MOVIE-LENS CASE STUDY

|  |
| --- |
| import pandas as pd import numpy as np import matplotlib.pyplot as PI t %matplotlib inline |

In [70] :

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | movie\_col\_names=[ "MovieID" , "title" , "Genre" ] ratings\_col\_names=["User1D" , "Movie1D " , "Rating" , "Timestamp"] users\_col\_names=["User1D " , "Gender" , "Age" , "Occupation" , "Zip-code"] | |

# READING DATA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | movie\_data=pd . read\_csv( ' movies.dat ' , delimiter= ' • ' movie data | names-movie col names, | engine= ' python ' ) | |

Out [ 3 ] : MovielD title Genre

1. Toy story (1995) AnimationlChildrenIslComedy
2. Jumanji (1995) AdventurelChildrenIslFantasy

2 3 Grumpier Old Men (1995) ComedylRomance 3 4 Waiting to Exhale (1995) ComedylDrama

4 5 Father of the Bride Part Il (1995) Comedy

3878 3948 Meet the Parents (2000) Comedy 3879 3949 Requiem for a Dream (2000) Drama 3880 3950 Tigerland (2000) Drama 3881 3951 Two Family House (2000) Drama 3882 3952 Contender, The (2000) DramalThriller

3883 rows x 3 columns

|  |  |
| --- | --- |
| ratings\_data=pd . read\_csv( ' ratings . dat ' , delimiter= ' ratings data  \_ | engine= ' python ' ) |

In

Out [4] : UserID MovielD Rating Timestamp

## 1 1193 5 978300760

1. 1 661 3 978302109
2. 1 914 3 978301968
3. 1 3408 4 978300275
4. 1 2355 5 978824291
5. 6040 1091 1 956716541
6. 6040 1094 5 956704887
7. 6040 562 5 956704746
8. 6040 1096 4 956715648
9. 6040 1097 4 956715569
10. rows x 4 columns

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | users\_data=pd . read\_csv( ' users.dat' , delimiter=' • users data | names-users col names, | engine= ' python ' ) | |

Out [ 5 ] : UserID Gender Age Occupation Zip-code

1.  1 10 48067
2. M 56 16 70072
3. 3 M 25 15 55117
4. 4 M 45 7 02460

## 4 5 M 25 20 55455

6035 6036  25 15 32603 UserID Gender Age Occupation Zip-code

|  |  |  |  |
| --- | --- | --- | --- |
| 6036 6037 F | 45 | 1 | 76006 |
| 6037 6038 F | 56 | 1 | 14706 |
| 6038 6039 F | 45 | o | 01060 |
| 1. 6040 M 2. rows x 5 columns | 25 | 6 | 11106 |
|  | |  |  |  |  | | --- | --- | --- | --- | | movie\_ratings\_combo=pd . merge (movie movie\_ ratings\_ combo | data, | ratings \_data, | 'Movie1D' ) | | | | | | |

Out [6] : MovielD title Genre UserID Rating Timestamp

1 Toy story (1995) AnimationlChildrenIslComedy 1 5 978824268

* 1. 1 Toy story (1995) AnimationlChildrenIslComedy 6 4 978237008
  2. 1 Toy story (1995) AnimationlChildrenIslComedy 8 4 978233496
  3. 1 Toy story (1995) AnimationlChildrenIslComedy 9 5 978225952
  4. 1 Toy story (1995) AnimationlChildrenIslComedy 10 5 978226474

1. 3952 Contender, The (2000) DramalThriller 5812 4 992072099
2. 3952 Contender, The (2000) DramalThriller 5831 3 986223125
3. 3952 Contender, The (2000) DramalThriller 5837 4 1011902656
4. 3952 Contender, The (2000) DramalThriller 5927 1 979852537
5. 3952 Contender, The (2000) DramalThriller 5998 4 1001781044
6. rows x 6 columns

# CONCAT ALL 3 DATASETS TO MASTER DATA

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | master\_data=pd . merge (movie\_ ratings \_combo, master data | users data, | on: ' UserID | |

Out [7] : Zip-

MovieID title Genre UserlD Rating Timestamp Gender Age Occupation

code

1 Toy story (1995) AnimationlChildrenIslComedy 1 5 978824268 10 48067

48 Pocahontas (1995) AnimationlChildrenIslMusicallRomance 1 5 978824351 10 48067

1. 150 Apollo 13 (1995) Drama 1 5 978301777 10 48067

Star Wars: Episode IV

1. 260 ActionlAdventurelFantasylSci-Fi 1 4 978300760 10 48067

A New Hope (1977)

1. 527 Schindler's List (1993) DramalWar 1 5 978824195 10 48067

Rules of Engagement

1. 3513 DramalThriller 5727 4 958489970 M 25 4 92843

(2000)

American Psycho

1. 3535 ComedylHorrorlThriller 5727 2 958489970 M 25 4 92843

(2000)

Keeping the Faith

1. 3536 ComedylRomance 5727 5 958489902 M 25 4 92843

(2000)

1. 3555 U-571 (2000) ActionlThriller 5727 3 958490699 M 25 4 92843
2. 3578 Gladiator (2000) ActionlDrama 5727 5 958490171 M 25 4 92843
3. rows x 10 columns

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | #master\_data=master\_data. drop master data | ' Timestamp ' | 'Zip-code ', 'Genre 'l, axis=l) | |

Out [8] : Zip-

MovieID title Genre UserID Rating Timestamp Gender Age Occupation

code

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 1 | 5 | 978824268 | 10 | 48067 |
|  | 48 | Pocahontas (1995) | AnimationlChildrenIslMusicallRomance | 1 | 5 | 978824351 | 10 | 48067 |
| 2 | 150 | Apollo 13 (1995) | Drama | 1 | 5 | 978301777 | 10 | 48067 |
| 3 | 260 | Star Wars: Episode IV | ActionlAdventurelFantasylSci-Fi | 1 | 4 | 978300760 | 10 | 48067 |

A New Hope (1977)

Zip-

MovielD title Genre UserID Rating Timestamp Gender Age Occupation

code

4 527 Schindler's List (1993) DramalWar 1 5 978824195 F 1 10 48067

Rules of Engagement

1. 3513 DramalThriller 5727 4 958489970 M 25 4 92843

(2000)

American Psycho

1. 3535 ComedylHorrorlThriller 5727 2 958489970 M 25 4 92843

(2000)

Keeping the Faith

1. 3536 ComedylRomance 5727 5 958489902 M 25 4 92843

(2000)

1. 3555 U-571 (2000) ActionlThriller 5727 3 958490699 M 25 4 92843 1000208 3578 Gladiator (2000) ActionlDrama 5727 5 958490171 M 25 4 92843

1000209 rows x 10 columns

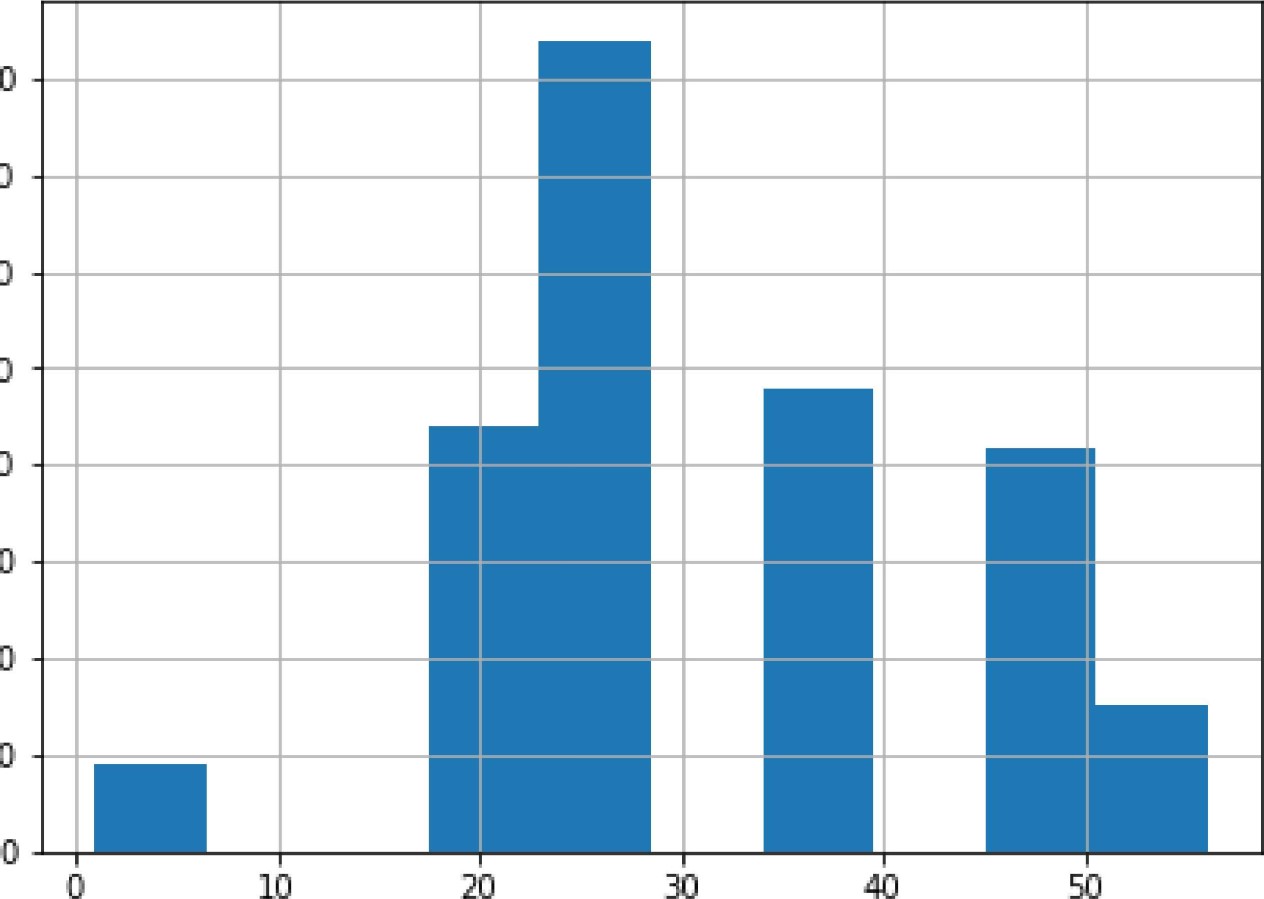
|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | #master\_data[master\_data[ 'Occupation ' ]==10]. value\_ counts ( ) . | ' bar ' ) | |

VISUAL REPRESENTATION

 # user age distribution plt.figure(figsize=(7,5)) users \_data . Age. hist( ) plt.title( 'User Age Distribution ' ) plt.xlabel( ' Age' ) plt.ylabel( ' Number of Users ' ) #plt. show()

Out [10] : Text(O, 0. 5, ' Number of Users ' )

user Age Distribution

2000

1750

1500

1250

1000

750

500

250

Age

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | #movies\_ratings\_group=movie\_ratings\_combo. groupby( ' title ' ) toy\_story=movie\_ratings\_combo[movie\_ratings\_combo[ ' title ' | 'Toy Story (1995) ' ) ] | |

# toy\_story[toy\_story[ ' Rating'

Out [12] : MovieID title Genre UserlD Rating Timestamp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | Toy Story (1995) | AnimationlChildren'slComedy | 1 | 5 | 978824268 |
|  | 1 | Toy Story (1995) | AnimationlChildren'slComedy | 6 | 4 | 978237008 |
| 2 | 1 | Toy story (1995) | AnimationlChildren'slComedy | 8 | 4 | 978233496 |
| 3 | 1 | Toy story (1995) | AnimationlChildren'slComedy | 9 | 5 | 978225952 |
| 4 | 1 | Toy story (1995) | AnimationlChildren'slComedy | 10 | 5 | 978226474 |
| 2070 | 1 | Toy Story (1995) | AnimationlChildren'slComedy | 6016 | 4 | 956778750 |
| 2072 | 1 | Toy story (1995) | AnimationlChildren'slComedy | 6022 | 5 | 956755763 |

In [13] :

Out[13] :

I n 14 ] :

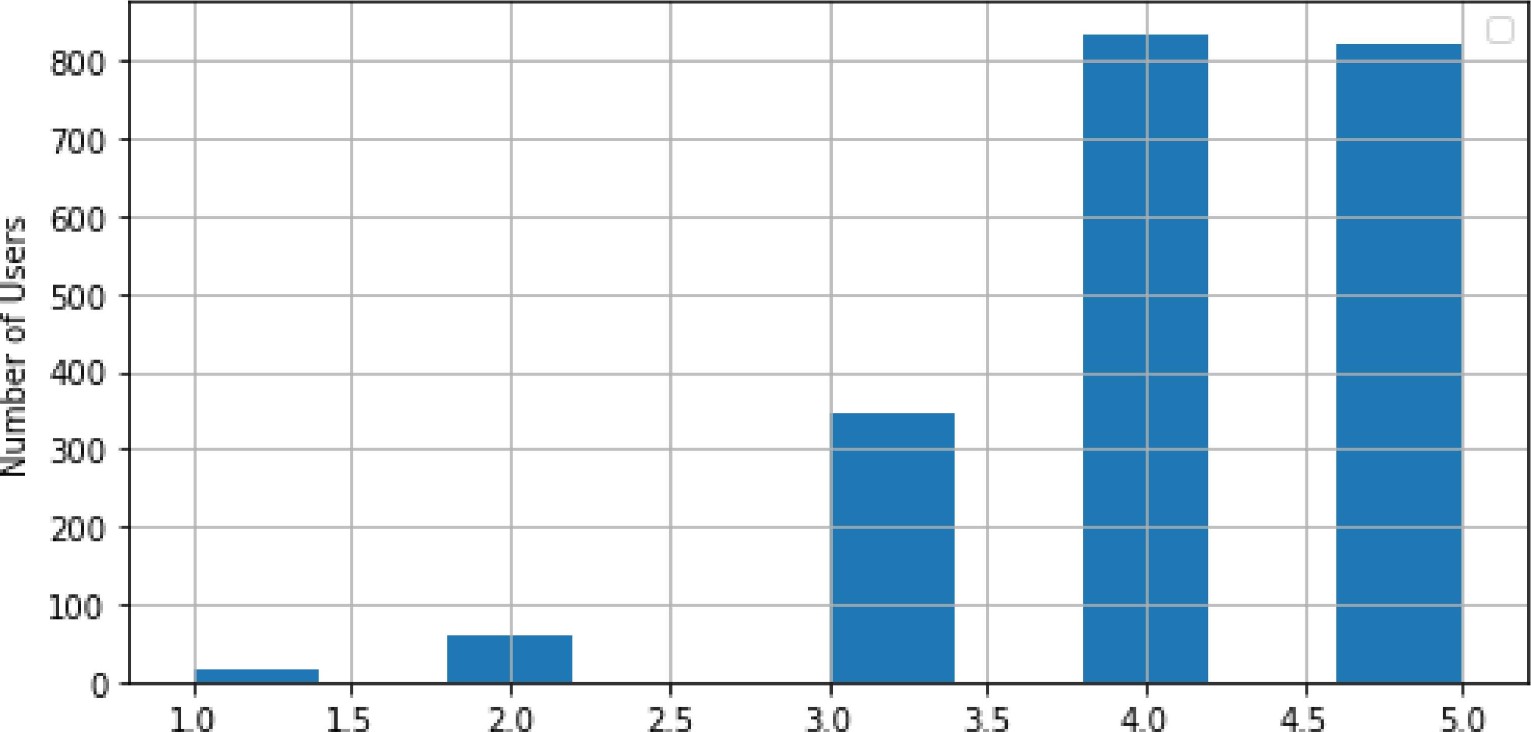
|  |  |
| --- | --- |
| plt.figure(figsize=(8,4)) toy\_story[ Rating' ] . hist() plt.title( 'User rating of the movie plt.xlabel( Rating' ) plt.ylabel( ' Number of Users' ) PIt . legend() | "Toy Story" |

No handles with labels found to put in legend.

<matplotlib.legend. Legend at Ox213dc4a2d90>

User rating of the movie "Toy Story"

3.0



25

Rating

|  |  |
| --- | --- |
| rating\_count rating\_count | = movie\_ratings\_combo.groupby( 'title ' ) [ 'Rating' ]  = rating\_count.count() |

NUMBER OF RATINGS PER MOVIE

MovielD title Genre UserlD Rating Timestamp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2073 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 6025 | 5 | 956812867 |
| 2074 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 6032 | 4 | 956718127 |
| 2075 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 6035 | 4 | 956712849 |

1655 rows x 6 columns

|  |
| --- |
| head( ) rating\_ count [ : 25] |

# Out [14] : title

American Beauty (1999) 3428

Star Wars: Episode IV - A New Hope (1977) 2991

Star Wars: Episode V - The Empire Strikes Back (1980) 2990

Star Wars: Episode VI - Return of the Jedi (1983) 2883

Jurassic Park (1993) 2672

Saving Private Ryan (1998) 2653

Terminator 2: Judgment Day (1991) 2649

## Matrix, The (1999) 2590

Back to the Future (1985) 2583

Silence of the Lambs, The (1991) 2578

Men in Black (1997) 2538

## Raiders of the Lost Ark (1981) 2514

Fargo (1996) 2513

Sixth Sense, The (1999) 2459 Braveheart (1995) 2443

Shakespeare in Love (1998) 2369 Princess Bride, The (1987) 2318 Schindler' s List (1993) 2304 L.A. Confidential (1997) 2288 Groundhog Day (1993) 2278

# E.T. the Extra-Terrestrial (1982) 2269

Star Wars: Episode I - The Phantom Menace (1999) 2250

Being John Malkovich (1999) 2241

Shawshank Redemption, The (1994) 2227

Godfather, The (1972) 2223 Name: Rating, dtype: int64

|  |
| --- |
| . groupby( 'title ' ) [ ' Rating' ] . mean ( ) avg\_ rating-avg\_ rating. sort\_values (ascending-False) avg\_ rating. head( ) |

In [15] :

# Out [15] : title Ulysses (Ulisse) (1954) 5.0

Lured (1947) 5.0

Follow the Bitch (1998) 5.0 Bittersweet Motel (2000) 5.0

## Song of Freedom (1936) 5.0 Name: Rating, dtype: float64

|  |
| --- |
| avg\_ rating\_count -- pd.DataFrame(data=avg\_rating) avg\_rating\_count[ '  - pd . DataFrame ( rating\_count ) avg\_ rating\_ count . head ( ) |

In [16] :

Out[16] : Rating number\_of\_ratings

title

|  |  |  |
| --- | --- | --- |
| Ulysses (Ulisse) (1954) | 5.0 | 1 |
| Lured (1947) | 5.0 | 1 |
| Follow the Bitch (1998) | 5.0 | 1 |
| Bittersweet Motel (2000) | 5.0 | 1 |
| Song of Freedom (1936) | 5.0 | 1 |

### grp=avg\_rating\_count . groupby( 'title ' ) grp.get\_group( ' Lured (1947)

Out [17] : Rating number\_of\_ratings

title

Lured (1947) 5.0 1

|  |  |  |
| --- | --- | --- |
| filter data - avg\_rating\_count[avg\_rating\_count[ ' number\_of\_ ratings ' ] > 10]  ' Rating' ] . sort\_values (ascending=Fa1se) | | |
| title  Sanjuro (1962)  Seven Samurai (The Magnificent Seven) (Shichinin no  Shawshank Redemption, The (1994)  Godfather, The (1972)  Close Shave, A (1995)  Carnosaur 2 (1995)  Amityville 3-D (1983)  3 Ninjas: High Noon On Mega Mountain (1998)  Turbo: A Power Rangers Movie (1997)  Carnosaur 3: Primal Species (1996)  Name: Rating, Length: 3233, dtype: float64 | samurai) (1954) | 4. 608696  4. 560510  4.554558  4. 524966  4. 520548  1.461538  1 . 372093  1.361702  1.318182  1.058824 |
| filter\_data [ : 25] |  |  |

In [18] :

Out [18] :

In [19] :

Out [19] : Rating number\_of\_ratings

|  |  |  |
| --- | --- | --- |
| Sanjuro (1962) | 4.608696 | 69 |
| Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954) | 4.560510 | 628 |
| Shawshank Redemption, The (1994) | 4.554558 | 2227 |
| Godfather, The (1972) | 4.524966 | 2223 |
| Close Shave, A (1995) | 4.520548 | 657 |
| Usual Suspects, The (1995) | 4.517106 | 1783 |
| Schindler's List (1993) | 4.510417 | 2304 |
| Wrong Trousers, The (1993) | 4.507937 | 882 |
| Sunset Blvd. (a.k.a. Sunset Boulevard) (1950) |  | 470 |
| Raiders of the Lost Ark (1981) | 4.477725 | 2514 |
| Rear Window (1954) | 4.476190 | 1050 |
| Paths of Glory (1957) | 4.473913 | 230 |
| Star Wars: Episode IV - A New Hope (1977) | 4.453694 | 2991 |
| Third Man, The (1949) | 4.452083 | 480 |
| Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963) | 4.449890 | 1367 |
| For All Mankind (1989) | 4.444444 | 27 |
| Wallace & Gromit: The Best of Aardman Animation (1996) | 4.426941 | 438 |
| To Kill a Mockingbird (1962) | 4.425647 | 928 |
| Double Indemnity (1944) | 4.41 5608 | 551 |
| Casablanca (1942) | 4.412822 | 1669 |
| World of Apu, The (Apur Sansar) (1959) | 4.410714 | 56 |
| Sixth Sense, The (1999) | 4.406263 | 2459 |
| Yojimbo (1961) | 4.404651 | 215 |
| Pather Panchali (1955) | 4.404255 | 47 |

Rating number\_of\_ratings

Lawrence of Arabia (1962) 4.401925 831

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | user 2696 user 2696 | = movie | ratings | combo | ' UserID ' ] | - 2696] | |

Out [20] : MovielD title Genre UserlD Rating Timestamp

95261 350 Client, The (1994) DramalMysterylThriller 2696 3 973308886 200536 800 Lone Star (1996) DramalMystery 2696 5 973308842

270342 1092 Basic Instinct (1992) MysterylThriller 2696 4 973308886

274576 1097 E.T. the Extra-Terrestrial (1982) ChildrenIslDramalFantasylSci-Fi 2696 3 973308690

349333 1258 Shining, The (1980) Horror 2696 4 973308710

360382 1270 Back to the Future (1985) ComedylSci-Fi 2696 2 973308676

448293 1 589 cop Land (1997) CrimelDramalMystery 2696 3 973308865

457193 1617 L.A. Confidential (1997) CrimelFilm-NoirlMysterylThriller 2696 4 973308842 459835 1625 Game, The (1997) MysterylThriller 2696 4 973308842

464473 1644 I Know What You Did Last Summer (1997) HorrorlMysterylThriller 2696 2 973308920 465069 1645 Devil's Advocate, The (1997) CrimelHorrorlMysterylThriller 2696 4 973308904 480658 1711 Midnight in the Garden of Good and Evil (1997) ComedylCrimelDramalMystery 2696 4 973308904

493231 1783 Palmetto (1998) Film-NoirlMysterylThriller 2696 4 973308865 496926 1805 Wild Things (1998) CrimelDramalMysterylThriller 2696 4 973308886

507001 1892 Perfect Murder, A (1998) MysterylThriller 2696 4 973308904

631867 2338 I Still Know What You Did Last Summer (1998) HorrorlMysterylThriller 2696 2 973308920

645948 2389 Psycho (1998) CrimelHorrorlThriller 2696 4 973308710

732119 2713 Lake Placid (1999) HorrorlThriller 2696 1 973308710

851313 3176 Talented Mr. Ripley, The (1999) DramalMysterylThriller 2696 4 973308865

MovielD Genre UserlD Rating Timestamp

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  | | --- | --- | --- | | 889623 3386 | JFK (1991) | DramalMystery 2696 1 973308842 | | movie\_ratings\_combo |  |  | |

Out [21] : MovielD title Genre UserID Rating Timestamp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 1 | 5 | 978824268 |
| 1 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 6 | 4 | 978237008 |
| 2 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 8 | 4 | 978233496 |
| 3 | 1 | Toy story (1995) | AnimationlChildrenIslComedy | 9 | 5 | 978225952 |
| 4 | 1 | Toy story (1995) | AnimationlChildren'slComedy | 10 | 5 | 978226474 |
|  | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 1. 3952 2. 3952 3. 3952 4. 3952 5. 3952 6. rows x 6 | Contender, The (2000)  Contender, The (2000)  Contender, The (2000)  Contender, The (2000)  Contender, The (2000)  columns | DramalThriller  DramalThriller  DramalThriller  DramalThriller  DramalThriller | 5812  5831  5837  5927  5998 | 4  3  4  1  4 | 992072099  986223125  1011902656  979852537  1001781044 | | movie\_ratings\_combo[ ' Genre' ] . value\_counts ( ) | |  |  |  |  | | | | | | | | | |

Out [22] : Comedy 116883

Drama 111423

Comedy I Romance 42712 Comedy I Drama 42245 Drama I Romance 29170

Drama I Romance I Western 29 Children ' s Fantasy 27

Comedy I Film-Noir I Thriller 5

Film-Noir I Horror 2

Fantasy 1

Name: Genre, Length: 301, dtype: int64

Genre category with a one-hot encoding ( 1 and 0)

In [23] :

master data

Out [23] : Zip-

MovieID title Genre UserID Rating Timestamp Gender Age Occupation code

1 Toy story (1995) AnimationlChildrenIslComedy 1 5 978824268 F 1 10 48067

48 Pocahontas (1995) AnimationlChildrenIslMusicallRomance 1 5 978824351 F 1 10 48067

1. 150 Apollo 13 (1995) Drama 1 5 978301777 F 1 10 48067

Star Wars: Episode IV

1. 260 ActionlAdventurelFantasylSci-Fi 1 4 978300760 F 1 10 48067

A New Hope (1977)

1. 527 Schindler's List (1993) DramalWar 1 5 978824195 F 1 10 48067

Rules of Engagement

1. 3513 DramalThriller 5727 4 958489970 M 25 4 92843

(2000)

American Psycho

1. 3535 ComedylHorrorlThriller 5727 2 958489970 M 25 4 92843

(2000)

Keeping the Faith

1. 3536 ComedylRomance 5727 5 958489902 M 25 4 92843

(2000)

1. 3555 U-571 (2000) ActionlThriller 5727 3 958490699 M 25 4 92843
2. 3578 Gladiator (2000) ActionlDrama 5727 5 958490171 M 25 4 92843
3. rows x 10 columns

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | | movie\_ratings\_selected\_features\_df  'Gender ' ,  ' Age' ,  'Occupation ' ,  ' Rating ' ,  I Genre' | = master\_data[[ | |

### movie\_ratings\_selected\_features\_df

Out [25] : Gender Age Occupation Rating Genre

F 1 10 5 AnimationlChildrenIslComedy

F 1 10 5 AnimationlChildrenIslMusicallRomance

1. F 1 10 5 Drama
2. F 1 10 4 ActionlAdventurelFantasylSci-Fi
3. F 1 10 5 DramalWar
4. M 25 4 4 DramalThriller
5. M 25 4 2 ComedylHorrorlThriller
6. M 25 4 5 ComedylRomance
7. M 25 4 3 ActionlThriller
8. M 25 4 5 ActionlDrama
9. rows x 5 columns

|  |
| --- |
| Genre=movie\_ratings\_selected\_features\_df[ ' Genre ]  Genre=Genre. str.  . add \_prefix( ' Genre\_ ) |

In [35] :

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  | | --- | --- | --- | --- | --- | | movie ratings\_Genres\_df - pd . concat( |  | , Genre] , axis-I ) |  |  | |  |  |  |  |  | | movie• \_ ratings\_Genres\_df |  |  |  |  | | Gender Age Occupation Rating | Genre | Genre Action Genre Adventure | Genre Animation | Genre Child | |

In [37] :

Out[37] :

|  |  |  |
| --- | --- | --- |
| 1 10 | 5 AnimationlChildrenIslComedy | 1 |
| 1 1 10 | 5 AnimationlChildrenIslMusicallRomance | 1 |

Gender Rating Genre Genre Action Genre Adventure Genre Animation Genre Child



10

Drama

10 ActionlAdventurelFantasylSci-Fi

10 DramalWar

1000204 M 25 DramalThriller 1000205 M 25 ComedylHorrorlThriller 1000206 M 25 ComedylRomance 1000207 M 25 ActionlThriller 1000208 M 25 ActionlDrama

1000209 rows x 23 columns



|  |  |
| --- | --- |
| movie\_ ratings\_Genres\_df=pd . get \_dummies (movie\_ratings\_Genres\_df , columns: [ movie\_ ratings\_Genres\_df | Gender |

In [38] :

Out [38] :

Age Occupation Rating Genre Genre Action Genre Adventure Genre Animation Genre Children's

10 AnimationlChildrenIslComedy

10 5 AnimationlChildrenIslMusicallRomance

10 Drama

10 ActionlAdventurelFantasylSci-Fi

10 DramalWar

1. 25 DramalThriller
2. 25 ComedylHorrorlThriller 1000206 25 ComedylRomance

Rating Genre Genre Action Genre Adventure Genre Animation Genre Children's



25

ActionlThriller

1. 25 4 5 ActionlDrama 1
2. rows x 24 columns



ln [39] :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | .oooooo | 0.078371 | 0.056869 | -0.030975 | -O.016730 | -0.047020 | -0.052858 | -0.044046 |
| Occupation | 0.078371 | .oooooo | 0.006753 | 0.018347 | 0.014309 | -0.003834 | -0.006906 | -0.006149 |
| Rating | 0.056869 | 0.006753 | .oooooo | -0.047633 | -0.036718 | O.019670 | -0.039829 | -0.039622 |
| Genre Action | -0.030975 | O.OI 8347 | -0.047633 | .oooooo | 0.374961 | -0.110294 | -0.141314 | -0.268092 |
| Genre Adventure | -O.OI 6730 | O.014309 | -0.036718 | 0.374961 | .oooooo | 0.004732 | 0.098283 | -o. 124960 |
| Genre Animation | -0.047020 | -0.003834 | 0.019670 | -0.110294 | 0.004732 | .oooooo | 0.576204 | 0.018544 |
| Genre Children•s | -0.052858 | -0.006906 | -0.039829 | -0.141314 | 0.098283 | 0.576204 | .oooooo | 0.058711 |
| Genre\_Comedy | -0.044046 | -0.006149 | -0.039622 | -0268092 | -o. 124960 | 0.018544 | 0.058711 | .oooooo |
| Genre Crime | -0.007931 | 0.002821 | 0.033446 | 0.088519 | -0.045924 | -0.062520 | -0.081977 | -0.078030 |
| Genre\_Documentary | 0.004407 | -0.002689 | 0.028098 | -0.052565 | -0.035109 | -O.OI 8991 | -0.024901 | -0.040697 |

### movie\_ratings\_Genres df.columns

## Out[39] :'Age ' , 'Occupation ' , ' Rating' , 'Genre ' , 'Genre Action '

' Genre Adventure' , 'Genre Animation ' , 'Genre Children ' s ' ,

' Genre\_Comedy 'Genre Crime' ' Genre\_Documentary ' 'Genre Drama 

Genre Fantasy' 'Genre Film-Noir i , 'Genre Horror' 'Genre Musical ' ,

' Genre\_Mystery ' , 'Genre Romance' , 'Genre Sci- Fi ' , 'Genre Thriller ' ,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| movie\_ratings\_Genres\_df . corr( ) |  |  |  |  |  |  |  |
| Age Occupation | Rating | Genre Action | Genre Adventure | Genre Animation | Genre Children's | Genre\_Comedy | Gen |

' Genre War' , 'Genre Western' , 'Gender F ' , 'Gender M dtype= object ' ) ln [41] :

Out [41] :

Rating Genre Action Genre Adventure Genre Animation Genre Children's Genre\_Comedy Gen

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Genre Drama | 0.063856 | -0.012326 | 0.122561 | -0.202415 | -0.194570 | -0.1 54479 | -0.135707 | -0.249840 | | Genre\_Fantasy | -0.024222 | 0.001299 | -0.023312 | 0.014551 | 0.227046 | 0.012025 | 0.263280 | -0.006010 | | Genre Film-Noir | 0.033495 | 0.005246 | 0.060259 | -0.080288 | -0.014178 | 0.037013 | -0.038033 | -0.101425 | | Genre Horror | -0.023901 | 0.001439 | -0.094353 | -0.042733 | -0.057256 | -0.049730 | -0.077099 | -0.093064 | | Genre Musical | 0.005158 | -0.007312 | 0.015643 | -o. 100432 | -0.022327 | 0.335231 | 0.312567 | 0.030566 | | Genre\_Mystery | 0.024308 | 0.002421 | 0.015848 | -0.054084 | -0.043503 | -0.042488 | -0.052786 | -o. 105346 | | Genre Romance | 0.017503 | -0.014018 | O. 009644 | -0.067830 | -0.024389 | -0.054540 | -0.084550 | o. 1 12843 | | Genre Scl-Fi• | -0.010879 | 0.026250 | -0.044487 | 0.319117 | 0.284190 | -0.055526 | -0.038844 | -o. 187079 | | Genre Thriller | -O.014100 | 0.008981 | -0.004806 | 0.202756 | -0.038423 | -0.085713 | -o. 132642 | -0.299501 | | Genre War | 0.038446 | 0.010264 | 0.075688 | o. 135872 | 0.016647 | ¯O.0461 | -0.066539 | -o. 127101 | | Genre Western | 0.038177 | 0.005924 | 0.00731 1 | 0.022242 | -0.01 1964 | -0.030908 | -0.031269 | 0.007927 | | Gender F | 0.003189 | -0.1 14974 | 0.019861 | -0.094380 | -0.038645 | 0.017719 | 0.031662 | 0.040758 | |  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gender M -0.003189 0.1 14974 -0.019861 0.094380  23 rows x 23 columns    Features affecting the ratings of any particular movie. | 0.038645 | -0.017719 | -0.031662 | -0.040758 |
| movie\_ratings\_Genres\_df=movie\_ratings\_Genres\_df. drop ( [ I Genre I ] , | axis=l) |  |  |  |
|  |  |  |  |  |
| movie\_ratings\_Genres\_df. dtypes |  |  |  |  |

In [46] :

ln [47] :

Out [47] : Age int64

Occupation int64

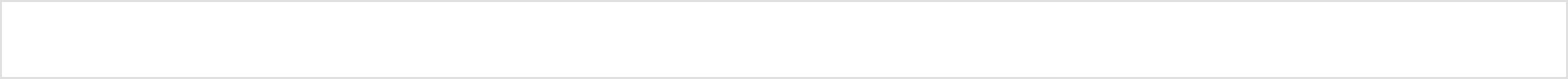
Rating int64

Genre Action int64

Genre Adventure int64

Genre Animation int64

Genre Children ' s int64

In [48] : 

|  |  |  |
| --- | --- | --- |
|  | |  | | --- | | movie\_ratings\_Genres\_df\_samp1e=movie\_ratings\_Genres\_df. s ample ( n=seaea, random state=ê | |

|  |  |
| --- | --- |
| Genre\_Comedy | int64 |
| Genre Crime | int64 |
| Genre\_Documentary | int64 |
| Genre Drama | int64 |
| Genre Fantasy | int64 |
| Genre Film-Noir | int64 |
| Genre Horror | int64 |
| Genre Musical | int64 |
| Genre\_Mystery | int64 |
| Genre Romance | int64 |
| Genre Sci-Fi | int64 |
| Genre Thriller | int64 |
| Genre War | int64 |
| Genre Western | int64 |
| Gender F | uint8 |
| Gender M dtype: object | uint8 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| movie\_ratings\_Genres\_df\_samp1e |  |  |  |  |  |  |
| Age Occupation Rating Genre\_Action | Genre Adventure | Genre Animation | Genre Children's | Genre\_Comedy | Genre Crime | Genre Do |

Out [61] :

324271 18 4 4 1

818637 18 4 3 1 1

148677 18 14 5

778790 50 7 4 1 1

525489 25 2 5

106711 25 20 2 1

Age Occupation Rating Genre Action Genre Adventure Genre Animation Genre Children's Genre\_Comedy Genre Crime Genre Do

792958 25 14 4 1

735824 56 1 4

42804 18 14 4 1

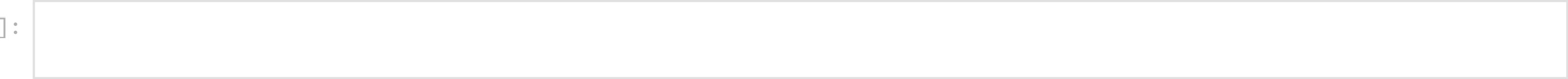
177795 35 1 3 1 1

50000 rows x 23 columns



|  |
| --- |
| x=movie\_ratings\_Genres\_df\_samp1e . drop( [ ' Rating ' ] , axis-I) y=movie\_ratings\_Genres\_df\_samp1e[ Rating ' ] |

In [62] :



MODEL TRAIN

|  |
| --- |
| from sklearn.linear\_model import LinearRegression from sklearn.model\_selection import train\_test\_split from sklearn import metrics  Ir=LinearRegression(copy\_X=True, fit \_intercept-True, n\_jobs=None, normalize=Fa1se) |

In [63] :

In [64] : x\_train , x\_test , y\_train , y\_test=train\_test\_split (x, y , test\_size=@.2, random\_state=@)

In [65] :

Ir. fit(x\_train,y\_train)

|  |  |
| --- | --- |
| Out [65] : | LinearRegression() |

|  |
| --- |
| predict (x\_test) |

In [66] :

METRICS EVALUATION

|  |
| --- |
| MSE-metrics . y\_pred)  MASE=metrics . mean\_absolute\_error(y\_test,y\_pred)  RMSE=np. sq rt (metrics . mean \_ squared\_error (y\_test, y\_pred) )    . r2\_score (y\_test, y\_pred) |

In [67] :

In [68] : print ( "MSE : " , MSE) print ( "MASE : " , MASE) print( " RMSE : " , RMSE) print ( " r2\_score : " , r2\_score)

|  |
| --- |
| result-pd .  ' ACTUAL : y\_test, 'PREDICTED' : y\_pred}) result |

MSE: 1.1977731707567232 MASE: e. 8978299534841195 RMSE: 1.0944282391992282 r2 score: 0.03795269985311833 In [69] :

|  |  |  |
| --- | --- | --- |
| Out [69] : | ACTUAL | PREDICTED |

|  |  |  |
| --- | --- | --- |
| 187446 | 4 | 4.322363 |
| 69421 | 4 | 3.439548 |
| 941725 | 3 | 3.408593 |
| 841836 | 4 | 3.652663 |
| 869012 | 4 | 3.559433 |
| 178851 | 5 | 3.733527 |
| 374575 | 5 | 3.915716 |
| 96738 | 5 | 3.784465 |
| 727461 | 3 | 3.454760 |
| 28123 | 1 | 3.521620 |

10000 rows x 2 columns

