VIDEO GAME SALES ANALYSIS USING PYTHON

Prepared by,

Swarnim Jambhule,

Neel Savla.

California State University, Los Angeles.

A. Data set:

https://www.kaggle.com/gregorut/videogamesales/version/2#vgsales.csv/

Data description:

This dataset contains a list of video games with sales greater than 100,000 copies.

Rank - Ranking of overall sales

- **Name** - The games name
- **Platform** - Platform of the games release (i.e. PC, PS4, etc.)
- **Year** - Year of the game's release
- **Genre** - Genre of the game
- **Publisher** - Publisher of the game
- **NA_Sales** - Sales in North America (in millions)
- **EU_Sales** - Sales in Europe (in millions)
- **JP_Sales** - Sales in Japan (in millions)
- **Other_Sales** - Sales in the rest of the world (in millions)
- **Global Sales** - Total worldwide sales.

Field Name	Type	Sample Value	Range of values	Attributes/Format	Comments
Year	Textual	1-01-1994	1-01-1985 to 1-01-2016	dd-mm-yyyy	OUT OF DATE
rank	Quantitative	3526	1 to 9303	Whole numbers	Some erroneous values
Name	Textual	FIFA 14	NA	Textual FORMAT	
PLATFORM	Textual	X360	NA	Textual FORMAT	MISSING VALUES
PUBLISHER	Textual	UBisoft	NA	Textual FORMAT	
Genre	Textual	Action	NA	Textual FORMAT	
NA_SALES	Quantitative	3.19 MILLION	0.1 - 41.49 (MILLION)	Decimal numbers	Shows no of active users

EU_SALES	Quantitative	0.92 MILLION	0.1 – 29.02 (MILLION)	Decimal numbers	Some values missing
JP_SALES	Quantitative	0.01 MILLION	0.1 – 10.22 (MILLION)	Decimal numbers	Some values missing
OTHER_SALES	Quantitative	0.42 MILLION	0.1 – 10.57 (MILLION)	Decimal numbers	MISSING VALUES
GLOBAL_SALES	Quantitative	4.53 MILLION	0.05 – 82.75 (MILLION)	Decimal numbers	MISSING VALUES

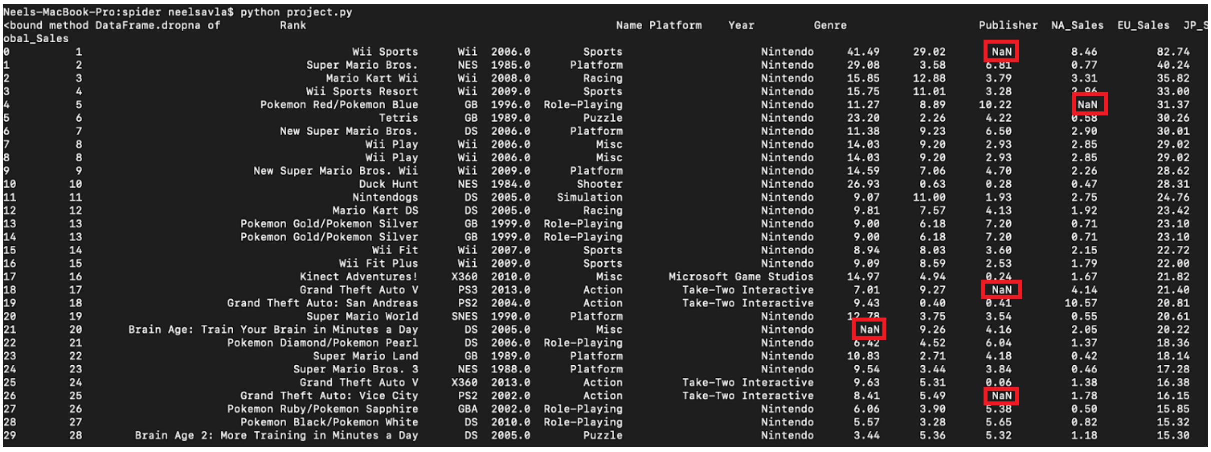
B. Data cleaning

1. Removing rows with null values:

File contained Nan values as shown below in before screenshot, these Nan values was not in one single column but multiple columns through out the data set. To get rid of this null values we used **Drop function** and along with this function we have used **File function**.

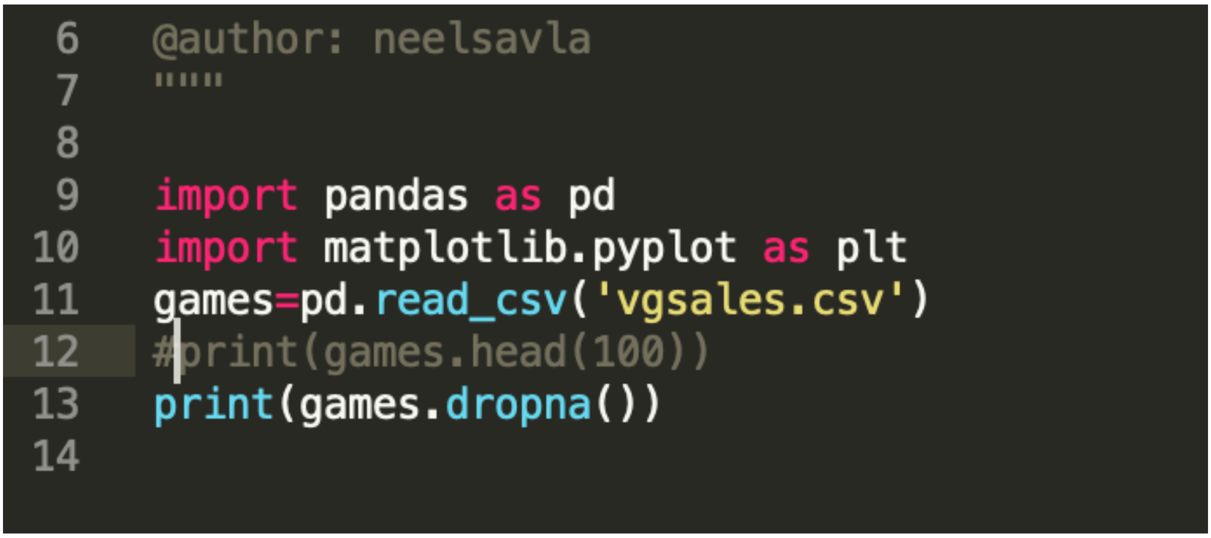
The second screenshot displays the data set without Null values, all the Nan/null values in all of the column are been removed.

Before:



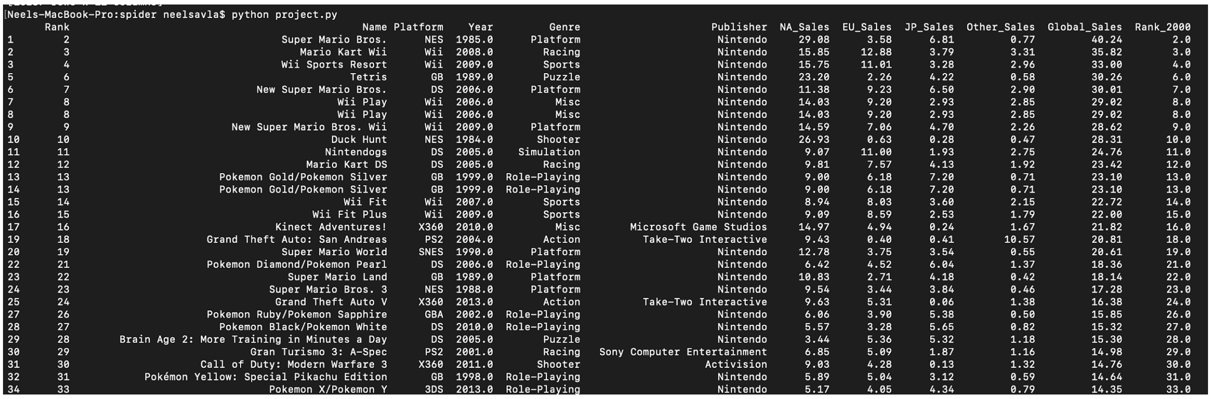
```
Neels-MacBook-Pro:spider neelsavla$ python project.py
<bound method DataFrame.dropna of         Rank                                       Name Platform    Year        Genre                  Publisher  NA_Sales  EU_Sales  JP_S
obal_Sales
0          1                                 Wii Sports      Wii  2006.0       Sports                   Nintendo     41.49     29.02      NaN       8.46      82.74
1          2                          Super Mario Bros.      NES  1985.0     Platform                   Nintendo     29.08      3.58     6.81       0.77      40.24
2          3                             Mario Kart Wii      Wii  2008.0       Racing                   Nintendo     15.85     12.88     3.79       3.31      35.82
3          4                          Wii Sports Resort      Wii  2009.0       Sports                   Nintendo     15.75     11.01     3.28       2.96      33.00
4          5                   Pokemon Red/Pokemon Blue       GB  1996.0  Role-Playing                  Nintendo     11.27      8.89    10.22        NaN      31.37
5          6                                     Tetris       GB  1989.0       Puzzle                   Nintendo     23.20      2.26     4.22       0.58      30.26
6          7                      New Super Mario Bros.       DS  2006.0     Platform                   Nintendo     11.38      9.23     6.50       2.90      30.01
7          8                                   Wii Play      Wii  2006.0         Misc                   Nintendo     14.03      9.20     2.93       2.85      29.02
8          8                                   Wii Play      Wii  2006.0         Misc                   Nintendo     14.03      9.20     2.93       2.85      29.02
9          9                  New Super Mario Bros. Wii      Wii  2009.0     Platform                   Nintendo     14.59      7.06     4.70       2.26      28.62
10        10                                  Duck Hunt      NES  1984.0      Shooter                   Nintendo     26.93      0.63     0.28       0.47      28.31
11        11                                 Nintendogs       DS  2005.0   Simulation                   Nintendo      9.07     11.00     1.93       2.75      24.76
12        12                              Mario Kart DS       DS  2005.0       Racing                   Nintendo      9.81      7.57     4.13       1.92      23.42
13        13                Pokemon Gold/Pokemon Silver       GB  1999.0  Role-Playing                  Nintendo      9.00      6.18     7.20       0.71      23.10
14        13                Pokemon Gold/Pokemon Silver       GB  1999.0  Role-Playing                  Nintendo      9.00      6.18     7.20       0.71      23.10
15        14                                    Wii Fit      Wii  2007.0       Sports                   Nintendo      8.94      8.03     3.60       2.15      22.72
16        15                               Wii Fit Plus      Wii  2009.0       Sports                   Nintendo      9.09      8.59     2.53       1.79      22.00
17        16                         Kinect Adventures!     X360  2010.0         Misc     Microsoft Game Studios     14.97      4.94     0.24       1.67      21.82
18        17                         Grand Theft Auto V      PS3  2013.0       Action       Take-Two Interactive      7.01      9.27      NaN       4.14      21.40
19        18              Grand Theft Auto: San Andreas      PS2  2004.0       Action       Take-Two Interactive      9.43      0.40     0.41      10.57      20.81
20        19                          Super Mario World     SNES  1990.0     Platform                   Nintendo     12.78      3.75     3.54       0.55      20.61
21        20   Brain Age: Train Your Brain in Minutes a Day       DS  2005.0         Misc                   Nintendo       NaN      9.26     4.16       2.05      20.22
22        21              Pokemon Diamond/Pokemon Pearl       DS  2006.0  Role-Playing                  Nintendo      6.42      4.52     6.04       1.37      18.36
23        22                           Super Mario Land       GB  1989.0     Platform                   Nintendo     10.83      2.71     4.18       0.42      18.14
24        23                        Super Mario Bros. 3      NES  1988.0     Platform                   Nintendo      9.54      3.44     3.84       0.46      17.28
25        24                         Grand Theft Auto V     X360  2013.0       Action       Take-Two Interactive      9.63      5.31     0.06       1.38      16.38
26        25                Grand Theft Auto: Vice City      PS2  2002.0       Action       Take-Two Interactive      8.41      5.49      NaN       1.78      16.15
27        26              Pokemon Ruby/Pokemon Sapphire      GBA  2002.0  Role-Playing                  Nintendo      6.06      3.90     5.38       0.50      15.85
28        27                Pokemon Black/Pokemon White       DS  2010.0  Role-Playing                  Nintendo      5.57      3.28     5.65       0.82      15.32
29        28    Brain Age 2: More Training in Minutes a Day       DS  2005.0       Puzzle                   Nintendo      3.44      5.36     5.32       1.18      15.30
```

