

Step1:Import Neccessary Libraries

```
In [1]: import pandas as pd  
import numpy as np
```

```
In [2]: import matplotlib.pyplot as plt  
import seaborn as sns
```

Step2: Import Datasets

```
In [3]: df=pd.read_csv("disney_movies.csv")
df
```

Out[3]:

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	1937-12-21	Musical	G	184925485	5228953251
1	Pinocchio	1940-02-09	Adventure	G	84300000	2188229052
2	Fantasia	1940-11-13	Musical	G	83320000	2187090808
3	Song of the South	1946-11-12	Adventure	G	65000000	1078510579
4	Cinderella	1950-02-15	Drama	G	85000000	920608730
5	20,000 Leagues Under the Sea	1954-12-23	Adventure	NaN	28200000	528279994
6	Lady and the Tramp	1955-06-22	Drama	G	93600000	1236035515
7	Sleeping Beauty	1959-01-29	Drama	NaN	9464608	21505832
8	101 Dalmatians	1961-01-25	Comedy	G	153000000	1362870985
9	The Absent Minded Professor	1961-03-16	Comedy	NaN	25381407	310094574
10	Babes in Toyland	1961-12-14	Musical	G	10218316	124841160
11	Bon Voyage!	1962-05-17	Comedy	Not Rated	9230769	109581646
12	The Sword in the Stone	1963-12-25	Adventure	NaN	22182353	153870834
13	The Jungle Book	1967-10-18	Musical	Not Rated	141843000	789612346
14	Blackbeard's Ghost	1968-02-08	Comedy	NaN	21540050	138612686
15	The Aristocats	1970-04-24	Musical	G	55675257	255161499
16	The Boatniks	1970-07-01	Comedy	NaN	18607492	101200742
17	Bedknobs and Broomsticks	1971-10-07	Musical	NaN	17871174	91305448
18	The Apple Dumpling Gang	1975-07-01	Comedy	NaN	31916500	131246872
19	Freaky Friday	1977-01-21	Comedy	NaN	25942000	98067733
20	The Many Adventures of Winnie the Pooh	1977-03-11	NaN	NaN	0	0

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross
21	The Rescuers	1977-06-22	Adventure	NaN	48775599	159743914
22	Herbie Goes to Monte Carlo	1977-06-24	NaN	NaN	28000000	105847527
23	The Black Hole	1979-12-21	NaN	NaN	35841901	120377374
24	Midnight Madness	1980-02-08	NaN	NaN	2900000	9088096
25	The Last Flight of Noah's Ark	1980-06-25	NaN	NaN	11000000	34472116
26	The Devil and Max Devlin	1981-01-01	NaN	NaN	16000000	48517980
27	Amy	1981-03-20	Drama	NaN	0	0
28	The Fox and the Hound	1981-07-10	Comedy	NaN	43899231	133118889
29	Condorman	1981-08-07	Action	NaN	0	0
...
549	Guardians of the Galaxy	2014-08-01	Adventure	PG-13	333172112	343771168
550	The Hundred-Foot Journey	2014-08-08	Romantic Comedy	PG	54235441	55961409
551	Alexander and the Terrible, Horrible,...	2014-10-10	Comedy	PG	66954149	69055550
552	Big Hero 6	2014-11-07	Adventure	PG	222527828	229249222
553	Into the Woods	2014-12-25	Musical	PG	128002372	130894237
554	Strange Magic	2015-01-23	Adventure	PG	12429583	12429583
555	McFarland, USA	2015-02-20	Drama	PG	44480275	44480275
556	Cinderella	2015-03-13	Drama	PG	201151353	201151353
557	Monkey Kingdom	2015-04-17	Documentary	G	16432322	16432322
558	Avengers: Age of Ultron	2015-05-01	Action	PG-13	459005868	459005868
559	Tomorrowland	2015-05-22	Adventure	PG	93436322	93436322
560	Inside Out	2015-06-19	Adventure	PG	356461711	356461711
561	Ant-Man	2015-07-17	Action	PG-13	180202163	180202163
562	Bridge of Spies	2015-10-16	Thriller/Suspense	PG-13	72313754	72313754
563	The Good Dinosaur	2015-11-25	Adventure	PG	123087120	123087120

