Jeff Bertucci

12/12/19

Final Project Report

IST 652

**Unmasking the basement billionaire: analyzing the mobile games market with data science**

**Introduction**

The mobile games industry has become increasingly popular in the past decade for both consumers as well as designers. In a platform where anyone has the ability to join for only a minor buy-in, it is now easier than ever for creators to showcase their work and for it to gain viral popularity. On top of that, with revenue from ads and microtransactions (in game purchases), this model has become significantly profitable to those who can leverage their popularity wisely.

In this state of the market, a common trope has developed among consumers, creators, and critics: The basement billionaire. In short, there are 3 elements to the basement billionaire model:

1. They work alone or in a small condensed group
2. Their apps can be “low effort” or simplistic
3. They have a low amount of financial risk

Utilizing these three elements, the trope follows that these creators have a high potential to make a fortune on the low budget, simplistic app that they do not have to share in the profits of.

The promise of this trope has led to a flood of apps in the Appstore that are all making their own attempt to claw to the top of the charts in an attempt to make this dream a reality. The purpose of this research is to investigate this common myth to determine if there is some validity to the claims and if not, to determine a more reasonable and realistic view of a difficult market to pin down.

**About the Data**

The data used in this analysis was taken from the Kaggle data science page titled ‘17k Apple Apps Store Strategy Games’. As the title suggests, there are approximately 17 thousand rows of data. Each row is a unique app store game with 18 columns of data. These columns consist of both structured and unstructured data. The structured data that was collected for this data set are categories such as price, rating, number of ratings, and app size. The unstructured data contains features such as app names, descriptions of the apps, and types of in app purchases. Although not utilized in this research, the dataset also contains columns of semi-structured web link data for both the game on the Appstore and the display picture for the app store game.

Utilizing this data, four questions were formed:

1. What are the hot button terms that will cause apps to be sold/downloaded?
2. What games tend to do well in the app market?
3. What is the consumer impact of well and poorly reviewed games?
4. Does app size factor in to the number of reviews/overall rating for the app?

Each of these questions homes in on various aspects of the aforementioned “basement billionaire” trope. The first question calls to the care and effort used in naming conventions in order to properly game the system with simplistic eye-catching terms. The second question will help determine success of games and what kind of teams create said success. The third question is would ideally determine if there is some sort of inertia that comes with positive or negative reviews. The last question calls into question just how “low-effort” the most popular games are by metric of how large the game is.

**Process**

**Cleaning the Data**

The first step in cleaning the data in order to prepare it for analysis takes place in the excel spreadsheet where all but six of the columns are removed as they would not be helpful to include in the analysis. The remaining columns that are to be utilized are App Name, App description, app price, overall rating, number of reviews, and app size. In these categories, there are not any null values to contend with.

The data is then loaded into python by means of a list. The categories are properly named and all data is put into a pandas data frame in order to do additional processing that would not be as easily done in the list format. All structured, numerical data is then coerced into proper numerical formatting. The description and title columns are then separated and tokenized for additional processing. Standard English stop words are removed from both the titles and the descriptions. Both sets of tokens are processed separately for comparison purposes.

**Analyzing the Data**

**Text Data**

For the text data, word frequency charts were created to determine what sort of words were common throughout the strategy games analyzed. It is important to note that the descriptions field was extraordinarily messy and unclean. Many app developers utilize the game description in order to showcase new updates, events, and special deals that are only tangentially related to describing the games. In addition, some developers utilize non-standard text data such as emojis and foreign (non-english) characters for emphasis, formatting, or it is simply the developer’s natural language.

A term frequency chart is shown below, showing the top 30 words common throughout all 17 thousand games:

A close up of text on a white background

Description automatically generated

In the above chart, one can see simple common terms such as game, play, and players. But certain eye-catching words such as new, time, fun, different, and free are also present. While not conclusive it is important to note what game developers decide is important to mention in their descriptions of their games.

The next point of analysis is for the game title data. This is processed in much the same way as the app descriptions:

A close up of text on a white background

Description automatically generated

Since the game title and the game icon are going to generally be the first and only things that potential customers see in order to judge whether they should check out a game, it is important for titles to be descriptive and concise. It may go without saying, however they also need to sound fun and interesting. Words such as puzzle, war, defense, battle, match, tower, escape, and idle have all of these four qualities. They give a concise view of what the app is selling while also giving strong action words that would indicate the game being exciting and fun. (Note: the term idle is generally used in apps to describe games where the user does not need to put in as much effort as other games. While a slight exception, it still markets the appeal to consumers seeking that quality out). Additionally, there are also numbers in the top thirty words used in titles indicating that sequels have a significant weight in the market. One last insight is that game development terms such as 3D and HD also stand out as common, potentially to display a game to have a higher quality than its peers.

**Numerical Data**

To process the numerical data the numerical columns are organized and summarized in order to get a better picture of games that are successful in the app market. One note to make here is that the term “sales” corresponds to the number of reviews that a game gets. As number of downloads and game sale statistics are not included in the dataset, these numbers had to be estimated using the number of reviews as a base. In processing the data, the set was ordered/arranged by two categories: Firstly, the star rating and secondly the number of reviews:

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Rating | Sales | Size |
| Cash Inc. Fame & Fortune Game | 5 | 374772 | 245957632 |
| Egg Inc. | 5 | 174591 | 74891264 |
| AFK Arena | 5 | 156766 | 225711104 |
| South Park: Phone Destroyer\u2122 | 5 | 156044 | 130186240 |
| From Zero to Hero: Cityman | 5 | 146729 | 296638464 |
| Sushi Bar Idle | 5 | 123606 | 257325056 |
| Fire Emblem Heroes | 5 | 120283 | 175634432 |
| Bloons TD 5 | 5 | 97776 | 133326848 |
| Naval Warfare | 5 | 90214 | 43198464 |
| Idle Roller Coaster | 5 | 88855 | 234342400 |

These were determined to be the top 10 games by this metric. It is important to note that these games only pull a small number of words from the top 30 list of words in game titles. In fact, the only similar words are ‘Game’ and ‘idle’. Another notable observation of this list is that 4 of the games come from big publishers or franchises. AFK Arena is one of the largest mobile game franchises, South Park: Phone Destroyer is a South Park franchise game, Fire Emblem Heroes is derived from Nintendo and the massively popular Fire Emblem series, and Bloons TD 5 is from a long held series that goes back at least a decade.

The next feature to be examined was the idea of app sales, or potential revenue the app generated. This was an estimated calculation that was derived from the number of reviews an app had multiplied by the price of the app. This feature does not take into account any free to play games however, which limits the data range.

* + In the 4 to 5 star range: Bloons TD 5
    - Total dollars earned from sale of app: $292,350.24
  + In the 3 to 4 star range: Traffic Rush
    - Total dollars earned from sale of app: $210,635.37
  + In the 2 to 3 star range: Omar Sharif Bridge
    - Total dollars earned from sale of app: $2,150.41
  + In the 1 to 2 star range: Waste Land : savage survival of fear world games
    - Total dollars earned from sale of app: $95.68

The breakdown above shows the highest earning game within each of the star categories. All partial stars were round up to the next star category since partial stars were not conducive to this analysis. What these numbers show is a logarithmic decrease in revenue as star rating goes down. While a decrease was expected in this case, it was not clear by how much the star rating had an impact. With a five star game earning nearly 100 thousand dollars more than the next category which earned over 200 thousand dollars than the next one after that.

Lastly, the size of each app was taken into account. The average game size was determined to be 115 million bytes of data. Of the games previously listed as being the top ten rated games in the app store, eight of them were larger than the average game size. Additionally, five of those eight games had nearly twice the average size. This shows a greater amount of effort in some variety of game updates, features, levels, and graphical capabilities of these apps. While these games may imply that there is a low effort required to play them, the exact opposite may be true in order to develop them.

**Conclusions**

While many of the Appstore games use similar descriptive words, the best games tended to be the ones that set themselves apart from those common words and phrases. This could potentially be due to the impact of brand appeal as the top games also seemed to be either part of a franchise such as Fire Emblem or South Park or part of a long held series such as AFK Arena or Bloons. These types of games tend to have an already developed core fan base which allows them to get more sales earlier on as well as more people who are more likely to enjoy the games as they already know they like the characters. This may yield higher review scores which allows for the games to maintain a positive inertia moving forward and netting more and better reviews and resulting in higher profits. Games that do not have this kind of solid base have a substantially more difficult journey. Even middling reviews can be enough to sink a fledgling app. Lastly, the games reaching the top of the charts in general have a significant size advantage over the average game. This allows them to have more and better content to sell to their consumers.

The results of these metrics point to refute the common idea of a basement billionaire. It does not seem like individuals have the capacity to cobble a simple app together in order to make their fortune. Instead, games with high financial support behind them with a significant amount of production time seem to net the best results. Additionally, this shows the team effort involved in app design as it is simply not generally doable by the average person alone, there are simply too many moving pieces.

Future steps in this analysis would include better refining the text analysis by removing punctuation from the token list. Additional details are necessary to find what exactly larger size in games correlates to in each case, be it levels, graphics, features, etc to get a better idea of how best to utilize app space.