

STAT 403

IOS Mobile App

Zhiying Xie, Yifei Chen, Seoyoung Park, Ziqi (Katie) Chen

Dataset-IOS Mobile Apps

- Obtained from *Kaggle* <https://www.kaggle.com/ramamet4/app-store-apple-data-set-10k-apps>
- Contains details for 7198 apps from 16 aspects (16 variables in the dataset) from iTunes Search API



—	Education
—	Entertainment
—	Health_And_Fitness
—	Lifestyle
—	Music
—	Games
—	Photo_And_Video
—	Productivity
—	Social_Networking
—	Utilities

Variables in the dataset

Id App ID	track_name App Name	size_bytes Size(in bytes)	Currency Currency Type
price Price amount	rating_count_tot User rating counts(for all ver.)	rating_count_ver User rating counts(current ver.)	user_rating Average user rating(for all ver.)
user_rating_ver Average user rating(current ver.)	ver Latest version code	cont_rating Content rating	prime_genre Primary genre
sup_devices.num # of supporting devices	ipadSc_urls.num # of screenshots shown	lang.num # of supported languages	vpp_lic Vpp device based licensing enabled

< Apps



HotSchedules

HotSchedules

price

\$2.99



cont_rating

4+

Age

4.7 ★★★★★

6,561 Ratings

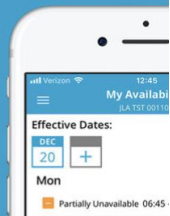
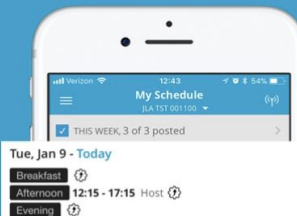
#1

Business

prime_genre

View your work
schedule from
anywhere, any time

Let your manager
know when you
are available



ipadSc_urls.num



Today



Games



Apps



Updates



Search

< Apps



\$2.99

currency

Ratings & Reviews

See All

4.7

out of 5



65,610 Ratings

rating_count_tot

crappy app that could be great

Feb 4



Dns2k

token and password issues EVERY update.
plus the location is always messed up (you
should make it with option to set location)
so we can set a location for calendar sync.
{and speaking of sync every time app is
foreground it resyncs pushing calendar
entries}

more

ver What's New

Version History

Version 4.108.0

1w ago

- My Schedule - Show Total Scheduled Hours
- Various bug fixes and improvements
- Thought of the week - "Bean bags are ju more



Today



Games



Apps



Updates



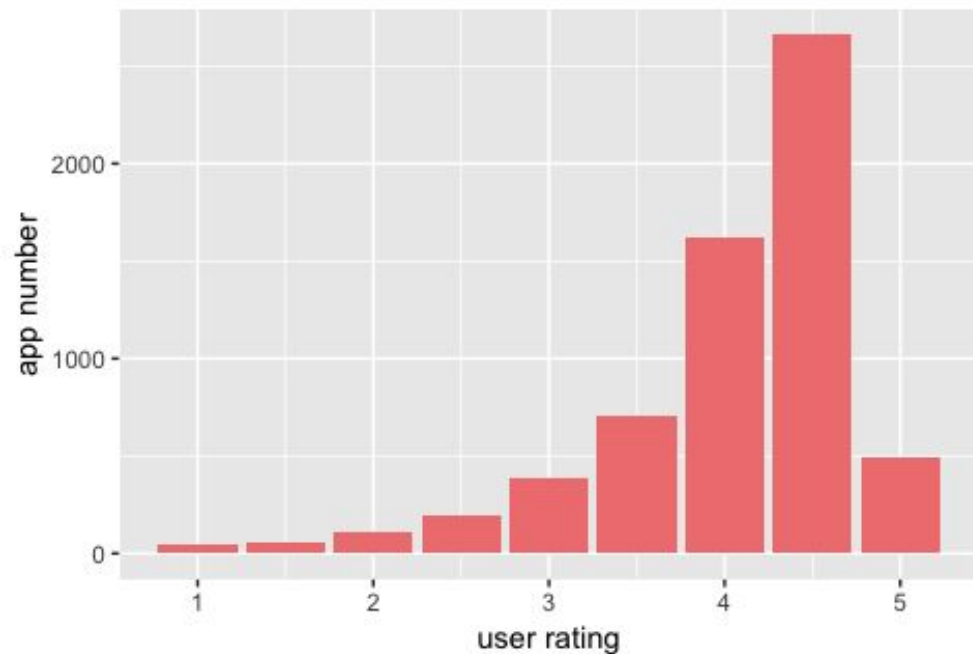
Search

Data Analysis Procedures

- Exploratory data analysis (EDA) of the Original Data
- Process the Data
 - Delete the apps with average rating and total rating counts equal to 0
- Resampling
 - Use bootstrap methods to gain 95% CI of the variables
- Prediction Models

General Information

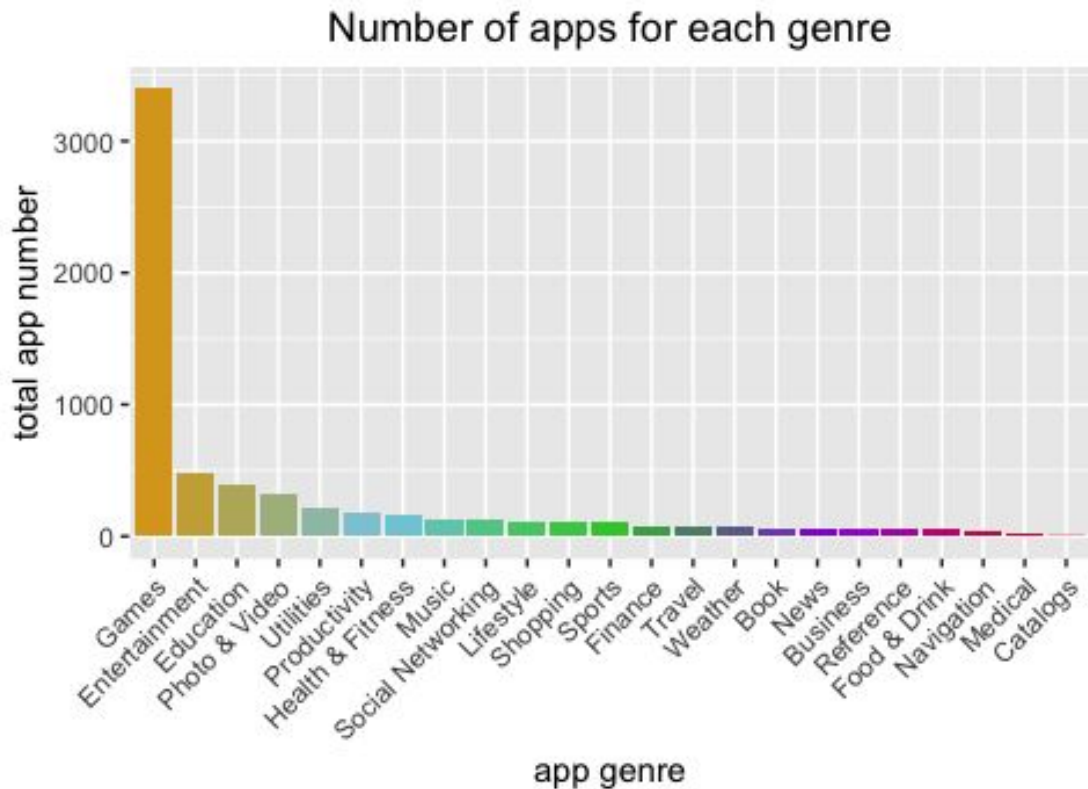
Number of apps by rating



Top 10 apps with the highest rating count

track_name	prime_genre	rating_count_tot	user_rating
Facebook	Social Networking	2974676	3.5
Instagram	Photo & Video	2161558	4.5
Clash of Clans	Games	2130805	4.5
Temple Run	Games	1724546	4.5
Pandora - Music & Radio	Music	1126879	4.0
Pinterest	Social Networking	1061624	4.5
Bible	Reference	985920	4.5
Candy Crush Saga	Games	961794	4.5
Spotify Music	Music	878563	4.5
Angry Birds	Games	824451	4.5

General Information (cont)



Research Questions



How is user rating categorized by genres affected by variables such as prices, sizes, languages and screenshots of the apps?

