tripleten

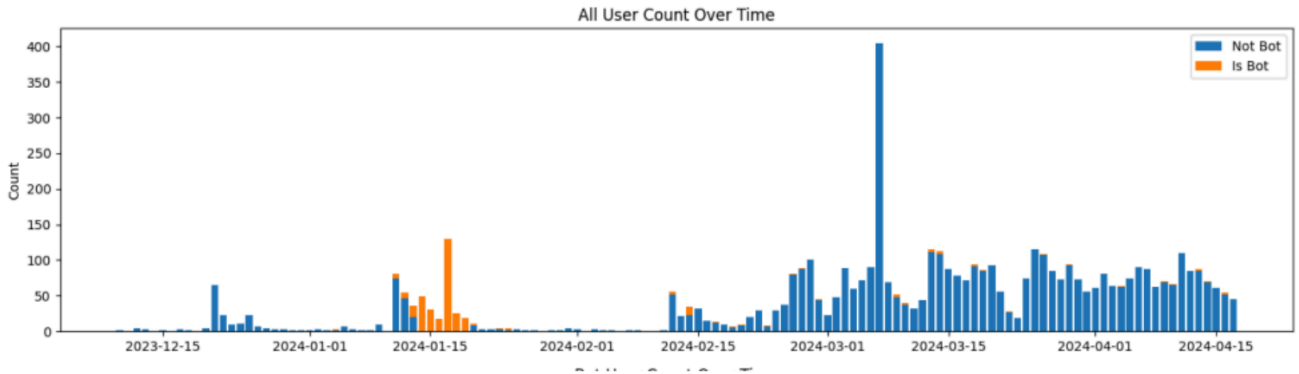
Machine learning to stop cyberbots

Partner: Besample, a platform enabling researchers to collect behavioral data across underrepresented world regions.

Website: https://besample.app/

Project description: Besample's platform was experiencing a cyberbot attack. Students analyzed user data to determine bot behavior and activity. They also built machine learning models to predict possible bot activity in user data.

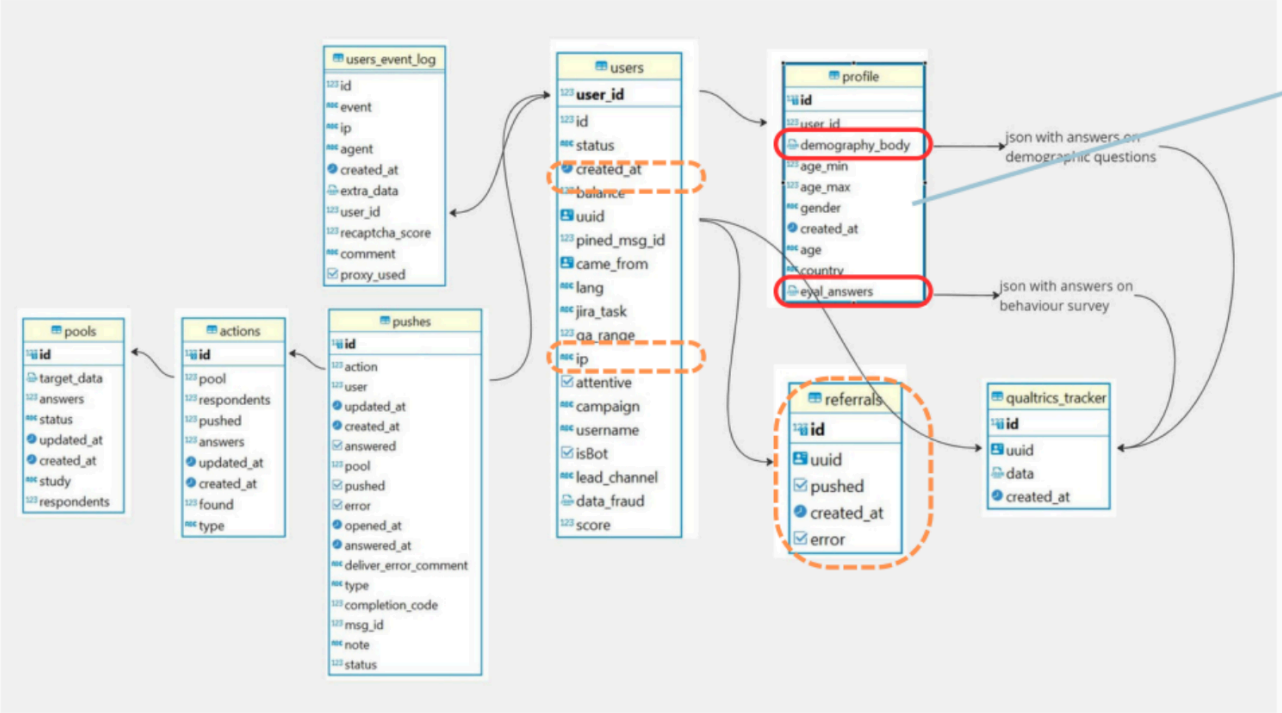
The Origin and Methodology of the Project



Challenges

- Realistic Bot users
- “isBot” limitation

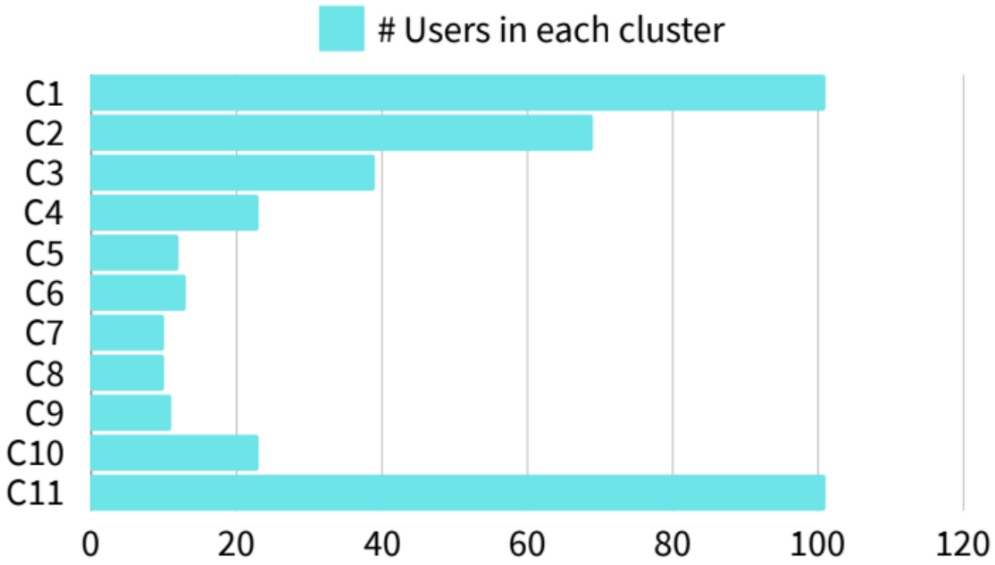
Determine the dataset for researching

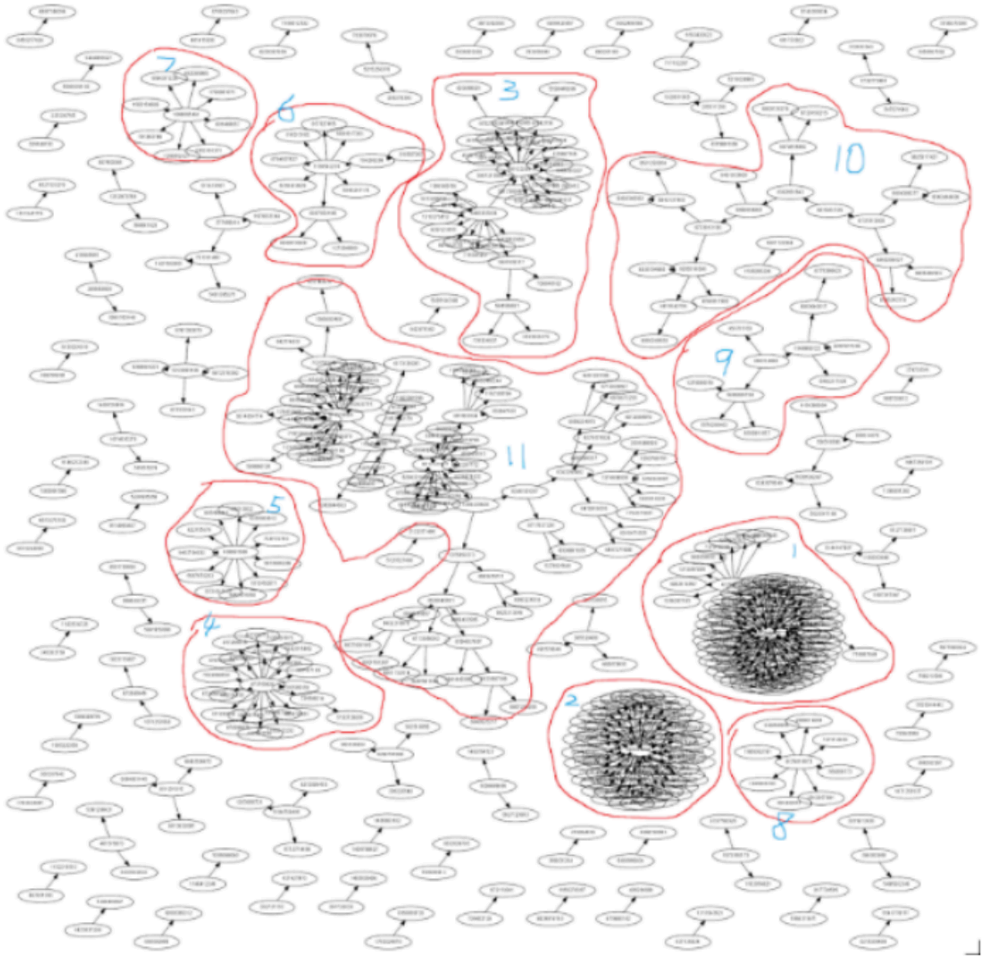


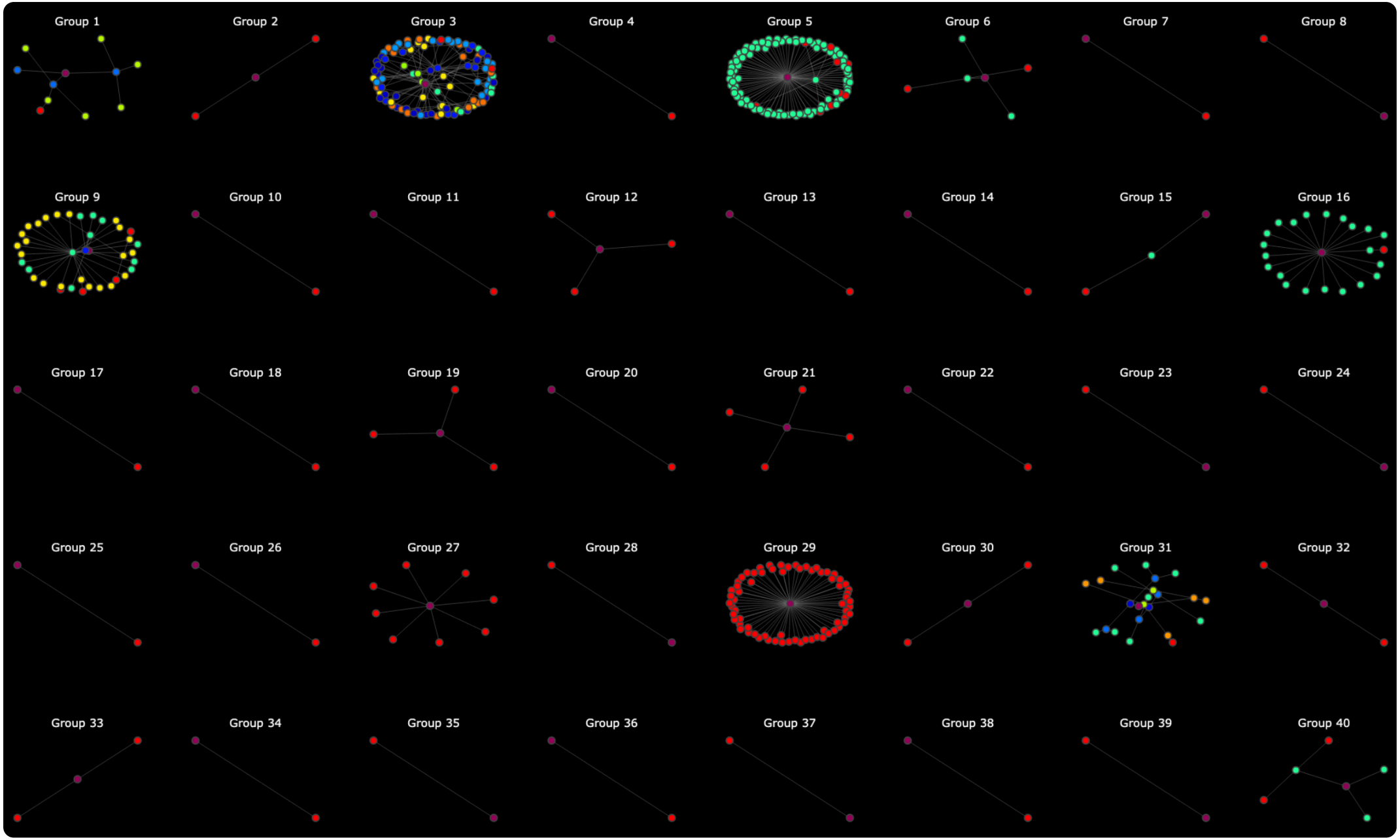
	Demography_body	eyal_answers
Non-Null Values	4184 (82.5%)	1998
open-end questions	QID11, 14, 18...	QID81, 154...
Chosen Question	QID11	QID154

