```
--1.数据清洗
--1.1 查看发行年份的组成数据值，并保留发行年份在2000年至2017年之间的数据

--查看数据总行数
SELECT COUNT(*) FROM video_games;
/* 结果：16719*/

--查看发行年份列数据组成成分
SELECT Year_of_Release, COUNT(*) FROM video_games
    GROUP BY Year_of_Release;
/*结果：
2006    1006
1985    14
2008    1427
2009    1426
1996    263
1989    17
1984    14
2005    939
1999    338
2007    1197
2010    1255
2013    544
2004    762
1990    16
1988    15
2002    829
2001    482
2011    1136
1998    379
2015    606
2012    653
2014    581
1992    43
1997    289
1993    62
1994    121
1982    36
2016    502
2003    775
1986    21
2000    350
N/A 269
1995    219
1991    41
1981    46
1987    16
1980    9
1983    17
2020    1
2017    3

可以看到有一个N/A的异常值，等等一并删除 */

--查询1999年游戏总数（后续有用）
SELECT COUNT(*) FROM video_games
    WHERE Year_of_Release = 1999 AND Name IS NOT NULL;
/*结果：338*/

--删除异常值
DELETE FROM video_games WHERE Year_of_Release = 'N/A';
DELETE FROM video_games WHERE Year_of_Release < 2000 OR Year_of_Release > 2016;

--查看剩余数据行数
SELECT COUNT(*) FROM video_games;
/* 结果：14470*/

--1.2 清洗姓名列（Name）

--查看姓名列是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Name IS NULL;
/* 结果：0，无空值*/
```

```

--查看姓名列是否有重复值
SELECT * FROM video_games
    WHERE Name IN (SELECT Name FROM video_games
                    GROUP BY Name
                    HAVING COUNT(*) > 1)
    ORDER BY Name;
/*结果：(由于结果太长，故截取部分)

 Frozen: Olaf's Quest   3DS 2013    Platform    Disney Interactive Studios  0.27    0.27    0   0.05    0.6
 Frozen: Olaf's Quest   DS  2013    Platform    Disney Interactive Studios  0.21    0.26    0   0.04    0.52
[Prototype 2]   PC  2012    Action  Activision  0.07    0.03    0   0.01    0.11    76  12  6.4 389 Radical Entertainment   M
[Prototype 2]   X360    2012    Action  Activision  0.48    0.24    0   0.07    0.79    74  69  7   173 Radical Entertainment   M
[Prototype 2]   PS3 2012    Action  Activision  0.36    0.28    0   0.1 0.74    79  39  6.8 179 Radical Entertainment   M
[Prototype] X360    2009    Action  Activision  0.84    0.35    0   0.12    1.31    78  83  7.8 356 Radical Entertainment   M
[Prototype] PS3 2009    Action  Activision  0.65    0.4 0   0.19    1.24    79  53  7.7 308 Radical Entertainment   M
007: Quantum of Solace  PC  2008    Action  Activision  0.01    0.01    0   0   0.03    70  18  6.3 55  Treyarch    T
007: Quantum of Solace  Wii 2008    Action  Activision  0.29    0.28    0.01    0.07    0.65    54  11  7.5 26  Treyarch    T
007: Quantum of Solace  DS  2008    Action  Activision  0.11    0.01    0   0.01    0.13    65  10  tbd     Vicarious Visions   T
007: Quantum of Solace  PS2 2008    Action  Activision  0.17    0   0   0.26    0.43
007: Quantum of Solace  X360    2008    Action  Activision  0.82    0.51    0.01    0.14    1.48    65  69  7.1 71  Treyarch    T
007: Quantum of Solace  PS3 2008    Action  Activision  0.43    0.51    0.02    0.19    1.14    65  42  6.6 47  Treyarch    T
007: The World is not Enough    N64 2000    Action  Electronic Arts 1.13    0.38    0.02    0.03    1.55
007: The World is not Enough    PS  2000    Action  Electronic Arts 0.51    0.35    0   0.06    0.92    61  11  6.7 44  Black Ops Entertainment T

可以看出是因为同一个游戏发行在不同的平台造成的，故我们不做删除处理，并将不同平台的同一游戏视作不同的游戏*/

--1.3 清洗平台列（Platform）

--查看平台列是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Platform IS NULL;
/* 结果：0，无空值*/

--查看平台列组成成分
SELECT DISTINCT(Platform) FROM video_games;
/*结果：
Wii
DS
X360
PS3
PS2
GBA
PS4
3DS
XB
PC
PSP
XOne
WiiU
GC
GB
PS
N64
PSV
DC
WS
```

```sql
无异常值*/

--1.4 清洗游戏类型列（Genre）

--查看游戏类型列是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Genre IS NULL;
/* 结果：0，无空值*/

--查看游戏类型列组成成分
SELECT DISTINCT(Genre) FROM video_games;
/*结果：
Sports
Racing
Platform
Misc
Simulation
Action
Role-Playing
Puzzle
Shooter
Fighting
Adventure
Strategy
分为运动类、竞速类、平台跳跃类、杂项、模拟类、动作类、角色扮演类、益智类、射击类、格斗类、冒险类、策略类游戏*/

--1.5 清洗发行商列（Publisher）

--查看发行商列是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Publisher IS NULL;
/* 结果：0，无空值*/

--查看游戏类型列组成成分
SELECT DISTINCT(Publisher) FROM video_games;
/*结果过多，不一一展示，但是可以看出有些公司命名不规范或经过收购，一家公司出现两种说法，如Activision和Activision Value统一改为Activision Value;
Ascaron Entertainment和Ascaron Entertainment GmbH统一改为后者等等。*/

--修改不规范名称
UPDATE video_games SET Publisher = 'Activision Value'
    WHERE Publisher = 'Activision';
UPDATE video_games SET Publisher = 'Ascaron Entertainment GmbH'
    WHERE Publisher = 'Ascaron Entertainment';
UPDATE video_games SET Publisher = 'ASCII Entertainment'
    WHERE Publisher = 'ASCII Media Works';
UPDATE video_games SET Publisher = 'Avanquest Software'
    WHERE Publisher = 'Avanquest';
UPDATE video_games SET Publisher = 'Big Ben Interactive'
    WHERE Publisher = 'Bigben Interactive';
UPDATE video_games SET Publisher = 'Codemasters'
    WHERE Publisher = 'Codemasters Online';
UPDATE video_games SET Publisher = 'Daedalic Entertainment'
    WHERE Publisher = 'Daedalic';
UPDATE video_games SET Publisher = 'FuRyu Corporation'
    WHERE Publisher = 'FuRyu';
UPDATE video_games SET Publisher = 'Hudson Soft'
    WHERE Publisher = 'Hudson Entertainment';
UPDATE video_games SET Publisher = 'Idea Factory'
    WHERE Publisher = 'Idea Factory International';
UPDATE video_games SET Publisher = 'Kadokawa Games'
    WHERE Publisher = 'Kadokawa Shoten';
UPDATE video_games SET Publisher = 'Marvelous Entertainment'
    WHERE Publisher = 'Marvelous Games' OR Publisher = 'Marvelous Interactive';
UPDATE video_games SET Publisher = 'Milestone'
    WHERE Publisher = 'Milestone S.r.l' OR Publisher = 'Milestone S.r.l.';
UPDATE video_games SET Publisher = 'Nippon'
    WHERE Publisher = 'Nippon Amuse' OR Publisher = 'Nippon Columbia' OR Publisher = 'Nippon Ichi Software';
UPDATE video_games SET Publisher = 'Paon Corporation'
    WHERE Publisher = 'Paon';
UPDATE video_games SET Publisher = 'Paradox'
```

```sql
    WHERE Publisher = 'Paradox Development' OR Publisher = 'Paradox Interactive';
UPDATE video_games SET Publisher = 'Rebellion'
    WHERE Publisher = 'Rebellion Developments';
UPDATE video_games SET Publisher = 'Revolution Software'
    WHERE Publisher = 'Revolution (Japan)';
UPDATE video_games SET Publisher = 'SNK'
    WHERE Publisher = 'SNK Playmore';
UPDATE video_games SET Publisher = 'Sony Computer Entertainment'
    WHERE Publisher = 'Sony Computer Entertainment America' OR Publisher = 'Sony Computer Entertainment Europe'
       OR Publisher = 'Sony Music Entertainment' OR Publisher = 'Sony Online Entertainment';
UPDATE video_games SET Publisher = 'Square'
    WHERE Publisher = 'Square Enix' OR Publisher = 'Square Enix ' OR Publisher = 'SquareSoft';
UPDATE video_games SET Publisher = 'System 3'
    WHERE Publisher = 'System 3 Arcade Software';
UPDATE video_games SET Publisher = 'Takara'
    WHERE Publisher = 'Takara Tomy';
UPDATE video_games SET Publisher = 'TDK'
    WHERE Publisher = 'TDK Core' OR Publisher = 'TDK Mediactive';
UPDATE video_games SET Publisher = 'Ubisoft'
    WHERE Publisher = 'Ubisoft Annecy';
UPDATE video_games SET Publisher = 'Valve Software'
    WHERE Publisher = 'Valve';

--1.6 清洗销量列（NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales）

--查看销量列是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE NA_Sales IS NULL;
SELECT COUNT(*) FROM video_games
    WHERE EU_Sales IS NULL;
SELECT COUNT(*) FROM video_games
    WHERE JP_Sales IS NULL;
SELECT COUNT(*) FROM video_games
    WHERE Other_Sales IS NULL;
SELECT COUNT(*) FROM video_games
    WHERE Global_Sales IS NULL;
/* 结果：全部为0，无空值*/

--1.7 清洗媒体评分和媒体总数列（Critic_Score, Critic_Count）

--查看是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Critic_Score IS NULL;
SELECT COUNT(*) FROM video_games
    WHERE Critic_Count IS NULL;
/* 结果：均为6586，因数量过多，故删除这两列*/
ALTER TABLE video_games DROP COLUMN Critic_Score,  DROP COLUMN Critic_Count;

--1.8 删除开发商列（Devloper）此项目不研究此项
ALTER TABLE video_games DROP COLUMN Developer;

----1.9 清洗用户评分和用户总数列（User_Score, User_Count）

--通过观察发现评分列不仅有空值，还有tbd值
--查看空值和tbd值数量
SELECT COUNT(*) FROM video_games
    WHERE User_Score IS NULL OR User_Score = 'tbd';
SELECT COUNT(*) FROM video_games
    WHERE User_Count IS NULL;
/* 结果：均为7102，因数量过多，故删除这两列*/
ALTER TABLE video_games DROP COLUMN User_Score,  DROP COLUMN User_Count;

--1.10 清洗评级列（Rating）

--查看是否有NULL值或空值
SELECT COUNT(*) FROM video_games
    WHERE Rating IS NULL;
/* 结果：4810，根据美国评级系统，填入RP*/
UPDATE video_games SET Rating = 'RP'
    WHERE Rating IS NULL;
```

```
--2.数据分析（MySQL+Tableau）
--2.1 游戏数量-增长率-年份关系
SELECT Year_of_Release AS '发行年份', COUNT(*) AS '游戏总数',
       CASE a.Year_of_Release
       WHEN 2000 THEN ((COUNT(*)/338)-1)
       ELSE ((COUNT(*)/(SELECT COUNT(*) FROM video_games AS b
                                    WHERE b.Year_of_Release = a.Year_of_Release-1
                                    GROUP BY b.Year_of_Release))-1)
       END AS '增长率'
    FROM video_games AS a
    GROUP BY Year_of_Release
    ORDER BY Year_of_Release;
--2.2 各平台发行游戏数-平均销量-年份关系
SELECT Year_of_Release AS '发行年份', Platform AS '发行平台', COUNT(*) AS '游戏总数', SUM(Global_Sales)/COUNT(*) AS '平均销量'
    FROM video_games
    GROUP BY Year_of_Release, Platform;
--2.3 游戏类型发行数-年份关系
SELECT Year_of_Release AS '发行年份', Genre AS '游戏类型', COUNT(*) AS '发行数量'
    FROM video_games
    GROUP BY Year_of_Release, Genre;
--2.4 各游戏类型销量-年份关系
SELECT Year_of_Release AS '发行年份', Genre AS '游戏类型', SUM(Global_Sales)/COUNT(*) AS '平均销量'
FROM video_games
GROUP BY Year_of_Release, Genre
--2.5 2000-2016年间各发行商总发行游戏数
SELECT ROW_NUMBER(), Publisher AS '发行商', COUNT(*) AS '发行总数'
    FROM video_games
    GROUP BY Publisher;
--2.6 发行总数前十的发行商及总发行游戏数
SELECT Publisher AS '发行商', COUNT(*) AS '总发行游戏数'
    FROM video_games
    GROUP BY Publisher
    ORDER BY COUNT(*) DESC LIMIT 10;
--2.7 发行总数前十的发行商的平均销量-总游戏数关系
SELECT Publisher AS '发行商', COUNT(*) AS '总发行游戏数', SUM(Global_Sales)/COUNT(*) AS '平均销量'
    FROM video_games
    GROUP BY Publisher
    ORDER BY COUNT(*) DESC LIMIT 10;
--2.8 发行总数前十的发行商游戏发行数占比/平均销量占比-年份关系
CREATE VIEW TOP10 AS
    (SELECT Publisher FROM video_games
                    GROUP BY Publisher
                    ORDER BY COUNT(*) DESC LIMIT 10) a);
CREATE VIEW EveryYear AS
(SELECT Year_of_Release, COUNT(*) AS C, SUM(Global_Sales) AS S
        FROM video_games
        GROUP BY Year_of_Release);
SELECT P_C.Year_of_Release AS '发行年份', P_C.Publisher AS '发行商',
       P_C.C1/EveryYear.C AS '发行数量占比', P_C.S1/everyyear.S AS '销量占比'
    FROM
    (SELECT Publisher, Year_of_Release, COUNT(*) AS C1, SUM(Global_Sales) AS S1
        FROM video_games
        WHERE Publisher IN (SELECT Publisher FROM TOP10)
        GROUP BY Year_of_Release, Publisher) P_C
    LEFT JOIN EveryYear
    ON P_C.Year_of_Release = EveryYear.Year_of_Release;
--2.9 各地区销量占比-年份关系
CREATE VIEW AreaSales AS
(SELECT Year_of_Release,
       SUM(NA_Sales) a,
       SUM(EU_Sales) b,
       SUM(JP_Sales) c,
       SUM(Other_Sales) d,
       SUM(Global_Sales) e
    FROM video_games
    GROUP BY Year_of_Release);
SELECT Year_of_Release AS '发行年份',
       a/e AS '北美',
       b/e AS '欧洲',
```

```sql
       c/e AS '日本',
       d/e AS '其他地区'
    FROM AreaSales;
--2.11 各地区喜爱游戏类型-年份关系
SELECT m.Year_of_Release AS '发行年份', m.Genre AS '游戏类型',
       m.na/AreaSales.a AS '北美', m.eu/AreaSales.b AS '欧洲',
       m.jp/AreaSales.c AS '日本', m.other/AreaSales.d AS '其他地区'
FROM
(SELECT Year_of_Release, Genre,
        SUM(NA_Sales) na,
        SUM(EU_Sales) eu,
        SUM(JP_Sales) jp,
        SUM(Other_Sales) other
    FROM video_games
    GROUP BY Year_of_Release, Genre) AS m
LEFT JOIN AreaSales
ON m.Year_of_Release = AreaSales.Year_of_Release;

```