

In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, StratifiedKFold
```

In [3]:

```
df = pd.read_csv("IMDb database.csv")
df
```

Out[3]:

	Title	Release Date	Color	Genre	Language	Country	Rating	Lead Actor	Director Name
0	Over the Hill to the Poorhouse	9-15-1920	Black and White	Crime	English	USA	Not Rated	Stephen Carr	Harry F. Millarde
1	Metropolis	1-26-1927	Black and White	Drama	German	Germany	Not Rated	Brigitte Helm	Fritz Lang
2	The Broadway Melody	11-11-1929	Black and White	Musical	English	USA	Passed	Anita Page	Harry Beaumont
3	42nd Street	8-29-1933	Black and White	Comedy	English	USA	Unrated	Ginger Rogers	Lloyd Bacon
4	Top Hat	4-15-1935	Black and White	Comedy	English	USA	Approved	Ginger Rogers	Mark Sandrich
...
3728	Sugar Rush	11-11-2016	Color	Comedy	English	USA	PG	Karen Trout	Humberto Guava
3729	Why is Grandma Laughing?	8-29-2016	Color	Mystery	English	USA	PG-13	Yulia Velasquez	Purd Haply
3730	Luke Warm Soup	4-15-2016	Color	Horror	English	USA	R	Skeeter McGavin	Nick Winkle
3731	Clown Parade III	6-1-2016	Color	Comedy	English	USA	PG	Stan Pancake	Todd Waffle
3732	Gene's Genes	8-11-2016	Color	Sci-Fi	Spanish	Spain	PG	Randy Killnose	August Clementine

3733 rows × 19 columns



In [4]:

```
df = df.dropna()  
df = df.reset_index(drop=True)
```

In [5]:

```
df1 = df.copy()  
df1["color_bool"] = df1["Color"].map({'Color':True , 'Black and White':False})
```

In [6]:

```
df1.color_bool = df1.color_bool.fillna(False)
```

In [7]:

```
df1["color_bool"].isnull().sum()
```

Out[7]:

0

In [8]:

```
X = df1.drop(["color_bool"], axis=1)  
Y = df1.color_bool  
X_Train, X_Test, y_train, y_test = train_test_split(X, Y, test_size = 0.2, random_state=42,
```

In [9]:

X_Train

Out[9]:

	Title	Release Date	Color	Genre	Language	Country	Rating	Lead Actor	Director Name
1697	A Home at the End of the World	5-13-2004	Color	Drama	English	USA	R	Robin Wright	Michael Mayer
3124	Fun Size	10-24-2012	Color	Adventure	English	USA	PG-13	Thomas McDonell	Josh Schwartz
3325	Oblivion	10-4-2013	Color	Action	English	USA	PG-13	Morgan Freeman	Joseph Kosinski
1190	All the Queen's Men	12-18-2001	Color	Action	English	Germany	PG-13	Eddie Izzard	Stefan Ruzowitzky
2288	Hostel: Part II	7-13-2007	Color	Horror	English	USA	R	Jay Hernandez	Eli Roth
...
2595	Brothers	4-5-2009	Color	Drama	English	USA	R	Natalie Portman	Jim Sheridan
572	Executive Decision	3-29-1996	Color	Action	English	USA	R	Oliver Platt	Stuart Baird
1617	Mona Lisa Smile	8-12-2003	Color	Drama	English	USA	PG-13	Julia Roberts	Mike Newell
2133	Littleman	6-30-2006	Color	Comedy	English	USA	PG-13	Chazz Palminteri	Keenen Ivory Wayans
612	Sling Blade	9-22-1996	Color	Drama	English	USA	R	Robert Duvall	Billy Bob Thornton

2944 rows × 19 columns



In [10]:

X_Test

Out[10]:

	Title	Release Date	Color	Genre	Language	Country	Rating	Lead Actor	Director Name
3099	Blue Like Jazz	6-12-2012	Color	Comedy	English	USA	PG-13	Jason Marsden	Steve Taylor
3473	Mr. Peabody & Sherman	2-27-2014	Color	Adventure	English	USA	PG	Ty Burrell	Rob Minkoff
2459	Mirrors	7-30-2008	Color	Horror	English	USA	R	Jason Flemyng	Alexandre Aja
1773	Little Black Book	12-14-2004	Color	Comedy	English	USA	PG-13	Holly Hunter	Nick Hurran
2919	A Very Harold & Kumar 3D Christmas	1-14-2011	Color	Adventure	English	USA	R	Patton Oswalt	Todd Strauss-Schulson
...
542	The Quick and the Dead	4-11-1995	Black and White	Action	English	Japan	R	Leonardo DiCaprio	Sam Raimi
3552	Aloha	2-7-2015	Color	Comedy	English	USA	PG-13	Emma Stone	Cameron Crowe
317	Mo' Better Blues	3-3-1990	Color	Drama	English	USA	R	Denzel Washington	Spike Lee
1148	The Ladies Man	10-27-2000	Color	Comedy	English	USA	R	Will Ferrell	Reginald Hudlin
1712	Beyond the Sea	1-7-2004	Black and White	Biography	English	USA	PG-13	Kevin Spacey	Kevin Spacey

736 rows × 19 columns



In [11]:

```
y_train
```

Out[11]:

```
1697    True
3124    True
3325    True
1190    True
2288    True
...
2595    True
572     True
1617    True
2133    True
612     True
Name: color_bool, Length: 2944, dtype: bool
```

In [12]:

```
y_test
```

Out[12]:

```
3099    True
3473    True
2459    True
1773    True
2919    True
...
542     False
3552    True
317     True
1148    True
1712    False
Name: color_bool, Length: 736, dtype: bool
```

In []: