m504_gh1017889

March 24, 2022

1 AI and Applications M504

1.1 Analysis of Video Game Sales

2 Understanding of Business Task

2.1 Introduction

Game is one of the oldest activities. Since ancient times, games have educated the younger generation, taught team skills, and supported primary socialization. However, the main function of the game has always been to provide joy and meet leisure needs. Today, tens of thousands of years later, nothing has changed, except that the game has received a completely new format of video game.

This digital game format means using computer technology during the game. This actually creates a video game environment where players are free to act.

The modern gaming industry spans all continents of the globe, speaks dozens of languages, is worth billions of dollars, and is an integral part of the world economy. Around the world, young and old play video games. And every year, more and more fans are "cutting themselves" in computer games.

2.1.1 Aim of the Notebook

I was hired by Video Games Selling Platform and I need to analyze the market. Moreover I need to generate questions that are relevant in my case and consenquentally to find answers using Python.

3 Data Collection

3.1 Importing Libraries

[1]:
```
import pandas as pd
import numpy as np
```

3.2 Loading the dataset

[17]:
```
df = pd.read_csv ("/home/abdullamuradov/vgsales.csv")
print(df.shape)
df.head(10)
```

```
(16598, 11)
```

[17]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	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	1.00	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	9	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.59	7.06	4.70	2.26	28.62
9	10	Duck Hunt	NES	1984.0	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31

3.3 Data Overview

Rank : The rank of the game sort on global sales

Name : Name of the Game

Platform : Name of the Platform

Year : The year of release

Genre : The genre of the game

Pubslisher : Name of the Publisher

NA_Sales : Sales in North America (million $)

EU_Sales : Sales in Europe (million $)

JP_Sales : Sales in Japan (million $)

Other Sales : Sales in other Continents (million $)

Global Sales : Overall Sales (million $)

4 Data Preprocessing

4.1 Detailed description of Dataset

[57]:
```
df.info()
df.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Rank          16598 non-null  int64
 1   Name          16598 non-null  object
 2   Platform      16598 non-null  object
 3   Year          16327 non-null  float64
 4   Genre         16598 non-null  object
 5   Publisher     16540 non-null  object
 6   NA_Sales      16598 non-null  float64
 7   EU_Sales      16598 non-null  float64
 8   JP_Sales      16598 non-null  float64
 9   Other_Sales   16598 non-null  float64
 10  Global_Sales  16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

[57]:

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	\
count	16598.000000	16327.000000	16598.000000	16598.000000	16598.000000	
mean	8300.605254	2006.406443	0.264667	0.146652	0.077782	
std	4791.853933	5.828981	0.816683	0.505351	0.309291	
min	1.000000	1980.000000	0.000000	0.000000	0.000000	
25%	4151.250000	2003.000000	0.000000	0.000000	0.000000	
50%	8300.500000	2007.000000	0.080000	0.020000	0.000000	
75%	12449.750000	2010.000000	0.240000	0.110000	0.040000	
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	

	Other_Sales	Global_Sales
count	16598.000000	16598.000000
mean	0.048063	0.537441
std	0.188588	1.555028
min	0.000000	0.010000

```
25%         0.000000      0.060000
50%         0.010000      0.170000
75%         0.040000      0.470000
max        10.570000     82.740000
```

As we can see we have some missing values. I will use 'dropna' function to drop missing values. By running describe function I got a general information about the dataset

[18]:
```
df = df.dropna()
print ("Missing Values has been Dropped")
```

```
Missing Values has been Dropped
```

[20]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16291 entries, 0 to 16597
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Rank          16291 non-null  int64
 1   Name          16291 non-null  object
 2   Platform      16291 non-null  object
 3   Year          16291 non-null  float64
 4   Genre         16291 non-null  object
 5   Publisher     16291 non-null  object
 6   NA_Sales      16291 non-null  float64
 7   EU_Sales      16291 non-null  float64
 8   JP_Sales      16291 non-null  float64
 9   Other_Sales   16291 non-null  float64
 10  Global_Sales  16291 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.5+ MB
```

The Value count in our dataset has been equalized. Now we can use it to answer business questions

5 Questions and Answers

What is the best selling platform?

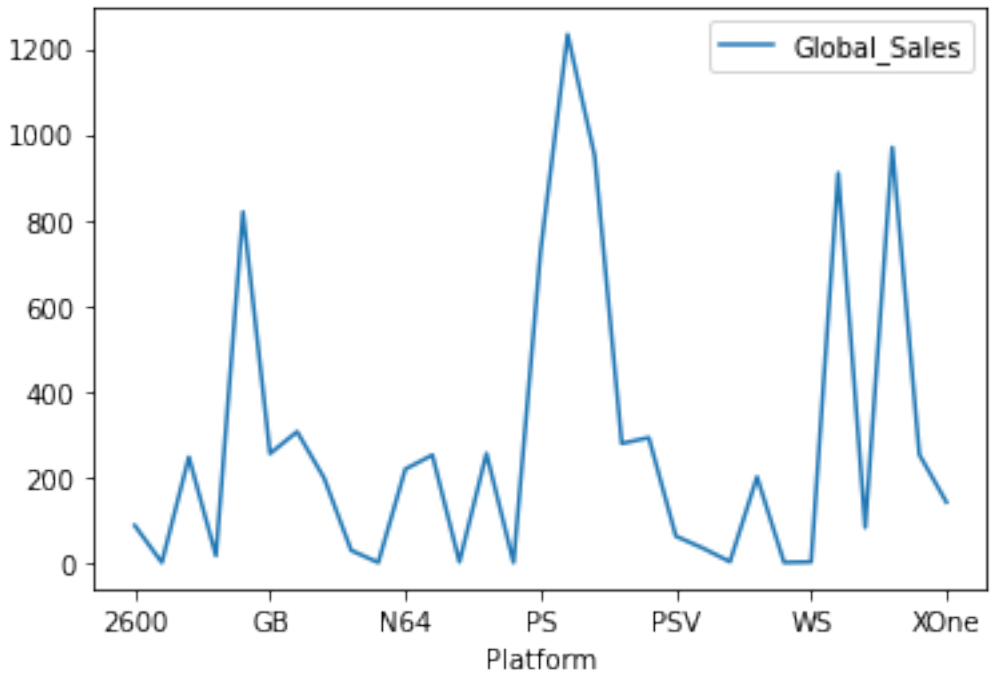
[59]:
```
df.groupby("Platform")["Global_Sales"].sum()
```

[59]:
```
Platform
2600      86.57
3DO        0.10
3DS      246.27
DC        15.97
```

```
DS       818.91
GB       254.42
GBA      305.62
GC       197.14
GEN       28.36
GG         0.04
N64      218.21
NES      251.07
NG         1.44
PC       254.70
PCFX       0.03
PS       727.39
PS2     1233.46
PS3      949.35
PS4      278.10
PSP      291.71
PSV       61.60
SAT       33.59
SCD        1.87
SNES     200.05
TG16       0.16
WS         1.42
Wii      909.81
WiiU      81.86
X360     969.60
XB       252.09
XOne     141.06
Name: Global_Sales, dtype: float64
```

[60]:
```
df.set_index('Platform', inplace=True)
df.groupby('Platform').sum()['Global_Sales'].plot(legend=True)
```

[60]:
```
<AxesSubplot:xlabel='Platform'>
```



After grouping Platform and Global Sales I answered the question. As it is shown in the graph 'Playstation' was the best selling platform. For more precise answer is PS2 with Global Sales of 1233 million $

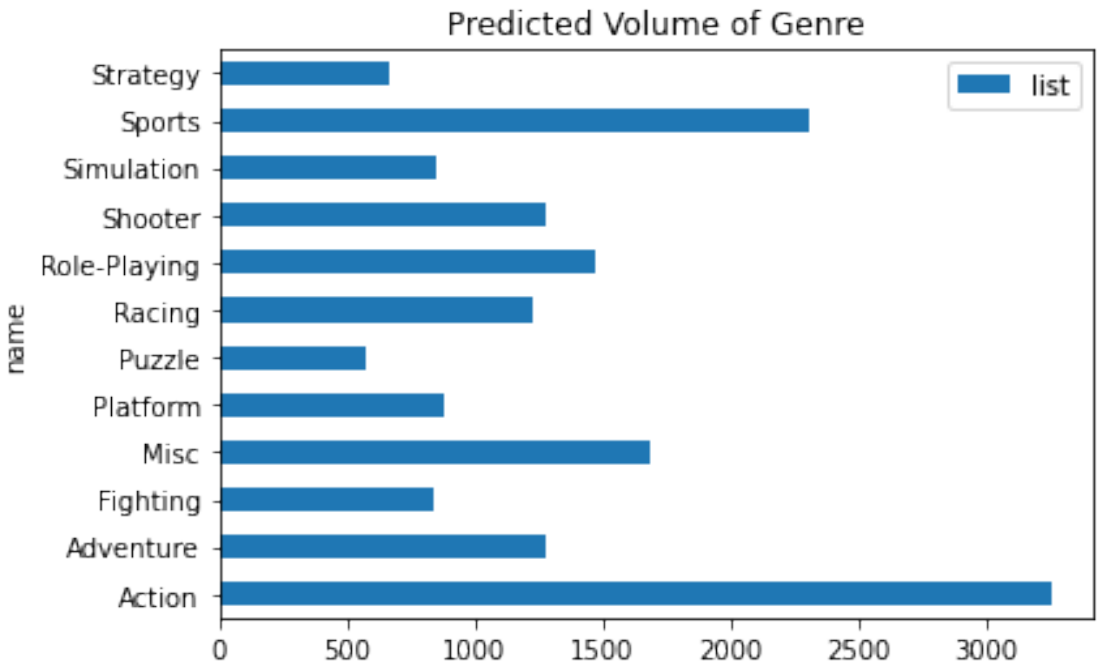
5.1 Which genre has more games than the others?

[21]:
```
genre = df["Genre"]
genre_counts = np.unique(genre, return_counts=True)
print(genre_counts)
```

```
(array(['Action', 'Adventure', 'Fighting', 'Misc', 'Platform', 'Puzzle',
       'Racing', 'Role-Playing', 'Shooter', 'Simulation', 'Sports',
       'Strategy'], dtype=object), array([3251, 1274,  836, 1686,  875,  570,
1225, 1470, 1282,  848, 2304,
        670]))
```

[62]:
```
df_tmp = pd.DataFrame({'name':genre_counts[0], 'list':genre_counts[1]})
df_tmp.plot.barh(x='name', y='list', title="Predicted Volume of Genre")
```

[62]: <AxesSubplot:title={'center':'Predicted Volume of Genre'}, ylabel='name'>



Obviously 'Action' genre has the best sales in the market. Because it covers wide range of most played games.

5.2 Which games sold the most globally?

[22]:
```
df.sort_values(by=['Global_Sales'], ascending=False).head(5)
```

[22]:

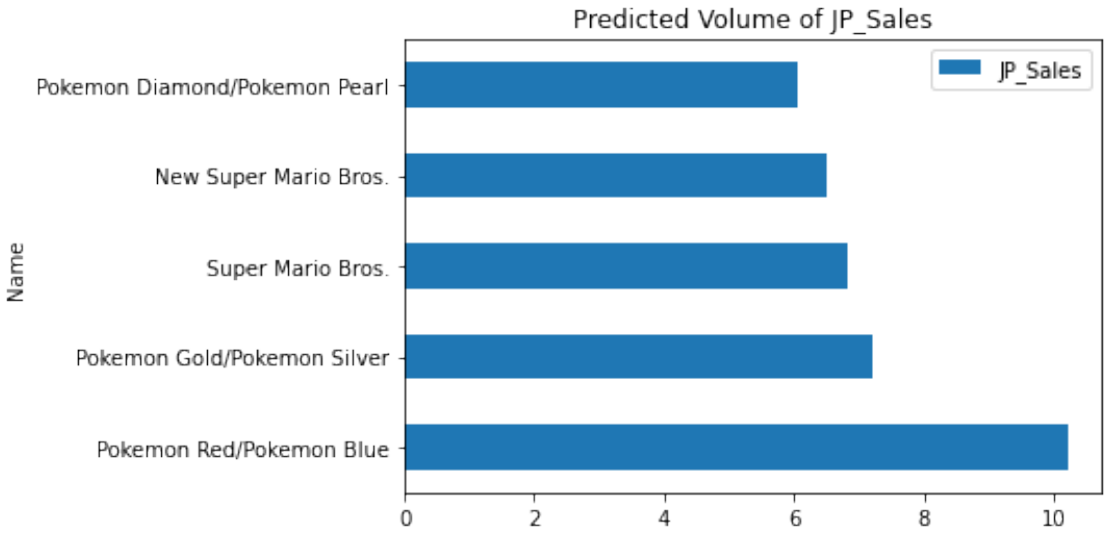
	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	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	1.00	31.37

Wii Sports is a leader of global sales and in top 5 we have 3 products of 'Wii'

5.3 What is the top selling 5 games in Japan?

[24]:
```
df.sort_values(by=["JP_Sales"], ascending=False).head(5)
df.sort_values(by=["JP_Sales"], ascending=False)[:5].plot.barh(x='Name',␣
↪y="JP_Sales", title="Predicted Volume of JP_Sales")
```

[24]:
```
<AxesSubplot:title={'center':'Predicted Volume of JP_Sales'}, ylabel='Name'>
```

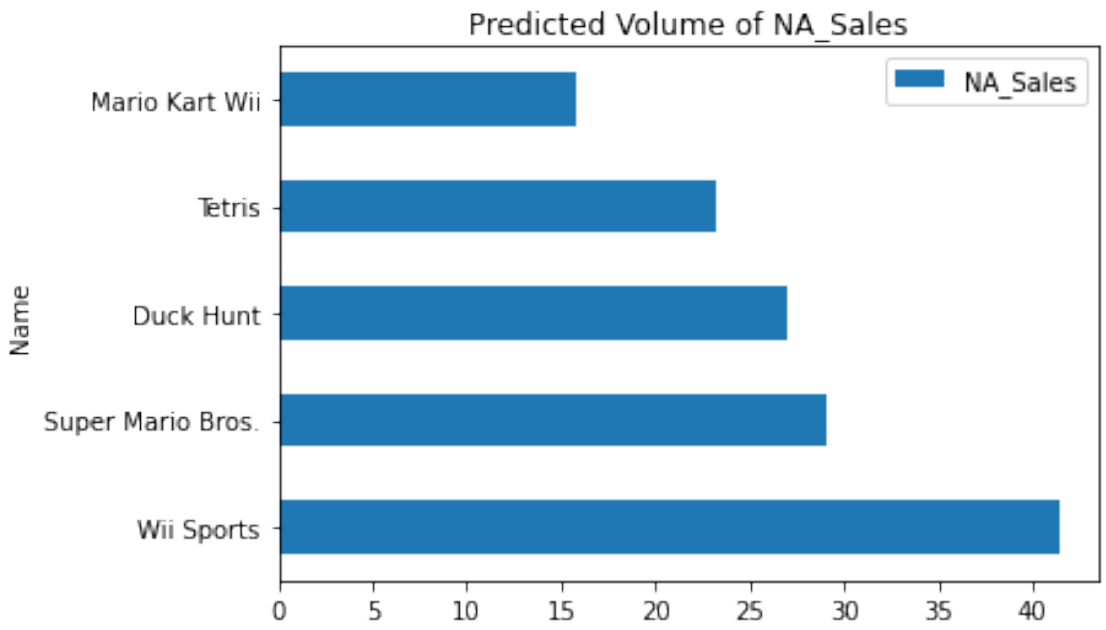


Pokemon Red/Pokemon Blue that is in 5 th places in Global Sales, selling the most in Japan. In top 5 Japan Sales Pokemon took 3 places. First,second and fith.

5.4 What is the top selling 5 games in North America?

[26]:
```
df.sort_values(by=["NA_Sales"], ascending=False).head(5)
df.sort_values(by=["NA_Sales"], ascending=False)[:5].plot.barh(x='Name',␣
↪y="NA_Sales", title="Predicted Volume of NA_Sales")
```

[26]:
```
<AxesSubplot:title={'center':'Predicted Volume of NA_Sales'}, ylabel='Name'>
```



Both Japan and North America have 'Mario' games in top 5 selling list.

5.5 How many Mario games are there?

[25]:
```
df[df["Name"].str.contains(r'Mario(?!$)')]
```

[25]:

	Rank	Name	Platform	\
1	2	Super Mario Bros.	NES	
2	3	Mario Kart Wii	Wii	
6	7	New Super Mario Bros.	DS	
8	9	New Super Mario Bros. Wii	Wii	
11	12	Mario Kart DS	DS	
…	…	…	…	
12373	12375	Mario vs. Donkey Kong: Tipping Stars	3DS	
12770	12772	Dance Dance Revolution: Mario Mix (JP sales)	GC	
13273	13275	Detective Conan: Marionette Symphony	3DS	
16357	16360	Mario vs. Donkey Kong: Tipping Stars	WiiU	
16542	16545	Mario & Luigi: Paper Jam & Mario Kart 7 Double…	3DS	

	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	\
1	1985.0	Platform	Nintendo	29.08	3.58	6.81	
2	2008.0	Racing	Nintendo	15.85	12.88	3.79	
6	2006.0	Platform	Nintendo	11.38	9.23	6.50	
8	2009.0	Platform	Nintendo	14.59	7.06	4.70	
11	2005.0	Racing	Nintendo	9.81	7.57	4.13	

```
…       …          …                    …        …       …       …
12373  2015.0      Puzzle               Nintendo      0.00      0.00      0.06
12770  2005.0  Simulation               Nintendo      0.00      0.00      0.05
13273  2013.0   Adventure  Namco Bandai Games      0.00      0.00      0.05
16357  2015.0      Puzzle               Nintendo      0.00      0.00      0.01
16542  2015.0        Misc               Nintendo      0.00      0.00      0.01

       Other_Sales  Global_Sales
1             0.77         40.24
2             3.31         35.82
6             2.90         30.01
8             2.26         28.62
11            1.92         23.42
…              …             …
12373         0.00          0.06
12770         0.00          0.06
13273         0.00          0.05
16357         0.00          0.01
16542         0.00          0.01

[107 rows x 11 columns]
```

In this dataset we have 10 games of Mario. First game of Mario was released in 1985 and last one in 2015

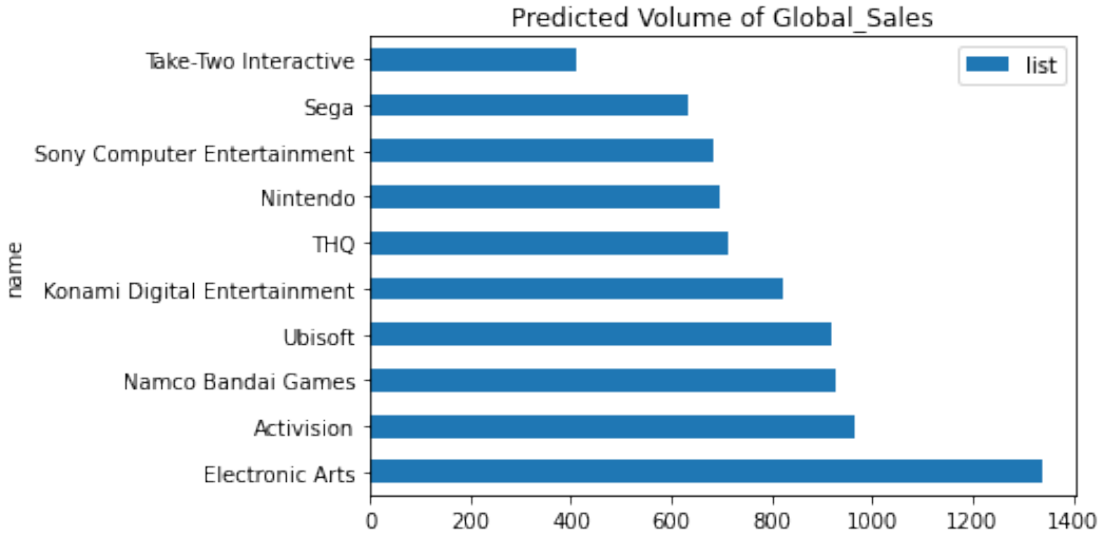
5.6 Which Publisher has more total sales (Global sales)?

[65]:
```
df_pub = df.groupby('Publisher')['Global_Sales'].count()
df_pub = df_pub.sort_values(ascending=False)
df_pub
```

[65]:
```
Publisher
Electronic Arts                 1339
Activision                       966
Namco Bandai Games               928
Ubisoft                          918
Konami Digital Entertainment     823
                                 …
Elite                              1
Mystique                           1
NDA Productions                    1
Navarre Corp                       1
Locus                              1
Name: Global_Sales, Length: 576, dtype: int64
```

[66]:
```
df_tmp = pd.DataFrame({'name':df_pub.index, 'list':df_pub.values})
df_tmp[:10].plot.barh(x='name', y="list", title="Predicted Volume of␣
↪Global_Sales")
```

[66]: <AxesSubplot:title={'center':'Predicted Volume of Global_Sales'}, ylabel='name'>



Ea games is undoubted leader in Global Sales.

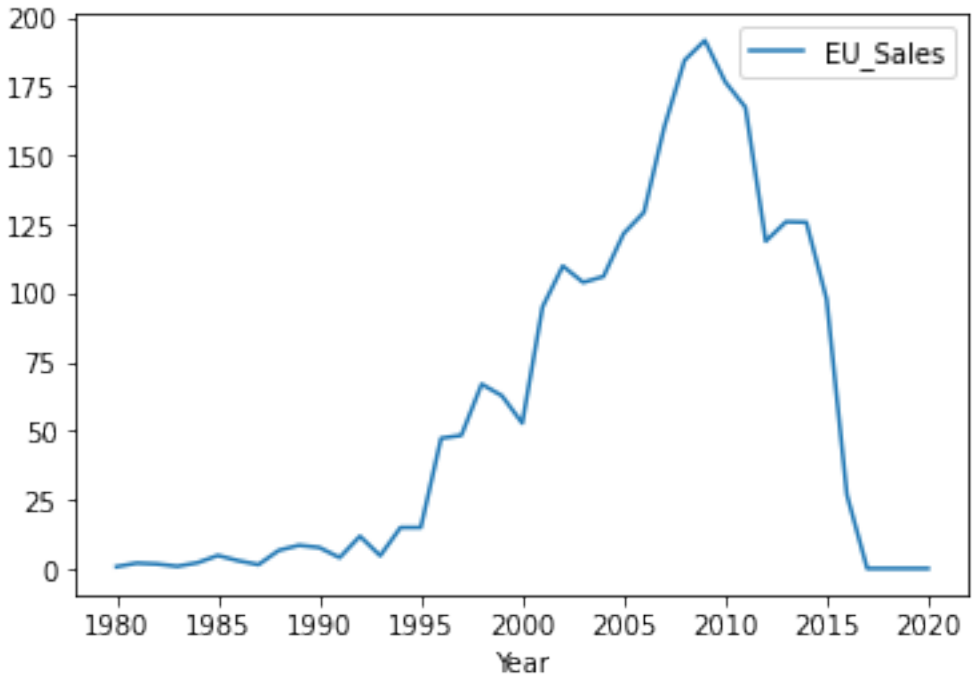
5.7 In which year are more sales in EU?

[67]:
```
df['Year'] = df['Year'].astype(int)
df.head()

EU = df[['Year','EU_Sales']]
EU = EU.groupby('Year').sum()

print(EU['EU_Sales'].idxmax())
print(max(EU.EU_Sales))
```

```
2009
191.5899999999998
```

[68]:
```
df.set_index('Year', inplace=True)
df.groupby('Year').sum()['EU_Sales'].plot(legend=True)
```

[68]: <AxesSubplot:xlabel='Year'>



2009 was a year with the best sales in Europe. According to the graph the sales were raising from 1985 to 2009 and then started to decrease.

6 Conclusion

According to the analysis, I can make suggestions and to state my personal opinion :

1. 'EA' is the best selling platform and action genre is most popular one. As a gamer i know that most of the games released by EA are action games. In my opinion other publishers have to take that fact into account.
2. 'Playstation' is the best selling platform and they have official partnership with 'EA'. It means that 'Playstation' is getting 'EA' games before all other platforms. That might be a reason of success, collaboration of 2 giants works well.
3. In my opinion 'Mario' should release a new game that will be similar to old 'Mario' but with new graphical effects. Mario had really huge volume of sales back in 1985 that was almost 41 million, however new 'Mario' series have almost "no sales". In my opinion it is related with their wish to produce something new but from the other hand they are not producing same type of games for which audience loved them. Restoring collaboration with 'Nintendo' is a good idea as well.

In this assigment I tried to present answers with graphs and without them. The dataset that I used is really handfull, however some records seems to be not so fresh.