Dharma Productions Movie Analysis

June 2, 2022

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```python
d_m_df =pd.read_csv('dharma.csv')
```

```python
# checking what the dataset contain
d_m_df.head()
```

```
                 Title  Year         Director Producer(s)  \
0              Dostana  1980      Raj Khosla   Yash Johar
1               Duniya  1984   Ramesh Talwar   Yash Johar
2   Muqaddar Ka Faisla  1987   Prakash Mehra   Yash Johar
3            Agneepath  1990  Mukul S. Anand   Yash Johar
4               Gumrah  1993    Mahesh Bhatt   Yash Johar


                                             Cast  \
0     Amitabh Bachchan, Shatrughan Sinha, Zeenat Aman
1   Dilip Kumar, Rishi Kapoor, Amrita Singh, Saira...
2   Raaj Kumar, Raakhee, Raj Babbar, Meenakshi Ses...
3   Amitabh Bachchan, Mithun Chakraborty, Danny De...
4                   Sridevi, Sanjay Dutt, Rahul Roy


                                       Synopsis  Ref.
0  A police officer and a lawyer are best friends...  [18]
1  An honest man is wrongly imprisoned by crimina...  [19]
2  A Hindu priest is wrongly imprisoned for rape...  [20]
3  A common man seeks revenge from an unscrupulou...  [21]
4  An aspiring singer is wrongly accused of drug ...  [22]
```

```python
d_m_df.tail()
```

```
                          Title  Year          Director  \
56                     Dostana 2  2022   Collin D'Cunha
57                       Bedhadak  2022  Shashank Khaitan
58  Rocky Aur Rani Ki Prem Kahani  2023       Karan Johar
```

```
59   Anand Tiwari's untitled next  2023        Karan Johar
60                        Selfiee  2023          Raj Mehta

                                  Producer(s)  \
56           Karan JoharHiroo Johar,Apoorva Mehta
57       Karan JoharHiroo Yash Johar,Apoorva Mehta
58       Karan JoharHiroo Yash Johar,Apoorva Mehta
59       Karan JoharHiroo Yash Johar,Apoorva Mehta
60  Karan JoharPrithviraj SukumaranAruna Bhatia,Hi...

                                           Cast Synopsis        Ref.
56               Janhvi Kapoor, Laksh Lalwani      TBA       [111]
57  Laksh Lalwani, Gurfateh Singh Pirzada, Shanaya...    TBA  [112][113]
58  Ranveer Singh, Alia Bhatt, Dharmendra, Shabana...    TBA       [114]
59          Vicky Kaushal, Tripti Dimri, Ammy Virk      TBA       [115]
60  Akshay Kumar, Emraan Hashmi, Nushrratt Bharucc...    TBA  [116][117]
```

[7]: `d_m_df.columns`

[7]:
```
Index(['Title', 'Year', 'Director', 'Producer(s)', 'Cast', 'Synopsis', 'Ref.'],
dtype='object')
```

[8]: `d_m_df.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
```

[9]: `d_m_df.describe()`

[9]:
```
               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
```

[10]: ```python
#checking for duplicates
d_m_df.duplicated().sum()
```

[10]: 0

[11]: ```python
#checking for null values in the dataset
d_m_df.isnull().sum()
```

[11]: ```
Title   0
Year   0
Director   0
Producer(s)   0
Cast   0
Synopsis   0
Ref.   0
dtype: int64
```

[12]: ```python
#looking at the columns after checking for null values on the dataset the
 ↪producers column needs renaming
d_m_df.rename(columns = {'Producer(s)':'Producers'}, inplace = True)
```

[13]: ```python
so from the look of the dataset it is okay but we still need to drop some
 ↪columns which are not useful e.g Ref column
d_m_df= d_m_df.drop('Ref.', axis = 1)
d_m_df.head()
```

[13]: ```
                  Title  Year         Director   Producers  \
0               Dostana  1980       Raj Khosla  Yash Johar
1                Duniya  1984    Ramesh Talwar  Yash Johar
2     Muqaddar Ka Faisla  1987    Prakash Mehra  Yash Johar
3             Agneepath  1990  Mukul S. Anand  Yash Johar
4                Gumrah  1993     Mahesh Bhatt  Yash Johar

                                            Cast  \
0     Amitabh Bachchan, Shatrughan Sinha, Zeenat Aman
1   Dilip Kumar, Rishi Kapoor, Amrita Singh, Saira...
2   Raaj Kumar, Raakhee, Raj Babbar, Meenakshi Ses...
3   Amitabh Bachchan, Mithun Chakraborty, Danny De...
4                 Sridevi, Sanjay Dutt, Rahul Roy

                                        Synopsis
0  A police officer and a lawyer are best friends...
1  An honest man is wrongly imprisoned by crimina...
2  A Hindu priest is wrongly imprisoned for rape...
```

```
3  A common man seeks revenge from an unscrupulou...
4  An aspiring singer is wrongly accused of drug ...
```

```
[14]:  # time to check for the unique number of columns we have
       d_m_df.nunique()
```

```
[14]:  Title       59
       Year        25
       Director    35
       Producers   32
       Cast        60
       Synopsis    50
       dtype: int64
```
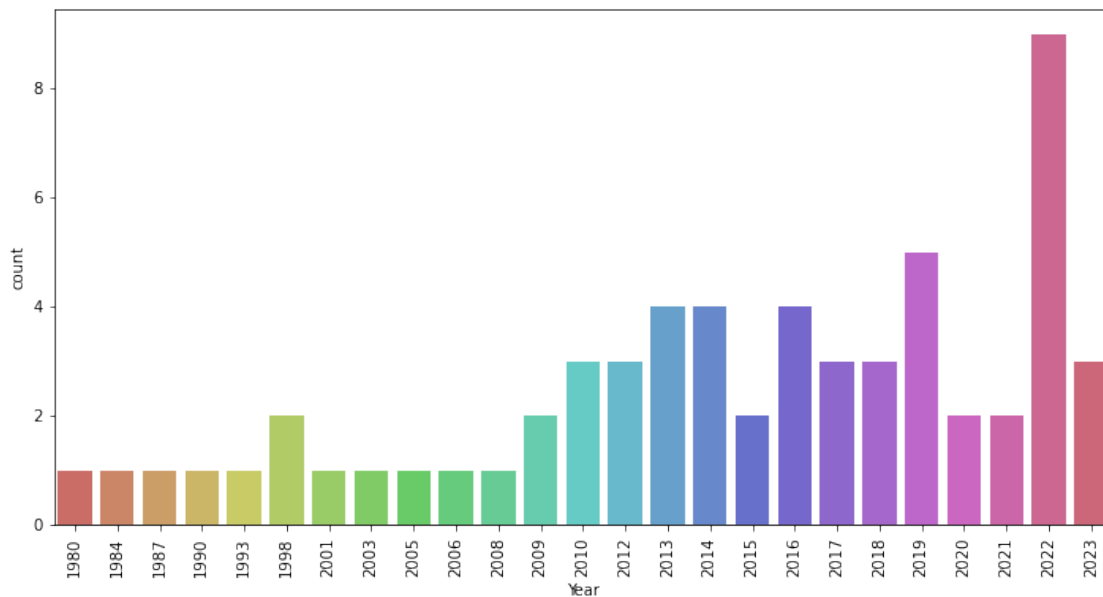
```
[15]:  # so lets check the value count of the years
       d_m_df.Year.value_counts()
```

```
[15]:  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
```

```
[27]:  #now lets do some visualisation by plotting graph
       plt.figure(figsize = (12,6))
       sns.countplot('Year', data = d_m_df, palette = 'hls')
```

```
plt.xticks(rotation = 90)
plt.show()
```


[19]:
```
#so here i tried getting the value counts of unique directors
d_m_df.Director.value_counts().unique()
```

[19]: `array([8, 5, 3, 2, 1], dtype=int64)`

[20]:
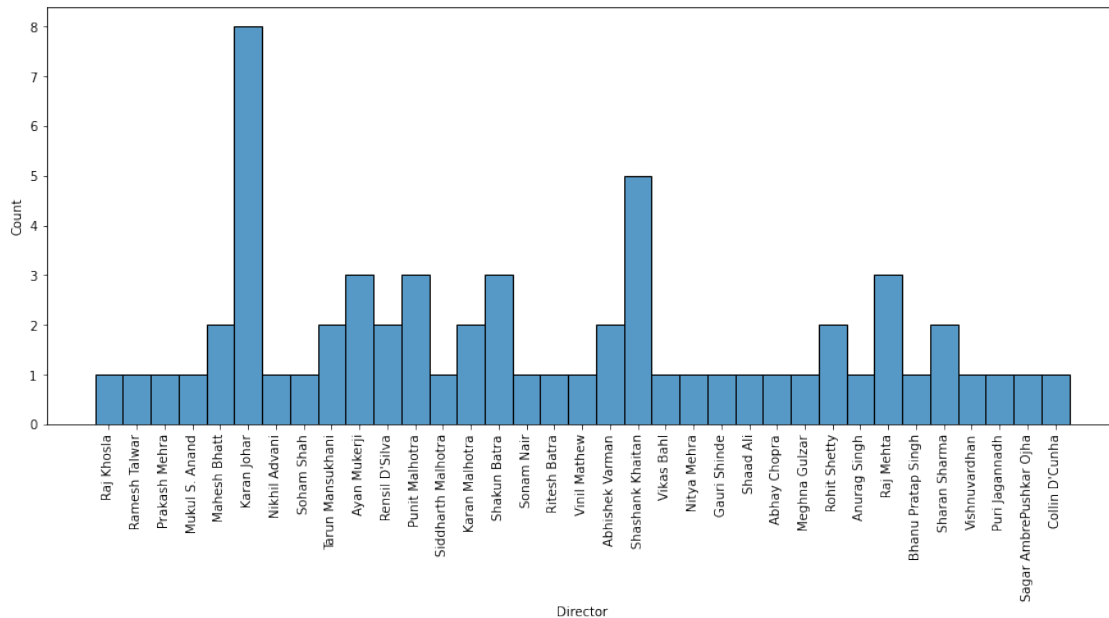```
#lets just see the value counts of the directors normally
d_m_df.Director.value_counts()
```

[20]: 
```
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
```

[24]:
```python
# so now lets do some visualisation of our directors
plt.figure(figsize = (15,6))
sns.histplot(d_m_df.Director, bins = 60)
plt.xticks(rotation = 90)
plt.show()
```

```python
# now lets check our producers and their counts too
d_m_df.Producers.value_counts().unique()
```

```
array([11,  8,  3,  2,  1], dtype=int64)
```

```python
d_m_df.Producers.value_counts()
```

```
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 Gandhi
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Zee Studios
1
Karan Johar,Hiroo Johar,Apoorva Mehta,Shashank Khaitan
1
```
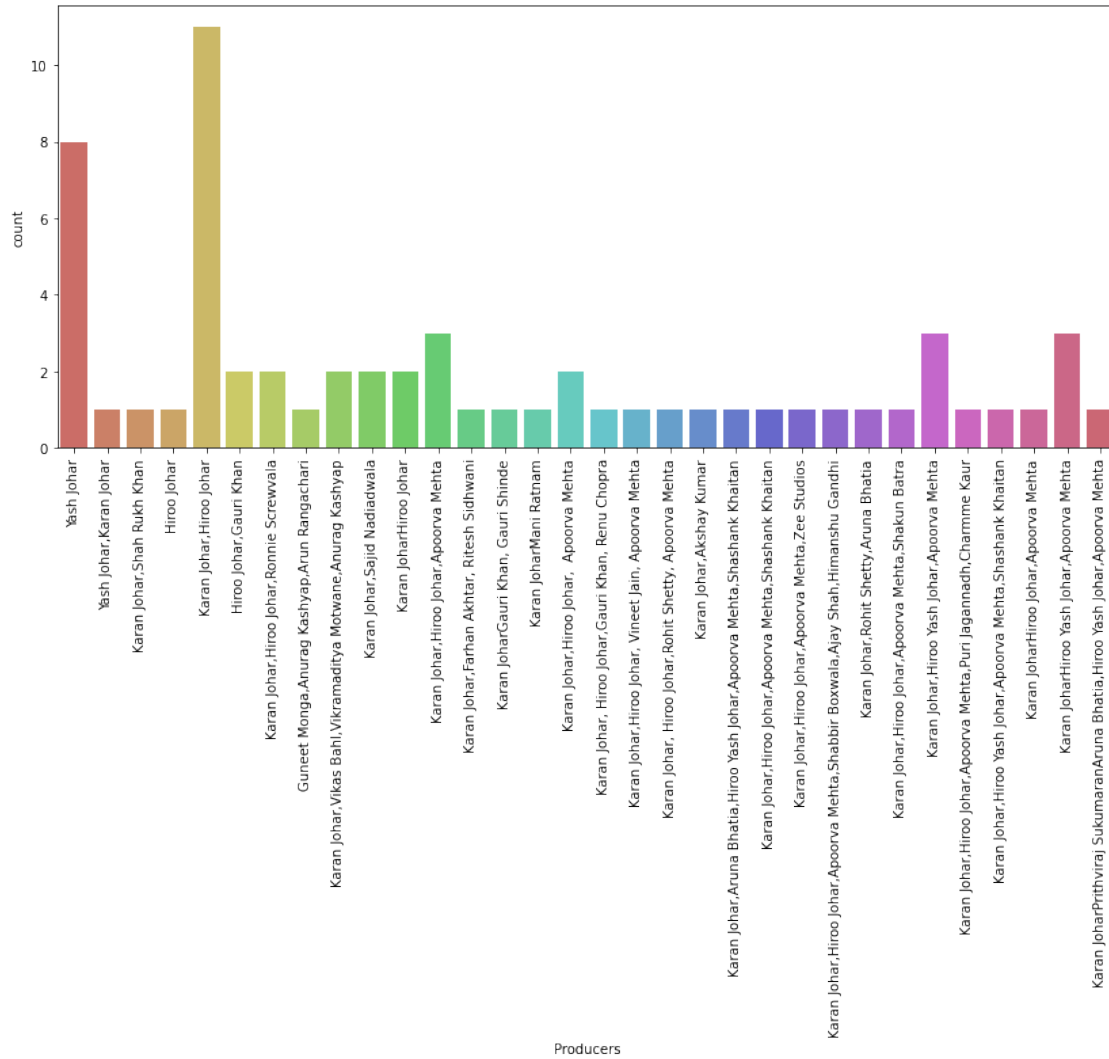
```
Karan Johar, Hiroo Johar,Gauri Khan, Renu Chopra
1
Karan Johar,Akshay Kumar
1
Karan Johar, Hiroo Johar,Rohit Shetty, Apoorva Mehta
1
Karan Johar,Hiroo Johar, Vineet Jain, Apoorva Mehta
1
Yash Johar,Karan Johar
1
Karan JoharMani Ratnam
1
Karan JoharGauri Khan, Gauri Shinde
1
Karan Johar,Farhan Akhtar, 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
```

[32]:
```python
#let's do some Visualisation for the producers
plt.figure(figsize = (14,6))
sns.countplot('Producers', data = d_m_df, palette = 'hls')
plt.xticks(rotation = 90)
plt.show()
```

Producers

```
[1]: jupyter nbconvert --to webpdf --allow-chromium-download Dharma Productions Movie␣
     ↪Analysis.ipynb
```

```
  Input In [1]
    jupyter nbconvert --to webpdf --allow-chromium-download Dharma Productions␣
↪Movie Analysis.ipynb
                 ^
SyntaxError: invalid syntax
```