Final Report

Brianna Li, Junjie Zhou, Chandler Glat, Harrison Locklear, Sunghwu Song

May 05, 2021

INTRODUCTION

Applications and technology provide the convenient ability to control critical aspects of our lives today. From fashion to language, the world is so often swayed by this demand for convenience, and as a result, the app development market is expected to reach a trillion dollars by 2023. Applications have grown from small calculators and games on handheld devices to developing entire companies and cultures. Social media conglomerates such as Instagram and Facebook are now platforms for content creators and individuals to launch businesses; however, for every successful app, there have been thousands of failures attempting to create something unique.

Our group was interested to see if we could use the Google Application Store data to understand what drives success in the mobile apps market. After examining the Google Play Store Apps dataset, we were curious: how can we predict whether apps will succeed based on the app's qualities such as category, rating, price, and installations? Applications provide the foundations of numerous cultural and pecuniary advancements, and by predicting drivers for success, we can help new and established app developers alike. Our group defined such success of an app by two main factors: installations and ratings. We were curious if popular apps drew large numbers of installations and high marks on their ratings and decided to explore the two factors in further detail.

The first question we investigated was what factors lead to high ratings for apps? For less established app developers or companies, downloads are an important metric, but they must first build a reputation of creating a high quality product. As such, ratings must be investigated. The second question we investigated was what factors lead to high installations for apps? Given that success is dependent on high ratings and installations, downloads may prove to be a more important driver for established developers and businesses who already have a strong brand name and track record.

Our group analyzes installations and ratings in depth to predict whether or not an app will succeed and become the next relevant trend, or fail to leave an impact and join the ranks of the vast majority of other apps that are not "important." In cryptocurrency such as DogeCoin, people have made as much as a billion dollars because it had high interaction and coverage despite the fact that it is inherently worthless. Likewise, our group hopes that by learning how apps are popularized we can identify critical patterns and utilize predictive modeling techniques to help developers find long term success.

DATA

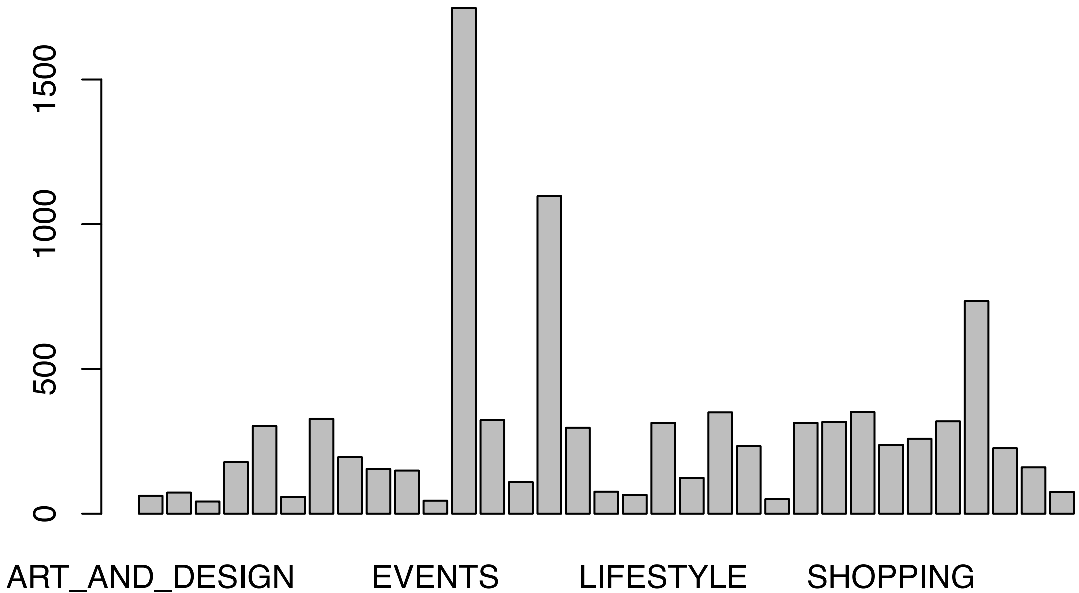
While many datasets on Kaggle and other websites contain information on the Apple App Store, there are not as many similar datasets that contain data for the Google Play Store apps. The data we used was compiled by Lavanya Gupta, a Machine Learning Engineer at Housing.com. She noticed in her research on app stores that while the iTunes App Store deploys a nicely indexed appendix-like structure, the Google Play Store uses sophisticated modern-day techniques, such as dynamic page load, using JQuery. The difference between the organizational structure of the two app stores made web scraping the iTunes App Store much more simple and easy to do whereas the techniques used in the Google Play Store made web scraping more challenging.

Gupta was able to overcome these challenges and compiled the dataset to scrape straight from the Google Play Store to create a dataset with around 10000 observations. Each row contains 13 variables:

- App (application name)
- Category (category of apps that it belongs to)
- Rating (overall user rating of the app)
- Reviews (number of user reviews for the app)
- Size (size of the app)
- Installs (number of user downloads/installs for the app)
- Type (paid or free)
- Price (price of app)
- Content Rating (age group the app is targeted at - Children / Mature 21+ / Adult)
- Genres (an app can belong to multiple genres apart from its main category)
- Last Updated (date when the app was last updated on Play Store)
- Current Version (current version of the app available on Play Store)
- Android Version (min required Android version)

Among these variables, App was the only character-based free response variable, whereas rating, size, installs, price were defined numerically, and all other variables were limited to certain types of character-based responses. The variables we examined more thoroughly in our report are Category, Rating, Reviews, Size, Installs, Type, Price, and Content Rating. Below we have displayed the first 10 rows of apps with the relevant variables that we looked at along with a figure to show the amount of apps in each category to show the relationships between each category that we try to uncover within our variables.

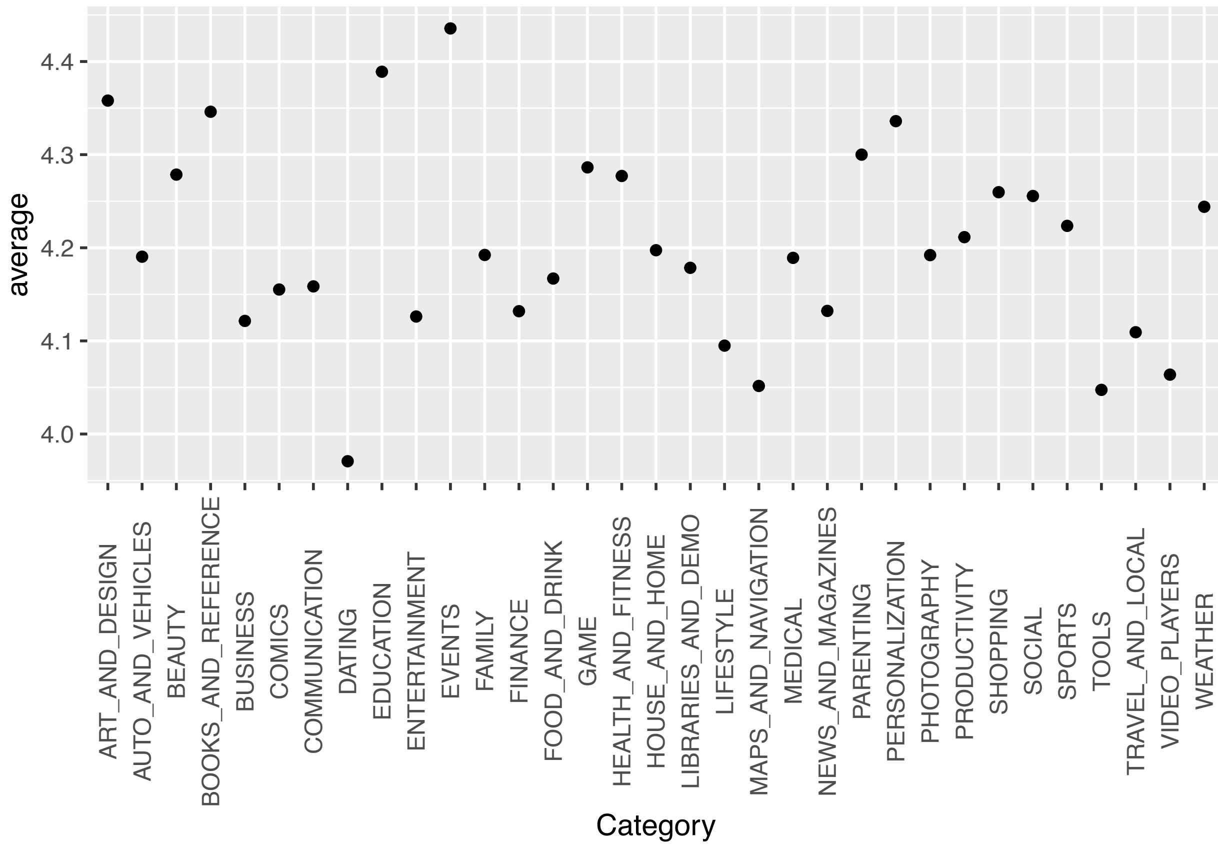
Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating
ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone
ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone
ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone
ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen
ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone
ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free	0	Everyone
ART_AND_DESIGN	3.8	178	19M	50,000+	Free	0	Everyone
ART_AND_DESIGN	4.1	36815	29M	1,000,000+	Free	0	Everyone



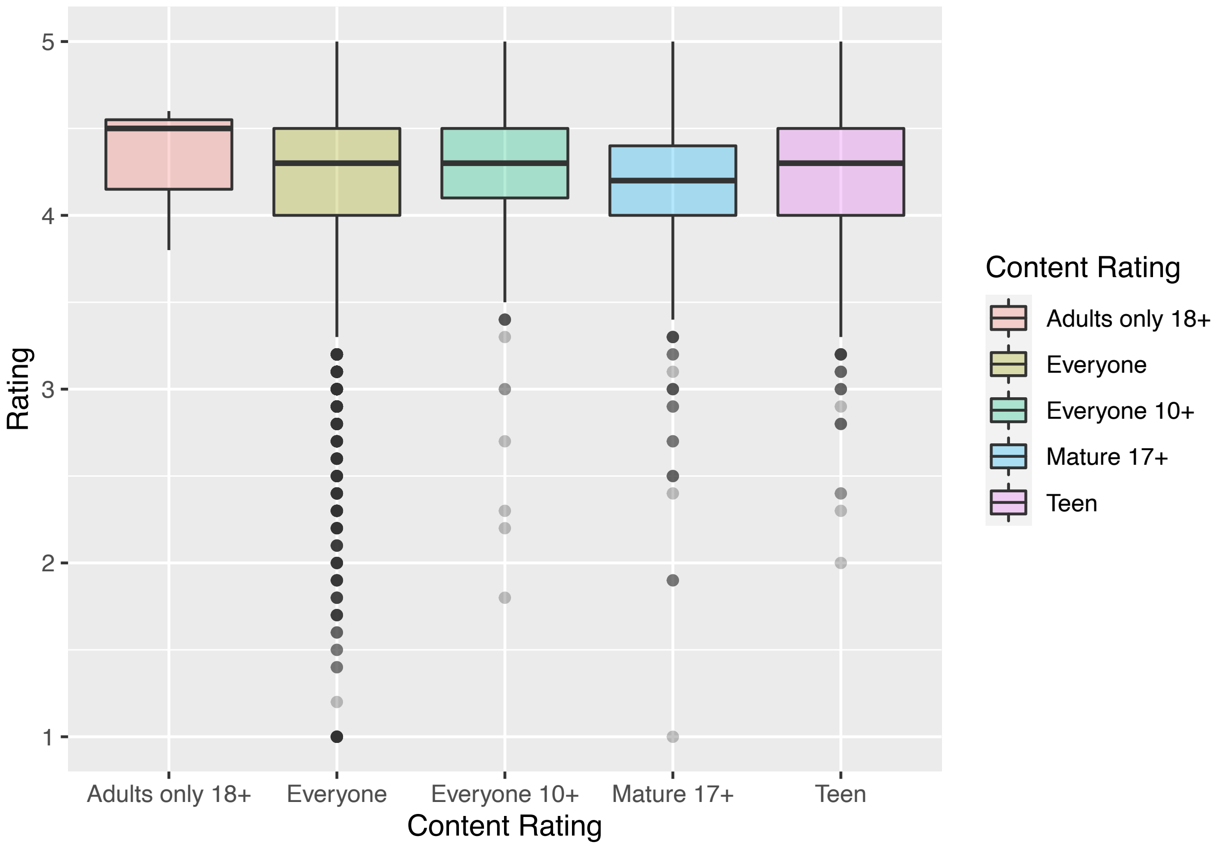
RESULTS

Question 1:

In answering our first question, we focused on building a series of visuals and utilizing predictive modeling to see how we could drive higher user satisfaction. This part of the analysis is predominantly geared towards new app developers who would prioritize user ratings over installations, especially as they try to build a brand and strong track record. First, we were curious if specific app categories outperformed in terms of ratings. To explore this, we grouped our data by Category and calculated the mean rating by category. Then we plotted the average ratings by Category and were able to see that the best performing categories were Entertainment, Education, Art and Design, Books and Reference, and Personalization. We then conducted a series of t-tests to see if these average ratings were significantly different than the rest of the apps, and the table shows that these apps all had significant p-values. As a result of this analysis, we would suggest that developers predominantly target these specific application categories.

