

# Phân loại tính cách từ dữ liệu hành vi



# Data

Nguồn: [Click](#)

About Data: Tập dữ liệu này được xây dựng nhằm phân loại tính cách người dùng thành hai nhóm chính:

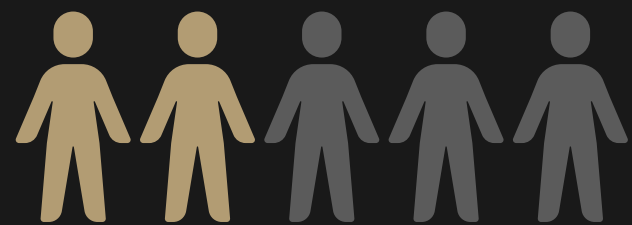
- Extrovert (Hướng ngoại)
- Introvert (Hướng nội)

Thông qua hành vi và xu hướng cá nhân như: mức độ tham gia xã hội, thói quen giao tiếp, quy mô bạn bè, tần suất đăng bài...



# Quy Mô Dữ Liệu

03



## Thông tin về dữ liệu

Dữ liệu gồm 5000 bản ghi và 8 cột

## Variable Description

	Tên Biến	Mô tả	Kiểu Dữ Liệu
1	Time_spent_Alone	Số giờ trung bình một người dành cho bản thân mỗi ngày	Float
2	Stage_fear	Người đó có sợ khi nói trước đám đông hoặc trình bày không	Object
3	Social_event_attendance	Người đó có sợ khi nói trước đám đông hoặc trình bày không	Float
4	Going_out	Tần suất đi ra ngoài	Float
5	Drained_after_socializing	Người đó có cảm thấy mệt mỏi, mất năng lượng sau khi giao tiếp xã hội hay không	Object
6	Friends_circle_size	Số lượng vòng bạn bè	Float
7	Post_frequency	Tần suất đăng bài lên mạng xã hội	Float
8	Personality	Nhân mục tiêu: phân loại người thành Extrovert(hướng ngoại), Introvert(hướng nội)	Object

03

# Khám Phá Dữ Liệu

## 1/Độ phân bố dữ liệu của categorical

**a/Stage\_fear with Personality**

Stage_fear	No: 2759
	Yes: 2241
Drained_after_socializing	No:2247
	Yes: 2753
Personality	Introvert 2502
	Extrovert 2498

