Predicting the Chance of Second Date Using Gender Influence in Mate Selection

Kamesh Munuswamy
Msc in Data Analytics
x19199562@student.ncirl.ie
National College of Ireland, Dublin

***Abstract*—Initial interaction with loved ones or crush has profound implications for the success of personal relationships. Dating is one such culture and environment which bond people together to share their love and mutual interest. From the researches it is been found that numerous companies investing significantly on online dating platform and dating apps.However, there are less known researches about the effect of personality variations on the initial sexual attraction. We have investigated dis-positional mindfulness and anticipated an eventual sexual desire beyond the influence of physical appearance in a speed-dating experiment. This study on gender influence in mate selection with deeper analyses found that, women were more attracted to men by their dis-positional intelligence, despite the influence of physical attraction. Men were more attracted to women who were sexually more attractive, but female mindfulness did not affect male initial desire. With gender influencing factors of both male and female, Multi-linear regression a data mining model has been implemented to predict the chance of second date. The Multi-linear regression model fit to identify influencing gender factors of male and female has be validated using chosen evaluation metrics.**

***Index Terms*—Dating, Gender, Data mining, Second date, Gender Influence, Speed-dating.**

I. Introduction

Today , society continues to value attractiveness, intellect, and money. The media emphasizes on those who are considered attractive, educated individuals are constantly celebrated and admired, and the affluent are envied for their lavish lifestyles. Many individuals have set their overall aspirations to be skilled in one or more of these fields, and in the quest for a lifelong partner, men and women continue to put priority on these qualities. Data from a speed dating experiment was used to investigate this assertion.

Choosing of a long term partner is a commitment that has significant consequences. Many study has focused on the role of physical attraction in the early romantic attachment. More recently , researchers have started to investigate the human variation predictors of emotional bond. In this present study , we investigated how the behavioral mindset anticipated an immediate decision of second date in a speed-dating experiment.

Dating culture has paved a platform for matching industries in market which made the business to invest significant resource and time on online dating. Numerous dating sites and dating mobile applications has been setup which is a great success to see people getting involved in online dating. This study mainly focuses on predicting the gender influence factors and mindfulness of each individual on choosing there partner for second date.

The data set used in this research was collected by Columbia Business School. Subjects were collected for experimental speed dating activities held in the confined room environment in which each participant would participate in four-minute interactions like a ”first date” with another participant of the opposite sex. Using the speed dating method, the study participants offered both experimental guidance and subsequent results on decisions that were the same as real-life decisions as to what a person really desires in a potential partner. All participants were graduate students from Columbia University who responded to e-mail and flier request are to take part in the study. Until engaging in the speed dating experience, each participant was required to complete a pre-event survey in which they were asked about characteristics such as their name, age, the undergraduate school they attended, and the zip code of where they grew up. We were also asked to provide scores to the six qualities of attractiveness, sincerity, intellect, fun, motivation, and mutual interest as these attributes are considered important in choosing their future dates.

In this analysis of predicting the chance of choosing for second date using speed experimental data, in dept insights of gender influencing attributes are compared and visualized by uni-variate, bi-variate and multivariate analyses. We provide empirical findings on two perspectives of gender influence. Next, we mentioned about the value of men’s and women ’s attributes. Women put more emphasis on intellect than men do, while men place more emphasis on physical attractiveness. Females also placed further emphasis on the race of the partner. We find that a man ’s desire for intellect and determination does not extend to women who are more educated or motivated than men. In reality, a man is far less likely to consider a woman who is more ambitious than him. At last, women choose men who grow up in wealthy neighborhoods, while men show no such interest.

The remainder of the paper is structured as follows. Section II is Literature review, Section III reports experimental design and analysis and Section IV explains about model and implementation and Section V is conclusion.

II. Literature Review

From a recent study [1] to explore and understand the influencing factors that provoke college students to get involved in "online dating" has been performed. In this paper, study has been conducted considering some of the important feature such as personality, gender and social impact. As a result of this research, it is been concluded that college students who have been emotionally impacted, personally attractive and due to peer pressure get involved in online dating. From statistic point of view, it is been found that there is no statistically significant difference exist between male and female students with respective to online dating.

In considering the philosophical perspective, the study has been made on "online dating" impact on traditional moral ethics. This analyzes addresses three kinds of ethical problems, such as role playing, sincerity, online dating games and fidelity issues and finally sexual related problems due to online dating. With an insight and deeper research it is evident that though online dating problems are being inconsistence with traditional relationship ethics which doesn't portrait role playing and online dating fun games are immortal [2].

Given the desire to attract both sexes, 84% of homosexual engage in same sex relationships. The principles underlying these dating decisions remain uncertain. This paper [3] discusses three hypotheses of differences : (1) the urge for reproduction, (2) the need to conform with social standards, and (3) the specific level of the accessible dating pool. Bisexuals (n=132) aged 18 to 49 conducted an online study of their bisexual views, sexual reproduction beliefs, and dating behaviors.

This paper [4] discusses shifts in consumer behaviour induced by the introduction of a phone app in terms of interaction and related outcomes in the sense of online dating. They further recognize three processes that are quite special to the mobile world yet not identified in the literature that cause this change in behavior: ubiquitousness, impulse control, and disinhibition. Our key recognition technique uses inclination score balancing paired with variance societal inequalities, along with a stringent falsification check to validate the validity of our identification strategy.

This article [5] discusses the effect of behavioral modeling theories upon on Complex Adaptive System (CAS) model. The author assume that ethical theories can be optimistic in their forecasts due to the complexities of modeling behavior. Supporting this theory, this study describes the obstacles of modeling behaviour, introduces the CAS illustration question of creating an online dating app, designs the dating app as a CAS, and examines the effects of various behavioral models on architecture. It also illustrates how specific the behavioral models can have big effect on the effects of the simulation.

Detailed research [6] has concentrated on social anxiety in various aspects of interpersonal relationships. Nevertheless, only a limited number of experiments have been undertaken to clarify social anxieties in the sense of online dating. The present research therefore looked at factors relating to social uncertainty among online daters. A group of 494 users of internet dating platforms and apps completed a survey evaluating social anxiety, socio-demographic factors and three general conclusions concerning: the environment (environment assessments); self-efficacy; and others (recognizing problems). Multiple regression model was used to classify predictors of social anxiety. The results showed that social anxiety between online daters was predicted by negative world evaluations, weak self-efficacy, and high identification issues. The new research highlights the significance of executive mechanisms in an ambiguous and insecure situation, such as online dating. Actually, taking into consideration the fundamental expectations of the environment, the self, and others, arises as crucial considerations in attempting to understand social anxieties in the form of online dating.

This research [7] analyzes how/why online daters are separated by class, and deliberately self-present in online dating profiles while following two opposing goals: gaining prospective daters and preventing recognition as liars. The results indicate that the purpose of presenting an appealing appearance in online dating was substantially correlated with an insatiable self-presentation. Online daters have embraced falsity more than any technique, and women have been more likely than males to exaggerate their self-presentation, particularly their external attractiveness.

In this article [8] , the author proposed a strategies to classify illicit activity based on early communication credibility, behavioral attributes and interaction content. In the context of statistical approaches, we suggest a Time Series (TIME) model based on the study of user experience characteristics. From the Application Session environment, the enhanced Latent Dirichlet Allocation (LDA) malicious environment architecture is designed to boost the performance of the application to detect malicious themes.

In this article [9], the author argues that online dating has the ability to reshape the dating phase and to bring improvements in the cultural structures of our community. Such developments will result in the adoption to a globalized dating community as online dating apps expand worldwide acceptance. This paper further discusses the literature on actions in the online dating world, underlining the benefit that computer-mediated contact is considered to provide.

A fascinating research found that people who are prone to rejection are more likely to utilize dating websites. The purpose of this research was to understand the connection between rejection sensitivity and the use of the online dating platform. Study 1 investigated how authentic desires-mediated the relationship between rejection vulnerability and online dating. Study 2 tried to replicate the results of Study 1 and analyze how self-disclosure moderated the interaction between true self and online dating in the intervention paradigm. Results showed the true self-mediated association between rejection vulnerability and the usage of the online dating platform [10].

III. Experimental design and Analysis

A. Data description

The data used for this research was extracted from Kaggle.com an open source data repository. This data was constructed by Columbia Business School by conducting speed dating experiment. The participants were asked to rate their date on six attributes such as Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests. The data is gathered through a questionnaire survey. The entire dataset contain 8378 records with 195 attributes.

B. Data Visualization

Detailed data analysis has been performed using data visualization at the initial stage to understand the data structure. All the important attributes is been analyzed and visualized as a part of uni-variate, bi-variate and multivariate analysis. The important attributes has been visualized and compared with each other to understand the relationship and correlation between them.

We have performed uni-variate and bi-variate analysis on the data to discover the distribution of the values and the data structural behavior. This data visualization gives better insights about the data at all levels by which data preprocessing and implementation is carried out.

Univariate Analysis: Univariate analysis is a simple and effective method which deal with only one attribute as a part of analysis. The main purpose of this analysis is to describe the attribute, summarize and find patterns.

- **Attribute Distribution: Age**
  The below histogram is plotted for both male and female gender to infer the age distribution of both gender category. Not to the surprise most of the participants fall between early twenties and mid thirties which shows that teenagers are pretty much into dating in search of their partners.

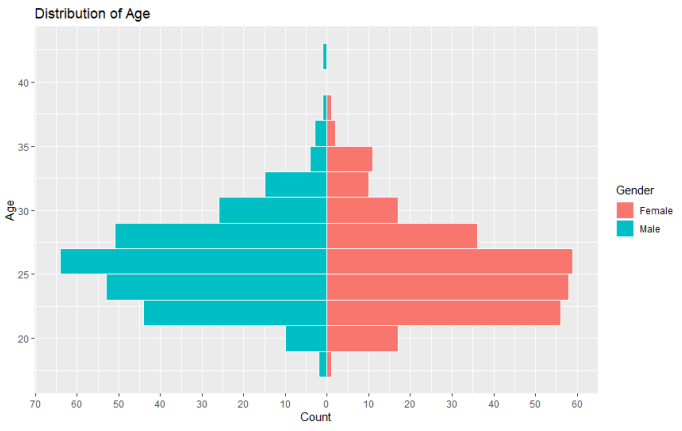

Fig. 1. Age distribution

- **Attribute Distribution: Race**
  From the below plot it is evident that, the dataset consists of almost 50% of Europeans in both male and female category. Followed by second large count of Asian/Asian American occupying 25%. Whereas, Other races category share 5% each in both male and female gender.

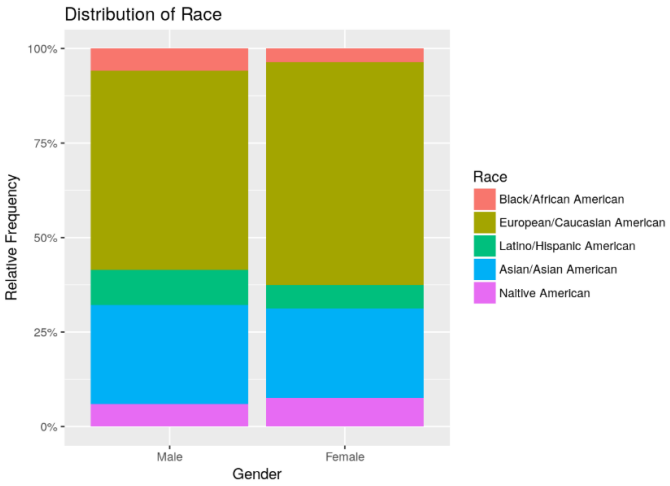

Fig. 2. Race distribution

- **Attribute Distribution: Goal re-partition**
  The histogram (Fig 3) is been plotted to understand the both male and female mentality or their goal while dating their partner. Unsurprisingly, Both the male and female gender comparatively likes to have fun night out on their date. On the other hand, females shows interest on dating to meet new people. From this graph it is been clear that both the gender prefer dating for a casual meet and hangouts in their initial meet ups.

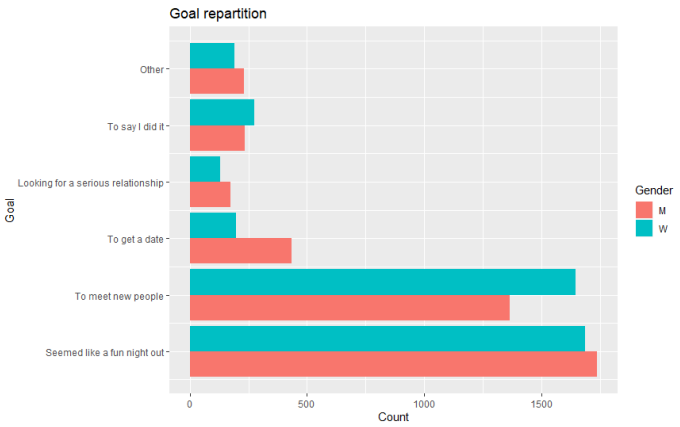

Fig. 3. Goal distribution

- **Attribute Distribution: Dates repartition**
  The below histogram distribution (Fig 4) is plotted to understand the time interval of each participants of both the gender willing to date. It was evident to see both male and female are interested to date at least twice a month. Surprisingly, females are more enthusiastic than males to date several times a year. At the same time majority males willing to go on date at least once a month.

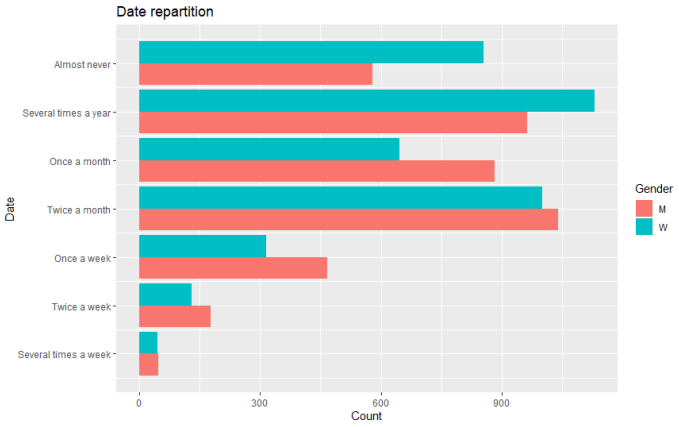

Fig. 4. Dating time interval distribution

- **Attribute Distribution: Matches by Gender**
  The below histogram is plotted to identify the number of matches count for both male and female. The Graph shows a strong evidence that over 4000 participants including male and female fails to find a partner for a date. On the other end only less than 1000 participants are successful in finding their right match for a date.

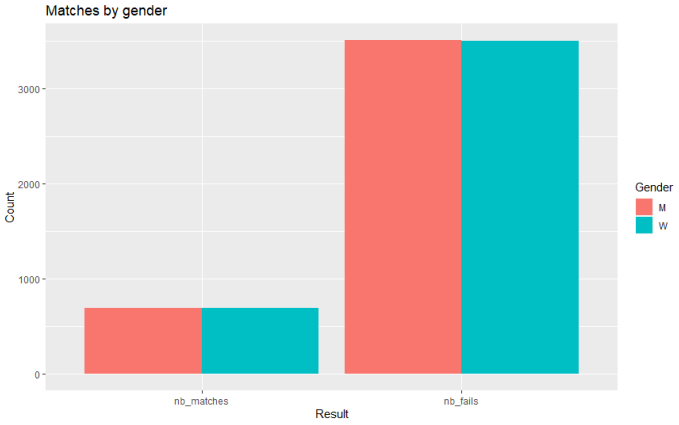

Fig. 5. Dating time interval distribution

Bi-variate Analysis:

Bi-variate analysis is one of the most common forms of quantitative (statistical) analysis. This requires the study of two attribute (often referred to as X , Y) in order to establish the empirical relationship between them. Bi-variate evaluation can be helpful in analyzing basic correlation theories.

In this analysis we have considered the most important factors of people interests considered when choosing a match for a date. The gender influencing factors such as Attractive, Sincere, intelligent, ambitious, shared interest and fun. All the participants were requested to rate their interest and expectation while choosing their date based on these 6 attributes. These Influencing factors on comparative analysis helps immensely in identify the people mentality and expectation when choosing their match for dating.

- **Attribute Distribution: Male Interest factors**
  The below graph is plotted against males interest over female as an influencing factor in choosing a date partner. The radar graph gives a clear picture of male perspective on female on considering six important factors as part of this speed dating experiment. It is observed that males greatly emphasize on physical attractiveness of female other than intelligent or ambitious. It is been understood that men are looking to date attractive females as a matter of fact that most of the men are looking for causal dating.

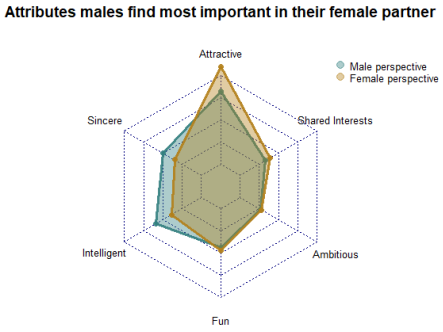

Fig. 6. Male radar of interest over female

- **Attribute Distribution: Female Interest factors:**
  The below radar graph is plotted against female interest over males. From the radar of female perspective of interest in males it is been found that females are interested in intelligent men. At the same time females expect men to be ambitious. Unlike men put more weight on physical attractiveness, females do expect important qualities from men to be intellect and highly motivated in nature when considering for a date. In fact, females are less significantly to accept men with no intellect and greater ambition about the future.

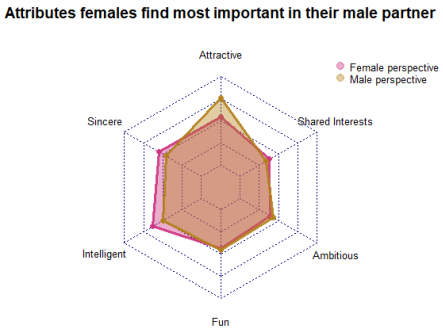

Fig. 7. female radar of interest over male

IV. Model and Implementation

From the above findings using visualizations, the best fit model for predicting the chance of second date considering the gender influential factors has been decided. In this research, considering the data insights we have chosen to implement Multi-Linear regression model.

A. Multi-Linear Regression model

Multi-Linear regression model is specially designed for establishing the relationship between two or more variables by changing the linear equation to the data being studied. In linear regression, target variable or dependent variable is predicted using number of independent variables.

Multi-linear regression is a method that helps an analyst or researcher to make assumptions about one variable depending on knowledge about other variable. Linear regression should only be used where there are two continuous variables — an independent variable and a dependent variable. The independent variable is the variable used to determine the target variable or the result. A multi-regression analysis is applied to a variety of independent variable [6].

$$y_i = \beta_0 + \beta_1 x_1 + \beta_{2x2} + ... + \beta_{pxp} + c \qquad (1)$$

where, for i=n observations
y – dependent variable.
B0 – y-intercept
Bp – slope coefficient for each independent variable.
x – independent variable
c – residual error.

B. Model Assumptions

There are some important assumptions must be satisfied before implementing the Linear regression model.

- There should be existence of linearity between target and independent variables.
- The independent variables should not be closely associated with each other.
- Observations of y variables are selected individually and uniformly among the populations.
- It is important for the variables to be Normal distribution.

Linearity Assumption Check

The below graph is plotted to check the linear relationship among the dependent variable and each of independent variable. It is evident that independent variables shows good linearity among them. This assumption will help in determining the linear relationship of the target variable.

If the variables shows non-linearity then it is mandatory to perform non-linear regression in order to transfer the data to linear form.Non-linear regression analysis is carried out based on the data characteristics by performing series of analysis and checks. Outliers in the data causes major impact in effecting the linearity of the dependent and independent variables.

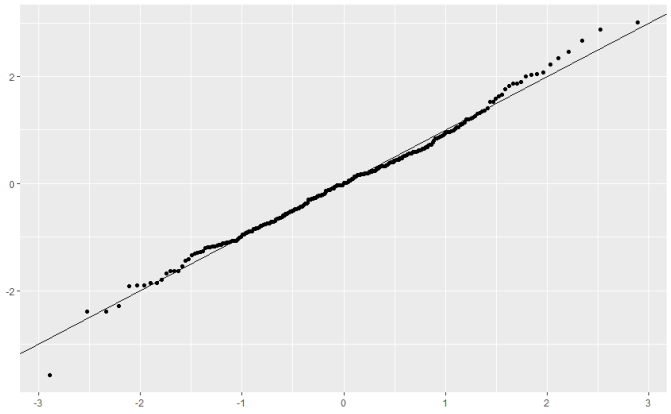

Fig. 8. Linearity Plot

Normal Distribution

Followed by linearity check one of the important assumption is to consider is normal distribution of residuals. The below histogram is plotted to understand the normal distribution of residuals. It is satisfying to see the residuals in our data follow very clear normal distribution.

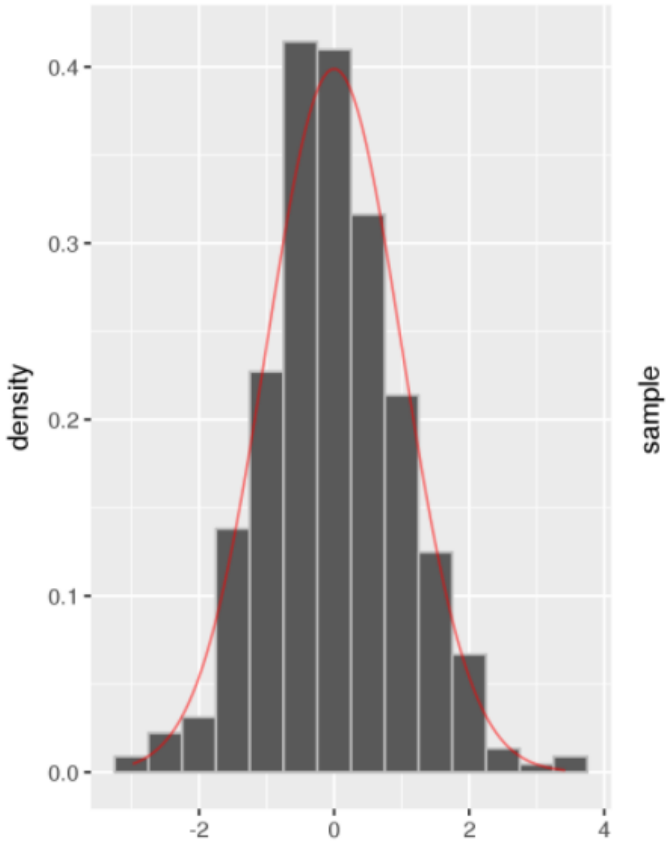

Fig. 9. Normal Distribution of residuals

V. Evaluation and Interpretation

In this section the implemented Multi-linear regression model outcome is interpreted and validated for its performance using chosen evaluation metrics.

In this research to forecast the probability of second date using gender influential factors, the implementation is performed in two stages for male gender and female gender. The main aim of this research is to identify the influencing factors individually which is the driving factor for both male and female to decide their partner for second date.

Male Gender Model Fit

At first, the linear regression model is implemented to each attributes of the male participants against female influencing factors. Each model fit using different attributes is evaluated using evaluation metrics.

According to the provided attributes in the data such as attractiveness, shared interest, ambition, intelligent and fun for both male and female, it is important to understand the correlation between each other attributes.

The below correlation graph is plotted for all male gender attributes. Relying on the matrix, the attributes with the largest relation to the percent of the choices were attractiveness and fun. Linear regressions of these properties have been established. Since the initial assumption claimed that women appear to consider intelligence to be the most valuable trait of their male partners. When the pair matrix apparently showed, attraction has the highest positive correlation with the number of choices with a correlation coefficient of 0.79. Intelligence has the produces the least with percentage of distributions of the three qualities chosen with a correlation coefficient of 0.31.

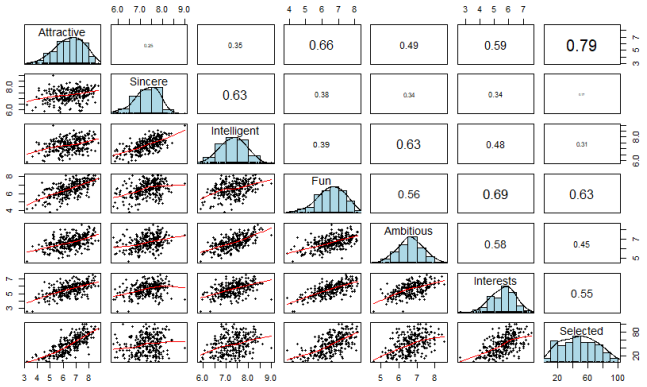

Fig. 10. Correlation plot for male attributes

Multiple regression analysis reveals that using a mix of all parameters, there was a relatively good significant relationship of percent of the choices. The modified R-square value was 0.6552, which means that 64.74 percent of the variance in the relationship between all parameters and the percent of the choices can be compensated for by the best-fit line in this regression study. The most statistically important variables were attractiveness with the minimum p-value and fun with a p-value of 0.000358. Honesty still shows a degree of importance, but the qualities of honesty, intellect, and ambition had no meaning in the p-value model above 0.05.

```
Call:
lm(formula = selected_perc ~ attr + sinc + intel + fun + amb +
    shar, data = regression_men)

Residuals:
    Min      1Q  Median      3Q     Max
-41.349  -9.549  -0.860   9.289  48.347

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -66.4721    12.6617  -5.250 3.14e-07 ***
attr         13.8817     1.0900  12.735  < 2e-16 ***
sinc         -4.8869     2.0821  -2.347 0.019662 *
intel         2.4104     2.4407   0.988 0.324261
fun           5.7296     1.5845   3.616 0.000358 ***
amb          -0.1195     1.8422  -0.065 0.948308
shar          1.0978     1.5512   0.708 0.479775
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.86 on 263 degrees of freedom
Multiple R-squared:  0.6552,    Adjusted R-squared:  0.6474
F-statistic:  83.3 on 6 and 263 DF,  p-value: < 2.2e-16
```

Fig. 11. Multi-linear regression outcome for male attributes

The below scatter plot was plotted for against attractiveness score and percentage of selection of male attribute.

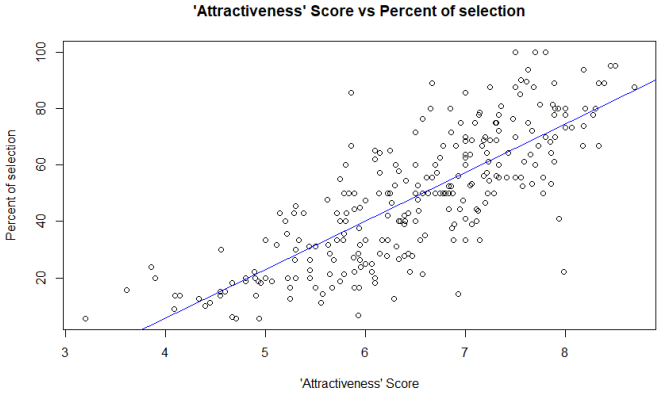

Fig. 12. Attractiveness score Vs percentage of selection male

Female Gender Model Fit

As a part of second part of implementation, the linear regression model is implemented to each attributes of the female participants against male influencing factors. As part of evaluation, each model fit using different attributes is evaluated using evaluation metrics.

The same process was replicated in the case of females. A pair matrix was created to represent the relationship between the level of each female subject for the six attributes and the percentage of the choices of the subject. Based on the graph, the attributes with the strongest association with the percent of the preferences were, respectively, attractiveness and fun. Linear regressions have been conducted using these attributes. Attractiveness has the highest positive correlation with the ratio of the choices with a correlation coefficient of 0.74.

Multiple regression analysis reveals that using a mix of all parameters, there was a fairly good positive relationship for

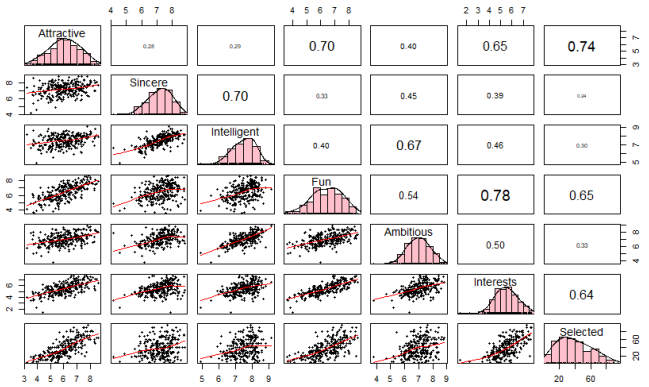

Fig. 13. Correlation plot for female attributes

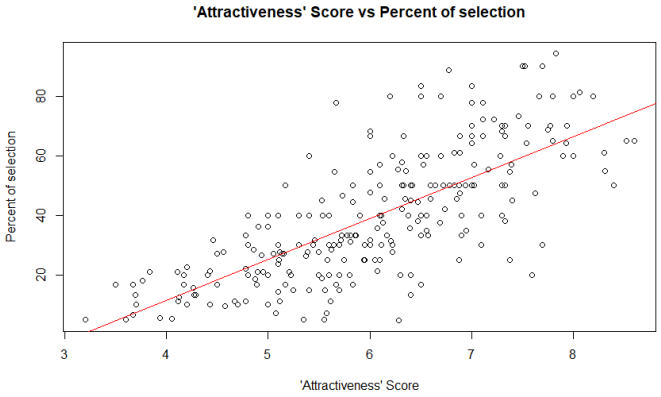

Fig. 15. Attractiveness score Vs percentage of selection female

percent of the selections. The modified R-square value was 0.5989, which means that 59.09 percent of the variance in the relationship between all parameters and the number of the choices can be compensated for by the best-fit line in this regression study. The most statistically important attribute was the quality of the lowest p-value. Shared desires and pleasure also displayed modest levels of importance, while other variables of honesty, intellect, and motivation had p-values greater than 0.05.

```
Call:
lm(formula = selected_perc ~ attr + intel + fun + amb + shar,
    data = regression_women)

Residuals:
    Min      1Q  Median      3Q     Max
-48.671  -9.285   0.199   7.787  41.445

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -57.086      9.885  -5.775 2.25e-08 ***
attr           9.595      1.067   8.991  < 2e-16 ***
intel          1.649      1.755   0.940  0.34833
fun            3.118      1.451   2.148  0.03265 *
amb           -2.673      1.566  -1.707  0.08906 .
shar           4.529      1.436   3.154  0.00181 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.64 on 253 degrees of freedom
Multiple R-squared:  0.5989,    Adjusted R-squared:  0.5909
F-statistic: 75.54 on 5 and 253 DF,  p-value: < 2.2e-16
```

Fig. 14. Multi-linear regression outcome for female attributes

The below scatter plot was plotted for against attractiveness score and percentage of selection of female attribute.

VI. Conclusion

As per the radar plots previously built, men seemed to place the most emphasis on the attributes of attraction, sincerity, and intelligence, while women seemed to place the most emphasis on the characteristics of intellect and honesty. It was noteworthy that both men and women overestimated the value of attraction for the opposite gender – women thought that men put a very significant weight in appearance, and men still assumed that women put the majority of weight in attractiveness.

The primary assumption for men is true. Men prefer to consider beauty as the most important aspect in choosing a date partner. The initial thesis for women appears to be a false. People are likely not to consider intellect to be the most important factor in finding a romantic partner, but instead, they often seem to consider attractiveness to be the most important factor in finding a romantic partner.

References

[1] J. Li, ”An Empirical Analysis of Influencing Factors of College Students’ Online Dating,” 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 2020, pp. 87-90, doi: 10.1109/ICCEA50009.2020.00025.

[2] S. Lomborg, ”Personal internet archives and ethics”, Research Ethics, vol. 9, no. 1, pp. 20-31, 2012. Available: 10.1177/1747016112459450.

[3] A. Wu, M. Marks, T. Young and M. Beasley, ”Predictors of Bisexual Individuals’ Dating Decisions”, Sexuality Culture, vol. 24, no. 3, pp. 596-612, 2019. Available: 10.1007/s12119-019-09651-1.

[4] J. Jung, R. Bapna, J. Ramaprasad and A. Umyarov, ”Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating”, MIS Quarterly, vol. 43, no. 1, pp. 47-72, 2019. Available: 10.25300/misq/2019/14289..

[5] A. J. Turner and J. Kuczynski, ”Impacts of Behavioral Modeling Assumptions for Complex Adaptive Systems: An Evaluation of an Online Dating Model,” 2019 Winter Simulation Conference (WSC), National Harbor, MD, USA, 2019, pp. 668-679, doi: 10.1109/WSC40007.2019.9004665.

[6] S. Pitcho-Prelorentzos, C. Heckel and L. Ring, ”Predictors of social anxiety among online dating users”, Computers in Human Behavior, vol. 110, p. 106381, 2020. Available: 10.1016/j.chb.2020.106381.

[7] K. Peng, ”To be attractive or to be authentic? How two competing motivations influence self-presentation in online dating”, Internet Research, vol. 30, no. 4, pp. 1143-1165, 2020. Available: 10.1108/intr-03-2019-0095.

[8] J. Qiu, X. Shen, Y. Guo, J. Yao and R. Fang, ”Detecting Malicious Users in Online Dating Application,” 2019 5th International Conference on Big Data Computing and Communications (BIGCOM), QingDao, China, 2019, pp. 255-260, doi: 10.1109/BIGCOM.2019.00043.

[9] Maria, Stoicescu G, ”The globalized online dating culture: Reframing the dating process through online dating,” 2019 Journal of Comparative Research in Anthropology and Sociology. Romania, 2019, pp. 100-113.

[10] M. Hance, G. Blackhart and M. Dew, ”Free to be me: The relationship between the true self, rejection sensitivity, and use of online dating sites”, The Journal of Social Psychology, vol. 158, no. 4, pp. 421-429, 2017. Available: 10.1080/00224545.2017.1389684.