EDA on Anime Dataset

This notebook aims to do basic data exploration using pandas. This is an anime dataset and has been obtained from Kaggle. We will use pandas to perform basic operations and see how the dataset can be wrangled to fit our needs. We will be taking a look at top rated anime, most used genres and how they are distributed across top rated anime. Let's begin!

In [1]:
```
#Importing the required Libraries
import pandas as pd
import numpy as np
import os
from scipy import stats
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [2]:
```
#Changing Directory to the file location
os.chdir("C:\\Users\\rakbansal\\Desktop\\Datasets\\anime")
```

In [3]:
```
#Reading the files
genre_data= pd.read_csv("datagenre-all-share-new.csv", delimiter= "|")
score_data= pd.read_csv("datascorehist-all-share-new.csv", delimiter= "|")
title_data= pd.read_csv("datatitle-all-share-new.csv", delimiter= "|")
```

In [4]:
```
#Let's look at the shape and size of the datasets
print('genre data shape:', genre_data.shape, '\nscore data shape:', score_
data.shape, '\ntitle data shape:', title_data.shape)
```

```
genre data shape: (4029, 2)
score data shape: (4029, 12)
title data shape: (4029, 2)
```

Looks like we have 4029 observations with 2 columns in genres, 2 columns in Titles Df and 12 columns in score Df. Let's take a look what these columns are.

In [5]:
```
#Let's take a look at each file one by one
title_data.head()
```

Out[5]:

	Anime_ID	Anime_name
0	11770	SteinsGate (TV)
1	10216	Fullmetal Alchemist: Brotherhood (TV)
2	9701	Clannad After Story (TV)
3	210	Rurouni Kenshin: Trust & Betrayal (OAV)
4	9173	Code Geass: Lelouch of the Rebellion R2 (TV)

In [6]:
```
score_data.head()
```

Out[6]:

	Anime_ID	Master-piece	Excellent	Very good	Good	Decent	So-so	Not really good	Weak	B
0	11770	0.570695	0.253150	0.096360	0.040597	0.009333	0.009799	0.003966	0.004900	0.0032
1	10216	0.508027	0.302085	0.108323	0.040598	0.014209	0.008120	0.004429	0.003691	0.0020
2	9701	0.624157	0.174336	0.080722	0.044823	0.017255	0.009718	0.005553	0.009520	0.0029
3	210	0.540228	0.220380	0.111179	0.058099	0.020532	0.012624	0.006540	0.005779	0.0022
4	9173	0.538169	0.240147	0.113526	0.050543	0.022522	0.007071	0.007464	0.004976	0.0040

In [7]:
```
score_data.describe()
```

Out[7]:

	Anime_ID	Master-piece	Excellent	Very good	Good	Decent	So-so
count	4029.000000	4029.000000	4029.000000	4029.000000	4029.000000	4029.000000	4029.000000
mean	6961.484736	0.069873	0.129875	0.202301	0.249826	0.161148	0.076609
std	5715.554618	0.077659	0.096083	0.092755	0.091914	0.088321	0.059734
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1472.000000	0.015385	0.056291	0.142857	0.193750	0.095890	0.034384
50%	5841.000000	0.050000	0.116168	0.209877	0.244318	0.153846	0.065934
75%	11983.000000	0.096045	0.197138	0.266667	0.297735	0.212121	0.106667
max	18052.000000	0.624157	0.666667	0.529412	0.777778	0.642857	0.416667

For the score data, we see that scores for ach anime ID vary from 'Master-Piece' to 'Awful'. Also interesting to note that the scores seem to be provided from a scale of 0-1 with none of the anime rated 1 but a lot of the have been given zeroes.

In [8]:
```
#Let's have a look at the genre data now
genre_data.head()
```

Out[8]:

	Anime_ID	Genres
0	11770	adventure;comedy;drama;mystery;psychological;r...
1	10216	adventure;comedy;drama;fantasy;thriller;alchem...
2	9701	comedy;drama;psychological;romance;supernatura...
3	210	action;drama;romance;historical;revenge;samura...
4	9173	action;drama;magic;psychological;science ficti...

Looks like each anime has a low of genres associated to it! We will come back to it. But first, let's take a look at the null values.

In [9]:
```
#checking for nulls
print('Null in Genre Data:', genre_data.isnull().sum(),
      '\nNull values in score data:', score_data.isnull().sum(),
      '\nNull values in title data:', title_data.isnull().sum())
```

```
Null in Genre Data: Anime_ID      0
Genres      277
dtype: int64
Null values in score data: Anime_ID           0
Master-piece        0
Excellent           0
Very good           0
Good                0
Decent              0
So-so               0
Not really good     0
Weak                0
Bad                 0
Awful               0
Worst ever          0
dtype: int64
Null values in title data: Anime_ID      0
Anime_name    0
dtype: int64
```

Title and Score Datasets do not have any null values. However, there are 277 missing genres. We can either drop them or fill them. Let's fill them with most occurring anime genres. But first, let's take a look at what kind of genres exist and their frequency.

In [11]:
```
#Checking unique count of genre data
genre_data['Genres'].value_counts()
```

Out[11]:
```
comedy
                                                              76
comedy;slice of life
                                                              61
adventure;fantasy
                                                              39
adventure
                                                              27
action;science fiction;mecha
                                                              27
drama;romance
                                                              26
drama
                                                              25
comedy;romance
                                                              24
action;adventure;science fiction;mecha
                                                              24
action;science fiction
```

```
                                                                  23
romance
                                                                  22
adventure;comedy;fantasy
                                                                  22
science fiction
                                                                  20
adventure;science fiction
                                                                  19
action;comedy;science fiction
                                                                  18
action
                                                                  18
tournament;sports
                                                                  17
action;adventure;science fiction
                                                                  17
comedy;slice of life;school
                                                                  17
slice of life
                                                                  17
action;comedy;science fiction;mecha
                                                                  16
adventure;comedy
                                                                  15
drama;science fiction
                                                                  15
drama;historical
                                                                  15
adventure;drama;fantasy
                                                                  14
fantasy
                                                                  14
adventure;comedy;science fiction
                                                                  14
comedy;magic
                                                                  13
comedy;fantasy
                                                                  13
adventure;science fiction;mecha
                                                                  12

                                                                  ..
drama;horror;mystery;psychological;thriller;idols
                                                                   1
comedy;slice of life;café;fanservice;maid;moe
                                                                   1
action;cyberpunk;mecha;police;post-apocalyptic
                                                                   1
action;drama;horror;magic;gore;historical;ninja
                                                                   1
adventure;drama;mystery;romance;amnesia;bishounen;male harem
                                                                   1
comedy;romance;slice of life;love polygon;school;school life
                                                                   1
comedy;mystery;supernatural;folklore
                                                                   1
```

```
action;slice of life;fanservice;girls with guns;school life;survival game
                                                               1
action;adventure;drama;romance;science fiction;aliens;military;space;space
 navy                                                          1
action;drama;psychological;science fiction;assassins;cyborg;girls with gun
s;terrorists                                                   1
comedy;drama;romance;card games;sports
                                                               1
drama;high school;music
                                                               1
comedy;drama;romance;supernatural;harem;music;otaku;parody;school life
                                                               1
adventure;comedy;drama;science fiction;thriller;assassins;conspiracy;crime
;gangs                                                         1
adventure;horror;supernatural;amnesia;demons;multiple personality;superpow
ers                                                            1
science fiction;post-apocalyptic
                                                               1
drama;romance;Moe
                                                               1
comedy;drama;psychological;romance;supernatural;bishoujo;coming of age;fam
ily;moe;parenting;tragedy                                      1
action;drama;assassins;military
                                                               1
action;comedy;romance;aliens
                                                               1
horror;supernatural;demons
                                                               1
romance;slice of life;love triangle;school
                                                               1
action;comedy;drama;magic;gothic lolita;hikikomori;living dolls
                                                               1
comedy;slice of life;fanservice;idols;music
                                                               1
comedy;magic;dark comedy;magical girl;parody
                                                               1
drama;magic;fighting;Magical Girl
                                                               1
action;comedy;drama;fantasy;romance;science fiction;fanservice;harem;mecha
;superpowers                                                   1
drama;historical;samurai
                                                               1
action;comedy;drama;mystery;supernatural;thriller;coming of age;conspiracy
;crime;fighting;Gangs;growing up;social-networking;spirits     1
science fiction;cyberpunk;tragedy
                                                               1
Name: Genres, Length: 2376, dtype: int64
```

'Comedy' and 'slice of life' look like clear winners and have the maximum frequency. Let's fill out the missing values with them.

In [12]:
```python
#Filling Null values
genre_data['Genres'].fillna(value='comedy;slice of life', inplace= True)
```

In [13]:
```python
#Let's have a look at our new dataset and see if missing values have been
```

```
take care of
genre_data.isnull().sum()
```

Out[13]:
```
Anime_ID    0
Genres      0
dtype: int64
```

In [19]:
```
#Let's check the unique value count
genre_data['Genres'].value_counts()
```

Out[19]:
```
comedy;slice of life
                                   338
comedy
                                    76
adventure;fantasy
                                    39
action;science fiction;mecha
                                    27
adventure
                                    27
drama;romance
                                    26
drama
                                    25
comedy;romance
                                    24
action;adventure;science fiction;mecha
                                    24
action;science fiction
                                    23
romance
                                    22
adventure;comedy;fantasy
                                    22
science fiction
                                    20
adventure;science fiction
                                    19
action;comedy;science fiction
                                    18
action
                                    18
tournament;sports
                                    17
comedy;slice of life;school
                                    17
slice of life
                                    17
action;adventure;science fiction
                                    17
action;comedy;science fiction;mecha
                                    16
drama;science fiction
                                    15
drama;historical
                                    15
adventure;comedy
```

```
                               15
adventure;comedy;science fiction
                               14
fantasy
                               14
adventure;drama;fantasy
                               14
comedy;magic
                               13
comedy;fantasy
                               13
science fiction;mecha
                               12

                             ...
adventure;comedy;drama;romance;science fiction;aliens;time travel
                                1
action;drama;gangs;yakuza
                                1
romance;slice of life;tournament;school;tennis
                                1
action;fantasy;demons;middle ages;military;war
                                1
action;drama;supernatural;special abilities
                                1
adventure;comedy;fantasy;talking animals
                                1
action;comedy;romance;supernatural;demons;fanservice;harem
                                1
comedy;romance;moe
                                1
action;drama;science fiction;military;monsters;post-apocalyptic;superpower
s                               1
action;comedy;espers;superpowers
                                1
action;drama;psychological
                                1
action;drama;horror;science fiction;alien;mecha;military
                                1
action;drama;romance;science fiction;tournament;fighting;martial arts;mech
a;super robot                   1
adventure;fantasy;slice of life
                                1
comedy;drama;psychological;aliens
                                1
action;mystery;psychological;science fiction;mecha;noir;robot girl
                                1
action;adventure;comedy;romance;science fiction;superpowers
                                1
action;horror;science fiction;aliens;fighting;military;technology
                                1
adventure;tournament;racing;sports
                                1
tournament;martial arts;sports
                                1
comedy;magic;bishounen;parody
                                1
```

```
comedy;drama;mystery;romance;supernatural
                                     1
action;supernatural;vampires
                                     1
comedy;drama;psychological;romance;supernatural;bishoujo;coming of age;fam
ily;moe;parenting;tragedy            1
action;comedy;science fiction;Mecha;politics;real robot;religion;war
                                     1
action;comedy;fantasy;crossdressing;online computer gaming;otaku;parody;vi
rtual reality                        1
action;drama;science fiction;bounty hunters;space;terrorists
                                     1
action;drama;horror;science fiction;supernatural;post-apocalyptic
                                     1
action;drama;science fiction;mecha;military;real robot;space navy;war
                                     1
adventure;comedy;drama;fantasy;romance;fanservice;harem;magical girl;tsund
ere                                  1
Name: Genres, Length: 2376, dtype: int64
```

We can see that initially 'comedy' alone was the leading genre but after we imputed the missing values, the spot has been taken away by 'Comedy; slice of life'. But that was expected since there were 277 missing values. Besides, we will be seperating the strings, so that effect will be accounted for.

Let us now turn our attention to the score dataset which has a lot of rating columns. We don't need that many ratings, so we average them out and create another column with mean ratings.

In [14]:
```
#trying getting mean of each column in score data- let's view the Df again
score_data.head()
```

Out[14]:

	Anime_ID	Master-piece	Excellent	Very good	Good	Decent	So-so	Not really good	Weak	B
0	11770	0.570695	0.253150	0.096360	0.040597	0.009333	0.009799	0.003966	0.004900	0.0032
1	10216	0.508027	0.302085	0.108323	0.040598	0.014209	0.008120	0.004429	0.003691	0.0020
2	9701	0.624157	0.174336	0.080722	0.044823	0.017255	0.009718	0.005553	0.009520	0.0029
3	210	0.540228	0.220380	0.111179	0.058099	0.020532	0.012624	0.006540	0.005779	0.0022
4	9173	0.538169	0.240147	0.113526	0.050543	0.022522	0.007071	0.007464	0.004976	0.0040

In [15]:
```
#Creating another columns called mean_score
score_data['mean_score']= score_data.iloc[:,1:11].mean(axis=1)
score_data.head()
```

Out[15]:

	Anime_ID	Master-piece	Excellent	Very good	Good	Decent	So-so	Not really good	Weak	B

0	11770	0.570695	0.253150	0.096360	0.040597	0.009333	0.009799	0.003966	0.004900	0.0032
1	10216	0.508027	0.302085	0.108323	0.040598	0.014209	0.008120	0.004429	0.003691	0.0020
2	9701	0.624157	0.174336	0.080722	0.044823	0.017255	0.009718	0.005553	0.009520	0.0029
3	210	0.540228	0.220380	0.111179	0.058099	0.020532	0.012624	0.006540	0.005779	0.0022
4	9173	0.538169	0.240147	0.113526	0.050543	0.022522	0.007071	0.007464	0.004976	0.0040

In [22]:

```
#Let's drop all other columns
score_data= score_data.drop(['Master-piece', 'Excellent', 'Very good', 'Go
od', 'Decent', 'So-so', 'Not really good','Weak','Bad', 'Awful','Worst eve
r'], axis=1)
score_data.head()
```

Out[22]:

	Anime_ID	mean_score
0	11770	0.099393
1	10216	0.099354
2	9701	0.097362
3	210	0.098144
4	9173	0.099057

Let's now merge Title data and score data to get a better picture.

In [23]:

```
#merging title and score
title_score= pd.merge(title_data, score_data)
title_score.head()
```

Out[23]:

	Anime_ID	Anime_name	mean_score
0	11770	SteinsGate (TV)	0.099393
1	10216	Fullmetal Alchemist: Brotherhood (TV)	0.099354
2	9701	Clannad After Story (TV)	0.097362
3	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144
4	9173	Code Geass: Lelouch of the Rebellion R2 (TV)	0.099057

Let's further have a look at the top rated anime by sorting the DF via 'mean_score'.

In [24]:

```
#Arranging them in decreasing order of the mean
title_score.sort_values(['mean_score'], ascending= False)
```

Out[24]:

	Anime_ID	Anime_name	mean_score
3587	2671	Guardian Hearts (OAV)	0.100000

3013	6358	One Piece: Defeat The Pirate Ganzak! (special)	0.100000
274	10952	Broken Blade (movie series)	0.100000
198	7468	Lovely Complex (TV)	0.100000
318	817	Daicon films (special)	0.100000
1954	1018	Little Snow Fairy Sugar (TV)	0.100000
2468	662	Flame of Recca (TV)	0.100000
1927	12547	Aquarion Evol (TV)	0.100000
1937	5842	Gala (movie)	0.100000
2558	17795	Ōya-san wa Shishunki! (TV)	0.100000
3109	631	Alakazam the Great (movie)	0.100000
3373	6359	Getsumen to Heiki Mina (TV)	0.100000
3691	3139	Harukanaru Toki no Naka de: Ajisai Yumegatari ...	0.100000
2475	15109	Mushibugyō (TV)	0.100000
662	7631	Oh! Edo Rocket (TV)	0.100000
2911	903	Ajimu - Kaigan Monogatari (ONA)	0.100000
1687	2839	Maison Ikkoku: Prelude When the Cherry Blosso...	0.100000
3681	1029	Ys (OAV)	0.100000
1610	13980	Ginga e Kickoff!! (TV)	0.100000
1686	11191	Psychic Detective Yakumo (TV)	0.100000
2552	10668	Sōten Kōro (TV)	0.100000
2541	7420	Princess Resurrection (TV)	0.100000
1982	15493	Yamishibai: Japanese Ghost Stories (TV)	0.100000
3124	17647	Prince of Stride: Alternative (TV)	0.100000
1116	11796	Fractale (TV)	0.100000
2491	14610	Macross FB7: Ore no Uta o Kike! (movie)	0.100000
2293	8767	Detective Conan: A Challenge from Agasa (OAV)	0.100000
1908	2387	Jing King of Bandits: Seventh Heaven (OAV)	0.100000
2231	2934	Di Gi Charat Natsuyasumi Special (special)	0.100000
1429	5560	Lupin III: Angel Tactics (special)	0.100000
...	...	...	...
4023	645	Tekken: The Motion Picture (OAV)	0.084339
3979	3085	Bobobo-bo Bo-bobo (TV)	0.083904
3637	15052	Vanquished Queens (OAV)	0.083333
2732	7202	Hal & Bons (OAV)	0.083333
3808	2057	Ariel Deluxe (OAV)	0.083333

3704	17153	Venus Project: Climax (TV)	0.083333
1366	8423	Evangelion Shin Gekijō-ban: ? (movie)	0.083333
4018	186	Ninja Resurrection (OAV)	0.082072
3376	856	Cipher (OAV)	0.081818
4005	11242	Isshoni Sleeping: Sleeping with Hinako (OAV)	0.081176
2173	1461	Ikkyū-san (TV 1975)	0.080952
4010	3875	Dracula: Sovereign Of The Damned (special)	0.080645
4021	3429	Duel Masters (TV)	0.080614
4020	791	MD Geist (OAV)	0.079911
3811	2241	Gregory Horror Show (TV)	0.078571
4015	951	Garzey's Wing (OAV)	0.077876
1643	1229	Bosco Daibōken (TV)	0.076471
4019	2276	Reign: The Conqueror (TV)	0.076437
4024	957	MD Geist II - Death Force (OAV)	0.075926
4014	1019	Super Milk-chan (TV)	0.075075
4012	2	Apocalypse Zero (OAV)	0.073810
3589	2957	PoPoLoCrois (TV)	0.073684
4016	1020	Super Milk-chan Show (TV)	0.073036
1085	2314	Manga Sekai Mukashi Banashi (TV)	0.072727
3924	3934	Kyōfu Densetsu: Kaiki! Frankenstein (special)	0.072727
4026	4653	Skelter+Heaven (OAV)	0.071739
3998	6452	Musashi (TV)	0.067143
4027	2922	SD Gundam Force (TV)	0.065306
1945	3631	Sekai Meisaku Dōwa: Mori wa Ikiteiru (movie)	0.064286
4028	5907	Hametsu No Mars (OAV)	0.061622

4029 rows × 3 columns

In [25]:
```
#Let's take a look at the top 25 anime
top_rated_25= title_score.sort_values(['mean_score'], ascending= False)[:2
5]
top_rated_25
```

Out[25]:

	Anime_ID	Anime_name	mean_score
3587	2671	Guardian Hearts (OAV)	0.1
3013	6358	One Piece: Defeat The Pirate Ganzak! (special)	0.1
274	10952	Broken Blade (movie series)	0.1

198	7468	Lovely Complex (TV)	0.1
318	817	Daicon films (special)	0.1
1954	1018	Little Snow Fairy Sugar (TV)	0.1
2468	662	Flame of Recca (TV)	0.1
1927	12547	Aquarion Evol (TV)	0.1
1937	5842	Gala (movie)	0.1
2558	17795	Ōya-san wa Shishunki! (TV)	0.1
3109	631	Alakazam the Great (movie)	0.1
3373	6359	Getsumen to Heiki Mina (TV)	0.1
3691	3139	Harukanaru Toki no Naka de: Ajisai Yumegatari ...	0.1
2475	15109	Mushibugyō (TV)	0.1
662	7631	Oh! Edo Rocket (TV)	0.1
2911	903	Ajimu - Kaigan Monogatari (ONA)	0.1
1687	2839	Maison Ikkoku: Prelude When the Cherry Blosso...	0.1
3681	1029	Ys (OAV)	0.1
1610	13980	Ginga e Kickoff!! (TV)	0.1
1686	11191	Psychic Detective Yakumo (TV)	0.1
2552	10668	Sōten Kōro (TV)	0.1
2541	7420	Princess Resurrection (TV)	0.1
1982	15493	Yamishibai: Japanese Ghost Stories (TV)	0.1
3124	17647	Prince of Stride: Alternative (TV)	0.1
1116	11796	Fractale (TV)	0.1

Looks like all of them have been rated at 0.1. Let's see what the worst ranked anime are.

In [26]:

```
#Checking lowest 25 rated anime
bottom_25= title_score.sort_values(['mean_score'], ascending= True)[:25]
bottom_25
```

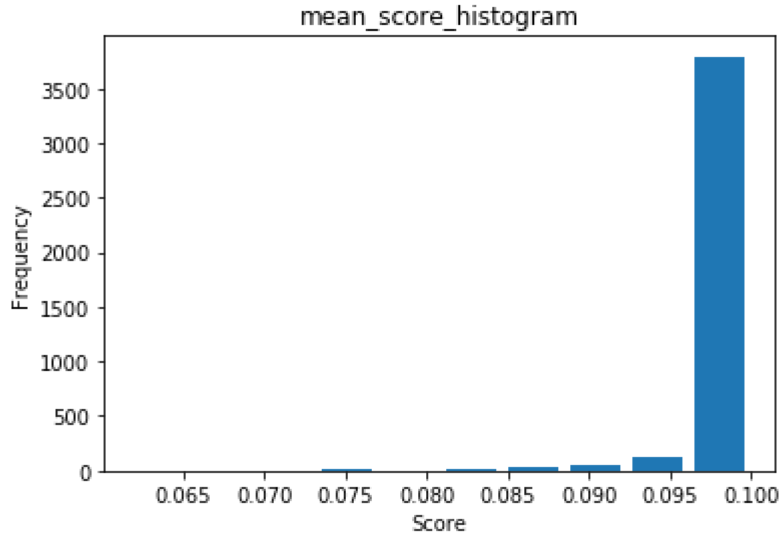
Out[26]:

	Anime_ID	Anime_name	mean_score
4028	5907	Hametsu No Mars (OAV)	0.061622
1945	3631	Sekai Meisaku Dōwa: Mori wa Ikiteiru (movie)	0.064286
4027	2922	SD Gundam Force (TV)	0.065306
3998	6452	Musashi (TV)	0.067143
4026	4653	Skelter+Heaven (OAV)	0.071739
3924	3934	Kyōfu Densetsu: Kaiki! Frankenstein (special)	0.072727
1085	2314	Manga Sekai Mukashi Banashi (TV)	0.072727

4016	1020	Super Milk-chan Show (TV)	0.073036
3589	2957	PoPoLoCrois (TV)	0.073684
4012	2	Apocalypse Zero (OAV)	0.073810
4014	1019	Super Milk-chan (TV)	0.075075
4024	957	MD Geist II - Death Force (OAV)	0.075926
4019	2276	Reign: The Conqueror (TV)	0.076437
1643	1229	Bosco Daibōken (TV)	0.076471
4015	951	Garzey's Wing (OAV)	0.077876
3811	2241	Gregory Horror Show (TV)	0.078571
4020	791	MD Geist (OAV)	0.079911
4021	3429	Duel Masters (TV)	0.080614
4010	3875	Dracula: Sovereign Of The Damned (special)	0.080645
2173	1461	Ikkyū-san (TV 1975)	0.080952
4005	11242	Isshoni Sleeping: Sleeping with Hinako (OAV)	0.081176
3376	856	Cipher (OAV)	0.081818
4018	186	Ninja Resurrection (OAV)	0.082072
1366	8423	Evangelion Shin Gekijō-ban: ? (movie)	0.083333
2732	7202	Hal & Bons (OAV)	0.083333

Looks like the worst rated anime are in the range of 0.06 to 0.08 implying that all anime have been ranked in the range of 0.06 to 0.1. Let's see that visually.

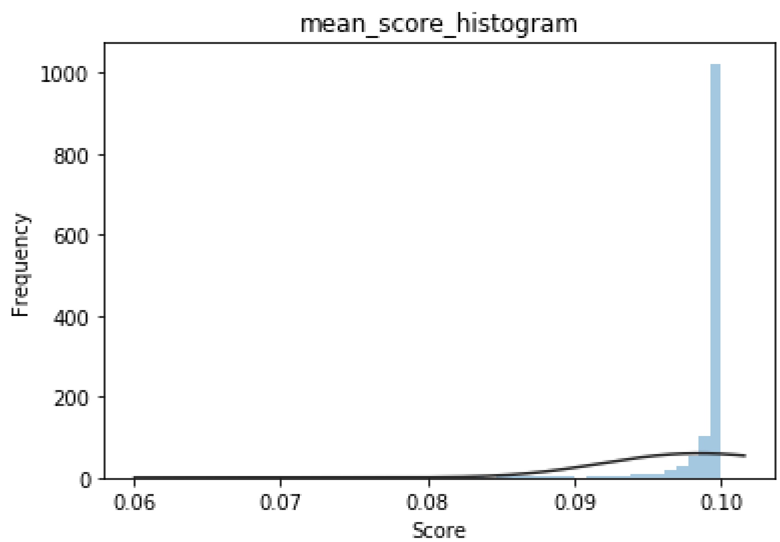
In [29]:
```
#visualizing Mean scores
plt.hist(x='mean_score', data= title_score, histtype= 'bar', rwidth=0.8)
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.title("mean_score_histogram")
plt.show()
```



That's odd!Most anime seems to have been rated well with most of them being rated from 0.096 to 0.1 Maybe we have too many anime stans! Let's confirm this via a KDE.

In [30]:
```
#Let's try visualizing it via a KDE with seaborn
score= title_score[['mean_score']]

sns.distplot(score, kde=False, fit=stats.gamma)
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.title("mean_score_histogram")
plt.show()
```



Just as we thought! Most anime seems to have been rated well!

Let's move to genres now. As we noted above, one Anime seems to be tagged to a lot of genres that are seperated by a ';'. Let's first sepearte all the genres respevtive to their Anime IDs.

In [31]:
```
#Quickly take a look at the DF
genre_data.head()
```

Out[31]:

	Anime_ID	Genres
0	11770	adventure;comedy;drama;mystery;psychological;r...
1	10216	adventure;comedy;drama;fantasy;thriller;alchem...
2	9701	comedy;drama;psychological;romance;supernatura...
3	210	action;drama;romance;historical;revenge;samura...
4	9173	action;drama;magic;psychological;science ficti...

Let us now seperate the 'Genres'. We will start with creating a new DataFrame with 'Anime_ID' as index.

In [32]:
```
#Seperating Genres with 'Str.split() and expanding the DF'
new_df = pd.DataFrame(genre_data.Genres.str.split(';', expand= True), inde
x=genre_data.Anime_ID).stack()
new_df.head()
```

Out[32]:
```
Anime_ID
210       0         mystery
          1    supernatural
          2       bishounen
          3     dark comedy
          4         demons
dtype: object
```

We see that even though Genres have now been split, we have two-indices in our DF. We need to get rid of one of these indices and create a more holistic Dataframe. To do this, we will make AnimeId as a column (it can't be an index since the values will be duplicated).

In [33]:
```
#Resetting index
new_df = new_df.reset_index([0, 'Anime_ID'])
```

In [36]:
```
#Let's set the columns names now
new_df.columns = ['Anime_ID', 'Genres']
new_df.head()
```

Out[36]:

	Anime_ID	Genres
0	210	mystery
1	210	supernatural
2	210	bishounen
3	210	dark comedy
4	210	demons

Looks good! But we still don't know what Anime_ID corresponds to which anime and it's respective score. Let's now merge the datasets.

In [38]:
```
#Merging genres with ratings data
anime_df= pd.merge(title_score, new_df)
anime_df.head()
```

Out[38]:

	Anime_ID	Anime_name	mean_score	Genres
0	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144	mystery
1	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144	supernatural
2	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144	bishounen
3	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144	dark comedy
4	210	Rurouni Kenshin: Trust & Betrayal (OAV)	0.098144	demons

In [39]:
```
#Let's check the shape
anime_df.shape
```

Out[39]:
```
(7629, 4)
```

The total number of observations has increased which was expected since we stacked each genre individually. This way, every 'Anime_ID' corresponds to multiple genres per row.

Time to take a look at our most used genres i.e. genres which have maximum number of Animes associated to them.

In [40]:
```
anime_df['Genres'].value_counts()
```

Out[40]:
```
comedy                          822
drama                           685
action                          543
science fiction                 501
adventure                       411
romance                         374
slice of life                   301
fantasy                         241
supernatural                    227
mecha                           197
mystery                         143
school                          134
psychological                   129
magic                           113
military                        112
tournament                      102
historical                       92
horror                           80
space                            70
harem                            67
sports                           64
superpowers                      63
tragedy                          58
aliens                           58
fanservice                       54
war                              53
demons                           53
```

```
magical girl                    51
parody                          51
thriller                        46
                              ...
cyborgs                          1
Unrequited Love                  1
dolls                            1
maid                             1
Mecha                            1
Growing Up                       1
Cross dressing                   1
Dark comedy                      1
middle ages                      1
Vampire                          1
Military                         1
neet                             1
fan service                      1
Love triangle                    1
immortal                         1
swordsman                        1
Bishounen                        1
Crossdressing                    1
student teacher relationship     1
butterfly effect                 1
disaster                         1
Alternate history                1
Family                           1
wizards/witches                  1
mind control                     1
wolf girls                       1
host club                        1
single parents                   1
ren'ai                           1
Growing up                       1
Name: Genres, Length: 270, dtype: int64
```

We can see that there are 270 different types of genres but let's not consider them all since most of them have a frequency of 1. What's interesting is that 'Comedy' is back on top whereas 'Slice of life' seems to have take a backseat. What's also evident is that there is a a stark difference in frequencies in the topmost genres and the ones that trail. For the sake of simplicity, we will only be dealing with genres that have a frequency of over 100.

Let us start by creating a new DataFrame called 'genre_count' which will have genres against their frequency, represent by 'Count' variable.

In [41]:
```
#Genres and their Frequencies
genre_count = pd.DataFrame(anime_df.groupby(by = ['Genres']).size(),column
s = ['count'])
genre_count = genre_count.reset_index()
genre_count.head()
```

Out[41]:

	Genres	count
0	Aliens	2

1	Alternate history	1
2	Another world	2
3	Bishojo	2
4	Bishounen	1

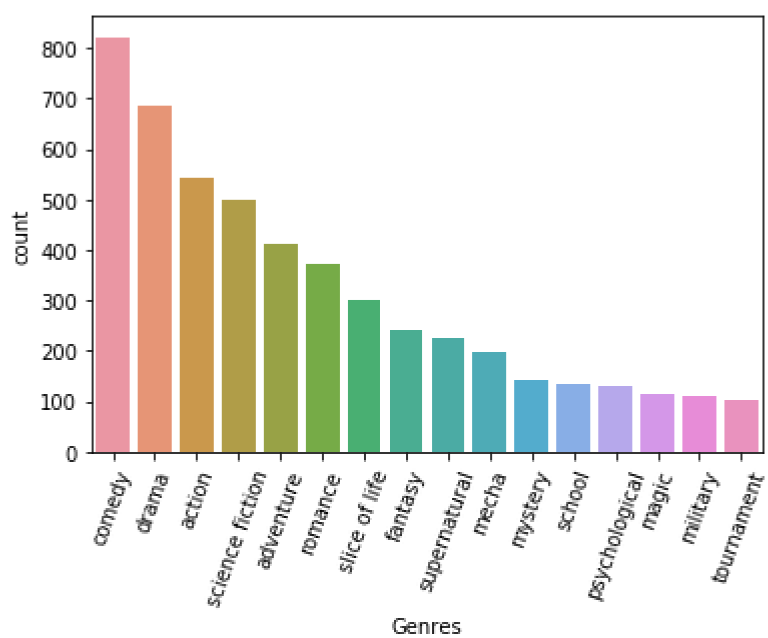
In [44]:
```
#Now that we have a DF with their count, let's see the most used genres-co
nsider top 100
sorted_genres= genre_count.sort_values('count', ascending= False)
above_100=sorted_genres[sorted_genres['count']>100]
above_100.head()
```

Out[44]:

	Genres	count
91	comedy	822
113	drama	685
62	action	543
214	science fiction	501
64	adventure	411

Plotting time! Let's take a look at this visually to get a better picture.

In [48]:
```
sns.barplot(x= 'Genres', y= 'count', data= above_100)
plt.xticks(rotation=70)
plt.rcParams['xtick.labelsize']=5
```



Let's take a look at the score Dataframe. Earlier, we noted that the highest rated movies have a mean score of 0.1. Let's make a seperate Dataframe for the same.

In [49]:
```
#let's make a DF with mean_score=0.1
```

```
top_rated= title_score[title_score.mean_score==0.1]
```

In [50]:
```
top_rated.head()
```

Out[50]:

	Anime_ID	Anime_name	mean_score
18	12398	Bunny Drop (TV)	0.1
84	2340	Sword of the Stranger (movie)	0.1
89	14995	Garden of Words (movie)	0.1
91	4115	Major (TV)	0.1
102	10654	Maison en Petits Cubes (movie)	0.1

In [51]:
```
#Checking the length here
len(top_rated)
```

Out[51]: 1014

This means there are 1014 movies with 0.1 rating. Wouldn't be possible to visualize them all. Let's take a look at them in ascending and descending order.

In [53]:
```
#Let's visualize the top rated anime
top_25_anime= top_rated.sort_values(['mean_score', 'Anime_name'], ascendin
g= [False, True])[:25]
top_25_anime
```

Out[53]:

	Anime_ID	Anime_name	mean_score
3129	11962	30-sai no Hoken Taiiku (TV)	0.1
3036	369	801 T.T.S. Airbats (OAV)	0.1
455	14873	AKB0048 next stage (TV)	0.1
3742	3748	AM Driver (TV)	0.1
829	16555	AOKANA: Four Rhythm Across the Blue (TV)	0.1
1376	11951	Aa Megami-sama (OAV 2011)	0.1
648	8679	Aa Megami-sama: Tatakau Tsubasa (special)	0.1
3549	15970	Abarenbō Rikishi!! Matsutarō (TV)	0.1
2180	405	Ace o Nerae! (TV)	0.1
3765	17703	Active Raid (TV)	0.1
2236	16733	Actually I Am… (TV)	0.1
2287	2129	Adventures of Pinocchio (TV)	0.1
3093	1360	Adventures of the Little Prince (TV)	0.1
1609	16113	Age 12 (OAV)	0.1
535	6426	Ah! My Goddess: Flights of Fancy (TV)	0.1

2062	17082	Ai Yori Aoshi ~Enishi~ Miyuki (OAV)	0.1
1000	15597	Aikatsu! (TV 2)	0.1
3393	7022	Akahori Gedō Hour Rabuge (TV)	0.1
2495	556	All Purpose Cultural Cat Girl Nuku Nuku (OAV)	0.1
3392	557	All Purpose Cultural Cat Girl Nuku Nuku (TV)	0.1
642	16133	Amagi Brilliant Park (TV)	0.1
3183	5012	Ame to Shoujo to Watashi no Tegami (ONA)	0.1
885	11511	And Yet the Town Moves (TV)	0.1
1874	1079	Andes Shōnen Pepero no Bōken (TV)	0.1
1230	3807	Angel Densetsu (OAV)	0.1

In [54]:

```
#Another possible way of doing it could be:
top_25_anime_desc= top_rated.sort_values(['mean_score', 'Anime_name'], asc
ending= [False, False])[:25]
top_25_anime_desc
```

Out[54]:

	Anime_ID	Anime_name	mean_score
1346	15689	Ōkii Ichinensei to Chiisana Ninensei (movie)	0.1
2479	3373	éX-D: Danger Zone (OAV)	0.1
3401	15178	sole penetra le illusioni ~ Day Break Illusion...	0.1
1340	13373	hack//The Movie (movie)	0.1
2757	709	hack//Liminality (OAV)	0.1
2689	9179	hack//G.U. Returner (OAV)	0.1
2009	13451	gdgd Fairies (TV)	0.1
2093	14199	blossom (special)	0.1
2920	5900	Zettai Seigi Love Pheromone (TV)	0.1
1124	11244	Zettai Karen Children (OAV)	0.1
3505	333	Zenki (TV)	0.1
883	6145	Zegapain (TV)	0.1
401	11478	Zan Sayonara Zetsubō Sensei Bangai-chi (OAV)	0.1
2299	1957	Yūsha Tokkyū Might Gaine (TV)	0.1
2054	1458	Yōsei Florence (movie)	0.1
1655	14576	Yuyushiki (TV)	0.1
2189	11219	Yutori-chan (ONA)	0.1
553	16339	Yuruyuri Nachu Yachumi! (OAV)	0.1
2953	13801	Yurumates3Dei (TV)	0.1
2095	13201	Yurumates Ha? (OAV)	0.1

2930	13017	Yuri Seijin Naoko-san (OAV)	0.1
771	10949	Yumeiro Pâtissière (TV)	0.1
524	16171	Yuki Yuna Is a Hero (TV)	0.1
2587	15764	Yu-Gi-Oh! Arc-V (TV)	0.1
2704	9816	Yozakura Quartet (TV)	0.1

We can also visualize the genres of these anime. We already saw which are the most populous genres. It's time to see what kind of genres do these top rated anime belong to.

Let's begin by making a Dataframe of genres of top rated anime called 'most_rated_genre'. We will follow a similar approach and add a 'Count' column to it.

In [56]:
```
most_rated_genre = pd.DataFrame(anime_df.groupby(by = ['Genres']).size(),c
olumns = ['count'])
most_rated_genre = most_rated_genre.reset_index()
most_rated_genre.head()
```

Out[56]:

	Genres	count
0	Aliens	2
1	Alternate history	1
2	Another world	2
3	Bishojo	2
4	Bishounen	1

In [57]:
```
#Let's filter out the top 25 genres in it
top_25= most_rated_genre.sort_values('count', ascending= False)[:25]
top_25
```

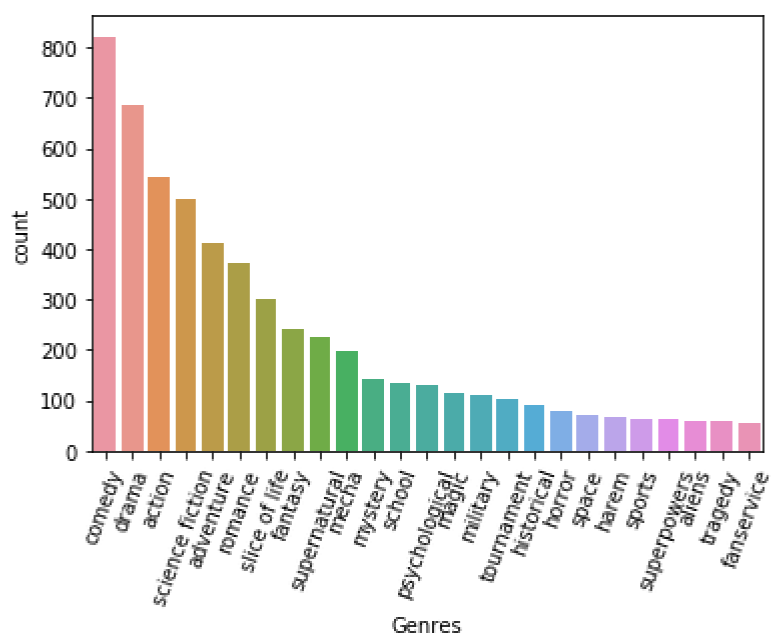
Out[57]:

	Genres	count
91	comedy	822
113	drama	685
62	action	543
214	science fiction	501
64	adventure	411
210	romance	374
223	slice of life	301
124	fantasy	241
239	supernatural	227
172	mecha	197

184	mystery	143
212	school	134
200	psychological	129
163	magic	113
177	military	112
251	tournament	102
146	historical	92
148	horror	80
226	space	70
143	harem	67
231	sports	64
240	superpowers	63
68	aliens	58
253	tragedy	58
123	fanservice	54

There doesn't seem to be much difference between genres belonging to top rated anime and genres with maximum frequency. Let's see it in a plot.

In [60]:
```
#Let's visualize and we can see that comedy, drama, sci-fi, action and adv
enture are clear winners.
sns.barplot(x= 'Genres', y= 'count', data= top_25)
plt.xticks(rotation=70)
plt.rcParams['xtick.labelsize']=10
```



As expected, the most preferred anime are also the one which have most common genres, with comedy leading and closely followed by drama and action.

That's it for this notebook. Stay tuned for more!

In [ ]:

```

```