.
- Đầu hộp: Phân vị 25% và 75%.
- Chấm tròn ngoài: Ngoại lệ (outliers).

Bộ lọc dữ liệu

Giới tính

Male × Female × × v

Cụm khách hàng

0 × 1 × 2 × 3 × × v

4 ×

Khoảng tuổi

18 70

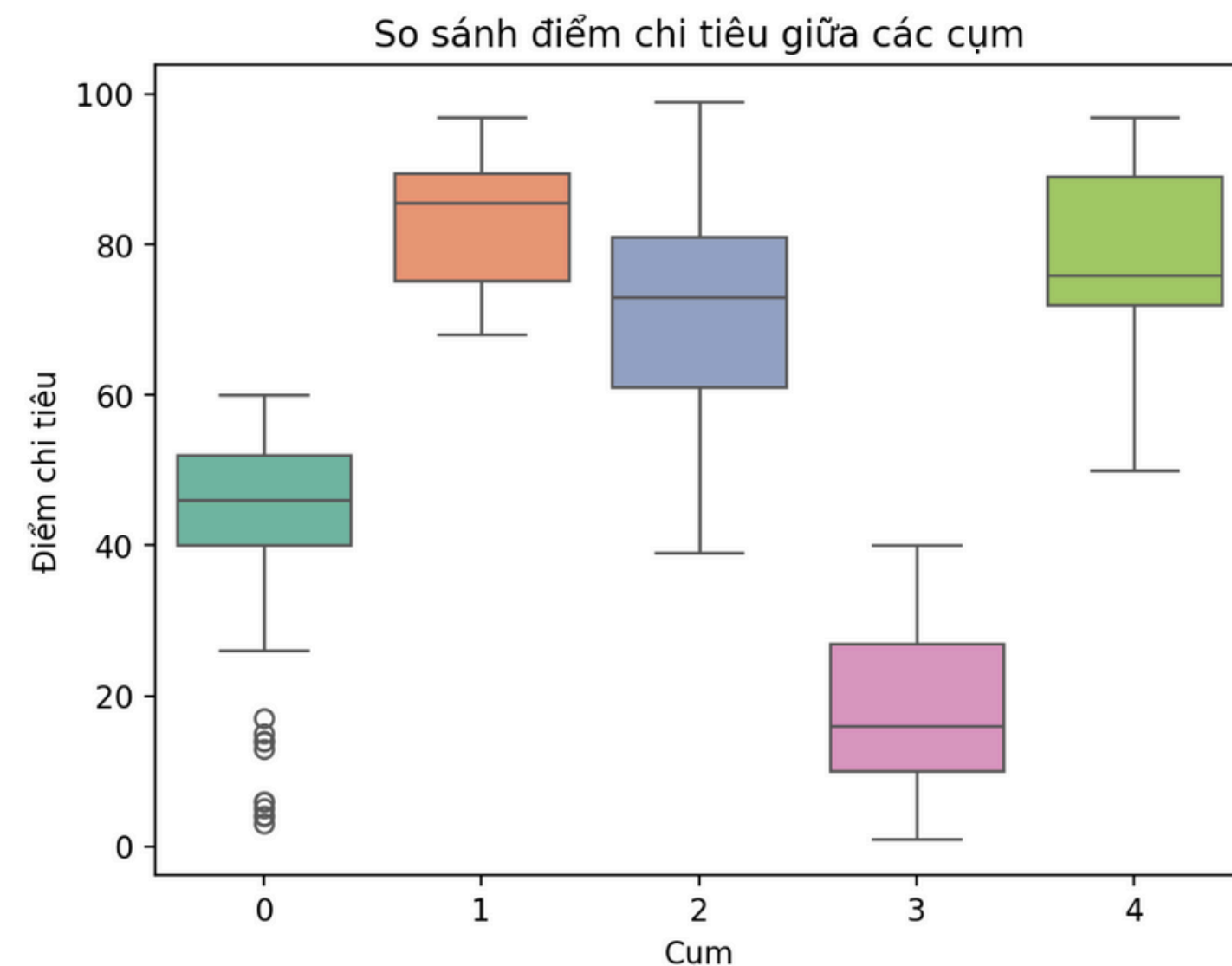
18 70

Thu nhập (k\$)

