vgsales

2020 年 6 月 2 日

读取电子游戏销售数据，导入所需库

In [243]:
```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression as LR
from sklearn import ensemble
os.chdir("C:\\Users\\acer_pc\\Downloads\\284_618_bundle_archive")
data = pd.read_csv("vgsales.csv")
data = data.dropna()
data.loc[data.Genre=='Role-Playing','Genre'] = 'Role_Playing'
```

1. 电子游戏市场分析：受欢迎的游戏、类型、发布平台、发行人等
受欢迎的游戏：按全球销量排序，可找出受欢迎的游戏排行：

In [173]:
```
df = data.copy()
df.sort_values(by='Global_Sales')
df.head(50)
```

Out[173]:

	Rank	Name	Platform	Year	\
0	1	Wii Sports	Wii	2006.0	
1	2	Super Mario Bros.	NES	1985.0	
2	3	Mario Kart Wii	Wii	2008.0	
3	4	Wii Sports Resort	Wii	2009.0	
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	
5	6	Tetris	GB	1989.0	
6	7	New Super Mario Bros.	DS	2006.0	
7	8	Wii Play	Wii	2006.0	
8	9	New Super Mario Bros. Wii	Wii	2009.0	
9	10	Duck Hunt	NES	1984.0	
10	11	Nintendogs	DS	2005.0	

11	12	Mario Kart DS	DS	2005.0
12	13	Pokemon Gold/Pokemon Silver	GB	1999.0
13	14	Wii Fit	Wii	2007.0
14	15	Wii Fit Plus	Wii	2009.0
15	16	Kinect Adventures!	X360	2010.0
16	17	Grand Theft Auto V	PS3	2013.0
17	18	Grand Theft Auto: San Andreas	PS2	2004.0
18	19	Super Mario World	SNES	1990.0
19	20	Brain Age: Train Your Brain in Minutes a Day	DS	2005.0
20	21	Pokemon Diamond/Pokemon Pearl	DS	2006.0
21	22	Super Mario Land	GB	1989.0
22	23	Super Mario Bros. 3	NES	1988.0
23	24	Grand Theft Auto V	X360	2013.0
24	25	Grand Theft Auto: Vice City	PS2	2002.0
25	26	Pokemon Ruby/Pokemon Sapphire	GBA	2002.0
26	27	Pokemon Black/Pokemon White	DS	2010.0
27	28	Brain Age 2: More Training in Minutes a Day	DS	2005.0
28	29	Gran Turismo 3: A-Spec	PS2	2001.0
29	30	Call of Duty: Modern Warfare 3	X360	2011.0
30	31	Pokémon Yellow: Special Pikachu Edition	GB	1998.0
31	32	Call of Duty: Black Ops	X360	2010.0
32	33	Pokemon X/Pokemon Y	3DS	2013.0
33	34	Call of Duty: Black Ops 3	PS4	2015.0
34	35	Call of Duty: Black Ops II	PS3	2012.0
35	36	Call of Duty: Black Ops II	X360	2012.0
36	37	Call of Duty: Modern Warfare 2	X360	2009.0
37	38	Call of Duty: Modern Warfare 3	PS3	2011.0
38	39	Grand Theft Auto III	PS2	2001.0
39	40	Super Smash Bros. Brawl	Wii	2008.0
40	41	Call of Duty: Black Ops	PS3	2010.0
41	42	Animal Crossing: Wild World	DS	2005.0
42	43	Mario Kart 7	3DS	2011.0
43	44	Halo 3	X360	2007.0
44	45	Grand Theft Auto V	PS4	2014.0
45	46	Pokemon HeartGold/Pokemon SoulSilver	DS	2009.0
46	47	Super Mario 64	N64	1996.0
47	48	Gran Turismo 4	PS2	2004.0
48	49	Super Mario Galaxy	Wii	2007.0
49	50	Pokemon Omega Ruby/Pokemon Alpha Sapphire	3DS	2014.0

	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	\
0	Sports	Nintendo	41.49	29.02	3.77	
1	Platform	Nintendo	29.08	3.58	6.81	
2	Racing	Nintendo	15.85	12.88	3.79	
3	Sports	Nintendo	15.75	11.01	3.28	
4	Role_Playing	Nintendo	11.27	8.89	10.22	
5	Puzzle	Nintendo	23.20	2.26	4.22	
6	Platform	Nintendo	11.38	9.23	6.50	
7	Misc	Nintendo	14.03	9.20	2.93	
8	Platform	Nintendo	14.59	7.06	4.70	
9	Shooter	Nintendo	26.93	0.63	0.28	
10	Simulation	Nintendo	9.07	11.00	1.93	
11	Racing	Nintendo	9.81	7.57	4.13	
12	Role_Playing	Nintendo	9.00	6.18	7.20	
13	Sports	Nintendo	8.94	8.03	3.60	
14	Sports	Nintendo	9.09	8.59	2.53	
15	Misc	Microsoft Game Studios	14.97	4.94	0.24	
16	Action	Take-Two Interactive	7.01	9.27	0.97	
17	Action	Take-Two Interactive	9.43	0.40	0.41	
18	Platform	Nintendo	12.78	3.75	3.54	
19	Misc	Nintendo	4.75	9.26	4.16	
20	Role_Playing	Nintendo	6.42	4.52	6.04	
21	Platform	Nintendo	10.83	2.71	4.18	
22	Platform	Nintendo	9.54	3.44	3.84	
23	Action	Take-Two Interactive	9.63	5.31	0.06	
24	Action	Take-Two Interactive	8.41	5.49	0.47	
25	Role_Playing	Nintendo	6.06	3.90	5.38	
26	Role_Playing	Nintendo	5.57	3.28	5.65	
27	Puzzle	Nintendo	3.44	5.36	5.32	
28	Racing	Sony Computer Entertainment	6.85	5.09	1.87	
29	Shooter	Activision	9.03	4.28	0.13	
30	Role_Playing	Nintendo	5.89	5.04	3.12	
31	Shooter	Activision	9.67	3.73	0.11	
32	Role_Playing	Nintendo	5.17	4.05	4.34	
33	Shooter	Activision	5.77	5.81	0.35	
34	Shooter	Activision	4.99	5.88	0.65	
35	Shooter	Activision	8.25	4.30	0.07	
36	Shooter	Activision	8.52	3.63	0.08	

37	Shooter	Activision	5.54	5.82	0.49
38	Action	Take-Two Interactive	6.99	4.51	0.30
39	Fighting	Nintendo	6.75	2.61	2.66
40	Shooter	Activision	5.98	4.44	0.48
41	Simulation	Nintendo	2.55	3.52	5.33
42	Racing	Nintendo	4.74	3.91	2.67
43	Shooter	Microsoft Game Studios	7.97	2.83	0.13
44	Action	Take-Two Interactive	3.80	5.81	0.36
45	Action	Nintendo	4.40	2.77	3.96
46	Platform	Nintendo	6.91	2.85	1.91
47	Racing	Sony Computer Entertainment	3.01	0.01	1.10
48	Platform	Nintendo	6.16	3.40	1.20
49	Role_Playing	Nintendo	4.23	3.37	3.08

	Other_Sales	Global_Sales
0	8.46	82.74
1	0.77	40.24
2	3.31	35.82
3	2.96	33.00
4	1.00	31.37
5	0.58	30.26
6	2.90	30.01
7	2.85	29.02
8	2.26	28.62
9	0.47	28.31
10	2.75	24.76
11	1.92	23.42
12	0.71	23.10
13	2.15	22.72
14	1.79	22.00
15	1.67	21.82
16	4.14	21.40
17	10.57	20.81
18	0.55	20.61
19	2.05	20.22
20	1.37	18.36
21	0.42	18.14
22	0.46	17.28
23	1.38	16.38

24	1.78	16.15
25	0.50	15.85
26	0.82	15.32
27	1.18	15.30
28	1.16	14.98
29	1.32	14.76
30	0.59	14.64
31	1.13	14.64
32	0.79	14.35
33	2.31	14.24
34	2.52	14.03
35	1.12	13.73
36	1.29	13.51
37	1.62	13.46
38	1.30	13.10
39	1.02	13.04
40	1.83	12.73
41	0.88	12.27
42	0.89	12.21
43	1.21	12.14
44	2.02	11.98
45	0.77	11.90
46	0.23	11.89
47	7.53	11.66
48	0.76	11.52
49	0.65	11.33

受欢迎的类型：首先分别按类型将数据集划分为多个子数据集。结果显示共有 12 个不同类别，在这里展示了 Sports 类别的子数据集

```
In [138]: Genres = df['Genre'].unique()
          print(Genres)
          Genres_split = []
          for temp_G in Genres:
              temp_data = df[df['Genre'].isin([temp_G])]
              exec("df_%s = temp_data" %temp_G)
              Genres_split.append(temp_data)
          df_Sports
```

```
['Sports' 'Platform' 'Racing' 'Role_Playing' 'Puzzle' 'Misc' 'Shooter'
 'Simulation' 'Action' 'Fighting' 'Adventure' 'Strategy']
```

Out[138]:

	Rank	Name	Platform	Year	Genre	\
0	1	Wii Sports	Wii	2006.0	Sports	
3	4	Wii Sports Resort	Wii	2009.0	Sports	
13	14	Wii Fit	Wii	2007.0	Sports	
14	15	Wii Fit Plus	Wii	2009.0	Sports	
77	78	FIFA 16	PS4	2015.0	Sports	
...	...	...	...	...	...	
16576	16579	Rugby Challenge 3	XOne	2016.0	Sports	
16578	16581	Outdoors Unleashed: Africa 3D	3DS	2011.0	Sports	
16579	16582	PGA European Tour	N64	2000.0	Sports	
16581	16584	Fit & Fun	Wii	2011.0	Sports	
16587	16590	Mezase!! Tsuri Master DS	DS	2009.0	Sports	

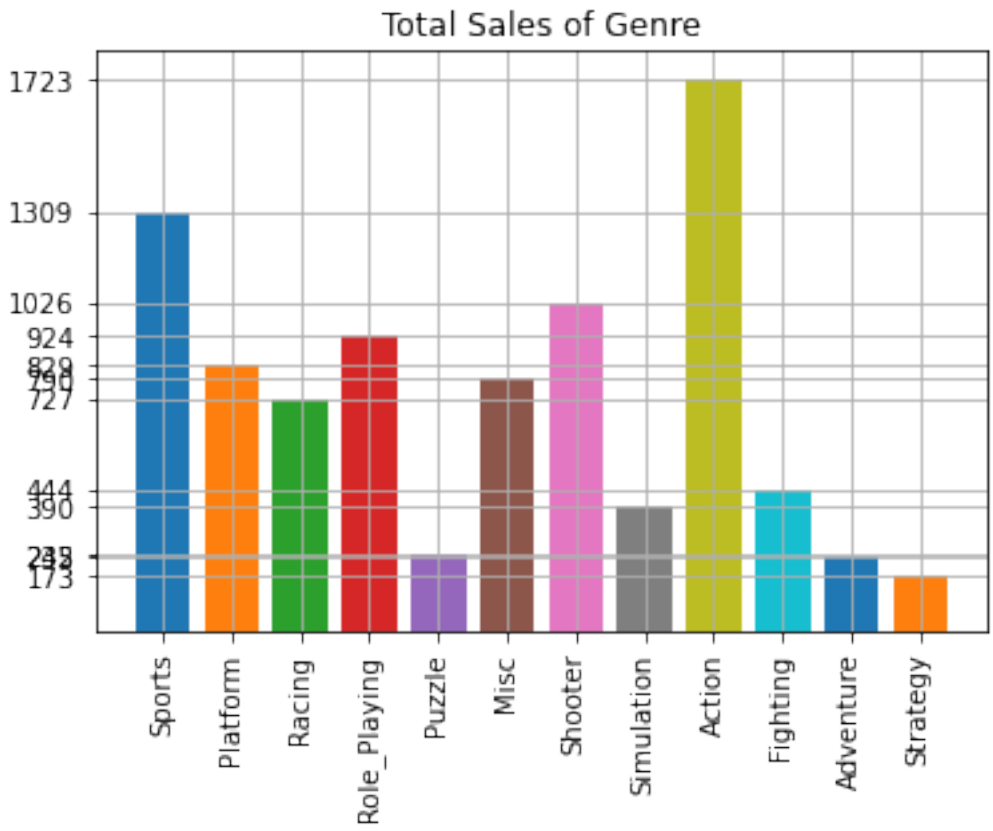
	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	\
0	Nintendo	41.49	29.02	3.77	8.46	
3	Nintendo	15.75	11.01	3.28	2.96	
13	Nintendo	8.94	8.03	3.60	2.15	
14	Nintendo	9.09	8.59	2.53	1.79	
77	Electronic Arts	1.11	6.06	0.06	1.26	
...	...	...	...	...	...	
16576	Alternative Software	0.00	0.01	0.00	0.00	
16578	Mastiff	0.01	0.00	0.00	0.00	
16579	Infogrames	0.01	0.00	0.00	0.00	
16581	Unknown	0.00	0.01	0.00	0.00	
16587	Hudson Soft	0.00	0.00	0.01	0.00	

	Global_Sales
0	82.74
3	33.00
13	22.72
14	22.00
77	8.49
...	...
16576	0.01
16578	0.01
16579	0.01
16581	0.01
16587	0.01

```
[2304 rows x 11 columns]
```

然后计算每一类型的销售总额，并将其可视化。从可视化结果可以看出最受欢迎的游戏类别是 Actions

```
In [139]: sale_of_genre = {}
          for sub_df in Genres_split:
              global_sale = np.sum(sub_df.Global_Sales)
              Genre_name = sub_df['Genre'].to_list()[0]
              sale_of_genre[Genre_name]=global_sale
          # print(sale_of_genre)
          keys = list(sale_of_genre.keys())
          values = list(sale_of_genre.values())
          # print(keys)
          # print(values)
          # plt.bar(keys,values)
          for i,key in enumerate(sale_of_genre):
              plt.bar(i,sale_of_genre[key])
          plt.xticks(np.arange(len(values)),keys,rotation=90)
          plt.yticks(values)
          plt.grid('on')
          plt.title('Total Sales of Genre')
          plt.show()
```



受欢迎的游戏平台：和之前操作类似，按照游戏平台将数据划分为多个子集。这里展示 PS4 平台的数据子集：

```
In [140]: Platforms = df['Platform'].unique() #sort_values(by=['Genre','Global_Sales'],ascending=F
          print(Platforms)
          Platforms_split = []
          for temp_G in Platforms:
              temp_data = df[df['Platform'].isin([temp_G])]
              exec("df_%s = temp_data" %temp_G)
              Platforms_split.append(temp_data)
          df_PS4
```

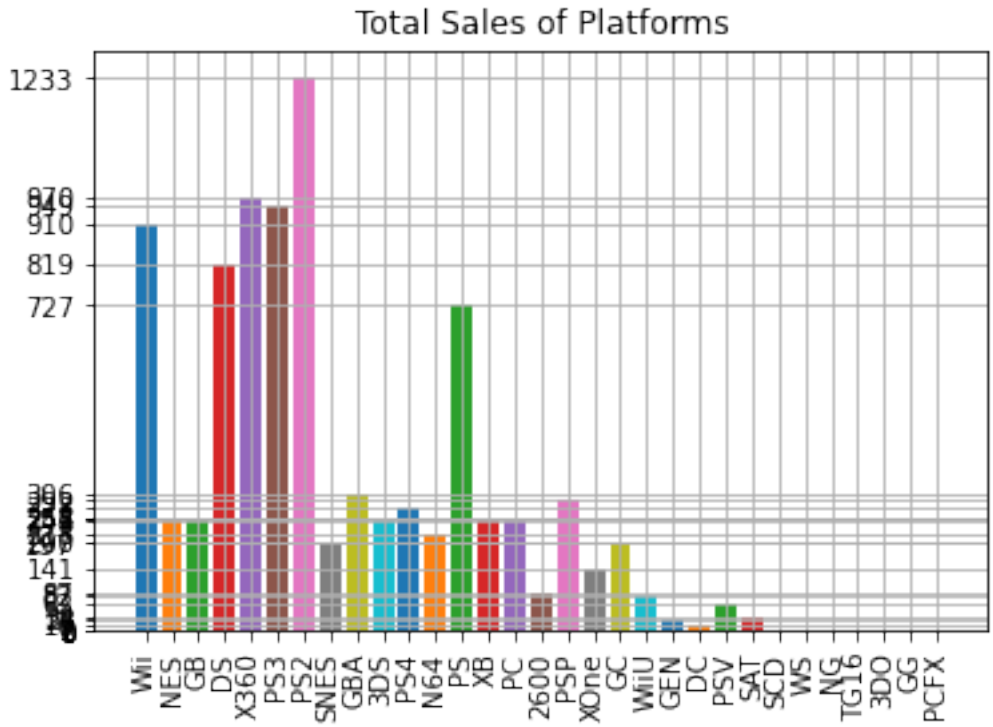
```
['Wii' 'NES' 'GB' 'DS' 'X360' 'PS3' 'PS2' 'SNES' 'GBA' '3DS' 'PS4' 'N64'
 'PS' 'XB' 'PC' '2600' 'PSP' 'XOne' 'GC' 'WiiU' 'GEN' 'DC' 'PSV' 'SAT'
 'SCD' 'WS' 'NG' 'TG16' '3DO' 'GG' 'PCFX']
```

```
Out[140]:         Rank                                          Name Platform  \
          33        34                       Call of Duty: Black Ops 3      PS4
```

```
44        45                                  Grand Theft Auto V      PS4
77        78                                          FIFA 16         PS4
92        93                    Star Wars Battlefront (2015)          PS4
93        94                  Call of Duty: Advanced Warfare          PS4
...      ...                                             ...         ...
16220  16223                    Dynasty Warriors: Eiketsuden          PS4
16260  16263                                        Paragon           PS4
16333  16336                                    Chaos;Child           PS4
16550  16553  God Eater Off Shot: Tachibana Sakuya-hen Twin ...       PS4
16570  16573                  Farming 2017 - The Simulation           PS4

         Year       Genre             Publisher  NA_Sales  EU_Sales  JP_Sales  \
33     2015.0     Shooter            Activision      5.77      5.81      0.35
44     2014.0      Action  Take-Two Interactive      3.80      5.81      0.36
77     2015.0      Sports       Electronic Arts      1.11      6.06      0.06
92     2015.0     Shooter       Electronic Arts      2.93      3.29      0.22
93     2014.0     Shooter            Activision      2.80      3.30      0.14
...       ...         ...                   ...       ...       ...       ...
16220  2016.0      Action            Tecmo Koei      0.00      0.00      0.01
16260  2016.0      Action            Epic Games      0.01      0.00      0.00
16333  2015.0   Adventure                  5pb       0.00      0.00      0.01
16550  2016.0      Action    Namco Bandai Games      0.00      0.00      0.01
16570  2016.0  Simulation     UIG Entertainment      0.00      0.01      0.00

       Other_Sales  Global_Sales
33            2.31         14.24
44            2.02         11.98
77            1.26          8.49
92            1.23          7.67
93            1.37          7.60
...            ...           ...
16220         0.00          0.01
16260         0.00          0.01
16333         0.00          0.01
16550         0.00          0.01
16570         0.00          0.01

[336 rows x 11 columns]
```

可视化各个平台的销售总额，可发现 PS2 平台最受欢迎。

```
In [141]: sale_of_platform = {}
          for sub_df in Platforms_split:
              global_sale = np.sum(sub_df.Global_Sales)
              platform_name = sub_df['Platform'].to_list()[0]
              sale_of_platform[platform_name]=global_sale
          # print(sale_of_genre)
          keys = list(sale_of_platform.keys())
          values = list(sale_of_platform.values())

          for i,key in enumerate(sale_of_platform):
              plt.bar(i,sale_of_platform[key])
          plt.xticks(np.arange(len(values)),keys,rotation=90)
          plt.yticks(values)
          plt.grid('on')
          plt.title('Total Sales of Platforms')
          plt.show()
```



受欢迎的发布者：和之前操作类似，按照不同发布者将数据集划分为多个子集。不同的发布者数量较多，这里不单独展示了。

```
In [156]: Publishers = df['Publisher'].unique() #sort_values(by=['Genre','Global_Sales'],ascending
```

```
print(Publishers)
Publishers_split = []
for idx,temp_G in enumerate(Publishers):
    temp_data = df[df['Publisher'].isin([temp_G])]

    exec("df_%d = temp_data" %idx)
    Publishers_split.append(temp_data)
# idx = Publishers.tolist().index('Nintendo')
# exec('print(df_%d)' %idx)
```

```
['Nintendo' 'Microsoft Game Studios' 'Take-Two Interactive'
 'Sony Computer Entertainment' 'Activision' 'Ubisoft' 'Bethesda Softworks'
 'Electronic Arts' 'Sega' 'SquareSoft' 'Atari' '505 Games' 'Capcom'
 'GT Interactive' 'Konami Digital Entertainment'
 'Sony Computer Entertainment Europe' 'Square Enix' 'LucasArts'
 'Virgin Interactive' 'Warner Bros. Interactive Entertainment'
 'Universal Interactive' 'Eidos Interactive' 'RedOctane' 'Vivendi Games'
 'Enix Corporation' 'Namco Bandai Games' 'Palcom' 'Hasbro Interactive'
 'THQ' 'Fox Interactive' 'Acclaim Entertainment' 'MTV Games'
 'Disney Interactive Studios' 'Majesco Entertainment' 'Codemasters'
 'Red Orb' 'Level 5' 'Arena Entertainment' 'Midway Games' 'JVC'
 'Deep Silver' '989 Studios' 'NCSoft' 'UEP Systems' 'Parker Bros.' 'Maxis'
 'Imagic' 'Tecmo Koei' 'Valve Software' 'ASCII Entertainment' 'Mindscape'
 'Infogrames' 'Unknown' 'Square' 'Valve' 'Activision Value' 'Banpresto'
 'D3Publisher' 'Oxygen Interactive' 'Red Storm Entertainment'
 'Video System' 'Hello Games' 'Global Star' 'Gotham Games'
 'Westwood Studios' 'GungHo' 'Crave Entertainment' 'Hudson Soft' 'Coleco'
 'Rising Star Games' 'Atlus' 'TDK Mediactive' 'ASC Games' 'Zoo Games'
 'Accolade' 'Sony Online Entertainment' '3DO' 'RTL' 'Natsume'
 'Focus Home Interactive' 'Alchemist' 'Black Label Games'
 'SouthPeak Games' 'Mastertronic' 'Ocean' 'Zoo Digital Publishing'
 'Psygnosis' 'City Interactive' 'Empire Interactive' 'Success' 'Compile'
 'Russel' 'Taito' 'Agetec' 'GSP' 'Microprose' 'Play It'
 'Slightly Mad Studios' 'Tomy Corporation' 'Sammy Corporation'
 'Koch Media' 'Game Factory' 'Titus' 'Marvelous Entertainment' 'Genki'
 'Mojang' 'Pinnacle' 'CTO SpA' 'TalonSoft' 'Crystal Dynamics' 'SCi'
 'Quelle' 'mixi, Inc' 'Rage Software' 'Ubisoft Annecy' 'Scholastic Inc.'
 'Interplay' 'Mystique' 'ChunSoft' 'Square EA'
 '20th Century Fox Video Games' 'Avanquest Software'
 'Hudson Entertainment' 'Nordic Games' 'Men-A-Vision' 'Nobilis'
```

'Big Ben Interactive' 'Touchstone' 'Spike' 'Jester Interactive'
'Nippon Ichi Software' 'LEGO Media' 'Quest' 'Illusion Softworks'
'Tigervision' 'Funbox Media' 'Rocket Company' 'Metro 3D'
'Mattel Interactive' 'IE Institute' 'Rondomedia'
'Sony Computer Entertainment America' 'Universal Gamex' 'Ghostlight'
'Wizard Video Games' 'BMG Interactive Entertainment' 'PQube'
'Trion Worlds' 'Laguna' 'Ignition Entertainment' 'Takara'
'Kadokawa Shoten' 'Destineer' 'Enterbrain' 'Xseed Games' 'Imagineer'
'System 3 Arcade Software' 'CPG Products' 'Aruze Corp' 'Gamebridge'
'Midas Interactive Entertainment' 'Jaleco' 'Answer Software' 'XS Games'
'Activision Blizzard' 'Pack In Soft' 'Rebellion' 'Xplosiv'
'GameMill Entertainment' 'Wanadoo' 'NovaLogic' 'Telltale Games' 'Epoch'
'BAM! Entertainment' 'Knowledge Adventure' 'Mastiff' 'Tetris Online'
'Harmonix Music Systems' 'ESP' 'TYO' 'Telegames' 'Mud Duck Productions'
'Screenlife' 'Pioneer LDC' 'Magical Company' 'Mentor Interactive' 'Kemco'
'Human Entertainment' 'Avanquest' 'Data Age' 'Electronic Arts Victor'
'Black Bean Games' 'Jack of All Games' '989 Sports' 'Takara Tomy'
'Media Rings' 'Elf' 'Starfish' 'Zushi Games' 'Jorudan'
'Destination Software, Inc' 'New' 'Brash Entertainment'
'ITT Family Games' 'PopCap Games' 'Home Entertainment Suppliers'
'Ackkstudios' 'Starpath Corp.' 'P2 Games' 'BPS' 'Gathering of Developers'
'NewKidCo' 'Storm City Games' 'CokeM Interactive' 'CBS Electronics'
'Magix' 'Marvelous Interactive' 'Kalypso Media'
'Nihon Falcom Corporation' 'Wargaming.net' 'Angel Studios'
'Arc System Works' 'Playmates' 'SNK Playmore' 'Hamster Corporation'
'From Software' 'Nippon Columbia' 'Nichibutsu' 'Little Orbit'
'Conspiracy Entertainment' 'DTP Entertainment' 'Hect' 'Mumbo Jumbo'
'Pacific Century Cyber Works' 'Indie Games' 'Liquid Games' 'NEC' 'Axela'
'ArtDink' 'Sunsoft' 'Gust' 'SNK' 'NEC Interchannel' 'FuRyu'
'Xing Entertainment' 'ValuSoft' 'Victor Interactive' 'Detn8 Games'
'American Softworks' 'Nordcurrent' 'Bomb' 'Falcom Corporation'
'AQ Interactive' 'CCP' 'Milestone S.r.l.' 'JoWood Productions'
'Seta Corporation' 'On Demand' 'NCS' 'Aspyr' 'Gremlin Interactive Ltd'
'Agatsuma Entertainment' 'Compile Heart' 'Culture Brain' 'Mad Catz'
'Shogakukan' 'Merscom LLC' 'Rebellion Developments' 'Nippon Telenet'
'TDK Core' 'bitComposer Games' 'Foreign Media Games' 'Astragon' 'SSI'
'Kadokawa Games' 'Idea Factory' 'Performance Designed Products'
'Asylum Entertainment' 'Core Design Ltd.' 'PlayV' 'UFO Interactive'
'Idea Factory International' 'Playlogic Game Factory' 'Essential Games'

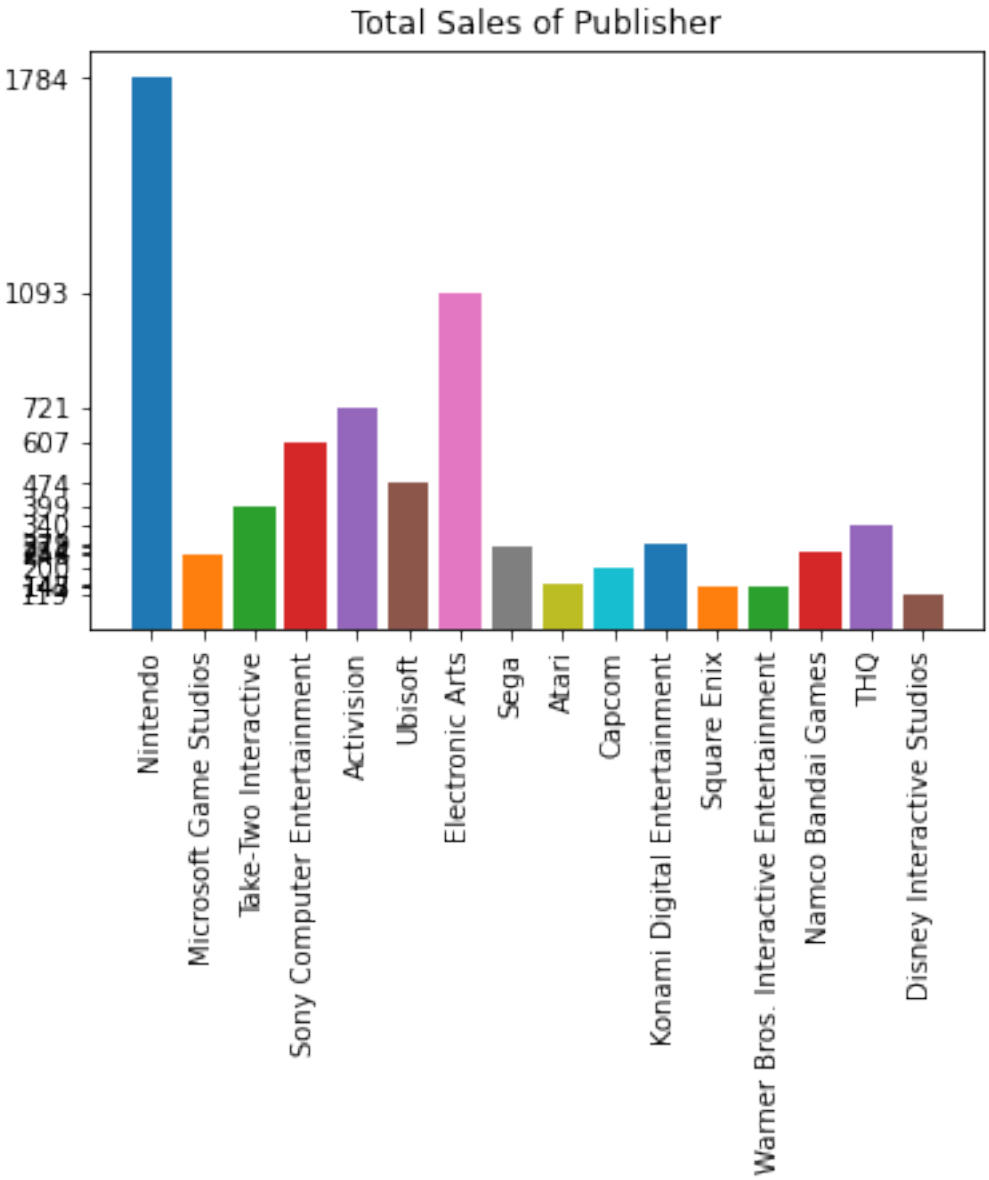
'Adeline Software' 'Funcom' 'Panther Software' 'Blast! Entertainment Ltd'
'Game Life' 'DSI Games' 'Avalon Interactive' 'Popcorn Arcade'
'Neko Entertainment' 'Vir2L Studios' 'Aques' 'Syscom'
'White Park Bay Software' 'System 3' 'Vatical Entertainment' 'Daedalic'
'EA Games' 'Media Factory' 'Vic Tokai' 'The Adventure Company'
'Game Arts' 'Broccoli' 'Acquire' 'General Entertainment'
'Excalibur Publishing' 'Imadio' 'Swing! Entertainment'
'Sony Music Entertainment' 'Aqua Plus' 'Paradox Interactive'
'Hip Interactive' 'DreamCatcher Interactive' 'Tripwire Interactive'
'Sting' 'Yacht Club Games' 'SCS Software' 'Bigben Interactive'
'Havas Interactive' 'Slitherine Software' 'Graffiti' 'Funsta' 'Telstar'
'U.S. Gold' 'DreamWorks Interactive' 'Data Design Interactive' 'MTO'
'DHM Interactive' 'FunSoft' 'SPS' 'Bohemia Interactive'
'Reef Entertainment' 'Tru Blu Entertainment' 'Moss' 'T&E Soft' 'O-Games'
'Aksys Games' 'NDA Productions' 'Data East' 'Time Warner Interactive'
'Gainax Network Systems' 'Daito' 'O3 Entertainment' 'Gameloft'
'Xicat Interactive' 'Simon & Schuster Interactive' 'Valcon Games'
'PopTop Software' 'TOHO' 'HMH Interactive' '5pb' 'Cave'
'CDV Software Entertainment' 'Microids' 'PM Studios' 'Paon' 'Micro Cabin'
'GameTek' 'Benesse' 'Type-Moon' 'Enjoy Gaming ltd.' 'Asmik Corp'
'Interplay Productions' 'Asmik Ace Entertainment' 'inXile Entertainment'
'Image Epoch' 'Phantom EFX' 'Evolved Games' 'responDESIGN'
'Culture Publishers' 'Griffin International' 'Hackberry' 'Hearty Robin'
'Nippon Amuse' 'Origin Systems' 'Seventh Chord' 'Mitsui' 'Milestone'
'Abylight' 'Flight-Plan' 'Glams' 'Locus' 'Warp' 'Daedalic Entertainment'
'Alternative Software' 'Myelin Media' 'Mercury Games'
'Irem Software Engineering' 'Sunrise Interactive' 'Elite'
'Evolution Games' 'Tivola' 'Global A Entertainment' 'Edia' 'Athena'
'Aria' 'Gamecock' 'Tommo' 'Altron' 'Happinet' 'iWin' 'Media Works'
'Fortyfive' 'Revolution Software' 'Imax' 'Crimson Cow' '10TACLE Studios'
'Groove Games' 'Pack-In-Video' 'Insomniac Games'
'Ascaron Entertainment GmbH' 'Asgard' 'Ecole' 'Yumedia' 'Phenomedia'
'HAL Laboratory' 'Grand Prix Games' 'DigiCube' 'Creative Core'
'Kaga Create' 'WayForward Technologies' 'LSP Games' 'ASCII Media Works'
'Coconuts Japan' 'Arika' 'Ertain' 'Marvel Entertainment' 'Prototype'
'Phantagram' '1C Company' 'The Learning Company' 'TechnoSoft' 'Vap'
'Misawa' 'Tradewest' 'Team17 Software' 'Yeti' 'Pow' 'Navarre Corp'
'MediaQuest' 'Max Five' 'Comfort' 'Monte Christo Multimedia'
'Pony Canyon' 'Riverhillsoft' 'Summitsoft' 'Milestone S.r.l' 'Playmore'

```
 'MLB.com' 'Kool Kizz' 'Flashpoint Games' '49Games' 'Legacy Interactive'
 'Alawar Entertainment' 'CyberFront' 'Cloud Imperium Games Corporation'
 'Societa' 'Virtual Play Games' 'Interchannel' 'Sonnet' 'Experience Inc.'
 'Zenrin' 'Iceberg Interactive' 'Ivolgamus' '2D Boy' 'MC2 Entertainment'
 'Kando Games' 'Just Flight' 'Office Create' 'Mamba Games' 'Fields'
 'Princess Soft' 'Maximum Family Games' 'Berkeley' 'Fuji'
 'Dusenberry Martin Racing' 'imageepoch Inc.' 'Big Fish Games'
 'Her Interactive' 'Kamui' 'ASK' 'TopWare Interactive' 'Headup Games'
 'KSS' 'Cygames' 'KID' 'Quinrose' 'Sunflowers' 'dramatic create' 'TGL'
 'Encore' 'Extreme Entertainment Group' 'Intergrow' 'G.Rev' 'Sweets'
 'Kokopeli Digital Studios' 'Number None' 'Nexon' 'id Software'
 'BushiRoad' 'Tryfirst' 'Strategy First' '7G//AMES' 'GN Software' "Yuke's"
 'Easy Interactive' 'Licensed 4U' 'FuRyu Corporation'
 'Lexicon Entertainment' 'Paon Corporation' 'Kids Station' 'GOA'
 'Graphsim Entertainment' 'King Records' 'Introversion Software'
 'Minato Station' 'Devolver Digital' 'Blue Byte' 'Gaga'
 'Yamasa Entertainment' 'Plenty' 'Views' 'fonfun' 'NetRevo'
 'Codemasters Online' 'Quintet' 'Phoenix Games' 'Dorart' 'Marvelous Games'
 'Focus Multimedia' 'Imageworks' 'Karin Entertainment' 'Aerosoft'
 'Technos Japan Corporation' 'Gakken' 'Mirai Shounen' 'Datam Polystar'
 'Saurus' 'HuneX' 'Revolution (Japan)' 'Giza10' 'Visco' 'Alvion' 'Mycom'
 'Giga' 'Warashi' 'System Soft' 'Sold Out' 'Lighthouse Interactive'
 'Masque Publishing' 'RED Entertainment' 'Michaelsoft'
 'Media Entertainment' 'New World Computing' 'Genterprise'
 'Interworks Unlimited, Inc.' 'Boost On' 'Stainless Games'
 'EON Digital Entertainment' 'Epic Games' 'Naxat Soft'
 'Ascaron Entertainment' 'Piacci' 'Nitroplus' 'Paradox Development'
 'Otomate' 'Ongakukan' 'Commseed' 'Inti Creates' 'Takuyo'
 'Interchannel_Holon' 'Rain Games' 'UIG Entertainment']
```

由于发布者数量过多，因此只将销售总额排名靠前的发布者信息进行可视化。结果显示：任天堂不愧是世界的主宰

```
In [167]: sale_of_publisher = {}
          for sub_df in Publishers_split:
              global_sale = np.sum(sub_df.Global_Sales)
              if global_sale>100:
                  publisher_name = sub_df['Publisher'].to_list()[0]
                  sale_of_publisher[publisher_name]=global_sale
```

```
keys = list(sale_of_publisher.keys())
values = list(sale_of_publisher.values())

for i,key in enumerate(sale_of_publisher):
    plt.bar(i,sale_of_publisher[key])
plt.xticks(np.arange(len(values)),keys,rotation=90)
plt.yticks(values)
# plt.grid('off')
plt.title('Total Sales of Publisher')
plt.show()
```



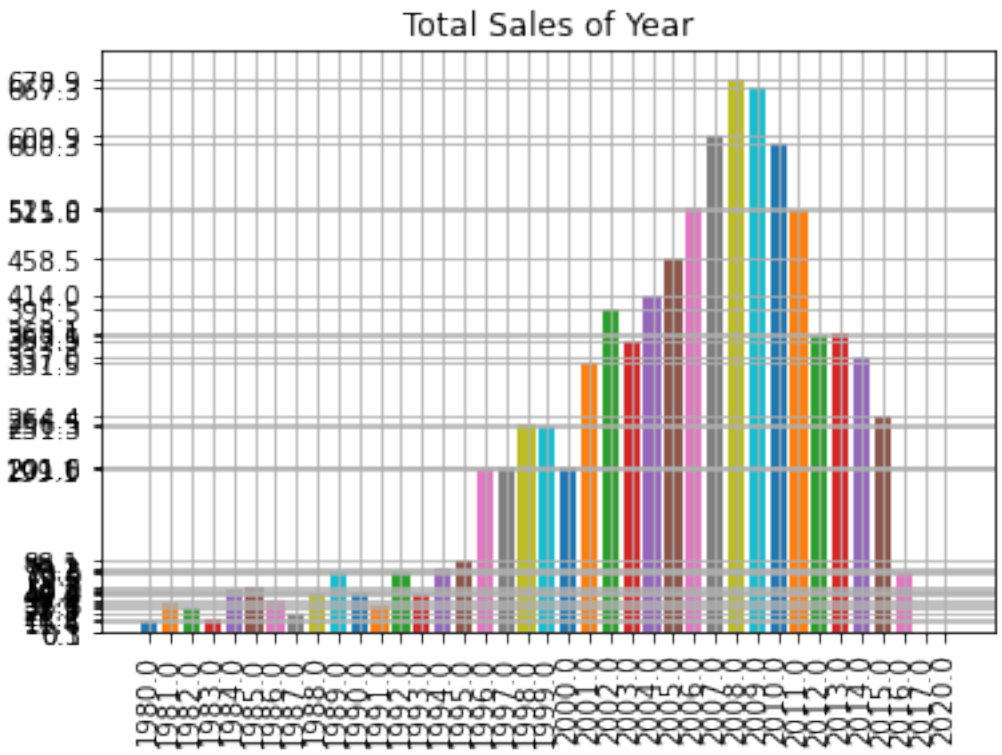
2. 预测每年电子游戏销售额：首先依然是按照年份划分子集，可以看到年份是从 1980 到 2020（缺少 2018，2019）

```
In [174]: df = df.sort_values(by='Year')
          years = df['Year'].unique() #sort_values(by=['Genre','Global_Sales'],ascending=False)
          print(years)
          years_split = []
          for idx,temp_G in enumerate(years):
              temp_data = df[df['Year'].isin([temp_G])]

              exec("df_%d = temp_data" %temp_G)
              years_split.append(temp_data)
          # idx = Publishers.tolist().index('Nintendo')
          # print(df_2015)
```

```
[1980. 1981. 1982. 1983. 1984. 1985. 1986. 1987. 1988. 1989. 1990. 1991.
 1992. 1993. 1994. 1995. 1996. 1997. 1998. 1999. 2000. 2001. 2002. 2003.
 2004. 2005. 2006. 2007. 2008. 2009. 2010. 2011. 2012. 2013. 2014. 2015.
 2016. 2017. 2020.]
```

可视化各年的所有游戏总销量：

```
In [175]: sale_of_year = {}
          for sub_df in years_split:
              global_sale = np.sum(sub_df.Global_Sales)
              year_name = sub_df['Year'].to_list()[0]
              sale_of_year[year_name]=global_sale
          # print(sale_of_genre)
          keys = list(sale_of_year.keys())
          values = list(sale_of_year.values())

          for i,key in enumerate(sale_of_year):
              plt.bar(i,sale_of_year[key])
          plt.xticks(np.arange(len(values)),keys,rotation=90)
          plt.yticks(values)
          plt.grid('on')
          plt.title('Total Sales of Year')
          plt.show()
```



下面对每年的电子游戏销售额进行回归与预测。这里分别使用线性 Logistic 模型和 AdaBoost 模型进行回归与预测。回归模型的输入 x 为年份，输出 y 为总销售额分别计算两个回归模型的拟合误差

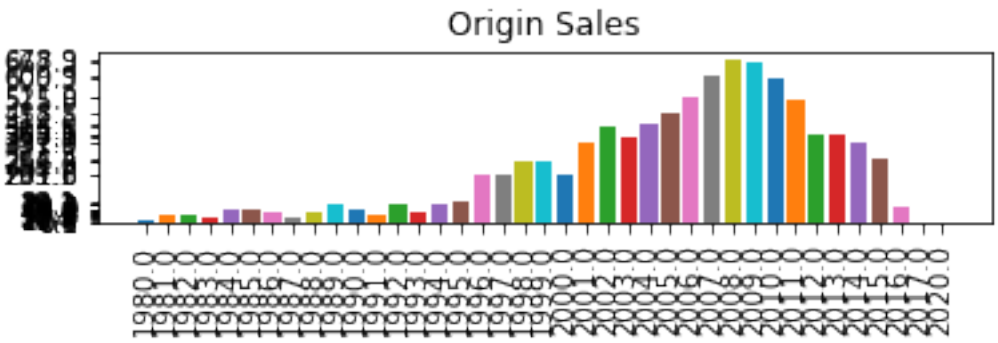
```
In [251]: list_ = list(zip(keys,values))
          df_years_sales = pd.DataFrame(list_,columns=['Year','Sales'])
          # df_years_sales.info()
          # print(list_)
          x = df_years_sales[['Year']].astype('int')
          y = df_years_sales['Sales'].astype('int')
          lr=LR()
          ada = ensemble.AdaBoostRegressor(n_estimators=30)
          ada.fit(x,y)
          lr.fit(x,y)
          # print("train complete")
          # print("correction rate:"+str(r2.score(x,y)))
          score = lr.score(x,y)
          predictions = lr.predict(x)
          error = predictions - y
          mae = abs(error).mean()
          print("mae_logistic = %d"%mae)
```

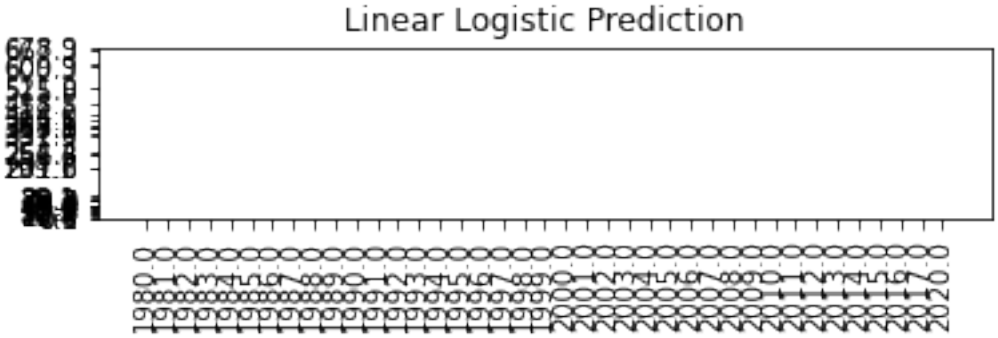
```
        score_ada = ada.score(x,y)
        predictions_ada = ada.predict(x)
        error_ada = predictions_ada - y
        mae_ada = abs(error_ada).mean()
        print("mae_adaboost = %d"%mae_ada)
```

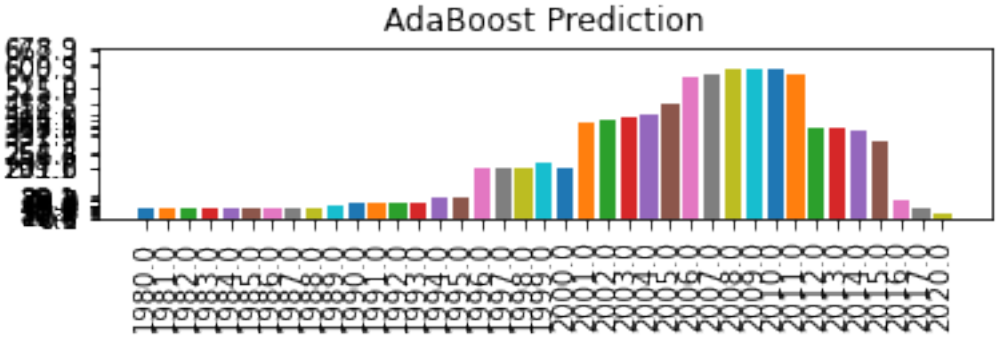
```
mae_logistic = 225
mae_adaboost = 21
```

从误差来看，线性回归模型误差较大，而 adaboost 模型误差较小。下面将预测结果进行可视化，可发现 logistic 模型不适用于此问题的回归分析，而 adaboost 对原数据的拟合度较高。

In [247]:
```
plt.subplot(311)
for i,key in enumerate(y.to_list()):
    plt.bar(i,key)
plt.xticks(np.arange(len(y)),df_years_sales['Year'].to_list(),rotation=90)
plt.yticks(values)
# plt.grid('on')
plt.title('Origin Sales')
plt.show()

plt.subplot(312)
for i,key in enumerate(predictions):
    plt.bar(i,key)
plt.xticks(np.arange(len(y)),df_years_sales['Year'].to_list(),rotation=90)
plt.yticks(values)
# plt.grid('on')
plt.title('Linear Logistic Prediction')
plt.show()

plt.subplot(313)
for i,key in enumerate(predictions_ada):
    plt.bar(i,key)
plt.xticks(np.arange(len(y)),df_years_sales['Year'].to_list(),rotation=90)
plt.yticks(values)
# plt.grid('on')
plt.title('AdaBoost Prediction')
plt.show()
```







下面分别用两种拟合的模型来预测原数据中缺少的 2018 年和 2019 年的销量：

```
In [248]: print(lr.predict([[2018]]))
          print(ada.predict([[2018]]))
```

```
[0]
[35.]
```

```
In [249]: print(lr.predict([[2019]]))
          print(ada.predict([[2019]]))
```

```
[0]
[23.33333333]
```

可以看出 Logistic 预测失效，而 Adaboost 预测出了较为可信的结果。

总体结论：从销售总额来看，1980 年至 2020 年的年销量呈现类正态分布趋势，最高销量年份为 2009 年。在此期间，最受欢迎的游戏为任天堂于 2006 年发行的 Wii Sports；人们普遍倾向于 PS 系列平台，以及 DS，X360，Wii 等平台；任天堂是众多游戏发行商的一枝独秀，其总销量难以被超越。