Animation Dataset using numpy,pandas and data visualization

In [78]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```
#Load dataset
data=pd.read_csv("Anime.csv")
```

In [5]:
```
#To see first 5 rows of the DataFrame
data.head()
```

Out[5]:

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_year	De
0	1	Demon Slayer: Kimetsu no Yaiba - Entertainment...	Kimetsu no Yaiba: Yuukaku-hen	TV	NaN	ufotable	Fall	Action, Adventure, Fantasy, Shounen, Demons, H...	4.60	2021.0	NaN	'T h ac
1	2	Fruits Basket the Final Season	Fruits Basket the Final	TV	13.0	TMS Entertainment	Spring	Drama, Fantasy, Romance, Shoujo, Animal Transf...	4.60	2021.0	NaN	ar
2	3	Mo Dao Zu Shi 3	The Founder of Diabolism 3	Web	12.0	B.C MAY PICTURES	NaN	Fantasy, Ancient China, Chinese Animation, Cul...	4.58	2021.0	NaN	M
3	4	Fullmetal Alchemist: Brotherhood	Hagane no Renkinjutsushi: Full Metal Alchemist	TV	64.0	Bones	Spring	Action, Adventure, Drama, Fantasy, Mystery, Sh...	4.58	2009.0	2010.0	f o is
4	5	Attack on Titan 3rd Season: Part II	Shingeki no Kyojin Season 3: Part II	TV	10.0	WIT Studio	Spring	Action, Fantasy, Horror, Shounen, Dark Fantasy...	4.57	2019.0	NaN	' V be

In [6]:
```
#To see last 5 rows of the DataFrame
data.tail()
```

Out[6]:

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_year	Des
18490	18491	Qin Shi Mingyue: Canghai Hengliu Xiaomeng Spec...	NaN	Web	2.0	Sparkly Key Animation Studio	NaN	Action, Ancient China, Chinese Animation, Hist...	NaN	2020.0	NaN	epi M Can
18491	18492	Yi Tang Juchang: Sanguo Yanyi	NaN	TV	108.0	NaN	NaN	Chinese Animation	NaN	2010.0	NaN	No ye ba
18492	18493	Fenghuang Ji Xiang Yu Qingming Shanghe Tu	NaN	TV	13.0	NaN	NaN	Chinese Animation, Family Friendly, Short Epis...	NaN	2020.0	NaN	No ye ba
18493	18494	Chengshi Jiyi Wo Men de Jieri	NaN	TV	NaN	NaN	NaN	Chinese Animation, Family Friendly, Short Epis...	NaN	2020.0	NaN	No ye ba
18494	18495	Heisei Inu Monogatari Bow: Genshi Inu Monogata...	NaN	Movie	NaN	Nippon Animation	NaN	Comedy, Slice of Life, Dogs	NaN	1994.0	NaN	No ye ba

In [7]:
```
#To find no of rows & columns
data.shape
```

Out[7]:
```
(18495, 17)
```

In [8]:
```
#To find total cell count
data.size
```

Out[8]:
```
314415
```

In [9]:
```
#To find Dimension
data.ndim
```

Out[9]:
```
2
```

In [10]:
```
#To find column names
data.columns
```

Out[10]:
```
Index(['Rank', 'Name', 'Japanese_name', 'Type', 'Episodes', 'Studio',
       'Release_season', 'Tags', 'Rating', 'Release_year', 'End_year',
       'Description', 'Content_Warning', 'Related_Mange', 'Related_anime',
       'Voice_actors', 'staff'],
      dtype='object')
```

In [11]:
```
#To find information about the DataFrame
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18495 entries, 0 to 18494
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Rank             18495 non-null  int64  
 1   Name             18495 non-null  object 
 2   Japanese_name    7938 non-null   object 
 3   Type             18495 non-null  object 
 4   Episodes         9501 non-null   float64
 5   Studio           12018 non-null  object 
 6   Release_season   4116 non-null   object 
 7   Tags             18095 non-null  object 
 8   Rating           15364 non-null  float64
 9   Release_year     18112 non-null  float64
 10  End_year         2854 non-null   float64
