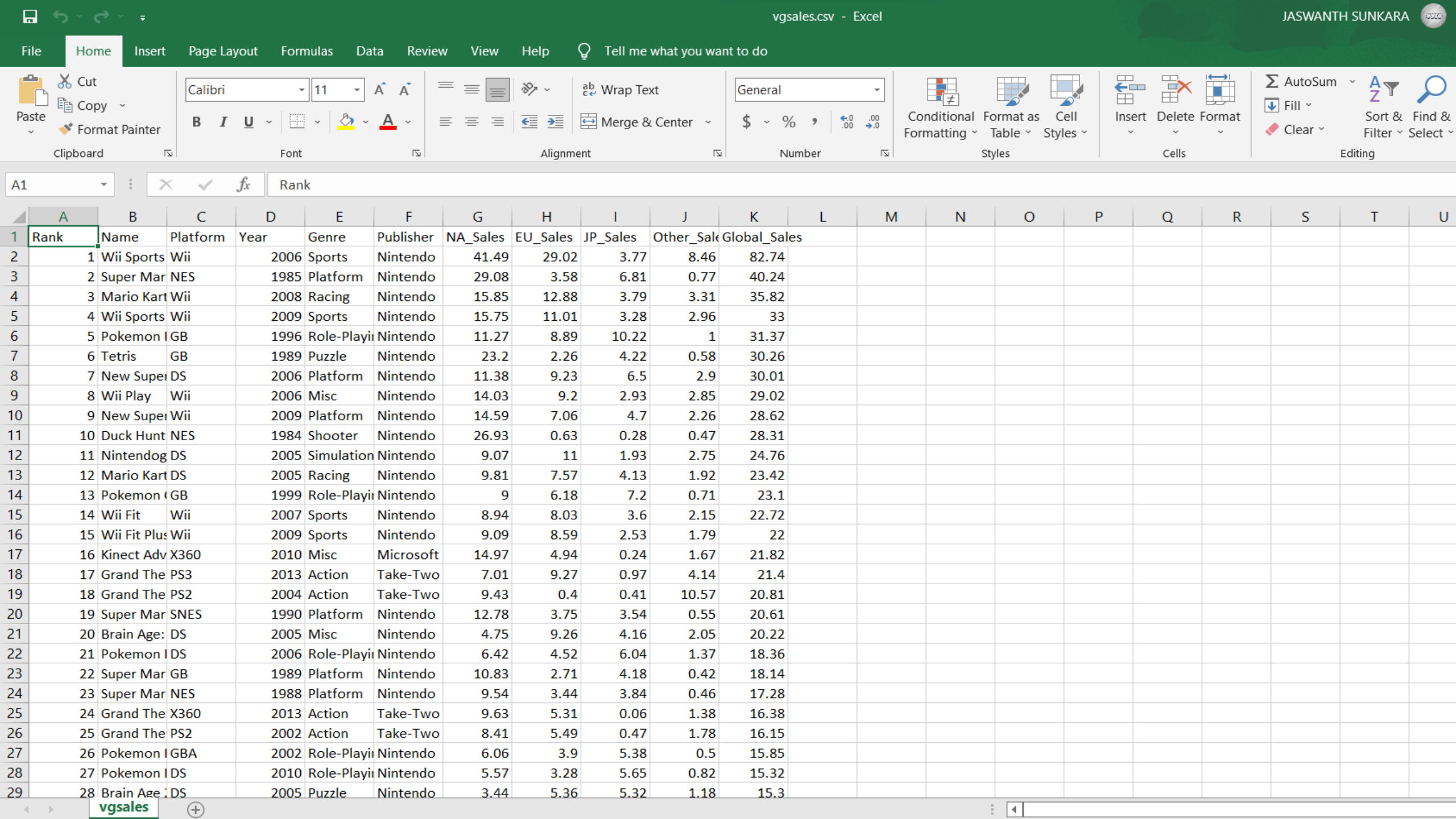
DATA ANALYSIS ON VEDIOGAME SALES

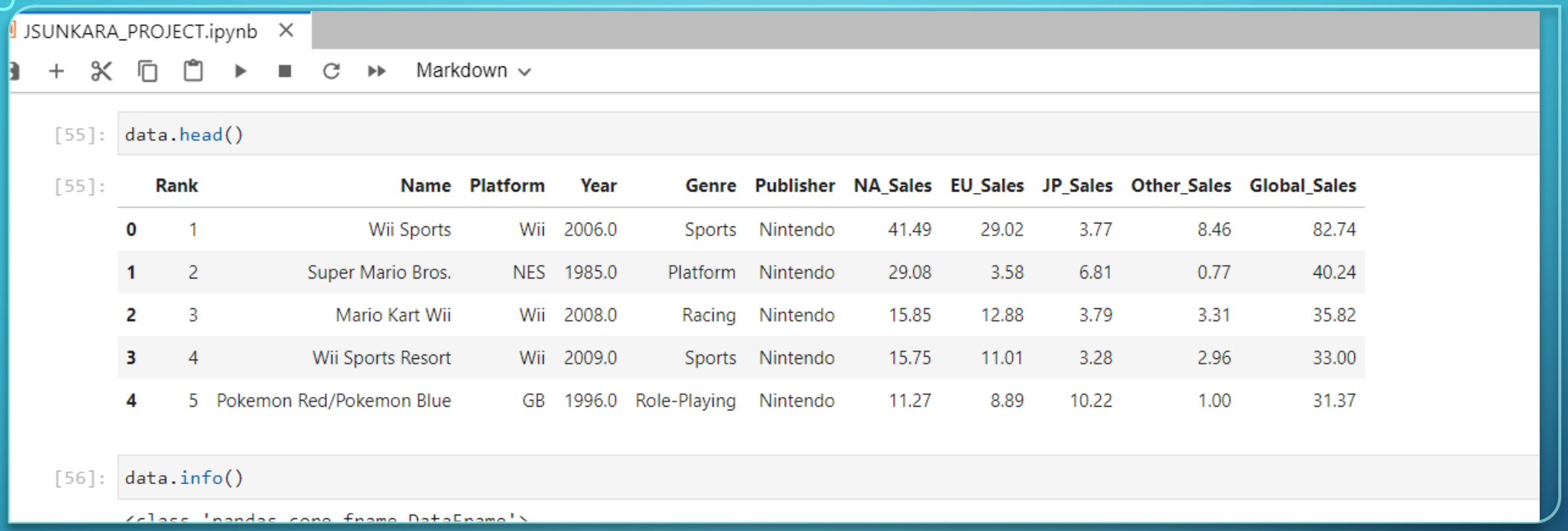
JASWANTH SUNKARA

STEPS INVOLVED

- 1.DATA CLEANING AND PROCESSING
- 2.DATA VISUALIZATION
- 3.TRAINING A MODEL



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sale	Global_Sales										
2	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74										
3	2	Super Mar	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24										
4	3	Mario Kart	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82										
5	4	Wii Sports	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33										
6	5	Pokemon I	GB	1996	Role-Playi	Nintendo	11.27	8.89	10.22	1	31.37										
7	6	Tetris	GB	1989	Puzzle	Nintendo	23.2	2.26	4.22	0.58	30.26										
8	7	New Super	DS	2006	Platform	Nintendo	11.38	9.23	6.5	2.9	30.01										
9	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.2	2.93	2.85	29.02										
10	9	New Super	Wii	2009	Platform	Nintendo	14.59	7.06	4.7	2.26	28.62										
11	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31										
12	11	Nintendog	DS	2005	Simulation	Nintendo	9.07	11	1.93	2.75	24.76										
13	12	Mario Kart	DS	2005	Racing	Nintendo	9.81	7.57	4.13	1.92	23.42										
14	13	Pokemon (	GB	1999	Role-Playi	Nintendo	9	6.18	7.2	0.71	23.1										
15	14	Wii Fit	Wii	2007	Sports	Nintendo	8.94	8.03	3.6	2.15	22.72										
16	15	Wii Fit Plus	Wii	2009	Sports	Nintendo	9.09	8.59	2.53	1.79	22										
17	16	Kinect Adv	X360	2010	Misc	Microsoft	14.97	4.94	0.24	1.67	21.82										
18	17	Grand The	PS3	2013	Action	Take-Two	7.01	9.27	0.97	4.14	21.4										
19	18	Grand The	PS2	2004	Action	Take-Two	9.43	0.4	0.41	10.57	20.81										
20	19	Super Mar	SNES	1990	Platform	Nintendo	12.78	3.75	3.54	0.55	20.61										
21	20	Brain Age:	DS	2005	Misc	Nintendo	4.75	9.26	4.16	2.05	20.22										
22	21	Pokemon I	DS	2006	Role-Playi	Nintendo	6.42	4.52	6.04	1.37	18.36										
23	22	Super Mar	GB	1989	Platform	Nintendo	10.83	2.71	4.18	0.42	18.14										
24	23	Super Mar	NES	1988	Platform	Nintendo	9.54	3.44	3.84	0.46	17.28										
25	24	Grand The	X360	2013	Action	Take-Two	9.63	5.31	0.06	1.38	16.38										
26	25	Grand The	PS2	2002	Action	Take-Two	8.41	5.49	0.47	1.78	16.15										
27	26	Pokemon I	GBA	2002	Role-Playi	Nintendo	6.06	3.9	5.38	0.5	15.85										
28	27	Pokemon I	DS	2010	Role-Playi	Nintendo	5.57	3.28	5.65	0.82	15.32										
29	28	Brain Age	DS	2005	Puzzle	Nintendo	3.44	5.36	5.32	1.18	15.3										



	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006.0	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

WE USE PANDAS TO READ THE CSV FILE TO DATA AND PERFORM VARIOUS OPERATIONS ON THE DATA

The initial commands are data.head(), data.info(), data.tail(), etc.. To represent the data

Dropping null values

```
[6]: data=data.dropna()
     data.isnull().sum()
```

DROPPING THE NULL VALUES FROM DATA TO AVOID CONFLICTS AND CLEAN USELESS DATA

Dropna() drops all the data values which are not available i.e NULL

Correlation (Pearson's)

```
[10]: data.corr()
```

[10]:

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
Rank	1.000000	0.178027	-0.400315	-0.379137	-0.269323	-0.332735	-0.426975
Year	0.178027	1.000000	-0.091285	0.006108	-0.169387	0.041128	-0.074647
NA_Sales	-0.400315	-0.091285	1.000000	0.768923	0.451283	0.634518	0.941269
EU_Sales	-0.379137	0.006108	0.768923	1.000000	0.436379	0.726256	0.903264

TENDENCY AND DISPERSION TO FIND MEAN, STD.DEVIATION AND OTHER FACTORS REQUIRED FOR ANALYSING DATA

VISUALIZATIONS

[11]:
```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plot
%matplotlib inline
```

COUNTPLOT

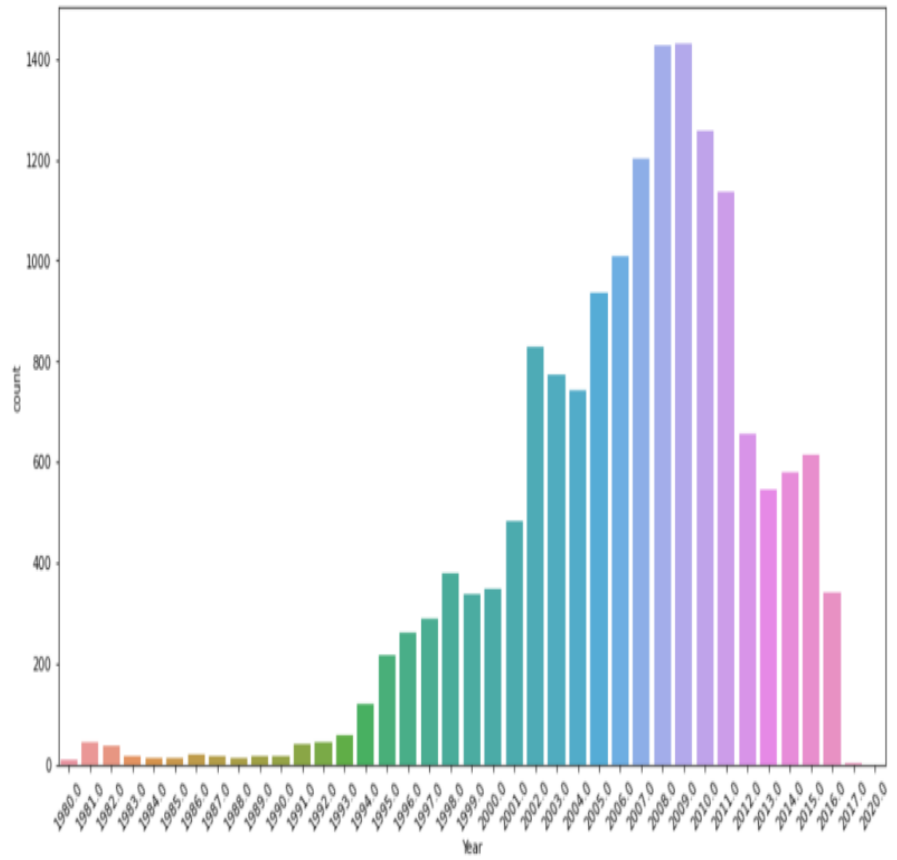
[12]:
```
plot.figure(figsize=(15,8))
sns.countplot(data = data, x='Platform')
```

[12]:
```
<AxesSubplot:xlabel='Platform', ylabel='count'>
```

WE ARE USING SEABORN AND MATPLOTLIB LIBRARIES FOR PERFORMING VISUALIZATION METHODS ON DATA

These include histograms, barplots, graphs, heatmaps, piecharts, etc. for visualizing data

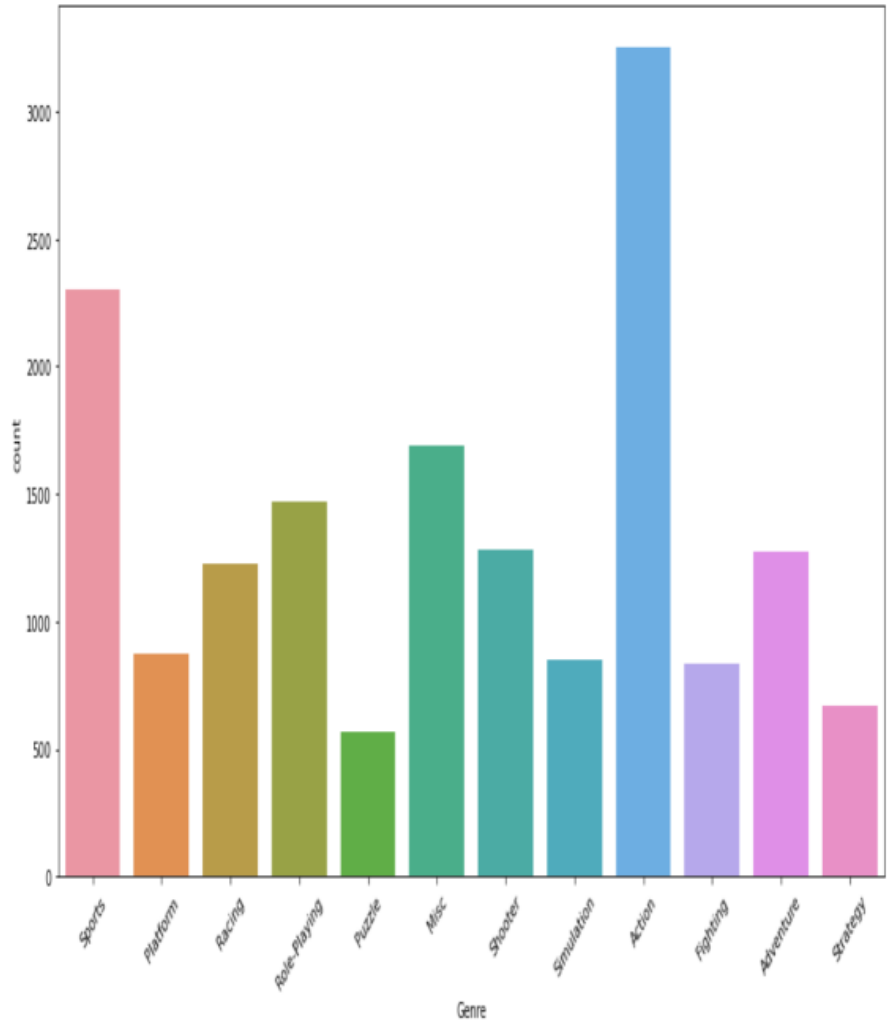
[13]:
```
plot.figure(figsize=(15,8))
sns.countplot(data=data, x='Year')
plot.xticks(rotation =45);
```



The above plot shows the number of games released in different years i.e 1980-2020

The years 2006 and 2007 has the most number of releases among all

```
sns.countplot(data=data, x='Genre')
plot.xticks(rotation =45);
```



The plot shows different genres and their game count

The action genre have highest count compared to other genres

GRAPHPLOTS

[24]:
```
ax=plot.figure(figsize=(15,6))
data.groupby(['Year'])['Global_Sales'].sum().plot()
plot.grid()
plot.ylabel('Global Sales')
plot.title('Global Sales over the years')
```

[24]:
```
Text(0.5, 1.0, 'Global Sales over the years')
```

TRAINING MODEL

[35]:
```
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import r2_score
from sklearn.model_selection import cross_val_score

from sklearn.tree import DecisionTreeRegressor
from sklearn.linear_model import LinearRegression
```

[26]:
```
from sklearn.preprocessing import LabelEncoder

data1 = data.copy()

le = LabelEncoder()
```

USING SKLEARN WE WILL TRAIN A MODEL USING LINEAR REGRESSION AND DECISIONTREE

LINEAR REGRESSION WORKS BETTER THAN DECISIONTREE FOR PREDICTING RESULTS

]:

	Actual	Predicted
0	0.57	0.570374
1	0.07	0.060217
2	0.19	0.190317
3	0.67	0.670245
4	0.22	0.220396
...	...	...
4068	0.05	0.040179
4069	0.06	0.050394
4070	2.20	2.189877
4071	0.06	0.060322

CONCLUSION

- Most Games produced in Specific Gaming Platform is DS then PS2
- The most popular type of game is Action then Sports. Lowest is Puzzle
- Top Publisher is Electronic Arts
- Most games Produced in between 2002 to 2016 specially in 2010
- Most published games are Action Genre
- Most of sales are between 2005 and 2010 and less sales in 1980 to 1990