Deepak sai Paty

Info 5810

Fall 2020

Trends in video game sales across different regions

**About the dataset**

* I have chosen this dataset because I do play lot of video games that comes in the market. I got curious to know about the market stats of some of my favorite games like COD, Mario etc.
* The data has been collected from [vgchartz](https://www.vgchartz.com/) website using web scraping libraries available in python programming language.
* The scrapped data describing video game sales is available at [Kaggle](https://www.kaggle.com/gregorut/videogamesales)
* The data has been collected by ‘Grogory Smith’ to make analysis on video games sales worldwide.
* Data dictionary-

|  |  |
| --- | --- |
| **Rank** | Rank column describes the ranking of video game sales on whole. |
| **Name** | Name assigned to a video game. |
| **Platform** | Platform column describes on which column the game has been released. For example, a game can be released to play on a PC or on a PlayStation etc. |
| **Year** | Defines the game release date |
| **Genre** | Games Genre, for example it can be sports or action or shooting etc. |
| **Publisher** | Publisher column defined the owner of the game. |
| **NA\_Sales** | This column defines the sales in millions in North America continent. |
| **EU\_Sales** | This column defines the sales in millions in Europe continent. |
| **JP\_Sales** | This column defines the sales in Japan country. |
| **Other\_Sales** | This column defines the sales in remaining countries other than North America, Europe, and Japan. |
| **Global Sales** | This column defines the sales took place in whole globe. (Sales in NA+ Sales in EU+ Sales in Japan+ Sales in other countries) |

* + The data consists of 11 columns and 16598 rows in it.



Figure 1 Showing all the fields of the video game dataset

**Data cleaning-**

* I will be considering all the columns to answer the questions mentioned in the data goals.
* The ‘publisher’ column has few NULL fields. However, it has been filled with ‘N/A’.

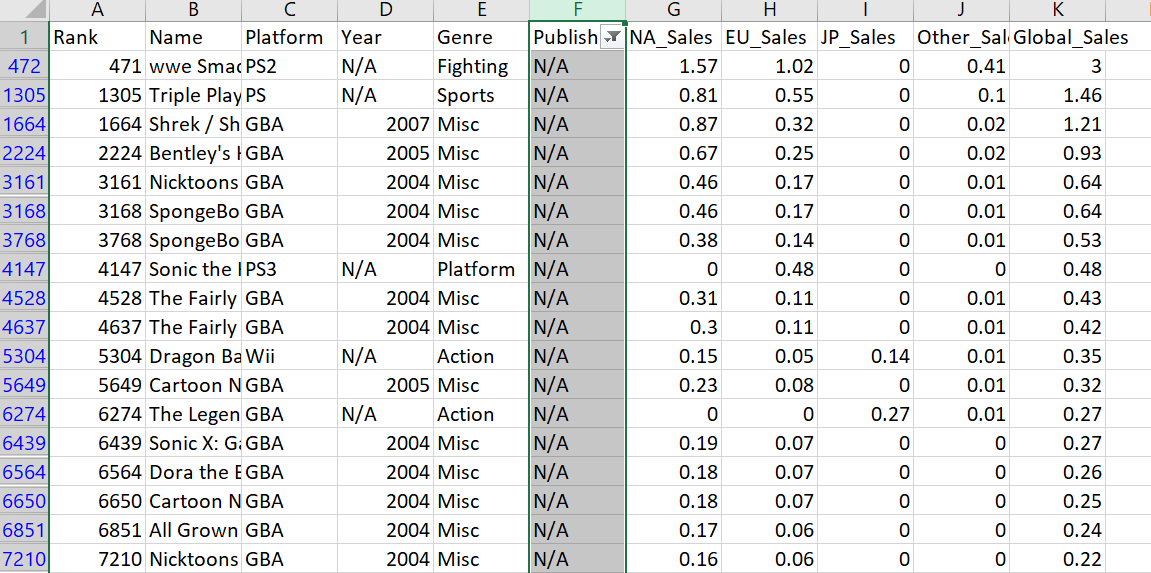


Figure 2 showing Null updated with 'N/A' in Publisher field

* The ‘year’ column also has NULL fields, it is also filled with ‘N/A’.

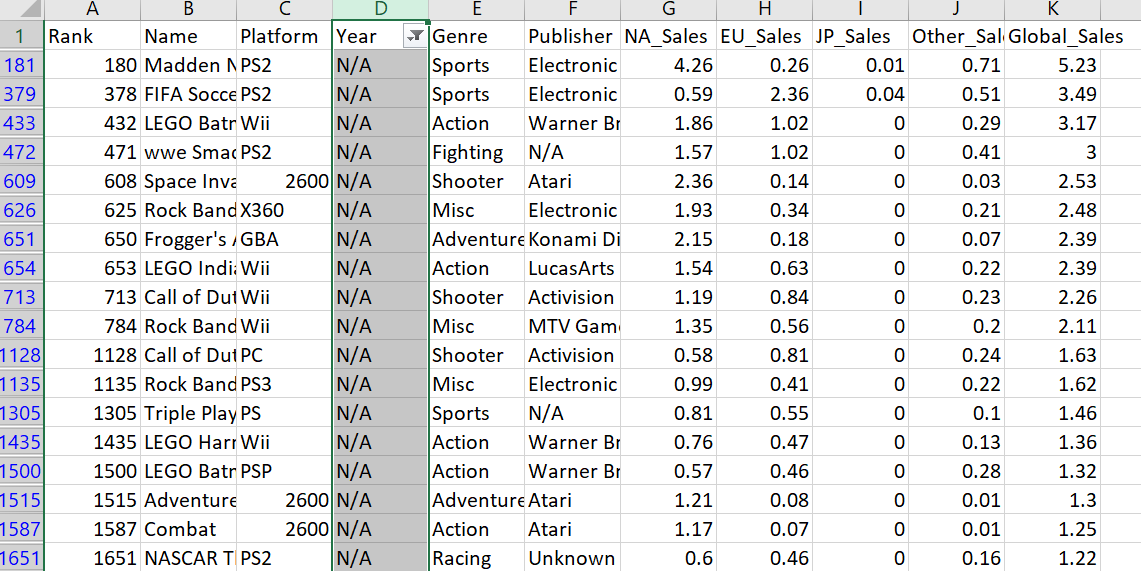


Figure 3 Showing the Null values updated with 'N/A' in Year field.

* No missing values observed in Platform, Genre, NA\_Sales, EU\_Sales, JP\_Sales, Other\_Sales, Global\_sales columns.
* Now I am good to perform the analysis on the data and answer all the data goals.

**Analysis Goals-**

1.In year 2006, Which publisher made lowest global sales?

2. The publisher ‘Take-Two Interactive’ got highest global sales in which year?

3. Which video game release platform made good Sales in North America till date?

4.For different years, Does Global sales closely spread or far apart?

5.What is the average global sales of ‘Grand Theft Auto’ game till date?

6.How many sales are done in North America for game ranked 50?

7.What is the rank of game named ‘Call of Duty 4: Modern Warfare’?

**Results of the analysis-**

**1.In year 2006, Which publisher made lowest global sales?**

* In year 2006, publisher named ‘Michaelsoft’ made lowest global sales.
* Initially used advanced filter option with,
  + List range= Whole data (A1:K16599)
  + Criteria range= Selected the criteria cells (Criteria used is year=2006)
  + Copy to= Cells in the active worksheet. The result of this step is,

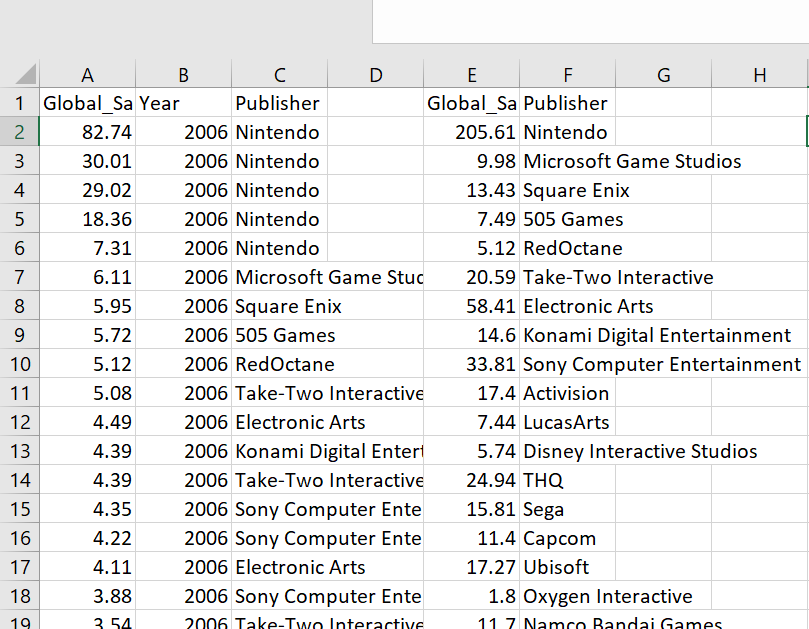
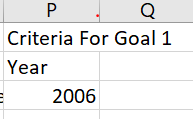
 

Figure 4 Showing the filtered year 2006

* Now to know global sales of each publisher, remove duplicates on publisher and add global sales

E2=SUMIF($C$2:$C$1009,F2:F113,$A$2:$A$1009)

* To know the publisher with lowest global sales in 2006,

I2=VLOOKUP(MIN(E2:E113),E2:F113,2,FALSE)

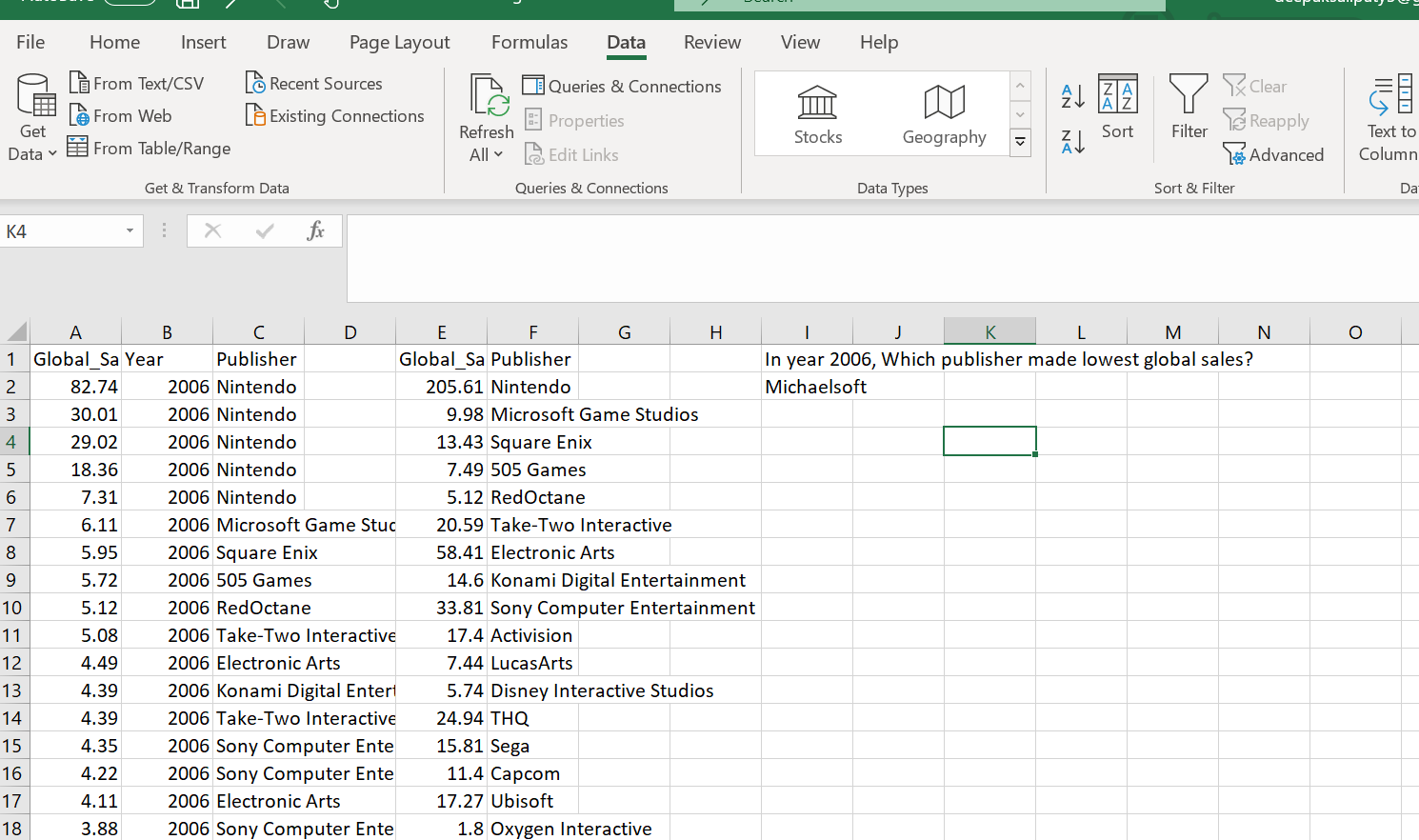


Figure 5 Goal 1 Result

**2. The publisher ‘Take-Two Interactive’ got highest global sales in which year?**

* The publisher ‘Take-Two Interactive’ got highest global sales in year 2013.
* Initially used advanced filter option with,
  + List range=Whole data (A1:K16599)
  + Criteria range= Selected the criteria cells (Criteria used is Publisher= 'Take-Two Interactive)
  + Copy to= Cells in the active worksheet. The result of this step is,

