Analysis of the Effects COVID-19 on Steam's PC Gaming Population

Riley Barklow

Department of Computer Science
University of Tennessee
Knoxville, TN, USA
rbarklow@vols.utk.edu

Abstract

Gaming was one of the industries that saw a major increase in users during the onset of the COVID-19 global lockdowns across the world. Steam, an online game distributor and pseudo social network for video games, is also the largest platform for PC games in the world. In this paper, I use clustering to determine whether there were quantifiable trends in the types of games people chose to play during the COVID-19 lockdowns of 2020, and how much COVID-19 influenced people's choice of games compared to pre COVID times.

I. Introduction

During the onset of the global pandemic in 2020, global lockdowns occurred in countries across the planet. During this time, people sought out virtual entertainment in higher numbers than ever before. Steam, the most popular platform for the distribution of PC games in the world, continuously broke player number records throughout the year 2020 [1]. While it is known that COVID-19 was beneficial for the gaming industry as a whole, there is little information about what specific games people played, and whether or not COVID-19 caused noticeable trends in the popularity of certain game genres.

In this paper, I show an analysis of how the COVID-19 lockdown influenced the types of games people chose to play on Steam, and to what extent these choices were influenced by the lockdowns. To do this, I determine the average year over year growth of the population of 1000 games on Steam through the years of 2016 through 2020. I then use K-Means clustering to group the games into distinct groups, and determine what kind of effect COVID-19 had on the population of the cluster. Finally, I analyze the genre tags most common for each cluster to determine what effect the COVID-19 lockdowns had on certain types of games.

II. Dataset

The dataset used in this analysis is a publicly available dataset on the populations of the top 1000 Steam from 2016-2020 [2]. For each of the 1000 games, Steam's API was queried to get the current player count for the game every 15 minutes for around 4 years. Each entry in the data contains the timestamp, accurate to the minute, and the current number of players playing that game. Each game is represented by its own separate CSV file.

III. Methodology

In this section, I detail the process and workflow used to transform the dataset's data into an analyzable format.

Data Cleaning and Preprocessing: The first step of cleaning the data involves dropping all entries that had a consistent population of zero. Steam's API does not differentiate between games and other applications on the platform, and some non-games were included in the dataset.

Data Processing. The next step involves turning the dataset's format into one more suited for data analysis. Because each game's player count was recorded every 15 minutes for 4 years, and there were 1000 games, a data processing system that could handle 10s of millions of lines of data was required. To process the data in a reasonable amount of time, I utilized Apache Spark's implementation of the MapReduce paradigm to transform the data [3]. Apache Spark's MapReduce implementation runs in parallel, and greatly reduces the amount of time needed to process the data.

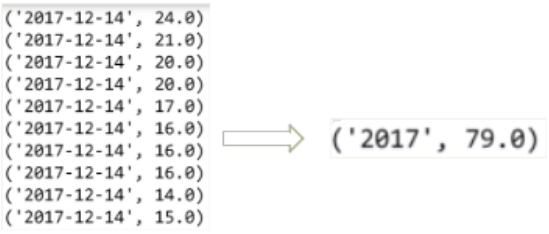

Figure 1: MapReduce key reduction example

The first map operation involves stripping each line of data of its timestamp, leaving only the date. A reduce operation follows this, which reduces each entry down to the maximum player count recorded for each day. This process is then repeated, but the map operation strips all of the date out except for the year. By the end, I have the daily average player count peak of the game for each year of recorded data. With this, I calculate the average yearly growth for two distinct time periods; the years of 2017-2019, and from 2019-2020. This shows a clear numeric value for each game's player growth before and during the transition into COVID-19 and its lockdowns.

Clustering: Before clustering, I utilize the elbow method to determine the most effective number of clusters for the analysis. I then use KMeans clustering with this K value to cluster the games on two variables,

the average yearly population growth before and during COVID-19.

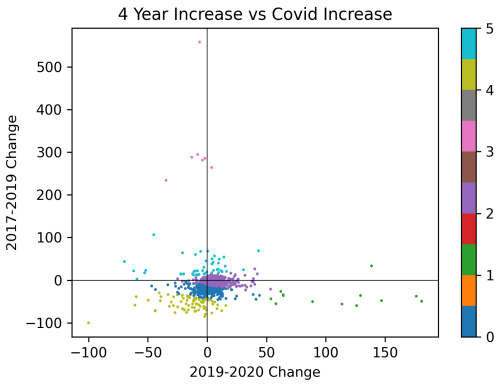

Figure 2: Visual representation of the clusters

Additional Processing: With the games all having their own identifiable cluster, I perform additional processing on the available data. Using a comprehensive list of the genre tags associated with each game, I determine what the genre percentage makeup of each distinct cluster is. By the end of this step, I have a detailed breakdown of the types of games most associated with each cluster, as well as how the average population change of each cluster was affected by COVID-19. Finally, I use a graphing system that can take the data for any specified combination of tags and graph the genre makeup for each cluster.

IV. Results

The three major genre groupings I analyze are whether the game was singleplayer or multiplayer, whether the game was classified as difficult or casual, and what the defining genre of the game was.

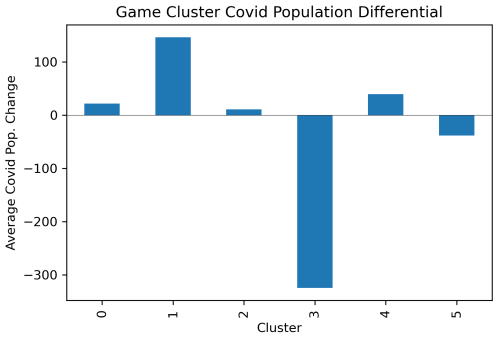

Figure 3: COVID-19 population differential by cluster

The resulting genre distributions show how these two sets of binary genre tags have a population growth that correlates with the onset of COVID-19 and its lockdowns. In the clusters that experienced positive population growth during COVID-19 compared to previous years, multiplayer games were more common than single player games, and casual games were more common than difficult games. There was no clear correlation between whether or not the game's defining stylistic genre had an effect on the games COVID-19 population differential.

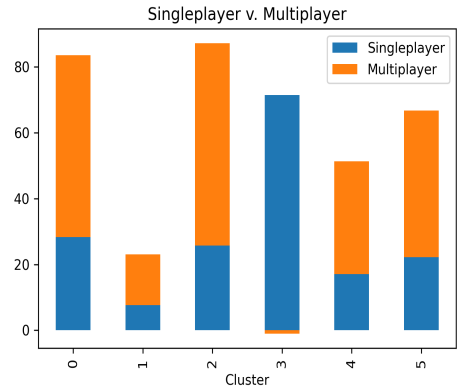

Figure 4: Single v. multiplayer distribution

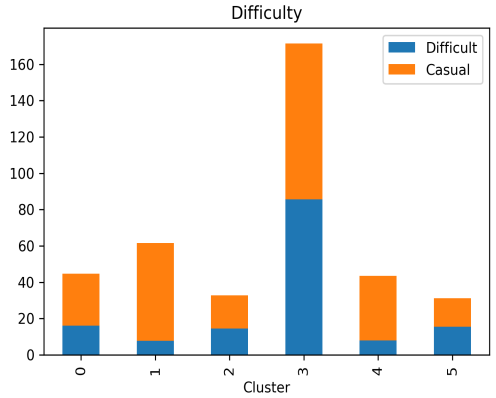

Figure 5: Difficult v. casual distribution

V. Conclusion and Future Work

From the analysis performed, I conclude that during the onset of the global lockdowns during the COVID-19 pandemic, PC gamers were more drawn to multiplayer games than single player games, and were more drawn to more casual gaming experiences than they were to difficult ones.

Future work on this topic could include a more robust and comprehensive tagging system to classify the games' many genres, since steam allows seemingly conflicting tags to exist on a single game.

VI. References

[1] N. Smith. " The giants of the video game industry have thrived in the pandemic. Can the success continue?" *The Washington Post*. May 12, 2020. [Online]. Available: https://www.washingtonpost.com/video-games/2020/05/12/video-game-industry-coronavirus/. [Accessed May 12, 2022]

[2] Wannigamage, D., Barlow M., Lakshika E., Kasmarik K., *Steam Games Dataset : Player count history, Price history and data about games*, Mendeley. [Dataset]. Available: https://doi.org/10.17632/ycy3sy3vj2.1. [Accessed: May 12, 2022]

[3] IBM Cloud Education, "Apache Spark" *IBM CLoud Learn Hub*. June 2, 2020. [Online] Available: https://www.ibm.com/cloud/learn/apache-spark