What Determines an App's Rating?

Kieran Cottrell, Nicolas Izurrategui, Alex Johnson, and Max Jones
Stat 3220, Statistics Department, University of Virginia

Background

High Usage

- According to *Statista*, mobile phones generated 52.2% of all website traffic worldwide in 2018.

High Profitability

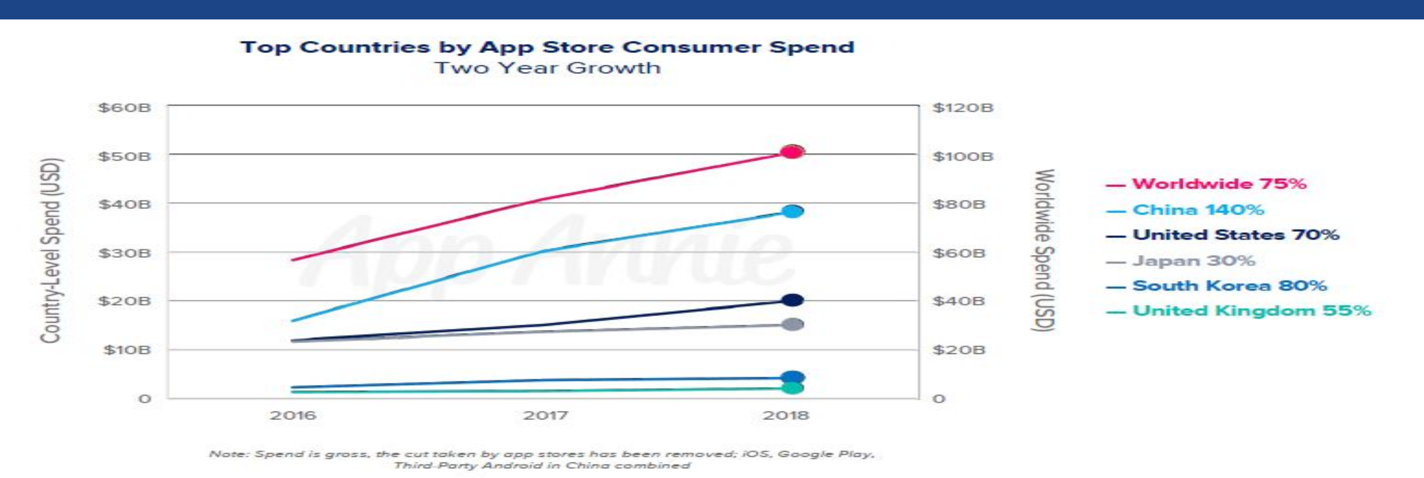
- As a result of this growth, the profitability of apps has also increased significantly. Just as of April 2019, global app revenue reached $19.5 billion, an increase of 17% from April 2018 (Nelson).

User Rating

- It is clear that app creation has the potential to be a highly lucrative venture, and one of the factors playing into whether an app becomes successful is its overall user rating. In fact, user ratings are often the first thing new customers look at to determine if an app is worth installing.

Useful for any App Developers

- For large companies or small developers looking to create successful apps, our findings could be greatly useful. Even developers just looking to make useful apps with no intention of making money could use our findings to better satisfy app users.



Data Summary

Source

- Our dataset comes from the Google Play Store, obtained from Kaggle.
- The dataset contains data for over 267,000 apps from the Google Play Store from April 2019.
- The data was extracted from the Google Play Store using Python web scraping tools. Since the data comes directly from Google, we can be reasonably sure of its validity.

Response Variable

To determine an app's success, we wish to predict an app's average user rating on the Google Play Store (0-5 stars). Since the response variable is an average of each app's ratings, the variable is continuous.

Explanatory Variables

Installs: Number of installations for the app, split into <1 million, 1-10 million, and 10 million+ installs
Price: Price of the app in dollars
Reviews: Total number of reviews for the app
Genre: Primary genre of the app, ranging between entertainment, lifestyle, media, and productivity
Content Rating: Age recommendation for the app, ranging between E (Everyone) and Teen

E vs Teen

We chose to only to investigate apps that had a content rating of E or Teen,to make our large dataset more manageable.

Analysis

Three Stage Model

Stage 1 Model (Quantitative)

Predicted Rating = $\beta_0 + \beta_1$*ln(price) + β_2*ln(reviews)

Global F Test	0.0007 < .1
Adjusted R-sq	0.0611

Predicted Rating = 4.15688 +(0.02632)*ln(price) + (0.03950)*ln(reviews)

Stage 2 Model (Stage 1 + Qualitative)

Predicted Rating = $\beta_0 + \beta_1$*ln(price) + β_2*ln(reviews) + β_3*DummyContentB + β_4*DummyInstallB β_5*DummyInstallC + β_6*DummyCatB + β_7*DummyCatC + β_8*DummyCatD

Global F Test	0.0177 < .1
Adjusted R-sq	0.0530

Nested F- Test
F = 0.71859505548
P=.63607

Stage 3 Model (Stage 2 + Interaction)

Predicted Rating = $\beta_0 + \beta_1$*ln(price) + β_2*ln(reviews)

Global F Test	0.0007 < .1
Adjusted R-sq	0.0611

Predicted Rating = 4.15688 +(0.02632)*ln(price) + (0.03950)*ln(reviews)

Final Model

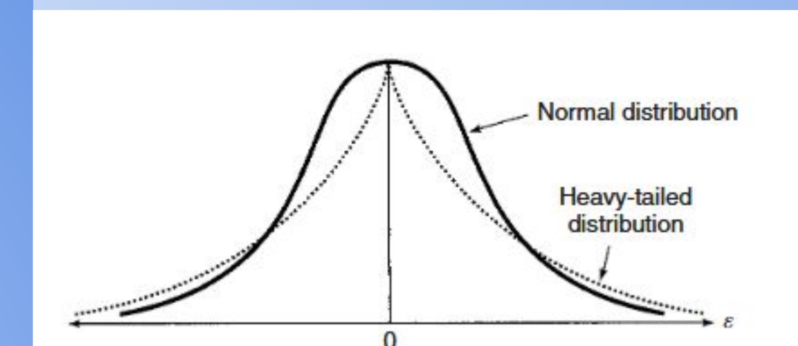
Predicted Rating = **4.15688 +(0.02632)*ln(price) + (0.03950)*ln(reviews)**

Global F Test	0.0007 < .1
Adjusted R-sq	0.0611

Robust Model

Why Use Robust?

- Data presents non-normality of errors, outliers on x or y axis.



Advantages and Disadvantages

- Robust regression yields estimates for the β's that are nearly as good as the least squares estimates when the assumption of normality is satisfied
- Significantly better results for a heavy-tailed distribution

Goodness-of-Fit	
Statistic	Value
R-Square	0.0395
AICR	225.0712
BICR	236.3716
Deviance	20.5832

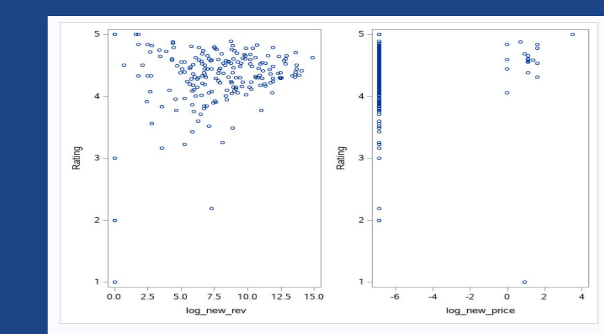
Biggest Differences

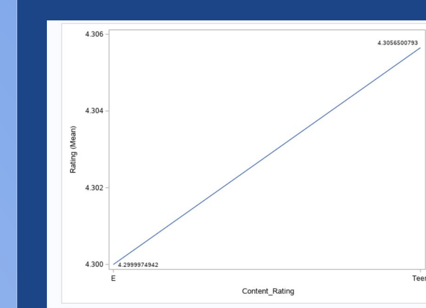
- Robust regression is better at dealing with outliers than traditional Multiple Linear Regression.

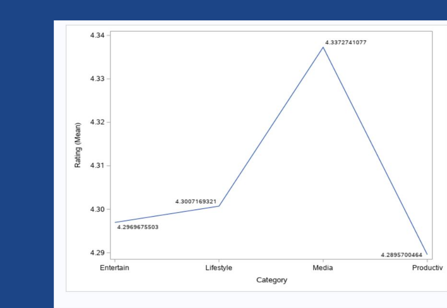
Robust Model:
R-Square = 0.0395

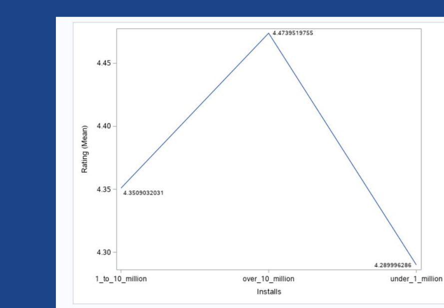
Multiple Linear Regression Model:
R-Square = 0.0705

Early Data Analysis

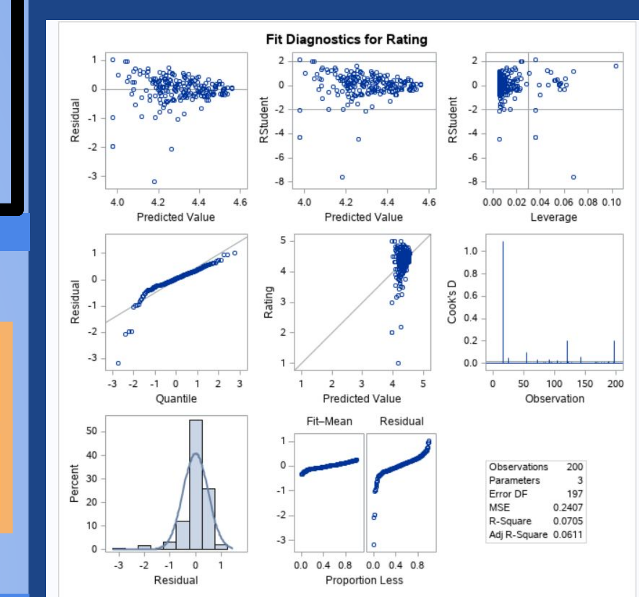


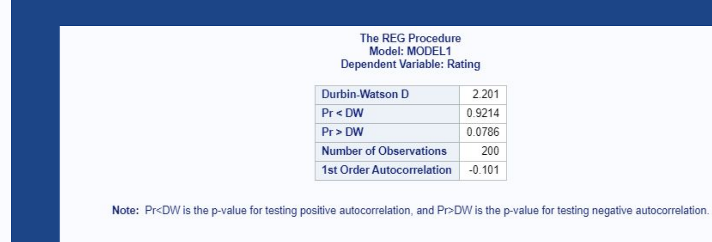






Regression Assumptions





Conclusion

- **Utilizing the log transformation on our quantitative variables favored our overall statistical significance in the model**
- **We were not able to produce a good fit for our model**
- **This might have been due to missing variables / model misspecification**

Was Our Research Question Answered?

- **The model was not substantially useful in predicting user rating**
- **Our beta estimation don't explain the user ratings of apps**