RESEARCH

Playing Cupid: Improving Online Dating Matchmaking through Machine Learning Models

Jason Catacutan*, Jason Catacutan, Paul Nepomuceno, Anish Pati, Joshua San Juan and Ana Vasquez

*Correspondence:
jcatacutan.msds2024@aim.edu
Full list of author information is available at the end of the article
†Equal contributor

Abstract

This study delves into enhancing the matchmaking process in online dating platforms through machine learning models, with a focus on predicting successful first dates after matching. Recognizing the prevalent user frustration associated with current online dating experiences, this research aims to address these challenges by developing a predictive model that leverages both user information and partner feedback. Employing Shapley values for feature importance assessment and DICE for counterfactual analysis, the study identifies key factors such as mutual 'like' ratings, perceived attractiveness, and shared interests, as significant predictors of match success. The findings suggest that mutual perception plays a more critical role in match formation than objective similarities. Furthermore, the study proposes integrating the model into existing dating platforms to guide users towards more meaningful interactions, thereby enhancing user satisfaction and engagement. This research not only contributes to the understanding of online dating dynamics but also offers practical implications for improving matchmaking algorithms and user experience on dating platforms.

Keywords: dating; online dating; matchmaking; love; relationships

Highlights

1. Shapley values show perceived similarity in hobbies outweighs actual similarity.
2. Self-confidence matters, but partner's perception is more crucial for a match.
3. DICE indicates that matching hobbies best turns a 'not match' into a match.

1 Introduction

Online technologies have significantly transformed modern matchmaking, with apps like Tinder, Bumble, and Hinge steadily increasing their user base. However, this rise in popularity has also led to a surge in users experiencing frustration with the overall online dating experience.

A 2022 survey found that four out of five people aged 18-54 experience some degree of emotional fatigue or burnout due to the use of online dating apps ("Emotional Fatigue and Burnout in Online Dating," 2022). This indicates that the increased usage of these apps does not always lead to satisfying outcomes and can often be a tiresome endeavor.

Furthermore, those who do match and go on to meet offline have reported low satisfaction scores with their experiences (Vera Cruz et al., 2023). This suggests a common trend of dissatisfaction or disappointment within even when a match is made and an offline meetup is secured. Despite these challenges, the platforms continue to attract new users (Statista, 2023).

The study aimed to explore methods to alleviate the frustration experienced by online dating users. By examining factors that guarantee a successful first date and minimize the frustration of spending time and money to meet someone offline. The goal is to minimize the time wasted and improve the overall experience.

1.1 Problem Statement

Users are now overwhelmed with the sheer number of potential matches, making the decision making process increasingly difficult. Consequently, current algorithms employed in these platforms are unable to facilitate the transition from online chat to first in-person date for users, likely due to the inability of these systems to fully account for user preferences and match-specific dynamics.

This results in a vicious cycle of disappointment for most users brought by the missed opportunities in potentially compatible partners or the time wasted in interactions that did not yield meaningful relationships. These circumstances highlight the need for a more nuanced approach, as addressing said gaps can significantly enhance user experience and improve online dating's overall product offering.

This raises the question of how to resolve the underlying need for a mechanism that not only assesses the compatibility of profiles but also peers into user-to-user interaction to predict the likelihood of a successful in-person date.

1.2 Objectives

The primary objective of this study is to engineer a machine learning model that transcends the current boundaries of conventional dating applications. The created model aims to draw from two primary sources, user information, and self and partner feedback, to forecast the likelihood of a successful first meeting.

Furthermore, the study intends to yield robust methodologies that offer actionable insights to guide users into converting more matches into meaningful interactions. This demands the incorporation of interpretability techniques in the machine learning process.

Lastly, the project intends to offer potential implementation of the model to existing platforms, taking into consideration the current online dating pipeline. This aims to bridge the gap from theory to practice, and bring the study's findings directly into the hands of users.

2 Related Works

Several studies have tried to explore the topic of dating, relationships and online dating. One such study conducted by Joel et. al made use of Machine Learning models to identify if attraction could be predicted prior to interaction in speed dating(2017). They were unable to predict if people were a match or not given their self-reported traits, and this suggested compatibility was difficult to predict without the two parties meeting first.

While similar this is where the current study is distinct. The current study has a clear predictor due to the presence of a feature indicating both parties desire to go on a second date. Another benefit was that the current study took the exploration of traits further as the current study compliments the findings of Joel and colleagues by exploring the scenario where the potential couple meets and also takes into account the perception of the partner and not just the self-reported traits.

3 Data and Methods

3.1 Dataset

The original dataset, sourced from Kaggle, consisted of 123 features and 8,378 rows. These features described the demographics of the pair (eg. age, race, field of study/work), their expectations from the speed dating event (eg. expected number of matches), and their interests or hobbies (eg. sports, museum, art). A preview can be found in Table 1.

It also contained the pair's rating for themselves and their partners on each of the listed traits (attractiveness, being funny, intelligence, sincerity, shared interests, and ambition), their preference and level of importance for each trait, and derived features from the trait ratings.

3.2 Exploratory Data Analysis

Match

Out of the 8,378 pairs of participants analyzed, 1,380 pairs (16.47%) resulted in a match, while the remaining 6,998 pairs (83.53%) did not (Appendix 1).

Demographics

The gender distribution among participants was evenly balanced (4184 male, 4194 female). Regarding racial similarity, 3,316 pairs shared the same race, while 5,062 pairs did not. Participants primarily consisted of young adults, with ages from 18 to 55, with a mean age of 26. The pairs showed a mean age gap of 3 years (Appendix 2).

There is a lack of consistency in field labeling, where variations such as "Business," "MBA," "Business[MBA]," and "business" are treated as distinct fields (Appendix 3). Subsequently, a cleaning process was implemented to eliminate duplicates, and similar fields were consolidated.

Traits
Due to the extensive array of features within the dataset, particularly concerning personality traits (e.g., attractiveness, intelligence, humor, shared interests, ambition, and sincerity), individual correlation maps were generated for each trait (Appendix 4).

Twenty-eight features were chosen to account for the pair's base ratings of the traits, their demographics, and expectations of the event. From the correlation map (Appendix 5), it is observable that 'Like' is highly correlated to the person's rating of attractiveness, humor, and shared interests (attractive_partner, funny_partner, and shared_interests_partner).

3.3 Preprocessing

Aggregating 'field' Categories
To address diversity in the 'field' feature, items were categorized into broader academic domains to facilitate better analysis and interpretation.

These fields were Arts and Humanities, Sciences, Law, Social Sciences and Policy, Health and Medicine, Engineering and Technology, Education, and Business and Economics. The fields were then organized into broader categories based on their respective domains.

Feature Selection
For the purposes of the study, derived features of the traits were removed from the dataset. The base features for hobbies (eg. sport) were also removed, but what remained was a score to determine the pair's correlation of interests for the sake of simplicity.

Null Values
Null values were identified in the dataset, particularly within features associated with ratings, such as 'shared_interests_o' and 'attractive_partner'. Rows with null values were removed from the dataset.

Final Dataset
The processed dataset consisted of 28 features (Appendix 6) and 5,025 rows. This serves as the final dataset used for creating the model.

3.4 Baseline

Proportional Chance Criterion(PCC)
The Proportional Chance Criterion (PCC) was calculated to establish a baseline against which the performance of the model could be evaluated. This criterion provides an indication of the expected accuracy if predictions were made at random, considering the distribution of the target variable within the dataset.

The PCC yielded a result of 72.48%. Multiplying this value by 1.25 yields 90.60%, suggesting that the model's accuracy should surpass 90.60%.

Previous Work

Kaggle user 'MicaelD' previously developed a classification model that predicted the target variable "match" with an accuracy score of 84.29% and a precision score of 56.67% using a Gradient Boosting Classifier Model. These scores served as the benchmark against which the classifier model was compared

3.5 Modeling

Various base models underwent grid search to identify the optimal model, without tuning parameters. Ten holdouts were conducted. The models tested included Random Forest Classifier, Linear Support Vector Classifier, Gradient Boosting Classifier, Logistic Regression, XGB Classifier, K-Neighbors Classifier, and Gaussian Naive Bayes (Table 1).

Table 1 Results of Model Search Without Hyperparameter Tuning

Classifier	Mean Test Accuracy	Mean Test Precision
RandomForestClassifier	0.8582	0.7136
LinearSVC	0.8516	0.6579
GradientBoostingClassifier (Default)	0.8597	0.6432
Baseline From Kaggle - GradientBoostingClassifier (Max Depth 4)	0.8587	0.6307
LogisticRegression	0.8524	0.6285
XGBClassifier	0.8551	0.6010
KNeighborsClassifier	0.8395	0.5315
GaussianNB	0.7686	0.3902

3.6 AUROC

Considering an imbalanced target variable (i.e., more non-match than match), ROC curves were plotted for each resampling method. The original dataset achieved an AUC score of 85.8% (Figure 1), which outperformed the various resampled datasets. Hence, no resampling method was necessary in this study.

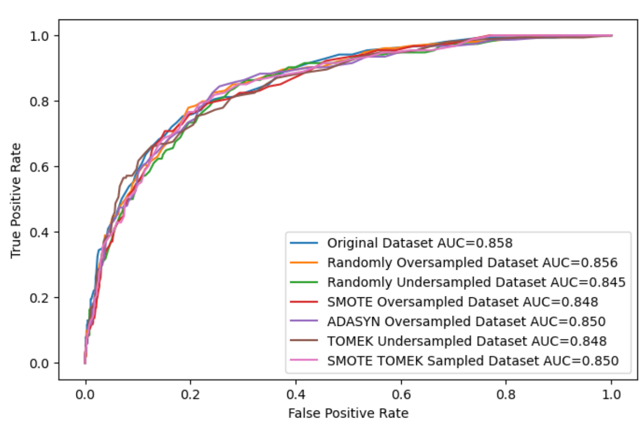


Figure 1 ROC Plot. The ROC curve is used to evaluate the performance of a classification model, providing insights into how well the model can distinguish between two classes under varying thresholds.

3.7 Hyperparameter Tuning

The Random Forest Classifier underwent grid search with various hyperparameters tested. The optimal hyperparameters were determined to be n_estimators=300, max_depth=15, and max_features=24. This resulted in a validation accuracy score of 86.34% and validation precision score of 69.30% (Table 2).

Table 2 Results of Model Search With Hyperparameter Configurations

ID	Classifier Configuration	Mean Test Accuracy	Mean Test Precision
225	RFC (300 Estimators, Depth=15, Features=24)	0.8634	0.6930
226	RFC (300 Estimators, Depth=15, Features=25)	0.8633	0.6898
262	RFC (300 Estimators, Depth=16, Features=26)	0.8632	0.6905
569	RFC (300 Estimators, Depth=None, Features=19)	0.8632	0.7008
221	RFC (300 Estimators, Depth=15, Features=20)	0.8632	0.6955

Similar hyperparameters were applied to the holdout dataset, resulting in an accuracy score of 86.49% and a precision score of 70.78%.

3.8 Statistical Tests

The t-test results showed that the Gradient Boosting Classifier model by MicaelD and the tuned Random Forest Classifier had similar accuracy scores but significantly different precision scores. The RandomForestClassifier outperformed the previous model in terms of precision. The accuracy t-test was 0.7655 and the precision t-test was 0.0357.

4 Results and Discussions

4.1 Results

While there was no statistical significance between the accuracy of the current study's model and user MicaelD, and even looking at there was a statistically significant change in the precision.The results of the current study yielded a precision score of 70.78%. This meant that the model when predicting a true positive, or in this case a match, the model was fairly confident in it's prediction. Such an improvement is very important in the case of predicting matches and in line with the objective of the study to reduce the possible frustration of users.

4.2 Interpretability

Considering the use of an ensemble model, interpretability techniques were utilized to explore the decision making process inside the black box. Particularly Shapley Values and counterfactuals, were used to better understand variable importance and explore alternative scenarios.

4.2.1 SHAP

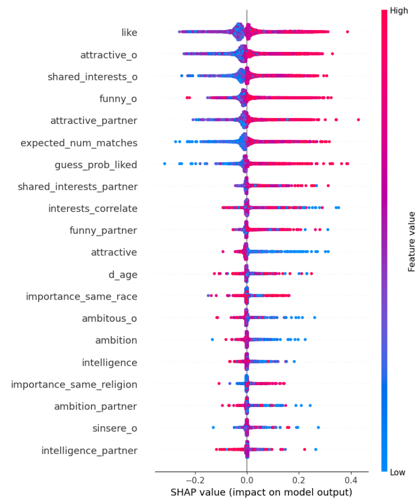


Figure 2 Beeswarm Plot. A Shapley beeswarm plot visually represents the impact of features on a model's prediction outcome.

The global importance of each feature within the predictive model was explored using Shapley values, a method from game theory adapted to assess individual feature contributions towards a model's predictions. This analysis revealed that the 'Like' variable is paramount; a higher rating of 'like' towards a partner substantially increases the probability of a match, as illustrated in Figure 2.

Moreover, the analysis highlighted the significance of mutual perception in the matchmaking process. Specifically, a partner's high rating on 'attractiveness' (attractive_o), 'shared interests' (shared_interests_o), and 'sense of humor' (funny_o) were strong predictors of a match. These findings underscore the importance of one's impression in the eyes of their partner concerning attractiveness, commonalities, and humor in the likelihood of forming a match.

Intriguingly, the perception of shared interests (shared_interests_o and shared_interests_partner) bore a greater influence on the prediction of a match than the actual correlation of hobbies (interests_correlate), indicating that subjective perceptions of similarity are more critical in the matching process than objective similarity metrics. This insight suggests that the feeling of connection based on shared interests plays a more vital role in determining compatibility than merely having similar hobbies.

4.2.2 DICE

Utilizing the DICE framework for model interpretability, Figure 3 lists the global importance of various features through counterfactual analysis. It is observed that common hobbies (interests_correlate) hold the most substantial influence, implying that shared interests significantly enhance match probability.

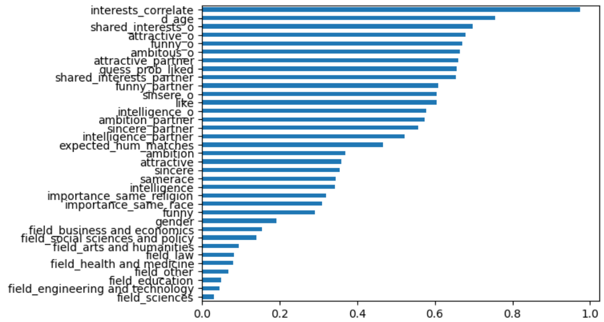


Figure 3 A vertical histogram of the weighted values given by DICE. The top weights are indicative of those that if changed will significantly contribute to changing the target into a 'match'.

The age difference (d_age) ranks as the second crucial feature, with the partner's ratings on traits such as 'shared interests' (shared_interests_o), 'attractiveness' (attractive_o), 'humor' (funny_o), and 'ambition' (ambition_o) following in importance. This suggests that both shared activities and perceived partner traits are vital components in the dynamics of match formation, similar inferences from SHAP.

Counterfactual analyses via DICE also allows to explore the influence of specific features on the transition from "non-match" to "match" outcomes. By manipulating features like partner ratings (e.g., attractive_o, intelligence_o) and interests_correlate, while holding one partner's ratings constant, the simulations assessed changes affecting match likelihood.

In Scenario 1 (Table 3), a male subject's low attractiveness and humor ratings precluded a match. Counterfactual adjustments to these ratings and a minor decrease in interests_correlate resulted in a favorable match outcome. This serve as an example of potential usage of the model and its interpretation to help users change outcomes.

Table 3 Counterfactual Analysis Outcomes

Feature	Initial Value	Counterfactual Value
gender	1	-
d_age	5.0000	-
samerace	1	-
importance_same_race	1.0000	-
importance_same_religion	1.0000	-
attractive_o	4.0000	8.0
sincere_o	6.0000	8.0
intelligence_o	7.0000	8.0
funny_o	7.0000	8.0
ambitious_o	5.0000	8.0
shared_interests_o	8.0000	-
interests_correlate	0.2600	-0.2
expected_num_matches	4.0000	6.0
match	0	1.0

4.3 Business Application

Part of the objectives to bring the created model into the hands of users. This can be done through potential implementation of the system into existing platforms. To achieve this, it is important to first understand the existing online dating process.

Users begin by creating profiles which include personal information and preferences. Based on these profiles, the algorithm provides a list of profiles for users to choose from based on shared interests, preferences, and even location. Once a match is made, the platform provides a channel for conversations through chat. However, the offering stagnates and the guidance of the platform ends.

After the match, users are left to decide whether or not to pursue an in-person meeting. This leaves plenty of room for uncertainty and hesitation which can hinder the chances of successful meetups.

Integrating the model into the existing pipeline can be done by collecting user feedback from both parties some time after a match. This information, combined with user details, are then used by the model to make the prediction of a good date or not.

Application of this new feature can vary between different platforms. It is possible to package the model as an online love coach or as a regular pop-up in the interface to guide users to initiate a date. Ultimately, the goal is to extend the platform's support for users. Thus, help them make informed decisions and improve user engagement.

5 Conclusions

In conclusion, the study's findings provide substantive insights into the intricate dynamics of match prediction within the realm of online dating. The precision of the developed model, at 70.78%, underscores its efficacy in predicting true matches. The application of Shapley values has been instrumental in unearthing the 'Like' variable as a critical determinant, reinforcing the role of mutual attraction and partner perception in predicting romantic compatibility.

Moreover, the DICE-based counterfactual analysis further illuminates the significance of shared interests and age disparity as influential factors, with a notable similarities with the inferences drawn from SHAP. These counterfactual scenarios not only highlight the crucial attributes leading to a match but also offer actionable insights for users seeking to improve their chances within the online dating landscape.

Discussion of the results ended with a suggested method to integrate the predictive model as a value-added feature in dating platforms. This holds the promise of enhancing user experience by facilitating more meaningful connections, creating a more attractive product, and guiding users in their journey from match to meaningful interaction.

6 Recommendations

Based on the findings and insights derived from the current study, the following recommendations are proposed to enhance the scope and applicability of future research:

1 Newer Datasets
Given the evolving social norms and technological advancements, there is a crucial need for utilizing more recent datasets in predictive modeling. Outdated datasets may not accurately reflect the complexities and dynamics of modern relationships and social interactions.
A dataset that encapsulates the current societal trends and preferences will likely yield more relevant and insightful predictions. Therefore, future studies should prioritize acquiring up-to-date data to potentially improve the precision and applicability of the predictive models.

2 Other Relationship Types
The current study predominantly focuses on heterosexual relationships, leaving a significant gap in understanding same-sex and non-binary relationships within the context of online dating.
Future research should aim to incorporate a broader spectrum of relationship types, including same-sex and non-binary partnerships, to ensure inclusivity and comprehensive representation. This expansion would not only enrich the dataset but also provide valuable insights into the varying dynamics across a wider audience.

3 Post-First-Date Dynamics
While the current model effectively predicts the likelihood of a successful first date, understanding the subsequent progression of relationships post-match remains an area ripe for investigation.
Future studies could focus on the outcomes following the initial date, examining factors that contribute to the sustainability of a relationship, repeated interactions, or eventual dissolution. This approach would offer a more holistic view of online dating success and could inform strategies to support users in navigating the complexities of relationship development beyond the first meeting.

7 References

1 Emotional fatigue and burnout in online dating - data study. Singles Reports. (2022, April 18).

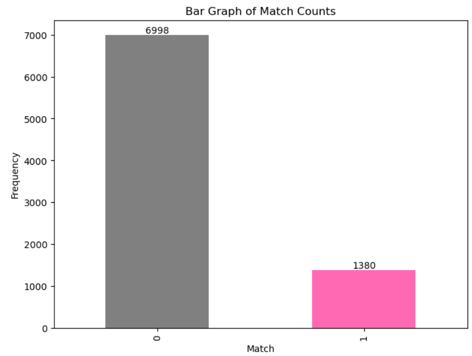
2 Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. The Quarterly Journal of Economics, 121(2), 673–697.

3 Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? machine learning applied to initial romantic attraction. Psychological Science, 28(10), 1478–1489.

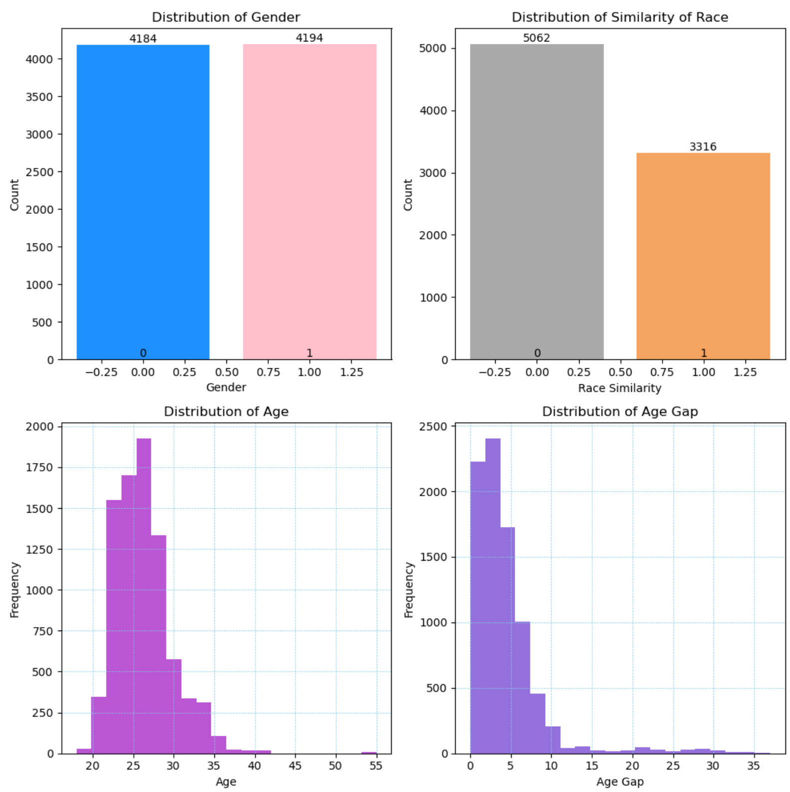
4 Online dating - worldwide: Statista market forecast. Statista. (2023, February).

5 Vera Cruz, G., Aboujaoude, E., Rochat, L., Bianchi-Demichelli, F., & Khazaal, Y. (2023). Finding intimacy online: A machine learning analysis of predictors of success. Cyberpsychology, Behavior, and Social Networking, 26(8), 604–612.

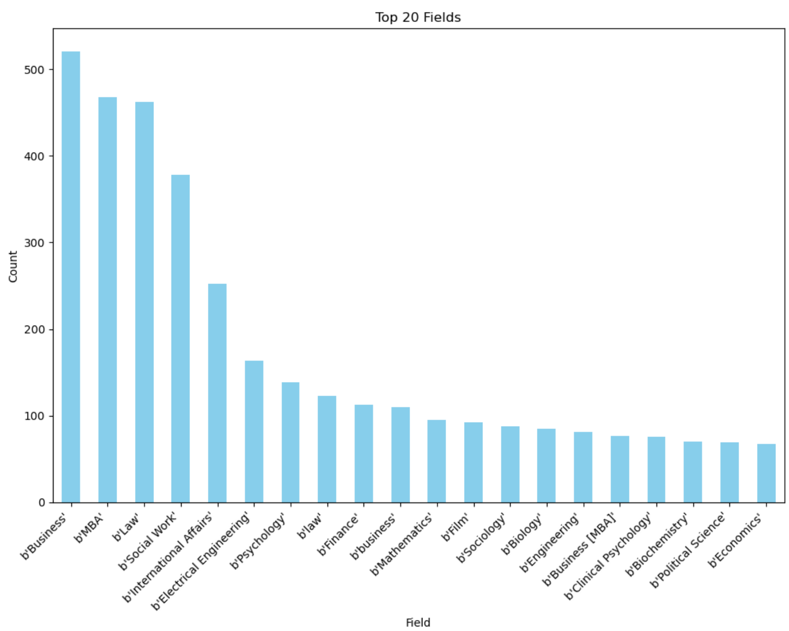
Appendix



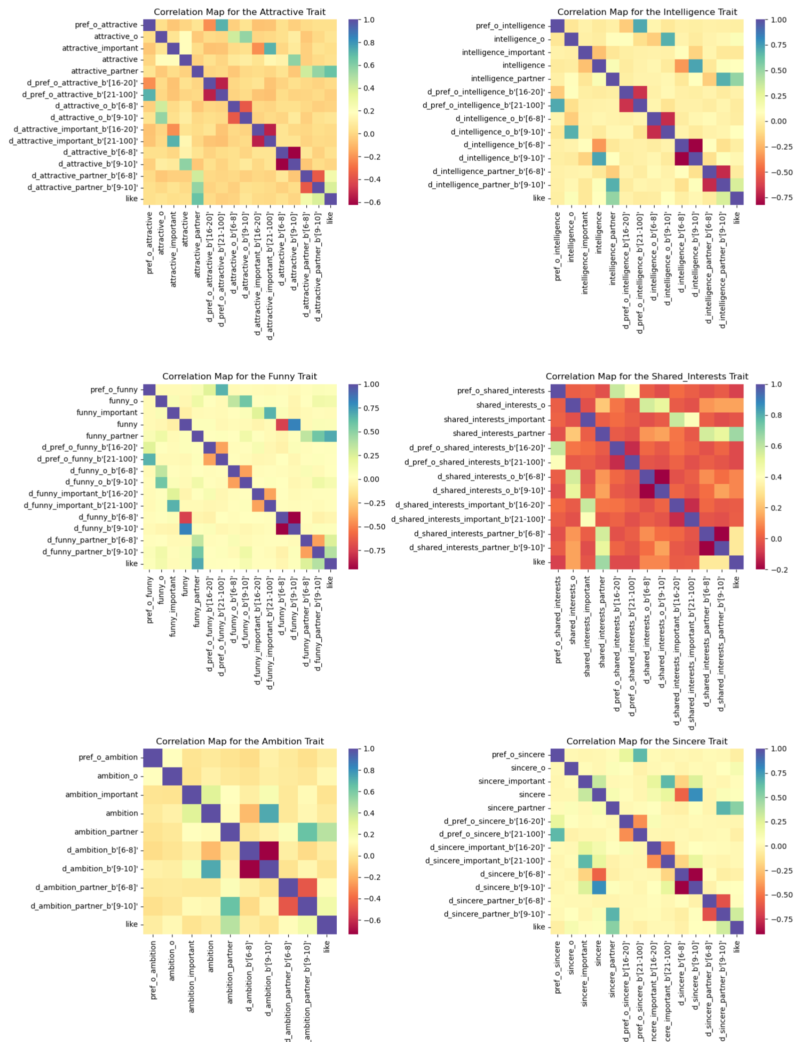
Appendix 1



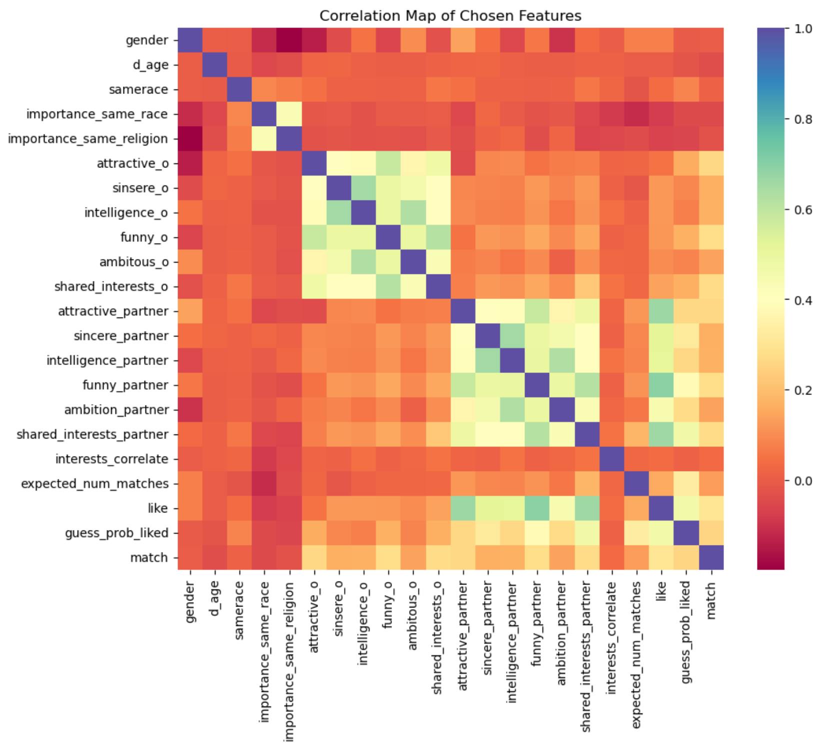
Appendix 2


Top 20 Fields
Count
500
400
300
200
100
0
b'Business'
b'MBA'
b'Law'
b'Social Work'
b'International Affairs'
b'Electrical Engineering'
b'Psychology'
b'law'
b'Finance'
b'business'
b'Mathematics'
b'Film'
b'Sociology'
b'Biology'
b'Engineering'
b'Business [MBA]'
b'Clinical Psychology'
b'Biochemistry'
b'Political Science'
b'Economics'
Field

Appendix 3


Correlation Map for the Attractive Trait
pref_o_attractive
attractive_o
attractive_important
attractive
attractive_partner
d_pref_o_attractive_b'[16-20]'
d_pref_o_attractive_b'[21-100]'
d_attractive_o_b'[6-8]'
d_attractive_o_b'[9-10]'
d_attractive_important_b'[16-20]'
d_attractive_important_b'[21-100]'
d_attractive_b'[6-8]'
d_attractive_b'[9-10]'
d_attractive_partner_b'[6-8]'
d_attractive_partner_b'[9-10]'
like
Correlation Map for the Intelligence Trait
pref_o_intelligence
intelligence_o
intelligence_important
intelligence
intelligence_partner
d_pref_o_intelligence_b'[16-20]'
d_pref_o_intelligence_b'[21-100]'
d_intelligence_o_b'[6-8]'
d_intelligence_o_b'[9-10]'
d_intelligence_b'[6-8]'
d_intelligence_b'[9-10]'
d_intelligence_partner_b'[6-8]'
d_intelligence_partner_b'[9-10]'
like
Correlation Map for the Funny Trait
pref_o_funny
funny_o
funny_important
funny
funny_partner
d_pref_o_funny_b'[16-20]'
d_pref_o_funny_b'[21-100]'
d_funny_o_b'[6-8]'
d_funny_o_b'[9-10]'
d_funny_important_b'[16-20]'
d_funny_important_b'[21-100]'
d_funny_b'[6-8]'
d_funny_b'[9-10]'
d_funny_partner_b'[6-8]'
d_funny_partner_b'[9-10]'
like
Correlation Map for the Shared_Interests Trait
pref_o_shared_interests
shared_interests_o
shared_interests_important
shared_interests_partner
d_pref_o_shared_interests_b'[16-20]'
d_pref_o_shared_interests_b'[21-100]'
d_shared_interests_o_b'[6-8]'
d_shared_interests_o_b'[9-10]'
d_shared_interests_important_b'[16-20]'
d_shared_interests_important_b'[21-100]'
d_shared_interests_partner_b'[6-8]'
d_shared_interests_partner_b'[9-10]'
like
Correlation Map for the Ambition Trait
pref_o_ambition
ambition_o
ambition_important
ambition
ambition_partner
d_ambition_b'[6-8]'
d_ambition_b'[9-10]'
d_ambition_partner_b'[6-8]'
d_ambition_partner_b'[9-10]'
like
Correlation Map for the Sincere Trait
pref_o_sincere
sincere_o
sincere_important
sincere
sincere_partner
d_pref_o_sincere_b'[16-20]'
d_pref_o_sincere_b'[21-100]'
d_sincere_important_b'[16-20]'
d_sincere_important_b'[21-100]'
d_sincere_b'[6-8]'
d_sincere_b'[9-10]'
d_sincere_partner_b'[6-8]'
d_sincere_partner_b'[9-10]'
like

Appendix 4


Correlation Map of Chosen Features
gender
d_age
samerace
importance_same_race
importance_same_religion
attractive_o
sinsere_o
intelligence_o
funny_o
ambitous_o
shared_interests_o
attractive_partner
sincere_partner
intelligence_partner
funny_partner
ambition_partner
shared_interests_partner
interests_correlate
expected_num_matches
like
guess_prob_liked
match
1.0
0.8
0.6
0.4
0.2
0.0
-0.0

Appendix 5