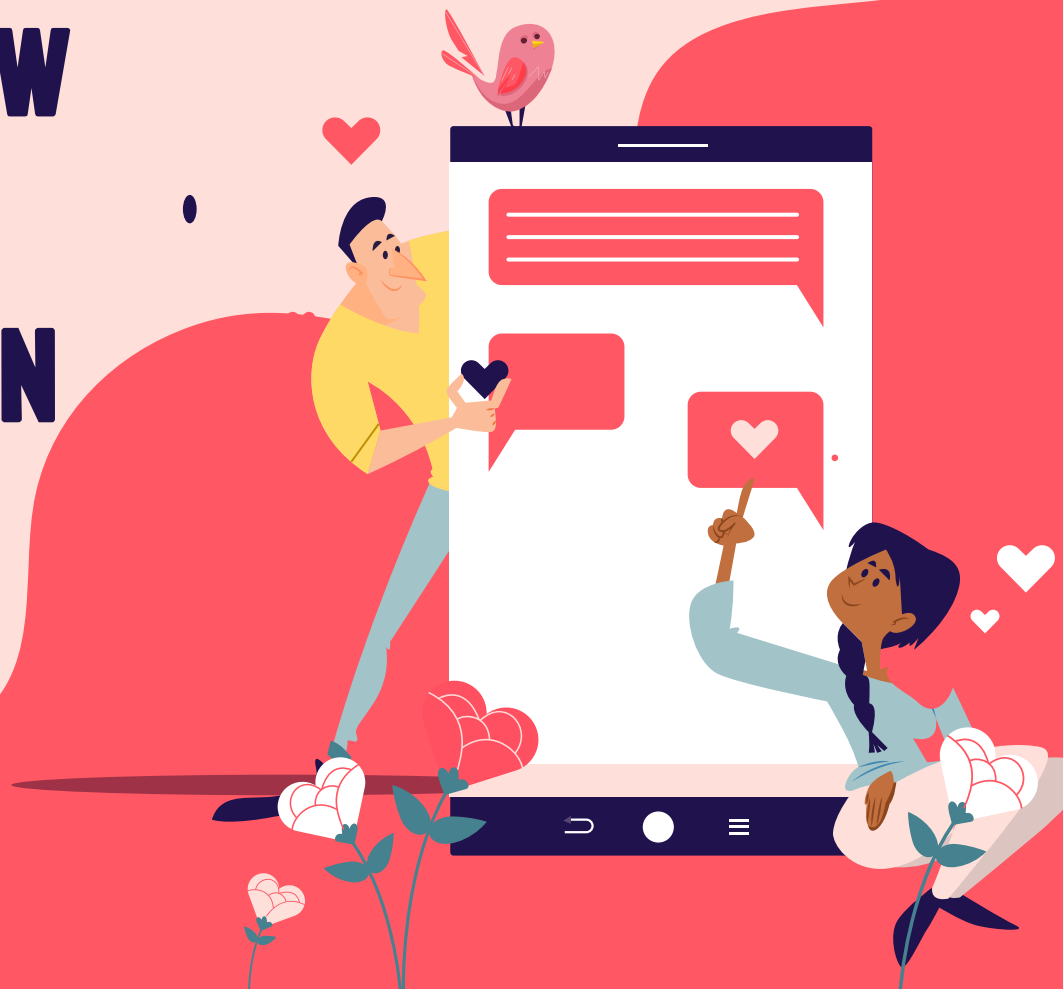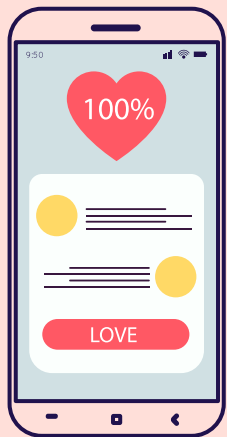# DATING APP REVIEW ANALYSIS & RECOMMENDATION

Columbia University
Spring 2022
APAN 5205 Final Project
Group 5
Huiyu Jiang, Ning Yang,
Yuting Zhu, Zixuan Wang

Background & Research Question

Over **323 million** people worldwide are using dating apps. Tinder, Bumble and Hinge are the top three players in the market, occupied a **69%** worldwide market share in total.