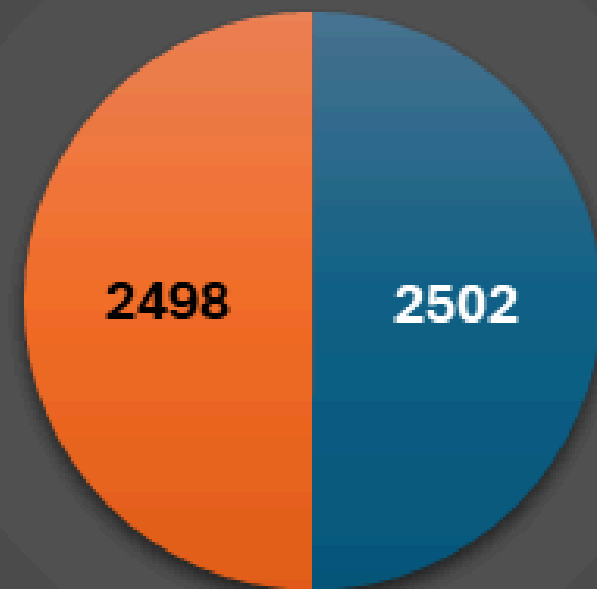


# Khám Phá Dữ Liệu

## 1/Độ phân bố dữ liệu của categorical

### a/ Personality

#### Personality



■ Extrovert ■ Introvert

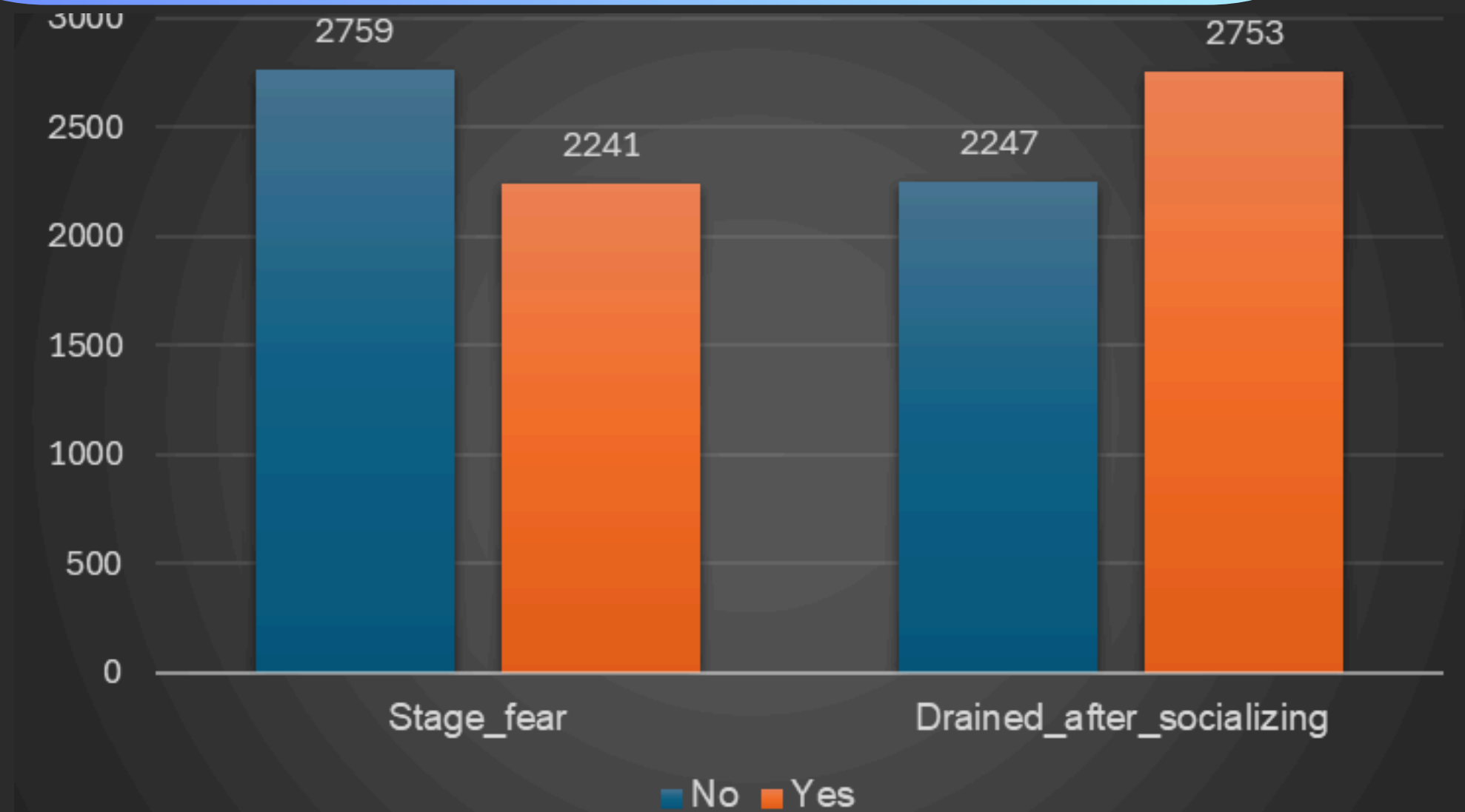
#### Biến mục tiêu

- Phân bố giữa hai lớp rất cân bằng, giúp mô hình học máy không bị lệch
- Không cần áp dụng kỹ thuật cân bằng lại dữ liệu (resampling)
- 2 lớp cân bằng ~50% cho Introvert ~50% cho Extrovert

# Khám Phá Dữ Liệu

## 1/Độ phân bố dữ liệu của categorical

### b/Stage\_fear with Drained\_after\_socializing



### Độ phân bố dữ liệu cho thấy

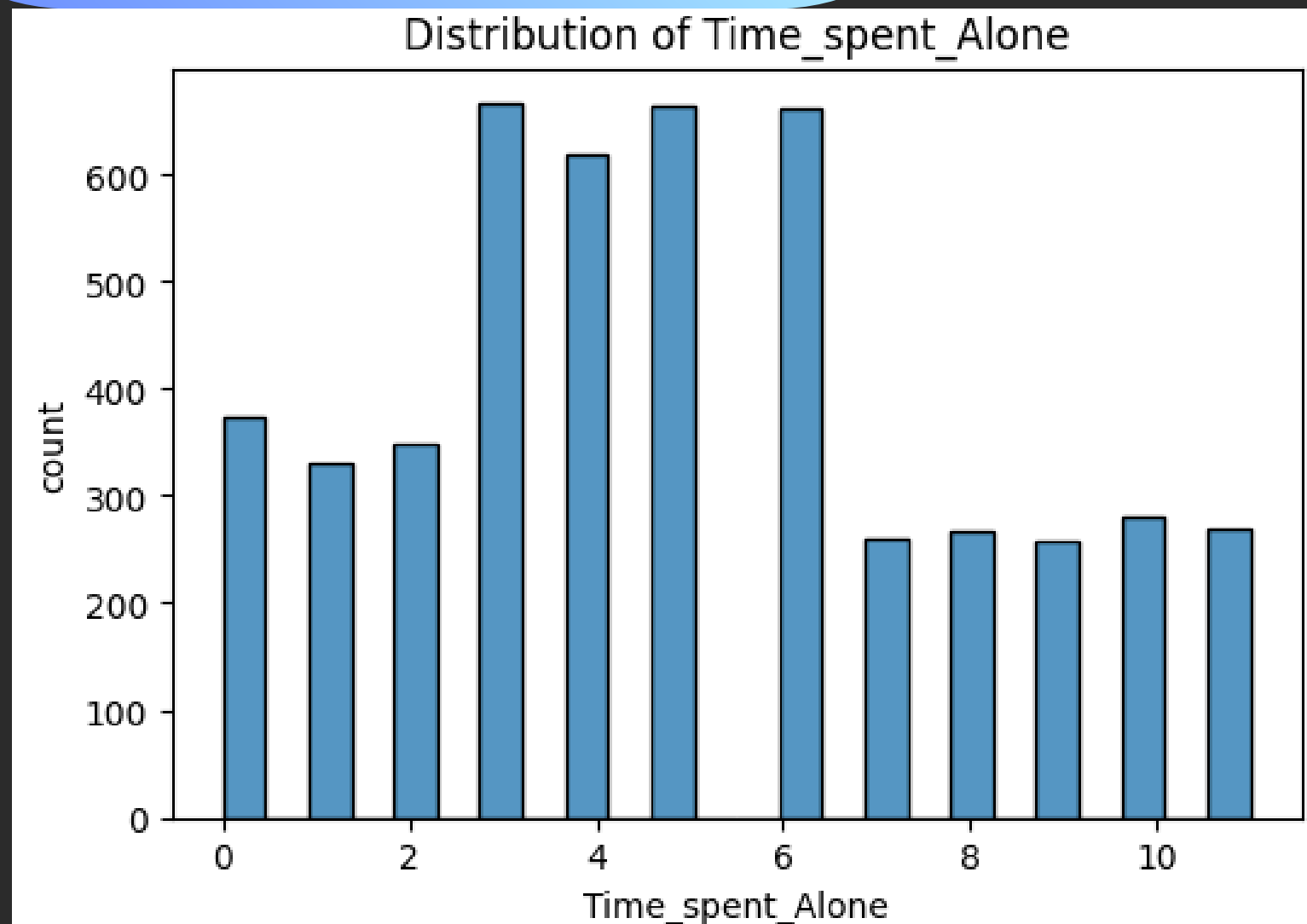
- Dữ liệu lệch nhẹ, chưa đến mức gây mất cân bằng nghiêm trọng, chưa cần xử lý oversampling/undersamp
- Stage\_fear có độ lệch trái nhẹ: Tỷ lệ: No ~ 55.2%, Yes ~ 44.8%
- Drained\_after\_socializing có độ lệch phải nhẹ: Tỷ lệ No ~ 44.9%, Yes ~ 55.1%



