



photocircle

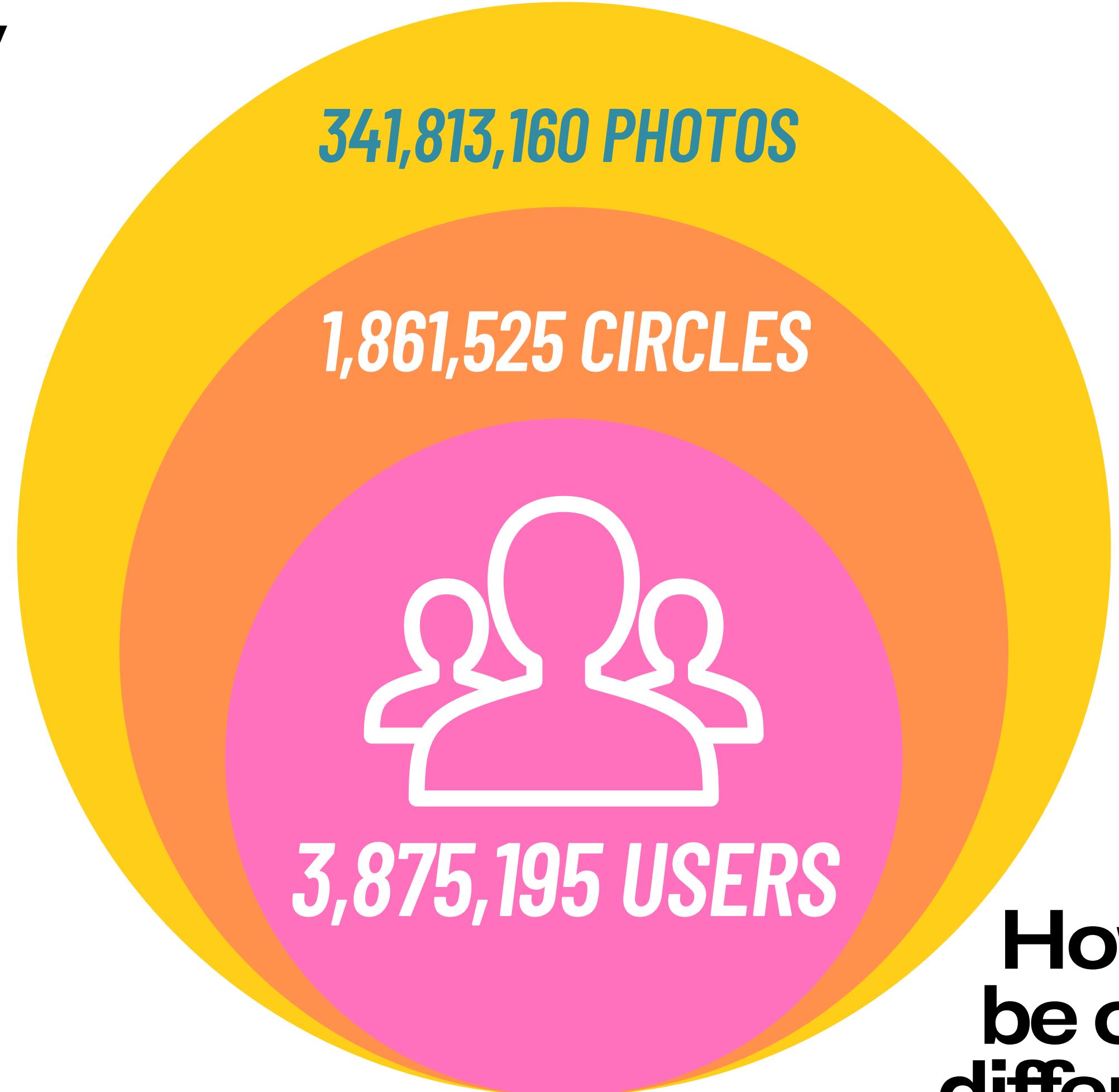
**Users Segmentation Analysis
by Bua Matthews**

DOWNLOAD |

Agenda

- 01** Overview + Data
- 02** Objectives and Goals
- 03** Methodologies
- 04** Insights from Keywords Extraction Method
- 05** Result from Topic Modeling
- 06** Limitations+ Future Studies

How many users are women?



How can circles be classified into different groups?

Objectives and Goals

Add relevant tags to users and understand user base.

Goal #1

Who is our user?

Goal #2

Who is our **ideal** user?

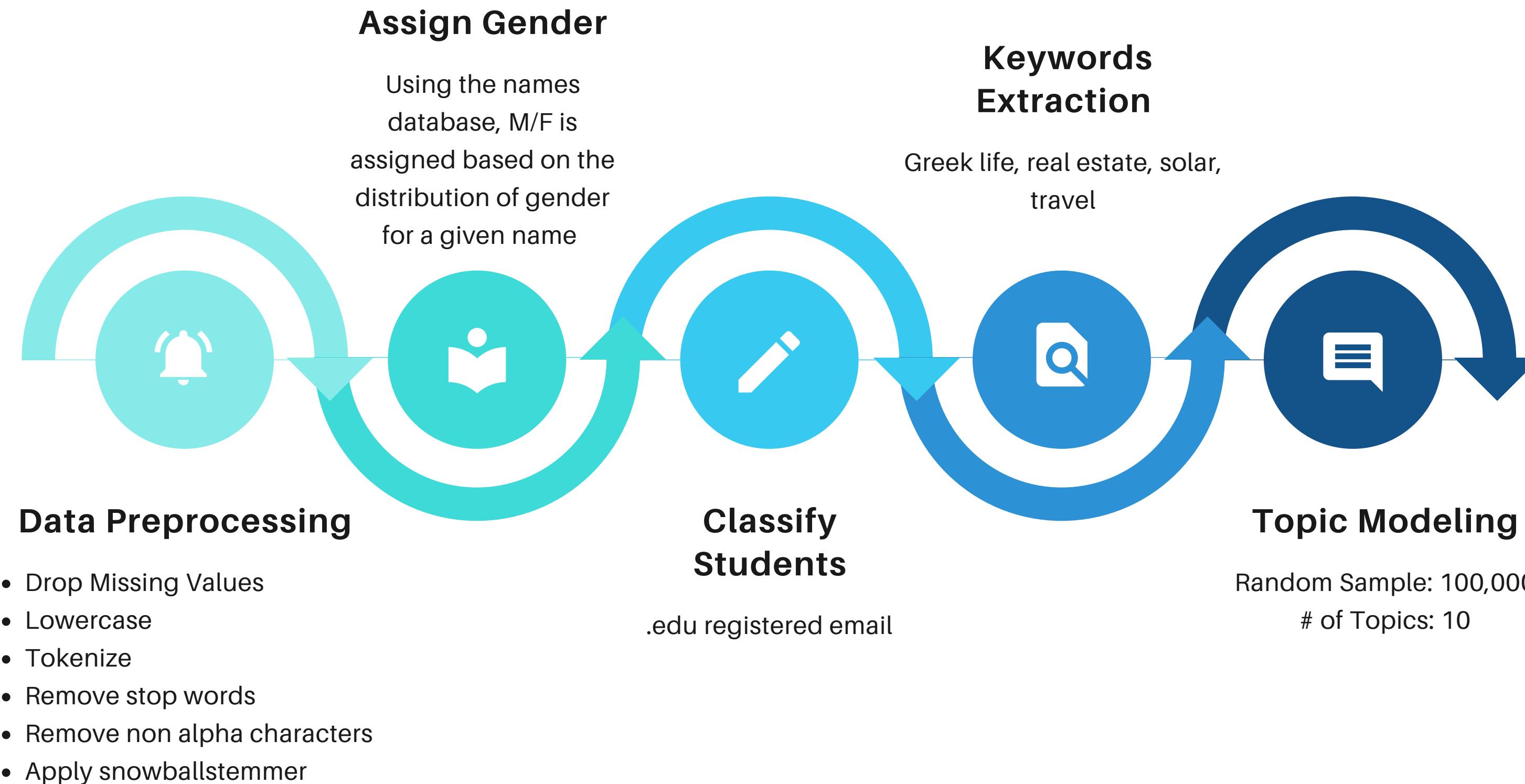
Goal #3

Who is most active?

Goal #4

Segment circles using titles

METHODOLOGIES



Keywords extraction using Real Estate / Home Development example

```
real_estate_tags = ['NN', 'NNS', 'NNP', 'NNPS', 'JJ', 'VB', 'VBD', 'VBG', 'VBN', 'VBP', 'VBZ']

real_estate_words = ['house', 'property', 'mortgage', 'market', 'rental', 'floor',
                     'wall', 'build', 'renovate', 'upgrade', 'remodel', 'refurbish']

pattern = r'\b(?:%s)\b' % '|'.join(real_estate_words)

circles['real_estate_keywords'] = circles['title'].str.findall(pattern, flags=re.IGNORECASE)

def is_real_estate(title):
    keywords = re.findall(pattern, title, flags=re.IGNORECASE)
    tokens = nltk.word_tokenize(title)
    pos_tags = nltk.pos_tag(tokens)
    if any(tag in real_estate_tags and (word in real_estate_words) for word, tag in pos_tags):
        return True
    elif any(word in real_estate_words for word in keywords):
        return True
    return False

circles['is_real_estate'] = circles['title'].apply(is_real_estate)
```

01

Define POS tags and real estate words.

02

Create a regex pattern for matching real estate keywords.

03

Extract real estate keywords from 'title' column into 'real_estate_keywords'.

04

Define a function to check if a title contains real estate keywords based on POS tags.

05

Apply the function to each title, storing results in 'is_real_estate' column.

06

'is_real_estate' column indicates real estate relevance based on keywords and POS tags.

INSIGHTS FROM KEYWORDS

EXTRACTION AND WORD FREQUENCY

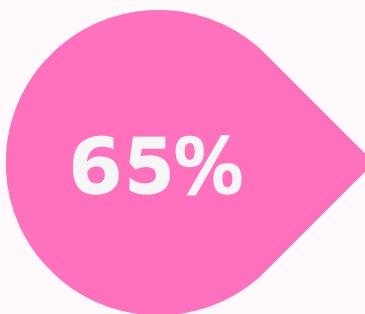


The analysis is based on prior assumptions from the app owner:

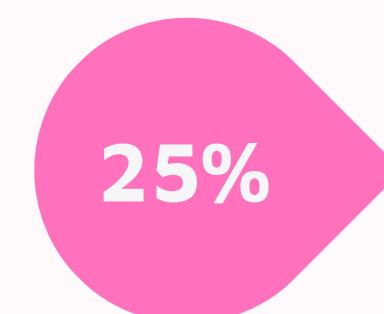
1. email domains registered with .edu tend to be student
2. photo album names with any Greek words tend to be connected with sorority/fraternity users

Aside from the assumptions above, more tags in home development, solar, and travel were also generated to help the app owner better understand marketing strategy.

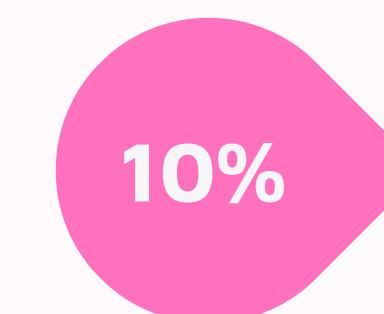
STATISTICS ON APP USAGE



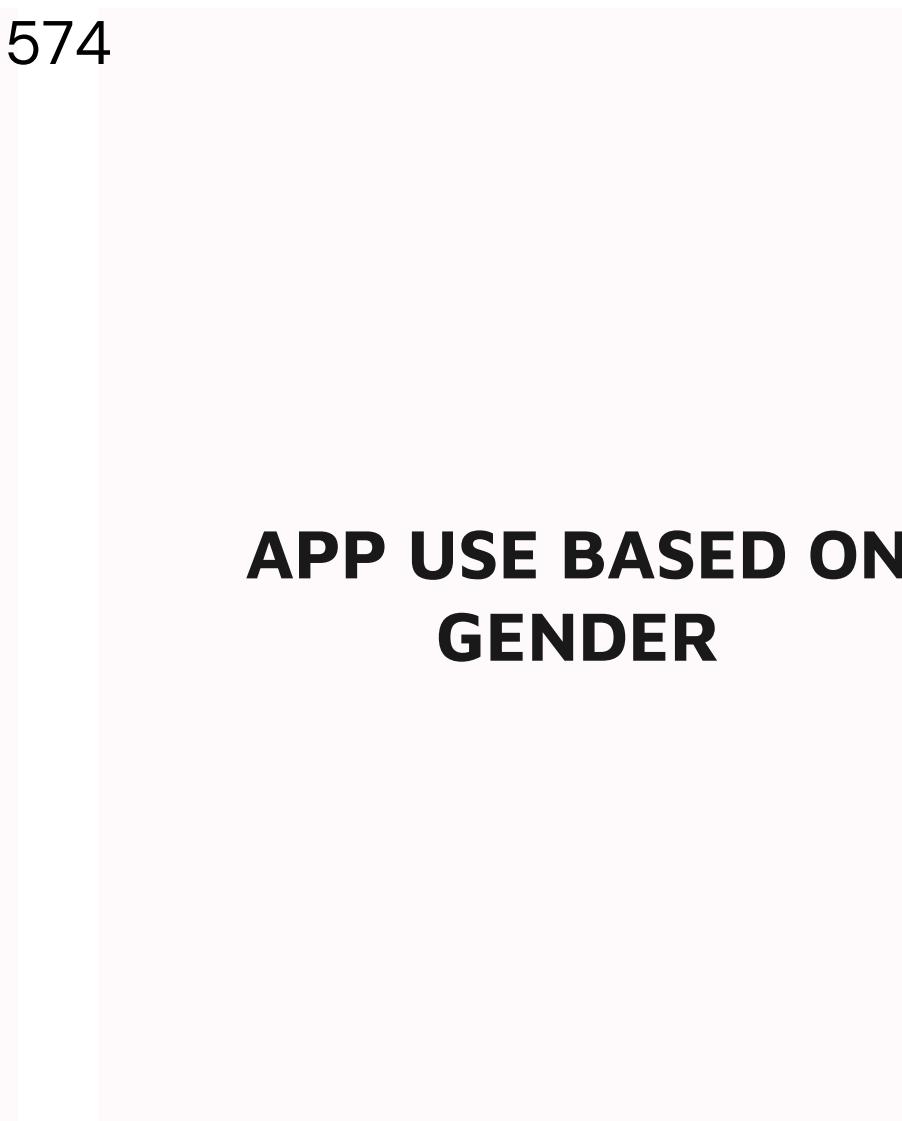
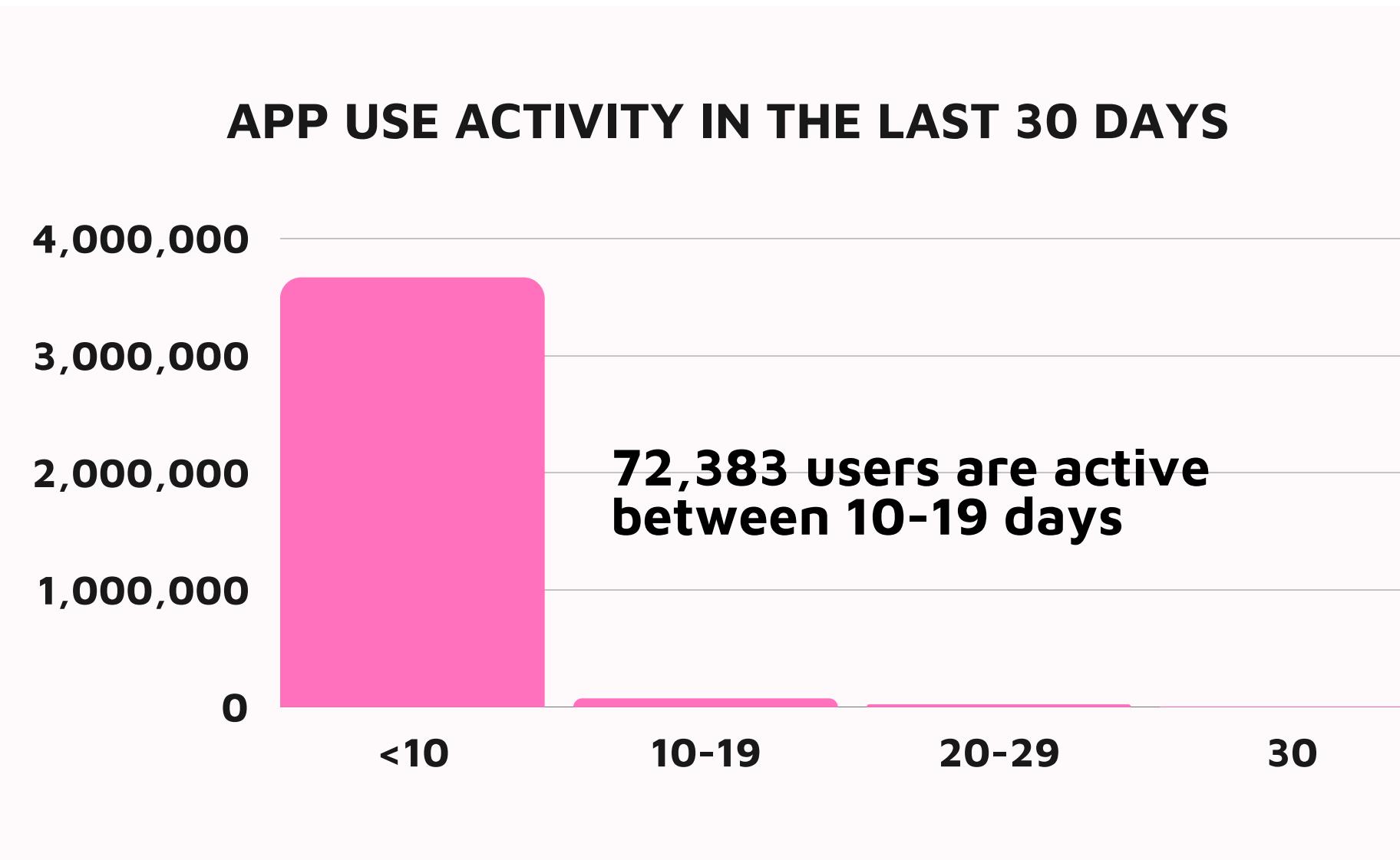
Use apps for recording personal travels and trips



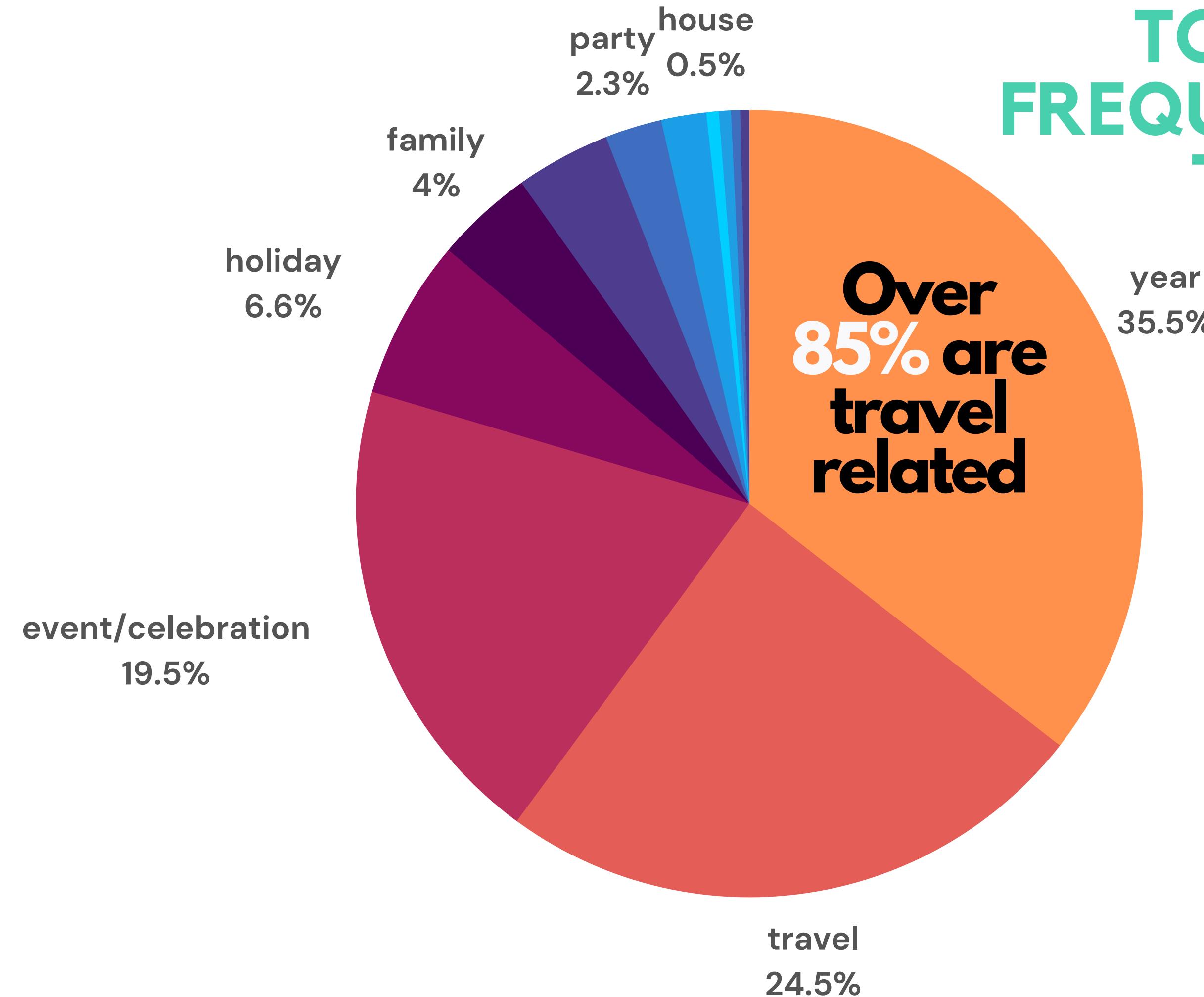
Use apps for recording life events and celebrations



Use apps for other personal and work related activities



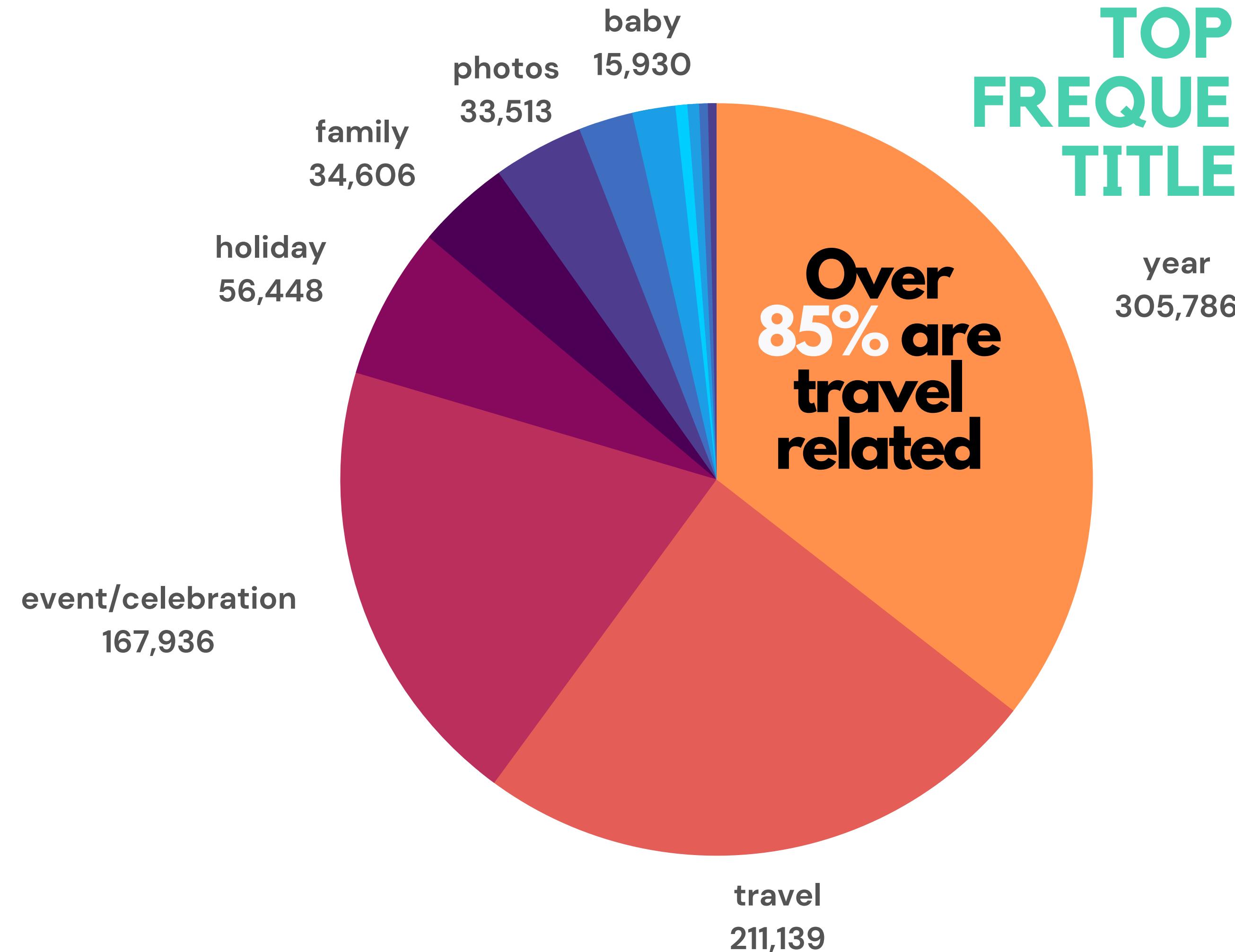
TOP 100 WORD FREQUENCY IN CIRCLE TITLES BY %



WORDS IN
'EVENT/CELEBRATION'
CATEGORY INCLUDES:

wedding	senior
birthday	bridal
weekend	prom
bday	bach
shower	50th
girls	reunion
graduation	grad
bachelorette	40th
big	first
4th	week

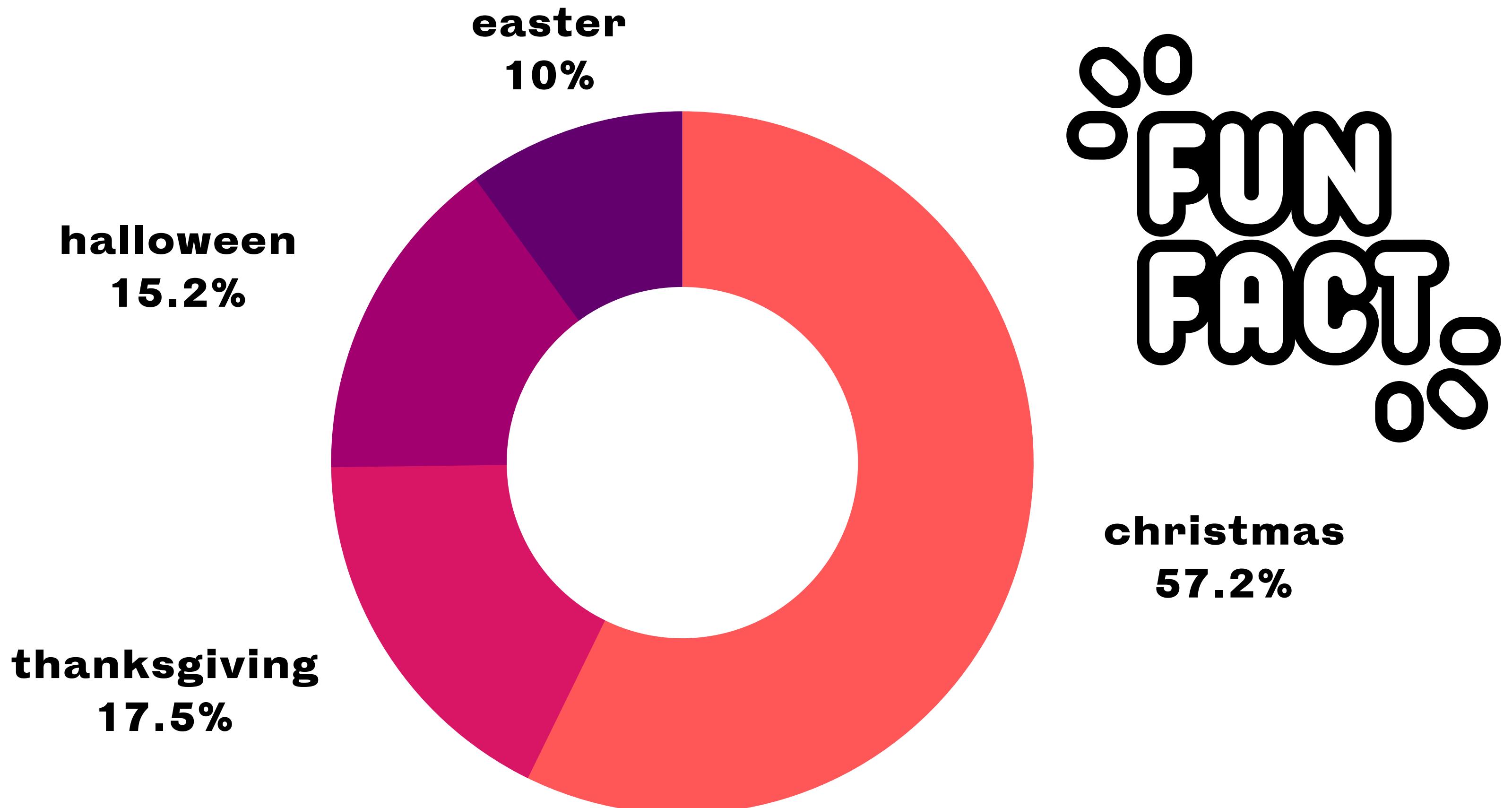
TOP 100 WORD FREQUENCY IN CIRCLE TITLES BY COUNT



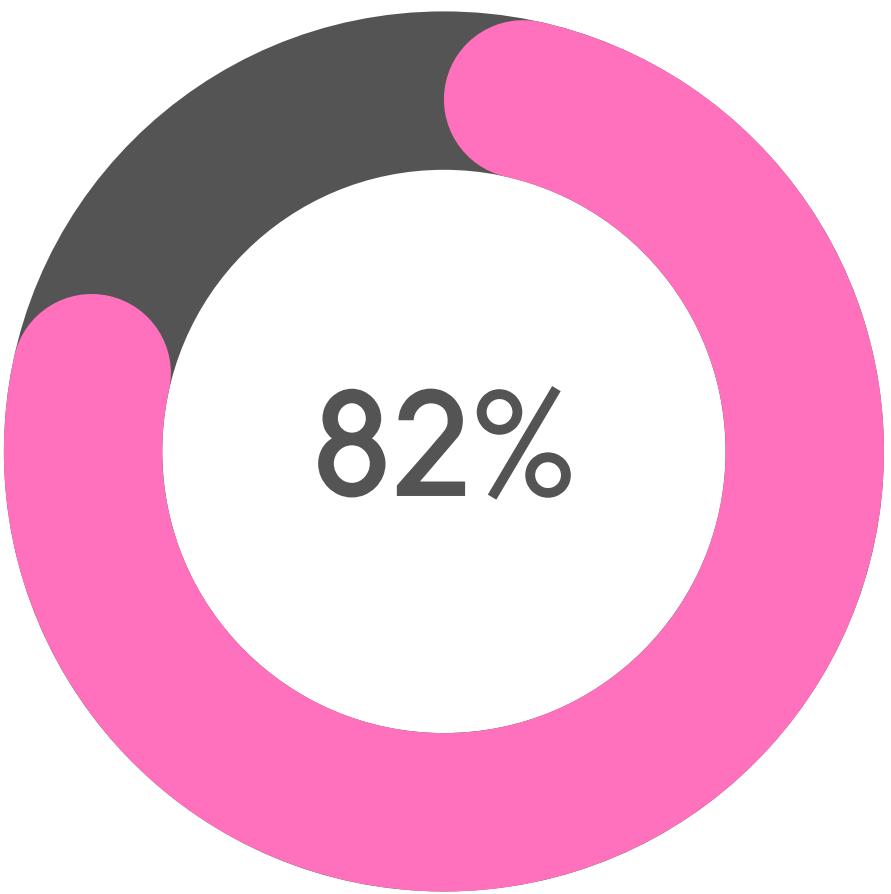
WORDS IN 'TRAVEL' CATEGORY INCLUDES:

trip	camp
new	florida
summer	june
july	mexico
beach	may
lake	adventures
vegas	march
disney	nyc
spring	la
cruise	camping
hawaii	cabo
san	park
fall	colorado
break	disneyland
vacation	city
	italy

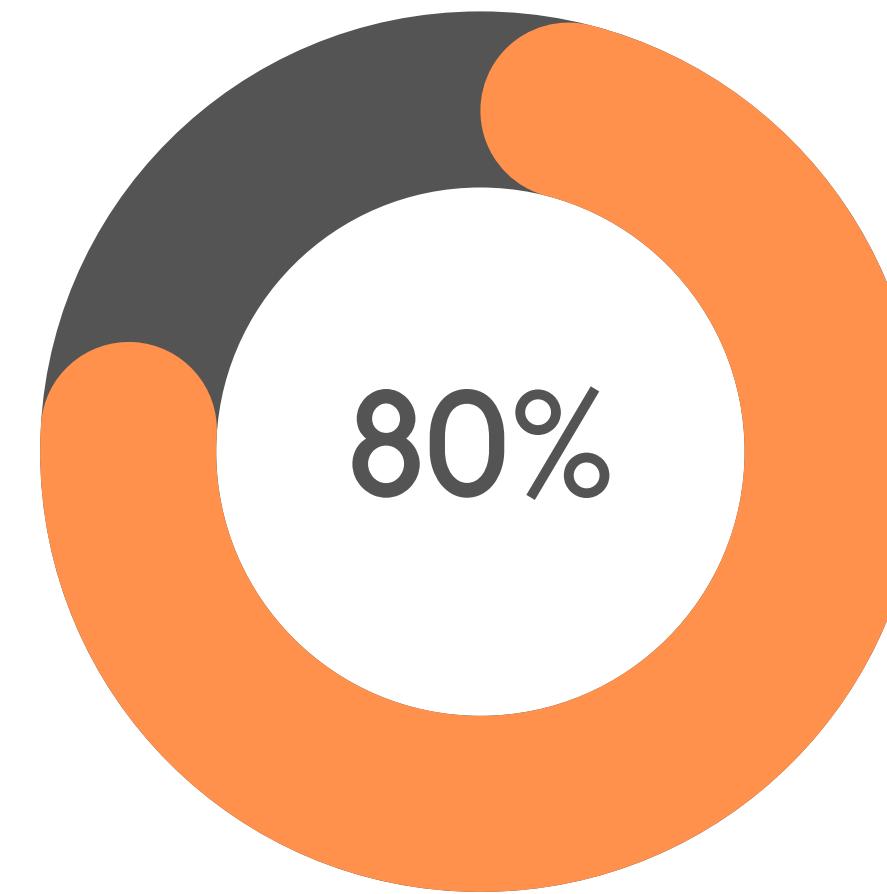
MOST POPULAR HOLIDAYS FOUND IN CIRCLE TITLES



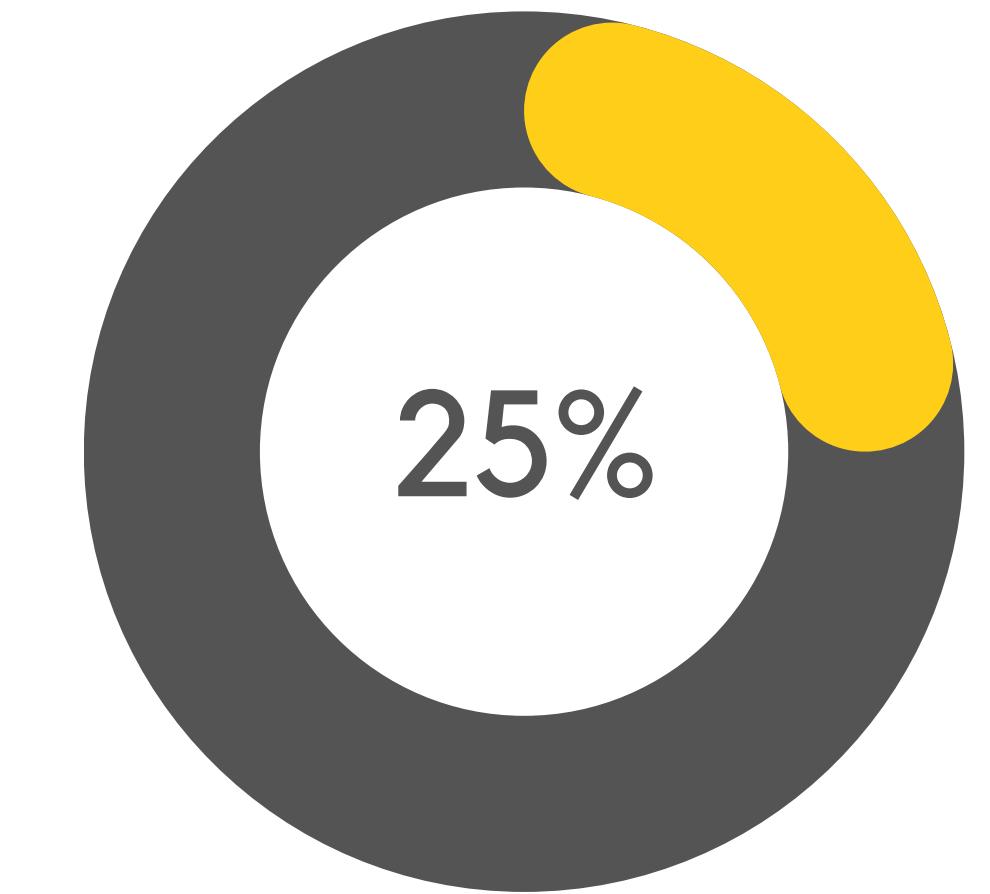
WHO IS OUR IDEAL USER?



WOMEN

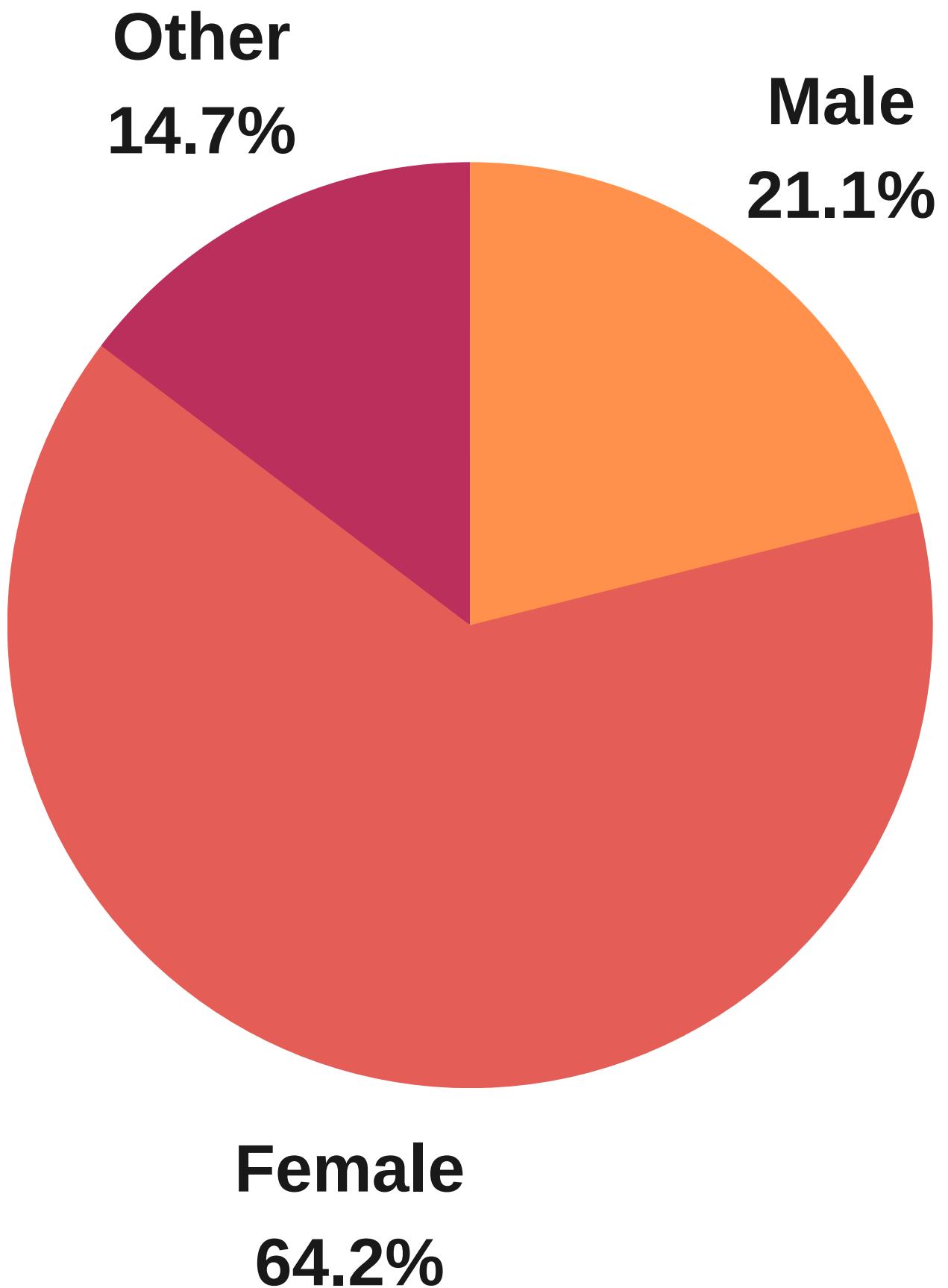


TRAVELRS



EVENT-CENTRIC

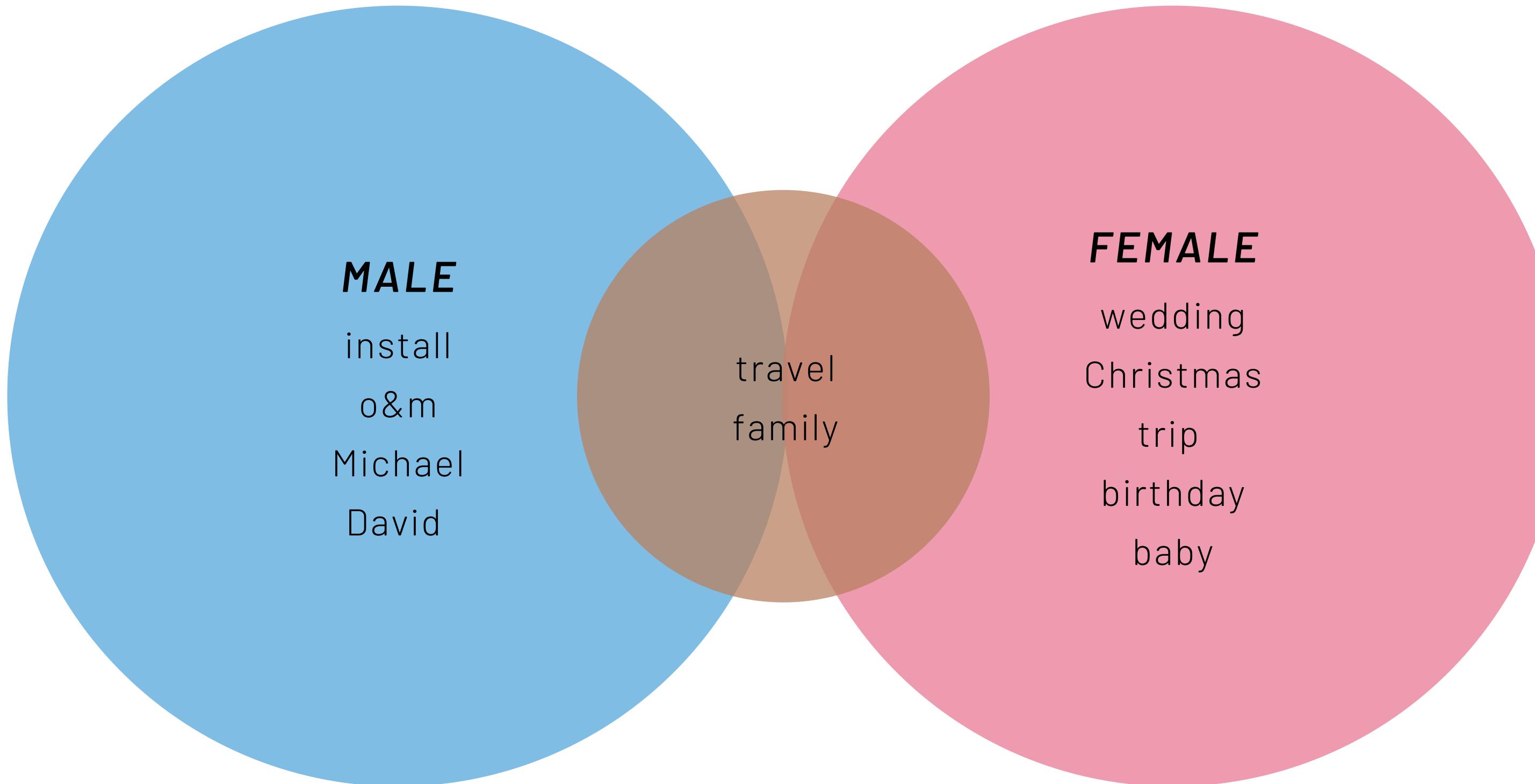
WHO IS MOST ACTIVE?



**USERS WHO
ARE ACTIVE
BETWEEN 20-
30 DAYS IN THE
LAST 30 DAYS
BY GENDER**

If we define active as users who are active between 20–30 days in the last 30 days, then over 50% of this subset of users are women.

ACTIVE MALE AND FEMALE USERS TEND TO CREATE CIRCLES RELATED TO TRAVEL AND FAMILY



TOPIC MODELING



RESEARCH LIMITATIONS AND FUTURE STUDIES

ADD MORE TEXT



POS TAGGING IN NLP

A circle title named 'Paris 2020' may be classified for 'travel' but it could contain photos of baby, dog, etc

STRATIFIED SAMPLING INSTEAD OF RANDOM SAMPLING

Preserve the variance of the population

COMPUTER VISION

To achieve classification accurate, image classification, object detection, segmentation, and recognition is more appropriate

A large blue word 'thank' is centered. Around it, several other words representing 'thank you' in different languages are arranged in a cloud-like pattern. The words include 'tusind tak' (Danish), 'dakujem vám' (Czech), '謝謝' (Chinese), 'ありがとう' (Japanese), 'gracias' (Spanish), 'obrigada' (Portuguese), 'obrigado' (Brazilian Portuguese), 'teşekkür ederim' (Turkish), 'شكراً' (Arabic), 'tack så mycket' (Swedish), and 'tack' (Swedish). The colors of the words vary, including blue, orange, yellow, red, and green.

EXTRA EXTRA!



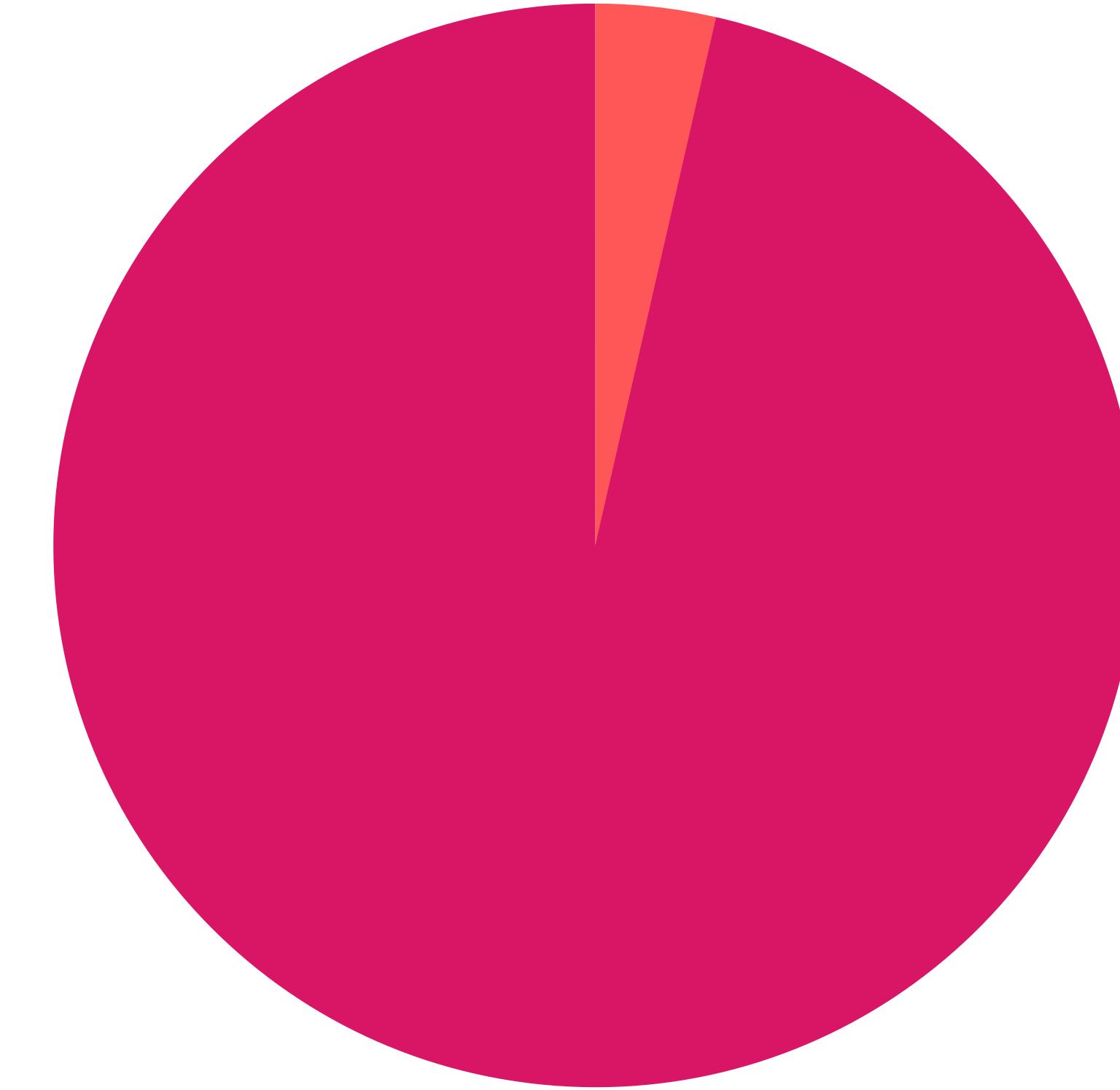
Greek Life Deep Dive



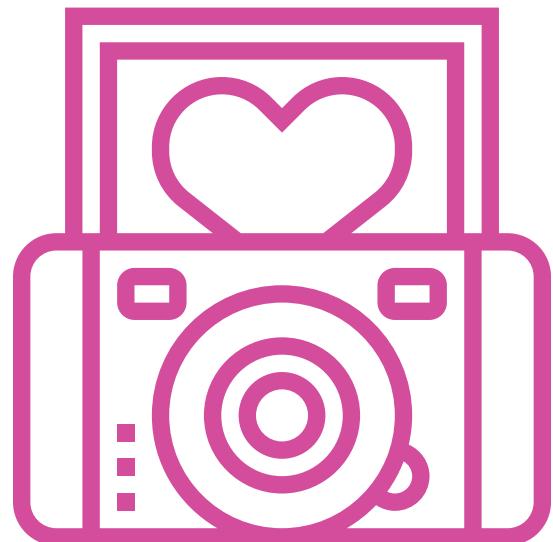
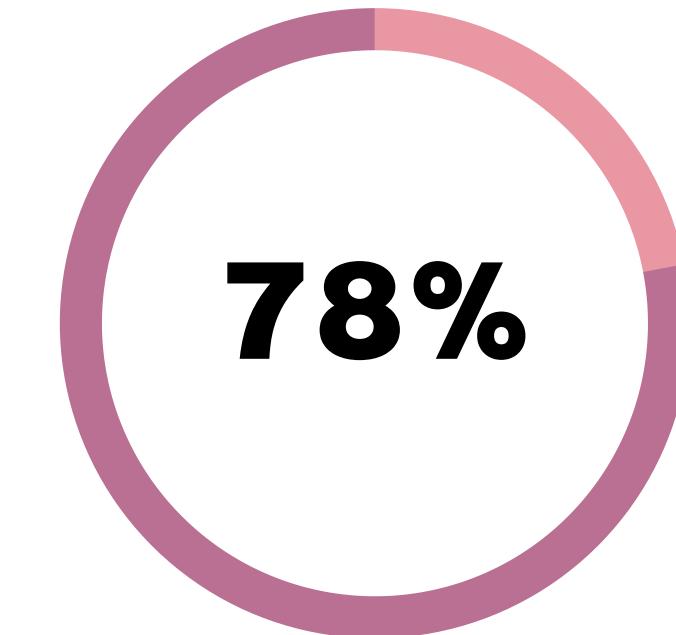
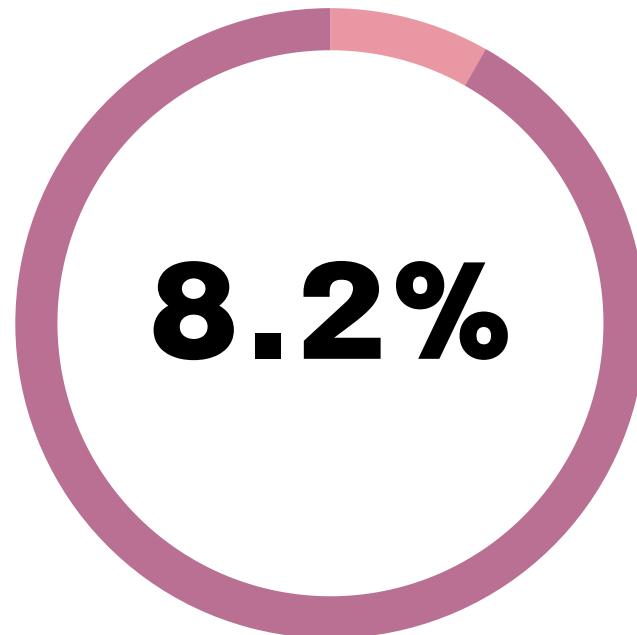
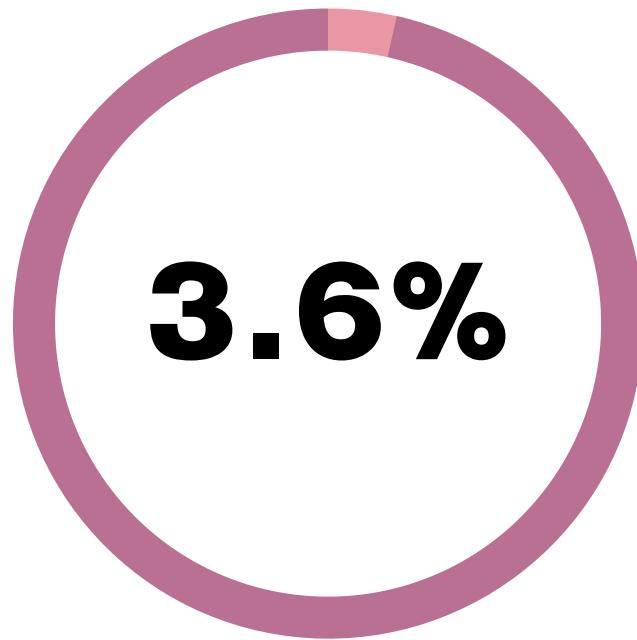




Total is_Greek Media
3.6%



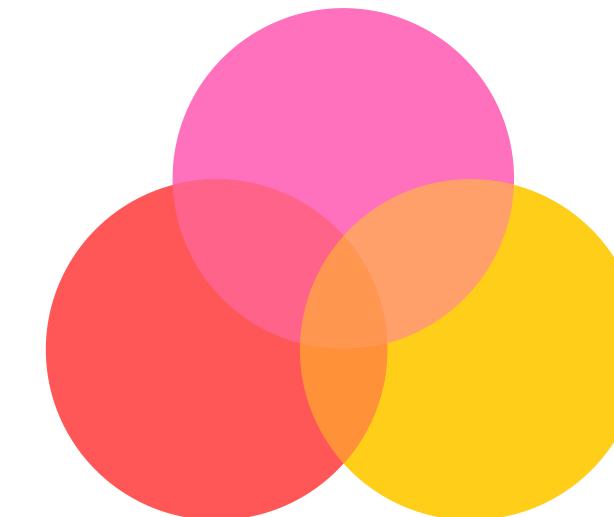
Other
96.4%



12,240,969
Photos



309,184
Users

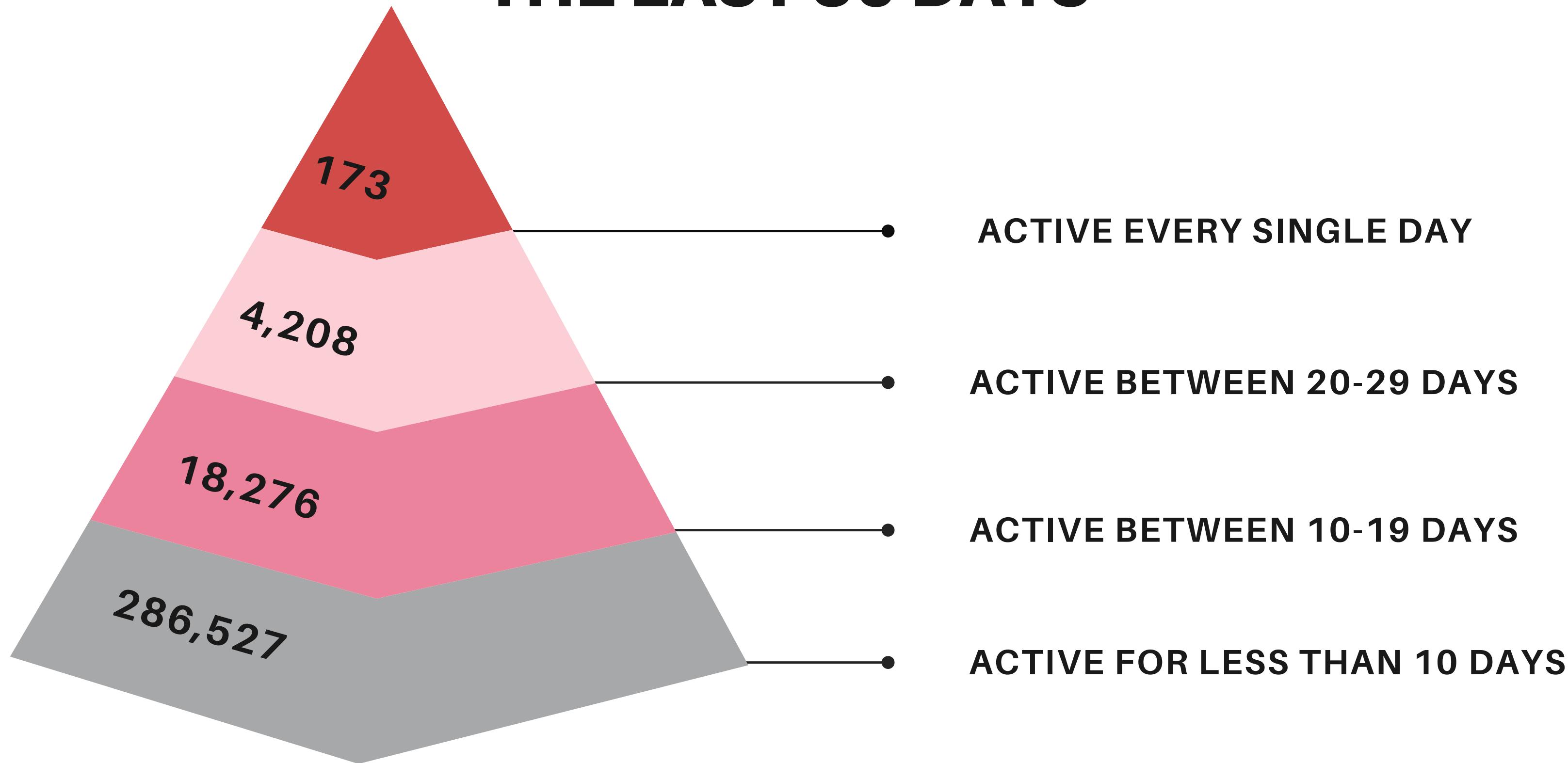


9,439
Circles



Female
Users

IS_GREEK USERS ACTIVITY IN THE LAST 30 DAYS



TOP GREEK MEDIA CONTRIBUTORS

Top 5 most media uploads circles

Sorority	Media Added
Alpha Phi	207,110
Kappa Delta	165,351
Alpha Chi Omega	108,826
Gamma Phi Beta	108,436
Alpha Delta Pi	99,205

Top 5 highest media contributors

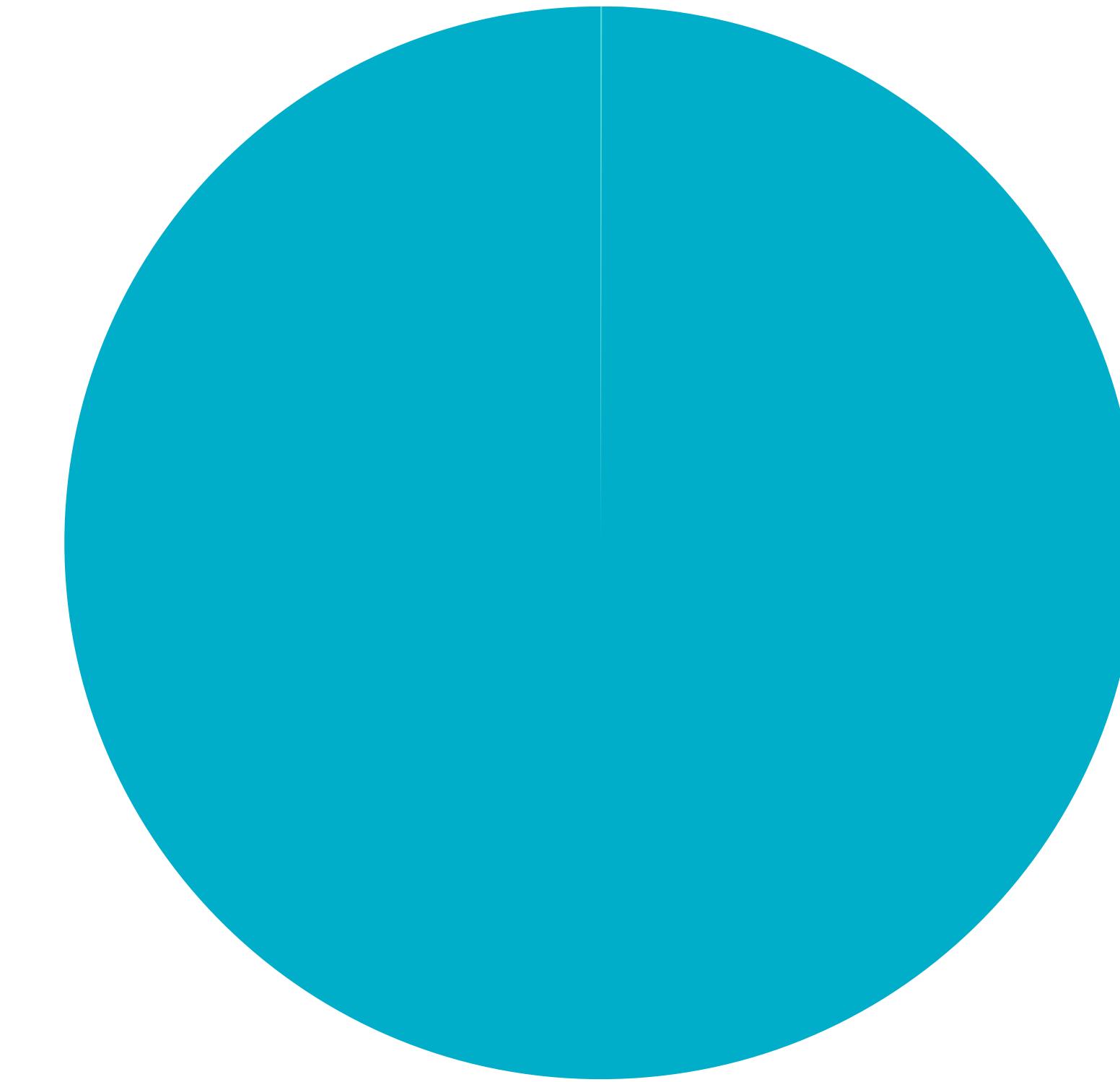
Name	Total Media
Abby	13,442
Kappa	8,444
Emma	7,923
Carly	6,802
Julia	6,730



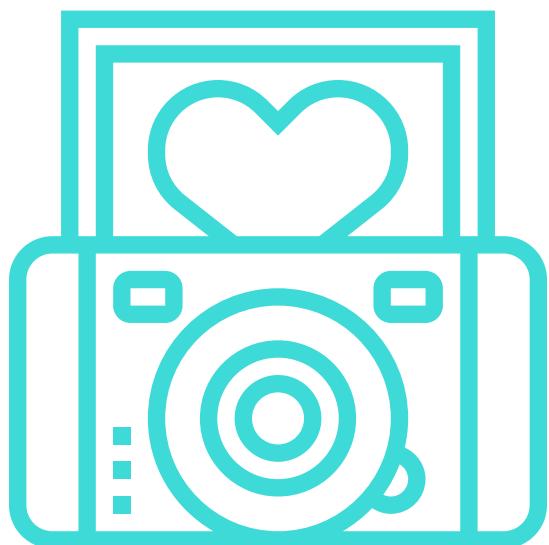
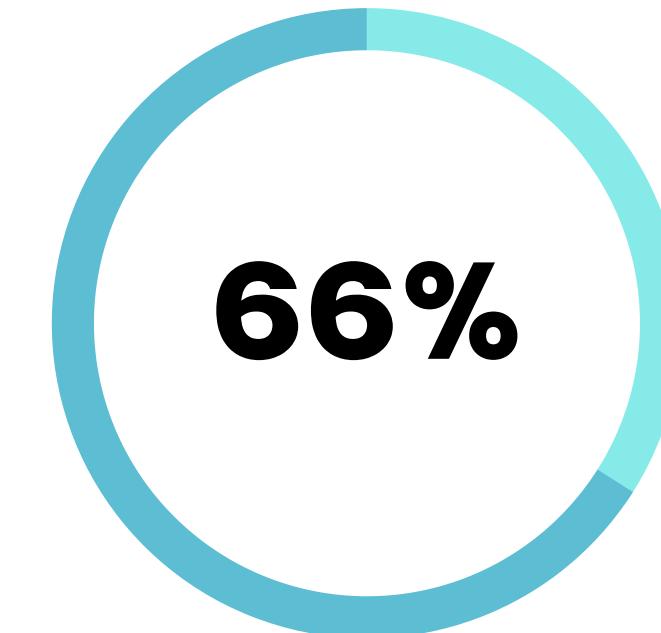
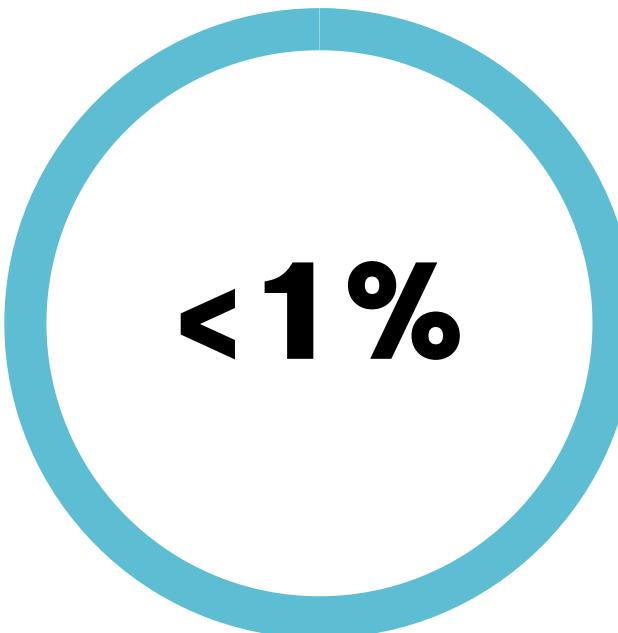
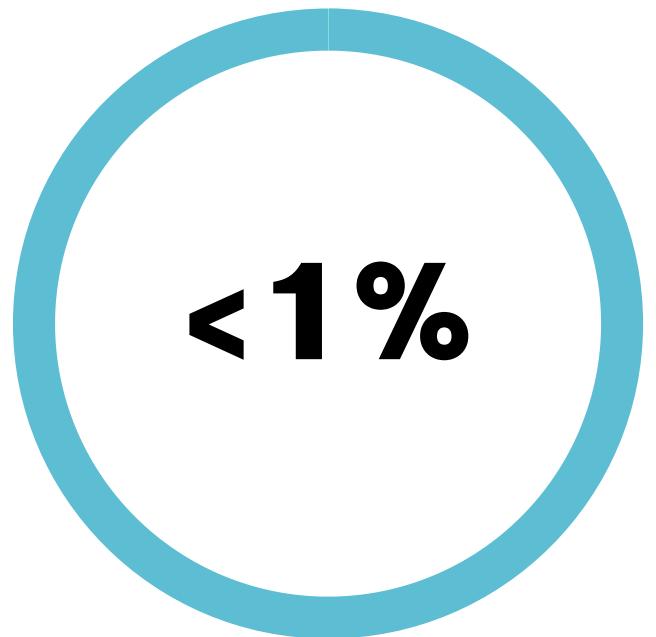


Home Media Contribution

141,542



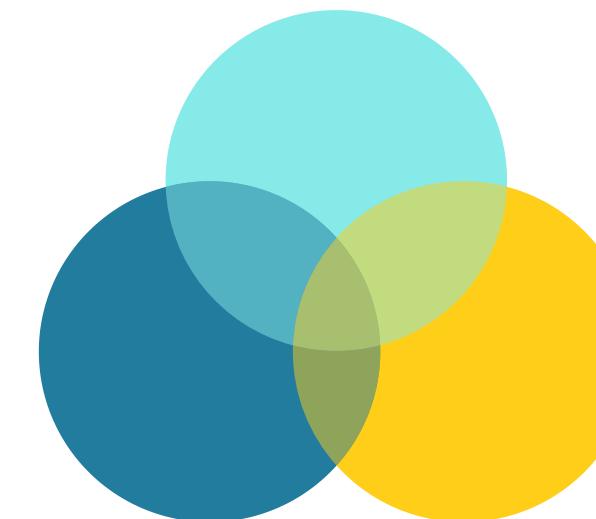
Other
341,671,738



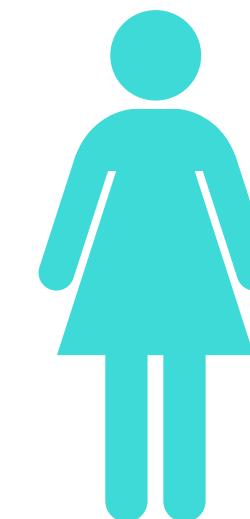
141,542
Photos



5,075
Users

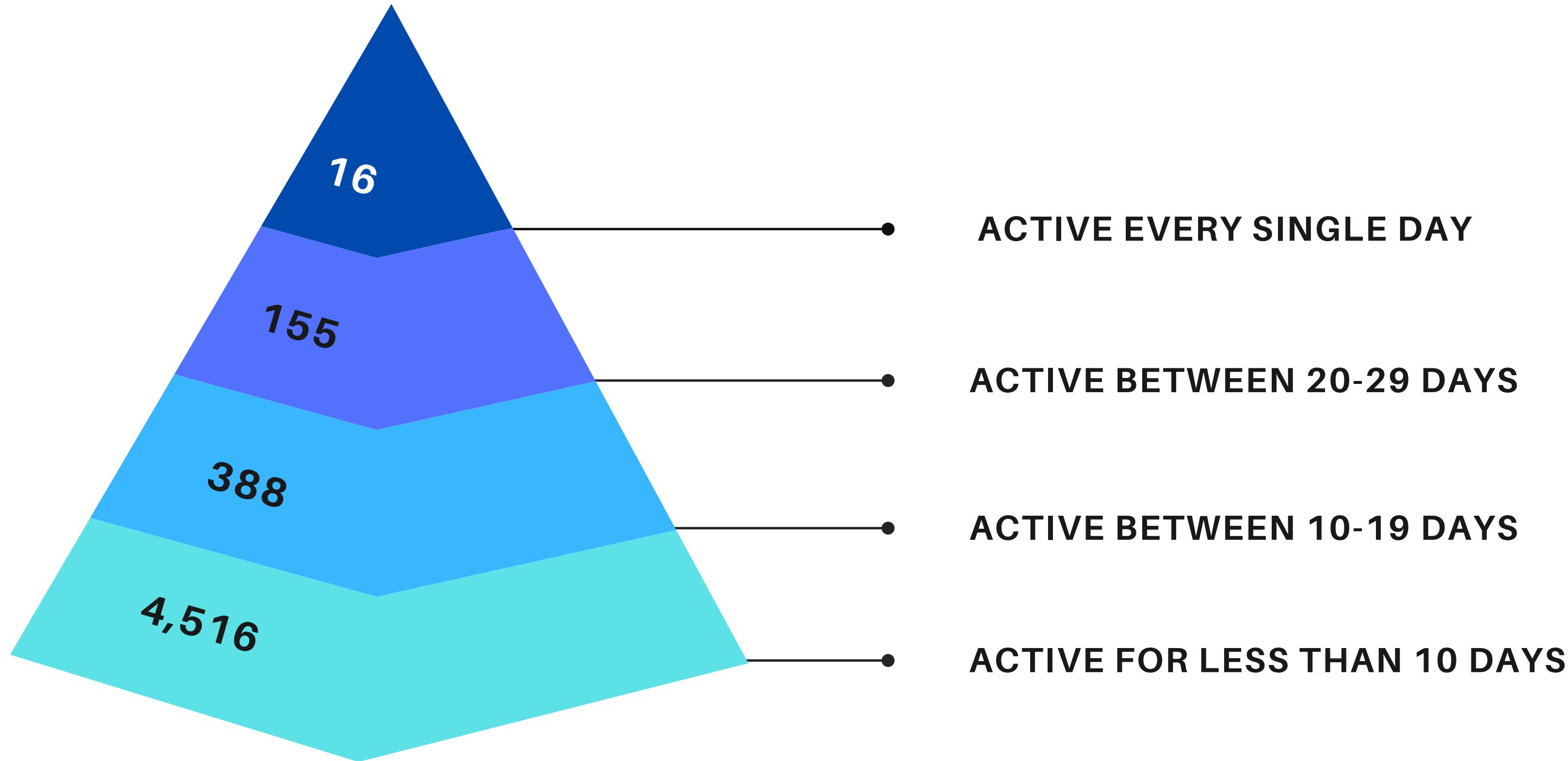


1,103
Circles



Female
Users

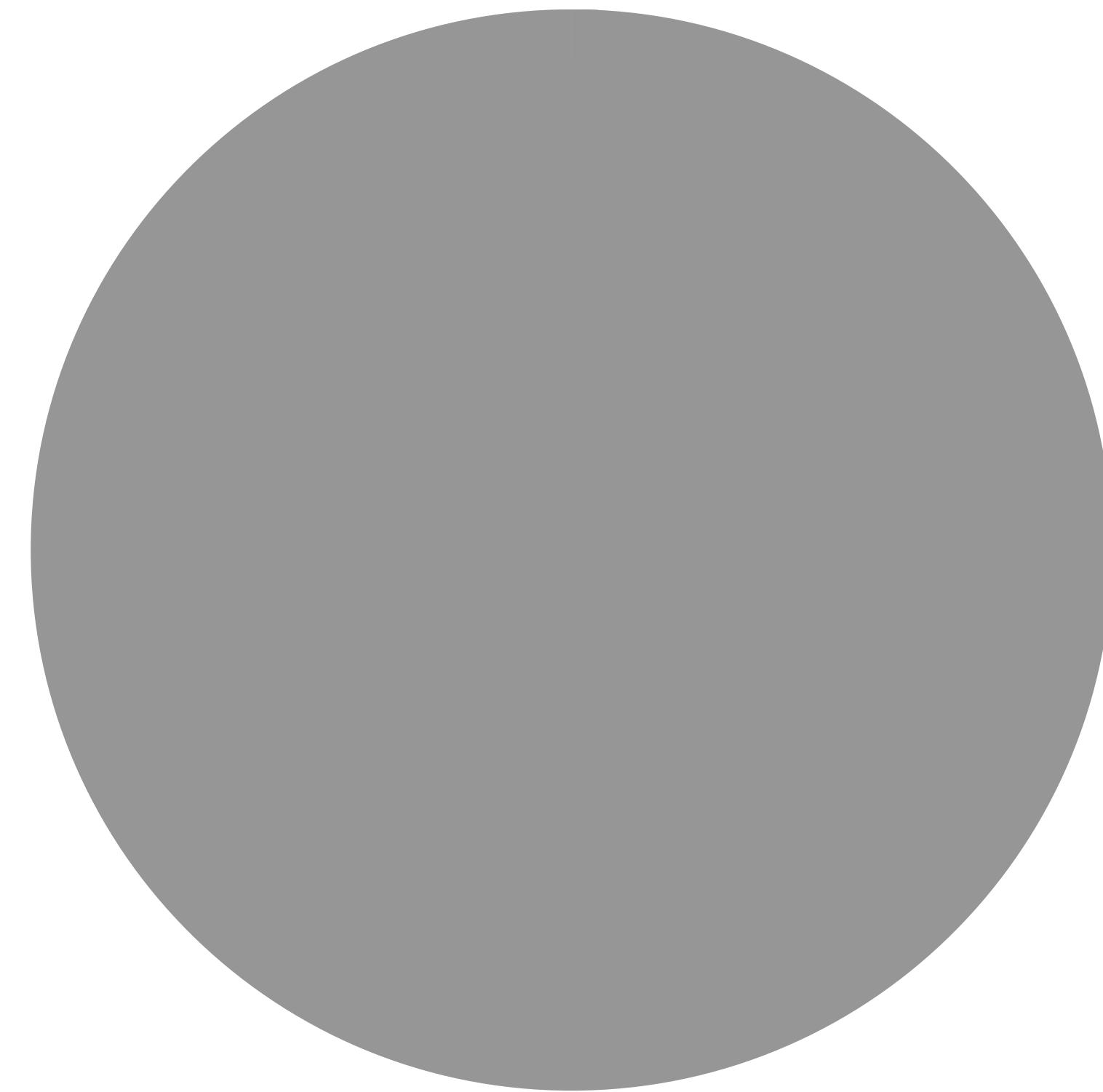
HOME IMPROVEMENT / REAL ESTATE USERS ACTIVITY IN THE LAST 30 DAYS



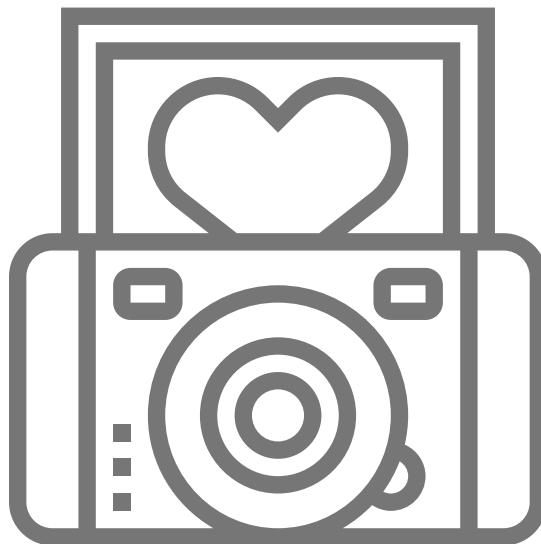
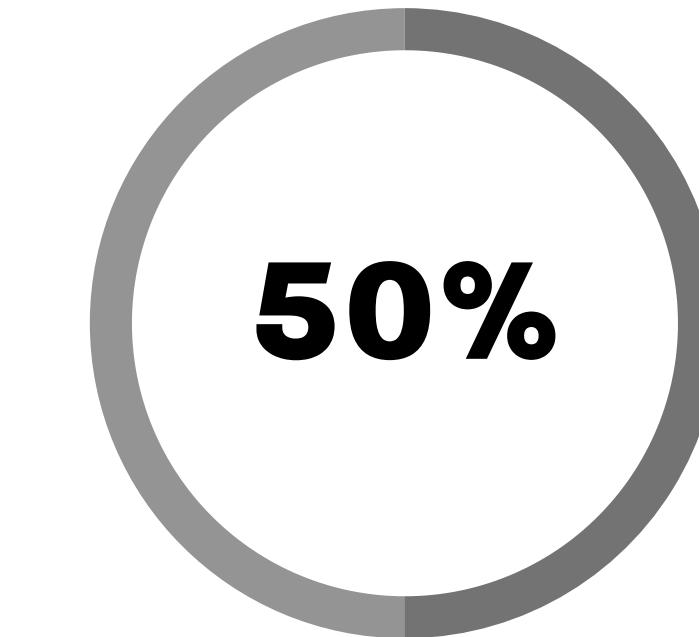
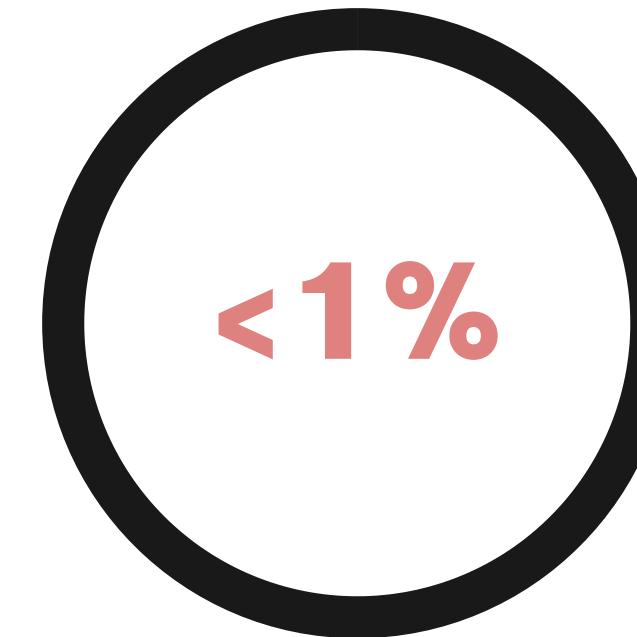
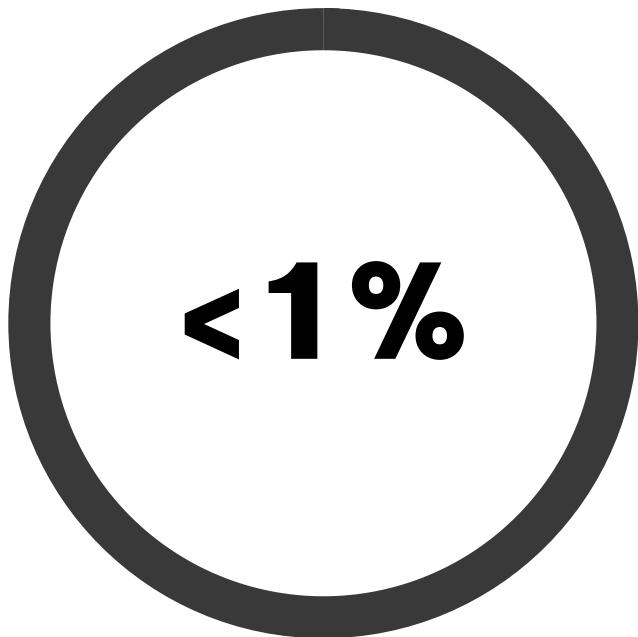
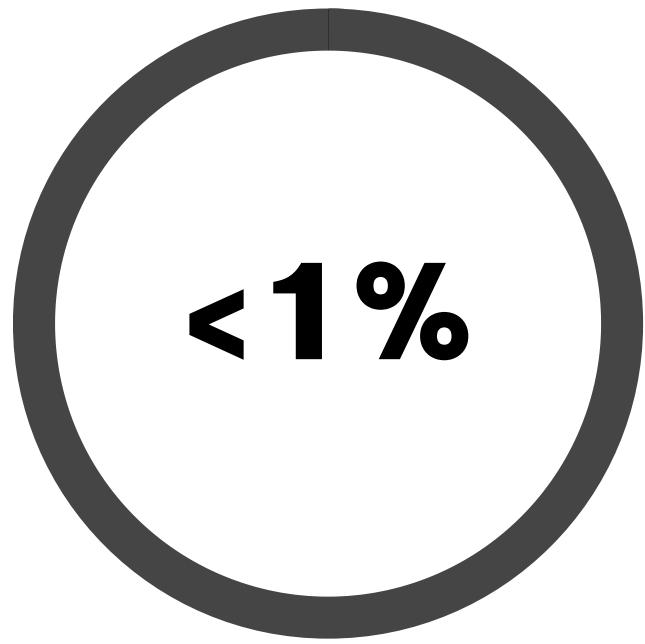


Solar Media Contribution

9,462



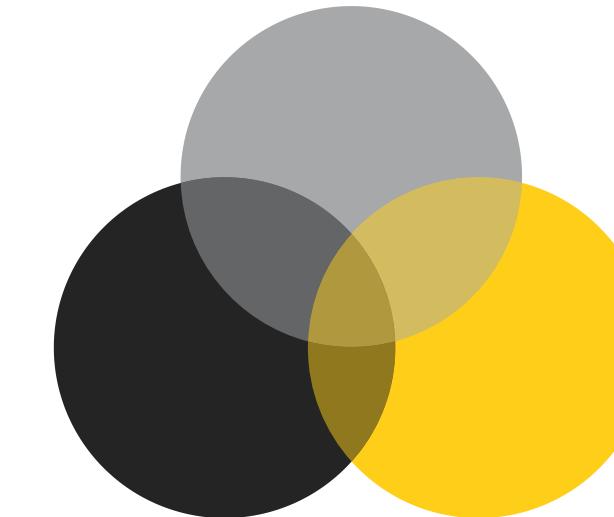
Other
341,803,818



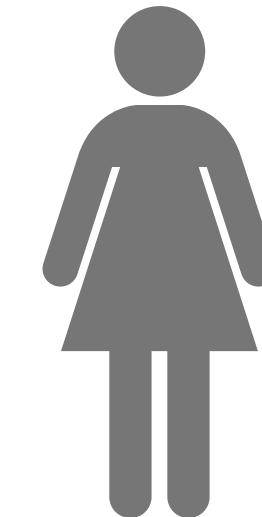
9,462
Photos



355
Users

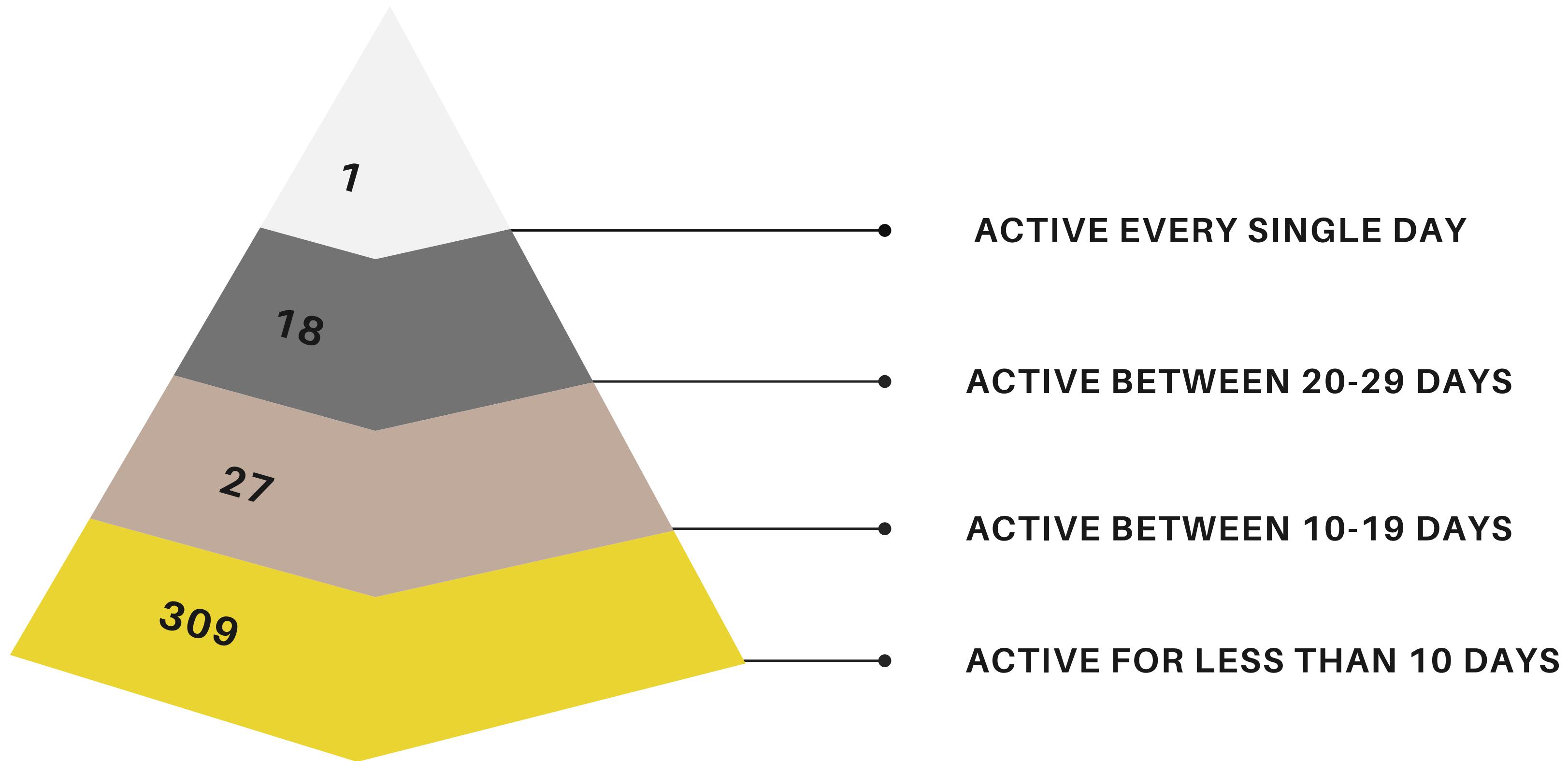


114
Circles



Female
Users

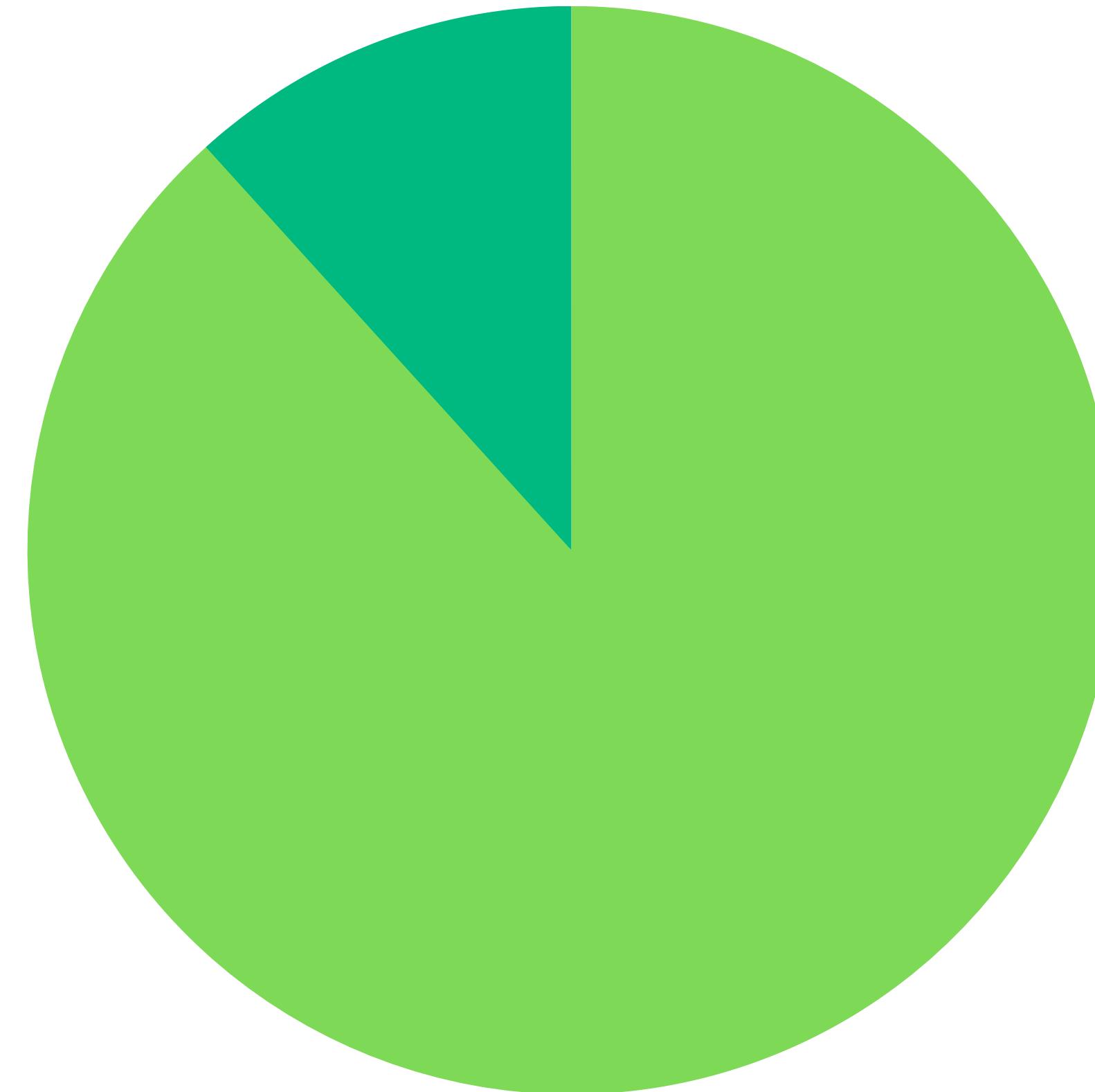
SOLAR USERS ACTIVITY IN THE LAST 30 DAYS



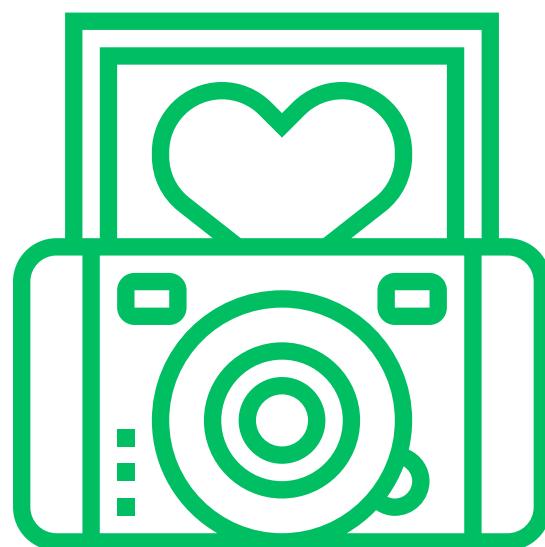
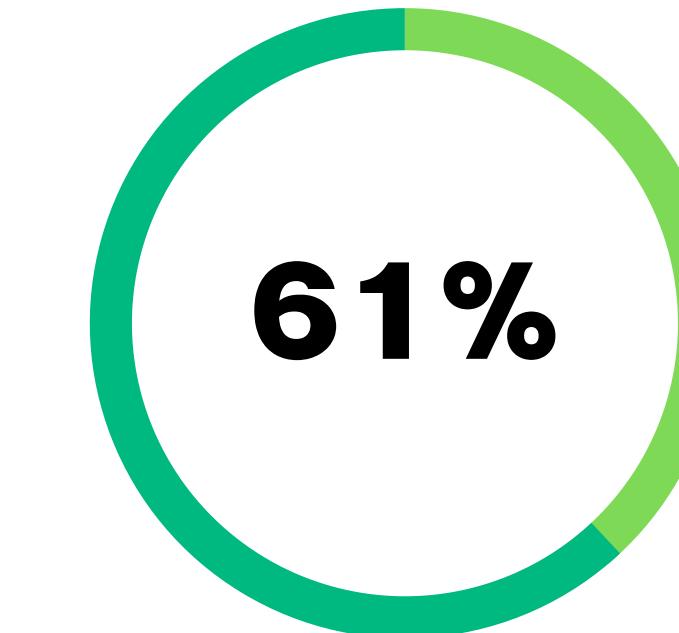
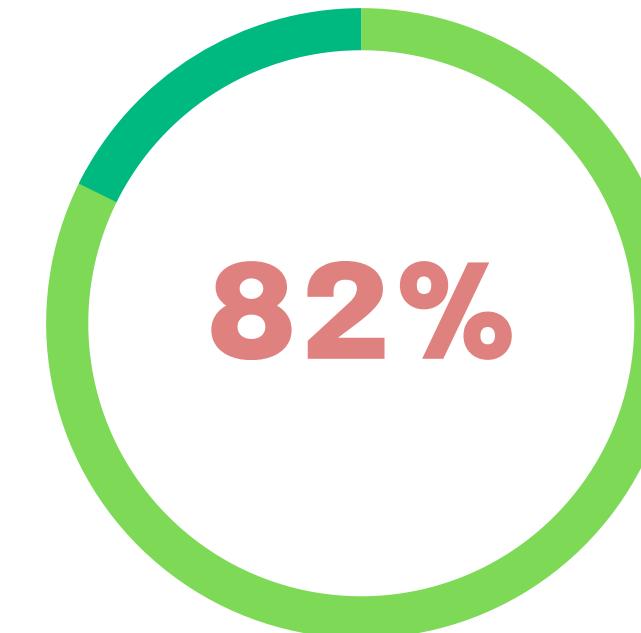
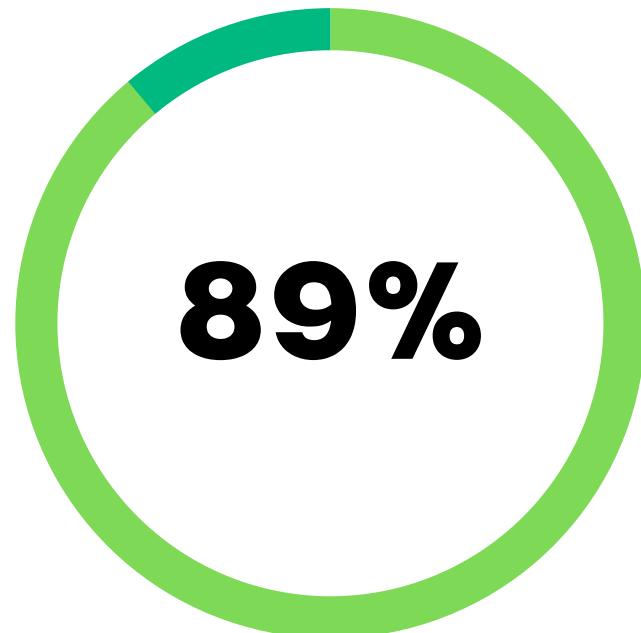
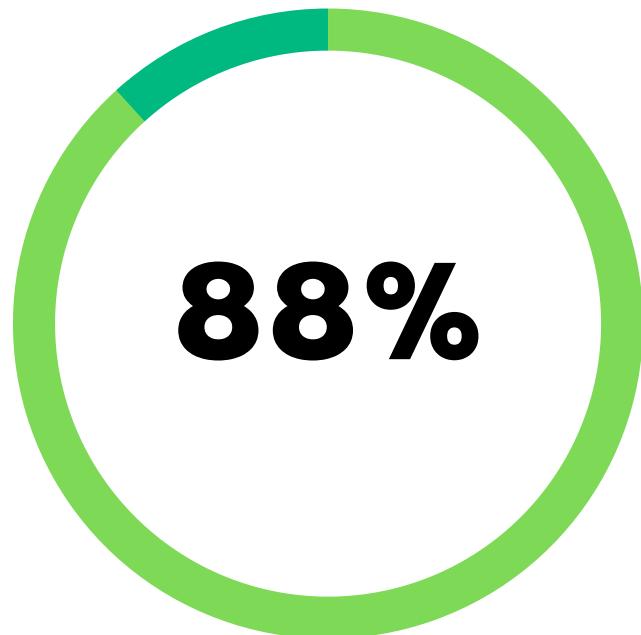




**Other
11.7%**



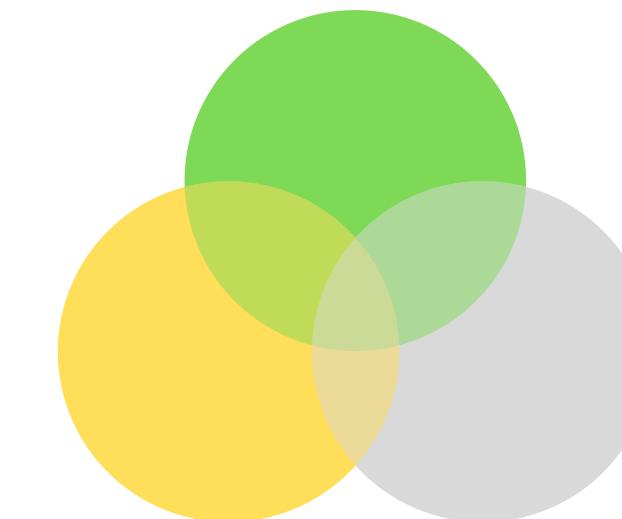
**Travel Photos
88.3%**



301,739,352
Photos



3,345,312
Users

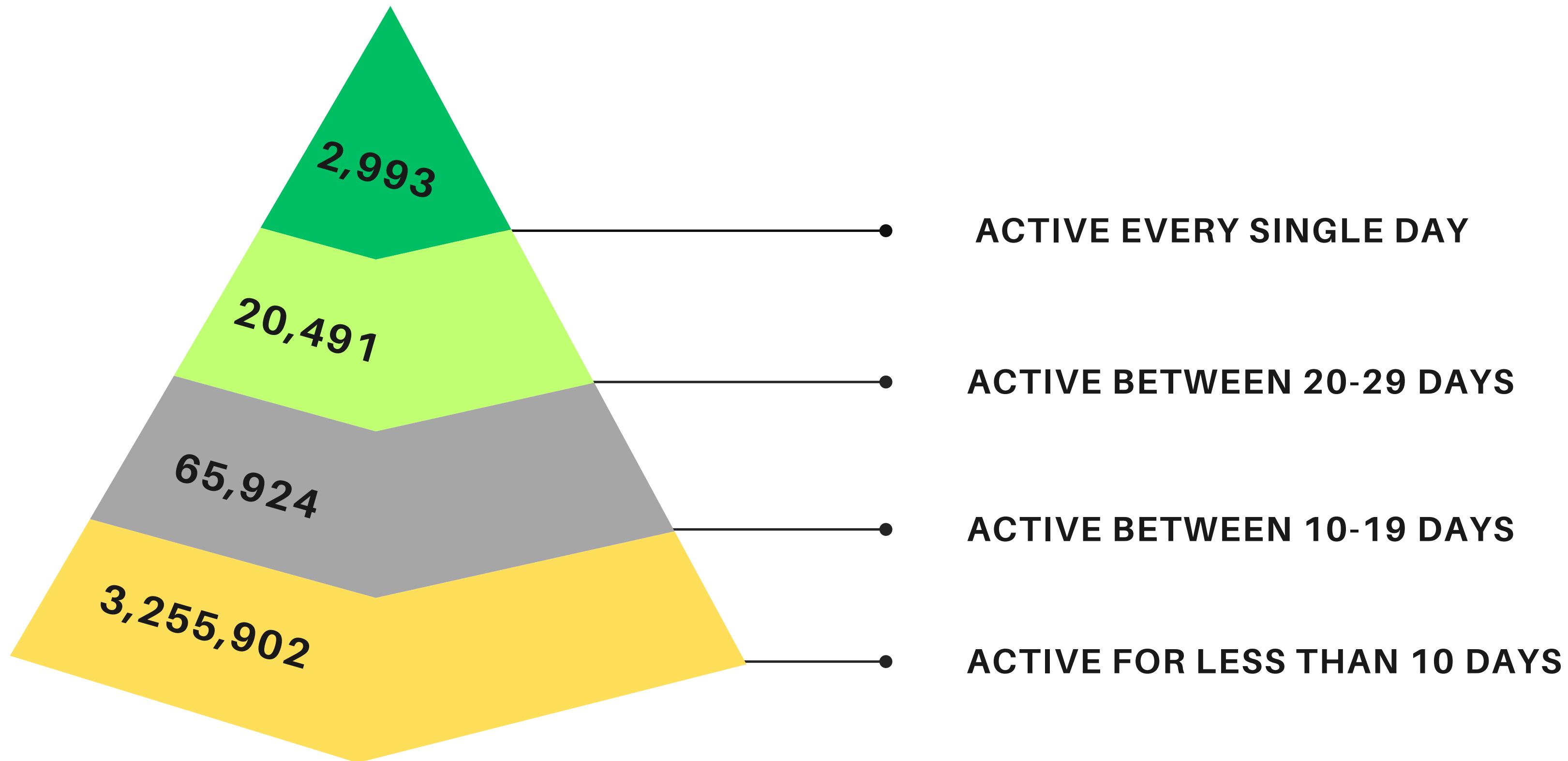


1,681,235
Circles



Female
Users

TRAVEL RELATED USERS ACTIVITY IN THE LAST 30 DAYS



KEYWORDS

REAL ESTATE AND HOME IMPROVEMENT-RELATED WORDS FOUND IN CIRCLE TITLES

Floor	House	house	Build	HOUSE
Wall	Property	Market	floor	Upgrade
FLOOR	House, Remodel	Remodel	build	House, Build
MARKET	house, build	WALL	Rental	wall
market	remodel	Floor, Market	House, Property	WAll
Build, House	House, Wall	Property, Remodel	House, Market	remodel, house
Wall, Build	Floor, Floor	House, House	Build, Wall	upgrade
property	RENTAL	Market, Market	Mortgage	houSe
House, remodel	Build, Build	Wall, House	Floor, Remodel	Wall, Wall
Rental, Property	rental	house, house	house, wall	Rental, House
Renovate	House, Rental	build, wall	HouSe	Remodel, House
Floor, Wall	WALL, HOUSE	FLOOR, WALL		

SOLAR-RELATED WORDS FOUND IN CIRCLE TITLES

Power	Sun	Panel
System	SOLAR	Energy
GRID	Solar	Sun, power
Grid	solar	power
Electricity	PANEL	Solar, System
Sun, Energy	panel	POWER, PANEL
Renewable, Energy	Solar, Panel	Sun, Sun
GRID, PHOTOVOLTAIC	eNeRgY	EnerGy,
Renewable	solar, system	Power, System