

# Tinder Review Text Analysis

Presented by Group White:  
Xinde Huang, Chiayi Lin, Liang Chi Liu, Jessica Xu, Zijian Yan



**tinder**



# Contents

---

**01**



**Industry & Business  
Analysis**

**02**



**Goal**

**03**



**Implementation  
Roadmap**

**04**



**Models**

**05**



**Business  
Recommendations**

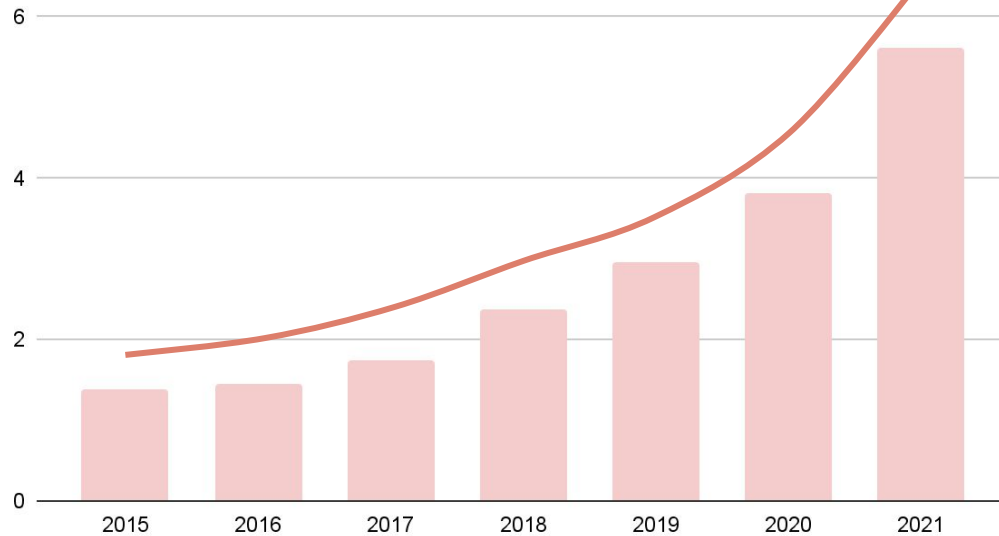




# **/01 Industry & Business Analysis**

# Dating App Industry: Business of Love

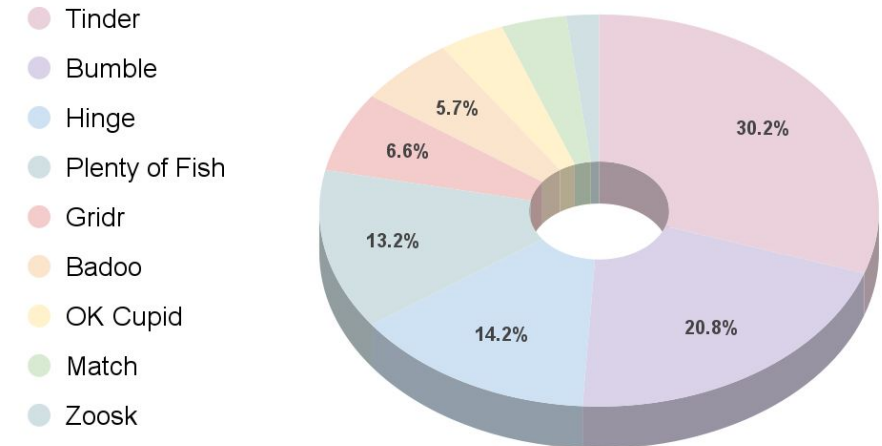
Global Dating App Revenue (\$bn)



**2021 global revenue : \$ 5.61 billion**

**5-year average growth rate : 33.8 %**

US Dating App Market Share



**US Market leader : Tinder**

**Bumble and Hinge also catching up**

# Tinder: worldwide No.1 online dating app

## Star Feature

Swipe to Match

## Matching System

"Double Opt-In" System

## Business Models

Tinder Plus\*: 4.99\$/m  
Tinder Gold\*: 14.99\$/m  
Tinder Platinum: 19.99\$/m  
Sponsored Profiles\*: 9\$  
Boosts: 1.99~3.99\$/boost

\*range depending on age or location



## Customer Acquisition

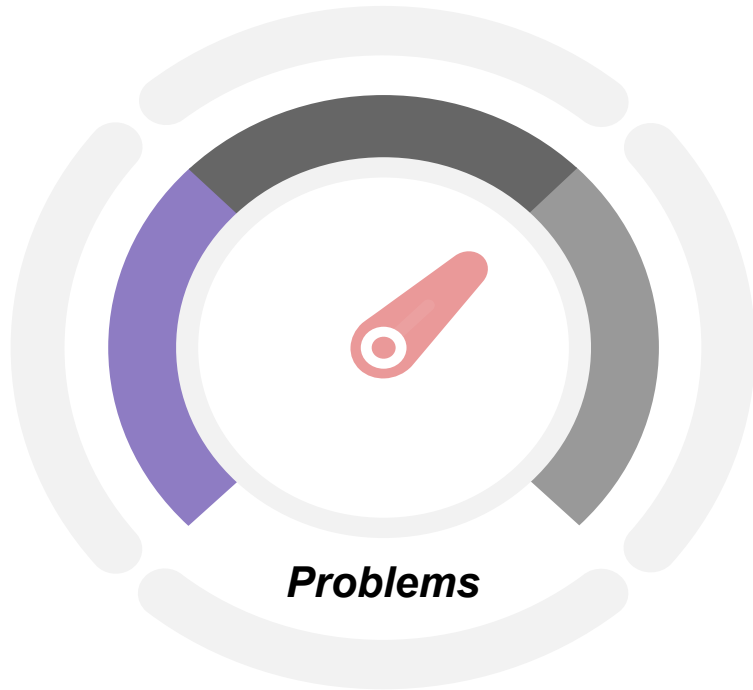
Start from  
'Campus to Campus'  
Expand by  
'City to City '

## Milestones

2015 – 5th highest-grossing mobile app  
2020 – 75 million Monthly Active Users  
2021 – 65 billion matches worldwide

# Business Problems

---



*Problems*



## Low Customer Growth Rate

Number of customers registered for the app is decreasing

---



## Low Subscription Rate

Users are not willing to pay money to subscribe and enjoy member services

---



## Low Retention Rate

Most people who go on Tinder don't stay on Tinder as consistent users



tinder





# **/02 Our Goals**

## What do we aim to achieve?

Is the review  
positive or negative?

---

What are customers  
complaining about?

Prediction and classification for  
future reviews

---

Based on identified problems, what  
improvement can tinder make?



Customer  
Review

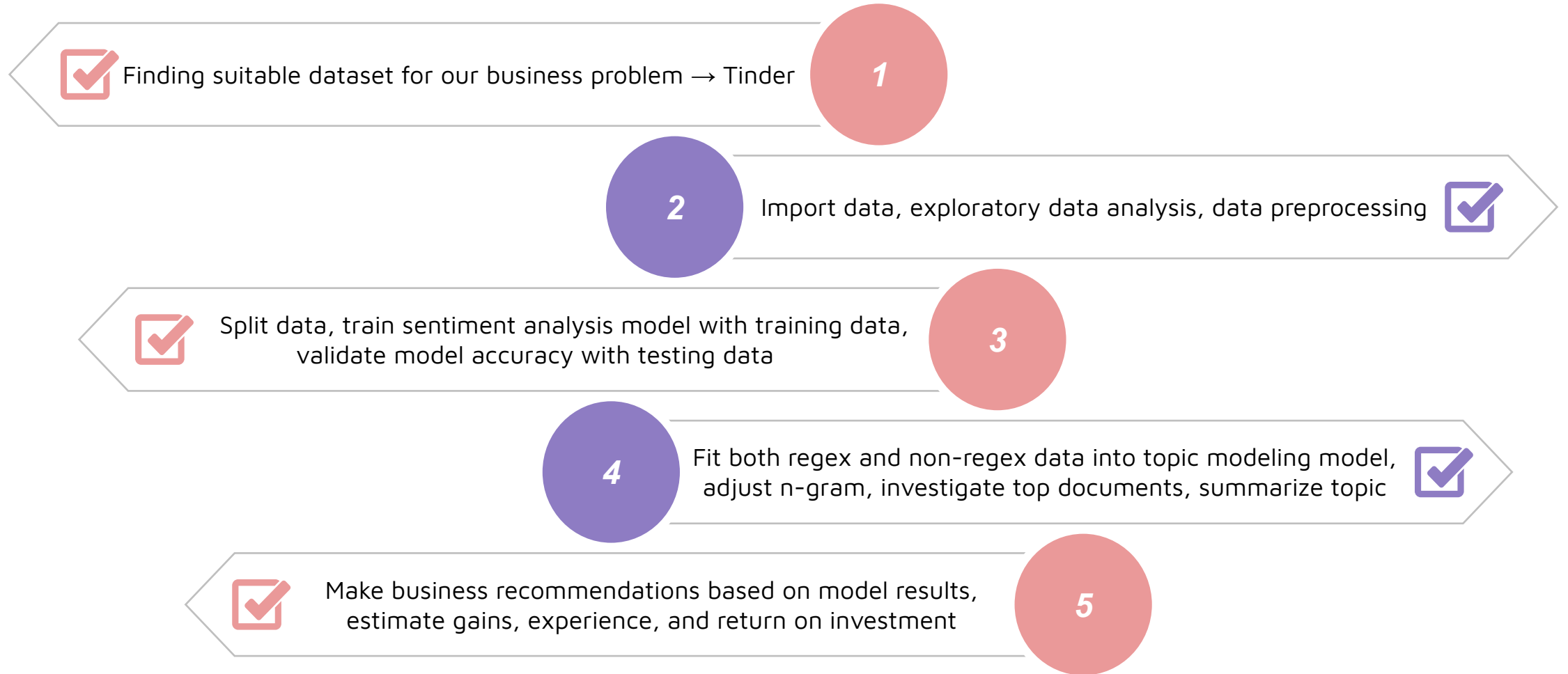
Business  
Improvement





# **/03 Implementation Roadmap**

# Implementation Roadmap



# Our Dataset

---

- We got our data from Kaggle “Tinder\_Google\_Play\_Reviews”
- <https://www.kaggle.com/code/dj67rockers/tinder-analysis/data>
- The full tinder review dataset consists of 423,747 customer reviews and ratings based on the tinder app user experience

## Columns:

- **userName:** Name of a user
- **userImage:** Profile image that a user has
- **content:** Comments made by a user
- **score:** Scores/Rating between 1 to 5
- **thumbsUpCount:** Number of Thumbs up received by a person
- **reviewCreatedVersion:** Version number on which the review is created
- **at:** Time when review was created
- **replyContent:** Reply to the comment by the Company
- **repliedAt:** Date and time of the above reply
- **reviewId:** Unique identifier



# Our Dataset

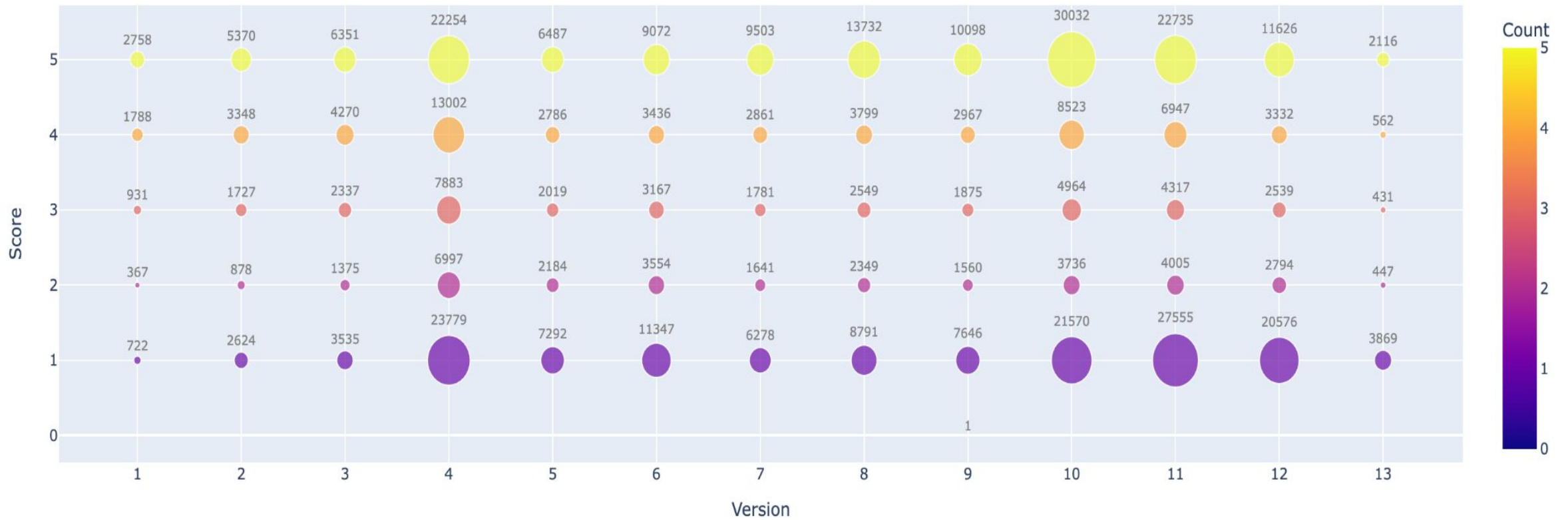


reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
gp:AOqpTOG	Yatin Raval	https://play-	Excellent	5	0	13.6.1	5/7/22 04:58		
gp:AOqpTOF	T 1000	https://play-	Messages co	1	0		5/7/22 04:49		
gp:AOqpTOF	Richard Mist	https://play-	Should be on	1	0	13.6.1	5/7/22 04:45		
gp:AOqpTOH	Shubham Sa	https://play-	Most of the a	1	0	13.6.1	5/7/22 04:38		
gp:AOqpTOE	swaggy jatt	https://play-	Better than c	4	1		5/7/22 04:37		
gp:AOqpTOF	Mpumi nxum	https://play-	The last upda	2	0	13.6.1	5/7/22 04:33		
gp:AOqpTOH	Danny Japan	https://play-	I meet with s	5	0		5/7/22 04:28		
gp:AOqpTOG	Will_i_am E	https://play-	I keep getting	1	0		5/7/22 04:15		
gp:AOqpTOF	Bone Daddy	https://play-	Mostly bots.	1	0	13.6.1	5/7/22 04:11		
gp:AOqpTOG	A F	https://play-	I can't stand	1	0	13.6.1	5/7/22 04:10		
gp:AOqpTOE	Trek Sisters	https://play-	Blocked and	1	0		5/7/22 04:05		
gp:AOqpTOF	Debashish Sa	https://play-	Only money	1	0		5/7/22 03:39		
gp:AOqpTOH	Sutria Utama	https://play-	bad apps. i'v	1	0		5/7/22 03:36		
gp:AOqpTOG	Jason Simpse	https://play-	T E R R I B L	1	0	8.1.0	5/7/22 03:31		
gp:AOqpTOF	Tanzeelur Ra	https://play-	Worst app	1	0	7.2.1	5/7/22 03:30		
gp:AOqpTOE	Sam Cayen	https://play-	Permanently	1	0		5/7/22 03:17		
gp:AOqpTOG	Boyko Chelik	https://play-	Money and n	1	0		5/7/22 03:17		
gp:AOqpTOH	Orion Rutley	https://play-	Where's my	1	0	13.6.1	5/7/22 02:58		
gp:AOqpTOF	‡‡û‡‡£‡‡	https://play-	Terrible	1	0	13.6.1	5/7/22 02:52		
gp:AOqpTOF	John Henry	https://play-	Fun times ar	5	0	13.6.1	5/7/22 02:15		



# Exploratory Data Analysis

App version vs Score review



tinder

# Exploratory Data Analysis

---

“

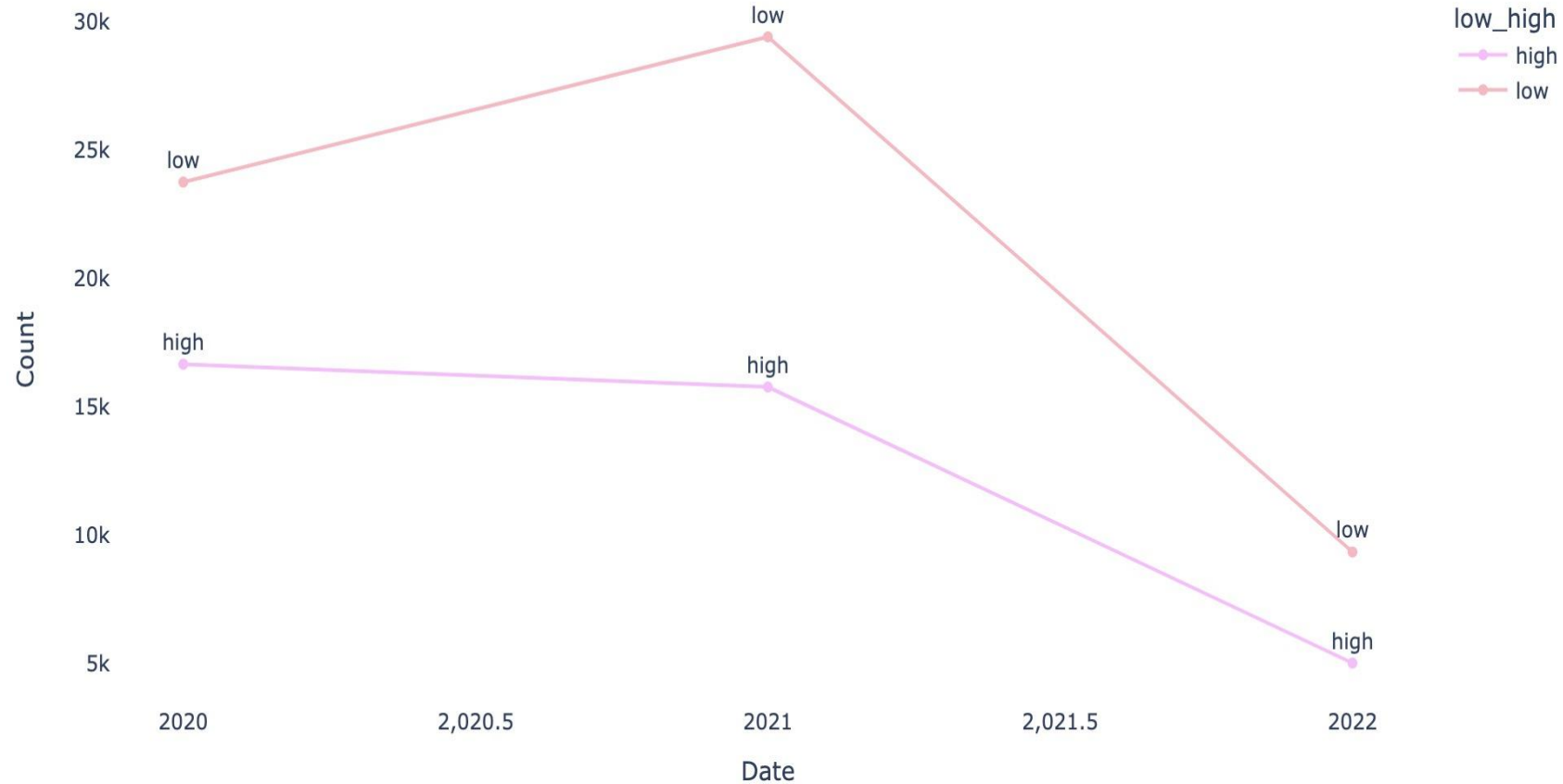
Number of daily  
customer reviews  
from April 2020 to  
April 2022

”



tinder

# Exploratory Data Analysis



Number of semi-annual  
low-score and  
high-score customer  
reviews from April 2020  
to April 2022

# Data Preprocessing

## Drop

- Drop Rows with null Content and Version
- Drop redundant columns

## Stopwords

- Remove stopwords (E.g. the, is, they)
- Additional stopwords (tinder, app, literally, really, very, definitely)

## Lemmatization

- Convert all words to original form
- Why not stemming: need to consider sentiment of the content, need to reserve original words for analysis

Drop

Filter

Stopwords

Punctuation

Lemmatization

Regex

## Filter

- Filter out outdated data
- Select data from 2020 to 2022

## Punctuation

- Separate punctuations from words

## Regex

- Substitute words with similar meaning
- group constantly occurring themes together



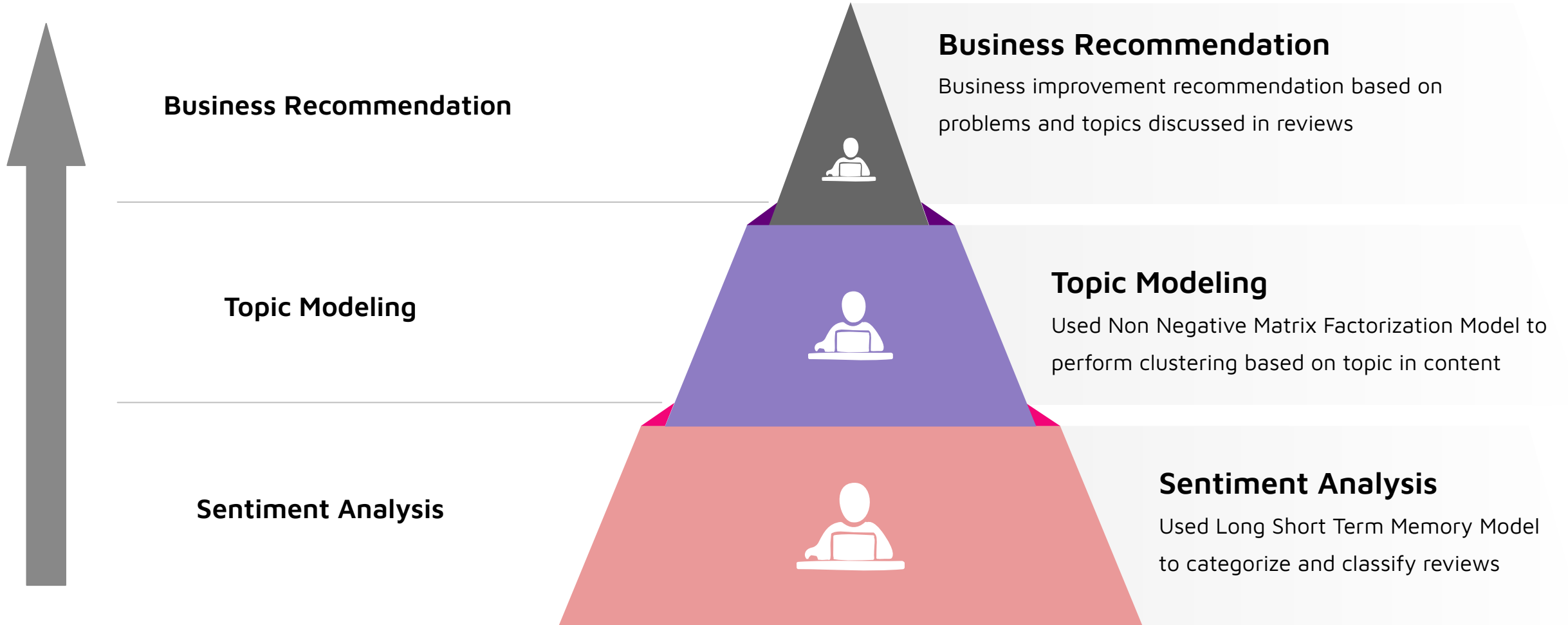
tinder





# **/04 Models**

# Models



# Why Sentiment Analysis?

## Classification

- LSTM algorithm can help customer support department to classify whether comments are positive and negative based on the accuracy of algorithm

## Increase Efficiency

- Machine learning still under the guidance of domain experts
- Domain experts have more time to think about how to cooperate with algorithm and improve it



## Reduce Cost

- With the help of LSTM classification algorithm, customer support department can reduce the cost from labors who need to manually check the comments

## Automation

- Streaming the classification of positive comments and negative comments by utilizing algorithm, reduce manual intervention

# Sentiment Analysis - Long Short Term Memory

01

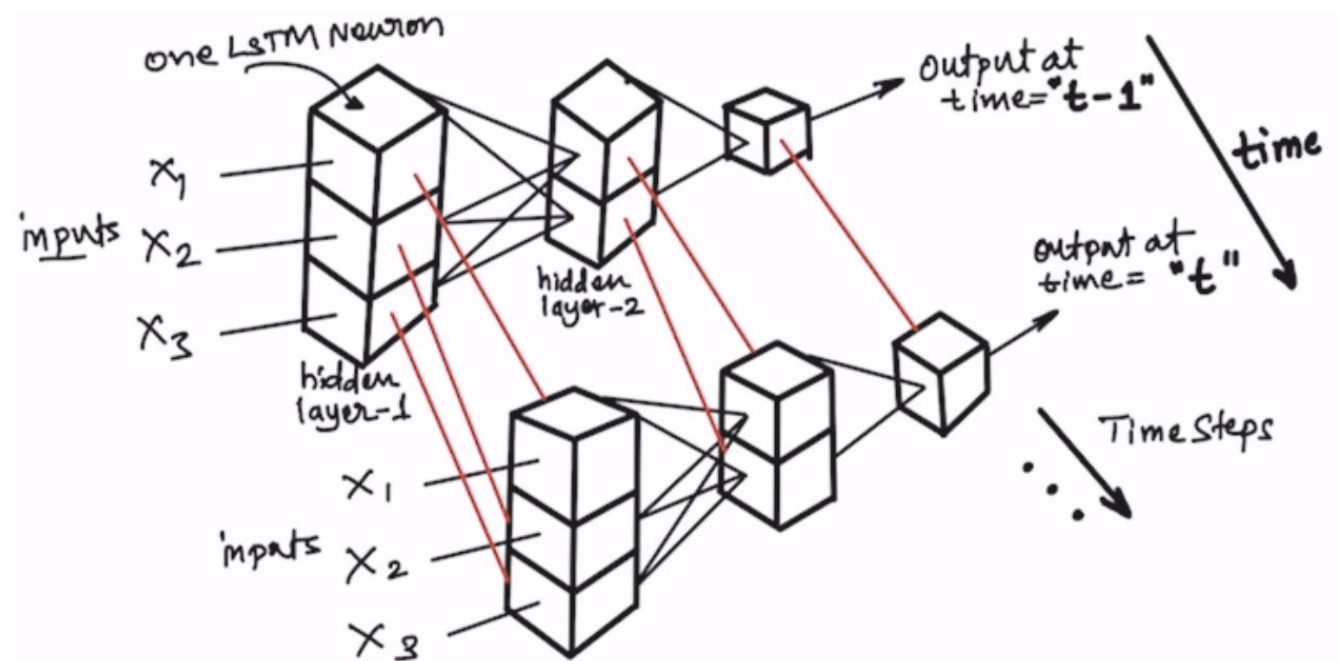
LSTM is artificial neural networks used in the fields of AI and deep learning

02

LSTM networks are well-suited to classifying, processing, and making predictions

03

LSTM is better solutions for gradient disappearance and gradient explosion in the process of long sequence training



tinder

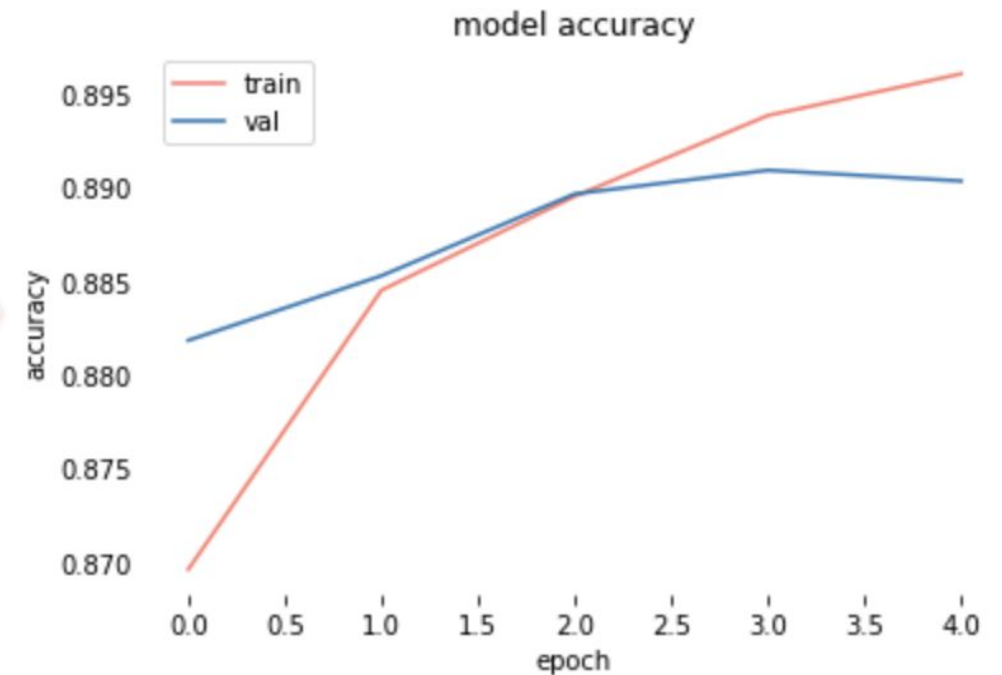


# Sentiment Analysis - Long Short Term Memory

Testing dataset accuracy : 89.21%

```
loss, accuracy = model.evaluate(X_test, y_test, verbose=1)
```

```
loss: 0.2785 - accuracy: 0.8921
```

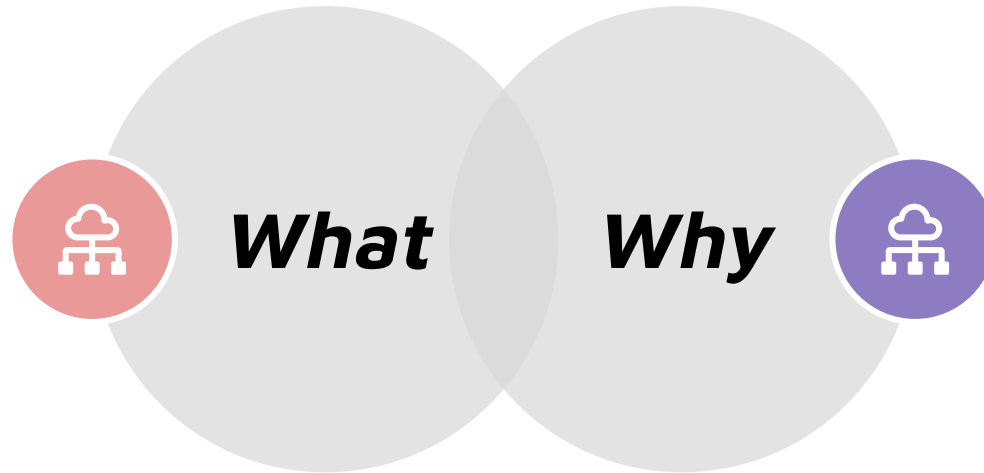


# Why Topic Modeling?

---

Identify product issues by reviewing comments on Google Play Store

What are the product features that our customers are complaining about?



Why are our customers complaining these product features?

# Topic Modeling - Non-Negative Matrix Factorization

01

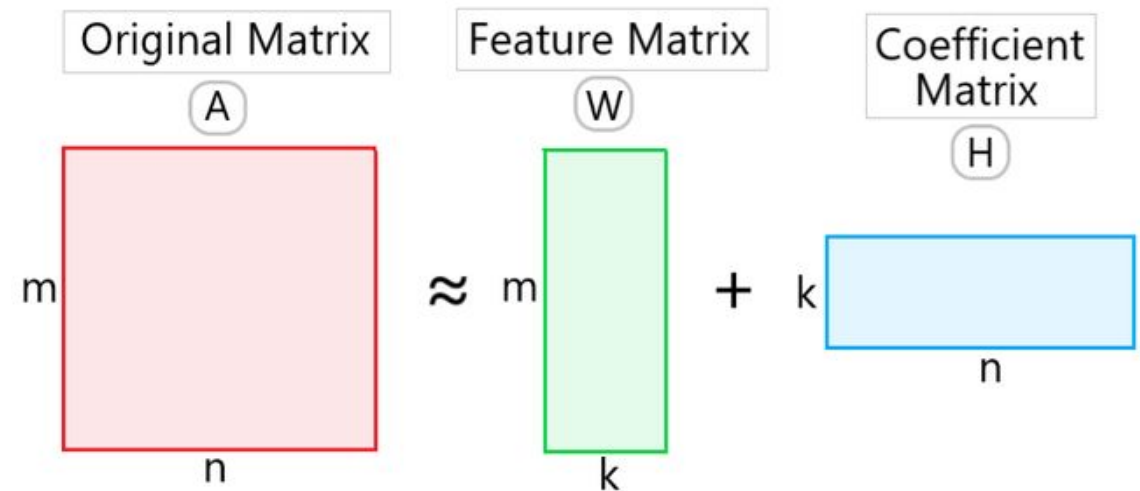
NMF is a group of algorithms in multivariate analysis and linear algebra where a matrix  $V$  is factorized into (usually) two matrices  $W$  and  $H$ , with the property that all three matrices have no negative elements.

02

NMF extracts topics and discovers the underlying relationships between texts.

03

We utilize NMF to help identify major product issues from Google Play Store reviews.



# Topic Modeling - Topic Summary

---

## Topic 0

---

not \_waste\_ time (58.8%)  
\_ban\_ without reason (1.2%)  
\_fake\_ not \_waste\_ (1.1%)  
please not \_waste\_ (1.1%)  
\_waste\_ time not (0.8%)

## Topic 1

---

account \_ban\_ reason (52.8%)  
\_ban\_ reason not (6.4%)  
\_ban\_ reason give (1.8%)  
say account \_ban\_ (1.4%)  
got \_ban\_ reason (1.2%)

## Topic 2

---

\_number\_ mile away (37.5%)  
show \_number\_ mile (4.3%)  
\_number\_ \_number\_ mile (3.6%)  
\_matching\_ \_number\_ mile (3.5%)  
set \_number\_ mile (2.8%)

## Topic 3

---

\_waste\_ time money (63.0%)  
dont \_waste\_ time (3.6%)  
complete \_waste\_ time (1.8%)  
time money not (1.6%)  
total \_waste\_ time (1.5%)





# **/05 Business Recommendations**

# Major Complaints



## Banned Account for No Reason

- "It banned me for doing nothing"

.....

## Distance

- "Matching with people who where 1-2 miles away and when i match with the person they will be 2,000-9,000 miles away"

.....

## Waste of Money

- "I paid for the gold service and literally for 50 dollars you get nothing."

.....

## Fake Accounts

- "You guys definitely need to start verifying your users you have so many fake profiles it's ridiculous!!!"

.....



# Business Recommendations

## Banned Account for No Reason

- Improve fraud detection algorithm
- Add appeal service for users

## Distance

- Increase frequency of updating user location
- Add a filter to only match people within certain distance

## Waste of Money

- Provide advice from experts for users with subscription to increase matching chance
- Filter by personal tags for members only

## Fake Accounts

- Give users the option to perform facial recognition and increase exposure rate for these users
- Adopt report service for fake matching
- Based on report system to improve fraud detection algorithm



# Return on Investment

---

## Earnings and Revenues

For each business model (m\$):

Revenue = MAU \* (Retention Rate + Expansion Rate + Upgrade Rate)

Rev(Total) = Rev(Platinum) + Rev(Gold) + Rev(plus)

= 0.57 + 0.50 + 0.43 = **1.50**

**\$ 1.50  
Million**

## Expenditures

For all business models (m\$):

Capital expenditure = 0.7

Direct Labor expenditure (domain expert team) = 0.04 \* 12 = 0.48

Expenditure (Total) = 0.7 + 0.48 = **1.18**

**\$ 1.18  
Million**

## Return on Investment

Total Earnings (1.5) - Total Expenditures (1.18)

Total Expenditures (1.18)

= Return on investment (**0.27**)

**ROI = 27%**



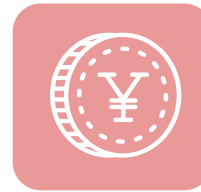
# Conclusions

## “Swipe Right”

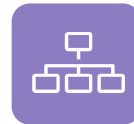
Through google play reviews for Tinder, we aim to identify the root causes behind Tinder’s business problems.



Advanced NLP techniques allow us to investigate deeply into review contents, classify reviews, and then automatically summarize topics covered in reviews;



By adopting Sentiment Analysis, Tinder can use the classification model as an automatic tool for splitting reviews and therefore reduce the labor costs and achieve automation in its initial screening process;



With Topic Modeling, Tinder will be able to extract major complaints and root problems behind its low customer growth, retention, and subscription rates. Tinder can therefore timely adjust their business model and app functions to address the problems and increase profit margin.





**Thanks**

**“Swipe Right”**

Group White: Xinde Huang, Chiayi Lin, Liang Chi Liu,  
Jessica Xu, Zijian Yan





# **/06 Appendix**

# Return on Investment Derivation

- Monthly Active Users (MAU) = 7.8m
  - data cited from *Data Source: businessofapps.com; keyua.org*
- Retention rate = 5%, Expansion rate = 2.5%
  - data derived from <https://datingzest.com/tinder-statistics/>
- We estimate that mean monthly fee for platinum, gold and plus are respectively 30\$, 17.5\$, and 7.5\$
- Revenue = MAU \* (Retention Rate + Expansion Rate + Upgrade Rate)
- Expansion Rev:
  - PLATINUM =  $7.8 \times 5\% \times (1 - 15\%) + 7.8 \times 2.5\% \times 10\% \times 10\% \times 30 = 0.39$
  - GOLD =  $7.8 \times 5\% \times (1 - 15\%) + 7.8 \times 2.5\% \times 10\% \times 20\% \times 17.5 = 0.40$
  - PLUS =  $7.8 \times 5\% \times (1 - 15\%) + 7.8 \times 2.5\% \times 10\% \times 70\% \times 7.5 = 0.43$
- Upgrade Rev:
  - PLATINUM =  $7.8 \times 5\% \times 15\% \times 10\% \times 30 = 0.18$
  - GOLD =  $7.8 \times 5\% \times 15\% \times 10\% \times 17.5 = 0.10$
- Total Rev: 1.5m

