Regno:23911

Name: Sriram Ashrith Raj

VIDEO GAME TRENDS ACROSS THE YEARS

Dataset

This dataset contains sales data for video games worldwide, across different platforms, genres, and regions. This database provides insight into what constitutes a hit game in today's gaming industry. Through its comprehensive analysis of various game titles, genres, and platforms, this dataset displays detailed insights into how video games can achieve global success and provides a wonderful window into the ever-changing trends of gaming culture.

Features

- Index   int
- Rank   int
- Game Title   object
- Platform   object
- Year   float
- Genre   object
- Publisher   object
- North America   float
- Europe   float
- Japan   float
- Rest of the World float
- Global   float
- Review   float

Rows

501 Entries

Pre Processing

- **Dropping**
  - The first step is to keep the necessary features and drop the unnecessary features.
  - Here we drop "**index**" and "**Rank**".

- **Imputing Null Values**
  - In the Data frame, Null values had to be removed/filled to avoid wrong inferences.
  - We had to fill few values to "**Year**" and "**Publisher**"

- **Manipulating Data**
  - Data can not be worked with as it is, it has to be manipulated for ease.
  - As there were many Publishers to deal with, they cut down to the top 5 publishers and the rest were grouped under "**Others**".
  - The same was done with platforms, they were cut down to the top 5 and the rest were grouped under "**Others**".

Plotting

## Line Plot

Total Sales in Each Region Over Time
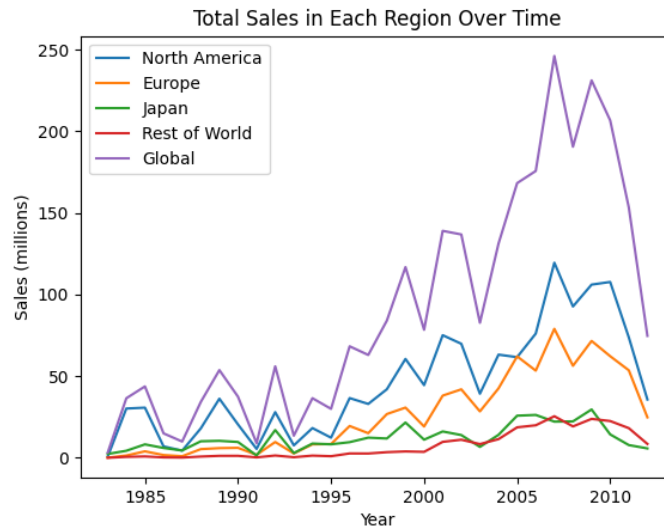
From the line plots, we can infer how game sales have changed over the years. It is visible that Sales are changing in each region differently.

