DATA MANAGEMENT PROJECT REPORT

ON

VIDEO GAME SALES AND RATINGS

Submitted By:

Shubham Madhwal

11716826

KM065

Under the guidance of:

Maneet Kaur – 15709



School of Computer Science and Engineering Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Shubham Madhwal** bearing Registration number **11716826** has completed **Data Management (INT217)** project titled, "**Video Game Sales and Ratings**" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor School of Computer Science and Engineering Lovely Professional University

Date: 13th November, 2019

DECLARATION

I, **Shubham Madhwal**, student of **Computer Science and Engineering** under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 13th November, 2019 Signature

Registration No. 11716826 Shubham Madhwal

ACKNOWLEDGEMENT

I take this opportunity to present our votes of thanks to all those guideposts who really acted as lightening pillars to enlighten my way throughout this Project that has led to successful and satisfactory completion of this Project. I am grateful to **Lovely Professional University** for providing us with an opportunity to undertake this Project and providing us with all the facilities. I am highly thankful to All for their active support, valuable time and advice, whole-hearted guidance, sincere cooperation and painstaking involvement during the project and in completing the assignment of preparing the said project within the time stipulated. Lastly, I am thankful to all those, particularly the various friends, who have been instrumental in creating proper, healthy and conductive environment and including new and fresh innovative ideas for me during the project, without their help, it would have been extremely difficult for me to complete the project in a time bound framework.

INDEX

- 1. Introduction
 - a. About Data Analysis
 - b. Background of the project
 - c. Data Defintion
- 2. Scope of Analysis
 - a. Objectives of the project
 - b. Scope
- 3. ETL Process
 - a. What is ETL?
 - b. ETL on the dataset
- 4. Analysis of Dataset
- 5. Dashboard
- 6. Gantt Chart
- 7. Bibliography

INTRODUCTION

If your business is not growing, then you must look back and acknowledge your mistakes and plan again without repeating those mistakes. And even if your business is growing, then you must look forward to making the business to grow more. All you need to do is analyze your business data and business processes.

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. Whenever we take any decision in our day-to-day life is by thinking about what happened last time or what will happen by choosing that decision. This is nothing but analyzing our past or future and making decisions based on it. For that, we gather memories of our past or dreams of our future. So that is nothing but data analysis. Now same thing analyst does for business purposes, is called Data Analysis.

Video game is interactive digital entertainment that you "play" via a computer, a game console (like the Xbox or PlayStation) or a phone or tablet. There, you can go home now because class is over, and we're done. That really does sum it all up — but it's not a very good explanation of tom clancy's the division everything going on here.

As the new person covering video games for TheWrap, I aim to provide good explanations for you, our readers. So, I'm going to dive into the more complicated explanation here.

Very much from the occurrence of <u>colossal computers</u> that filled rooms in the early 1970s to the Ultrabook's super thin laptops we have nowadays, video games have existed to fill our needs for entertainment and maybe even learning. Video games have increasingly become sophisticated as newly operating software were produced and better-performing hardware were invented. Indeed, the digital information boom at the end of the 20th century engendered a series of ultrafast developments that led from the creation of multi-pixel 8-bit video games such as Pacman, to the open world non-linear games such as <u>Grand Theft Auto</u>, which take on several gigabytes on the computer's hardware storage capacity.

With the sophistication and proliferation of games, people have engendered more complex and mixed reviews about them. Computer games were originally for entertainment for those very few who could afford computers only. As software became cheaper to manufacture, the word PC("Personal Computers") emerged, and families were already buying PCs and software (including video games) in numbers.

Before we go deep into the societal impact video games have for the generations around this time, a choice must be made of whether video games are mediums or not. A simple look-up in the dictionary tells us that mediums are "an agency or means of doing something." Ian Bogost, in his book *How To Do things with Videogames*, claimed that: "games are models of experiences...we operate these models...our actions [are] constrained by their rules...we take on a role in a videogame, putting ourselves in the shoes of someone else..." (Bogost 04) Simply said, video games are a means for people to immerse themselves in information models to assume a role in a certain environment. Therefore, according to Bogost, (and I would strongly agree) video games are a medium.

