Speed Dating Experiment: A Statistical Analysis of Speed Dating Data

OMIS 324

Brandon Hauck, Chris Jernigan, Allison Marshall, Phillip Wilson, Thomas Yi

Purpose

As the amount of data collected by businesses and organizations continues to grow, it becomes increasingly important for individuals from all backgrounds to understand how to read, analyze, interpret, and share the most important factors within this massive amount of data. Throughout the course of the fall semester, the students in OMIS 324 have learned a variety of techniques for performing these tasks, including but not limited to: descriptive statistics, data visualization, and hypothesis testing using multiple linear regression models. In this project, our group will explore one an extensive dataset—the Speed Dating Experiment—and show how this data can be described and analyzed using a variety of statistical analysis techniques learned in this course. We will begin by providing an understanding of the dataset in terms of background and variables. Then we will explore the known and theorized relationships in the dataset, including some of the major correlations. From these relationships, will choose and explore a single implicit relationship and undergo a set of hypothesis tests for regression analysis. Finally, we will use the results of these tests to draw conclusions and provide recommendations based on our overall analysis.

Dataset

Background: What is Speed Dating?

Before the era of online dating sites and dating mobile applications, Speed Dating was a highly popular method of finding a partner and meeting new people. Speed dating generally involves an equal or near equal number of men and women who are rotated around to meet every other participant over the course of the event. The dates themselves vary from two minutes to ten minutes depending on the organizer and the overall goal of the event. (Brown, 2003) At the end of each date, each participant indicates which partners with whom they wish to exchange contact information. The organizers will then send a list of matches to every participant following the event.

Experimental Method

In early 2002, Columbia Business School Professors Ray Fisman and Sheena Iyengar devised an experiment to gauge how men and women perceive the opposite gender during the dating process. (Montoya, 2016) To do so, the experimenters held twenty-one speed dating sessions from 2002 to 2004. These sessions were comprised of four-minute speed dates with all the participants of the opposite sex meeting each other only once, which varied between nine and twenty-two dates depending on the session. (Montoya, 2016) Prior to the first date, each participant filled out a survey including demographics, attribute preferences, self-assessments, interests, goals, and other identifying information relevant to the dating process. Following each individual date, the both participants filled out an additional survey scorecard. In this scorecard, each participant ranked their partner on attractiveness, sincerity, intelligence, fun, ambitiousness, shared interests/hobbies, how well liked they were, partner's probability of saying yes, and their overall match decision. The day after participating in the speed dating event, the experimenters sent the participants another survey, requesting additional information on the overall perceptions of the event. One final follow-up survey was sent three to four weeks after receiving a list of their matches, which included questions about any contacts or dates caused by the event.

Types of Variables

Our group obtained the dataset from Kaggle.com, which has all the data compiled into a .csv file spanning 119 total variables and 8,379 observations. As the dataset has an extensive number of variables, we focused our analysis on a few key variables, which we will detail below based on variable type.

In our data exploration and analysis, we examined a variety of categorical variables. While the dataset includes more variables of this type that could be analyzed, doing so went beyond the scope of this project. Each of the categorical variables listed below were assigned dummy values so that these categorical variables could be included in the statistical analysis. The categorical variables and definitions used in the analysis are as follows: gender, which indicates the participants gender as male (1) or female (0); match, which indicates if both participants indicated yes (1) or one or both of the participants indicated no (0); race, which indicated the race of the participants; dec and dec_o, which indicates a yes (1) or no (0) for the participant and the participant's partner respectively; goal, which was coded to survey questions relating to the goal of attending the event; and field_cd, which provides a coded field of study for each of the participants. For some of the visualizations and tables in our analysis, we recreated the aliases for these categorical variables to prevent confusion in the presentation of the data. We used the data key provided with the dataset to accomplish this manually.

We also employed a variety of ratio variables in our analysis. While the dataset includes more variables of this type that could be analyzed, doing so went beyond the scope of this project. The ratio variables and definitions used in the analysis are as follows: age, which indicates the age of the participant; income, which indicates the income of the participate attr_o, amb_o, fun_o, intel_o, shar_o, and sinc_o, which indicate how the participant was ranked in the attributes of attractive, ambition, fun, intelligent, shared interests/hobbies, and sincere respectively by each respective partner; attr1_1, amb1_1, fun1_1, intel1_1, shar1_1, and sinc1_1, which have a combined total of 100 and indicate the importance of attractive, ambition, fun, intelligent, shared interests/hobbies, and sincere respectively for the participant; and like_o, which indicates how much the partner liked the participant overall.

