Predicting The Future Of Gaming ?
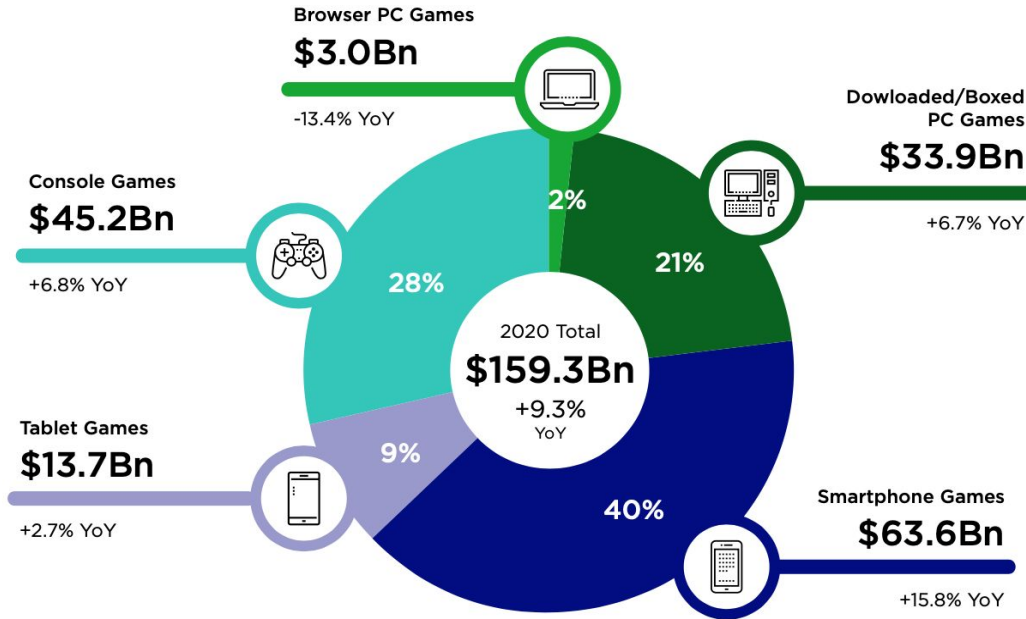
By:
Maria Castellanos
Blake Kennedy
Michelle Johnson
Joseph Thompson
Nicholas Singh

Topic

Video games have been around since the mid-1900s, but it was not until milestones such as the release of PacMan and Nintendo's creation of Donkey Kong in 80s that the industry really started to boom. Through the eras of arcade gaming, the development of multiplayer games, and then to the creation of home gaming consoles -the population of global players has steadily continued to rise [1].
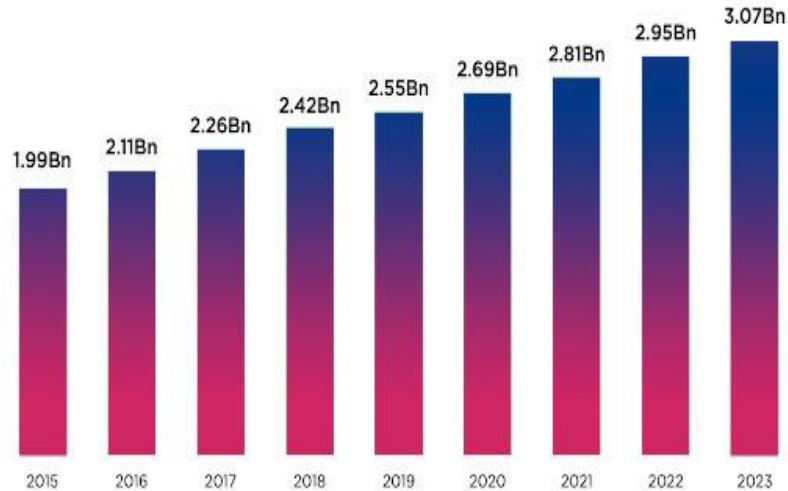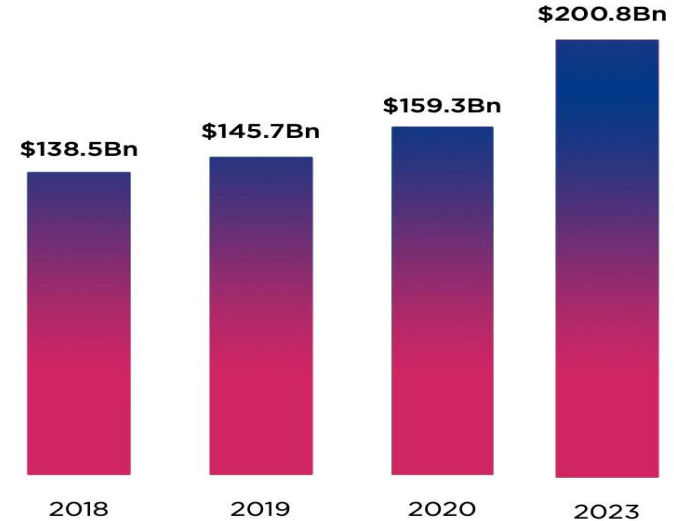
2020 Global Games Market
Per Segment

