



**CHANDIGARI  
UNIVERSITY**  
Discover. Learn. Empower.

**Proof of Concept  
On  
Video Game Sales Analysis**

**Submitted for the requirement of  
Big Data Engineering Course  
BACHELOR OF ENGINEERING  
COMPUTER SCIENCE & ENGINEERING  
Big Data and Analytics  
CST-334  
Semester-5**

**Submitted to:**

Ms. Gurpreet Kaur  
Project Supervisor

**Submitted By:**

Manraj Singh - 18BCS3752  
Sudiksha - 18BCS3774

## **ACKNOWLEDGEMENT**

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. A special gratitude we give to our 5th semester B.D.E project supervisor, Ms. Gurpreet Kaur, whose contribution in stimulating suggestions and encouragement, helped us to coordinate our project and especially in writing this report. Furthermore, we would also like to acknowledge with much appreciation her crucial role, in giving the permission to use all required equipment and the necessary materials to complete the task '**Video Game Sales Analysis**' using Apache Hadoop (**Hive**), and gave the required suggestion so as to complete the task. The motivation of this project is to visualize the data set and practice exploratory analysis of data set using Hive. For the better understanding, We have analyzed the data by some plots, which will help us to know the trend of the industry.

## **OVERVIEW**

Video Game is an electronic game that is played on electronic medium devices such as personal computer, TV screen, gaming console or mobile phone. Some time the Video Game industry is called the interactive entertainment industry. The input device used for games, the game controller, varies across platforms. Common controller includes game pad, joysticks, mouse, keyboard, the touchscreens of mobile devices and buttons. Players typically view the game on a video screen or television and there are often game sounds from loudspeakers. Video Game development has a long history since 1970's and in recent past with the revolution of the smartphones and tablets introduced new categories of video games such as mobile and social games. Developers introduced various technology and methodology in the computing system to popularize and make the video game more interesting and interactive such as “virtual reality”.

## **OBJECTIVES**

- Present the Data Analysis of the dataset
- Comparison of Sales between different countries
- Calculation of Global Sales of the Video Games
- Top-Selling Games and Gaming Consoles of all-time
- Use of different Visualisations for better understanding

## **COLUMNS AND DATA-TYPES**

• Rank - Ranking of overall sales	<b>Int</b>
• Name - The games name	<b>String</b>
• Platform - Platform of the games release (i.e. PC,PS4, etc.)	<b>String</b>
• Year - Year of the game's release	<b>Int</b>
• Genre - Genre of the game	<b>String</b>
• Publisher - Publisher of the game	<b>String</b>
• NA_Sales - Sales in North America (in millions)	<b>Float</b>
• EU_Sales - Sales in Europe (in millions)	<b>Float</b>
• JP_Sales - Sales in Japan (in millions)	<b>Float</b>
• Other_Sales - Sales in the rest of the world (in millions)	<b>Float</b>
• Global_Sales - Total worldwide sales.	<b>Float</b>

## **PROBLEM STATEMENTS**

**The Dataset:** <https://www.kaggle.com/gregorut/videogamesales>

1. Which genre has been made the most with respect to video games ?
2. Which year had the most game releases ?
3. Top 10 video game publishers with amount of video games published ?
4. Comparison of sales between Japan and Europe.
5. Which is the best selling gaming console of all-time ?
6. Which year had the highest sales worldwide ?
7. Top 10 video games which are the most frequent in the dataset ?

# Shell Script

## Hive

### Step 1:

Create and use a database:

### Query & Result:

```
[training@192 ~]$ hive
Hive history file=/tmp/training/hive_job_log_training_202011302154_920496821.txt
hive> create database bdeproject;
OK
Time taken: 2.904 seconds
hive> use bdeproject;
OK
Time taken: 0.036 seconds
hive>
```

### Step 2:

- Create the table and load the data into the table
- View the table using the describe command.

### Query & Result:

```
hive> create table vgsales(rank int,name string,platform string,year int,genre string,publisher string,na_sales float,
eu_sales float,jp_sales float,other_sales float,global_sales float)
> row format delimited
> fields terminated by ','
> lines terminated by '\n'
> stored as textfile;
```

Time taken: 0.52 seconds

```
hive> load data local inpath '/home/training/Desktop/vgsales.csv' overwrite into table vgsales;
```

```
hive> describe vgsales;
OK
rank      int
name      string
platform  string
year      int
genre     string
publisher string
na_sales  float
eu_sales  float
jp_sales  float
other_sales float
global_sales float
Time taken: 0.306 seconds
hive>
```

## Problem Statement 1:

\* Which genre has been made the most with respect to video games ?

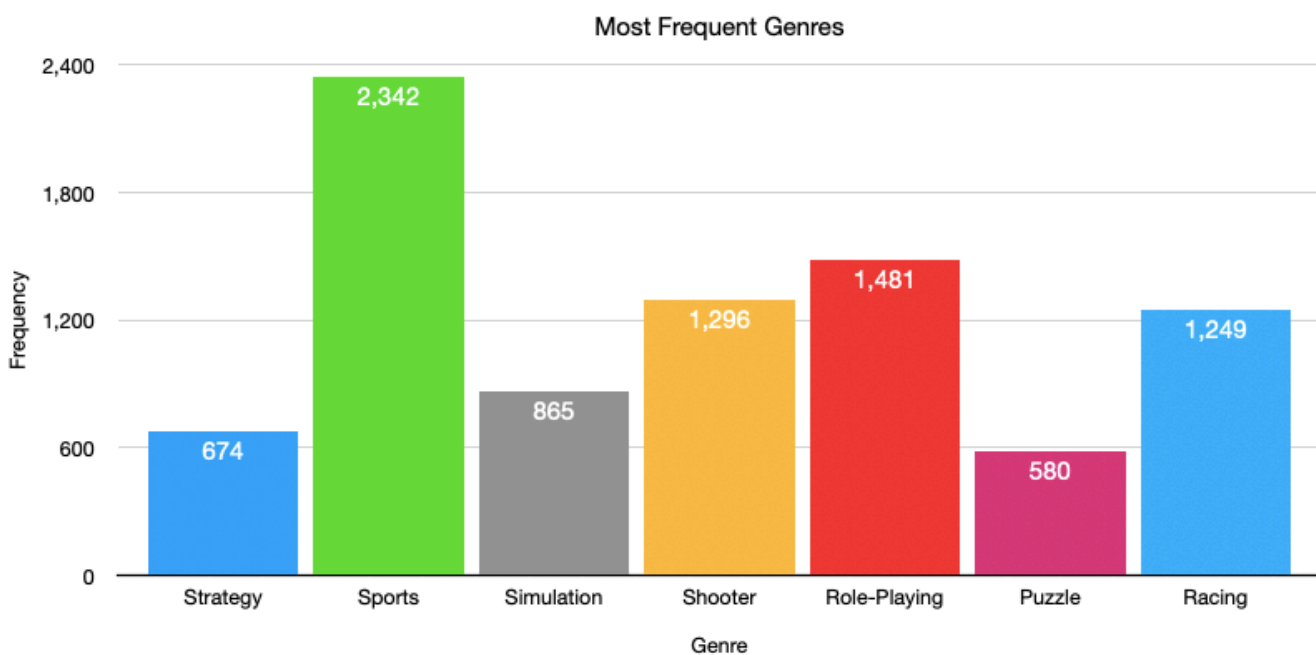
## Query & Result:

```
hive> select genre,COUNT(genre) from vgsales GROUP BY genre ORDER BY genre acse limit 10;
```

```
Strategy          674
Sports 2342
Simulation         865
Shooter 1296
Role-Playing      1481
Racing  1249
Puzzle   580
Time taken: 26.006 seconds
hive> █
```

■ **Sports** genre has been used the most with respect to video games.

## Visualization:



## Problem Statement 2:

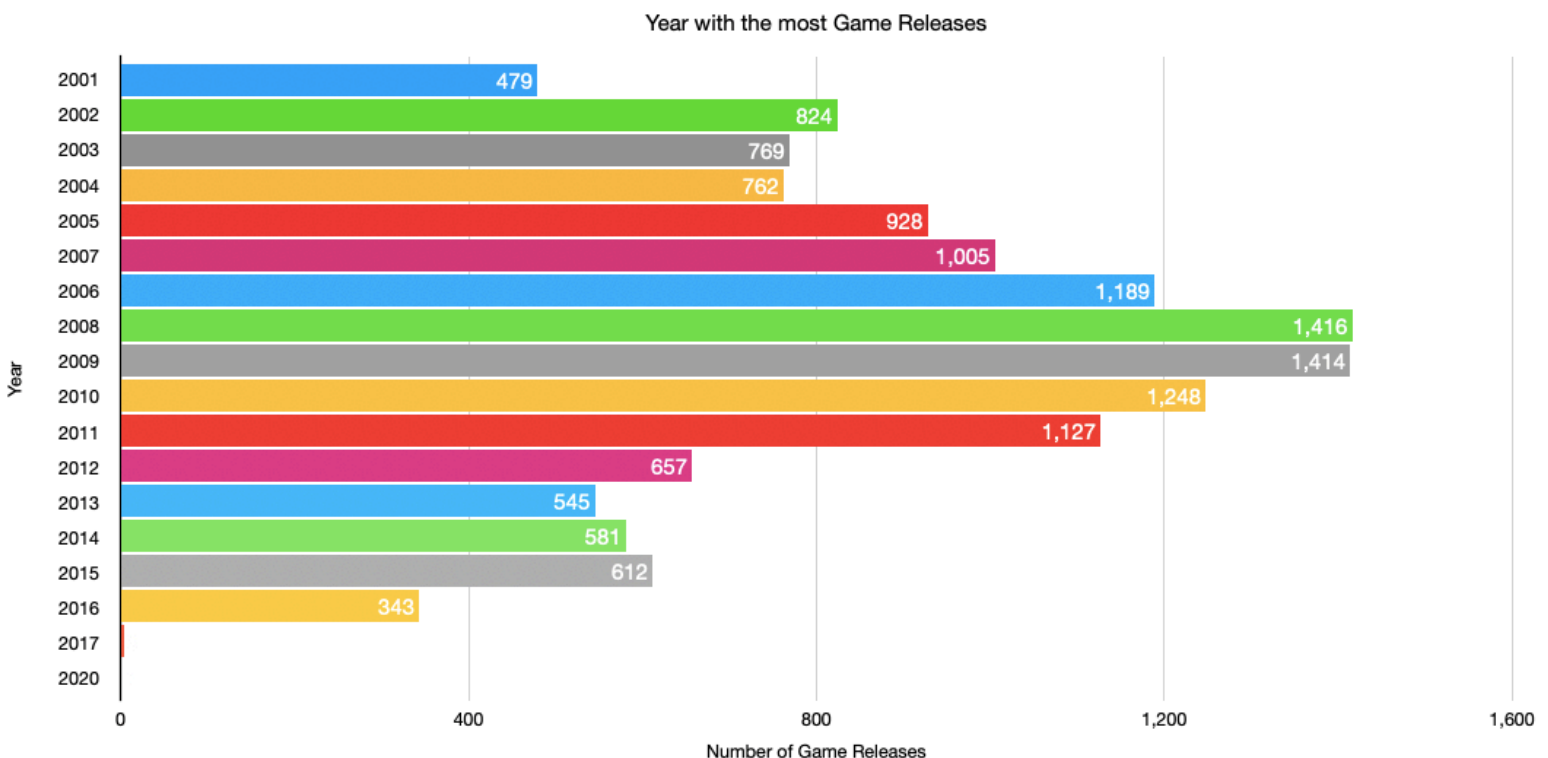
\* Which year had the most game releases ?

## Query & Result

```
hive> select year,COUNT(year) from vgsales where year>2000 group by year;
Ended Job = job_202011302151_0061
OK
2001      479
2002      824
2003      769
2004      762
2005      928
2006     1005
2007     1189
2008     1416
2009     1414
2010     1248
2011     1127
2012      657
2013      545
2014      581
2015      612
2016      343
2017        3
2020        1
Time taken: 15.132 seconds
hive>
```

■ The year **2008** had the most Video Game Releases.

## Visualization:



### Problem Statement 3:

- Top 10 video game publishers with amount of games published ?

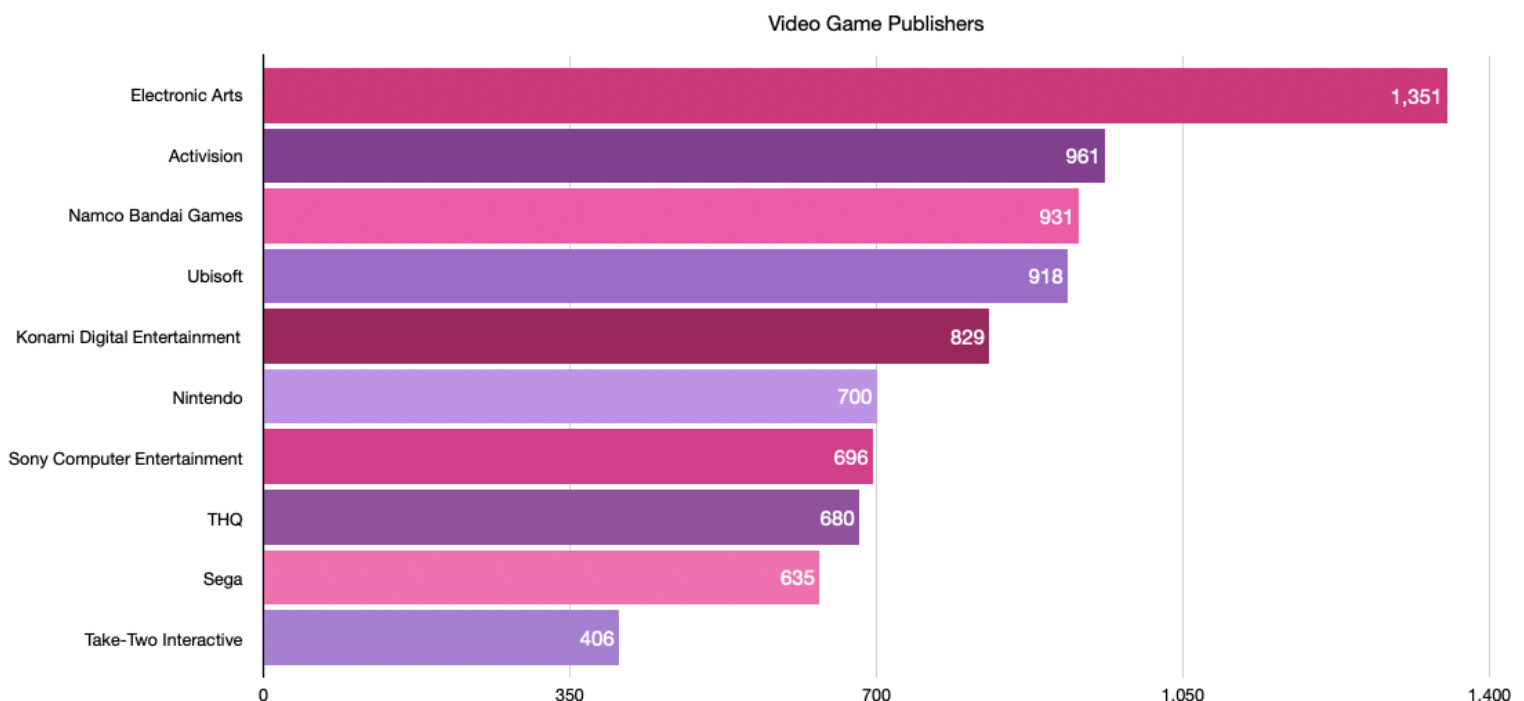
### Query & Result:

```
hive> select publisher, COUNT(publisher) as Frequency from vgsales group by publisher order by Frequency desc limit 10;

Electronic Arts 1351
Activision      961
Namco Bandai Games 931
Ubisoft 918
Konami Digital Entertainment 829
Nintendo 700
THQ 696
Sony Computer Entertainment 680
Sega 635
Take-Two Interactive 406
Time taken: 32.396 seconds
hive>
```

- Electronic Arts published 1351 video games.

### Visualization





#### Problem Statement 4:

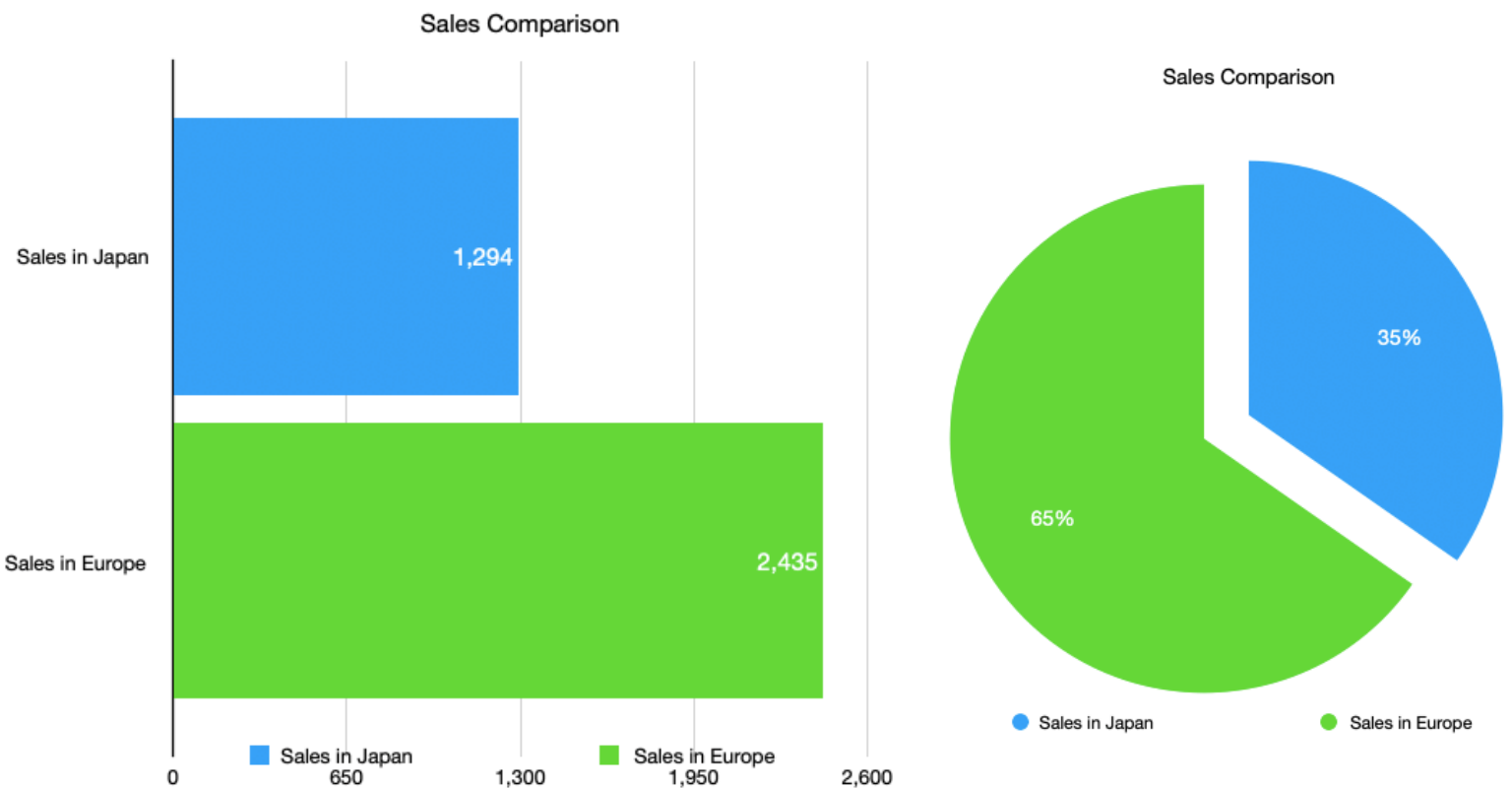
- Comparison of sales between Japan and Europe.

#### Query & Result:

```
Time taken: 14.63 seconds
hive> select SUM(jp_sales),SUM(eu_sales) from vgsales;
```

```
Ended Job = job_202011302151_0024
OK
1293.9699985310435      2435.8699983768165
Time taken: 14.645 seconds
hive> █
```

#### Visualization:



### Problem Statement 5:

■ Which is the best selling gaming console of all-time ?

### Query & Result:

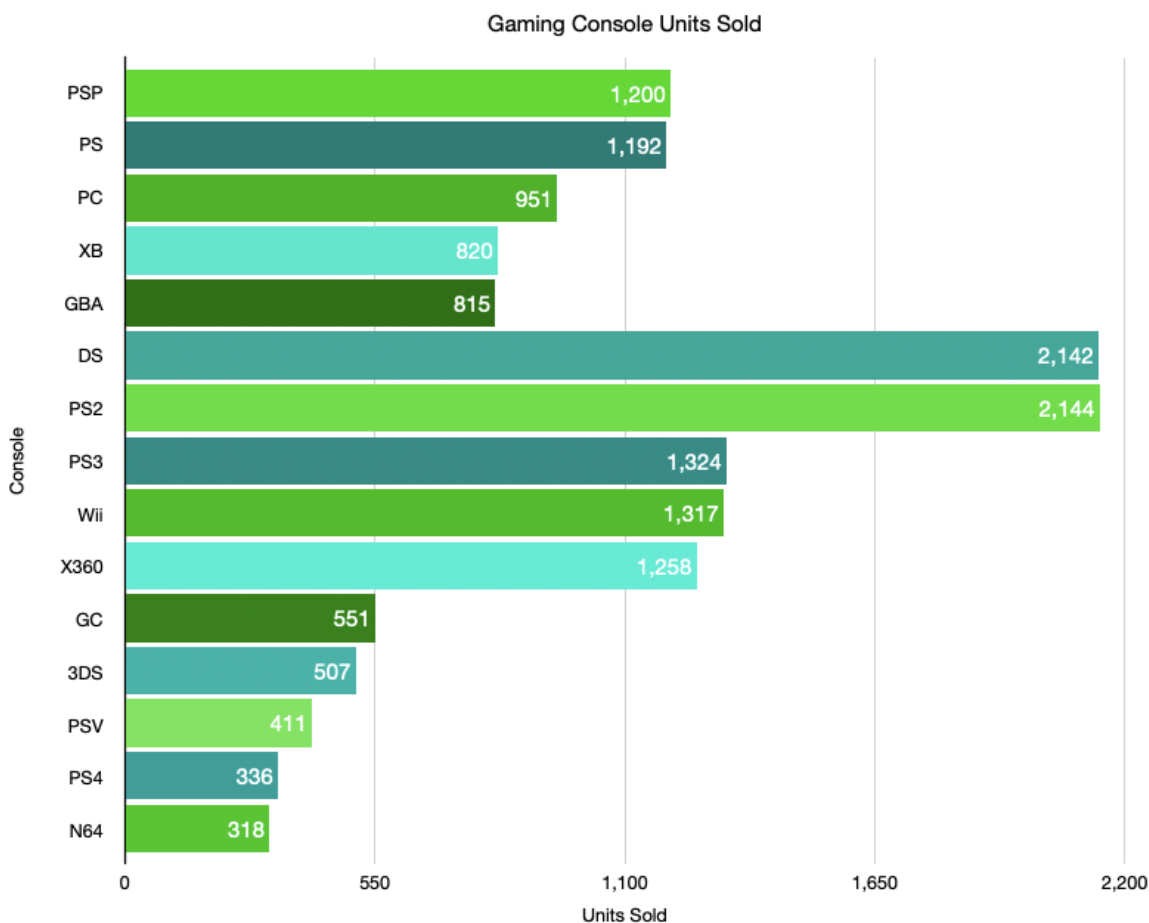
```
hive> select platform, COUNT(platform) as Frequency from vgsales group by platform order by Frequency desc limit 15;
```

PS2	2144
DS	2142
PS3	1324
Wii	1317
X360	1258
PSP	1200
PS	1192
PC	951
XB	820
GBA	815
GC	551
3DS	507
PSV	411
PS4	336
N64	318

Time taken: 29.559 seconds

■ PS2 is the best selling gaming console of all time.

### Visualization:



### **Problem Statement 6:**

- Which year had the highest sales worldwide ?

### **Query & Result:**

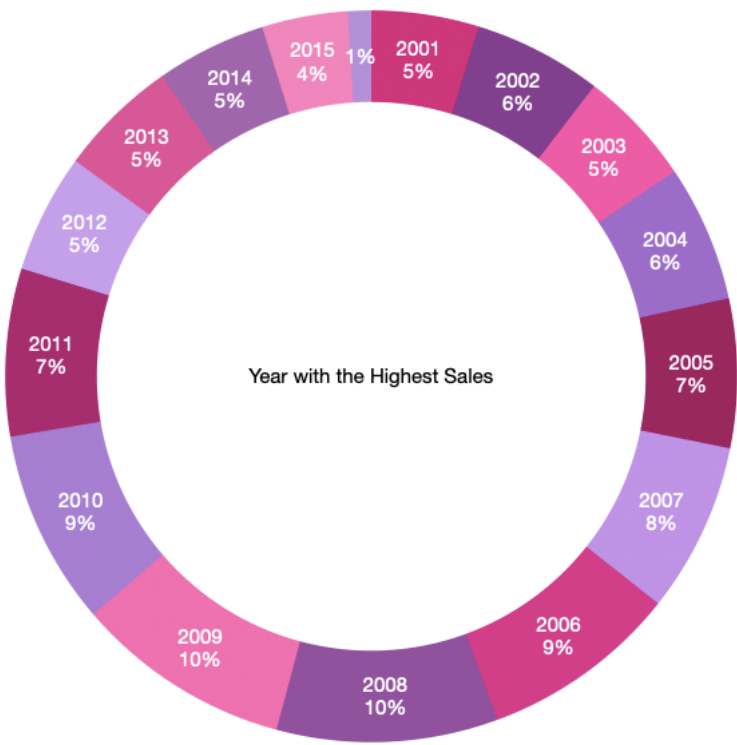
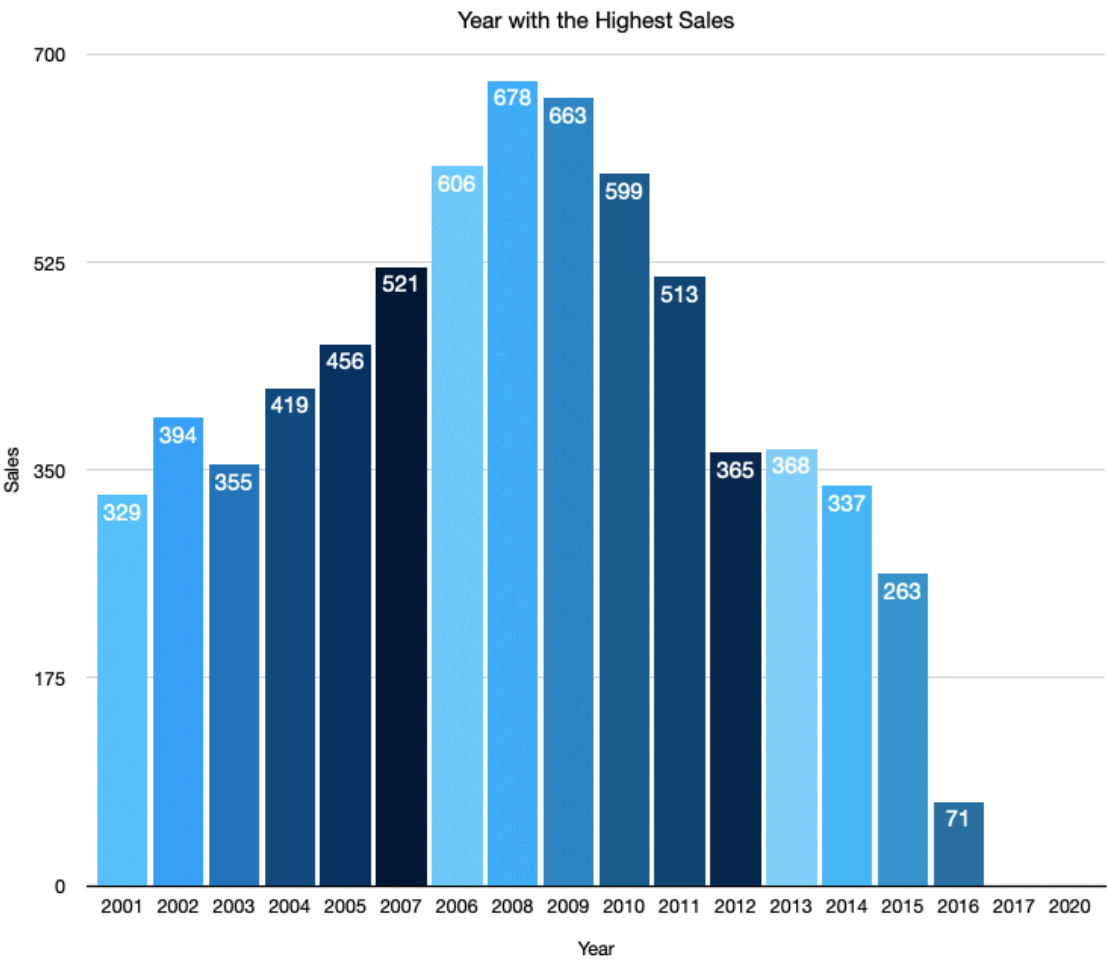
```
hive> select year,SUM(global_sales) from vgsales where year>2000 group by year;
```

2001	329.0699998643249
2002	394.18999859690666
2003	355.53999938070774
2004	419.2199985664338
2005	456.23999912291765
2006	520.7899984251708
2007	606.4800008945167
2008	677.6999990902841
2009	662.7000000644475
2010	598.7699987422675
2011	513.330000154674
2012	363.5299986060709
2013	368.08999936282635
2014	337.0400005541742
2015	263.5600003376603
2016	70.90999971143901
2017	0.04999999888241291
2020	0.28999999165534973

Time taken: 15.174 seconds

- The year **2008** had the highest sales worldwide.
- Year **2008** also had the highest video game releases
- Hence, The year **2008** was quite a revolution for the Video Gaming Industry.

**Visualization:**



## Problem Statement 7:

■ Top 10 video games which are the most frequent in the dataset ?

## Query & Result:

```
hive> select name, COUNT(name) as Frequency from vgsales group by name order by Frequency desc limit 10;■
```

```
Need for Speed: Most Wanted      12
"Warhammer 40                    11
LEGO Marvel Super Heroes         9
Ratatouille                      9
Madden NFL 07                   9
FIFA 14                          9
LEGO The Hobbit                  8
LEGO Jurassic World              8
LEGO Harry Potter: Years 5-7     8
LEGO Star Wars II: The Original  8
Trilogy                          8
Time taken: 28.544 seconds
hive>
```

## Visualization:

