Mobile APP Analysis

Shang-Shiun (Brian) Tsai

ALY 6110 Data Management & Big Data

Northeastern University

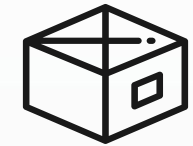
Table of contents

1 Introduction

2 Analysis

3 Conclusion

01 Introduction



Introduction

Background

1、Digitization, Intelligentization
2、Increased ownership of smart devices,
3、The expansion of app stores(App Store1, Google Play)

Goal

software data in the Apple App Store.

Method

Clustering, Heatmap, Correlation

Data Set

Applications published in the Apple IOS store.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7197 entries, 0 to 7196
Data columns (total 17 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   id                7197 non-null   int64
 1   track_name        7197 non-null   object
 2   size_bytes        7197 non-null   int64
 3   currency          7197 non-null   object
 4   price             7197 non-null   float64
 5   rating_count_tot  7197 non-null   int64
 6   rating_count_ver  7197 non-null   int64
 7   user_rating       7197 non-null   float64
 8   user_rating_ver   7197 non-null   float64
 9   ver               7197 non-null   object
 10  cont_rating       7197 non-null   object
 11  prime_genre       7197 non-null   object
 12  sup_devices.num   7197 non-null   int64
 13  ipadSc_urls.num   7197 non-null   int64
 14  lang.num          7197 non-null   int64
 15  vpp_lic           7197 non-null   int64
 16  app_desc          7197 non-null   object
dtypes: float64(3), int64(8), object(6)
memory usage: 956.0+ KB
```

02

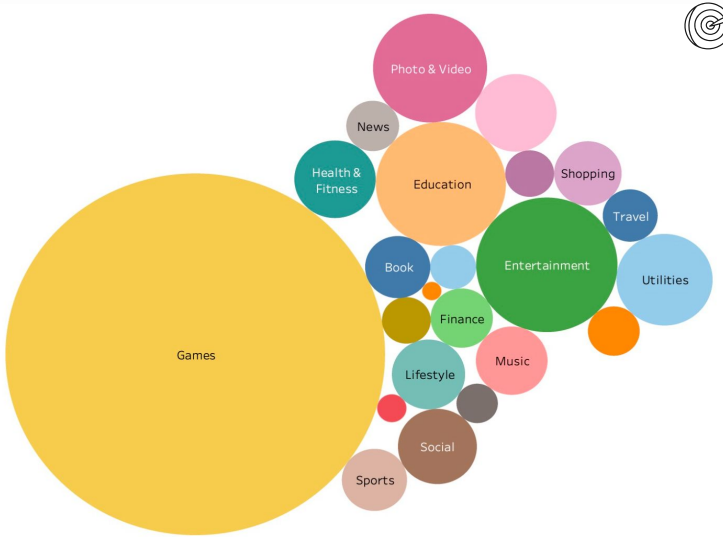
Analysis

Genre

NEW!

Top 5 App category

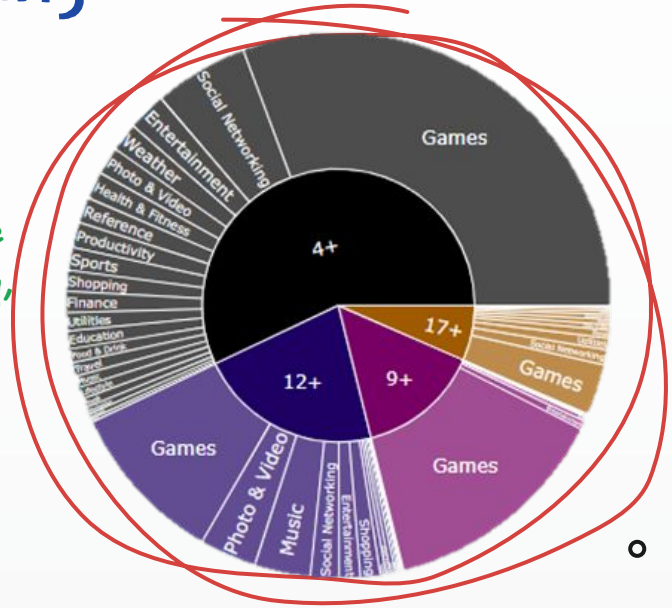
prime_genre	count
Games	3862
Entertainment	535
Education	453
Photo & Video	349
Utilities	248

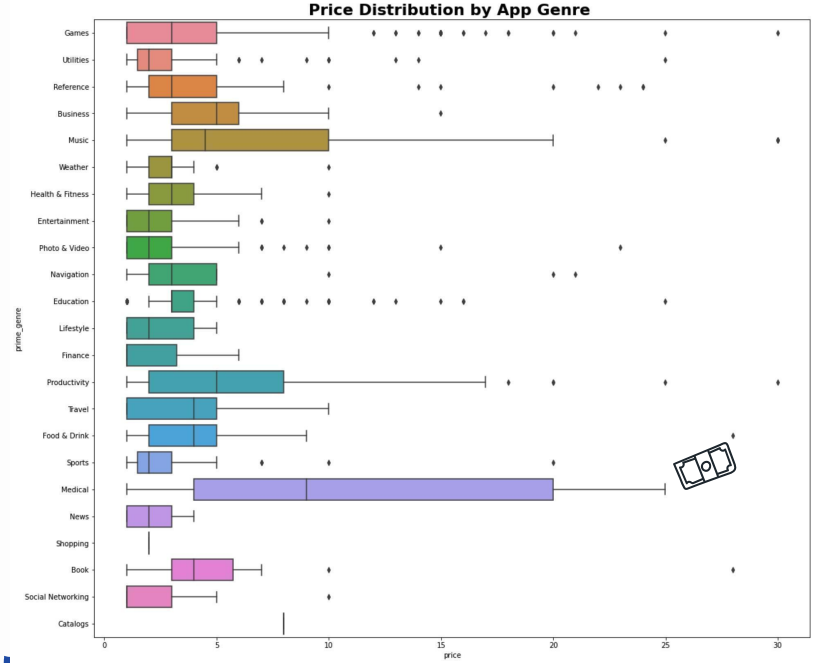


Game is dominant

Content Rating

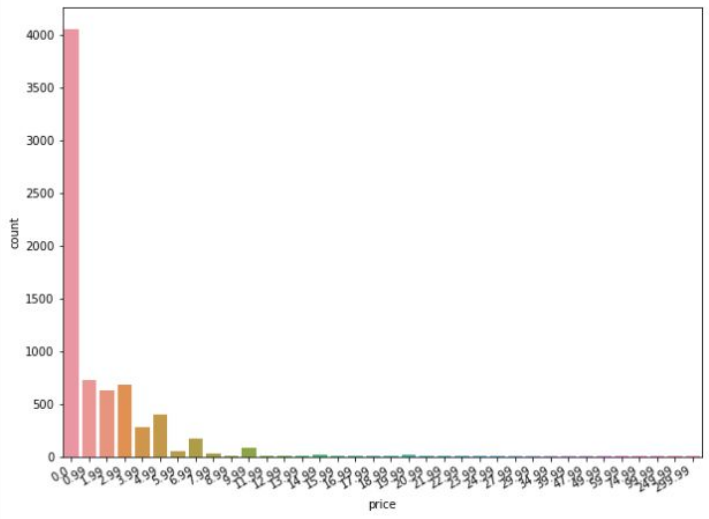
- 4+ content is the most
- Each age group has corresponding game
- prevalence of mild violence, pornography, gore and other horrific content





Medical has the greatest price range and highest average price

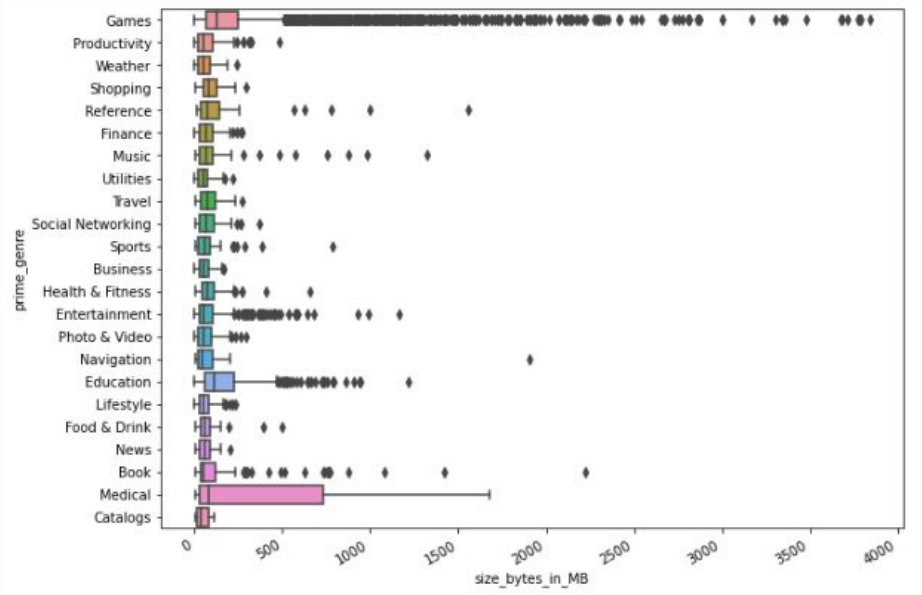
Price-count plot



- Most of the apps are FREE
- The distribution of price count is between

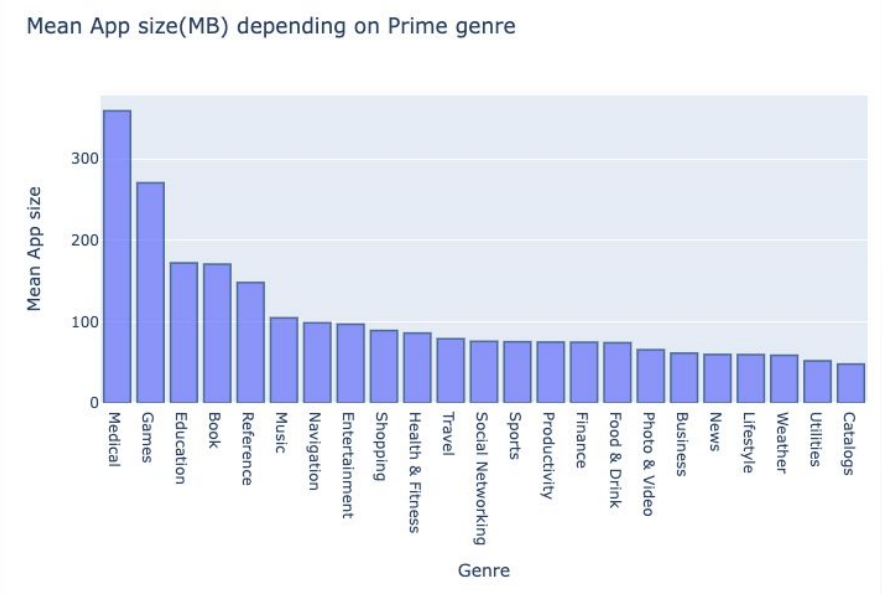
  0 to 9.99 USD

Size bytes plot



- **Games** have the most outliers
- Most of the genres' sizes are less than 500 MB.

Average Size of app in different genres



Medical is the first, go with **Games**, and **Education** and Book are about the same.

Paid APP and Free APP Rating

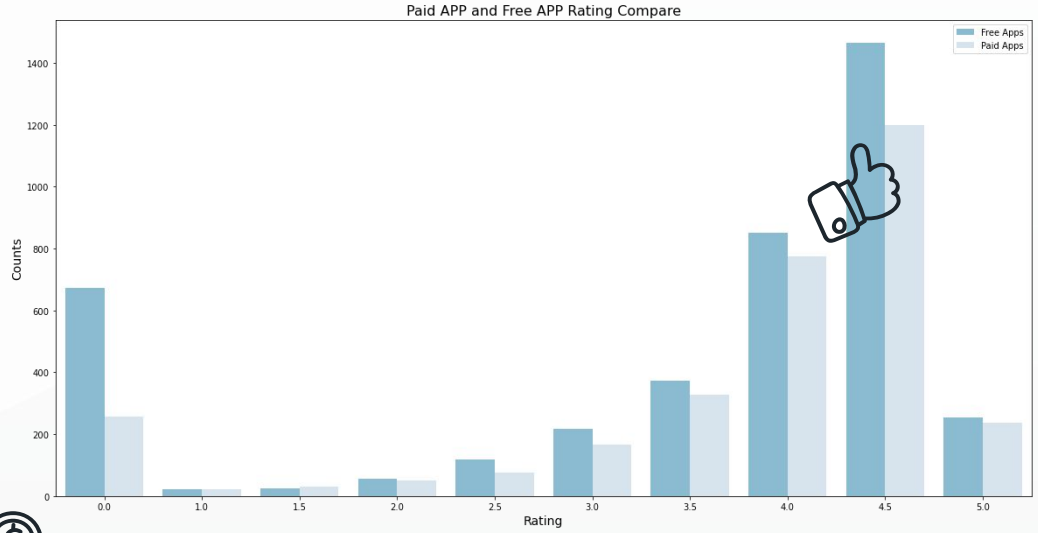
Higher than 4.0

Better user experience

0 Rating: 700

Bad user experience

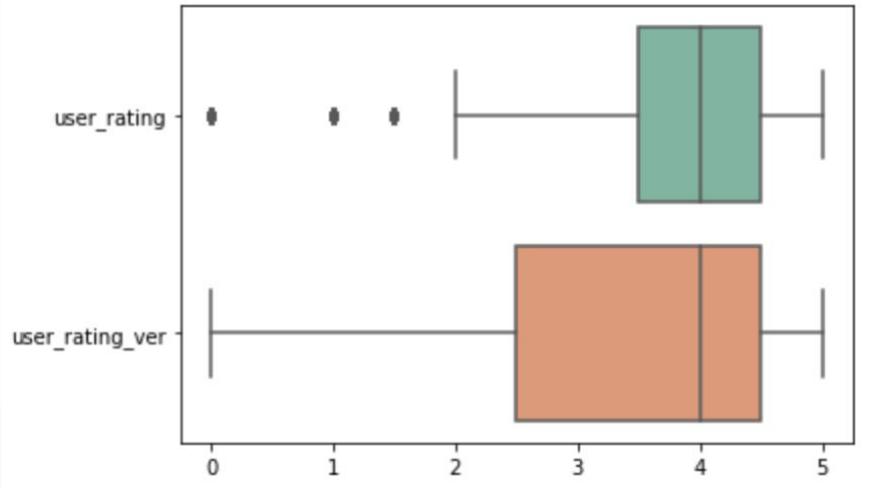
- keeps crashing
- too many ads



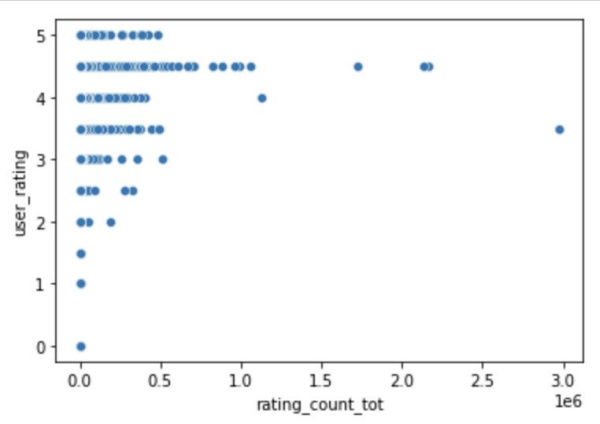
Comparison between Current Version & all Versions

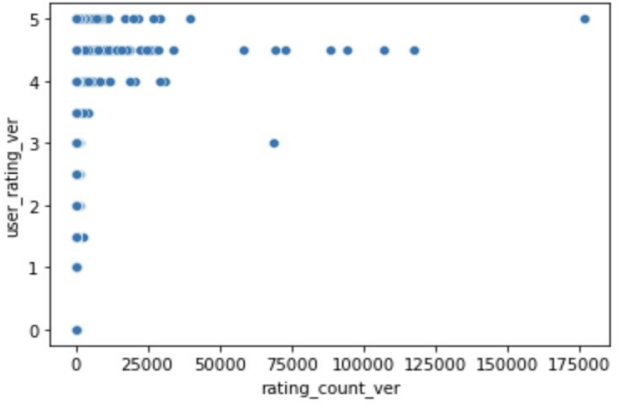
All versions ratings in 3.5-4.5 star

current version ratings in 2.5—4.5 star



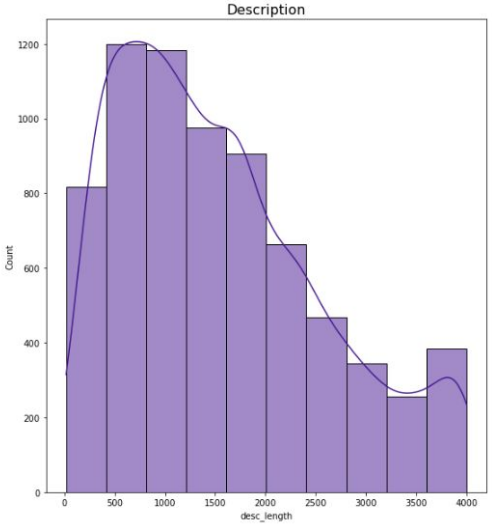
Relationship between ratings and the number of ratings

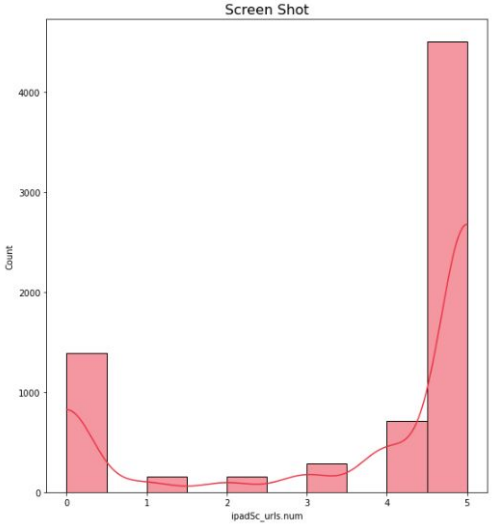




When the quantity is **larger**, the grades is high
but for the **less** quantity, the ratings can vary.

App text length and number of displayed images





Most apps display interface details with text length between 750-1500 words and 5 images

Word Cloud

App

Game

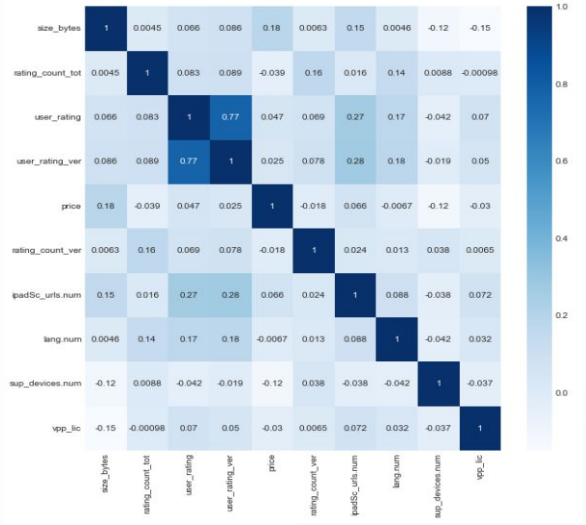
Play

New

Use

Friend

Correlation



- **The most relevant**
  user rating of all versions
  and user rating of current versions
- **The correlation between**
  the number of languages supported
  and user rating
- Screenshot **correlates with** user rating
- Pricing **relative** APP_Size

Clustering

optimal value: K=3

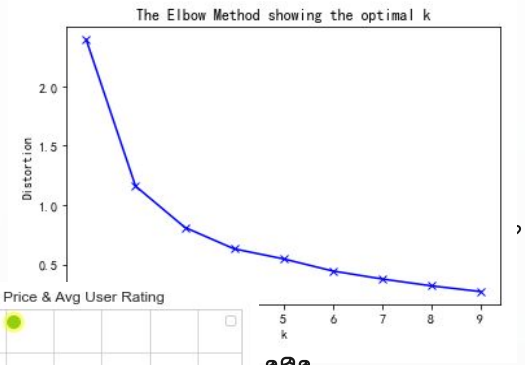
Clusters of orange:

populated by users who download

free APPS or fewer than $5.

Clusters of GREEN:

more expensive APPS



Conclusion 03

Conclusion

- Considerable competition for **Game** App development .
- Better download for free, specialized Apps need to be paid (ie. medical)
- App prices are overwhelmingly focused on $9.99 or less.
- The content rating range of most types of apps is 4+.
- Current version ratings mostly are lower than all versions ratings.
- There is a significant positive correlation between price and size.
- Most of the genres' sizes are less than 500 MB.

Reference

Apple. (2022, May 31). Report finds third-party apps see global success on the App Store. Apple Newsroom.https://www.apple.com/newsroom/2022/04/report-finds-third-party-apps-see-global-success-on-the-app-store/

Botelho, B., & Bigelow, S. J. (2022, January 5). big data. SearchDataManagement. https://www.techtarget.com/searchdatamanagement/definition/big-data

Thanks!

Does anyone have any questions?