MovieLens Team5
Fall 2017

BY

HARSH PATHAK

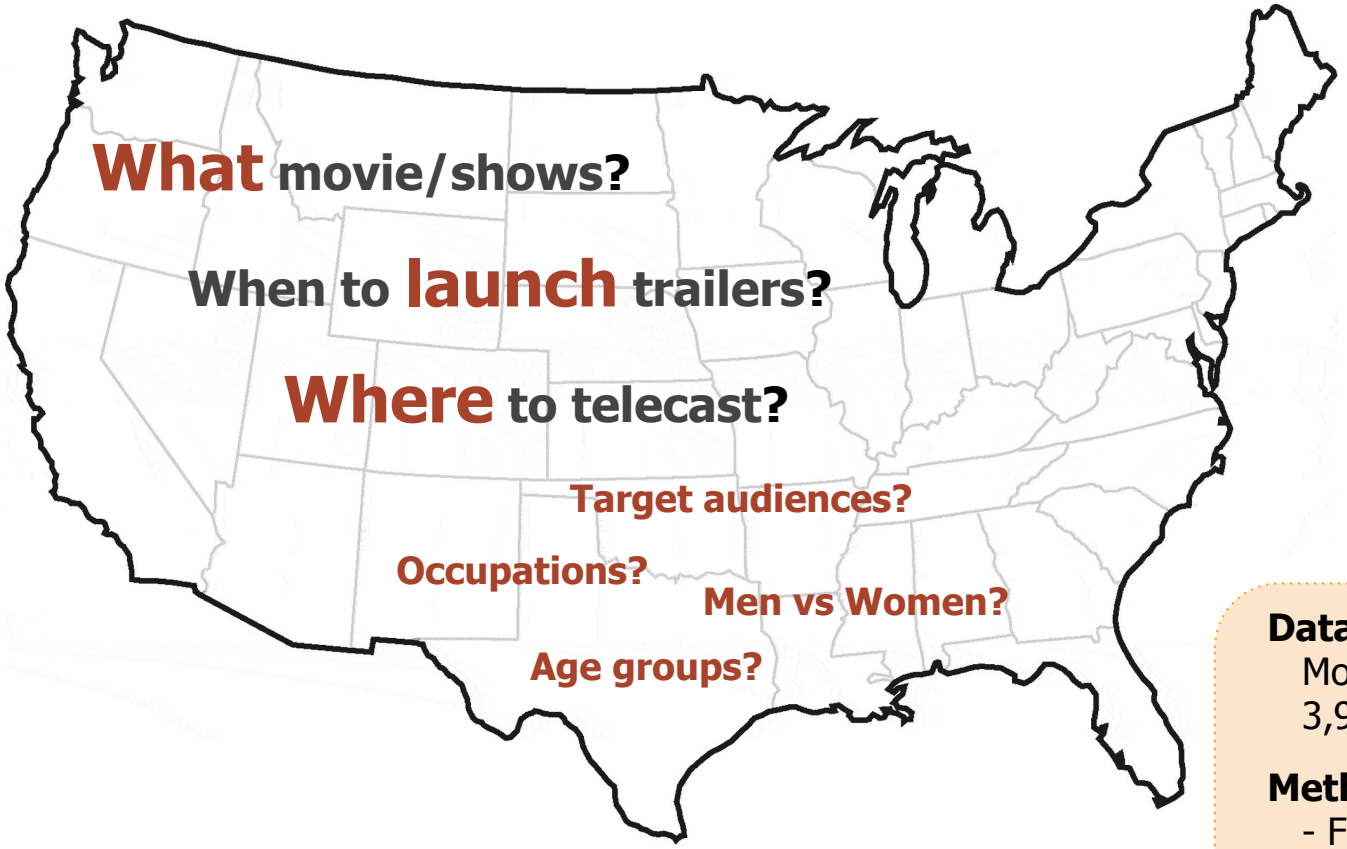
JIDAPA THADAJARASSIRI

KRUSHIKA TAPEDIA

PITCHAYA WIRATCHOTISATIAN

PRINCE SHIVA CHAUDHARY

What America's watching?



Data:
MovieLens 1M ratings
3,900 Movies / 6,040 Users

Methodology:
- Frequency analysis
- Histogram/Scatter plot
- Correlation

Data Preparation

Data Download → Downloaded dataset from http://grouplens.org/datasets/movielens/.

Data Merging → Left join on rating and movies, called [ratings and movies]
Left join on [ratings and movies] with user's data frame.

Data storage → Stored resultant data frame in HDF5 file.

	UserID	MovieID	Rating	Timestamp	Title	Genres	Gender	Age	Occupation	Zip-code
0	1	1193	5	978300760	One Flew Over the Cuckoo's Nest (1975)	Drama	F	1	10	48067
1	1	661	3	978302109	James and the Giant Peach (1996)	Animation|Children's|Musical	F	1	10	48067
2	1	914	3	978301968	My Fair Lady (1964)	Musical|Romance	F	1	10	48067
3	1	3408	4	978300275	Erin Brockovich (2000)	Drama	F	1	10	48067
4	1	2355	5	978824291	Bug's Life, A (1998)	Animation|Children's|Comedy	F	1	10	48067

I. Basic details (1/2)

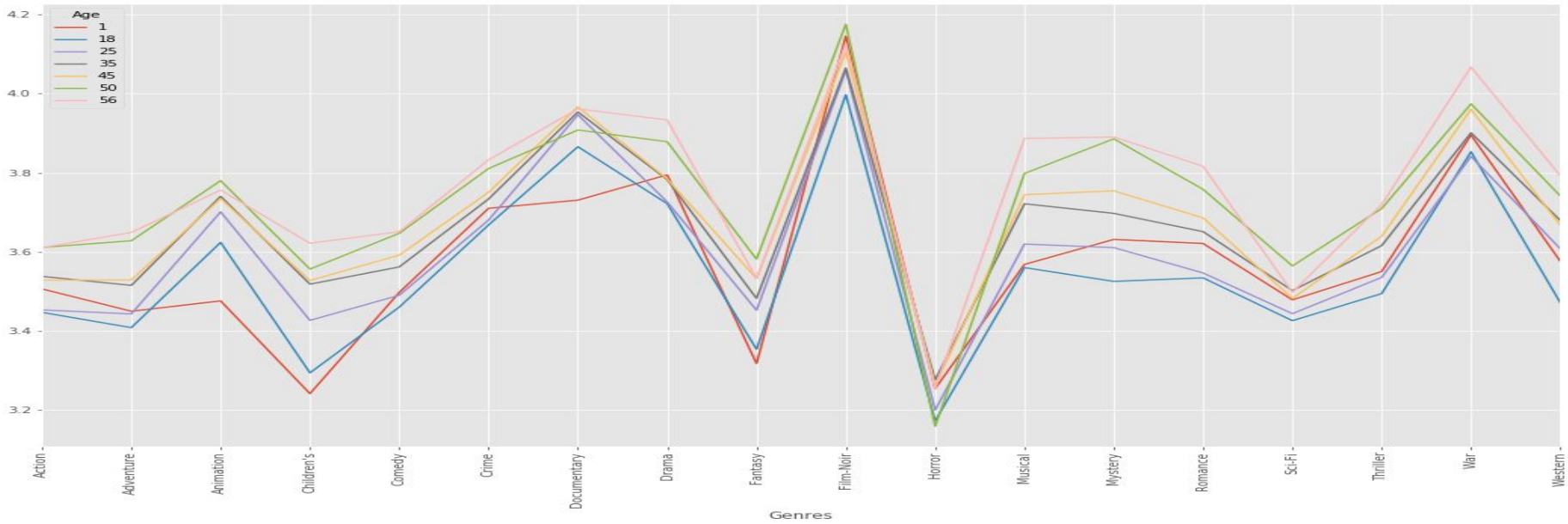
- How many movies have an average rating over 4.5 overall? → **21 movies**
- How many movies have an avg rating over 4.5 among men and women? → **men: 23, women: 51**
- How many movies have a median rating over 4.5 among men over age 30? How about women over age 30? → **men: 86, women: 149**
- What are the ten most popular movies?
  Popular movies: A popular movie is one, which received highest number of ratings. However, ratings can be a subjective to individual, depending on individual's taste or liking of cinema.

	Title	Rating
0	American Beauty (1999)	3428
1	Star Wars: Episode IV - A New Hope (1977)	2991
2	Star Wars: Episode V - The Empire Strikes Back...	2990
3	Star Wars: Episode VI - Return of the Jedi (1983)	2883
4	Jurassic Park (1993)	2672

	Title	Rating
5	Saving Private Ryan (1998)	2653
6	Terminator 2: Judgment Day (1991)	2649
7	Matrix, The (1999)	2590
8	Back to the Future (1985)	2583
9	Silence of the Lambs, The (1991)	2578

I. Basic details (2/2)

Conjecture: People in different age groups have different rating behavior. Some specific age group might be easier to please.



People in aged group (18-24 and 25-34) generally give ‘low’ ratings to almost each segment of movies when compared against old people of age group belonging to 56+ generally give ‘high’ ratings and they are easy to please.

II. Histograms (1/3)

Rating of all movies

misleading because of different number of rating for each movie

Figure 2.1: The ratings of all movies

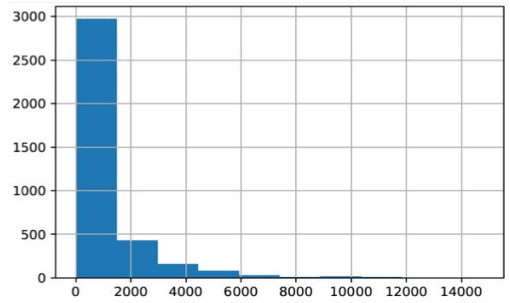
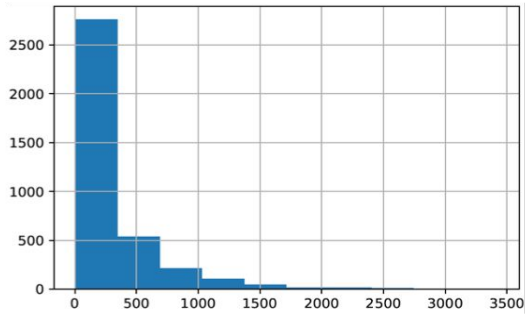


Figure 2.2: The number of ratings each movie received



Average rating for each movie

tails for rated movie > 100 times - less spread imply less variance → preferable

Figure 2.3: The average rating for each movie

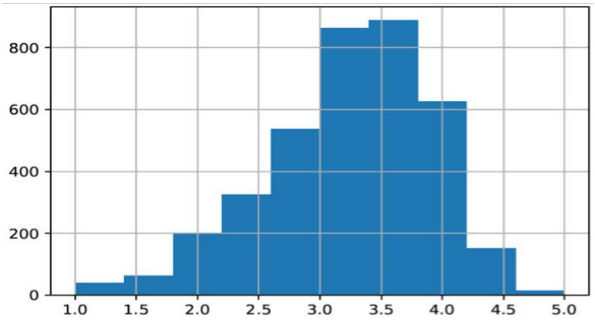
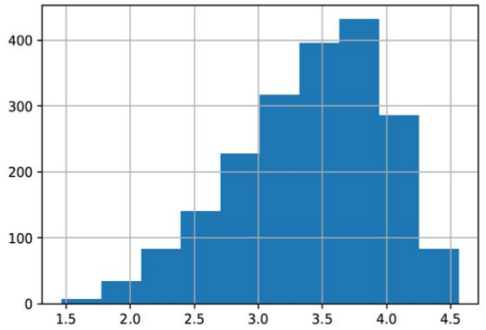
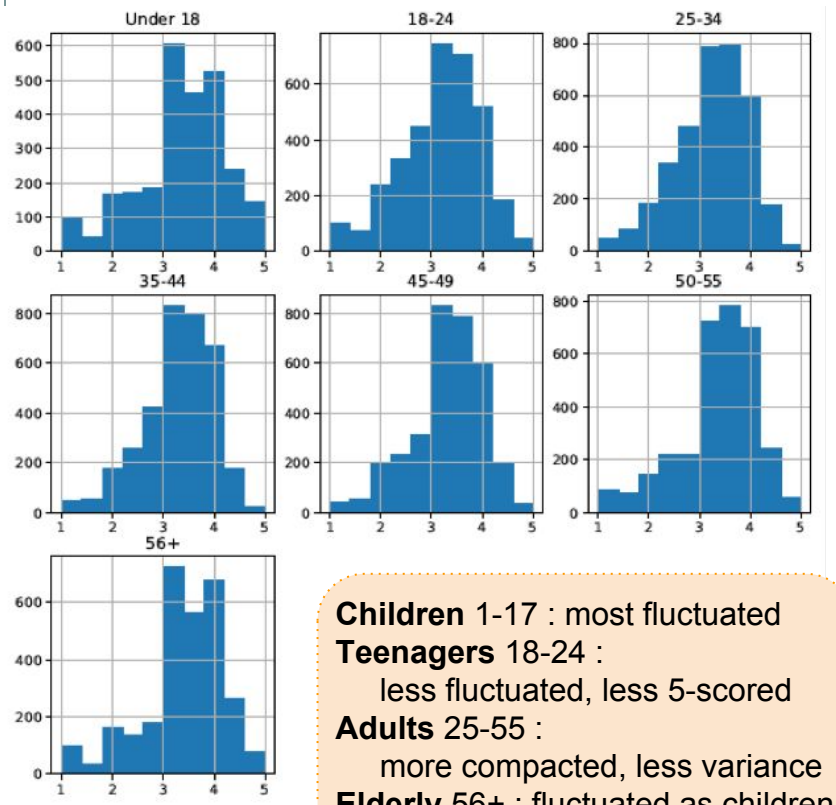


Figure 2.4: The average rating for movies which are rated more than 100 times



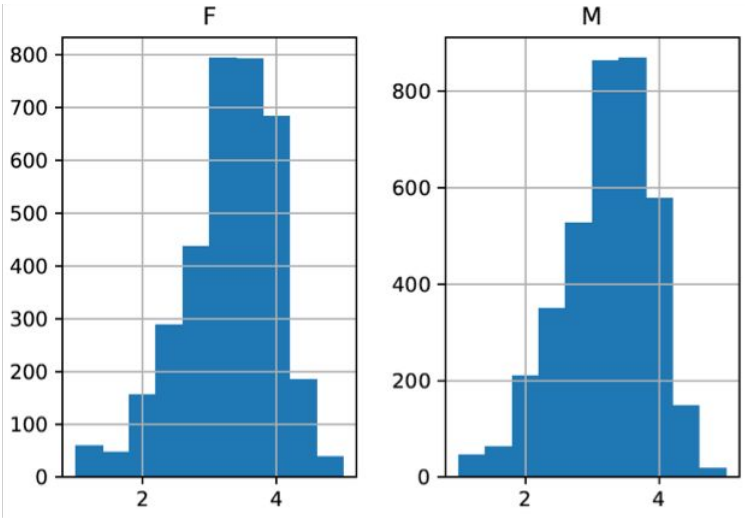
II. Histograms (2/3)

Conjecture 1: The distribution of movie rating may be different according to **age groups**



Children 1-17 : most fluctuated
Teenagers 18-24 :
less fluctuated, less 5-scored
Adults 25-55 :
more compacted, less variance
Elderly 56+ : fluctuated as children

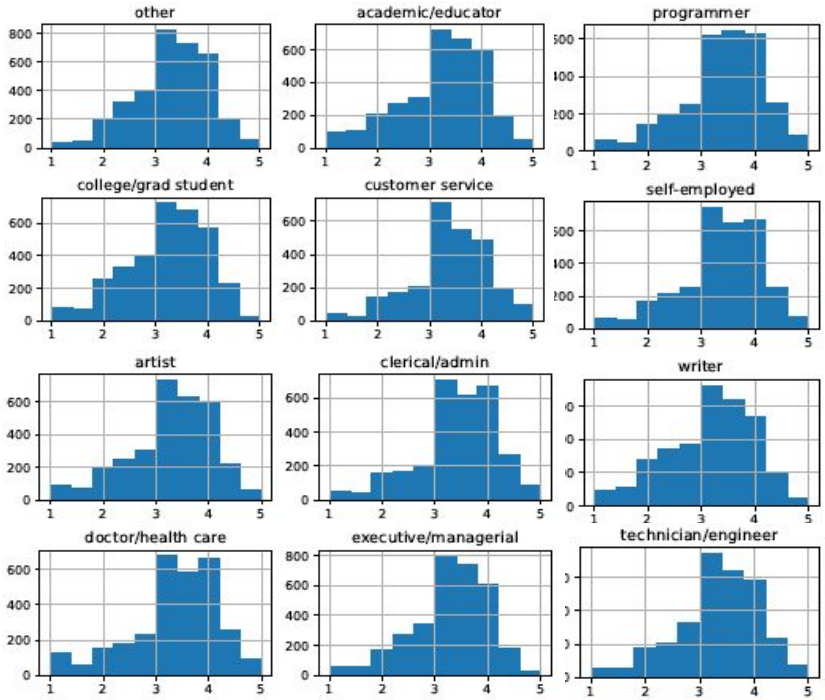
Conjecture 2: The distribution of movie rating may be different according to **gender**

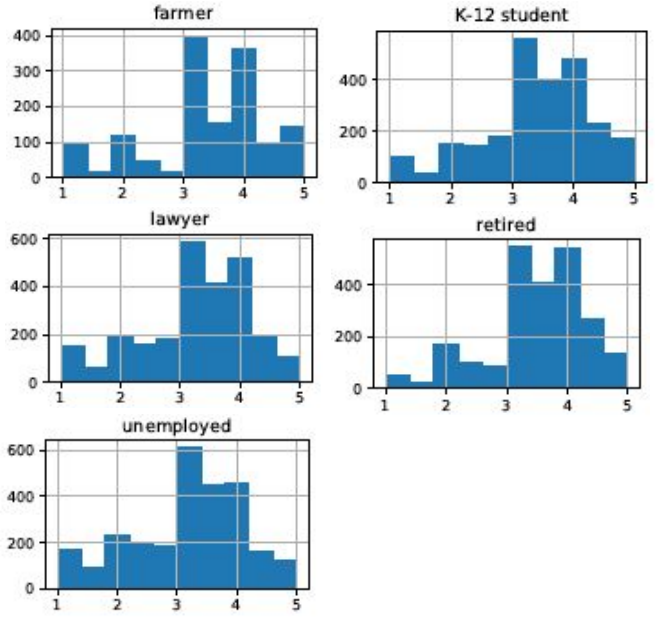


Women - more left-skewed than men
Women - easier to rate 1-scored

II. Histograms (3/3)

Conjecture 3: The distribution of movie rating may differ according to **occupations**





Most occupations

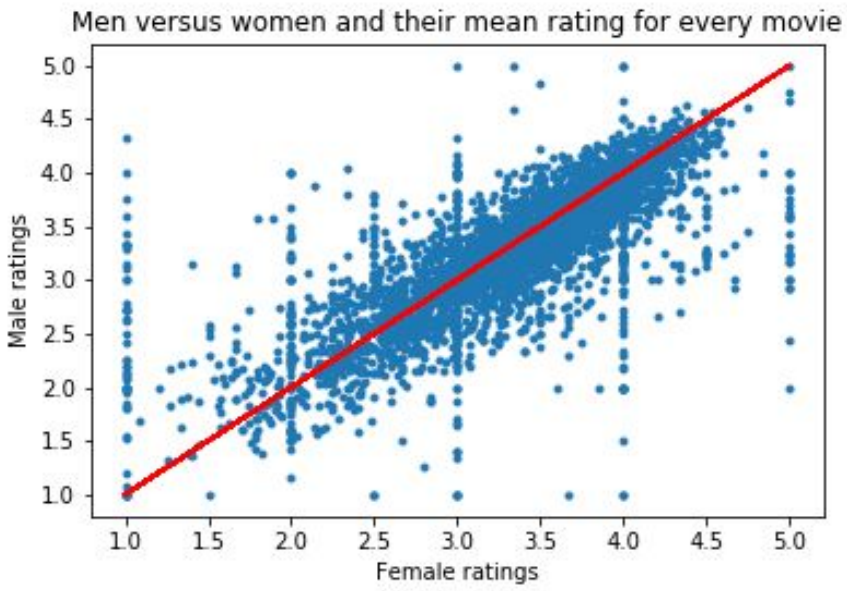
left-skewed and peak between 3 to 4

Farmer, K-12 student, lawyer, retired, unemployed

easier to give extreme ratings

III. Correlation - men vs women (1/3)

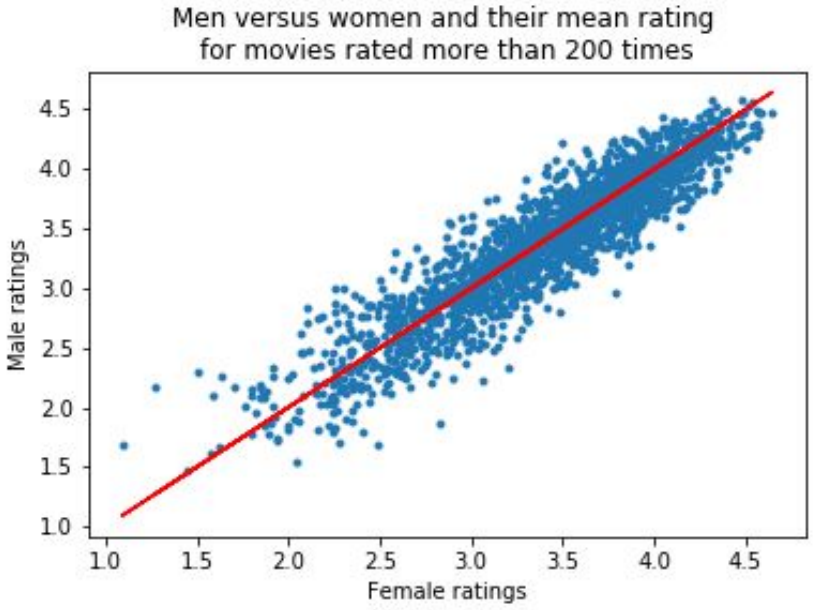
Every Movies



Correlation Coefficient : **0.7602**

- Positive correlation
- Many departed scatter points

Movies rated > 200 times



Correlation Coefficient : **0.9204**

- More predictability
- Even strongly positive correlation
- Both genders tends to have similar opinions towards popular movies

III. Correlation - men vs women (2/3)

Conjecture 1: Men and women are more similar for specific age groups, considered their rating over the same genres

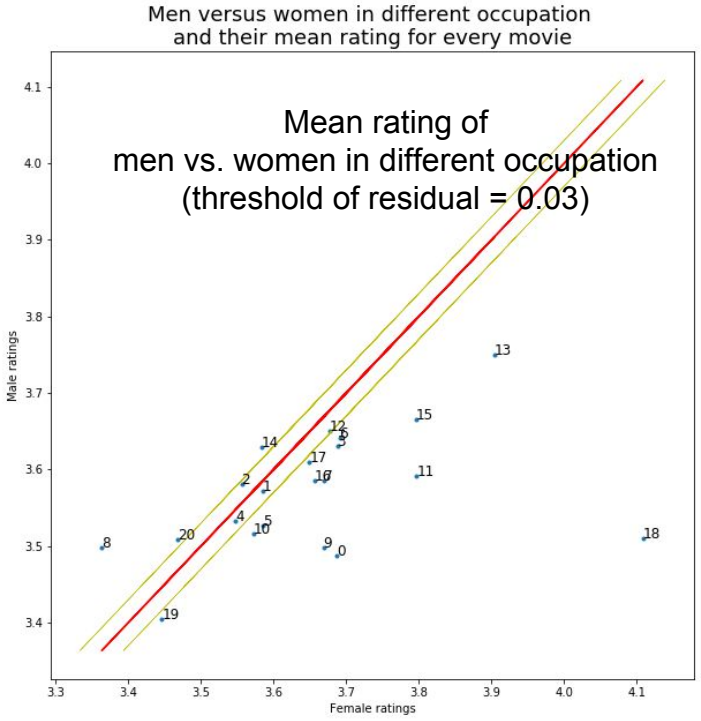
Correlation of male vs female over genres

Age group	
Children	0.6824
Teenager	0.8764
Adult	0.9135
Elderly	0.8821

Men and women more similar when they are older (larger positive correlation)

Age groups
- Children: under 18
- Teenager: 18-24
- Adult: 25-55
- Elderly: 56+

Conjecture 2: Men and women are more similar for some occupations

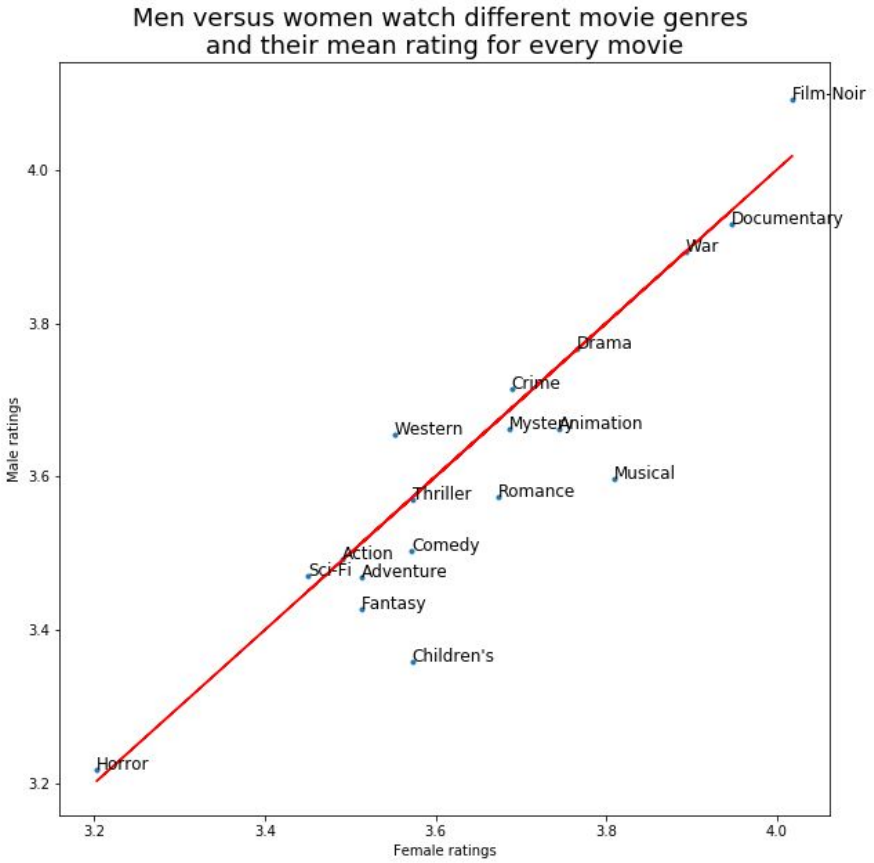


Men and women are more similar when they work as academic/educator (1), artist (2), college/grad student (4), and programmer (12)

III. Correlation - men vs women (3/3)

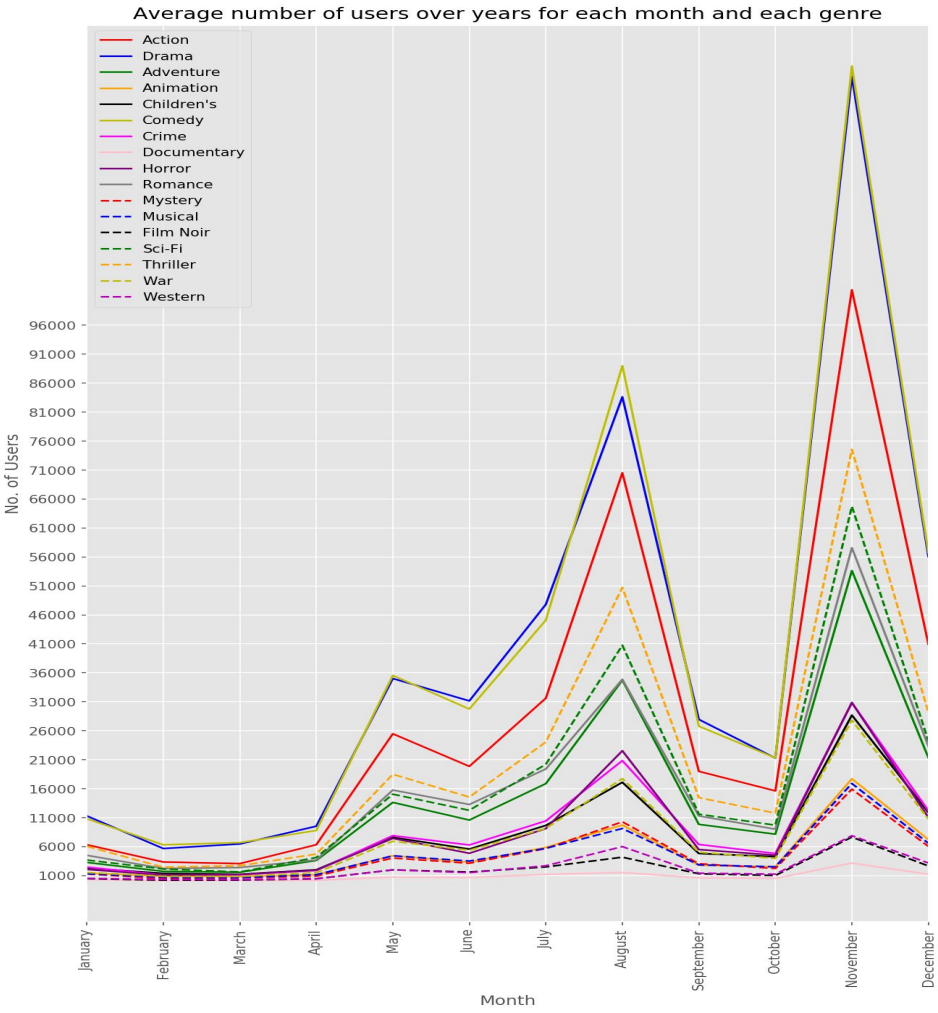
Conjecture 3: **Men and women are more similar for some genres**

The following movie genres: Action, Drama, War, and Triller, lie on the straight line and, thus, men and women are more similar (giving the same ratings) when they watch movies in those genres.



IV. Business Questions (1/3)

Which movies/tv shows to make/telecast and when to do that?



Most

watched months → **Nov & Aug**

watched genres → **Comedy & Drama**

Business Strategy

Make Comedy & Drama movies

- all time fav. types of movie
- valued investment b/c can telecast any time
- long run profit increase

Strategize Telecasting

⬆ variety of comedy and drama in Nov & Aug
  - make audiences to stick to our channel

⬆ variety of type of movies in less different genres such as Oct, Apr
  - target audience preferences

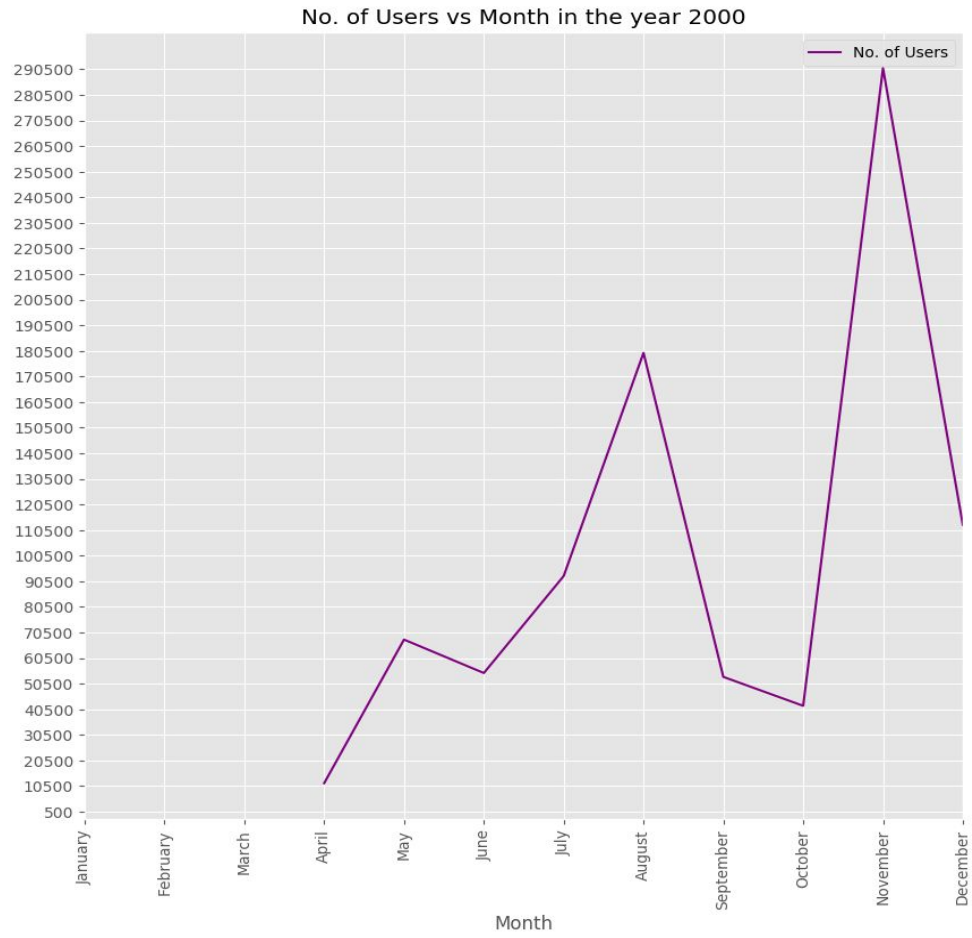
IV. Business Questions (2/3)

When to launch upcoming movie trailers?

year **2000** - max. information

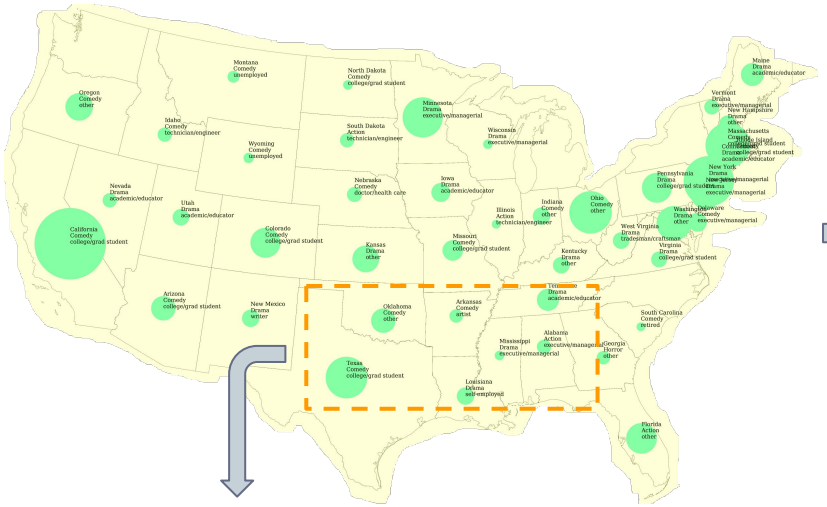
preferred time watching movie → **Nov & Aug**

releasing or advertising movie trailers → **Before/During Nov & Aug**

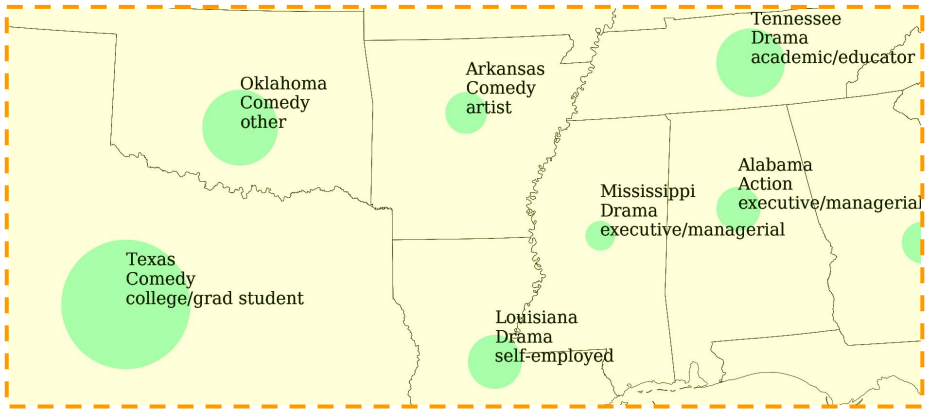


IV. Business Questions (3/3)

Which movies/tv shows to telecast at which location considering occupation?



In which part of the country what occupation rates the most. We define this term as User Rate Frequency (URF).



Here is the focused view of the map visual, where we can see popular genre in among Texas people with occupation student is comedy and being a movie company we want feedbacks for our movies and local TV-shows.

Thank you

Q & A