In [97]:

```
#EDS Mini Project
#Batch D3
#Members
#448-Kaustubh Patil (202201090057)
#453-Prathamesh Tiwari (202201050016)
#458-Yash Sawant (202201070032)
```

In [148]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
print(data)
```

```
                            Name Platform  Year_of_Release         Genre  \
0                     Wii Sports      Wii           2006.0        Sports
1                Super Mario Bros.      NES           1985.0      Platform
2                   Mario Kart Wii      Wii           2008.0        Racing
3                Wii Sports Resort      Wii           2009.0        Sports
4          Pokemon Red/Pokemon Blue       GB           1996.0  Role-Playing
...                             ...      ...              ...           ...
16714  Samurai Warriors: Sanada Maru      PS3           2016.0        Action
16715               LMA Manager 2007     X360           2006.0        Sports
16716        Haitaka no Psychedelica      PSV           2016.0     Adventure
16717               Spirits & Spells      GBA           2003.0      Platform
16718            Winning Post 8 2016      PSV           2016.0    Simulation

          Publisher  NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales  \
0          Nintendo     41.36     28.96      3.77         8.45         82.53
1          Nintendo     29.08      3.58      6.81         0.77         40.24
2          Nintendo     15.68     12.76      3.79         3.29         35.52
3          Nintendo     15.61     10.93      3.28         2.95         32.77
4          Nintendo     11.27      8.89     10.22         1.00         31.37
...             ...       ...       ...       ...          ...           ...
16714    Tecmo Koei      0.00      0.00      0.01         0.00          0.01
16715   Codemasters      0.00      0.01      0.00         0.00          0.01
16716  Idea Factory      0.00      0.00      0.01         0.00          0.01
16717       Wanadoo      0.01      0.00      0.00         0.00          0.01
16718    Tecmo Koei      0.00      0.00      0.01         0.00          0.01

       Critic_Score  Critic_Count User_Score  User_Count Developer Rating
0              76.0          51.0          8       322.0  Nintendo      E
1               NaN           NaN        NaN         NaN       NaN    NaN
2              82.0          73.0        8.3       709.0  Nintendo      E
3              80.0          73.0          8       192.0  Nintendo      E
4               NaN           NaN        NaN         NaN       NaN    NaN
...             ...           ...        ...         ...       ...    ...
16714           NaN           NaN        NaN         NaN       NaN    NaN
16715           NaN           NaN        NaN         NaN       NaN    NaN
16716           NaN           NaN        NaN         NaN       NaN    NaN
16717           NaN           NaN        NaN         NaN       NaN    NaN
16718           NaN           NaN        NaN         NaN       NaN    NaN

[16719 rows x 16 columns]
```

In [149]:

```
#1
mean_global_sales = np.mean(df['Global_Sales'])
print(mean_global_sales)
```

```
0.7775897435897436
```

In [150]:

```
#2
median_user_score = np.median(df['Critic_Score'])
print(median_user_score)
```

```
72.0
```

In [151]:

```
#3
max_critic_score = np.max(df['Critic_Score'])
print(max_critic_score)
```

```
98.0
```

In [152]:

```
#4
min_year_of_release = np.min(df['Year_of_Release'])
print(min_year_of_release)
```

```
1985.0
```

In [153]:

```
#5
td_global_sales = np.std(df['Global_Sales'])
print(td_global_sales)
```

```
1.9632989450006317
```

In [154]:

```
#6
released_in_21st_century = np.all(df['Year_of_Release'] >= 2000)
print(released_in_21st_century)
```

```
False
```

In [155]:

```
#7
released_on_wii = np.any(df['Platform'] == 'Wii')
print(released_on_wii)
```

```
True
```

In [156]:

```
#8
random_games = np.random.choice(df['Name'], 10)
print(random_games)
```

```
['Enemy Territory: Quake Wars' "Baldur's Gate: Dark Alliance"
 'MySims Agents' 'Call of Duty: Black Ops II' 'OutRun 2006: Coast 2 Coast'
 'Hyperdimension Neptunia' 'WarioWare: Smooth Moves'
 'Grand Theft Auto III' 'Left 4 Dead 2'
 'Plants vs. Zombies: Garden Warfare 2']
```

In [157]:

```
#9
unique_platforms = np.unique(df['Platform'])
print(unique_platforms)
```

```
['3DS' 'DC' 'DS' 'GBA' 'GC' 'PC' 'PS' 'PS2' 'PS3' 'PS4' 'PSP' 'PSV' 'Wii'
 'WiiU' 'X360' 'XB' 'XOne']
```

In [158]:

```
#10
percentile_user_count = np.percentile(df['User_Count'], 75)
print(percentile_user_count)
```

```
89.0
```

In [159]:

```
data=pd.read_csv("Video_Games_Sales.csv")
dtf=pd.DataFrame(data)
```

In [160]:

```
#Panda
```

In [161]:

```
#1
print("The index is:")
print(dtf.index)
```

```
The index is:
RangeIndex(start=0, stop=16719, step=1)
```

In [162]:

```
#2
print("The columns are:")
print(dtf.columns)
```

```
The columns are:
Index(['Name', 'Platform', 'Year_of_Release', 'Genre', 'Publisher', 'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales', 'Critic_Score',
       'Critic_Count', 'User_Score', 'User_Count', 'Developer', 'Rating'],
      dtype='object')
```

In [163]:

```
#3
print("The axes are:")
print(dtf.axes)
```

```
The axes are:
[RangeIndex(start=0, stop=16719, step=1), Index(['Name', 'Platform', 'Year_of_Release', 'Genre', 'Pub
lisher', 'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales', 'Critic_Score',
       'Critic_Count', 'User_Score', 'User_Count', 'Developer', 'Rating'],
      dtype='object')]
```

In [164]:

```
#4
print("The Datatypes are:")
print(dtf.dtypes)
```

```
The Datatypes are:
Name              object
Platform          object
Year_of_Release  float64
Genre             object
Publisher         object
NA_Sales         float64
EU_Sales         float64
JP_Sales         float64
Other_Sales      float64
Global_Sales     float64
Critic_Score     float64
Critic_Count     float64
User_Score        object
User_Count       float64
Developer         object
Rating            object
dtype: object
```

In [165]:

```
#5
print("The size is:")
print(dtf.size)
```

```
The size is:
267504
```

In [166]:

```
#6
print("The shape is:")
print(dtf.shape)
```

```
The shape is:
(16719, 16)
```

In [167]:

```
#7
print("The Number of dimensions is")
print(dtf.ndim)
```

```
The Number of dimensions is
2
```

In [168]:

```
#8
print("The total number of entries in each column is")
print(dtf.count())
```

```
The total number of entries in each column is
Name               16717
Platform           16719
Year_of_Release    16450
Genre              16717
Publisher          16665
NA_Sales           16719
EU_Sales           16719
JP_Sales           16719
Other_Sales        16719
Global_Sales       16719
Critic_Score        8137
Critic_Count        8137
User_Score         10015
User_Count          7590
Developer          10096
Rating              9950
dtype: int64
```

In [169]:

```
df.head()
#9
dtf = df.dropna()
```

In [170]:

```
#10
dtf = df.drop_duplicates()
dtf
```

Out[170]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
6	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80
7	Wii Play	Wii	2006.0	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92
...	...	...	...	...	...	...	...	...	...	...
16667	E.T. The Extra-Terrestrial	GBA	2001.0	Action	NewKidCo	0.01	0.00	0.00	0.00	0.01
16677	Mortal Kombat: Deadly Alliance	GBA	2002.0	Fighting	Midway Games	0.01	0.00	0.00	0.00	0.01
16696	Metal Gear Solid V: Ground Zeroes	PC	2014.0	Action	Konami Digital Entertainment	0.00	0.01	0.00	0.00	0.01
16700	Breach	PC	2011.0	Shooter	Destineer	0.01	0.00	0.00	0.00	0.01
16706	STORM: Frontline Nation	PC	2011.0	Strategy	Unknown	0.00	0.01	0.00	0.00	0.01

6825 rows × 16 columns

In [171]:

```
#11
df = df.rename(columns={'User_Score': 'User_Rating'})
dtf
```

Out[171]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
6	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80
7	Wii Play	Wii	2006.0	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92
...	...	...	...	...	...	...	...	...	...	...
16667	E.T. The Extra-Terrestrial	GBA	2001.0	Action	NewKidCo	0.01	0.00	0.00	0.00	0.01
16677	Mortal Kombat: Deadly Alliance	GBA	2002.0	Fighting	Midway Games	0.01	0.00	0.00	0.00	0.01
16696	Metal Gear Solid V: Ground Zeroes	PC	2014.0	Action	Konami Digital Entertainment	0.00	0.01	0.00	0.00	0.01
16700	Breach	PC	2011.0	Shooter	Destineer	0.01	0.00	0.00	0.00	0.01
16706	STORM: Frontline Nation	PC	2011.0	Strategy	Unknown	0.00	0.01	0.00	0.00	0.01

6825 rows × 16 columns

In [172]:

```
#12
df_sorted = df.sort_values('Global_Sales', ascending=False)
dtf
```

Out[172]:

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
6	New Super Mario Bros.	DS	2006.0	Platform	Nintendo	11.28	9.14	6.50	2.88	29.80
7	Wii Play	Wii	2006.0	Misc	Nintendo	13.96	9.18	2.93	2.84	28.92
...	...	...	...	...	...	...	...	...	...	...
16667	E.T. The Extra-Terrestrial	GBA	2001.0	Action	NewKidCo	0.01	0.00	0.00	0.00	0.01
16677	Mortal Kombat: Deadly Alliance	GBA	2002.0	Fighting	Midway Games	0.01	0.00	0.00	0.00	0.01
16696	Metal Gear Solid V: Ground Zeroes	PC	2014.0	Action	Konami Digital Entertainment	0.00	0.01	0.00	0.00	0.01
16700	Breach	PC	2011.0	Shooter	Destineer	0.01	0.00	0.00	0.00	0.01
16706	STORM: Frontline Nation	PC	2011.0	Strategy	Unknown	0.00	0.01	0.00	0.00	0.01

6825 rows × 16 columns

In [173]:

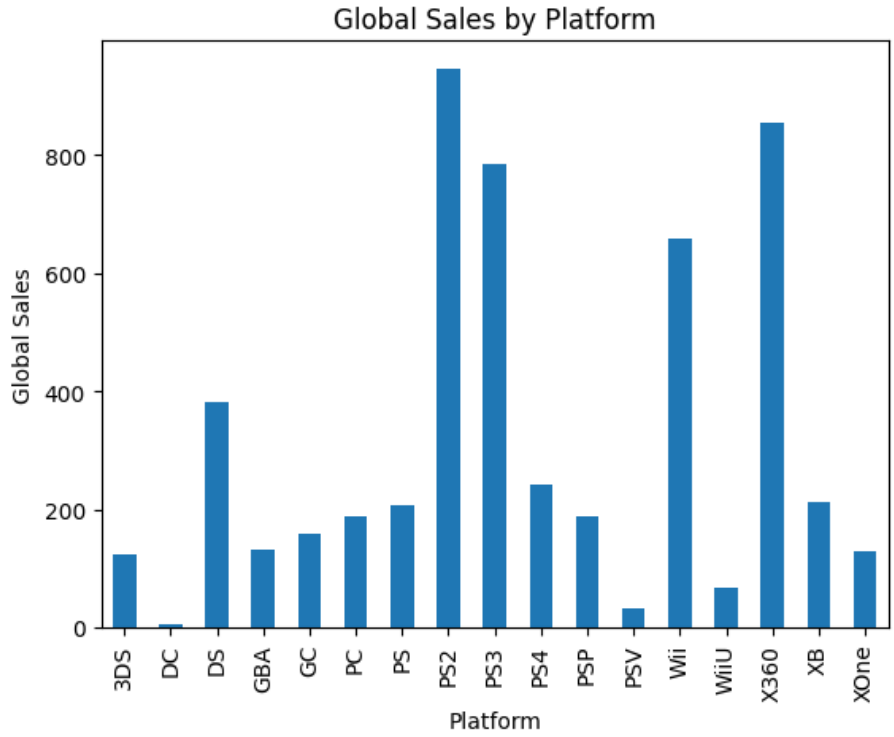
```
#MATPLOTLIB
```

In [174]:

```
#1
df_by_platform = df.groupby('Platform')['Global_Sales'].sum()
df_by_platform.plot(kind='bar', xlabel='Platform', ylabel='Global Sales', title='Global Sales by Platform')
```

Out[174]:

```
<Axes: title={'center': 'Global Sales by Platform'}, xlabel='Platform', ylabel='Global Sales'>
```

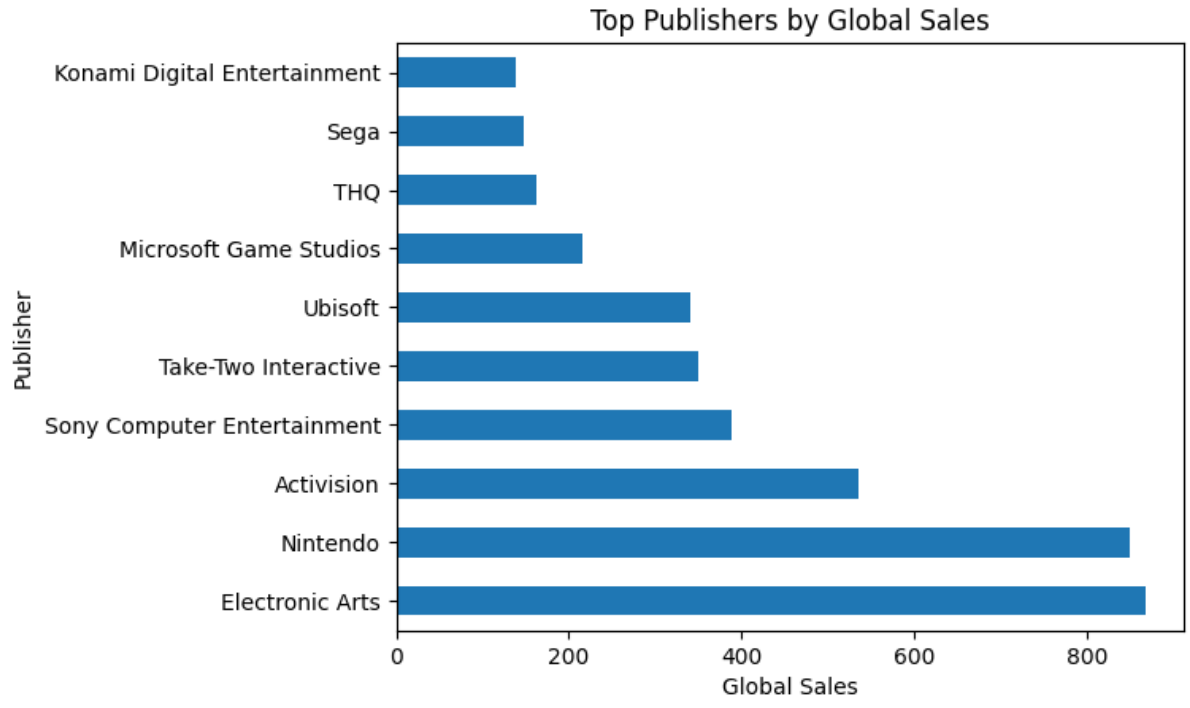


In [175]:

```
df_by_publisher = df.groupby('Publisher')['Global_Sales'].sum().nlargest(10)
df_by_publisher.plot(kind='barh', xlabel='Global Sales', ylabel='Publisher', title='Top Publishers by Global Sales'
```

Out[175]:

```
<Axes: title={'center': 'Top Publishers by Global Sales'}, xlabel='Global Sales', ylabel='Publisher'>
```

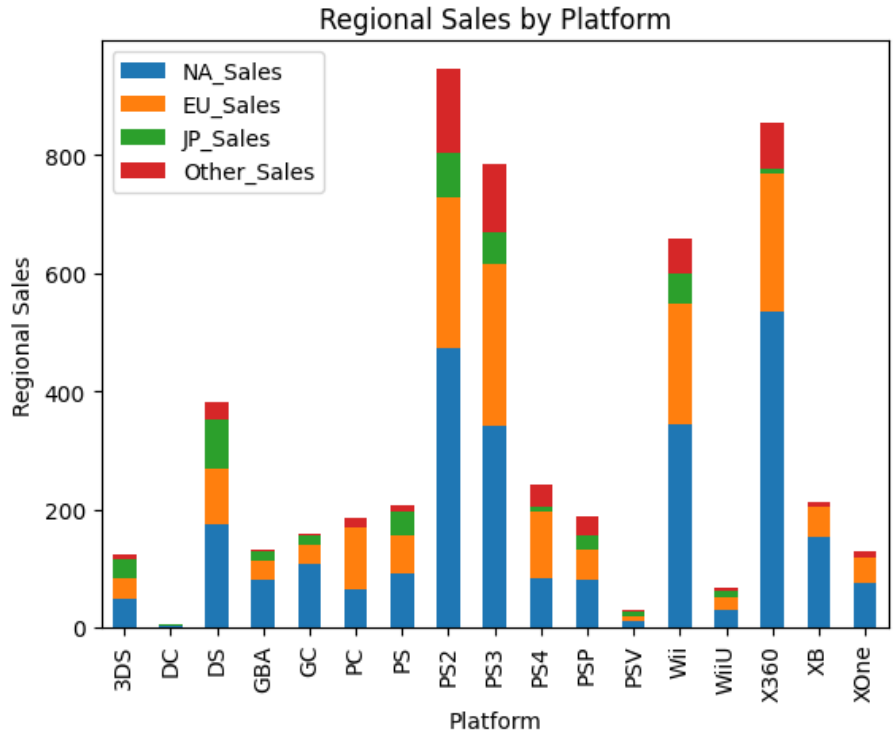


In [176]:

```
df_regional_sales = df.groupby('Platform')[['NA_Sales', 'EU_Sales', 'JP_Sales','Other_Sales']].sum()
df_regional_sales.plot(kind='bar', stacked=True, xlabel='Platform', ylabel='Regional Sales', title='Regional Sales
```

Out[176]:

```
<Axes: title={'center': 'Regional Sales by Platform'}, xlabel='Platform', ylabel='Regional Sales'>
```

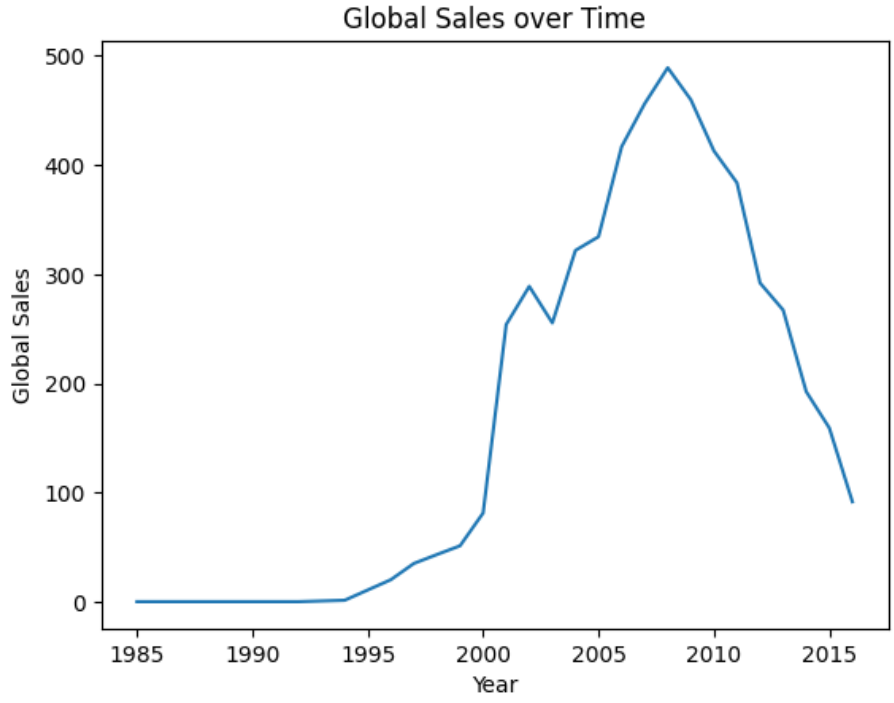


In [177]:

```
df_by_year = df.groupby('Year_of_Release')['Global_Sales'].sum()
df_by_year.plot(kind='line', xlabel='Year', ylabel='Global Sales', title='Global Sales over Time')
```

Out[177]:

```
<Axes: title={'center': 'Global Sales over Time'}, xlabel='Year', ylabel='Global Sales'>
```

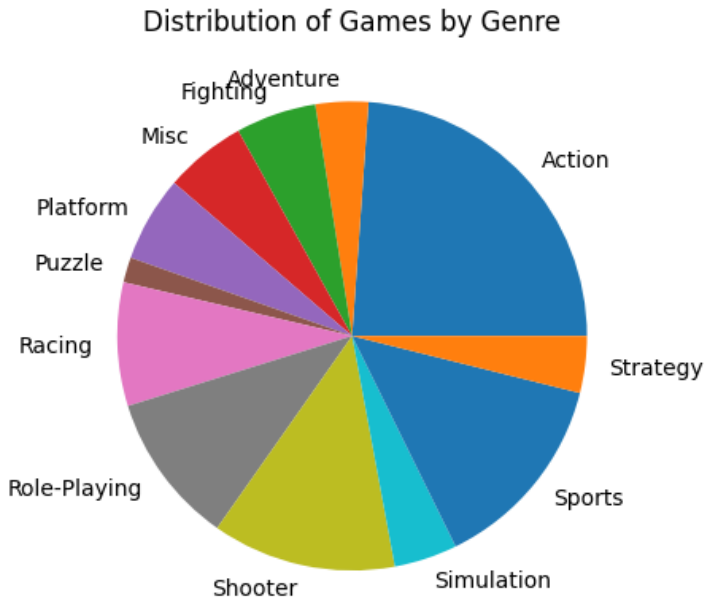


In [178]:

```
df_by_genre = df.groupby('Genre')['Name'].count()
df_by_genre.plot(kind='pie', ylabel='', title='Distribution of Games by Genre')
```

Out[178]:

```
<Axes: title={'center': 'Distribution of Games by Genre'}>
```



In [179]:

```
df_by_platform = df.groupby('Platform')['Name'].count()
df_by_platform.plot(kind='bar', xlabel='Platform', ylabel='Number of Games', title='Number of Games by Platform')
```

Out[179]:

```
<Axes: title={'center': 'Number of Games by Platform'}, xlabel='Platform', ylabel='Number of Games'>
```