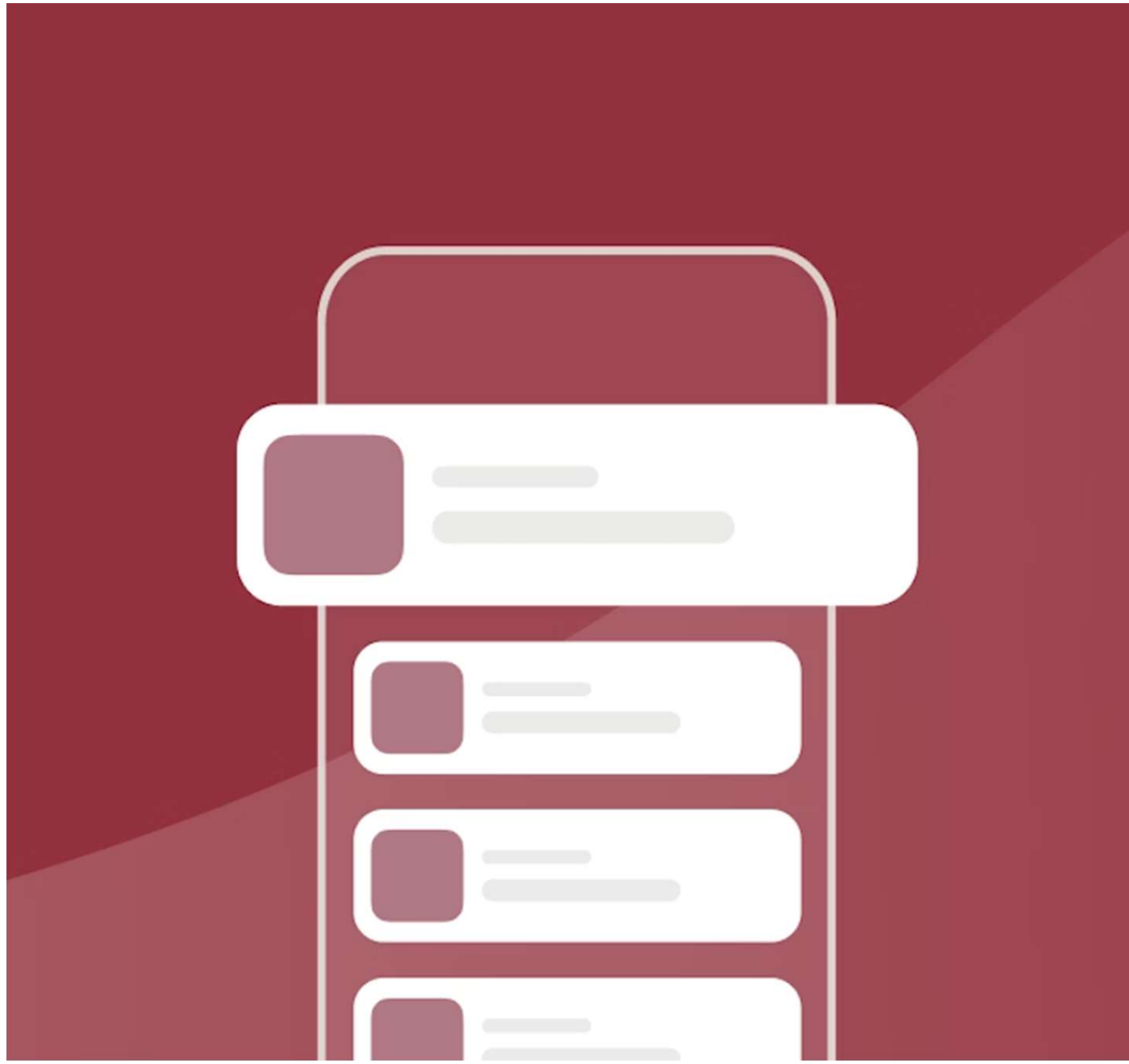
Pizzaz.com Analysis

Speed Dating Analytics & Online Dating Optimization

Natalie LePera | 24 Apr 2025



Introduction

What Makes Pizzaz.com Unique?

Real-World Data for Real-World Matches

Real-world data from in person speed dating events is leveraged to enhance user experience and improve match compatibility.

Predictive Matching Systems

Predictive matching is utilized to offer users high-quality potential matches based on gender specific identified preferences.

Overall Goals

- Gender-specific prediction model creation allowing for improved personalization in potential matches
- Real world data driven preference rankings add a pseudo peer-review of a dater’s perceived characteristics
- User customization options allow for tailored match results based on the user’s preferences

User Customization Features

Dater Characteristics Rankings

Users can first record their rankings through an initial review period. Their rankings of sample profiles are recorded to contribute to the overall characteristic data for the user pool.

- Attractiveness
- Fun Score
- Sincerity
- Shared Interests
- Intelligence
- Ambitiousness

Like Score Cutoff

Users can select a like score cutoff to filter potential matches, enhancing their experience.

Speed Dating Data Overview

Data Collection and Characteristics

Participant Demographics

- 276 heterosexual couples, randomly matched for a short speed date
- General demographics collected: Age & Race

Characteristic Ratings

- Participants rated their matched partners on key characteristics
  - Attractiveness, sincerity, intelligence, fun, ambitiousness, & shared interests
- Rankings captured in a 1 – 10 scale (1 = lowest / 10 = highest)

Second Date Rankings

- Participant interest in a second date (1 = Yes 2nd date / 0 = No 2nd date)
- Assumed partner's interest in a second date (1 – 10 scale; 1 = lowest likelihood / 10 = highest likelihood)
- Data export errors omitted these rankings from utilization
  - Recommendations for future use included in further sections

Missing Values and Summary Statistics

Variable	N Miss	N
LikeM	2	274
LikeF	4	272
AgeM	3	273
AgeF	5	271
AttractiveM	3	273
AttractiveF	2	274
SincereM	5	271
SincereF	3	273
IntelligentM	8	268
IntelligentF	3	273
FunM	6	270
FunF	6	270
AmbitiousM	17	259
AmbitiousF	10	266
SharedInterestsM	27	249
SharedInterestsF	30	246
smrace	0	276
closea	0	276

Table 1: Count of missing vs populated values in speed dating data

N Miss = number of missing observations
N = number of complete observations

Identification of Missing Values

Analysis revealed numerous missing values in the dataset, captured in Table 1

- Implications and recommendations discussed in later sections

Dater Demographic Similarities

- 56.16% of pairs were identified to be of the same race
- 43.84% of pairs were considered close in age.
  - (Close age: ages within 2 years of one another)

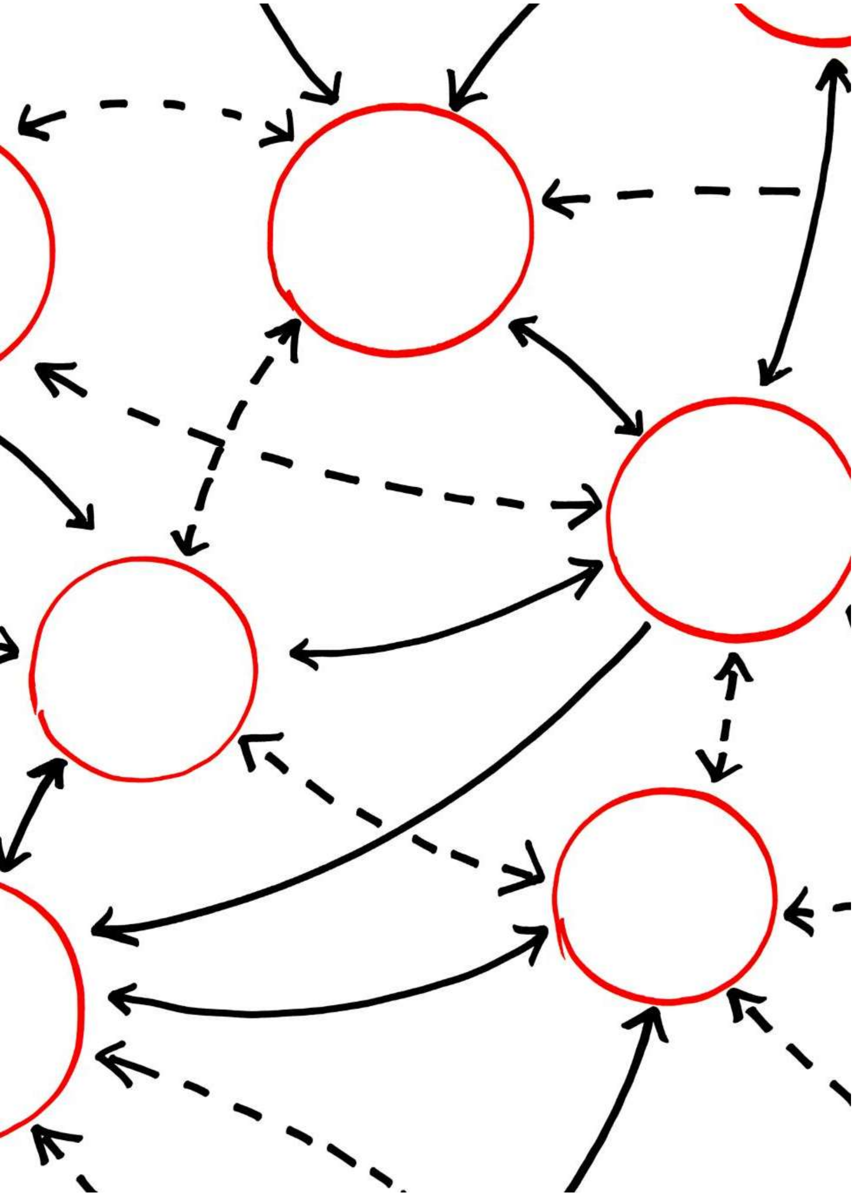
Correlation Findings

Correlation was calculated to determine potential impact of dater demographic similarities on reported 'like' scores.

- Less than 0.10 correlation between race match & reported 'like' scores
- Less than 0.02 correlation between age gap & reported 'like' scores

These low correlation values indicate dater age gaps and race did not significantly affect reported 'like' scores.

Model Creation and Analysis



Data Splitting and Cross Validation

Data Splitting Methodology

- Initial split: Male dater responses & Female dater responses
- Second split: 80:20 split into training and testing data
  - 221 Training observations
  - 55 Test observations

NOTE: Totals do not account for observations dropped due to missing values

Cross Validation Process

- Prediction models developed on training data sets
- Models tested on test data sets
- Ensures model is not over-fitted to data
  - Allows for adequate predictions after introduction of new data

Prediction Model Development Process

Regression Assumptions $\boxed{\hat{E}\ \widetilde{iid}\ N(\mu, \sigma^2)}$**:**
- Residuals Normally Distributed
- Residual Mean = 0
- Residual Variance is Constant
- All Predictor Variables Independent

MALE MODEL

- Pre-model creation checks
  - Attractiveness indicated as potentially present in 2nd order
  - Pairwise interactions included
- Initial model
  - Violated regression assumption of normal residual distribution & homoscedasticity (standard variance)
- Final Model
  - Like score squared to ensure standard variance & improve normal distribution
  - Stepwise forward selection utilized
    - Ensures variables dropped if significance changes as model grows
  - Calculated shrinkage (0.0623) indicates reliable model
  - No collinearity identified

Like_M^2 ~ Fun + (Attractiveness * Sincerity) + (Attractiveness * Shared Interests) + Intercept

FEMALE MODEL

- Pre-model creation checks
  - Only single order terms identified
  - Pairwise interactions included
- Initial Model
  - Violated regression assumption of normal residual distribution & homoscedasticity (standard variance)
- Final Model
  - Like score squared to ensure standard variance & improve normal distribution
  - Stepwise forward selection utilized
    - Ensures variables dropped if significance changes as model grows
  - Calculated shrinkage (-0.0659) indicates reliable model
  - No collinearity identified

Like_F^2 ~ (Attractiveness * Fun) + (Sincerity * Intelligence) (Intelligence * Shared Interests) + Intercept

Men vs. Women

The age-old comparison

Purpose of splitting data by participant gender:

- Identified clear distribution differences in all characteristic ranking values
- Allows for unique matching approaches
- Acknowledges preferences are not one-size fits all

Male preferences:

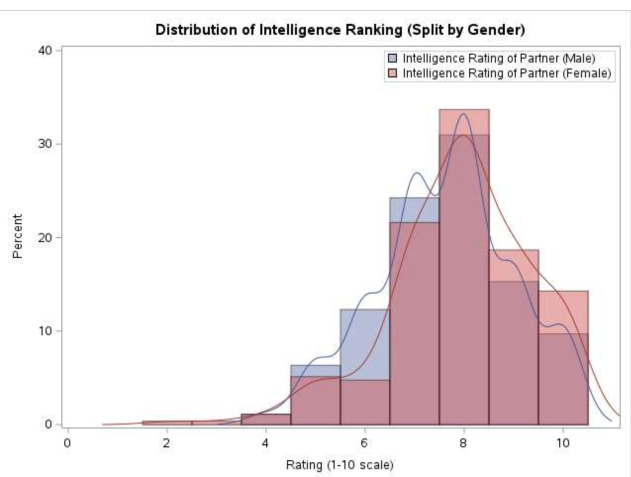
- **Fun, Attractiveness** (twice!), **Sincerity, Shared Interests**

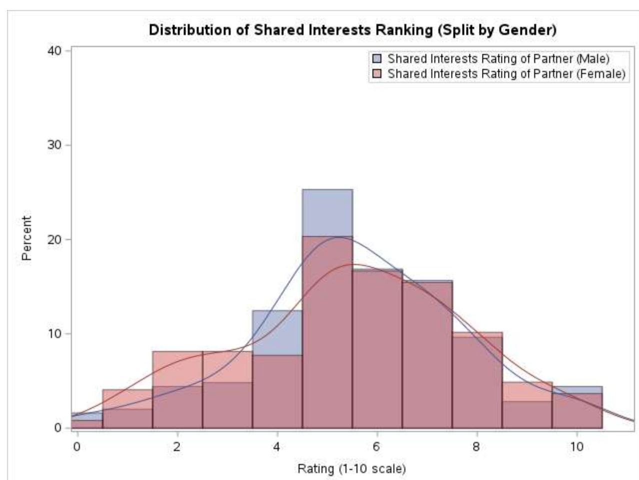
Female Preferences:

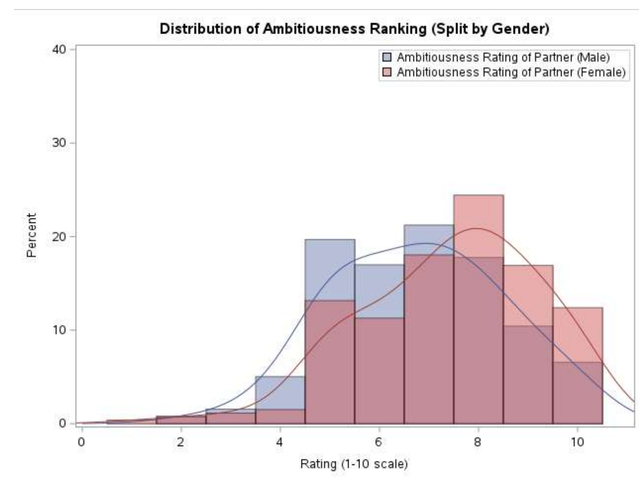
- **Attractiveness, Fun, Sincerity**, Intelligence (twice!), **Shared Interests**

Takeaway:

- Both genders value: fun, attractiveness, sincerity, & shared interests
- Men value attractiveness more than women (present in 2 pairwise interactions)
- Women value intelligence more than men (present in 2 pairwise interactions & missing from male model)







Summary of Findings

Findings and Recommendations

Gender-Specific Preferences

- Males value attractiveness at a greater rate than females
- Females value intelligence while males do not

Data Integrity Recommendations

Blank values can skew prediction models resulting in inaccurate predictions & incompatible dater pairings.

Options to fix:

- Remove options for participants to leave responses blank
- Develop imputation procedures to predict observations for blank values

Enhancing User Experience

Enhance user experience using the following approaches:

- Aggregate characteristic rankings across multiple evaluations for pseudo peer-review
- Allow for minimum predicted like ranking cutoffs for matches
- Re-evaluate age groupings to determine if dater age impacts desired characteristics