



Problem to Solve

Did Apple Store apps receive better reviews than Google Play apps?

Approach

01 – Data Sourcing & Cleaning

- Download the mobile app statistics datasets for Google and Apple Stores from Kaggle
- Examine the datasets to fix errors and remove outliers
 - Remove the records in the datasets where the app price has an error
 - Remove the records in the datasets where the app review count is 0
 - Convert the review counts in the datasets to numerical numbers

Google Play Store Apps

Web scraped data of 10k Play Store apps for analysing the Android market.



Data Code (808) Discussion (79) Metadata

About Dataset

Context

While many public datasets (on Kaggle and the like) provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, I found out that iTunes App Store page deploys a nicely indexed appendix-like structure to allow for simple and easy web scraping. On the other hand, Google Play Store uses sophisticated modern-day techniques (like dynamic page load) using JQuery making scraping more challenging.

Usability 7.06

License Unknown

Expected update frequency Not specified

Mobile App Store (7200 apps)

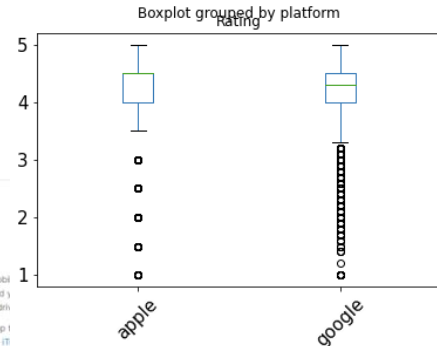
Analytics for Mobile Apps

Data Code (104) Discussion (24) Metadata

About Dataset

Mobile App Statistics (Apple iOS app store)

The ever-changing mobile landscape is a challenging space to navigate. The percentage of mobile holds about 53.2% of the smartphone market, while iOS is 43%. To get more people to download, easily find your app. Mobile app analytics is a great way to understand the existing strategy to drive. With million of apps around nowadays, the following data set has become very key to getting top 1 contains more than 7000 Apple iOS mobile application details. The data was extracted from the IT and Linux web scraping tools were used for this study.





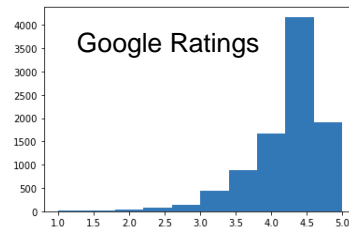
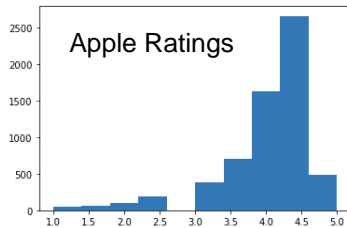
Approach

02 – Modeling (Hypothesis Test: Significance Level of 0.05)

- Hypothesis test formulation
 - H_0 : the observed difference in the mean rating of Apple Store and Google Play apps is due to chance (and thus not due to the platform).
 - H_a : the observed difference in the mean ratings of Apple and Google users is not due to chance (and is actually due to platform)
- Get the distribution of the data
 - Check whether the ratings for Apple and Google apps are in normal distribution
- Create 10,000 permutations
 - Calculate the mean ratings for Google and Apple apps and the difference between these for each one
 - Take the average of all of these differences
 - The p-value of our observed data is the proportion of the data given the Null that's at least as extreme as that observed data
 - If p-value ≤ 0.05 , we reject the null hypothesis; otherwise, we accept the null hypothesis.

Results

- Both the ratings of Apple and Google apps are not in normal distribution



- Observed mean difference in the ratings between Apple and Google apps is 0.14

	count	mean	std	min	25%	50%	75%	max
platform								
apple	6268.0	4.049697	0.726943	1.0	4.0	4.5	4.5	5.0
google	9366.0	4.191757	0.515219	1.0	4.0	4.3	4.5	5.0

- Calculate the number of the differences which are at least as extreme as the observed mean difference
 - There are zero for both positive and negative extremes compared to the observed mean difference from the permutation test
- Therefore, the p-value of our observed data is 0 and we reject the null hypothesis



Conclusion

- Our findings suggest that platform does impact on ratings
- Specifically, we suggest to integrate **only Google Play** into their operating system interface