15 135

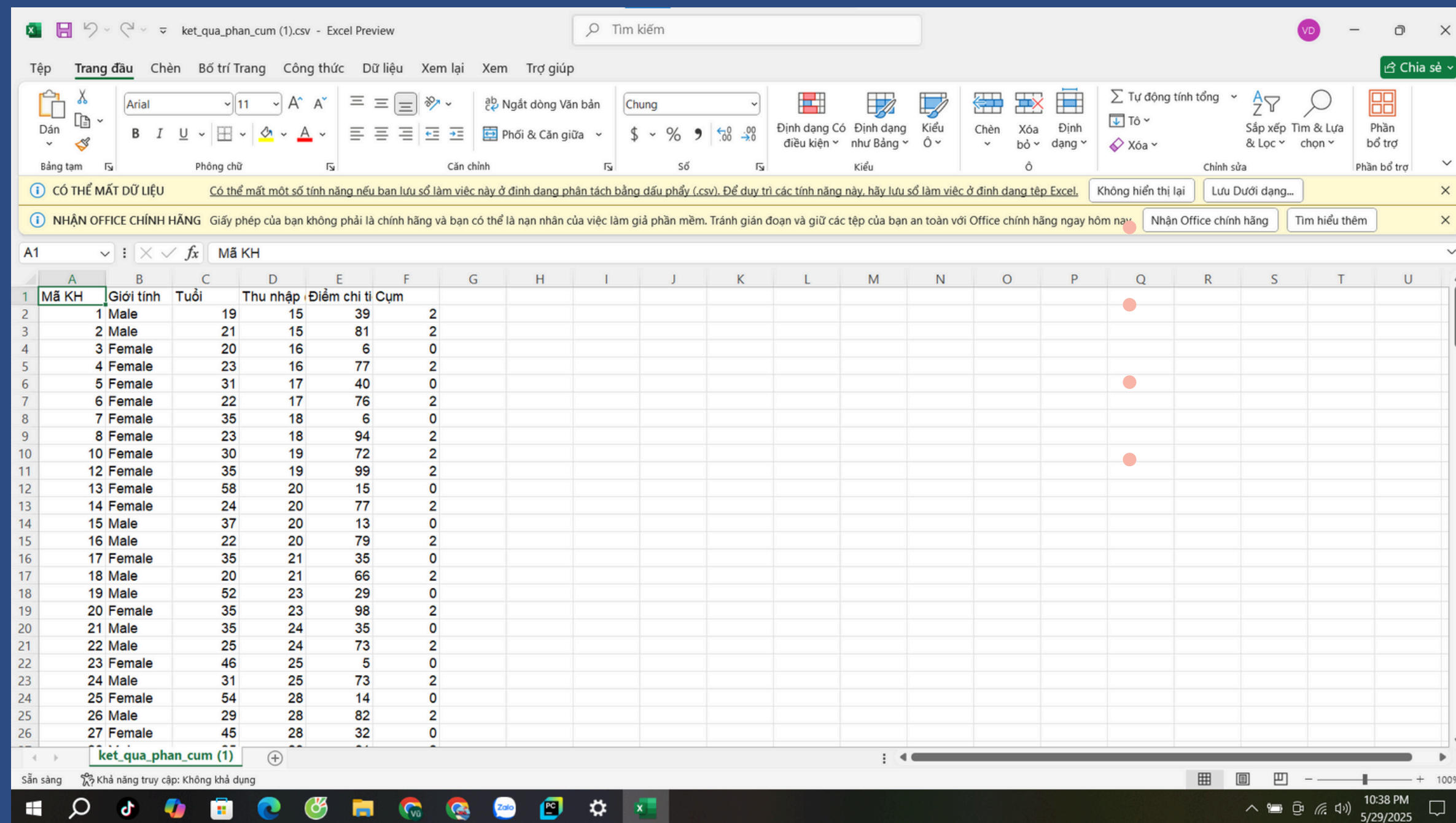
15 137

So sánh điểm chi tiêu giữa các cụm



Kết quả chạy thực nghiệm

- Đã phân cụm 200 khách hàng thành 5 nhóm
- Hệ thống chạy tốt, nhanh, lọc và hiển thị tương tác
- Cho phép người dùng phân tích sâu từng cụm



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Mã KH	Giới tính	Tuổi	Thu nhập	Điểm chi ti	Cụm															
2	1	Male	19	15	39	2															
3	2	Male	21	15	81	2															
4	3	Female	20	16	6	0															
5	4	Female	23	16	77	2															
6	5	Female	31	17	40	0															
7	6	Female	22	17	76	2															
8	7	Female	35	18	6	0															
9	8	Female	23	18	94	2															
10	10	Female	30	19	72	2															
11	12	Female	35	19	99	2															
12	13	Female	58	20	15	0															
13	14	Female	24	20	77	2															
14	15	Male	37	20	13	0															
15	16	Male	22	20	79	2															
16	17	Female	35	21	35	0															
17	18	Male	20	21	66	2															
18	19	Male	52	23	29	0															
19	20	Female	35	23	98	2															
20	21	Male	35	24	35	0															
21	22	Male	25	24	73	2															
22	23	Female	46	25	5	0															
23	24	Male	31	25	73	2															
24	25	Female	54	28	14	0															
25	26	Male	29	28	82	2															
26	27	Female	45	28	32	0															

Tính năng mạnh mẽ:
Giao diện đẹp, dễ dùng
Lưu kết quả dễ dàng
Phù hợp cho các cửa hàng/quản lý khách hàng

Kết luận

- Những gì đã thực hiện:
- Đã xây dựng ứng dụng web bằng Python + Streamlit để phân loại khách hàng từ dữ liệu Mall Customers.
- Áp dụng thuật toán KMeans để chia khách hàng thành các nhóm theo hành vi tiêu dùng (thu nhập, độ tuổi, điểm chi tiêu).
- Tạo ra biểu đồ phân tích trực quan giúp người dùng dễ dàng hiểu được đặc điểm từng nhóm.

Đã phân cụm khách hàng thành 5 nhóm chính, trong đó xác định rõ các nhóm như:

- Nhóm chi tiêu cao - thu nhập cao (khách VIP)
- Nhóm chi tiêu thấp - thu nhập thấp (ít tiềm năng)
- Nhóm chi tiêu cao - thu nhập trung bình

