

# Capstone Project 1 Proposal

**By Pan Chen**

## **1. What is the problem you want to solve?**

On a high level, I want to see if critics' review scores and users' review score would affect a game's sales. More specifically, I want to explore things like:

- a. Did critic scores and user scores have the same impacts on a game's sales?
  - b. Are the review scores more important to sales now as compared to it was before, due to the prevalence of the internet/smartphones/social media?
  - c. Did critic scores have the same impacts on the sales of multi-platform games vs. platform-exclusive games?
  - d. Did critic scores from Western media have the same impacts on the sales of games in all areas in the world?
  - e. Did certain genre sell particular well, despite of critic scores? What are their characteristics?
- ## **2. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?**

My target audiences include game developers and their PR team and marketing team. They could decide whether or not to spend more effort to increase a game's quality and/or making better relations with reviewing websites and player community in order to get better review scores.

## **3. What data are you going to use for this? How will you acquire this data?**

I would use a dataset downloaded from kaggle by Rush Kirubi that included game sales and their review scores scraped from Metacritic.

(<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings/>).

**4. In brief, outline your approach to solving this problem (knowing that this might change later).**

- a. Inspect the data and decide which rows to include (e.g. games from 80s? 90s?) and how to tackle the NAs in the data.
- b. Wrangle the data, create new columns needed for the further analysis (e.g. Multiplatform game or exclusive games, game franchise).
- c. Create exploratory visualizations to see if there are any obvious trends.
- d. Propose some hypotheses according to these trends, and test them with the techniques I am going to learn later.
- e. Report the findings.

**5. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.**

I will submit a jupyter notebook file that documents my code, visualizations and findings, and a slide deck that focuses on my approaches and findings.