Using resampling methods (bootstrap)



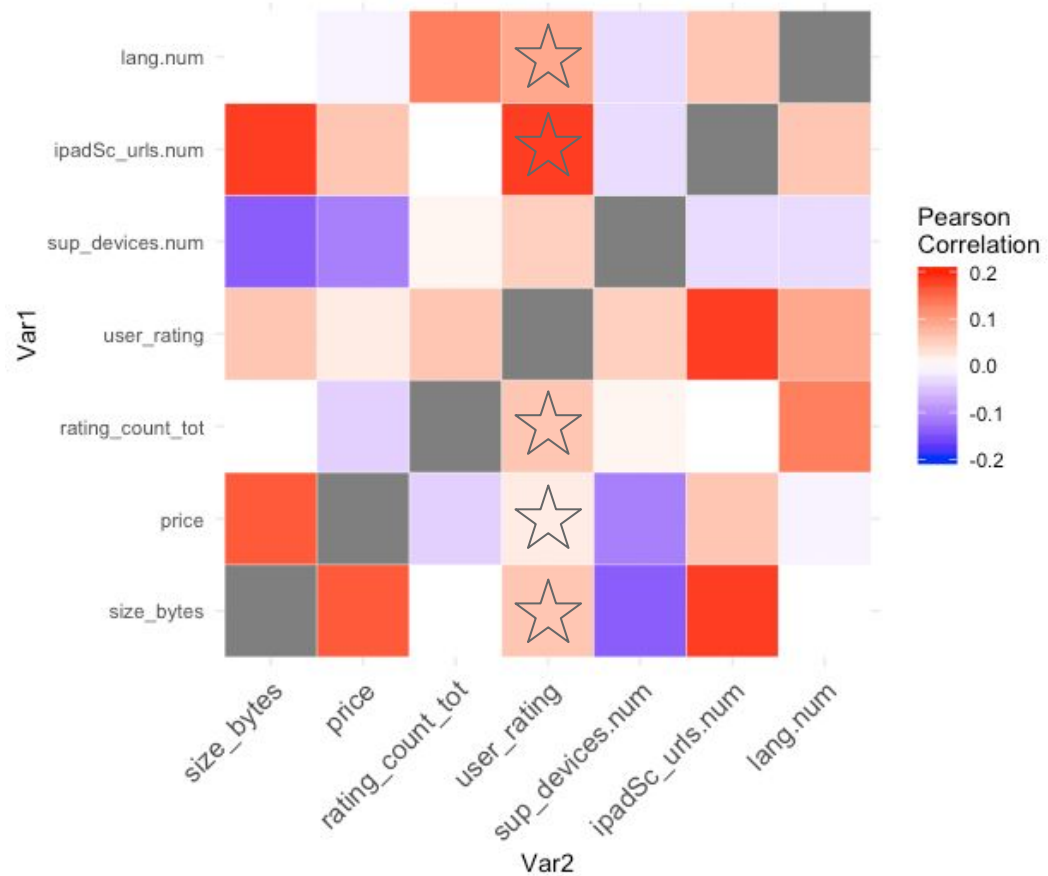
What could be some important factors to predict the rating of an APP?

AIC: predicting the rating of a new app using the model coefficients we got.

Analysis of Data

- Main variables:
 - Dependent variable
 - **User rating** (continuous)
 - Independent variable
 - Total rating count (discrete)
 - Ipad Screenshot (discrete)
 - Number of Languages (discrete)
 - Number of Supporting Devices (discrete)
 - Price (continuous)
 - Size in bytes (continuous)

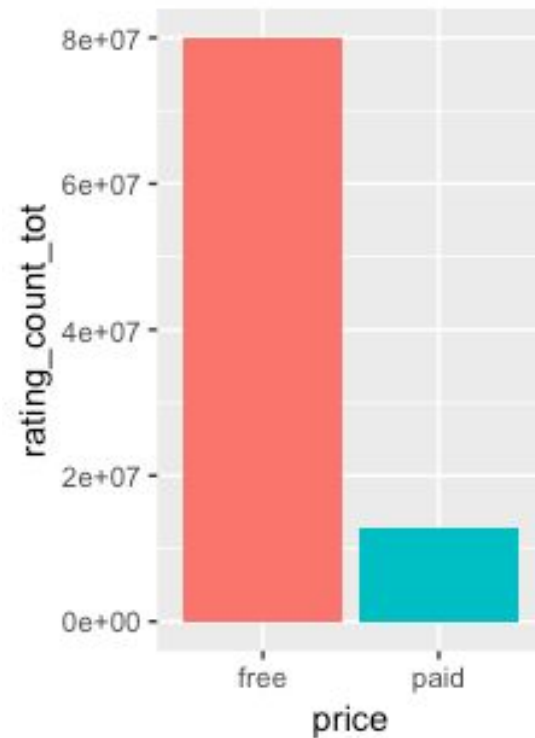
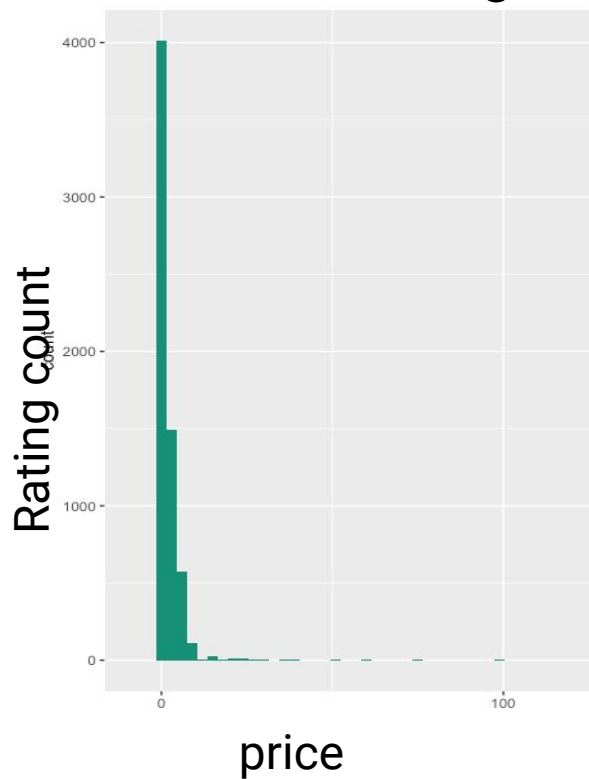
Correlation between variables



4 variables that are most correlated to the average user rating count:
of screenshots, # of languages, size in bytes, and total rating counts

