

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [11]:

```
dharma_data = pd.read_csv('dharma.csv')
```

In [12]:

```
dharma_data.head()
```

Out[12]:

|   | Title              | Year | Director       | Producer(s) | Cast  | Synopsis  | Ref. |
|---|--------------------|------|----------------|-------------|---|---|------|
| 0 | Dostana            | 1980 | Raj Khosla     | Yash Johar  | Amitabh Bachchan, Shatrughan Sinha, Zeenat Aman   | A police officer and a lawyer are best friends... | [18] |
| 1 | Duniya             | 1984 | Ramesh Talwar  | Yash Johar  | Dilip Kumar, Rishi Kapoor, Amrita Singh, Saira... | An honest man is wrongly imprisoned by crimina... | [19] |
| 2 | Muqaddar Ka Faisla | 1987 | Prakash Mehra  | Yash Johar  | Raaj Kumar, Raakhee, Raj Babbar, Meenakshi Ses... | A Hindu priest is wrongly imprisoned for rape.... | [20] |
| 3 | Agneepath          | 1990 | Mukul S. Anand | Yash Johar  | Amitabh Bachchan, Mithun Chakraborty, Danny De... | A common man seeks revenge from an unscrupulou... | [21] |
| 4 | Gumrah             | 1993 | Mahesh Bhatt   | Yash Johar  | Sridevi, Sanjay Dutt, Rahul Roy                   | An aspiring singer is wrongly accused of drug ... | [22] |

In [13]:

```
dharma_data.tail()
```

Out[13]:

|    | Title                         | Year | Director         | Producer(s)                                       | Cast  | Synopsis  | Ref.           |
|----|-------------------------------|------|------------------|---|---|---|----------------|
| 56 | Dostana 2                     | 2022 | Collin D'Cunha   | Karan JoharHiroo Johar,Apoorva Mehta              | Janhvi Kapoor, Laksh Lalwani                      | TBA   | [111]          |
| 57 | Bedhadak                      | 2022 | Shashank Khaitan | Karan JoharHiroo Yash Johar,Apoorva Mehta         | Laksh Lalwani, Gurfateh Singh Pirzada, Shanaya... | TBA   | [112]<br>[113] |
| 58 | Rocky Aur Rani Ki Prem Kahani | 2023 |                  | Karan Johar                                       | Karan JoharHiroo Yash Johar,Apoorva Mehta         | Ranveer Singh, Alia Bhatt, Dharmendra, Shabana... | TBA            |
| 59 | Anand Tiwari's untitled next  | 2023 | Karan Johar      | Karan JoharHiroo Yash Johar,Apoorva Mehta         | Vicky Kaushal, Tripti Dimri, Ammy Virk            | TBA   | [115]          |
| 60 | Selfiee                       | 2023 | Raj Mehta        | Karan JoharPrithviraj SukumaranAruna Bhatia,Hi... | Akshay Kumar, Emraan Hashmi, Nushrratt Bharucc... | TBA   | [116]<br>[117] |

In [14]:

```
dharma_data.shape
```

Out[14]:

(61, 7)

In [15]:

```
dharma_data.columns
```

Out[15]:

Index(['Title', 'Year', 'Director', 'Producer(s)', 'Cast', 'Synopsis', 'Ref.'], dtype='object')

In [16]:



```
dharmadata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61 entries, 0 to 60
Data columns (total 7 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Title           61 non-null    object
 1   Year            61 non-null    int64
 2   Director        61 non-null    object
 3   Producer(s)     61 non-null    object
 4   Cast            61 non-null    object
 5   Synopsis        61 non-null    object
 6   Ref.            61 non-null    object
dtypes: int64(1), object(6)
memory usage: 3.5+ KB
```

In [17]:



```
dharmadata.describe()
```

Out[17]:

|       | Year        |
|-------|-------------|
| count | 61.000000   |
| mean  | 2012.918033 |
| std   | 10.028784   |
| min   | 1980.000000 |
| 25%   | 2010.000000 |
| 50%   | 2016.000000 |
| 75%   | 2020.000000 |
| max   | 2023.000000 |

In [18]:



```
dharmadata.isnull().sum()
```

Out[18]:

|             |   |
|-------------|---|
| Title       | 0 |
| Year        | 0 |
| Director    | 0 |
| Producer(s) | 0 |
| Cast        | 0 |
| Synopsis    | 0 |
| Ref.        | 0 |

dtype: int64

In [19]:



```
dharma_data.rename(columns = {'Producer(s)': 'Producers'}, inplace = True)
```

