A photograph of two people from the waist up, standing side-by-side against a solid blue background. On the left, a person with dark skin is wearing a bright green V-neck sweater over a pink collared shirt and an orange tie, paired with blue jeans. On the right, another person with dark skin is wearing a bright yellow V-neck sweater over a green t-shirt, paired with a magenta skirt.

MICHAEL CHARARA 10.2023

# EXPLORATORY ANALYSIS OF OKCUPID USERS

Understanding User Profiles for Effective Marketing

# AGENDA

- ▶ Project Overview
- ▶ Data Source & Ethics
- ▶ Data Cleaning & Pre-processing
- ▶ Data Analysis Outcomes
  - > Age & Orientation
  - > Income & Relationship Status
  - > Geographical Distribution
  - > Orientation & Drinking Habits
- ▶ Key Takeaways
- ▶ Acknowledgments & References
- ▶ Questions & Feedback





okcupid

# GOAL

**Identify micro-segments and personas for future campaigns.**



# DATA SOURCE AND ETHICS

- ▶ Scraped from user profiles.
- ▶ Data approved for academic & business use after anonymization.
- ▶ Source: [ResearchGate](#)



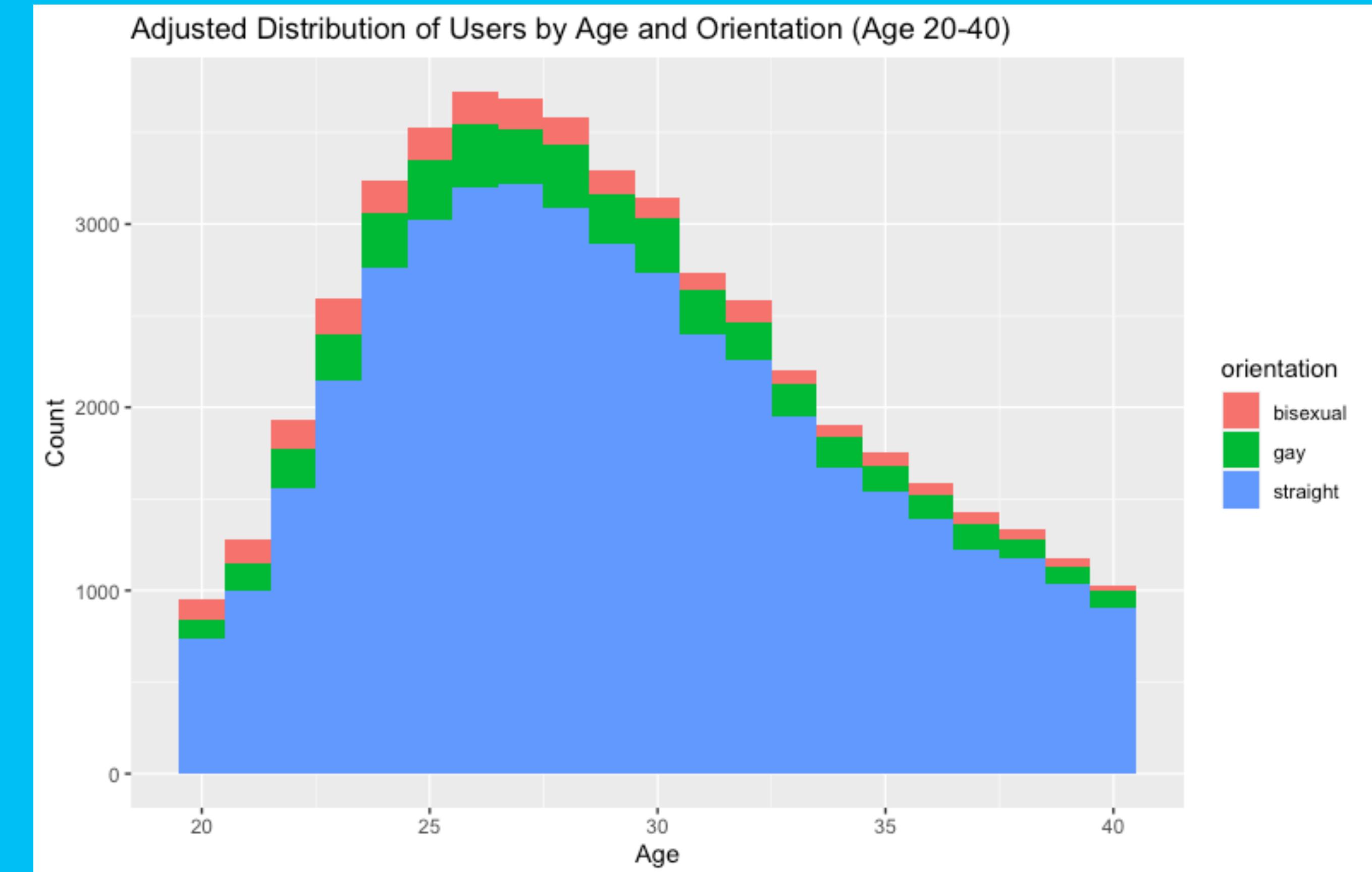


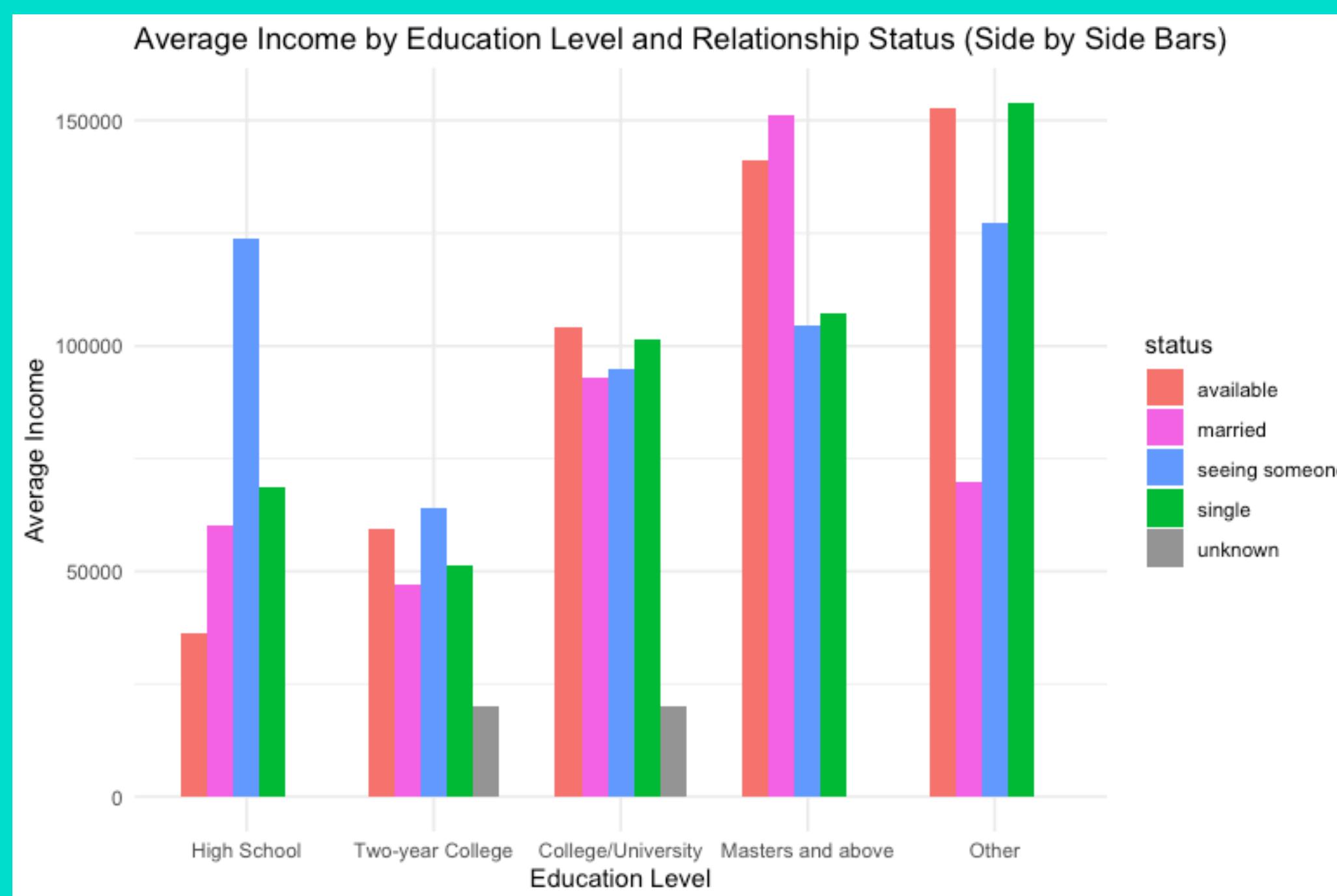
# DATA CLEANING & PRE-PROCESSING

- ▶ Loaded profiles, LatLon, addr, and sharedCensus datasets.
- ▶ Merged datasets on the 'location' column.
- ▶ Removed age outliers (e.g., 109, 110).
- ▶ Corrected data points like "drinks desperately".
- ▶ Handled missing values.
- ▶ Recategorized 'education' for clarity.
- ▶ Set boundaries for US geographic data.

# AGE & ORIENTATION

- The **age distribution peaks between 25-30**, indicating most users are within this age range.
- Most users identify as **straight**, followed by gay and bisexual.





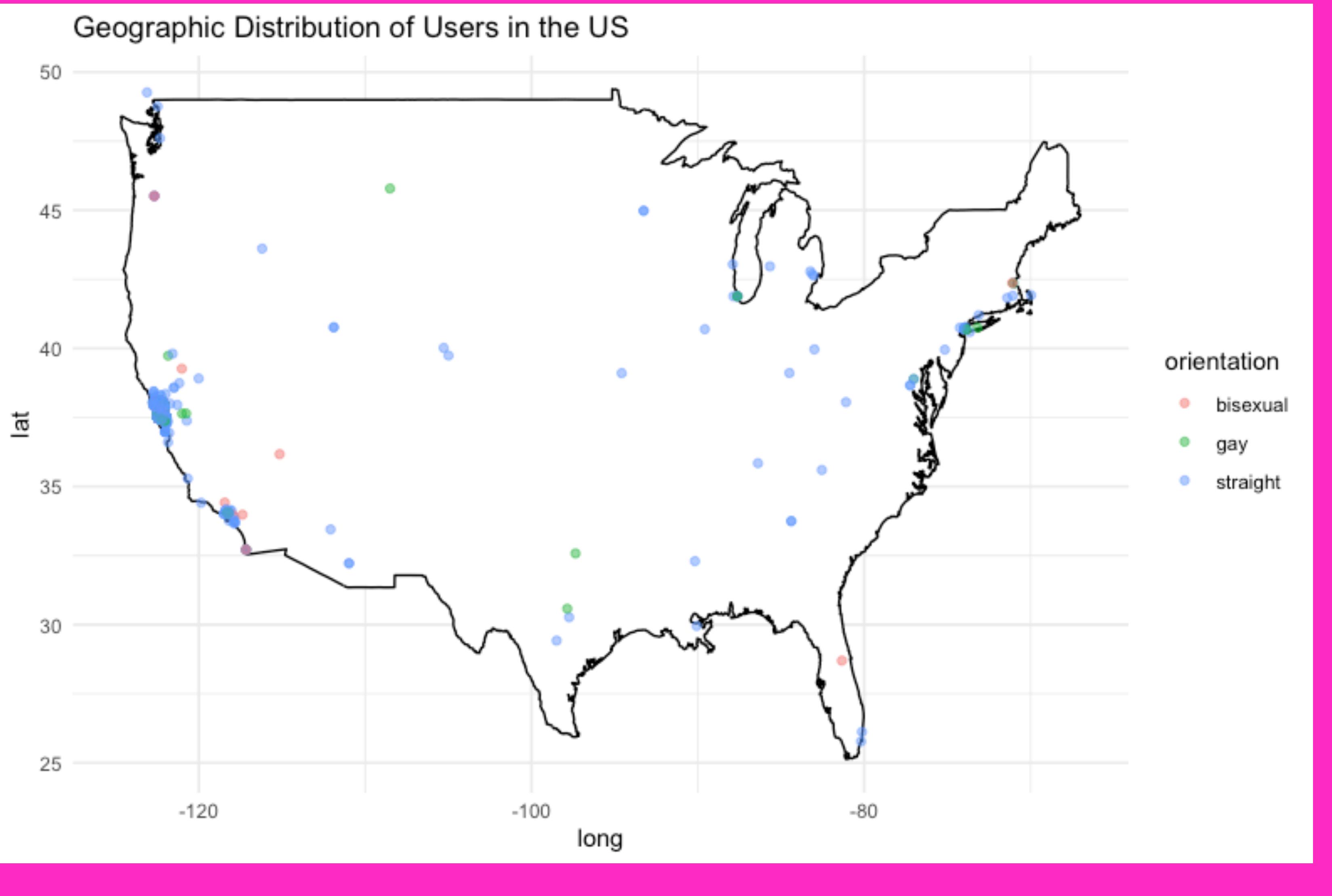
# INCOME, EDUCATION & RELATIONSHIP STATUS: UNCOVERING PATTERNS

## Income & Relationship Status:

- "Available" and "Single" users have higher average incomes
- Potential Strategy: Target marketing campaigns toward "Available" and "Single" users due to their higher spending potential.

## Income & Education:

- "Masters and above" generally have a higher income than those with only a "High School" education.
- Potential Strategy: "Available" and "Single" users, especially those with higher education, can be the key target audience for premium offerings or campaigns.



- Concentrations of users in certain regions.
  - Hotspots can inform targeted advertising, events, or promotions in those areas.
- Orientation and Location:
  - Geographic distribution of orientations (bisexual, gay, straight) can help customize marketing messages.
  - Regions with more gay and bisexual users can have LGBTQ+-friendly campaigns or events.

# GEOGRAPHICAL DISTRIBUTION

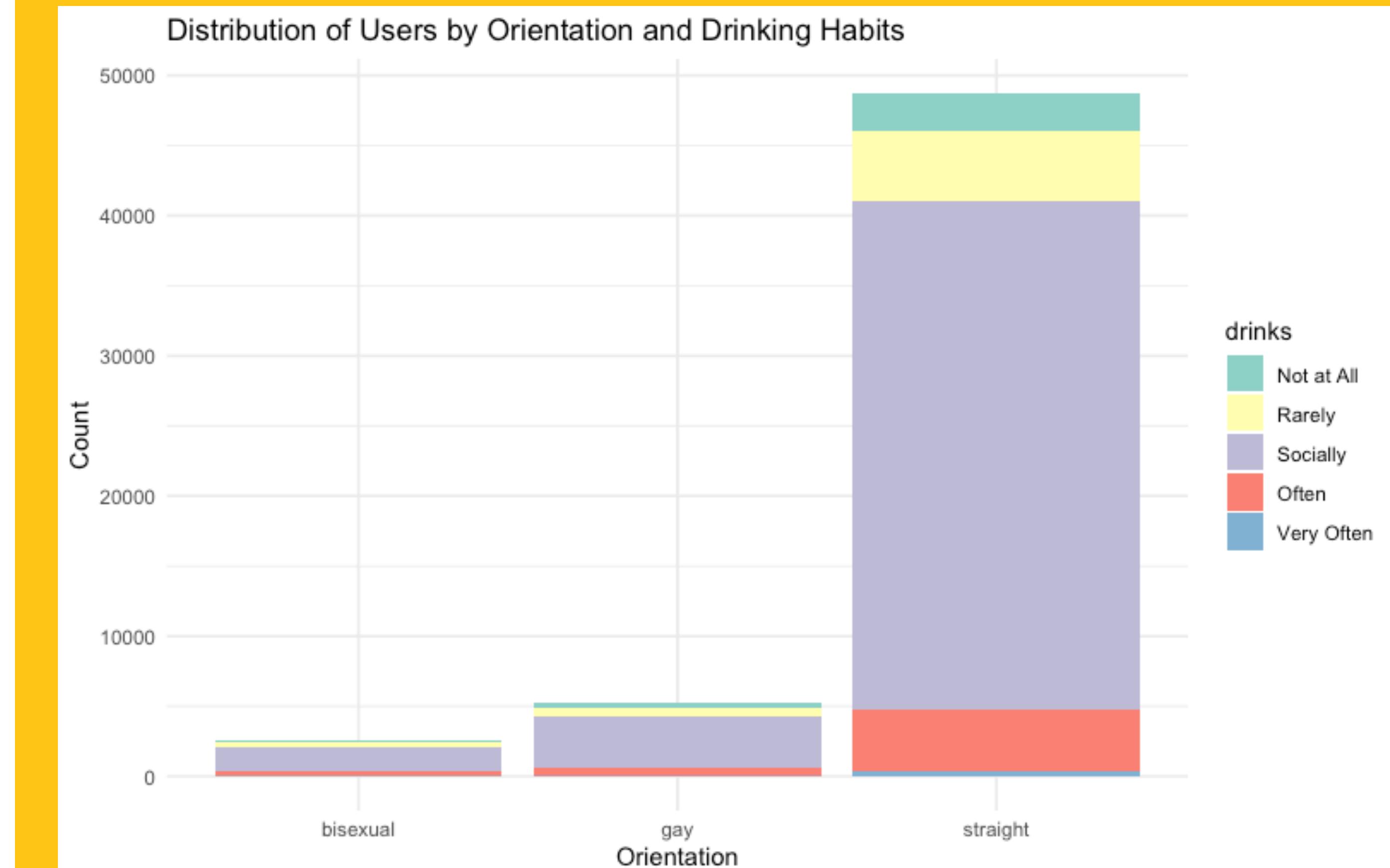
# ORIENTATION & DRINKING HABITS

## ● Social Drinkers Predominant:

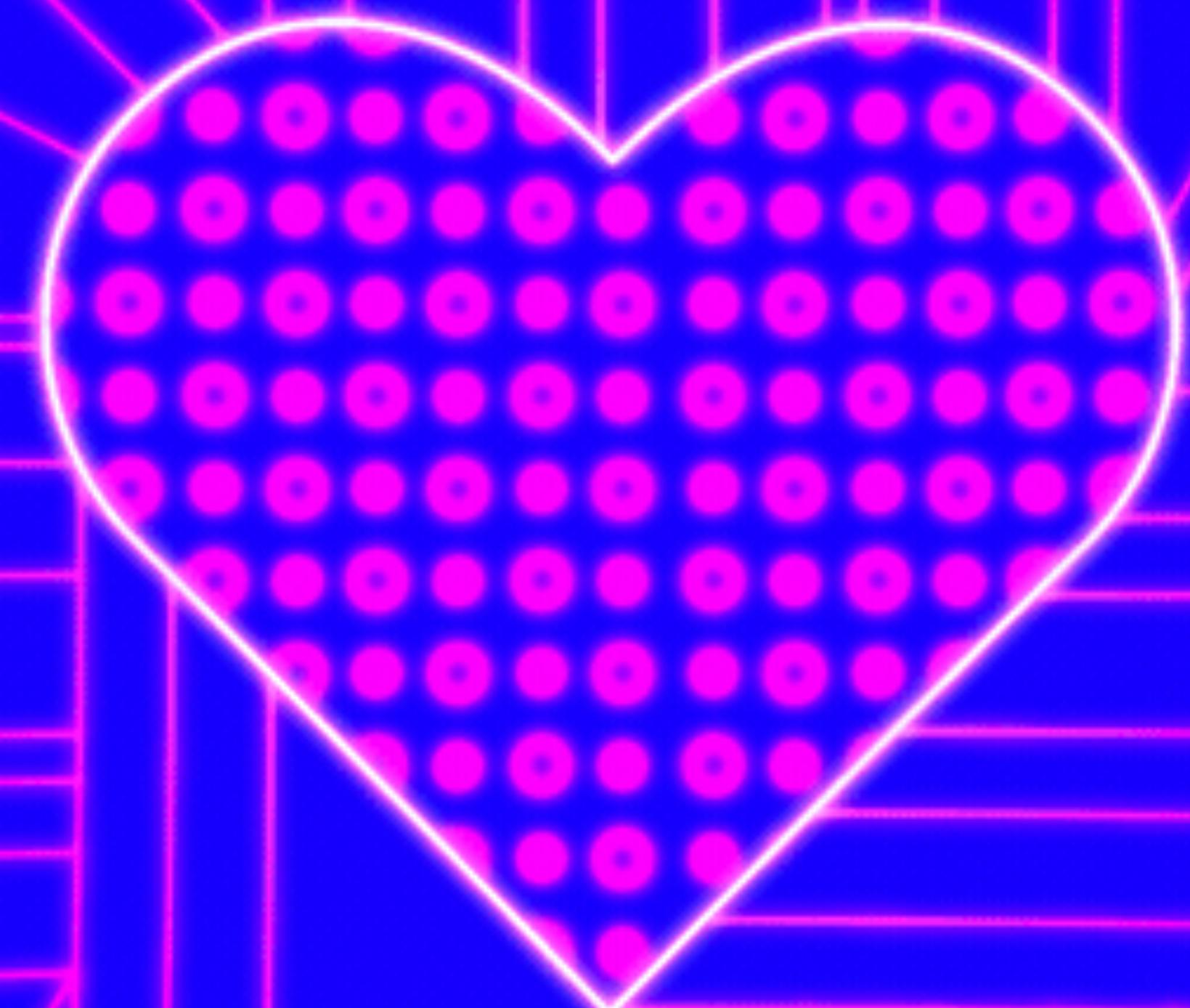
- A significant majority of straight users identify as social drinkers. Tailored campaigns highlighting social settings or events might appeal to this segment.

## ● Lifestyle Choices among LGBTQ+:

- A noticeable portion of gay and bisexual users abstain from alcohol or drink rarely. Understanding these lifestyle choices can lead to more inclusive marketing efforts.



# KEY TAKEAWAYS



## Age & Orientation:

- > The majority of users fall within the 25-30 age range.
- > Orientation distribution: Most users identify as straight, followed by gay and then bisexual.

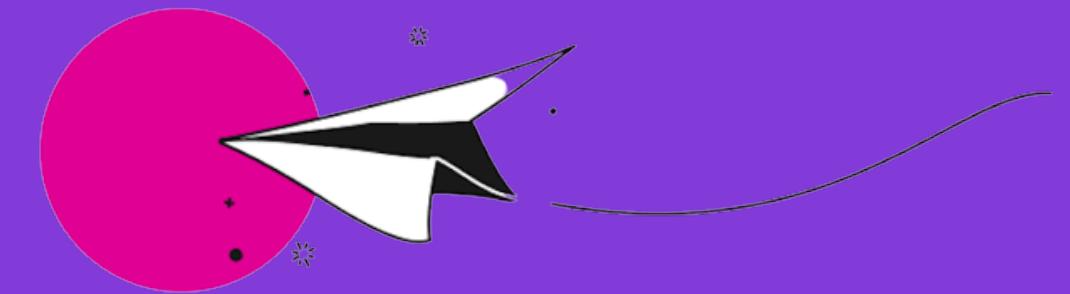
## Income Dynamics



- > "Available" and "Single" users showcase higher average incomes, hinting at their potential spending power.
- > Education is pivotal in income; advanced degrees often lead to higher earnings.

## Geographical Concentration:

- > Certain states or cities exhibit higher user concentration, suggesting potential for geographically targeted campaigns.
- > Areas with denser concentrations offer opportunities for local events or partnerships.



## Orientation & Lifestyle Choices:

- > Most straight users predominantly drink socially, presenting opportunities for social events or experiences.
- > A sizable segment of gay and bisexual users abstain from alcohol, highlighting diverse lifestyles within our user base.





# OVERALL IMPLICATIONS:

- Segment-focused campaigns can enhance user engagement and response rates.
- Uncovering diverse lifestyles within our community allows for more inclusive and targeted marketing.
- Leveraging geographical data can optimize localized campaigns and partnerships.



## FINAL NOTE

Maximizing outreach means deeply understanding our users – from socio-economic factors to lifestyle choices. These insights pave the path for nuanced marketing campaigns.

# QUESTIONS & FEEDBACK



# ACKNOWLEDGMENTS & REFERENCES

**Data Source: ResearchGate**

**R-Studio Copyright (C) 2022 by Posit Software, PBC**

**Includes Images from OkCupid**

**Special thanks to the OkCupid team.**