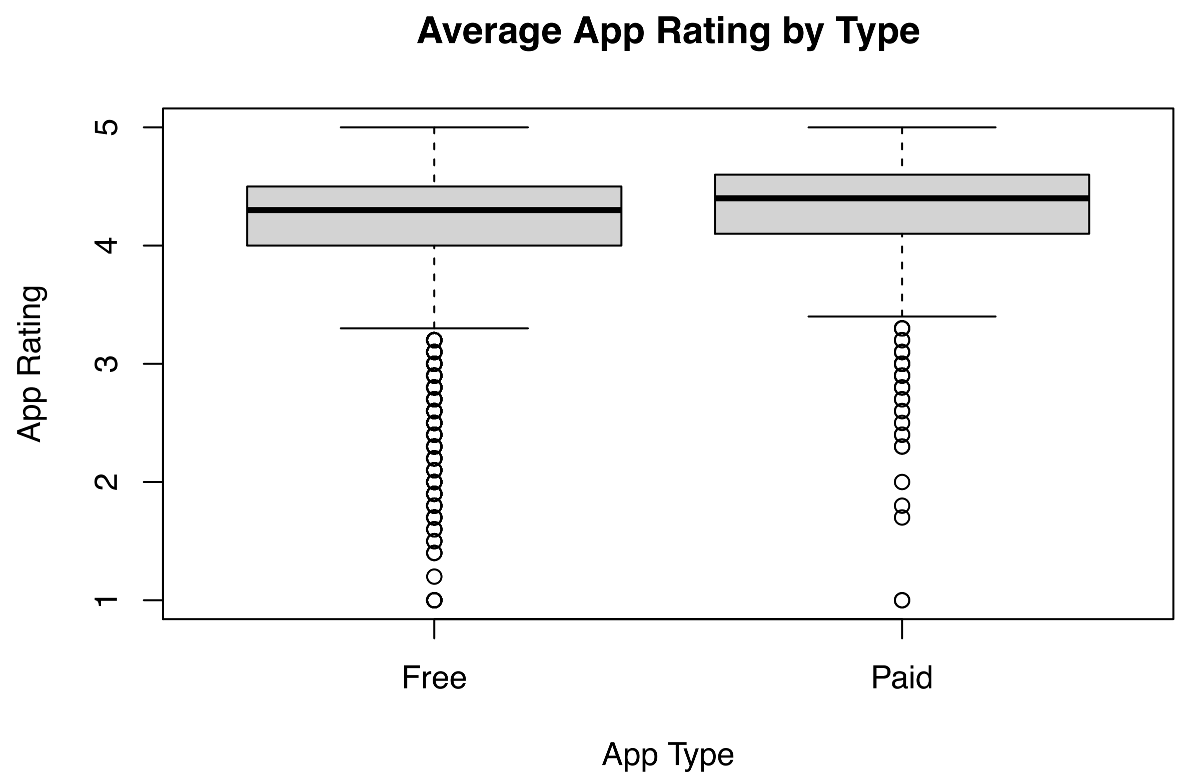


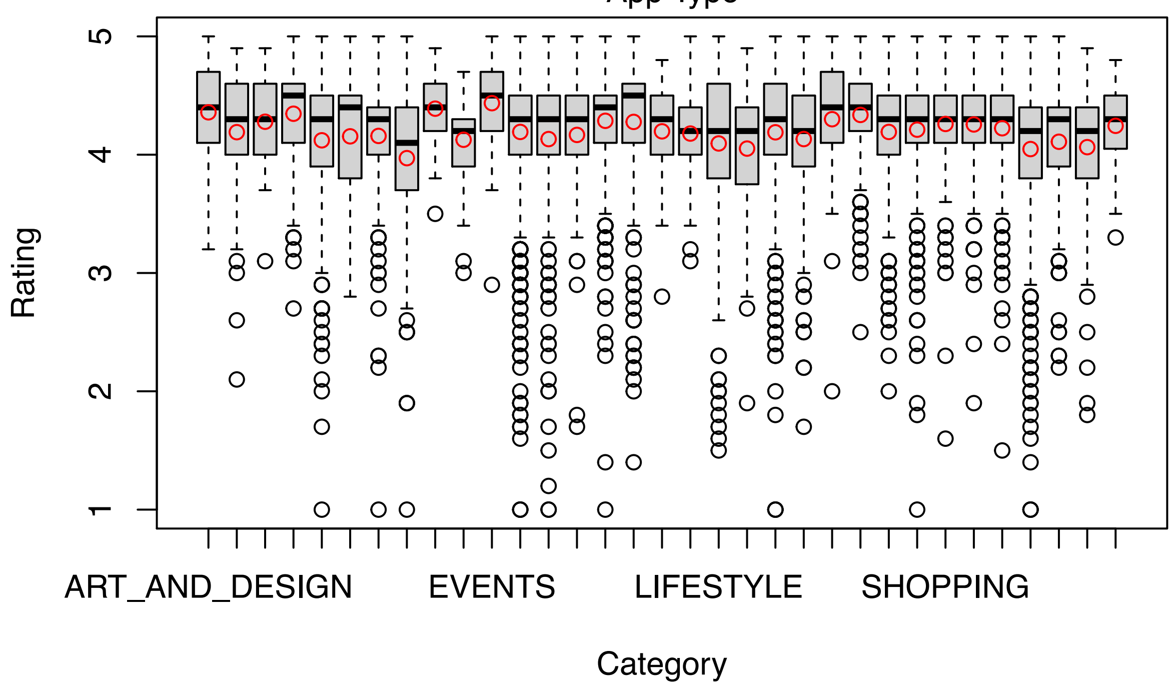
We also explored the relationship between the average user rating and the content rating of each app. Using a series of boxplots, we plotted this relationship between rating and the 6 content ratings: Adults only 18+, Everyone, Everyone 10+, Mature 17+, Teen.



We found that the Adults only 18+ content rating had the highest average rating out of all the age groups/demographics with no outliers. The other content ratings had similar average ratings to each other. We also found that because more apps fell under the Everyone content rating, there were a higher number of outliers for ratings. Thus, through this analysis we recommend that app developers create apps that are geared towards an Adults only 18+ audience to yield a higher average rating with a lower chance of it falling outside of the confidence interval.

We were also curious to look into the impact of pricing strategy on app ratings, specifically seeing if there is a significant difference between free and paid apps by rating.

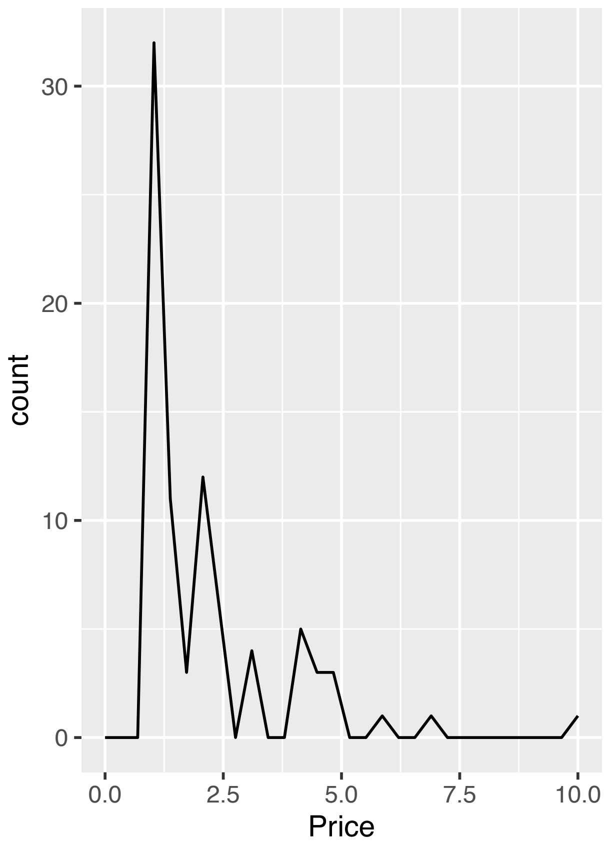


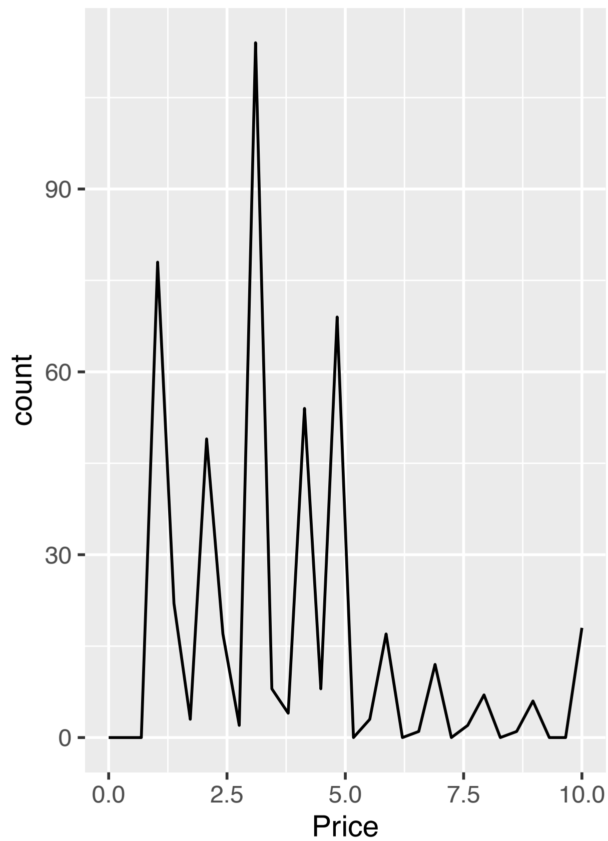


	P-Val
Entertainment	8.988e-03
Education	2.200e-16
Art and Design	6.000e+00
Books and Reference	5.480e-06
Personalization	2.883e-11

From the boxplot we can see that there is not a huge difference in terms of rating distribution. However, after running a one-tailed independent t-test of ratings of free apps vs paid apps, we can see that paid apps are actually rated significantly higher than free apps. Our t-test yielded in a p-value of 0.0003149 which is significant at the 5% level. Therefore, we reject the null-hypothesis that the ratings of the free and paid apps are equal. Logically, this would imply that individuals who buy apps are more inclined to give higher ratings since they purchase with an expectation of being satisfied. Therefore, for new app developers, we recommend utilizing a paid model as opposed to a free version of the app.

We also wanted to analyze the relationship between the top rated categories for apps and their price distribution. This would give us, as well as developers a better understanding of how apps at different price points perform in terms of user satisfaction and reviews in the app store. The frequency plot on the left contains the apps within the most popular categories by rating while the graph on the left contains the rest of the app categories. When looking at the graphs you can see more popular apps have a lower price distribution, primarily targeting $1.99. Therefore, in addition to using a paid model, we suggest new developers price their apps around $1.99 specifically to maximize ratings.





Lastly, in analyzing ratings, we wanted to build a predictive model to see what variables were optimal when predicting app ratings. We found that there was a linear relationship between ratings vs reviews + size, with both predictors variables have significant p-values. Therefore, we can conclude that reviews and app size can predict a lot of the variation in ratings. The implications of having such a significant model are that developers can set specific rating targets once their app is on the market and have an understanding of how they can structure their app in terms of reviews and size to best meet that target.

Observations	7471 (3370 missing obs. deleted)
Dependent variable	Rating
Type	OLS linear regression

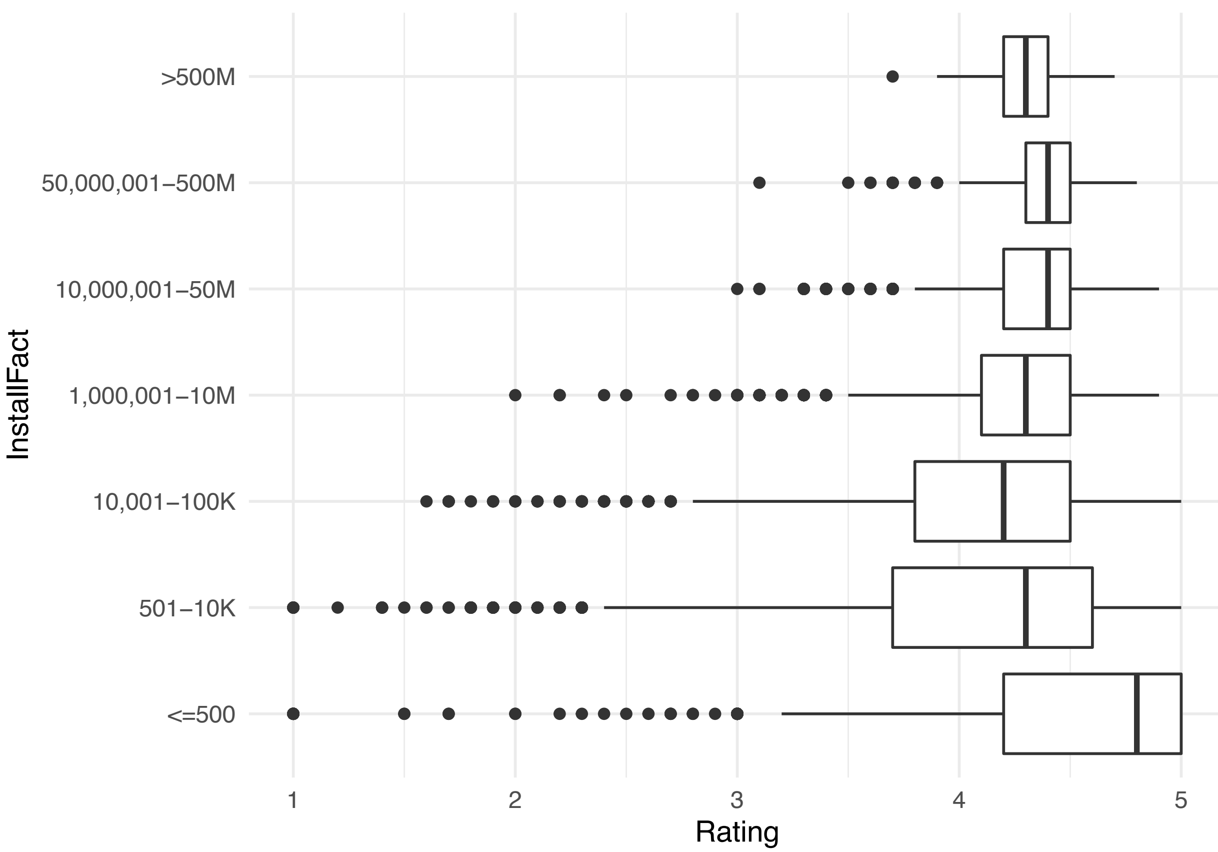
F(3,7467)	25.19
R^2	0.01
Adj. R^2	0.01

	Est.	S.E.	t val.	p
(Intercept)	4.14	0.01	462.94	0.00
Reviews	0.00	0.00	5.49	0.00
Size	0.00	0.00	5.04	0.00
Price	-0.00	0.00	-1.44	0.15

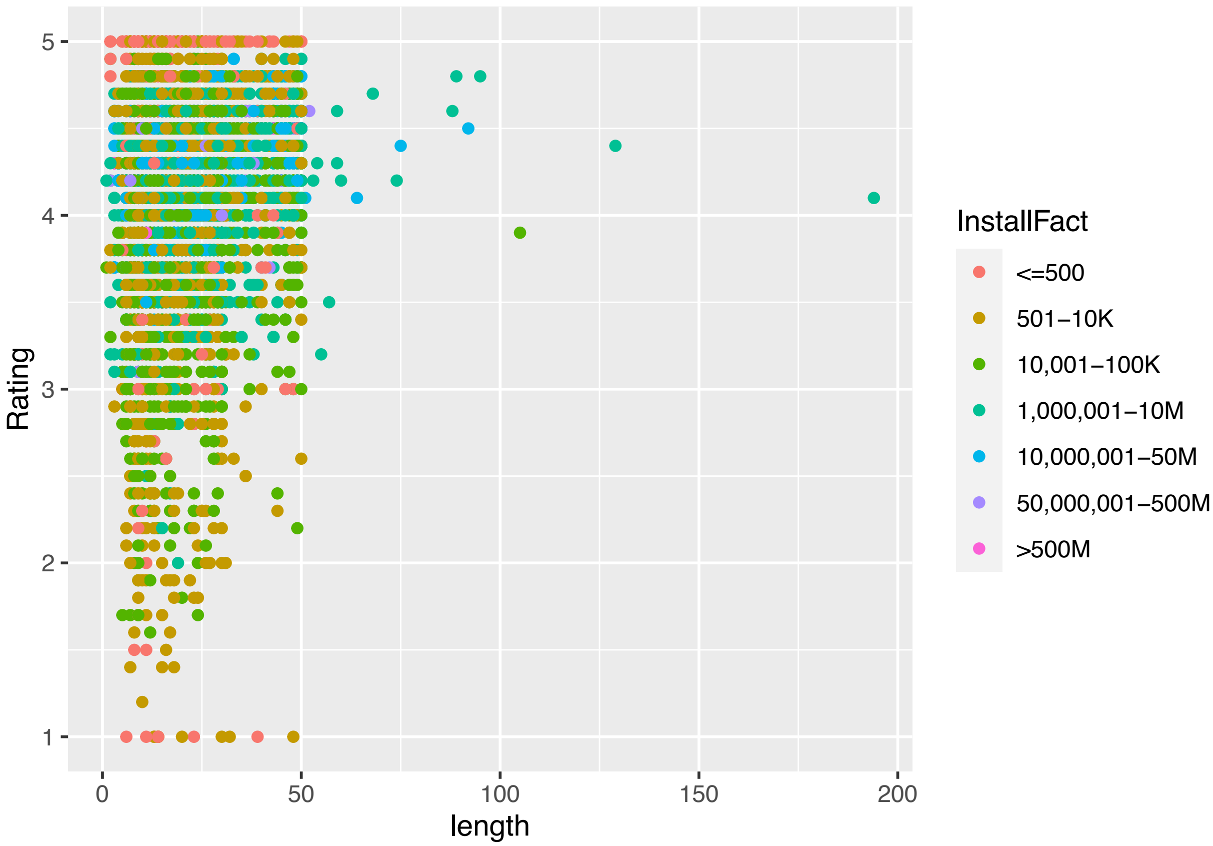
Standard errors: OLS

Question 2:

The next question we looked into was identifying some of the key drivers for installations of apps. The first thing we looked into was the relationship between ratings and installations. To our surprise, there was a slight negative correlation between these two success metrics. Therefore, we decided that we should treat these success factors separately and make the argument that new developers should prioritize ratings while established developers should focus on installations.



Another aspect we considered when looking at what factors contribute to a higher number of installations is the length of the app's name. Now-a-days it is rare to see popular apps have a name that is longer than one word. This is something we wanted to take a closer look at when evaluating the data from the google play store. Does a shorter name for an app lead to a higher number of installations or is this just a coincidence? In the graph below, we can see that it is rare for an app's name to be longer than fifty characters and the apps with higher installs tend to be much lower than that.



Lastly, for predicting installations we used a best subset selection method to find the variables that are significant when predicting the number of installations an app has. We created a new binary variable called viral for apps with over and under 1,000,000 installations. Then, we used the bestglm function to find whether rating, reviews, size, and price are good predictors of number of installs. We found that size and price have a p-value of 0.92 and 0.68 respectively which are both greater than the alpha value of 0.05 so they are not good predictors. Meanwhile, rating and reviews both had an extremely low p-value of less than 0.01 which is less than 0.05 so they are good predictors of installations. This model is significant because it can help developers structure their apps while predicting whether it will go viral or not.

	Estimate	P-value
Rating	-0.4900	0.01
Reviews	0.0005	0.00
Size	0.0005	0.92
Price	-9.0600	0.68

CONCLUSION

To say technology is interwoven in our modern society might be an understatement. Apple's "There's an app for that" campaign was twelve years ago and now more than ever is that statement true. There are currently over 2.8 million apps on the Google Play Store at this moment and over 3000 new apps added daily. Even though those numbers are staggering, the average person's phone has only 40 apps installed and a majority of their time on their phones is centered around just 18 apps. This means that while the mobile applications market is expected to reach over $1T by 2023, the revenue is vastly skewed to those dominant few apps that seem to be on everyone's phone. But, our group wanted to look at what new and established app developers can do to try and rival some of the top apps currently on the market. Drawing on our prior analysis, we set out to answer two questions: 1) How can app developers best structure their applications to yield high ratings and user satisfaction; 2) How can we best predict installations for developers. As mentioned earlier, there appears to be a negative relationship between ratings and installations, which is why we treat these two success metrics separately. Focusing on the conclusion for our first question: We found a significant difference using a series of t-tests between the top 5 most popular app categories by rating relative to the rest of the apps. Therefore, we would suggest new app developers should target the following categories: Entertainment, Education, Art and Design, Books and Reference, and Personalization. We also found that paid apps tend to outperform free apps in terms of ratings, and accordingly we'd suggest that new developers should utilize a paid model, enabling them to build a more tight-knit community of users as they build their brand. In line with this pricing model, we looked into the price distribution of the top-rated categories relative to the other categories. From this, we found that the most popular app categories have a lower price distribution, primarily revolving around the $1.99 price range. Therefore, we would suggest new developers should use a paid model and with a price point of around $1.99. We also found that for the relationship between ratings and content ratings, the Adults only 18+ content rating group yielded the highest average ratings out of all the age groups and no outliers outside of the confidence interval. Thus, we suggest app developers target this select age group of Adults only 18+ for their apps. Lastly, we built a linear model that could be used to predict ratings. This model had a significant F-Value, and we selected quantitative predictors with significant p-values – reviews and size. This has significant implications for developers who can leverage the model to understand going forward how they can best obtain a certain rating.

Focusing next on our second question: The first thing we analyzed was the relationship between ratings and installations. We found that apps that have higher ratings are correlated with lower installations. Therefore, we conclude that established app developers who already have a track record should focus on installation as their primary value driver. One of the main things we looked at for installation was seeing how the length of the app name is correlated with installations. We found that the apps with the highest number of downloads tend to be below 50 characters, and so we would suggest developers focus on shorter app names in terms of length as a way to optimize their installations. For installations, we also focused on predictive modeling, specifically to see what factors can predict whether an app will go viral or not. We defined a viral app as having over 1 million downloads and built an indicator variable called 'Viral'. We used the best GLM for our model selection method and ultimately determined that the best model for predicting virality includes ratings and reviews. This is significant because it can help an established developer structure their app in order to predict whether it will exceed million downloads. (After an app has over 10,000 installs, the average rating steadily increases towards 5. However, this trend may be caused by the fact that popular apps get higher ratings. Something interesting that we found is that apps with less than 500 installs have the highest average rating of over 4.75. We speculate that the high average rating is due to companies buying fake reviews.)

Ultimately, drawing on our conclusions for predicting high ratings and high reviews, we essentially looked into the relationship between variables from a visual perspective, providing a roadmap for new and established developers in terms of what factors they should focus on. We also placed a lot of emphasis on predictive modeling and model selection methods to determine what are the best drivers for ratings and installations respectively. Our goal for this analysis is that developers can thoroughly understand what drivers are highly correlated with their success, enabling them to structure their apps appropriately in terms of app name, size, pricing model, category etc. Additionally, developers can leverage our predictive models to hit critical targets for both ratings and installations as their app hits the market.