511 Project Abstract
Yifen Chen, Hanming Li, Changwen Li

1. Introduction and Background

The global pandemic of COVID-19 has undoubtedly led to many trends in the global gaming market and driven a spike in global player engagement. These increases in engagement will also be reflected in the soaring revenue in the gaming market. Even after the global pandemic subsides, many expect that most of the additional increase in player engagement and revenue due to this pandemic will remain the same. During the ongoing global quarantine, gaming has become more frequent in people's lives, and some of the investments (gaming setup) for gaming entertainment are not easily shelved. However, the great growth brought by the specificity of 2020 to the gaming market cannot be replicated. Among the different platforms, PC and console have a higher barrier to entry, but therefore the customer stickiness is also stronger. According to Newzoo's "Global Gaming Market" report, 2.8 billion gamers worldwide will bring $189.3 billion in revenue to the market by 2021. With the continued growth of infrastructure settings and economies in Southeast Asia, the Middle East and North Africa, these emerging markets will drive revenue growth in the global market.

With the growing trend of the game market, we as gamers want to focus on what factors can affect the overall sales of a game. The theme of our project will focus on whether the players' evaluation of each game affects the overall sales of the game. Nowadays there are many different video game digital distribution platforms such as Steam, Google Play Store, Epic Games, etc., and these platforms allow people to rate and review a game after purchasing it. The project will focus on whether positive and negative game reviews can influence other people's decision to buy a game, which reflects on the sales number of the game.

2. Dataset

A. Steam dataset

The Steam platform is one of the world's largest comprehensive digital distribution platforms for video games. It was launched in September 2003 as a standalone software client as a way for Valve to provide automatic updates for its games, and expanded to include games from third-party publishers. Players can buy, download, discuss, upload and share games and software on the platform. Popular PC games,

whether they are big titles like the Half-Life series or the Call of Duty series, or casual mini-games like Plants vs. Zombies can be purchased on Steam. The Steam platform is now the largest digital distribution platform for PC games, with about 75% of the market share in 2013. By 2017, users who purchased games through Steam totaled approximately $4.3 billion, accounting for at least 18 percent of global PC game sales. By 2019, the service had more than 34,000 games and more than 95 million monthly active users.

The huge number of users makes steam our first choice of data source. Steam is a platform encouraging intersection between developers and consumers by forum discussion per game and gamer rating. The rating system created by Steam is more effective than other gaming rating site and platforms because the user can refund the games if they do not like, users actually buy the game than review bombing the things they don't like without actually buying them.

B. Google Play Store

Google as the world's largest search engine owner. Their own game sales platform has been bagging a large part of the market share. Many casual gamers go to the search engine first to search for the games they want. Therefore, we have also obtained the evaluation data and sales of Google Play Store.

The data we obtain includes the name of the game, the number of sales and the total number of reviews. Among them, the positive and negative ratings of the games are based on the star rating given by the players. We base positive ratings on five to four stars; three stars are average; two and one stars are negative.

3. Statistical Methods for Analysis

a. Linear regression
   A linear regression model will be implemented over the number of owners vs. positive review ratio

b. T-test
   An t-test will be conducted to see if the mean sale of games having positive rating more than 0.5 is different from those lower than 0.5

c. Chi-square Test
   A chi-square test will be conducted to see if sale is independent of rating

d. Permutation Test
A permutation test will be conducted to see if sale is independent of price

4. Expected Result

We expect that the review would play a huge role in determining the total sales besides price, gaming genre, brand recognition after the first week of sales because sales in the first week is predominately hype-induced due to media exposure, and the first week of sales usually dictates less than 50% of the total sales.