 11  Description      18491 non-null  object 
 12  Content_Warning  1840 non-null   object 
 13  Related_Mange    7627 non-null   object 
 14  Related_anime    10063 non-null  object 
 15  Voice_actors     15309 non-null  object 
 16  staff            13005 non-null  object 
dtypes: float64(4), int64(1), object(12)
memory usage: 2.4+ MB
```

In [12]:
```
#To find min,max,count,etc...
data.describe()
```

Out[12]:

	Rank	Episodes	Rating	Release_year	End_year
count	18495.00000	9501.000000	15364.000000	18112.000000	2854.000000
mean	9248.00000	20.920850	3.355133	2006.520318	2004.256132
std	5339.19095	37.990858	0.400624	15.189537	13.257484
min	1.00000	1.000000	0.960000	1907.000000	1962.000000
25%	4624.50000	2.000000	3.130000	2001.000000	1996.000000
50%	9248.00000	12.000000	3.360000	2012.000000	2007.000000
75%	13871.50000	26.000000	3.590000	2017.000000	2015.000000
max	18495.00000	800.000000	4.600000	2023.000000	2022.000000

In [13]:
```
#Slicing of rows
data.loc[75:85]
```

Out[13]:

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_year
75	76	March Comes in like a Lion 2nd Season	3-gatsu no Lion 2nd Season	TV	22.0	SHAFT	Fall	Drama, Seinen, Slice of Life, Board Games, Mel...	4.36	2017.0	2018.0
76	77	Natsume's Book of Friends Season 3	Natsume Yuujinchou San	TV	13.0	Brain's Base	Summer	Drama, Shoujo, Slice of Life, Bodyguards, Cats...	4.36	2011.0	NaN
77	78	Attack on Titan 2nd Season	Shingeki no Kyojin 2nd Season	TV	12.0	WIT Studio	Spring	Action, Fantasy, Horror, Shounen, Dark Fantasy...	4.36	2017.0	NaN
78	79	I Want to Eat Your Pancreas	Kimi no Suizou wo Tabetai	Movie	NaN	Studio VOLN	NaN	Drama, Romance, Seinen, Coming of Age, Illness...	4.36	2018.0	NaN
79	80	Rascal Does Not Dream of a Dreaming Girl	Seishun Buta Yarou wa Yumemiru Shoujo no Yume...	Movie	NaN	CloverWorks	NaN	Drama, Romance, Illness, Senpai-Kouhai Relatio...	4.36	2019.0	NaN
80	81	Given Movie	NaN	Movie	NaN	Lerche	NaN	BL, Drama, Romance, Shounen-ai, Adult Couples,...	4.36	2020.0	NaN
81	82	Mushishi Zoku Shou	Mushishi -Next Passage-	TV	10.0	Artland	Spring	Fantasy, Seinen, Episodic, Iyashikei, Supernat...	4.36	2014.0	NaN

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_year
82	83	Princess Mononoke	Mononoke Hime	Movie	NaN	Studio Ghibli	NaN	Action, Adventure, Fantasy, Curse, Environment...	4.36	1997.0	NaN
83	84	Kuroko's Basketball 3	Kuroko no Basket 3	TV	25.0	Production I.G	Winter	Shounen, Sports, Basketball, School Club, Tour...	4.36	2015.0	NaN
84	85	BTS: We are Bulletproof - the Eternal	NaN	Music	NaN	Studio Pivote	NaN	Chibi, Korean Animation	4.36	2020.0	NaN
85	86	Haikyuu!! Movie 4: Battle of Concepts	Haikyuu!! Movie 4: Concept no Tatakai	Movie	NaN	Production I.G	NaN	Shounen, Sports, Recap, School Club, Tournamen...	4.36	2017.0	NaN

In [14]:
```
#To find the total count of a specific variable
data["Type"].value_counts()
```

Out[14]:
```
TV       5446
Movie    3577
Web      2488
OVA      2235
Music    2165
Other     990
DVD S     911
TV Sp     683
Name: Type, dtype: int64
```

In [22]:
```
print("Mean Value : ",data[data["Release_season"]=="Spring"]["Episodes"].mean())
print("Median Value : ",data[data["Release_season"]=="Spring"]["Episodes"].median())
print("Count : ",data[data["Release_season"]=="Spring"]["Episodes"].count())
```