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross
564	Star Wars Ep. VII: The Force Awakens	2015-12-18	Adventure	PG-13	936662225	936662225
565	The Finest Hours	2016-01-29	Thriller/Suspense	PG	27569558	27569558
566	Zootopia	2016-03-04	Adventure	PG	341268248	341268248
567	The Jungle Book	2016-04-15	Adventure	PG	364001123	364001123
568	A Beautiful Planet	2016-04-29	Documentary	G	7895708	7895708
569	Captain America: Civil War	2016-05-06	Action	PG-13	408084349	408084349
570	Alice Through the Looking Glass	2016-05-27	Adventure	PG	77042381	77042381
571	Finding Dory	2016-06-17	Adventure	PG	486295561	486295561
572	The BFG	2016-07-01	Adventure	PG	55483770	55483770
573	Pete's Dragon	2016-08-12	Adventure	PG	76233151	76233151
574	The Light Between Oceans	2016-09-02	Drama	PG-13	12545979	12545979
575	Queen of Katwe	2016-09-23	Drama	PG	8874389	8874389
576	Doctor Strange	2016-11-04	Adventure	PG-13	232532923	232532923
577	Moana	2016-11-23	Adventure	PG	246082029	246082029
578	Rogue One: A Star Wars Story	2016-12-16	Adventure	PG-13	529483936	529483936

579 rows × 6 columns



In [4]: `df.head(11)`

Out[4]:

	movie_title	release_date	genre	mpaa_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	1937-12-21	Musical	G	184925485	5228953251
1	Pinocchio	1940-02-09	Adventure	G	84300000	2188229052
2	Fantasia	1940-11-13	Musical	G	83320000	2187090808
3	Song of the South	1946-11-12	Adventure	G	65000000	1078510579
4	Cinderella	1950-02-15	Drama	G	85000000	920608730
5	20,000 Leagues Under the Sea	1954-12-23	Adventure	NaN	28200000	528279994
6	Lady and the Tramp	1955-06-22	Drama	G	93600000	1236035515
7	Sleeping Beauty	1959-01-29	Drama	NaN	9464608	21505832
8	101 Dalmatians	1961-01-25	Comedy	G	153000000	1362870985
9	The Absent Minded Professor	1961-03-16	Comedy	NaN	25381407	310094574
10	Babes in Toyland	1961-12-14	Musical	G	10218316	124841160

Display some information about data

In [5]: `print(df.info())`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 579 entries, 0 to 578
Data columns (total 6 columns):
movie_title          579 non-null object
release_date         579 non-null object
genre                562 non-null object
mpaa_rating          523 non-null object
total_gross          579 non-null int64
inflation_adjusted_gross 579 non-null int64
dtypes: int64(2), object(4)
memory usage: 27.2+ KB
None
```

Step 3: Data Processing

```
In [6]: print(df.isna().sum()) #print NaN values
```

```
movie_title      0
release_date     0
genre            17
mpaa_rating      56
total_gross      0
inflation_adjusted_gross  0
dtype: int64
```

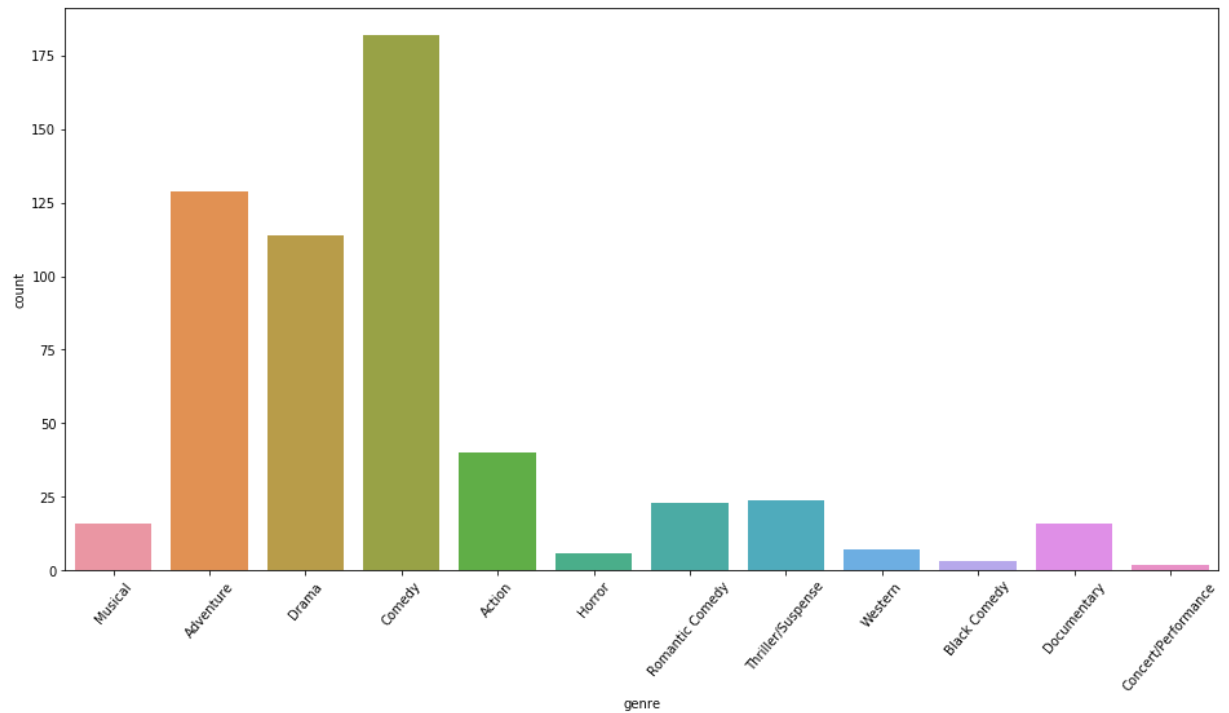
```
In [7]: #What is the most common movie genre made by Walt Disney Studios?
df['genre'].value_counts()
```

```
Out[7]: Comedy      182
Adventure    129
Drama        114
Action       40
Thriller/Suspense  24
Romantic Comedy  23
Documentary   16
Musical       16
Western       7
Horror        6
Black Comedy   3
Concert/Performance  2
Name: genre, dtype: int64
```

Step 4: Data Visualization

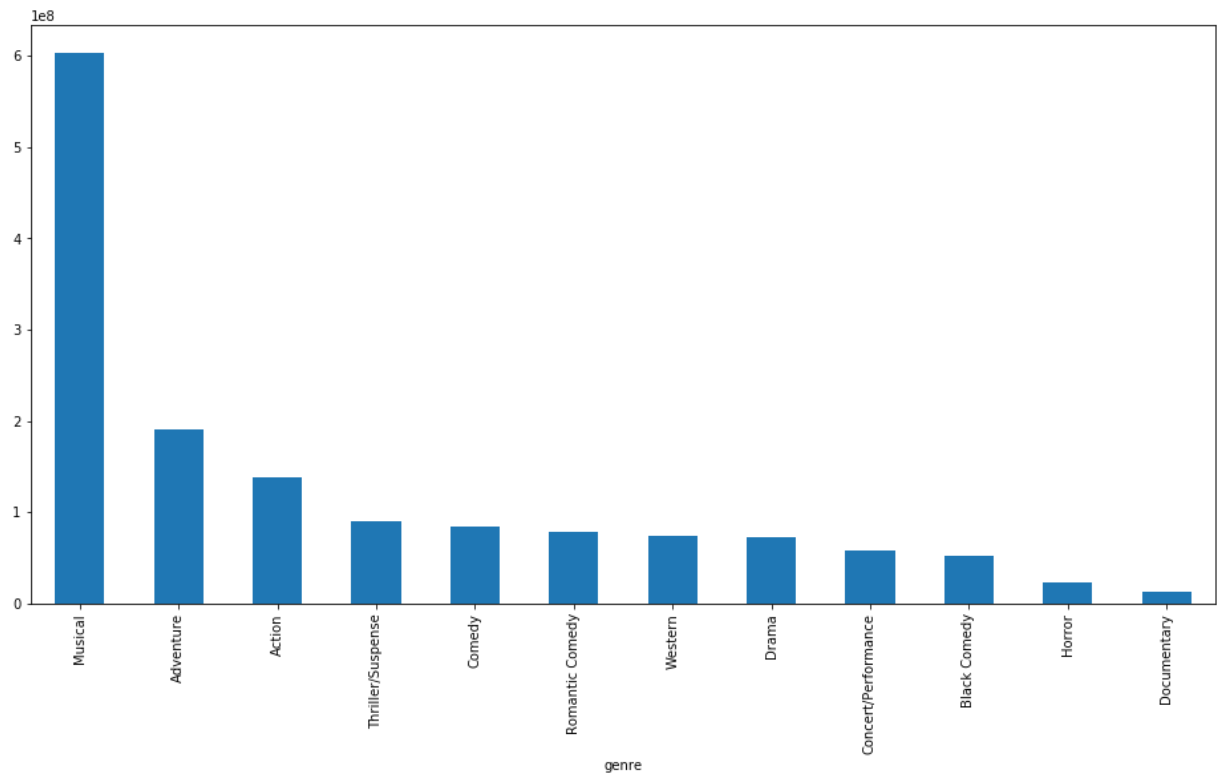
```
In [46]: #visualization
sns.countplot(x=df['genre'])
data=df
order=df.genre.value_counts().index
palette='magma'
plt.xticks(rotation=50)
```

