Steam Sales Data

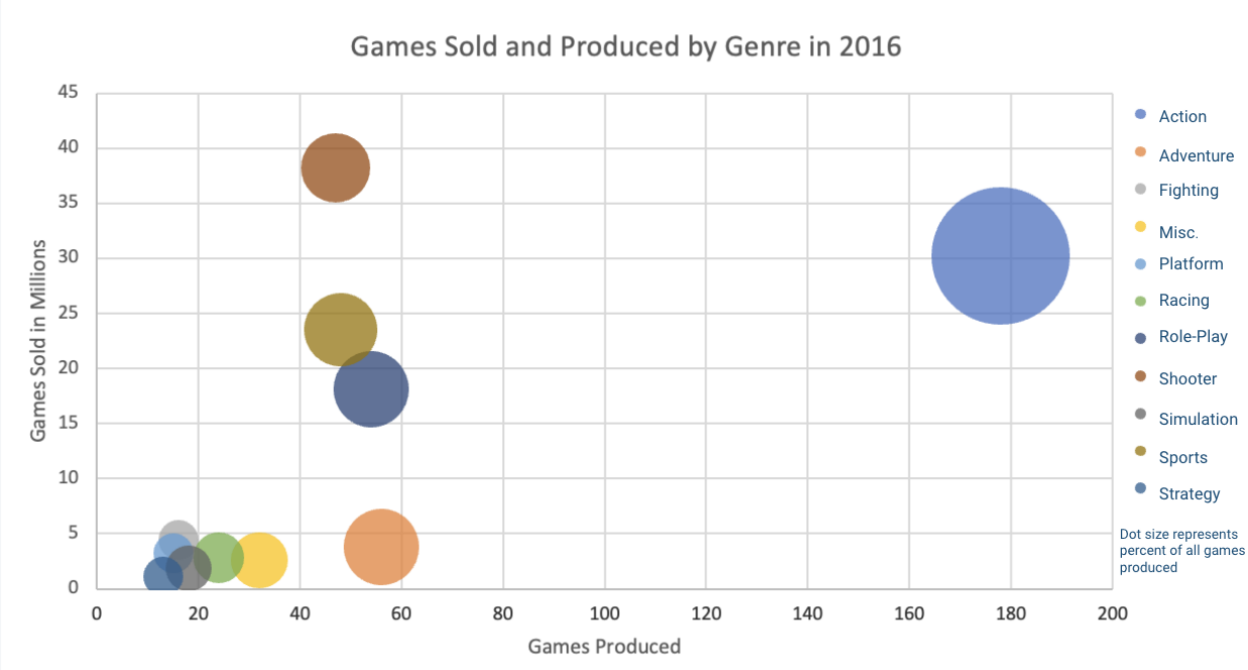


Figure 1: Games Sold & Produced by Genre in 2016. Visualization by Will Cromer.

Team Fighting Badgers

Name	**Purdue Email address**
Lev Working	lworking@purdue.edu
Will Cromer	wcromer@purdue.edu
Giovanni Tuzzio	gtuzzio@purdue.edu
Rudy Manian	rmanian@purdue.edu

Table of Contents

Steam Sales Analysis

Introduction

Game developers often overlook the crucial step of analyzing the market before picking the genre of game they choose to develop. This often leads to developers entering oversaturated markets; their titles, consequently, flounder. This project seeks to provide a breakdown of revenue statistics from the popular platform Steam, and determine what genres are oversaturated and which are underutilized by examining the revenue associated with each tag. The results should be enlightening for game companies and independent developers alike, with the results most pertinent to smaller studios.

Background

The team chose two datasets to analyze. One was a collection of information gathered by games industry veteran Daniel Weinbaum, which provided a valuable amount of raw data and some intermediate analysis about game sales on steam from 2016 to 2021. The methodology of estimating revenue this dataset used was more sophisticated, detailed, and compelling than others available online. The other data set looked at video games from 1980 up to 2016 looking at their sales, genres, game ratings, and what platforms they are available on. Research for the second dataset was conducted by Minh Hoang Dang.

No additional data was acquired, although a lack of data about in-game purchases, cosmetics, and DLCs limits the scope of our claim.

Questions

Our target audience is game developers who want to better understand what types of games are the most profitable and what games are produced at the highest rates. Further questions include:

What genres of video games are produced at the highest rate?

What genres create the most sales?

How much money is being brought in by different genres?

When is it best to release games for the most revenue?

What games are most enjoyed by the users?

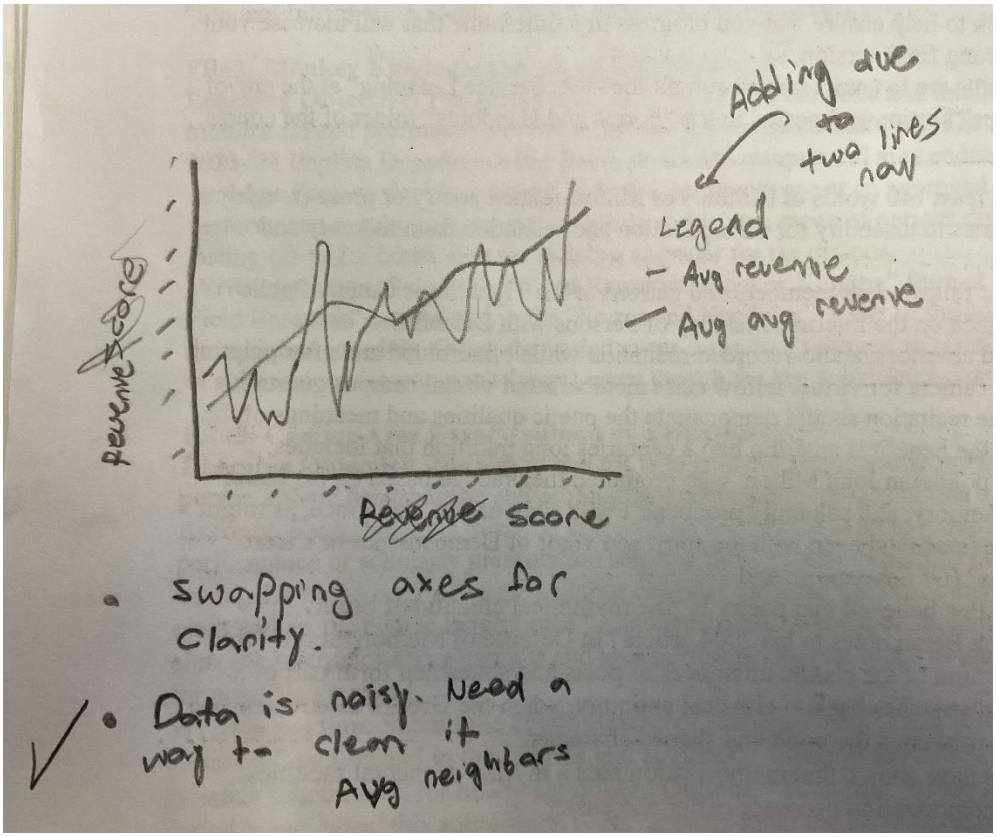
Problem Statement

Many dashboards exist to provide insights into steam sales, with some sites such as VG Insights (https://vginsights.com/steam-analytics) looking at prices, ratings, and revenue distributions. Other sites such as https://comparecamp.com/steam-statistics/ compile all sorts of statistics from revenue to show player count on some of the larger games and other odds and ends. Other sites show the breakdown of revenue by geography on the Steam platform, such as Game World Observer.

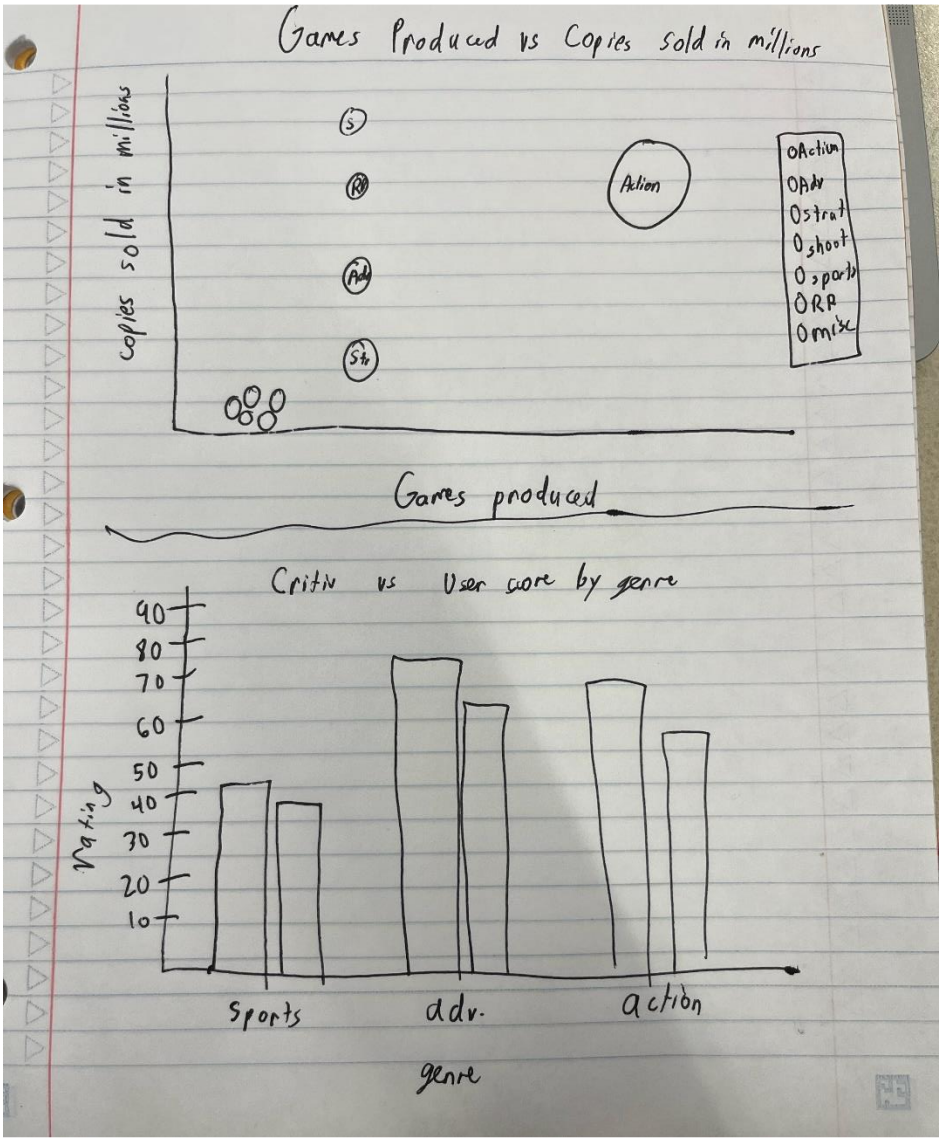
Methodology

Lev Working: For the visualization with the best and worst performing tags, I removed tags that were not pertinent to game development such as Software, Documentary, or Tool. I also removed tags that had less than 50 games of that tag on steam. This was done with the assumption that such tags are more likely to be distorted by one or two successful games that had that tag, and was thus not indicative of consistent

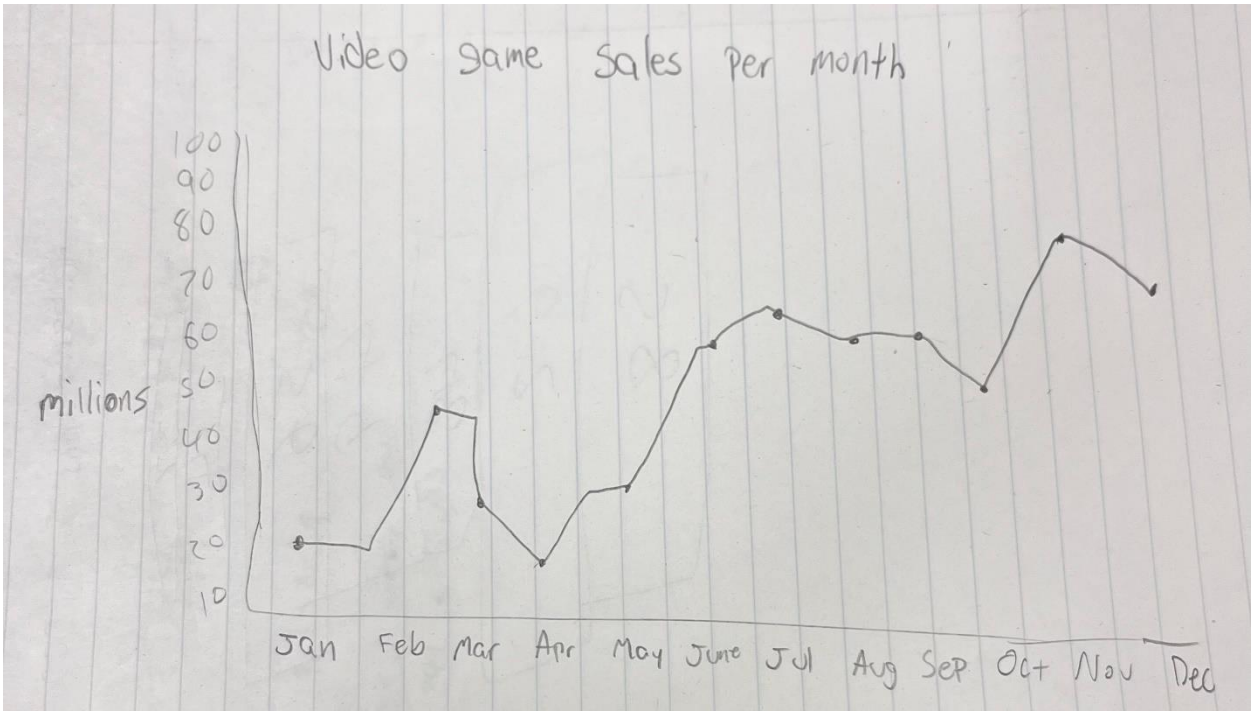
success. For the other visualization, I sorted the data by User Review score, created a chart showing the average revenue for all of those scores, and then averaged those values with their neighbors to make seeing the trend easier to see amongst the noise.



Will Cromer: For my question, I was mostly looking at the games in their respective genres as a whole. The first and most important step for me was to filter the data into different genres. From there, I used the same filters for each: organizing them in order of copies sold, user ratings, critic ratings, etc. to get a big picture and represent those values by each genre.



Giovanni Tuzzio: For my visualization showing global video game sales by month in 2016, I decided to only use games from 2016 in my dataset because it was the most recent year that had the most accurate data with minimal duplicates or missing values. I sorted all the data into tables so it would only show data for games in 2016, then I sorted them all by their release date and added all the sales for each game by each month starting from January. Once I had all the sales for each month I created my visualization. This was done under the assumption that the data given was accurate for the games in the data set and that it included the most relevant games that were released during 2016.



Rudy: The question that I sought to answer with my analysis and visualization of the data was what is the price and volume of video game sales by genre. Analyzing these two factors would yield key insights about expected revenue, sales brackets, and what developers could expect as a return for their efforts. In particular, I wanted to see how much independent, or ‘indie’ developers could earn for their efforts. Although I have not worked in the games industry, helping develop art and assets for games is something that I am working towards and knowing what I can expect for that work if I was working outside an established studio as part of an independent team would be extremely helpful.

Results

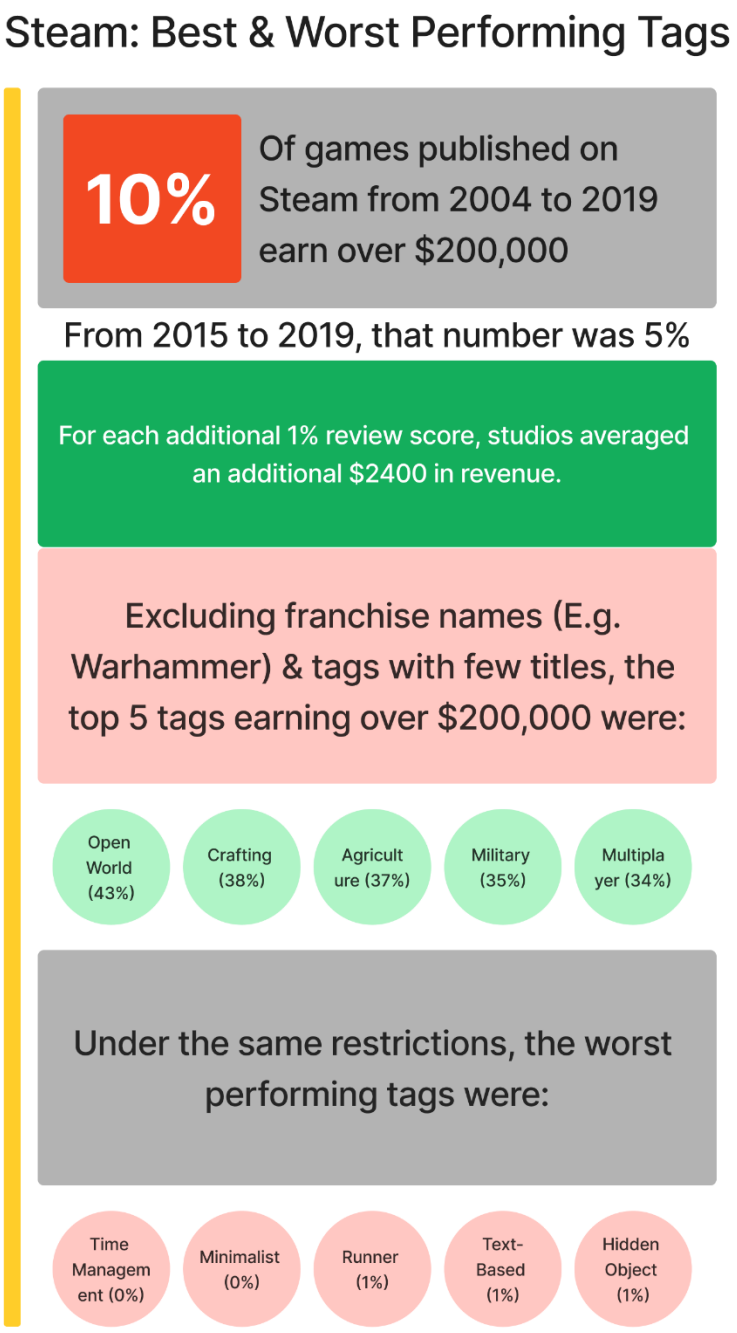


Figure 2: The best and worst performing tags on Steam. Visualization by Lev Working.

Discussion and Conclusion

Exploring the two datasets, we came to a number of conclusions:

- Despite some genres having more games produced, they do not always bring in the most sales.
- Games that received a higher user review score tended to earn more money.
- May, November, and December are the best months to publish a game in terms of sales.
- Even ‘indie’ games can make similar revenue to industry standards.
- The best performing tags on Steam are Open World, Crafting, Agriculture, Military, and Multiplayer. The worst are Time Management, Minimalist, Runner, Text-Based, and Hidden Object.

References

https://www.gamedeveloper.com/business/genre-viability-on-steam-and-other-trends---an-analysis-using-review-count

https://data.world/mhoangvslev/steam-games-dataset/workspace/intro

Appendix A – Resources Used

Datasets

Steam Sales Data – Research by Daniel Weinbaum. Contains estimated revenue on Steam by user reviews, year, tag, and revenue category (e.g. > $50k).

Steam Games Data - Research by Minh Hoang Dang. Contains sales, critic and user ratings, year of release, game, publisher, and the games age rating.

Tools used

Tool/Application	**Description**
Excel	Data cleaning & Visualization.
Figma	Data visualization
Photoshop	Data visualization.
HTML	Web development
Hitfilm Express	Video Editing
Google Slides	Presentation design.
Discord	Communication
Zoom	Presentation Recording

Appendix B – Percent Contribution

Group Contributions

- Contributed to the data visualization process.
- Brainstormed topic ideas.
- Served as rotating team leader.
- Contributed content to the summary.
- Discussed visualizations in the 5-minute video.
- Explained content in the presentation.
- Answered audience questions.

Individual Contributions

In the table below list each team member's full name, their contribution (body of work) and their % of the work completed. The total must add up to 100%

Team Member	Description	Contribution
Lev Working	*Edited all project videos.* *Filled out Team Paper.* *Analyzed & created Best & Worst Tags and User Review Score vs. Revenue visualizations.*	*30%*
Rudy Manian.	*Recorded 5-minute video.* *Designed and added content to website.* *Analyzed & created Revenue segmentation by Genre visualizations.*	*30%*
Will Cromer	*Analyzed and created visualization for Games Sold & Produced by Genre, as well as Critic vs. User Ratings by game.*	*25%*
Giovanni Tuzzio	*Created Sales per Month visualization.*	*15%*
	Total	100%

Appendix C – Individual Contributions

In this appendix each team member must contribute a one-page document relating the team's topic/data to their home town or home country. The one-page document must contain: (1) a description of the problem, (2) a comparison of the team's findings with insights about your home town/country related to the teams group project data (3) a visualization to support items (1) and (2).

Each person should create their individual page **(1-page only)** and make it available to the designated team member who will upload the final document.

This will be viewed and assessed as part of each person's individual contribution.

Leave this page as is.

Start adding individual page content on the next page.

REMOVE any blank pages before submitting.

Team Member #1: Lev Working

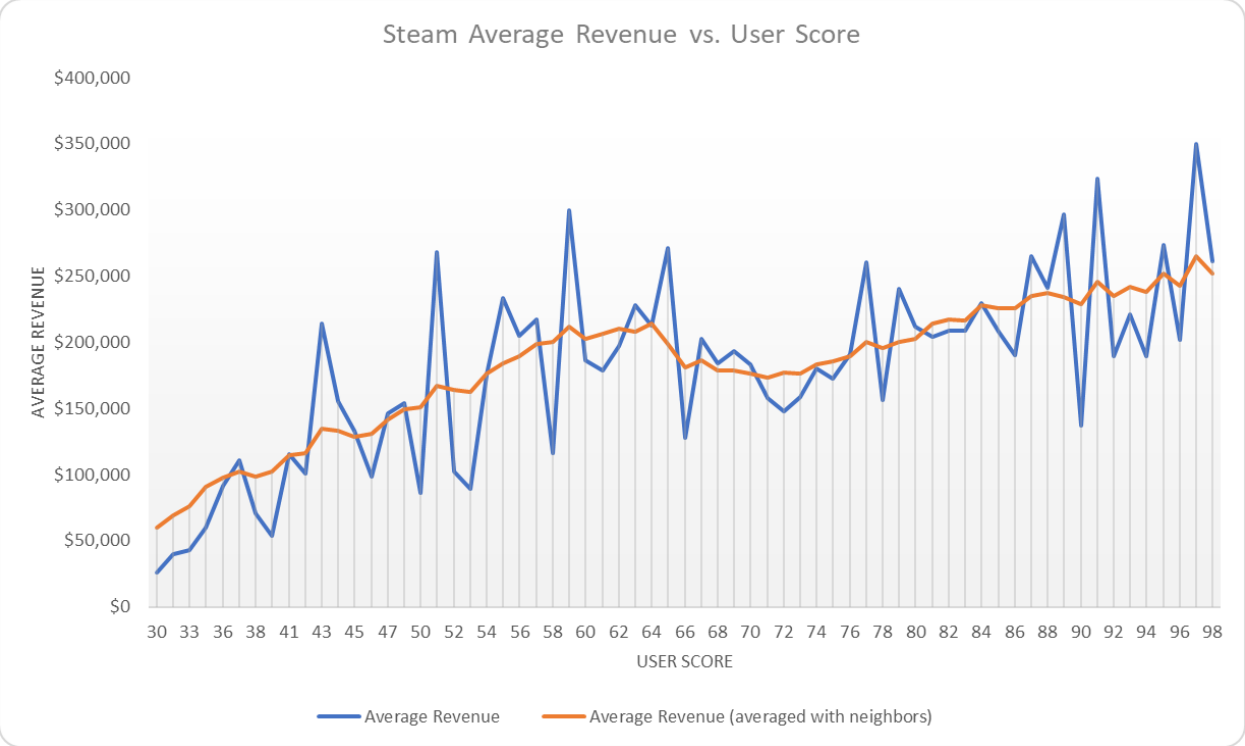
My Hometown/City/Country: Chicago, Illinois Group Topic (dataset): Steam Sales Data

I've been interested in video games for the majority of my life, and I've been interested in developing games since Freshman year of high school. Thus, the results of this project were especially of interest to me because there are good odds that I'll try to release my own games on Steam at some point in time.

The two problems I was addressing in this project were the relationship between User Review Scores and sales, as well as which specific tags were the best and worst performing on Steam. User Review Scores are an integral part of our formula for estimating sales, given that Steam keeps the actual numbers under lock and key. Because of this, I got curious whether the User Score was correlated with sales on its own. Many AAA developers embrace consumer unfriendly practices that lower the score of their game, and I wanted to see if this had the downside one might intuitively assume it had.

Regarding the best and worst performing tags on Steam, it's important to know what features of a game are correlated with success and failure, as this was central to our project.

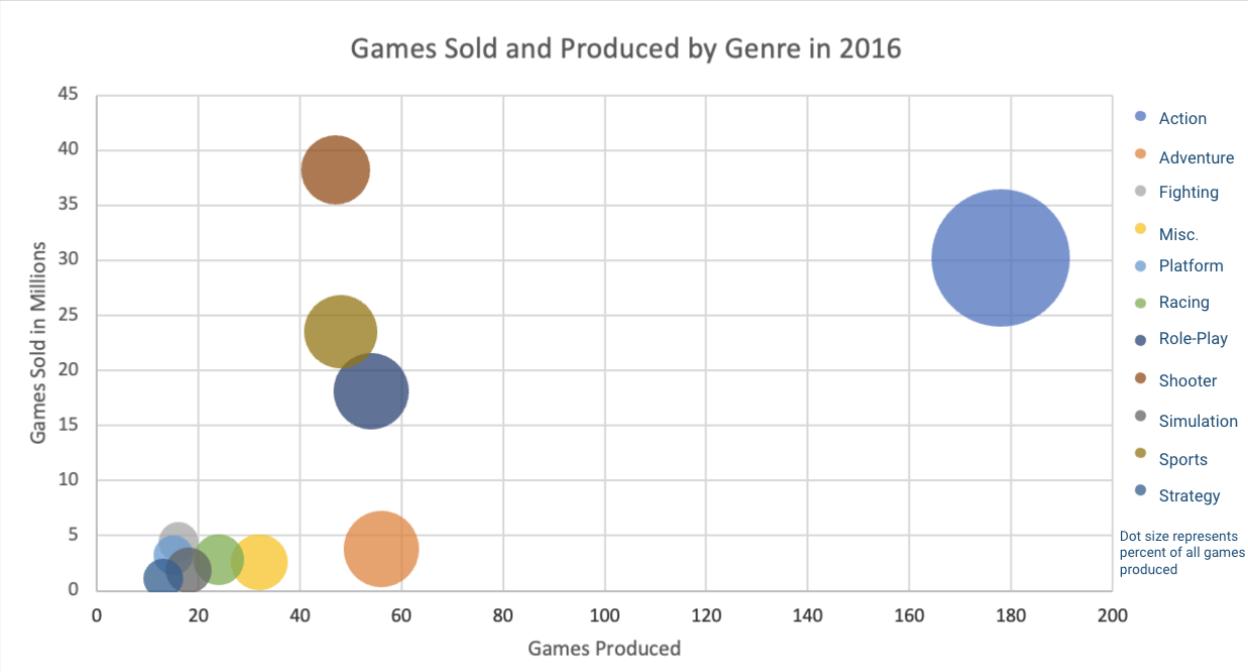
Our project wasn't geographically, oriented, but I took some time to try applying our findings to studios based in Illinois. My home town of Chicago has a couple dozen game studios as either branch offices or full headquarters for some of the smaller ones. Some of the more noteworthy studios include Gameloft, responsible for the Dungeon Hunter series, King (developers of Candy Crush), and Netherrealm Studios (developers of the Mortal Kombat and Injustice titles). These huge and highly successful studios naturally use many of the most profitable tags of Crafting, Open World, Multiplayer, except for King, which does not sell games on Steam. This opens up the intriguing question of whether other platforms, such as IOS or console, have a different set of genres and tags that are more profitable.