```
Mean Value :  29.045864045864047
Median Value :  22.0
Count :  1221
```

In [24]:
```
print("Mean Value : ",data.groupby("Type")["Episodes"].mean())
```

```
Mean Value :  Type
DVD S          NaN
Movie          NaN
Music          NaN
OVA       2.547101
Other          NaN
TV       33.931782
TV Sp          NaN
Web       9.949669
Name: Episodes, dtype: float64
```

In [25]:
```
#Sort value based upon some condition
data.sort_values(["Episodes","Release_year"],ascending=False)
```

Out[25]:

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_ye
17902	17903	Shuimu Baobao Kan Shijie	NaN	TV	800.0	NaN	NaN	Chinese Animation, Family Friendly, Short Epis...	NaN	2008.0	Na
7759	7760	Kotowaza House	NaN	TV	773.0	Eiken	NaN	Slice of Life, Short Episodes	3.35	1987.0	1994
8986	8987	Asa da yo! Kaishain	Kaishain: Shellfish Employees	TV	744.0	DLE	Spring	Comedy, Animal Protagonists, Anthropomorphic, ...	3.29	2016.0	2019
6529	6530	Shima Shima Tora no Shimajirou	NaN	TV	726.0	St. Signpost	NaN	Comedy, Fantasy, Slice of Life, Animal Protago...	3.42	1993.0	2008
5426	5427	Ninja Hattori-kun	Ninja Hattori	TV	694.0	Shin-Ei Animation	NaN	Comedy, Shounen, Ninja, Based on a Manga	3.48	1981.0	1987
...	...	...	...	...	...	...	...	...	...	...	
18464	18465	Reincarnated as a Sword	Tensei Shitara Ken deshita	TV	NaN	NaN	NaN	Action, Adventure, Fantasy, Animal Characteris...	NaN	NaN	Na
18479	18480	Baise Shandian	NaN	Web	NaN	HuaMei Animation	NaN	Sports, Chinese Animation, Ping Pong	NaN	NaN	Na
18481	18482	Goblin Slayer 2	NaN	TV	NaN	NaN	NaN	Action, Adventure, Fantasy, Seinen, Dark Fanta...	NaN	NaN	Na

	Rank	Name	Japanese_name	Type	Episodes	Studio	Release_season	Tags	Rating	Release_year	End_yea
18483	18484	Peleliu: Rakuen no Guernica		NaN	Other	NaN	NaN	NaN	Action, Shounen, Island, War, World War 2, Bas...	NaN	NaN	Na
18487	18488	Make My Day		NaN	Web	NaN	Studio 5	NaN	Horror, Monsters, Original Work	NaN	NaN	Na

18495 rows × 17 columns

In [27]:
```
data.groupby("Release_season")["Rating"].agg(["mean","median","count","max","min"])
```

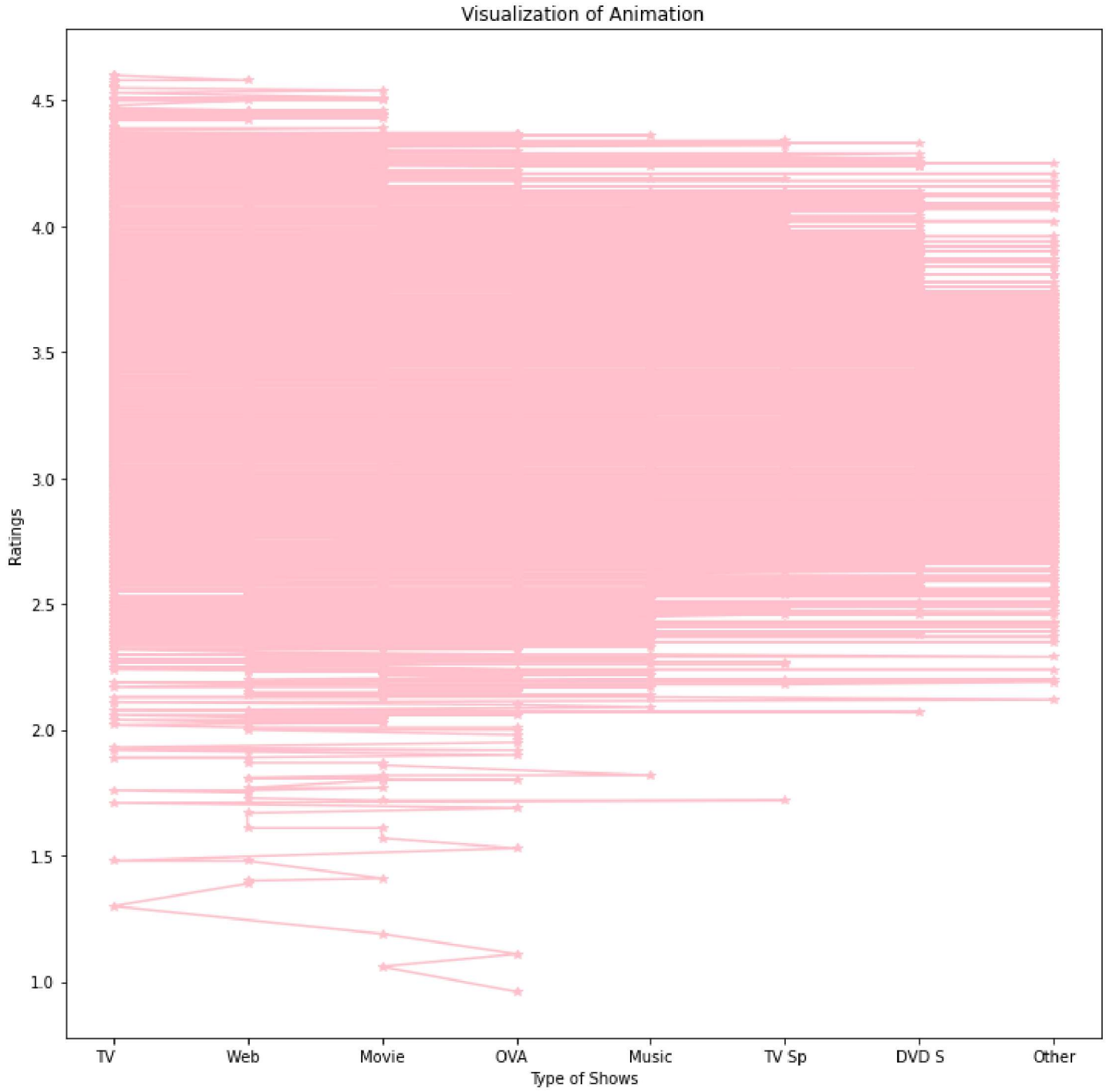
Out[27]:

Release_season	mean	median	count	max	min
Fall	3.474794	3.470	1143	4.60	2.06
Spring	3.476873	3.475	1228	4.60	1.48
Summer	3.442440	3.440	791	4.48	1.71
Winter	3.455766	3.450	829	4.56	1.30

Line plot

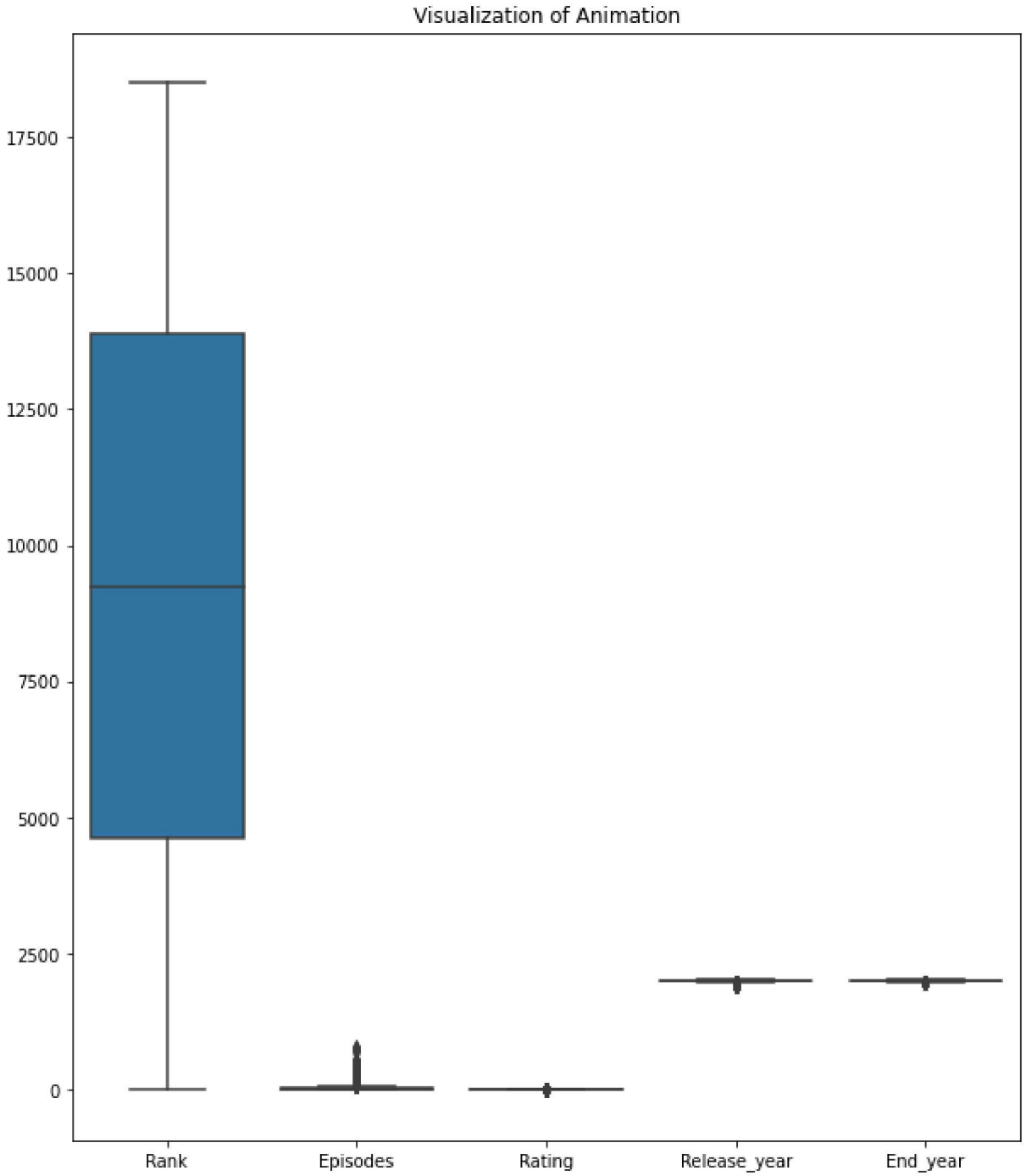
In [37]:

```
plt.figure(figsize=(12,12))
plt.title("Visualization of Animation")
plt.xlabel("Type of Shows")
plt.ylabel("Ratings")
plt.plot(data["Type"],data["Rating"],marker="*",color="Pink");
```


Visualization of Animation
Ratings
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
TV
Web
Movie
OVA
Music
TV Sp
DVD S
Other
Type of Shows

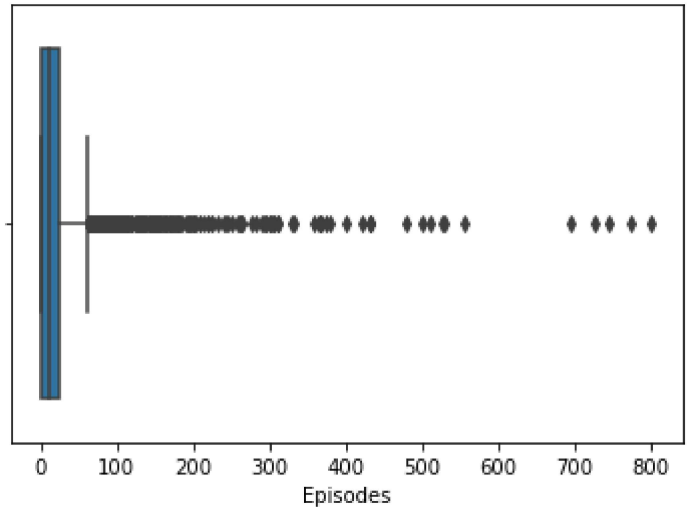
To find Outlier

In [61]:
```
plt.figure(figsize=(10,12))
plt.title("Visualization of Animation")
sns.boxplot(data=data);
```


Visualization of Animation
17500
15000
12500
10000
7500
5000
2500
0
Rank
Episodes
Rating
Release_year
End_year

In [62]:
```
sns.boxplot(data["Episodes"]);
```

```
C:\Users\admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variab
le as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing othe
r arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

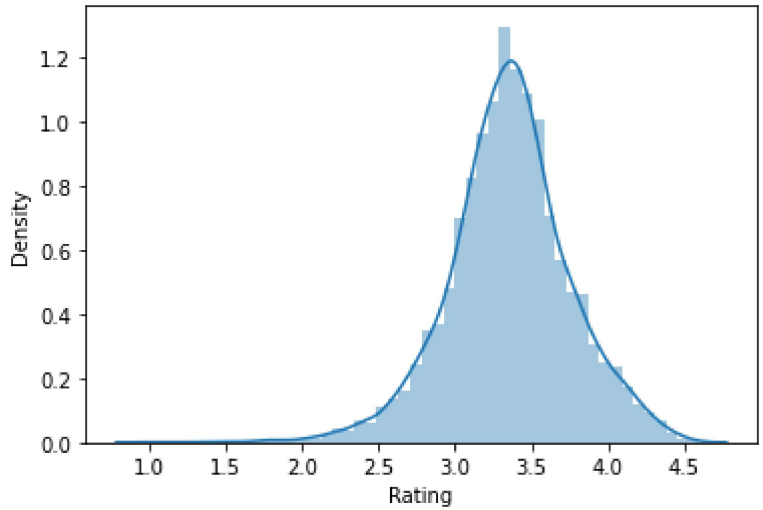


Distribution of the Value

In [69]:
```
sns.distplot(data["Rating"]);
print("Mean Value : ",data["Episodes"].mean())
print("Median Value : ",data["Episodes"].median())
```

```
C:\Users\admin\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a depre
cated function and will be removed in a future version. Please adapt your code to use either `displot` (a figu
re-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

Mean Value :  20.920850436796126
Median Value :  12.0
```



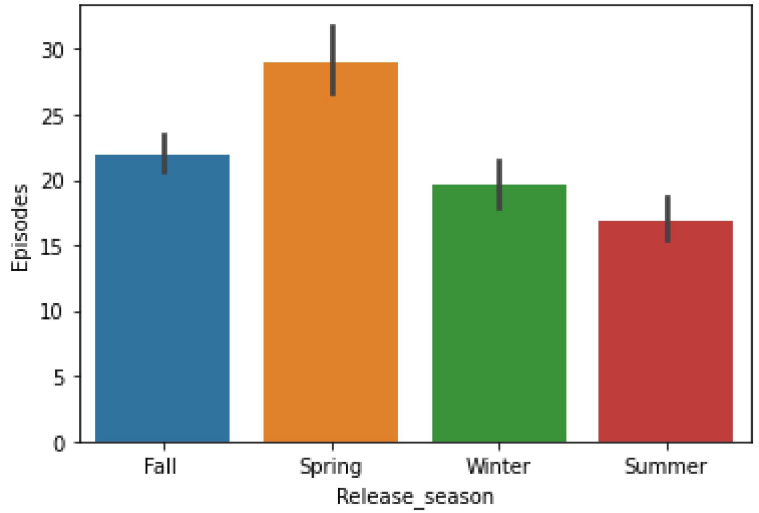
In [ ]:
```
Here the data is not normally distributed it contains left skew
```

In [70]:
```
data["Release_season"].value_counts()
```

Out[70]:
```
Spring    1291
Fall      1164
Winter     854
Summer     807
Name: Release_season, dtype: int64
```

In [72]:
```
sns.barplot(data["Release_season"],data["Episodes"]);
```

```
C:\Users\admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variab
les as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing o
ther arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

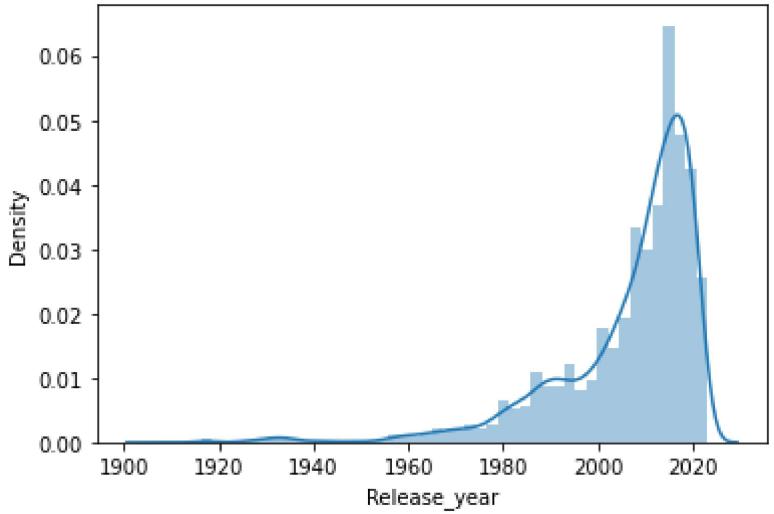


In [75]:
```
data["Release_year"].value_counts()
```

Out[75]:
```
2017.0    1010
2018.0     997
2016.0     957
2019.0     903
2020.0     885
           ...
1951.0       2
1949.0       1
1944.0       1
1945.0       1
1907.0       1
Name: Release_year, Length: 103, dtype: int64
```

In [76]:
```
sns.distplot(data["Release_year"]);
```

```
C:\Users\admin\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a depre
cated function and will be removed in a future version. Please adapt your code to use either `displot` (a figu
re-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



In [ ]: