Releasing A Successful Video Game

Capstone Project By:
Martell Tardy
Noor Bahr Al Uloom

Valve Corporation

Is a video game developer, publisher and digital distributor.

OVERVIEW

Valve Corporation made video games until 2013 then transitioned their focus

Half-Life

A horror first-person shooter video game series where a physicist battles an alien invasion.

Team Fortress

A first-person shooter video game series that pits two teams against each other

Counter-Strike

A multiplayer first-person shooter video game series where teams of terrorists and counter-terrorists battle

Day of Defeat

A team-based multiplayer first-person shooter video game series set in WWII

Portal

A first-person puzzle-platform video games series set in the Half-Life universe

Left 4 Dead

A first-person shooter, survival horror video game series set after a zombie apocalyptic event

Dota 2

A free-to-play multiplayer online battle arena video game.

TBD

Unknown

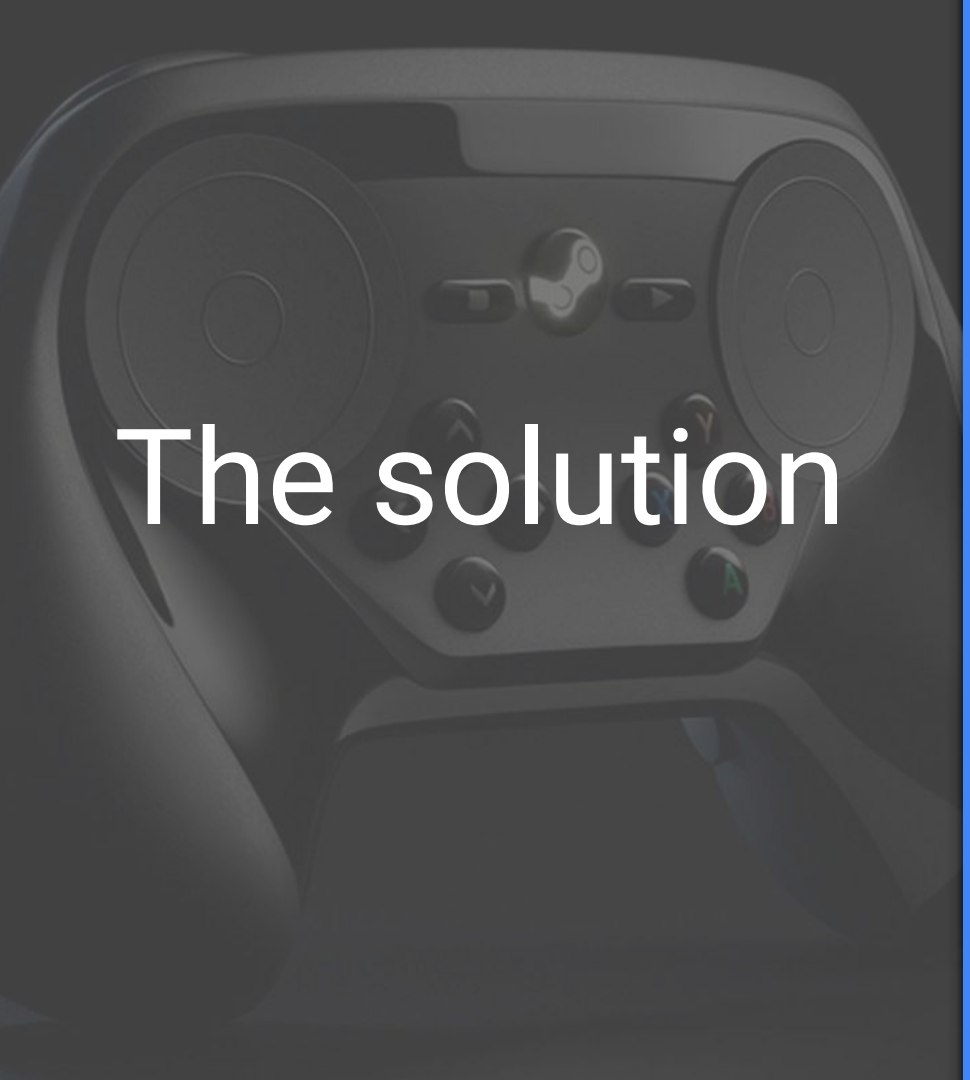
The Business Problem

Can Valve Corporation release a video game that is popular and profitable?

With Valve Corporation's unique composition of being a digital content distribution channel, an international tournament sponsor, a producer of consumer electronic devices and prior video game developer - Gabe Newell is confident there is no better time than now for Valve Corporation to produce the next big video game to complete their domination of the video game industry.

The Analytical Problem

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?



The solution

Check historic data to understand more about:

- User Interaction
- Sales
- Tournament Retention
- Ratings

The Objectives

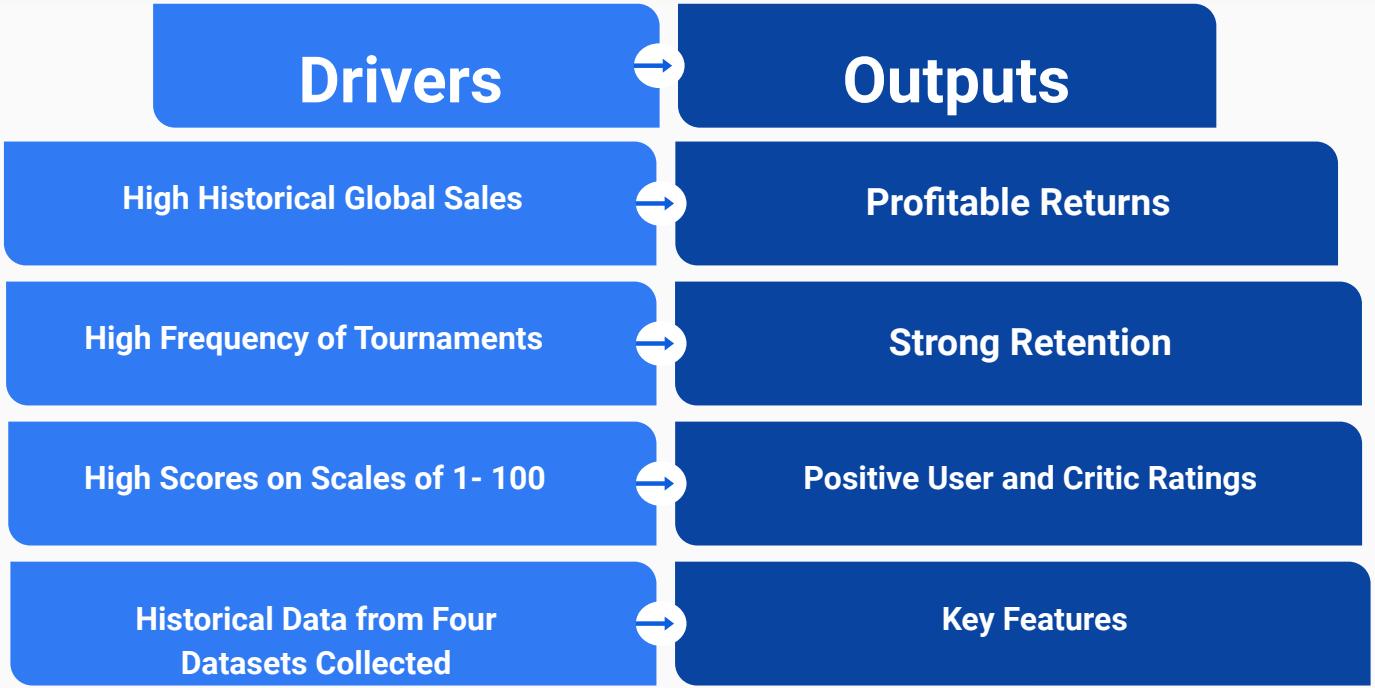
To identify which combination of the twelve genre characteristics of video games historically results in maximized sales, user interaction, ratings, and tournament retention

To identify which of or the combination of the 30 types of Consoles historically show positive correlation to maximized sales, user interaction, ratings, and tournament retention

To determine the probability of a video game's frequency of play by users on an online platform and at tournaments

To determine if the eight types of the ratio variable ESRB Ratings historically correlates positively to maximized sales and ratings

Drivers and Outputs



The Methodology

Data Management

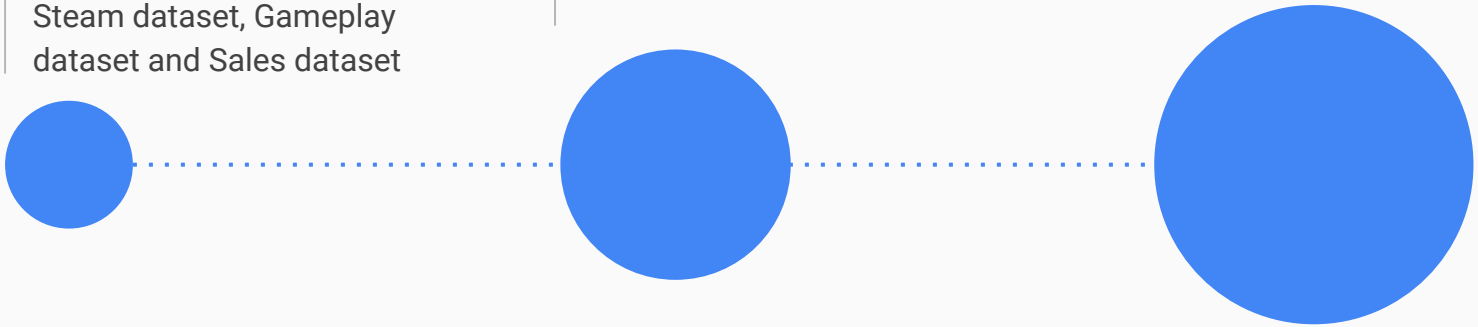
Clean, combine and manage datasets: Tournament dataset, Steam dataset, Gameplay dataset and Sales dataset

Descriptive Statistics

Find our relationships between: sales, genre, tournaments, player ratings and critic scores

Prescriptive Statistics

Provide recommendations for the new game release using: Decision Analysis and Time Series Forecasting



Tools

Here are a few of the tools used to solve our analytical problems

Excel

WEKA
The University of Waikato

SilverDecisions

Visio

Studio®

+ableau®

Data Cleaning and Preparation

Dataset Categories

The datasets selected derive from multiple sources relating to the video gaming industry and its historical performance.

Tournament

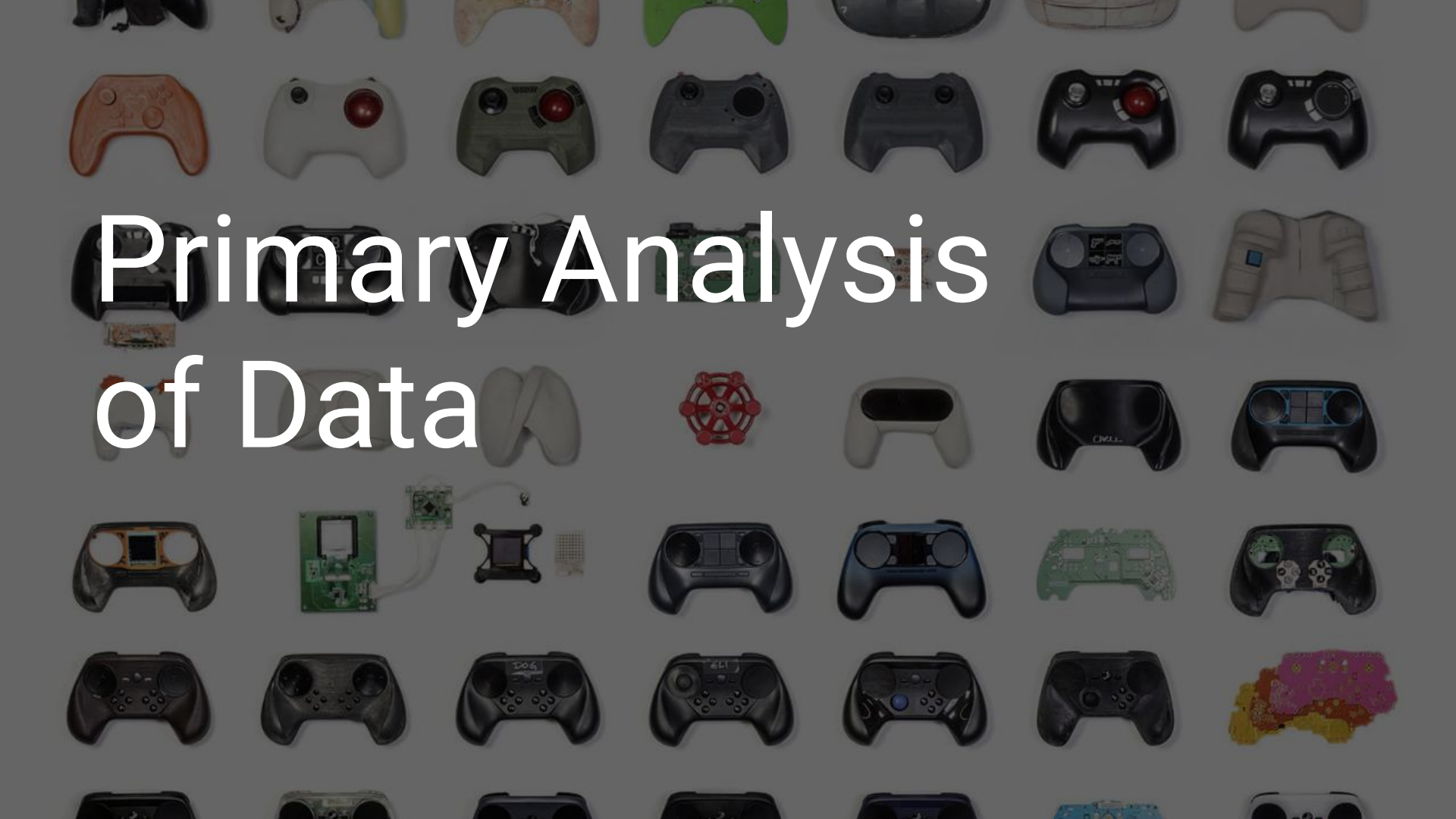
Steam

Gameplay

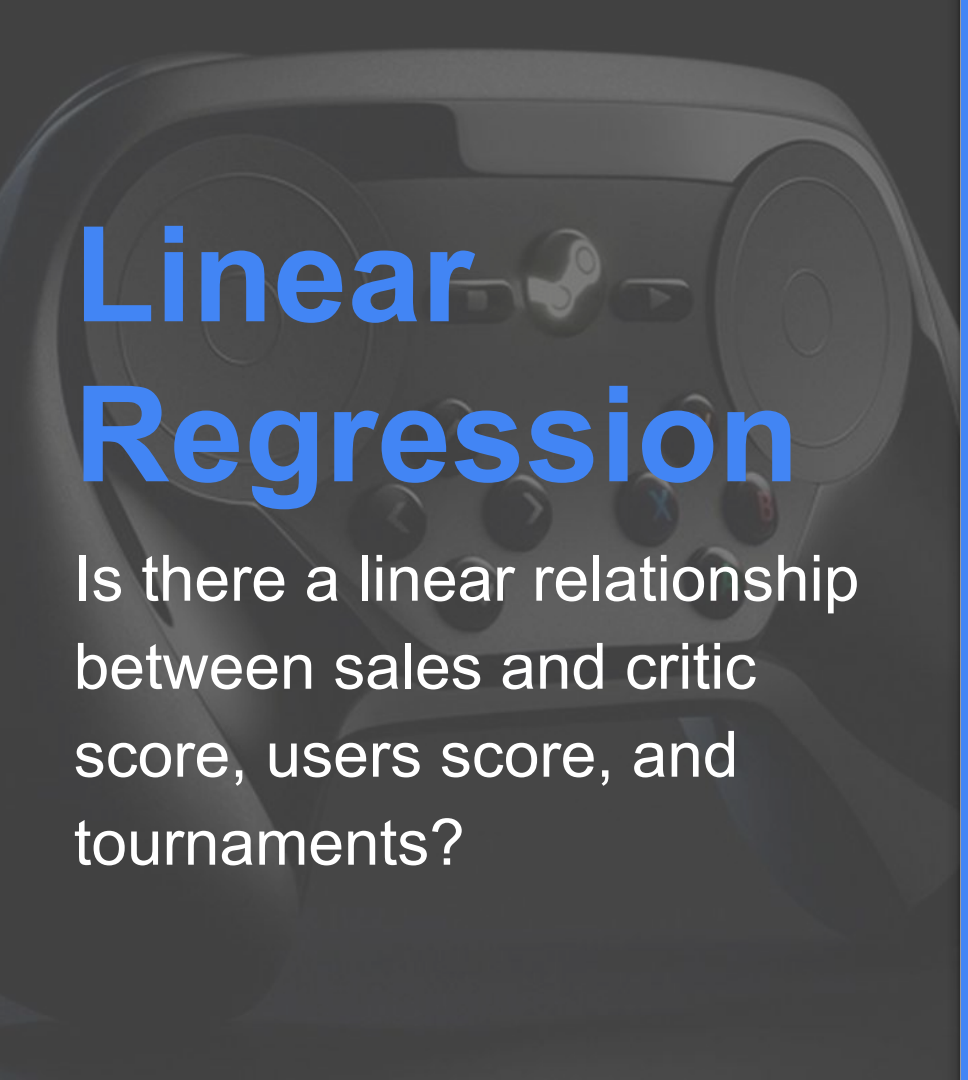
Sales

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



Primary Analysis of Data



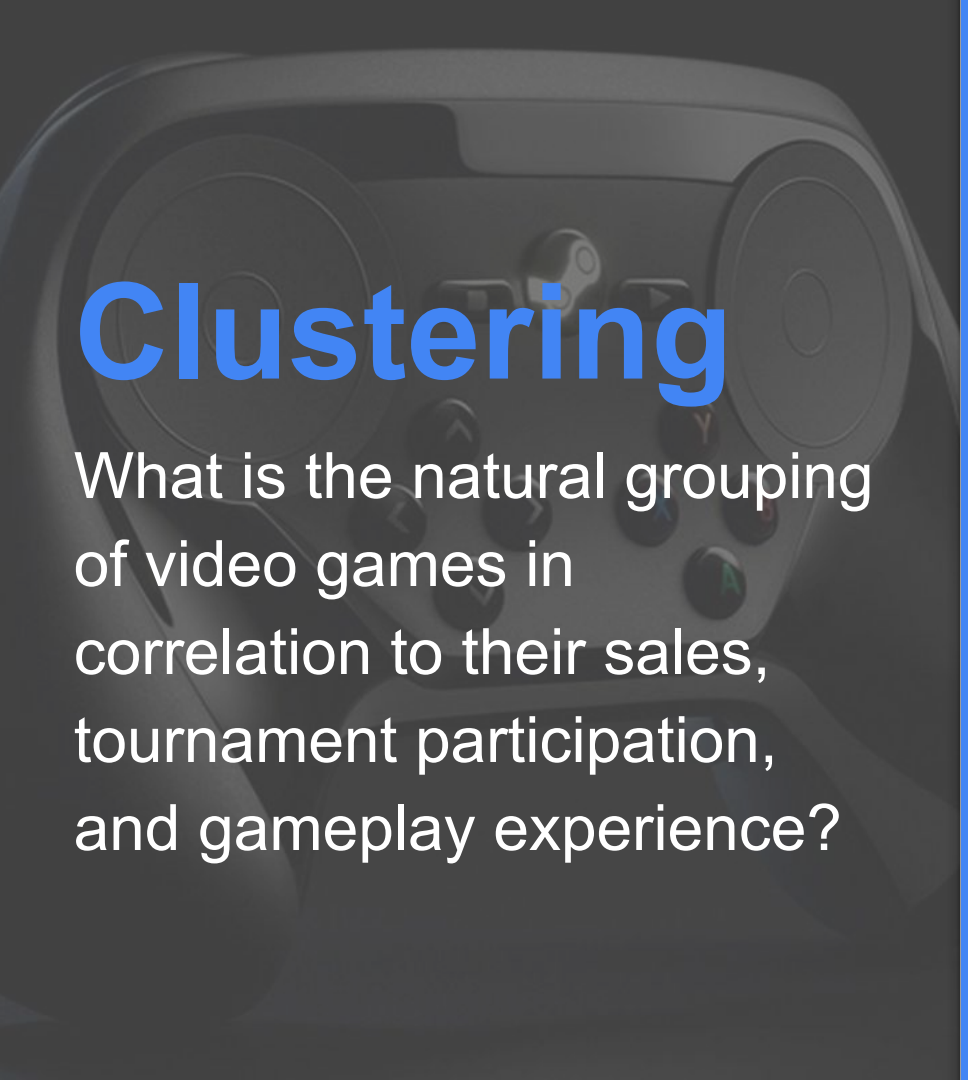
Linear Regression

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.



Clustering

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

Variables Used:

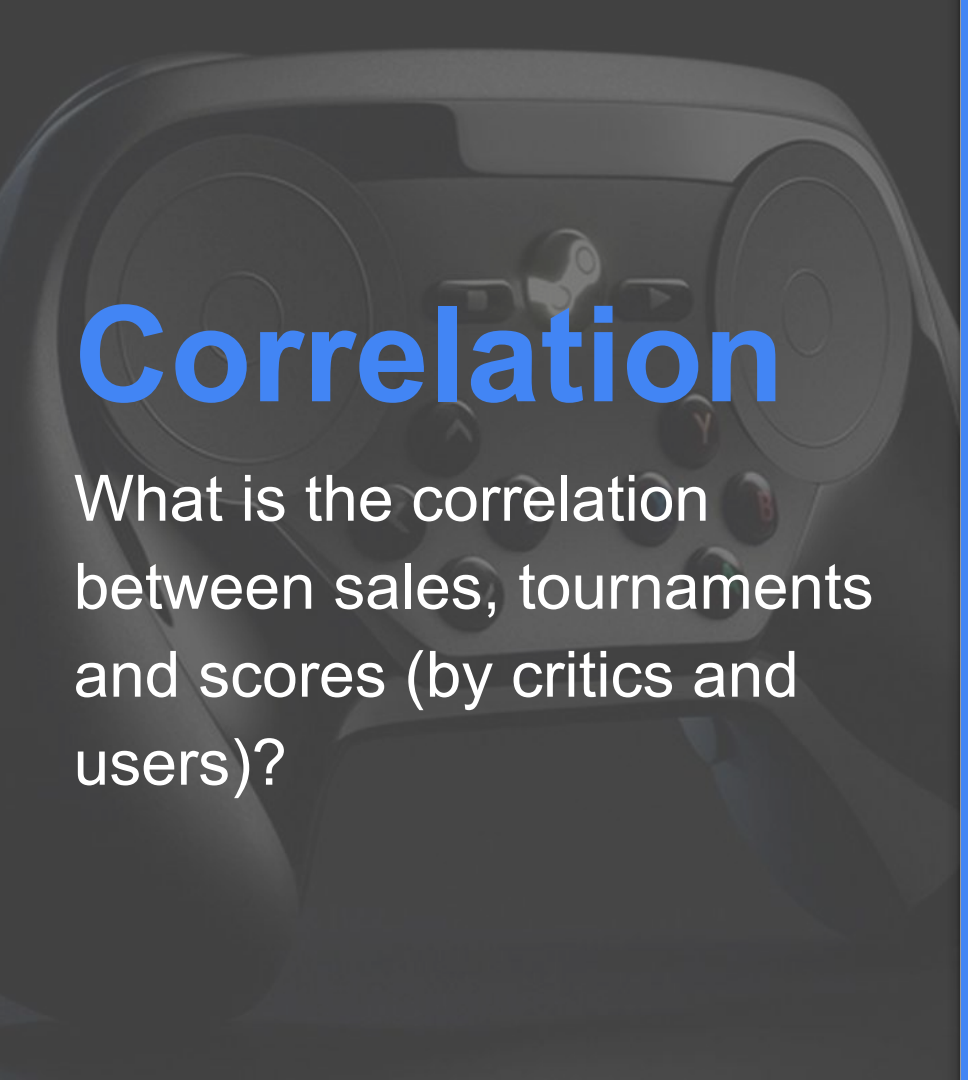
- GameTitle
- Released
- NA Sales
- Global sales
- User score
- Critic score
- ESRB rating
- Tournament Money Awarded
- Tournament Total Players
- Total number of Tournaments
- 22 types of Genre (logical variables)

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

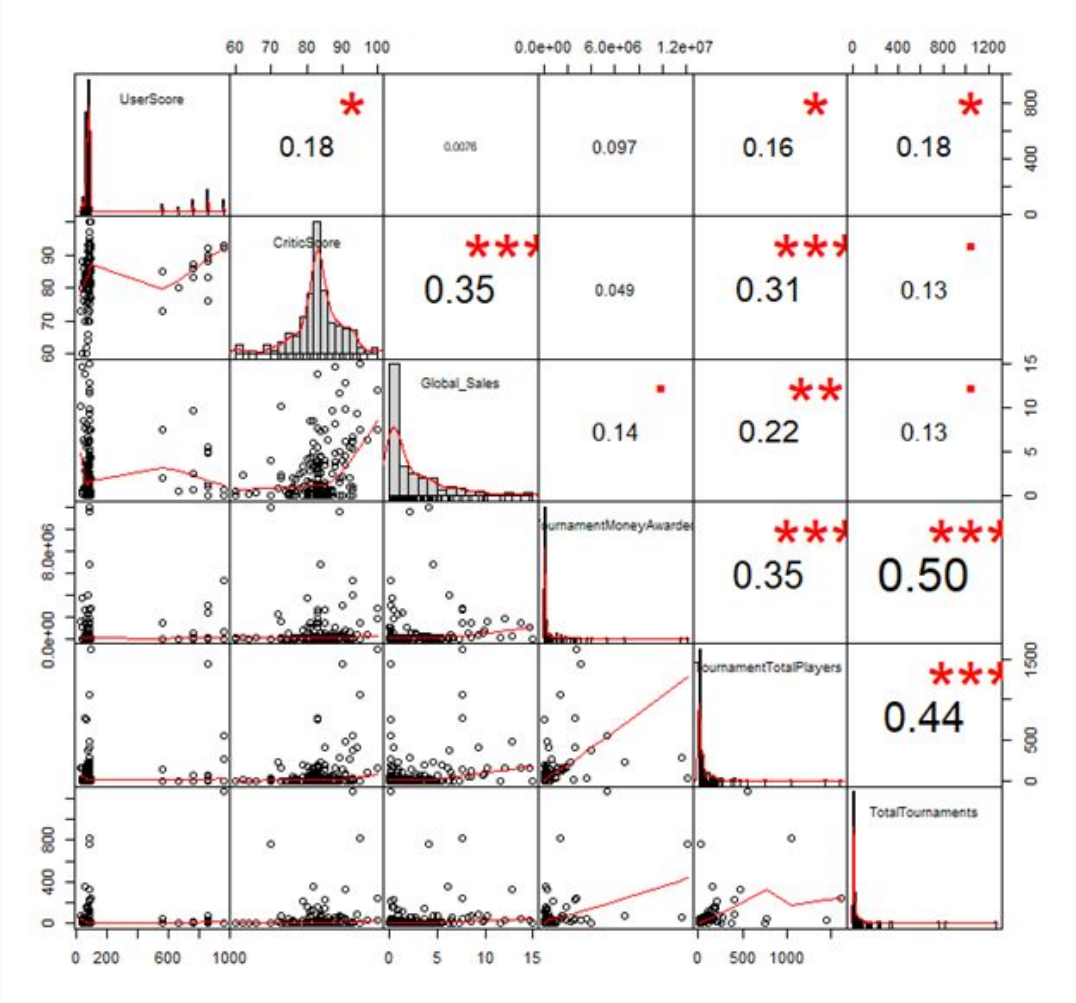
What is the correlation between sales, tournaments and scores (by critics and users)?

