**1. Background Reading Introduction**

In this project I will try to answer some questions based on video game data collected from Steam (digital distribution platform) in 2016. Going forward, I decided to concentrate on those two questions:

“Are there any significant differences that can be measured between games that received a Metacritic score and those that didn’t?”

“Does the number of game screenshots have a measurable effect on the sales of the game?”

Question that I still want to address, but most likely not in depth:

“Is there an association between how many copies of the game were sold and the price of the game?”

And finally, one of my initial questions that I realized is basically impossible to assess:

“Does relative success of a game means that most likely it’s localized in most popular languages and are there any exceptions for this assumption?”

I won’t be trying to answer it because success of a game might be subjective thing to different people. And even if I wanted to evaluate game’s success from a financial stand point, since I don’t have data on how much was spent by each developer and how many people worked on it to develop a game and for time etc. - I can’t calculate the profit that developer made. In addition to that, the data required for this is most likely confidential and can’t be found for many games. Thus, I decided not to tackle this particular question.

**2. Topics of research:**

1. ***Metacritic score***

Since one of my questions deals with Metacritic score, I had to research how it works and basically dig any information that might be useful in my understanding of it, since it will aid in answering the question(s) later.

Official “About” and “FAQ” pages on Metacritic website turned out to be very useful in my research and cleared out most uncertainties I had about it [1]. Here are my findings:

First, Metacritic website covers virtually all new game releases in United States and other English-speaking territories, as long as they are reviewed by multiple publications (at least four).

There are two main scores on the website: Metacritic score (based on weighted calculation of scores from different publications/critics) and a User score (based on user reviews). User scores are not used in Metacritic score calculation and in contrast to publication/critic reviews, user reviews may be changed down the line (Metacritic only accepts first review from a publication to avoid situation when publication’s score changes down the line by means of potential external pressure on the publication). Also, some of publication/critic reviews might not provide a discrete value for their score, so it’s up to Metacritic staff to assign a numeric value to such reviews. Usually they work in increments of 10 (0-100), but sometimes might fall somewhere in between (for instance: 75).

Speaking of Metacritic score, weightings for it are kept in secret (User score is not weighted). And lastly, low score for Metacritic score doesn’t necessarily mean that game is that bad in terms of experience (but it certainly might); it means that most of publication/critic reviews were generally negative.

1. ***Finding associations/correlations***

Most sources I found didn’t state any clear questions that they wanted to be answered and were more like articles than scientific papers. Only one paper [2] used regression and some other Data Mining techniques to find correlation between two variables. Other ones used more generic analysis techniques, such as comparing averages of different categories and not much else [3] [4]. One of the articles I stumbled upon [5], while didn’t really do much in terms of analysis of video games data, provided and interesting outlook on why someone might be interested in doing their own research on video games if they are planning to get into game development or are already one.

**3. Background Reading Conclusion**

The most useful information was found for Metacritic score, which will certainly help me with answering my question. In terms of research in this area, there is quite a bit of research more based on implications of gaming in general and more sophisticated studies on specific aspects of the games. Couldn’t find much in terms of good studies on how different parameters of the video games correlate/associate with each other. Most of articles that I found did a more rudimentary analysis – comparing averages of certain categories or just analyzing distributions of certain parameters. One of the articles used regression for finding correlation between variables, it might be a good fit for some of my questions.

**4. Methodology**

***Data Collection:***

For this project I’m using a preexisting dataset [6]. It consists of all listings (mostly games, but also some non-game software) from Steam (digital, game-centered distribution platform) dating December 12 of 2016. Most of the data was collected from Steam via Steam API and some extra data (for instance owner count and players estimate) was acquired from Steam Spy (service that collects the kind of data mentioned before from game on Steam).

***Data Analysis:***

“Are there any qualitative differences that can be measured between games that received a Metacritic score and those that didn’t?” For this question I’ll probably compare the averages for different features (those that are numeric) and will look at the distributions of categorical variables.

“Does length of game description or number of game screenshots have a measurable effect on the sales of the game?” For this one, most likely I’ll use some sort of regression analysis, do determine if there is any correlation.

“Is there an association between how many copies of the game were sold and the price of the game?”, for this one regression seems like the most appropriate choice from what we learned so far in DwD1.

Because I have all the games from Steam, and Steam was and currently is number one game distribution service on PC, I will be able to generalize my conclusions to all games that were released for PC at that time (2016).

***Data Visualization:***

All the relevant visualizations will be done in R.

**Works Cited**

[1] About Metascore, <https://www.metacritic.com/about-metascores>

FAQ, <https://www.metacritic.com/faq#item18>

[2] “Using Steam data to tell if your game will sink or swim”, <https://venturebeat.com/2017/06/28/using-steam-data-to-tell-you-if-your-game-will-sink-or-swim/>

[3] “Steam – What’s your Game?”,<https://nycdatascience.com/blog/student-works/web-scraping/steam-whats-game/>

[4] “What’s in the Name? Data analysis of 5,820 Steam Games”, <https://gamedevelopment.tutsplus.com/articles/whats-in-a-name-data-analysis-of-5820-steam-games--cms-30101>

[5] “Understanding your game through data”, <https://galyonk.in/understanding-your-game-through-data-8b09ca93ec11>

[6] Steam data, <https://github.com/CraigKelly/steam-data>