```
Out[46]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11]),
<a list of 12 Text xticklabel objects>)
```



```
In [54]: # Visualize the adjusted gross mean by genre
df.groupby('genre')['inflation_adjusted_gross']
        .mean().sort_values(ascending=False).plot(kind='bar')
```

```
Out[54]: <matplotlib.axes._subplots.AxesSubplot at 0xbb7df8f8be0>
```



How Earnings changed with time

```
In [48]: df.index
```

```
Out[48]: RangeIndex(start=0, stop=579, step=1)
```

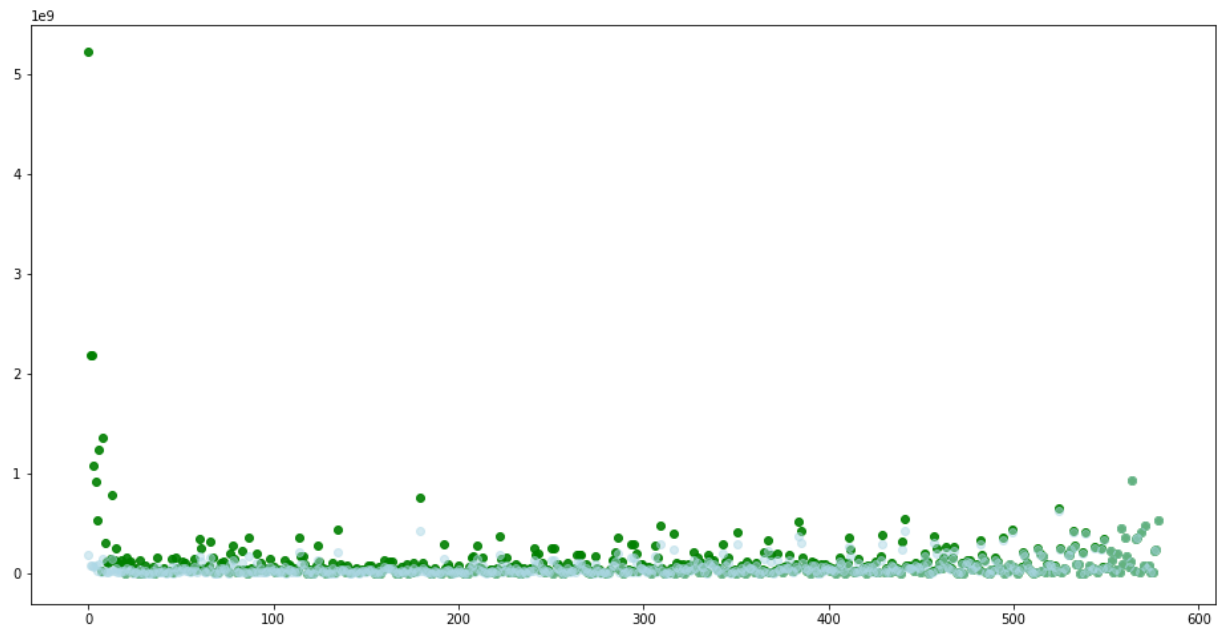
```
In [49]: dates = df.index.values
inflation_adjusted = df['inflation_adjusted_gross']
gross = df['total_gross']
```



```
In [53]: #Scatterplot
plt.rcParams['figure.figsize']=[16,8]
print(plt.scatter(x=dates, y=inflation_adjusted, color='green', alpha=0.9))
print(plt.scatter(x=dates, y=gross, color='lightblue', alpha = 0.5))
```

<matplotlib.collections.PathCollection object at 0x000000BB7DAD47B8>

<matplotlib.collections.PathCollection object at 0x000000BB7DAD4320>



```
In [52]: #lineplot using seaborn
sns.lineplot(x=dates, y=inflation_adjusted)
sns.lineplot(x=dates, y=gross, alpha = 0.6)
```

Out[52]: <matplotlib.axes._subplots.AxesSubplot at 0xbb7efd6128>