Finally, we analyzed a single ordinal variable from the dataset in our descriptive analysis—imprace, which indicates on a scale of 1-7 how important race is in the opposite sex.

Descriptive Statistics

Given the size of the dataset, we could analyze and include a significant number of variables using descriptive statistics, but for the scope of this project, we focused on the major demographic variables and the variables used in our later analysis. Below is a breakdown of the mean, median, mode, standard deviation, variance, range, minimum, and maximum for age and income.

	Mean	Median	Mode	Std Dev	Variance	Range	Min	Max
Age	26.46	26	27	3.56	12.7	37	18	55
Income ($)	44,887	43,185	55,080	17,206	296,078,129	100,424	8,607	109,031

Per the data table above, there was a significant disparity in the age and income class for the participants in the study. The average participant was 26 years old with an income around $44,887, but the large range was cause for further analysis.

Analyzing age alone did not accurately capture the breakdown of the participants, so we divided this information again by gender. In the first table, we can see that the overall age range for the participants in the dataset was 37, which ranged from the minimum of 18 to the maximum of 55 years of age. After breaking down the dataset further into the table below, we can see that most of this range comes from the female participants, who span from 19-55 years of age. A 55-year-old female would be considered a significant outlier in this dataset, as it is nearly eight standard deviations away from the mean. It is important to understand some of these outliers within the dataset, as they have the potential to skew the analysis. For the sake of the project, we left any outliers in demographics within the dataset, as we were not exploring the impact of age.

Age by Gender	Mean	Median	Mode	Std Dev	Variance	Range	Min	Max
Female	25.99	25	24	3.68	13.54	36	19	55
Male	26.44	26	25	3.64	13.25	24	18	42

Aside from gender, we wanted to explore the dataset for racial breakdowns across genders. As seen in the table below, there were a disproportionate number of Caucasian participants across both males and females (56.85%). African American and Latino participants accounted for less than 15% of the total dataset. This is in line with the overall demographics of the United States, so this dataset can be considered a good sample representation of the United States in this regard.

Race by Gender	**African American**	**Caucasian**	**Latino**	**Asian/Pacific Islander**	**Other**
Female	202	1979	287	701	217
Male	158	1757	147	642	116

Finally, the table below explores the descriptive statistical measures pertaining to the survey question, “What do you look for in the opposite sex?” These values combined had to add up to 100, as detailed in the variable section above.

	Mean	**Median**	**Mode**	**Std Dev**	**Variance**	**Range**	**Min**	**Max**
Attr1_1	22.51	20	20	12.59	158.45	100	0	100
Intel1_1	20.27	20	20	6.78	46	50	0	50
Sinc1_1	17.4	18.18	20	7.05	49.66	60	0	60
Amb1_1	10.68	10	10	6.12	37.51	53	0	53
Fun1_1	17.46	18	20	6.09	37.03	50	0	50
Shar1_1	11.85	10	10	6.36	40.48	30	0	30

Based on this data table, we can see that shar (shared interests/hobbies) was the only attribute to never get distributed half of the available points (50) by a participant. Additionally, attractiveness has the highest variance across these six attributes, and it was the only attribute rated as the only important attribute by a participant. For attractiveness to be rated at 100, all other attributes on the survey would be rated at 0. On average, attractiveness is the most important attribute, followed by intelligence and fun. A visualization of this variable can be found in Appendix A**.**

For the full range of descriptive statistic tables and visualizations for this dataset, please see appendix B.

Relationships Among Variables

Explicit Relationships

While not discussed in the variables above as we do not use them for analysis, there is an explicit relationship between igd (the unique id within the session) and id (the unique subject number for the experiment) and as igd is determined by the id. Other explicit relationships may exist in the 119 variables within the dataset, but a full exploratory analysis of this data was beyond the scope of the project.

Implicit Relationships

Implicit relationships are not as simple to determine, and they often require statistical or visual analysis to understand and interpret the relationship. Our data set had a series of relationships such as these that could be explored through further analysis. For example, we can analyze the data in more depth to determine the extent of the implicit relationship between the individual's importance of race (imprace) and whether they said yes to a with a partner of a different race. In addition, the dataset includes many variables on the different rankings for the different attributes as detailed above. A relationship between attributes and gender, attributes and race, or attributes and age can tell us the different preferences and trends amongst different demographics. Furthermore, the relationship between position and decision can tell us if there is a better chance of getting a date at the beginning of an event, in the middle of an event, or at the end of an event. For the sake of our analysis, we will focus on the implicit relationship between the six major attributes in this study attractiveness (attr_o), sincerity (sinc_o), intelligence (intel_o), fun (fun_o), ambition (amb_o), and shared interests/hobbies (shar_o) and the overall likeability (like_o) of an individual.

Correlation Analysis

Before performing any in-depth analysis on the impact of the six attributes in the study on likeability, we first had to ensure there was no inherent high correlation between the independent variables. Any multicollinearity between the independent variables could skew the results of our analysis, as the independent variables may be a better predictor of each other than the dependent variable likeability. For this analysis, a high correlation is any value exceeding ±.7. Below is the result of the Excel correlation analysis:

	attr_o	sinc_o	intel_o	fun_o	amb_o	shar_o
attr_o	1					
sinc_o	0.399632	1				
intel_o	0.390744	0.657548	1			
fun_o	0.586334	0.491453	0.494097	1		
amb_o	0.359001	0.456194	0.627332	0.48957	1	
shar_o	0.480048	0.397943	0.400031	0.617995	0.43226	1

While there was a higher correlation between intelligence and sincerity, ambition and intelligence, and shared interest/hobbies and fun, there was no correlation exceeding the set value. Therefore, there is no multicollinearity between our independent variables. This reduces the likelihood of issues such as inflated p-values and sign swaps on coefficients, as well as minimizes any difficulties in isolating the effects of the independent variables.

Statistical Analysis

Hypothesis

In our analysis, we aimed to determine the nature of the relationship between the six independent variables attributes and the dependent variable, likeability. We devised the following hypothesis test to determine whether any one of our independent variables was statistically significant in determining the target variable:

$$\boldsymbol{H_0}: \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \boldsymbol{\beta_3} = \boldsymbol{\beta_4} = \boldsymbol{\beta_5} = \boldsymbol{\beta_6} = \mathbf{0}$$

$$\boldsymbol{H_1}: \boldsymbol{at\ least\ one\ \beta_j is\ not}\ \mathbf{0}$$

For the regression equation:

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon}$$

With likeability as Y and attractiveness as X_1, sincerity as X_2, intelligence as X_3, fun as X_4, ambition as X_5, and shared interests/hobbies as X_6.

This means that if we can reject the null hypothesis, at least one of the independent variables (X) is significant in determining the dependent variable (Y). For our analysis, used an alpha level of .05.

Multiple Linear Regression: Setup

One limitation of Excel's linear regression function is the lack of support for null values. Given the size of our dataset, we did not want to parse through the data to remove lines with null values depending on the analysis we ran. To counteract this, we used SAP Predictive Analytics to run our multiple linear regression model. An overview of our setup can be found in Appendix C.

In our overall setup, we had three identical regression models, one for all participants male and female, then two with additional filter nodes activated. These filters divided the dataset by females (0) only and males (1) only. We then ran the three separate regression models to compare the results. In each scenario, the alternative and null hypothesis was the same. After developing the regression model, we performed hypothesis testing, as well as analyzed whether each attribute alone was statistically significant in determining likeability. It is important to note that the variable used for the attributes (attribute_o), is how the subject is viewed by their partner. Therefore, filtering by male participants would yield the results for the female perspective, and vise versa.

Multiple Linear Regression: Combined Male and Female

First, we ran the multiple regression analysis for the full dataset, which included both male and female participants. The regression output line is as follows:

$$\hat{y} = .36695 + .26832x_1 + .10852x_2 + .0896x_3 + .25858x_4 - .05278x_5 + .25759x_6 + \varepsilon$$

As seen in the regression line above, each of the six attributes contributes positively to the overall likeability of an individual except for ambition, which has a negative slope. This means that ambition is seen as a detractor to overall likeability because for every one point granted towards ambition, the overall likeability is estimated to decrease by -.05 when all other variables remain constant. Conversely, attractiveness, fun, and shared interests each held a near equal weight in their positive influence on overall likeability. Using this model, we can estimate an individual's overall likeability rating based on the ratings he or she received from a partner. For example, an individual who was rated at a 5 across all six attributes would have an estimated likeability rating of 5.01.

To perform our hypothesis test, we had to determine whether the model significance, or the p-value, was less than our set alpha of .05. Per the results output in Appendix D, the p-value for the model was 2.2e-16, which is below our set alpha level .05 for a 95% confidence interval. Because the p-value of the model is less than alpha, we can reject the null hypothesis. This means that there is a relationship between likeability (Y) and at least of the one independent variables. Following our hypothesis testing, we determined whether each independent variable alone was significant in determining the target variable—likeability (like_o). The p-value of each of these independent variables had to also be below .05 to be considered statistically significant. As seen in the regression output, each of the independent variables proved to be significant in this model. In fact, most of these variables were significant even beyond our set confidence interval of .05. Additionally, our adjusted R^2 for this model was .6326, which means that 63.26% of the variation in the dependent variable is explained by the independent variables. This level of adjusted R^2 indicates a good model.