"Is terrible- there are 50 different tabs to show you the same people- but in different ways and it never syncs- go back to chronological"

—USER###

100%

LOVE

In the research, we aims to analyze comments and ratings of dating apps and **investigate what features** of the dating apps **users care the most about.**

The research means to **help the dating app companies** to understand the users' demand and favor to improve the satisfaction by developing the apps.
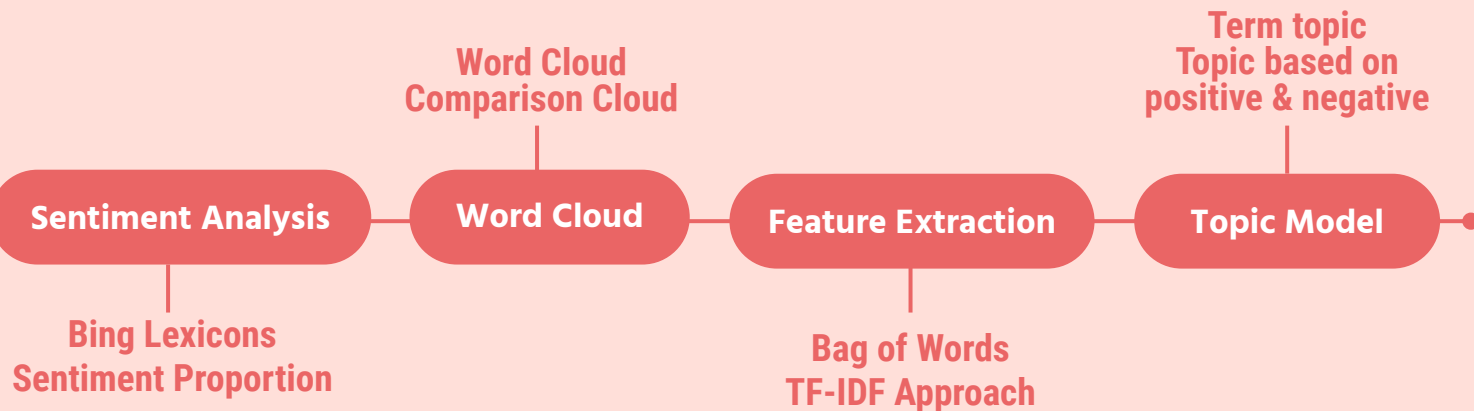
Data Source & Analysis Techniques

DATA EXTRACTION

- 'Apprler' package to extract review and rating data from Apple App Store.
- Collected reviews were posted from January 20th 2022 to April 14th 2022.
- Nearly 1000 reviews for each app.

DATA PREPARATION

- Create Corpus
- Clean Text (particularly, remove meaningfulness stopwords)
- Tokenize - Document Term Matrix

Word Cloud
Comparison Cloud

Term topic
Topic based on positive & negative

Sentiment Analysis — **Word Cloud** — **Feature Extraction** — **Topic Model**

Bing Lexicons
Sentiment Proportion

Bag of Words
TF-IDF Approach

SENTIMENT ANALYSIS

We use bing lexicons to yield sentiment score for each word, determining the positive and negative words of user reviews.
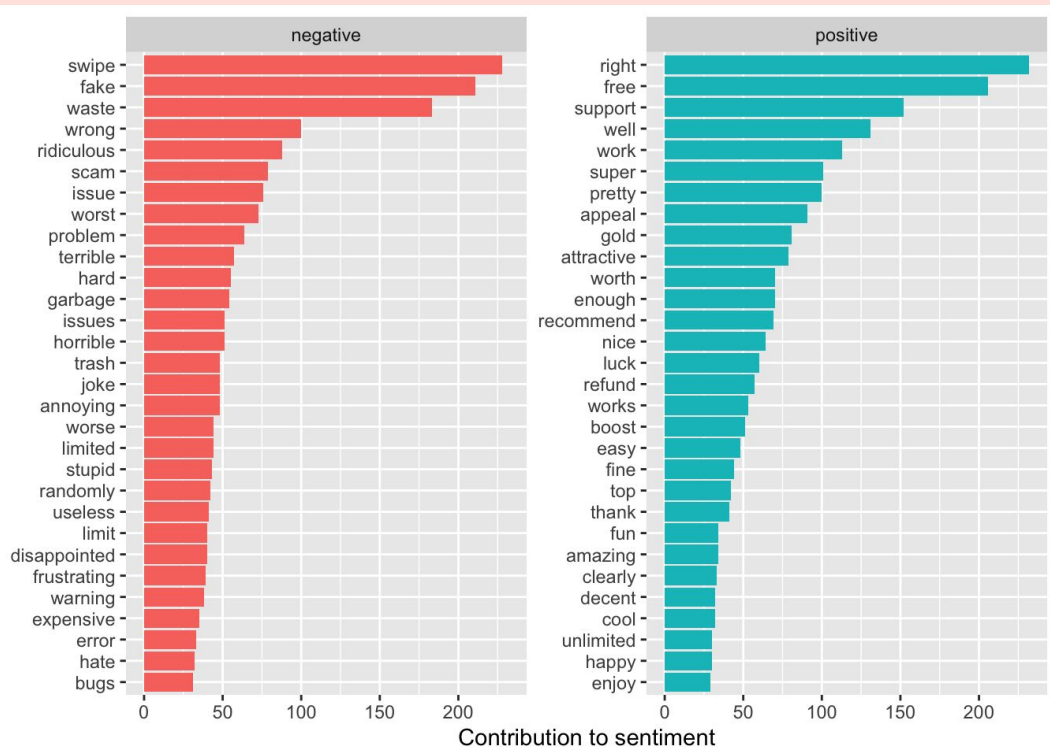
Proportion of positive (and negative words) for each rating

Words such as "swipe", "fake", "scam","expensive", and" error" show negative sentiments.

Features that need to be improved: user profile accuracy audit, pricing, system operation, etc.

positive

negative

ly corresponds to

right, free - waste, fake

attractive, work, recommend - wrong, limited, issue

...ole's

...ey, dating, time

...se , someone, free

...swipe, message, waste

FEATURE EXTRACTION (TF-IDF)

- **TF (term frequency):**
  Term frequency works by analyzing the frequency of a particular term relative to the document.

- **IDF (inverse document frequency):**
  Inverse document frequency works by looking at how common (or uncommon) a word is in a corpus.

Next, we built **the decision tree** diagram for prediction. The selected tree contains **4 variables and 4 splits.**
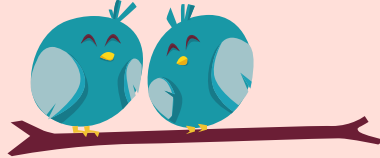
CART is based on the known probability of occurrence of various situations, by forming a decision tree to find the probability that the expected value is greater than or equal to zero.
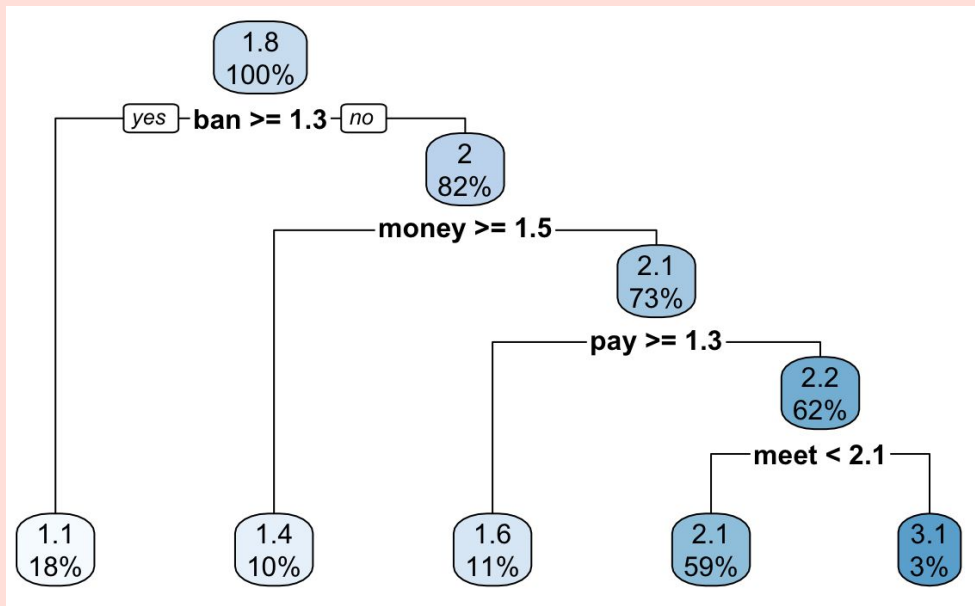
Topic Modeling(LDA)

Topic modeling is an useful way to discover topics that customers care about from reviews. It is an extension to detect customer attitudes based on sentiment analysis. We can calculate and visualize the probability that each review is associated with each topic.
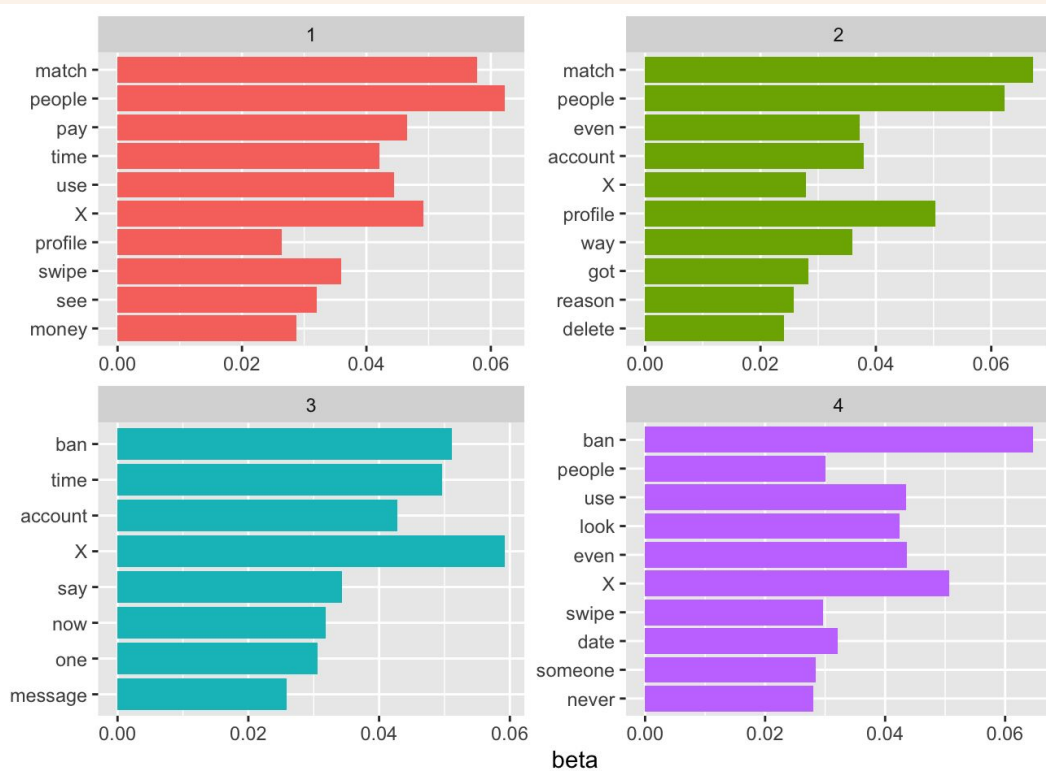
LDA analysis and categorized all words into four topics.
Topic 1:Payment system, swipe feature
Topic 2: User Accounts, Profiles
Topic 3: User account, message feature, Prohibited Features
Topic 4: Prohibited Features, swipe feature

TOPIC MODEL BASED ON POSITIVE AND NEGATIVE

Two-topic models of positive and negative reviews are created to find out which aspects of the customer experience come up in positive and negative reviews.

Positive Reviews:
Topic1: Customer support
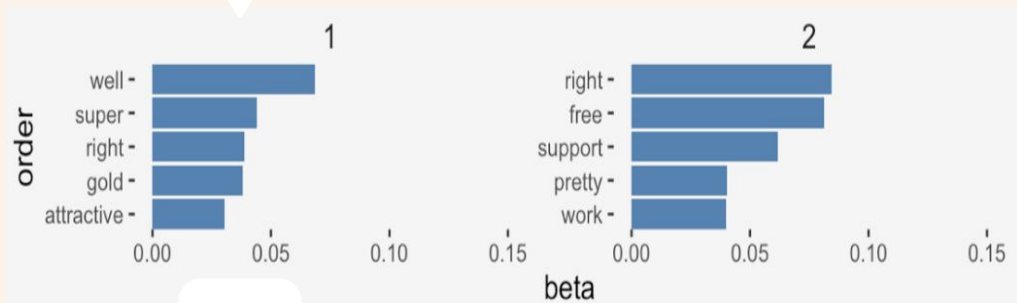Topic2: Free and premium features (Gold subscription)

Negative Reviews:
Topic1: User Information authenticity
Topic2: Swipe design of the app

Positive review topics

Negative review topics

CONCLUSION AND RECOMMENDATIONS

Conclusion

Positive emotion: The positive emotions of users come from the commercial use of social software, and the functions of matching users, promoting social interaction with strangers, customer service support, free download and use.

Negative emotion: The negative sentiment comes from other users' untrue information, receiving spam, the design of the app's sliding function, paid-for-use features of apps, payment systems, and account-related activities.

Recommendation

- Enhance users' identity reviews to prevent identity theft and fraud
- Improve the security optimization of the background, especially for specious words
- Develop the questionnaire that is sent to users in the beginning to upgrade the satisfaction of the "swipe" and match function