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?

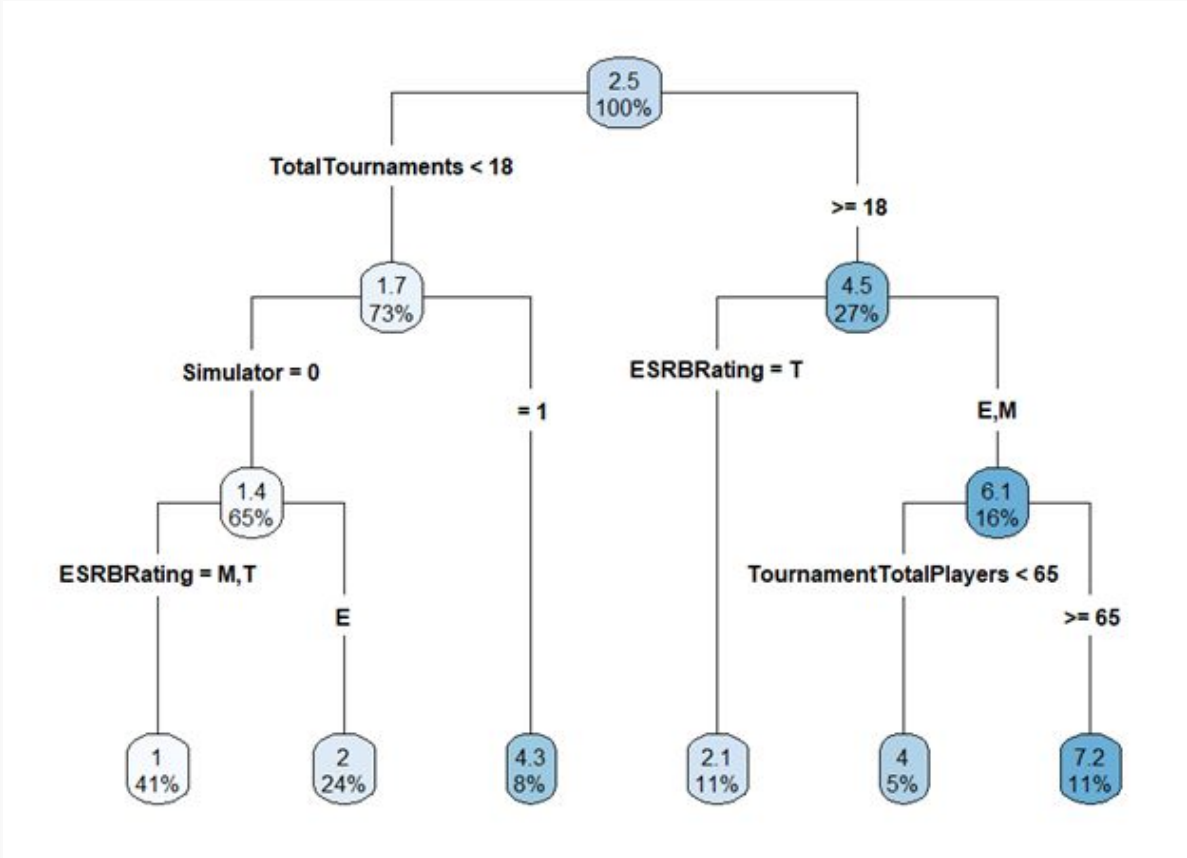
Variables Used:

- 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

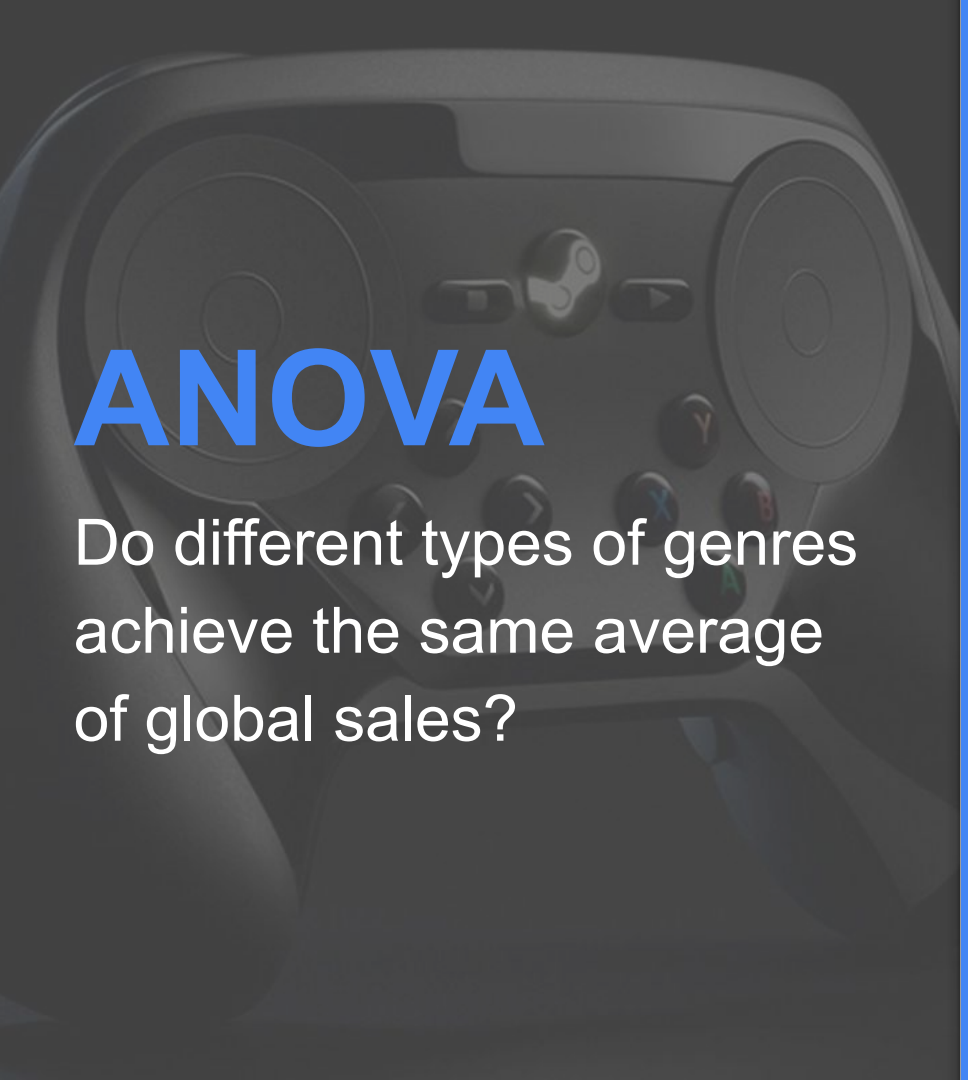
With p-value less than the significant alpha (0.05), the model suggest that number of tournament players is affected by some genres.

