

# Mobile App Development

Phoebe Wu

## Research Background & Question

The number of smartphone user are almost 3 billion around the world in 2018. → the importance of the smartphone App nowadays

- What can make an app become a top trending (top installation/downloads) app in the app store?
- How to make an App get high rating (rating  $\geq 4.0$ ) from users?

Google Android(26.8%) and Apple iOS(24%) are the two most popular smartphone operating systems in the smartphone industry.  
→ Google Android App & Apple iOS App

## Datasets - from Kaggle

- Dataset 1 : Google Play App Store
  - The data was scraped from Google Play Store.
  - This dataset contains more than 10 thousands app information and 37 thousands app reviews.
  
- Dataset 2 : Apple iOS Apps
  - The data was extracted from the iTunes Search API at the Apple Inc website.
  - This dataset contains about 7200 Apple IOS mobile application details and descriptions.

## The Google App

### • App information dataset

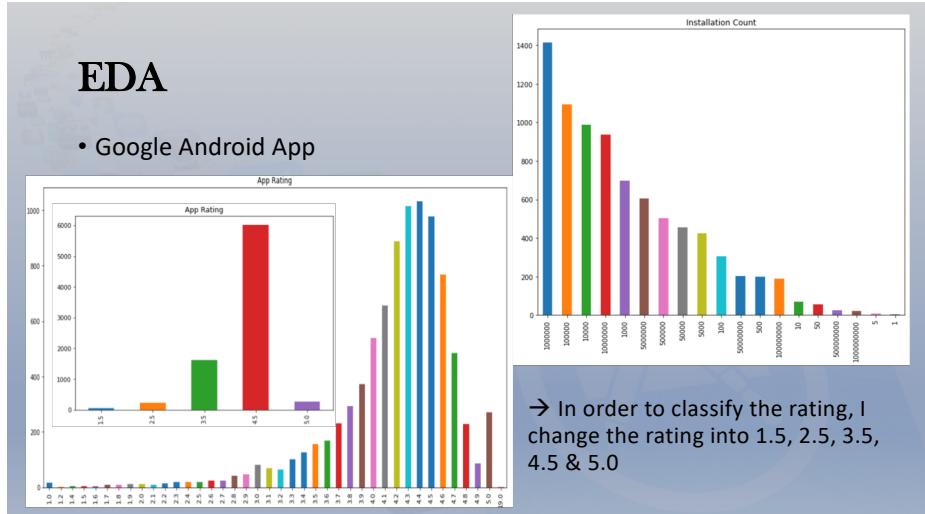
1. App: App name
2. Category: Category the app belongs to
3. Rating: Overall user rating of the app
4. Reviews: Number of reviews
5. Size: Size of the app
6. Price
7. Type: Free or Paid
8. Installs: Number of user downloads/install the app
9. Content Rating: the target age group of the app (children/ mature 21+/ adult)

### • App reviews dataset

1. App: App name
2. Translated review: User review
3. Sentiment: Positive/Negative/Neutral

# EDA

- Google Android App



# The Apple App

- App information dataset

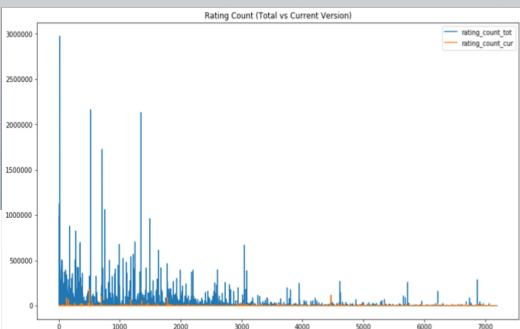
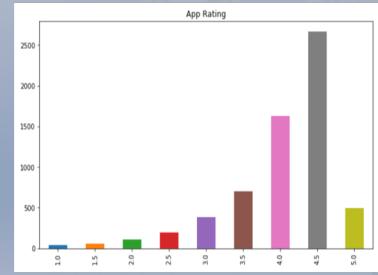
1. App Name
2. size\_bytes: Size (in Bytes)
3. price
4. prime\_genre
5. rating\_count\_tot: User Rating counts (for all version)
6. rating\_count\_cur: User Rating counts (for current version)
7. user\_rating: Average User Rating value (for all version)
8. user\_rating\_cur: Average User Rating value (for current version)
9. cont\_rating: Content Rating( 4+: all age, 9+: age over 9, 12+: age over 12, 17+: age over 17)
10. sup\_devices\_num: Number of supporting devices
11. screenshot\_num: Number of screenshots showed for display
12. lang\_num: Number of supported languages

- App description dataset

1. App Name
2. app\_desc: Application description

## EDA

- Apple iOS App



## Challenge & Adjustment

- Firstly, I try to classify the rating which are classes 1.0, 1.5, 2.0, ..., 4.5, 5.0. → The accuracy was around 0.3
- I tried cross validation, PCA but the results did not improve.
- Finally, I redefined the classes into rating  $\geq 4.0$  &  $< 4.0$
- Moreover, in Apple app dataset, I create a new column: 'rating count before'  
(‘total rating count’ – ‘current version rating count’)

## Analysis

- **Classification & Regression** - App information (Apple App & Google Store App)
  - Rating  $\geq 4.0$  vs  $< 4.0$
  - Installation count & Rating count
  - Classifying the Positive, Neutral and Negative reviews
- **Natural Language Processing** → developers' emphasis match users' concern
  1. Google Store App reviews
    - Find the most frequent words in the reviews
  2. Apple App description
    - Find the most frequent words in the descriptions

## User Rating - Google App

- Classification
  - Random Forest, KNN, Logistic Regression, Decision Tree
  - Features (feature importance)
- 'Size', 'Category', 'Price'

Train-test-split: 70/30

	Classifier_name	train_score	test_score	F1_score
0	Random Forest Classification	0.854428	0.717771	0.698182
1	KNN	0.786262	0.738105	0.699861
2	Logistic Regression	0.766911	0.766978	0.665833
3	Decision Tree	0.863668	0.676291	0.667992

## User Rating - Apple App

- Classification

- Random Forest, KNN, Logistic Regression, Decision Tree

- Features(feature importance)

'size', 'rating count before', 'language number', 'prime genre' and 'support devices number'

Train-test-split: 70/30

	Classifier_name	train_score	test_score	F1_score
0	Random Forest Classification	0.98108	0.69697	0.696173
1	KNN	0.753818	0.695375	0.683077
2	Logistic Regression	0.702074	0.695906	0.613539
3	Decision Tree	1	0.651249	0.653317

## User rating count (Installation) - Apple App

- Regression

- Features

'size', 'rating count before', 'content rating', 'support devices number' , 'language number'

Train-test-split: 70/30

	Regression_name	train_score	R square
0	Linear Regression	0.978083	0.906962
1	Random Forest Regressor	0.906868	0.873986

# Google APP Reviews

NLP & Classification – Positive, Neutral and Negative

- Random Forest Classifier

- KNN

- Decision Tree

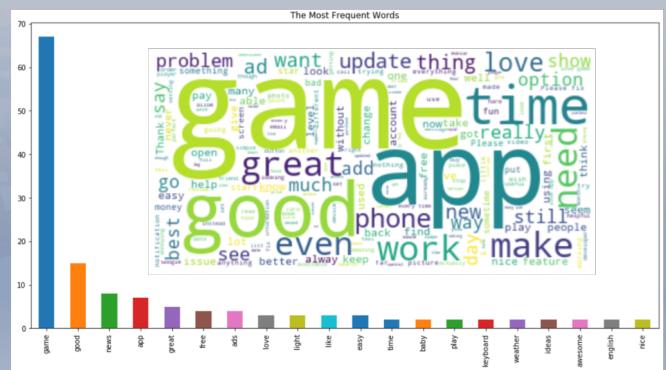
Train-test-split: 70/30

	Classifier_name	train_score	test_score	F1_score
0	Random Forest Classifier	0.989503	0.8812	0.88006
1	KNN	0.783991	0.702467	0.719462
2	Decision Tree	0.993549	0.869089	0.869035

# NLP - Google APP Reviews

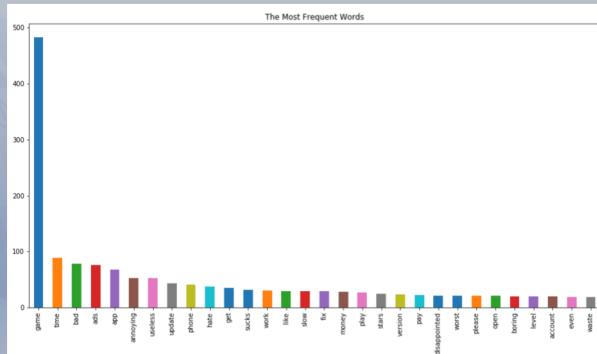
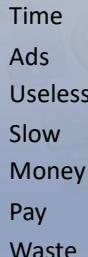
The most frequent words

Game  
News  
Weather  
Good  
Free  
Ads  
Light  
Easy  
Time



NLP - Google APP Negative Reviews

The most frequent words – except the emotional words



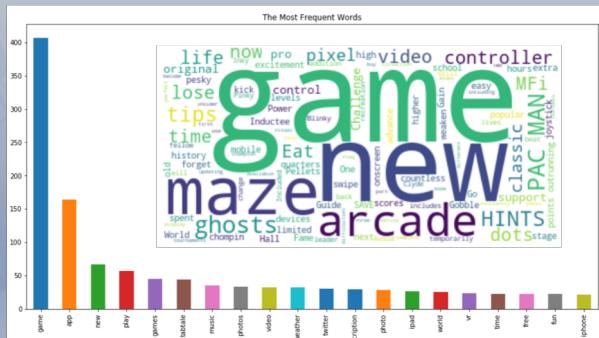
# NLP - Apple App Descriptions

## The most frequent words –

- Game
  - Music
  - Photo
  - Video
  - Weather

---

  - New
  - Play
  - Time
  - Free



# NLP - Apple High Rating App Description

## The most frequent words —

## Game

New

Play

Music

## Photo

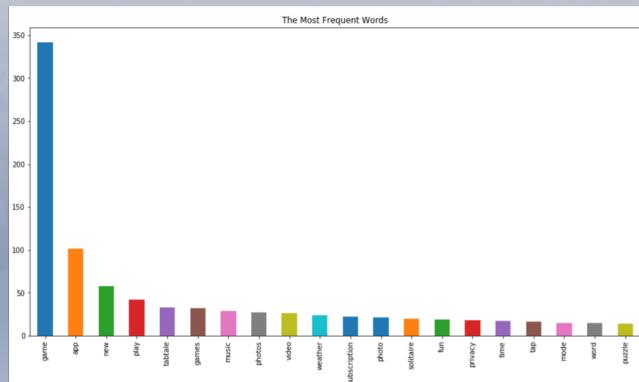
## Video

Weather

Sub

Fun

Private



# NLP - Top Trending App Descriptions

## The most frequent words –

## Game

Weather

Free

## Photo

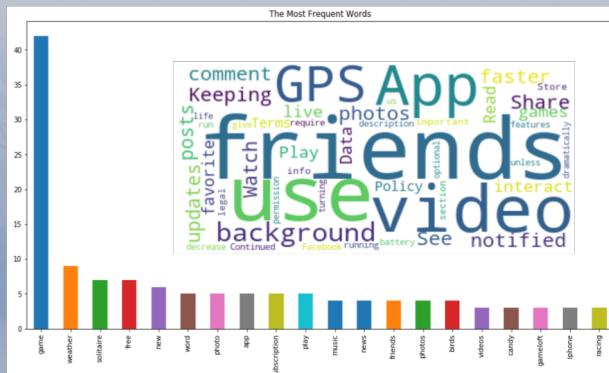
## Subscription

Music

News

Erienc

## Videos



## Conclusion

- **Getting High Rating**

- Google Android App – Size, Category and Price
- Apple iOS App – Size, Prime genre, number of Supporting Devices and Language

- **Become Top Trending** – Size, Content rating, number of Supporting Devices and Language

- **Most Frequent words in both reviews & descriptions – Game, Time and Free**

- Game → The most popular app category in both system
- Time & Free → The concern by both developers and users
- Reviews: Useless & Ads → Useful or not and ads disturbance
- Descriptions: Friends and share → related to the top trending Social media apps

## Conclusion - Apply to Real World

- For both Operating System:

- Size, Price, Number of Supporting Devices and Languages for developing High Rating & Top Trending Apps

- Google Android App

- Improving the app operating speed and make the app easier to use

- Apple iOS App

- Put more focus on the number of supporting devices and language of the app.