It is not untrue that video games caused quite a dilemma for families in the 80s and 90s. In fact, many families reported that their children were virtually addicted to video games and did not put enough attention on the family. The problem persists till today as a main family and societal issue. This is also why "All-too-familiar questions arise about whether games promote violent action or whether they make us fat through inactivity." (Bogost 05) In his bestseller, however, Bogost talks about how parents and people alike have simply misjudged video games as a dichotomous choice of good or bad, which he dubbed as the "media ecological approach", rather than seeing games as a medium which is able to influence culture in numerous ways (microecology). I generally agree with Bogost's idea. Games act as a medium by impacting people's daily lives continuously, both in communication and perception. I will explore this along with examples in the next three paragraphs. YouTube, a large video-sharing website as you might know it, has a very large gamer community. And many game commentators post game walkthroughs and reviews for the large audience on YouTube for a living. In fact, according to YouTube statistics, gaming commentators and reviewers alike will upload up to 75 gigabytes of video data to the website every ten seconds. Each gaming video might have more than a million comments (many of which the commentators rely to) and there are plenty of private discussion and public Q&A sessions. From this perspective, I believe that these videos undoubtedly have a large impact on the lives of millions of people who are watching the videos on a daily basis in terms of communication. Again, the gaming content of those videos are irrelevant compared to the impact the videos have on collective communication in gaming communities, as Bogost would have it: "The things a medium does to a culture are more important than the content it conveys." (Bogost 04)

On the other hand, video games can alter our perceptions dramatically. How our perceptions are changed depend on the type of the video games and our perceived cosmopolitan view of the world. When engaging in video games, we are both acting out the role of the protagonist according to our general perception of the world while simultaneously abiding by the rules of the "model" (the gaming environment) that were created by the game developers.

A good example would be the <u>Portal series created by Valve Corporation</u>. The protagonist in the game is a test subject who has to navigate across numerous test chambers with her portal gun, which can create interdimensional space. Her goal is to flee the "unethical" testing facility but is constantly stalled by the facility's main AI computer, <u>GLaDOS</u>. Each test chamber is unique, and there are several ways to finish a level. It all comes down to how the video game player perceives the level. Also, there are many moral decisions to make in the game, further altering the gamers perceptions about certain aspects. In one level, for instance, after using the Companion Cube extensively, the player must make a choice of whether incinerating the cube and pass the level or get stuck in the level with the cube. And it again depends on how the player perceives the game. In fact, many players on the Steam Community Hub reported feeling extremely emotional at that moment.

To acquire a more comprehensive view about video games in our society, we must think more critically about them, not just dismiss them as superficial objects that someone might get addicted to. Parents and families, along with other people who are in presence around video games, need to regard games as a medium which has multifaceted uses rather than only one or two. That said, games are currently used not only in entertainment, but also in medicine, psychiatry training sessions, tools for soldiers to simulate real combat and even placebo means in hospitals etc..In terms of communication and perception, as aforementioned, video games acts as an indispensable means to a medium by encouraging all sorts of discussions and perceptual alterations. The various uses of games cannot be overstated, and most of them have profound impacts in different sections of our society.

Finally, I believe that it is vital for people to study gaming behaviour. There are myriads of reasons for doing so. Social-cognitive psychologists could research brain pattern behaviour when people are playing games. I personally have always wondered why people's body do physical movements when they are actually playing games set in virtual reality. Furthermore, researching possible changes in perception of thought would be a great basis for developing our understanding of human behavior. The bottom line is, as games become more and more a

part of people's daily lives, the necessity to study them extensively is ultimately of extreme importance for the comprehension of human physical and psychological behavior to our community.

The database keeps track of the following data fields:

Alongside the fields: Name, Platform, Year_of_Release, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, India_Sales, Global_Sales, we have:-

- Critic_score Aggregate score compiled by Metacritic staff
- Critic_count The number of critics used in coming up with the Critic_score
- User score Score by Metacritic's subscribers
- User_count Number of users who gave the user_score
- Developer Party responsible for creating the game
- Rating The ESRB ratings

SCOPE OF ANALYSIS

Recent years have seen important advances in the study of video games, with many scholars examining how video games impact emotional experiences of players. While this research borrows heavily from traditional media effects paradigms, it has proceeded without effectively integrating developments in emotion and emotion regulation. Therefore, this paper first reviews the nature of emotion and emotion regulation, and then reviews the extant video game literature. We next draw on theoretical accounts of emotion to frame three broad research questions that may advance future work on video games and emotions: (1) What is the nature of video game-evoked emotional experiences? (2) What are the processes through which video games produce emotions? (3) How are video games used to regulate emotions?

Game development project managers and product owners, such as directors, producers, and studio heads, rely on experience-based tacit knowledge to decide how much content to create for players. However, they could be operating on a misunderstanding of the way their players consume game content and how much game content is desirable to players.

