Seen any good movies lately?

Understanding Film Preferences for Recommendations Solely on Background Information

Motivation

Seen any good movies lately?

Have you seen any movies
I would like to see?

What movies do they *like*?

What movies have they *seen*?

Sign up

Name

Email

Birth date

Male ◉ Female ○

Sign up

By clicking Sign up, I accept the Terms of Service and Privacy Policy.

“

Based solely on their personal information, how do we recommend movies to others?

6k
users

gender
age
occupation
zip code

3.8k
movies

title
genre

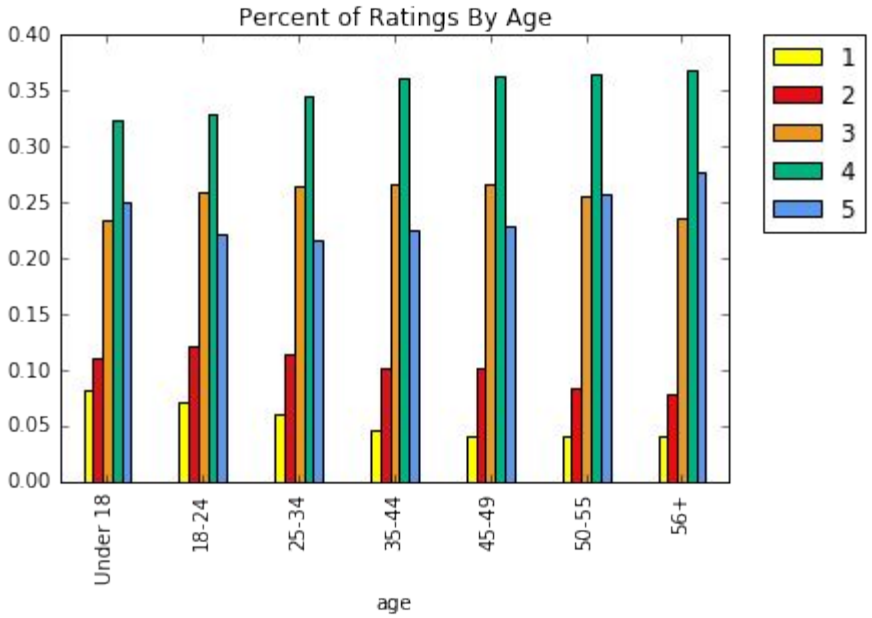
1m
reviews

movie
user
rating
timestamp

Objectives

1. Analyze Basic Statistics

Conjecture: The older a person gets, the more difficult they are to please.



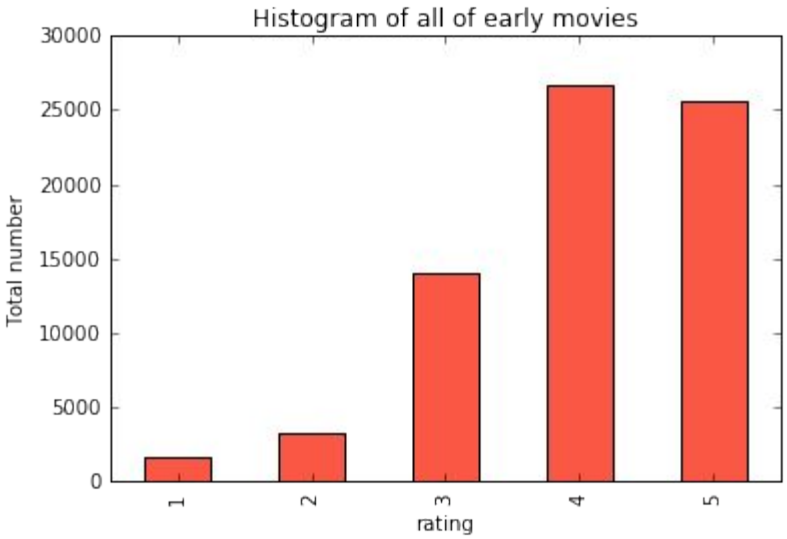
rating	1	2	3	4	5
age					
Under 18	0.082246	0.109625	0.234464	0.323693	0.249972
18-24	0.071174	0.120265	0.259355	0.328224	0.220981
25-34	0.060416	0.113301	0.263647	0.345903	0.216733
35-44	0.045562	0.101772	0.266277	0.361718	0.224670
45-49	0.040761	0.100881	0.266773	0.362704	0.228881
50-55	0.040668	0.082673	0.254725	0.365347	0.256587
56+	0.039995	0.077385	0.236282	0.368669	0.277669

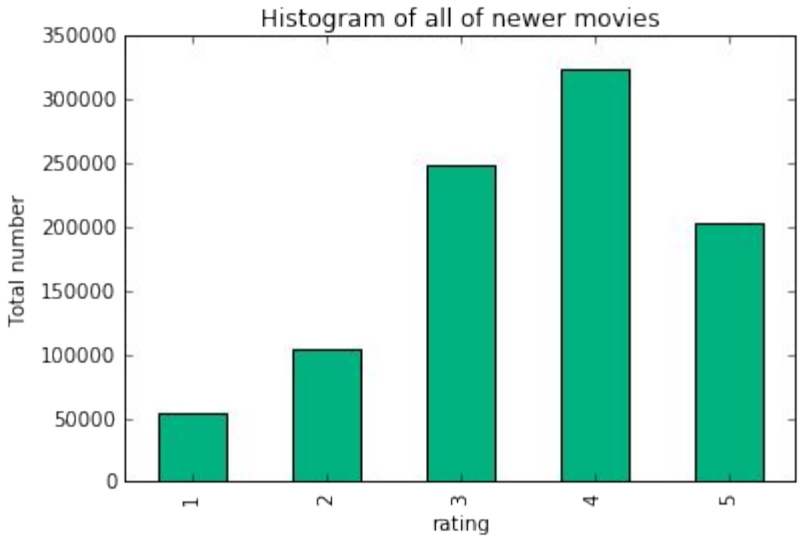
Conclusion

- False, given the data
- Older people are:
  - more likely than younger people to give a rating of 5
  - and less likely to give a rating of 1
- Increase in average rating in higher age groups

2. Expand Our Investigation to Histograms

Conjecture: The distribution of all ratings of older movies is less normally-distributed than that of newer movies.



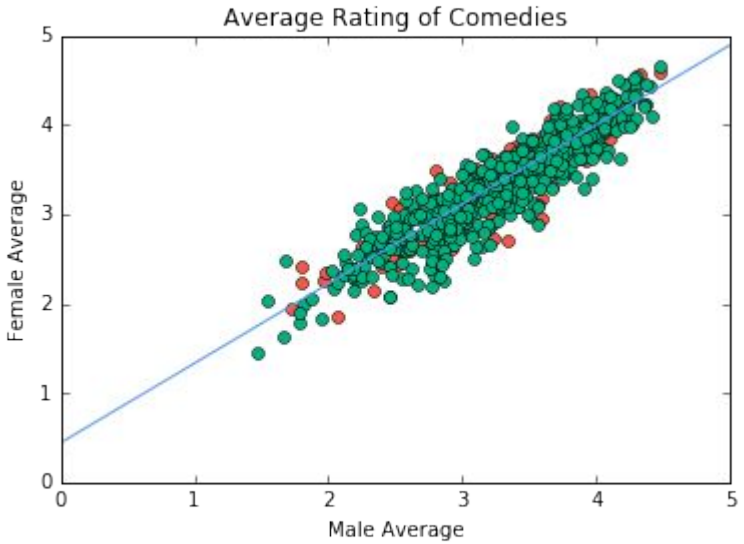


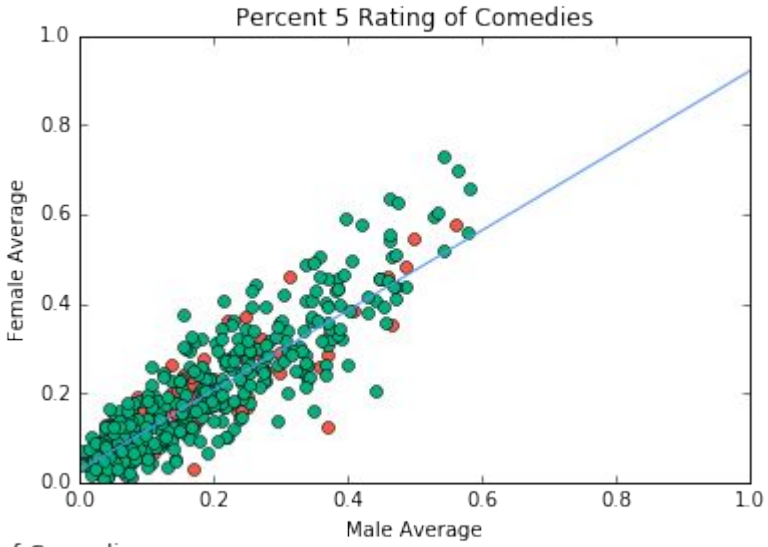
Conclusion

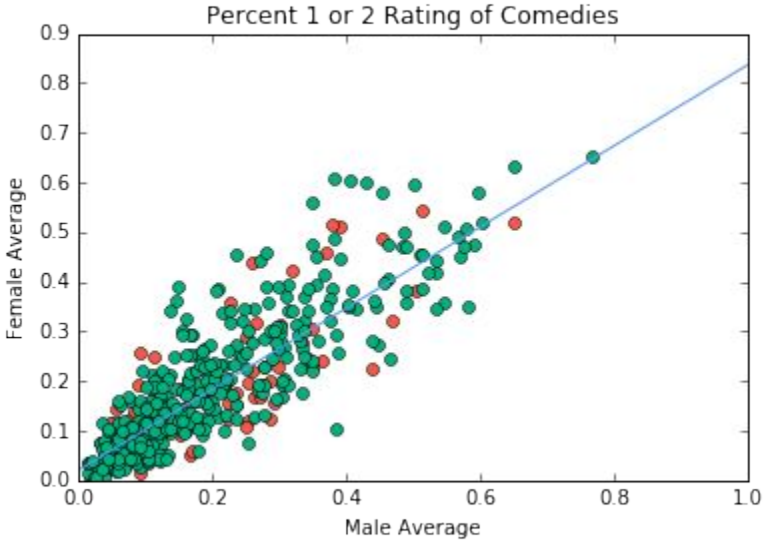
- Framed in terms of distributions as there is a gap in quantity of ratings: new movies rated more frequently
- Supported by the data
- *Nostalgia factor*, older movies receive disproportionately more high-end ratings

3. Explore Male Versus Female Ratings

Conjecture: Genders agree on what is "funny"







Training data

Testing data

Linear model

Conclusion

- Given that the average rating per movie:
  - percent 5 rating per movie
  - and percent 1 or 2 rating per movie are all reliably (and linearly) predictable between genders
- Supported by the data (movies with at least 100 ratings)
- Genders tend to agree on what is funny

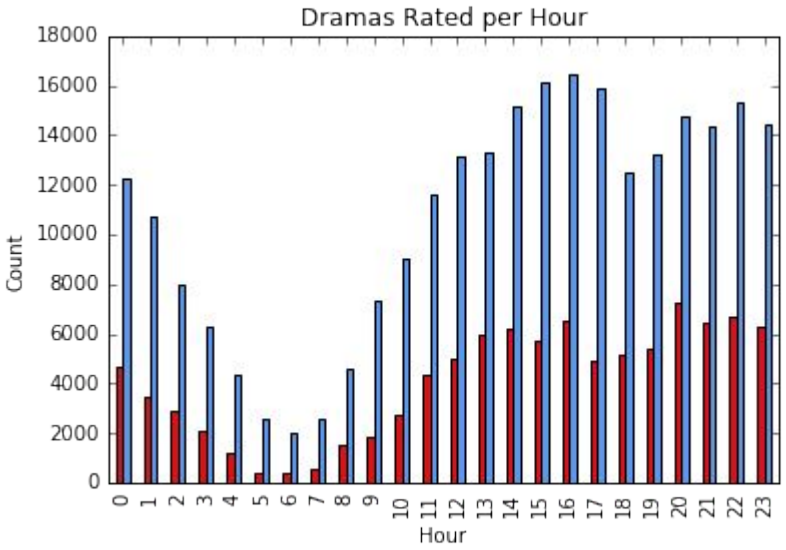
4. Relate to Business Intelligence

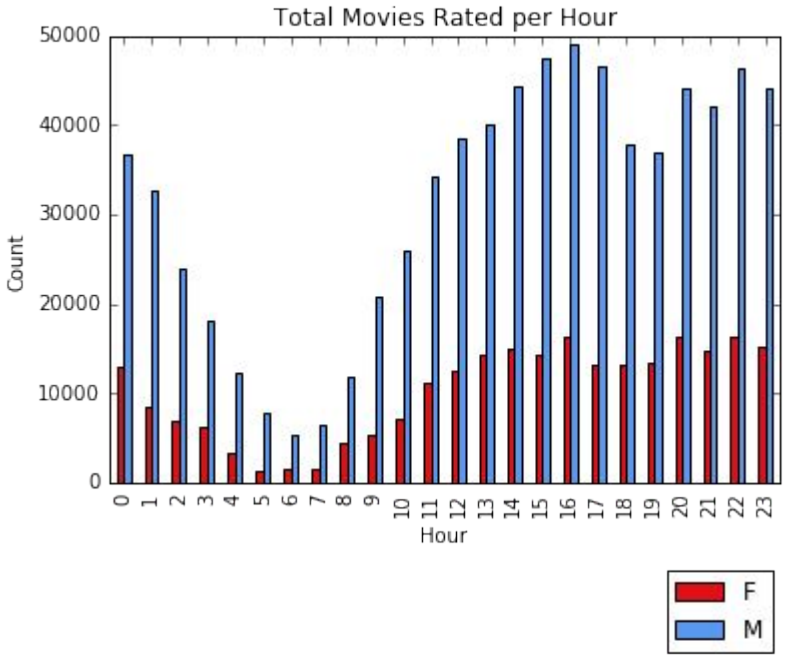
A What is the best time to recommend a drama for each gender?

B Which occupation is most likely to enjoy a comedy?

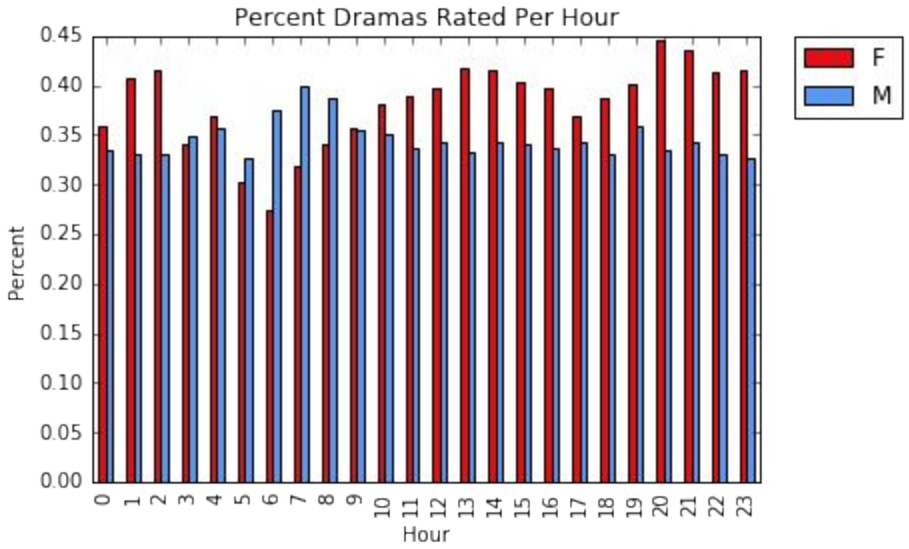
C What age group watches the most adventure movies?

A. What is the best time to recommend a drama for each gender?





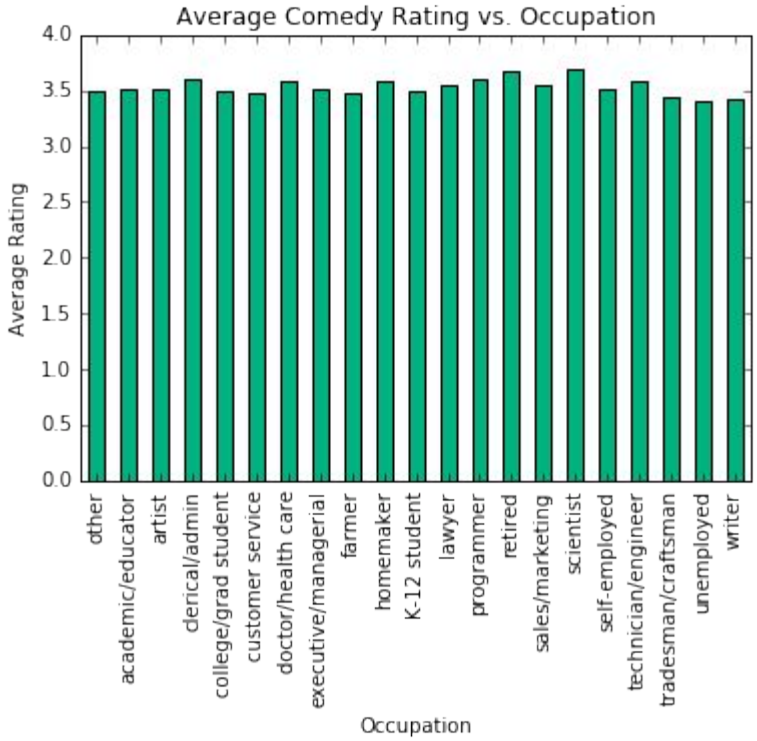
A. What is the best time to recommend a drama for each gender?

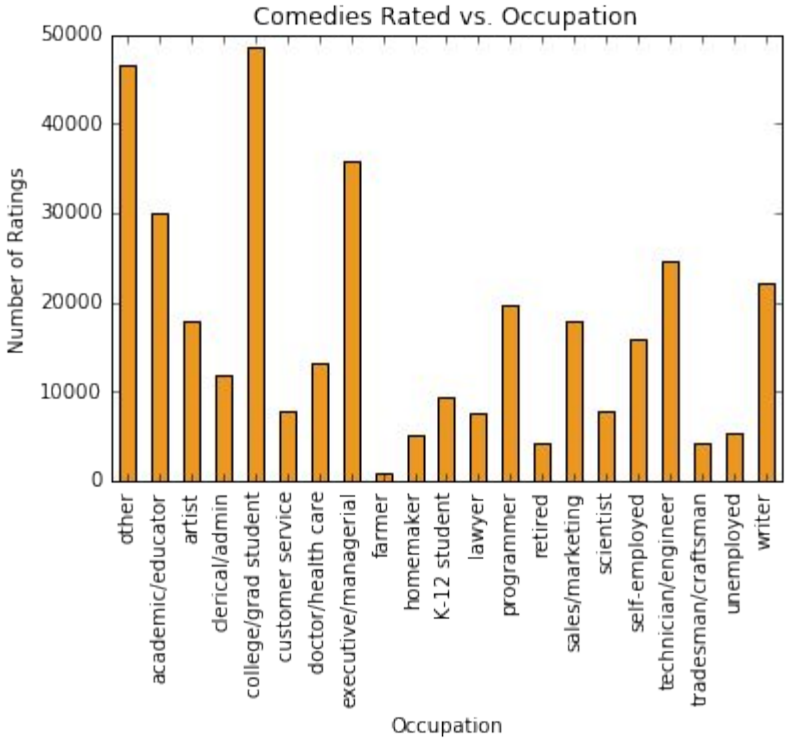


Conclusion

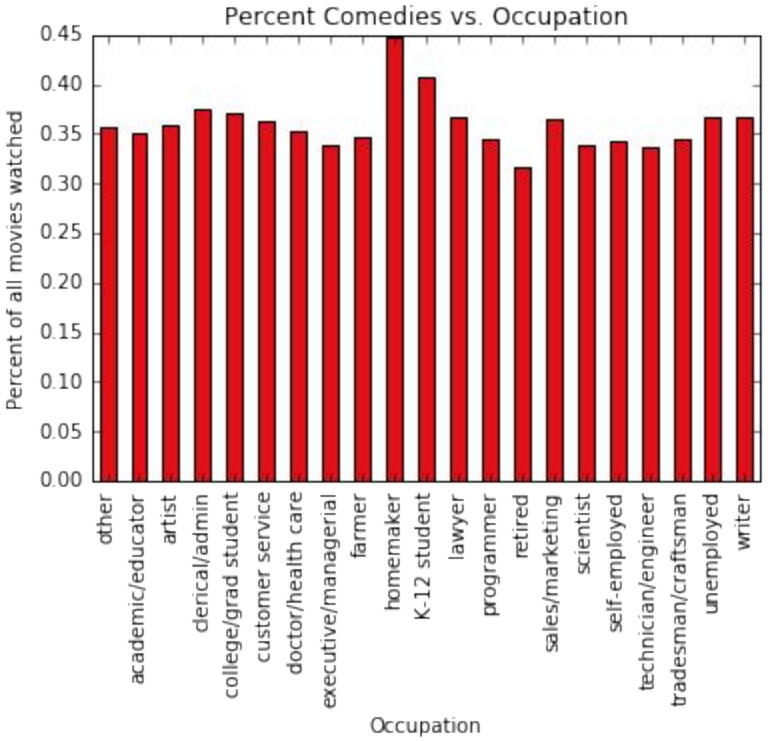
- Women show significantly higher interest in Dramas in the evening than in the morning
- Men are more interested in morning Drama-viewing compared to women
- **Further analysis** Movie length, run-time

B. Which occupation is most likely to enjoy a comedy?





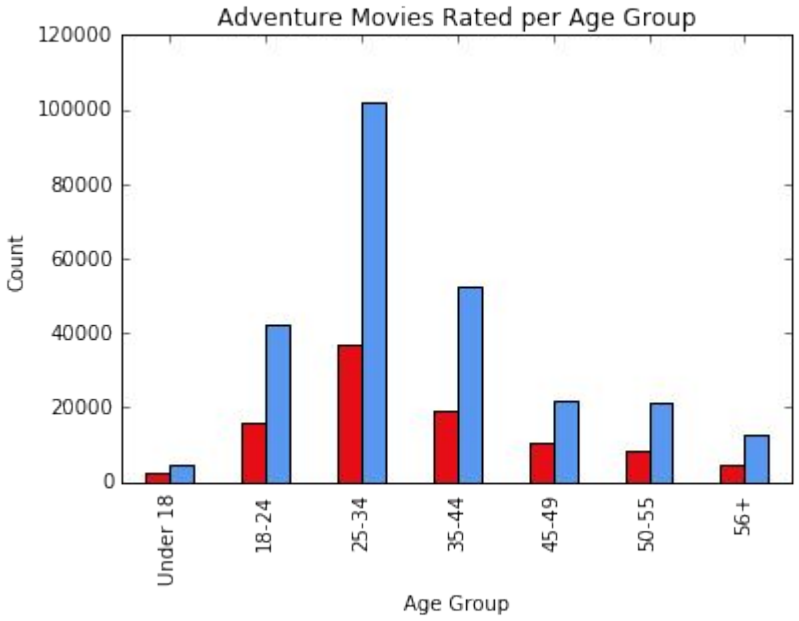
B. Which occupation is most likely to enjoy a comedy?

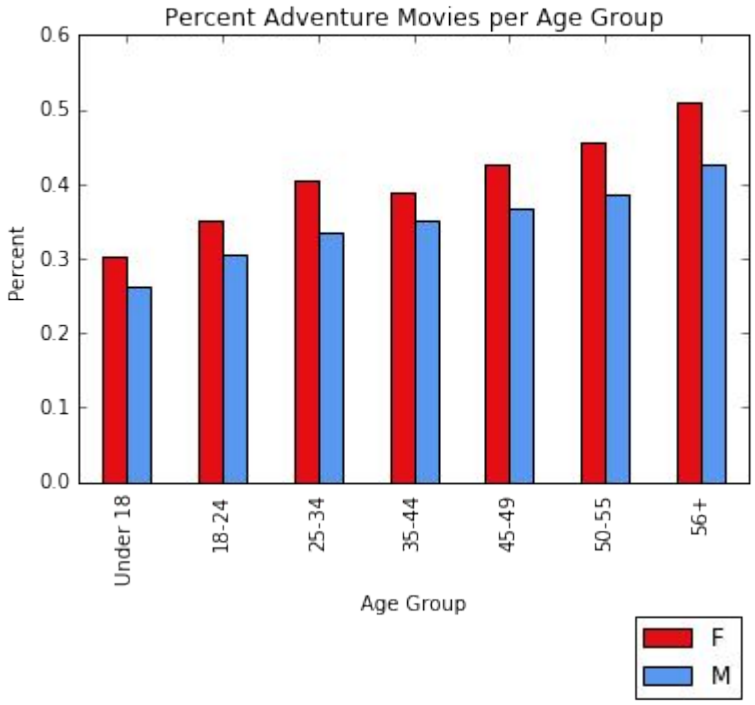


Conclusion

- College students, homemakers, and children all show high interest in comedies compared to other occupations
- **Further analysis** Greater stratification within the “other” occupation

C. What age group watches the most adventure movies?





Conclusion

- Women show a particularly strong interest in Adventure movies
  - Nearly half of movies watched by older women were adventure movies
- Older people are more interested in adventure movies than younger individuals
- **Further analysis** Greater representation among women, finer age-group segmentation

Data Limitations

Data Limitations

- All movies released before 2000
- Under representation (rating-response) bias
  - *Only* 6k user ratings
  - All from United States
- Sampling bias
  - 3:1 male to female ration

Conclusions

Conclusions

- Older people are easier to please (among “popular” movies)
- Nostalgia factor
- Genders tend to agree what is funny
- Unwise to “bet-the-farm” recommending a single genre
  - Instead determine multiple genres (4 or 5) that each individual is likely to show interest in
- No “first-time” recommendation silver bullet given the data

Questions