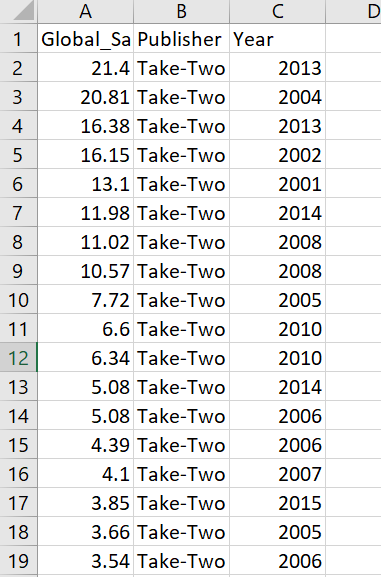
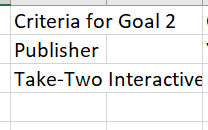
 

Figure 6 Showing the result of advanced filter

* Remove the duplicates on Year field and added global sales.

E2=SUMIF($C$2:$C$414,F2:F22,$A$2:$A$414)

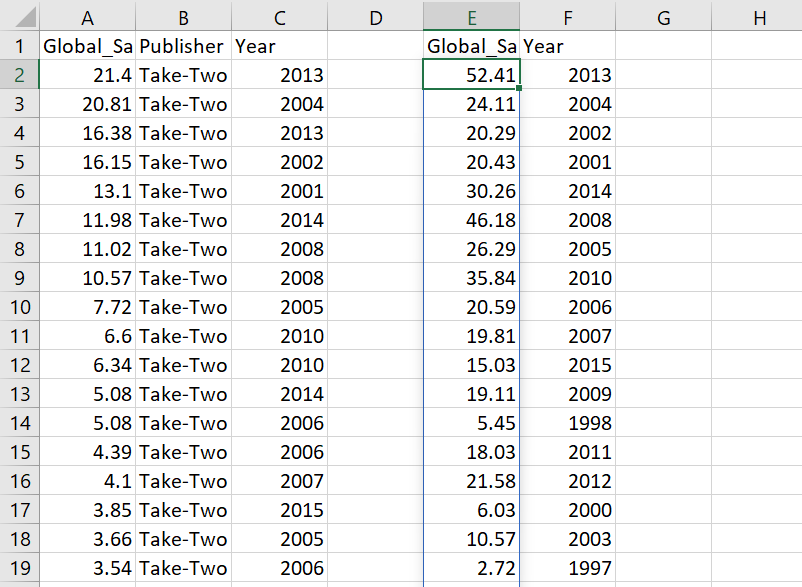


Figure 7 Removed duplicates and added global sales

* Now to know the year of highest global sales of ‘Take two Interactive’,

I2=VLOOKUP(MAX(E2:E22),E2:F22,2,FALSE)

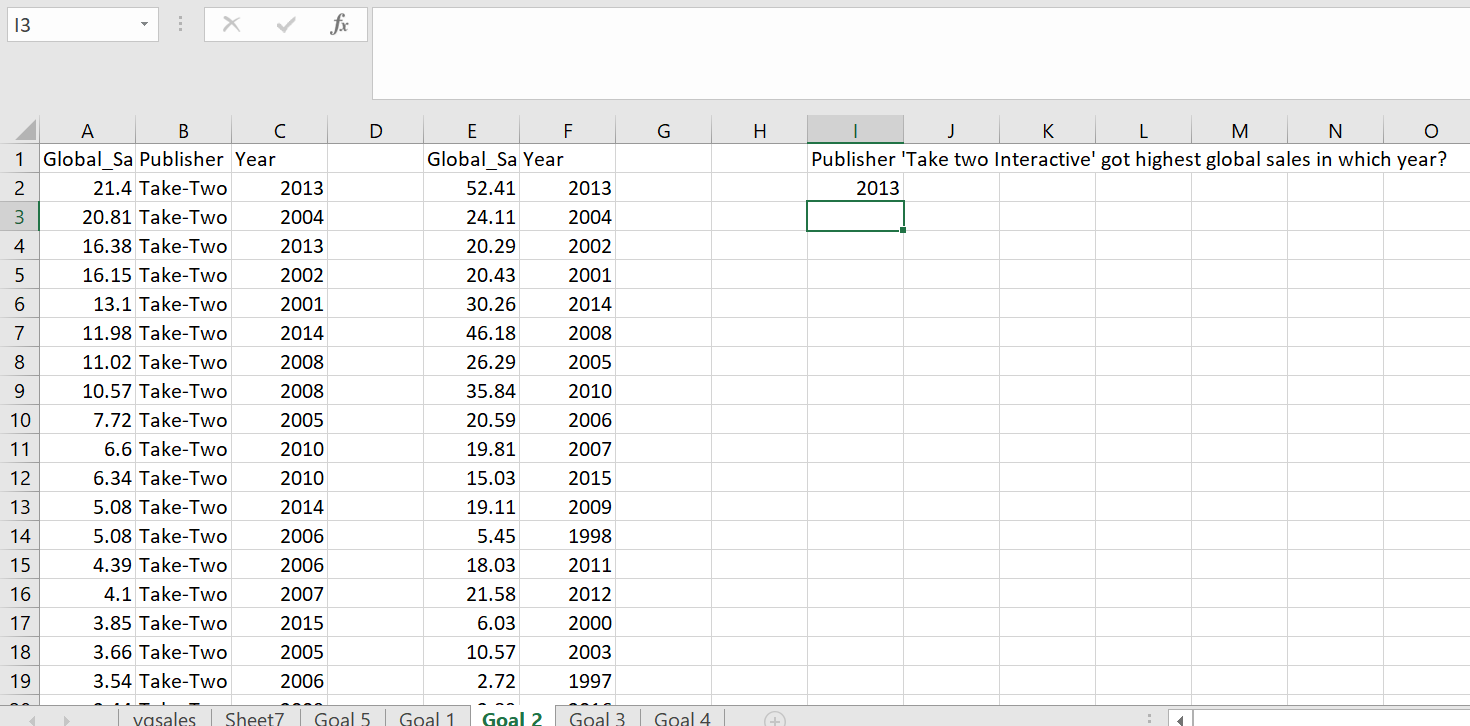


Figure 8 Result of Goal 2

**3. Which video game release platform made good Sales in North America till date?**

* ‘X360’ platform made good sales in North America till date.
* Initially used advanced filter option with,
  + List range= Whole data (A1:K16599)
  + Criteria range= No criteria
  + Copy to= Cells in the active worksheet. The result of this step is,

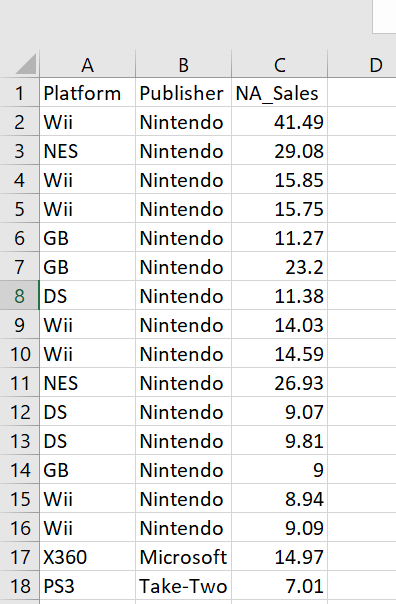


Figure 9 Shows Platform, publisher, and NA sales

* Removed duplicates on platform and added North America sales.

E2=SUMIF($A$2:$A$16599,F2:F32,C2:C16599)

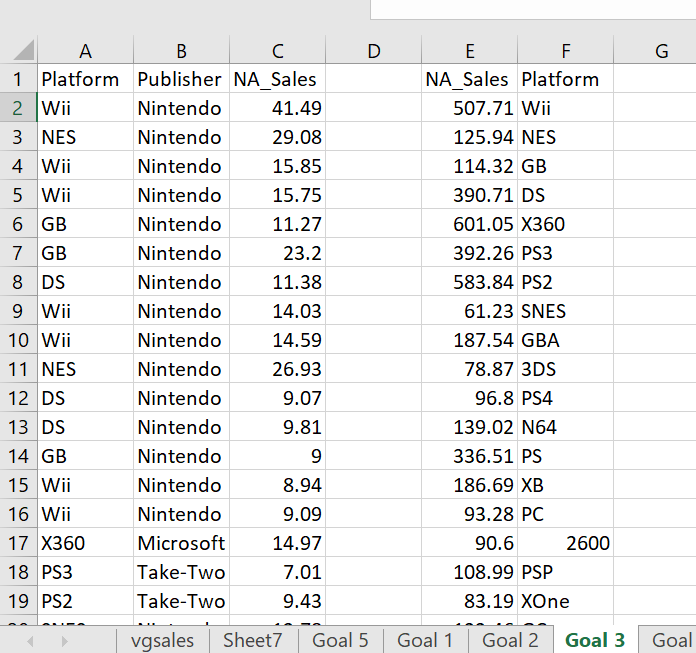


Figure 10 Removed duplicates on platform and added NA\_Sales

* Now, In H2= VLOOKUP(MAX(E2:E32),E2:F32,2,FALSE)

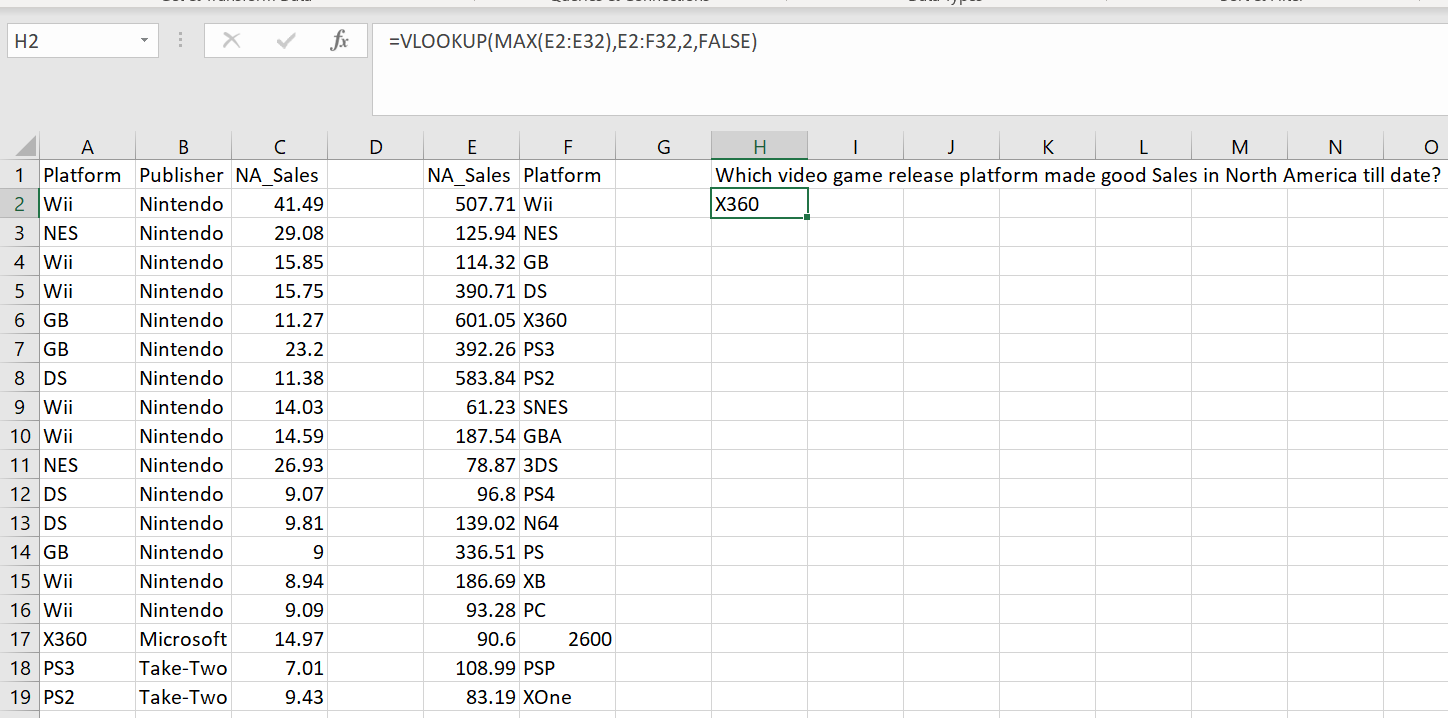


Figure 11 Result of goal 3

**4.For different years, Does Global sales closely spread or far apart?**

* The global sales value for different years are spread far apart.
* Initially used advanced filter option with,
  + List range= Whole data (A1:K16599)
  + Criteria range= No criteria
  + Copy to= Cells in the active worksheet. The result of this step is,

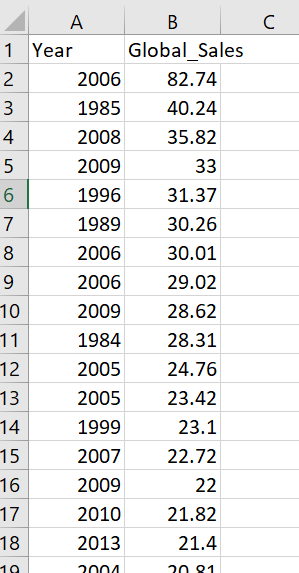


Figure 12 Showing result of advanced filter

* Removed duplicates on year and added global sales.

F2=SUMIF(A2:A16599,E2:E41,B2:B16599)

K2= MAX(F2:F41)

K3=MIN(F2:F41)

K4=k2-k3 Here range value is calculated.

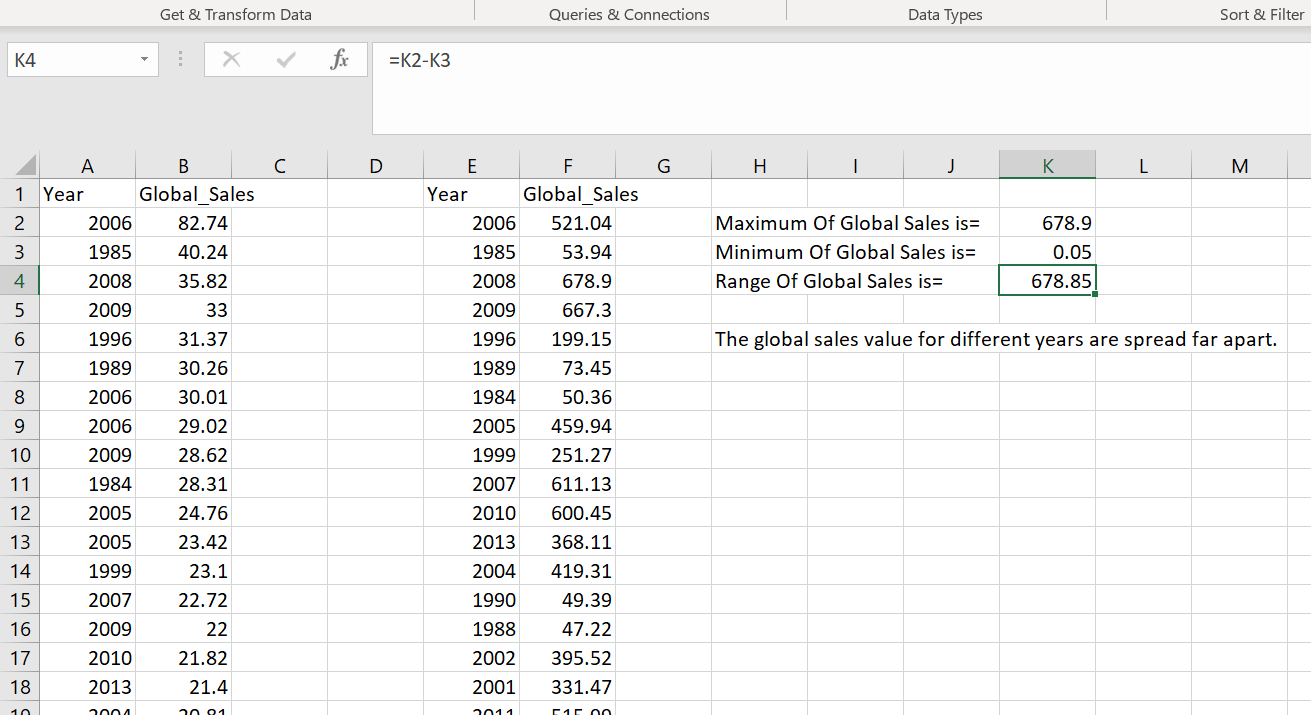


Figure 13 Result of Goal 4

**5.What is the average global sales of ‘Grand Theft Auto’ game till date?**

* Average global sales of different versions of ‘Grand Theft Auto’ game is ‘6.053077’.
* Initially used advanced filter option with,
  + List range= Whole data (A1:K16599)
  + Criteria range= selected criteria cells (Name=grand theft Auto, Global\_Sales=Null).
  + Copy to= Cells in the active worksheet. The result of this step is,

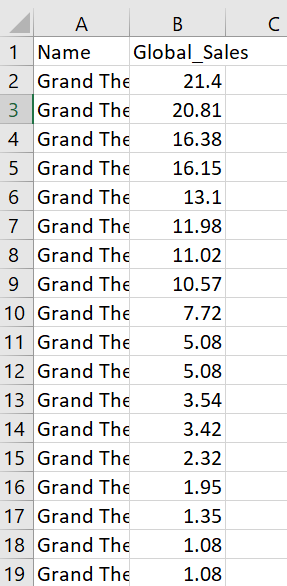
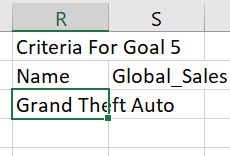
 

Figure Result of advanced filter option

* Now F3= AVERAGE(B2:B27)

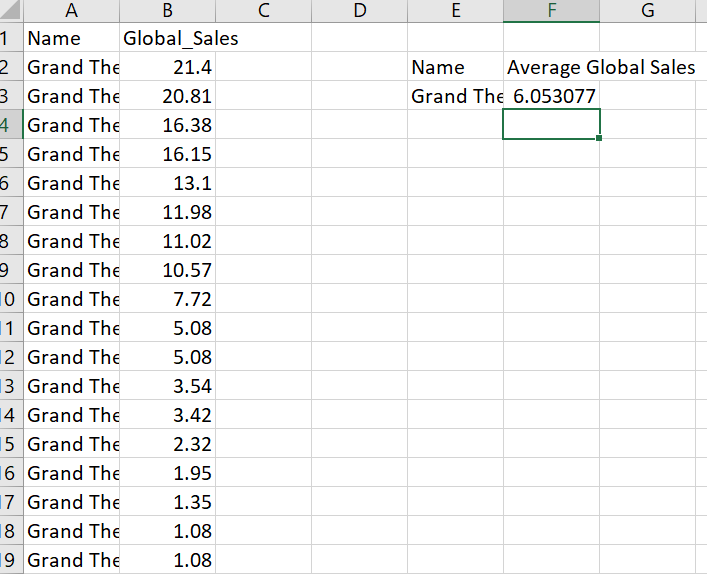


Figure Result of Goal 5

**6.How many sales are done in North America for game ranked 50?**

* Sales done in North America for game ranked 50 is 4.23 million.

N3=HLOOKUP(M3,A1:K100,51,FALSE)

**7.What is the rank of game named ‘Call of Duty 4: Modern Warfare’?**

* Rank of game named ‘Call of Duty: Modern Warfare’ is 71.

N6= MATCH("Call of Duty 4: Modern Warfare",B2:B16599,0)

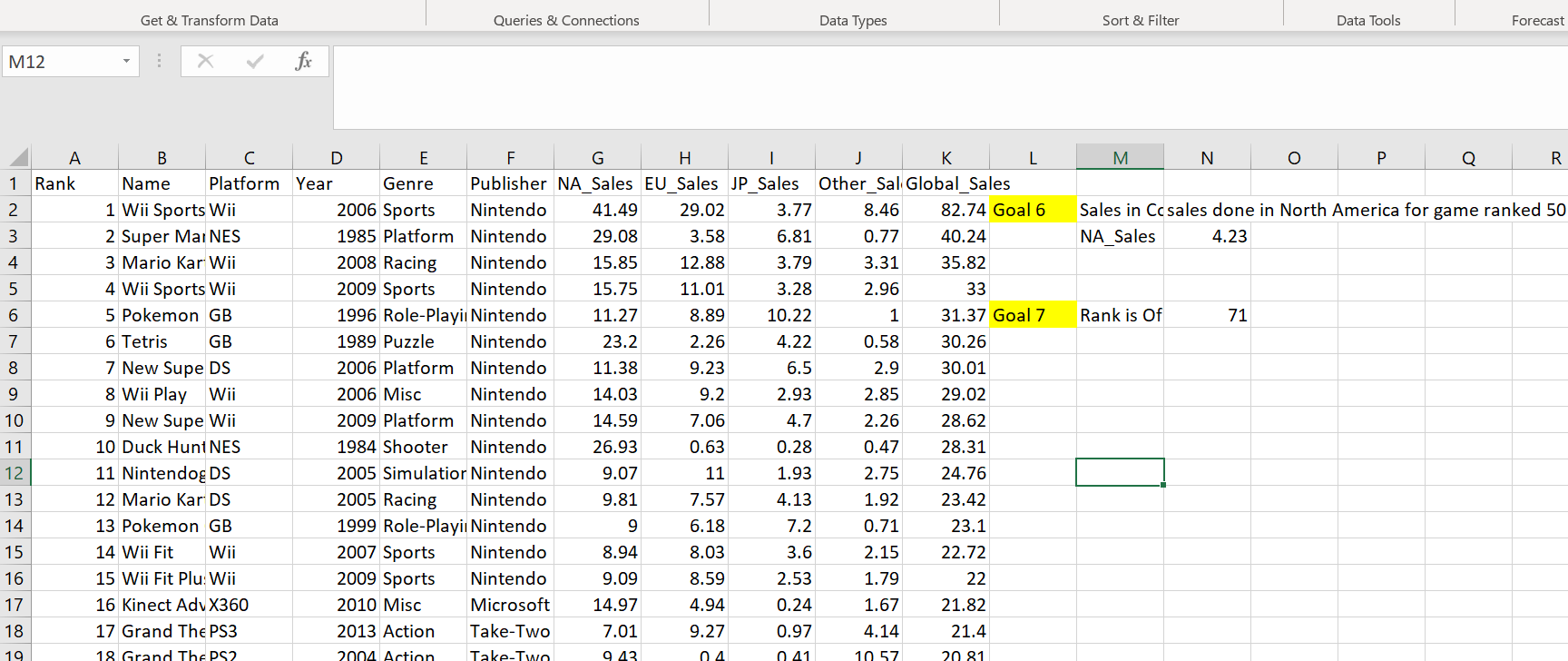


Figure 16 Showing the results of Goal 6 and 7