Lover Recommendations Using Unsupervised Learning
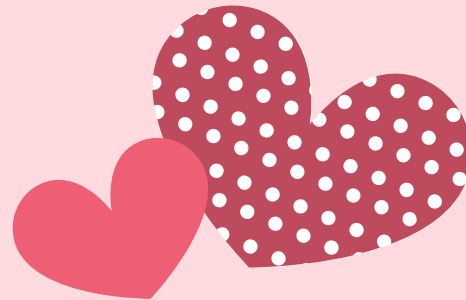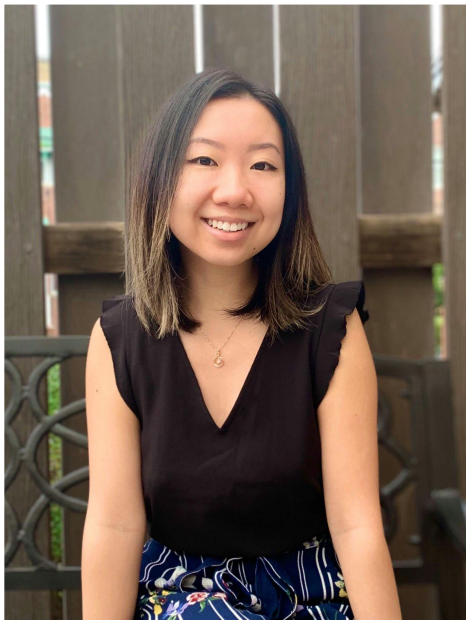
Veronica Leong

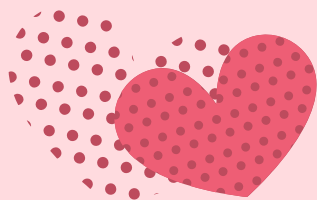# About Me

VERONICA LEONG
DATA SCIENTIST

LinkedIn | Portfolio | GitHub

Agenda ♥

01 BACKGROUND

02 PROBLEM STATEMENT

03 DATA CLEANING

04 DATA VISUALIZATIONS

05 MODELING ANALYSES

06 STREAMLIT DEMO

07 CONCLUSION

08 FUTURE PROJECT REFINEMENTS
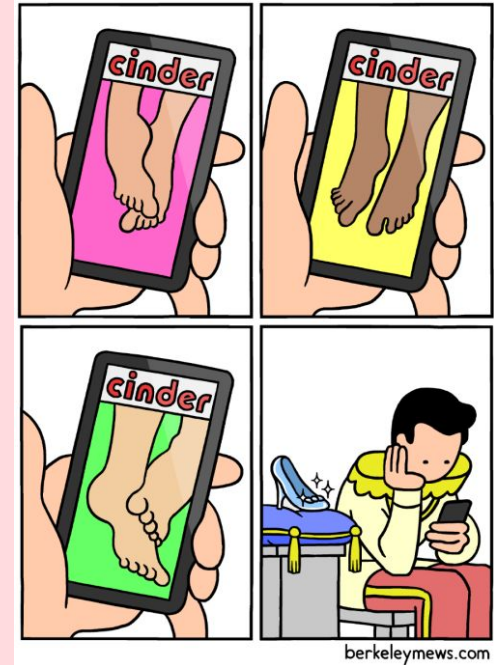
09 APPENDIX
- Reference Links
- Elaborated Data Resources

How COVID Affected Online Dating

- +15% new subscribers for Match Group apps Q1'20
- Aided loneliness and isolation
- Virtual dating = new normal
- Apps adopted to COVID guidelines

Problem Statement

Using unsupervised learning, can we build a model to recommend potential matches for a user who's looking for a partner?

Data Collection & Cleaning *

- Kaggle - OkCupid Profile Dataset
  - 14.2+ MB file → 3.7 MB file

- Dropped
  - "Essay", "Last Online", "Sign", "Height", "Ethnicity", "Job", "Income", "Speaks"

- Filtered status (single / available)

- Filled missing values
  - Offspring → Doesn't have kids
  - Pets → Dislikes dogs and cats

- Grouped / Generalized values
  - Ex) Drinks

Original Value	"Generalized" Value
Often	Yes
Desperately	
Socially	Sometimes
Rarely	
Not at all	No

* Further detailed in Appendix -- Missing values; Grouped values

Data Viz

Age Distribution by Sex

* Data Source: Kaggle - OkCupid Profiles

Data Viz (cont'd)

Diet by Religion
(excludes "anything" diet & "doesnt matter" religion)

8
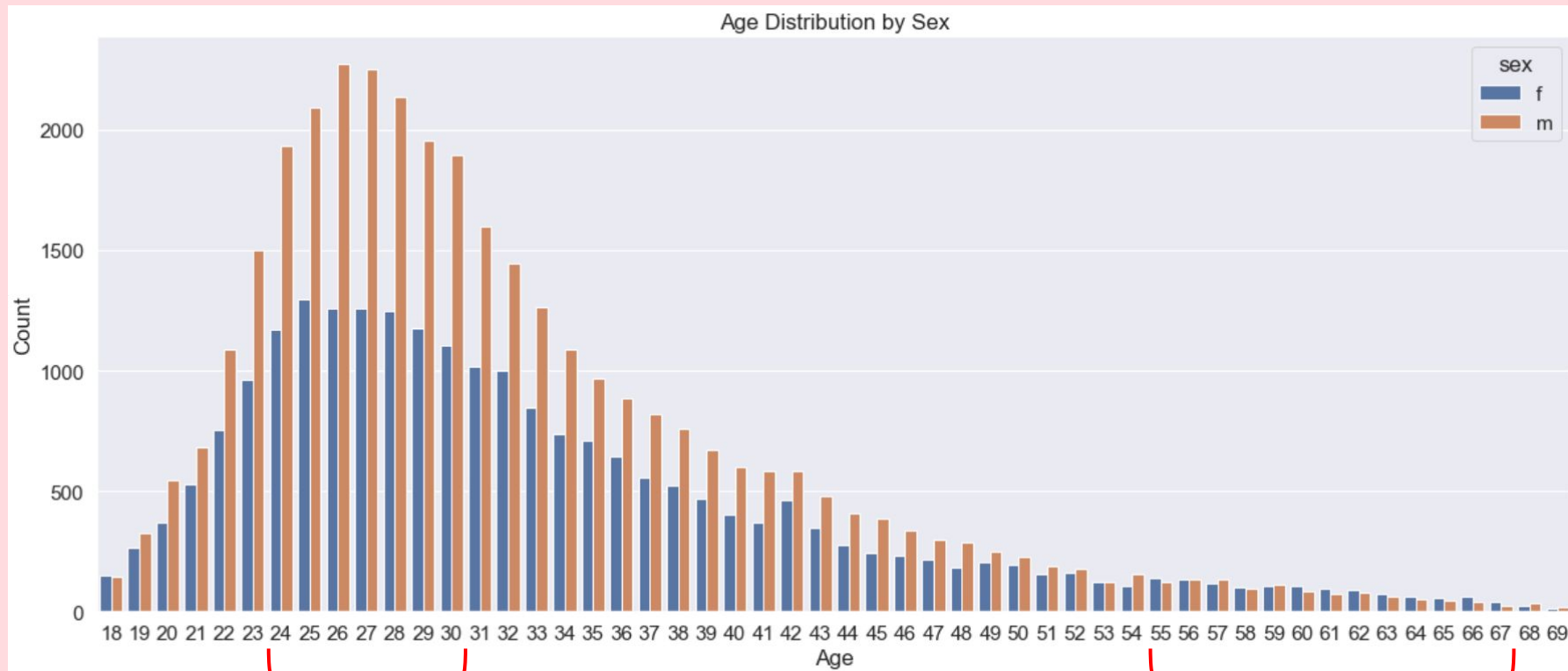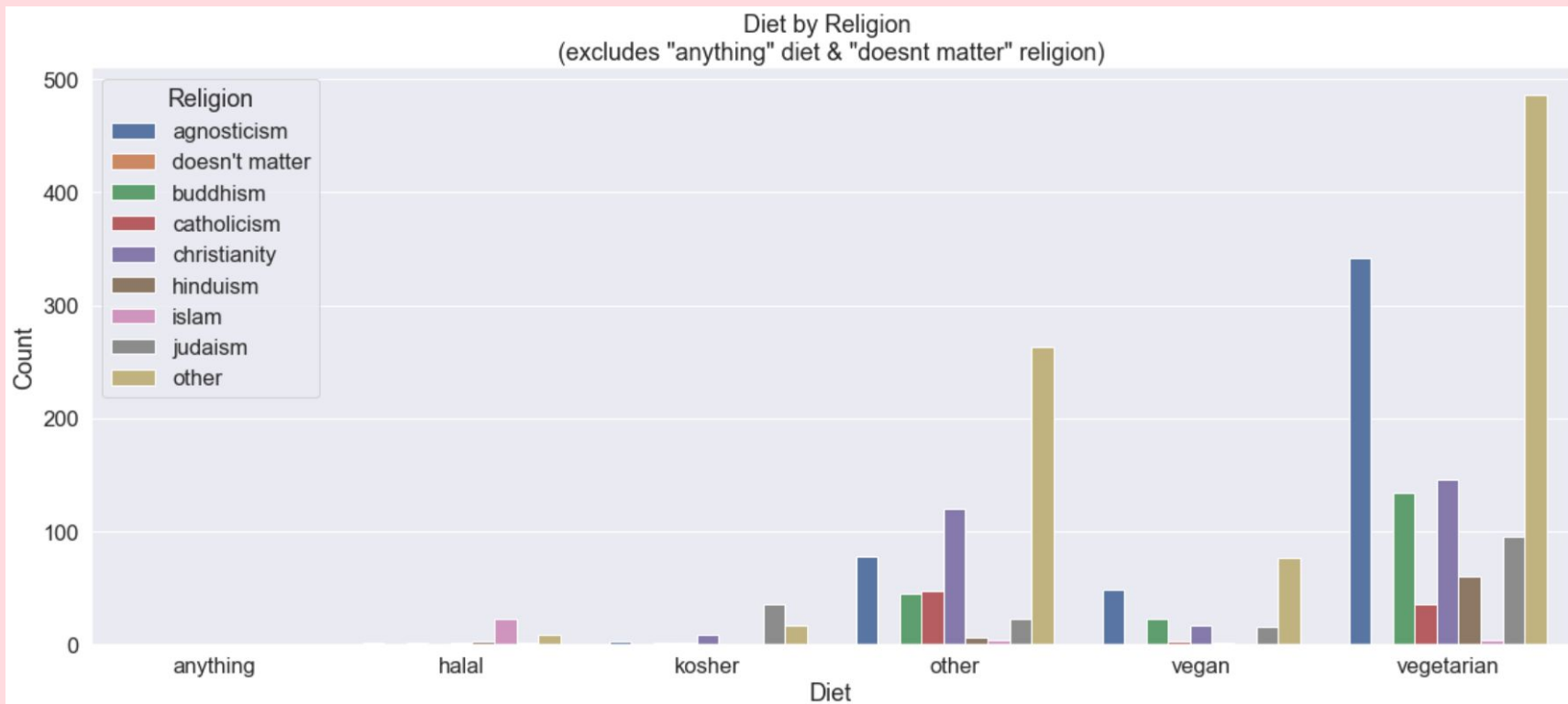
Modeling Approach – Recommender System

Profile liked by User

Similar profiles

Recommend profile to User
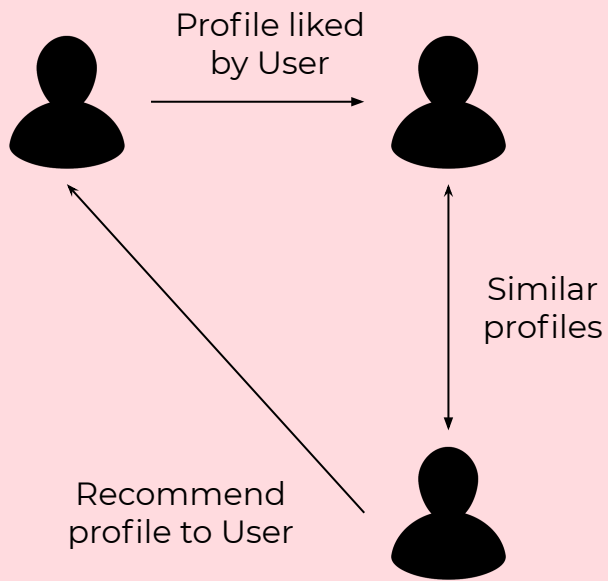
Content-Based Filtering

- Recommend similar profiles that user liked in the past

➔ Metric: **Cosine Similarity**

- Measures the similarity between vectors

- cosine_similarity(X, Y) → 0.00 - 1.00

- ↑ cosine similarity = ↑ vectors similarity

Matching Criteria - Orientation
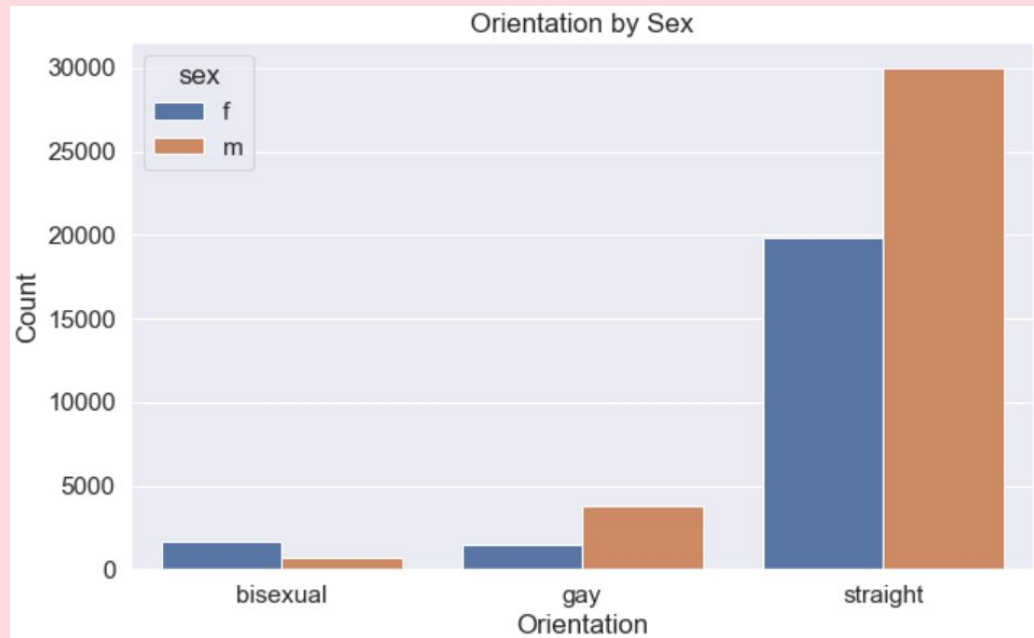
Orientation by Sex

You Are...	Matching Criteria
Straight M	Straight F / Bisexual F
Straight F	Straight M / Bisexual M
Gay M	Gay M / Bisexual M
Gay F	Gay F / Bisexual F
Bisexual M	Bisexual M/F / Gay M / Straight F
Bisexual F	Bisexual M/F / Gay F / Straight M

* Data Source: Kaggle - OkCupid Profiles

Matching Criteria (cont'd) *

- Age range

- Religion
  - Religion matters = Same religion
  - Religion doesn't matter = Atheism / not serious about religion

- Offspring sentiment
  - Want kid(s) → Has kid(s)
  - Don't want kid(s) → Doesn't have kid(s)

Modeling Process

1. User input (X input)
   a. Age Range → Lowest Age: 32 | Highest Age: 45

Age	Sex	Orientation	Body Type	Diet	Drinks	Drugs	Offspring	Pets	Religion	Smokes
30	f	gay	Rather not say	vegetarian	yes	yes	Doesn't have kid(s), but wants kid(s)	Dislikes dogs and cats	christianity	yes

2. Filter table of potential matches based on matching criteria (Y input)
   a. Age Range
   b. Orientation = Gay F + Gay F / Bisexual F
   c. Religion = christianity + christianity
   d. Offspring sentiment = doesn't have kid(s), but wants kid(s) + have kid(s)

user_id	age	sex	orientation	body type	diet	drinks	drugs	offspring	pets	religion	smokes
6254	44	f	gay	full figured	other	sometimes	no	doesn't want kids	dislikes dogs and cats	christianity	no
24990	32	f	bisexual	average	anything	sometimes	no	has kid(s)	dislikes dogs and cats	christianity	sometimes
25664	35	f	bisexual	full figured	anything	yes	no	has kid(s) and wants more	likes dogs	christianity	yes
34449	41	f	bisexual	full figured	anything	sometimes	sometimes	doesn't have kids	likes dogs and cats	christianity	sometimes
42052	33	f	gay	average	anything	sometimes	no	doesn't want kids	likes dogs	christianity	no

Modeling Process (cont'd)

3. Encode X + Y inputs into 0/1

X →

user_id	age_44	age_30	age_32	age_35	sex_f	orientation_gay	orientation_bisexual	body_type_full figured
0	0	1	0	0	1	1	0	0

Y →

user_id	age_44	age_30	age_32	age_35	sex_f	orientation_gay	orientation_bisexual	body_type_full figured
6254	1	0	0	0	1	1	0	1
24990	0	0	1	0	1	0	1	0
25664	0	0	0	1	1	0	1	1
34449	0	0	0	0	1	0	1	1
42052	0	0	0	0	1	1	0	0

4. cosine_similarity(X, Y)
   a. X = user info
   b. Y = lookup table

5. Sort by similarity

Modeling Process (cont'd)

6. Return top matches

User Input: Age Range → Lowest Age: 32 | Highest Age: 45

Age	Sex	Orientation	Body Type	Diet	Drinks	Drugs	Offspring	Pets	Religion	Smokes
30	f	gay	Rather not say	vegetarian	yes	yes	Doesn't have kid(s), but wants kid(s)	Dislikes dogs and cats	christianity	yes

Top Matches:

user_id	age	sex	orientation	body_type	diet	drinks	drugs	offspring	pets	religion	smokes
25664	35	f	bisexual	curvy	mostly anything	often	never	has kids, and might want more	has dogs	christianity	yes
49211	32	f	gay	average	anything	socially	never	has a kid	dislikes dogs and dislikes cats	christianity	no
24990	32	f	bisexual	average	anything	rarely	never	has kids	dislikes dogs and dislikes cats	christianity and somewhat serious about it	sometimes
42459	43	f	bisexual	athletic	mostly anything	socially	never	has kids, but doesn't want more	likes dogs and has cats	christianity and somewhat serious about it	no

Modeling Process (cont'd) – More Matches

7. Indicate liked profile(s) (X input)

user_id	age	sex	orientation	body_type	diet	drinks	drugs	offspring	pets	religion	smokes
49211	32	f	gay	average	anything	socially	never	has a kid	dislikes dogs and dislikes cats	christianity	no
24990	32	f	bisexual	average	anything	rarely	never	has kids	dislikes dogs and dislikes cats	christianity and somewhat serious about it	sometimes

➔ Repeat steps 2 - 6

8. Return additional matches

user_id	age	sex	orientation	body_type	diet	drinks	drugs	offspring	pets	religion	smokes
58513	42	f	gay	a little extra	mostly anything	socially	never	doesn't have kids	dislikes dogs and dislikes cats	christianity	no
42052	33	f	gay	average	anything	socially	never	doesn't want kids	has dogs	christianity	no
52501	34	f	gay	average	anything	not at all	never	doesn't have kids, but wants them	dislikes dogs and dislikes cats	christianity	no
6254	44	f	gay	curvy	strictly other	socially	never	doesn't want kids	dislikes dogs and dislikes cats	christianity and very serious about it	no
47403	44	f	gay	average	anything	rarely	never	doesn't have kids	likes dogs and has cats	christianity and somewhat serious about it	sometimes

15

Model Analyses

➔ User input -- Straight Female
  ◆ Age Range → Lowest Age: 32 | Highest Age: 40

Age	Sex	Orientation	Body Type	Diet	Drinks	Drugs	Offspring	Pets	Religion	Smokes
37	f	straight	full figured	anything	yes	no	doesn't have kid(s), but wants kid(s)	likes dogs and cats	doesn't matter	no

➔ Model returns top matches based on matching criteria

user_id	age	sex	orientation	body type	diet	drinks	drugs	offspring	pets	religion	smokes
7206	34	m	straight	fit	strictly anything	often	never	has a kid	likes dogs and likes cats	catholicism and laughing about it	no
3497	35	m	straight	curvy	mostly anything	not at all	never	has a kid, and might want more	likes dogs and likes cats	atheism	no
15047	39	m	straight	fit	mostly anything	often	never	has a kid, and wants more	likes dogs and likes cats	atheism and somewhat serious about it	no
54828	39	m	straight	average	anything	very often	never	has kids, but doesn't want more	likes dogs and likes cats	atheism	no
56272	38	m	straight	fit	mostly anything	often	never	has kids	likes dogs and has cats	atheism	no

Model Analyses (cont'd)

➜ User liked profile(s) '7206' and '15047'

user_id	age	sex	orientation	body type	diet	drinks	drugs	offspring	pets	religion	smokes
7206	34	m	straight	fit	strictly anything	often	never	has a kid	likes dogs and likes cats	catholicism and laughing about it	no
15047	39	m	straight	fit	mostly anything	often	never	has a kid, and wants more	likes dogs and likes cats	atheism and somewhat serious about it	no

➜ Model returns top matches based on liked profile(s)

user_id	age	sex	orientation	body type	diet	drinks	drugs	offspring	pets	religion	smokes
42530	34	m	straight	athletic	strictly anything	socially	never	has a kid	likes dogs and likes cats	judaism and laughing about it	no
4532	34	m	straight	fit	anything	socially	never	has a kid	likes dogs and likes cats	agnosticism and laughing about it	no
43965	34	m	straight	athletic	mostly anything	rarely	never	has a kid	has dogs and has cats	catholicism but not too serious about it	no
41942	34	m	straight	fit	anything	not at all	never	has a kid	likes dogs and likes cats	catholicism but not too serious about it	no
56272	38	m	straight	fit	mostly anything	often	never	has kids	likes dogs and has cats	atheism	no
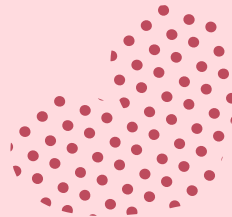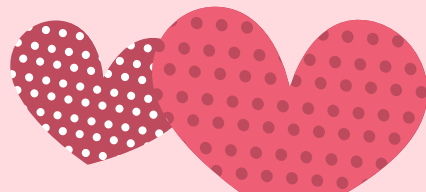
Conclusion & Limitations

- Recommendations might become too similar over time
- Additional recommendations might stray from user's sentiment
- Data was imbalanced
- Values were ambiguous
  - What does a null value mean? (i.e. "pets" or "offspring" columns)

Future Project Refinements

- Add a sort of "rating" for each user profile
- Refine additional recommendations based on liked profiles
- Handle duplicate profiles that would be recommended again
- Incorporate additional relationship statuses
  - i.e. married, seeing someone

Thank You! Questions?

. . . .

VERONICA LEONG | DATA SCIENTIST
LinkedIn | Portfolio | GitHub

Appendix

Reference Links

- [OkCupid Kaggle Dataset](#)
- [My Streamlit Demo - Lover Recommender App](#)
- [Simulated Monster Match Dating Game](#)
  - Layman terms of how recommender systems work in dating apps - https://monstermatch.hiddenswitch.com/algorithms
- [Recommender System Types](#)

Data Cleaning – Null Values

Feature	Value Filled
Body Type	Rather not say
Diet	Anything
Drinks	Not at all
Drugs	Never
Offspring	Doesn't have kids
Pets	Dislikes dogs and cats
Religion	Atheism
Smokes	No

Data Cleaning - Grouping Values

Feature	Original Value	Grouped / Generalized Value
Drinks	Desperately; Often	Yes
	Socially; Rarely	Sometimes
Smokes	Trying to quit; When drinking	Sometimes
Pets	Likes/has dogs; likes/has dogs and dislikes cats	Likes dogs
	Likes/has cats; likes/has cats and dislikes dogs	Likes cats
	Likes/has dogs and likes/has cats	Likes dogs and cats
Body Type	Average; A little extra	Average
	Fit; Athletic; Jacked	Fit
	Thin; Skinny	Thin
	Full figured; Overweight	Full figured

* Jump to original slide

Data Cleaning - Grouping Values (cont'd)

Feature	Original Value	Grouped / Generalized Value
Offspring	Has kids ; Has a kid	Has kid(s)
	Wants kids; Might want kids	Wants kids
	Has a kid, and [might] wants more	Has kid(s) and wants more
	Doesn't have kids, but [might] want them	doesn't have kid(s), but wants kids
	Has a kid/has kids, but doesn't want more	has kid(s), but doesn't want more
Religion	Atheism; [Religion] but not too serious about it; [Religion] and laughing about it	Doesn't matter

* Jump to <u>original slide</u>

Matching Criteria - Orientation

You Are...	Matching Criteria
Straight M	Straight F / Bisexual F
Straight F	Straight M / Bisexual M
Gay M	Gay M / Bisexual M
Gay F	Gay F / Bisexual F
Bisexual M	Bisexual M/F / Gay M / Straight F
Bisexual F	Bisexual M/F / Gay F / Straight M

Matching Criteria (cont'd)

Feature	You Are...	Matching Criteria
Age Range	---	Lowest / Highest Age
Offspring	doesn't have kid(s), but wants kid(s)	has kid(s) / has kid(s) and wants more / has kid(s), but doesn't want more
	doesn't have kids, and doesn't want any	doesn't have kids / doesn't want kids / doesn't have kids, and doesn't want any
	has kid(s), but doesn't want more	doesn't have kids / doesn't have kid(s), but wants kid(s) / wants kid(s)
	has kid(s) and wants more	has kid(s) / wants kid(s) / has kid(s) and wants more
Religion	[religion]	[religion] and somewhat / very serious about it
	Religion doesn't matter	[religion], but not too serious about it / and laughing about it