Video Game Sales

- ABDULRHMAN ALQURASHI
- KHALID ALHARBI

OUTLINE:

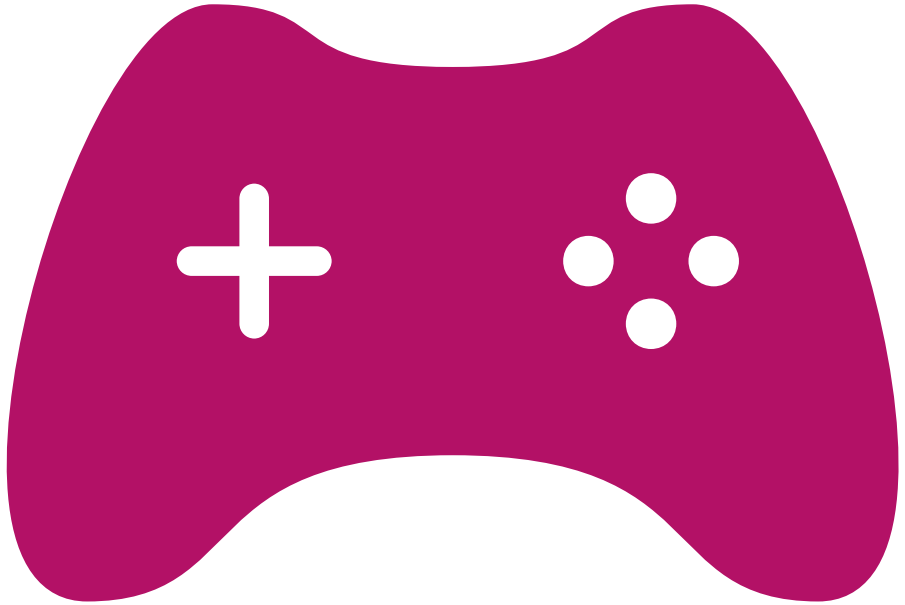
- **INTRODUCTION**
- **DESIGN**
- **AIM AND OBJECTIVES**
- **PROCESSOR AND STEPS**
- **CONCLUSION**

Introduction

Video Game Industry:

- A video game is an electronic game that can be played on a computing device, such as a personal computer, gaming console, or mobile phone. Depending on the platform, video games can be subcategorized into computer games and console games

- Video games have come a long way since the first games emerged in the 1970s.



Design

We have a company working in the field of video games and we want to create a marketing campaign for the company's brand

By organizing an online tournament for video game fans

We need to determine which platform, genre and game name to by organizing an online tournament

Aim and Objectives

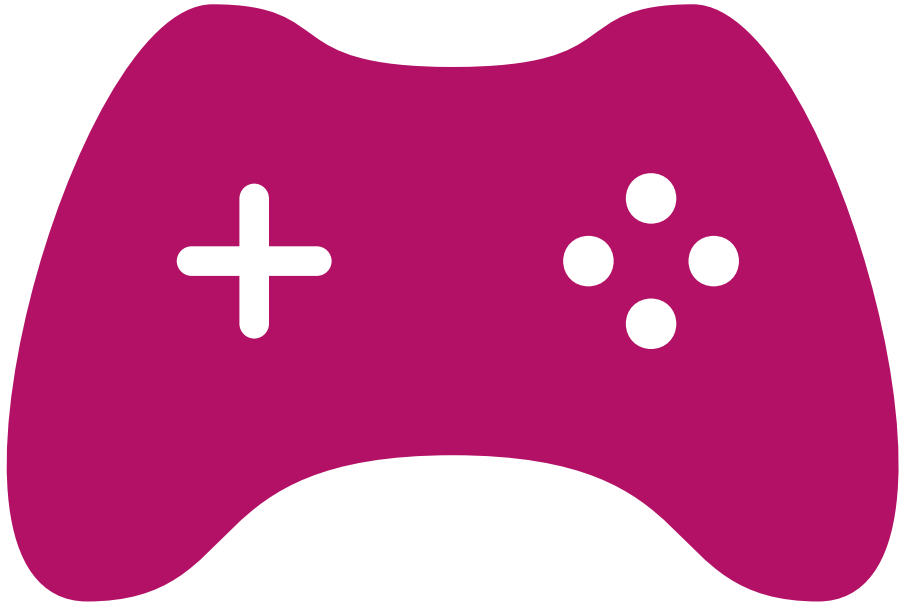
What is the video game Sales by year?

What is the video game Sales by Platform?

What is the video game Sales by Genre?

What is the most popular game?

Through this answer, we will determine the genre of online tournament



Processor and Steps

- Load data
- Reading & Understanding the data
- Data Cleaning
- EDA

Step1: Load data

About Dataset:

- The data set is provided in .csv format contains information of Motivated by Gregory Smith's web scrape of VGChartz Video Games Sales
- The data set was extracted from Kaggle
- Number of rows = 16598 rows
- Number of columns = 11 columns

Step 2: Reading & Understanding the data

- Dataset Dimension

```
DataSet dimension     : (16598, 11)
DataSet size          : 182578
```

- Checking Numerical Columns Statistics

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16598.000000	16327.000000	16598.000000	16598.000000	16598.000000	16598.000000	16598.000000
mean	8300.605254	2006.406443	0.264667	0.146652	0.077782	0.048063	0.537441
std	4791.853933	5.828981	0.816683	0.505351	0.309291	0.188588	1.555028
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4151.250000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8300.500000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12449.750000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.470000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

Field Description

Field Name	Description
Rank	Ranking of overall sales
Name	Name of the game
Platform	Console on which the game is running
Year	Year of the game released
Genre	Game's category
Publisher	Publisher
NA_Sales	Game sales in North America (in millions of units)
EU_Sales	Game sales in the European Union (in millions of units)
JP_Sales	Game sales in Japan (in millions of units)
Other_Sales	Game sales in the rest of the world, i.e. Africa, Asia excluding Japan, Australia, Europe excluding the E.U
Global_Sales	Total sales in the world (in millions of units)

Step 3: Data Cleaning

Max and Min year

```
Mim Year Value:  1980.0
Max Year Value:  2020.0
```

Show Max Year:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
5957	5959	Imagine: Makeup Artist	DS	2020.0	Simulation	Ubisoft	0.27	0.0	0.0	0.02	0.29

Imagine: Makeup Artist game was launched on 16th April 2010. Thus, we will change 2020 to 2010

By code [.replace(2020.0, 2010.0)]

Info About The Column Types

Insight: We can see there are some null values in the column. Let's inspect the null values first

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Rank          16598 non-null  int64
 1   Name          16598 non-null  object
 2   Platform      16598 non-null  object
 3   Year          16327 non-null  float64
 4   Genre         16598 non-null  object
 5   Publisher     16540 non-null  object
 6   NA_Sales      16598 non-null  float64
 7   EU_Sales      16598 non-null  float64
 8   JP_Sales      16598 non-null  float64
 9   Other_Sales   16598 non-null  float64
 10  Global_Sales  16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

Total Null

	Column Name	Null Values	Null Values Percentage
0	Rank	0	0.000000
1	Name	0	0.000000
2	Platform	0	0.000000
3	Year	271	1.632727
4	Genre	0	0.000000
5	Publisher	58	0.349440
6	NA_Sales	0	0.000000
7	EU_Sales	0	0.000000
8	JP_Sales	0	0.000000
9	Other_Sales	0	0.000000
10	Global_Sales	0	0.000000

Drop All Null

```
Rank            0
Name            0
Platform        0
Year            0
Genre           0
Publisher       0
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

Insight:

We have successfully imputed or removed null values

Change years type

```
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Rank          16598 non-null  int64
 1   Name          16598 non-null  object
 2   Platform      16598 non-null  object
 3   Year          16327 non-null  float64
 4   Genre         16598 non-null  object
 5   Publisher     16540 non-null  object
 6   NA_Sales      16598 non-null  float64
 7   EU_Sales      16598 non-null  float64
 8   JP_Sales      16598 non-null  float64
 9   Other_Sales   16598 non-null  float64
 10  Global_Sales  16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

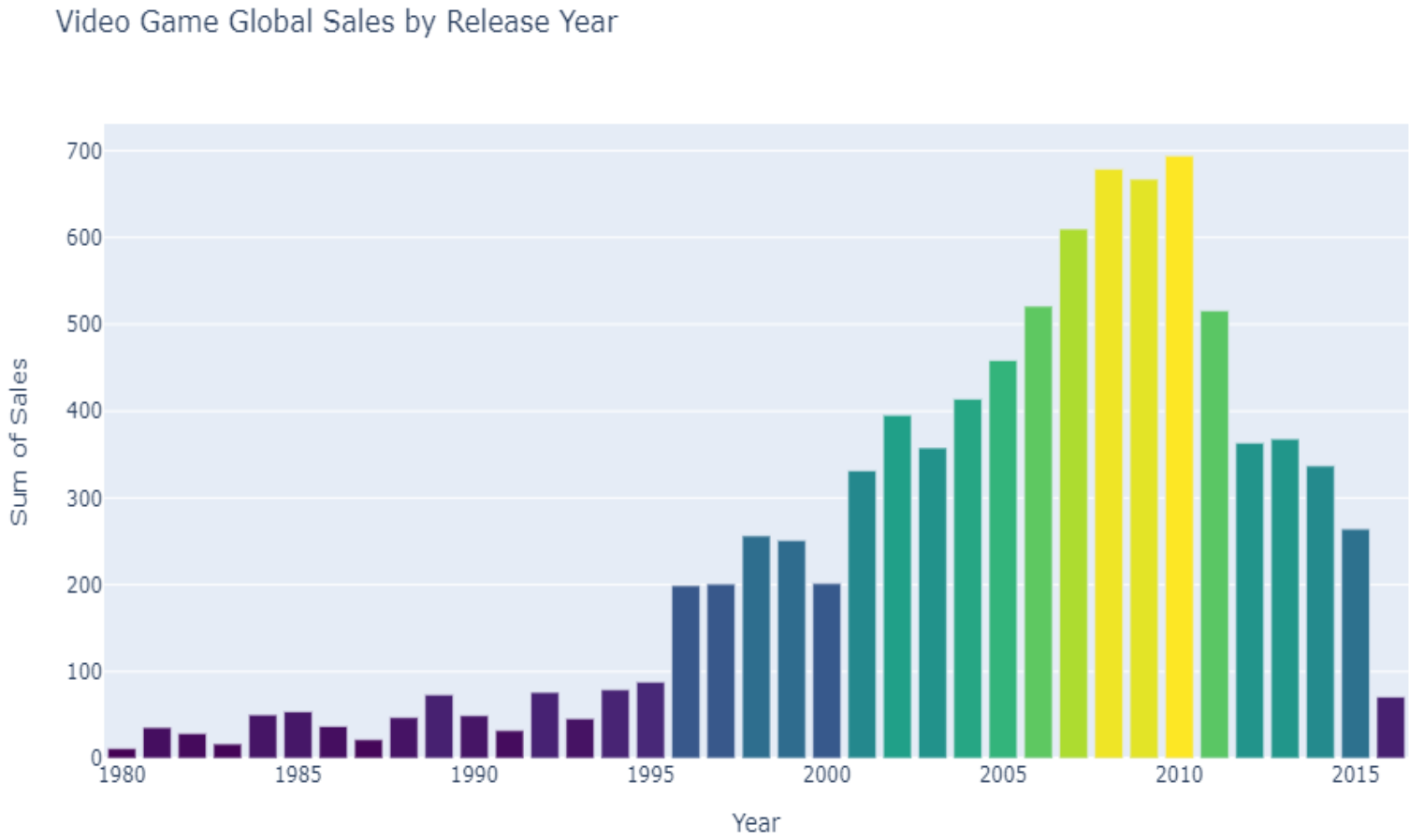
```
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Rank          16537 non-null  int64
 1   Name          16537 non-null  object
 2   Platform      16537 non-null  object
 3   Year          16537 non-null  int32
 4   Genre         16537 non-null  object
 5   Publisher     16537 non-null  object
 6   NA_Sales      16537 non-null  float64
 7   EU_Sales      16537 non-null  float64
 8   JP_Sales      16537 non-null  float64
 9   Other_Sales   16537 non-null  float64
 10  Global_Sales  16537 non-null  float64
dtypes: float64(5), int32(1), int64(1), object(4)
memory usage: 2.0+ MB
```

Nsight :

We have successfully imputed the Year column. Now let's change it to an integer column