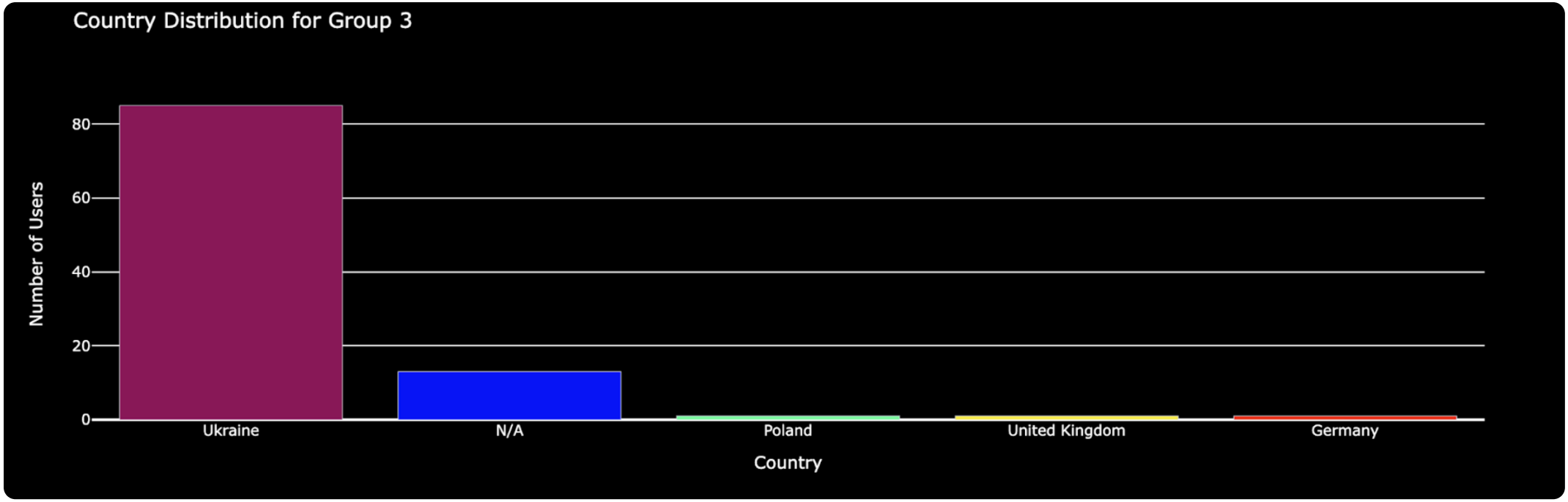
- Enough data
- Open-ended questions

Explore the user answers data: Referral Ancestors





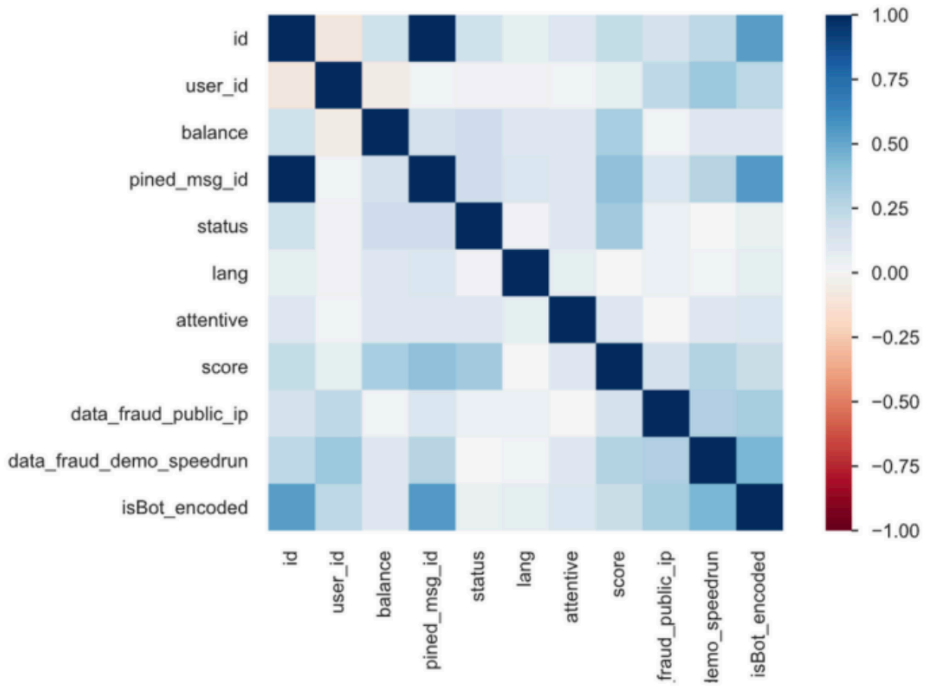

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Group 9
Group 10
Group 11
Group 12
Group 13
Group 14
Group 15
Group 16
Group 17
Group 18
Group 19
Group 20
Group 21
Group 22
Group 23
Group 24
Group 25
Group 26
Group 27
Group 28
Group 29
Group 30
Group 31
Group 32
Group 33
Group 34
Group 35
Group 36
Group 37
Group 38
Group 39
Group 40

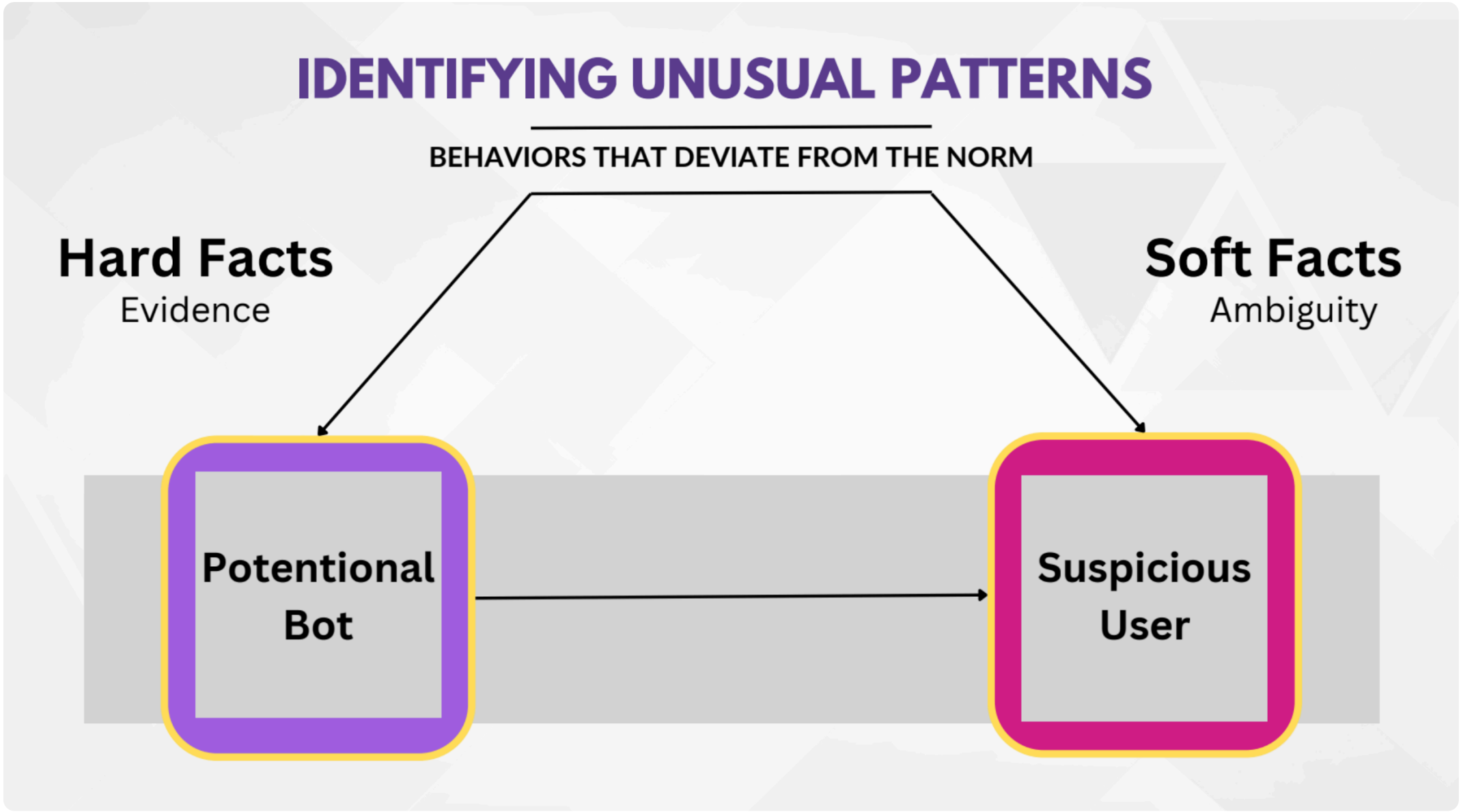

Country Distribution for Group 3
Number of Users
0
20
40
60
80
Ukraine
N/A
Poland
United Kingdom
Germany
Country

CORRELATION ANALYSIS

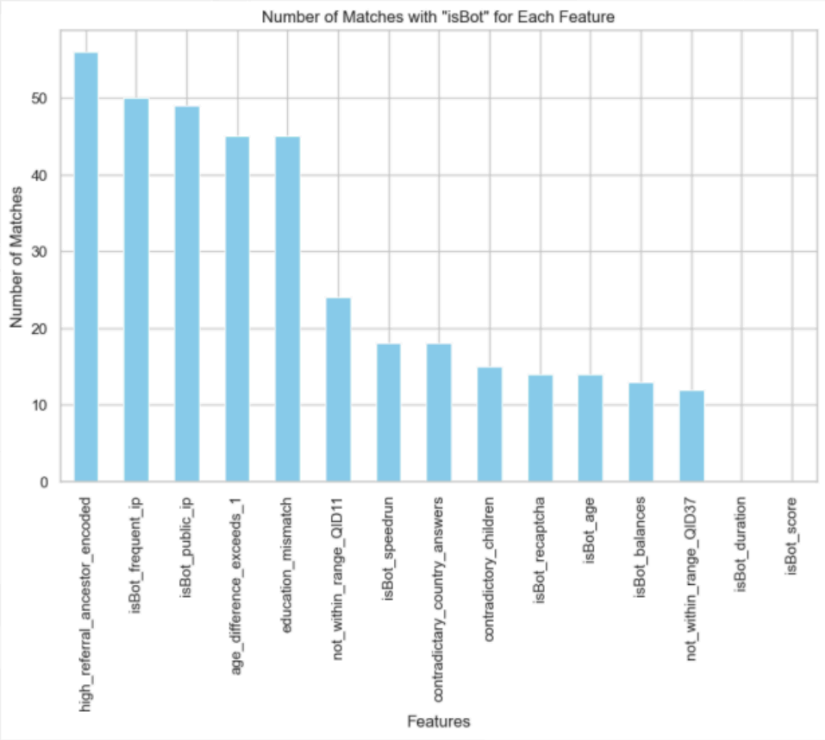
The "**isBot**" column is analyzed as the initial step in identifying potential bots among users to find patterns of automated activity.

data_fraud_public_ip and **data_fraud_demo_speedrun** have notable positive correlations with **isBot_encoded,** suggesting predictions for bot behavior.




IDENTIFYING UNUSUAL PATTERNS
BEHAVIORS THAT DEVIATE FROM THE NORM
Hard Facts
Evidence
Soft Facts
Ambiguity
Potentional Bot
Suspicious User

CONDIITIONS TO ‘ISBOT’ MATCHES



380 SUSPISCIOUS USERS

71 USERS ARE BOT MATCHES

UUID from top matches provided