# Video-Games: Metacritic

![image](https://hips.hearstapps.com/hmg-prod.s3.amazonaws.com/images/gridoffset-videogames-1-1585583517.jpg)

## Entertainment Project Plan and Presentation

Link to presentation can be found [Here](https://docs.google.com/presentation/d/1Ufqt-ugDe9Q2NsigC36nNvBCTmWIoBLgSVdLJVUhqxk/edit#slide=id.gc6f980f91\_0\_0)

Data Source: https://www.kaggle.com/brunovr/metacritic-videogames-data

### Overview

#### Why We Chose This Topic?

For the final project, the team has decided to analyze the video game industry based on our shared interest in Entertainment. Video games are continuously evolving in the tech industry. It has become a popular culture and provides a fun and social form of entertainment. Its relevance to the tech will be applicable for our tech career. The team has sourced the video game dataset from Kaggle (found in the link above). Based on the dataset, our objective is to predict the most anticipated video game franchise based on sales and scores.

#### Questions To Answer:

- Based on how many metacritics (user) and critics, are we able to predict the score for an anticipated video game?

- Based on the review scores (critic and users) and franchise, are we able to to predict video game sales and its projected growth?

- Does genre, review scores, platform, developer, franchise, and release dates have any impact on video game (vg) sales?

The team will analyze which is the most anticipated game/franchise release and by which developer, which game would have the highest rating in the future, which game developer has the best score, as well as what genre of games is most popular by number of players (single vs. multiplayer).

#### Technologies Used:

Python, Pandas, Jupyter Notebook, SQL, SQLAlchemy, Tableau, Excel

For further information on what techonology we'll be using and how, please see the techonology.md file.

### Module 20 Deadlines:

- Module 20 First Segment Project Deliverable due March 6th, 2022

- Module 20 Second Segment Project Deliverable Due March 20th, 2022

- Module 20 Third Segment Project Deliverable Due March 27th, 2022

- Module 20 Final Segment Project Deliverable Due March 31st 2022

### Entertainment Project Team and Roles for Deliverable 1:

Team Member Role Description

- Rasheem G.- X: The X role will focus on the technology side of the project \*uploaded readMe on his branch for project plan\*

- Caitlin B.- Circle: The circle role will oversee the mockup database \*uploaded mockup database as .csv file on main branch\*

- Edin C.- Triangle: The triangle role is responsible for creating a simple machine learning model. \*please see Edin's branch for machine learning model\*

- Tasnuva M.- Square: The square role will be responsible for setting up the repository. \*pushed and uploaded README and technologiy file to main branch\*

The roles above will apply for the first segment of the module 20. As required, we will plan to rotate the X role among team members. In the upcoming week the team will determine the rotation of the X role to align with the deliverable deadlines.

### Entertainment Project Team and Roles for Deliverable 2:

Team Member Role Description

- Rasheem G.- Square: The square role will refine the machine learning model using (train and test) \*uploaded readMe on his branch for project plan\*

- Caitlin B.- Circle: The circle role will clean and analyze data and create visuals to accompany data story \*uploaded mockup database as .csv file on main branch\*

- Edin C.- Triangle: The triangle role is responsible for transforming the mockup database into a full database that integrates with data. \*please see Edin's branch for machine learning model\*

- Tasnuva M.- X: The X role will outline and conceptualize dashboard and presentation as well as update the repository. \*pushed and uploaded README and technologiy file to main branch\*

The roles above will apply for the second segment of the module 20. As required, we will plan to rotate the X role among team members. In the upcoming week the team will determine the rotation of the X role to align with the deliverable deadlines.

\*all members have 4 commits in their branches.

### Data Exploratory and Analysis:

#### Initial Clean up:

Sample data set for machine learning module:

1) Dropping Player column.

2) Clean & use 2021 Sales per game csv data for additional info / join requirement (postgres).

3) All potential duplicate values are an exact match for easy filtering, reducing numerical/grammatical errors.

4) "Release Date" column formayted mm-dd-yy.

5) Format "Genre" column for grouping.

6) Create new column determining "Franchise"

7) Verify "Critics" column usage.

8) Line or Logistic Regression model for determining Developer/Franchise success rate.

#### Understanding our raw data file - Data columns:

\* name: The name of the game

\* platform: Platform it was released

\* r-date: date it was released

\* score: average score given by critics (metascore)

\* user score: average score given by users in the website

\* developer: game developer

\* genre: genre of the game (can be multiple)

\* players: Number of players (DROP THIS COLUMN due to missing information)

\* critics: number of critics reviewing the game

\* users: Number of metacritic users that reviewed the game

#### Table Data:

\* Critic Scores/User Scores by Developer

\* Critic Scores/User Scores by Platform

\* Developer by Release Date

\* Franchise Scores by Release Date

\* Franchise by Critics

\* Critics/Score by Genere

### Machine Learning:

Link to Machine Learning description [Here](https://github.com/taznuva/Entertainment\_project/blob/Rasheem/Module%2020%20Second%20Segment.docx)

### Database Integration:

We used Pandas and SQLAlchemy to load the csv files that we started with, and merged tables that we created into a SQL database. Our database is then able to interact with our machnie learning model, which allows us to make predictions and analyze our results.

### Dashboard:

We will be using Tableau and Pandas to build visualizations of our analysis which we will showcase on Google Slides.

For the interactive portion, we will build the options to manipulate the features for the graphs.

### Communication Plan:

The entire team will mainly communicate via the entertainment Slack group chat. The team will also use scheduled meeting times via Zoom on Tuesday and Thursday 02-22-2022 – 03-31-2022. Additional meeting to be scheduled via Zoom as needed. Emails have been exchanged in case we're not active on Slack. The team has also created a shared Notion page to checklist tasks and to keep track of the timeline.