BUMBLE

GOOGLE PLAY STORE - APP REVIEWS

Sameera
June 2022

INTRODUCTION

- Bumble is a dating application.

- Second most popular dating app, after tinder.

- Profiles of potential matches are displayed to users, who can **swipe left** to reject a candidate, or **swipe right** to indicate interest.

- Users can sign-up using their phone number or Face Book profile.

- Paid subscription (3 types) or free subscription available.

- The company reports having over 55 million users in 150 countries as of 2019.

GOAL

- To analyze the reviews –
  - User reviews give a chance to make the application better.

- Topic Modeling -
  - Divide the reviews into clusters of different topics
  - Know the percentage of each topic

DATA

- Data Set
  - Initially there were – 110,031 entries
  - 10 columns – reviews, ratings, date, …
  - Feature of Interest – **content**, which has reviews by users
  - Reviews were in different languages with script in English and Non-English
  - Extracted only the reviews in English language using **langdetect**
  - 89,472 rows – Only English language entries

TOOLS

Pandas – Clean, Explore and Feature Engineering

Scikit-Learn – Build different Classification models and perform cross validation, variable selection and regularization

Matplotlib/ Seaborn – Visualizing data exploration, modeling and results

Python 3.8.5 – to run all of the above

nltk - Natural language toolkit, to work with human language data.

Langdetect – to detect the language of the text

Scattertext – to detect the positive and negative words

RESULTS

- Five different topics were discovered –

  - Bad reviews About Paid Subscriptions --   69%

  - Profile Match   --   22.7 %

  - Good Reviews About The App   --   3.4%

  - Good Reviews About Other Users --   2.3%

  - Easy To Use   --   2.2%

- Over-all more negative reviews were discovered than positive reviews using **Scattertext**

  - **Positive** document count: 31,428 and word count: 563,035

  - **Negative** document count: 58,078 and word count: 1,947,201

TOP 30 WORDS OF EACH TOPIC

Topic: Bad Reviews For Paid Subscriptions
match, app, like, get, time, message, profile, people, pay, woman, even, one, first, money, waste, see, swipe, make, day, never, fake, account, want, got, work, would, month, back, many, say

Topic: Profile Match
great, app, people, meet, love, new, best, way, friend, met, awesome, far, work, concept, nice, meeting, really, experience, worst, idea, woman, lot, think, connection, recommend, amazing, thanks, user, place, interface

Topic: Good Reviews
good, app, pretty, far, experience, really, people, nice, concept, best, friend, meet, overall, love, site, met, luck, looking, interface, work, idea, application, used, meeting, awesome, date, new, enough, quality, thing

Topic: Good Reviews About People On The App
easy, use, fun, facebook, nice, people, simple, app, love, really, meet, lot, super, navigate, interface, friendly, free, set, friend, make, new, user, account, sign, site, way, feature, well, option, without

Topic: Easy To Use
better, tinder, cool, much, way, apps, people, meet, far, pretty, nice, quality, site, could, like, lot, best, new, others, love, experience, met, definitely, seems, woman, girl, would, really, fun, one
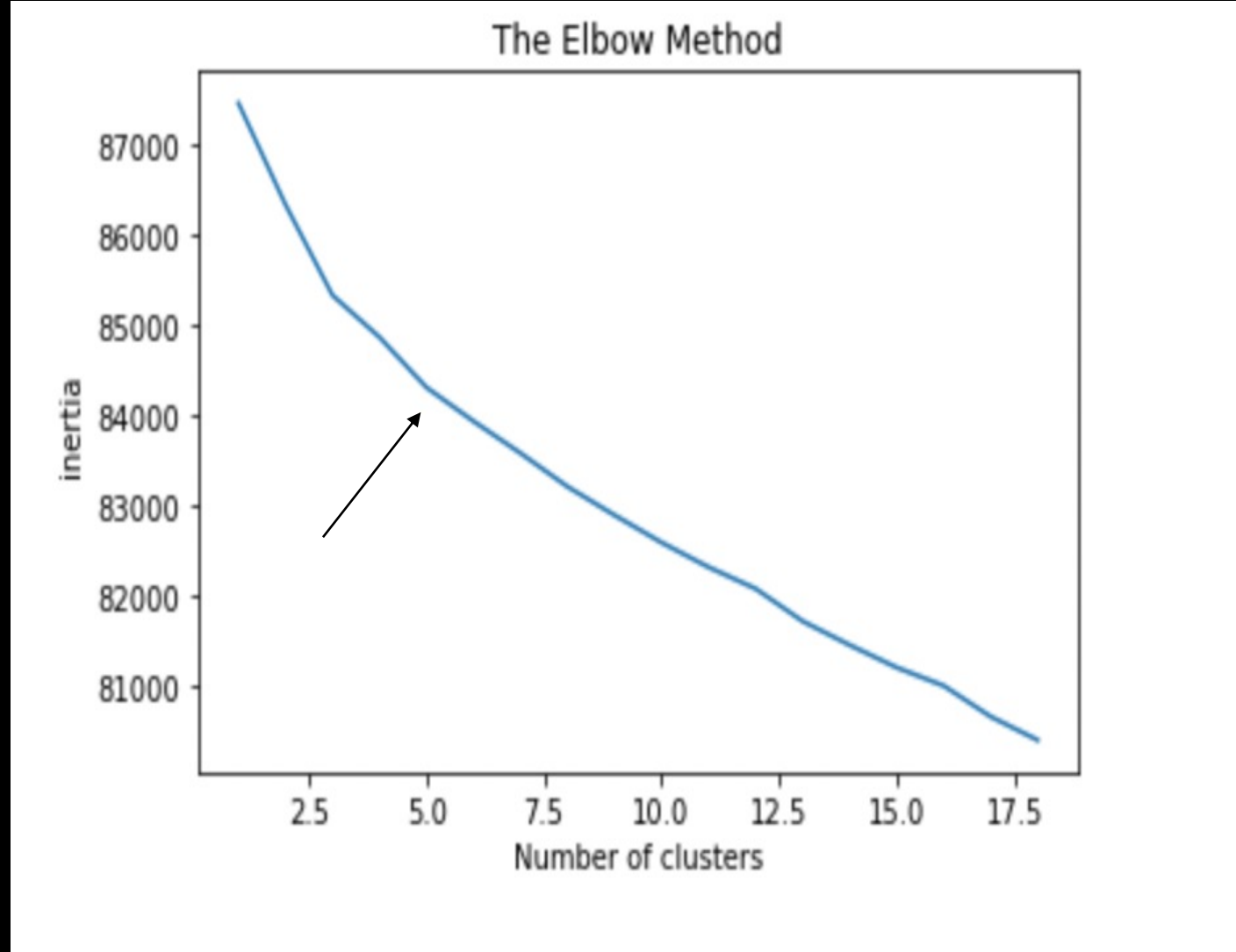
METHODOLOGY

- WordNetLemmatizer was used.

- Custom stop words were added to the stopword set.

- Transform the text data into a matrix, which represents the weights of each word in the corpus.

- Algorithms Used -
  - K-Means Algorithm
  - NMF – Non-Negative Matrix Factorization
  - Naïve Bayes

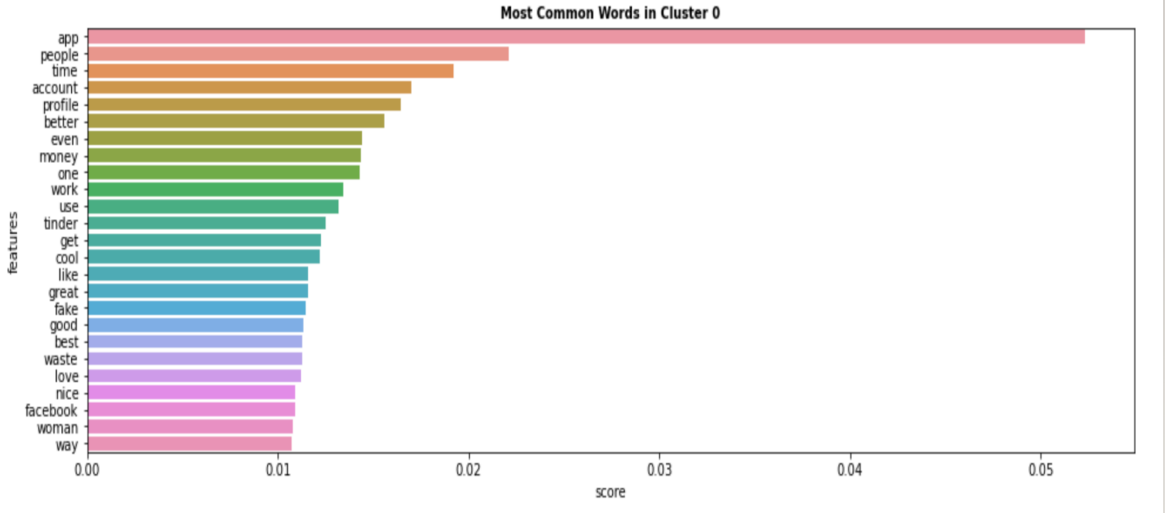- Use the elbow method to decide the number of clusters.

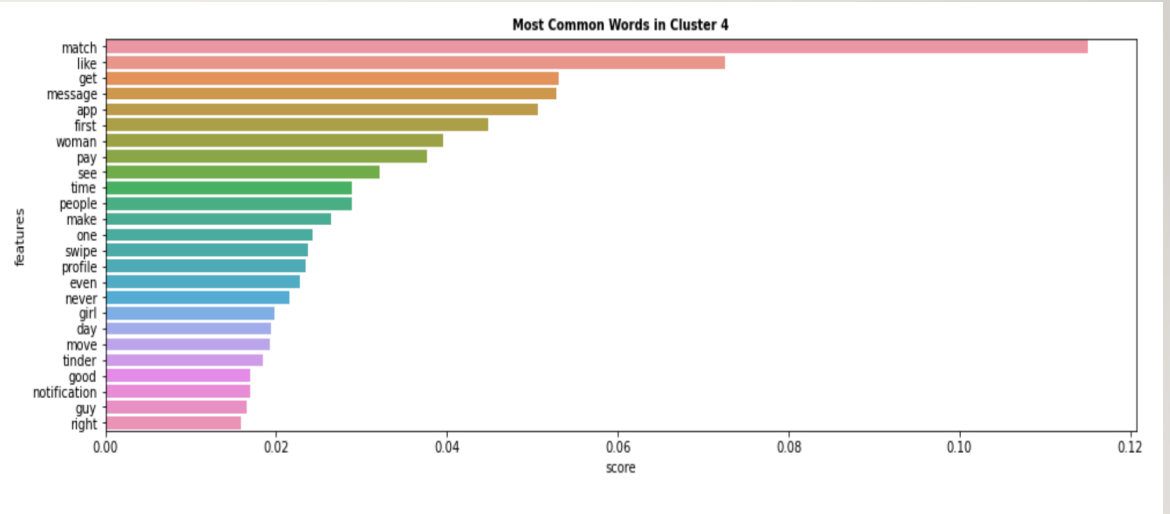# ELBOW METHOD –

NUMBER OF CLUSTERS VS INERTIA

VISUALIZATIONS

GRAPHS AND WORDCLOUD

(cluster0)-Bad Reviews For Paid Subscriptions

Most Common Words in Cluster 0

CLUSTER 0 – BAD REVIEWS ABOUT PAID SUBSCRIPTIONS

(cluster4)-Profile Match

Most Common Words in Cluster 4

CLUSTER 4 – PROFILE MATCH

(cluster1)-Good Reviews

Most Common Words in Cluster 1
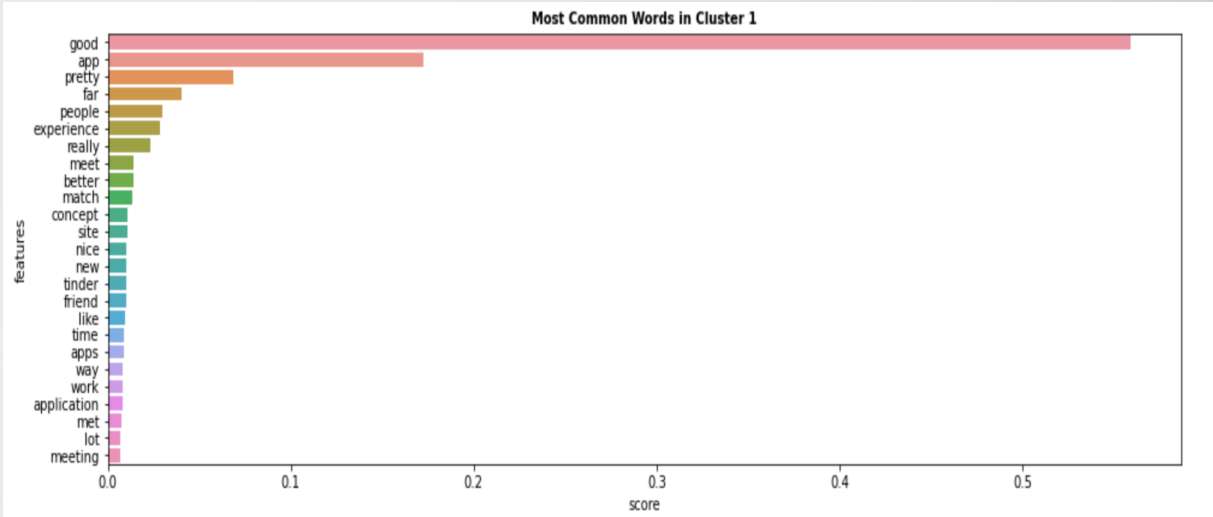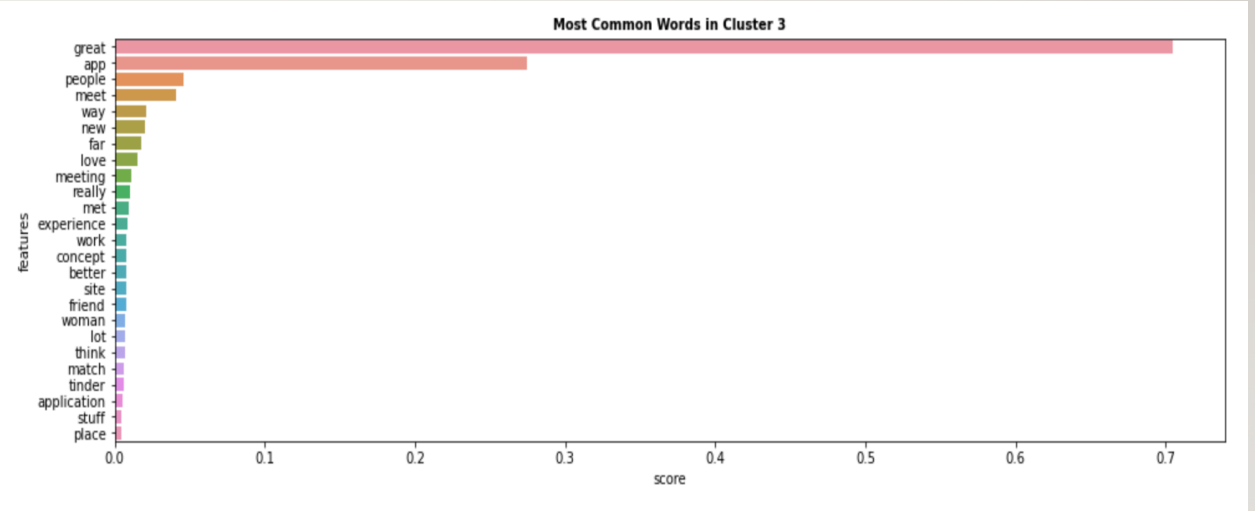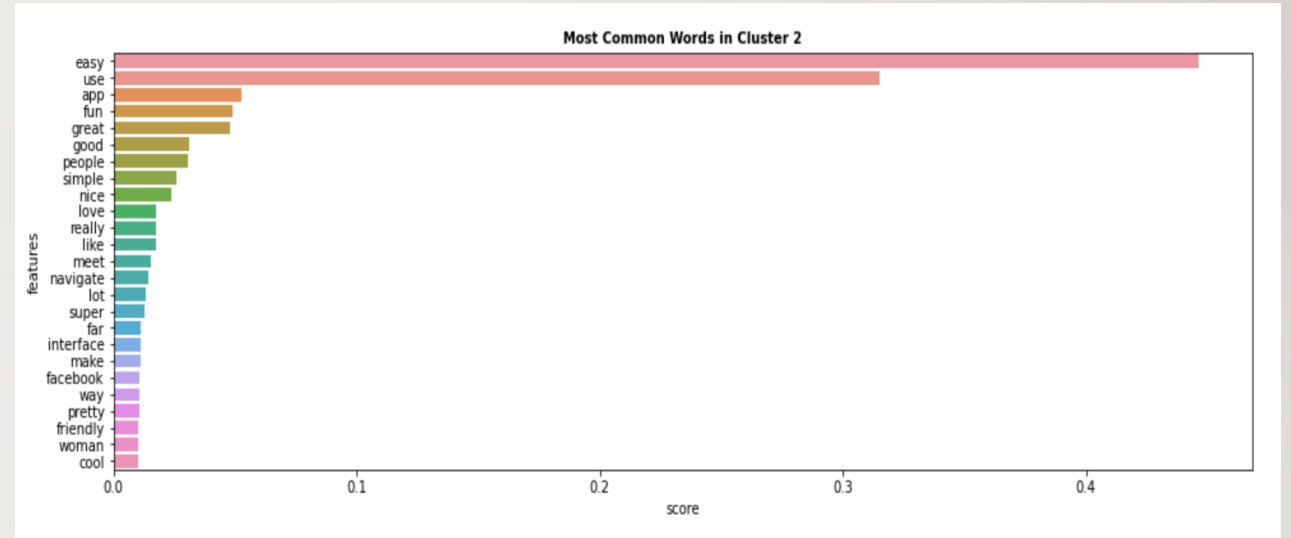
CLUSTER 1 – GOOD REVIEWS ABOUT THE APP

(cluster3)-Good Reviews About People On The App

Most Common Words in Cluster 3

CLUSTER 3 – GOOD REVIEWS ABOUT OTHER USERS

(cluster2)-Easy To Use

Most Common Words in Cluster 2

CLUSTER 2 – EASY TO USE

Axis labels:
- Vertical axis: Positive Frequency (Frequent / Average / Infrequent); also labeled "highly recommended", "phenomenal", "thankyou", "swell", "spark", "kool", "intimidating", "kiss", "thst", "keen", "worldwide", "rhyme", "capitalize"
- Horizontal axis: Negative Frequency (Infrequent / Average / Frequent)

Top Positive
- awesome
- fun
- amazing
- easy
- met
- cool
- thank
- best
- 👍
- quality
- meeting
- love
- interesting
- genuine

Top Negative
- worst
- waste
- cancel
- blocked
- useless
- charged
- facebook
- refund
- scam
- terrible
- deleted
- email
- banned
- log

Characteristic
- bumble
- app
- tinder
- swipe
- facebook
- swiping
- swiped
- apps
- bff
- fake
- match
- wa
- scam
- uninstalled
- ha
- useless
- blocked
- instagram
- doe
- bot
- conversation
- android
- backtrack
- matched
- messaged
- notification
- glitch
- dating
- paywall
- banned

Search the chart
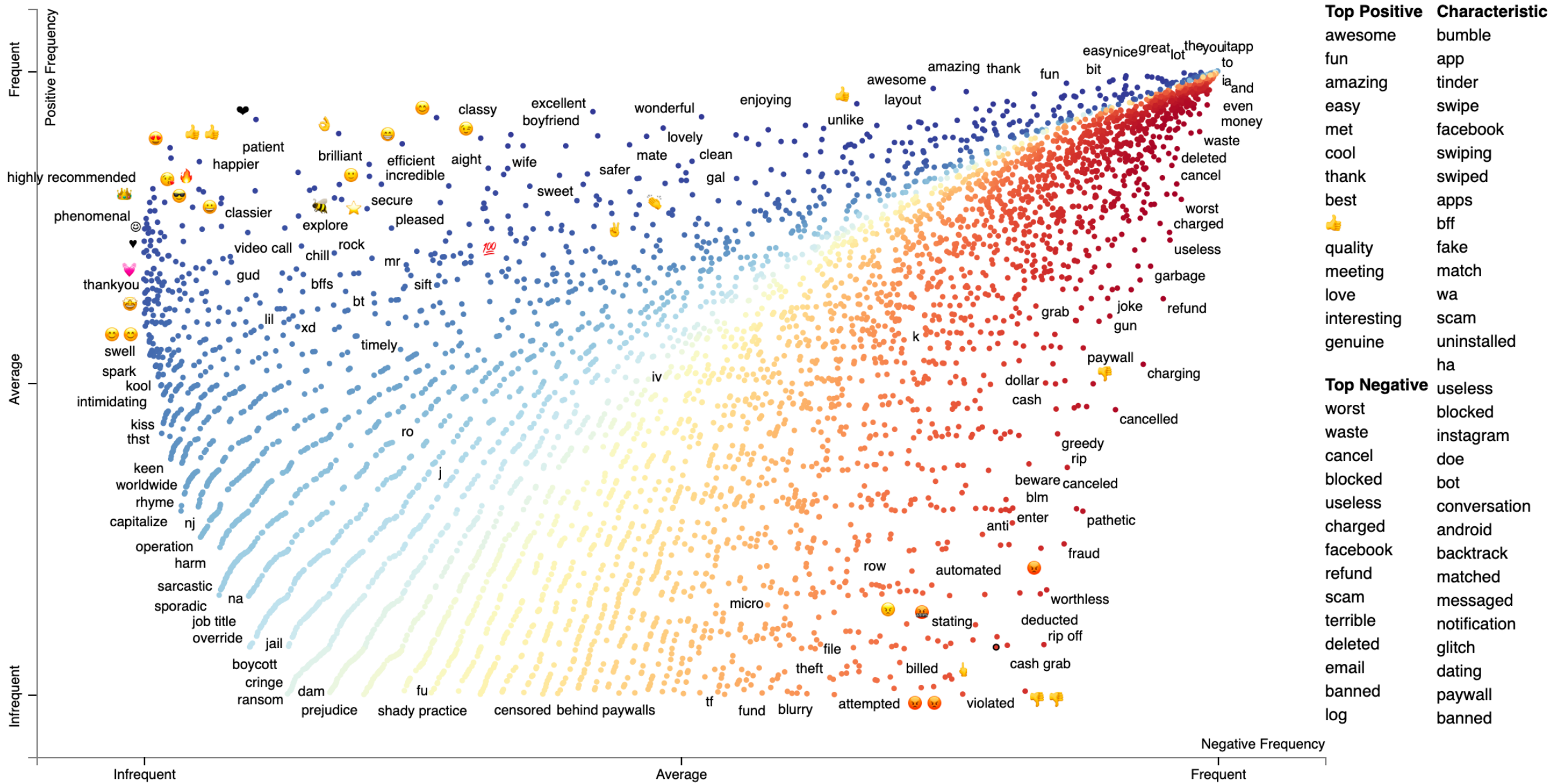
Positive document count: 31,428; word count: 563,035

NAÏVE BAYES MODEL

- Multinomial NB Score – 0.855

  - Target – Sentiment ('Positive' or 'Negative')

  - Positive - Rating of 4 and 5

  - Negative - Rating of 1, 2 and 3

VADER SENTIMENT

- Using SentimentIntensityAnalyzer method

- Best Review –
  - Score -   0.9062
  - Review - good app thanks dear woman you are beautiful

- Worst Review –
  - Score -   - 0.7997
  - Review - after being a premium user i m not able to login to my id pathetic experience no customer support is provided

- Vader is not very accurate, as it could not detect the word fraudulent, and this review has a score of 0.9.

english	text_lemmatized	cluster_name	month_year	sentiment	vader_score
1	after being a premium user i m not able to log...	0	2022-05-30	Negative	-0.7997
1	fraudulent app if you install a basic version ...	4	2022-05-30	Negative	0.9011
1	it s a lot better than hinge but it s still no...	0	2022-05-30	Negative	-0.6163

FUTURE WORK

- Translate the remaining 20 K non-English entries to English language

- Calculate the best Coherence Score for NMF model

- Retrain the models

- Use LSA and LDA topic modeling approach

- Deploy the model