code:



```
@author: neelsavla
"""

import pandas as pd
import matplotlib.pyplot as plt
games=pd.read_csv('vgsales.csv')
#print(games.head(100))
print(games.dropna())
```

After:

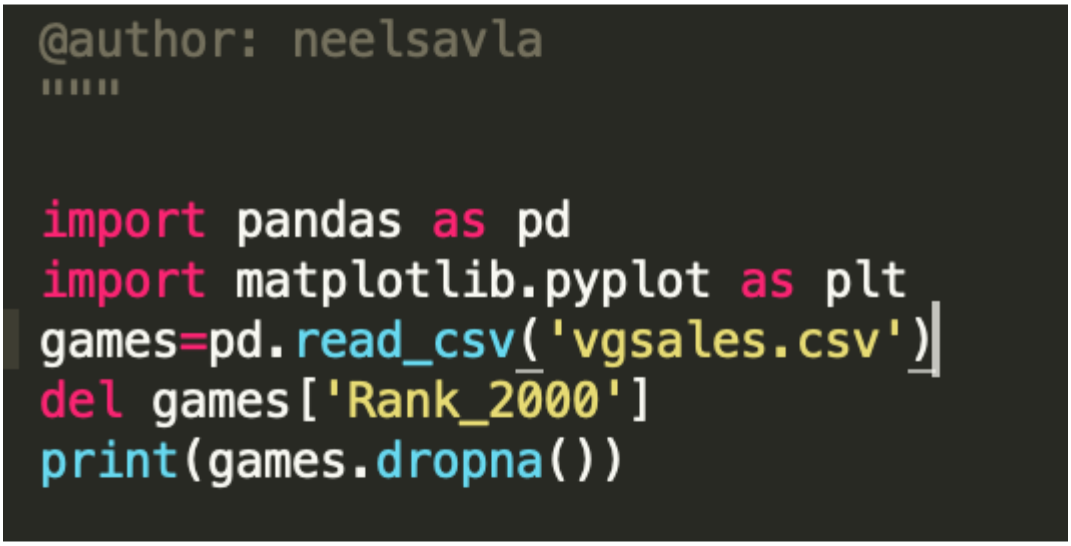


```
[Neels-MacBook-Pro:spider neelsavla$ python project.py
      Rank                                         Name Platform   Year        Genre                   Publisher NA_Sales EU_Sales JP_Sales Other_Sales Global_Sales Rank_2000
1        2                            Super Mario Bros.      NES 1985.0     Platform                    Nintendo    29.08     3.58     6.81        0.77        40.24       2.0
2        3                               Mario Kart Wii      Wii 2008.0       Racing                    Nintendo    15.85    12.88     3.79        3.31        35.82       3.0
3        4                            Wii Sports Resort      Wii 2009.0       Sports                    Nintendo    15.75    11.01     3.28        2.96        33.00       4.0
5        6                                       Tetris       GB 1989.0       Puzzle                    Nintendo    23.20     2.26     4.22        0.58        30.26       6.0
6        7                        New Super Mario Bros.       DS 2006.0     Platform                    Nintendo    11.38     9.23     6.50        2.90        30.01       7.0
7        8                                     Wii Play      Wii 2006.0         Misc                    Nintendo    14.03     9.20     2.93        2.85        29.02       8.0
8        8                                     Wii Play      Wii 2006.0         Misc                    Nintendo    14.03     9.20     2.93        2.85        29.02       8.0
9        9                    New Super Mario Bros. Wii      Wii 2009.0     Platform                    Nintendo    14.59     7.06     4.70        2.26        28.62       9.0
10      10                                    Duck Hunt      NES 1984.0      Shooter                    Nintendo    26.93     0.63     0.28        0.47        28.31      10.0
11      11                                   Nintendogs       DS 2005.0   Simulation                    Nintendo     9.07    11.00     1.93        2.75        24.76      11.0
12      12                                Mario Kart DS       DS 2005.0       Racing                    Nintendo     9.81     7.57     4.13        1.92        23.42      12.0
13      13                  Pokemon Gold/Pokemon Silver       GB 1999.0 Role-Playing                    Nintendo     9.00     6.18     7.20        0.71        23.10      13.0
14      13                  Pokemon Gold/Pokemon Silver       GB 1999.0 Role-Playing                    Nintendo     9.00     6.18     7.20        0.71        23.10      13.0
15      14                                      Wii Fit      Wii 2007.0       Sports                    Nintendo     8.94     8.03     3.60        2.15        22.72      14.0
16      15                                 Wii Fit Plus      Wii 2009.0       Sports                    Nintendo     9.09     8.59     2.53        1.79        22.00      15.0
17      16                           Kinect Adventures!     X360 2010.0         Misc      Microsoft Game Studios    14.97     4.94     0.24        1.67        21.82      16.0
19      18                Grand Theft Auto: San Andreas      PS2 2004.0       Action        Take-Two Interactive     9.43     0.40     0.41       10.57        20.81      18.0
20      19                            Super Mario World     SNES 1990.0     Platform                    Nintendo    12.78     3.75     3.54        0.55        20.61      19.0
22      21                Pokemon Diamond/Pokemon Pearl       DS 2006.0 Role-Playing                    Nintendo     6.42     4.52     6.04        1.37        18.36      21.0
23      22                             Super Mario Land       GB 1989.0     Platform                    Nintendo    10.83     2.71     4.18        0.42        18.14      22.0
24      23                          Super Mario Bros. 3      NES 1988.0     Platform                    Nintendo     9.54     3.44     3.84        0.46        17.28      23.0
25      24                           Grand Theft Auto V     X360 2013.0       Action        Take-Two Interactive     9.63     5.31     0.06        1.38        16.38      24.0
27      26                Pokemon Ruby/Pokemon Sapphire      GBA 2002.0 Role-Playing                    Nintendo     6.06     3.90     5.38        0.50        15.85      26.0
28      27                  Pokemon Black/Pokemon White       DS 2010.0 Role-Playing                    Nintendo     5.57     3.28     5.65        0.82        15.32      27.0
29      28  Brain Age 2: More Training in Minutes a Day       DS 2005.0       Puzzle                    Nintendo     3.44     5.36     5.32        1.18        15.30      28.0
30      29                       Gran Turismo 3: A-Spec      PS2 2001.0       Racing Sony Computer Entertainment     6.85     5.09     1.87        1.16        14.98      29.0
31      30               Call of Duty: Modern Warfare 3     X360 2011.0      Shooter                  Activision     9.03     4.28     0.13        1.32        14.76      30.0
32      31      Pokémon Yellow: Special Pikachu Edition       GB 1998.0 Role-Playing                    Nintendo     5.89     5.04     3.12        0.59        14.64      31.0
34      33                          Pokemon X/Pokemon Y      3DS 2013.0 Role-Playing                    Nintendo     5.17     4.05     4.34        0.79        14.35      33.0
```

2. Remove irrelevant columns:

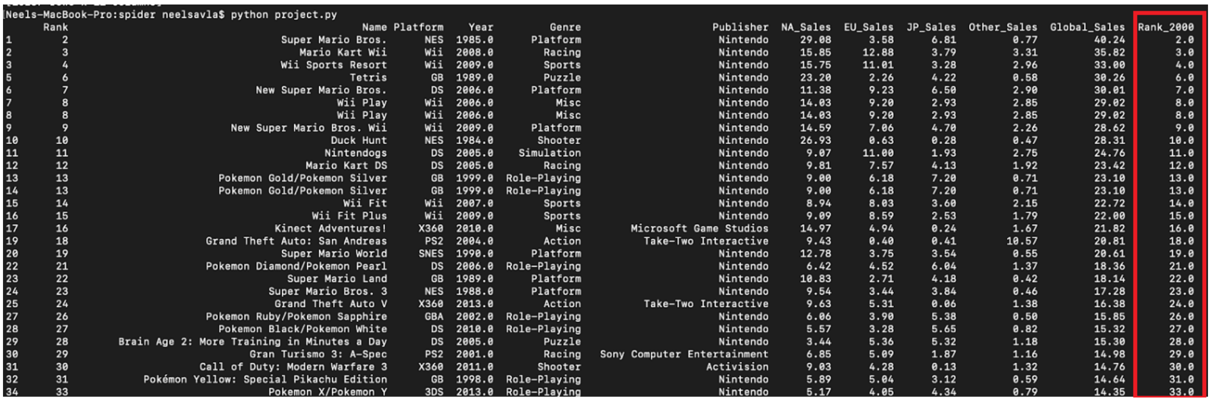
In the data set their was an irrelevant column which was the copy of the first column ‘Rank’ named differently as ‘Rank_2000’ which was not of any use, thus, the column was removed using **del function**, which is a inbuilt function used to delete the column.

The code we used is del games[‘Rank_2000’], here the game is the instance where we stored the csv and Rank_2000 is the column to be deleted and finally used del function to delete the irrelevant column from our data set. We have used **File function** for this code



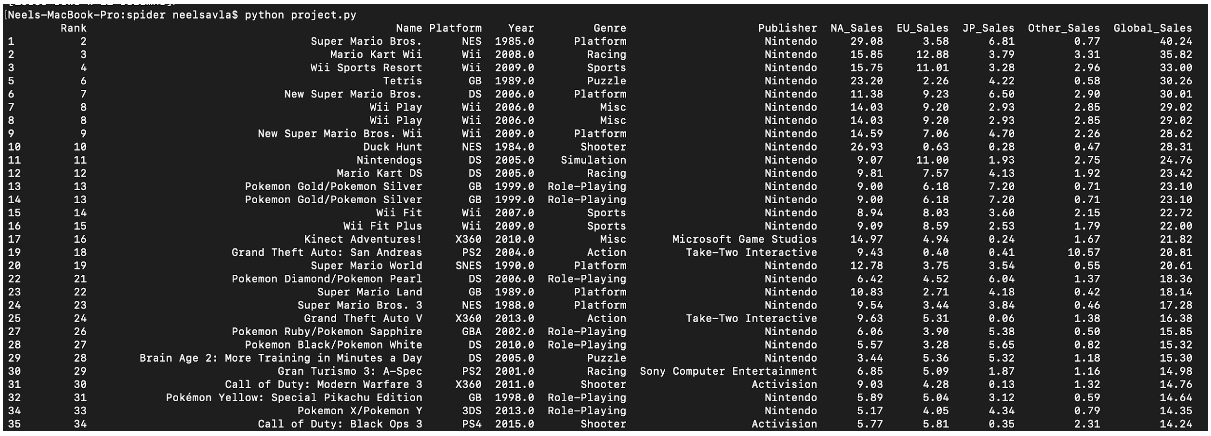
```
@author: neelsavla
"""


import pandas as pd
import matplotlib.pyplot as plt
games=pd.read_csv('vgsales.csv')
del games['Rank_2000']
print(games.dropna())
```

Before:



```
[Neels-MacBook-Pro:spider neelsavla$ python project.py
      Rank                                         Name Platform    Year        Genre                    Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales  Rank_2000
1        2                            Super Mario Bros.      NES  1985.0     Platform                     Nintendo     29.08      3.58      6.81         0.77         40.24        2.0
2        3                               Mario Kart Wii      Wii  2008.0       Racing                     Nintendo     15.85     12.88      3.79         3.31         35.82        3.0
3        4                            Wii Sports Resort      Wii  2009.0       Sports                     Nintendo     15.75     11.01      3.28         2.96         33.00        4.0
5        6                                       Tetris       GB  1989.0       Puzzle                     Nintendo     23.20      2.26      4.22         0.58         30.26        6.0
6        7                        New Super Mario Bros.       DS  2006.0     Platform                     Nintendo     11.38      9.23      6.50         2.90         30.01        7.0
7        8                                     Wii Play      Wii  2006.0         Misc                     Nintendo     14.03      9.20      2.93         2.85         29.02        8.0
8        8                                     Wii Play      Wii  2006.0         Misc                     Nintendo     14.03      9.20      2.93         2.85         29.02        8.0
9        9                    New Super Mario Bros. Wii      Wii  2009.0     Platform                     Nintendo     14.59      7.06      4.70         2.26         28.62        9.0
10      10                                    Duck Hunt      NES  1984.0      Shooter                     Nintendo     26.93      0.63      0.28         0.47         28.31       10.0
11      11                                   Nintendogs       DS  2005.0   Simulation                     Nintendo      9.07     11.00      1.93         2.75         24.76       11.0
12      12                                Mario Kart DS       DS  2005.0       Racing                     Nintendo      9.81      7.57      4.13         1.92         23.42       12.0
13      13                  Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                     Nintendo      9.00      6.18      7.20         0.71         23.10       13.0
14      13                  Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                     Nintendo      9.00      6.18      7.20         0.71         23.10       13.0
15      14                                      Wii Fit      Wii  2007.0       Sports                     Nintendo      8.94      8.03      3.60         2.15         22.72       14.0
16      15                                 Wii Fit Plus      Wii  2009.0       Sports                     Nintendo      9.09      8.59      2.53         1.79         22.00       15.0
17      16                           Kinect Adventures!     X360  2010.0         Misc       Microsoft Game Studios     14.97      4.94      0.24         1.67         21.82       16.0
19      18                Grand Theft Auto: San Andreas      PS2  2004.0       Action         Take-Two Interactive      9.43      0.40      0.41        10.57         20.81       18.0
20      19                            Super Mario World     SNES  1990.0     Platform                     Nintendo     12.78      3.75      3.54         0.55         20.61       19.0
22      21                Pokemon Diamond/Pokemon Pearl       DS  2006.0 Role-Playing                     Nintendo      6.42      4.52      6.04         1.37         18.36       21.0
23      22                             Super Mario Land       GB  1989.0     Platform                     Nintendo     10.83      2.71      4.18         0.42         18.14       22.0
24      23                          Super Mario Bros. 3      NES  1988.0     Platform                     Nintendo      9.54      3.44      3.84         0.46         17.28       23.0
25      24                           Grand Theft Auto V     X360  2013.0       Action         Take-Two Interactive      9.63      5.31      0.06         1.38         16.38       24.0
27      26                Pokemon Ruby/Pokemon Sapphire      GBA  2002.0 Role-Playing                     Nintendo      6.06      3.90      5.38         0.50         15.85       26.0
28      27                  Pokemon Black/Pokemon White       DS  2010.0 Role-Playing                     Nintendo      5.57      3.28      5.65         0.82         15.32       27.0
29      28  Brain Age 2: More Training in Minutes a Day       DS  2005.0       Puzzle                     Nintendo      3.44      5.36      5.32         1.18         15.30       28.0
30      29                       Gran Turismo 3: A-Spec      PS2  2001.0       Racing  Sony Computer Entertainment      6.85      5.09      1.87         1.16         14.98       29.0
31      30               Call of Duty: Modern Warfare 3     X360  2011.0      Shooter                   Activision      9.03      4.28      0.13         1.32         14.76       30.0
32      31      Pokémon Yellow: Special Pikachu Edition       GB  1998.0 Role-Playing                     Nintendo      5.89      5.04      3.12         0.59         14.64       31.0
34      33                          Pokemon X/Pokemon Y      3DS  2013.0 Role-Playing                     Nintendo      5.17      4.05      4.34         0.79         14.35       33.0
```

After:



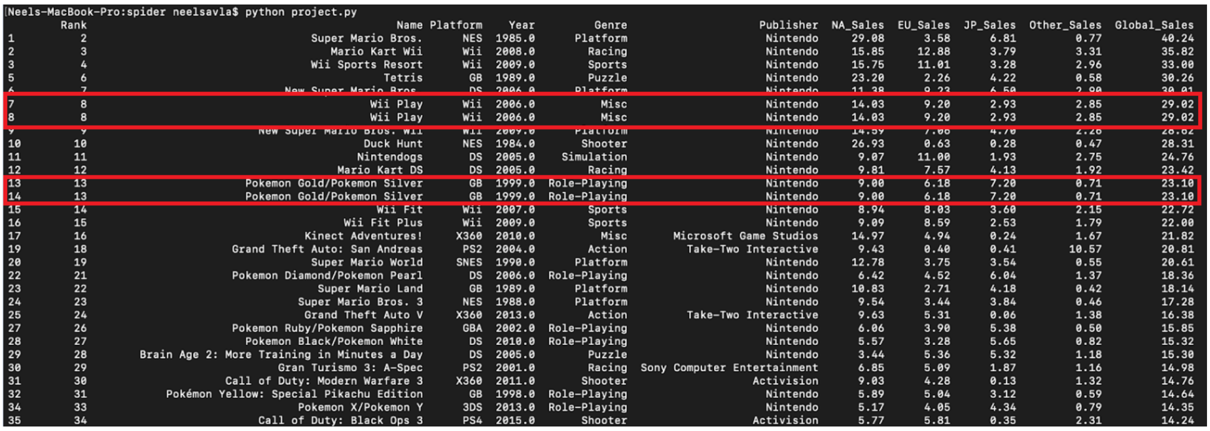
```
[Neels-MacBook-Pro:spider neelsavla$ python project.py
         Rank                                        Name Platform    Year        Genre                      Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
1           2                           Super Mario Bros.      NES  1985.0     Platform                       Nintendo     29.08      3.58      6.81         0.77         40.24
2           3                              Mario Kart Wii      Wii  2008.0       Racing                       Nintendo     15.85     12.88      3.79         3.31         35.82
3           4                           Wii Sports Resort      Wii  2009.0       Sports                       Nintendo     15.75     11.01      3.28         2.96         33.00
5           6                                      Tetris       GB  1989.0       Puzzle                       Nintendo     23.20      2.26      4.22         0.58         30.26
6           7                       New Super Mario Bros.       DS  2006.0     Platform                       Nintendo     11.38      9.23      6.50         2.90         30.01
7           8                                    Wii Play      Wii  2006.0         Misc                       Nintendo     14.03      9.20      2.93         2.85         29.02
8           8                                    Wii Play      Wii  2006.0         Misc                       Nintendo     14.03      9.20      2.93         2.85         29.02
9           9                   New Super Mario Bros. Wii      Wii  2009.0     Platform                       Nintendo     14.59      7.06      4.70         2.26         28.62
10         10                                   Duck Hunt      NES  1984.0      Shooter                       Nintendo     26.93      0.63      0.28         0.47         28.31
11         11                                  Nintendogs       DS  2005.0   Simulation                       Nintendo      9.07     11.00      1.93         2.75         24.76
12         12                               Mario Kart DS       DS  2005.0       Racing                       Nintendo      9.81      7.57      4.13         1.92         23.42
13         13                 Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                       Nintendo      9.00      6.18      7.20         0.71         23.10
14         13                 Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                       Nintendo      9.00      6.18      7.20         0.71         23.10
15         14                                     Wii Fit      Wii  2007.0       Sports                       Nintendo      8.94      8.03      3.60         2.15         22.72
16         15                                Wii Fit Plus      Wii  2009.0       Sports                       Nintendo      9.09      8.59      2.53         1.79         22.00
17         16                          Kinect Adventures!     X360  2010.0         Misc         Microsoft Game Studios     14.97      4.94      0.24         1.67         21.82
19         18               Grand Theft Auto: San Andreas      PS2  2004.0       Action           Take-Two Interactive      9.43      0.40      0.41        10.57         20.81
20         19                           Super Mario World     SNES  1990.0     Platform                       Nintendo     12.78      3.75      3.54         0.55         20.61
22         21               Pokemon Diamond/Pokemon Pearl       DS  2006.0 Role-Playing                       Nintendo      6.42      4.52      6.04         1.37         18.36
23         22                            Super Mario Land       GB  1989.0     Platform                       Nintendo     10.83      2.71      4.18         0.42         18.14
24         23                         Super Mario Bros. 3      NES  1988.0     Platform                       Nintendo      9.54      3.44      3.84         0.46         17.28
25         24                          Grand Theft Auto V     X360  2013.0       Action           Take-Two Interactive      9.63      5.31      0.06         1.38         16.38
27         26               Pokemon Ruby/Pokemon Sapphire      GBA  2002.0 Role-Playing                       Nintendo      6.06      3.90      5.38         0.50         15.85
28         27                 Pokemon Black/Pokemon White       DS  2010.0 Role-Playing                       Nintendo      5.57      3.28      5.65         0.82         15.32
29         28 Brain Age 2: More Training in Minutes a Day       DS  2005.0       Puzzle                       Nintendo      3.44      5.36      5.32         1.18         15.30
30         29                      Gran Turismo 3: A-Spec      PS2  2001.0       Racing    Sony Computer Entertainment      6.85      5.09      1.87         1.16         14.98
31         30              Call of Duty: Modern Warfare 3     X360  2011.0      Shooter                     Activision      9.03      4.28      0.13         1.32         14.76
32         31     Pokémon Yellow: Special Pikachu Edition       GB  1998.0 Role-Playing                       Nintendo      5.89      5.04      3.12         0.59         14.64
34         33                         Pokemon X/Pokemon Y      3DS  2013.0 Role-Playing                       Nintendo      5.17      4.05      4.34         0.79         14.35
35         34                   Call of Duty: Black Ops 3      PS4  2015.0      Shooter                     Activision      5.77      5.81      0.35         2.31         14.24
```

3. Remove duplicate data:

The data set contained multiple duplicate values as show in screenshot for before in the second column called name contains duplicate data such as 'Wii Play' and 'Pokemon Gold/ Pokemon Silver' which are occuring twice in the data set. The column name 'Name' is a column containing unique values thus, it should not have duplicate values. Thus, we are removing the rows with duplicate value occurring only once.

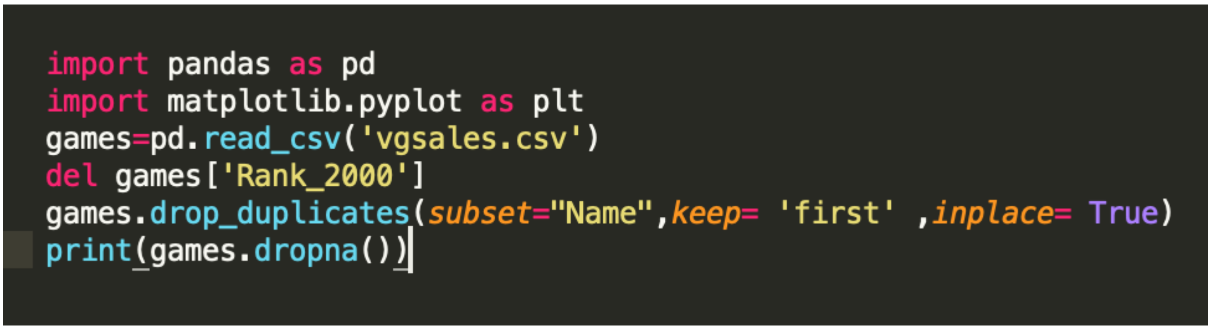
Here we have used '**drop_duplicates**' functions to remove the row with duplicate value. As show in the second screenshot the row with duplicate values is removed. Now the value which was duplicate has only one occurence in the data set making the column unique which is exactly as per the requirement. We have used **Pandas Data Frame** for this code

Before:



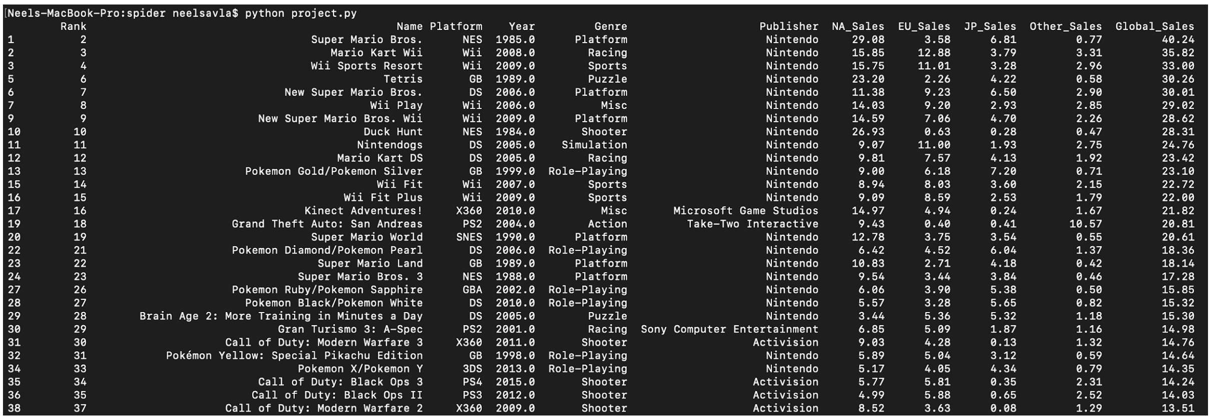
```
[Neels-MacBook-Pro:spider neelsavla$ python project.py
         Rank                                        Name Platform    Year        Genre                      Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
1           2                           Super Mario Bros.      NES  1985.0     Platform                       Nintendo     29.08      3.58      6.81         0.77         40.24
2           3                              Mario Kart Wii      Wii  2008.0       Racing                       Nintendo     15.85     12.88      3.79         3.31         35.82
3           4                           Wii Sports Resort      Wii  2009.0       Sports                       Nintendo     15.75     11.01      3.28         2.96         33.00
5           6                                      Tetris       GB  1989.0       Puzzle                       Nintendo     23.20      2.26      4.22         0.58         30.26
6           7                       New Super Mario Bros.       DS  2006.0     Platform                       Nintendo     11.38      9.23      6.50         2.90         30.01
7           8                                    Wii Play      Wii  2006.0         Misc                       Nintendo     14.03      9.20      2.93         2.85         29.02
8           8                                    Wii Play      Wii  2006.0         Misc                       Nintendo     14.03      9.20      2.93         2.85         29.02
9           9                   New Super Mario Bros. Wii      Wii  2009.0     Platform                       Nintendo     14.59      7.06      4.70         2.26         28.62
10         10                                   Duck Hunt      NES  1984.0      Shooter                       Nintendo     26.93      0.63      0.28         0.47         28.31
11         11                                  Nintendogs       DS  2005.0   Simulation                       Nintendo      9.07     11.00      1.93         2.75         24.76
12         12                               Mario Kart DS       DS  2005.0       Racing                       Nintendo      9.81      7.57      4.13         1.92         23.42
13         13                 Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                       Nintendo      9.00      6.18      7.20         0.71         23.10
14         13                 Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                       Nintendo      9.00      6.18      7.20         0.71         23.10
15         14                                     Wii Fit      Wii  2007.0       Sports                       Nintendo      8.94      8.03      3.60         2.15         22.72
16         15                                Wii Fit Plus      Wii  2009.0       Sports                       Nintendo      9.09      8.59      2.53         1.79         22.00
17         16                          Kinect Adventures!     X360  2010.0         Misc         Microsoft Game Studios     14.97      4.94      0.24         1.67         21.82
19         18               Grand Theft Auto: San Andreas      PS2  2004.0       Action           Take-Two Interactive      9.43      0.40      0.41        10.57         20.81
20         19                           Super Mario World     SNES  1990.0     Platform                       Nintendo     12.78      3.75      3.54         0.55         20.61
22         21               Pokemon Diamond/Pokemon Pearl       DS  2006.0 Role-Playing                       Nintendo      6.42      4.52      6.04         1.37         18.36
23         22                            Super Mario Land       GB  1989.0     Platform                       Nintendo     10.83      2.71      4.18         0.42         18.14
24         23                         Super Mario Bros. 3      NES  1988.0     Platform                       Nintendo      9.54      3.44      3.84         0.46         17.28
25         24                          Grand Theft Auto V     X360  2013.0       Action           Take-Two Interactive      9.63      5.31      0.06         1.38         16.38
27         26               Pokemon Ruby/Pokemon Sapphire      GBA  2002.0 Role-Playing                       Nintendo      6.06      3.90      5.38         0.50         15.85
28         27                 Pokemon Black/Pokemon White       DS  2010.0 Role-Playing                       Nintendo      5.57      3.28      5.65         0.82         15.32
29         28 Brain Age 2: More Training in Minutes a Day       DS  2005.0       Puzzle                       Nintendo      3.44      5.36      5.32         1.18         15.30
30         29                      Gran Turismo 3: A-Spec      PS2  2001.0       Racing    Sony Computer Entertainment      6.85      5.09      1.87         1.16         14.98
31         30              Call of Duty: Modern Warfare 3     X360  2011.0      Shooter                     Activision      9.03      4.28      0.13         1.32         14.76
32         31     Pokémon Yellow: Special Pikachu Edition       GB  1998.0 Role-Playing                       Nintendo      5.89      5.04      3.12         0.59         14.64
34         33                         Pokemon X/Pokemon Y      3DS  2013.0 Role-Playing                       Nintendo      5.17      4.05      4.34         0.79         14.35
35         34                   Call of Duty: Black Ops 3      PS4  2015.0      Shooter                     Activision      5.77      5.81      0.35         2.31         14.24
```

Code:



```
import pandas as pd
import matplotlib.pyplot as plt
games=pd.read_csv('vgsales.csv')
del games['Rank_2000']
games.drop_duplicates(subset="Name",keep= 'first' ,inplace= True)
print(games.dropna())
```

After:

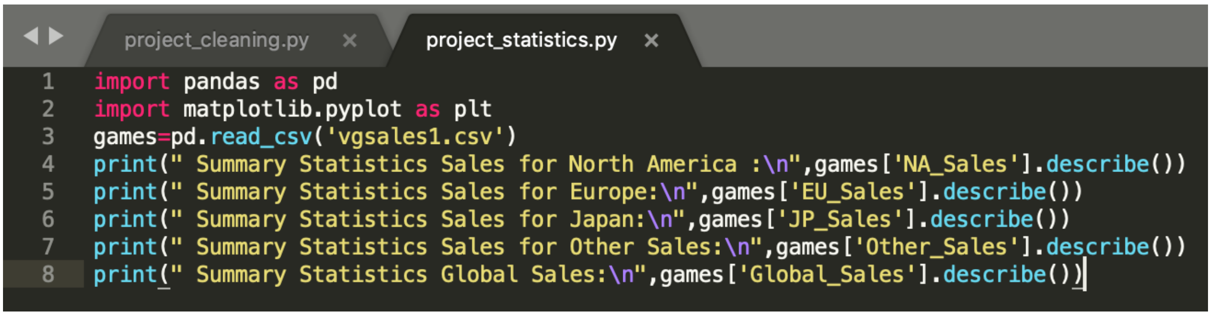


```
Neels-MacBook-Pro:spider neelsavla$ python project.py
      Rank                                        Name Platform    Year        Genre                     Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
1        2                           Super Mario Bros.      NES  1985.0     Platform                      Nintendo     29.08      3.58      6.81         0.77         40.24
2        3                              Mario Kart Wii      Wii  2008.0       Racing                      Nintendo     15.85     12.88      3.79         3.31         35.82
3        4                           Wii Sports Resort      Wii  2009.0       Sports                      Nintendo     15.75     11.01      3.28         2.96         33.00
5        6                                      Tetris       GB  1989.0       Puzzle                      Nintendo     23.20      2.26      4.22         0.58         30.26
6        7                       New Super Mario Bros.       DS  2006.0     Platform                      Nintendo     11.38      9.23      6.50         2.90         30.01
7        8                                    Wii Play      Wii  2006.0         Misc                      Nintendo     14.03      9.20      2.93         2.85         29.02
9        9                   New Super Mario Bros. Wii      Wii  2009.0     Platform                      Nintendo     14.59      7.06      4.70         2.26         28.62
10      10                                   Duck Hunt      NES  1984.0      Shooter                      Nintendo     26.93      0.63      0.28         0.47         28.31
11      11                                  Nintendogs       DS  2005.0   Simulation                      Nintendo      9.07     11.00      1.93         2.75         24.76
12      12                               Mario Kart DS       DS  2005.0       Racing                      Nintendo      9.81      7.57      4.13         1.92         23.42
13      13                 Pokemon Gold/Pokemon Silver       GB  1999.0 Role-Playing                      Nintendo      9.00      6.18      7.20         0.71         23.10
15      14                                     Wii Fit      Wii  2007.0       Sports                      Nintendo      8.94      8.03      3.60         2.15         22.72
16      15                                Wii Fit Plus      Wii  2009.0       Sports                      Nintendo      9.09      8.59      2.53         1.79         22.00
17      16                          Kinect Adventures!     X360  2010.0         Misc        Microsoft Game Studios     14.97      4.94      0.24         1.67         21.82
19      18               Grand Theft Auto: San Andreas      PS2  2004.0       Action          Take-Two Interactive      9.43      0.40      0.41        10.57         20.81
20      19                           Super Mario World     SNES  1990.0     Platform                      Nintendo     12.78      3.75      3.54         0.55         20.61
22      21               Pokemon Diamond/Pokemon Pearl       DS  2006.0 Role-Playing                      Nintendo      6.42      4.52      6.04         1.37         18.36
23      22                            Super Mario Land       GB  1989.0     Platform                      Nintendo     10.83      2.71      4.18         0.42         18.14
24      23                         Super Mario Bros. 3      NES  1988.0     Platform                      Nintendo      9.54      3.44      3.84         0.46         17.28
27      26               Pokemon Ruby/Pokemon Sapphire      GBA  2002.0 Role-Playing                      Nintendo      6.06      3.90      5.38         0.50         15.85
28      27                 Pokemon Black/Pokemon White       DS  2010.0 Role-Playing                      Nintendo      5.57      3.28      5.65         0.82         15.32
29      28 Brain Age 2: More Training in Minutes a Day       DS  2005.0       Puzzle                      Nintendo      3.44      5.36      5.32         1.18         15.30
30      29                      Gran Turismo 3: A-Spec      PS2  2001.0       Racing   Sony Computer Entertainment      6.85      5.09      1.87         1.16         14.98
31      30              Call of Duty: Modern Warfare 3     X360  2011.0      Shooter                    Activision      9.03      4.28      0.13         1.32         14.76
32      31     Pokémon Yellow: Special Pikachu Edition       GB  1998.0 Role-Playing                      Nintendo      5.89      5.04      3.12         0.59         14.64
34      33                         Pokemon X/Pokemon Y      3DS  2013.0 Role-Playing                      Nintendo      5.17      4.05      4.34         0.79         14.35
35      34                   Call of Duty: Black Ops 3      PS4  2015.0      Shooter                    Activision      5.77      5.81      0.35         2.31         14.24
36      35                  Call of Duty: Black Ops II      PS3  2012.0      Shooter                    Activision      4.99      5.88      0.65         2.52         14.03
38      37              Call of Duty: Modern Warfare 2     X360  2009.0      Shooter                    Activision      8.52      3.63      0.08         1.29         13.51
```

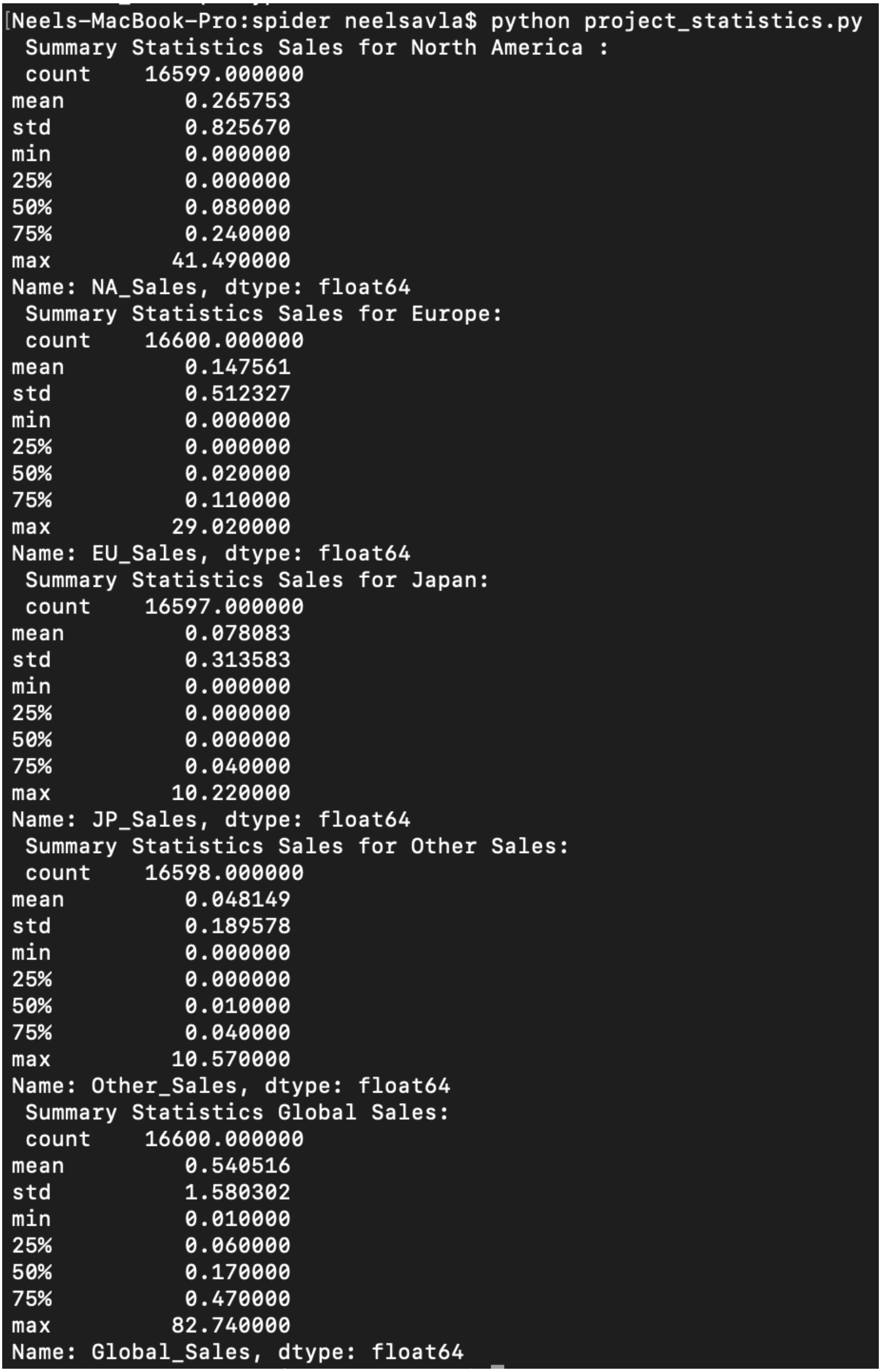
C. Statistical Summary:

To find the statistical summary **.describe()** function is used and is applied to the required columns such as for sales in North America, Sales in Europe, Sales in Japan, Other sales and Global Sales. Using .describe() we found the count, mean, standard deviation, minimum, 25%, 50%, 75% and maximum value for each of the columns specified above.

For example, if we see the Sales for North America in the screenshot below we can see that the count of sale is 16599, mean of sales is 0.265753, standard deviation is 0.825670, minimum sales if 0.0000, 25% of the sales is 0.00000, 50% of the sales is 0.80000, 75% 0f the sales is 0.240000 and finally maximum value of sales for North America is 41.490000.



```python
import pandas as pd
import matplotlib.pyplot as plt
games=pd.read_csv('vgsales1.csv')
print(" Summary Statistics Sales for North America :\n",games['NA_Sales'].describe())
print(" Summary Statistics Sales for Europe:\n",games['EU_Sales'].describe())
print(" Summary Statistics Sales for Japan:\n",games['JP_Sales'].describe())
print(" Summary Statistics Sales for Other Sales:\n",games['Other_Sales'].describe())
print(" Summary Statistics Global Sales:\n",games['Global_Sales'].describe())
```



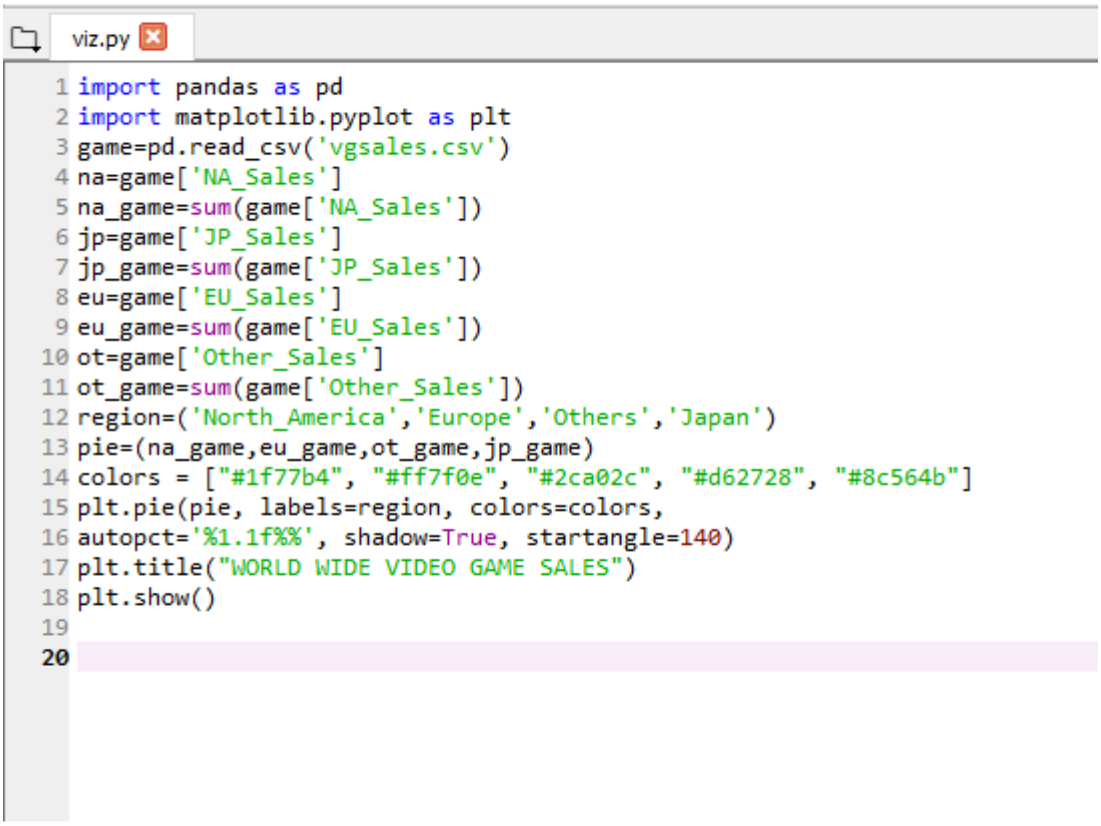
```
[Neels-MacBook-Pro:spider neelsavla$ python project_statistics.py
 Summary Statistics Sales for North America :
 count    16599.000000
mean         0.265753
std          0.825670
min          0.000000
25%          0.000000
50%          0.080000
75%          0.240000
max         41.490000
Name: NA_Sales, dtype: float64
 Summary Statistics Sales for Europe:
 count    16600.000000
mean         0.147561
std          0.512327
min          0.000000
25%          0.000000
50%          0.020000
75%          0.110000
max         29.020000
Name: EU_Sales, dtype: float64
 Summary Statistics Sales for Japan:
 count    16597.000000
mean         0.078083
std          0.313583
min          0.000000
25%          0.000000
50%          0.000000
75%          0.040000
max         10.220000
Name: JP_Sales, dtype: float64
 Summary Statistics Sales for Other Sales:
 count    16598.000000
mean         0.048149
std          0.189578
min          0.000000
25%          0.000000
50%          0.010000
75%          0.040000
max         10.570000
Name: Other_Sales, dtype: float64
 Summary Statistics Global Sales:
 count    16600.000000
mean         0.540516
std          1.580302
min          0.010000
25%          0.060000
50%          0.170000
75%          0.470000
max         82.740000
Name: Global_Sales, dtype: float64
```

D. Data cleaning Analysis & Visualization

Question 1:

What is the Video game sales pattern in entire world?

1) Chart use- Pie Chart



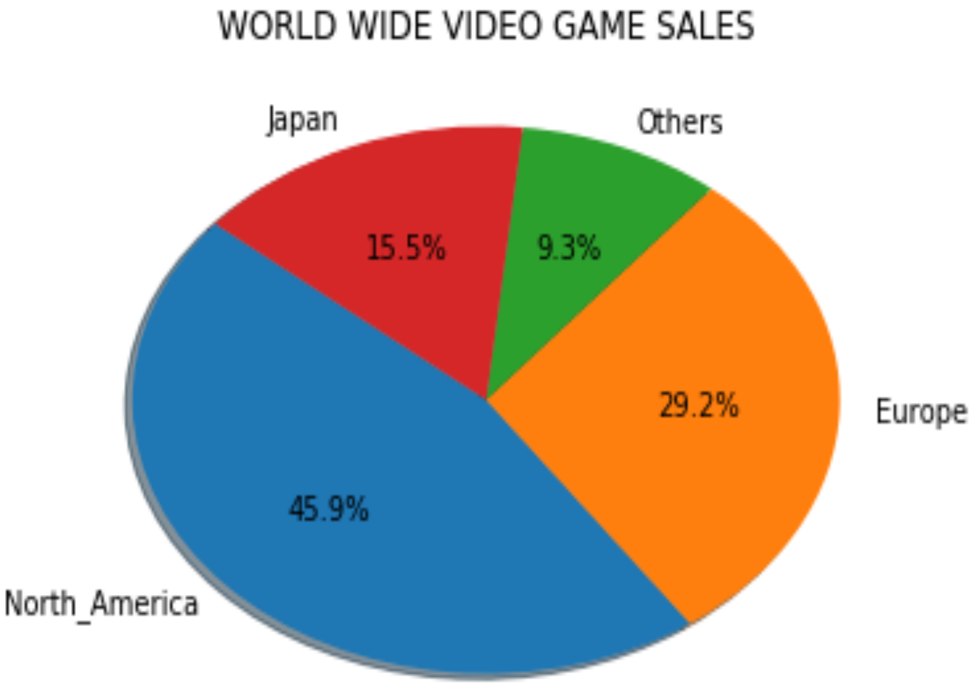
```python
import pandas as pd
import matplotlib.pyplot as plt
game=pd.read_csv('vgsales.csv')
na=game['NA_Sales']
na_game=sum(game['NA_Sales'])
jp=game['JP_Sales']
jp_game=sum(game['JP_Sales'])
eu=game['EU_Sales']
eu_game=sum(game['EU_Sales'])
ot=game['Other_Sales']
ot_game=sum(game['Other_Sales'])
region=('North_America','Europe','Others','Japan')
pie=(na_game,eu_game,ot_game,jp_game)
colors = ["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#8c564b"]
plt.pie(pie, labels=region, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=140)
plt.title("WORLD WIDE VIDEO GAME SALES")
plt.show()


```

To get Insight of the data, we need to compare some important statistic of the data. This Visualization helps us to understand the the global sales pattern. We have divided world into four major region, those region are North America, Japan, Europe and Other part.

Not plotting the **Pie chart** with the attributes of region which is mentioned previously,

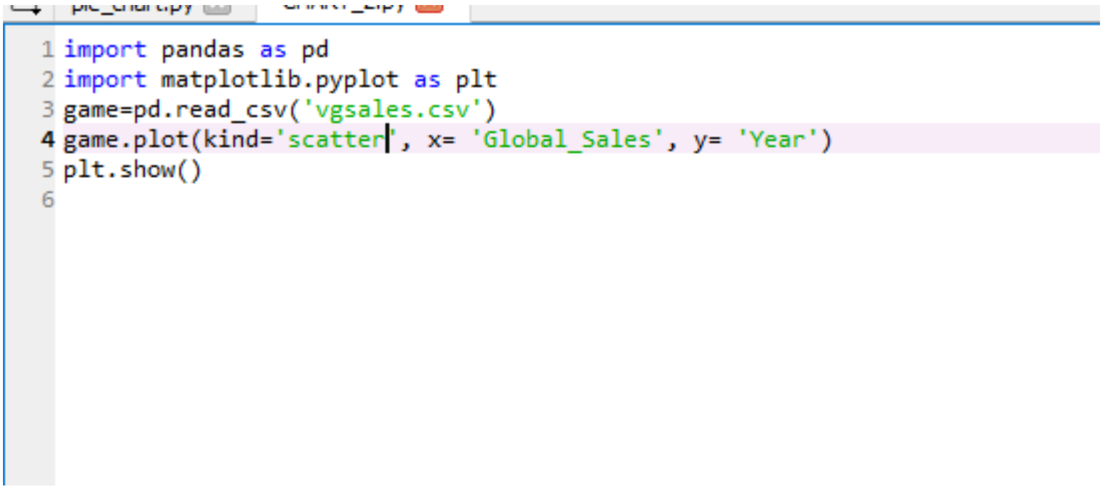
We see the best region to concentrate that's the region with major sales.

North America being the dominant sales region in the world containing 45.9% of the worlds game sales, followed by Europe region with 29.2%, Japan having 15.5% of the worlds sales share. We have used **Dictionary and List** for this code

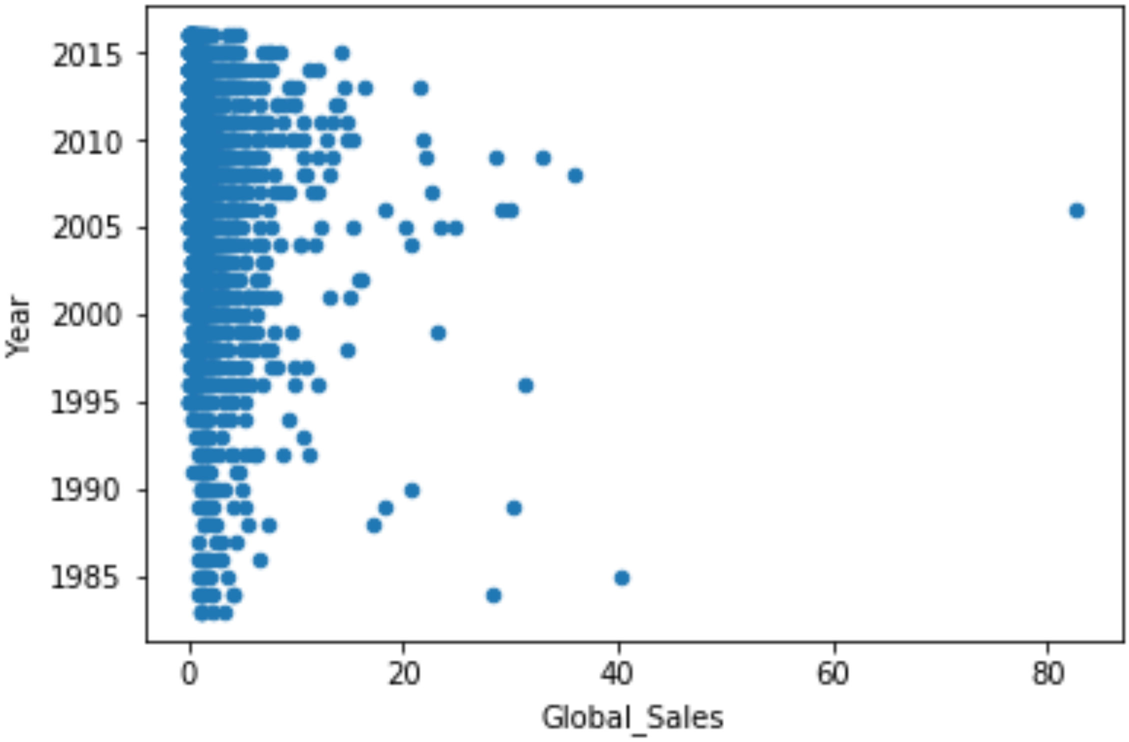


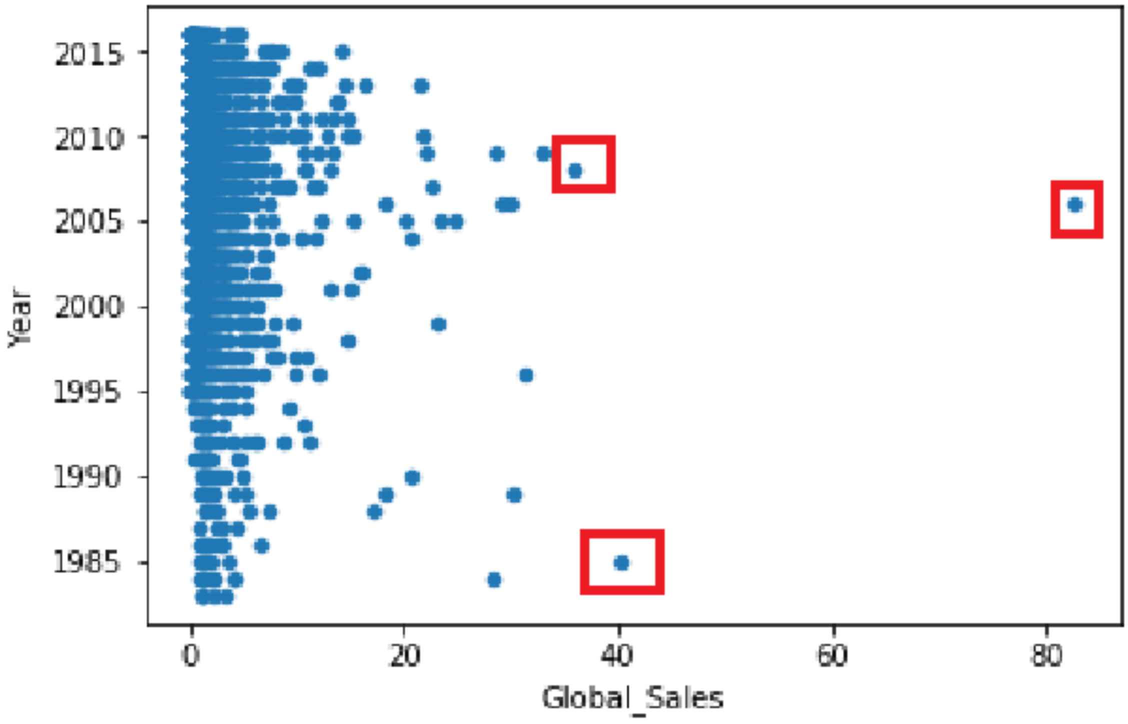
Question 2:

What insights one might get by comparing top three performing games of all time?



```
import pandas as pd
import matplotlib.pyplot as plt
game=pd.read_csv('vgsales.csv')
game.plot(kind='scatter', x= 'Global_Sales', y= 'Year')
plt.show()
```





Top Performing games -

Wii Sports - 80.74 (2006), Super Mario Bros -40.24 (1985), Wii Sports-82.74 (2008)

Interesting fact - Top three games have the same publisher Nintendo!

Question 3:

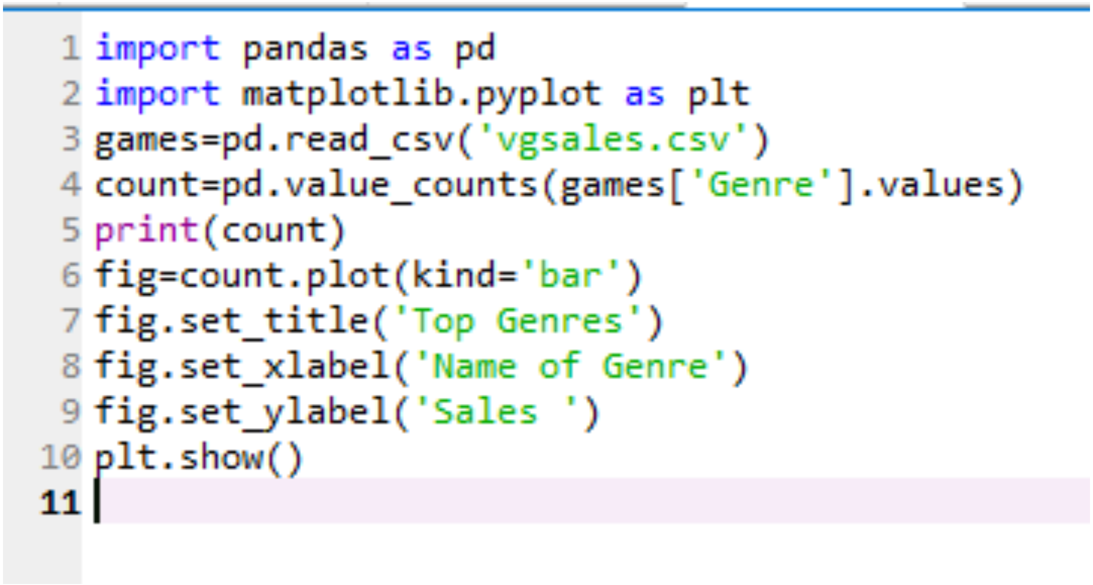
Does different type of game has impact on it's sales?

If person trying to make a debut in gaming industries, one of main things that person wants to choose is what type of video game will generate him more profits. Thus, now by looking at the bar chart visualization below one will have a clear view about the sales that video Genre in making, for instance if a person is planning to make a puzzle game which has only 20 millions compared to the sports games which has around 230 millions of sales worldwide.

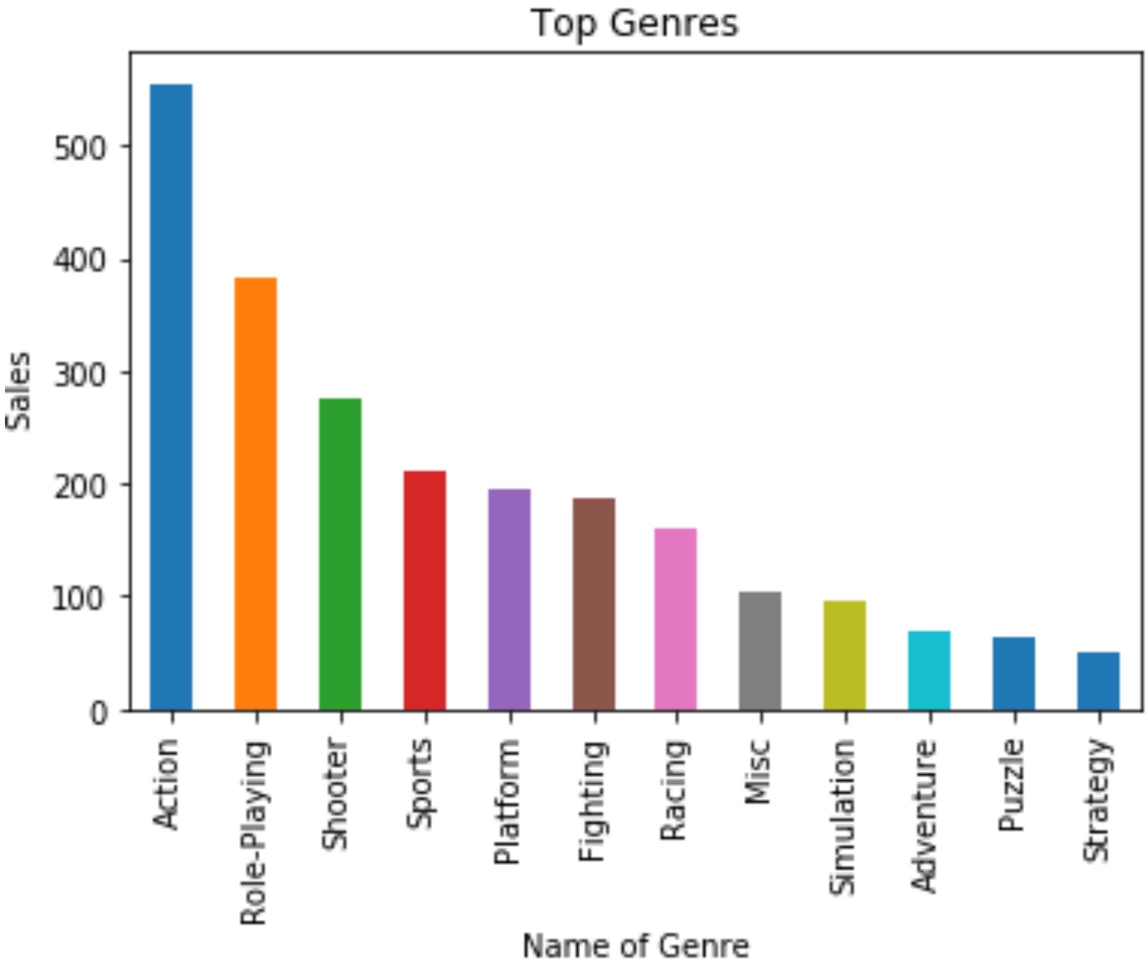
One can also make a conclusion, that if action genre is the most sold video game in world then, perhaps the action video gaming has much more competition and when there is a less sales in strategic video game, then the competition is fever compared to action game.

We have used **String** for this code

Code:



```
import pandas as pd
import matplotlib.pyplot as plt
games=pd.read_csv('vgsales.csv')
count=pd.value_counts(games['Genre'].values)
print(count)
fig=count.plot(kind='bar')
fig.set_title('Top Genres')
fig.set_xlabel('Name of Genre')
fig.set_ylabel('Sales ')
plt.show()

```



Top Genres
Sales
0
100
200
300
400
500
Action
Role-Playing
Shooter
Sports
Platform
Fighting
Racing
Misc
Simulation
Adventure
Puzzle
Strategy
Name of Genre

E. Consolidated Code:

Code for testing:

```
import pandas as pd

games=pd.read_csv('vgsales.csv')

del games['Rank_2000']

games.drop_duplicates(subset="Name",keep= 'first' ,inplace= True)

print(games.dropna())
```

Code for visuals:

Question 1:

```
import pandas as pd

import matplotlib.pyplot as plt

game=pd.read_csv('vgsales.csv')

na=game['NA_Sales']

na_game=sum(game['NA_Sales'])

jp=game['JP_Sales']

jp_game=sum(game['JP_Sales'])

eu=game['EU_Sales']

eu_game=sum(game['EU_Sales'])

ot=game['Other_Sales']

ot_game=sum(game['Other_Sales'])

region=('North_America','Europe','Others','Japan')

pie=(na_game,eu_game,ot_game,jp_game)
```

```
colors = ["#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#8c564b"]

plt.pie(pie, labels=region, colors=colors,

autopct='%1.1f%%', shadow=True, startangle=140)

plt.title("WORLD WIDE VIDEO GAME SALES")

plt.show()
```

Question 2:

```
import pandas as pd

import matplotlib.pyplot as plt

game=pd.read_csv('vgsales.csv')

game.plot(kind='line', x= 'Global_Sales', y= 'Year')

plt.show()
```

Question 3:

```
import pandas as pd

import matplotlib.pyplot as plt

games=pd.read_csv('vgsales.csv')

count=pd.value_counts(games['Genre'].values)

print(count)

fig=count.plot(kind='bar')

fig.set_title('Top Genres')

fig.set_xlabel('Name of Genre')

fig.set_ylabel('Sales ')

plt.show()
```

Code for summary statistics:

```
import pandas as pd

import matplotlib.pyplot as plt

games=pd.read_csv('vgsales1.csv')

print(" Summary Statistics Sales for North America :\n",games['NA_Sales'].describe())

print(" Summary Statistics Sales for Europe:\n",games['EU_Sales'].describe())

print(" Summary Statistics Sales for Japan:\n",games['JP_Sales'].describe())

print(" Summary Statistics Sales for Other Sales:\n",games['Other_Sales'].describe())

print(" Summary Statistics Global Sales:\n",games['Global_Sales'].describe())
```