Title: Integrated Assessment (HDip for DA)

- Student Name: Eoin Cantwell
- Student Number: sba21230
- Submission Date: 02/06/2021

Assignment Introduction

The data set was found on Kaggle and it can be accessed through the following link;
https://www.kaggle.com/gregorut/videogamesales

- The data set is on the sales of worldwide video games that each sold 100k copies or more. Overall there are 11 feature columns that have both quantitative/qualitative data, there are a total of 16,598 observations and this data set has not been used by me before. All of which confirms the assignments specifications and requirements at the outset are met.

- This integrated assignment is based on a video game sales data set. The assignment covers two modules 1) Statistics 2) Data Preparation and there are a set of questions for each module. Overlap may occur but each question heading will specify what module/question is included for discussion.

- I chose this data set as when I was younger I was into playing computer games and so I found it very interesting to dig into what insights could be extracted from the data set.

Q1. Statistics / Q2 Data Preparation: Characterization of the Data Set

In [2]:
```
#Importing required libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import warnings
warnings.filterwarnings("ignore")
```

In [3]:
```
#Reading in the data set from where it is stored on my PC
#The data is housed in a .csv file (comma separated values)

df = pd.read_csv("data/vgsales.csv")
```

In [4]:
```
#Shows us the first 5 rows, this can be increased by inserting a number inside bracke
df.head()
```

Out[4]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Globa
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	

In [5]:
```
#Renaming columns into lower and snake case to simplify code writing and aid accuracy
new_names =  {'Rank': 'rank', 'Name':'name', 'Platform':'platform',
              'Year':'year', 'Genre':'genre', 'Publisher':'publisher',
              'NA_Sales': 'north_american_sales', 'EU_Sales': 'european_sales',
              'JP_Sales': 'japanese_sales', 'Other_Sales': 'other_sales',
              'Global_Sales': 'global_sales'}
```

In [6]:
```
#Passing through the renaming of the columns so it is in place
df.rename(columns=new_names, inplace=True)
```

In [7]:
```
#Increasing the frame width to the max width, this spaces columns nicely and data can
pd.set_option('display.max_colwidth', None)

df.head()
```

Out[7]:

	rank	name	platform	year	genre	publisher	north_american_sales	european_sales	japanese_sa
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10

The above output confirms the columns were successfully renamed

In [8]:
```
#Shows the number of observations and features the data set has
df.shape
```

Out[8]: (16598, 11)

Data Dictionary

- A data dictionary is a practical depiction of the metadata connected with a data object, such as a database, table, or column (OCTOPAI, 2019).The data dictionary helps give an understanding as to what each feature means and is very useful at the start of any data analysis project, for both the user and any person using/reading the data at a later date.

Feature	Description
rank	The order of each video game based on total sales
name	The name of the video game
platform	The machine the game can be played on
year	The year of release
genre	The type of the video game
publisher	The company that created the video game
north_american_sales	The total of North American video game sales
european_sales	The total of European video game sales
japanese_sales	The total of Japanese video game sales
other_sales	The total of other regions/countries video game sales
global_sales	The overall global total of video games sales

In [9]:
```
df.info()
```

```
&lt;class 'pandas.core.frame.DataFrame'&gt;
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   north_american_sales  16598 non-null  float64
 7   european_sales        16598 non-null  float64
 8   japanese_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
```

The above code shows the number of observations (rows), the number of columns (variables/features), what data type each feature is and finally the memory usage the data set demands from your pc.

In [10]:
```
#Creating new df to enable numeric values be used in the statistical function below
stats_df1 = df[['north_american_sales', 'european_sales', 'japanese_sales','other_sale

stats_df1.describe()
```

Out[10]:

	north_american_sales	european_sales	japanese_sales	other_sales	global_sales
count	16598.000000	16598.000000	16598.000000	16598.000000	16598.000000
mean	0.264667	0.146652	0.077782	0.048063	0.537441
std	0.816683	0.505351	0.309291	0.188588	1.555028
min	0.000000	0.000000	0.000000	0.000000	0.010000
25%	0.000000	0.000000	0.000000	0.000000	0.060000
50%	0.080000	0.020000	0.000000	0.010000	0.170000
75%	0.240000	0.110000	0.040000	0.040000	0.470000
max	41.490000	29.020000	10.220000	10.570000	82.740000

The above table shows key statistical information regarding our quantitative data columns.

In [156…
```
#Show data types that are objects
df.describe(include='object')
```

Out[156…

	name	platform	genre	publisher
count	16598	16598	16598	16540
unique	11493	31	12	578
top	Need for Speed: Most Wanted	DS	Action	Electronic Arts
freq	12	2163	3316	1351

The above code gives an understanding of the categorical data in the dataset and is useful for future analysis. Insights gained, 31 different platforms/devices the video games are played on. There are 12 different genre types. Need for Speed: Most Wanted is the game that appears most.

Mode

- The following mode functions confirm what we is seen in the above table

In [10]:
```
trend = df["platform"].mode()
print(trend)
```

```
0    DS
dtype: object
```

In [11]:
```
trend = df["year"].mode()
print(trend)
```

```
0    2009.0
dtype: float64
```

In [12]:
```
trend = df["genre"].mode()
print(trend)
```

```
0    Action
dtype: object
```

In [13]:
```
trend = df["publisher"].mode()
print(trend)
```

```
0    Electronic Arts
dtype: object
```

The above functions show the mode of each categorical data type, the results are that Nintendo DS is the most frequent platform for the video games. The year where most video games are sold is 2009. The favourite genre is action and the the most frequent publisher of video games is the company Electronic Arts.

In [158…
```
#This code reveals if there is any correlation between numerical variables
stats_df2 = df[['north_american_sales', 'european_sales', 'japanese_sales','other_sale
stats_df2.corr()
```

Out[158…

	north_american_sales	european_sales	japanese_sales	other_sales	global_sales
north_american_sales	1.000000	0.767727	0.449787	0.634737	0.941047
european_sales	0.767727	1.000000	0.435584	0.726385	0.902836
japanese_sales	0.449787	0.435584	1.000000	0.290186	0.611816
other_sales	0.634737	0.726385	0.290186	1.000000	0.748331
global_sales	0.941047	0.902836	0.611816	0.748331	1.000000

In [159…
```
#Shows how spread out the data is
skewness = df.skew()
print(skewness)
```

```
rank                     0.000066
year                    -1.002560
north_american_sales    18.799627
european_sales          18.875535
japanese_sales          11.206458
other_sales             24.233923
global_sales            17.400645
dtype: float64
```

In [160…
```
df['year'].min()
```

Out[160… 1980.0

In [161…
```
df['year'].max()
```

Out[161… 2020.0

The min/max function reveals the dataset covers 40 years of video game sales which supports why there are so many observations.

In [162…
```
#Z Score

from scipy import stats
z = df[['north_american_sales', 'european_sales', 'japanese_sales','other_sales', 'gl
print(z)
```

```
       north_american_sales  european_sales  japanese_sales  other_sales  \
0                     41.49           29.02            3.77         8.46
1                     29.08            3.58            6.81         0.77
2                     15.85           12.88            3.79         3.31
3                     15.75           11.01            3.28         2.96
4                     11.27            8.89           10.22         1.00
...                     ...             ...             ...          ...
16593                  0.01            0.00            0.00         0.00
16594                  0.01            0.00            0.00         0.00
16595                  0.00            0.00            0.00         0.00
16596                  0.00            0.01            0.00         0.00
16597                  0.01            0.00            0.00         0.00

       global_sales
0             82.74
1             40.24
2             35.82
3             33.00
4             31.37
...             ...
16593          0.01
16594          0.01
16595          0.01
16596          0.01
16597          0.01

[16598 rows x 5 columns]
```

A Z-score is a numerical measurement that calls out a value's relationship to the mean of a group of values. Z-score is calculated in terms of standard deviations from the mean.

A Z-score equalling to 0 indicates that the data point's score is identical to the mean score. A Z-score of 1 means a value that is one standard deviation from the mean.

Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean (Hayes, 2021).

Q1. Statistics / Q2 Data Preparation: Exploratory Data Analysis

In [18]:
```
#Investigate if there is any null values present
df.isnull().sum()
```

Out[18]:
```
rank                      0
name                      0
platform                  0
year                    271
genre                     0
publisher                58
north_american_sales      0
european_sales            0
japanese_sales            0
other_sales               0
global_sales              0
dtype: int64
```

Two features have null values, year and publisher, there are multiple ways to deal with null values, I am going to drop the rows that contain them.

In [163…
```
#Dropping Null values
df.dropna(subset=['year', 'publisher'], inplace=True)
```

In [164…
```
#Confirming null values have been dropped
df.isnull().sum()
```

Out[164…
```
rank                    0
name                    0
platform                0
year                    0
genre                   0
publisher               0
north_american_sales    0
european_sales          0
japanese_sales          0
other_sales             0
global_sales            0
dtype: int64
```

In [166…
```
#Verifying the new shape of the dataset after null values and rows were dropped
df.shape
```

Out[166… (16291, 11)

In [167…
```
#Check for duplicate data
duplicate = df.duplicated()
print(duplicate.sum())
df[duplicate]
```

0

Out[167…

rank	name	platform	year	genre	publisher	north_american_sales	european_sales	japanese_sales	other_s

There are no duplicates recorded, therefore nothing further needs to be done

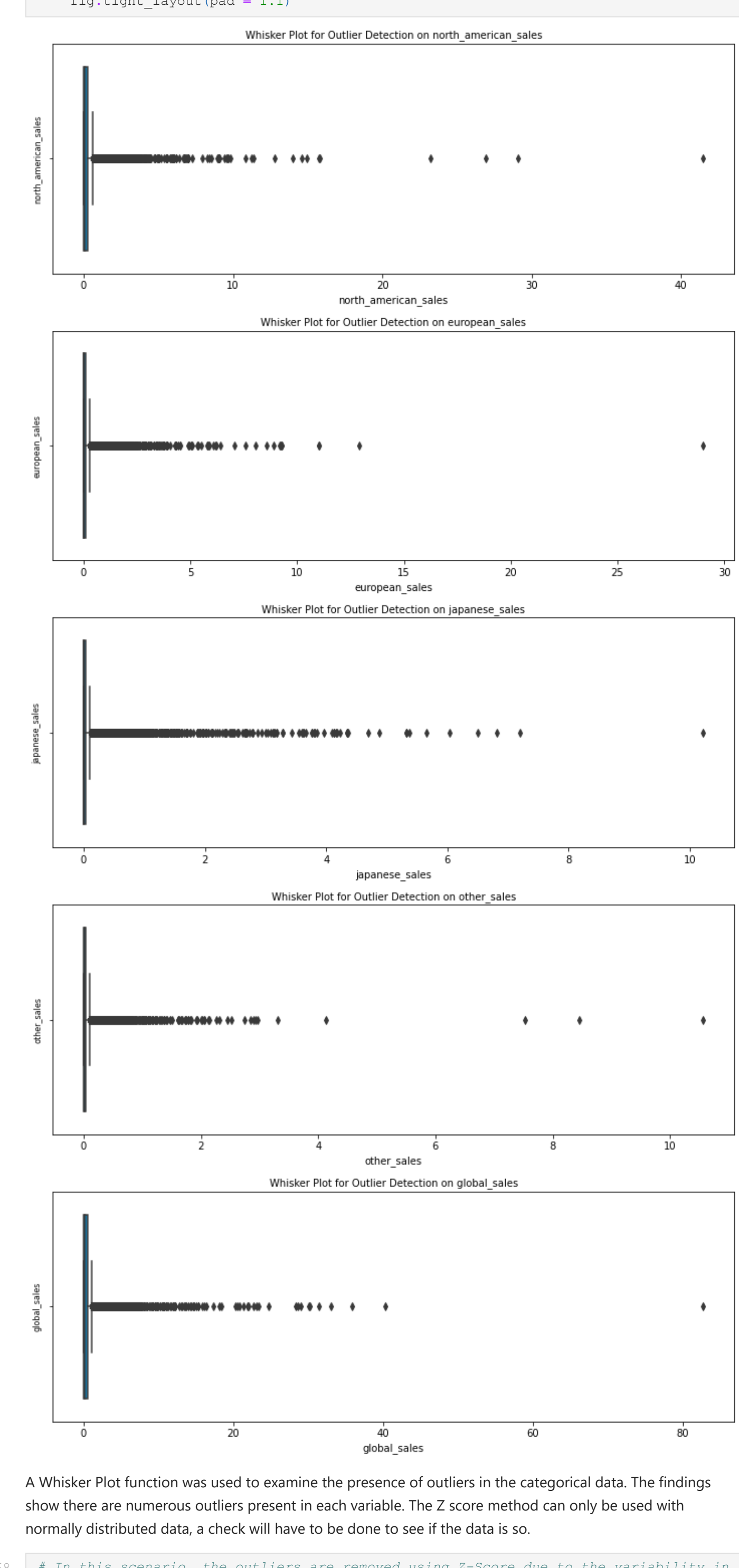
Outlier Detection and Handling

- Feature Engineering

Feature engineering is the initial stage in a machine learning channel and it comprises all the techniques utilised to clean existing datasets, improve their signal-noise ratio, and reduce their dimensionality (Bonaccorso, 2017).

In [22]:
```
# Create a list for numerical columns that are to be visualized
Column_List = ['north_american_sales', 'european_sales', 'japanese_sales','other_sales
```

In [23]:
```
# Generate whisker plots to detect the presence of any outliers
fig, ax = plt.subplots (len(Column_List), figsize = (10, 20))

for i, col_list in enumerate(Column_List):
    sns.boxplot(df[col_list], ax = ax[i], palette = "winter", orient = 'h')
    ax[i].set_title("Whisker Plot for Outlier Detection on" + " " + col_list, fontsiz
    ax[i].set_ylabel(col_list, fontsize = 8)
    fig.tight_layout(pad = 1.1)
```



A Whisker Plot function was used to examine the presence of outliers in the categorical data. The findings show there are numerous outliers present in each variable. The Z score method can only be used with normally distributed data, a check will have to be done to see if the data is so.

In [168…
```
# In this scenario, the outliers are removed using Z-Score due to the variability in
#df = df[(np.abs(stats.zscore(df[['north_american_sales', 'european_sales',
                                  #'japanese_sales','other_sales',
                                 # 'global_sales']])) < 3).all(axis = 1)] # a
#df = df.reset_index() # Due to elimination of rows, index has to be reset
```

In [169…
```
#Compare the new dimension with the old one, the old shape of data set was (16598, 11
#so we see 777 values have been removed.
#df.shape
```

In [26]:
```
#Change dates to remove decimal points, maybe truncate is better and then to pd datet
df['year']=df['year'].astype(int)
```

In [27]:
```
#Placing Rank as the Index
#df.set_index('rank', inplace=True)
#df.head()
```

In [28]:
```
df1 = df[['genre', 'platform', 'north_american_sales', 'european_sales', 'japanese_sa
```

In [29]:
```
results1 = df1.groupby(['genre']).sum()
results1
```

Out[29]:

genre	north_american_sales	european_sales	japanese_sales	other_sales	global_sales
Action	861.77	516.48	158.65	184.92	1722.84
Adventure	101.93	63.74	51.99	16.70	234.59
Fighting	220.74	100.00	87.15	36.19	444.05
Misc	396.92	211.77	106.67	73.92	789.87
Platform	445.99	200.65	130.65	51.51	829.13
Puzzle	122.01	50.52	56.68	12.47	242.21
Racing	356.93	236.31	56.61	76.68	726.76
Role-Playing	326.50	187.57	350.29	59.38	923.83
Shooter	575.16	310.45	38.18	101.90	1026.20
Simulation	181.78	113.02	63.54	31.36	389.98
Sports	670.09	371.34	134.76	132.65	1309.24
Strategy	67.83	44.84	49.10	11.23	173.27

In [30]:
```
#df1 = df1[['Platform', 'Year', 'NA_Sales', 'EU_Sales', 'JP_Sales', 'Other_Sales', 'G
```

In [171…
```
#Examining platform sales
results2 = df1.groupby(['platform']).sum()
results2
```

Out[171…

platform	north_american_sales	european_sales	japanese_sales	other_sales	global_sales
2600	80.78	4.86	0.00	0.84	86.57
3DO	0.00	0.00	0.10	0.00	0.10
3DS	78.03	58.29	97.30	12.55	246.27
DC	5.43	1.69	8.56	0.27	15.97
DS	388.53	194.05	175.02	60.29	818.91
GB	113.64	47.51	85.12	8.16	254.42
GBA	178.43	72.49	46.56	7.51	305.62
GC	131.94	38.32	21.34	5.13	197.14
GEN	19.27	5.52	2.67	0.89	28.36
GG	0.00	0.00	0.04	0.00	0.04
N64	138.91	41.03	33.76	4.31	218.21
NES	125.94	21.15	98.65	5.31	251.07
NG	0.00	0.00	1.44	0.00	1.44
PC	92.04	137.35	0.17	24.33	254.70
PCFX	0.00	0.00	0.03	0.00	0.03
PS	334.71	212.38	139.78	40.69	727.39
PS2	572.92	332.63	137.54	190.47	1233.46
PS3	388.90	340.47	79.21	140.81	949.35
PS4	96.80	123.70	14.30	43.36	278.10
PSP	107.09	67.16	75.89	41.52	291.71
PSV	16.07	16.27	20.86	8.41	61.60
SAT	0.72	0.54	32.26	0.07	33.59
SCD	1.00	0.36	0.45	0.05	1.87
SNES	61.23	19.04	116.55	3.22	200.05
TG16	0.00	0.00	0.16	0.00	0.16
WS	0.00	0.00	1.42	0.00	1.42
Wii	497.37	264.35	68.28	79.20	909.81
WiiU	38.32	24.23	12.79	6.45	81.86
X360	594.33	278.00	12.30	84.67	969.60
XB	182.06	59.65	1.38	8.48	252.09
XOne	83.19	45.65	0.34	11.92	141.06

In [170…
```
#Examining year with highest global sales
df2 = df[['year', 'global_sales']]
```

In [33]:
```
results3 = df2.groupby(['year'])['global_sales'].sum().sort_values(ascending=False)
results3 = results3.to_frame()
results3
```

Out[33]:

year	global_sales
2008	678.90
2009	667.30
2007	609.92
2010	600.29
2006	521.04
2011	515.80
2005	458.51
2004	414.01
2002	395.52
2013	368.11
2012	363.49
2003	357.85
2014	337.03
2001	331.47
2015	264.44
1998	256.47
1999	251.27
2000	201.56
1997	200.98
1996	199.15
1995	88.11
1994	79.17
1992	76.16
1989	73.45
2016	70.90
1985	53.94
1984	50.36
1990	49.39
1988	47.22
1993	45.98
1986	37.07
1981	35.77
1991	32.23
1982	28.86
1987	21.74
1983	16.79
1980	11.38
2020	0.29
2017	0.05

In [172…
```
#Using the .loc function to select one year and reveal that years data
df.loc[df.year == 2009]
```

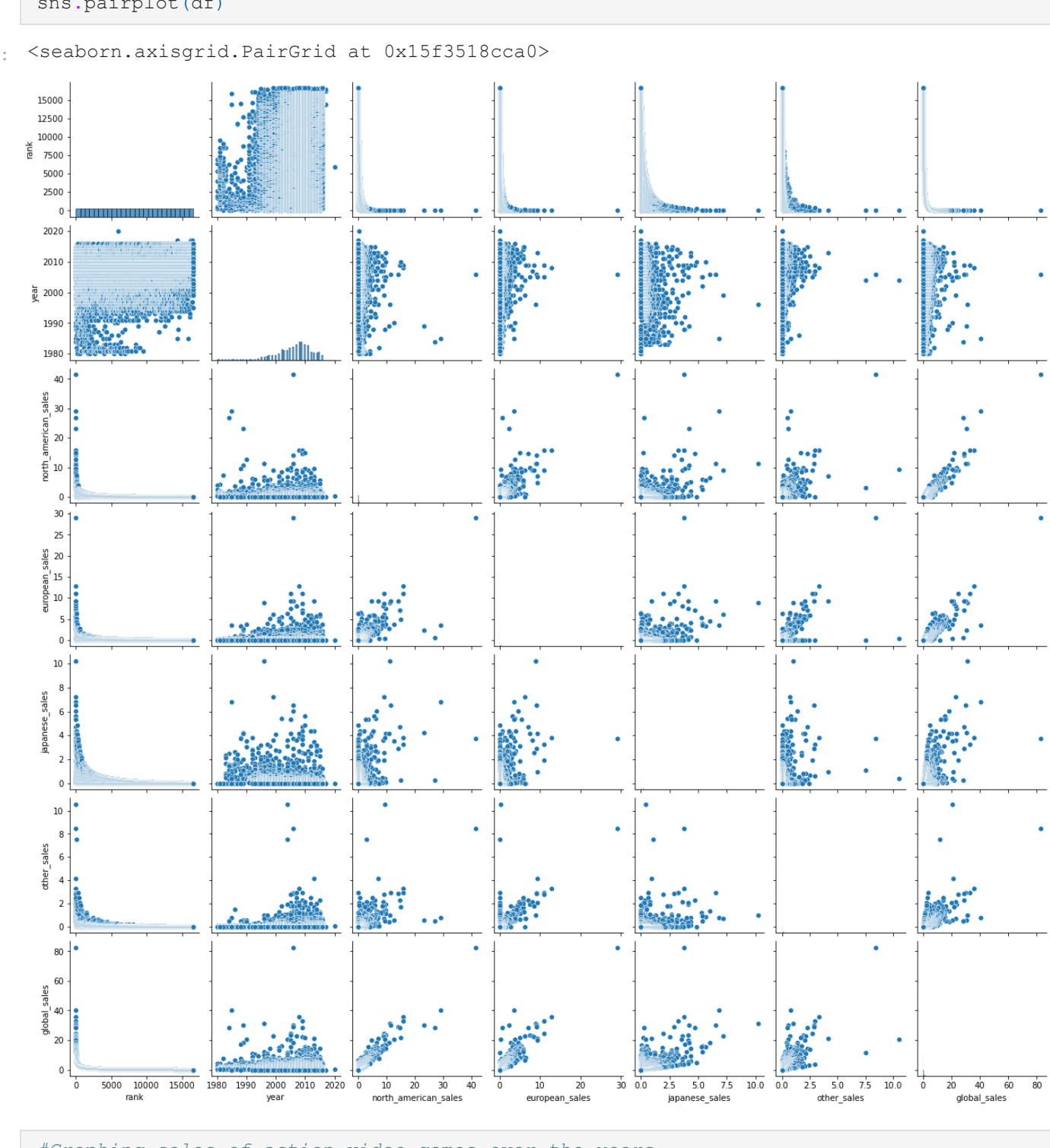
Out[172…

	rank	name	platform	year	genre	publisher	north_american_sales	european_s
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	1
8	9	New Super Mario Bros. Wii	Wii	2009.0	Platform	Nintendo	14.59	
14	15	Wii Fit Plus	Wii	2009.0	Sports	Nintendo	9.09	
36	37	Call of Duty: Modern Warfare 2	X360	2009.0	Shooter	Activision	8.52	
45	46	Pokemon HeartGold/Pokemon SoulSilver	DS	2009.0	Action	Nintendo	4.40	
...	...	...	...	...	...	...	...	
16540	16543	15 Days	PC	2009.0	Adventure	DTP Entertainment	0.00	
16551	16554	Sweet Honey Coming	PS2	2009.0	Adventure	Kadokawa Shoten	0.00	
16577	16580	Damnation	PC	2009.0	Shooter	Codemasters	0.00	
16586	16589	Secret Files 2: Puritas Cordis	DS	2009.0	Adventure	Deep Silver	0.00	
16587	16590	Mezase!! Tsuri Master DS	DS	2009.0	Sports	Hudson Soft	0.00	

1431 rows × 11 columns

In [35]:
```
#pairwise scatter plot
sns.pairplot(df)
```

Out[35]: <seaborn.axisgrid.PairGrid at 0x15f3518cca0>

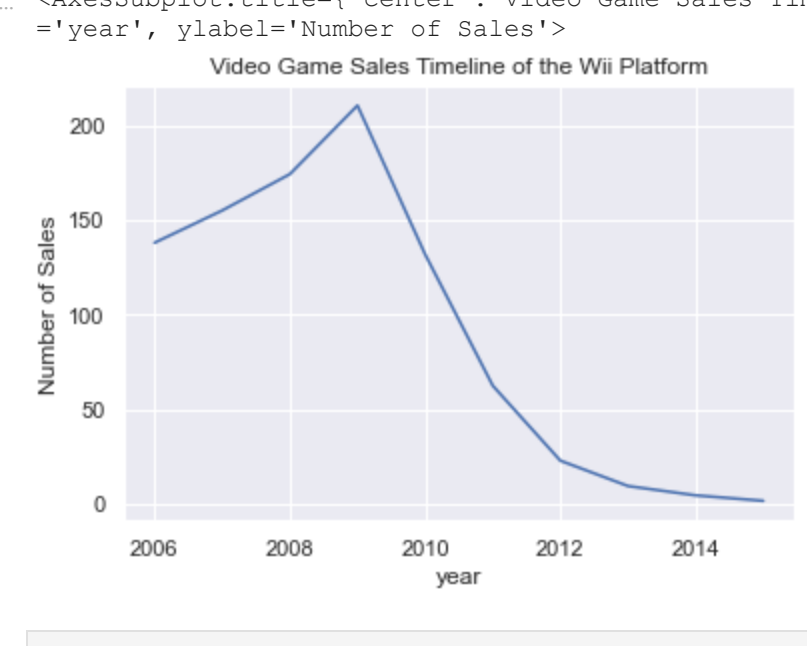


In [173…
```
#Graphing sales of action video games over the years
plt.title("Yearly Sales of Action Video Games")
plt.ylabel("Number of Sales")
df[df["genre"] == "Action"].groupby("year")["global_sales"].sum().plot()
```

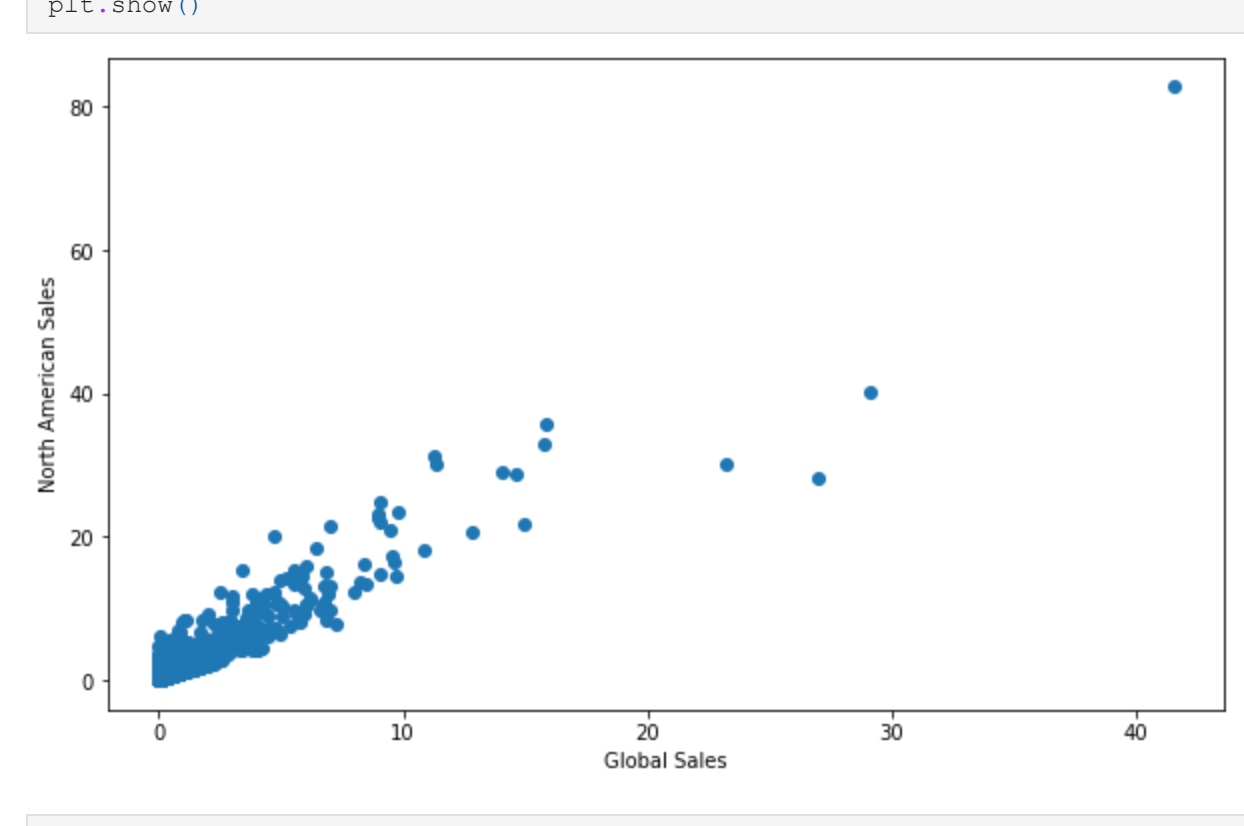
Out[173… <AxesSubplot:title={'center':'Yearly Sales of Action Video Games'}, xlabel='year', ylabel='Number of Sales'>

In [174…
```
#Selecting the platform "Wii" and examining its sales timeline
plt.title("Video Game Sales Timeline of the Wii Platform")
plt.ylabel("Number of Sales")
df[df["platform"] == "Wii"].groupby("year")["global_sales"].sum().plot()
```

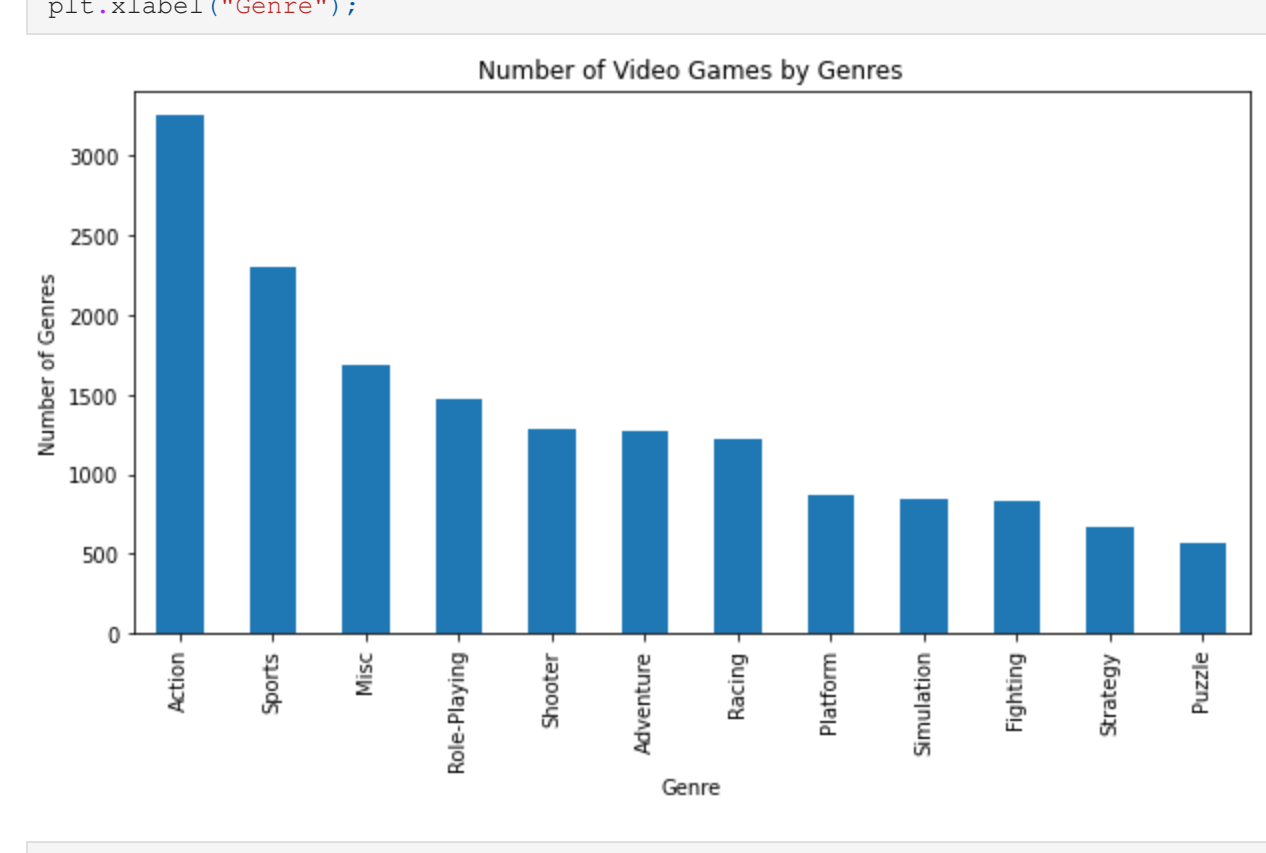
Out[174… <AxesSubplot:title={'center':'Video Game Sales Timeline of the Wii Platform'}, xlabel='year', ylabel='Number of Sales'>



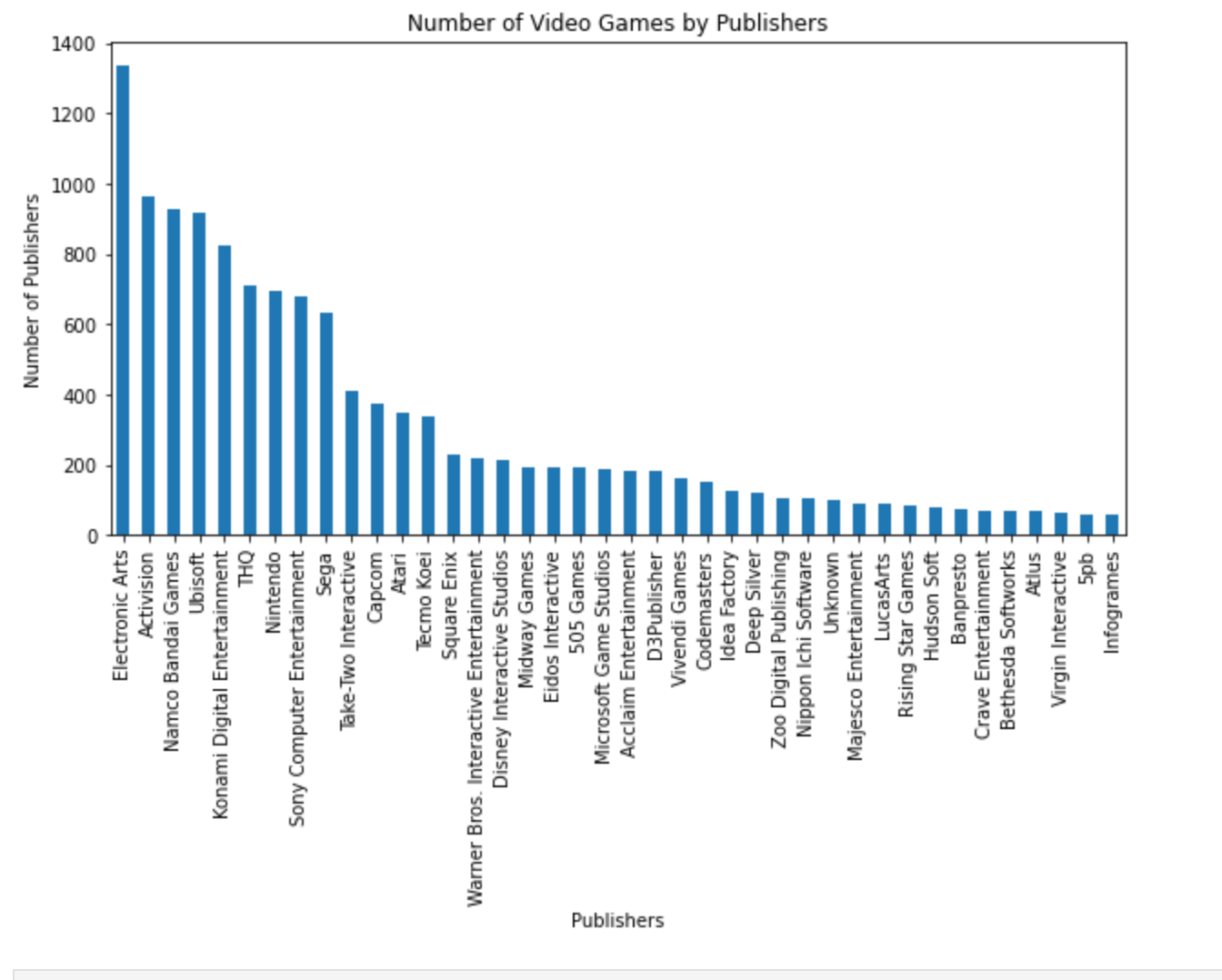
In [38]:
```
#Understanding the relationship of north american sales and overall global sales
fig, ax = plt.subplots(figsize=(10,6))
ax.scatter(df["north_american_sales"], df["global_sales"])
ax.set_xlabel("Global Sales")
ax.set_ylabel("North American Sales")
plt.show()
```



In [39]:
```
#Displaying the differences in total sales between genres
df.genre.value_counts().nlargest(40).plot(kind="bar", figsize=(10,5))
plt.title("Number of Video Games by Genres")
plt.ylabel("Number of Genres")
plt.xlabel("Genre");
```

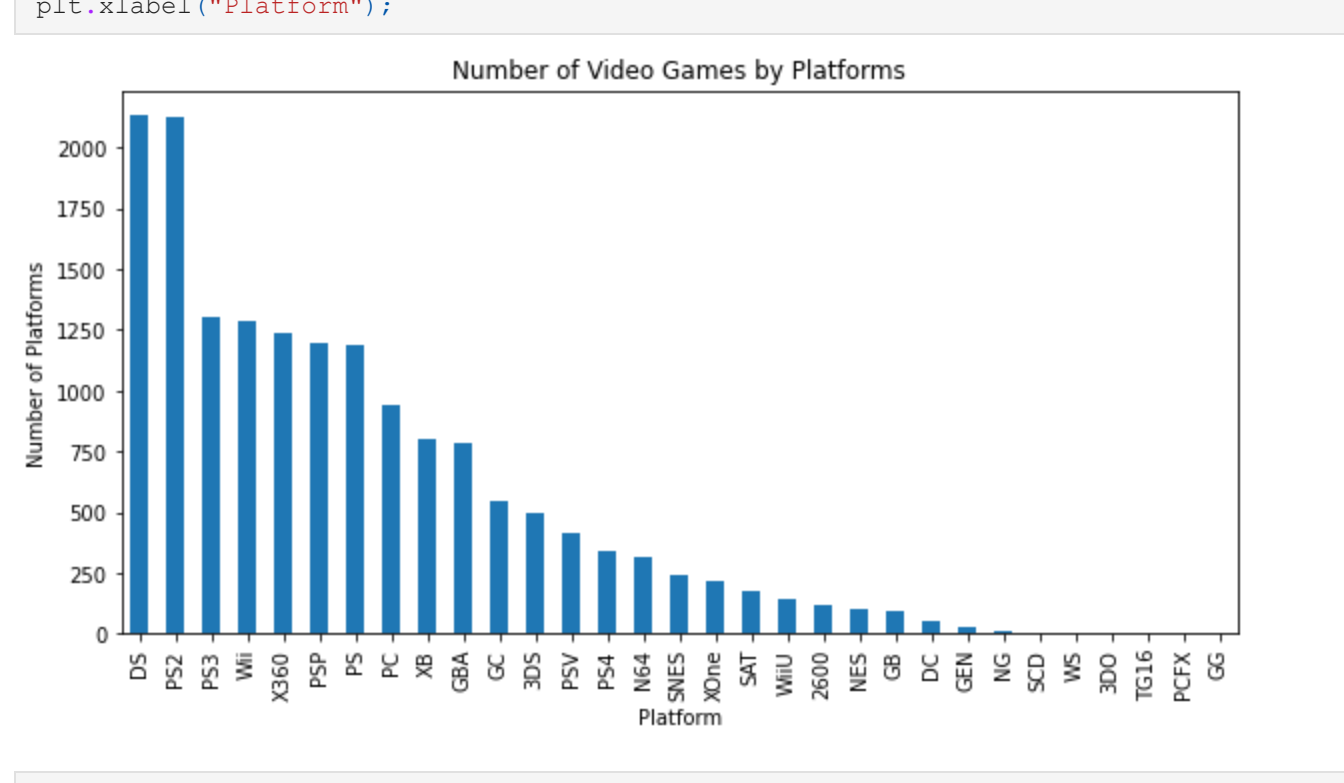


In [40]:
```
#Displaying the differences in total sales between publishers
df.publisher.value_counts().nlargest(40).plot(kind="bar", figsize=(10,5))
plt.title("Number of Video Games by Publishers")
plt.ylabel("Number of Publishers")
plt.xlabel("Publishers");
```



In [41]:

```
#Displaying the differences in total sales between platforms
df.platform.value_counts().nlargest(40).plot(kind="bar", figsize=(10,5))
plt.title("Number of Video Games by Platforms")
plt.ylabel("Number of Platforms")
plt.xlabel("Platform");
```



In [42]:

```
#Video Game Names: first 20 entries
game_name = df[['name']]
game_name[:20]
```

Out[42]:

	name
0	Wii Sports
1	Super Mario Bros.
2	Mario Kart Wii
3	Wii Sports Resort
4	Pokemon Red/Pokemon Blue
5	Tetris
6	New Super Mario Bros.
7	Wii Play
8	New Super Mario Bros. Wii
9	Duck Hunt
10	Nintendogs
11	Mario Kart DS
12	Pokemon Gold/Pokemon Silver
13	Wii Fit
14	Wii Fit Plus
15	Kinect Adventures!
16	Grand Theft Auto V
17	Grand Theft Auto: San Andreas
18	Super Mario World
19	Brain Age: Train Your Brain in Minutes a Day

In [175…

```
#Attempt at extracting video game names

#Method 1: str.split() by space
video_game_names = df['name'].apply(lambda x: x.split(" ")[0])
video_game_names[:20]
```

Out[175…

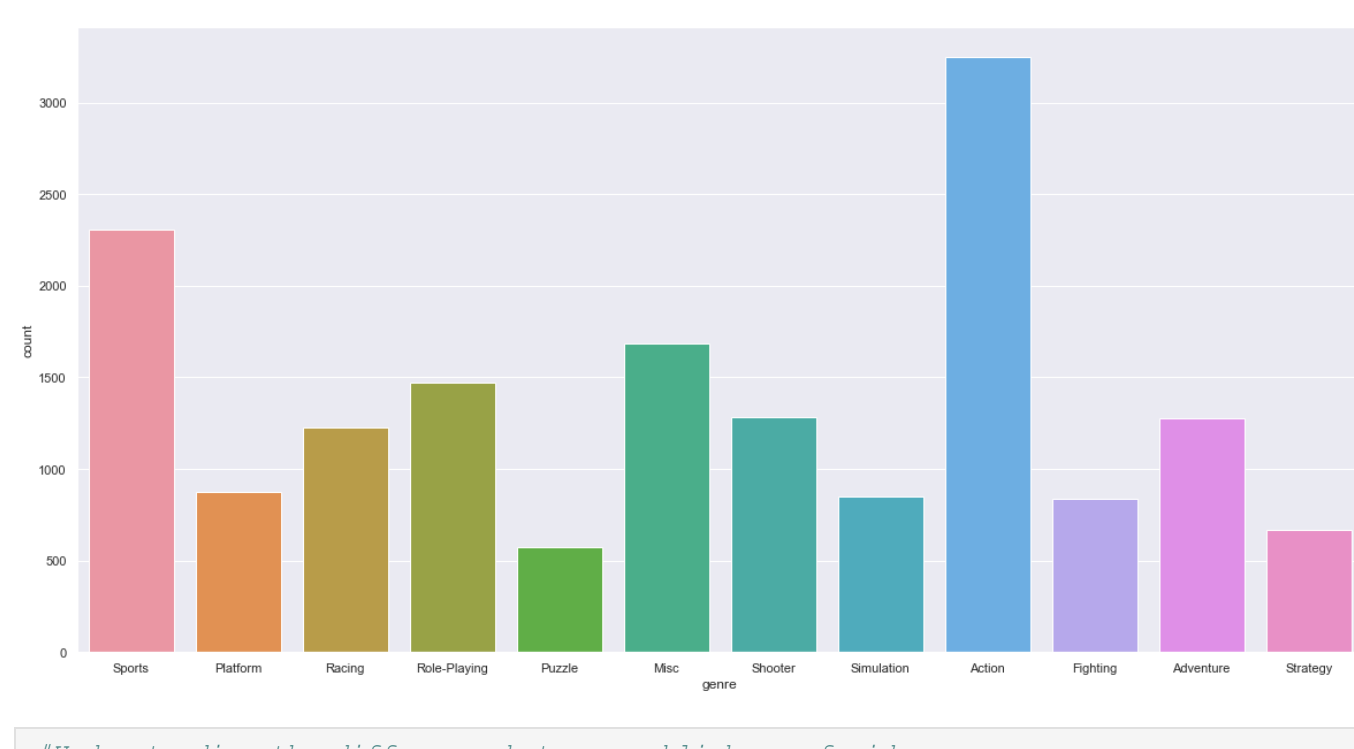
```
0            Wii
1          Super
2          Mario
3            Wii
4        Pokemon
5         Tetris
6            New
7            Wii
8            New
9           Duck
10    Nintendogs
11         Mario
12       Pokemon
13           Wii
14           Wii
15        Kinect
16         Grand
17         Grand
18         Super
19         Brain
Name: name, dtype: object
```

Result = Fail. The results show it is not possible to identify and understand the games name this way

In [180…

```
#Displaying the differences in total sales between genres using count function
plt.figure(figsize=(20,10))
sns.countplot(x='genre', data=df)
```

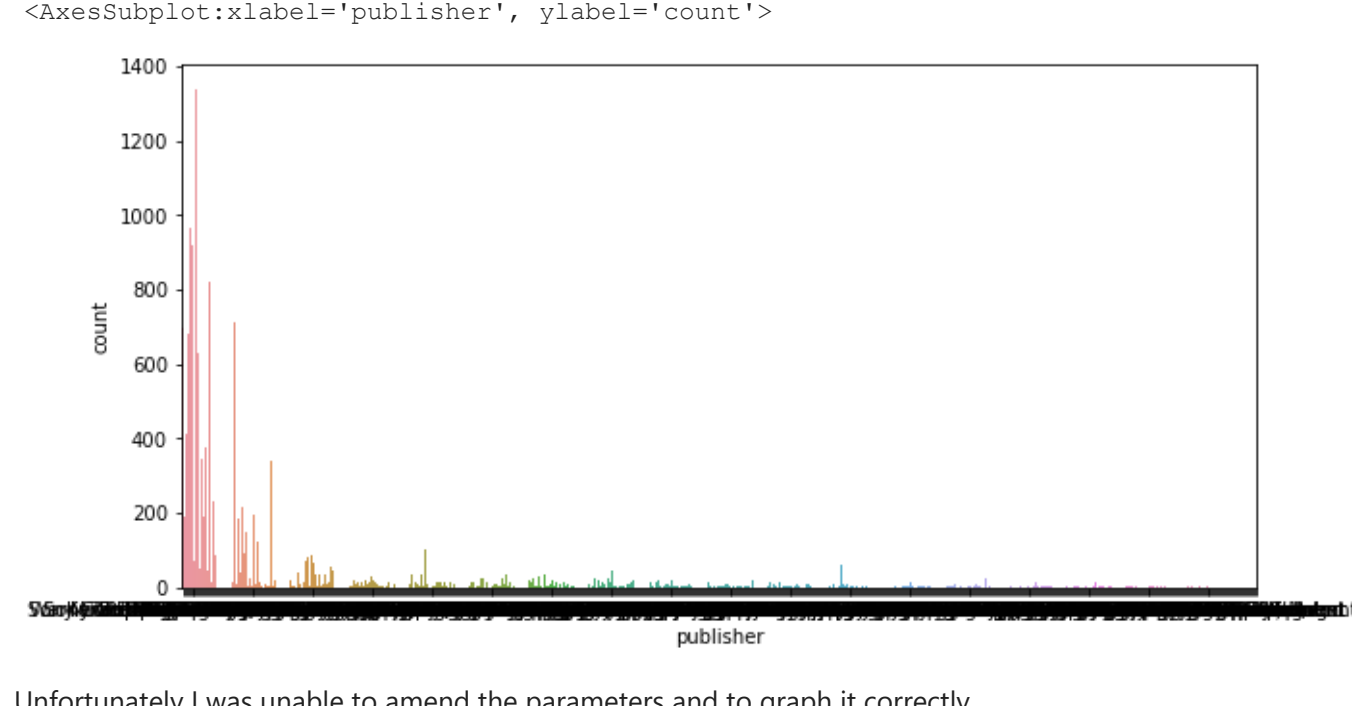
Out[180… `<AxesSubplot:xlabel='genre', ylabel='count'>`



In [45]:

```
#Understanding the difference between publishers of video games
plt.figure(figsize=(10,5))
sns.countplot(x='publisher', data=df)
```

Out[45]: `<AxesSubplot:xlabel='publisher', ylabel='count'>`



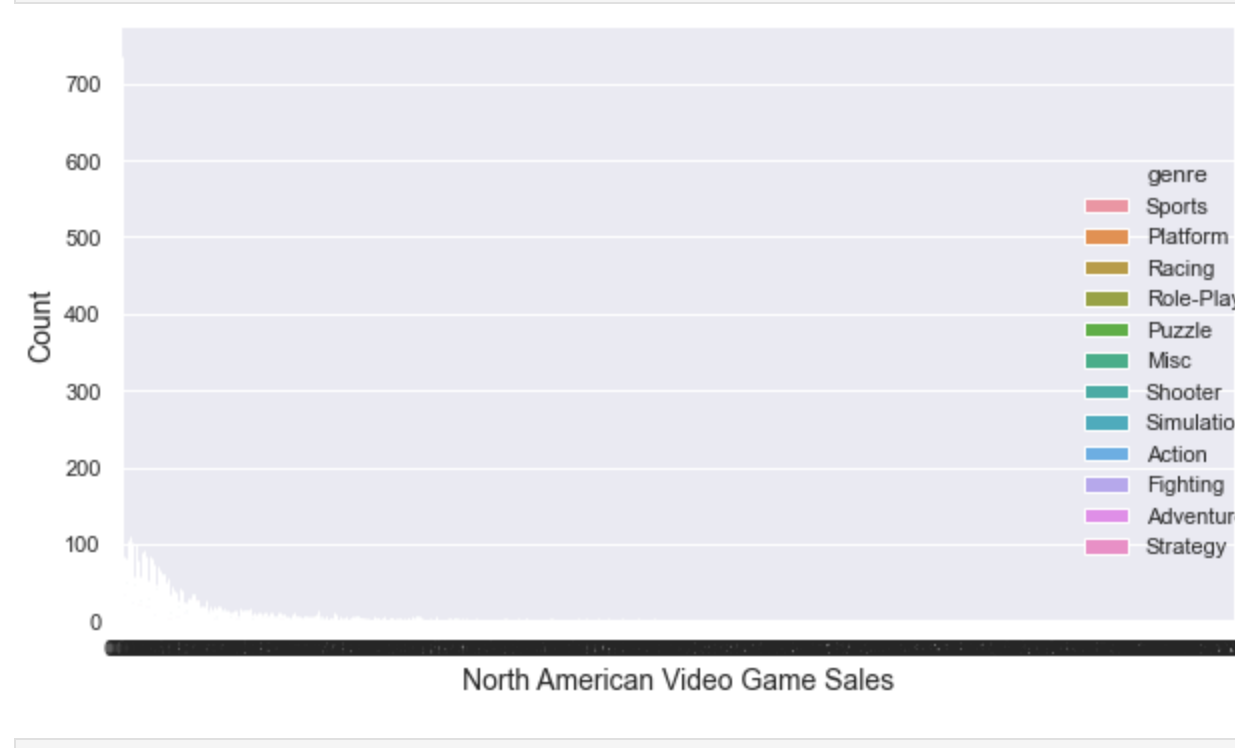
Unfortunately I was unable to amend the parameters and to graph it correctly

In [181…

```
df3 = df[['genre', 'north_american_sales', 'european_sales', 'japanese_sales', 'other_
```

In [186…

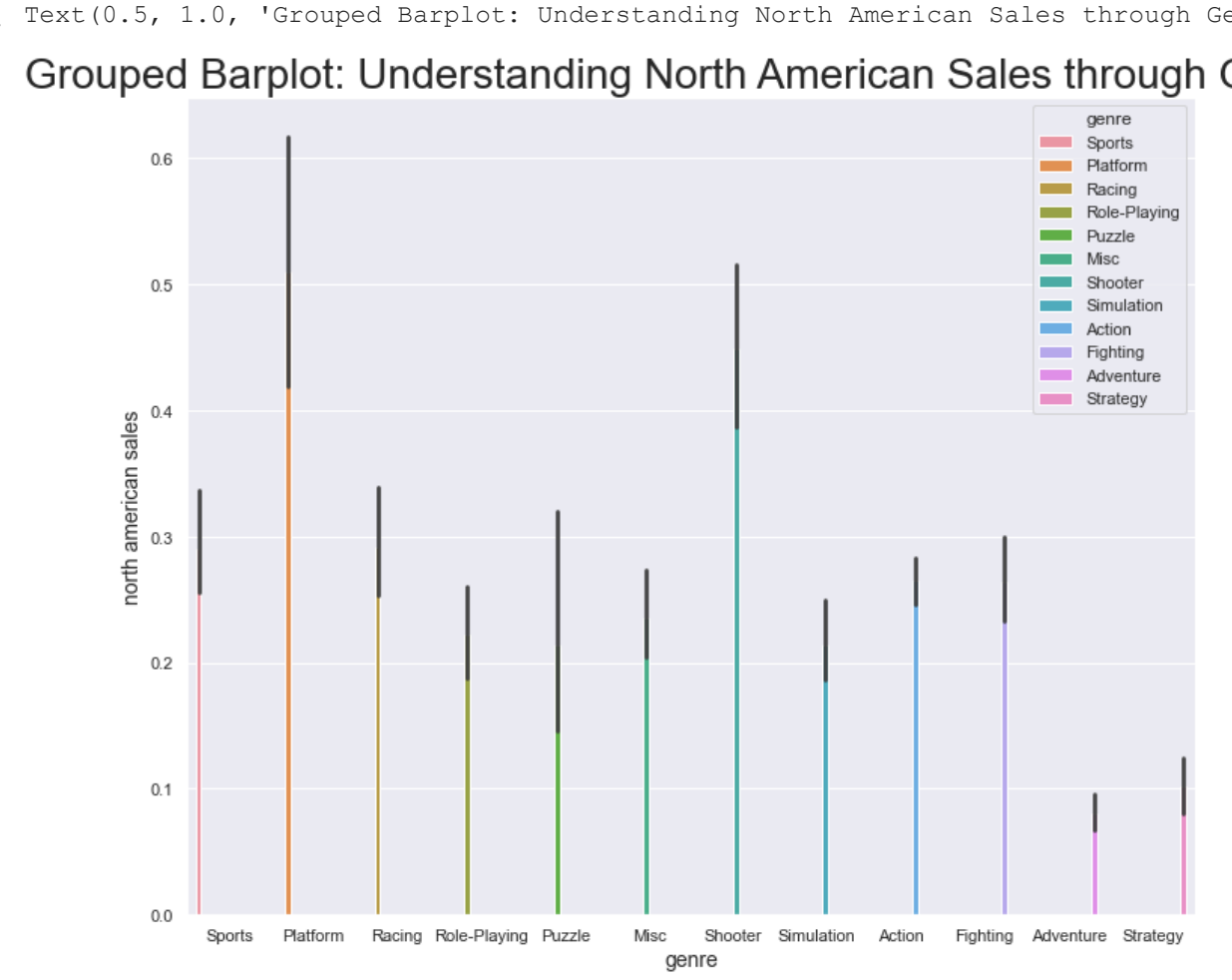
```
sns.catplot( x='north_american_sales',
             kind="count",
             hue="genre",
             height=5,
             aspect=1.5,
             data=df3)
plt.xlabel("North American Video Game Sales", size=14)
plt.ylabel("Count", size=14)
plt.tight_layout()
```



In [185…

```
plt.figure(figsize=(12, 10))
sns.barplot(x="genre",
            y="north_american_sales",
            hue="genre",
            data=df3)
plt.ylabel("north american sales", size=14)
plt.xlabel("genre", size=14)
plt.title("Grouped Barplot: Understanding North American Sales through Genres", size=
```

Out[185… `Text(0.5, 1.0, 'Grouped Barplot: Understanding North American Sales through Genres')`



Normalizing & Scaling

In [49]:

```
#Scales the data. Returns the z-scores of every attribute
# z-score:z-score (also called a standard score) gives you an idea of how far from the
# Nut more technically its a measure of how many standard deviations below or above t
#from sklearn.preprocessing import StandardScaler
#std_scale = StandardScaler()
#std_scale
```

In [50]:

```
#vg_df['north american sales'] = std_scale.fit_transform(vg_df[['north american sales
#vg_df['european sales'] = std_scale.fit_transform(vg_df[['european sales']])
#vg_df['japanese sales'] = std_scale.fit_transform(vg_df[['japanese sales']])
#vg_df['other sales'] = std_scale.fit_transform(vg_df[['other sales']])
#vg_df['global sales'] = std_scale.fit_transform(vg_df[['global sales']])
```

In [51]:

```
#vg_df.head()
```

In [52]:

```
# Visualzing total volume of sales year-wise
#check = vg_df.groupby('year')['global sales'].sum()
#plt.figure(figsize = (30, 4))
#ax1 = plt.subplot(121)
#check.plot(y = "global sales", legend = False, fontsize = 12, sharex = False, title
#ax1.ticklabel_format(useOffset = False, style = 'plain')
#ax1.set_ylabel("Total Global Sales")
#ax1.yaxis.grid(True)
```

In [53]:

```
sns.countplot(x='global_sales', hue='genre', data=df)
```

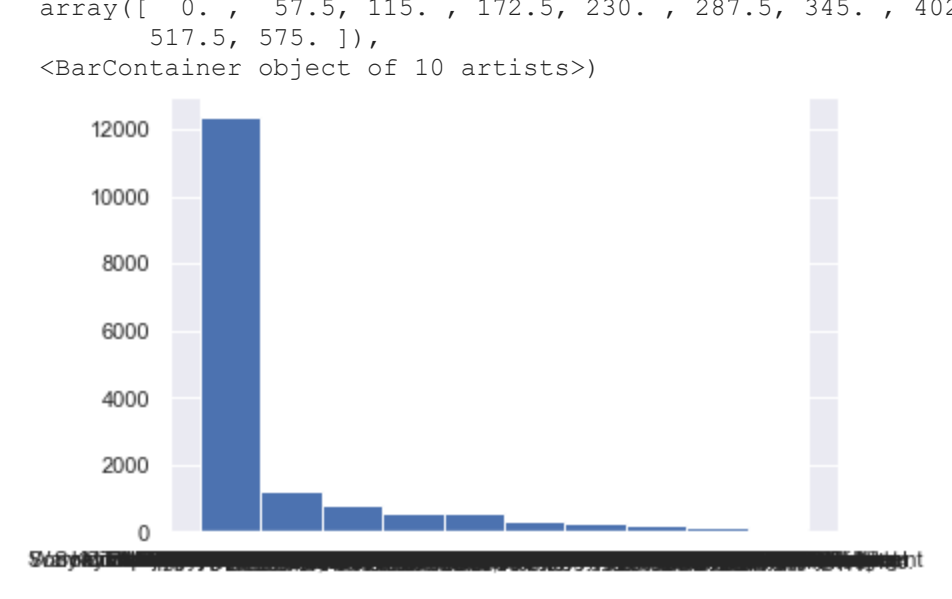
Out[53]: `<AxesSubplot:xlabel='global_sales', ylabel='count'>`

In [187…

```
#Graphical understanding of publishers of video games
plt.hist(df['publisher'])
```

Out[187…

```
(array([12345.,  1182.,   786.,   564.,   533.,   295.,   227.,   165.,
          117.,    77.]),
 array([  0. ,  57.5, 115. , 172.5, 230. , 287.5, 345. , 402.5, 460. ,
        517.5, 575. ]),
 <BarContainer object of 10 artists>)
```



In [55]:

```
df4 = df.filter(['name'],axis=1)
df4.shape
```

Out[55]: `(16291, 1)`

In [188…

```
#Game with the highest sales
df4['name']. value_counts(). idxmax()
```

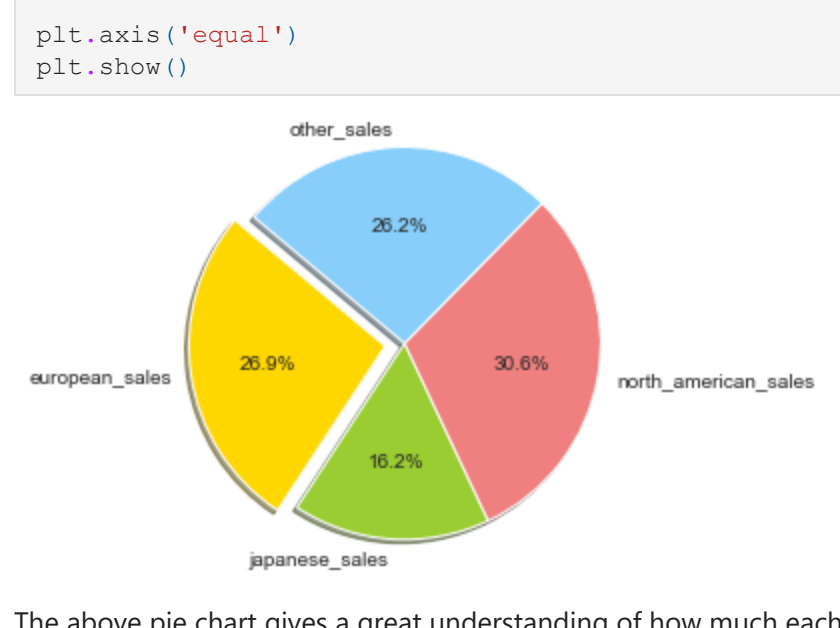
Out[188… `'Need for Speed: Most Wanted'`

In [57]:

```
import pandas as pd
pd.set_option('display.max_columns',100)
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set_style('darkgrid')
```

In [60]:

```
import matplotlib.pyplot as plt

# Data to plot
labels = 'european_sales', 'japanese_sales', 'north_american_sales', 'other_sales'
sizes = [215, 130, 245, 210]
colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue']
explode = (0.1, 0, 0, 0)  # explode 1st slice

# Plot
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=140)

plt.axis('equal')
plt.show()
```



The above pie chart gives a great understanding of how much each region/continent makes up of the total sales of video games

Q2. Statistics: Correlation Analysis

(also useful for Q2 Data Preparation)

In [190…

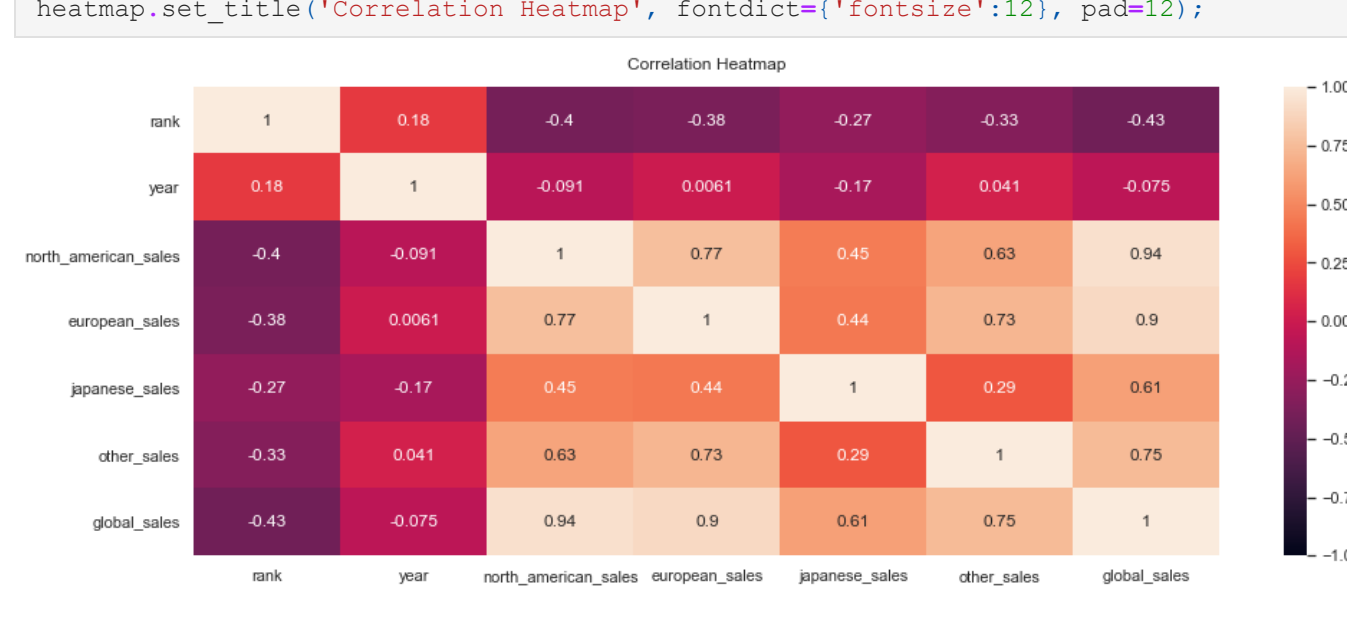
```
#Correlation Matrix: This code reveals if there is any correlation between numerical
stats_df2 = df[['north_american_sales', 'european_sales', 'japanese_sales','other_sal
stats_df2.corr()
```

Out[190…

	north_american_sales	european_sales	japanese_sales	other_sales	global_sales
north_american_sales	1.000000	0.768923	0.451283	0.634518	0.941269
european_sales	0.768923	1.000000	0.436379	0.726256	0.903264
japanese_sales	0.451283	0.436379	1.000000	0.290559	0.612774
other_sales	0.634518	0.726256	0.290559	1.000000	0.747964
global_sales	0.941269	0.903264	0.612774	0.747964	1.000000

In [192…

```
#Heat map showing correlation
plt.figure(figsize=(16, 6))

# Set the range of values to be displayed on the colormap from -1 to 1
heatmap = sns.heatmap(df.corr(), vmin=-1, vmax=1, annot=True)

# Give a title to the heatmap. Pad defines the distance of the title from the top of
heatmap.set_title('Correlation Heatmap', fontdict={'fontsize':12}, pad=12);
```



Correlation between European Sales & Global Sales

Correlation measures the strength of linear dependence between X and Y and lies between -1 and 1 (Gollapudi & Laxmikanth, 2016). Carrying out a correlation analysis between two variables will show if there a relationship between them and how dependent they are of each other.

In [193…

```
# calculate the spearmans's correlation between two variables
from numpy.random import randn
from numpy.random import seed
from scipy.stats import spearmanr
# seed random number generator
seed(1)
# prepare data
data1 = 20 * randn(1000) + 100
data2 = data1 + (10 * randn(1000) + 50)
# calculate spearman's correlation
corr, _ = spearmanr(data1, data2)
print('Spearmans correlation: %.3f' % corr)
```

```
Spearmans correlation: 0.872
```

In [63]:

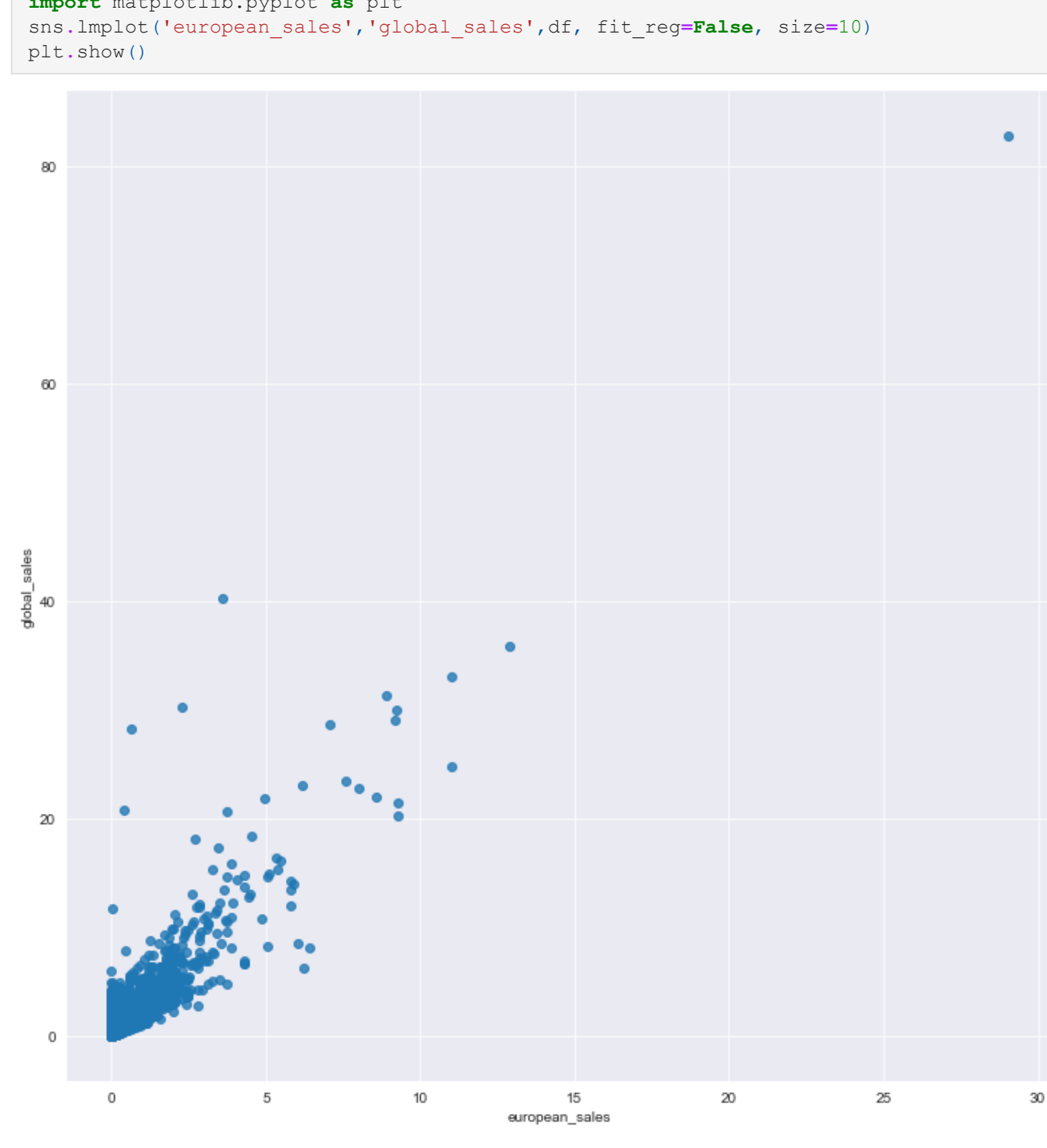
```
from scipy.stats import pearsonr
# Convert dataframe into series
list1 = df['european_sales']
list2 = df['global_sales']

# Apply the pearsonr()
corr, _ = pearsonr(list1, list2)
print('Pearsons correlation: %.3f' % corr)
```

```
Pearsons correlation: 0.903
```

The correlation coefficient or Pearson's r, is a measure of both the strength and direction of the linear relationship between two variables. The statistic will be a number between -1 and 1 where -1 is the total negative correlation and +1 is the total positive correlation (Lesmeister, 2015). Therefore the result above of 0.851 shows there is a strong positive relationship between EU Sales and Global Sales.

In [64]:

```
#Scatter plot is utilised to get a better feel of relationship
import seaborn as sns
import matplotlib.pyplot as plt
sns.lmplot('european_sales','global_sales',df, fit_reg=False, size=10)
plt.show()
```



In [65]:

```
import pingouin as pg
pg.corr(x=df['global_sales'], y=df['european_sales'])
```

Out[65]:

	n	r	CI95%	r2	adj_r2	p-val	BF10	power
pearson	16291	0.903264	[0.9, 0.91]	0.815885	0.815863	0.0	inf	1.0

The above output is broken down as follows (Vallat, 2019);

n

- n is the sample size, i.e. how many observations were included in the calculation of the correlation coefficient

r

- r is the correlation coefficient, 0.85 in this case, this is high. CI95% are the 95% confidence intervals around the correlation coefficient

r2 & adj_r2

- r2 and adj_r2 are the r-squared and ajusted r-squared respectively. It is r squared which is a measure of the proportion of the variance in the first variable that is predictable from the second variable.

p-val

- p-val is the p-value of the test. The general rule is that you can reject the hypothesis that the two variables are not correlated if the p-value is below 0.05, which is the case in the above, the value is 0. Therefore there is a significant correlation between the two variables.

BF10

- BF10 is the Bayes Factor of the test, which also measure the statistical significance of the test. While they are conceptually different, the Bayes Factor and p-values will in practice often reach the same conclusion.

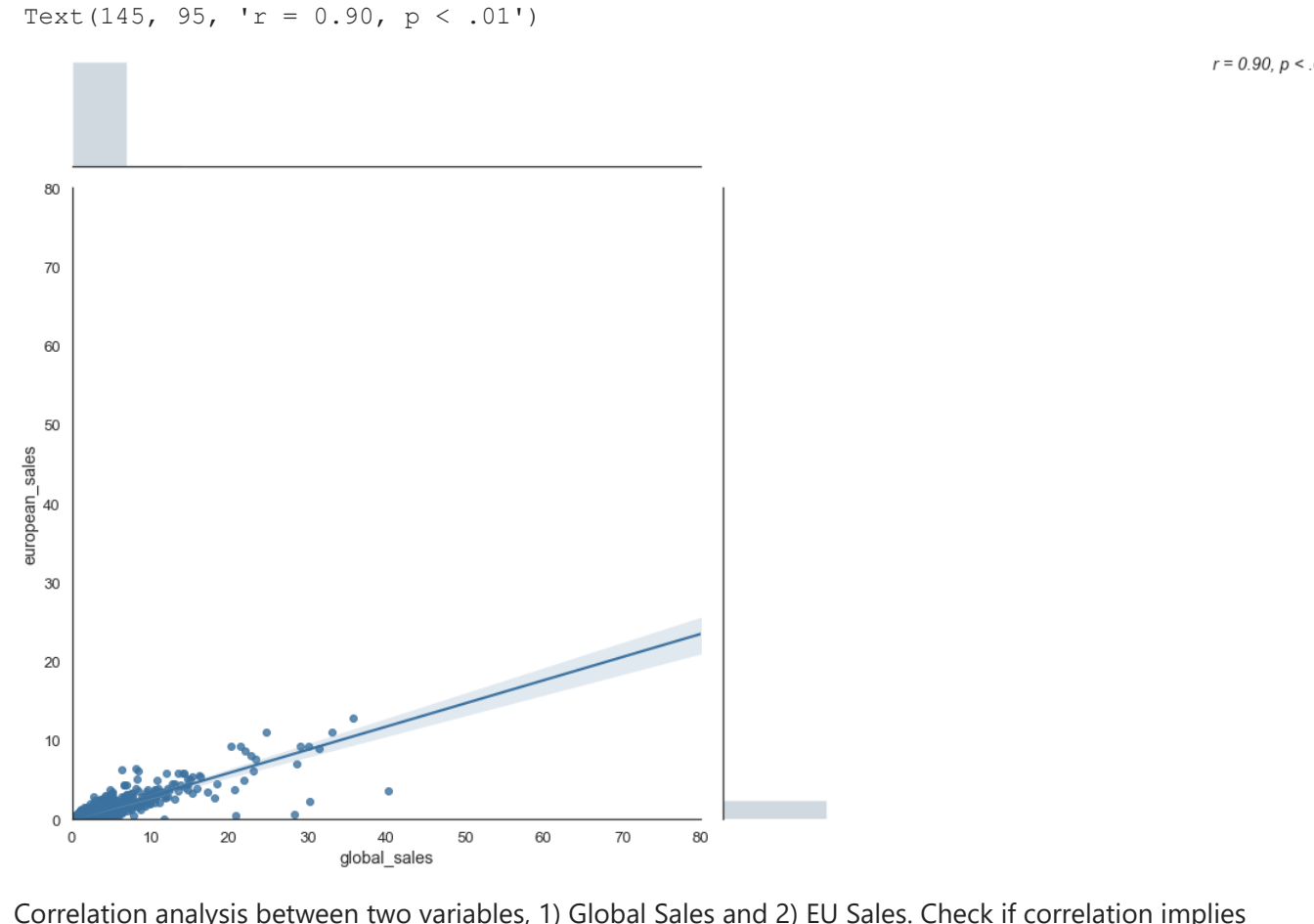
power

- power is the achieved power of the test, which is the likelihood that we will detect an effect when there is indeed an effect there to be detected. The higher this value is, the more robust our test is. In that case, a value of 1 means that we can be greatly confident in our ability to detect the significant effect.

In [66]:

```
import warnings
warnings.filterwarnings("ignore")

sns.set(style='white', font_scale=1.2)

g = sns.JointGrid(data=df, x='global_sales', y='european_sales', xlim=(0, 80), ylim=(
g = g.plot_joint(sns.regplot, color="xkcd:muted blue")
g = g.plot_marginals(sns.distplot, kde=False, bins=12, color="xkcd:bluey grey")
g.ax_joint.text(145, 95, 'r = 0.90, p < .01', fontstyle='italic')
```

Out[66]: `Text(145, 95, 'r = 0.90, p < .01')`



Correlation analysis between two variables, 1) Global Sales and 2) EU Sales. Check if correlation implies causation. Short Explanation and conclusion based on findings

Q3. Statistics: Construct Predictive Model for Two Variables

Regression Analysis

Regression models are used to predict a continuous value and it is a supervised technique (Dave, 2018). Regression models define the connection between variables by fitting a line to the observed data. Regression enables the user to approximate how a dependent variable changes as the independent variable(s) change (Bevans, 2020).

Simple linear regression is used to estimate the relationship between two quantitative variables. You can use simple linear regression when you want to know: The goal of linear regression is to uncover how strong the relationship is between two variables and also highlight the value of the dependent variable at a certain value of the independent variable.

In [67]:

```
#Create X array
x = df[['global_sales']]
```

In [68]:

```
y = df['other_sales']
```

In [69]:

```
from sklearn.model_selection import train_test_split
```

In [70]:

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.3)
```

In [71]:

```
from sklearn.linear_model import LinearRegression
```

In [72]:

```
model = LinearRegression()
```

In [73]:

```
model.fit(x_train, y_train)
```

Out[73]: `LinearRegression()`

In [74]:

```
print(model.coef_)
```

```
[0.09349656]
```

In [75]:

```
print(model.intercept_)
```

```
-0.0016593725699553946
```

In [76]:

```
pd.DataFrame(model.coef_, x.columns, columns = ['Coeff'])
```

Out[76]:

	Coeff
global_sales	0.093497

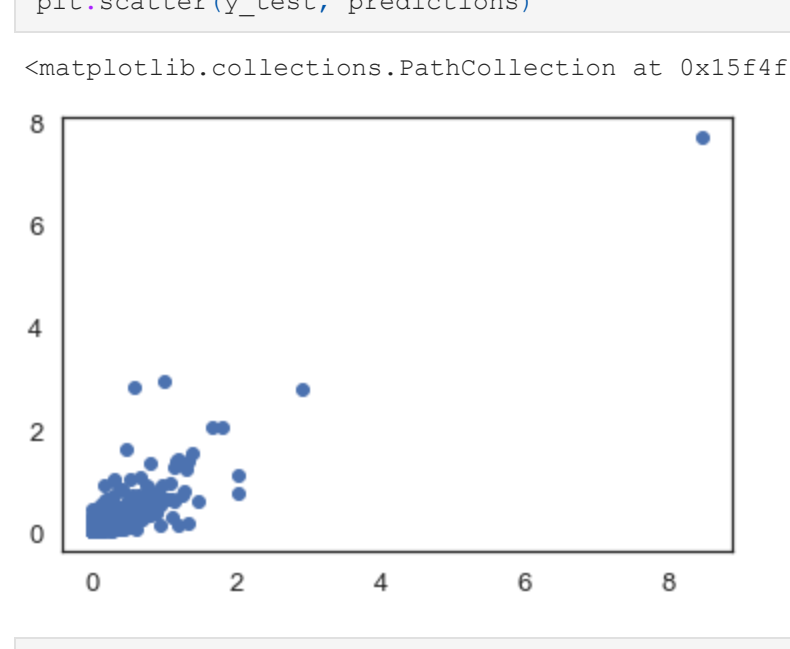
In [77]:

```
predictions = model.predict(x_test)
```

In [78]:

```
plt.scatter(y_test, predictions)
```

Out[78]: `<matplotlib.collections.PathCollection at 0x15f4ffc5820>`

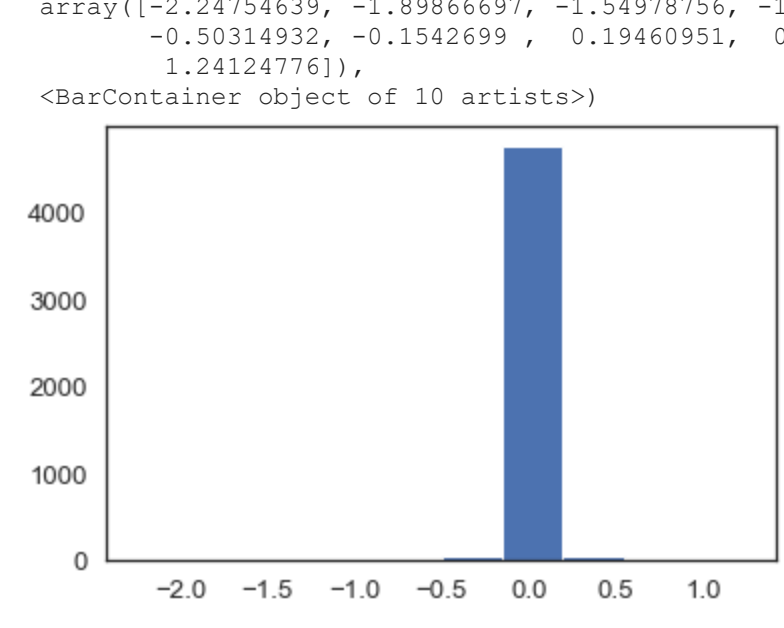


In [79]:

```
plt.hist(y_test - predictions)
```

Out[79]:

```
(array([2.00e+00, 0.00e+00, 0.00e+00, 1.00e+00, 4.00e+00, 5.70e+01,
        4.76e+03, 5.60e+01, 4.00e+00, 4.00e+00]),
 array([-2.24754639, -1.89866697, -1.54978756, -1.20090814, -0.85202873,
        -0.50314932, -0.1542699 ,  0.19460951,  0.54348893,  0.89236834,
         1.24124776]),
 <BarContainer object of 10 artists>)
```



In [80]:

```
from sklearn import metrics
```

In [81]:

```
metrics.mean_absolute_error(y_test, predictions)
```

Out[81]: `0.025412822662528492`

In [82]:

```
metrics.mean_squared_error(y_test, predictions)
```

Out[82]: `0.00713100462375786`

In [83]:

```
np.sqrt(metrics.mean_squared_error(y_test, predictions))
```

Out[83]: `0.0844452759114319`

In [84]:

```
#Create X array
X = df[['global_sales']]

y = df['other_sales']

# Call the train_test_split method to split the data and the default splitting is 75%
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state = 42)

# Train the LinearRegression mode by using a method fit() function/ method
lr = LinearRegression().fit(X_train, y_train)
```

In [85]:

```
# Display the coefficient and intercept values
print("lr.coef_: {}".format(lr.coef_))
print("lr.intercept_: {}".format(lr.intercept_))
```

```
lr.coef_: [0.08621736]
lr.intercept_: 0.0015315143175630153
```

In [86]:

```
# Display the score values by calling lr
print("Training set score: {:.2f}".format(lr.score(X_train, y_train)))
print("Test set score: {:.2f}".format(lr.score(X_test, y_test)))
```

```
Training set score: 0.49
Test set score: 0.69
```

As R^2 of 0.69 is okay but not great to be honest. It shows the model is not performing very well, it could be further impproved. As the training and testing scores are not too close,with the testing result being better, there could be slight underfitting present. ie the model is too simple and can generalize on any new test data, not giving good insights.

Q2. Data Preparation: Further look at Data Preparation Methods

In [87]:

```
import nltk
from nltk.corpus import stopwords
import re
import networkx
from collections import Counter
import string
import nltk
import numpy as np # linear algebra
import pandas as pd

import warnings
warnings.filterwarnings("ignore")

sns.set(font_scale=1.5)
sns.set_style("whitegrid")
```

In [88]:

```
video_game_name = df[df['name'] == 1]['name'].str.lower()
```

In [89]:

```
video_game_name = video_game_name.values.tolist()
```

In [90]:

```
video_game_name_break = [nltk.tokenize.wordpunct_tokenize(text) for text in video_game
```

In [91]:

```
stopwords = nltk.corpus.stopwords.words('english')
```

In [92]:

```
from itertools import groupby

def clear_multiple_char(comment):
    ti = []
    for words in comment:
        t = [''.join(["".join(i) for i, _ in groupby(word)]) if len(word)>10 else word
        ti.append(t)
    return ti
```

In [93]:

```
video_game_name_break = clear_multiple_char(video_game_name_break)
```

In [94]:

```
import string
from subprocess import check_output
from nltk.stem.lancaster import LancasterStemmer
stemmer = LancasterStemmer()
punctuation = string.punctuation # '!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
# Add numbers
punctuation += '0123456789'

def comment_raiz(comment):
    text = []
    for lista in comment:
        valids = [stemmer.stem(word) for word in lista if word not in stopwords and wo
                  and len(word)>2]
        valids_true = [''.join([char for char in word if char not in punctuation]) for
                       len(''.join([char for char in word if char not in punctuation])
        text.append(valids_true)
    return text
```

In [95]:

```
video_game_name_clear = comment_raiz(video_game_name_break)
```

In [96]:

```
from collections import Counter
def counter(comment_clear):
    cnt = Counter()
    for words in comment_clear:
        for word in words:
            cnt[word] += 1
    return cnt
```

In [97]:

```
video_game_name_cnt = counter(video_game_name_clear)
```

In [98]:

```
video_game_name_cnt.most_common(10)
```

Out[98]: `[]`

In [99]:

```
#import matplotlib.pyplot as plt
#for column in df.select_dtypes(include='object'):sns.countplot(y=column, df=df)
#plt.show()
```

In [100…

```
text_data = df['name']
```

In [101…

```
set(text_data)
```

Out[101…

```
{"Disney's Chicken Little: Ace In Action",
 'Crash Bandicoot',
 'Xenosaga Episode III: Also sprach Zarathustra',
 'Dreamer Series: Top Model',
 'Boktai: The Sun is in Your Hand',
 'Pro Yakyuu Famista DS 2010',
 'Animal Crossing: City Folk',
 'Brothers In Arms: Road to Hill 30',
 'Shukufuku no Campanella Portable',
 "nail'd",
 "Germany's Next Topmodel 2011",
 'Uchida Yasuou Mystery: Meitantei Senken Mitsuhiko Series: Fukutoshin Renzoku Satsuji
n Jiken',
 'Nova Usagi no Game de Ryuugaku!? DS',
 'Final Fantasy Tactics',
 'Rumble Racing',
 'Donkey Kong Classics',
 'Skylanders SWAP Force',
 'Worms: Battle Islands',
 'Speed Kings',
 'Kuusen II',
 'Yokai Watch 2 Shinuchi',
 'Pokémon I Choose You / Squirtle Squad Game Boy Advance Video',
 'Hi Hi Puffy AmiYumi: The Genie & the Amp',
 'Sonic Riders',
 'Dreamworks Madagascar Kartz',
 'Skeleton Warriors',
 'SEGA Bass Fishing Duel',
 'The Chronicles of Narnia: Prince Caspian',
 'Sakura Wars 1 & 2',
 'Star Wars: Empire at War',
 'Bangai-O Spirits',
 'Akatsuki no Amaneka to Aoi Kyojin',
 'Need for Speed: V-Rally',
 'Silpheed: The Lost Planet',
 'PGR3 - Project Gotham Racing 3',
 'Virtua Fighter CG Portrait Series Vol.6: Lau Chan',
 'NASCAR Racing',
 'Global Defence Force',
 'MVP Baseball',
```

```
'Dynasty Warriors 2',
'BioShock 2',
'BattleTanx',
'Fate/Tiger Colosseum Upper',
'Moujuutsukai to Oujisama',
'Just Sing!',
'Chicken Run',
'Fishing Master (jp sales)',
'Fab 5 Soccer',
'The Conveni 4',
'Katekyoo Hitman Reborn! DS Fate of Heat III - Yuki no Shugosha Raishuu!',
'Franklin: A Birthday Surprise',
'Ready 2 Rumble Revolution',
'Triple Play 2002',
'Spyro Orange: The Cortex Conspiracy',
'Konductra',
'Worms: WMD',
'Namco Classic Fighter Collection',
'Brain Boost: Beta Wave',
'Hiiro no Kakera 3: Aoikuro no Kusabi',
'The Sims 3: Ambitions',
'Lady Sia',
'ISS Pro Evolution',
'Really? Really! DS',
'Disney Sports Soccer',
'Yu-Gi-Oh! Duel Monsters II: Dark Duel Stories',
'Risen 3: Titan Lords',
'Sorcery Saga: The Curse of the Great Curry God',
'Monster Trucks',
'Rocksmith',
'Galaxy Fight',
'Gekijouban Macross F: Itsuwarino Utahime - Hybrid Pack',
'Sega Superstars',
'CSI: 3 Dimensions of Murder',
'New Legends',
'Tennis no Oji-Sama Gyutto! DokiDoki Survival - Umi to Yama no Love Passion',
'Atelier Ayesha: The Alchemist of Dusk',
'Crusader: No Remorse',
"Jikkyou Powerful Pro Yakyuu '98 Kaimakuban",
'Patrician IV',
'Mystery Legends: Beauty and the Beast',
'Summer Athletics: The Ultimate Challenge',
'The Ant Bully',
'Pro Yakyuu Team o Tsukurou!',
'Star Wars: Starfighter',
'Kidz Bop Dance Party! The Video Game',
'Teenage Mutant Ninja Turtles (3DS)',
'Saltwater Sportfishing',
'Imagine: Wedding Designer',
'NASCAR Rumble',
'Jampack Volume 12',
'Petz: Dogz Family',
'Pro Yaky? Spirits 6',
'Amazing Island',
'Kangaroo',
'Kinetica',
'Age of Mythology: The Titans',
'FIFA Soccer 2005',
'Mario Bros.',
'Assassination Classroom: Grand Siege on Kuro-sensei',
'NFL GameDay 2001',
'Every Extend Extra',
'Amped 3',
'Invizimals: Shadow Zone',
'Aquapazza: Aquaplus Dream Match',
'Mobil 1 Rally Championship',
'Rudra no Hihou',
'Buster Bros. Collection',
'Pool Hall Pro',
'Onechanbara Z2: Chaos',
'Duke Nukem: Zero Hour',
'Astro Boy: The Video Game',
'Corvette Evolution GT',
'Guitar Freaks V2 & DrumMania V2',
'Invizimals: The Lost Kingdom',
'Nuclear Strike 64',
'NFL Xtreme 2',
'Code Geass: Hangyaku no Lelouch - Lost Colors',
'ICO',
'Pac-Man Party 3D',
'Castlevania: Legacy of Darkness',
'Brothers in Arms: Furious 4',
'Ratchet & Clank: Quest for Booty',
'The Urbz: Sims in the City (all regions sales)',
'The Sims 3: Generations',
'NCAA Final Four 2004',
'Littlest Pet Shop: Beach Friends',
"Banushi Life Game: Winner's Circle",
'International Superstar Soccer 64',
'Rebel Raiders: Operation Nighthawk',
'Planet 51',
'Airline Tycoon 2',
"Pro Yakyuu Family Stadium '87",
'Project Sylpheed: Arc of Deception',
'Steins;Gate',
'Mister Slime',
'Arc the Lad II',
'MLB 2002',
'Harvey Birdman: Attorney at Law',
'Stadium Games',
'Soldier of Fortune II: Double Helix',
'Major League Baseball 2K11',
'Rock Band 4',
'Katekyoo Hitman Reborn! Battle Arena',
'WorldShift',
'Baja: Edge of Control',
'Moto Racer',
'Sengoku Musou 3 Z Special',
'Away: Shuffle Dungeon',
'Naruto: Ninja Destiny (US sales)',
'Harvest Moon: A Wonderful Life',
'Rock Band 3',
'Galactic Attack',
'Saint Seiya: Sanctuary Battle',
'MonHun Nikki: Poka Poka Ailu Mura',
'Final Fantasy Explorers',
'The King of Fighters Collection: The Orochi Saga (JP sales)',
'Gothic Universe',
"Jumper: Griffin's Story",
'SD Gundam G Generation: Cross Drive',
'Dig Dug: Digging Strike',
'Shadow of the Colossus',
'The House of the Dead: Overkill',
'Katekyoo Hitman Reborn! Kindan no Yami no Delta',
'Magical Vacation',
'MLB 15: The Show',
'Knights in the Nightmare',
'Ghostbusters (2016)',
'Dungeon Siege: Throne of Agony',
'Super Robot Spirits',
'Monster Trux Arenas: Special Edition',
'Kong: King of Atlantis',
'Gex 3: Deep Cover Gecko',
'Shin Sangoku Musou Online: Kamishou Ranbu',
'Valkyrie Profile: Lenneth',
'Choujikuu Yousai Macross: Ai Oboete Imasu ka',
"I'm A Celebrity: Get Me Out of Here!",
'Super Scope 6',
'Chocolatier',
'Story of Seasons: Good Friends of the Three Villages',
'Galaxy Angel II: Zettairyouiki no Tobira',
'Cloudy With a Chance of Meatballs',
'God Eater 2: Rage Burst',
'R: Racing Evolution',
'Naruto Shippuden: Legends: Akatsuki Rising',
'Fate/Extra CCC',
'Kirby Super Star',
'Beyond Good & Evil',
'Worms Reloaded',
'Record of Lodoss War: Eiyuu Sensou',
"Disney's Chicken Little",
'The Dark Spire',
'Taiko no Tatsujin Wii: Minna de Party * 3-Yome!',
'Mortal Kombat: Tournament Edition',
'Ratatouille: Food Frenzy',
'SoulCalibur II',
'Dragon Ball Z',
'999: Nine Hours, Nine Persons, Nine Doors',
'Road Trip',
'Sonic Lost World',
'Def Jam Rapstar',
'Body and Brain Connection',
'Go! Go! Hypergrind',
'Hot Shots Golf 3',
'Katekyoo Hitman Reborn! DS: Flame Rumble Hyper - Moeyo Mirai',
'Supermodel Makeover by Lauren Luke',
'Legend of Kay',
'Def Jam Vendetta',
'Higurashi no Naku Koro ni Matsuri: Kakera Asobi',
'Dokuro',
'Hello Kitty Party',
'Horse Life 4: My Horse, My Friend, My Champion',
'Treasure Chase',
'Aoishiro',
'Forgotten Realms: Demon Stone',
'Clock Tower II: The Struggle Within',
'Castlevania: Portrait of Ruin',
'Super Baseball Simulator 1.000',
'D.C. III: Da Capo III',
'Blue Roses: Yousei to Aoi Hitomi no Senshitachi',
'Space Invaders Extreme 2',
'Jewels of the Tropical Lost Island',
'Pet Pals: New Leash on Life',
'Advance Wars 2: Black Hole Rising',
'Bloody Roar',
'Fuzion Frenzy 2',
'Kidou Senshi Gundam UC',
'Star Driver: Kagayaki no Takuto - Ginga Bishounen Densetsu',
'Rugby 08',
'Zumba Fitness: World Party',
'Battle Stadium D.O.N',
"Legend of the Guardians: The Owls of Ga'Hoole",
'Shinseiki Evangelion: Koutetsu no Girlfriend',
'Wacky Races: Crash & Dash',
'The Godfather (old US sales)',
'Mountain King',
'SD Gundam G Generation Wars',
'Vacation Sports',
'Rayman Raving Rabbids 2',
'Mega Man X8',
'Musashi: Samurai Legend',
'Ultimate Board Game Collection',
'Shin Sangoku Musou Online: Souten Ranbu',
'Kingdom Hearts: Birth by Sleep',
'Ochaken no Heya DS 2',
'Farming Simulator 2014',
"Disney's 102 Dalmatians: Puppies to the Rescue",
'Suzumiya Haruhi no Yakusoku',
'Burstrick: Wake Boarding!!',
'Heavy Fire: Afghanistan',
'Bakumatsu Rock: Ultra Soul',
'Family Fest presents: Movie Games',
"Smile PreCure! Let's Go! Marchen World",
'Pitfall II: Lost Caverns',
'WinBack: Covert Operations',
'Saki Portable',
'Diabolik Lovers: Vandead Carnival',
'Super Bomberman 4',
'Forever Kingdom',
'Jak X: Combat Racing',
'How to Train Your Dragon 2',
'Constantine',
'The X Files: Resist or Serve',
"Nobunaga's Ambition: Sphere of Influence - Sengoku Risshiden",
'Hard Rock Casino',
'Party Pigs: Farmyard Games',
'Advance Wars: Dual Strike',
'TransFormers Animated: The Game',
'The Walking Dead: Season One',
'Panzer Dragoon Saga',
'The Gunstringer',
'Frogs And Flies',
'DropCast',
'SingStar Pop Hits',
'Stolen Song',
'Mario Golf',
'SimCity 2000',
'NHL 2002',
'Vacation Isle: Beach Party',
'Risk: Global Domination',
'London 2012: The Official Video Game of the Olympic Games',
'Breeding Stud 2',
'Metal Gear Ac!d 2',
'Dorabase DS: Dramatic Stadium',
'Tetris Party Deluxe',
'Go Play: Circus Star',
'Dead Ball Zone',
'Genkai Tokki: Moero Crystal',
'Contender 2',
'Canada Hunt',
"Cabela's Big Game Hunter",
'Shawn Johnson Gymnastics',
'Vampire Legends: Power of Three',
'NHL 2K11',
"Tony Hawk's Underground 2 Remix",
'Afro Samurai',
'All Star Tennis 99',
'Mega Man Battle Network',
'White Album 2: Shiawase no Mukougawa',
'Fantavision',
'Kart Racer',
'Beyblade G-Revolution',
'Dragon Quest IX: Sentinels of the Starry Skies',
'Yu-Gi-Oh! The Eternal Duelist Soul',
'Super Bust-A-Move',
'Doctor Who: Return to Earth',
'Lilie no Atelier: Salburg no Renkinjutsushi 3',
'Naruto: Ninja Council',
'Project: Horned Owl',
'The Elder Scrolls Online',
'12Riven: The Psi-Climinal of Integral',
"Dragon's Lair 3D: Return to the Lair",
'Darius Burst: Chronicle Saviours',
'Atelier Shallie: Alchemists of the Dusk Sea',
'Yakuza Zero: The Place of Oath',
'National Geographic Panda (JP sales)',
'Ape Escape Academy',
'The Idolmaster: Gravure For You! Vol.8',
'Shinkyoku Soukai Polyphonica',
'Bust-A-Move DS',
'Animal Paradise',
'Homie Rollerz',
'Code R',
'Resistance: Fall of Man',
'Battle Arena Toshinden 2',
'Ferrari Challenge Trofeo Pirelli',
'Short Track Racing: Trading Paint',
'Twister Mania',
'Cross Treasures',
'Kao the Kangaroo Round 2',
'Madden NFL 2000',
'Neighborhood Games',
'Guilty Gear Xrd: Sign',
'Mind Your Language: Learn English! / French! / German! / Spanish! /Japanese!',
'Marvel Trading Card Game',
'Prinny 2: Dawn of Operation Panties, Dood!',
'Winning Post 8',
'Pachitte Chonmage Tatsujin 15: Pachinko Fuyu no Sonata 2',
'Harry Potter and the Goblet of Fire',
'Sharin no Kuni, Himawari no Shoujo',
'Culdcept',
'Bomberman II',
'Mushishi: Amefuru Sato',
'Gretzky NHL 06',
'SD Gundam: Gashapon Wars',
'Mystery Case Files: The Malgrave Incident',
'Shin Super Robot Taisen',
'Doom',
'Motor Toon Grand Prix (Japan)',
'Tenchu Z',
'EVE Online',
'Family Fun Football',
'Diabolik Lovers: More Blood',
'Keystone Kapers',
'Muv-Luv Alternative: Total Eclipse',
'Jeep Thrills',
'Jumble Madness',
'Tales of Graces',
'Test Drive Off-Road Wide Open',
'Metroid Fusion',
"Foster's Home for Imaginary Friends: Imagination Invaders",
'Gegege no Kitarou 2: Youkai Gundan no Chousen',
'Knights Contract',
'My First Dollhouse',
'River King: Mystic Valley (US sales)',
'Forza Motorsport',
'Major League Baseball 2K8 Fantasy All-Stars',
'Cities in Motion Collection',
'Ace Combat 3: Electrosphere',
'Story Hour: Adventures',
'Civilization II',
'SBK X: Superbike World Championship',
'Homefront: The Revolution',
'Darkstone',
'Outlaw Volleyball Remixed',
'The Cheetah Girls',
'Donkey Konga',
'Scaler',
'Gradius III and IV',
"Luxor: Pharaoh's Challenge",
'Sly Cooper: Thieves in Time',
'Tonic Trouble',
'Daito Giken Koushiki Pachi-Slot Simulator: Shake II',
'Mystereet',
'Tales of Hearts',
'Harvest Moon: Hero of Leaf Valley',
'Frontier Gate',
'Knights of Honor',
'Achtung Panzer: Kharkov 1943',
'Martian Panic',
'Azure Striker Gunvolt: Striker Pack',
'Tobal No.1',
'Yu-Gi-Oh! The Sacred Cards (American and Others sales)',
'Carmen Sandiego: The Secret of the Stolen Drums',
'Get Up and Dance',
'Mega Man Battle Network 3 Blue / White Version',
'No More Heroes 2: Desperate Struggle',
'Dragon Ball Z: Idainaru Dragon Ball Densetsu',
'WWE SmackDown vs. Raw 2009',
'Darkened Skye',
'Bust-A-Move 2 Arcade Edition',
'AirForce Delta Strike',
'Cake Mania',
'My Boyfriend',
'UFC Personal Trainer: The Ultimate Fitness System',
'Ookiku Furikabutte: Honto no Ace ni Nareru kamo',
'WWF No Mercy',
'Jissen Pachi-Slot Hisshouhou: Hokuto no Ken',
'Aikatsu! My No.1 Stage!',
'Mobile Suit Gundam Seed',
'Final Fantasy XI: Treasures of Aht Urhgan',
'Spy Hunter / Super Sprint',
'Beatdown: Fists of Vengeance',
'Trinity Universe',
'Snowboard Kids 2',
'F1 Racing Championship',
'Breakdown',
"World Class Rugby 2: Kokunai Gekitou Hen '93",
'Lovely x Cation 1 & 2',
'From TV Animation One Piece: Yume no Lufy Kaizokudan Tanjou!',
'Romancing SaGa 2',
'Jewel Master: Cradle of Athena',
'Gex: Enter the Gecko',
"Cabela's Alaskan Adventure",
'2 Games in 1: Sonic Battle & ChuChu Rocket!',
'Yu-Gi-Oh! World Championship 2008',
'Jake Power: Fireman',
'Namco Museum DS',
'Medabots 9: Metabee / Rokusho',
'Killer Loop',
'Luminous Arc',
'Derby Stallion 98',
'7 Wonders of the Ancient World',
'Dragon Ball: Origins',
'American Chopper',
'NFL Quarterback Club 98',
'Command & Conquer: Red Alert',
'Wreckin Crew',
'D.C. II P.S.: Da Capo II Plus Situation',
'G.I. Joe: The Rise of Cobra',
'Chaos Legion',
'Jak 3',
'Blast Corps',
'K-1 Grand Prix',
'Mahjongg Mysteries: Ancient Egypt',
'World Soccer Winning Eleven 7 International',
'Brothers In Arms: D-Day',
"Yoshi's Island DS",
'Charm Girls Club: My Perfect Prom',
'Astro Boy: Omega Factor',
'The Sims 3: Town Life Stuff',
'E.T.: The Extra Terrestrial',
'TOCA 2: Touring Cars',
'Total War: Shogun 2',
"Gabrielle's Ghostly Groove 3D",
'RockMan & Forte',
'Real Robots Final Attack',
'Disgaea 2: Dark Hero Days',
'NASCAR',
'Otogi: Myth of Demons',
'Sports Champions',
'Tear Ring Saga Yutona Eiyuu Senki',
'Portal Runner',
'Full Spectrum Warrior: Ten Hammers',
'BeatMania IIDX 3rd Style',
'Crazy Taxi: Fare Wars',
'Soul Reaver 2',
"Broken Sword: Shadow of the Templars - The Director's Cut",
'Top Gun: Fire at Will!',
'Pro Evolution Soccer 2008',
'Famista Returns',
'Minecraft: Story Mode',
'300: March to Glory',
'Wipeout 2',
'Super Dragon Ball Z',
'Puzzle Quest: Challenge of the Warlords',
'Gotouchi Tetsudou: Gotouchi Chara to Nihon Zenkoku no Tabi',
'Bubble Bobble Double Shot',
'Captain Tsubasa J: The Way to World Youth',
"Master Jin Jin's IQ Challenge",
'My Fun Facts Coach',
'Psycho-Pass',
"Backyard NFL Football '09",
'Star Trek: The Game',
'Naruto Shippuden: Dairansen! Kage Bunshin Emaki',
'PopCap Arcade Vol 2',
'Enkaku Sousa: Sana e no 23 Hiai',
'Shining Soul',
'Street Sk8er',
'Grand Theft Auto III',
'Hakuoki Zuisouroku Omokage Hana',
"Disney's Kim Possible: Global Gemini",
'Rugby 2005',
'Imagine Figure Skater (JP sales)',
'Logic Machines',
'Pro Yakyuu Greatest Nine 98',
'Mary-Kate and Ashley: Winners Circle',
"Mickey's Speedway USA",
'Babysitting Mania',
'Doshin the Giant',
'Commando',
'The Con',
'Jade Cocoon: Story of the Tamamayu',
'Skate it',
'Crash Bandicoot 2: Cortex Strikes Back',
'Apocalypse',
'Mega Man ZX',
'Exit DS',
'Rocket Power: Zero Gravity Zone',
'Evil Zone',
'FlatOut',
'World Snooker Challenge 2005',
'Mega Man Anniversary Collection',
'Downforce',
'One Piece: Daikaizoku Coliseum',
'Ultra Bust-A-Move',
'Tennis no Oji-Sama: Genius Boys Academy',
'Sengoku Basara 3 Utage',
'Rally Fusion: Race of Champions',
'Stranger of Sword City',
'Hisshou 777 Fighter: Pachi Slot Eiyu Densetsu',
'NBA Street',
'Super Puyo Puyo',
"Lemony Snicket's A Series of Unfortunate Events",
'Sing4: The Hits Edition',
'Gravity Games Bike: Street Vert Dirt',
'Uta Kumi 575',
'Chibi-Robo! Plug into Adventure!',
'Pokémon Emerald Version',
'Rival Schools: United By Fate',
'.hack//G.U. Vol.2//Reminisce',
'World of Warcraft: Cataclysm',
'Jump Ultimate Stars',
'Mount & Blade',
'Mass Effect',
'The Guy Game',
'Paper Mario: The Thousand-Year Door',
'J-League Soccer: Prime Goal',
'Darkest of Days',
'Castlevania: Lords of Shadow 2',
'GT Pro Series',
'Football Manager Handheld 2008',
'Bullet Soul: Tama Tamashii',
'Virtua Tennis: World Tour (US & Others sales)',
'Rule of Rose',
'World Championship Poker 2: Featuring Howard Lederer',
'Fighting Fantasy: The Warlock of Firetop Mountain',
'Rocket Power: Beach Bandits',
'Games Around the World',
'Farming 2017 - The Simulation',
'My Baby: First Steps',
'And1 Streetball',
'Dig Dug',
'Namco Tennis Smash Court',
'Rock Band Unplugged',
'Fashionable Puppy: Oshare na Koinu DS',
'Block Kuzushi',
'Shin Hayarigami',
'Shugo Chara! Amunonijiro Chara Change',
'Zero no Tsukaima: Muma ga Tsumugu Yokaze no Nocturne',
'Knockout Kings 2002',
'Super Adventure Island',
'Kekkon Zenya',
'Super Robot Taisen: Original Generation',
'The Warriors',
'Secret Agent Clank',
'Choro Q3',
'My Baby Boy',
'Tetris Worlds',
'Killzone 3',
'Furu Furu Park',
'Infinite Undiscovery',
"Pac-Man Player's Choice Bundle",
'Metal Slug 3',
'Oshiri Kajiri Mushi no Rhythm Lesson DS: Kawai Ongaku Kyoushitsu Kanshuu',
'MTV Music Generator 2',
'Gallop Racer 2004',
'Hisshou Pachinko*Pachi-Slot Kouryaku Series Vol.10: CR Shinseiki Evangelion: Kiseki no Kachi',
'Outdoors Unleashed: Africa 3D',
'Kawasaki Snowmobiles',
'Disney Fairies: Tinker Bell and the Great Fairy Rescue',
'Barbarian',
'The Idolmaster 2',
'LEGO Harry Potter: Years 1-4',
'Naruto: Uzumaki Chronicles 2',
'World of Warcraft: Wrath of the Lich King',
'Scooby-Doo and the Cyber Chase',
'Inazuma Eleven',
'Princess Natasha: Student, Secret Agent, Princess',
'Action Bass',
'The Awakened Fate Ultimatum',
'Resident Evil Archives: Resident Evil',
'SingStar Pop Edition',
'Major League Baseball 2K7',
'Tom and Jerry Tales',
'Samantha Swift and the Hidden Roses of Athena',
'ESPN: Sports Connection',
'Jam With the Band',
'Bleach: Dark Souls',
'The Legend of Heroes: A Tear of Vermillion',
'Goat Simulator',
'LEGO Racers',
'Zoo Tycoon 2',
'Strawberry Shortcake: Strawberryland Games',
'Batman: Arkham Asylum',
'Ogre Battle Saga Episode Five: The March of the Black Queen',
'Yoake Yori Ruriiro na: Brighter than Dawning Blue',
'Pro Yakyuu Famista DS',
'Crayon Shin-Chan: Arashi o Yobu Enji',
'Arkanoid DS',
'Tekken',
'1942',
'My Weight Loss Coach',
'Rooms: The Main Building',
'Prey',
'Tiger Woods PGA Tour 07',
'Major League Baseball 2K9',
'Nicktoons: Battle for Volcano Island',
'Suikoden Tactics',
"Dynasty Warriors DS: Fighter's Battle",
'Durarara!! 3-way Standoff',
'Pro Evolution Soccer 2010',
'Anarchy Reigns',
'Cruise Ship Vacation Games',
'RollerCoaster Tycoon 2',
'Trade & Battle: Card Hero',
'Caesars Palace 2000: Millennium Gold Edition',
'Jampack: Summer 2003 (RP-T)',
'Rock Band Country Track Pack',
'Maximum Racing: Crash Car Racer',
'Naruto: Ninja Council 3',
'Bloody Roar 4',
'Jet Moto',
'Sonic Advance & Sonic Pinball Party Combo Pack',
'Routes PE',
'Ener-G Horse Riders',
'Swingerz Golf',
'Transworld Snowboarding',
'Chaos;Head Noah',
'Mobile Suit Gundam: Zeonic Front',
'iCarly 2: iJoin The Click!',
'Gundam Assault Survive',
'Mortimer Beckett and the Secrets of Spooky Manor',
'Makai Senki Disgaea Portable: Tsuushin Taisen Hajime Mashita',
'Rocky: Legends',
'Kirarin * Revolution: Minna de Odorou Furi Furi Debut!',
'NHL FaceOff',
'Go Vacation',
'Pro Wrestling',
'Puzzle Challenge: Crosswords and More!',
'All Kamen Rider: Rider Generation 2',
'Final Fantasy Adventure',
'SD Gundam Force: Showdown!',
"Uncharted 3: Drake's Deception",
'Mario Party 3',
'Funky Barn',
'Deca Sports DS',
'Hoshigami: Ruining Blue Earth Remix',
'Sphinx and the Cursed Mummy',
'StarCraft II: Legacy of the Void',
'Max Payne',
'The Misadventures of Tron Bonne',
'We Sing Robbie Williams',
'Mobile Suit Z Gundam: AEUG vs. Titans',
'Major Dream: Major DS Dream Baseball',
'Mad Dog McCree: Gunslinger Pack',
'Virtua Tennis 4: World Tour',
'Silent Hill: Book of Memories',
'Dark Summit',
'Shin Sangoku Musou: Multi Raid 2',
'Winning Post 5',
'Mobile Suit Gundam Side Story II: Aoi o Uketsugu Mono',
'Entaku no Seito: Students of Round',
'Disney Infinity 2.0: Marvel Super Heroes',
'Driver: Parallel Lines',
'Chevrolet Camaro: Wild Ride',
'Chaos Wars',
'Kidou Senshi Gundam 00',
"Women's Volleyball Championship",
'Lunar: Silver Star Harmony',
'Dark Angel: Vampire Apocalypse',
'Jet Moto 2',
'S.Y.K Renshouden Portable',
'WWE Survivor Series',
'Castle Shikigami 2',
'Babysitting Mama',
'Formula 1 Championship Edition',
'Jikkyou Powerful Major League 2009',
'Seaman',
'King Kong',
"Jampack Winter '98",
'Frequency',
'Grand Prix Legends',
'Shadow The Hedgehog',
'Nicktoons: MLB',
'Godzilla Unleashed: Double Smash',
'The Orange Box',
'Star Wars: Rebel Assault II - The Hidden Empire',
'Ski Region Simulator',
'Star Ocean: The Last Hope International',
'Kaidan Restaurant: Ura Menu 100-Sen',
'Momu-chan Diet Wii: Figurobics by Chon Dayon',
'Catz',
'Star Wars Battlefront (2015)',
'MorphX',
"Sesame Street: Cookie's Counting Carnival",
'Crash City Mayhem',
'Blazing Angels 2: Secret Missions of WWII',
'Destiny: The Taken King',
'Super Football',
'Railroad Tycoon 3',
'North American Hunting Extravaganza',
'Super Robot Taisen Compact 2 Dai-1-Bu',
'Gauntlet Legends',
"Scooby-Doo! Who's Watching Who?",
'Naruto: Rise of a Ninja',
"VR Golf '97",
'World Tour Soccer 2005',
'WTF: Work Time Fun',
'Gran Turismo 4 Prologue',
'Mytran Wars',
'SD Gundam Sangokuden Brave Battle Warriors: Shin Mirisha Taisen',
'Kiniro no Corda 3',
'Trivial Pursuit: Bet You Know It',
'Trick x Logic: Season 1',
'Rome: Total War',
'Neo Angelique Special',
'Imagine: Babysitters',
'Chicken Shoot',
'Hello Kitty: Happy Party Pals',
'Dragon Ball: Advanced Adventure',
"The Legend of Zelda: Link's Awakening",
'Vitamin X',
'S.Y.K Renshouden',
'Gaist Crusher God',
'BeatMania Append 3rdMix',
'Pro Yaky? Spirits 4',
'Speedball 2100',
'Bass Strike',
'Lord of the Rings: Battle for Middle-Earth',
'NHL FaceOff 2000',
'SingStar Mallorca Party',
'Darius II',
'Food Network: Cook or Be Cooked',
'Endless Ocean: Blue World',
'Phantasy Star Online 2',
'Nightmare Creatures II',
'Mind Zero',
'Mr. Driller',
'Hakuouki: Yuugi Roku DS',
'PachiPara 13: Super Umi no Pachi-Pro Fuuunroku',
'Jampack Winter 2003 (RP-M)',
'Jewel Legends: Tree of Life',
"GoGo's Crazy Bones",
'Scene It? Box Office Smash',
'Pet in TV',
'Lufia: The Ruins of Lore',
"Drakan: The Ancients' Gates",
'Risen',
'Tennis no Oji-Sama: Doubles no Oji-Sama - Boys, Be Glorious!',
'DS Kageyama Method: Dennou Hanpuku - Tadashii Kanji Kakitori-Kun',
'Ishin Renka: Ryouma Gaiden',
'Juusanshi Engi Engetsu Sangokuden 2',
'Polaris',
'Tenchu: Shadow Assassins',
'Wings of War',
'Justice League: Chronicles',
'The Legend of Zelda: Oracle of Ages',
'Virtua Fighter CG Portrait Series Vol.2: Jacky Bryant',
"Pop'n Music Portable 2",
"Texas Hold 'Em Poker",
'Rygar: The Legendary Adventure',
'Mobile Suit Gundam 00: Gundam Meisters',
'Shadow Master',
'Monster Hunter',
'Densha de Go! Shinkansen EX: Sanyou Shinkansen-hen',
'Heroes of the Pacific',
"Oddworld: Munch's Oddysee",
'Emma at the Farm',
'Mario Party Advance',
'Pocket Pool',
'Twin Strike: Operation Thunder',
'NBA ShootOut 2004',
'Onimusha Essentials',
'Final Fantasy XIV: A Realm Reborn',
'Die Hard Trilogy',
'Gray Matter',
'Stronghold Crusader Extreme',
'ZombiU',
'Tekken 6',
'Ghostbusters: The Video Game (DS Version)',
'Small Soldiers',
'Driver 2 Advance',
'uDraw Studio: Instant Artist',
'Thief: Deadly Shadows',
'Shockwave Assault',
'Just Dance 3',
'Puzzle & Dragons Z + Super Mario Bros. Edition',
'DemiKids: Light Version',
'Chibi Maruko-Chan: Harikiri 365-Nichi no Maki',
'Saints Row 2',
'Marvel Nemesis: Rise of the Imperfects',
'NeverDead',
'R.B.I. Baseball 2016',
'Major League Baseball Featuring Ken Griffey Jr',
'Syndicate',
'Warriors of Might and Magic',
'GTR Evolution',
'Batman Begins',
'Dreamcast Collection',
'Scrabble (Others sales)',
'Rock Band Track Pack Volume 1',
'S.L.A.I.: Steel Lancer Arena International',
'Sorcery',
'Memories Off #5: Togireta Film',
'Mega Man II',
"NHL Powerplay '96",
'Monster Hunter Frontier Online: Season 10',
'Konami Classics Vol. 2',
'Perfect Weapon',
'Ringling Bros. and Barnum & Bailey Circus',
'Rugrats Studio Tour',
'Scene It? Bright Lights! Big Screen!',
'WipEout 64',
'Too Human',
'Castlevania: Lords of Shadow - Mirror of Fate',
'NBA 2K7',
'Highschool Terra Story',
'Unsolved Crimes',
'Frogger: Ancient Shadow',
'One Piece Mansion',
'Tom and Jerry in Fists of Furry',
'Otome wa Oanesama ni Koi Shiteru Portable: 2-Jin no Elder',
'Fishing Master World Tour',
'SoulCalibur Legends (JP sales)',
'WWII Aces',
'Gallop Racer (JP)',
'Suzunone Seven: Rebirth Knot',
'Virtua Pro Football',
"Frankenstein's Monster",
'Automobili Lamborghini',
'The Train Giant',
'World of Warcraft',
'NOëL: NOT DiGITAL',
'Bomberman Tournament',
'Twisted Metal: Head On',
'Pokémon Trading Card Game',
'The Sims 2: Family Fun Stuff',
'Duke Nukem Trilogy: Critical Mass',
'Phantasy Star Universe',
'MediEvil',
'Professor Layton and the Azran Legacy',
"Pretty Cure All Stars Everyone Gather ? Let's Dance!",
'Kisuato',
"Kuroko's Basketball: Ties to Future",
'Keiba Eight Special',
'FIFA Soccer 13',
'Super Mario 64',
'The Princess and the Frog',
'Snowboarding',
'Rugrats: Search For Reptar',
'Summer Athletics',
'Attack of the Movies 3D',
'DEATH NOTE: L o Tsugu Mono',
'Naruto Shippuuden: Shinobi Retsuden III',
'Super Puzzle Fighter II Turbo',
'Digimon World 2',
'1 vs. 100',
'Kuma-Tomo',
'DDRMAX2: Dance Dance Revolution',
'Kao Challengers',
'.hack//Mutation Part 2',
'Command & Conquer 4: Tiberian Twilight',
'Kung Fu Panda 2',
'Dodge Racing: Charger vs Challenger',
'Shadow Hearts',
'New Tokyo Legacy: Operation Babel',
"FunkMaster Flex's Digital Hitz Factory",
'WWF SmackDown!',
'Mission: Impossible - Operation Surma',
'Jikkyou Powerful Pro Yakyuu 8',
'Jinsei Game Wii',
'BoomBots',
"Rock 'N Roll Racing",
'Wheel of Fortune',
'Ms. Pac-Man',
'Solitaire Overload',
'Gears of War 2',
'Family Party: 30 Great Games Winter Fun',
"Famista '92",
'Puzzle de Harvest Moon',
'Dai-2-Ji Super Robot Taisen Z: Hakai-hen',
'Guilty Gear Judgment',
'Bad Apple Wars',
'Super Bomberman 5',
'Doraemon: Nobita no Kyouryuu 2006 DS',
'Sunny Garcia Surfing',
'Kung Fu Panda',
'Camping Mama: Outdoor Adventures',
'Cabbage Patch Kids: The Patch Puppy Rescue',
'Command & Conquer',
'The Cat in the Hat',
'Big Brain Academy: Wii Degree',
'Tamagotchi no Doki Doki Dream Omisecchi',
'Death Connection Portable',
'Brain Boost: Gamma Wave',
'Vancouver 2010 - The Official Video Game of the Olympic Winter Games',
'Tecmo Classic Arcade',
'Hot Shots Tennis: Get a Grip',
'Battlefield: Hardline',
'White Knight Chronicles II',
'Defiance',
'Star Soldier: Vanishing Earth',
'Guild Wars',
'Guild Wars Trilogy',
'Ninja: Shadow of Darkness',
'Mahjong Fight Club DS: Wi-Fi Taiou',
'Super Mario Land',
'Hot Wheels Velocity X',
"Cabela's Big Game Hunter: Hunting Party",
'Machi: Unmei no Kousaten',
'Refrain no Chika Meikyuu to Majo no Ryodan',
'LEGO Dimensions',
'Football Manager 2014',
'Etrian Mystery Dungeon',
'Fatal Fury: Wild Ambition',
'Maximum Force',
'Goldeneye 007: Reloaded',
'Digimon Battle Spirit 2',
'N3: Ninety-Nine Nights',
'Dead Rising 2: Off the Record',
'Sub Rebellion',
'Moto Racer World Tour',
'Downtown no Gaki no Tsukai Yaarahen de!! Zettai ni Tsukamatte Haikenai Gasu Kurobikari Land',
'Saint Seiya Omega: Ultimate Cosmo',
'Backyard Baseball 2006',
'The Legend of Zelda',
'Kimi ni Todoke: Tsutaeru Kimochi',
'WCW Mayhem',
'Knights in the Nightmare: DHE Series Special Pack',
'Gex',
'Hot Wheels: World Race',
'Spirits and Demons: 5 Game Pack',
'CandyLand / Chutes & Ladders / Memory',
'R4: Ridge Racer Type 4',
'Samurai Warriors',
'Fantastic 4',
'Shin Hisui no Shizuku: Hiiro no Kakera 2 Portable',
'Kamen Rider: Battride War',
'NFL Street 3',
"Tony Hawk's American Sk8land",
"Disney's Treasure Planet",
'Renai Banchou: Inochi Meishi, Koiseyo Otome! Love is Power!!!',
"WWE '13",
'Shutter Island',
'Super Mario Sunshine',
'Mischief Makers',
'EyeToy: Operation Spy',
'Serious Sam',
'Magna Carta 2',
'Star Wars: Rogue Squadron',
'Bratz: Rock Angelz',
'My Make-Up',
'Tokimeki no Houkago',
'Don King Presents: Prizefighter',
'Zero4 Champ RR-Z',
'Rally Championship',
'Army Men: Air Combat',
'Backstreet Billiards',
'The Sims 4: Spa Day',
'Time Crisis II',
'Puzzle City',
'Makai Senki Disgaea 4: Fuuka & Desco-hen Hajime Mashita',
'Inazuma Eleven 3',
'World Championship Rugby',
"A Witch's Tale",
'Casper: Spirit Dimensions',
'Crime Killer',
'Adventure Time: Finn & Jake Investigations',
'Blue Breaker: Ken Yorimo Hohoemi o',
'Burnout Revenge',
'Disgaea 2: Cursed Memories',
'Naruto Shippuden: Ultimate Ninja Storm Collection',
"Broken Sword 5: The Serpent's Curse",
'BloodRayne 2',
'Top Spin',
'Deathrow: Underground Team Combat',
'Nanotek Warrior',
'TimeSplitters: Future Perfect',
'Rad Racer',
'Star Trek: Conquest',
'Harvest Moon: The Lost Valley',
'The Granstream Saga',
'Blazblue: Continuum Shift Extend',
'Portal 2',
'Cosmic Ark',
'MySims Party',
"Peter Pan in Disney's Return to Never Land",
```

```
 'the evil within',
 'ddrmax2: dance dance revolution',
 'crazy taxi',
 'virtua fighter 4',
 'sly 2: band of thieves',
 'guitar hero: world tour',
 'sonic advance 2',
 'madden nfl 07',
 'nba 2k11',
 'nfl gameday 2000',
 'call of duty: black ops 3',
 "disney's the lion king",
 'beyond: two souls',
 'kinect star wars',
 'midnight club ii',
 'dragon quest monsters: joker',
 'spongebob squarepants: supersponge',
 'the getaway: black monday',
 'professor layton and the mask of miracle',
 'just cause 2',
 "dragon's dogma",
 'the legend of zelda: the wind waker',
 '50 cent: bulletproof',
 'high school musical: sing it!',
 'wii party u',
 'madden nfl 25',
 'final fantasy ii',
 'kirby 64: the crystal shards',
 'dead or alive 3',
 'ufc 2009 undisputed',
 'metroid ii: return of samus',
 'wwf attitude',
 'the spongebob squarepants movie',
 'golden sun',
 'sonic the hedgehog 3',
 'kid icarus',
 'def jam: fight for ny',
 "tom clancy's ghost recon",
 'state of emergency',
 'bioshock infinite',
 'hitman: absolution',
 '2 games in 1 double pack: the incredibles / finding nemo: the continuing adventure
s',
 'call of duty: black ops 3',
 'the urbz: sims in the city',
 'fifa soccer 08',
 'star wars: starfighter',
 'madden nfl 2000',
 'kingdom hearts',
 'star wars: battlefront',
 'pro evolution soccer',
 'crackdown',
 'yoshi',
 "disney's aladdin",
 ...]
```

In [104]:
```
#lower_case = lower_case.values.tolist()
```

In [105]:
```
# Strip whitespaces
strip_whitespace = [string.strip() for string in lower_case]
# Show text
strip_whitespace
```

Out[105]:
```
['wii sports',
 'super mario bros.',
 'mario kart wii',
 'wii sports resort',
 'pokemon red/pokemon blue',
 'tetris',
 'new super mario bros.',
 'wii play',
 'new super mario bros. wii',
 'duck hunt',
 'nintendogs',
 'mario kart ds',
 'pokemon gold/pokemon silver',
 'wii fit',
 'wii fit plus',
 'kinect adventures!',
 'grand theft auto v',
 'grand theft auto: san andreas',
 'super mario world',
 'brain age: train your brain in minutes a day',
 'pokemon diamond/pokemon pearl',
 'super mario land',
 'super mario bros. 3',
 'grand theft auto v',
 'grand theft auto: vice city',
 'pokemon ruby/pokemon sapphire',
 'pokemon black/pokemon white',
 'brain age 2: more training in minutes a day',
 'gran turismo 3: a-spec',
 'call of duty: modern warfare 3',
 'pokémon yellow: special pikachu edition',
 'call of duty: black ops',
 'pokemon x/pokemon y',
 'call of duty: black ops 3',
 'call of duty: black ops ii',
 'call of duty: black ops ii',
 'call of duty: modern warfare 2',
 'call of duty: modern warfare 3',
 'grand theft auto iii',
 'super smash bros. brawl',
 'call of duty: black ops',
 'animal crossing: wild world',
 'mario kart 7',
 'halo 3',
 'grand theft auto v',
 'pokemon heartgold/pokemon soulsilver',
 'super mario 64',
 'gran turismo 4',
 'super mario galaxy',
 'pokemon omega ruby/pokemon alpha sapphire',
 'super mario land 2: 6 golden coins',
 'grand theft auto iv',
 'gran turismo',
 'super mario 3d land',
 'gran turismo 5',
 'call of duty: modern warfare 2',
 'grand theft auto iv',
 'super mario all-stars',
 'pokemon firered/pokemon leafgreen',
 'super mario 64',
 'just dance 3',
 'call of duty: ghosts',
 'halo: reach',
 'mario kart 64',
 'new super mario bros. 2',
 'halo 4',
 'final fantasy vii',
 'call of duty: ghosts',
 'just dance 2',
 'gran turismo 2',
 'call of duty 4: modern warfare',
 'donkey kong country',
 'minecraft',
 'animal crossing: new leaf',
 'mario party ds',
 'the elder scrolls v: skyrim',
 'super mario kart',
 'fifa 16',
 'wii party',
 'halo 2',
 'mario party 8',
 'pokemon black 2/pokemon white 2',
 'fifa soccer 13',
 'the sims 3',
 'goldeneye 007',
 'mario & sonic at the olympic games',
 'final fantasy x',
 'final fantasy viii',
 'pokémon platinum version',
 'pac-man',
 'grand theft auto: liberty city stories',
 'super mario galaxy 2',
 'star wars battlefront (2015)',
 'call of duty: advanced warfare',
 'the legend of zelda: ocarina of time',
 'crash bandicoot 2: cortex strikes back',
 'super mario bros. 2',
 'super smash bros. for wii u and 3ds',
 'call of duty: world at war',
 'battlefield 3',
 'the legend of zelda: twilight princess',
 'call of duty: black ops 3',
 'just dance',
 'battlefield 3',
 'need for speed underground',
 'tekken 3',
 'crash bandicoot 3: warped',
 'super smash bros. melee',
 'mario kart 8',
 'fallout 4',
 'mario kart: double dash!!',
 'just dance 4',
 'fifa 14',
 'need for speed underground 2',
 'medal of honor: frontline',
 "uncharted 3: drake's deception",
 'crash bandicoot',
 'zumba fitness',
 'gears of war 2',
 'uncharted 2: among thieves',
 'call of duty 4: modern warfare',
 'fifa 12',
 'big brain academy',
 'red dead redemption',
 'fifa 15',
 'donkey kong country returns',
 'the elder scrolls v: skyrim',
 'the legend of zelda',
 "assassin's creed iii",
 'halo: combat evolved',
 'pokémon emerald version',
 'kingdom hearts',
 'pokémon crystal version',
 'halo 3: odst',
 'red dead redemption',
 'super mario sunshine',
 'street fighter ii: the world warrior',
 'world of warcraft',
 'driver',
 'kinect sports',
 'gears of war 3',
 'gears of war',
 'metal gear solid 2: sons of liberty',
 'sonic the hedgehog 2',
 'metal gear solid 4: guns of the patriots',
 'metal gear solid',
 'the last of us',
 'final fantasy xii',
 'littlebigplanet',
 'dragon quest ix: sentinels of the starry skies',
 'lego star wars: the complete saga',
 'resident evil 2',
 'tekken 2',
 'cooking mama',
 'destiny',
 'tetris',
 "assassin's creed ii",
 'super smash bros.',
 "assassin's creed",
 'batman: arkham city',
 'forza motorsport 3',
 'monster hunter freedom unite',
 'super mario advance',
 'mario kart: super circuit',
 'super mario world',
 'pokemon stadium',
 'call of duty: world at war',
 'crash bandicoot: the wrath of cortex',
 'minecraft',
 'final fantasy xiii',
 'dr. mario',
 'pokemon pinball',
 "assassin's creed iii",
 'final fantasy ix',
 'final fantasy x-2',
 'donkey kong 64',
 "assassin's creed ii",
 'professor layton and the curious village',
 'tomb raider ii',
 'dragon quest viii: journey of the cursed king',
 'diablo iii',
 'super mario bros. 3',
 'new super mario bros. u',
 'super mario land 3: wario land',
 'fifa soccer 13',
 'the legend of zelda: phantom hourglass',
 "donkey kong country 2: diddy's kong quest",
 'tomodachi life',
 'mario & sonic at the olympic games',
 'medal of honor: rising sun',
 "kirby's dream land",
 'fable iii',
 'call of duty: advanced warfare',
 'microsoft flight simulator',
 'guitar hero ii',
 'resident evil 5',
 'grand theft auto v',
 'grand theft auto: vice city stories',
 'fifa soccer 11',
 'super mario bros.',
 'resident evil',
 "tony hawk's pro skater",
 'super smash bros. for wii u and 3ds',
 'warzone 2100',
 'spyro the dragon',
 "link's crossbow training",
 'guitar hero iii: legends of rock',
 'fallout 3',
 'pokemon mystery dungeon: explorers of time/explorers of darkness',
 "uncharted: drake's fortune",
 'madden nfl 06',
 'lego star wars: the complete saga',
 'diddy kong racing',
 'monster hunter freedom 3',
 'dr. mario',
 'god of war iii',
 'starcraft ii: wings of liberty',
 "assassin's creed",
 'fifa 15',
 'crash team racing',
 'fifa 17',
 'batman: arkham city',
 'driver 2',
 'the simpsons: hit & run',
 "tony hawk's pro skater 2",
 'animal crossing: city folk',
 'the lord of the rings: the two towers',
 'forza motorsport 4',
 'tomb raider',
 'guitar hero iii: legends of rock',
 'the legend of zelda: a link to the past',
 'the legend of zelda: the wind waker',
 'mario & sonic at the olympic winter games',
 "luigi's mansion: dark moon",
 'splatoon',
 'the last of us',
 'guitar hero iii: legends of rock',
 'madden nfl 2005',
 'pitfall!',
 'madden nfl 07',
 'spider-man: the movie',
 'dragon quest vii: warriors of eden',
 'uncharted: the nathan drake collection',
 'call of duty: advanced warfare',
 'god of war',
 'nintendo land',
 'michael jackson: the experience',
 "tony hawk's pro skater 3",
 'winning eleven: pro evolution soccer 2007',
 'the elder scrolls iv: oblivion',
 'zelda ii: the adventure of link',
 'need for speed: most wanted',
 'resistance: fall of man',
 'sonic the hedgehog',
 'kingdom hearts ii',
 'fifa 14',
 'call of duty: advanced warfare',
 'asteroids',
 'fable ii',
 'halo 5: guardians',
 'super mario 3d world',
 'batman: arkham asylum',
 "assassin's creed: revelations",
 'namco museum',
 'metal gear solid 3: snake eater',
 'daxter',
 'warcraft ii: tides of darkness',
 'the legend of zelda: ocarina of time',
 'fifa soccer 06',
 'eyetoy play',
 "uncharted 4: a thief's end",
 'gran turismo 5 prologue',
 "assassin's creed: revelations",
 'street fighter iv',
 'fifa 12',
 'teenage mutant ninja turtles',
 'excitebike',
 'frogger',
 'madden nfl 2003',
 'watch dogs',
 'half-life',
 "super mario world 2: yoshi's island",
 'fifa soccer 07',
 'the legend of zelda: skyward sword',
 'street fighter ii turbo',
 'carnival games',
 'fallout 4',
 'world of warcraft: the burning crusade',
 'fallout: new vegas',
 'god of war ii',
 'forza motorsport 2',
 'world soccer winning eleven 9',
 'namco museum vol.3',
 'tekken tag tournament',
 'far cry 4',
 'star fox 64',
 'minecraft',
 'golf',
 'fallout 3',
 'professor layton and the diabolical box',
 'left 4 dead 2',
 'namco museum: 50th anniversary',
 'ea sports active',
 "assassin's creed: unity",
 'donkey kong land',
 "tony hawk's underground",
 "assassin's creed iv: black flag",
 'monster hunter 4 ultimate',
 'motorstorm',
 'the elder scrolls v: skyrim',
 'tekken 5',
 'dragon warrior iii',
 'sports champions',
 'nba 2k16',
 'world soccer winning eleven 8 international',
 'flash focus: vision training in minutes a day',
 'namco museum vol.1',
 "the legend of zelda: link's awakening",
 "mario & luigi: bowser's inside story",
 'big brain academy: wii degree',
 'nintendogs + cats',
 'batman: arkham knight',
 'mario & sonic at the london 2012 olympic games',
 'lego indiana jones: the original adventures',
 "resident evil director's cut",
 'ratchet & clank: size matters',
 'just dance 2014',
 'the sims: unleashed',
 'super paper mario',
 "harry potter and the sorcerer's stone",
 'the witcher 3: wild hunt',
 'resident evil 3: nemesis',
 'spyro: year of the dragon',
 'fifa soccer 2005',
 'pokémon trading card game',
 'battlefield 4',
 'mysims',
 'friend collection',
 "tony hawk's pro skater 4",
 'midnight club 3: dub edition',
 'banjo-kazooie',
 'socom: u.s. navy seals',
 'guitar hero: world tour',
 'gran turismo 6',
 'jak and daxter: the precursor legacy',
 'pokemon snap',
 'fifa soccer 10',
 'pro evolution soccer 2008',
 'call of duty: ghosts',
 'resident evil 4',
 "tom clancy's the division",
 'star wars: battlefront',
 'doom ii: hell on earth',
 'the simpsons: road rage',
 'cooking mama 2: dinner with friends',
 "luigi's mansion",
 'star wars: battlefront ii',
 'wwf smackdown!',
 'croc: legend of the gobbos',
 'grand theft auto: liberty city stories',
 'the getaway',
 'tomb raider iii: adventures of lara croft',
 'james bond 007: agent under fire',
 'lego star wars: the video game',
 'battlefield 4',
 'fifa soccer 11',
 'rugrats in paris: the movie',
 "spyro 2: ripto's rage!",
 'left 4 dead',
 'mario & sonic at the olympic winter games',
 'resident evil 5',
 "donkey kong country 3: dixie kong's double trouble!",
 "assassin's creed: brotherhood",
 'kung fu',
 'batman: arkham asylum',
 'clubhouse games',
 'star wars battlefront (2015)',
 'battlefield: bad company 2',
 'guitar hero: on tour',
 "assassin's creed: unity",
 'far cry 3',
 'monster hunter 4',
 'borderlands',
 'tekken 4',
 'battlefield 4',
 'littlebigplanet 2',
 'grand theft auto 2',
 'final fantasy iii',
 'spider-man 2',
 'f-1 race',
 'fifa soccer 2003',
 'crash bash',
 'lego batman: the videogame',
 'metal gear solid v: the phantom pain',
 'the legend of zelda: spirit tracks',
 'wwf war zone',
 "the legend of zelda: majora's mask",
 'professor layton and the unwound future',
 'rugrats: search for reptar',
 "assassin's creed iv: black flag",
 'english training: have fun improving your skills!',
 'james bond 007: nightfire',
 'ratchet & clank',
 'mario party 9',
 'star wars episode iii: revenge of the sith',
 'max payne',
 'the lord of the rings: the return of the king',
 "assassin's creed syndicate",
 'destiny',
 'true crime: streets of la',
 'wii music',
 'gran turismo (psp)',
 'monster hunter x',
 'tekken',
 'fifa 16',
 'l.a. noire',
 'madden nfl 16',
 'yokai watch 2 ganso/honke',
 '007: tomorrow never dies',
 'baseball',
 'killer instinct',
 'wwf smackdown! 2: know your role',
 'crisis core: final fantasy vii',
 'god of war: chains of olympus',
 'dragon quest vi: maboroshi no daichi',
 'dance central',
 'enter the matrix',
 'super mario maker',
 'fallout: new vegas',
 'rhythm heaven',
 'ace combat 04: shattered skies',
 'club penguin: elite penguin force',
 'animal crossing',
 'lego batman: the videogame',
 'sonic rush',
 'resident evil 6',
 'halo: the master chief collection',
 'fifa soccer 08',
 'spider-man',
 'need for speed iii: hot pursuit',
 'star wars episode i racer',
 'personal trainer: cooking',
 'the elder scrolls iv: oblivion',
 'dragon warrior iv',
 'heavy rain',
 'mass effect 2',
 'fifa 15',
 'dragon ball z: budokai',
 'madden nfl 2002',
 'world class track meet',
 'borderlands 2',
 'the legend of zelda: a link between worlds',
 'donkey kong',
 'the sims: vacation',
 'mass effect 3',
 'kung fu panda',
 'far cry 3',
 'epic mickey',
 'saints row 2',
 'rayman',
 'killzone 2',
 "tom clancy's splinter cell",
 "mike tyson's punch-out!!",
 'wwe smackdown! shut your mouth',
 'gears of war: ultimate edition',
 'nba 2k13',
 'dragon ball z: budokai tenkaichi 3',
 'middle-earth: shadow of mordor',
 'star fox',
 'the sims: livin large',
 'world soccer winning eleven 6 international',
 'devil may cry',
 'animal crossing: happy home designer',
 'battlefield: bad company 2',
 'the sims 4',
 'star wars: the old republic',
 'sonic heroes',
 'kirby super star ultra',
 'hitman 2: silent assassin',
 'street fighter iv',
 'wave race 64',
 'socom ii: u.s. navy seals',
 'wwe smackdown! vs. raw 2006',
 'infamous',
 'pac-man collection',
 'mass effect',
 'warioware: smooth moves',
 'cooking mama: cook off',
 "super mario world 2: yoshi's island",
 'fifa 14',
 'titanfall',
 'sonic mega collection plus',
 'sega superstars tennis',
 'world soccer winning eleven 7 international',
 'need for speed: most wanted',
 'hot shots golf 3',
 'call of duty: finest hour',
 'zumba fitness 2',
 'syphon filter',
 'imagine: babyz',
 'colin mcrae rally',
 'call of duty: ghosts',
 'madden nfl 13',
 'the elder scrolls iii: morrowind',
 'mortal kombat',
 'f-zero',
 "yoshi's story",
 'command & conquer: red alert',
 "skylanders: spyro's adventure",
 'fifa 16',
 "assassin's creed: brotherhood",
 'bioshock',
 'madden nfl 10',
 'medal of honor',
 'wwe smackdown! vs. raw',
 'metroid prime',
 'finding nemo',
 'myst',
 'need for speed: hot pursuit',
 'imagine: fashion designer',
 'wwe smackdown! here comes the pain',
 'star wars: the force unleashed',
 "yoshi's island ds",
 'hot shots golf',
 'killzone 3',
 "a bug's life",
 'wwf smackdown! just bring it',
 'dragon quest v: tenkuu no hanayome',
 'mortal kombat trilogy',
 'battlefield 3',
 'nba street',
 'professor layton and the last specter',
 'monster hunter tri',
 'jak ii',
 'infamous: second son',
 'ratchet & clank: going commando',
 'tekken 6',
 'the sims',
 'sonic and the secret rings',
 'saints row: the third',
 'missile command',
 'mario paint',
 "assassin's creed iv: black flag",
 'pokémon stadium 2',
 'metroid',
 'l.a. noire',
 'call of duty 3',
 'the legend of zelda: a link to the past',
 'onimusha: warlords',
 'mortal kombat x',
 'mario party',
 'madden nfl 12',
 'lego star wars ii: the original trilogy',
 'madden nfl 11',
 'mario vs. donkey kong: mini-land mayhem!',
 'skate 3',
 'burnout 3: takedown',
 'yokai watch 2 shinuchi',
 'need for speed: shift',
 'final fantasy xiii-2',
 'mortal kombat',
 'call of duty 2: big red one',
 'guitar hero: world tour',
 'medal of honor',
 'fable',
 'hannah montana',
 "tom clancy's splinter cell",
 'skylanders giants',
 'medal of honor heroes',
 'super mario bros.: the lost levels',
 'star wars: shadows of the empire',
 'lego star wars: the complete saga',
 'midnight club: street racing',
 'pac-man: adventures in time',
 'halo wars',
 'god of war collection',
 'espn nfl 2k5',
 'mario strikers charged',
 'medal of honor',
 'fifa soccer 10',
 'nba 2k12',
 'style savvy',
 'harry potter and the chamber of secrets',
 'super mario all-stars: limited edition',
 'streets of rage',
 'fifa soccer 09',
 'ratchet & clank future: tools of destruction',
 'nba live 2005',
 'littlebigplanet',
 'killzone: shadow fall',
 'ghostbusters: the video game (ds version)',
 'monster hunter freedom 2',
 'wwe smackdown vs. raw 2007',
 'ratchet & clank: up your arsenal',
 'art academy',
 'dragon age: origins',
 'madden nfl 08',
 'sonic adventure 2 battle',
 'dragon warrior ii',
 'madden nfl 13',
 'pokémon colosseum',
 'deca sports',
 'madden nfl 09',
 'crazy taxi',
 'perfect dark',
 'game party',
 'dragon warrior',
 'tomb raider (2013)',
 'marvel: ultimate alliance',
 'pes 2009: pro evolution soccer',
 'mario party 2',
 'saints row: the third',
 'rockstar games double pack: grand theft auto iii & grand theft auto vice city',
 'lego harry potter: years 1-4',
 'mortal kombat: deadly alliance',
 'resistance 2',
 'pure',
 'tomb raider: the last revelation',
 'nba 2k13',
 'nba 2k14',
 'mario party 4',
 'udraw studio',
 'midnight club: los angeles',
 'final fantasy v',
 'final fantasy tactics',
 'twisted metal 2',
 'halo: combat evolved anniversary',
 'need for speed carbon: own the city',
 "the sims: bustin' out",
 'riven: the sequel to myst',
 "tom clancy's rainbow six: vegas 2",
 'sonic adventure',
 'ice hockey',
 'pro evolution soccer 2011',
 'dragon ball z: budokai 2',
 'atv offroad fury',
 'dino crisis',
 'minecraft',
 'need for speed: prostreet',
 'madden nfl 16',
 'tetris plus',
 'pro wrestling',
 'theme hospital',
 'need for speed (2015)',
 'zumba fitness',
 'monopoly',
 'batman: arkham origins',
 'dead island',
 'guitar hero',
 'lego star wars: the complete saga',
 'kinect: disneyland adventures',
 'wcw/nwo revenge',
 'bloodborne',
 'nba 2k15',
 'madden nfl 11',
 'yu-gi-oh! the duelists of the roses',
 'madden nfl 08',
 'half-life 2',
 'the incredibles',
 "tony hawk's underground 2",
 'madden nfl 10',
 'kinect sports: season two',
 "kirby's dream land 2",
 'lego harry potter: years 1-4',
 'donkey kong land ii',
 'dragon warrior monsters',
 'midnight club: los angeles',
 'skylanders swap force',
 'wwe smackdown vs raw 2008',
 'my fitness coach',
 'resident evil - code: veronica x',
 'the sims 3',
 "the legend of zelda: majora's mask 3d",
 'nba 2k16',
 'just dance kids',
 'mario tennis',
 'grand theft auto',
 'james bond 007: everything or nothing',
 'warioware touched!',
 'need for speed: high stakes',
 'chrono trigger',
 'cool boarders 3',
 'lego marvel super heroes',
 'guild wars 2',
 'pokemon ranger: shadows of almia',
 'dead island',
 'tomb raider ii',
 'god of war: ascension',
 'need for speed underground 2',
 'pro evolution soccer 2010',
 'just dance 2015',
 'yokai watch busters',
 'mario bros.',
 'fifa soccer 09',
 'star wars: the force unleashed',
 'borderlands 2',
 'toca touring car championship',
 'paper mario: sticker star',
 "sonic's ultimate genesis collection",
 'super monkey ball: banana blitz',
 'kirby: nightmare in dream land',
 'wario land 4',
 'guitar hero iii: legends of rock',
 'resident evil 4',
 'minecraft',
 "onimusha 2: samurai's destiny",
 'cool boarders 2',
 'paper mario: the thousand-year door',
 'madden nfl 12',
 'monster hunter tri',
 'dissidia: final fantasy',
 'sonic advance',
 'spyro: season of ice',
 'teenage mutant ninja turtles ii: the arcade game',
 'lego indiana jones: the original adventures',
 'dance dance revolution x2',
 'new super luigi u',
 'prince of persia: the sands of time',
 'scribblenauts',
 "the legend of zelda: link's awakening dx",
 'need for speed: hot pursuit',
 'bioshock 2',
 "assassin's creed iv: black flag",
 "devil may cry 3: dante's awakening special edition",
 'rock band 2',
 'frogger',
 'guitar hero ii',
 'wario land 3',
 'pro yakyuu family stadium',
 'sonic unleashed',
 "high school musical: makin' the cut!",
 'pro evolution soccer 2012',
 'saints row',
 'star wars: knights of the old republic',
 "kirby's pinball land",
 'donkey kong country',
 'hitman: absolution',
 'yu-gi-oh: duel monsters 4',
 'atv offroad fury 2',
 'forza motorsport 5',
 'madden nfl 25',
 'pokemon ranger',
 'mario & luigi: superstar saga',
 'final fantasy xiii',
 'star wars: rogue squadron',
 'tennis',
 'the sims: house party',
 'your shape featuring jenny mccarthy',
 'syphon filter 2',
 "tom clancy's rainbow six: siege",
 'volleyball',
 'fifa 15',
 'call of duty: roads to victory',
 'super mario rpg: legend of the seven stars',
 'mahjong',
 'r.c. pro-am',
 'simcity 2000',
 'demon attack',
 'dance dance revolution extreme',
 'tetris ds',
 'final fantasy tactics advance',
 'far cry: primal',
 'toy story 3: the video game',
 'final fantasy iii',
 "kirby's epic yarn",
 'project gotham racing',
 'golf',
 'nba 2k12',
 'borderlands',
 "tony hawk's pro skater",
 'pes 2009: pro evolution soccer',
 'need for speed: hot pursuit 2',
 'mario sports mix',
 'dynasty warriors 4',
 'mortal kombat',
 'parasite eve',
 'scooby-doo! night of 100 frights',
 'project gotham racing (jp weekly sales)',
 'need for speed: most wanted 5-1-0',
 'portal 2',
 "devil may cry 3: dante's awakening",
 'derby stallion',
 'dead rising',
 'finding nemo',
 'nba 2k11',
 'wipeout: the game',
 'the sims 2: pets',
 'battlefield: hardline',
 'metal gear solid: peace walker',
 'world of warcraft: cataclysm',
 'tetris worlds',
 'need for speed: undercover',
 'dance dance revolution: hottest party',
 'mario party 5',
 'ghostbusters: the video game (ds version)',
 'sim theme park',
 'warcraft: orcs & humans',
 'skate 3',
 'spec ops: airborne commando',
 'kingdom hearts: birth by sleep',
 'ufc 2009 undisputed',
 'dark souls',
 'yu-gi-oh! the eternal duelist soul',
 'need for speed: prostreet',
 'the sims 3',
 "tom clancy's splinter cell: conviction",
 'pokemon mystery dungeon: red/blue rescue team',
 'the incredibles',
 'r4: ridge racer type 4',
 'soulcalibur ii',
 'mario & luigi: dream team',
 'need for speed: prostreet',
 'carnival games',
 'nba 2k14',
 'pokémon mystery dungeon: blue rescue team',
 'sonic mega collection',
 'diablo iii',
 'nba jam',
 'pgr4 - project gotham racing 4',
 'kingdom hearts 358/2 days',
 'dance central 2',
 'rayman raving rabbids: tv party',
 'socom: u.s. navy seals: combined assault',
 "peter jackson's king kong: the official game of the movie",
 'donkey kong country',
 'need for speed: the run',
 'madden nfl 15',
 '1080°: teneighty snowboarding',
 'just dance 3',
 'nba live 2004',
 'dead space',
 'call of duty 2',
 'madden nfl 15',
 'forza horizon',
 'touchmaster',
 "tom clancy's the division",
 'lego marvel super heroes',
 'fire emblem: awakening',
 'driveclub',
 'super street fighter ii',
 'mortal kombat vs. dc universe',
 'tenchu: stealth assassins',
 'tennis',
 'harry potter and the chamber of secrets',
 'harry potter and the chamber of secrets',
 'simcity',
 'battlefield 4',
 'resident evil 6',
 'dragon age: inquisition',
 'sonic classic collection',
 'spyro: enter the dragonfly',
 'call of duty: world at war',
 'devil may cry 2',
 'burnout paradise',
 'e.t.: the extra terrestrial',
 'bioshock infinite',
 'hot shots golf: open tee',
 'rad racer',
 'soccer',
 'kirby squeak squad',
 "tom clancy's rainbow six: vegas",
 'destiny',
 'crash nitro kart',
 'grand theft auto: san andreas',
 'star wars: dark forces',
 'nba street vol. 2',
 'def jam vendetta',
 'fifa soccer 09',
 'who wants to be a millionaire',
 'scooby doo and the cyber chase',
 'alleyway',
 'dragon ball z: budokai 3',
 'jet moto 2',
 "tom clancy's ghost recon",
 'x-men legends',
 'batman: arkham origins',
 'guitar hero: world tour',
 'mortal kombat ii',
 'virtua fighter 2',
 'tetris dx',
 'mortal kombat vs. dc universe',
 'watch dogs',
 'mario super sluggers',
 'tomb raider (2013)',
 'the legend of zelda: oracle of ages',
 'need for speed: shift',
 'monopoly',
 'red faction',
 "the sims: makin' magic",
 'mx unleashed',
 'parappa the rapper',
 'toy story mania!',
 'mario party 3',
 'star wars: the force unleashed',
 'game party 2',
```

```
# Import necessary libraries
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Load the dataset
data = np.loadtxt('housing.csv', delimiter=',', skiprows=1)

# Split the data into features and target variable
X = data[:, :-1]
y = data[:, -1]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create a linear regression model
model = LinearRegression()

# Train the model on the training data
model.fit(X_train, y_train)

# Make predictions on the testing data
y_pred = model.predict(X_test)

# Calculate the mean squared error and R-squared score
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

# Print the results
print('Mean Squared Error:', mse)
print('R-squared Score:', r2)
```

```
'XXXXXXX XXXXXX',
'XXXXX & XXXXX: XXXXXXXXX XXXX',
'XXXXX XXXXXXX XXXX',
'XXXX XXXX: XXXXX XXXXXXXX',
'XXXXXX',
'XXX XXXX: XXXXX XXXXX',
'XXXX XXXXX XXXXXXXXX XXXXX XXXXXXXX',
'XXXXXX XXXXXX 2',
"XXX XXXXXX'X XXXXXXX XXX: XXXXX",
'XXXXXXXXXX',
'XXXX 15',
'XXXX XX XXXX: XXXXX XX XXXXXXX',
'XXXXX XXXXX XXX: XXXXXX XX XXX XXXXX XXXXX',
'XXXXXXX',
'XX XXX-XX',
'XXXXXXX 2000',
'XXXXX XXXXXX',
'XXXXX XXXXX XXXXXXXXXX XXXXXXX',
'XXXXXX XX',
'XXXXX XXXXXXX XXXXXXX XXXXXXX',
'XXX XXX: XXXXXX',
'XXX XXXXX 3: XXX XXXXX XXXX',
'XXXXX XXXXXXX XXX',
"XXXXX'X XXXX XXXX",
'XXXXXXX XXXXXX XXXXXX',
'XXXX',
'XXX 2X12',
'XXXXXXXXXXX',
"XXXX XXXX'X XXX XXXXXX",
'XXX 2009: XXX XXXXXXXXX XXXXXX',
'XXXX XXX XXXXX: XXX XXXXXXX 2',
'XXXXX XXXXXX XXX',
'XXXXXXX XXXXXXXX 4',
'XXXXXX XXXXXX',
'XXXXXXXX XXX',
'XXXXXX-XXX! XXXXX XX 100 XXXXXXX',
'XXXXXXX XXXXXX XXXXXX (XX XXXXXX XXXXX)',
'XXXX XXX XXXXX: XXXX XXXXXX 5-1-0',
'XXXXXX 2',
"XXXXX XXX XXX 3: XXXXX'X XXXXXXXXX",
'XXXXX XXXXXXXX',
'XXXX XXXXXX',
'XXXXXXX XXXX',
'XXX 2X11',
'XXXXXXX: XXX XXXX',
'XXX XXXX 2: XXXX',
'XXXXXXXXXXX: XXXXXXXX',
'XXXXX XXXX XXXXX: XXXXX XXXXXX',
'XXXXX XX XXXXXXXX: XXXXXXXXX',
'XXXXXX XXXXXX',
'XXXX XXX XXXXX: XXXXXXXXXX',
'XXXXX XXXXX XXXXXXXXXX: XXXXXXX XXXXX',
'XXXXX XXXXX 5',
'XXXXXXXXXXXX: XXX XXXXX XXXX (XX XXXXXXX)',
'XXX XXXXX XXXX',
'XXXXXXXX: XXXX & XXXXXX',
'XXXXX 3',
'XXXX XXX: XXXXXXXX XXXXXXXX',
'XXXXXXX XXXXXX: XXXXX XX XXXXX',
'XXX 2009 XXXXXXXXXX',
'XXXX XXXXX',
'XX-XX-XX! XXX XXXXXXX XXXXXXX XXXX',
'XXXX XXX XXXXX: XXXXXXXXX',
'XXX XXXX 3',
"XXX XXXXXX'X XXXXXXXX XXXX: XXXXXXXXXX",
'XXXXXXX XXXXXXX XXXXXXX: XXX/XXXX XXXXXX XXXX',
'XXX XXXXXXXXXXX',
'X4: XXXXX XXXXX XXXX 4',
'XXXXXXXXXXX XX',
'XXXXX & XXXXX: XXXXX XXXX',
'XXXX XXX XXXXX: XXXXXXXXX',
'XXXXXXXX XXXXX',
'XXX 2X14',
'XXXéXXX XXXXXXX XXXXXXX: XXXX XXXXXX XXXX',
'XXXXX XXXX XXXXXXXXXX',
'XXXXXX XXX',
'XXX XXX',
'XXX4 - XXXXXXX XXXXXX XXXXXX 4',
'XXXXXXX XXXXXX 358/2 XXXX',
'XXXXX XXXXXXX 2',
'XXXXXX XXXXXX XXXXXXX: XX XXXXX',
'XXXXX: XX XXXX XXXXX: XXXXXXXX XXXXXXX',
"XXXXX XXXXXXX'X XXXX XXXX: XXX XXXXXXXX XXXX XX XXX XXXXX",
'XXXXXX XXXX XXXXXXX',
'XXXX XXX XXXXX: XXX XXX',
'XXXXXX XXX 15',
'1080°: XXXXXXXXX XXXXXXXXXXXX',
'XXXX XXXXX 3',
'XXX XXXX 2004',
'XXXX XXXXX',
'XXXX XX XXXX 2',
'XXXXXX XXX 15',
'XXXXX XXXXXXX',
'XXXXXXXXXXX',
"XXX XXXXXX'X XXX XXXXXXXX",
'XXXX XXXXXX XXXXX XXXXXX',
'XXXX XXXXXX: XXXXXXXXX',
'XXXXXXXXX',
'XXXXX XXXXXX XXXXXXX XX',
'XXXXXX XXXXXX XX XX XXXXXXXX',
'XXXXXX: XXXXXXX XXXXXXXXX',
'XXXXXX',
'XXXXX XXXXXX XXX XXX XXXXXXX XX XXXXXXX',
'XXXXX XXXXXX XXX XXX XXXXXXX XX XXXXXXX',
'XXXXXXX',
'XXXXXXXXXXX 4',
'XXXXXXXX XXXX 6',
'XXXXXX XXX: XXXXXXXXXXX',
'XXXXX XXXXXXX XXXXXXXXXX',
'XXXXX: XXXXX XXX XXXXXXXXX',
'XXXX XX XXXX: XXXXX XX XXX',
'XXXXX XXX XXX 2',
'XXXXXXX XXXXXXXX',
'XX: XXX XXXXX XXXXXXXXXXX',
'XXXXXXXX XXXXXXXX',
'XXX XXXXX XXXX: XXXX XXX',
'XXX XXXXX',
'XXXXXX',
'XXXXX XXXXXX XXXXX',
"XXX XXXXXX'X XXXXXXX XXX: XXXXX",
'XXXXXXX',
'XXXXX XXXXX XXXX',
'XXXXX XXXXX XXXX: XXX XXXXXXX',
'XXXX XXXX: XXXX XXXXXX',
'XXX XXXXXX XXX 2',
'XXX XXX XXXXXXXX',
'XXXX XXXXXX 09',
'XXX XXXXX XX XX X XXXXXXXXXXX',
'XXXXXX XXX XXX XXX XXXXX XXXXX',
'XXXXXXXX',
'XXXXXX XXXX X: XXXXXXX 3',
'XXX XXXX 2',
"XXX XXXXXX'X XXXXX XXXXX",
'X-XXX XXXXXXX',
'XXXXXX: XXXXXX XXXXXXX',
'XXXXXX XXXX: XXXXX XXXX',
'XXXXXX XXXXXX XX',
'XXXXXX XXXXXXX 2',
'XXXXXX XX',
'XXXXXX XXXXXX XX XX XXXXXXXX',
'XXXXX XXXX',
'XXXXX XXXXX XXXXXXXX',
'XXXX XXXXXX (2013)',
'XXX XXXXXX XX XXXXX: XXXXXX XX XXXX',
'XXXX XXX XXXXX: XXXXX',
'XXXXXXXX',
'XXX XXXXXXX',
"XXX XXXX: XXXXX' XXXXX",
'XX XXXXXXXXX',
'XXXXXXX XXX XXXXXX',
'XXX XXXXX XXXXX!',
'XXXXX XXXXX 3',
'XXXX XXXX: XXX XXXXX XXXXXXXXX',
'XXXX XXXXX 2',
'XXXXXXX: XXX XXXXX XXXX',
'XXXX XXXX XXXX XXX: XXX XXXXX XXXX',
'XXXX XXXX XXXXX XXXXXX: XXXXX XXXXXXXX XX',
'XXX XXXXX 3',
'XXXX XXX XXXXX XXXXXX',
'XXXXXX XXX 09',
'XXXXX XXXXX',
'XXXXXXX & XXXXX XXXXXX: X XXXXX XX XXXX',
'XX-XX-XX! XXXXXXXXX XXXXXXXX',
'XXXXXX XXXXXX: XXXXXXXXX',
'XXX XXXXX XXXX 2',
'XXXXX XX XXXXX: XXXXXXXX XXXXXXX (XXX XXXXXX XXXXX)',
'XXX XXXXX',
'XXXX XXXXXX XXXXXXX: XXXX XX!',
'XXXX XXX XXXXXXXXXX',
'XXXX 2000',
'XXXXXXX XXXXXXXX 3',
'XXXXXX XXXX XXXXXXX XXXXXXX',
'XXX XXXXXX XX XXXXX: XXXXXX XX XXXXXXX',
'XXXXXX XXX 06',
'XXXXX 2',
'XXX XXXXXX XX XXXXXXX',
'XXXXX 2: XXXXX XX XXXX',
'XXXXXXXX',
'XXXXXXXXXXXXXXX 3',
'XXX XXXX 2002',
'XXXXXXX XXX XXXX',
'XXXXXX XXXXX',
'XXXXXXXXX 007 (2010)',
'XXXXX XXXXX: XXXXXX XXXX',
'XXX 3',
'XXXXXX XXXXXX XXXXXXX 2',
'XXXXXXX XXXXXX: XXXXX XX XXXXXXXX',
'XXXXXXX',
'XXXXXXXX 2',
'XXXX XXXXXXX XXXXXXX 2001 XXXXX',
"XXXXX'X XXXXX",
'XXXX XXXXXXXX 06',
'XXX XXX, XXXXXXX!',
'2 XXXXX XX 1: XXXXX XXXXXXX & XXXXXX XXXXXX!',
'XXXXX XX XXXXXXXX: XXXXX XX XXXXXXXX',
'XXX XXXX',
'XXXX XXXXX 3',
'XXXXXX XX XXXX',
'XXX XXXXXX XXX',
'XXX XXXXXX',
'XXXXX: XXXXXX XXXXXX',
'XXXXX & XXXXXXXX',
'XXXXX XXXXXX',
'XXX XXXX: XXX XXXX',
'XXXXXXX XXXXX 3: XXXXXXXXXX',
'XXXXXX XXXXX XX: XXXXXX XX XXXXXXXXXX',
'XXX XXXX XXXXXX',
'XXXXXX2: XXXXX XXXXX XXXXXXXXXX',
'XXXXX XXXX',
'XXXXXX XXXXXXX 4',
'XXX 2: XXXX XX XXXXXXX',
'XXXXXX XXXX: XXXXX XXXX',
'XXXXX XXXXXXX 2',
'XXXXXX XXX 07',
'XXX 2X11',
'XXX XXXXXXX 2000',
'XXXX XX XXXX: XXXXX XXX 3',
"XXXXXX'X XXX XXXX XXXX",
'XXXXXX: XXX XXXXX',
'XXXXXX XXXX XXXX',
'XXXXXXXX XXXX XX',
'XXXXXX XXXXX XXXXXXXX: XXXXX',
'XXXXXXXXX XXXXXXXXXXX: XXXXXXXXXXX',
'XXX XXXXXXX: XXXXX XXXXXX',
'XXXXXXXXX XXXXXX XXX XXX XXXX XX XXXXXXX',
'XXXX XXXXX 2',
"XXXXXX'X XXXXX",
'XXX XXXXXX XX XXXXX: XXX XXXX XXXXX',
'50 XXXX: XXXXXXXXXXX',
'XXXX XXXXXX XXXXXXX: XXXX XX!',
'XXX XXXXX X',
'XXXXXX XXX 25',
'XXXXX XXXXXXX XX',
'XXXXX 64: XXX XXXXXXX XXXXXX',
'XXXX XX XXXXX 3',
'XXX 2009 XXXXXXXXXX',
'XXXXXXX XX: XXXXXX XX XXXXX',
'XXX XXXXXXXX',
'XXX XXXXXXXXX XXXXXXXXXXX XXXXX',
'XXXXXX XXX',
'XXXXX XXX XXXXXXXX 3',
'XXX XXXXXX',
'XXX XXX: XXXXX XXX XX',
"XXX XXXXXX'X XXXXX XXXXX",
'XXXXX XX XXXXXXXXX',
'XXXXXXXX XXXXXXXX',
'XXXXXX: XXXXXXXXXX',
'2 XXXXX XX 1 XXXXXX XXXX: XXX XXXXXXXXXXX / XXXXXXX XXXX: XXX XXXXXXXXXX XXXXXXXXX
X',
'XXXX XX XXXX: XXXXX XXX 3',
'XXX XXXX: XXXX XX XXX XXXX',
'XXXX XXXXXX 08',
'XXXX XXXX: XXXXXXXXXXX',
'XXXXXX XXX 2000',
'XXXXXXX XXXXXX',
'XXXX XXXX: XXXXXXXXXXX',
'XXX XXXXXXXXX XXXXXX',
'XXXXXXXXX',
'XXXXX',
"XXXXXX'X XXXXXXX",
...]
```

In [110]:

```python
# Create a list of lists containing lowercase words for each tweet
words_in_remove_periods = [name.lower().split() for name in remove_periods]
words_in_remove_periods[:3]
```

Out[110]:

```
[['wii', 'sports'], ['super', 'mario', 'bros'], ['mario', 'kart', 'wii']]
```

In [111]:

```python
import itertools
import collections

clean_list = list(itertools.chain(remove_periods))

#Create counter
counts = collections.Counter(clean_list)

counts.most_common(15)
```

Out[111]:

```
[('need for speed: most wanted', 12),
 ('fifa 14', 9),
 ('lego marvel super heroes', 9),
 ('ratatouille', 9),
 ('fifa soccer 13', 8),
 ('fifa 15', 8),
 ('madden nfl 07', 8),
 ('lego star wars ii: the original trilogy', 8),
 ('madden nfl 08', 8),
 ('monopoly', 8),
 ('the lego movie videogame', 8),
 ('cars', 8),
 ('lego jurassic world', 8),
 ('lego batman 3: beyond gotham', 8),
 ('terraria', 8)]
```

In [112]:

```python
clean_remove_periods = pd.DataFrame(counts.most_common(15),
                                    columns=['words', 'count'])

clean_remove_periods.head()
```

Out[112]:

	words	count
0	need for speed: most wanted	12
1	fifa 14	9
2	lego marvel super heroes	9
3	ratatouille	9
4	fifa soccer 13	8

In [113]:

```python
text_data = df['name']
```

In [114]:

```python
#names_of_video_games = [line.replace("\n", "") for line in open("text_data")]
#print(names_of_video_games[:5])
```

In [115]:

```python
df["name"]= df["name"].str.split("t", n = 1, expand = True)
```

In [116]:

```python
f = lambda x: len(x["name"].split("error")) -1
df["error"] = df.apply(f, axis=1)
```

In [117]:

```python
#data_set = df["name"]

#split() returns list of all the words in the string
#split_it = data_set.split()

#Pass the split_it list to instance of Counter class.
#Counter = Counter(split_it)

#most_common() produces k frequently encountered
#input values and their respective counts.
#most_occur = Counter.most_common(4)

#print(most_occur)
```

In [118]:

```python
# Split the words from one tweet into unique elements
lower_case[0].split()
```

Out[118]:

```
['wii', 'sports']
```

In [119]:

```python
# Split the words from one tweet into unique elements
lower_case[0].lower().split()
```

Out[119]:

```
['wii', 'sports']
```

In [120]:

```python
n =  df['name']
print(n)
```

```
0                              Wii Spor
1                      Super Mario Bros.
2                              Mario Kar
3                               Wii Spor
4                 Pokemon Red/Pokemon Blue
                          ...
16593    Woody Woodpecker in Crazy Cas
16594    Men in Black II: Alien Escape
16595                           SCORE In
16596                        Know How 2
16597                             Spiri
Name: name, Length: 16291, dtype: object
```

In [121]:

```python
stacked = pd.DataFrame(n.str.split().tolist()).stack()
print (stacked)
```

```
0      0      Wii
       1     Spor
1      0    Super
       1    Mario
       2    Bros.
             ...
16288  1       In
16289  0     Know
       1      How
       2        2
16290  0    Spiri
Length: 42012, dtype: object
```

In [122]:

```python
#print (stacked).value_counts()
```

In [123]:

```python
#stacked = list(itertools.chain(*lower_case))

# Create counter
#counts_stacked = collections.Counter(lower_case)

#lower_case.most_common(15)
```

Q3. Data Preparation: Encoding Techniques & PCA Implementation

In [124]:

```python
import category_encoders as ce
import pandas as pd
train_df=pd.DataFrame({'Genre':['Action', 'Adventure', 'Fighting', 'Misc', 'Platform',

# create object of Ordinalencoding
encoder= ce.OrdinalEncoder(cols=['Genre'],return_df=True,
                           mapping=[{'col':'Genre',
'mapping':{'None':0,'Action':1, 'Adventure':2, 'Fighting':3, 'Misc':4, 'Platform':5,

#Original data
train_df
```

Out[124]:

	Genre
0	Action
1	Adventure
2	Fighting
3	Misc
4	Platform
5	Puzzle
6	Racing
7	Role-Playing
8	Shooter
9	Simulation
10	Sports
11	Strategy

In [125]:

```python
import category_encoders as ce
```

In [126]:

```python
#fit and transform train data
df_train_transformed = encoder.fit_transform(train_df)
df_train_transformed
```

Out[126]:

	Genre
0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9
9	10
10	11
11	12

One Hot Encoding

The idea behind this approach is to create a new dummy feature for each unique value in the nominal feature column (Raschka, 2015).

In [128]:

```python
ohe_df=pd.DataFrame({'Genre':['Action', 'Adventure', 'Fighting', 'Misc', 'Platform',

#Create object for one-hot encoding
encoder=ce.OneHotEncoder(cols='Genre',handle_unknown='return_nan',return_df=True,use_c

#Original Data
ohe_df
```

Out[128]:

	Genre
0	Action
1	Adventure
2	Fighting
3	Misc
4	Platform
5	Puzzle
6	Racing
7	Role-Playing
8	Shooter
9	Simulation
10	Sports
11	Strategy

In [132]:

```python
#Fit and transform Data
data_encoded = encoder.fit_transform(ohe_df)
data_encoded
```

Out[132]:

	Genre_Action	Genre_Adventure	Genre_Fighting	Genre_Misc	Genre_Platform	Genre_Puzzle	Genre_Racing
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	1.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	1.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	1.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0

In [133]:

```python
# Dummy Encoding
```

In [134]:

```python
dummy_data = pd.DataFrame({'Genre':['Action', 'Adventure', 'Fighting', 'Misc', 'Platfo

#Create object for one-hot encoding
encoder=ce.OneHotEncoder(cols='Genre',handle_unknown='return_nan',return_df=True,use_c

#Original Data
dummy_data
```

Out[134]:

	Genre
0	Action
1	Adventure
2	Fighting
3	Misc
4	Platform
5	Puzzle
6	Racing
7	Role-Playing
8	Shooter
9	Simulation
10	Sports
11	Strategy

In [135]:

```python
#encode the data
data_encoded=pd.get_dummies(data=dummy_data,drop_first=True)
data_encoded
```

Out[135]:

	Genre_Action	Genre_Adventure	Genre_Fighting	Genre_Misc	Genre_Platform	Genre_Puzzle	Genre_Racing
0	1	0	0	0	0	0	0
1	0	1	0	0	0	0	0
2	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0
5	0	0	0	0	0	1	0
6	0	0	0	0	0	0	1
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0

Principle Component Analysis (PCA)

Principal Component Analysis is a procedure used to uncover the dimensions that explain the majority of the variance in a dataset. PCA produces the principal components by computing the eigenvalues (special set of scalars of the covariance matrix of a dataset (Mayor, 2015).

In [136]:

```python
%matplotlib inline
import numpy as np
from sklearn.decomposition import PCA
import seaborn as sns;
import matplotlib.pyplot as plt
sns.set()
```

In [137]:

```python
from sklearn.decomposition import PCA
```

In [138]:

```python
df.head()
```

Out[138]:

	rank	name	platform	year	genre	publisher	north_american_sales	european_sales	japanese_sale
0	1	Wii Spor	Wii	2006	Sports	Nintendo	41.49	29.02	3.7
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.8
2	3	Mario Kar	Wii	2008	Racing	Nintendo	15.85	12.88	3.7
3	4	Wii Spor	Wii	2009	Sports	Nintendo	15.75	11.01	3.2
4	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.2

In [139]:

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 16291 entries, 0 to 16597
Data columns (total 12 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  int32
 4   genre                 16291 non-null  object
 5   publisher             16291 non-null  object
 6   north_american_sales  16291 non-null  float64
 7   european_sales        16291 non-null  float64
 8   japanese_sales        16291 non-null  float64
 9   other_sales           16291 non-null  float64
 10  global_sales          16291 non-null  float64
 11  error                 16291 non-null  int64
dtypes: float64(5), int32(1), int64(2), object(4)
memory usage: 2.2+ MB
```

In [140]:

```python
video_game = df.drop(["name", "platform", "year", "genre", "publisher", "global_sales'
pca = PCA().fit(video_game)
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance');
```


The above visual representation unfortunately appears incorrect and as such the outcome can not be confirmed.

References

- Bevans, R., 2020. Scibbr. [Online] Available at: https://www.scribbr.com/statistics/simple-linear-regression/ [Accessed 30th May 2021].
- Bonaccorso, G., 2017. Machine Learning Algorithms. Birmingham: Packt Publishing.
- Cuesta, H., 2013. Practical Data Analysis. Birmingham: Packt Publishing.
- Dave, A., 2018. Data Driven Investor. [Online] Available at: https://medium.datadriveninvestor.com/regression-in-machine-learning-296caae933ec [Accessed 30th May 2021].
- Gollapudi, S. & Laxmikanth, V., 2016. Practical Machine Learning. Birmingham: Packt Publishing.
- Hayes, A., 2021. Investopedia. [Online] Available at: https://www.investopedia.com/terms/z/zscore.asp#:~:text=A Z-score is a,identical to the mean score. [Accessed 30th May 2021].
- Lesmeister, C., 2015. Mastering Machine Learning with R.. Birmingham: Packt.
- Mayor, E., 2015. Learning Predictive Analytics with R.Vol.1. Birmingham: Packt Publishing.
- OCTOPAI, 2019. OCTOPAI. [Online] Available at: https://www.octopai.com/automatic-data-dictionary-mapping-using-machine-learning/ [Accessed 30th May 2021].
- Raschka, S., 2015. Python Machine Learning. Birmingham: Packt Publishing.
- Vallat, R., 2019. raphaelvallat.com. [Online] Available at: https://raphaelvallat.com/correlation.html [Accessed 25th May 2021].

In [ ]: