KOVAI.CO ASSESSMENT

Task 2: Video Game Sales

Exploratory Data Analysis:

1. Reading the Dataset

```python
data = pd.read_csv('/content/Video Games Sales.csv')
```

2. Examining the Dataset and checking for null values

```python
data.info()
```

```python
data.isnull().sum()
```

```
index              0
Rank               0
Game Title         0
Platform           0
Year              29
Genre              0
Publisher          2
North America      0
Europe             0
Japan              0
Rest of World      0
Global             0
Review             0
dtype: int64
```

3. Replacing the Null values
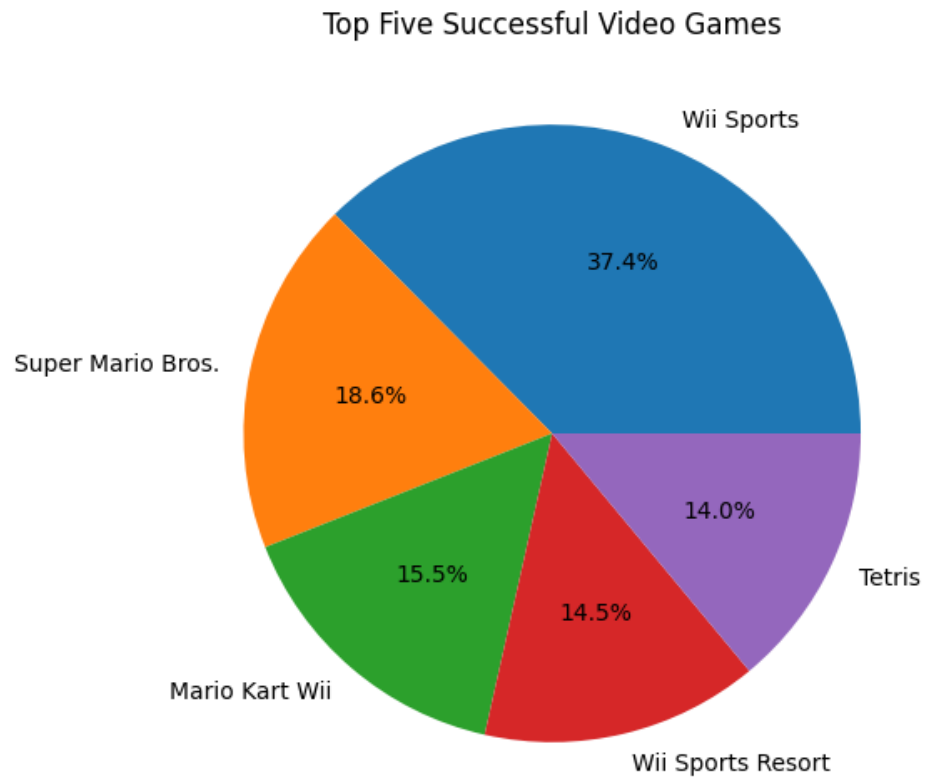
```python
data["Year"].fillna( "NA",inplace=True)
```

```python
data["Publisher"].fillna( "Unknown",inplace=True)
```

4. Games that are successful

```python
sorted = data.sort_values(by='Global', ascending=False)
Global_Successful_Games = sorted[['Game Title', 'Global']].head(5)
print(Global_Successful_Games)
```

```
        Game Title  Global
0         Wii Sports   81.12
1  Super Mario Bros.   40.24
2     Mario Kart Wii   33.55
3  Wii Sports Resort   31.52
4             Tetris   30.26
```

5. Visualizing the game that are successful

Top Five Successful Video Games

INSIGHTS:

1. Top 10 Genre with the highest rating:

```
       Genre  Global
      Sports  703.11
      Action  637.27
    Platform  595.24
     Shooter  557.20
Role-Playing  496.20
      Racing  451.80
        Misc  426.12
    Fighting  249.00
   Adventure  239.45
  Simulation  205.14
```

From above figure we can infer that "Sports" genre is liked by many people globally followed by "Action"

2. **The region which play most of the game**

```python
total_games_played = data[['North America', 'Europe', 'Japan', 'Rest of World']].sum()
most_games_region = total_games_played.idxmax()
print(f"The region that played the most games is: {most_games_region}")

The region that played the most games is: North America
```

The region which played most of the games is North America

3. **The game and the genre people are least interested**

```python
least_interested_game = data[data['Global'] == data['Global'].min()]['Game Title'].values[0]
least_interested_genre = data[data['Global'] == data['Global'].min()]['Genre'].values[0]
```

The least interested game is 'DiRT' in the genre 'Racing' with 0.83 Global sales.

4. **Statistical Analysis of given data**

	index	Rank	North America	Europe	Japan	Rest of World	Global	Review
count	1907.0000	1907.0000	1907.000000	1907.000000	1907.000000	1907.000000	1907.000000	1907.000000
mean	953.0000	954.0000	1.258789	0.706675	0.317493	0.206471	2.489240	79.038977
std	550.6478	550.6478	1.956560	1.148904	0.724945	0.343093	3.563159	10.616899
min	0.0000	1.0000	0.000000	0.000000	0.000000	0.000000	0.830000	30.500000
25%	476.5000	477.5000	0.510000	0.230000	0.000000	0.060000	1.110000	74.000000
50%	953.0000	954.0000	0.810000	0.440000	0.020000	0.130000	1.530000	81.000000
75%	1429.5000	1430.5000	1.375000	0.810000	0.300000	0.220000	2.540000	86.230000
max	1906.0000	1907.0000	40.430000	28.390000	7.200000	8.540000	81.120000	97.000000