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

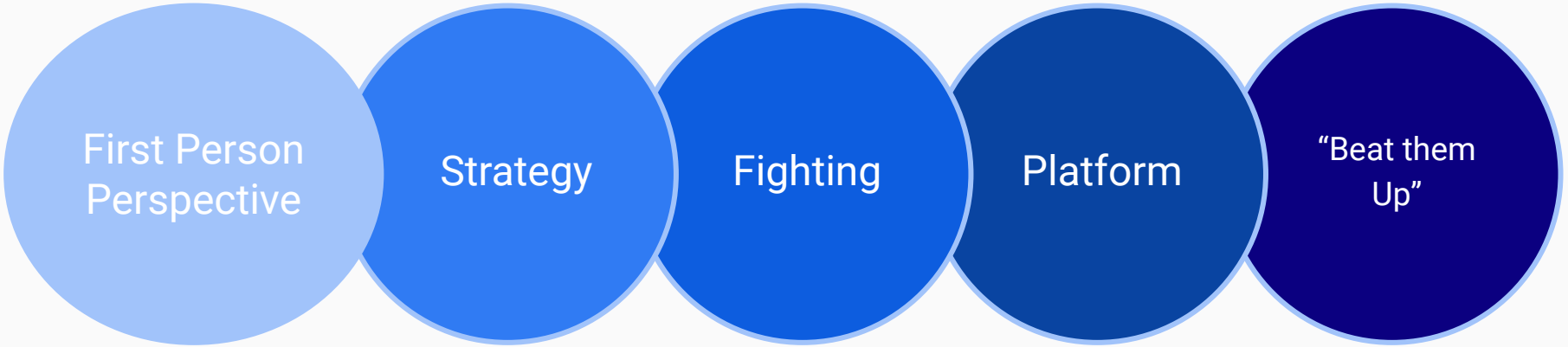
Do different types of genres achieve the same average of global sales?

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:



Conclusions

What Models Were Used?

Game Genre	Clusters (PAM)	Poisson regression	Anova	Decision Trees
Action	Yes	Yes	no	no
Science Fiction	Yes	no	no	no
First Person Perspective	Yes	Yes	Yes	no
Strategy	no	Yes	Yes	no
Fighting	no	no	Yes	no
Platform	no	no	Yes	no
Beat.emUp	no	Yes	Yes	no
Party	no	Yes	no	no
Role Playing Games	no	Yes	no	no
Comedy	no	Yes	no	no
Warfare	no	Yes	no	no
Adventure	no	Yes	no	no
Simulation	no	no	no	yes
Fantasy	no	Yes	no	no

Game ESRB Rating	Clusters (PAM)	Decision Tree
E: Everyone	no	Yes
T: Teen	no	no
M: Mature	yes	Yes

Recommendations

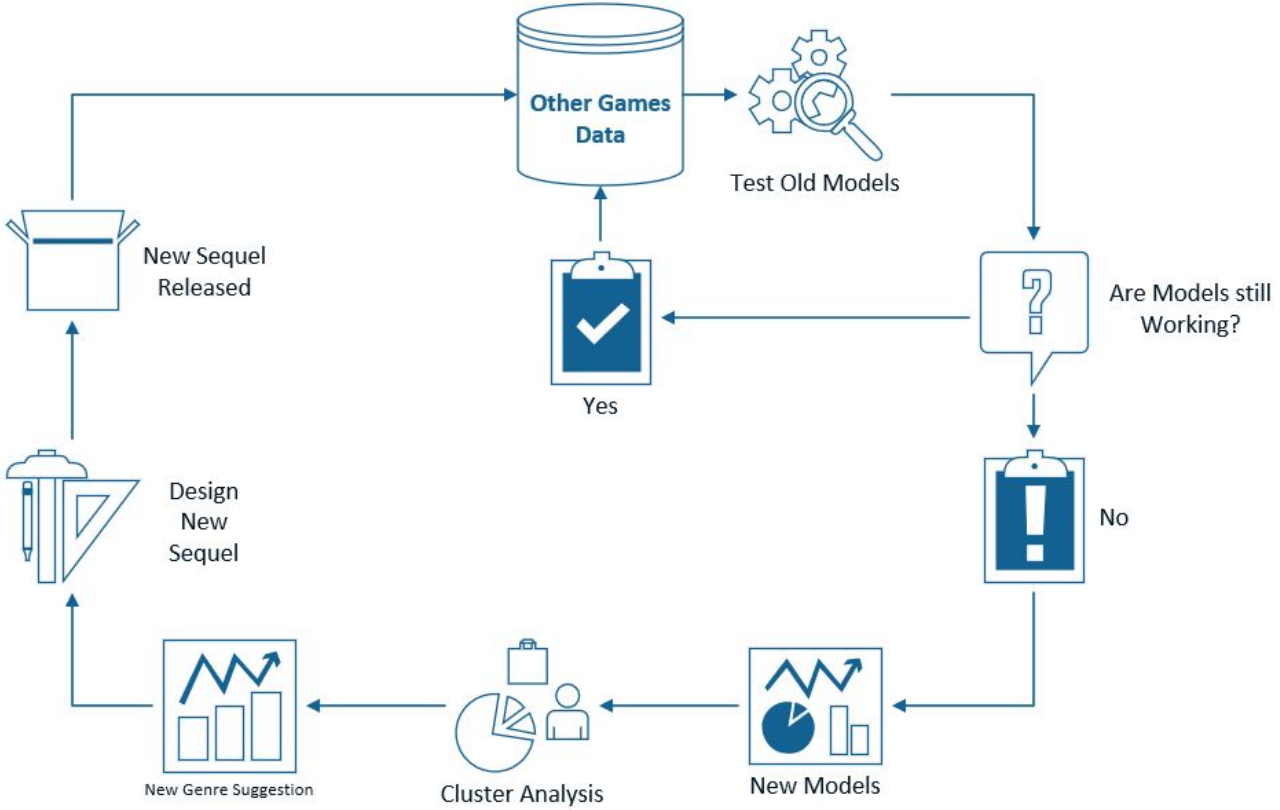
What is our Recommendation?

Video Game Profile:

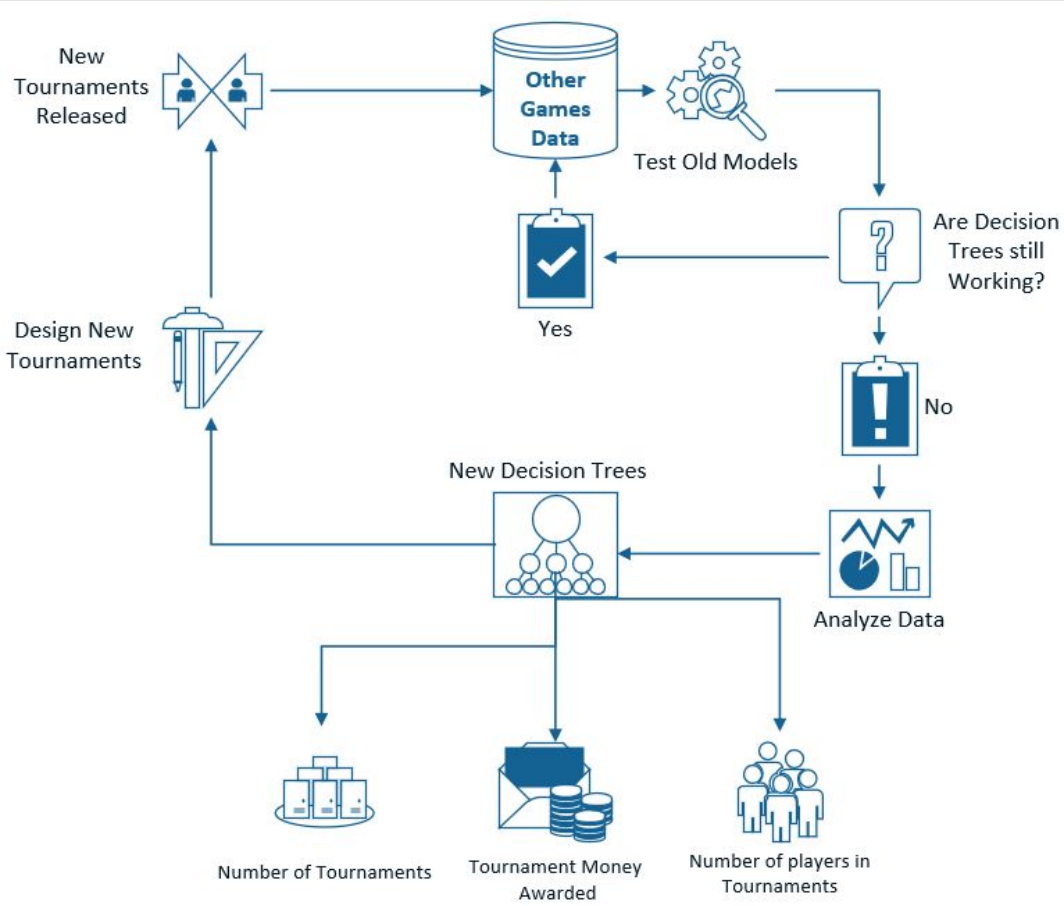
- ESRB rating M for mature
- Genres of Action, Strategy, and Beat.emUp
- Played from the First-Person Perspective
- Average of total money awarded in tournaments is $1,241,553
- Average number of players in tournaments is 114
- Average number of tournaments is 45

Model Lifecycle

Model Life Cycle (Genre for Sequel)



Model Life Cycle (Tournaments Planning)




Thank you!