



Predicting Speed Dating Matches

ARYAN, KARTHIK,
MIKALA, SANKEERTH

Agenda

1. Dataset Walkthrough
 2. EDA
 3. Feature Engineering
 4. Model Outputs
 5. Results & Next Steps
-

Speed dating experiment

Four-minute
"first date"

Only dates with
the opposite
gender

Pre-date
questionnaire

Asked if they
would see their
date again

Dataset predictors

General predictors

Age
Difference in age
Race
Same race
Interests correlation

Questionnaire: Rate Yourself (1-10)

Attractiveness - rate yourself
Sincerity - rate yourself
Intelligence - rate yourself
Funny - rate yourself
Ambition - rate yourself

Questionnaire: Importance Ratings

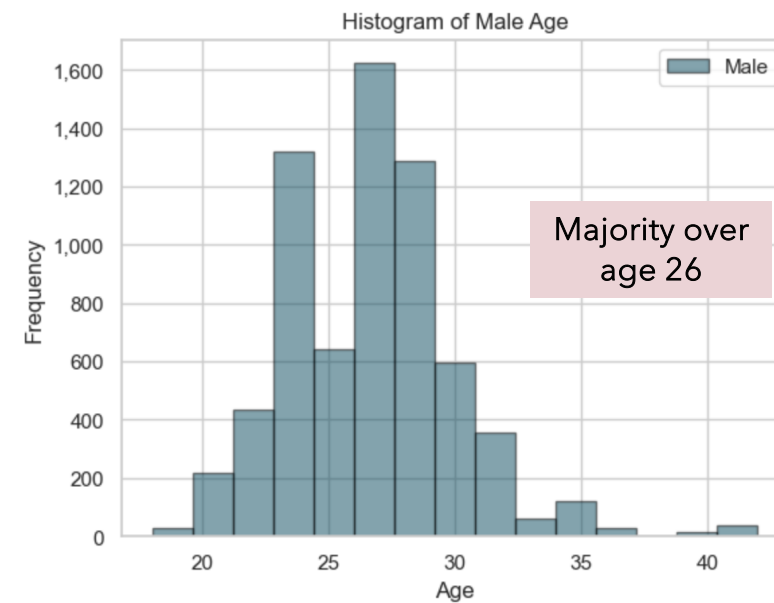
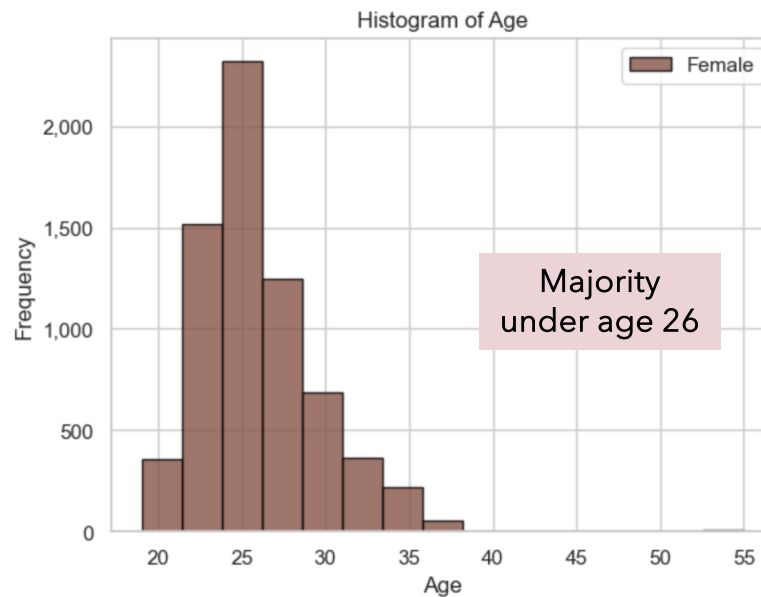
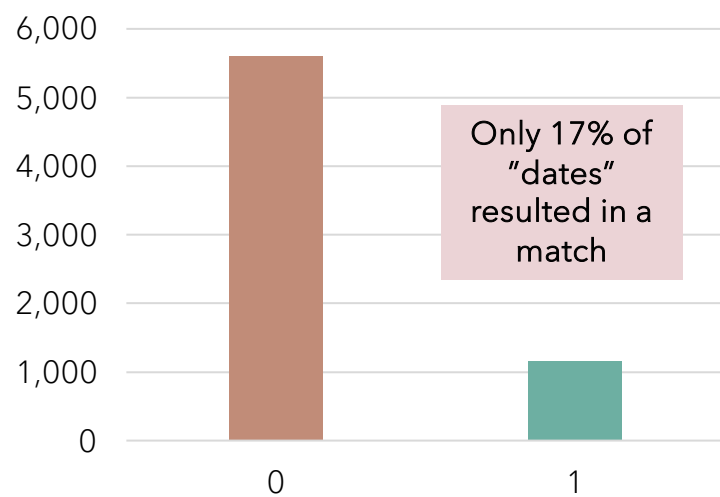
Attractive importance
Sincere importance
Intelligence importance
Funny importance
Ambitious importance
Same race importance
Same religion importance
Shared interests importance

Feature engineered variables

Similar values bucket
Importance deltas
Values deltas

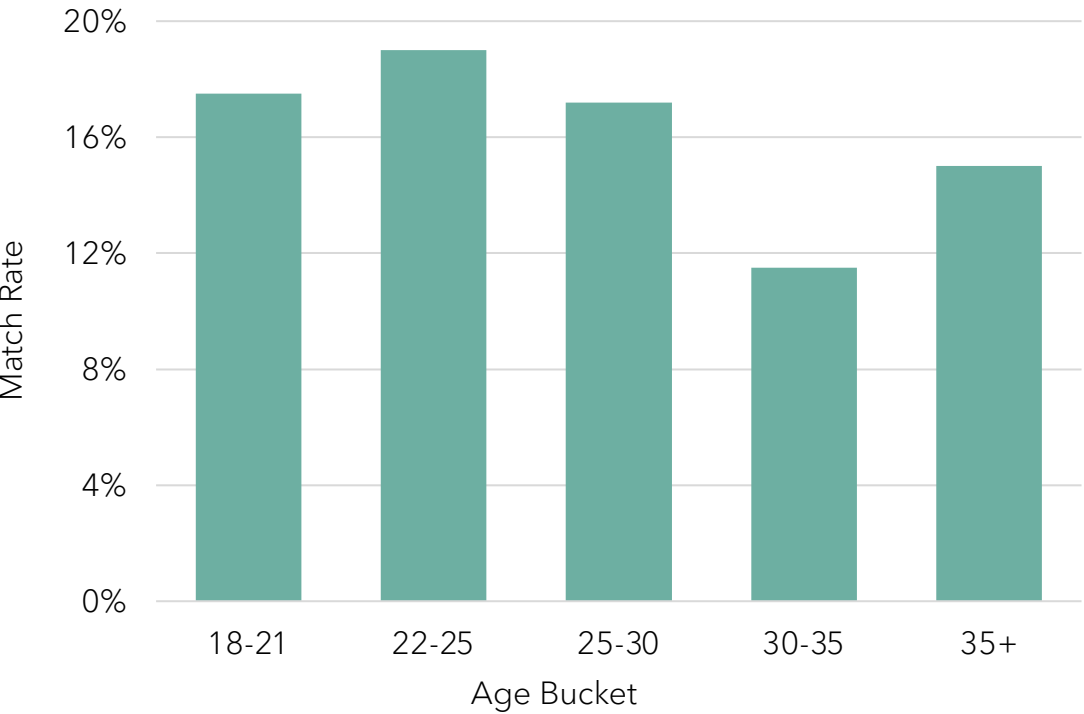
EDA

Target Variable
Distribution

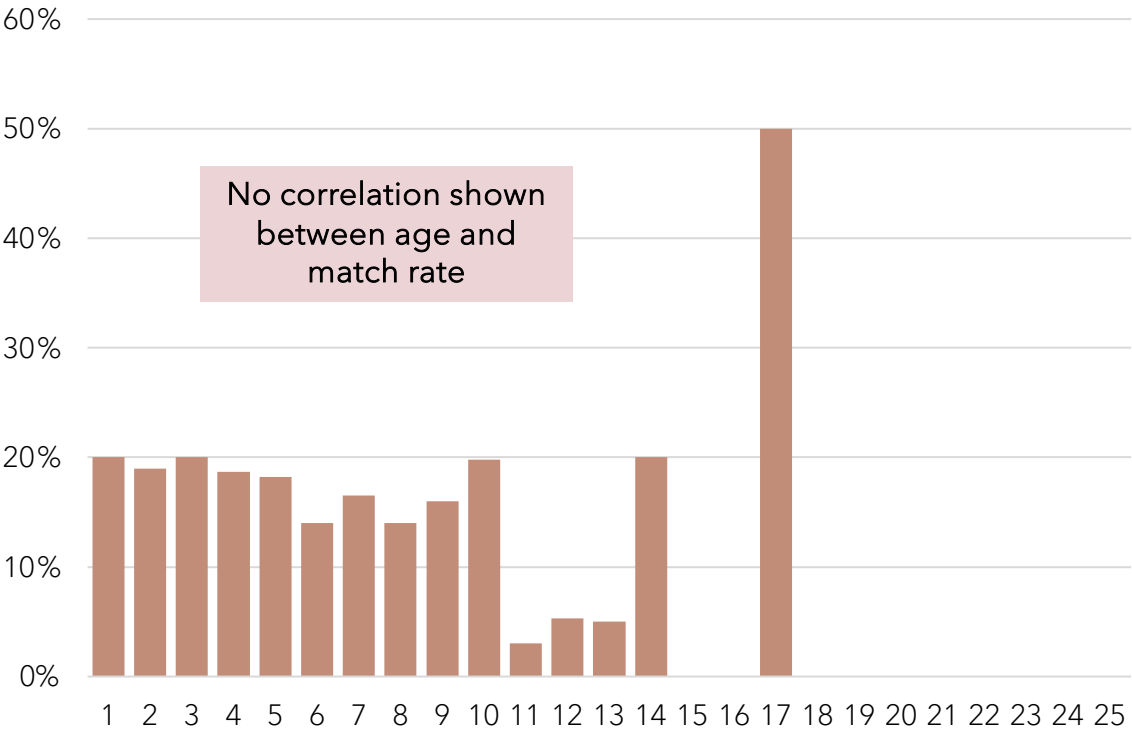


EDA – Age

Age Bucket & Match Rate

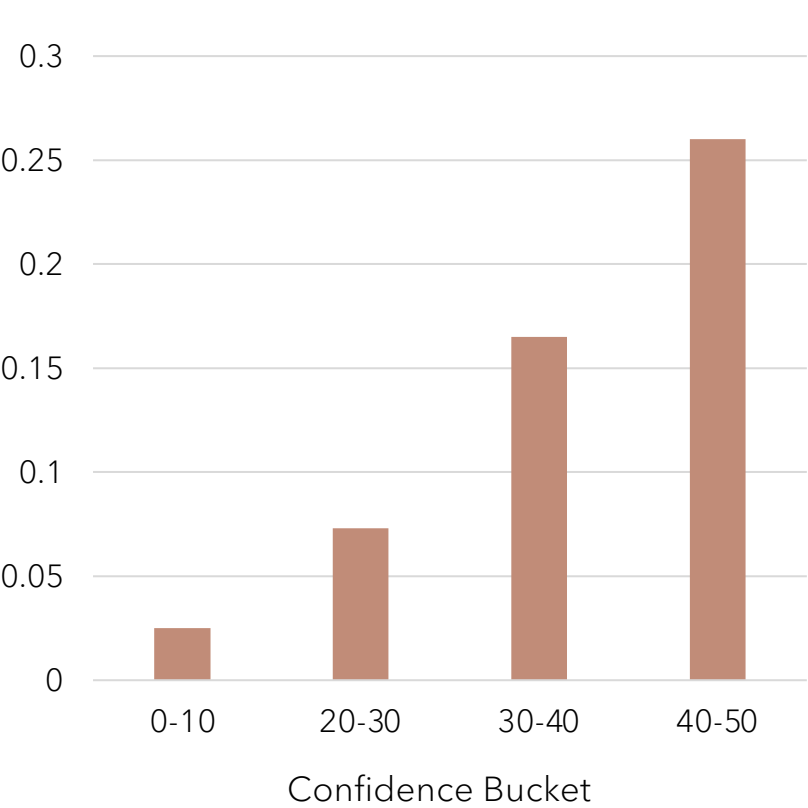


Age Difference & Match Rate

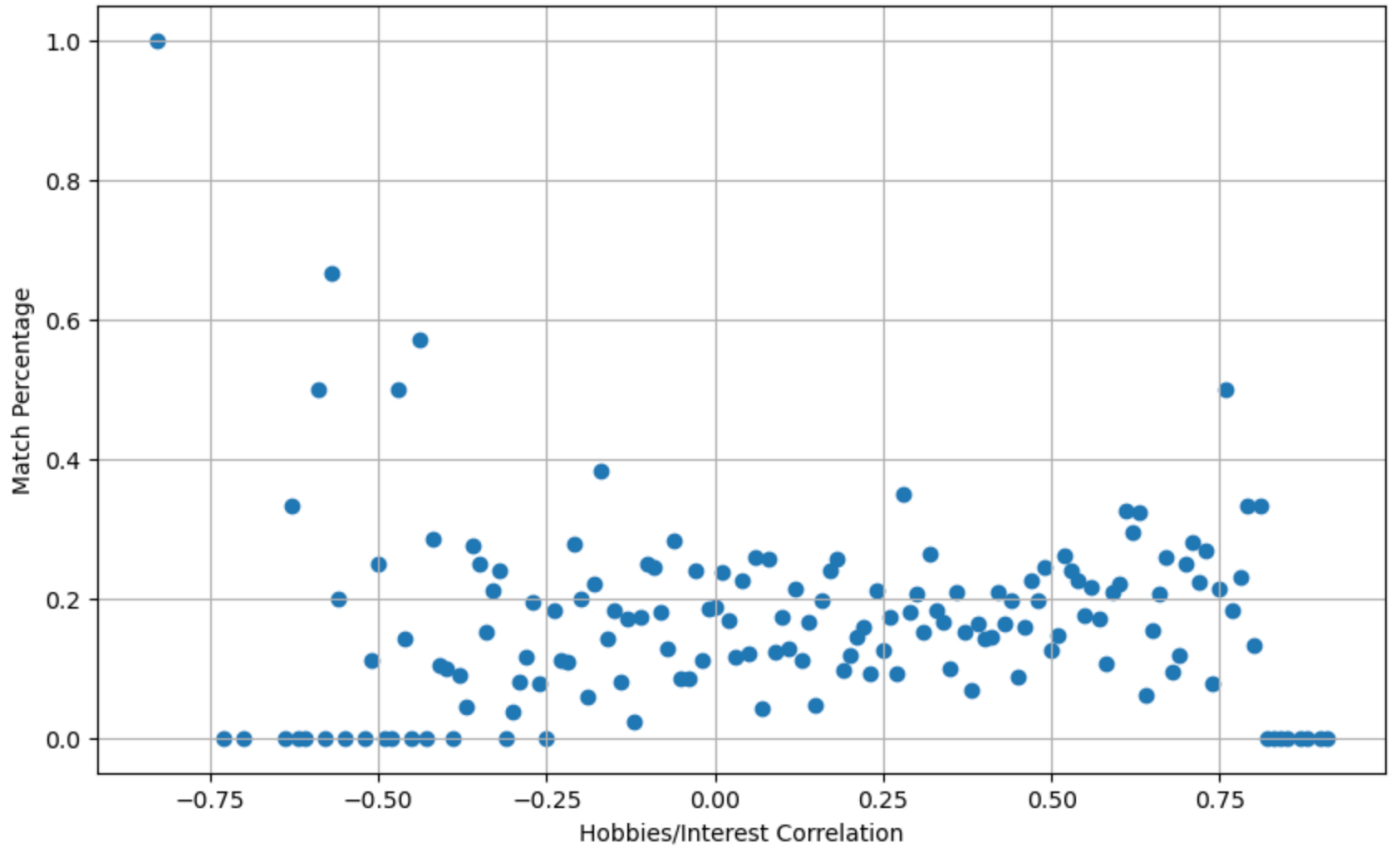


EDA – Self Confidence | Interest Correlation

Match Rates Across Buckets

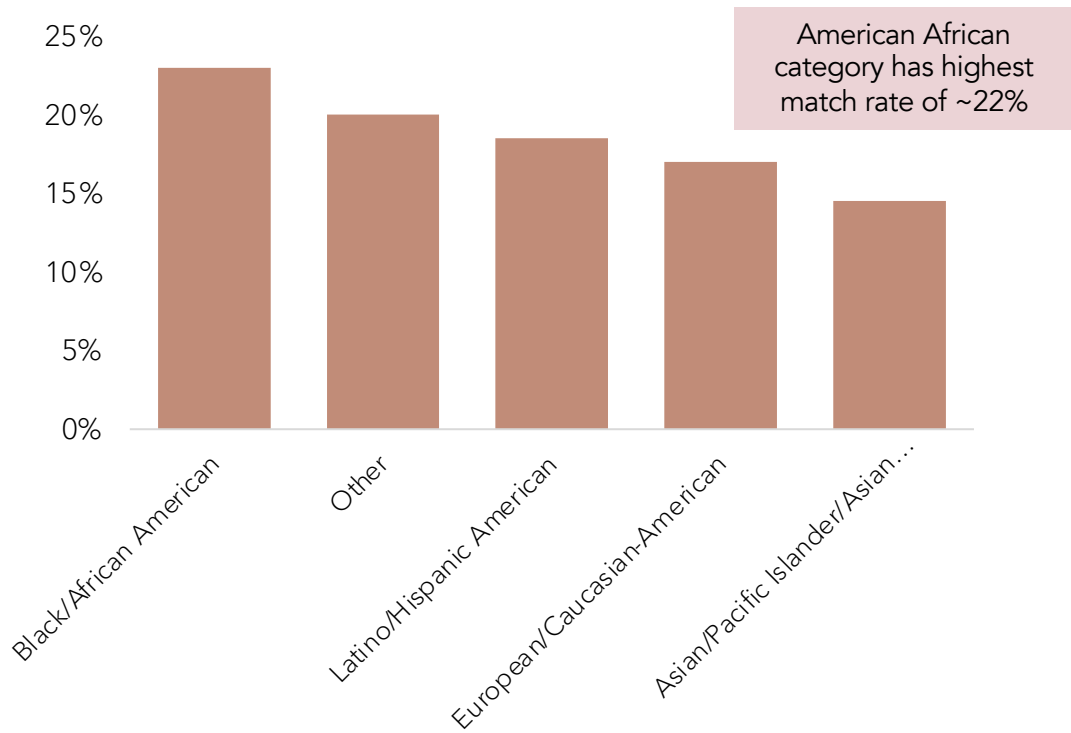


Interests Correlation & Match Percentage

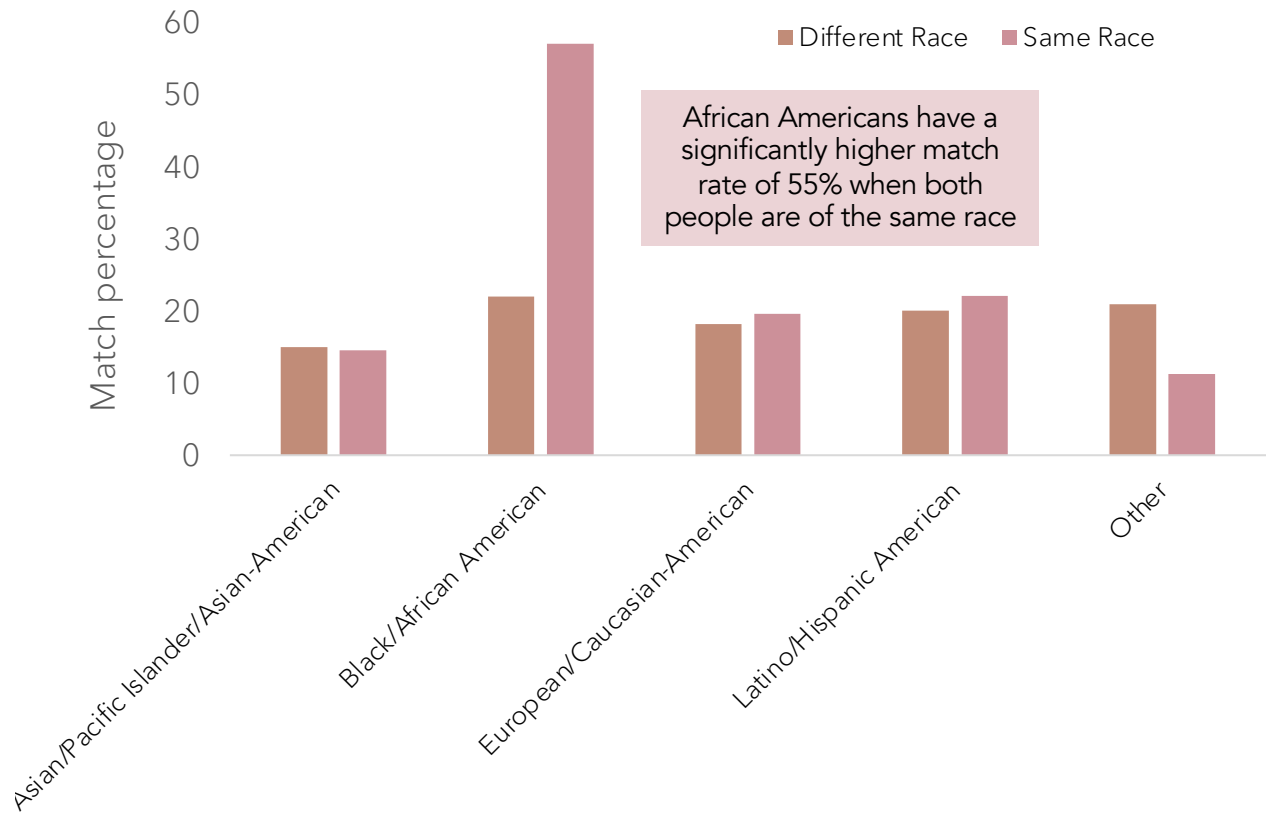


EDA - Race

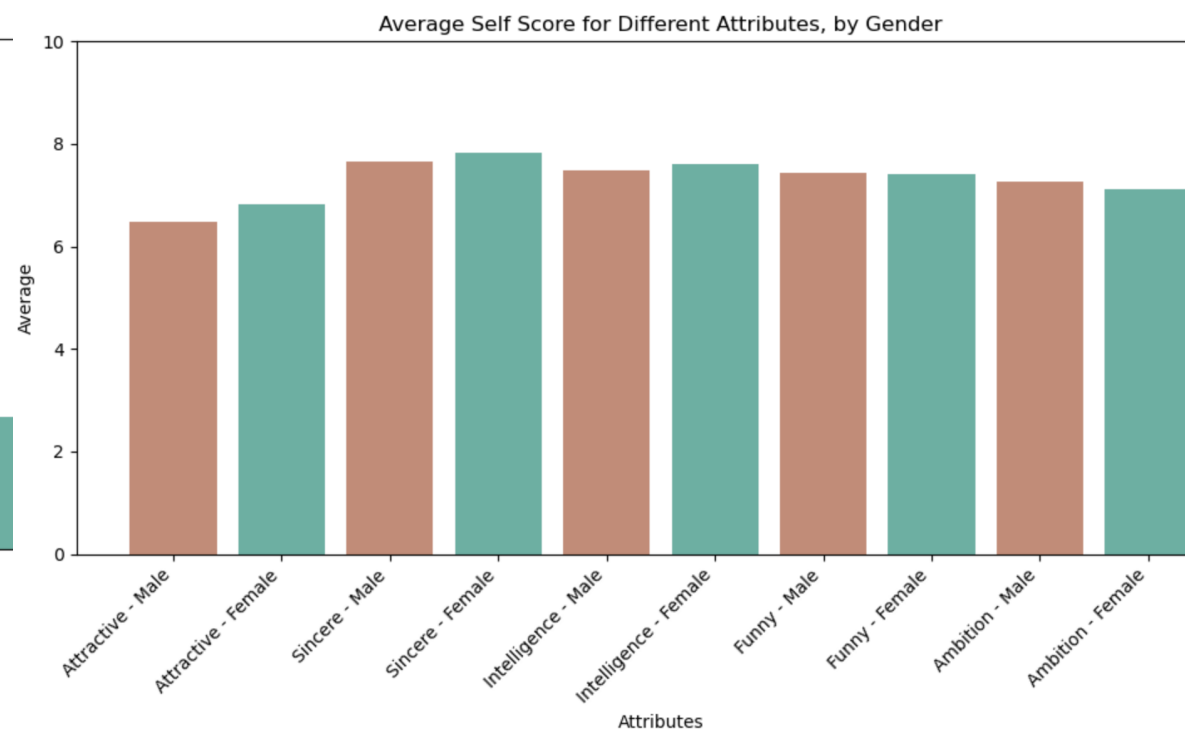
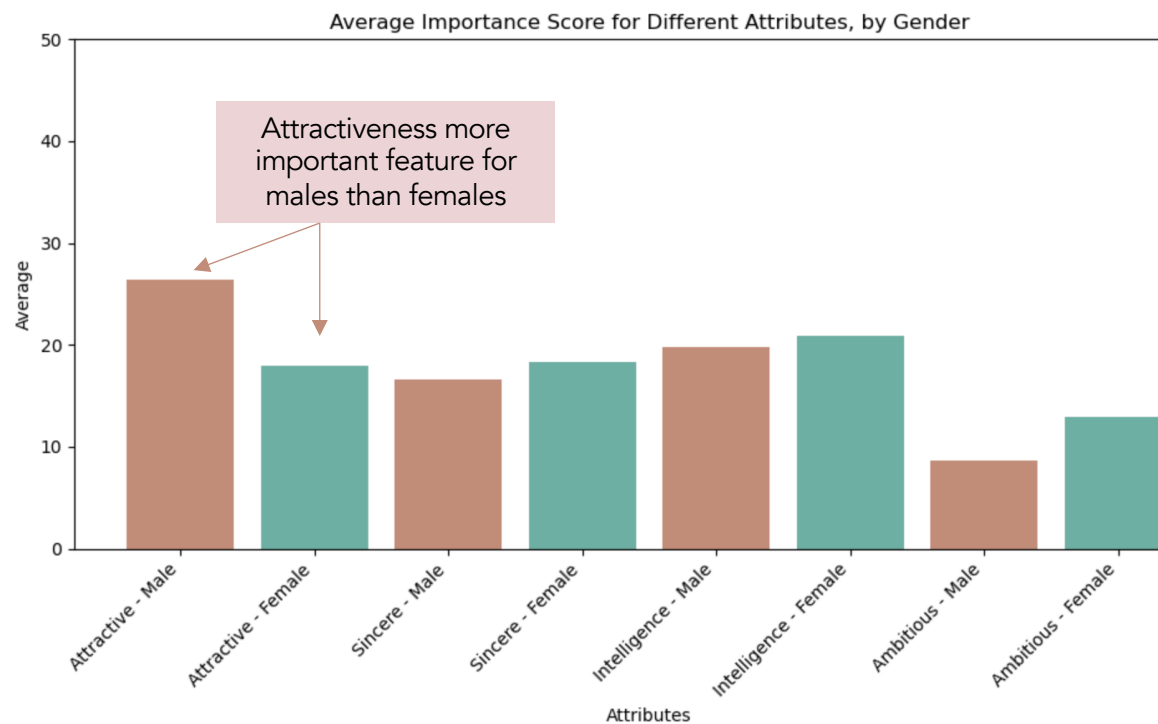
Match Rate Across Races



Match Percentage by Race

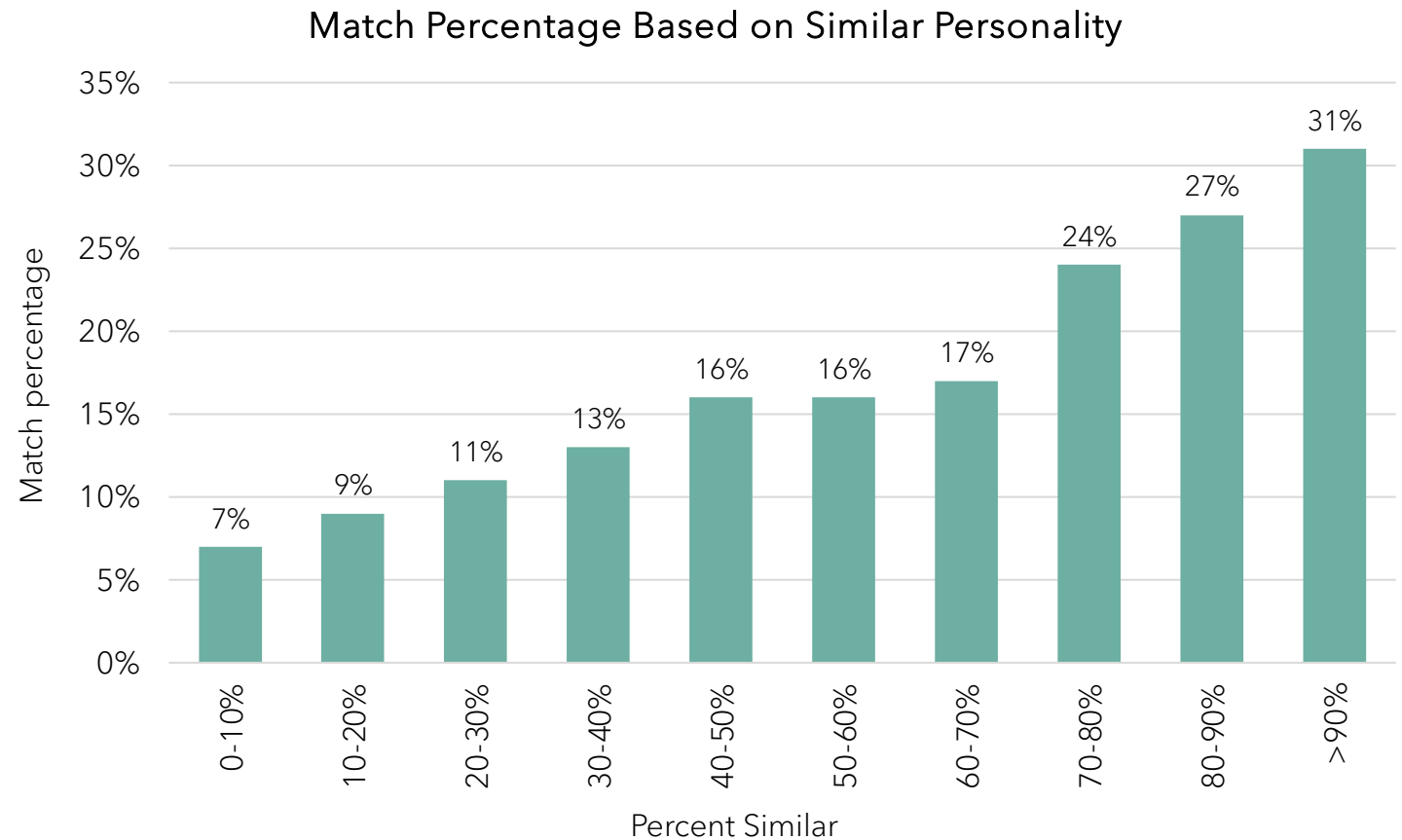


EDA - Attribute scores



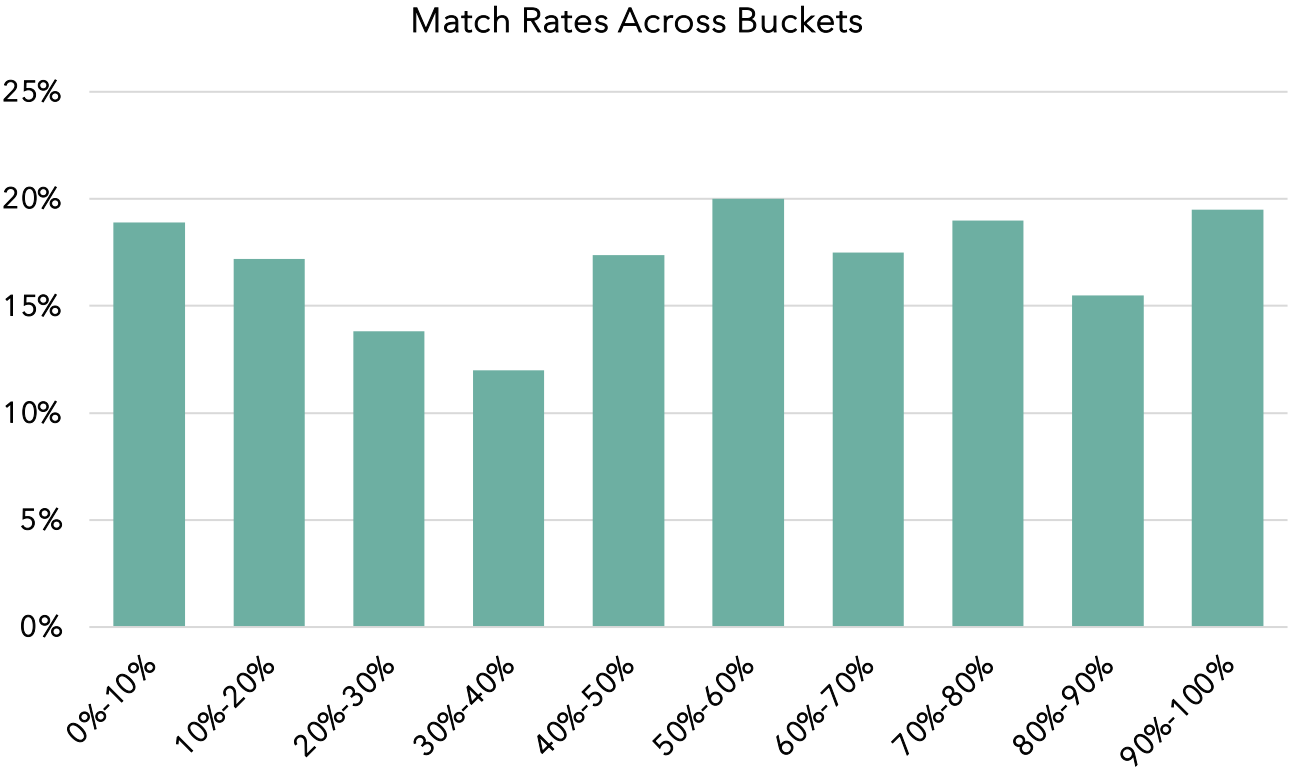
Feature Engineering – attribute similarity

5 Personality Traits (self-rated)
Attractiveness
Sincerity
Intelligence
Funny
Ambition



Feature Engineering – trait importance

Importance Ratings
Attractive importance
Sincere importance
Intelligence importance
Funny importance
Ambitious importance
Same race importance
Same religion importance
Shared interests importance

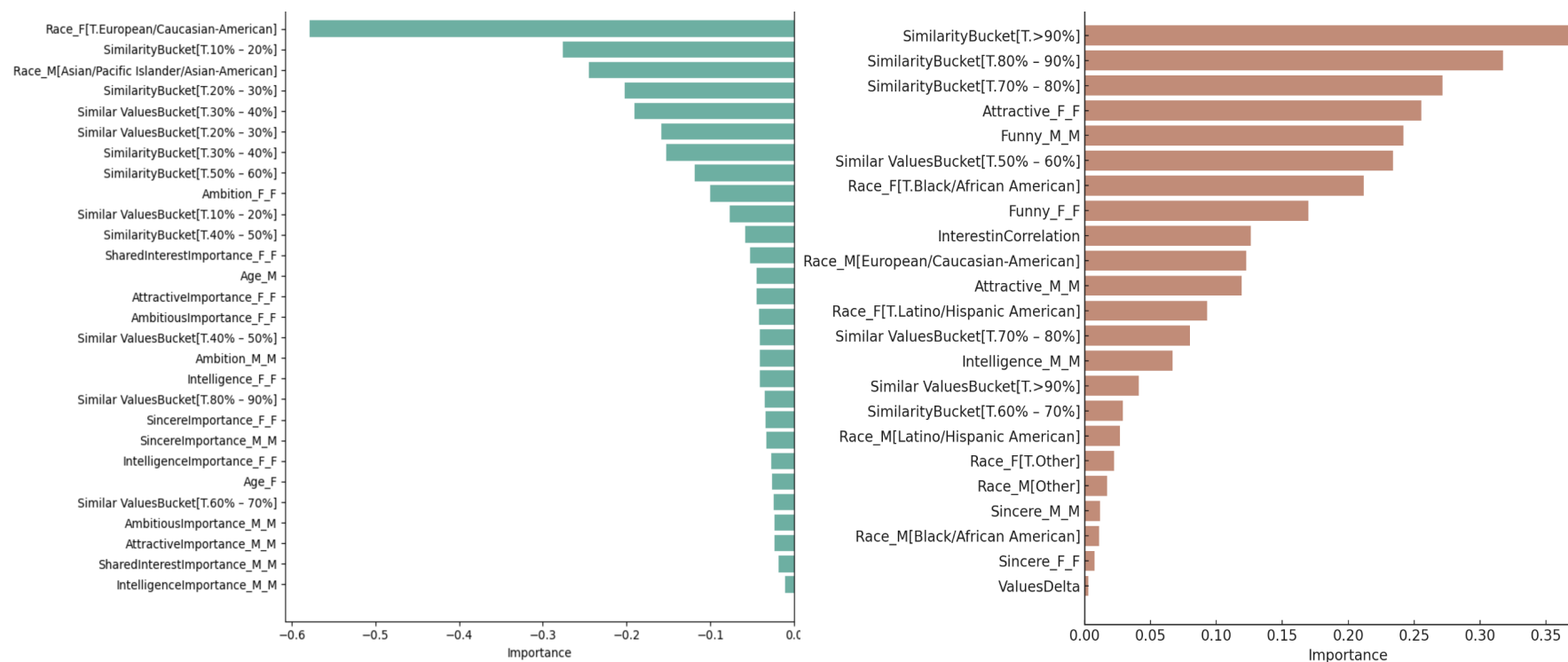


Logistic Regression

Train Accuracy	Test Accuracy
83%	82%

- Similarity Buckets have high weights
- Self ratings (confidence) of Funny, Attractiveness also has relatively high weights
- Not a great accuracy lift over baseline (82%)

Variable Importance



Decision Tree

Training Set

Depth	Training Accuracy	Cross Validated Accuracy
2	82.8%	82.8%
4	83%	82.9%
7	84.3%	81%
10	87%	79%
25	99.9%	76.3%

Test Set

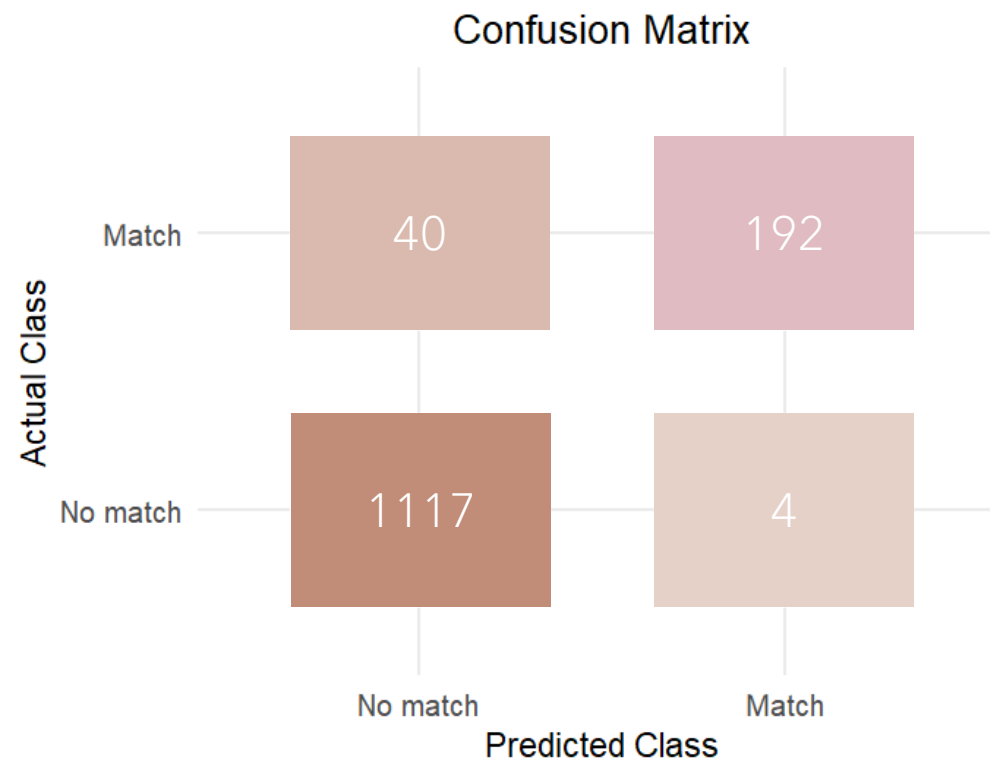
Depth	Test Accuracy
4	81%

Important Variables

- Humor for both males and females
 - Attractiveness for females
 - Importance of Ambition, Sincerity
-

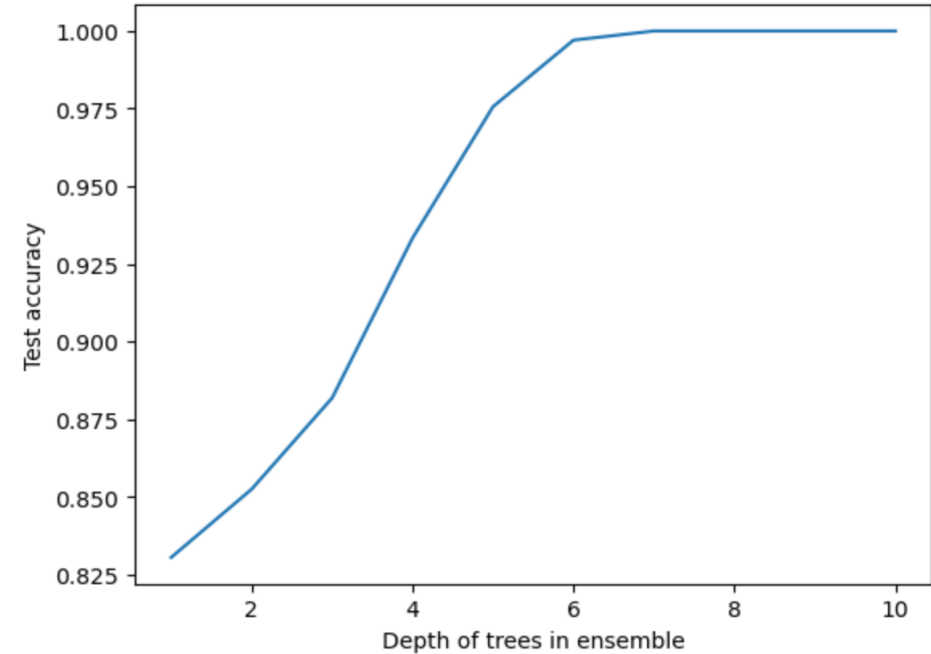
Random Forest

Train Accuracy	Test Accuracy
97.1%	96.75%

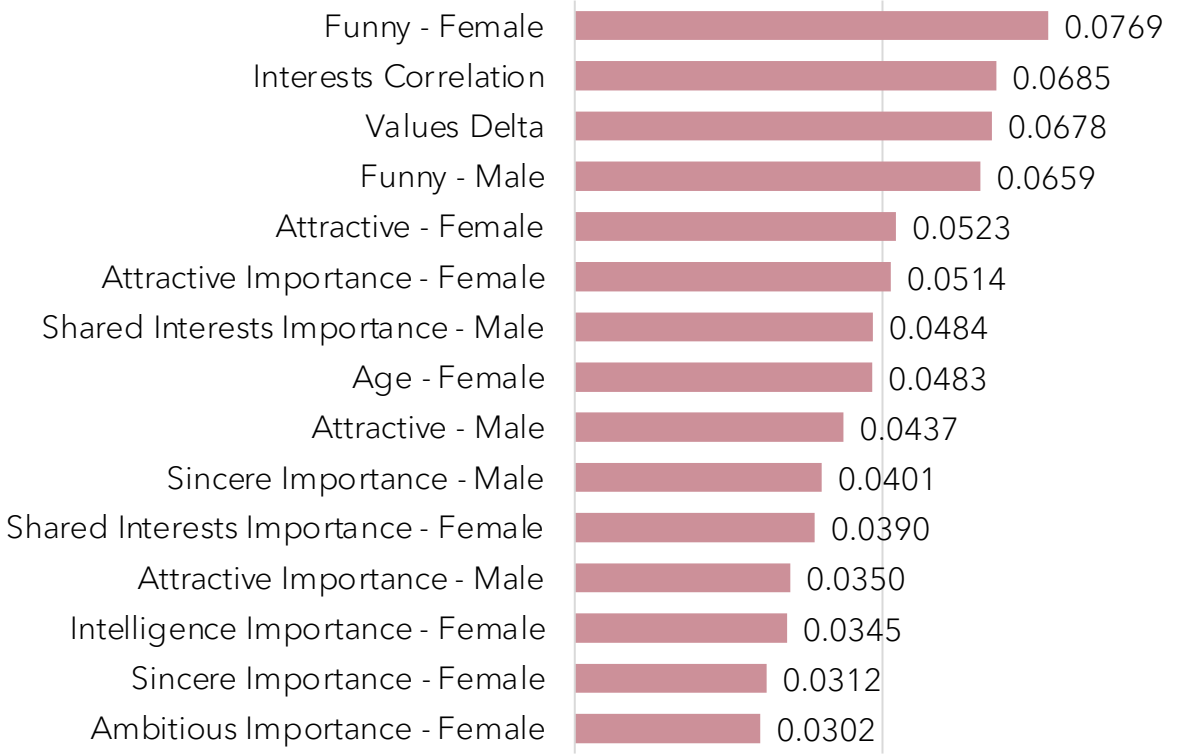


Boosting

Train Accuracy	Test Accuracy
93%	93%



Feature Importance



Results & next steps

Model	Train Accuracy	Test Accuracy
Logistic Regression	83%	82%
Decision Tree	83%	81%
Random Forest	97%	97%
Boosting	93%	93%

Next Steps

- Additional Feature Engineering
 - Weighted Personality Scores
- Increase number of Hyperparameters
- Add data about prior dating history

Thank you
