

Make Sure to Drop a Like

The Effect of Consumer Rating on
Application Success



Authors: Oleg Ananyev, Oren Carmeli, Romain Hardy, Sam Rosenberg



Research Question

The **Google Play Store** is one of the major hubs for applications for Android mobile phone and tablet users. Users can download any app for personal consumption.

Using a rich dataset of Google Play Store data, we want to answer the critical question of:

→ *Does having a higher consumer rating score lead to more downloads for Google Play Store applications?*



Data and Modeling

Data Origin:



We leveraged a rich dataset from **Kaggle** that was scraped directly from the Google Play Store

Sample Size:



-10,000 **unique data points** with 11 different variables to use

Initial Variables of Interest:



Consumer rating, Price, Category, Age, Download Size

Modeling Challenges:

Reverse Causality



Initial exploration of the data led us to believe there was a **reverse causal relationship** between # of downloads and # of reviews

Omitted Variables



We had to consider **omitted variables** such as **brand awareness**, **application rankings**, and **total addressable market** (and their impact on our results)

Transformations



We had to choose **appropriate transformations** to apply to each variable and decide **how to handle binned data**



Model Building

Modeled Options:



Small Model: $\text{Log}(\# \text{ Installs}) - \text{Rating}$



Medium Model: $\text{Log}(\# \text{ Installs}) - \text{Rating} + \text{Log}(\text{Size}) + \text{Log}(\text{Version}) + \text{Log}(\text{Last Updated}) + \text{Free/Paid} + \text{Family Category Dummy} + \text{Game Category Dummy} + \text{Tool Category Dummy} + \text{Content Rating: Everyone Dummy}$



Large Model: $\text{Log}(\# \text{ Installs}) - \text{Rating} + \text{Log}(\text{Size}) + \text{Log}(\text{Version}) + \text{Log}(\text{Last Updated}) + \text{Free/Paid} + \text{Content Rating: Everyone Dummy} + \text{Rating} * \text{Family Category} + \text{Rating} * \text{Game Category} + \text{Rating} * \text{Tool Category}$



Model Output

Dependent Variable: *Log(Installs)*

	(1) Model Small	(2) Model Medium	(3) Model Large
Rating	0.252*** (0.037)	0.141*** (0.036)	0.137*** (0.045)
Log(Size)		0.701*** (0.038)	0.702*** (0.038)
Log(Current Version)		0.769*** (0.083)	0.768*** (0.083)
Log>Last Updated)		-1.083*** (0.085)	-1.072*** (0.085)
Is Free		1.407*** (0.055)	1.409*** (0.055)
Family Category		0.026 (0.043)	1.113*** (0.382)
Game Category		0.877*** (0.056)	-0.837 (0.570)
Tools Category		0.296*** (0.063)	-0.765* (0.438)
Rating * Family Category			-0.260*** (0.092)
Rating * Game Category			0.402*** (0.137)
Rating * Tools Category			0.264** (0.110)
Rating: Everyone		-0.209*** (0.043)	-0.207*** (0.043)
Constant	3.968*** (0.149)	2.139*** (0.176)	2.150*** (0.209)

Model Statistics

	(1) Model Small	(2) Model Medium	(3) Model Large
Observations	7,226	7,226	7,226
R ²	0.007	0.248	0.251
Adjusted R ²	0.007	0.247	0.250
Residual Std. Error	1.598 (df=7,224)	1.392 (df=7,216)	1.389 (df=7,213)
F Statistic	54.454*** (df=1; 7,224)	263.871*** (df=9; 7,216)	201.697*** (df=12; 7,213)

Note: *p<0.1; **p<0.05; ***p<0.01



Conclusions

Interpreting the Outcome:

- As expected, **apps with higher ratings, free apps, and older apps** are more likely to be installed
 - Increasing consumer rating by 1 corresponds to a **14.1% increase** in the number of downloads
 - Free applications have 141% more downloads on average than paid applications
- It is surprising that download size has a positive relationship with download count
 - Given the limited space available on devices, we expected smaller apps to be more successful
 - Download size could be an **indicator of production quality**
- Apps with **frequent updates** are installed more
 - More responsive to consumer feedback
- **Interactions between rating and category** may take precedence over standalone variables
 - Games are positively correlated with downloads, while family apps are negatively correlated
- Our models have relatively low R^2 values, showing that there is **room for improvement**
 - New data could mitigate omitted-variable bias, irregularities in the data, and lack of precision in the outcome variable
- This study is valuable to **application developers** and **Google**



Thank You!