# Khám Phá Dữ Liệu

## 2/Độ phân bố dữ liệu của numerical

### a/ Time\_spent\_Alone



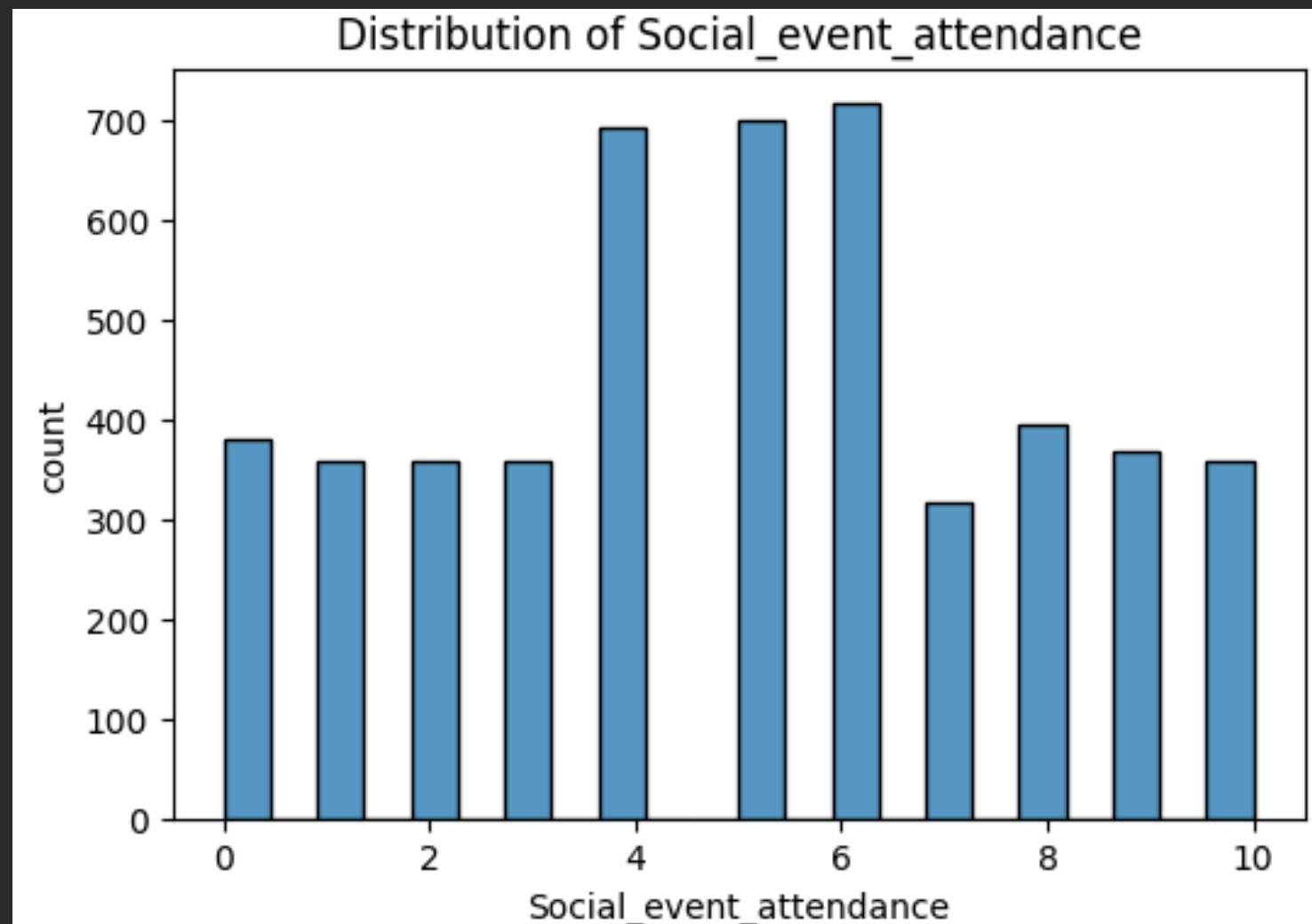
Biểu đồ hist cho thấy

- dữ liệu phân bố khá đồng đều hơi lệch trái.
- Đa số mọi người dành thời gian cho bản thân từ 3 đến 6 giờ

# Khám Phá Dữ Liệu

## 2/Độ phân bố dữ liệu của numerical

### b/Social\_event\_attendance



Biểu đồ hist cho thấy

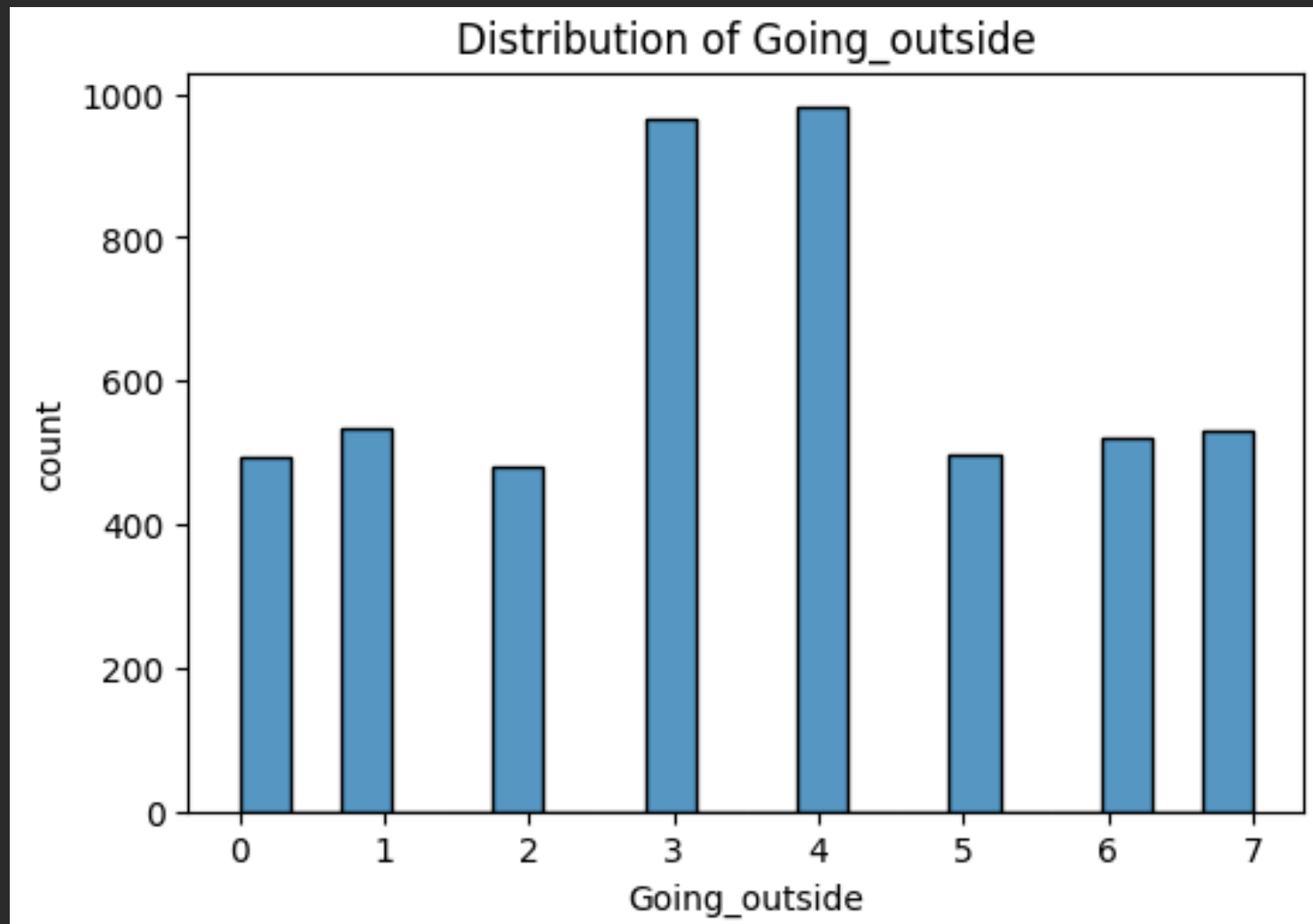
- Phân phối khá đối xứng, trung bình gần đúng giữa
- Dữ liệu tập trung nhiều quanh các mức 4–6
- Có vẻ là phân phối gần chuẩn nhẹ, rất cân bằng



# Khám Phá Dữ Liệu

## 2/Độ phân bố dữ liệu của numerical

**c/Going\_outside**



Biểu đồ hist cho thấy:

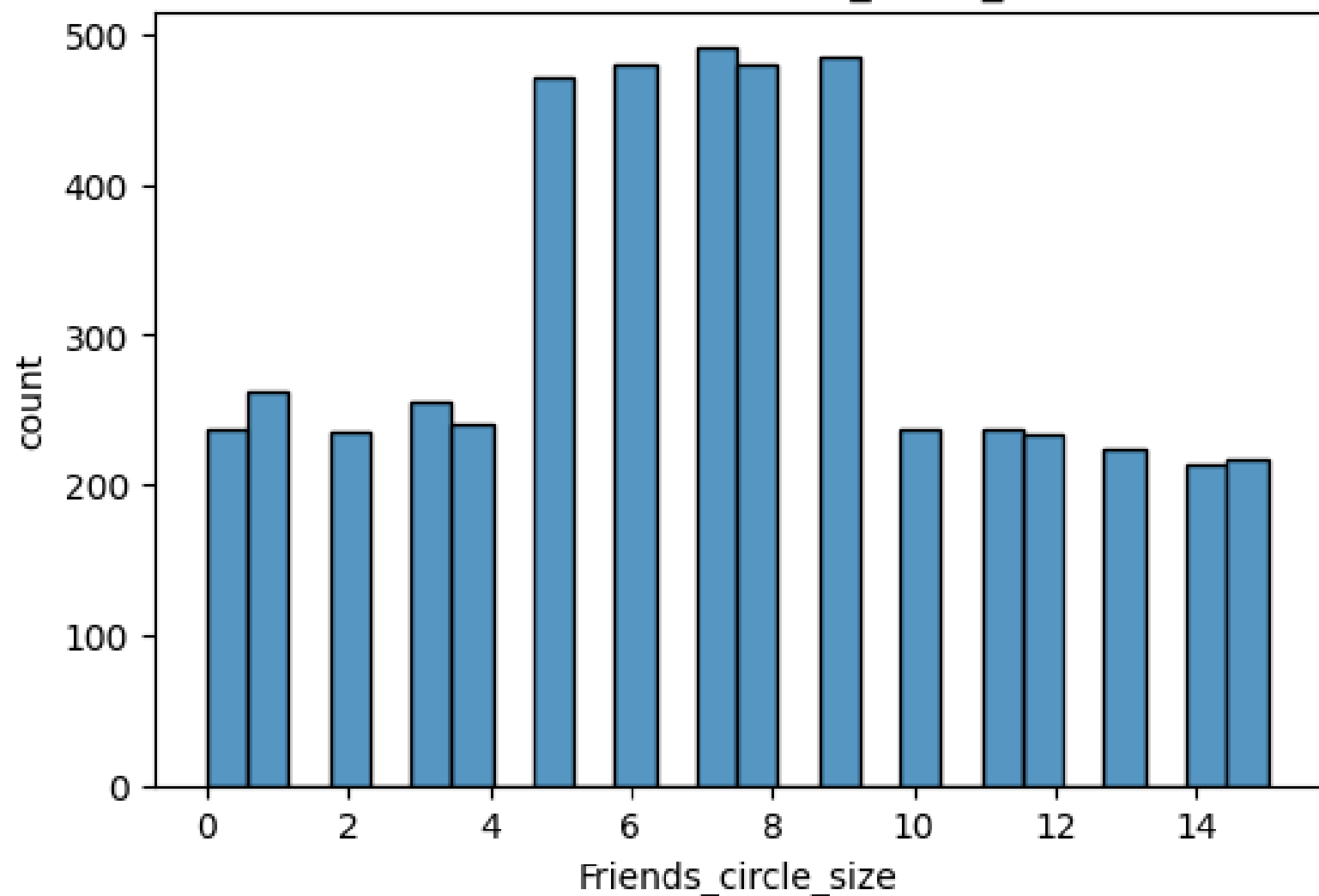
- Phản ánh tần suất hoặc mức độ đi ra ngoài của cá nhân
- Trung bình chỉ ở mức 3-4
- Dữ liệu hơi lệch trái → nhóm ít ra ngoài phổ biến hơn

# Khám Phá Dữ Liệu

## 2/Độ phân bố dữ liệu của numerical

### d/Friends\_circle\_size

Distribution of Friends\_circle\_size



Biểu đồ hist cho thấy:

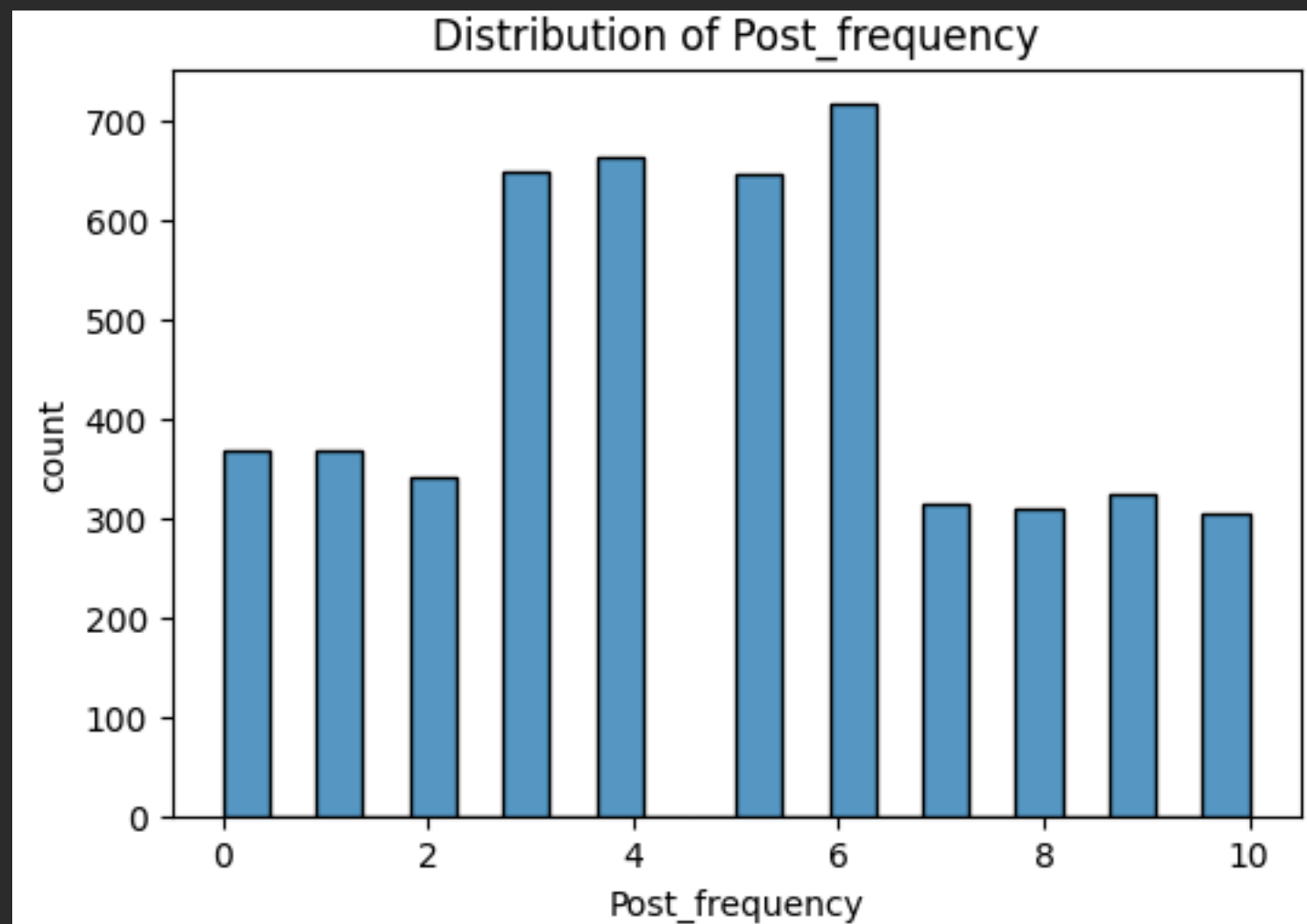
- Biến Friends\_circle\_size có phân phối rộng nhưng cân bằng
- Trung bình người tham gia khảo sát có khoảng 7 người bạn
- Dữ liệu này phù hợp cho:
- Phân tích mối quan hệ với tính cách (Personality)
- Tương quan với các yếu tố xã hội khác như: đi ra ngoài, dự sự kiện,...



# Khám Phá Dữ Liệu

## 2/Độ phân bố dữ liệu của numerical

### d/Post\_frequency

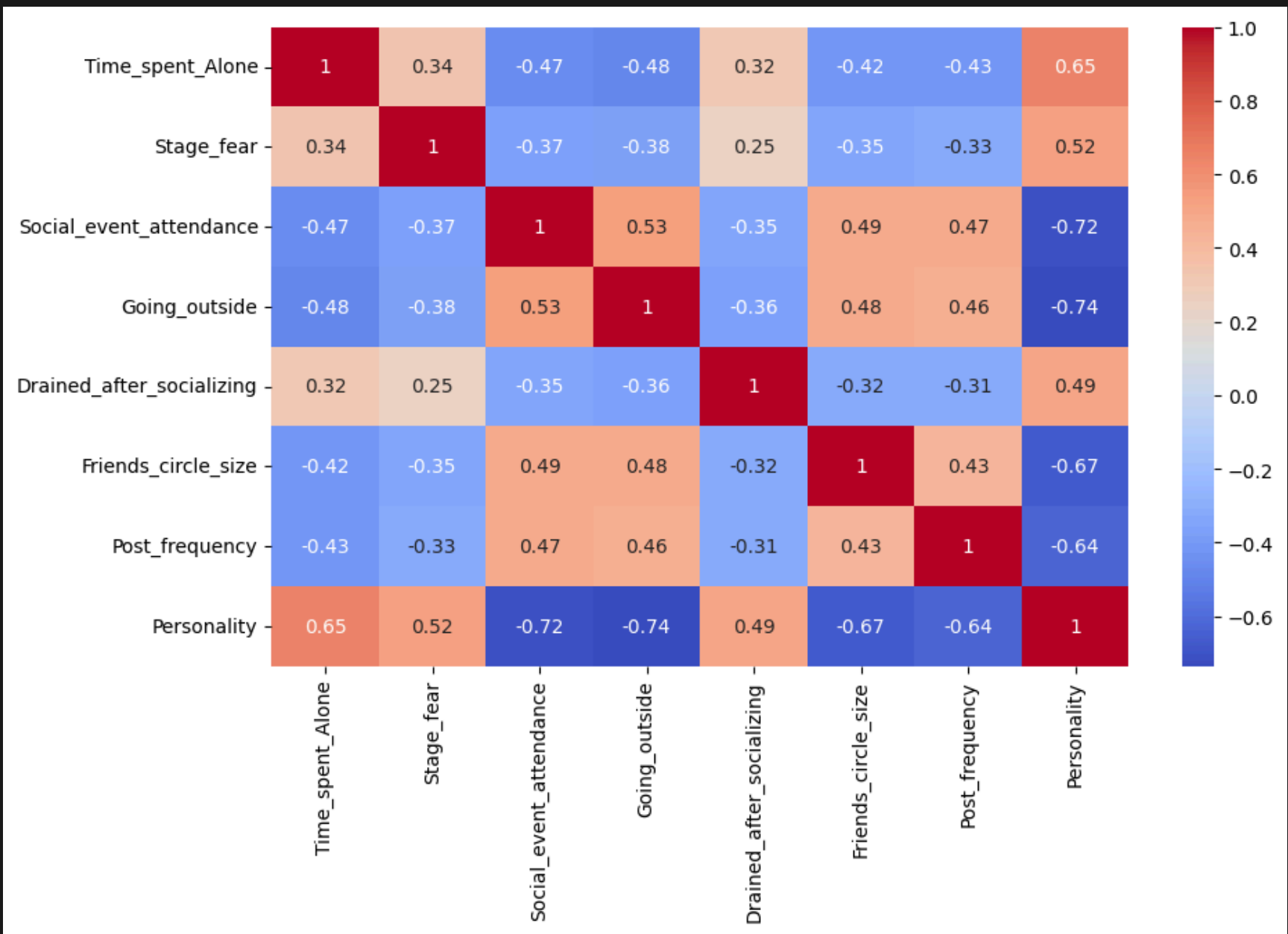


Biểu đồ hist cho thấy:

- Post\_frequency phản ánh tần suất đăng bài, trung bình khoảng 5 lần
- Phân phối cân đối, không thiên lệch rõ
- Thích hợp để:
- So sánh giữa các nhóm tính cách (Extrovert vs Introvert)
- Kiểm tra tương quan với các hành vi xã hội khác



# Tương Quan Dữ Liệu



## Đánh giá tổng quan:

Going\_outside (-0.74): Người càng thường ra ngoài càng có xu hướng hướng ngoại

Social\_event\_attendance (-0.72): Người càng dự nhiều sự kiện càng hướng ngoại

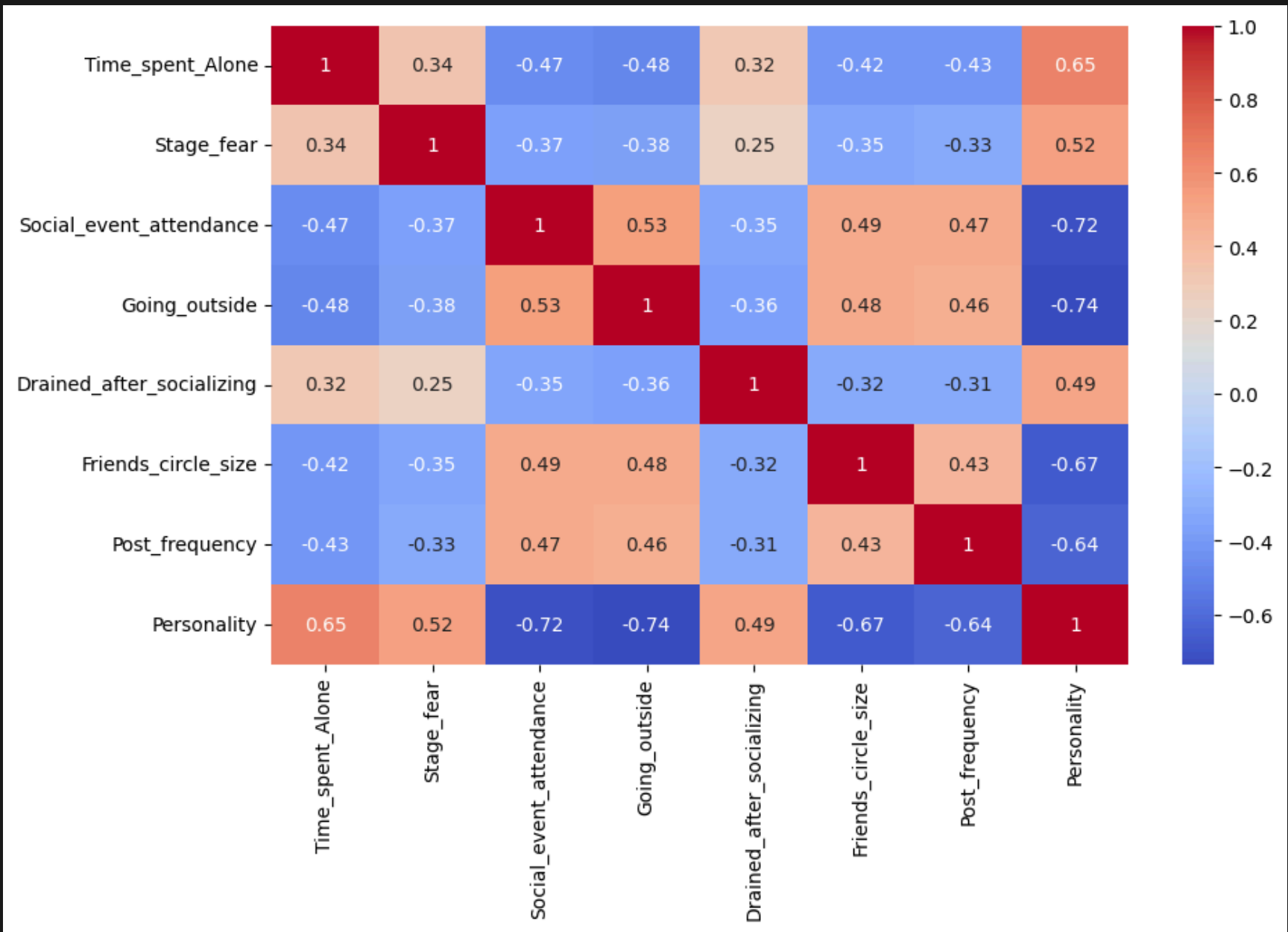
Friends\_circle\_size (-0.67): Số bạn bè nhiều hơn → hướng ngoại

Post\_frequency (-0.64): Đăng bài nhiều hơn → hướng ngoại

Stage\_fear (0.52): Sợ sân khấu cao hơn → xu hướng hướng nội

Time\_spent\_Alone (0.65): Ở một mình nhiều hơn → hướng nội





## Kết luận:

- Biểu đồ tương quan cho thấy mối quan hệ tuyến tính rõ rệt giữa các hành vi xã hội và đặc điểm tính cách của cá nhân. Cụ thể, biến mục tiêu Personality thể hiện tương quan âm mạnh với các biến hành vi xã hội như Going\_outside ( $r = -0.74$ ), Social\_event\_attendance ( $r = -0.72$ ), Friends\_circle\_size ( $r = -0.67$ ), và Post\_frequency ( $r = -0.64$ ). Điều này cho thấy những cá nhân có xu hướng tham gia hoạt động xã hội nhiều hơn thường nghiêng về đặc điểm hướng ngoại.
- Ngược lại, Personality có tương quan dương đáng kể với Time\_spent\_Alone ( $r = 0.65$ ) và Stage\_fear ( $r = 0.52$ ), gợi ý rằng các đặc điểm như thích ở một mình và lo lắng khi xuất hiện trước đám đông có liên hệ chặt chẽ với tính cách hướng nội.

-Ngoài ra, các biến hành vi xã hội cũng cho thấy sự đồng phát với nhau, thể hiện qua các hệ số tương quan dương từ mức trung bình đến mạnh (0.46–0.53). Điều này củng cố giả thuyết rằng các hành vi giao tiếp xã hội có xu hướng đồng thời xuất hiện ở những cá nhân hướng ngoại



# Chọn Mô Hình

Sử dụng LazyClassifier() lựa chọn mô hình cho Supervised.

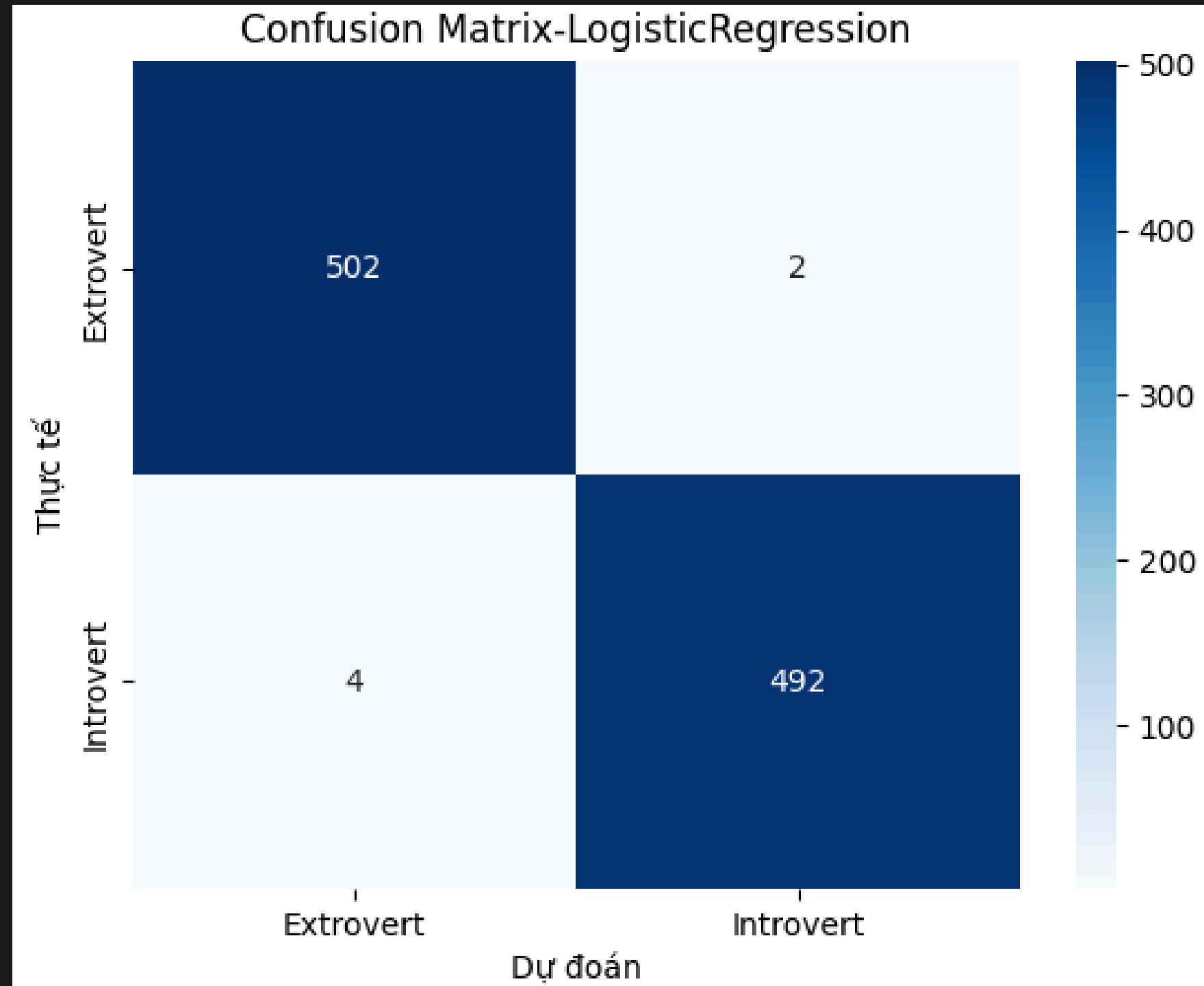
```
“from lazypredict.Supervised import LazyClassifier  
clt=LazyClassifier()  
pred_model=clt.fit(x_train,x_test,y_train,y_test)  
print(pred_model[0].sort_values(by='Accuracy', ascending=False).head(5))”
```

Sao đó lựa chọn mô hình LogisticRegression()

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score	Time Taken
LogisticRegression	0.994	0.993984	0.993984	0.994	0.0334337



# Confusion Matrix



Confusion Matrix của Logistic Regression cho thấy mô hình đạt độ chính xác 99.4% trên tập kiểm tra. Trong đó:

- 502/504 người Extrovert được dự đoán đúng

- 492/496 người Introvert được dự đoán đúng

Mô hình cho thấy khả năng phân loại cực kỳ tốt và ổn định, với tỷ lệ sai thấp, không thiên lệch. Điều này khẳng định rằng hành vi xã hội có thể dự đoán tính cách một cách đáng tin cậy bằng Logistic Regression



# Kết Luận

## Mối liên hệ giữa hành vi và tính cách

- Time\_spent\_Alone, Stage\_fear: càng dành thời gian ở một mình nhiều và sợ sân khấu sẽ có tính cách hướng nội
- Social\_event\_attendance, Going\_outside, Post\_frequency: thường xuyên đi tham gia các sự kiện xã hội, ra ngoài thường xuyên, hay post những bài trên các mạng xã hội sẽ có tính cách hướng ngoại

## Ý NGHĨA

- Dự vào hành vi xã hội có thể dự đoán tính cách của một cá nhân
- Có thể ứng dụng mô hình này trong:
  - +Tư vấn tâm lý đưa ra liệu trình phù hợp
  - +Dự đoán tính cách ứng viên cho vị trí làm việc (HR)
  - +Thăm dò gợi ý những sản phẩm phù hợp cho tính cách của mỗi khách hàng





# Mã Nguồn

**Link Github**

<https://github.com/Jven-DS/Project1>

**THANKS FOR LISTENING**