This research presents the results of data mining video game "achievement" data to discover trends in game completion rates and correlations to factors outside of the length of the game itself. Through tagging game achievements that signal a player has "finished" the primary single-player content, this research discovered that, for most games, few players will consume the provided content, although "higher quality" content does maintain a higher player retention rate. With an understanding of how players consume game content, project managers and product owners can make more informed decisions on project scope, which could reduce game budgets and improve production efficiency.

Data analysis in gaming companies has gone from being a half-remembered aside just a decade ago, to being the driving force behind some of the world's biggest studios. Every major studio and many smaller ones employ at least one, if not (ideally) a team of data scientists to help it extract information from its games, to analyse and quantify this data and to explain it in a way that can help the studio improve - be that through bringing in more money, keeping people playing longer, getting more downloads or purchases in the first place, or whatever else.

Basically analytics have become as important a factor to the success of a gaming company as just about any other more 'traditional' aspect. But we're still in relatively early days for this discipline - what does the future of the analytical approach hold for game developers, publishers and even players? And, really, what's the point?

The U.S. is projected to be one of the most prominent digital games market in the world.

Digital games have always been a favorite choice of entertainment for the people of the region. In recent years, there has been a transformation in the distribution channel of the games in the country. Research shows that game sales via digital means are dominating the U.S. games market while boxed games are slowly becoming the second choice.

The U.S. presents a robust market with preferences for all gaming platforms - PC, mobile, and console. Though gaming on a console is most popular in the country, the time spent on mobile gaming has also increased. Mobile gaming is gaining traction as some major companies in the nation like Take-Two Interactive, Activision Blizzard, etc., who were previously only into PC/Online games, are entering into the mobile gaming segment.

Overall, the total consumer spending on video gaming has increased rapidly in the last few years at ~16.48% CAGR (2014-16) in the U.S. Also, the high growth of interest in eSports and virtual-augmented reality, clubbed with the increasing prevalence of digitally downloadable games are expected to drive the Digital Games market forward in the country.

By platforms, the market is segmented into PC games, mobile games, and console games. Among these segments, console games is the most significant revenue generator while mobile games have exhibited the fastest growth rate.

The main goal of the project is to find out in depth about the Video Gaming Industry and to answer the following objectives:-

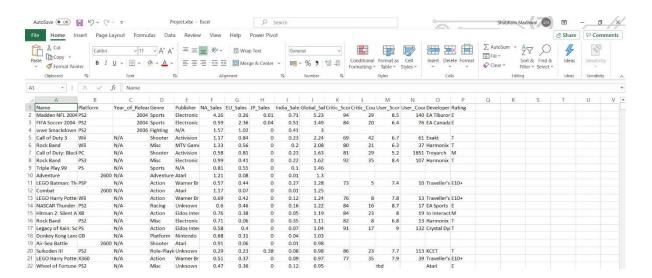
- Highest Grossing Games
- Top 10 Publishers Market Acquisition
- Critically Highest Rated Game
- Highly Rated User Game
- Yearly Game sales 2000-2018
- Most Preferred Genre Critic
- Most Preferred Genre User
- Comparison Indian Sales Vs Global
- Market influenced by different consoles
- Developers with Highest Profits

Aim of this project is to answer the above objectives in the form of visualization by creating a dashboard to convey the answers effectively.

ETL PROCESS

ETL is defined as a process that extracts the data from different RDBMS source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system. ETL full-form is Extract, Transform and Load.

Before ETL, the dataset looked like this. This data was taken from Kaggle.



Through the process of ETL, we are going to clean the dataset and bring all the entities to their proper data format.

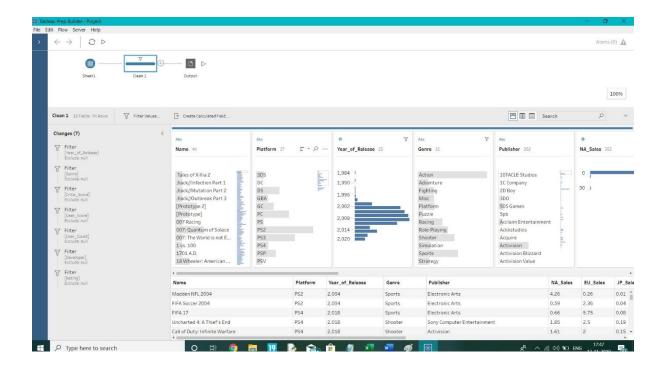
Step 1: Remove the blank cells from the dataset.

Open the file in tableau prep builder and select blank tab in any attribute or column name that has this particular keyword select that and choose the exclude option.

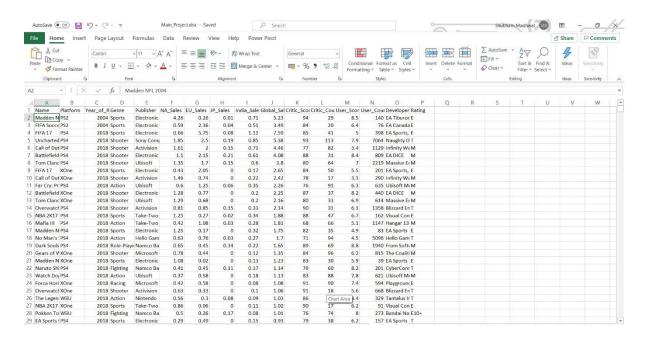
Step 2: Remove columns which are not properly defined or not crucial to our analysis.

Open the file in tableau prep builder and select the not defined tab in any attribute or column name that has this particular keyword select that and choose the exclude option.

Step 3: Removing the Na values from the Shipping cost columns to zero.



After cleaning the data and removing all the anomalies the data was ready, and now all the analysis can be performed with accuracy.



ANALYSIS OF DATASET

1. Highest Grossing Games

In this analysis I tried to showcase the top grossing games across all the platforms. Being it PC, PS3, PS2,Wii etc. The total number of games are sorted in top 10 format and the sum of global sales is in millions.

ANALYSIS		
Row Labels	Sum of Global_Sales	
Call of Duty: Black Ops	29.19	
Call of Duty: Black Ops II	29.4	
Call of Duty: Ghosts	26.7	
Call of Duty: Modern Warfare 3	30.59	
Grand Theft Auto V	56.57	
Mario Kart Wii	35.52	
New Super Mario Bros.	29.8	
New Super Mario Bros. Wii	28.32	
Wii Sports	82.53	
Wii Sports Resort	32.77	
Grand Total	381.39	



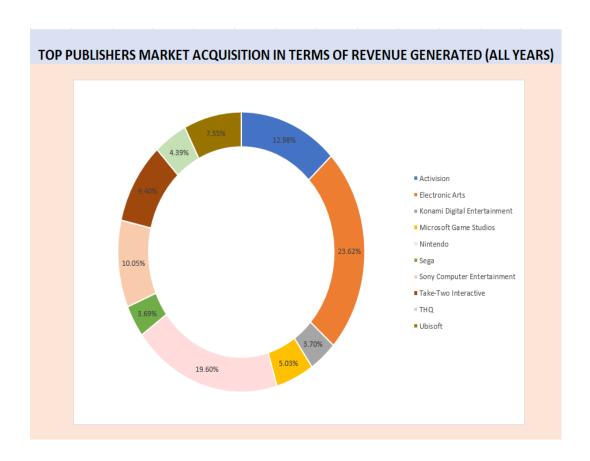
The highest grossing game across all platform is Wii Sports, followed by GTA V and Mario Kart Wii at the third spot

2. Top 10 Publishers Market Acquisition

In this analysis I tried to showcase the top Publishers of games. Being it PC, PS3, PS2, Wii etc. The total number of publishers are sorted in top 10 format according to the revenue generated by them in all the fiscal years included and the sum of global sales is in millions and converted it into percentage to show the market hold of that particular organization.

For example in the below analysis Electronic arts has a market acquisition of a solid 23.62% which makes it roughly 24% where as the leas market acquisition in the top 10 is of Konami of 3.70% roughly 4%

ANALYSIS			
Row Labels	Sum of Global_Sales		
Activision	12.98%		
Electronic Arts	23.62%		
Konami Digital Entertainment	3.70%		
Microsoft Game Studios	5.03%		
Nintendo	19.60%		
Sega	3.69%		
Sony Computer Entertainment	10.05%		
Take-Two Interactive	9.40%		
THQ	4.39%		
Ubisoft	7.55%		
Grand Total	100.00%		



	CONCLUSION			
Electronic arts is the highest money generating publisher,				
whereas Konami is the least amongst the top 10.				

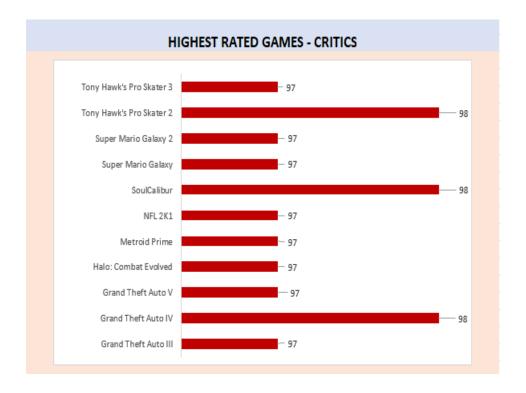
3. Highest Rated Games – Critic

In this analysis I tried to showcase the top highest rated games according to the critic across all platforms being it PC, PS3, PS2,Wii etc. Int the entire time stand that is from 1985-2018.

There is a tough competition between the top spot as there are three games with an equal rating out of 100 provided by the critic. The games are:

Grand Theft Auto IV, Soul Calibur, Tony Hawk's Pro Skater 2

ANALYSIS			
Row Labels	Max of Critic_Score		
Grand Theft Auto III	97		
Grand Theft Auto IV	98		
Grand Theft Auto V	97		
Halo: Combat Evolved	97		
Metroid Prime	97		
NFL 2K1	97		
SoulCalibur	98		
Super Mario Galaxy	97		
Super Mario Galaxy 2	97		
Tony Hawk's Pro Skater 2	98		
Tony Hawk's Pro Skater 3	97		
Grand Total	98		



Tony Hawk's Pro Skater 2, Soul Calibur, Grand Theft Auto IV is the highest critics rated games

4. Highest Rated Game – User

In this analysis I tried to showcase the top highest rated games according to the user across all platforms being it PC, PS3, PS2,Wii etc. Int the entire time stand that is from 1985-2018.

There is a tough competition between the top spot as there are three games with an equal rating out of 10 provided by the users. The games are:

Lego Star Wars II: The Original Trilogy, Madden NFL 07, Need For Speed Most Wanted.

ANALYSIS			
Row Labels ,T	Count of User_Count		
Harry Potter and the Order of the Phoenix	7		
LEGO Star Wars II: The Original Trilogy	8		
Madden NFL 07	8		
Madden NFL 08	7		
Need for Speed Carbon	7		
Need for Speed: Most Wanted	8		
Need for Speed: ProStreet	7		
Spider-Man 3	7		
Terraria	7		
The Sims 2	7		
Tomb Raider: Legend	7		
Grand Total	80		



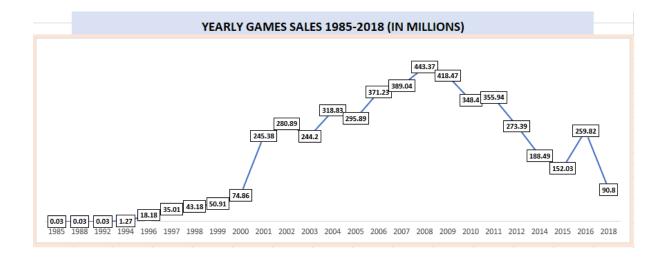
NFS Most Wanted, Madden NFL 08 and Madden NFL 07 is the highest rated games

5. Yearly Games Sales 1985-2015

In this analysis I tried to showcase the sales games across all the platforms. Being it PC, PS3, PS2,Wii etc. The years are filtered from the 1985-2018.

The Video Games Industry is currently a confusing market when it comes to the overall revenue generated by this market, because people now are preferring to go towards free games where as consoles games demands a, development of gaming in mobile phones and removal of CD section from various PC has led to a stigma. That's the reason for the sudden decline in sales

ANALYSIS		
Row Labels 🔻 🤄	Sum of Global_Sales	
1985	0.03	
1988	0.03	
1992	0.03	
1994	1.27	
1996	18.18	
1997	35.01	
1998	43.18	
1999	50.91	
2000	74.86	
2001	245.38	
2002	280.89	
2003	244.2	
2004	318.83	
2005	295.89	
2006	371.23	
2007	389.04	
2008	443.37	
2009	418.47	
2010	348.4	
2011	355.94	
2012	273.39	
2014	188.49	
2015	152.03	
2016	259.82	
2018	90.8	
Grand Total	4899.67	



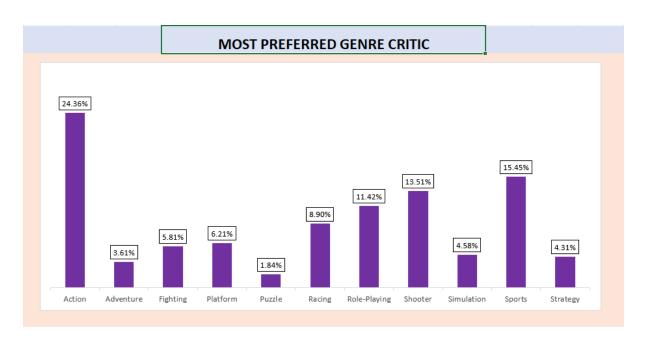
The Video Games Industry is currenlty a confusing market when it comes to the overall revenue generated by this market, because people now are prefering to go towards free games where as consoles games demands a, development of gaming in mobile phones and removal of CD section from various PC has led to a stigma. That's the reason for the sudden decline in sales

6. Most Preferred Genre Critic

In this analysis I tried to showcase the genre that is the critics favorite across all platforms being it PC, PS3, PS2,Wii etc. Int the entire time stand that is from 1985-2018.

According to the critics Action is preferred by them with a percentage of 23.36% and the least favorite category according to the critic is of Puzzle with 1.84%.

ANALYSIS		
Row Labels	Sum of Critic_Score	
Action	24.36%	
Adventure	3.61%	
Fighting	5.81%	
Platform	6.21%	
Puzzle	1.84%	
Racing	8.90%	
Role-Playing	11.42%	
Shooter	13.51%	
Simulation	4.58%	
Sports	15.45%	
Strategy	4.31%	
Grand Total	100.00%	



Action is the Genre preferred by the critic, whereas Puzzle is the least preferred

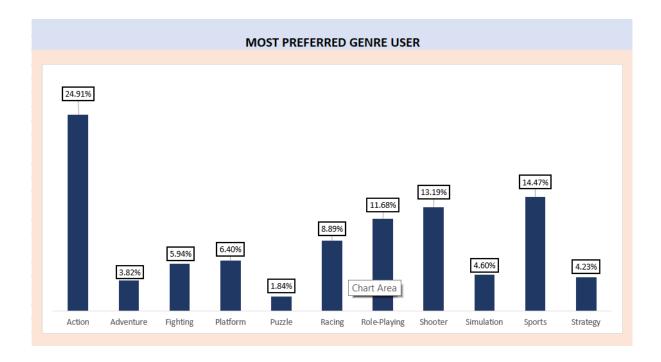
7. Most Preferred Genre User

In this analysis I tried to showcase the genre that is the Users favorite across all platforms being it PC, PS3, PS2, Wii etc. Int the entire time stand that is from 1985-2018.

According to the Users Action is preferred by them with a percentage of 24.91% and the least favorite category according to the critic is of Puzzle 1.84%.

There are a huge amount of similarities between the likings of genre between critics and users

ANALYSIS		
Row Labels	Sum of User_Score	
Action	24.91%	
Adventure	3.82%	
Fighting	5.94%	
Platform	6.40%	
Puzzle	1.84%	
Racing	8.89%	
Role-Playing	11.68%	
Shooter	13.19%	
Simulation	4.60%	
Sports	14.47%	
Strategy	4.23%	
Grand Total	100.00%	



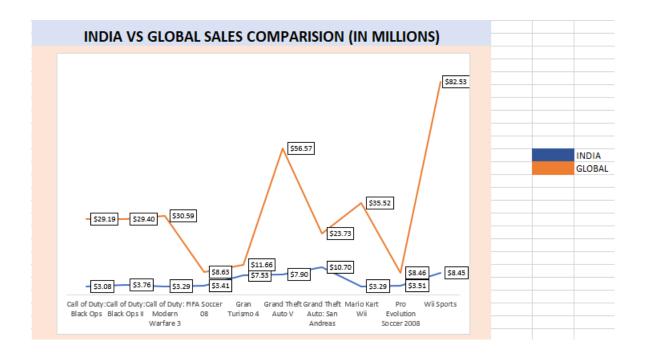
There is a similarity when it comes to the first and ast position of the genre selection the user and critic vhereas we see a slight diffrence in the other choice:

8. Comparison Indian Sales Vs Global

In this analysis I tried to showcase the top grossing games across all the platforms. Being it PC, PS3, PS2, Wii etc. The total number of games are sorted in top 10 format and the sum of global sales is in millions. And compared it with the Indian Sales so as to find out the difference in likings of the Indian culture and that of the other foreign countries.

In India GTA San Andreas is the highest grossing game of all time whereas the WII sports is the highest grossing game globally.

ANALYSIS	1	
Row Labels	Sum of India_Sales	Sum of Global_Sales
Call of Duty: Black Ops	3.08	29.19
Call of Duty: Black Ops II	3.76	29.4
Call of Duty: Modern Warfare 3	3.29	30.59
FIFA Soccer 08	3.41	8.63
Gran Turismo 4	7.53	11.66
Grand Theft Auto V	7.9	56.57
Grand Theft Auto: San Andreas	10.7	23.73
Mario Kart Wii	3.29	35.52
Pro Evolution Soccer 2008	3.51	8.46
Wii Sports	8.45	82.53
Grand Total	54.92	316.28

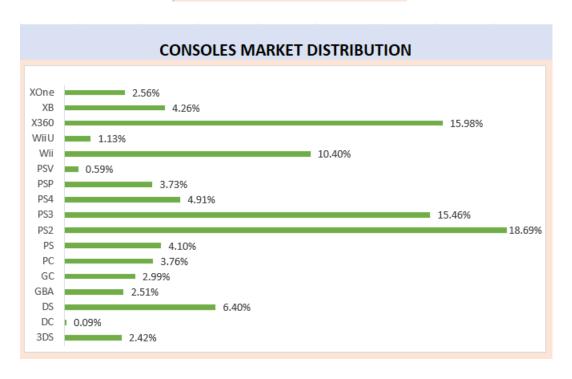


GTA San Andreas is the highest grossing game in India where Wii Sports Globally, Whereas COD: Black Ops has a weak hold in India while rest of countries has a bigger liking for it Hence, in order to gain the maximum profits The publishers of COD Black Ops should target the Indian market for a better revenue generation

9. Consoles Market Distribution

In this analysis I tried to showcase the consoles market distribution and how each consoles has contributed towards the sales of games in between the time period of 1985-2018 in terms of percentage of the total global sales.

ANALYSIS		
Row Labels 🔻	Sum of Global_Sales	
3DS	2.42%	
DC	0.09%	
DS	6.40%	
GBA	2.51%	
GC	2.99%	
PC	3.76%	
PS	4.10%	
PS2	18.69%	
PS3	15.46%	
PS4	4.91%	
PSP	3.73%	
PSV	0.59%	
Wii	10.40%	
WiiU	1.13%	
X360	15.98%	
XB	4.26%	
XOne	2.56%	
Grand Total	100.00%	

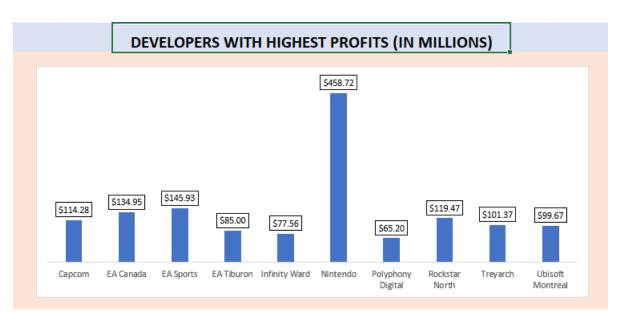


Its of no thinking that PS2 is the leading console as it's the revolutionory console that took the gaming industry by storm

10. Developers with Highest Profits

In this analysis I tried to showcase the top developers of games with the games across across all the platforms. Being it PC, PS3, PS2,Wii etc. The total number of games are sorted in top 10 format and the sum of global sales is in millions.

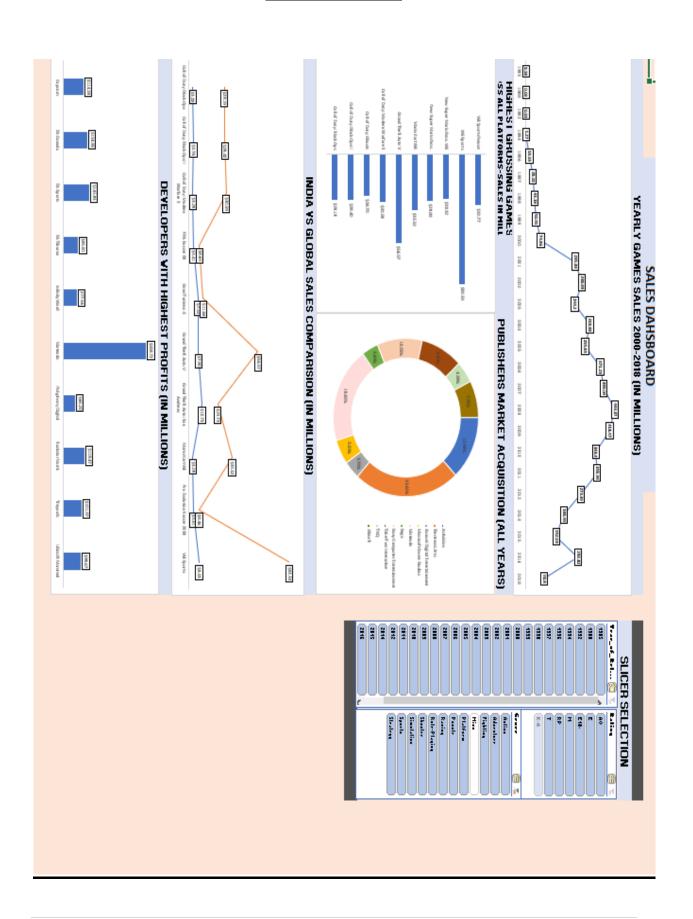
ANALYSIS		
Row Labels 🍑	Sum of Global_Sales	
Capcom	114.28	
EA Canada	134.95	
EA Sports	145.93	
EA Tiburon	85	
Infinity Ward	77.56	
Nintendo	458.72	
Polyphony Digital	65.2	
Rockstar North	119.47	
Treyarch	101.37	
Ubisoft Montreal	99.67	
Grand Total	1402.15	

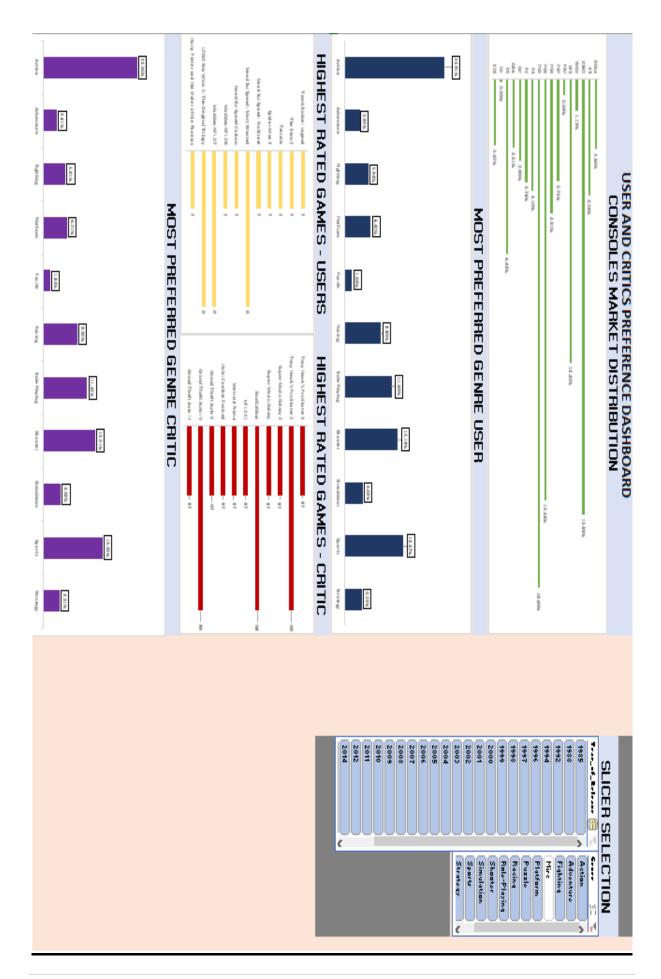


CONCLUSION

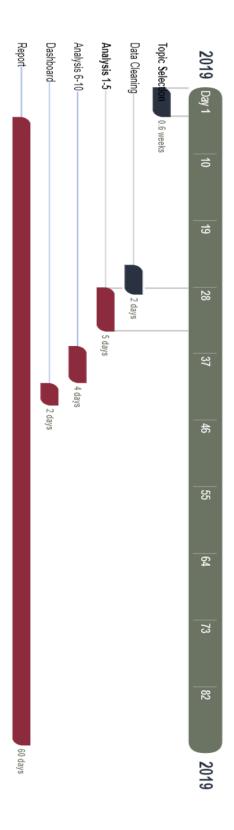
Nintendo is the highest earning games developing company. With Infinty Ward at the last amongs the top 10.

DASHBOARD





GANTT CHART



VIDEO GAMES SALES AND RATINGS

REFERENCES AND BIBLIOGRAPHY

- Wikipedia
- Coursera
- Kaggle
- Reddit