[insert viral app]

Baes 4 Bayes

**Joe Choo-Choy, Esperanza Hernandez,
Bhrij Patel, Zhixue (Mary) Wang**

Topic Background

- 2018 - 2.1 million apps available on Google PlayStore
- 62 % of app users have anywhere from 1-20 apps on their phone
- Hard to make a successful app
  - 59% of apps don't generate enough revenue to break even on development costs
  - 62% of users will use an app less than 11 times

Intro to the Data

- Around 10,000 apps
- 13 variables each
  - application name
  - category
  - rating
  - reviews
  - size
  - number of installs
  - type (paid or free)
  - price
  - content rating
  - genres
  - date last updated
  - current version
  - Android version

Research Question

Q: What makes a viral Google Play Store app?

Cost?

Free apps?

Content rating?

Everyone 10+?

Category?

Dating apps?

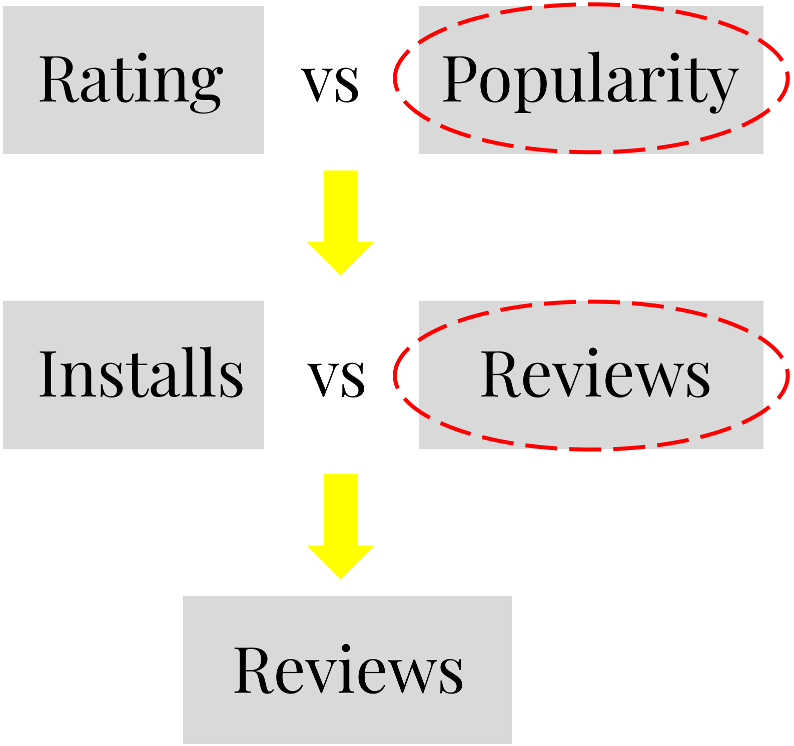
Beauty apps?

Last update year?

Download size?

Version?

Choosing a Metric



Rating vs Popularity

Installs vs Reviews

Reviews

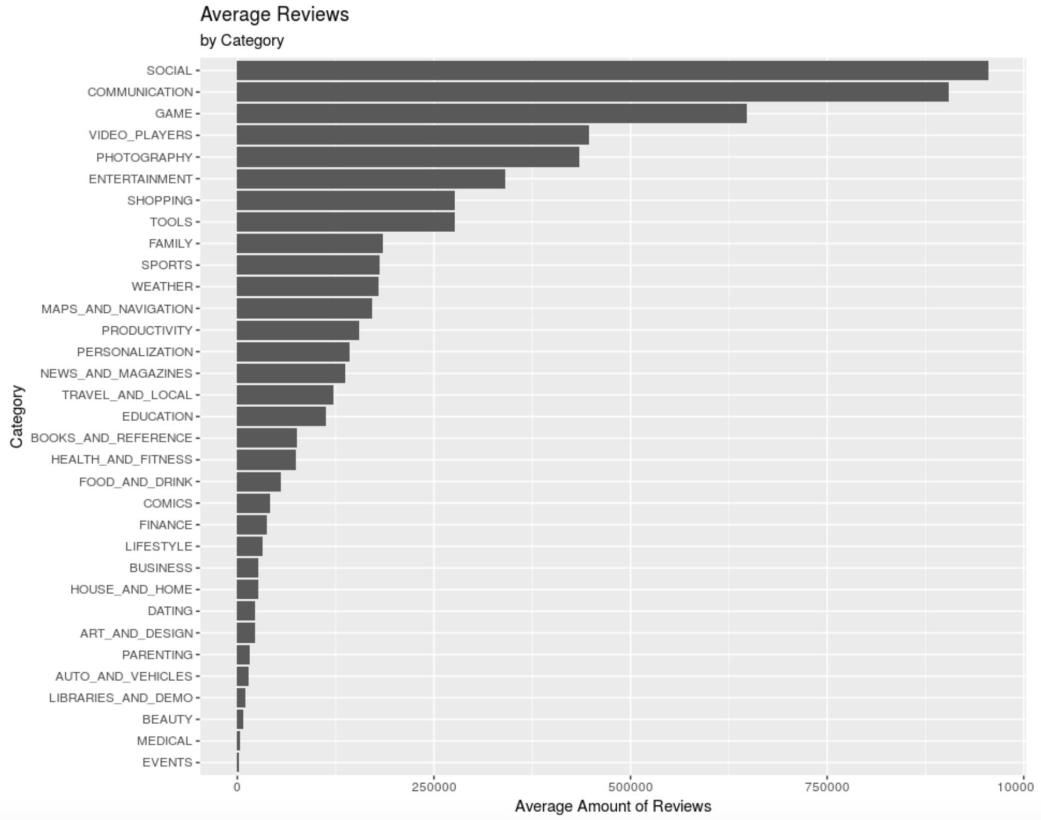
- Lack of variation in rating
  - Between 4 and 5
- Number of installs is categorical
  - E.g. 100,000+
- Correlation between installs and reviews

Cleaning the Data

- General
  - Varies with device ↣ NA
- Removing units to create numerical data
  - Size (60M ↣ 60)
  - Price ($0.99 ↣ 0.99)
- Duplicates
  - Averaged number of reviews

Data Analysis - Ranking Categories

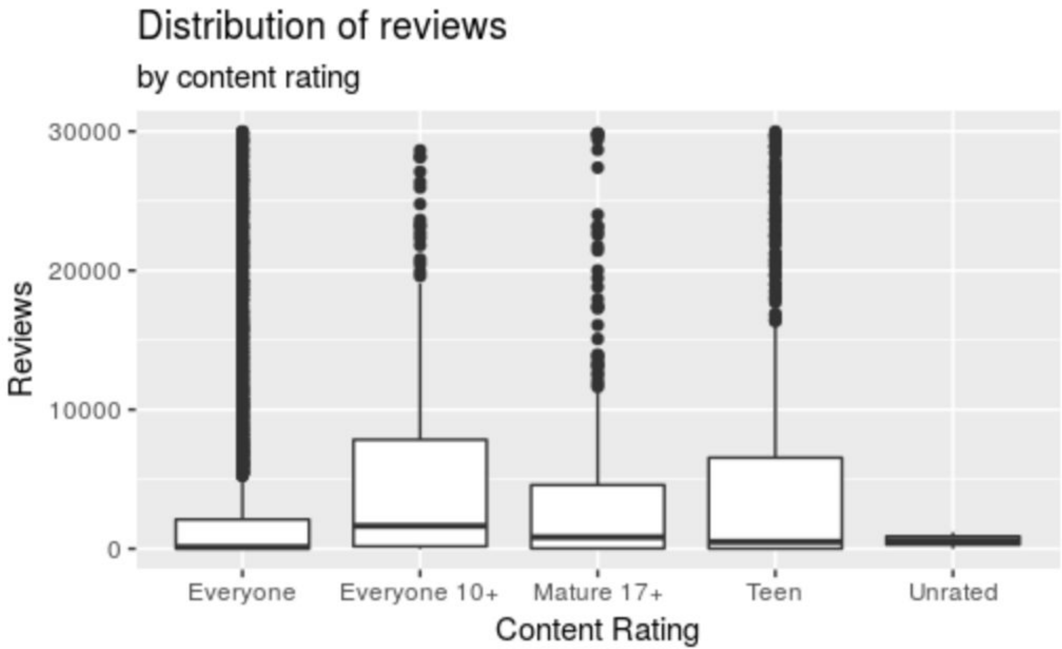
Social and Communication apps with the largest average amount of reviews



Data Analysis - Content Rating

Compared distribution of reviews by content ratings

Highest median: everyone 10+

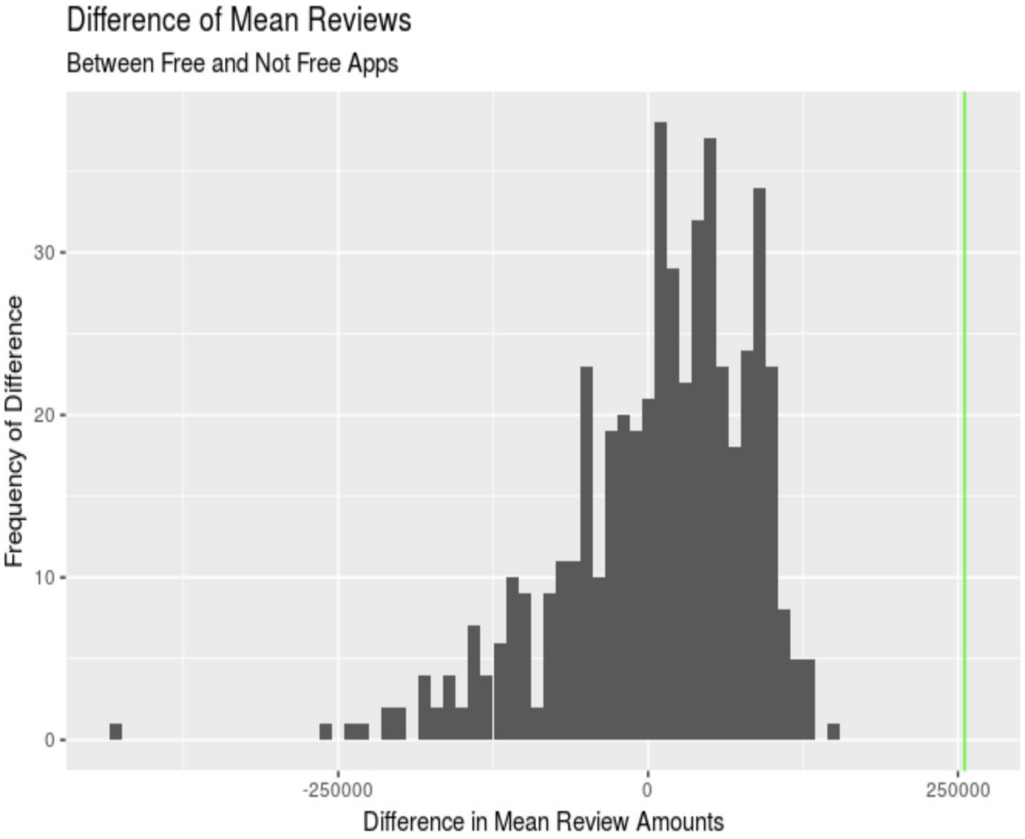


Data Analysis - Hypothesis Test

One-tailed test
Null H: No difference in mean amt of reviews between free and paid apps

Alt H: Greater mean amt of reviews for free apps

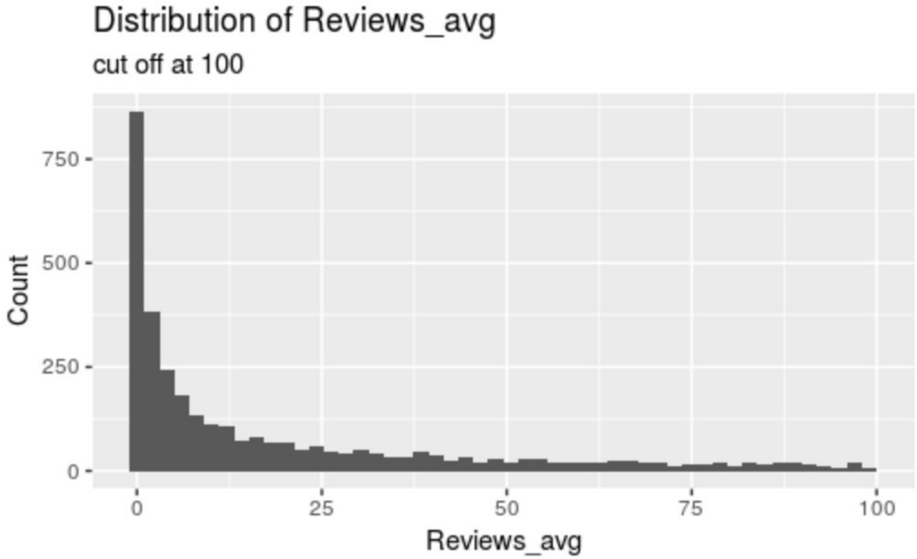
Sample Stats: 255,438
P-value: 0.



Regression Model

log(Reviews_avg) ~ Category + Size + Type + Price + Content_Rating + Type * Size

- No rating/installs
- No genre (collinearity)
- No current_ver



Regression Model

log(Reviews_avg) ~ Category + Size + Type + Content_Rating + Type * Size

- CategoryENTERTAINMENT (16.7) vs CategoryMEDICAL (0.13)
- Size (1.05)
- Paid (0.26)
- Everyone 10+ (4.32)

Conclusion

- What are the characteristics of a viral app?
    - Size = Large
    - Type = Free
    - Category = Entertainment
    - Content Rating = Everyone 10+

Discussion

- Strong right skew
    - Majority zero reviews
    - Few, very viral apps
        - Mean vs median
- Reliability
    - How were the 10,000/2 million apps chosen?
- Recency
    - Last updated 2 months ago
- Next steps: self-web scraping!

Thank you!