Step4: EDA

The video game Sales by Years



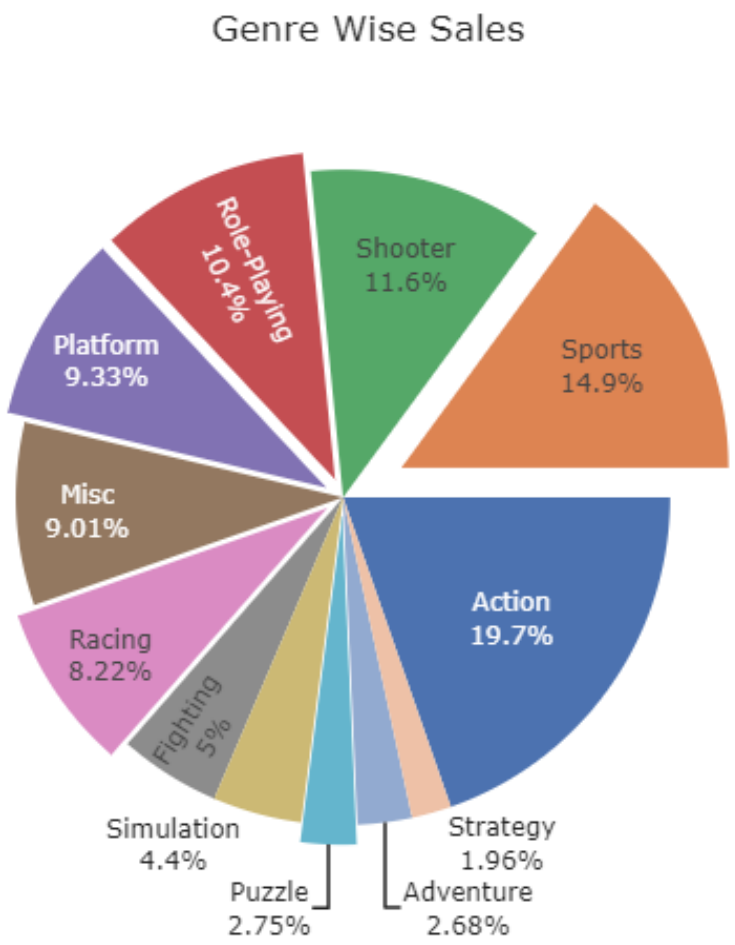
The video game Sales by Platform

Year Platform	2013	2014	2015	2016	total
total	546	580	614	342	2082
PS3	127	108	74	32	341
PSV	63	102	113	58	336
PS4	16	75	137	107	335
3DS	91	78	86	35	290
XOne	19	61	79	54	213
X360	75	64	40	8	187
PC	38	44	50	38	170
WiiU	42	31	28	10	111
PSP	54	10	3	0	67
Wii	12	6	4	0	22
DS	9	1	0	0	10

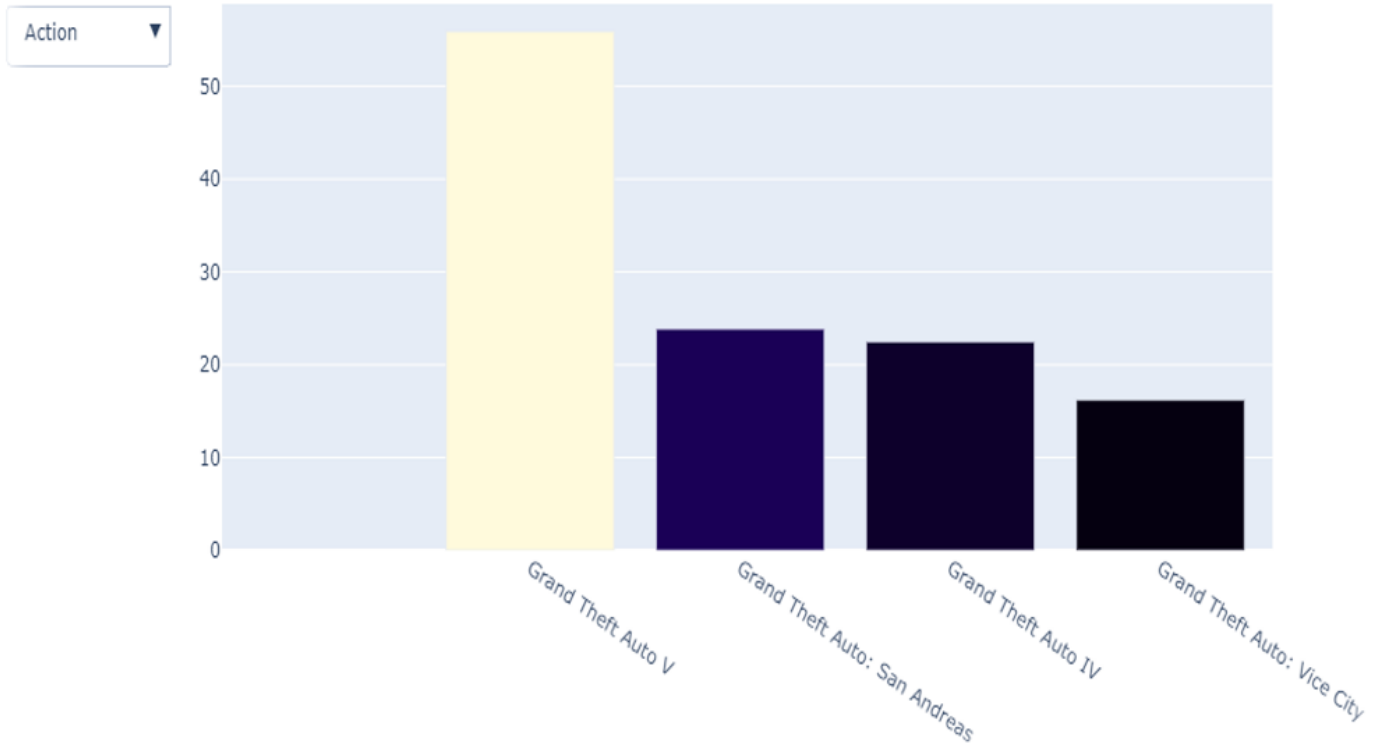
Here is a statistic about Platform Analysis of the number and sales of released games for 2013 to 2016

We see a decrease in the number of games created on the old generation consoles and an increase in production for new consoles: PS4, XBOne and WiiU.

The video game Sales by Genre



Popular Game



We see here the GTA 5 at the Top,
Based on that GTA 5 is a popular game

Conclusion

we decided after answering the questions that the target device for the tournament is PS4 and GTA 5 game.


thank you

Analysis of Video Game Sales (MVP)

The purpose of the analysis is to understand This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of vgchartz

We want to organize an online tournament for video game fans. Our first question:
* What is the Genre wise Game Sales? - Through this answer, we will determine the genre of online tournament

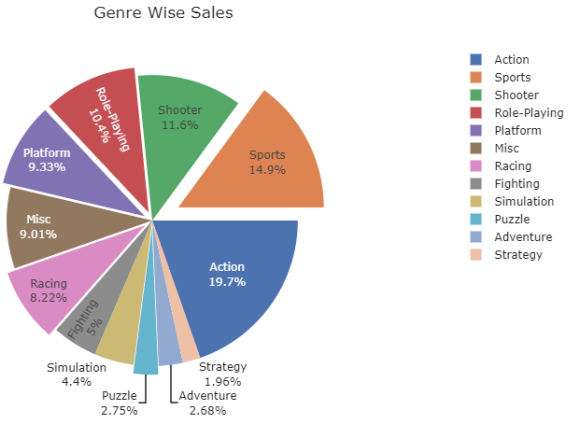


Figure 1: Genre Wise Sales

EDA_video_games

About Videos Games:

Globally, there are 2.3 billion players. They have an inexhaustible love for video games, much like the Universe!! One of the most addictive things gamers do is read about video games, particularly the history and interesting information about some of their favorite titles, gaming publishers, and game developers — it's just so fascinating how it all came together.

Motivation It would be interesting to see any machine learning techniques or continued data visualizations applied on this data set.

Aim and Objectives:

- What is the video game Sales by year?
- What is the video game Sales by Platform?
- What is the video game Sales by Genre?
- What is the most popular game?

Data Description:

The data set is provided in .csv format contains information of Motivated by Gregory Smith's web scrape of VGChartz Video Games Sales, this data set simply extends the number of variables with another web scrape from Metacritic. Unfortunately, there are missing observations as Metacritic only covers a subset of the platforms.

The data set was extracted from Kaggle

Number of rows = 16598 rows

Number of columns = 11 columns

Field Description:

Field Name	Description
Rank	Ranking of overall sales
Name	Name of the game
Platform	Console on which the game is running
Year	Year of the game released
Genre	Game's category

Field Name	Description
Publisher	Publisher
NA_Sales	Game sales in North America (in millions of units)
EU_Sales	Game sales in the European Union (in millions of units)
JP_Sales	Game sales in Japan (in millions of units)
Other_Sales	Game sales in the rest of the world, i.e. Africa, Asia excluding Japan, Australia, Europe exclu
Global_Sales	Total sales in the world (in millions of units)

Libraries:

- Pandas
- NumPy
- missingno

- Matplotlib
- Seaborn
- Counter from collections
- matplotlib
- matplotlib.pyplot
- matplotlib.style
- matplotlib.colors
- warnings

Tools:

- python
- jupyter notebook
- PowerPoint
- Excel

MVP Goal:

- The MVP goal is to answer of the questions we mentioned