Problem Description

The Video game industry is not as popular as before, as the computer and mobile game industry has grown rapidly in recent years. Global sales of video games decreased after 2008, and it has shown a decline until recently. In this aspect, we wonder which factors a relatively big influence on sales volume increase have, therefore, our project mainly focuses on analyzing the features that can affect sales volume by finding correlations between features. Using this information, we will provide insights into the video game industry and video game sales companies to help them achieve better development and promote their games, enabling them to grow their sales volume.

Conceptual Data Model

Our raw dataset is sourced from Kaggle (https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings) and covers features related to video games with 16 columns including release year, global sales, and publisher, and with 16720 rows between 1985 and 2016 from the website VGChartz and corresponding ratings from Metacritic. VGChartz is a video game sales tracking website that provides weekly sales figures of console software and hardware by region. Metacritic is a website that aggregates reviews of video games. To make the results of the data more accurate, we deleted rows including any empty cells. Besides, we removed 3 columns that we will not use, which were Critic_Count, User_Count, and Developer. Instead, we added a unique ID and ESRB rating description column.

The newly created dataset gathered data on 6,827 video games with 18 columns, consisting of new 5 columns (game, publisher, platform, rating ID and rating description) and the 13 original columns except for the three columns that were deleted. The below table describes each attribute:
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game ID</td>
<td>Numeric</td>
<td>Unique ID for each game</td>
</tr>
<tr>
<td>Game Name</td>
<td>Text</td>
<td>Name of the game</td>
</tr>
<tr>
<td>Release Year</td>
<td>Numeric</td>
<td>Year of the game released</td>
</tr>
<tr>
<td>Genre</td>
<td>Text</td>
<td>Game’s category</td>
</tr>
<tr>
<td>Publisher ID</td>
<td>Numeric</td>
<td>Unique ID for each publisher</td>
</tr>
<tr>
<td>Rating ID</td>
<td>Numeric</td>
<td>Unique ID for each rating grade</td>
</tr>
<tr>
<td>Platform</td>
<td>Text</td>
<td>Console on which the game is running</td>
</tr>
<tr>
<td>Platform ID</td>
<td>Numeric</td>
<td>Unique ID for each platform</td>
</tr>
<tr>
<td>Publisher</td>
<td>Text</td>
<td>The company name that publishes the game</td>
</tr>
<tr>
<td>Rating</td>
<td>Text</td>
<td>The ESRB ratings (E.g. Everyone, Teen…etc)</td>
</tr>
<tr>
<td>Description</td>
<td>Text</td>
<td>The definition of each rating grade</td>
</tr>
<tr>
<td>NA sales</td>
<td>Numeric</td>
<td>Game sales in North America</td>
</tr>
<tr>
<td>EU sales</td>
<td>Numeric</td>
<td>Game sales in the European Union</td>
</tr>
<tr>
<td>JP sales</td>
<td>Numeric</td>
<td>Game sales in Japan</td>
</tr>
<tr>
<td>Other sales</td>
<td>Numeric</td>
<td>Game sales in the rest of the world</td>
</tr>
<tr>
<td>Global sales</td>
<td>Numeric</td>
<td>Total sales in the world</td>
</tr>
<tr>
<td>Critic score</td>
<td>Numeric</td>
<td>Aggregate score compiled by Metacritic staff</td>
</tr>
<tr>
<td>User score</td>
<td>Numeric</td>
<td>Score by Metacritic’s subscribers</td>
</tr>
</tbody>
</table>

For our proposed database, there are five entities: Game, Sales & Score, ESRB rating, Publisher, and Platform. Each game, publisher, platform, and ESRB rating is identified by each unique ID. Video games and platforms have a many-to-many relationship. In other words, each game may be published on more than one type of platform, and each platform can have many video games.

In our proposed database, there are five entities: GAME, SALES & SCORE, ESRB RATING, PUBLISHER, and PLATFORM. Each game, publisher, platform, and rating ID is an
identifier attribute that uniquely references the instances. For SALES & SCORE entity, a combination of the game ID and platform ID is an identifier attribute. All attributes except for each entity’s identifier attribute in the game entity are considered as a simple attribute.

Each game is published by only one publisher (modality= mandatory, cardinality=one). Each publisher can have many games and must have at least one (modality=mandatory, cardinality=many). This logic applies equally to games and ESRB rating relationship. In the case of the relationship between GAME and SALES & SCORE, each game can have more than one sale and score record because there are some duplicated games that are run by different platforms. For example, there are two duplicated game names is called “call of Duty: Black Ops II” is run by PS3 and X360 platforms in our dataset. A game must have at least one sale & score record (modality= mandatory, cardinality=many). Each sale & score record exists by only one game, and the record must have a game (modality=mandatory, cardinality=one). For a relationship between SALE & SCORE and PLATFORM, each platform can have more than one sale and score record, and a platform must have at least one sale & score record (modality=mandatory, cardinality=many). Each sale & score record exists by the only platform, and the record must have a platform (modality=mandatory, cardinality=one).
Relational Schema

The original dataset was in 0 NF because there were some duplicated rows such as the same game name with different platforms and had no primary keys. To perform normalization to make sure that each relationship is in 3NF, we separated the table into 5 different tables by assigning each primary key. For GAME, game ID is the primary key, and publisher and rating ID are the foreign keys that connect PUBLISHER and ESRB rating, respectively. For PUBLISHER, PLATFORM, and RATING, publisher, platform, and rating ID are the primary key, respectively. In addition, to solve the duplicate problem, we assigned the composite key for SALES & SCORE. Thus, game ID and platform ID are the composite keys. As a result, every table has the primary key (PK) that uniquely identifies each row, and relations does not have any repeating groups, partial and transitive dependencies. This is a result for transforming the ERD into a relational schema by using short text statements:

GAME (GameID, GameName, ReleaseYear, Genre, PublisherID, RatingID)
PUBLISHER (PublisherID, Publisher)
ESRB RATING (RatingID, Rating, Description)
PLATFORM (PlatformID, Platform)
SALES & SCORE (GameID, PlatformID, NA Sales, EU Sales, JP Sales, Other Sales, Global Sales, Critic Score, User Score)

This is a result of graphical representation:
Database Implementation

These are the SQL commands to create the tables including INSERT command:

**GAME table:**

```
CREATE TABLE GAME (
    GAMEID NUMBER(5,0) NOT NULL,
    GAMENAME VARCHAR2(255),
    RELEASEYEAR NUMBER(5,0) NOT NULL,
    GENRE VARCHAR2(30),
    CONSTRAINT EMP_GAME_PK primary key (GAMEID),
    CONSTRAINT EMP_GAME_FK foreign key (PUBLISHERID) references PUBLISHER (PUBLISHERID)
    CONSTRAINT EMP_GAME_FK foreign key (RATINGID) references RATING (RATINGID)
);
```

```
INSERT INTO GAME (GAMEID, GAMENAME, RELEASEYEAR, GENRE, PUBLISHERID, RATINGID)
VALUES (1279, 'MySims Agents', 2009, Adventure, 8, 1);
```

<table>
<thead>
<tr>
<th>GAMEID</th>
<th>GAMENAME</th>
<th>RELEASEYEAR</th>
<th>GENRE</th>
<th>PUBLISHERID</th>
<th>RATINGID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1279</td>
<td>MySims Agents</td>
<td>2009</td>
<td>Adventure</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>2001</td>
<td>MySims Racing</td>
<td>2009</td>
<td>Racing</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>2303</td>
<td>NCAA March Madness 06</td>
<td>2003</td>
<td>Sports</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>1735</td>
<td>Need for Speed Shift</td>
<td>2009</td>
<td>Racing</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9007</td>
<td>Disney's Party</td>
<td>2002</td>
<td>Misc</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>2382</td>
<td>SimAnimals</td>
<td>2009</td>
<td>Simulation</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>2247</td>
<td>Mega Man Zero</td>
<td>2002</td>
<td>Platform</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>3667</td>
<td>Carnival Island</td>
<td>2011</td>
<td>Misc</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>852</td>
<td>Tiger Woods PGA Tour 2005</td>
<td>2004</td>
<td>Sports</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

**PLATFORM table:**

```
CREATE TABLE PLATFORM (
    PLATFORMID NUMBER(5,0) NOT NULL,
    PLATFORM VARCHAR2(30)
);
```

```
INSERT INTO PLATFORM (PLATFORMID, PLATFORM)
VALUES (1279, 'MySims Agents', 2009, Adventure, 8, 1);
VALUES (1, 'Will');

PUBLISHER table:
CREATE TABLE PUBLISHER (  
PUBLISHERID NUMBER(5,0) NOT NULL,  
PUBLISHER VARCHAR2(255)  
);  
INSERT INTO PUBLISHER (PUBLISHERID, PUBLISHER)  
VALUES (1, Nintendo);
For the RATING table, we created a description column to explain about each ESRB rating.

**RATING table:**

```sql
CREATE TABLE RATING (  
RATINGID NUMBER(5,0) NOT NULL,  
RATING VARCHAR2(30),  
DESCRIPTION VARCHAR2(30)  );

INSERT INTO RATING (RATINGID, RATING, DESCRIPTION)  
VALUES (1, 'E', 'Everyone');
```
For the SALES table, because one game can be published on more than one platform, and each platform can have many games, we created the composite key (Game ID + Platform ID) to uniquely identify each sale and score for each game on each platform. Previously, the entity name “SALES & SCORE” was changed to “SALES” in the table.

**SALES table:**

```sql
CREATE TABLE SALES (
    GAMEID NUMBER(5,0) NOT NULL,
    PLATFORMID NUMBER(5,0) NOT NULL,
    NA_SALES NUMBER(4,2),
    EU_SALES NUMBER(4,2),
    JP_SALES NUMBER(4,2),
    OTHER_SALES NUMBER(4,2),
    GLOBAL_SALES NUMBER(4,2),
    CRITIC_SCORE NUMBER(4,2),
    USER_SCORE NUMBER(4,2),
    CONSTRAINT EMP_SALES_PK primary key (GAMEID, PLATFORMID),
);

INSERT INTO SALES (GAMEID, PLATFORMID, NA_SALES, EU_SALES, JP_SALES, OTHER_SALES, GLOBAL_SALES, CRITIC_SCORE, USER_SCORE)
VALUES (6361, 13, 0.03, 0.01, 0, 0, 0.03, 79, 7.3);
```
### Web Design and Data Analysis

First of all, to understand the current state of the video game industry, we created a bar chart for sale volume changes from 2005 to 2016. The SQL query command we used is below:

```sql
SELECT GAME.RELEASEYEAR, SUM(SALES.GLOBAL_SALES) AS "Global Sales"
FROM GAME JOIN SALES ON GAME.GAMEID=SALES.GAMEID
WHERE game.releaseyear between 2005 and 2016
GROUP BY GAME.RELEASEYEAR
ORDER BY GAME.RELEASEYEAR;
```

![Global Sales Graph](image)

<table>
<thead>
<tr>
<th>GAMEID</th>
<th>PLATFORMID</th>
<th>NA_SALES</th>
<th>EU_SALES</th>
<th>JP_SALES</th>
<th>OTHER_SALES</th>
<th>GLOBAL_SALES</th>
<th>CRITIC_SCORE</th>
<th>USER_SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>6361</td>
<td>10</td>
<td>.08</td>
<td>.01</td>
<td>0</td>
<td>.00</td>
<td>79</td>
<td>7.3</td>
<td>7.3</td>
</tr>
<tr>
<td>6360</td>
<td>10</td>
<td>0</td>
<td>.22</td>
<td>0</td>
<td>.02</td>
<td>91</td>
<td>9.2</td>
<td>9.2</td>
</tr>
<tr>
<td>6359</td>
<td>10</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.05</td>
<td>68</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>6358</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>72</td>
<td>7.3</td>
<td>7.3</td>
</tr>
<tr>
<td>6357</td>
<td>10</td>
<td>.02</td>
<td>.01</td>
<td>0</td>
<td>.00</td>
<td>72</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>6356</td>
<td>7</td>
<td>.04</td>
<td>0</td>
<td>0</td>
<td>.01</td>
<td>77</td>
<td>6.6</td>
<td>6.6</td>
</tr>
<tr>
<td>6355</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>82</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>6354</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>71</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>6353</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>70</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>6352</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>68</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>6351</td>
<td>10</td>
<td>0</td>
<td>.03</td>
<td>0</td>
<td>.01</td>
<td>62</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>6350</td>
<td>9</td>
<td>.03</td>
<td>0</td>
<td>0</td>
<td>.00</td>
<td>86</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>6349</td>
<td>11</td>
<td>.02</td>
<td>0</td>
<td>0</td>
<td>.02</td>
<td>63</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>6348</td>
<td>9</td>
<td>.03</td>
<td>.01</td>
<td>0</td>
<td>.00</td>
<td>67</td>
<td>9.0</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Although sale volumes increased from 2006 to 2008 and peaked in 2008, the sales volumes have been in decline. The reason for this decline may be the growth of the computer games industry.

Next, to understand overall video game industry trends, we looked at which genre was the most popular globally and in each region (NA, EU, and JP). We found TOP 5.

**Top 5 genre sales in Global & NA**

```sql
SELECT GAME.genre, SUM(SALES.global_sales) AS Global
FROM GAME JOIN SALES ON GAME.gameID = SALES.gameID
GROUP BY GAME.genre
ORDER BY Global DESC
FETCH FIRST 5 ROWS ONLY;
```

```sql
SELECT GAME.GENRE, SUM(SALES.NA_SALES) AS "NORTH AMERICAN"
FROM GAME JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY GAME.GENRE
ORDER BY "NORTH AMERICAN" DESC
FETCH FIRST 5 ROWS ONLY;
```

<table>
<thead>
<tr>
<th>Most Popular 5 Genre</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genre</strong></td>
<td><strong>Global</strong></td>
</tr>
<tr>
<td>Action</td>
<td>1187.02</td>
</tr>
<tr>
<td>Sports</td>
<td>833.85</td>
</tr>
<tr>
<td>Shooter</td>
<td>810.44</td>
</tr>
<tr>
<td>Racing</td>
<td>476.11</td>
</tr>
<tr>
<td>Role-Playing</td>
<td>425.33</td>
</tr>
</tbody>
</table>

**Top 5 genre sales in EU & JP**

```sql
SELECT GAME.GENRE, SUM(SALES.EU_SALES) AS EUROPE
FROM GAME JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY GAME.GENRE
```

```sql
SELECT GAME.GENRE, SUM(SALES.JP_SALES) AS "JAPANESE"
FROM GAME JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY GAME.GENRE
```
ORDER BY EUROPE DESC
FETCH FIRST 5 ROWS ONLY;

SELECT GAME.GENRE, SUM(SALES.JP_SALES) AS JAPAN
FROM GAME JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY GAME.GENRE
ORDER BY JAPAN DESC
FETCH FIRST 5 ROWS ONLY;

As a result, Action, Sports, and Shooter genres were popular globally except for Japan. Exceptionally, Japan showed that the Role-playing genre was the most popular. If game producers only consider game genre regardless of other factors, they must target the action genre. However, if they try to address a new niche market in Japan, they must focus on the Role-playing genre.
Furthermore, we concentrated on the data analysis based on the publisher. We tried to find the top 5 publishers globally and in each region (NA, EU, and JP), and searched the most popular game with the corresponding publisher.

**-Top 5 publisher in Global & NA**

```sql
SELECT PUBLISHER.publisher, SUM(SALES.global_sales) AS GlobalSales
FROM PUBLISHER JOIN GAME ON PUBLISHER.publisherID = GAME.publisherID JOIN SALES ON GAME.gameID = SALES.gameID
GROUP BY PUBLISHER.publisher
ORDER BY GlobalSales DESC
FETCH FIRST 5 ROWS ONLY;
```

```sql
SELECT PUBLISHER.publisher, SUM(SALES.na_sales) AS NA
FROM PUBLISHER JOIN GAME ON PUBLISHER.publisherID = GAME.publisherID JOIN SALES ON GAME.gameID = SALES.gameID
GROUP BY PUBLISHER.publisher
ORDER BY NA DESC
FETCH FIRST 5 ROWS ONLY;
```

**-Top 5 publisher in EU & JP.**

```sql
SELECT PUBLISHER.publisher, SUM(SALES.EU_sales) AS EU
FROM PUBLISHER JOIN GAME ON PUBLISHER.publisherID = GAME.publisherID JOIN SALES ON GAME.gameID = SALES.gameID
GROUP BY PUBLISHER.publisher
```

<table>
<thead>
<tr>
<th>Publisher</th>
<th>GlobalSales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Arts</td>
<td>868.55</td>
</tr>
<tr>
<td>Nintendo</td>
<td>849.49</td>
</tr>
<tr>
<td>Activision</td>
<td>535.74</td>
</tr>
<tr>
<td>Sony Computer Entertainment</td>
<td>388.1</td>
</tr>
<tr>
<td>Take-Two Interactive</td>
<td>350.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publisher</th>
<th>EU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic Arts</td>
<td>465.45</td>
</tr>
<tr>
<td>Nintendo</td>
<td>371.36</td>
</tr>
<tr>
<td>Activision</td>
<td>306.82</td>
</tr>
<tr>
<td>Take-Two Interactive</td>
<td>187.82</td>
</tr>
<tr>
<td>Sony Computer Entertainment</td>
<td>176.37</td>
</tr>
</tbody>
</table>
ORDER BY EU DESC
FETCH FIRST 5 ROWS ONLY;

SELECT PUBLISHER.publisher, SUM(SALES.JP_sales) AS JP
FROM PUBLISHER JOIN GAME ON PUBLISHER.publisherID = GAME.publisherID JOIN SALES ON
GAME.gameID = SALES.gameID
GROUP BY PUBLISHER.publisher
ORDER BY JP DESC
FETCH FIRST 5 ROWS ONLY;

Top 10 popular games with publishers

SELECT GAME.gamename, PUBLISHER.publisher, SUM(SALES.global_sales) AS GlobalSales
FROM PUBLISHER JOIN GAME ON PUBLISHER.publisherID = GAME.publisherID JOIN SALES ON
GAME.gameID = SALES.gameID
GROUP BY GAME.gamename, PUBLISHER.publisher
ORDER BY GlobalSales DESC
Interesting results happened. Certain companies had a high market share such as Electronic Arts, Nintendo, Activision, and Sony Computer Entertainment. However, according to the result of the top 10 game sellers, Electronic Arts, Activision, and Sony Computer Entertainment were not included. After we researched Electronic Arts to find the reason for this situation, Electronic Arts were one of the representative companies that succeeded in producing the sports game series. We created one query to find game series including “FIFA” as shown below.

-Finding a popular series

Select GAME.GAMENAME, SUM(SALES.GLOBAL_SALES) AS "GLOBAL"
FROM PUBLISHER JOIN GAME ON PUBLISHER.PUBLISHERID = GAME.PUBLISHERID
JOIN SALES ON GAME.GAMEID = SALES.GAMEID
WHERE PUBLISHER.PUBLISHER = 'Electronic Arts' AND GAME.GAMENAME LIKE '%FIFA%
GROUP BY PUBLISHER.PUBLISHER, GAME.GAMENAME
ORDER BY "GLOBAL" DESC;
Now, we found there were many sports game series produced by Electronic Arts (EA). In terms of the game publisher, if investors are looking for game publishers with high returns and are interested in a sports game, they have to invest Electronic Arts. However, if they consider the Japanese market as well, investing Nintendo will be the best choice because they have an overwhelming high market share in Japan.

To identify the appropriate targets, we identified which age rating contributed greatly to sales globally and in each region (NA, EU, and JP).

-Who spend most? In Global & NA.

```
SELECT RATING.RATING, RATING.DESCRIPTION, SUM(SALES.global_sales) AS Global
FROM RATING JOIN GAME ON RATING.RATINGID = GAME.RATINGID
JOIN SALES ON GAME.GAMEID = SALES.GAMEID
GROUP BY RATING.RATING, RATING.DESCRIPTION
ORDER BY Global DESC
FETCH FIRST 4 ROWS ONLY;
```
SELECT RATING.RATING, RATING.DESCRIPTION, SUM(SALES.na_sales) AS NA
FROM RATING JOIN GAME ON RATING.RATINGID=GAME.RATINGID JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY RATING.RATING, RATING.DESCRIPTION
ORDER BY NA DESC
FETCH FIRST 4 ROWS ONLY;

-Who spend most? in EU & JP.

SELECT RATING.RATING, RATING.DESCRIPTION, SUM(SALES.EU_sales) AS Europe
FROM RATING JOIN GAME ON RATING.RATINGID=GAME.RATINGID JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY RATING.RATING, RATING.DESCRIPTION
ORDER BY Europe DESC
FETCH FIRST 4 ROWS ONLY;

SELECT RATING.RATING, RATING.DESCRIPTION, SUM(SALES.JP_sales) AS Japan
FROM RATING JOIN GAME ON RATING.RATINGID=GAME.RATINGID JOIN SALES ON GAME.GAMEID=SALES.GAMEID
GROUP BY RATING.RATING, RATING.DESCRIPTION
ORDER BY Japan DESC
FETCH FIRST 4 ROWS ONLY;
As expected, E rating, which is for everyone, had the highest sales in the world, followed by M and T rating.

To do a deeper analysis, we identified which age rating contributed to the high sales for the three most popular genres (Action, Sports, and Shooter) we got from the previous query.

- Top 7 Combination Rating with genres

```sql
SELECT RATING.rating, GAME.genre, SUM(SALES.global_sales) AS GlobalSales
FROM GAME JOIN SALES ON GAME.gameID = SALES.gameID JOIN RATING ON GAME.ratingID = RATING.ratingID
WHERE(SELECT DISTINCT(GENRE) FROM GAME WHERE GENRE='Action') = GAME.genre
GROUP BY RATING.rating, GAME.genre
UNION
SELECT RATING.rating, GAME.genre, SUM(SALES.global_sales) AS GlobalSales
FROM GAME JOIN SALES ON GAME.gameID = SALES.gameID JOIN RATING ON GAME.ratingID = RATING.ratingID
WHERE(SELECT DISTINCT(GENRE) FROM GAME WHERE GENRE='Sports') = GAME.genre
```
GROUP BY RATING.rating, GAME.genre
UNION
SELECT RATING.rating, GAME.genre, SUM(SALES.global_sales) AS GlobalSales
FROM GAME JOIN SALES ON GAME.gameID = SALES.gameID JOIN RATING ON GAME.ratingID = RATING.ratingID
WHERE(SELECT DISTINCT(GENRE) FROM GAME WHERE GENRE='Shooter') = GAME.genre
GROUP BY RATING.rating, GAME.genre
ORDER BY GlobalSales DESC
FETCH FIRST 7 ROWS ONLY;

Even though the sports genre with E rating showed the highest global sales at 710.28 million, the important thing we need to pay attention to is there is only one sports genre of top 7. Rather, the sum of the action genre is larger than the sports genre at 1185.07 million (568.75 + 314.31+...+ 140.09). Even, the sum of the Shooter genre is larger than the Sports genre. In terms of rating, the sum of M rating shows a higher sales volume than that of E rating (1172.4 vs. 850.37). In brief, all things considered together, the game industry stakeholders are most likely to succeed when they invest in games with M rating and Action genre.

Finally, we tried to make several combinations to explore a wider range of options we can suggest. The first one was about genres based on each top 3 publishers, and the second one was about genres based on each top 3 ratings. These are SQL commands:
- Sales by Publisher & Genre

```sql
SELECT PUBLISHER.publisher, GAME.GENRE, SUM(SALES.global_sales) AS Global
FROM PUBLISHER JOIN GAME ON PUBLISHER.PUBLISHERID=GAME.PUBLISHERID JOIN SALES ON
GAME.GAMEID=SALES.GAMEID
WHERE PUBLISHER.publisher = 'Electronic Arts'
GROUP BY PUBLISHER.publisher, GAME.GENRE
ORDER BY Global DESC
FETCH FIRST 3 ROWS ONLY
;
```

```sql
SELECT PUBLISHER.publisher, GAME.GENRE, SUM(SALES.global_sales) AS Global
FROM PUBLISHER JOIN GAME ON PUBLISHER.PUBLISHERID=GAME.PUBLISHERID JOIN SALES ON
GAME.GAMEID=SALES.GAMEID
WHERE PUBLISHER.publisher = 'Nintendo'
GROUP BY PUBLISHER.publisher, GAME.GENRE
ORDER BY Global DESC
FETCH FIRST 3 ROWS ONLY
;
```

```sql
SELECT PUBLISHER.publisher, GAME.GENRE, SUM(SALES.global_sales) AS Global
FROM PUBLISHER JOIN GAME ON PUBLISHER.PUBLISHERID=GAME.PUBLISHERID JOIN SALES ON
GAME.GAMEID=SALES.GAMEID
WHERE PUBLISHER.publisher = 'Activision'
GROUP BY PUBLISHER.publisher, GAME.GENRE
ORDER BY Global DESC
FETCH FIRST 3 ROWS ONLY
;
```

- Targeting Customer

```sql
SELECT RATING.RATING, SUM(SALES.global_sales) AS Global, GAME.GENRE
```

<table>
<thead>
<tr>
<th>Electronic Arts</th>
<th>Nintendo</th>
<th>Activision</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publisher</strong></td>
<td><strong>Genre</strong></td>
<td><strong>Global</strong></td>
</tr>
<tr>
<td>Electronic Arts</td>
<td>Sports</td>
<td>359.85</td>
</tr>
<tr>
<td>Electronic Arts</td>
<td>Shooter</td>
<td>146.8</td>
</tr>
<tr>
<td>Electronic Arts</td>
<td>Racing</td>
<td>108.83</td>
</tr>
<tr>
<td>Nintendo</td>
<td>Sports</td>
<td>178.88</td>
</tr>
<tr>
<td>Nintendo</td>
<td>Platform</td>
<td>169.52</td>
</tr>
<tr>
<td>Nintendo</td>
<td>Misc</td>
<td>132.78</td>
</tr>
<tr>
<td>Activision</td>
<td>Shooter</td>
<td>247.03</td>
</tr>
<tr>
<td>Activision</td>
<td>Action</td>
<td>95.68</td>
</tr>
<tr>
<td>Activision</td>
<td>Misc</td>
<td>60.99</td>
</tr>
</tbody>
</table>

1 - 3
Based on the information we got from previous and the first combination, one interesting finding is that Activation has the highest market leadership in the shooting genre, and the company with the highest market leadership in the sports genre was Electronic Arts as we saw in the previous results. Through the second combination, we found the best game genre for the Top
3 rating. As popular genres are different for each rating in general, video game industry stakeholders need to consider this information if they mainly focus on ESRB rating.

Conclusion

Given all the information we have analyzed, the best combination to activate the video game industry is to develop games with the action genre and E rating. If the video game industry stakeholders want to invest in a new market that is not yet explored, then targeting the Japanese market will be one of the best choices, as there were low market share tendencies and many exceptional results such as the top 5 popular genres, popular publisher ranking.

Application Information