


```

.....:                                     ='rating'
.....: print '\nEach movie average rating is: \n\n'
.....:
.....: # 2. Present the highest average rating movies list
.....:
.....:                                     =                                     .                                     'rating'                                     =False
.....:
.....: print '\nTop 20 average rating movies: \n\n'                                     .                                     20
.....:
.....: # 3. Calculate the each movie rating gender wise
.....:
.....:                                     =                                     .                                     ='title'
.....:                                     ='gender'                                     ='rating'
.....:
.....: print '\nGender wise movie ratings: \n\n'
.....:
.....: # 4. What is the most watched movie from the dataset?
.....:
.....: .
.....: list
.....:                                     =                                     .                                     'title'                                     .
.....:                                     .                                     =False
.....:                                     .
.....:                                     .
.....: print '\n\nThe most watched movie form the dataset:\n\n'                                     .
.....:
.....: # 5. What are the total number of unique movies in the dataset
.....: # and how many of them are rated atleast 3 on average
.....:
.....:                                     =                                     .                                     'title'                                     .                                     .
.....:
.....: print '\n\nTotal number of unique movies in dataset are: '
.....:
.....:                                     =                                     'rating' > 3
.....:                                     'rating' .
.....:
.....: print '\n\nNumber of movies with atleast 3 as average rating:
.....: 'rating' .
.....:
.....: # 6. how many movies were watched only by male and how many only by female?
.....:
.....: ''' As there are no zero values in any columns, I will consider movie was
.....: watched only female when the average male rating is not available,
.....: and vice-versa '''
.....:
.....:                                     =                                     'F'                                     .                                     .
.....:
.....: print '\n\nNumber of movies watched only by male: '
.....:
.....:                                     =                                     'M'                                     .                                     .
.....:
.....: print '\n\nNumber of movies watched only by female: '
.....:
.....: # 7. What is the top 20 rated movies in the age group 20 to 25 and most watched
movie for same group
.....:
.....:                                     =                                     'age' >= 20 &                                     'age' <= 25
.....:
.....:

```

```

....: # age_group dataset if 20 to 25 age group subset of total dataset
....:
....:
....:         =
....:         = 'title'         = 'rating' .         = 'rating'
= False .         20
....:
....: print '\n\nTop 20 rated movies in age group 20 to 25: \n'
....:
....: print '\n\nMost watched movie in 20 to 25 age group: \n'
....:         .         'title' .         .
....:
....: # 8. Plot a bargraph to know how the users are spread across zipcodes.
....:
....:         =         'user_id' 'zip' .
....:
....:         =         .         'zip' .
....: print '\n\n' .
....:
....: #zip_cnt.plot(kind='bar',figsize=(10,15))
....:
....: '''We can see that as the number of zip codes on x axis are many
....: the bar plot doesn't show any clear information
....: '''

```

Each movie average rating is:

title	rating
\$1,000,000 Duck (1971)	3.027027
'Night Mother (1986)	3.371429
'Til There Was You (1997)	2.692308
'burbs, The (1989)	2.910891
...And Justice for All (1979)	3.713568
...	...
Zed & Two Noughts, A (1985)	3.413793
Zero Effect (1998)	3.750831
Zero Kelvin (Kjærlighetens kjøtere) (1995)	3.500000
Zeus and Roxanne (1997)	2.521739
eXistenZ (1999)	3.256098

[3706 rows x 1 columns]

Top 20 average rating movies:

title	rating
Ulysses (Ulissee) (1954)	5.000000
Lured (1947)	5.000000
Follow the Bitch (1998)	5.000000
Bittersweet Motel (2000)	5.000000
Song of Freedom (1936)	5.000000
One Little Indian (1973)	5.000000
Smashing Time (1967)	5.000000
Schlafes Bruder (Brother of Sleep) (1995)	5.000000
Gate of Heavenly Peace, The (1995)	5.000000
Baby, The (1973)	5.000000
I Am Cuba (Soy Cuba/Ya Kuba) (1964)	4.800000
Lamerica (1994)	4.750000

Apple, The (Sib) (1998)	4.666667
Sanjuro (1962)	4.608696
Seven Samurai (The Magnificent Seven) (Shichini...	4.560510
Shawshank Redemption, The (1994)	4.554558
Godfather, The (1972)	4.524966
Close Shave, A (1995)	4.520548
Usual Suspects, The (1995)	4.517106
Schindler's List (1993)	4.510417

Gender wise movie ratings:

gender	F	M
title		
\$1,000,000 Duck (1971)	3.375000	2.761905
'Night Mother (1986)	3.388889	3.352941
'Til There Was You (1997)	2.675676	2.733333
'burbs, The (1989)	2.793478	2.962085
...And Justice for All (1979)	3.828571	3.689024
...
Zed & Two Noughts, A (1985)	3.500000	3.380952
Zero Effect (1998)	3.864407	3.723140
Zero Kelvin (Kjærlighetens kjøtere) (1995)	NaN	3.500000
Zeus and Roxanne (1997)	2.777778	2.357143
eXistenZ (1999)	3.098592	3.289086

[3706 rows x 2 columns]

The most watched movie form the dataset:

American Beauty (1999)

Total number of unique movies in dataset are: 3706

Number of movies with atleast 3 as average rating: 2454

Number of movies watched only by male: 225

Number of movies watched only by female: 35

Top 20 rated movies in age group 20 to 25:

title	rating
Inheritors, The (Die Siebtelbauern) (1998)	5.000000
Callejón de los milagros, El (1995)	5.000000
Seven Chances (1925)	5.000000
Savior (1998)	5.000000
Dingo (1992)	5.000000
Ayn Rand: A Sense of Life (1997)	5.000000
Return with Honor (1998)	5.000000
Collectionneuse, La (1967)	5.000000
Message to Love: The Isle of Wight Festival (1996)	5.000000
Late Bloomers (1996)	5.000000

King in New York, A (1957)	5.000000
Spiders, The (Die Spinnen, 1. Teil: Der Goldene...	5.000000
Belizaire the Cajun (1986)	5.000000
Feast of July (1995)	5.000000
Ulysses (Ulissee) (1954)	5.000000
Gate of Heavenly Peace, The (1995)	5.000000
Butterfly Kiss (1995)	5.000000
Harmonists, The (1997)	4.857143
Apple, The (Sib) (1998)	4.750000
T-Men (1947)	4.750000

Most watched movie in 20 to 25 age group:
American Beauty (1999)

(3439,)

Out[1]: "We can see that as the number of zip codes on x axis are many\nthe bar plot doesn't show any clear information\n"

In [2]: