Oracle

Aadish Chopra

10/20/2019

Loading Data

1. Links are used to extract the data
2. Column Names from the readme file

```
U.Data<-read.csv("http://files.grouplens.org/datasets/movielens/ml-100k/u.data",header = FALSE,sep = "\t",col.name
s = c("User_ID","Item_ID","Rating","TimeStamp"))

U.Item<-read.csv("http://files.grouplens.org/datasets/movielens/ml-100k/u.item",header=FALSE,sep="|",col.names = c
("movie_id","movie_title","release_date","video_release_date","IMDb URL", "unknown","Action","Adventure","Animatio
n","Children's","Comedy","Crime",    "Documentary","Drama","Fantasy","Film-Noir","Horror","Musical","Mystery","Ro
mance"     ,"Sci-Fi","Thriller","War","Western"),as.is = TRUE)

U.User<-read.csv("http://files.grouplens.org/datasets/movielens/ml-100k/u.user",header=FALSE,sep="|",col.names = c
("User_ID","age","gender","occupation","zip code"))


U.Genres<-read.csv("http://files.grouplens.org/datasets/movielens/ml-100k/u.genre",header=FALSE,sep="|",col.names
= c("genre","Number"))
```

1. Remove Video Release date as it has no data

```
#Remove "Video Release date" as it has no data

#sum(is.na(U.Item[,4]))
U.Item<-U.Item[,-4]
```

Number of missing values in video_release_date is 0

Programming Assignments

Assumptions:

> Average Rating to be the selection criterion (higher is better).

> Ties are broken by counts.

> Secondary level ties are ignored and the item which comes first is selected

Top 3 movies by Occupation

```
#merge the dataset

UData_User<-merge(U.Data,y = U.User,by = "User_ID")

# group by occupation,item_id
groupby<-by(UData_User,list(UData_User$occupation,UData_User$Item_ID),FUN = function(x)
  {
  data.frame(occupation=unique(x$occupation),
  Item_ID=unique(x$Item_ID),
  mean_rating=mean(x$Rating),
  Count_Item_ID=nrow(x)
  )
  })

Top3Occupation<-do.call(rbind,groupby)
Top3Occupation<-Top3Occupation[with(Top3Occupation,order(occupation,-mean_rating,-Count_Item_ID)),]


# pick top 3

Top3Occupation<-by(Top3Occupation,list(Top3Occupation$occupation),head,n=3)
Top3Occupation<-do.call(rbind.data.frame,Top3Occupation)
rownames(Top3Occupation)<-NULL

# Get the names of the movie

Top3Occupation<-merge(Top3Occupation,U.Item[,1:2],by.x = "Item_ID",by.y = "movie_id",sort = FALSE)
Top3Occupation<-Top3Occupation[with(Top3Occupation,order(occupation,-mean_rating,-Count_Item_ID)),]
knitr::kable(Top3Occupation,caption = "Top 3 Movies by Occupation")
```

Top 3 Movies by Occupation

	Item_ID	occupation	mean_rating	Count_Item_ID	movie_title
1	408	administrator	5	5	Close Shave, A (1995)
3	251	administrator	5	3	Shall We Dance? (1996)
4	359	administrator	5	2	Assignment, The (1997)
2	408	artist	5	4	Close Shave, A (1995)
6	169	artist	5	3	Wrong Trousers, The (1993)
9	505	artist	5	3	Dial M for Murder (1954)
10	474	doctor	5	2	Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963)
11	483	doctor	5	2	Casablanca (1942)
12	514	doctor	5	2	Annie Hall (1977)
13	963	educator	5	3	Some Folks Call It a Sling Blade (1993)
14	464	educator	5	2	Vanya on 42nd Street (1994)

	Item_ID	occupation	mean_rating	Count_Item_ID	movie_title
15	953	educator	5	2	Unstrung Heroes (1995)
16	611	engineer	5	2	Laura (1944)
17	1063	engineer	5	2	Little Princess, A (1995)
18	834	engineer	5	1	Halloween: The Curse of Michael Myers (1995)
19	179	entertainment	5	4	Clockwork Orange, A (1971)
20	99	entertainment	5	2	Snow White and the Seven Dwarfs (1937)
21	150	entertainment	5	2	Swingers (1996)
22	603	executive	5	4	Rear Window (1954)
24	316	executive	5	2	As Good As It Gets (1997)
25	490	executive	5	2	To Catch a Thief (1955)
26	67	healthcare	5	1	Ace Ventura: Pet Detective (1994)
27	148	healthcare	5	1	Ghost and the Darkness, The (1996)
30	320	healthcare	5	1	Paradise Lost: The Child Murders at Robin Hood Hills (1996)
31	319	homemaker	5	2	Everyone Says I Love You (1996)
32	845	homemaker	5	2	That Thing You Do! (1996)
33	22	homemaker	5	1	Braveheart (1995)
34	205	lawyer	5	4	Patton (1970)
35	488	lawyer	5	4	Sunset Blvd. (1950)
36	12	lawyer	5	3	Usual Suspects, The (1995)
37	533	librarian	5	2	Daytrippers, The (1996)
38	656	librarian	5	2	M (1931)
39	1099	librarian	5	2	Red Firecracker, Green Firecracker (1994)
40	496	marketing	5	3	It's a Wonderful Life (1946)
23	316	marketing	5	2	As Good As It Gets (1997)
41	275	marketing	5	2	Sense and Sensibility (1995)
42	64	none	5	3	Shawshank Redemption, The (1994)
43	333	none	5	3	Game, The (1997)
44	54	none	5	2	Outbreak (1995)
45	547	other	5	2	Young Poisoner's Handbook, The (1995)
46	601	other	5	2	For Whom the Bell Tolls (1943)
47	644	other	5	2	Thin Blue Line, The (1988)
48	114	programmer	5	6	Wallace & Gromit: The Best of Aardman Animation (1996)
29	320	programmer	5	1	Paradise Lost: The Child Murders at Robin Hood Hills (1996)
49	416	programmer	5	1	Old Yeller (1957)
50	498	retired	5	3	African Queen, The (1951)
8	169	retired	5	2	Wrong Trousers, The (1993)
51	306	retired	5	2	Mrs. Brown (Her Majesty, Mrs. Brown) (1997)
52	173	salesman	5	3	Princess Bride, The (1987)
53	340	salesman	5	3	Boogie Nights (1997)
54	346	salesman	5	3	Jackie Brown (1997)
55	19	scientist	5	2	Antonia's Line (1995)
56	81	scientist	5	2	Hudsucker Proxy, The (1994)
57	156	scientist	5	2	Reservoir Dogs (1992)
28	320	student	5	4	Paradise Lost: The Child Murders at Robin Hood Hills (1996)
58	279	student	5	2	Once Upon a Time… When We Were Colored (1995)
59	614	student	5	2	Giant (1956)
7	169	technician	5	5	Wrong Trousers, The (1993)
60	14	technician	5	3	Postino, Il (1994)
5	359	technician	5	2	Assignment, The (1997)
61	650	writer	5	3	Seventh Seal, The (Sjunde inseglet, Det) (1957)
62	652	writer	5	2	Rosencrantz and Guildenstern Are Dead (1990)
63	18	writer	5	1	White Balloon, The (1995)

```
write.csv(x = Top3Occupation,file = "Top3Occupation.csv")
```

Top 3 movies by Genre

```
#reversing one-hot encoding to match to the dataset

rev_one_hot<-as.data.frame(which(U.Item[,5:23]==1,arr.ind = T))
rev_one_hot$genre_transformed<-names(U.Item[,5:23])[rev_one_hot[order(rev_one_hot[,1]),2]]
Genre<-merge(U.Item,rev_one_hot,by.x = "movie_id",by.y = "row")
```

Selecting relevant columns from Genre

```
# merging Genre with UData on Item Id
UGenre<-Genre[,c(1,2,25)]
UGenreUser<-merge(UGenre,U.Data,by.x ="movie_id",by.y = "Item_ID" )

#group by Genre

groupby<-by(UGenreUser,list(UGenreUser$genre_transformed,UGenreUser$movie_id),FUN = function(x)
  {
  data.frame(genre_transformed=unique(x$genre_transformed),
  Item_ID=unique(x$movie_id),
  movie_title=unique(x$movie_title),
  mean_rating=mean(x$Rating),
  Count_Item_ID=nrow(x)
  )
  })

Top3Genre<-do.call(rbind,groupby)
Top3Genre<-Top3Genre[with(Top3Genre,order(genre_transformed,-mean_rating,-Count_Item_ID)),]


# pick top 3

Top3Genre<-by(Top3Genre,list(Top3Genre$genre_transformed),head,n=3)
Top3Genre<-do.call(rbind.data.frame,Top3Genre)
rownames(Top3Genre)<-NULL

knitr::kable(Top3Genre,caption = "Top 3 Movies by Genre")
```

Top 3 Movies by Genre

genre_transformed	Item_ID	movie_title	mean_rating	Count_Item_ID
Horror	318	Schindler's List (1993)	4.466443	298
Horror	114	Wallace & Gromit: The Best of Aardman Animation (1996)	4.447761	67
Horror	285	Secrets & Lies (1996)	4.265432	162
Romance	1122	They Made Me a Criminal (1939)	5.000000	1
Romance	169	Wrong Trousers, The (1993)	4.466102	118
Romance	513	Third Man, The (1949)	4.333333	72
Sci.Fi	127	Godfather, The (1972)	4.283293	413
Sci.Fi	493	Thin Man, The (1934)	4.150000	60
Sci.Fi	83	Much Ado About Nothing (1993)	4.062500	176
Comedy	1189	Prefontaine (1997)	5.000000	3
Comedy	1599	Someone Else's America (1995)	5.000000	1
Comedy	1653	Entertaining Angels: The Dorothy Day Story (1996)	5.000000	1
Documentary	113	Horseman on the Roof, The (Hussard sur le toit, Le) (1995)	4.111111	9
Documentary	165	Jean de Florette (1986)	4.109375	64
Documentary	490	To Catch a Thief (1955)	4.020000	50
Drama	1293	Star Kid (1997)	5.000000	3
Drama	814	Great Day in Harlem, A (1994)	5.000000	1
Drama	1642	Some Mother's Son (1996)	4.500000	2
Action	1293	Star Kid (1997)	5.000000	3
Action	1500	Santa with Muscles (1996)	5.000000	2
Action	119	Maya Lin: A Strong Clear Vision (1994)	4.500000	4
Thriller	1201	Marlene Dietrich: Shadow and Light (1996)	5.000000	1
Thriller	1398	Anna (1996)	4.500000	2
Thriller	408	Close Shave, A (1995)	4.491071	112
Crime	1293	Star Kid (1997)	5.000000	3
Crime	98	Silence of the Lambs, The (1991)	4.289744	390
Crime	172	Empire Strikes Back, The (1980)	4.204360	367
Musical	173	Princess Bride, The (1987)	4.172840	324
Musical	302	L.A. Confidential (1997)	4.161616	297
Musical	659	Arsenic and Old Lace (1944)	4.078261	115
Adventure	1293	Star Kid (1997)	5.000000	3
Adventure	963	Some Folks Call It a Sling Blade (1993)	4.292683	41
Adventure	313	Titanic (1997)	4.245714	350
Mystery	484	Maltese Falcon, The (1941)	4.210145	138
Mystery	923	Raise the Red Lantern (1991)	4.155172	58
Mystery	23	Taxi Driver (1976)	4.120879	182
War	1536	Aiqing wansui (1994)	5.000000	1
War	1594	Everest (1998)	4.500000	2
War	50	Star Wars (1977)	4.358491	583
Film.Noir	22	Braveheart (1995)	4.151515	297
Film.Noir	205	Patton (1970)	3.992647	136
Film.Noir	297	Ulee's Gold (1997)	3.960000	50
Children.s	1467	Saint of Fort Washington, The (1993)	5.000000	2
Children.s	1122	They Made Me a Criminal (1939)	5.000000	1

genre_transformed	Item_ID	movie_title	mean_rating	Count_Item_ID
Children.s	1449	Pather Panchali (1955)	4.625000	8
Western	166	Manon of the Spring (Manon des sources) (1986)	4.120690	58
Western	1505	Killer: A Journal of Murder (1995)	4.000000	1
Western	607	Rebecca (1940)	3.969697	66
Animation	89	Blade Runner (1982)	4.138182	275
Animation	56	Pulp Fiction (1994)	4.060914	394
Animation	59	Three Colors: Red (1994)	4.060241	83
unknown	152	Sleeper (1973)	3.634146	82
unknown	72	Mask, The (1994)	3.193798	129
Fantasy	921	Farewell My Concubine (1993)	3.978261	46
Fantasy	945	Charade (1963)	3.925000	40
Fantasy	1062	Four Days in September (1997)	3.750000	12

```
write.csv(x = Top3Genre,file = "Top3Genre.csv")
```

Top 3 movies by Occupation,Genre

Dataset created in **Top 3 Movies by Genre** is reused

```
# merging data

UGenreUser_Occupation<-merge(UGenreUser,U.User,by="User_ID")

#  group by occupation,genre

groupby<-by(UGenreUser_Occupation,list(UGenreUser_Occupation$occupation,UGenreUser_Occupation$genre_transformed,UG
enreUser_Occupation$movie_id),FUN = function(x)
  {
  data.frame(occupation=unique(x$occupation),
  genre_transformed=unique(x$genre_transformed),
  movie_id=unique(x$movie_id),
  mean_rating=mean(x$Rating),
  Count_Item_ID=nrow(x)
  )
  })

Top3OccupationGenre<-do.call(rbind,groupby)
Top3OccupationGenre<-Top3OccupationGenre[with(Top3OccupationGenre,order(occupation,genre_transformed,-mean_rating,
-Count_Item_ID)),]

Top3OccupationGenre<-aggregate(Top3OccupationGenre,by=list(Top3OccupationGenre$occupation,Top3OccupationGenre$genr
e_transformed),FUN = head,n=3)

kable(Top3OccupationGenre,caption = "Top 3 Movies by Occupation, Genre")
```

Top 3 Movies by Occupation, Genre

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
administrator	Horror	c(1, 1, 1)	c(1, 1, 1)	c(1147, 311, 318)	c(5, 4.66666666666667, 4.63157894736842)	c(1, 3, 19)
artist	Horror	c(2, 2, 2)	c(1, 1, 1)	c(729, 652, 317)	c(5, 4.75, 4.5)	c(1, 4, 2)
doctor	Horror	c(3, 3, 3)	c(1, 1, 1)	c(132, 183, 286)	c(5, 5, 4.6)	c(1, 1, 5)
educator	Horror	c(4, 4, 4)	c(1, 1, 1)	c(464, 1199, 114)	c(5, 5, 4.8)	c(2, 1, 5)
engineer	Horror	c(5, 5, 5)	c(1, 1, 1)	c(114, 464, 318)	c(4.5, 4.5, 4.44)	c(6, 2, 25)
entertainment	Horror	c(6, 6, 6)	c(1, 1, 1)	c(855, 285, 580)	c(5, 5, 5)	c(2, 1, 1)
executive	Horror	c(7, 7, 7)	c(1, 1, 1)	c(316, 131, 132)	c(5, 4.66666666666667, 4.57142857142857)	c(2, 3, 7)
healthcare	Horror	c(8, 8, 8)	c(1, 1, 1)	c(148, 316, 318)	c(5, 4.5, 4.4)	c(1, 2, 5)
homemaker	Horror	c(9, 9, 9)	c(1, 1, 1)	c(316, 148, 1)	c(4.5, 4, 4)	c(2, 3, 2)
lawyer	Horror	c(10, 10, 10)	c(1, 1, 1)	c(205, 293, 580)	c(5, 5, 5)	c(4, 1, 1)
librarian	Horror	c(11, 11, 11)	c(1, 1, 1)	c(285, 124, 318)	c(4.68421052631579, 4.5625, 4.52380952380952)	c(19, 16, 21)
marketing	Horror	c(12, 12, 12)	c(1, 1, 1)	c(316, 1242, 1358)	c(5, 5, 5)	c(2, 1, 1)
none	Horror	c(13, 13, 13)	c(1, 1, 1)	c(17, 132, 280)	c(5, 5, 5)	c(1, 1, 1)
other	Horror	c(14, 14, 14)	c(1, 1, 1)	c(1172, 776, 903)	c(5, 5, 5)	c(2, 1, 1)
programmer	Horror	c(15, 15, 15)	c(1, 1, 1)	c(114, 416, 837)	c(5, 5, 5)	c(6, 1, 1)
retired	Horror	c(16, 16, 16)	c(1, 1, 1)	c(902, 1169, 285)	c(5, 4.5, 4.42857142857143)	c(1, 2, 7)
salesman	Horror	c(17, 17, 17)	c(1, 1, 1)	c(82, 114, 132)	c(5, 5, 5)	c(1, 1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
scientist	Horror	c(18, 18, 18)	c(1, 1, 1)	c(652, 902, 855)	c(5, 5, 4.66666666666667)	c(1, 1, 6)
student	Horror	c(19, 19, 19)	c(1, 1, 1)	c(1367, 114, 958)	c(5, 4.71428571428571, 4.5)	c(1, 14, 2)
technician	Horror	c(20, 20, 20)	c(1, 1, 1)	c(652, 114, 318)	c(5, 4.66666666666667, 4.6)	c(2, 3, 10)
writer	Horror	c(21, 21, 21)	c(1, 1, 1)	c(652, 837, 1237)	c(5, 5, 5)	c(2, 1, 1)
administrator	Romance	c(1, 1, 1)	c(2, 2, 2)	c(308, 626, 769)	c(5, 5, 5)	c(1, 1, 1)
artist	Romance	c(2, 2, 2)	c(2, 2, 2)	c(169, 171, 362)	c(5, 5, 5)	3:1
doctor	Romance	c(3, 3, 3)	c(2, 2, 2)	c(187, 462, 523)	c(5, 5, 5)	c(1, 1, 1)
educator	Romance	c(4, 4, 4)	c(2, 2, 2)	c(1524, 169, 656)	c(5, 4.5, 4.5)	c(2, 8, 4)
engineer	Romance	c(5, 5, 5)	c(2, 2, 2)	c(1143, 945, 169)	c(5, 5, 4.64285714285714)	c(2, 1, 14)
entertainment	Romance	c(6, 6, 6)	c(2, 2, 2)	c(169, 434, 549)	c(5, 5, 5)	c(2, 2, 2)
executive	Romance	c(7, 7, 7)	c(2, 2, 2)	c(490, 584, 1041)	c(5, 5, 5)	c(2, 1, 1)
healthcare	Romance	c(8, 8, 8)	c(2, 2, 2)	c(1122, 1368, 187)	c(5, 5, 4.5)	c(1, 1, 4)
homemaker	Romance	c(9, 9, 9)	c(2, 2, 2)	c(845, 350, 181)	c(5, 5, 4.5)	c(2, 1, 6)
lawyer	Romance	c(10, 10, 10)	c(2, 2, 2)	c(180, 182, 485)	c(5, 5, 5)	c(3, 3, 3)
librarian	Romance	c(11, 11, 11)	c(2, 2, 2)	c(656, 741, 1024)	c(5, 5, 5)	c(2, 1, 1)
marketing	Romance	c(12, 12, 12)	c(2, 2, 2)	c(275, 378, 51)	c(5, 5, 5)	c(2, 2, 1)
none	Romance	c(13, 13, 13)	c(2, 2, 2)	c(684, 98, 270)	c(5, 5, 5)	c(4, 2, 2)
other	Romance	c(14, 14, 14)	c(2, 2, 2)	c(644, 1368, 1377)	c(5, 5, 5)	c(2, 2, 1)
programmer	Romance	c(15, 15, 15)	c(2, 2, 2)	c(818, 1260, 169)	c(5, 5, 4.76470588235294)	c(1, 1, 17)
retired	Romance	c(16, 16, 16)	c(2, 2, 2)	c(169, 306, 482)	c(5, 5, 4.66666666666667)	c(2, 2, 3)
salesman	Romance	c(17, 17, 17)	c(2, 2, 2)	c(346, 591, 147)	c(5, 5, 5)	3:1
scientist	Romance	c(18, 18, 18)	c(2, 2, 2)	c(525, 650, 652)	c(5, 5, 5)	c(1, 1, 1)
student	Romance	c(19, 19, 19)	c(2, 2, 2)	c(914, 1193, 1143)	c(5, 5, 4.66666666666667)	c(1, 1, 6)
technician	Romance	c(20, 20, 20)	c(2, 2, 2)	c(169, 14, 652)	c(5, 5, 5)	c(5, 3, 2)
writer	Romance	c(21, 21, 21)	c(2, 2, 2)	c(650, 652, 74)	c(5, 5, 5)	3:1
administrator	Sci.Fi	c(1, 1, 1)	c(3, 3, 3)	c(320, 947, 493)	c(5, 5, 4.8)	c(1, 1, 5)
artist	Sci.Fi	c(2, 2, 2)	c(3, 3, 3)	c(115, 334, 753)	c(5, 5, 5)	c(1, 1, 1)
doctor	Sci.Fi	c(3, 3, 3)	c(3, 3, 3)	c(83, 334, 753)	c(5, 5, 5)	c(1, 1, 1)
educator	Sci.Fi	c(4, 4, 4)	c(3, 3, 3)	c(1598, 753, 115)	c(5, 4.66666666666667, 4.5)	c(1, 3, 2)
engineer	Sci.Fi	c(5, 5, 5)	c(3, 3, 3)	c(1469, 246, 334)	c(5, 4.85714285714286, 4.33333333333333)	c(1, 7, 3)
entertainment	Sci.Fi	c(6, 6, 6)	c(3, 3, 3)	c(99, 1204, 127)	c(5, 5, 4.66666666666667)	c(2, 1, 6)
executive	Sci.Fi	c(7, 7, 7)	c(3, 3, 3)	c(778, 1110, 127)	c(5, 5, 4.5)	c(1, 1, 14)
healthcare	Sci.Fi	c(8, 8, 8)	c(3, 3, 3)	c(320, 898, 1218)	c(5, 5, 5)	c(1, 1, 1)
homemaker	Sci.Fi	c(9, 9, 9)	c(3, 3, 3)	c(1, 278, 281)	c(4, 4, 4)	c(2, 1, 1)
lawyer	Sci.Fi	c(10, 10, 10)	c(3, 3, 3)	c(703, 127, 91)	c(5, 4.5, 4.5)	c(1, 4, 2)
librarian	Sci.Fi	c(11, 11, 11)	c(3, 3, 3)	c(115, 389, 989)	c(5, 5, 5)	c(1, 1, 1)
marketing	Sci.Fi	c(12, 12, 12)	c(3, 3, 3)	c(83, 127, 423)	c(5, 4.41666666666667, 4.33333333333333)	c(1, 12, 6)
none	Sci.Fi	c(13, 13, 13)	c(3, 3, 3)	c(54, 278, 366)	c(5, 5, 5)	c(2, 1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
other	Sci.Fi	c(14, 14, 14)	c(3, 3, 3)	c(247, 1031, 753)	c(5, 5, 4.5)	c(1, 1, 4)
programmer	Sci.Fi	c(15, 15, 15)	c(3, 3, 3)	c(320, 721, 1012)	c(5, 4.5, 4.33333333333333)	c(1, 4, 6)
retired	Sci.Fi	c(16, 16, 16)	c(3, 3, 3)	c(91, 99, 423)	c(4.5, 4.5, 4.25)	c(2, 2, 4)
salesman	Sci.Fi	c(17, 17, 17)	c(3, 3, 3)	c(423, 83, 553)	c(5, 5, 5)	c(2, 1, 1)
scientist	Sci.Fi	c(18, 18, 18)	c(3, 3, 3)	c(703, 246, 127)	c(5, 4.5, 4.30769230769231)	c(1, 4, 13)
student	Sci.Fi	c(19, 19, 19)	c(3, 3, 3)	c(320, 115, 1116)	c(5, 5, 5)	c(4, 1, 1)
technician	Sci.Fi	c(20, 20, 20)	c(3, 3, 3)	c(1431, 246, 127)	c(5, 4.57142857142857, 4.38461538461539)	c(1, 7, 13)
writer	Sci.Fi	c(21, 21, 21)	c(3, 3, 3)	c(691, 721, 827)	c(5, 4.5, 4.5)	c(1, 2, 2)
administrator	Comedy	c(1, 1, 1)	c(4, 4, 4)	c(408, 574, 634)	c(5, 5, 5)	c(5, 2, 2)
artist	Comedy	c(2, 2, 2)	c(4, 4, 4)	c(408, 505, 921)	c(5, 5, 5)	4:2
doctor	Comedy	c(3, 3, 3)	c(4, 4, 4)	c(172, 483, 89)	c(5, 5, 5)	c(2, 2, 1)
educator	Comedy	c(4, 4, 4)	c(4, 4, 4)	c(851, 867, 1111)	c(5, 5, 5)	c(1, 1, 1)
engineer	Comedy	c(5, 5, 5)	c(4, 4, 4)	c(611, 1121, 1144)	c(5, 5, 5)	c(4, 1, 1)
entertainment	Comedy	c(6, 6, 6)	c(4, 4, 4)	c(188, 489, 572)	c(5, 5, 5)	c(2, 2, 2)
executive	Comedy	c(7, 7, 7)	c(4, 4, 4)	c(603, 501, 778)	c(5, 5, 5)	c(4, 2, 2)
healthcare	Comedy	c(8, 8, 8)	c(4, 4, 4)	c(489, 370, 392)	c(5, 5, 5)	c(2, 1, 1)
homemaker	Comedy	c(9, 9, 9)	c(4, 4, 4)	c(319, 1152, 255)	c(5, 5, 5)	c(2, 2, 1)
lawyer	Comedy	c(10, 10, 10)	c(4, 4, 4)	c(504, 12, 180)	c(5, 5, 5)	c(6, 3, 3)
librarian	Comedy	c(11, 11, 11)	c(4, 4, 4)	c(656, 108, 389)	c(5, 5, 5)	c(2, 1, 1)
marketing	Comedy	c(12, 12, 12)	c(4, 4, 4)	c(101, 530, 887)	c(5, 5, 5)	c(2, 2, 2)
none	Comedy	c(13, 13, 13)	c(4, 4, 4)	c(326, 33, 55)	c(5, 5, 5)	c(4, 2, 2)
other	Comedy	c(14, 14, 14)	c(4, 4, 4)	c(247, 360, 618)	c(5, 5, 5)	c(1, 1, 1)
programmer	Comedy	c(15, 15, 15)	c(4, 4, 4)	c(611, 494, 975)	c(5, 5, 5)	c(2, 1, 1)
retired	Comedy	c(16, 16, 16)	c(4, 4, 4)	c(512, 201, 271)	c(5, 5, 5)	c(2, 1, 1)
salesman	Comedy	c(17, 17, 17)	c(4, 4, 4)	c(423, 346, 483)	c(5, 5, 5)	4:2
scientist	Comedy	c(18, 18, 18)	c(4, 4, 4)	c(178, 524, 219)	c(5, 5, 5)	c(2, 2, 1)
student	Comedy	c(19, 19, 19)	c(4, 4, 4)	c(614, 888, 1189)	c(5, 5, 5)	c(2, 2, 2)
technician	Comedy	c(20, 20, 20)	c(4, 4, 4)	c(512, 611, 45)	c(5, 5, 5)	c(2, 2, 1)
writer	Comedy	c(21, 21, 21)	c(4, 4, 4)	c(18, 841, 851)	c(5, 5, 5)	c(1, 1, 1)
administrator	Documentary	c(1, 1, 1)	c(5, 5, 5)	c(308, 1262, 429)	c(5, 5, 4.71428571428571)	c(1, 1, 7)
artist	Documentary	c(2, 2, 2)	c(5, 5, 5)	c(707, 558, 165)	c(5, 5, 4.8)	c(2, 1, 5)
doctor	Documentary	c(3, 3, 3)	c(5, 5, 5)	c(286, 165, 179)	c(4.6, 4, 4)	c(5, 1, 1)
educator	Documentary	c(4, 4, 4)	c(5, 5, 5)	c(953, 1459, 490)	c(5, 5, 4.08333333333333)	c(2, 1, 12)
engineer	Documentary	c(5, 5, 5)	c(5, 5, 5)	c(707, 986, 490)	c(4.5, 4.5, 4.33333333333333)	c(4, 2, 3)
entertainment	Documentary	c(6, 6, 6)	c(5, 5, 5)	c(179, 549, 879)	c(5, 5, 4)	c(4, 2, 3)
executive	Documentary	c(7, 7, 7)	c(5, 5, 5)	c(490, 165, 707)	c(5, 5, 5)	c(2, 1, 1)
healthcare	Documentary	c(8, 8, 8)	c(5, 5, 5)	c(429, 637, 1123)	c(4.33333333333333, 4, 4)	c(3, 1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
homemaker	Documentary	c(9, 9, 9)	c(5, 5, 5)	c(879, 82, 588)	c(4.5, 4, 4)	c(2, 1, 1)
lawyer	Documentary	c(10, 10, 10)	c(5, 5, 5)	c(179, 2, 250)	c(5, 5, 5)	c(3, 1, 1)
librarian	Documentary	c(11, 11, 11)	c(5, 5, 5)	c(707, 429, 490)	c(4, 4, 4)	c(9, 5, 4)
marketing	Documentary	c(12, 12, 12)	c(5, 5, 5)	c(942, 490, 179)	c(5, 4.33333333333333, 4.2)	c(1, 3, 5)
none	Documentary	c(13, 13, 13)	c(5, 5, 5)	c(986, 2, 82)	c(5, 4.5, 4.33333333333333)	1:3
other	Documentary	c(14, 14, 14)	c(5, 5, 5)	c(113, 429, 82)	c(5, 4.09090909090909, 4)	c(1, 11, 29)
programmer	Documentary	c(15, 15, 15)	c(5, 5, 5)	c(165, 179, 588)	c(4.25, 4.2, 4.05882352941176)	c(8, 15, 17)
retired	Documentary	c(16, 16, 16)	c(5, 5, 5)	c(165, 610, 162)	c(4.5, 4.33333333333333, 4.2)	c(4, 3, 5)
salesman	Documentary	c(17, 17, 17)	c(5, 5, 5)	c(82, 162, 588)	c(5, 5, 5)	c(1, 1, 1)
scientist	Documentary	c(18, 18, 18)	c(5, 5, 5)	c(113, 707, 165)	c(5, 5, 4.2)	c(1, 1, 5)
student	Documentary	c(19, 19, 19)	c(5, 5, 5)	c(1262, 162, 558)	c(5, 4.22222222222222, 4.1)	c(1, 9, 10)
technician	Documentary	c(20, 20, 20)	c(5, 5, 5)	c(113, 942, 490)	c(5, 5, 4.5)	c(1, 1, 2)
writer	Documentary	c(21, 21, 21)	c(5, 5, 5)	c(1597, 165, 588)	c(5, 4.33333333333333, 4.22222222222222)	c(1, 3, 9)
administrator	Drama	c(1, 1, 1)	c(6, 6, 6)	c(935, 1020, 308)	c(5, 5, 5)	c(2, 2, 1)
artist	Drama	c(2, 2, 2)	c(6, 6, 6)	c(53, 169, 505)	c(5, 5, 5)	c(4, 3, 3)
doctor	Drama	c(3, 3, 3)	c(6, 6, 6)	c(483, 132, 170)	c(5, 5, 5)	c(4, 2, 2)
educator	Drama	c(4, 4, 4)	c(6, 6, 6)	c(963, 1405, 1558)	c(5, 5, 5)	c(3, 2, 2)
engineer	Drama	c(5, 5, 5)	c(6, 6, 6)	c(834, 945, 982)	c(5, 5, 5)	c(1, 1, 1)
entertainment	Drama	c(6, 6, 6)	c(6, 6, 6)	c(641, 150, 169)	c(5, 5, 5)	c(4, 2, 2)
executive	Drama	c(7, 7, 7)	c(6, 6, 6)	c(543, 603, 522)	c(5, 5, 5)	c(4, 4, 2)
healthcare	Drama	c(8, 8, 8)	c(6, 6, 6)	c(489, 571, 575)	c(5, 5, 5)	c(1, 1, 1)
homemaker	Drama	c(9, 9, 9)	c(6, 6, 6)	c(222, 22, 255)	c(5, 5, 5)	c(2, 1, 1)
lawyer	Drama	c(10, 10, 10)	c(6, 6, 6)	c(488, 12, 603)	c(5, 5, 5)	c(4, 3, 3)
librarian	Drama	c(11, 11, 11)	c(6, 6, 6)	c(533, 149, 339)	c(5, 5, 5)	c(2, 1, 1)
marketing	Drama	c(12, 12, 12)	c(6, 6, 6)	c(496, 275, 955)	c(5, 5, 5)	c(3, 2, 2)
none	Drama	c(13, 13, 13)	c(6, 6, 6)	c(385, 17, 64)	c(5, 5, 5)	c(4, 3, 3)
other	Drama	c(14, 14, 14)	c(6, 6, 6)	c(601, 645, 1194)	c(5, 5, 5)	c(2, 2, 2)
programmer	Drama	c(15, 15, 15)	c(6, 6, 6)	c(543, 1137, 1269)	c(5, 5, 5)	c(2, 2, 2)
retired	Drama	c(16, 16, 16)	c(6, 6, 6)	c(498, 271, 169)	c(5, 5, 5)	c(6, 3, 2)
salesman	Drama	c(17, 17, 17)	c(6, 6, 6)	c(28, 483, 173)	c(5, 5, 5)	c(6, 4, 3)
scientist	Drama	c(18, 18, 18)	c(6, 6, 6)	c(81, 156, 207)	c(5, 5, 5)	c(4, 2, 2)
student	Drama	c(19, 19, 19)	c(6, 6, 6)	c(1137, 1462, 968)	c(5, 5, 5)	c(4, 2, 1)
technician	Drama	c(20, 20, 20)	c(6, 6, 6)	c(169, 14, 641)	c(5, 5, 5)	c(5, 3, 2)
writer	Drama	c(21, 21, 21)	c(6, 6, 6)	c(1269, 74, 691)	c(5, 5, 5)	c(2, 1, 1)
administrator	Action	c(1, 1, 1)	c(7, 7, 7)	c(408, 548, 899)	c(5, 5, 5)	c(5, 1, 1)
artist	Action	c(2, 2, 2)	c(7, 7, 7)	c(408, 10, 207)	c(5, 5, 5)	c(4, 2, 2)
doctor	Action	c(3, 3, 3)	c(7, 7, 7)	c(10, 332, 689)	c(5, 5, 5)	c(1, 1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
educator	Action	c(4, 4, 4)	c(7, 7, 7)	c(119, 733, 856)	c(5, 5, 5)	c(1, 1, 1)
engineer	Action	c(5, 5, 5)	c(7, 7, 7)	c(1063, 945, 982)	c(5, 5, 5)	c(2, 1, 1)
entertainment	Action	c(6, 6, 6)	c(7, 7, 7)	c(952, 478, 517)	c(5, 5, 5)	c(2, 1, 1)
executive	Action	c(7, 7, 7)	c(7, 7, 7)	c(501, 969, 584)	c(5, 5, 5)	c(2, 2, 1)
healthcare	Action	c(8, 8, 8)	c(7, 7, 7)	c(376, 452, 1034)	c(5, 5, 5)	c(1, 1, 1)
homemaker	Action	c(9, 9, 9)	c(7, 7, 7)	c(284, 350, 313)	c(5, 5, 4.5)	c(1, 1, 4)
lawyer	Action	c(10, 10, 10)	c(7, 7, 7)	c(589, 337, 339)	c(5, 5, 5)	c(2, 1, 1)
librarian	Action	c(11, 11, 11)	c(7, 7, 7)	c(656, 1099, 119)	c(5, 5, 5)	c(2, 2, 1)
marketing	Action	c(12, 12, 12)	c(7, 7, 7)	c(663, 1084, 49)	c(5, 5, 5)	c(2, 2, 1)
none	Action	c(13, 13, 13)	c(7, 7, 7)	c(249, 69, 93)	c(5, 5, 5)	c(2, 1, 1)
other	Action	c(14, 14, 14)	c(7, 7, 7)	c(547, 904, 1293)	c(5, 5, 5)	c(2, 2, 2)
programmer	Action	c(15, 15, 15)	c(7, 7, 7)	c(850, 1103, 1166)	c(5, 5, 5)	c(1, 1, 1)
retired	Action	c(16, 16, 16)	c(7, 7, 7)	c(408, 664, 191)	c(5, 5, 4.5)	c(1, 1, 6)
salesman	Action	c(17, 17, 17)	c(7, 7, 7)	c(20, 71, 144)	c(5, 5, 5)	c(1, 1, 1)
scientist	Action	c(18, 18, 18)	c(7, 7, 7)	c(207, 525, 228)	c(5, 5, 4.66666666666667)	c(2, 1, 6)
student	Action	c(19, 19, 19)	c(7, 7, 7)	c(914, 1396, 1500)	c(5, 5, 5)	c(1, 1, 1)
technician	Action	c(20, 20, 20)	c(7, 7, 7)	c(312, 119, 501)	c(5, 5, 5)	c(2, 1, 1)
writer	Action	c(21, 21, 21)	c(7, 7, 7)	c(478, 487, 1019)	c(4.71428571428571, 4.66666666666667, 4.5)	c(7, 3, 2)
administrator	Thriller	c(1, 1, 1)	c(8, 8, 8)	c(408, 251, 1142)	c(5, 5, 5)	c(5, 3, 2)
artist	Thriller	c(2, 2, 2)	c(8, 8, 8)	c(408, 207, 41)	c(5, 5, 5)	c(4, 2, 1)
doctor	Thriller	c(3, 3, 3)	c(8, 8, 8)	c(514, 22, 79)	c(5, 5, 5)	c(2, 1, 1)
educator	Thriller	c(4, 4, 4)	c(8, 8, 8)	c(856, 1201, 1301)	c(5, 5, 5)	c(1, 1, 1)
engineer	Thriller	c(5, 5, 5)	c(8, 8, 8)	c(1063, 1203, 607)	c(5, 5, 4.66666666666667)	c(2, 1, 3)
entertainment	Thriller	c(6, 6, 6)	c(8, 8, 8)	c(188, 654, 497)	c(5, 5, 5)	c(2, 2, 1)
executive	Thriller	c(7, 7, 7)	c(8, 8, 8)	c(632, 715, 135)	c(5, 5, 4.77777777777778)	c(1, 1, 9)
healthcare	Thriller	c(8, 8, 8)	c(8, 8, 8)	c(67, 370, 608)	c(5, 5, 5)	c(1, 1, 1)
homemaker	Thriller	c(9, 9, 9)	c(8, 8, 8)	c(22, 350, 762)	c(5, 5, 5)	c(1, 1, 1)
lawyer	Thriller	c(10, 10, 10)	c(8, 8, 8)	c(480, 484, 485)	c(5, 5, 5)	c(6, 3, 3)
librarian	Thriller	c(11, 11, 11)	c(8, 8, 8)	c(253, 998, 1222)	c(5, 5, 5)	c(2, 1, 1)
marketing	Thriller	c(12, 12, 12)	c(8, 8, 8)	c(101, 856, 315)	c(5, 5, 4.75)	c(1, 1, 4)
none	Thriller	c(13, 13, 13)	c(8, 8, 8)	c(270, 326, 538)	c(5, 5, 5)	c(2, 2, 2)
other	Thriller	c(14, 14, 14)	c(8, 8, 8)	c(904, 1233, 408)	c(5, 5, 4.6)	c(2, 1, 10)
programmer	Thriller	c(15, 15, 15)	c(8, 8, 8)	c(416, 251, 408)	c(5, 4.75, 4.69230769230769)	c(1, 4, 13)
retired	Thriller	c(16, 16, 16)	c(8, 8, 8)	c(498, 902, 201)	c(5, 5, 5)	3:1
salesman	Thriller	c(17, 17, 17)	c(8, 8, 8)	c(153, 199, 45)	c(5, 5, 5)	c(2, 2, 1)
scientist	Thriller	c(18, 18, 18)	c(8, 8, 8)	c(207, 902, 1347)	c(5, 5, 5)	c(2, 2, 1)
student	Thriller	c(19, 19, 19)	c(8, 8, 8)	c(279, 1015, 1462)	c(5, 5, 5)	c(2, 1, 1)
technician	Thriller	c(20, 20, 20)	c(8, 8, 8)	c(253, 45, 270)	c(5, 5, 5)	c(2, 1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
writer	Thriller	c(21, 21, 21)	c(8, 8, 8)	c(1451, 654, 1129)	c(5, 4.6, 4.5)	c(1, 10, 2)
administrator	Crime	c(1, 1, 1)	c(9, 9, 9)	c(1278, 510, 173)	c(5, 4.58333333333333, 4.56521739130435)	c(1, 12, 23)
artist	Crime	c(2, 2, 2)	c(9, 9, 9)	c(242, 223, 362)	c(5, 5, 5)	c(2, 1, 1)
doctor	Crime	c(3, 3, 3)	c(9, 9, 9)	c(514, 132, 172)	c(5, 5, 5)	c(2, 1, 1)
educator	Crime	c(4, 4, 4)	c(9, 9, 9)	c(48, 124, 132)	c(4.69230769230769, 4.47368421052632, 4.43478260869565)	c(13, 19, 23)
engineer	Crime	c(5, 5, 5)	c(9, 9, 9)	c(984, 1278, 172)	c(5, 5, 4.45714285714286)	c(1, 1, 35)
entertainment	Crime	c(6, 6, 6)	c(9, 9, 9)	c(946, 98, 182)	c(5, 4.83333333333333, 4.66666666666667)	c(1, 6, 6)
executive	Crime	c(7, 7, 7)	c(9, 9, 9)	c(601, 1278, 1426)	c(5, 5, 5)	c(1, 1, 1)
healthcare	Crime	c(8, 8, 8)	c(9, 9, 9)	c(148, 615, 984)	c(5, 5, 4.5)	c(1, 1, 2)
homemaker	Crime	c(9, 9, 9)	c(9, 9, 9)	c(222, 740, 148)	c(5, 5, 4)	c(1, 1, 3)
lawyer	Crime	c(10, 10, 10)	c(9, 9, 9)	c(182, 250, 601)	c(5, 5, 5)	c(3, 1, 1)
librarian	Crime	c(11, 11, 11)	c(9, 9, 9)	c(1217, 430, 124)	c(5, 4.66666666666667, 4.5625)	c(1, 3, 16)
marketing	Crime	c(12, 12, 12)	c(9, 9, 9)	c(1217, 515, 601)	c(5, 4.6, 4.5)	c(1, 5, 2)
none	Crime	c(13, 13, 13)	c(9, 9, 9)	c(55, 98, 132)	c(5, 5, 5)	c(2, 2, 1)
other	Crime	c(14, 14, 14)	c(9, 9, 9)	c(601, 1293, 515)	c(5, 5, 4.43478260869565)	c(2, 2, 23)
programmer	Crime	c(15, 15, 15)	c(9, 9, 9)	c(172, 199, 98)	c(4.41666666666667, 4.375, 4.36666666666667)	c(36, 16, 30)
retired	Crime	c(16, 16, 16)	c(9, 9, 9)	c(604, 199, 20)	c(5, 4.66666666666667, 4.5)	c(1, 3, 2)
salesman	Crime	c(17, 17, 17)	c(9, 9, 9)	c(173, 199, 515)	c(5, 5, 5)	c(3, 2, 2)
scientist	Crime	c(18, 18, 18)	c(9, 9, 9)	c(1025, 172, 48)	c(5, 4.55555555555556, 4.5)	c(1, 9, 2)
student	Crime	c(19, 19, 19)	c(9, 9, 9)	c(341, 1540, 172)	c(5, 4.5, 4.38461538461539)	c(1, 4, 91)
technician	Crime	c(20, 20, 20)	c(9, 9, 9)	c(242, 312, 173)	c(5, 5, 4.66666666666667)	c(1, 1, 12)
writer	Crime	c(21, 21, 21)	c(9, 9, 9)	c(74, 1597, 487)	c(5, 5, 4.66666666666667)	c(1, 1, 3)
administrator	Musical	c(1, 1, 1)	c(10, 10, 10)	c(173, 19, 1022)	c(4.56521739130435, 4.5, 4.5)	c(23, 6, 2)
artist	Musical	c(2, 2, 2)	c(10, 10, 10)	c(1022, 201, 336)	c(5, 5, 4.5)	c(2, 1, 2)
doctor	Musical	c(3, 3, 3)	c(10, 10, 10)	c(173, 183, 689)	c(5, 5, 5)	c(1, 1, 1)
educator	Musical	c(4, 4, 4)	c(10, 10, 10)	c(361, 87, 6)	c(5, 4.2, 4.2)	c(1, 15, 5)
engineer	Musical	c(5, 5, 5)	c(10, 10, 10)	c(173, 302, 659)	c(4.33333333333333, 4.28571428571429, 4.16666666666667)	c(27, 21, 6)
entertainment	Musical	c(6, 6, 6)	c(10, 10, 10)	c(1022, 404, 461)	c(5, 5, 5)	c(2, 1, 1)
executive	Musical	c(7, 7, 7)	c(10, 10, 10)	c(391, 461, 993)	c(5, 4.5, 4.5)	c(1, 2, 2)
healthcare	Musical	c(8, 8, 8)	c(10, 10, 10)	c(1138, 420, 6)	c(5, 4, 4)	c(1, 3, 1)
homemaker	Musical	c(9, 9, 9)	c(10, 10, 10)	c(87, 873, 333)	c(5, 5, 4.75)	c(1, 1, 4)
lawyer	Musical	c(10, 10, 10)	c(10, 10, 10)	c(873, 173, 87)	c(5, 4.75, 4.5)	c(1, 4, 2)
librarian	Musical	c(11, 11, 11)	c(10, 10, 10)	c(1155, 302, 183)	c(5, 4.25, 4.25)	c(1, 20, 12)
marketing	Musical	c(12, 12, 12)	c(10, 10, 10)	c(299, 183, 302)	c(4.5, 4.33333333333333, 4.09090909090909)	c(2, 6, 11)
none	Musical	c(13, 13, 13)	c(10, 10, 10)	c(333, 361, 993)	c(5, 5, 5)	c(3, 1, 1)
other	Musical	c(14, 14, 14)	c(10, 10, 10)	c(361, 1629, 6)	c(5, 5, 4.66666666666667)	c(1, 1, 3)
programmer	Musical	c(15, 15, 15)	c(10, 10, 10)	c(156, 420, 173)	c(4.41176470588235, 4.33333333333333, 4.2)	c(17, 3, 30)
retired	Musical	c(16, 16, 16)	c(10, 10, 10)	c(201, 302, 420)	c(5, 4.2, 4)	c(1, 5, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
salesman	Musical	c(17, 17, 17)	c(10, 10, 10)	c(173, 156, 201)	c(5, 5, 5)	3:1
scientist	Musical	c(18, 18, 18)	c(10, 10, 10)	c(19, 156, 659)	c(5, 5, 4.5)	c(2, 2, 2)
student	Musical	c(19, 19, 19)	c(10, 10, 10)	c(302, 173, 659)	c(4.3125, 4.28048780487805, 4.16666666666667)	c(48, 82, 12)
technician	Musical	c(20, 20, 20)	c(10, 10, 10)	c(6, 19, 361)	c(5, 5, 5)	c(1, 1, 1)
writer	Musical	c(21, 21, 21)	c(10, 10, 10)	c(19, 173, 302)	c(4.2, 4.05882352941176, 4)	c(5, 17, 22)
administrator	Adventure	c(1, 1, 1)	c(11, 11, 11)	c(548, 574, 832)	c(5, 5, 5)	c(1, 1, 1)
artist	Adventure	c(2, 2, 2)	c(11, 11, 11)	c(520, 265, 1009)	c(5, 5, 5)	c(3, 2, 2)
doctor	Adventure	c(3, 3, 3)	c(11, 11, 11)	c(520, 313, 185)	c(5, 4.4, 4)	c(1, 5, 3)
educator	Adventure	c(4, 4, 4)	c(11, 11, 11)	c(963, 953, 1199)	c(5, 5, 5)	3:1
engineer	Adventure	c(5, 5, 5)	c(11, 11, 11)	c(1203, 963, 654)	c(5, 4.6, 4.46153846153846)	c(1, 5, 13)
entertainment	Adventure	c(6, 6, 6)	c(11, 11, 11)	c(99, 654, 963)	c(5, 5, 5)	c(2, 2, 1)
executive	Adventure	c(7, 7, 7)	c(11, 11, 11)	c(316, 963, 391)	c(5, 5, 5)	c(2, 2, 1)
healthcare	Adventure	c(8, 8, 8)	c(11, 11, 11)	c(452, 313, 316)	c(5, 4.55555555555556, 4.5)	c(1, 9, 2)
homemaker	Adventure	c(9, 9, 9)	c(11, 11, 11)	c(313, 316, 274)	c(4.5, 4.5, 4)	c(4, 2, 2)
lawyer	Adventure	c(10, 10, 10)	c(11, 11, 11)	c(520, 184, 538)	c(5, 5, 5)	c(2, 1, 1)
librarian	Adventure	c(11, 11, 11)	c(11, 11, 11)	c(1466, 538, 697)	c(4.66666666666667, 4.5, 4.5)	c(3, 2, 2)
marketing	Adventure	c(12, 12, 12)	c(11, 11, 11)	c(316, 51, 838)	c(5, 5, 5)	c(2, 1, 1)
none	Adventure	c(13, 13, 13)	c(11, 11, 11)	c(185, 538, 17)	c(5, 5, 5)	c(2, 2, 1)
other	Adventure	c(14, 14, 14)	c(11, 11, 11)	c(1194, 1293, 909)	c(5, 5, 5)	c(2, 2, 1)
programmer	Adventure	c(15, 15, 15)	c(11, 11, 11)	c(313, 902, 697)	c(4.58333333333333, 4.5, 4.5)	c(24, 4, 2)
retired	Adventure	c(16, 16, 16)	c(11, 11, 11)	c(498, 697, 902)	c(5, 5, 5)	c(3, 1, 1)
salesman	Adventure	c(17, 17, 17)	c(11, 11, 11)	c(214, 329, 265)	c(5, 5, 5)	c(2, 2, 1)
scientist	Adventure	c(18, 18, 18)	c(11, 11, 11)	c(902, 1025, 185)	c(5, 5, 4.66666666666667)	c(1, 1, 3)
student	Adventure	c(19, 19, 19)	c(11, 11, 11)	c(1375, 313, 902)	c(5, 4.54430379746835, 4.4)	c(1, 79, 5)
technician	Adventure	c(20, 20, 20)	c(11, 11, 11)	c(1009, 1203, 520)	c(5, 5, 4.5)	c(1, 1, 6)
writer	Adventure	c(21, 21, 21)	c(11, 11, 11)	c(130, 691, 963)	c(5, 5, 5)	c(1, 1, 1)
administrator	Mystery	c(1, 1, 1)	c(12, 12, 12)	c(923, 510, 490)	c(5, 4.58333333333333, 4.33333333333333)	c(2, 12, 3)
artist	Mystery	c(2, 2, 2)	c(12, 12, 12)	c(10, 653, 736)	c(5, 5, 5)	c(2, 1, 1)
doctor	Mystery	c(3, 3, 3)	c(12, 12, 12)	c(10, 510, 680)	c(5, 5, 5)	c(1, 1, 1)
educator	Mystery	c(4, 4, 4)	c(12, 12, 12)	c(23, 484, 194)	c(4.64705882352941, 4.46153846153846, 4.3)	c(17, 13, 30)
engineer	Mystery	c(5, 5, 5)	c(12, 12, 12)	c(246, 484, 490)	c(4.85714285714286, 4.35714285714286, 4.33333333333333)	c(7, 14, 3)
entertainment	Mystery	c(6, 6, 6)	c(12, 12, 12)	c(481, 602, 1039)	c(5, 5, 5)	c(1, 1, 1)
executive	Mystery	c(7, 7, 7)	c(12, 12, 12)	c(490, 481, 939)	c(5, 4.5, 4.5)	c(2, 2, 2)
healthcare	Mystery	c(8, 8, 8)	c(12, 12, 12)	c(602, 736, 204)	c(4.5, 4.33333333333333, 4.2)	c(2, 3, 5)
homemaker	Mystery	c(9, 9, 9)	c(12, 12, 12)	c(319, 274, 259)	c(5, 4, 4)	c(2, 2, 1)
lawyer	Mystery	c(10, 10, 10)	c(12, 12, 12)	c(484, 259, 881)	c(5, 5, 5)	c(3, 1, 1)
librarian	Mystery	c(11, 11, 11)	c(12, 12, 12)	c(881, 212, 923)	c(5, 4.33333333333333, 4.28571428571429)	c(1, 9, 7)
marketing	Mystery	c(12, 12, 12)	c(12, 12, 12)	c(101, 736, 490)	c(5, 5, 4.33333333333333)	c(1, 1, 3)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
none	Mystery	c(13, 13, 13)	c(12, 12, 12)	c(925, 1150, 96)	c(5, 5, 4.66666666666667)	c(1, 1, 3)
other	Mystery	c(14, 14, 14)	c(12, 12, 12)	c(923, 653, 481)	c(4.66666666666667, 4.33333333333333, 4.28571428571429)	c(6, 3, 7)
programmer	Mystery	c(15, 15, 15)	c(12, 12, 12)	c(1443, 736, 923)	c(5, 4.5, 4.25)	c(1, 2, 4)
retired	Mystery	c(16, 16, 16)	c(12, 12, 12)	c(194, 23, 481)	c(4.8, 4.5, 4.33333333333333)	c(5, 2, 3)
salesman	Mystery	c(17, 17, 17)	c(12, 12, 12)	c(484, 1039, 194)	c(5, 5, 4.33333333333333)	c(1, 1, 3)
scientist	Mystery	c(18, 18, 18)	c(12, 12, 12)	c(923, 23, 246)	c(4.75, 4.5, 4.5)	c(4, 6, 4)
student	Mystery	c(19, 19, 19)	c(12, 12, 12)	c(23, 1039, 1190)	c(4.37837837837838, 4.31578947368421, 4.25)	c(37, 19, 4)
technician	Mystery	c(20, 20, 20)	c(12, 12, 12)	c(246, 484, 490)	c(4.57142857142857, 4.57142857142857, 4.5)	c(7, 7, 2)
writer	Mystery	c(21, 21, 21)	c(12, 12, 12)	c(1174, 481, 96)	c(5, 4.5, 4.36363636363636)	c(1, 2, 11)
administrator	War	c(1, 1, 1)	c(13, 13, 13)	c(359, 173, 1537)	c(5, 4.56521739130435, 4.5)	c(2, 23, 2)
artist	War	c(2, 2, 2)	c(13, 13, 13)	c(1240, 653, 1232)	c(5, 5, 5)	c(2, 1, 1)
doctor	War	c(3, 3, 3)	c(13, 13, 13)	c(172, 173, 359)	c(5, 5, 5)	c(1, 1, 1)
educator	War	c(4, 4, 4)	c(13, 13, 13)	c(50, 137, 100)	c(4.52, 4.33333333333333, 4.31481481481481)	c(50, 24, 54)
engineer	War	c(5, 5, 5)	c(13, 13, 13)	c(1469, 1240, 50)	c(5, 4.5, 4.48)	c(1, 2, 50)
entertainment	War	c(6, 6, 6)	c(13, 13, 13)	c(188, 517, 1240)	c(5, 5, 5)	c(2, 1, 1)
executive	War	c(7, 7, 7)	c(13, 13, 13)	c(501, 228, 188)	c(5, 4.75, 4.6)	c(2, 4, 5)
healthcare	War	c(8, 8, 8)	c(13, 13, 13)	c(571, 357, 73)	c(5, 4.14285714285714, 4)	c(1, 7, 3)
homemaker	War	c(9, 9, 9)	c(13, 13, 13)	c(50, 327, 100)	c(4, 4, 4)	c(2, 2, 1)
lawyer	War	c(10, 10, 10)	c(13, 13, 13)	c(202, 642, 193)	c(5, 5, 5)	3:1
librarian	War	c(11, 11, 11)	c(13, 13, 13)	c(430, 100, 127)	c(4.66666666666667, 4.38461538461539, 4.36363636363636)	c(3, 26, 22)
marketing	War	c(12, 12, 12)	c(13, 13, 13)	c(357, 127, 100)	c(4.75, 4.41666666666667, 4.33333333333333)	c(8, 12, 9)
none	War	c(13, 13, 13)	c(13, 13, 13)	c(93, 94, 357)	c(5, 5, 5)	c(1, 1, 1)
other	War	c(14, 14, 14)	c(13, 13, 13)	c(247, 359, 1240)	c(5, 5, 5)	c(1, 1, 1)
programmer	War	c(15, 15, 15)	c(13, 13, 13)	c(50, 357, 172)	c(4.54, 4.47826086956522, 4.41666666666667)	c(50, 23, 36)
retired	War	c(16, 16, 16)	c(13, 13, 13)	c(100, 127, 327)	c(4.1, 4, 4)	c(10, 8, 3)
salesman	War	c(17, 17, 17)	c(13, 13, 13)	c(173, 137, 127)	c(5, 5, 4.5)	c(3, 1, 4)
scientist	War	c(18, 18, 18)	c(13, 13, 13)	c(228, 172, 345)	c(4.66666666666667, 4.55555555555556, 4.5)	c(6, 9, 2)
student	War	c(19, 19, 19)	c(13, 13, 13)	c(1594, 127, 50)	c(4.5, 4.48684210526316, 4.41666666666667)	c(2, 76, 132)
technician	War	c(20, 20, 20)	c(13, 13, 13)	c(359, 501, 137)	c(5, 5, 4.75)	c(2, 1, 4)
writer	War	c(21, 21, 21)	c(13, 13, 13)	c(1536, 127, 653)	c(5, 4.45454545454545, 4.25)	c(1, 22, 4)
administrator	Film.Noir	c(1, 1, 1)	c(14, 14, 14)	c(224, 1127, 1175)	c(5, 5, 5)	c(1, 1, 1)
artist	Film.Noir	c(2, 2, 2)	c(14, 14, 14)	c(22, 55, 150)	c(4.2, 4.2, 4.16666666666667)	c(5, 5, 6)
doctor	Film.Noir	c(3, 3, 3)	c(14, 14, 14)	c(22, 307, 990)	c(5, 5, 5)	c(1, 1, 1)
educator	Film.Noir	c(4, 4, 4)	c(14, 14, 14)	c(1198, 209, 22)	c(5, 4.22727272727273, 4.07407407407407)	c(1, 22, 27)
engineer	Film.Noir	c(5, 5, 5)	c(14, 14, 14)	c(297, 205, 22)	c(4.66666666666667, 4.1875, 4.10714285714286)	c(3, 16, 28)
entertainment	Film.Noir	c(6, 6, 6)	c(14, 14, 14)	c(150, 224, 916)	c(5, 5, 5)	c(2, 1, 1)
executive	Film.Noir	c(7, 7, 7)	c(14, 14, 14)	c(22, 55, 205)	c(4, 3.85714285714286, 3.71428571428571)	c(8, 7, 7)
healthcare	Film.Noir	c(8, 8, 8)	c(14, 14, 14)	c(990, 22, 209)	c(4, 3.75, 3.66666666666667)	c(1, 4, 3)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
homemaker	Film.Noir	c(9, 9, 9)	c(14, 14, 14)	c(22, 237, 307)	c(5, 4.5, 4)	c(1, 2, 2)
lawyer	Film.Noir	c(10, 10, 10)	c(14, 14, 14)	c(205, 353, 1175)	c(5, 5, 5)	c(4, 1, 1)
librarian	Film.Noir	c(11, 11, 11)	c(14, 14, 14)	c(741, 297, 248)	c(5, 4.125, 4.11111111111111)	c(1, 8, 9)
marketing	Film.Noir	c(12, 12, 12)	c(14, 14, 14)	c(297, 205, 22)	c(5, 4.6, 4.33333333333333)	c(1, 5, 6)
none	Film.Noir	c(13, 13, 13)	c(14, 14, 14)	c(55, 150, 257)	c(5, 5, 4.5)	c(2, 1, 2)
other	Film.Noir	c(14, 14, 14)	c(14, 14, 14)	c(205, 22, 224)	c(4.44444444444444, 4.24137931034483, 4)	c(9, 29, 5)
programmer	Film.Noir	c(15, 15, 15)	c(14, 14, 14)	c(297, 205, 365)	c(4.5, 4.38461538461539, 4.25)	c(2, 13, 4)
retired	Film.Noir	c(16, 16, 16)	c(14, 14, 14)	c(1463, 257, 22)	c(4.5, 4.16666666666667, 4)	c(2, 6, 4)
salesman	Film.Noir	c(17, 17, 17)	c(14, 14, 14)	c(205, 209, 990)	c(5, 5, 5)	c(1, 1, 1)
scientist	Film.Noir	c(18, 18, 18)	c(14, 14, 14)	c(224, 205, 22)	c(5, 4.66666666666667, 4.11111111111111)	c(1, 3, 9)
student	Film.Noir	c(19, 19, 19)	c(14, 14, 14)	c(297, 205, 224)	c(4.5, 4.33333333333333, 4.33333333333333)	c(6, 12, 6)
technician	Film.Noir	c(20, 20, 20)	c(14, 14, 14)	c(224, 150, 205)	c(5, 4.4, 4.28571428571429)	c(1, 5, 7)
writer	Film.Noir	c(21, 21, 21)	c(14, 14, 14)	c(297, 209, 307)	c(4.33333333333333, 4, 3.85714285714286)	c(3, 15, 14)
administrator	Children.s	c(1, 1, 1)	c(15, 15, 15)	c(835, 1282, 1449)	c(5, 5, 5)	c(2, 1, 1)
artist	Children.s	c(2, 2, 2)	c(15, 15, 15)	c(265, 201, 334)	c(5, 5, 5)	c(2, 1, 1)
doctor	Children.s	c(3, 3, 3)	c(15, 15, 15)	c(307, 474, 938)	c(5, 5, 5)	c(2, 2, 2)
educator	Children.s	c(4, 4, 4)	c(15, 15, 15)	c(868, 1449, 686)	c(5, 5, 4.6)	c(1, 1, 5)
engineer	Children.s	c(5, 5, 5)	c(15, 15, 15)	c(611, 834, 973)	c(5, 5, 5)	c(2, 1, 1)
entertainment	Children.s	c(6, 6, 6)	c(15, 15, 15)	c(52, 60, 1273)	c(5, 5, 5)	c(1, 1, 1)
executive	Children.s	c(7, 7, 7)	c(15, 15, 15)	c(1041, 1109, 265)	c(5, 5, 4.57142857142857)	c(1, 1, 7)
healthcare	Children.s	c(8, 8, 8)	c(15, 15, 15)	c(949, 1122, 133)	c(5, 5, 4.25)	c(1, 1, 4)
homemaker	Children.s	c(9, 9, 9)	c(15, 15, 15)	c(284, 307, 304)	c(5, 4, 4)	c(1, 4, 1)
lawyer	Children.s	c(10, 10, 10)	c(15, 15, 15)	c(266, 611, 328)	c(5, 5, 4.5)	c(2, 1, 4)
librarian	Children.s	c(11, 11, 11)	c(15, 15, 15)	c(1449, 1296, 537)	c(5, 4.5, 4.33333333333333)	1:3
marketing	Children.s	c(12, 12, 12)	c(15, 15, 15)	c(813, 31, 133)	c(5, 4.66666666666667, 4.6)	c(1, 3, 5)
none	Children.s	c(13, 13, 13)	c(15, 15, 15)	c(54, 938, 25)	c(5, 5, 5)	c(2, 2, 1)
other	Children.s	c(14, 14, 14)	c(15, 15, 15)	c(335, 1031, 61)	c(5, 5, 4.66666666666667)	c(1, 1, 3)
programmer	Children.s	c(15, 15, 15)	c(15, 15, 15)	c(611, 61, 474)	c(5, 4.28571428571429, 4.23809523809524)	c(1, 7, 21)
retired	Children.s	c(16, 16, 16)	c(15, 15, 15)	c(201, 568, 611)	c(5, 5, 4.5)	c(1, 1, 2)
salesman	Children.s	c(17, 17, 17)	c(15, 15, 15)	c(42, 201, 265)	c(5, 5, 5)	c(1, 1, 1)
scientist	Children.s	c(18, 18, 18)	c(15, 15, 15)	c(474, 52, 60)	c(4.71428571428571, 4.5, 4.33333333333333)	c(7, 2, 3)
student	Children.s	c(19, 19, 19)	c(15, 15, 15)	c(1467, 611, 474)	c(5, 4.33333333333333, 4.24242424242424)	c(1, 3, 33)
technician	Children.s	c(20, 20, 20)	c(15, 15, 15)	c(835, 537, 592)	c(5, 5, 5)	c(2, 1, 1)
writer	Children.s	c(21, 21, 21)	c(15, 15, 15)	c(1113, 1174, 821)	c(5, 5, 4.5)	c(1, 1, 2)
administrator	Western	c(1, 1, 1)	c(16, 16, 16)	c(262, 607, 1170)	c(4.33333333333333, 4.25, 4)	c(3, 4, 1)
artist	Western	c(2, 2, 2)	c(16, 16, 16)	c(1065, 166, 262)	c(4.75, 4.5, 4.5)	c(4, 2, 2)
doctor	Western	c(3, 3, 3)	c(16, 16, 16)	c(117, 47, 166)	c(4, 4, 4)	c(3, 1, 1)
educator	Western	c(4, 4, 4)	c(16, 16, 16)	c(791, 1198, 166)	c(5, 5, 4.42857142857143)	c(1, 1, 7)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
engineer	Western	c(5, 5, 5)	c(16, 16, 16)	c(607, 166, 431)	c(4.66666666666667, 4, 3.85714285714286)	c(3, 2, 21)
entertainment	Western	c(6, 6, 6)	c(16, 16, 16)	c(914, 1065, 47)	c(5, 5, 4.25)	c(1, 1, 4)
executive	Western	c(7, 7, 7)	c(16, 16, 16)	c(1605, 431, 166)	c(5, 4.33333333333333, 4)	c(1, 3, 1)
healthcare	Western	c(8, 8, 8)	c(16, 16, 16)	c(63, 262, 47)	c(4, 4, 3.75)	c(2, 2, 4)
homemaker	Western	9	16	117	4	2
lawyer	Western	c(10, 10, 10)	c(16, 16, 16)	c(607, 47, 431)	c(5, 4.5, 4.5)	c(1, 2, 2)
librarian	Western	c(11, 11, 11)	c(16, 16, 16)	c(262, 607, 1170)	c(4.5, 4, 4)	c(4, 7, 1)
marketing	Western	c(12, 12, 12)	c(16, 16, 16)	c(63, 607, 755)	c(4, 4, 4)	c(2, 1, 1)
none	Western	c(13, 13, 13)	c(16, 16, 16)	c(117, 106, 63)	c(4.33333333333333, 4, 3.5)	c(6, 1, 2)
other	Western	c(14, 14, 14)	c(16, 16, 16)	c(1065, 166, 914)	c(5, 4.14285714285714, 4)	c(1, 7, 1)
programmer	Western	c(15, 15, 15)	c(16, 16, 16)	c(1065, 166, 1106)	c(4.5, 4.33333333333333, 4)	c(2, 6, 1)
retired	Western	c(16, 16, 16)	c(16, 16, 16)	c(1065, 166, 262)	c(5, 4.33333333333333, 4)	c(1, 3, 1)
salesman	Western	c(17, 17, 17)	c(16, 16, 16)	c(262, 117, 755)	c(5, 4.2, 4)	c(1, 5, 1)
scientist	Western	c(18, 18, 18)	c(16, 16, 16)	c(1065, 166, 607)	c(4.5, 4, 4)	c(4, 2, 1)
student	Western	c(19, 19, 19)	c(16, 16, 16)	c(914, 166, 607)	c(5, 4.22222222222222, 4)	c(1, 9, 7)
technician	Western	c(20, 20, 20)	c(16, 16, 16)	c(166, 607, 1065)	c(5, 4, 4)	c(1, 1, 1)
writer	Western	c(21, 21, 21)	c(16, 16, 16)	c(47, 166, 607)	c(4.28571428571429, 4.25, 4)	c(7, 4, 7)
administrator	Animation	c(1, 1, 1)	c(17, 17, 17)	c(954, 1059, 59)	c(5, 5, 4.5)	c(1, 1, 4)
artist	Animation	c(2, 2, 2)	c(17, 17, 17)	c(954, 89, 56)	c(5, 4.8, 4.4)	c(1, 10, 10)
doctor	Animation	c(3, 3, 3)	c(17, 17, 17)	c(59, 89, 332)	c(5, 5, 5)	c(1, 1, 1)
educator	Animation	c(4, 4, 4)	c(17, 17, 17)	c(964, 1093, 56)	c(5, 4.25, 4.21875)	c(2, 4, 32)
engineer	Animation	c(5, 5, 5)	c(17, 17, 17)	c(59, 56, 89)	c(4.4, 4.13888888888889, 4.1)	c(5, 36, 30)
entertainment	Animation	c(6, 6, 6)	c(17, 17, 17)	c(99, 549, 56)	c(5, 5, 4.625)	c(2, 2, 8)
executive	Animation	c(7, 7, 7)	c(17, 17, 17)	c(969, 59, 549)	c(5, 4.33333333333333, 4)	c(2, 3, 3)
healthcare	Animation	c(8, 8, 8)	c(17, 17, 17)	c(1138, 99, 56)	c(5, 4.25, 4.14285714285714)	c(1, 4, 7)
homemaker	Animation	c(9, 9, 9)	c(17, 17, 17)	c(319, 332, 685)	c(5, 4.5, 3)	c(2, 2, 1)
lawyer	Animation	c(10, 10, 10)	c(17, 17, 17)	c(293, 681, 89)	c(5, 5, 4.33333333333333)	c(1, 1, 3)
librarian	Animation	c(11, 11, 11)	c(17, 17, 17)	c(1448, 59, 89)	c(5, 4.4, 4.25)	c(1, 10, 12)
marketing	Animation	c(12, 12, 12)	c(17, 17, 17)	c(89, 264, 59)	c(4.75, 4, 4)	c(4, 2, 1)
none	Animation	c(13, 13, 13)	c(17, 17, 17)	c(267, 293, 332)	c(5, 5, 4.5)	c(1, 1, 2)
other	Animation	c(14, 14, 14)	c(17, 17, 17)	c(59, 969, 267)	c(4.66666666666667, 4, 4)	c(3, 8, 1)
programmer	Animation	c(15, 15, 15)	c(17, 17, 17)	c(89, 56, 59)	c(4.375, 4.05555555555556, 4)	c(32, 36, 7)
retired	Animation	c(16, 16, 16)	c(17, 17, 17)	c(99, 89, 59)	c(4.5, 4, 4)	c(2, 3, 2)
salesman	Animation	c(17, 17, 17)	c(17, 17, 17)	c(59, 264, 56)	c(5, 5, 4)	c(1, 1, 5)
scientist	Animation	c(18, 18, 18)	c(17, 17, 17)	c(56, 89, 59)	c(4.46153846153846, 4.42857142857143, 4)	c(13, 7, 3)
student	Animation	c(19, 19, 19)	c(17, 17, 17)	c(56, 969, 89)	c(4.07446808510638, 3.94444444444444, 3.90566037735849)	c(94, 18, 53)
technician	Animation	c(20, 20, 20)	c(17, 17, 17)	c(1093, 89, 59)	c(5, 4.41666666666667, 4.33333333333333)	c(1, 12, 3)
writer	Animation	c(21, 21, 21)	c(17, 17, 17)	c(89, 969, 56)	c(4.45454545454545, 4.25, 4.17647058823529)	c(11, 4, 17)
administrator	unknown	c(1, 1)	c(18, 18)	c(72, 152)	c(3.42857142857143, 3.2)	c(14, 5)
artist	unknown	2	18	152	4	3
doctor	unknown	c(3, 3)	c(18, 18)	c(152, 72)	c(4, 3)	c(1, 1)

Group.1	Group.2	occupation	genre_transformed	movie_id	mean_rating	Count_Item_ID
educator	unknown	c(4, 4)	c(18, 18)	c(152, 72)	c(4.125, 3.7)	c(8, 10)
engineer	unknown	c(5, 5)	c(18, 18)	c(152, 72)	c(3.92307692307692, 3.15384615384615)	c(13, 13)
entertainment	unknown	c(6, 6)	c(18, 18)	c(152, 72)	c(3, 3)	c(3, 1)
executive	unknown	c(7, 7)	c(18, 18)	c(72, 152)	c(3, 2.66666666666667)	c(1, 3)
healthcare	unknown	c(8, 8)	c(18, 18)	c(152, 72)	c(3.5, 3)	2:1
lawyer	unknown	c(10, 10)	c(18, 18)	c(72, 152)	c(4, 1)	2:1
librarian	unknown	c(11, 11)	c(18, 18)	c(72, 152)	c(3, 2.6)	c(7, 5)
marketing	unknown	c(12, 12)	c(18, 18)	c(152, 72)	c(3, 2)	c(1, 1)
other	unknown	c(14, 14)	c(18, 18)	c(152, 72)	c(3.42857142857143, 3)	c(7, 11)
programmer	unknown	c(15, 15)	c(18, 18)	c(152, 72)	c(4.5, 3.69230769230769)	c(4, 13)
retired	unknown	c(16, 16)	c(18, 18)	c(152, 72)	c(4.33333333333333, 3.66666666666667)	c(3, 3)
salesman	unknown	17	18	72	2.5	2
scientist	unknown	c(18, 18)	c(18, 18)	c(152, 72)	c(3.66666666666667, 2.5)	3:2
student	unknown	c(19, 19)	c(18, 18)	c(152, 72)	c(3.8, 3)	c(10, 37)
technician	unknown	c(20, 20)	c(18, 18)	c(152, 72)	c(4.5, 3)	c(4, 6)
writer	unknown	c(21, 21)	c(18, 18)	c(152, 72)	c(3.16666666666667, 3)	c(6, 4)
administrator	Fantasy	c(1, 1, 1)	c(19, 19, 19)	c(1062, 1278, 1468)	c(5, 5, 5)	c(1, 1, 1)
artist	Fantasy	c(2, 2, 2)	c(19, 19, 19)	c(921, 385, 1278)	c(5, 4.5, 4)	c(2, 2, 1)
doctor	Fantasy	c(3, 3, 3)	c(19, 19, 19)	c(965, 748, 921)	c(4, 3, 3)	c(1, 2, 1)
educator	Fantasy	c(4, 4, 4)	c(19, 19, 19)	c(1062, 945, 965)	c(5, 4.25, 4.25)	c(1, 4, 4)
engineer	Fantasy	c(5, 5, 5)	c(19, 19, 19)	c(945, 1278, 921)	c(5, 5, 4.28571428571429)	c(1, 1, 7)
entertainment	Fantasy	c(6, 6, 6)	c(19, 19, 19)	c(1062, 385, 945)	c(5, 4, 4)	c(1, 3, 1)
executive	Fantasy	c(7, 7, 7)	c(19, 19, 19)	c(1109, 1278, 385)	c(5, 5, 4)	c(1, 1, 6)
healthcare	Fantasy	c(8, 8, 8)	c(19, 19, 19)	c(1615, 559, 748)	c(4, 3.5, 3.28571428571429)	c(1, 2, 7)
homemaker	Fantasy	c(9, 9)	c(19, 19)	c(748, 628)	c(3.5, 3.5)	c(6, 2)
lawyer	Fantasy	c(10, 10, 10)	c(19, 19, 19)	c(559, 945, 385)	c(5, 5, 4.5)	c(1, 1, 2)
librarian	Fantasy	c(11, 11, 11)	c(19, 19, 19)	c(945, 699, 1109)	c(4.2, 4, 4)	c(5, 5, 2)
marketing	Fantasy	c(12, 12, 12)	c(19, 19, 19)	c(1062, 385, 748)	c(5, 3.66666666666667, 3)	c(1, 3, 13)
none	Fantasy	c(13, 13, 13)	c(19, 19, 19)	c(385, 338, 1278)	c(5, 5, 5)	c(2, 1, 1)
other	Fantasy	c(14, 14, 14)	c(19, 19, 19)	c(921, 1062, 699)	c(4.75, 4.5, 4.27272727272727)	c(4, 2, 11)
programmer	Fantasy	c(15, 15, 15)	c(19, 19, 19)	c(965, 974, 1072)	c(4, 4, 4)	c(1, 1, 1)
retired	Fantasy	c(16, 16, 16)	c(19, 19, 19)	c(748, 921, 559)	c(4, 4, 4)	c(2, 2, 1)
salesman	Fantasy	c(17, 17, 17)	c(19, 19, 19)	c(385, 699, 748)	c(4, 4, 3.8)	c(1, 1, 5)
scientist	Fantasy	c(18, 18, 18)	c(19, 19, 19)	c(921, 628, 965)	c(4, 4, 4)	c(4, 2, 1)
student	Fantasy	c(19, 19, 19)	c(19, 19, 19)	c(921, 1069, 699)	c(4.44444444444444, 4, 3.68421052631579)	c(9, 4, 19)
technician	Fantasy	c(20, 20, 20)	c(19, 19, 19)	c(628, 1109, 699)	c(4, 4, 4)	3:1
writer	Fantasy	c(21, 21, 21)	c(19, 19, 19)	c(965, 945, 699)	c(4.5, 4.4, 4.11111111111111)	c(2, 5, 9)

For occupation and genre data has been formatted so that the top 3 movies are shown with their items ids.

Top 3 Movies by Age Group

```
# find the oldest user

max_age=max(U.User$age)

U.User$age_bracket<- cut(U.User$age, breaks = c(0,6, 12, 18, 30, 50,(max_age+1)),
      labels = c("0-6", "6-12", "12-18", "18-30","30-50","50+"),
      right = T)

# grouping by Age bracket


UAgeUser<-merge(U.User,U.Data,by.x ="User_ID",by.y = "User_ID" )
groupby<-by(UAgeUser,list(UAgeUser$age_bracket,UAgeUser$Item_ID),FUN = function(x)
  {
  data.frame(age_bracket=unique(x$age_bracket),
  Item_ID=unique(x$Item_ID),
  mean_rating=mean(x$Rating),
  Count_Item_ID=nrow(x)
  )
  })

Top3Age_Group<-do.call(rbind,groupby)
  Top3Age_Group<-Top3Age_Group[with(Top3Age_Group,order(age_bracket,-mean_rating,-Count_Item_ID)),]


# pick top 3

Top3Age_Group<-by(Top3Age_Group,list(Top3Age_Group$age_bracket),head,n=3)
Top3Age_Group<-do.call(rbind.data.frame,Top3Age_Group)
rownames(Top3Age_Group)<-NULL


# Get the names of the movie

Top3Age_Group<-merge(Top3Age_Group,U.Item[,1:2],by.x = "Item_ID",by.y = "movie_id",sort = FALSE)
Top3Age_Group<-Top3Age_Group[with(Top3Age_Group,order(age_bracket,-mean_rating,-Count_Item_ID)),]


knitr::kable(Top3Age_Group,caption = "Top 3 Movies by Age group")
```

Top 3 Movies by Age group

Item_ID	age_bracket	mean_rating	Count_Item_ID	movie_title
8	6-12	5	1	Babe (1995)
69	6-12	5	1	Forrest Gump (1994)
94	6-12	5	1	Home Alone (1990)
316	12-18	5	5	As Good As It Gets (1997)
134	12-18	5	4	Citizen Kane (1941)
686	12-18	5	4	Perfect World, A (1993)
113	18-30	5	2	Horseman on the Roof, The (Hussard sur le toit, Le) (1995)
119	18-30	5	2	Maya Lin: A Strong Clear Vision (1994)
1189	18-30	5	2	Prefontaine (1997)
851	30-50	5	2	Two or Three Things I Know About Her (1966)
1367	30-50	5	2	Faust (1994)
814	30-50	5	1	Great Day in Harlem, A (1994)
1558	50+	5	3	Aparajito (1956)
1512	50+	5	2	World of Apu, The (Apur Sansar) (1959)
253	50+	5	1	Pillow Book, The (1995)

```
write.csv(x = Top3Age_Group,file = "Top3Age_Group.csv")
```

Top 3 Genres released in Summer

Using the release date column to find movies released in Summer

```
#subsetting the data frame by selecting movies released in Summer

Genre$release_date<-as.Date(Genre$release_date,format="%d-%b-%y")
Genre$release_date<-format (Genre$release_date, "%b")
Genre_Summer<-Genre[Genre$release_date %in% c("May","June","July"),]

Genre_Summer<-Genre_Summer[,c(1:3,25)]

# merging data

Genre_Summer<-merge(Genre_Summer,U.Data,by.x = "movie_id","Item_ID")


groupby<-by(Genre_Summer,list(Genre_Summer$genre_transformed),FUN = function(x)
  {
  data.frame(genre_transformed=unique(x$genre_transformed),
  mean_rating=mean(x$Rating),
  Count_Item_ID=nrow(x)
  )
  })

Top3Genre_Summer<-do.call(rbind,groupby)
Top3Genre_Summer<-Top3Genre_Summer[with(Top3Genre_Summer,order(-mean_rating,-Count_Item_ID)),]


# pick top 3

Top3Genre_Summer<-head(Top3Genre_Summer,n = 3)
rownames(Top3Genre_Summer)<-NULL

knitr::kable(Top3Genre_Summer,caption = "Top 3 Genres released in Summer")
```

Top 3 Genres released in Summer

genre_transformed	mean_rating	Count_Item_ID
Crime	3.477157	197
War	3.455172	145
Documentary	3.455000	200

```
write.csv(x = Top3Genre_Summer,file = "Top3Genre_Summer.csv")
```

For each genre, co-occuring (top2) genre

```
# correlation matrix would give us what genres are closed to what genres
Correlation_Matrix<-cor(U.Item[,5:23])
Correlation_Matrix<-as.data.frame(Correlation_Matrix)

find_co_occuring<-function(cname)
{
  top2<-tail(head(with(Correlation_Matrix,order(-cname)),n=3),n = 2)
  rownames(Correlation_Matrix)[top2]
}

Top2CoOccuring<-apply(Correlation_Matrix, 2, find_co_occuring)

knitr::kable(Top2CoOccuring,caption = "Top 2 Co-Occuring Genres for each genre")
```

Top 2 Co-Occuring Genres for each genre

unknown	Action	Adventure	Animation	Children.s	Comedy	Crime	Documentary	Drama	Fantasy	Film.Noir	Horror	Musical	Mystery	Romance	Sci.Fi
Fantasy	Adventure	Action	Children.s	Animation	Romance	Film.Noir	unknown	War	Children.s	Crime	Sci.Fi	Animation	Thriller	Comedy	Action
Film.Noir	Sci.Fi	Children.s	Musical	Fantasy	Musical	Thriller	War	Crime	Adventure	Mystery	Thriller	Children.s	Film.Noir	War	Adventu

```
write.csv(x = Top2CoOccuring,file = "Top2CoOccuring.csv")
```

This means that whenever the movie has been rated for action it has also been rated for Adventure and Sci-Fi