In [20]:



```
dharma_data = dharma_data.drop('Ref.', axis = 1)
```

In [21]:



```
dharma_data.nunique()
```

Out[21]:

```
Title          59
Year           25
Director        35
Producers       32
Cast           60
Synopsis        50
dtype: int64
```

In [22]:



```
dharma_data.Year.unique()
```

Out[22]:

```
array([1980, 1984, 1987, 1990, 1993, 1998, 2001, 2003, 2005, 2006, 2008,
       2009, 2010, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020,
       2021, 2022, 2023], dtype=int64)
```

In [23]:

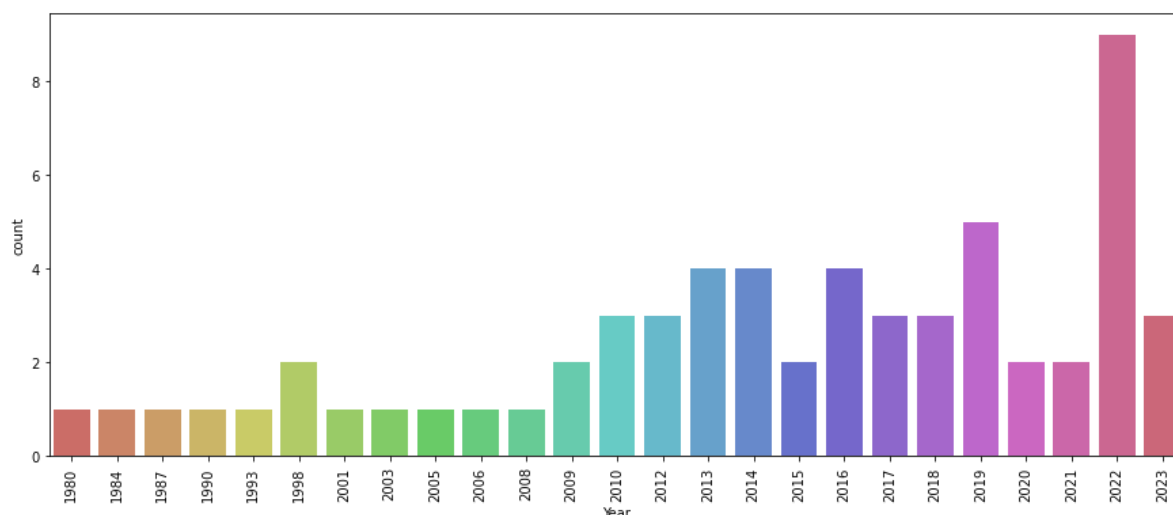
```
dharma_data.Year.value_counts()
```

Out[23]:

```
2022    9
2019    5
2016    4
2014    4
2013    4
2010    3
2012    3
2018    3
2017    3
2023    3
2009    2
2015    2
1998    2
2020    2
2021    2
2008    1
1984    1
2006    1
2005    1
2003    1
2001    1
1993    1
1990    1
1987    1
1980    1
Name: Year, dtype: int64
```

In [24]:

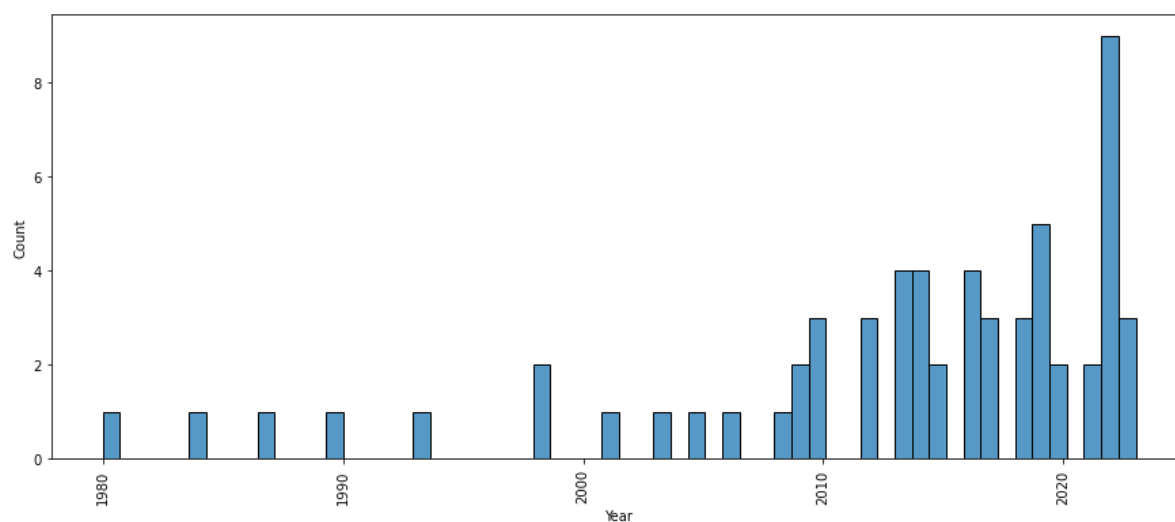
```
plt.figure(figsize=(15,6))
sns.countplot('Year', data = dharma_data, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



In [30]:



```
plt.figure(figsize=(15,6))
sns.histplot(dharma_data.Year, bins = 60)
plt.xticks(rotation = 90)
plt.show()
```



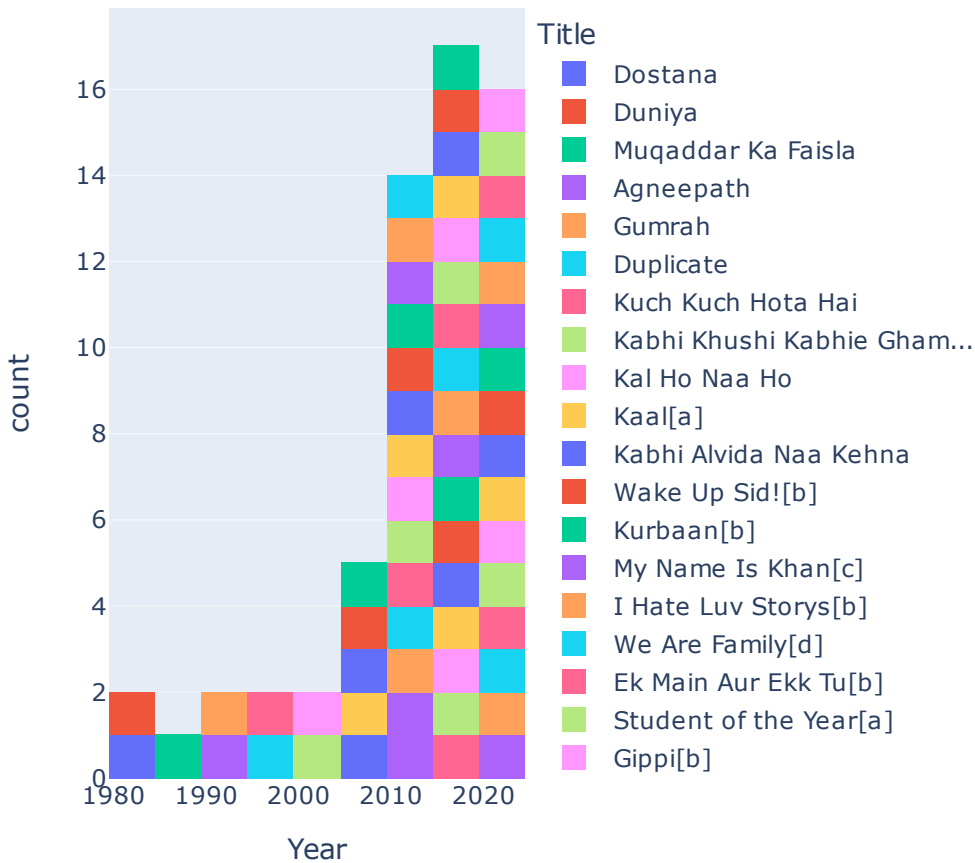
In [31]:



```
import plotly.express as px
```

In [32]:

```
fig1 = px.histogram(dharma_data, x = 'Year', color = 'Title')
fig1.show()
```



In [38]:

```
dharma_data.Director.unique()
```

Out[38]:

```
array(['Raj Khosla', 'Ramesh Talwar', 'Prakash Mehra', 'Mukul S. Anand',
      'Mahesh Bhatt', 'Karan Johar', 'Nikhil Advani', 'Soham Shah',
      'Tarun Mansukhani', 'Ayan Mukerji', 'Rensil D'Silva',
      'Punit Malhotra', 'Siddharth Malhotra', 'Karan Malhotra',
      'Shakun Batra', 'Sonam Nair', 'Ritesh Batra', 'Vinil Mathew',
      'Abhishek Varman', 'Shashank Khaitan', 'Vikas Bahl', 'Nitya Mehra',
      'Gauri Shinde', 'Shaad Ali', 'Abhay Chopra', 'Meghna Gulzar',
      'Rohit Shetty', 'Anurag Singh', 'Raj Mehta', 'Bhanu Pratap Singh',
      'Sharan Sharma', 'Vishnuvardhan', 'Puri Jagannadh',
      'Sagar AmbrePushkar Ojha', 'Collin D'Cunha'], dtype=object)
```

In [39]:



```
dharma_data.Director.value_counts()
```

Out[39]:

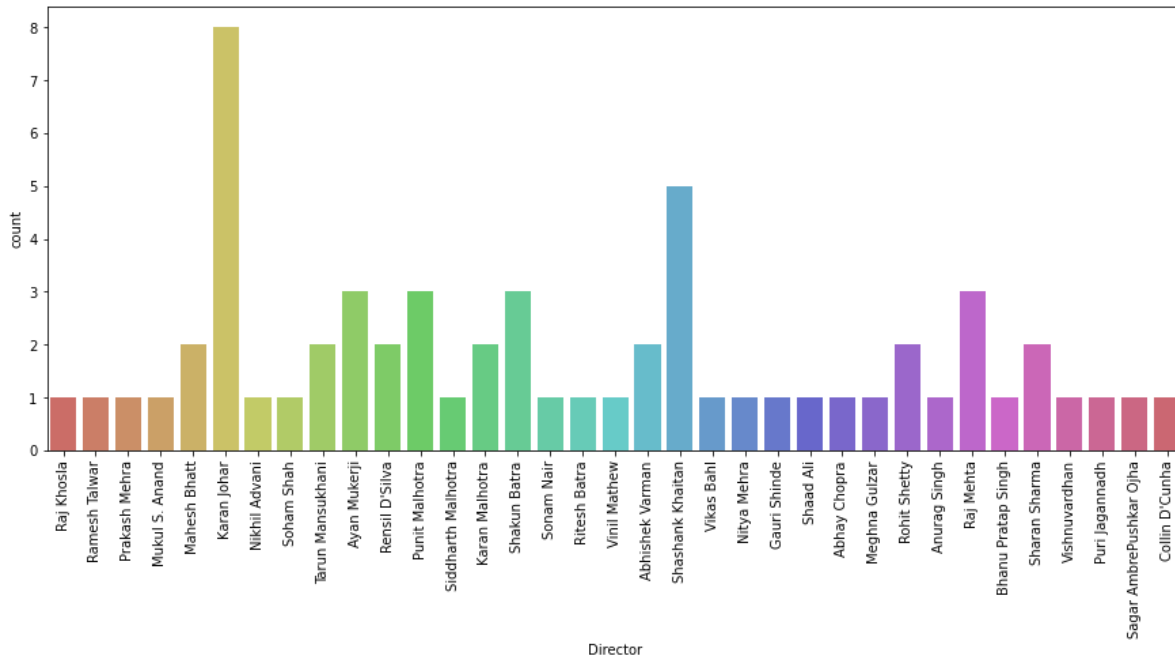
|                         |   |
|-------------------------|---|
| Karan Johar             | 8 |
| Shashank Khaitan        | 5 |
| Punit Malhotra          | 3 |
| Raj Mehta               | 3 |
| Ayan Mukerji            | 3 |
| Shakun Batra            | 3 |
| Mahesh Bhatt            | 2 |
| Abhishek Varman         | 2 |
| Sharan Sharma           | 2 |
| Tarun Mansukhani        | 2 |
| Rensil D'Silva          | 2 |
| Rohit Shetty            | 2 |
| Karan Malhotra          | 2 |
| Meghna Gulzar           | 1 |
| Anurag Singh            | 1 |
| Raj Khosla              | 1 |
| Bhanu Pratap Singh      | 1 |
| Shaad Ali               | 1 |
| Vishnuvardhan           | 1 |
| Puri Jagannadh          | 1 |
| Sagar AmbrePushkar Ojha | 1 |
| Abhay Chopra            | 1 |
| Vinil Mathew            | 1 |
| Gauri Shinde            | 1 |
| Nitya Mehra             | 1 |
| Vikas Bahl              | 1 |
| Ramesh Talwar           | 1 |
| Ritesh Batra            | 1 |
| Sonam Nair              | 1 |
| Siddharth Malhotra      | 1 |
| Soham Shah              | 1 |
| Nikhil Advani           | 1 |
| Mukul S. Anand          | 1 |
| Prakash Mehra           | 1 |
| Collin D'Cunha          | 1 |

Name: Director, dtype: int64



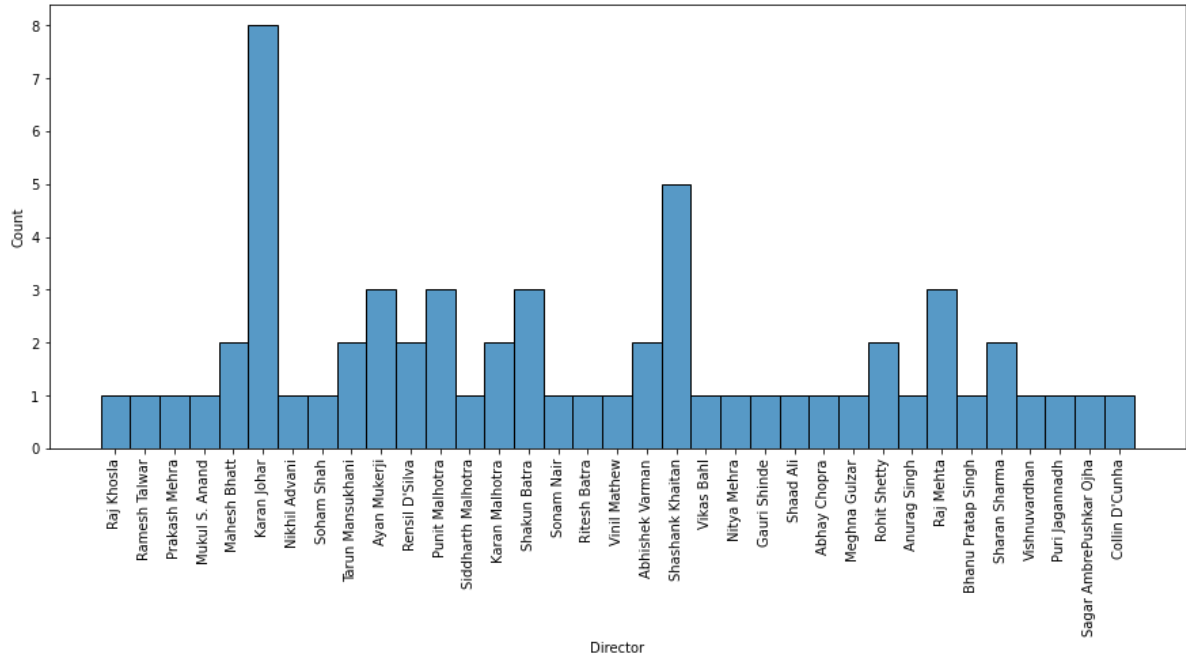
In [43]:

```
plt.figure(figsize=(15,6))
sns.countplot('Director', data = dharma_data, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



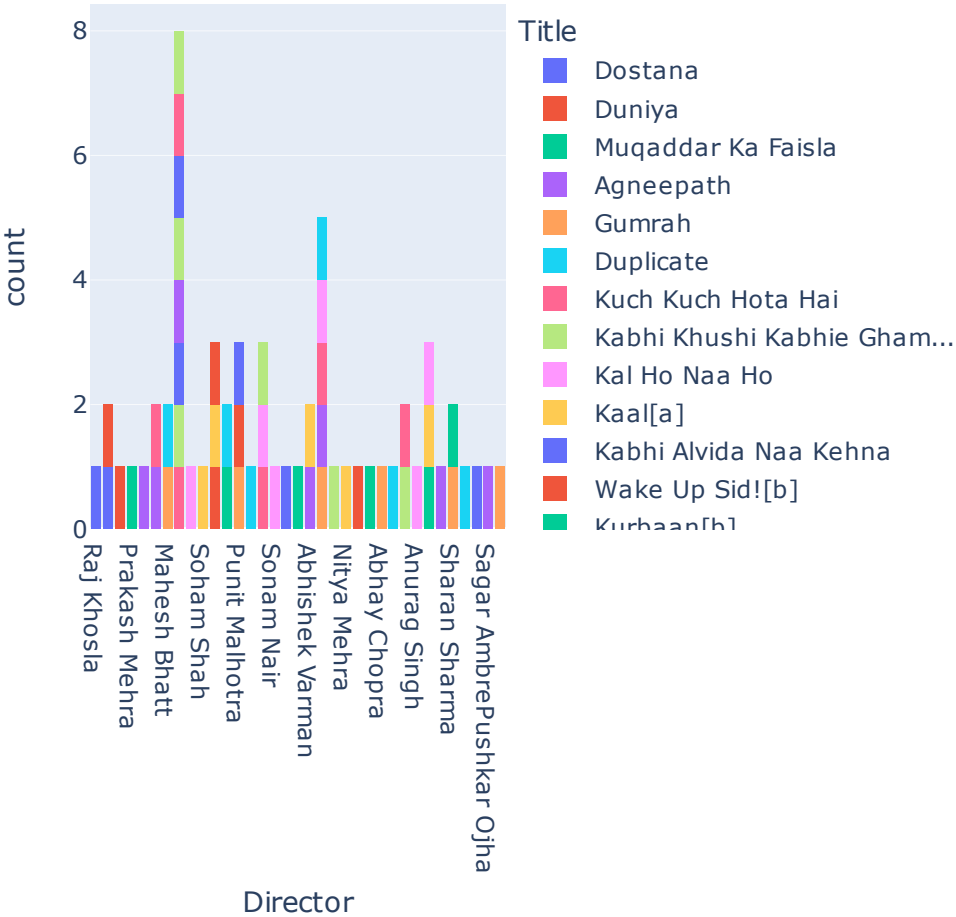
In [40]:

```
plt.figure(figsize=(15,6))
sns.histplot(dharma_data.Director, bins = 60)
plt.xticks(rotation = 90)
plt.show()
```



In [41]:

```
fig2 = px.histogram(dharma_data, x = 'Director', color = 'Title')
fig2.show()
```



In [44]:



```
dharma_data.Producers.unique()
```

Out[44]:

```
array(['Yash Johar', 'Yash Johar,Karan Johar',
      'Karan Johar,Shah Rukh Khan', 'Hiroo Johar',
      'Karan Johar,Hiroo Johar', 'Hiroo Johar,Gauri Khan',
      'Karan Johar,Hiroo Johar,Ronnie Screwvala',
      'Guneet Monga,Anurag Kashyap,Arun Rangachari',
      'Karan Johar,Vikas Bahl,Vikramaditya Motwane,Anurag Kashyap',
      'Karan Johar,Sajid Nadiadwala', 'Karan JoharHiroo Johar',
      'Karan Johar,Hiroo Johar,Apoorva Mehta',
      'Karan Johar,Farhan Akhtar, Ritesh Sidhwani',
      'Karan JoharGauri Khan, Gauri Shinde', 'Karan JoharMani Ratnam',
      'Karan Johar,Hiroo Johar, Apoorva Mehta',
      'Karan Johar, Hiroo Johar,Gauri Khan, Renu Chopra',
      'Karan Johar,Hiroo Johar, Vineet Jain, Apoorva Mehta',
      'Karan Johar, Hiroo Johar,Rohit Shetty, Apoorva Mehta',
      'Karan Johar,Akshay Kumar',
      'Karan Johar,Aruna Bhatia,Hiroo Yash Johar,Apoorva Mehta,Shashank K
haitan',
      'Karan Johar,Hiroo Johar,Apoorva Mehta,Shashank Khaitan',
      'Karan Johar,Hiroo Johar,Apoorva Mehta,Zee Studios',
      'Karan Johar,Hiroo Johar,Apoorva Mehta,Shabbir Boxwala,Ajay Shah,Hi
manshu Gandhi',
      'Karan Johar,Rohit Shetty,Aruna Bhatia',
      'Karan Johar,Hiroo Johar,Apoorva Mehta,Shakun Batra',
      'Karan Johar,Hiroo Yash Johar,Apoorva Mehta',
      'Karan Johar,Hiroo Johar,Apoorva Mehta,Puri Jagannadh,Charmme Kau
r',
      'Karan Johar,Hiroo Yash Johar,Apoorva Mehta,Shashank Khaitan',
      'Karan JoharHiroo Johar,Apoorva Mehta',
      'Karan JoharHiroo Yash Johar,Apoorva Mehta',
      'Karan JoharPrithviraj SukumaranAruna Bhatia,Hiroo Yash Johar,Apoor
va Mehta'],
      dtype=object)
```

In [45]:



```
dharma_data.Producers.value_counts()
```

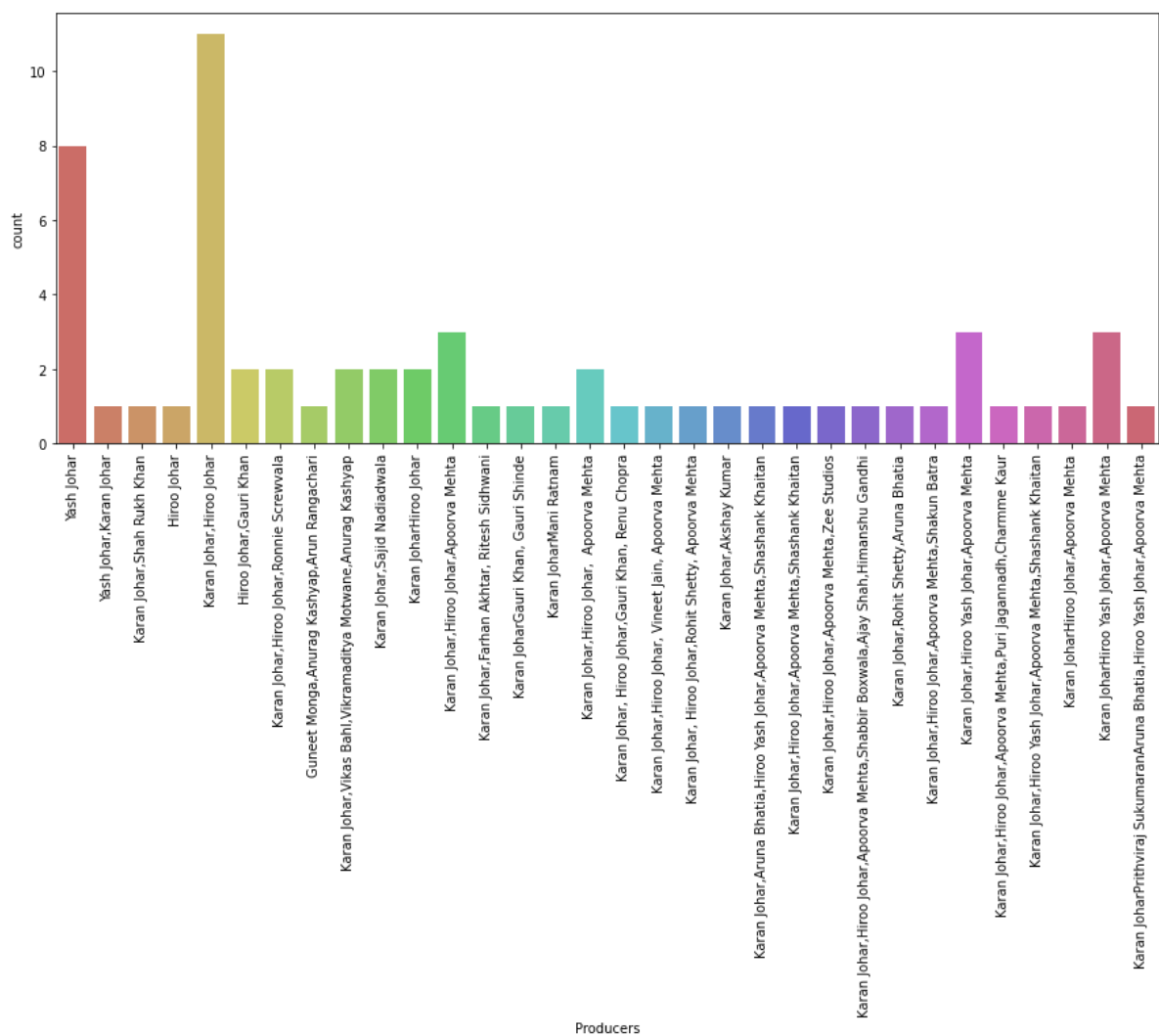
Out[45]:

```
Karan Johar,Hiroo Johar
11
Yash Johar
8
Karan JoharHiroo Yash Johar,Apoorva Mehta
3
Karan Johar,Hiroo Yash Johar,Apoorva Mehta
3
Karan Johar,Hiroo Johar,Apoorva Mehta
3
Karan Johar,Vikas Bahl,Vikramaditya Motwane,Anurag Kashyap
2
Karan Johar,Hiroo Johar, Apoorva Mehta
2
Karan Johar,Sajid Nadiadwala
2
Karan JoharHiroo Johar
2
Karan Johar,Hiroo Johar,Ronnie Screwvala
2
Hiroo Johar,Gauri Khan
2
Karan Johar,Hiroo Yash Johar,Apoorva Mehta,Shashank Khaitan
1
Karan Johar,Aruna Bhatia,Hiroo Yash Johar,Apoorva Mehta,Shashank Khaitan
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Puri Jagannadh,Charmme Kaur
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Shakun Batra
1
Karan JoharHiroo Johar,Apoorva Mehta
1
Karan Johar,Rohit Shetty,Aruna Bhatia
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Shabbir Boxwala,Ajay Shah,Himanshu G
andhi 1
Karan Johar,Hiroo Johar,Apoorva Mehta,Zee Studios
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Shashank Khaitan
1
Karan Johar, Hiroo Johar,Gauri Khan, Renu Chopra
1
Karan Johar,Akshay Kumar
1
Karan Johar, Hiroo Johar,Rohit Shetty, Apoorva Mehta
1
Karan Johar,Hiroo Johar, Vineet Jain, Apoorva Mehta
1
Yash Johar,Karan Johar
1
Karan JoharMani Ratnam
1
Karan JoharGauri Khan, Gauri Shinde
1
```

Karan Johar,Farhan Akhtar, Ritesh Sidhwani  
1  
Guneet Monga,Anurag Kashyap,Arun Rangachari  
1  
Hiroo Johar  
1  
Karan Johar,Shah Rukh Khan  
1  
Karan JoharPrithviraj SukumaranAruna Bhatia,Hiroo Yash Johar,Apoorva Mehta  
1  
Name: Producers, dtype: int64

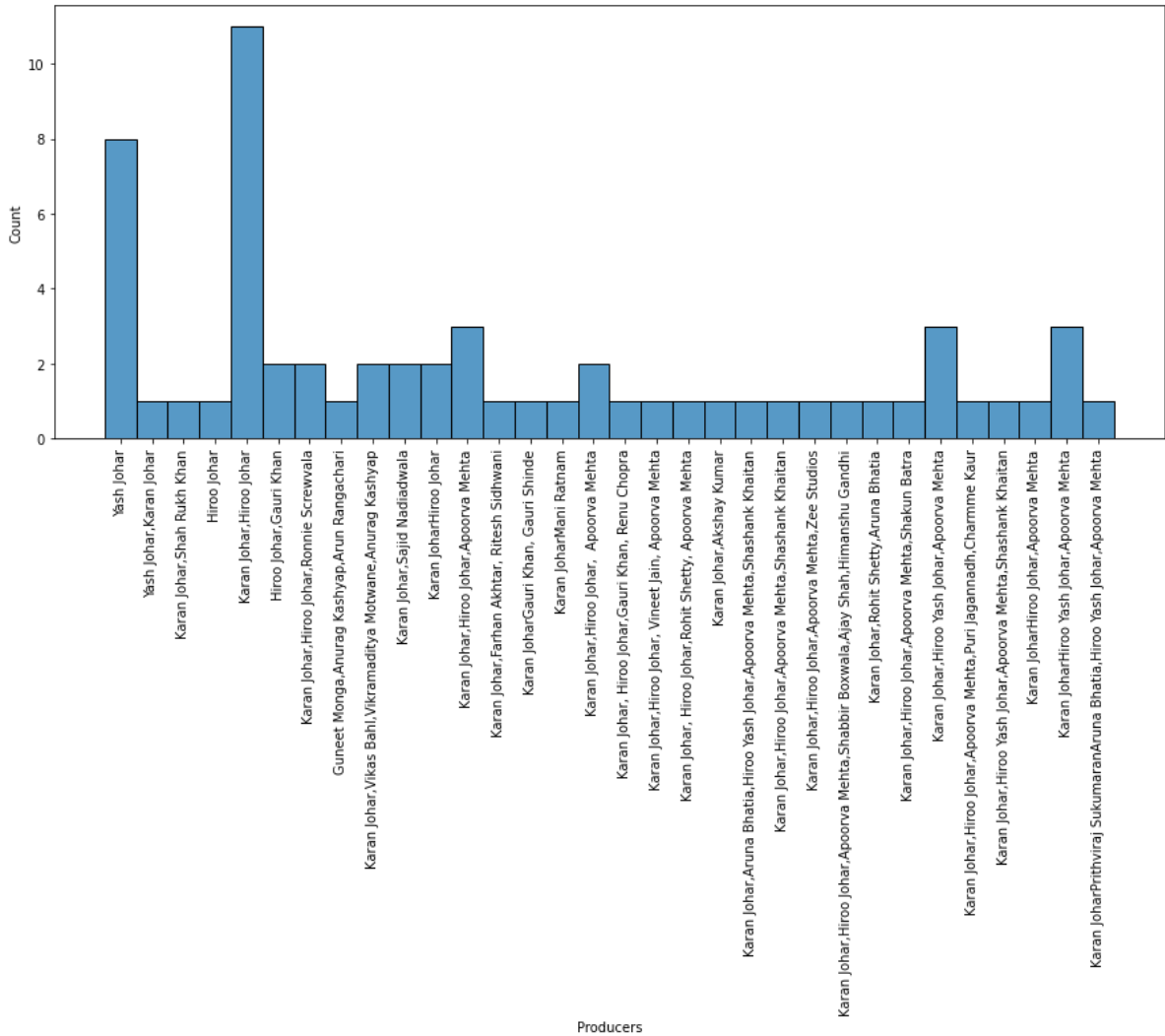
In [46]: ▶

```
plt.figure(figsize=(15,6))
sns.countplot('Producers', data = dharma_data, palette='hls')
plt.xticks(rotation = 90)
plt.show()
```



In [47]:

```
plt.figure(figsize=(15,6))
sns.histplot(dharma_data.Producers, bins = 60)
plt.xticks(rotation = 90)
plt.show()
```



In [49]:

```
fig3 = px.histogram(dharma_data, x = 'Producers', color = 'Title')
fig3.show()
```



<Figure size 1080x432 with 0 Axes>