The State of Dating Apps Report

Sabrina Sutlief, Jennifer Lee, Micah Chiang, Xiaoqian Xiang, Aditya Arora

We are interested in exploring patterns in online dating behavior. While this area of study is rich in anecdotal and cultural relevance, many available datasets in this domain are often sparse or limited in scope. We have a small, self-collected dataset of about 20 responses gathered out of personal interest in 2024. To expand this into a meaningful project and conduct a more robust analysis, we plan to compare trends in this sample with those identified in larger, publicly available datasets, such as those hosted on platforms like Kaggle.

I. INTRODUCTION

(added by Sabrina)

In the digital age, dating apps have transformed the way people form romantic connections. Platforms like Tinder, Bumble, Hinge, and OkCupid have not only changed dating behaviors but also generated large amounts of user data, offering unique insights into modern relationship trends. This report aims to explore patterns and trends within dating app data using both a small, personally collected dataset and publicly available datasets. By analyzing variables such as match preferences, profile characteristics, and messaging behavior, we hope to identify key factors that influence dating outcomes.

In addition to data analysis, this project draws upon insights from two well-regarded books: How to Not Die Alone by behavioral scientist Logan Ury, and Dataclysm by OKCupid co-founder Christian Rudder. These texts offer psychological and algorithmic perspectives on online dating, which we use to contextualize our findings.

The ultimate goal is to highlight consistencies and inconsistencies across datasets, generate new questions for further research, and offer both data-driven and practical perspectives on what helps or hinders romantic success online.

II. METHOD

(added by Sabrina)

Step-by-Step Plan: Initial Data Exploration Identify relevant large-scale online dating datasets. Analyze how different variables (e.g., appearance, occupation, education) correlate with user engagement metrics like number of likes or matches. Conduct text analysis on user-generated profile descriptions to detect thematic patterns or linguistic trends.

Comparison with Personal Dataset Apply a similar analysis pipeline to our smaller, self-collected dataset. Compare the frequency and context of self-descriptors, values, and preferences across both datasets. Identify areas of overlap and divergence

Use of Synthetic Data Generate synthetic data to simulate additional user profiles. Test whether observed trends hold across simulated scenarios. Utilize findings to formulate hypotheses for further study on optimizing user interactions on dating platforms.

(Optional Extension) If time permits, we will experiment with advanced techniques such as clustering to group user types or preferences. We also plan to compare our findings with general population statistics on dating in the U.S., investigating whether online dating diverges from offline trends.

III. RESULTS AND DISCUSSION

A. Dating Preferences in an In-Person Era

(Added by Micah)

Since dating apps take away emotional and physical connection during first-impressions, we wanted to see what dating was like when people met up in person for their first impressions of each other. Using a dataset called "Speed Dating", I plotted variables that give us information on speed dating participants, their preferences, their opinions on their date, and if they were a match.

- **Attractiveness:** While many participants claim that physical attractiveness is not a major factor in their dating decisions, the data shows that individuals who rate themselves higher in attractiveness tend to receive more matches. Those who place greater importance on attractiveness also tend to match more frequently.
- **Intelligence:** People who see themselves as more intelligent usually get more matches. This might suggest that confidence can be an attractive quality to many participants.
- **Race:** Two participants being the same race doesn't seem to have an effect on whether they match. Even among participants who claimed that being the same race is important, there wasn't a big difference in match rates.

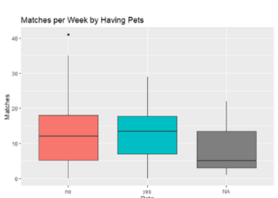
The data shows that looks and intelligence have a strong impact on match rates. This could be due to a participant's confidence and the way they present themselves. Race doesn't seem to matter much, even to those who say it's an important factor.

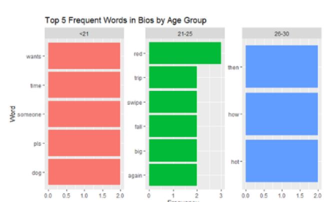
B. Personal Findings

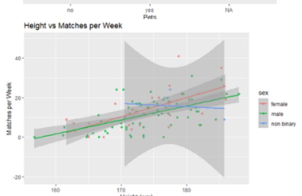
(added by Sabrina)

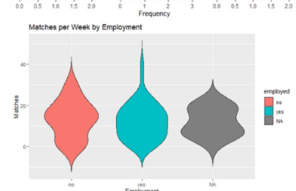
Our goal was to look at a personal data set that Sabrina collected to see if extra specificity in certain qualities in the dating apps could improve our ability to predict likes. The data set that was created was small and only included about 100 samples but it showed that the diversity in other aspects of a dating profile such as their bio could have influence over how a person was perceived on the apps.

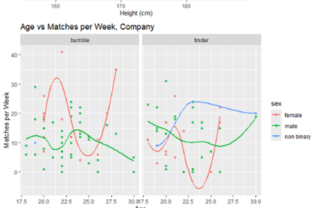
- **Sample Size:** 99 profiles.
- **Sex Distribution:** 68 male, 27 female, 4 non-binary.
- **Age:** Range 18–30 years, mean 22.4 ± 2.9 years.
- **Height:** Range 156.7–188.0 cm, mean 174.3 ± 6.4 cm.
- **Matches per Week:** Range 0–41, mean 12.5 ± 8.5.
- **Swipes per Week:** Range 0–1211, mean 260 ± 200.
- **Match Efficiency:** Average conversion rate 4.9%, with a maximum of 30%.
- **Exercise:** 46 yes, 30 no, 15 sometimes, 8 unknown.
- **Alcohol Use:** 48 yes, 32 sometimes, 19 no.
- **Employment:** 58 employed, 26 not employed, 15 unknown.
- **Bio Length:** Mean 5.9 ± 2.9 words, with a maximum of 13 words.

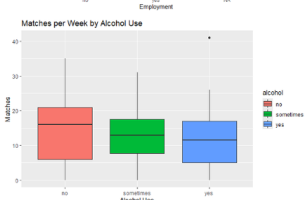












C. Expression of Effort and Personality in Profiles

(Added by Aditya)

To explore self-expression on dating platforms, I developed an "effort score" based on essay presence and word count across OkCupid profiles. I analyzed how this score varied across gender, age, job sector, and personality markers.

- **Gender:** Median effort is similar between men and women, though male users show greater variance and more extreme outliers—suggesting diverse engagement levels.
- **Age:** Users aged 36–45 and 56+ put slightly more effort into profiles, hinting at greater seriousness or reflection in older age groups.
- **Job Sector:** Creative, academic, and tech users had the highest effort scores, while business-related fields (sales, finance) showed lower averages—possibly reflecting differing communication styles.
- **Personality Traits:** Profiles referencing emotional, intellectual, funny, or creative qualities showed consistently higher effort. Emotional and intellectual traits in particular correlated with the most effortful profiles.

These trends suggest that personality and life context shape how much users invest in their profiles, offering insight into self-presentation behaviors in online dating.

D. User Ratings, Sentiment, and Developer Engagement

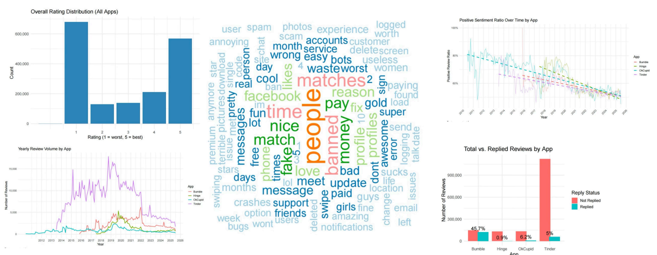
(Added by Xiaoqian)

To understand user satisfaction and platform responsiveness, we analyzed five key dimensions of app reviews across Tinder, Bumble, Hinge, and OkCupid:

- **Rating Distribution:** Review scores are highly polarized. The most common rating is 1 star, followed closely by 5 stars, suggesting users tend to leave feedback only when strongly dissatisfied or satisfied.
- **Word Frequency (Review Themes):** Common keywords across reviews include *people*, *matches*, *pay*, *fake*, and *banned*, highlighting recurring concerns around authenticity, monetization, and account moderation.
- **Sentiment Over Time:** All four apps exhibit a declining trend in positive review ratio. OkCupid started strong but dropped quickly, while Hinge maintained higher sentiment for longer. This suggests growing discontent across platforms.

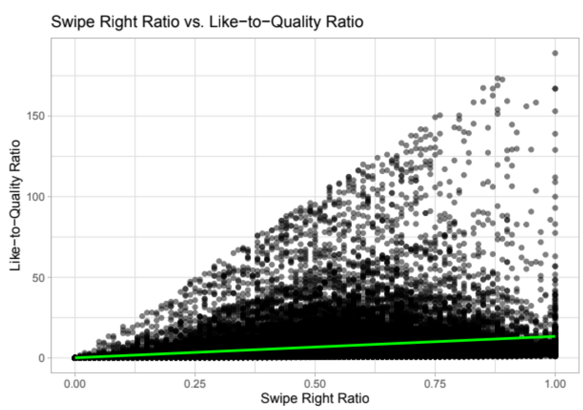
- **Review Volume Trends:** Tinder consistently received the highest review volume, peaking in 2018–2019. Review counts across all apps declined post-2020, possibly due to user fatigue or platform stabilization.
- **Developer Responsiveness:** Bumble stood out by replying to 45.7% of its reviews, while Hinge, Tinder, and OkCupid replied to less than 6%. This suggests Bumble places more emphasis on customer engagement.

These patterns reveal rising user frustration and shifting sentiment across dating platforms. While dissatisfaction appears widespread, active developer responses—especially from Bumble—may help rebuild trust. This analysis highlights evolving user expectations and strategic opportunities for app reputation management.



E. Linear Model

(Added by Sabrina and Jennifer) Part of the assignment requires a linear model, so here is our analysis:



There's a positive trend — users who like more tend to have higher like-to-quality ratios, meaning they receive fewer matches per like. The correlation coefficient was r = 0.67, indicating a moderate-to-strong positive relationship between swipe right ratio and like-to-quality ratio. The R^2 value was 0.45, meaning that 45% of the variation in like-to-quality ratio can be explained by swipe right ratio. This suggests that selectivity plays a substantial role in matching efficiency, although other profile, behavioral, or algorithmic factors also contribute to the remaining variation.

IV. CONCLUSIONS

Our analysis shows that online dating success depends on more than just swiping volume. Men tended to swipe more but were less efficient, while women often secured more matches with fewer swipes. Age also mattered: users 21–25 were the most active, while those 26–30 wrote longer, more detailed bios. Lifestyle factors such as exercise and employment correlated with higher match counts, and bio text revealed recurring themes of humor, authenticity, and shared interests.

Overall, both demographics and self-presentation play meaningful roles in dating outcomes. While limited in size, our personal dataset reflected patterns seen in larger studies, reinforcing that effort quality, lifestyle cues, and expressive profiles shape success on dating apps.

V. AI USAGE STATEMENT

We affirm that generative AI tools (including but not limited to ChatGPT, Copilot, Gemini, and Claude) were not used at any stage of this project. This includes idea generation, code writing, debugging, documentation, or any other form of assistance. All work presented here is the original product of our group members, developed through our own reasoning, collaboration, and programming efforts.

REFERENCES

1. Rudder, C. (2014). Dataclysm. HarperCollins Publishers.
2. URY, L. (2021). How to not die alone: The surprising science of finding love. PIATKUS BOOKS.