Team Member #2: Will Cromer

My Hometown: Indianapolis, IN Group Topic: Steam Sales Data

The problem that I wanted to address was to see what genres were most prominent in the gaming industry. Growing up video games was always something I did in my free time with my siblings, dad, and friends. Even though we all played video games, the genre of games we all played varied by who I was playing with at a certain time. My brother and dad loved to play sports games such as NHL, MLB, or FIFA. When I was with friends though we loved to play shooter games such as Call of Duty or action games like Rocket League. Growing up playing so many different genres of games made this project important to me as I got to see which of these genres really were the most popular. Coming into the project I thought the answer would've been shooters or sports because that is what I knew growing up, but it was fascinating to learn about action games being such a highly produced genre despite it being a genre I did not play as much growing up.



Team Member #3: Giovanni Tuzzio

My Hometown/City/Country: Group Topic (dataset): Steam Sales Data

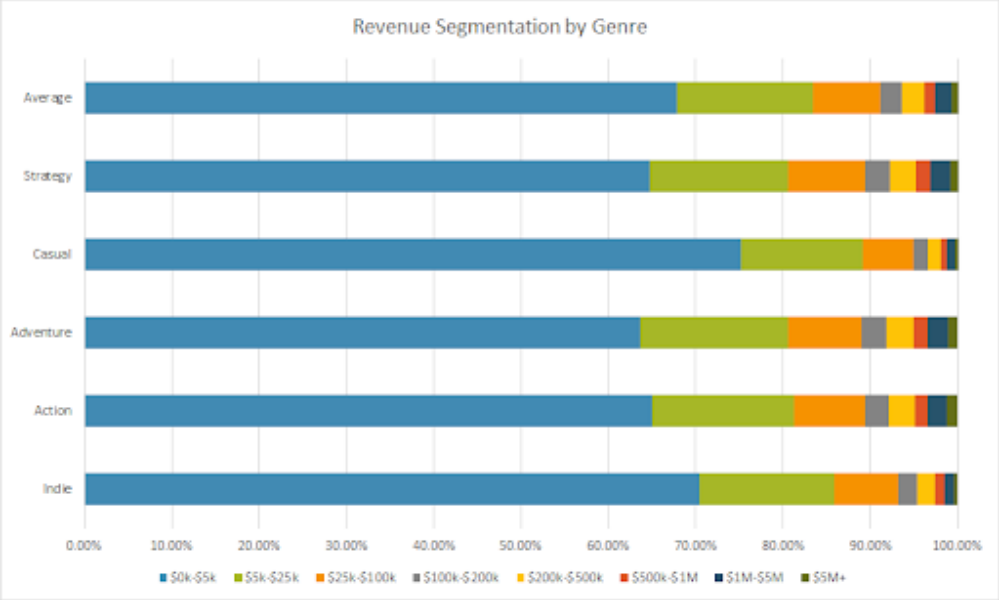
For this project, we were trying to help game developers by giving them information that lets them know what genres of games produce the most sales, what games are most enjoyed by users, and what time of the year is best to release a game. As someone who is currently majoring in game design, I found this project to contain a lot of useful data that could help game developers create better games for the players. For as long as I can remember, I would spend my free time playing video games at home. I started playing games like Super Mario Sunshine and Mario Kart and from that point I had already developed a passion for video games. Through this project I was able to learn that action games are the most produced genre within the game industry and since I grew up playing a lot of action games, I can see how this is the most produced genre of video games.

One problem that a lot of game developers face before publishing a game is deciding when is the best time to release their game. I wanted to help solve this problem by creating my own visualization that shows what months throughout the year see the most video game sales to help developers decide which month they should publish their game. Through my visualization I was able to find that May and November see the most video game sales throughout the year. The reason for this is because May is right before summer break for a lot of students in the U.S. and November is during black friday and cyber monday and its right before the winter holidays. As a game developer, this data can be useful because it shows you how many sales the video game industry sees on each month throughout the year. If I was a developer deciding when to release my game, I would publish it in May and then put it on a discount in November during black friday weekend while the game industry is seeing its most sales during the year.

Team Member #4: Rudy Mahan

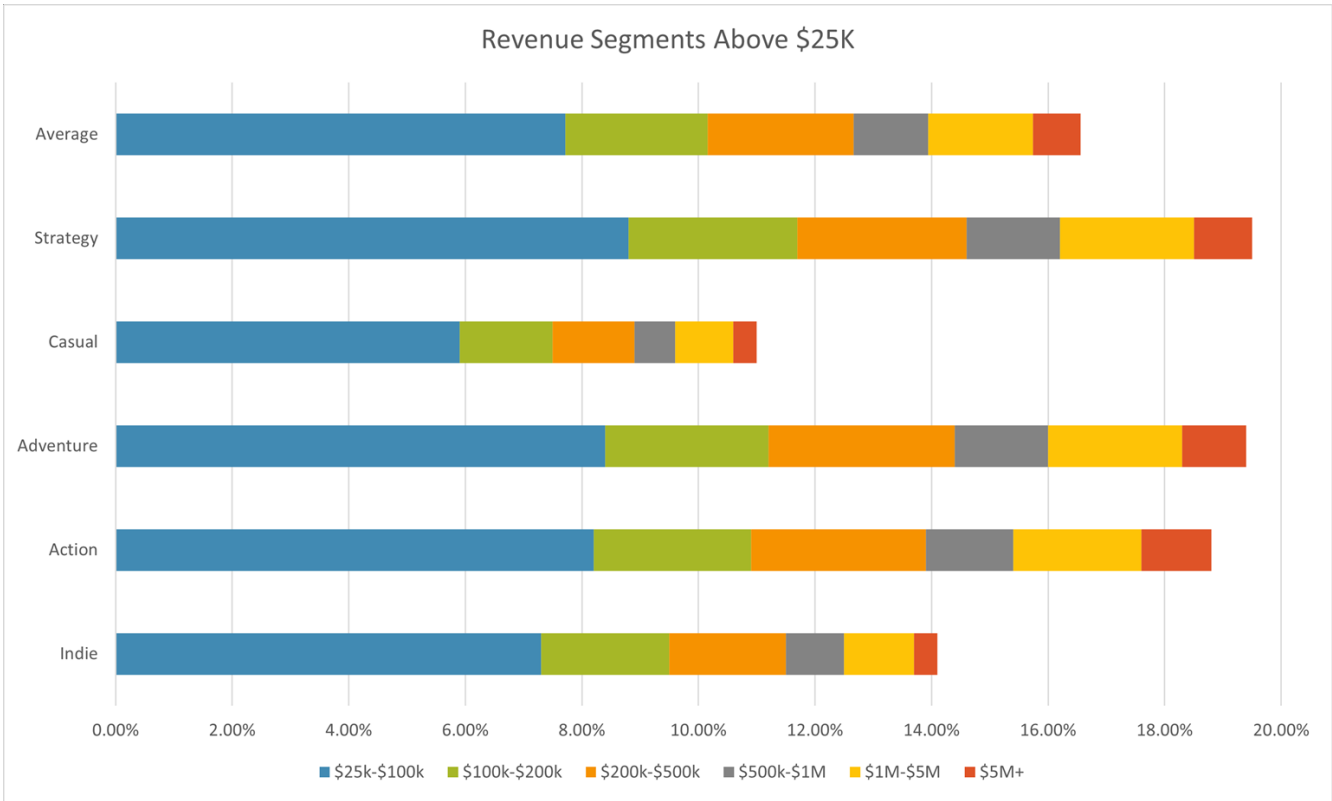
My Hometown/City/Country: Group Topic (dataset): Steam Sales Data

To this end, I used a few publicly available datasets and the platform 'steam,' an online retailer of various video games. While there was a quite a lot of data available about various aspects of the games on steam, such as number of games by descriptive tag, I found little that would help my research. One key set of data that I was able to obtain freely was for the price history of a game, which is useful as many games listed on steam change price a few years after their release and if I used this adjusted price in my calculations, I would not be able to draw accurate or useful conclusions. Although data was not directly available for what I needed, such as copies sold or revenue generated, Steam does list how many reviews a game has received. From there, I did some research and found a few articles that described the trend between copies sold and reviews of a game. Using the methodology of these articles, I scraped the review data from Steam and applied the equation to get an estimate of copies sold. Multiplying this by the initial price of the game, I was able to complete the data that I needed. Unfortunately, there was absolutely no geographic data that I was able to find or incorporate into my analysis. Thus, insights about my hometown or any other area based analysis could not be done.



The chart above is my broad visualization of my data. The colors represent brackets of income, and their size being proportional to the number of games of a particular genre that fall into that bracket. blue being the most prominent bracket at $0-$5k. Interestingly, although I expected that indie games would make far less revenue than other genres, I was surprised to see that the majority of games made saw similar revenues, with only casual games having significantly less revenue. Aside from games making over $100k, the analysis indicates that even indie games can expect similar revenue to other genres.

Knowing this, I was now curious as to the breakdown of games at the high end of revenue brackets. So I filtered out games that made less than $25k from my data. The chart below displays the result.



Here, the disparity in genres grows much more apparent. Casual games continue to generate far less revenue than standard and indie games too show a major dip. This is more along the lines of what I had expected, coming into this assignment. This was not too disheartening, however, as even though indie games generating high revenue were much less common than other genres, the cost of developing these games is also likely to be far lower than at large publishers or studios. Thus, a good return on investment for any games projects I work on in future is achievable.

Appendix E – Team Consensus

Team Consensus

I have read and approve of the content as a representation of the team's work and my contribution.

Team Member (full name)	Signature	Date
Lev Working	Lev Working	12/7/23
Will Cromer	Will Cromer	12/7/23
Giovanni Tuzzio	Giovanni Tuzzio	12/7/23
Rudy Manian	Rudy Manian	12/7/23

Attention:

- Make sure every member of the team has a copy of the final document.
- Save this document at **TeamName_GroupProject_F2023.pdf**
  - Replace TeamName with your team's name.
- This is a group submission. Only one person from the team needs to submit the short paper.
- Upload the final version in Brightspace.
- Make sure the final version is accessible from the team's project web page.