

Releasing A Successful Video Game


Capstone Project (Part II) By:

Martell Tardy

Noor Bahr Al Uloom



Problem Statement

A person wearing a VR headset and holding a controller is standing in the center of a room, surrounded by several people sitting in chairs. The room has large windows in the background. The entire image has a blue tint.

Can the use of historical data identify the key features of past profitable video games and predict the characteristics of a similar deliverable for the Valve Corporation to release in the last quarter of 2018?

Objectives

The Objectives



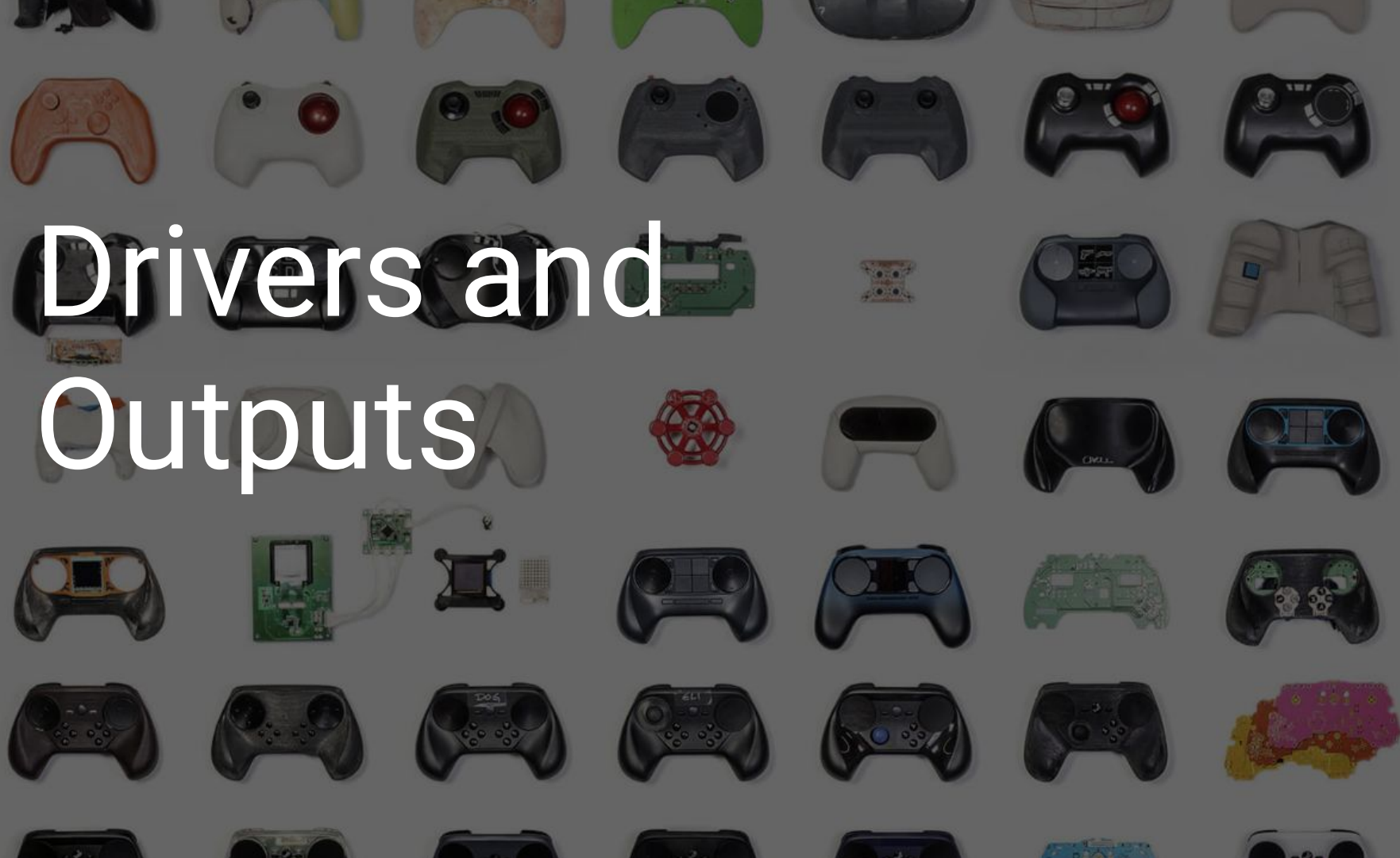
to identify which combination of genre characteristics and gameplay, correlates to profitable and popular video games historically



to determine if a video game's frequency in tournaments, player participation, and amount of award money at tournaments, correlates to profitable and popular video games historically

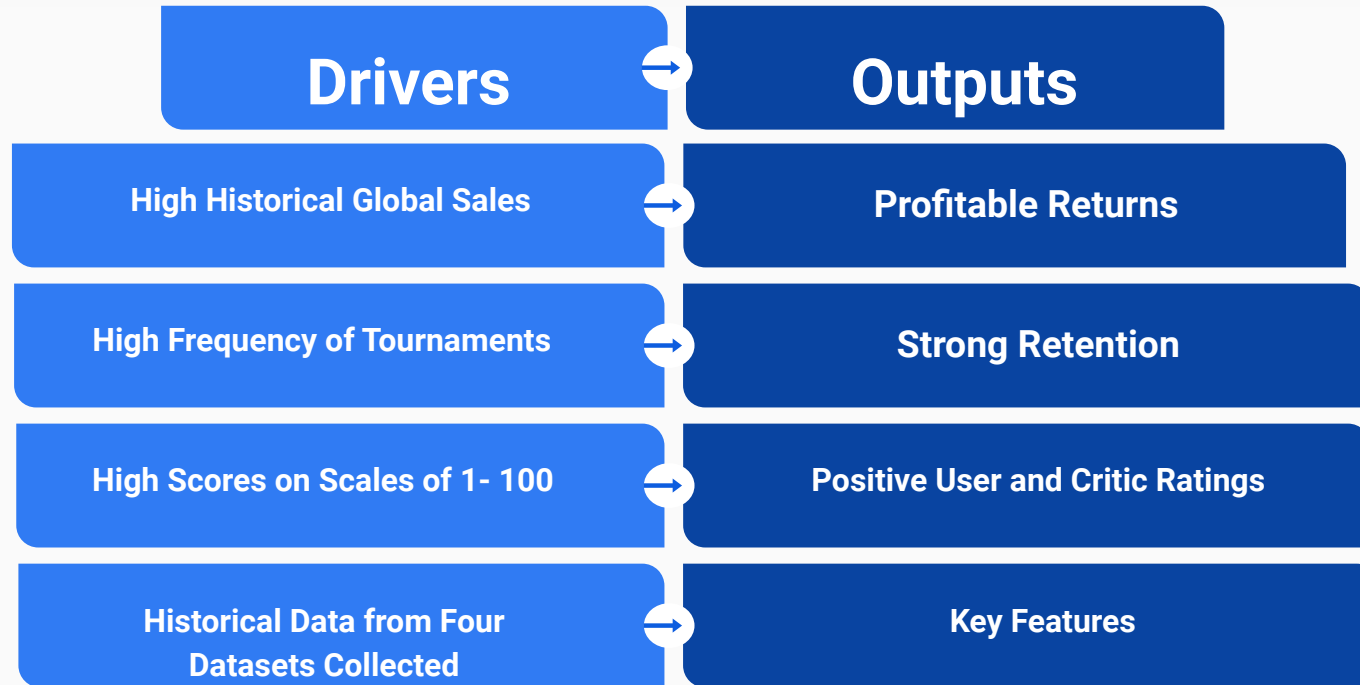


To determine if profitable sales of a video game, correlates to high tournament participation and certain genres and gameplay experiences



Drivers and Outputs

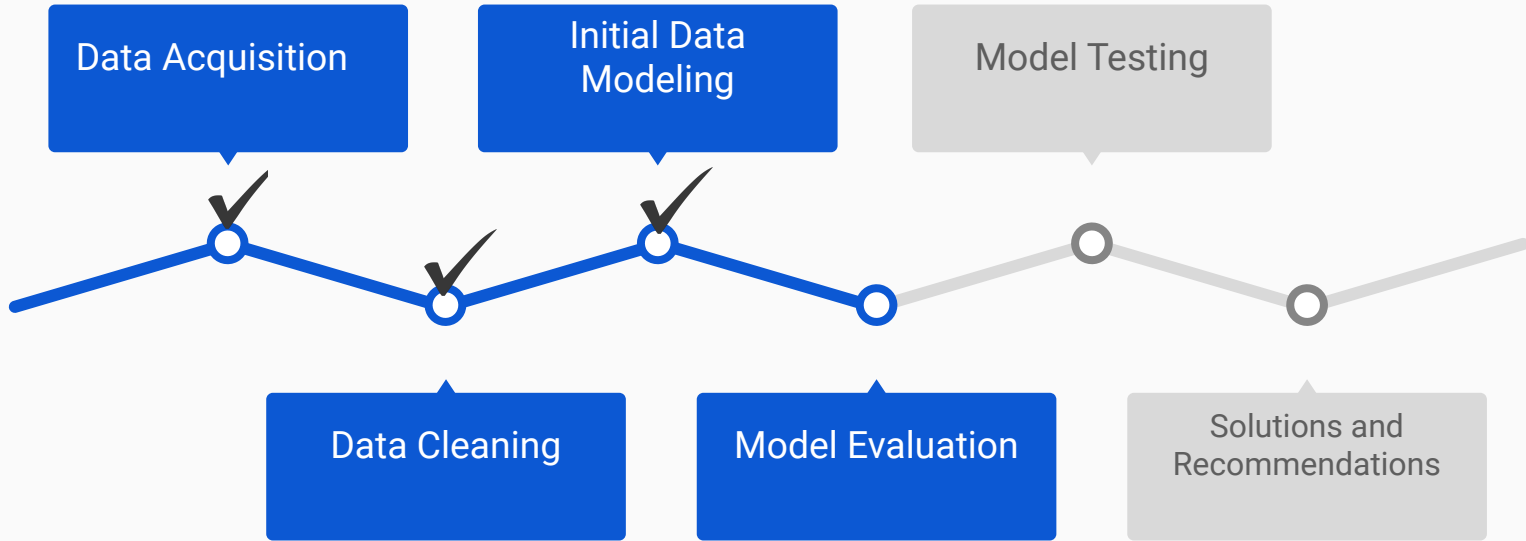
Drivers and Outputs





Project Status

Project Status

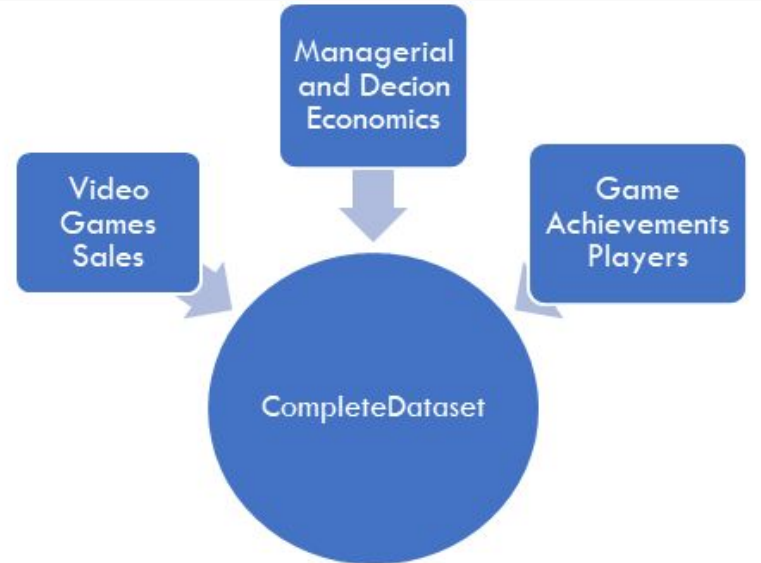




Data Cleaning and Preparation

Data Preparation

- All three datasets were merged into one single dataset
- Outliers were checked



Data Cleaning

Variable	Issue	Action	Result
Release	NA	Deletion (Row)	148 Rows Deleted
Global_Sales	NA	Deletion (Row)	23 Rows Deleted
Total Tournaments	NA	Deletion (Row)	6 Rows Deleted
NA_Sales	NA	Imputation (Mode)	10 Instances Changed to 0.1
User Score	Not Scaled 1-100; NA	Multiplied by 10; Imputation (Mean)	Data Range 1-100; 38 Instances Converted to 64
Critic Score	Not Scaled 1-100; NA	Multiplied by 10; Imputation (Mean)	Data Range 1-100; 27 Instances Converted to 83
Tournament Money Awarded	NA	Imputation (Min.)	17 Instances Converted to 25
Tournament Total Players	NA	Imputation (Min.)	18 Instances Converted to 1
Genre and Manufacturer Based Variables	Duplicates, NA	Deletion (Column)	317 Columns Deleted





Clustering

What is the natural grouping of video games in correlation to their sales, tournament participation, and gameplay experience?

Variables Used:

- **GameTitle**
- **Released**
- **NA Sales**
- **Global sales**
- **User score**
- **Critic score**
- **ESRB rating**
- **Tournament Money Awarded**
- **Tournament Total Players**
- **Total number of Tournaments**
- **22 types of Genre (logical variables)**

Findings:

A video game launch with a similar gameplay as Cluster 3 would prove successful for Valve Corporation. Cluster 3 is a video game rated M for mature. It is an Action focused gaming experience with a Science Fiction based setting and narrative, played from the First-Person Perspective.

Tab.3. Comparing Clusters

	Tournament (Avg)	Gameplay	Scores (Avg)	Sales (Avg)
CompleteDataset (Standard)	Money = \$46000 Players = 13 Tournaments = 5	E = 56 T = 74 M = 4 Action = 88 Fighting = 60 Shooter = 43 FirstPerson = 43 (34%)	UserScore = 75 CriticScore = 83	NA_Sales = 0.380 GlobalSales = 1.010
Cluster 1	Money = lower Players = higher Tournament = lower	T = 54 (87%) Fighting = 49 Action = 49 FirstPerson = 0	Similar	Lower
Cluster 2	Money = higher Players = higher Tournament = similar	T = 11 (69%) Strategy = 14 (88%) FirstPerson = 0	Similar	Lower
Cluster 3	Money = higher Players = higher Tournament = higher	M = 34 (79%) Action = 35 ScienceFiction = 18 FirstPerson = 43 (100%)	Similar	High Higher (double)
Cluster 4	Money = higher(double) Players = lower Tournament = lower	E = 37 (100%) Sports = 36 FirstPerson = 0	Lower Similar	Higher (double)
Cluster 5	Money = lower Players = lower Tournament = lower	E = 12 (92%) Racing = 13 (100%) Simulator = 5 FirstPerson = 0	Higher	Higher(quadruple) Higher(double)