Browser PC Games
$3.0Bn
-13.4% YoY

Dowloaded/Boxed
PC Games
$33.9Bn
+6.7% YoY

Console Games
$45.2Bn
+6.8% YoY

2%

21%

28%

2020 Total
$159.3Bn
+9.3%
YoY

Tablet Games
$13.7Bn
+2.7% YoY

9%

40%

Smartphone Games
$63.6Bn
+15.8% YoY

NewZoo

In 2020 alone, the gaming industry has yielded a total revenue of 159.3 billion -up 9.3% from 2019.

Global Player Forecast 2015-2023

Year	Players
2015	1.99Bn
2016	2.11Bn
2017	2.26Bn
2018	2.42Bn
2019	2.55Bn
2020	2.69Bn
2021	2.81Bn
2022	2.95Bn
2023	3.07Bn

Global Game Revenue Forecast Toward 2023

Year	Revenue
2018	$138.5Bn
2019	$145.7Bn
2020	$159.3Bn
2023	$200.8Bn

NewZoo

It has also accounted for 2.7 billion players globally -up 5.3% from 2019. In fact, it is forecasted that the gaming market revenue will grow to $200.8 billion and the global player count will rise to 3.07 billion by 2023. [2]

FunPlus Phoenix, winners of the 2019 League of Legends World Championship, took home $2.5 million in prize money (Joao Ferreira/ESPAT Media)

Man waits in line 30 hours at GameStop for a PS5

Among Us Available On Nintendo Switch With Cross-Play

EA To Buy Codemasters For $1.2 Billion

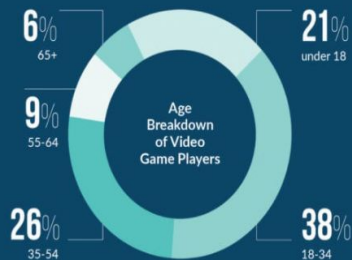**A Space Jam Game Is Coming To Xbox, And Fans Are Designing It**

PS5, Switch Sales Drive Sony And Nintendo Stock To Highest Point In Over A Decade

Who, Why, and How

Questions to Answer

What Makes a video game successful ?

- What are the most successful gaming platforms ?
- How does Metacritic's ratings correlate to video game sales ?

What is the best time to release a video game ?

- Do video games sell better in a certain time of the year ?
- How do the holidays affect video game sales ?
- How does world events affect video games ?

How does demographic / location effect video game sales?

- What part of the world plays / purchases the most games ?
- What demographic plays / purchases the most video games ?
- What genres are popular by regions?

Description of Data Source

- Data was imported from a Kaggle video game sales data set (vgsales.csv) [3] that had sales by region per game per console.

- We wrote scraping algorithms to gather further feature data, particularly metacritic and user score data and joined it with the original kaggle data.

- Metacritic.com is a leading review site for video games, as well as TV shows and movies.

- Data was then cleaned and uploaded to PostgreSQL using Sqlalchemy for storage.

ERD

game_title_genre
game_title	🔑	varchar
game_genre		varchar

game_platform_ratings
game_title	🔑	varchar
game_platform	🔑	varchar
user_score		float
user_positve		float
user_neutral		float
user_negative		float
critic_metacritic		float
critic_positive		float
critic_neutral		float
critic_negative		float

game_platform_release
game_title	🔑	varchar
game_platform	🔑	varchar
game_publisher		varchar
game_developer		varchar
maturity_rating		varchar
player_number		varchar
release_date		varchar
release_year		float
release_month		float
is_handheld		varchar
is_retro		varchar
is_depriciated		varchar

game_platform_sales
game_title	🔑	varchar
game_platform	🔑	varchar
sales_NA		float
sales_EU		float
sales_JP		float
sales_other		float
sales_global		float

game_platform_reviews
review_ID	🔑	int
game_title		varchar
game_platform		varchar
critic_name		varchar
critic_score		float
critic_review_summary		varchar

Description of Data Exploration Phase

1. Feature engineering was performed

   a. Random Forest was used to interpolate missing values in the dataset.

   b. Data was encoded for KMeans algorithm using label encoder.

   c. Data was clustered using KMeans to find significant groupings of data that could help us perform feature selection for the deep learning model.

2. A deep artificial neural network was created to predict Sales

   a. The developer and publisher columns had too many unique values to be used in a deep learning model even after binning so it was dropped. Binned multiplayer categories.

   b. Performed feature engineering on the release data column to get the month and year column.

   c. Created columns for console games vs handheld games and age of the game.

Description of Analysis Phase

Machine Learning (Neural Network)

- OOP (Object-Oriented Programming) was implemented in order to make easier to test different subsets of the data on the model to determine what works best without filling the notebook with repetitive blocks of code.

- After learning that the data we had was unfit to to predict sales, we chose to scale back and only predict sales for a certain region such is North America rather than globally which then came out with far better results.

Data Visualization (Tableau)

- In order to analyze and understand trends in the data, we used Tableau which is a interactive data visualization software.
- Using Tableau we were able to create different visualizations for video game analysis.
- The visualizations are primarily focused sales related to platform and metacritic scores on a regional and global scale.

Tools and Programs Used

- Tableau
- Jupyter Notebook
- Google Slides
- Pandas
- Matplotlib
- Natural Language Processing
- Spark
- Random Forest Regression
- Linear Regression models.

Future Recommendations + Group Assessment

- So to start off , we as a group realize that our machine learning model was not the most accurate nor the best way to populate the information we needed. While we were able to populate data and results from our algorithm. We realize it is flawed and we realize why.
- The reason our model did not work out quite the way we wanted it to is because of the fact that the original data we had was so skewed right meaning that the games that made alot of money tended to be outliers and the games that made around the same amount had the bulk of the data which meant that we were only able to accurately predict sales for the majority of the games rather than the minority in this case the most successful games based on sales.
- Next time around I believe that we would get not only a larger data set but a more balanced one so that the data is not majorly skewed which would allow for more accurate results

References

1. Chikhani, Riad. "The History Of Gaming: An Evolving Community." TechCrunch, TechCrunch, 31 Oct. 2015, techcrunch.com/2015/10/31/the-history-of-gaming-an-evolving-community/.
2. Tim Wijman. "2020 Global Games Market Report". NewZoo. Accessed 05 December 2020
3. Smith, Gregory. "Video Game Sales." *Kaggle*, 26 Oct. 2016, www.kaggle.com/gregorut/videogamesales/version/2?select=vgsales.csv.