Exploring Video Games Data

Libraries

In [57]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

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

In [58]:
```
df = pd.read_csv('datasets/vgsales.csv')
df.head()
```

Out[58]:

	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

In [59]:
```
df.info()
```

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

In [60]:
```
df.isnull().any()
```

Out[60]:
```
Rank            False
Name            False
Platform        False
Year             True
Genre           False
Publisher        True
NA_Sales        False
EU_Sales        False
JP_Sales        False
Other_Sales     False
Global_Sales    False
dtype: bool
```

Checking Null Values

In [61]:
```
df.isnull().sum()
```

Out[61]:
```
Rank            0
Name            0
Platform        0
Year          271
Genre           0
Publisher      58
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

Droping Null Values

In [62]:
```
df = df.dropna()
df.isnull().sum()
```

Out[62]:
```
Rank            0
Name            0
Platform        0
Year            0
Genre           0
Publisher       0
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

Transforming Data Types

In [63]:
```
df['Year'] = df['Year'].astype('int64')
```

In [64]:
```
df.head()
```

Out[64]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37

Statistical Analysis

In [65]:
```
df.describe()
```

Out[65]:

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
count	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000	16291.000000
mean	8290.190228	2006.405561	0.265647	0.147731	0.078833	0.048426	0.540910
std	4792.654450	5.832412	0.822432	0.509303	0.311879	0.190083	1.567345
min	1.000000	1980.000000	0.000000	0.000000	0.000000	0.000000	0.010000
25%	4132.500000	2003.000000	0.000000	0.000000	0.000000	0.000000	0.060000
50%	8292.000000	2007.000000	0.080000	0.020000	0.000000	0.010000	0.170000
75%	12439.500000	2010.000000	0.240000	0.110000	0.040000	0.040000	0.480000
max	16600.000000	2020.000000	41.490000	29.020000	10.220000	10.570000	82.740000

In [66]:
```
print("*"*60)
for i, col in enumerate(list(df.columns[:-1])):
    print(col)
    print(df[col].value_counts())
    print("*"*60)
```

```
************************************************************
Rank
2047     1
6790     1
12963    1
15010    1
8865     1
        ..
9534     1
15677    1
13628    1
3387     1
2049     1
Name: Rank, Length: 16291, dtype: int64
************************************************************
Name
Need for Speed: Most Wanted                    12
FIFA 14                                         9
Ratatouille                                     9
```

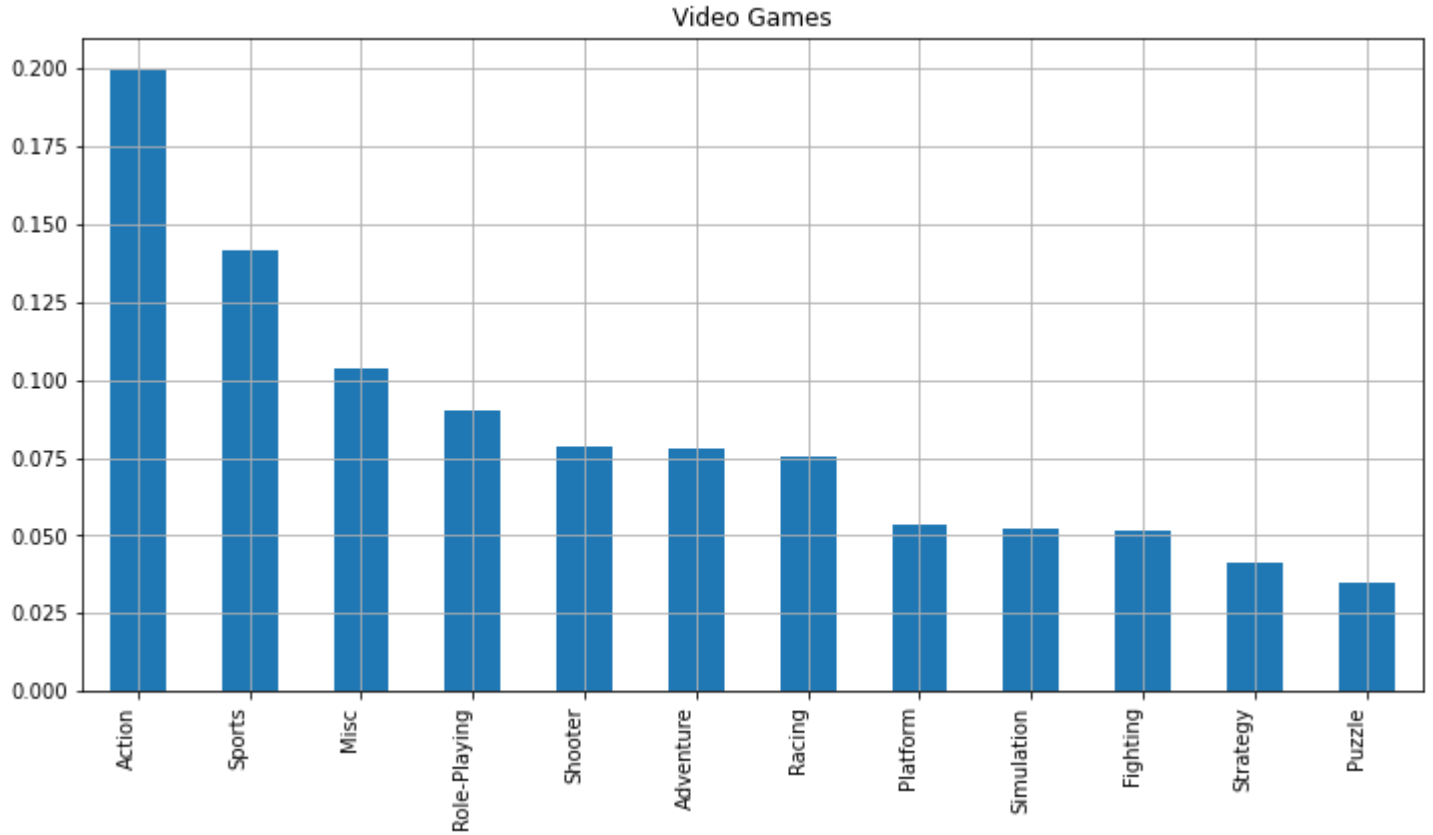
In [67]:
```
df['Genre'].value_counts()
```

Out[67]:
```
Action          3251
Sports          2304
Misc            1686
Role-Playing    1470
Shooter         1282
Adventure       1274
Racing          1225
Platform         875
Simulation       848
Fighting         836
Strategy         670
Puzzle           570
Name: Genre, dtype: int64
```

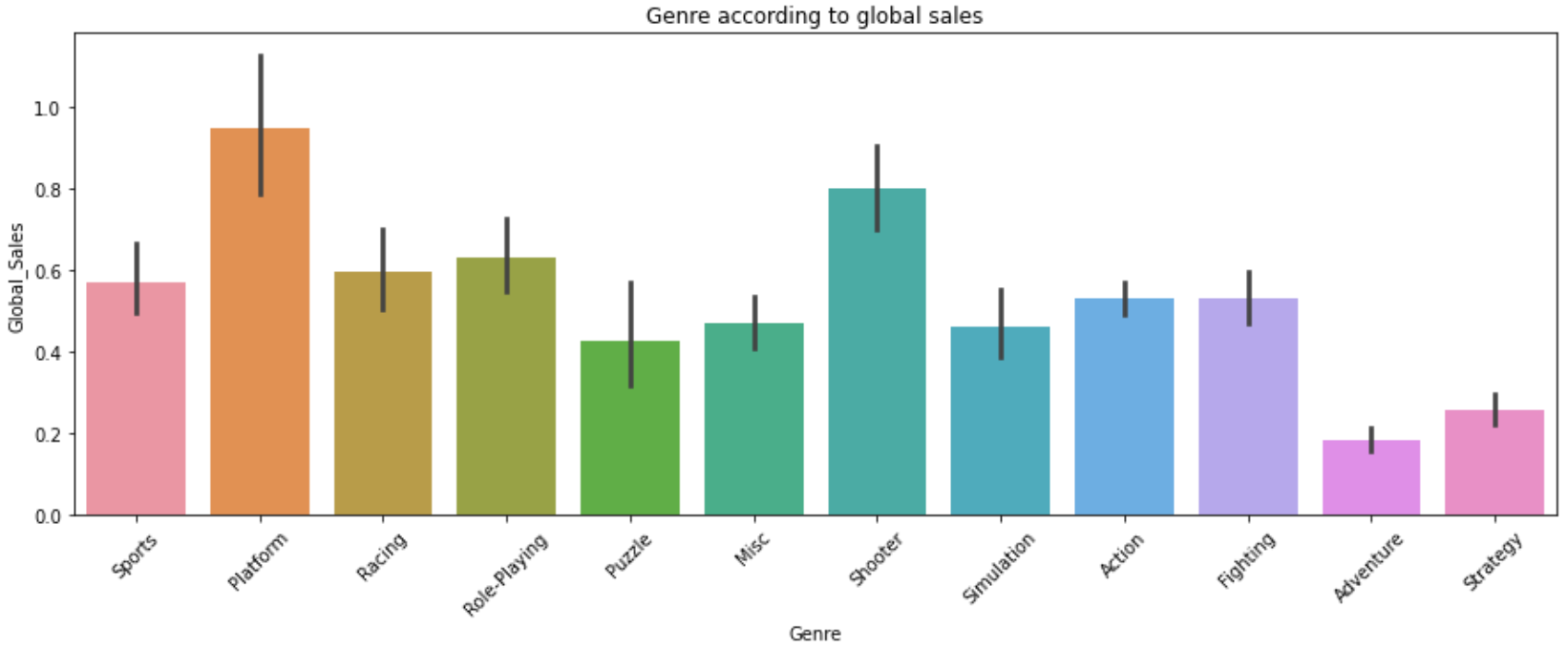
In [ ]:
```
### <font color="sky-blue">   </font>
```

In [68]:
```
games = df['Genre'].value_counts(normalize=True)
```

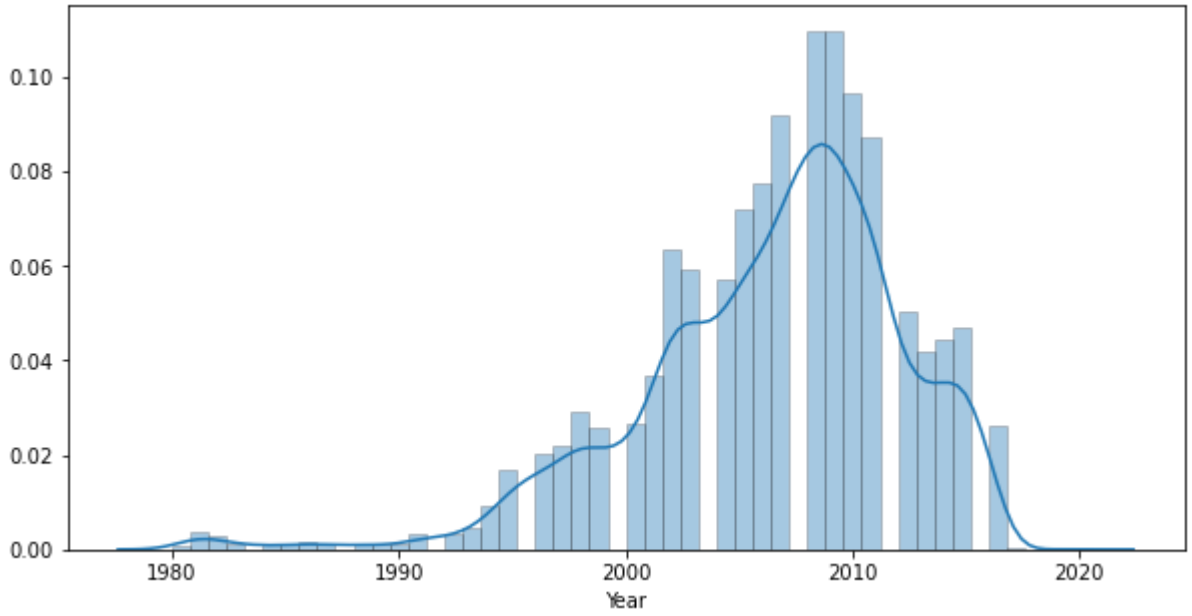
In [69]:
```
plt.figure(figsize=(12,6))
games.plot(kind='bar', grid=True)
plt.xticks(horizontalalignment='right')
plt.title('Video Games')
plt.show();
```



In [70]:
```
plt.figure(figsize=(15,5))
plt.title('Genre according to global sales')
sns.barplot(x="Genre", y="Global_Sales", data=df)
plt.xticks(
rotation=45)
plt.show();
```



In [81]:
```
plt.figure(figsize=(10,5))
sns.distplot(df.Year, kde=True, hist_kws=dict(edgecolor='k', linewidth=0.5))
plt.show();
```



Assembled & Compiled by : Hafiz Muhammad Waqas