Correlation

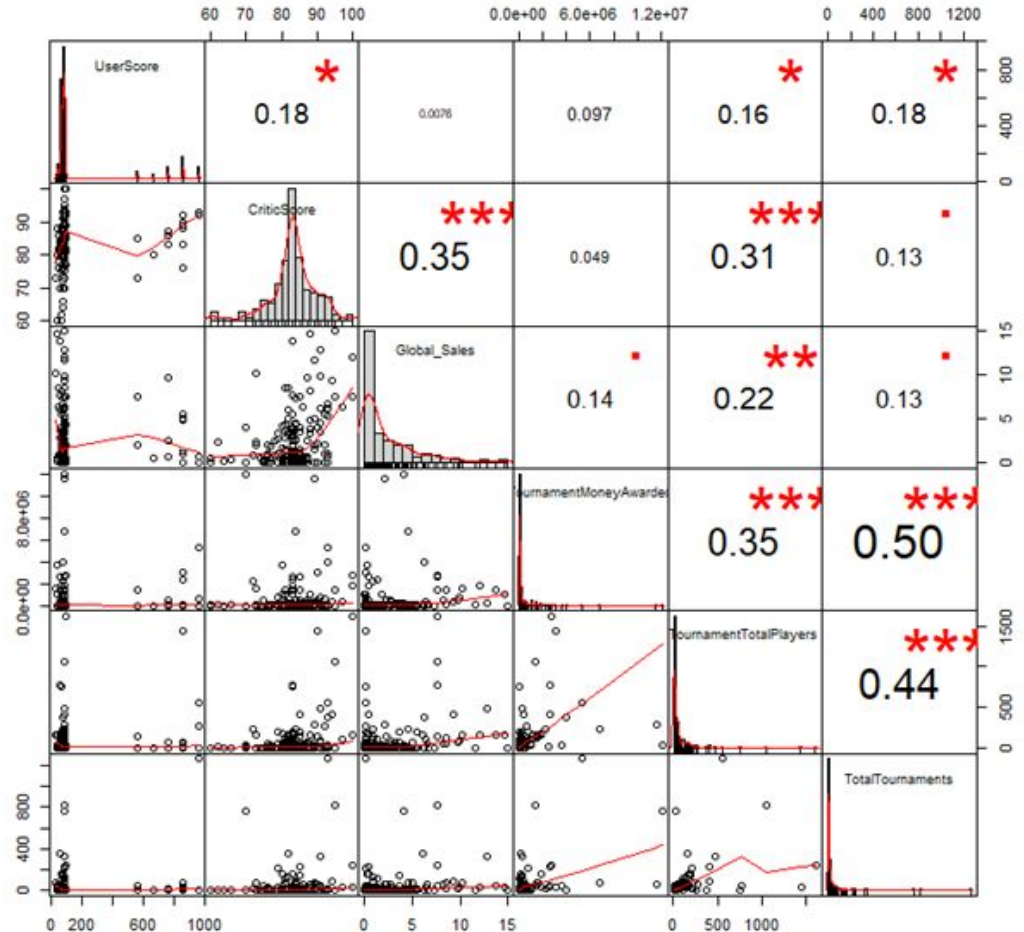
What is the correlation between sales, tournaments and scores (by critics and users)?

Variables Used:

- **Global sales (as Response)**
- **User score**
- **Critic score**
- **Tournament Money Awarded**
- **Tournament Total Players**
- **Total number of Tournaments**

Findings:

- Strong positive correlation between tournament money awarded and total tournament; $r = 0.5$
- Medium strength of positive correlation between total tournaments and total tournament players; $r = 0.44$
- Medium strength of positive correlation between tournament Money awarded and total tournament players; $r = 0.35$
- Medium strength of positive correlation between critics score and global sales; $r = 0.35$
- Medium strength of positive correlation between critics score and tournaments total players; $r = 0.31$





Decision Tree

How was global sales affected by genres, ESRB ratings, number of tournaments and number of players?

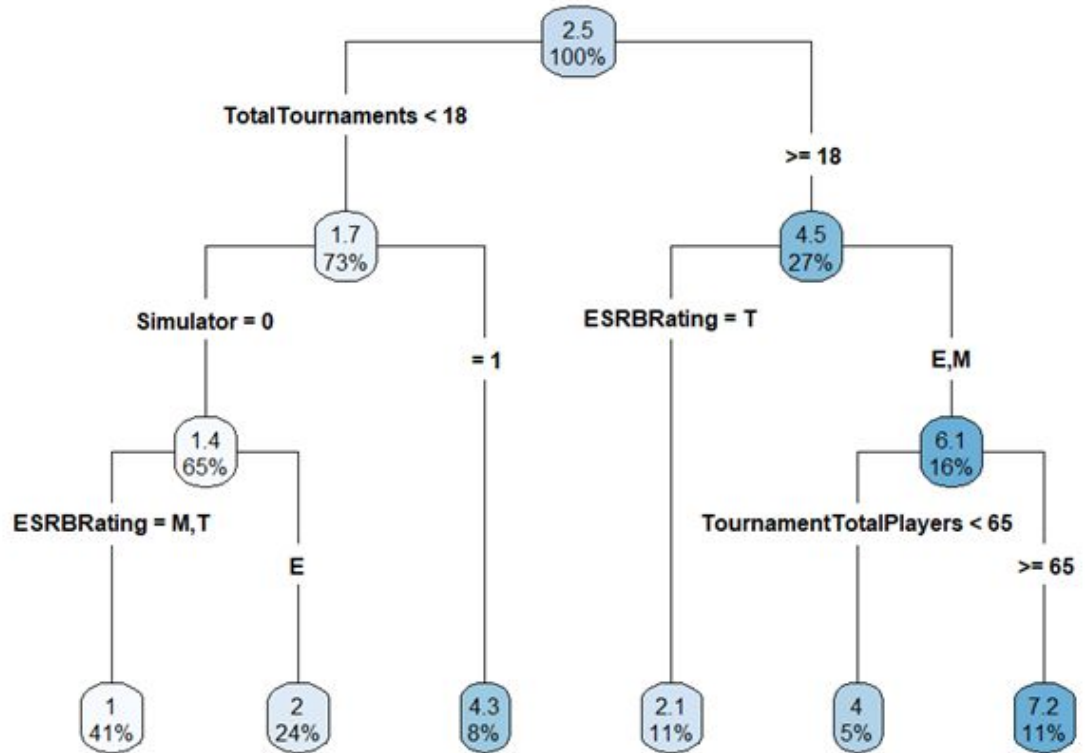
Variables Used:

- **Global Sales (as Response)**
- **ESRB ratings**
- **22 types of Genre (logical variables)**
- **Number of players**
- **Number of tournaments**

Findings:

The decision tree above suggests that with low number of tournaments (<18 tournaments), there is still chance to achieve higher global sales by choosing simulator genre.

With higher number of tournaments (≥ 18 tournaments), there is a chance to achieve higher global sales with ESRB ratings of E(Everyone) and M(Mature). If ESRB ratings of T(teen) was chosen, then there is a chance to have lower global sales.





Poisson Regression


How does games' genres
effect number of
tournaments players?

Variables Used:

- Number of players in tournaments (as Response)
- 22 explanatory variables that describes types of Genre (boolean variables)

Poisson Regression

With p-value less than the significant alpha (0.05), the model suggest that number of tournament players is affected by some genres.



Genres that have positive relationship with number of total players (in tournament) are (descending order by number of coefficient):

- Party
- RPG (Role-Playing games)
- Comedy
- Beat.emUp (Beat them Up)
- Strategy
- Fantasy
- Action
- FirstPersonPerspective
- Warfare
- Adventure

Genres that have negative relationship with number of total players (in tournament) are (descending order by number of coefficient):

- Sports
- Simulator
- Puzzle
- Science Fiction
- Racing
- Horror
- TBS (turn-based strategy)



ANOVA

Do different types of genres achieve the same average of global sales?

Variables Used:

- **Global Sales (as Response)**
- **22 explanatory variables that describes types of Genre (boolean variables)**

ANOVA

With p-value less than the significant alpha (0.05), we fail to reject the null hypothesis, we conclude that we have enough statistical evidence that not all genres has the same average of global sales. The games with genres with significant p-values are:



First Person
Perspective

Strategy

Fighting

Platform

“Beat them
Up”



Linear Regression

Is there a linear relationship between sales and critic score, users score, and tournaments?

Variables Used:

- **Global sales (as Response)**
- **User score**
- **Critic score**
- **Tournament Money Awarded**
- **Tournament Total Players**
- **Total number of Tournaments**

For global sales and critic scores, with p-value less than the stated alpha (0.05) we reject the null hypothesis and conclude that we have enough statistical evidence to confirm that there is a positive linear relationship between critic score and global sales. However, R² value suggests that this model only explains 15% of data variation.



Preliminary Conclusions

Preliminary Conclusions

Video Game Profile:

- ESRB Rating M, for Mature
- Action focused (genre)
- Science Fiction setting and story narrative
- First-Person Perspective for player

A grid of various video game controllers and components, including different models, colors, and internal parts, arranged in a grid pattern. The controllers include various models like PlayStation, Xbox, and custom builds. Some are disassembled, showing internal components like circuit boards and joysticks. The colors range from standard black and white to custom colors like orange, green, and red.

Thank You!