

```
import regex
import pandas as pd
import numpy as np
```

```
from google.colab import drive
drive.mount('/content/drive')
```

↳ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.r

```
cd /content/drive/MyDrive/Dataset
```

↳ /content/drive/MyDrive/Dataset

```
file_path = '/content/drive/MyDrive/Dataset/movie_data.csv'
df = pd.read_csv(file_path)
```

```
print(df['Month'].mean())
print(df['Month'].median())
print(df['Month'].mode())
```

↳ 7.259583789704272
7.0
0 7
Name: Month, dtype: int64

```
# Convert the 'genre1' column to a numerical representation before calculating the mean.
# One way to do this is to assign a unique numerical value to each genre.
```

```
# Create a dictionary mapping genres to numerical values
genre_mapping = {genre: i for i, genre in enumerate(df['genre1'].unique())}
```

```
# Create a new numerical column based on the mapping
df['genre1_numeric'] = df['genre1'].map(genre_mapping)
```

```
# Calculate the mean of the new numerical column
print(df['genre1_numeric'].mean())
print(df['genre1_numeric'].median())
print(df['genre1_numeric'].mode())
```

↳ 1.9353778751369113
1.0
0 0
Name: genre1_numeric, dtype: int64

```
print(df.columns)
```

```
print(df["Month"].corr(df["Year"]))
```

```

➞ Index(['RowNumber', 'Title', 'Movie Info', 'Distributor', 'Release Date',
        'Genre', 'Movie Runtime', 'License', 'WeekNum', 'Month', 'Year',
        'Domestic_Sales_mil_dollars', 'International_Sales_mil_dollars',
        'World_Sales_mil_dollars', 'genre1', 'genre2', 'genre3', 'genre4'],
        dtype='object')
-0.05885893774664353

```

```

print(df["Month"].unique())
print(df["Year"].unique())
print(df["genre1"].unique())
print(df["genre2"].unique())
print(df["genre3"].nunique())

```

```
print(df["genre3"].value_counts())
```

```

➞ [12  4  7  6  3 11  5  2 10  9  8  1]
[2015 2019 2009 2018 2021 1997 2012 2017 2008 2016 1999 1977 2004 1982
 2013 2006 1994 2010 2002 1993 2011 2003 2005 2014 2007 2001 1983 1996
 1980 1990 1975 2000 1989 1981 1984 1973 1991 1992 1998 1985 2020 1978
 1937 1995 1988 1987 1972 1986 1955 1953 1979 1970]
['Action' 'Drama' 'Adventure' 'Crime' 'Horror' 'Comedy' 'Animation'
 'Biography' 'Mystery' 'Documentary' 'Fantasy']
['Adventure' 'Romance' 'Animation' 'Crime' 'Family' 'Fantasy' 'Drama'
 'Comedy' 'Thriller' 'Biography' 'Horror' 'Mystery' 'Sci-Fi' 'War' 'Music'
 'Musical' 'Western' 'Sport' 'History']
21
genre3
Comedy          241
Sci-Fi          110
Thriller        103
Drama           95
Fantasy         78
Family          48
Romance         43
Crime           39
Action          36
Mystery         24
Animation       23
Horror          19
History         13
Biography        8
Sport           8
War             7
Adventure        5
Musical          5
Music            5
Western          2
Documentary      1
Name: count, dtype: int64

```

```
print(df["genre2"].value_counts())
```

```

➞ genre2
Adventure      247
Drama          146
Animation      111
Comedy         110

```

Crime	55
Romance	48
Mystery	42
Family	39
Thriller	27
Sci-Fi	20
Fantasy	17
Horror	13
Music	10
Sport	7
Biography	7
Musical	4
Western	4
History	4
War	2

Name: count, dtype: int64

```
print(df.sort_values(by=["Year"], ascending=False).head(15))
```



191

288.500000

501.138437

ACTION

	genre2	genre3	genre4
247	Adventure	Sci-Fi	Action
558	Adventure	Comedy	Fantasy
334	Adventure	Thriller	Action
702	Sci-Fi	Thriller	Action
600	Animation	Comedy	Family
777	Thriller	Horror	Horror
505	Comedy	Fantasy	Sci-Fi
164	Adventure	Fantasy	Sci-Fi
316	Adventure	Fantasy	Sci-Fi
770	Comedy	Drama	Family
286	Crime	Thriller	Action
5	Adventure	Fantasy	Sci-Fi
830	Comedy	Crime	Drama
631	Adventure	Drama	Sci-Fi
191	Adventure	Sci-Fi	Thriller

```
print(df.sort_values(by=["Year"], ascending=False))
```



898	Twentieth Century Fox	1970-06-01
761	Walt Disney Studios Motion Pictures	1955-07-01
821	Walt Disney Studios Motion Pictures	1953-07-01
243	Walt Disney Studios Motion Pictures	1937-07-01

898	81.6000000	81.6000000	Comedy
761	0.331068	93.933394	Adventure
821	87.404651	87.404651	Adventure
243	184.925486	184.925486	Adventure

	genre2	genre3	genre4
247	Adventure	Sci-Fi	Action
558	Adventure	Comedy	Fantasy
334	Adventure	Thriller	Action
702	Sci-Fi	Thriller	Action
600	Animation	Comedy	Family
..
439	Drama	Crime	Crime
898	Drama	War	Comedy
761	Animation	Comedy	Family
821	Animation	Family	Fantasy
243	Animation	Family	Fantasy

[913 rows x 18 columns]

```
print(df.nlargest(10, ["Year"]))
```

5	Sony Pictures Entertainment (SPE)	2021-07-01
164	Walt Disney Studios Motion Pictures	2021-07-01
191	Sony Pictures Entertainment (SPE)	2021-07-01
247	Walt Disney Studios Motion Pictures	2021-07-01
286	Universal Pictures	2021-06-01
316	Walt Disney Studios Motion Pictures	2021-07-01
334	Metro-Goldwyn-Mayer (MGM)	2021-11-01
505	Sony Pictures Entertainment (SPE)	2021-07-01
558	Walt Disney Studios Motion Pictures	2021-07-01
600	Universal Pictures	2021-06-01

Genre Movie Runtime License \

164	207.689718	432.233010	ACTION
191	288.500000	501.138437	Action
247	195.979696	379.631351	Action
286	553.223556	726.229501	Action
316	237.027887	401.847482	Action
334	613.262000	774.034007	Action
505	66.100000	191.378330	Adventure
558	103.901930	220.889446	Action
600	82.129000	192.669715	Adventure

	genre2	genre3	genre4
5	Adventure	Fantasy	Sci-Fi
164	Adventure	Fantasy	Sci-Fi
191	Adventure	Sci-Fi	Thriller
247	Adventure	Sci-Fi	Action
286	Crime	Thriller	Action
316	Adventure	Fantasy	Sci-Fi
334	Adventure	Thriller	Action
505	Comedy	Fantasy	Sci-Fi
558	Adventure	Comedy	Fantasy
600	Animation	Comedy	Family

```
print(df.nsmallest(10, ["Year"]))
```

243	Walt Disney Studios Motion Pictures	1937-07-01
821	Walt Disney Studios Motion Pictures	1953-07-01
761	Walt Disney Studios Motion Pictures	1955-07-01
898	Twentieth Century Fox	1970-06-01
439	Paramount Pictures	1972-03-15
152	Warner Bros.	1973-07-01
117	Universal Pictures	1975-06-20
19	Twentieth Century Fox	1977-05-25
752	Paramount Pictures	1977-07-01
232	Paramount Pictures	1978-06-16

243	184.925486	184.925486	Adventure
821	87.404651	87.404651	Adventure
761	0.331068	93.933394	Adventure
898	81.600000	81.600000	Comedy
439	111.154575	246.120986	Crime
152	136.017945	441.306145	Horror
117	210.653000	471.411300	Adventure
19	195.751992	775.398007	Action
752	142.900000	237.113184	Drama
232	206.200000	396.271103	Comedy

	genre2	genre3	genre4
243	Animation	Family	Fantasy
821	Animation	Family	Fantasy
761	Animation	Comedy	Family
898	Drama	War	Comedy
439	Drama	Crime	Crime
152	Horror	Horror	Horror
117	Thriller	Adventure	Adventure
19	Adventure	Fantasy	Sci-Fi
752	Music	Drama	Drama
232	Musical	Romance	Comedy

```
print(df.nlargest(20, ["Month"]))
```



	genre2	genre3	genre4
0	Adventure	Sci-Fi	Action
2	Adventure	Fantasy	Sci-Fi
6	Romance	Drama	Drama
9	Adventure	Fantasy	Sci-Fi
13	Adventure	Sci-Fi	Action
14	Adventure	Fantasy	Sci-Fi
36	Adventure	Comedy	Fantasy
45	Adventure	Drama	Fantasy
56	Biography	Drama	War
57	Adventure	Drama	Fantasy
73	Adventure	Comedy	Fantasy
78	Adventure	Drama	Fantasy
98	Family	Fantasy	Adventure
106	Romance	Comedy	Comedy
110	Comedy	Family	Musical
123	Drama	Sci-Fi	Thriller
151	Drama	Romance	Adventure
169	Animation	Comedy	Drama
173	Adventure	Drama	Romance
178	Animation	Comedy	Drama

```
# Check if 'genre' is in the DataFrame columns
print(df.columns)
```

```
if 'genre' not in df.columns:
    print("Column 'genre' not found in the DataFrame. Please add it or adjust your
else:
    print(df.groupby("genre")[val].sum())
    # The 'for' loop should be at the same indentation level as the print statement
    for genre, group in df.groupby("genre"):
        print(f"Genre: {genre}")
        print(group)
```

```
➞ Index(['RowNumber', 'Title', 'Movie Info', 'Distributor', 'Release Date',
        'Genre', 'Movie Runtime', 'License', 'WeekNum', 'Month', 'Year',
        'Domestic_Sales_mil_dollars', 'International_Sales_mil_dollars',
        'World_Sales_mil_dollars', 'genre1', 'genre2', 'genre3', 'genre4'],
        dtype='object')
Column 'genre' not found in the DataFrame. Please add it or adjust your code.
```