- **Japan –** Sales remained almost the same.
- **North America** - Sales increased very high.
- **Europe** - Sales increased normally.
- **Rest of World** - Sales remained almost the same.
- **Global** - Sales increased very high.
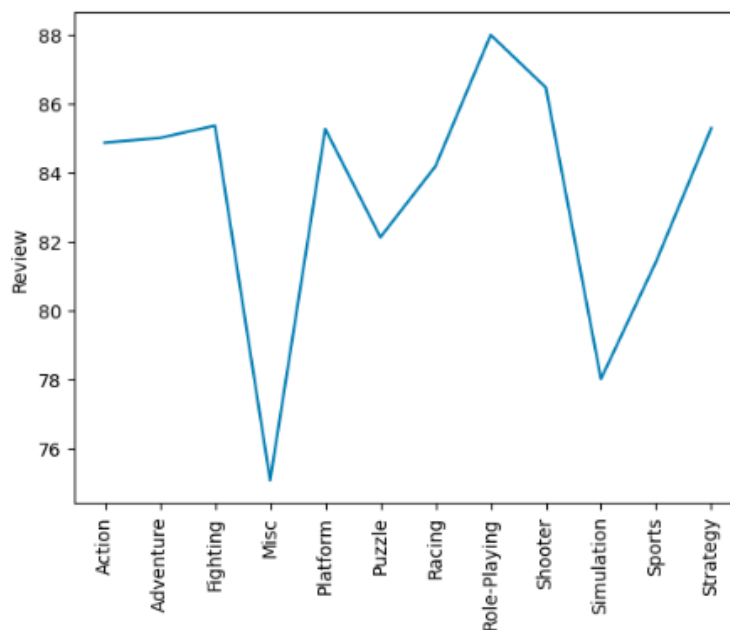
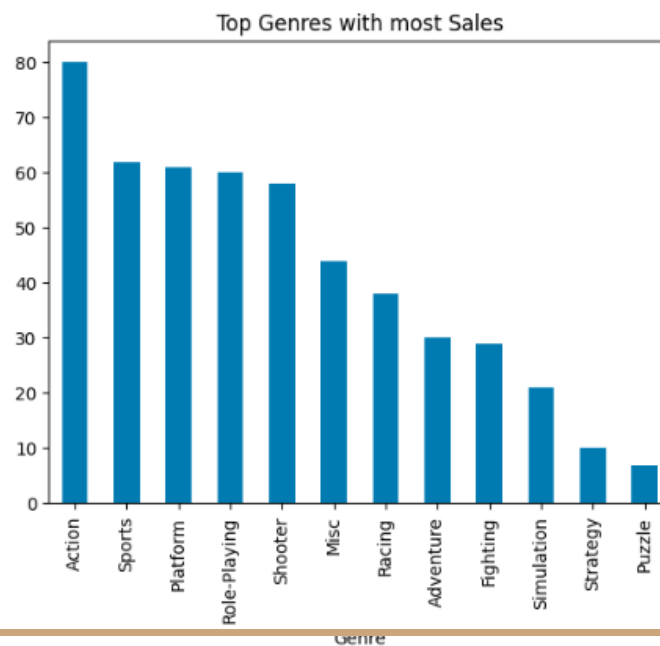From the above graph, we can infer overall reviews of each genre.
  - Genre with the most reviews is **Role-Playing**.
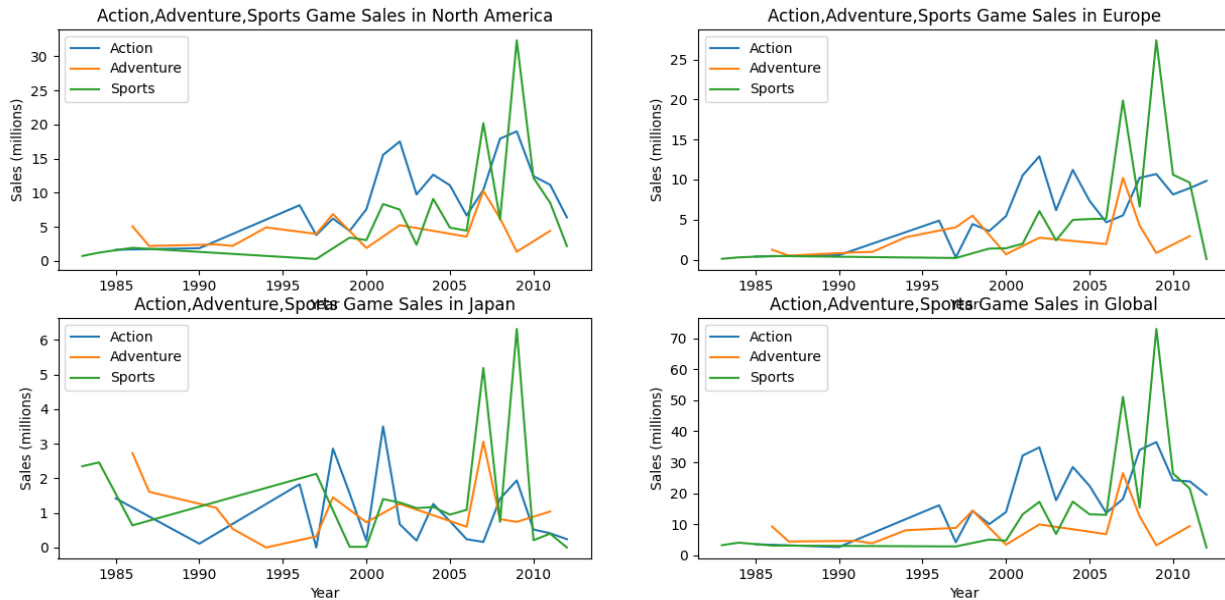
From the plots, we can see that we can see how sales of the most popular genres (**Action, Adventure, and Sports**) changed over the years.
  - Sports genre sales increased tremendously.
  - Compared to the other two, the Adventure genre had fewer sales.
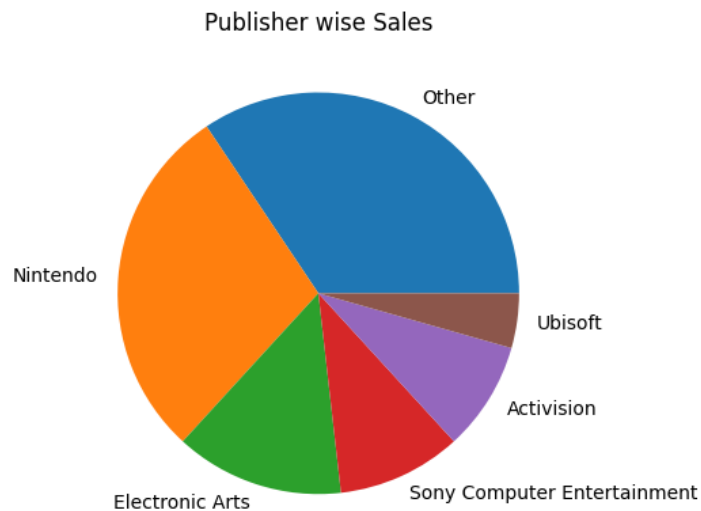
Bar plot

From the above bar graph, we can see how many games are there in each genre.
      - It is visible that **Action** is the one with the most games, followed by **Sports**.
      - **Puzzle** stands last with the least games, followed by **Strategy.**

Pie Chart

Publisher wise Sales
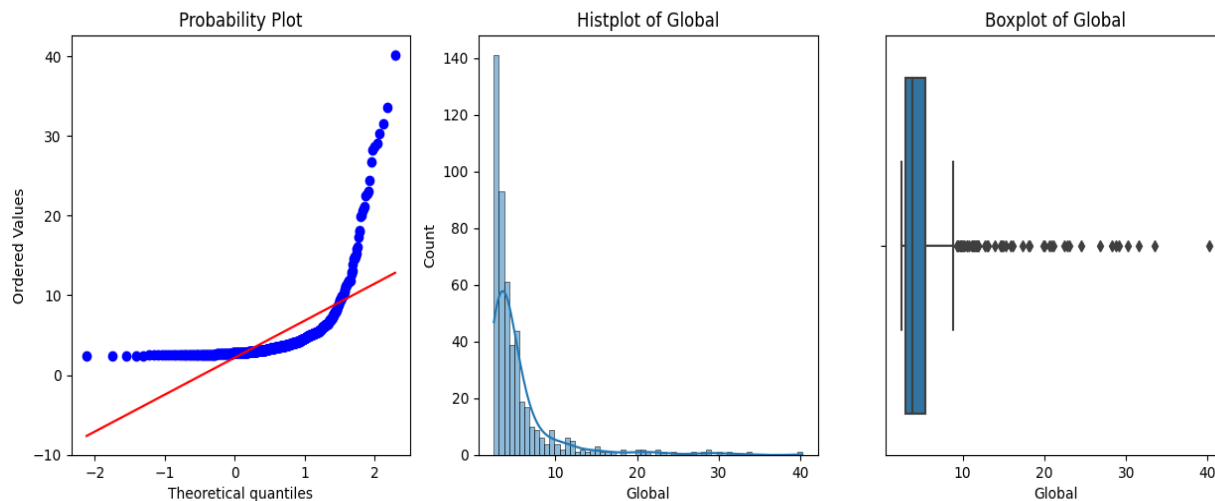
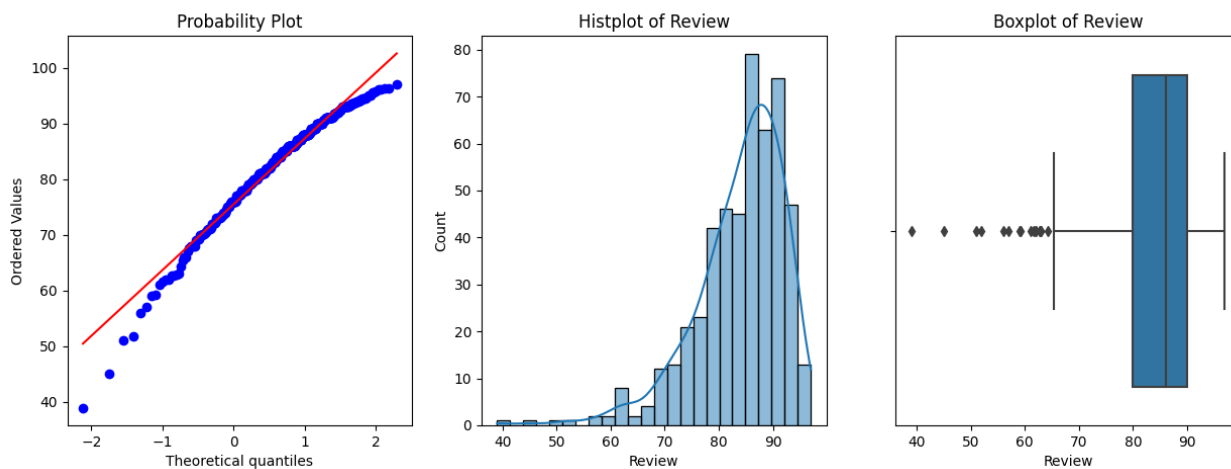The above pie chart shows which Publisher has having most game sales.
      - It is clear that Nintendo has most games.
      - Second stands Electronic Arts.

Hypothesis Testing

Original Data

Global Sales

Reviews

From the above plots, we can infer that data is skewed either to the left or right, with many outliers.

The data taken is not normalized to do Hypothesis Testing. It has to be normalized.
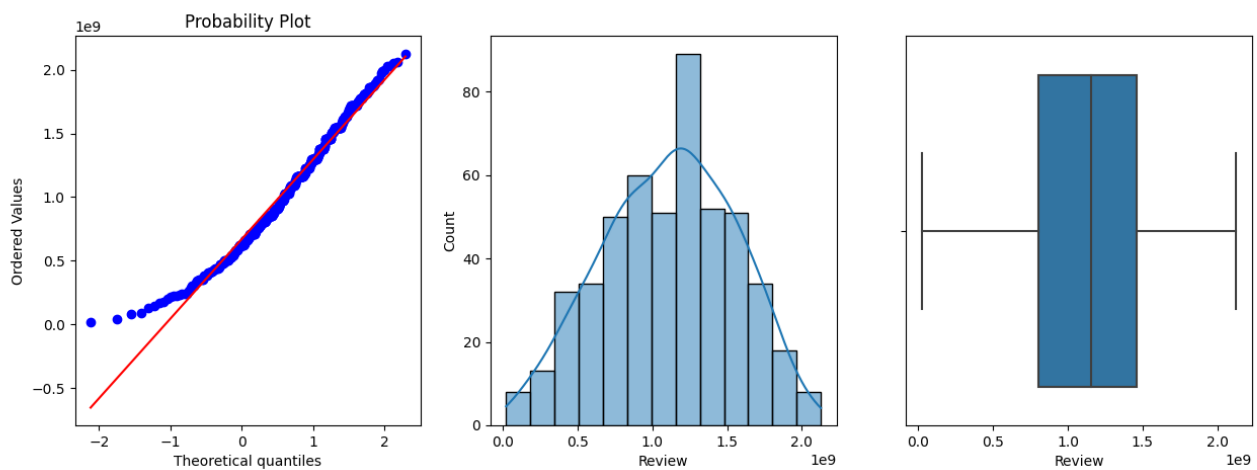
Normalized Data

Data is normalized using the Boxcox method

```python
def normalize(col):
    y, lam = stats.boxcox(col)
    normalcol=(col**lam-1)/lam
    return normalcol,lam
```
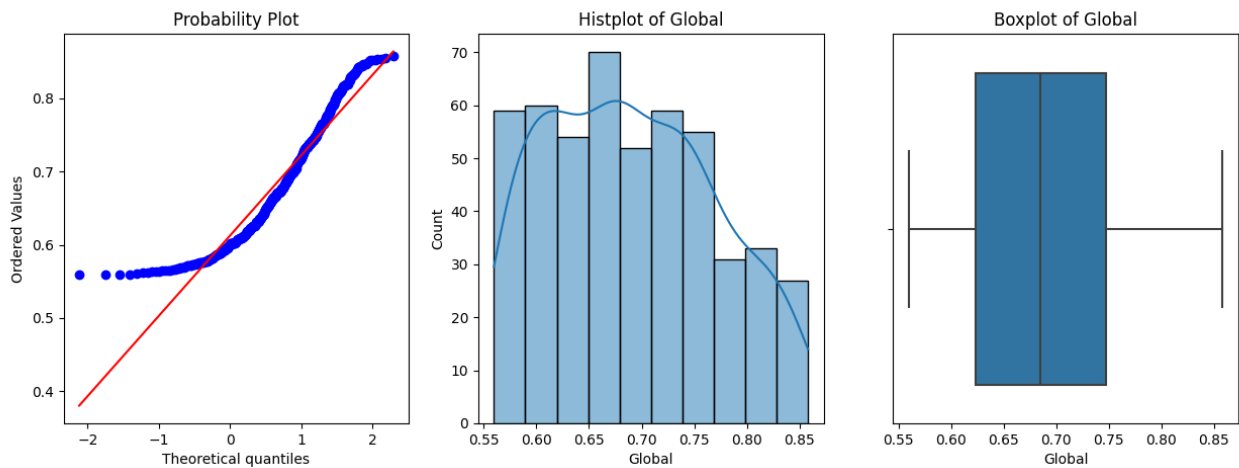

Global Sales

Reviews

Now that the data is normalized, Hypothesis testing can be performed.

Global Sales

$$H_0 = 4\ million\ VS\ H_1 \neq 4\ million$$

```python
mu0=4
alpha=0.05
mu0=(mu0**glam-1)/glam
mean=global_sales.mean()
var=global_sales.var(ddof=1)
n=len(global_sales)
Zcal=(mean-mu0)/(np.sqrt(var/n))
p=2*(1-stats.norm.cdf(abs(Zcal)))
```

Do not Reject H_0

Reviews

$$H_0 = 83\%\ VS\ H_1 \neq 83\%$$

```python
mu0=83
alpha=0.05
mu0=(mu0**rlam-1)/rlam
mean=reviews.mean()
var=reviews.var(ddof=1)
n=len(reviews)
Zcal=(mean-mu0)/(np.sqrt(var/n))
p=2*(1-stats.norm.cdf(abs(Zcal)))
```

Reject H_0