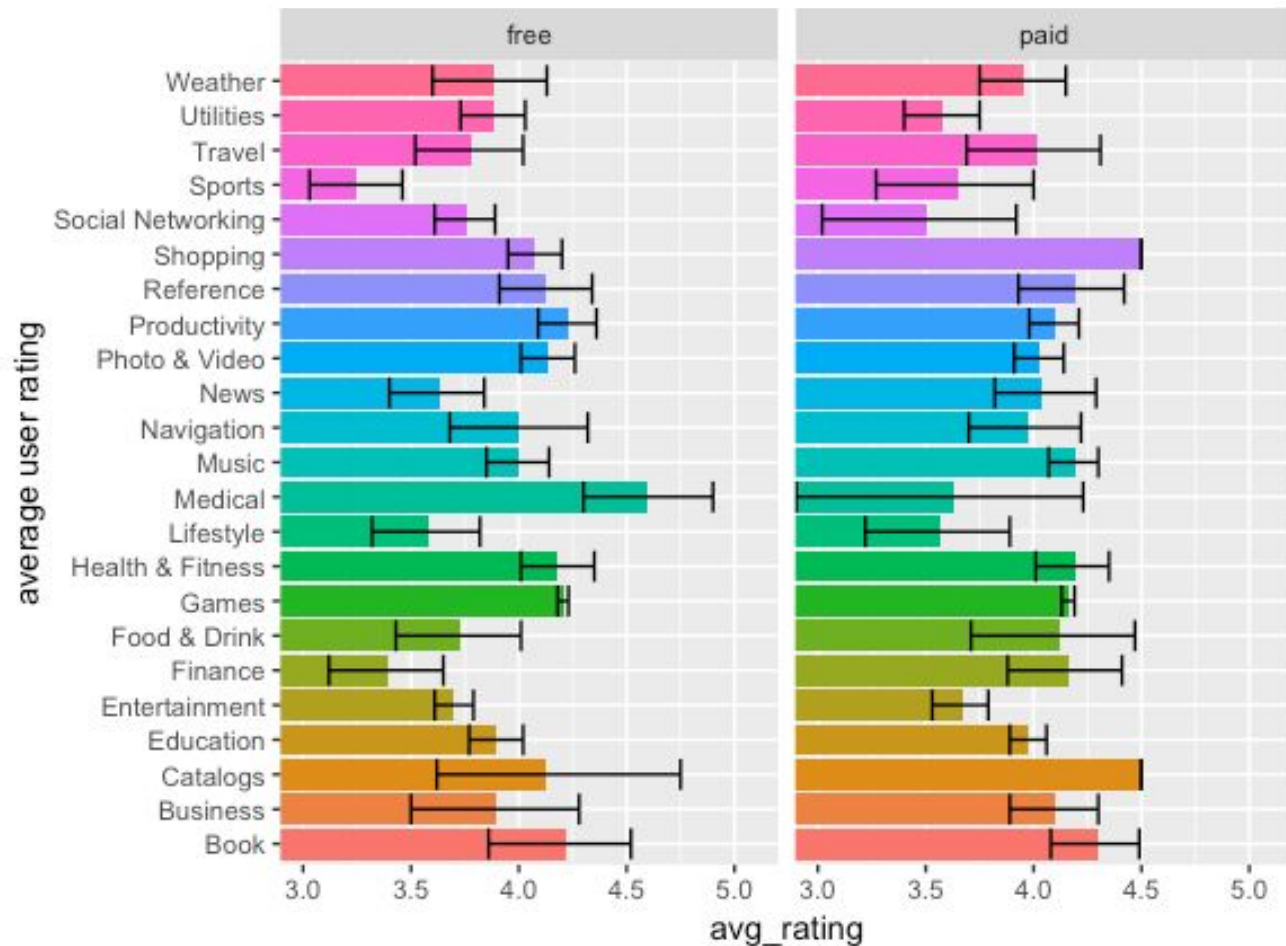
Price
in frequency

Rating count vs price

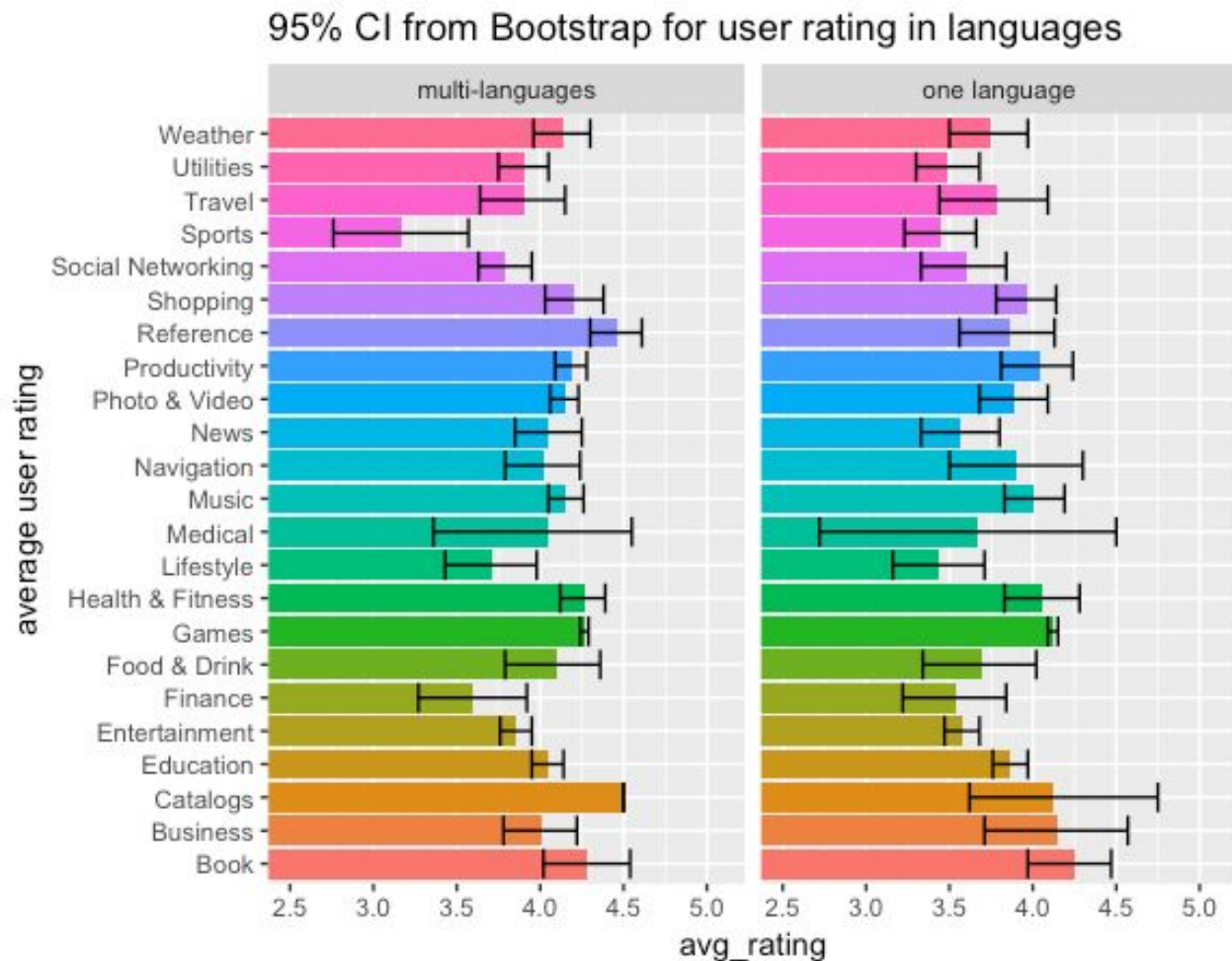


User Rating in Price

95% CI from Bootstrap for user rating in price

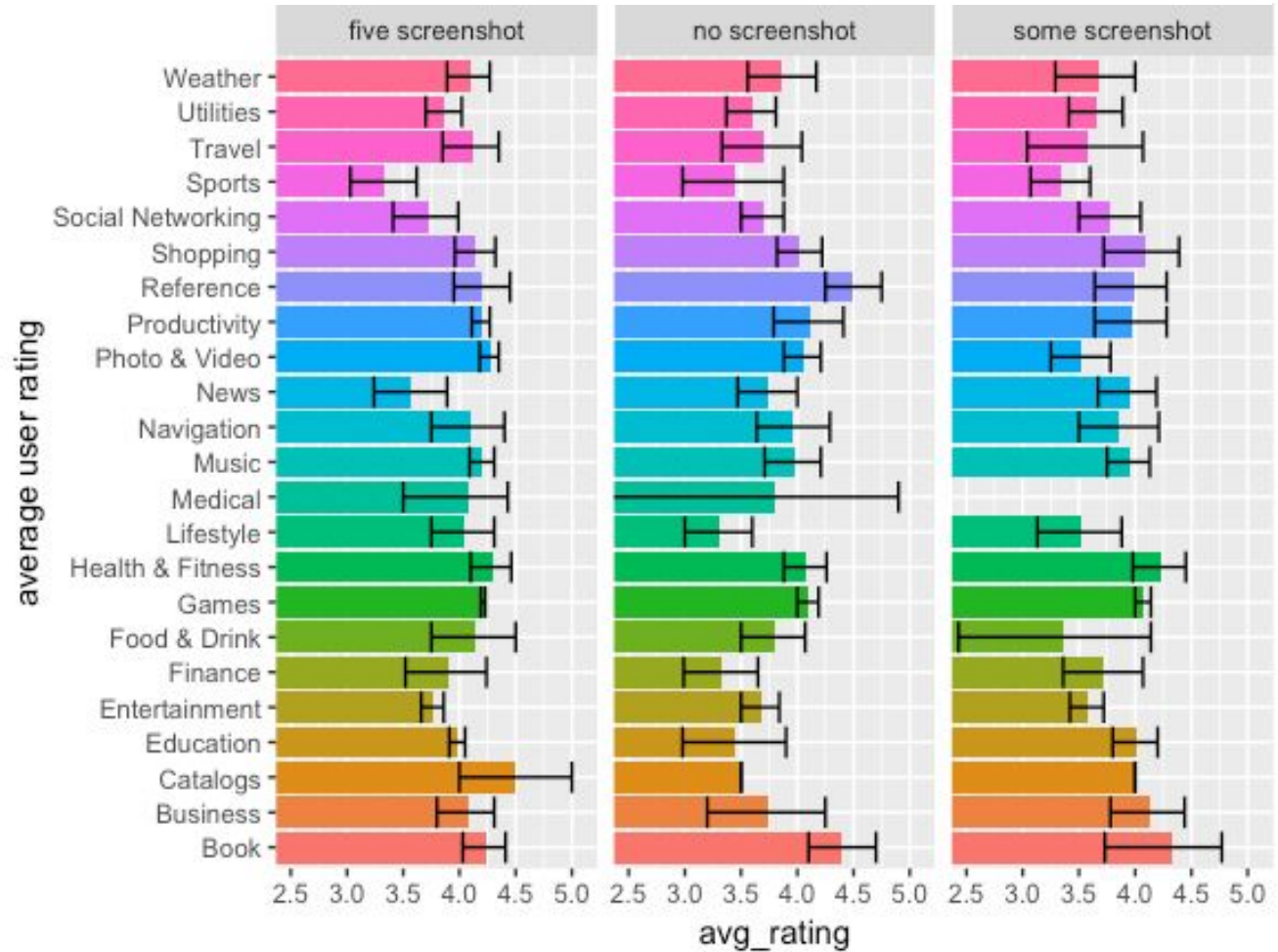


User Rating in languages



User Rating in screenshot

95% CI from Bootstrap for user rating in screenshots



Model of User Rating (overall)

$$y = -0.3251 + 0.0003848x_1^4 + 0.0749x_2^{1/5} + 5.445 * 10^{-37} x_3^{10} + 3.481x_4^{1/50} + 0.00000282x_5^3$$

y = user rating (0-5)
x_1 = number of screen shot
x_2 = number of languages
x_3 = size in MB
x_4 = total rating count
x_5 = number of supportive devices

Model of User Rating (Paid & Free)

Model for **free** apps:

$$y = -0.7561 + 0.0002714x_1^4 + 0.1543x_2^{1/8} + 0.06206\log(x_3) + 3.589x_4^{1/72} + 1.353 * 10^{-7}x_5^4$$

Model for **paid** apps:

$$y = -0.2885 + 0.002175x_1^3 + 0.05908x_2^{1/3} + 1.509 * 10^{-58}x_3^{16} + 3.557x_4^{1/49} + 0.03899x_6^{1/3}$$

y = user rating (0-5)

x_1 = number of screen shot

x_2 = number of languages

x_3 = size in MB

x_4 = total rating count

x_5 = number of supportive devices

x_6 = price

Predict the user rating of an app

- For example:

- For a *paid* game app (\$0.99) with the *3 screenshots* in ipad Itunes stores, *4 usable languages*, *size of 34.4 MB*, *total rating count of 6334*, and *10 supportive devices*, what is the predicted user rating disregarding its content?
- predicted user rating :

$$y = -0.2885 + 0.002175x_1^3 + 0.05908x_2^{1/3} + 1.509 * 10^{-58}x_3^{16} + 3.557x_4^{1/49} + 0.03899x_6^{1/3}$$

$$x_1 = 3, x_2 = 4, x_3 = 34.4, x_4 = 6334, x_5 = 10, x_6 = 0.99$$
$$\rightarrow y = 4.15561$$

y = user rating (0-5)
x_1 = number of screen shot
x_2 = number of languages
x_3 = size in MB
x_4 = total rating count
x_5 = number of supportive devices
x_6 = price

Thank you