Multiple Linear Regression: Male Perspective

Given these results, we wanted to determine if there was a statistical difference between how males and females rate the likeability of a partner. To do this, we reran the regression using the filtering detailed in the setup section. This analysis yielded the regression line below:

$$\hat{y} = .53379 + .31743x_1 + .0934x_2 + .06134x_3 + .25255x_4 - .07877x_5 + .26677x_6 + \varepsilon$$

This model mirrors the overall signs of the combined model with each of the six attributes contributing positively to the overall likeability except for ambition, which has a negative slope. One notable difference in this model and the combined model is the comparatively high slope for X_1—attractiveness. This means that when all other variables are held constant, the rating for attractiveness has a higher impact on likeability for males over the combined model (.31743 for every 1 point of attractiveness for males over .26832 for every 1 point of attractiveness for combined). In addition, males had the highest intercept of any model, meaning the base level of likeability for males was higher than the combined and female models.

To perform our hypothesis test, we had to determine whether the model significance, or the p-value, was less than our set alpha of .05. Per the results output in Appendix D, the p-value for the model was 2.2e-16, which is below our set alpha level .05 for a 95% confidence interval. Because the p-value of the model is less than alpha, we can reject the null hypothesis. This means that there is a relationship between likeability (Y) and at least of the one independent variables. Following our hypothesis testing, we determined whether each independent variable alone was significant in determining the target variable—likeability (like_o). The p-value of each of these independent variables had to also be below .05 to be considered statistically significant. As seen in the regression output, intelligence had a p-value higher than our alpha level, meaning it is not statistically significant in determining likeability. If we were trying to build the best model, we would drop this variable from the model and rerun. Additionally, our adjusted R^2 for this model was .625, which means that 62.5% of the variation in the dependent variable is explained by the independent variables. This level of adjusted R^2 indicates a good model.

Multiple Linear Regression: Female Perspective

Following the male perspective analysis, we reran the regression using the filtering detailed in the setup section. This analysis yielded the regression line below:

$$\hat{y} = .19429 + .20298x_1 + .11413x_2 + .12433x_3 + .27828x_4 - .02598x_5 + .2507x_6 + \varepsilon$$

This model mirrors the overall signs of the combined model and the male model with each of the six attributes contributing positively to the overall likeability except for ambition, which has a negative slope. One notable difference in this model and the combined model is the comparatively low slope for X_1—attractiveness and a higher slope for X_3—intelligence . This means that when all other variables are held constant, the rating for attractiveness has a lower impact on likeability for females over the combined and male models, and intelligence had a higher impact on likeability for females over the combined and male models.

To perform our hypothesis test, we had to determine whether the model significance, or the p-value, was less than our set alpha of .05. Per the results output in Appendix D, the p-value for the model was 2.2e-16, which is below our set alpha level .05 for a 95% confidence interval.

Because the p-value of the model is less than alpha, we can reject the null hypothesis. This means that there is a relationship between likeability (Y) and at least of the one independent variables. Following our hypothesis testing, we determined whether each independent variable alone was significant in determining the target variable—likeability (like_o). The p-value of each of these independent variables had to also be below .05 to be considered statistically significant. As seen in the regression output, ambition had a p-value higher than our alpha level, meaning it is not statistically significant in determining likeability. If we were trying to build the best model, we would drop this variable from the model and rerun. Additionally, our adjusted R^2 for this model was .6435, which means that 64.35% of the variation in the dependent variable is explained by the independent variables. This level of adjusted R^2 indicates a good model.

Conclusions

Based on the results of our hypothesis testing above, we can draw specific conclusions about how males and females differ in judging the opposite sex in a first meeting. First and foremost, there is a statistically significant relationship between how an individual is viewed in relation to each of the six attributes and how likeable he or she is to the opposite sex. Beginning with the initial survey data (attribute1_1) in the descriptive statistics section, we can see that attractiveness is an important attribute for most participants, and intelligence is ranked second across the combined dataset. Per the visualizations in Appendix A, males rated attractiveness significantly higher than females as the most important attribute, whereas females placed more weight on intelligence and sincerity. Following each date, the participants had the opportunity to rank their partner based on a variety of attributes as well as how much they liked the individual overall. This likeability factor was intended to reflect how well liked the person was based on personality traits that were not included in the individual ranking systems, but we found that for males and females, attractiveness played one of the largest roles in determining this likeability measurement. Our findings were consistent with the top rankings for males but not for females from the pre-event survey. When determining overall likeability, males placed a notably high weight on how attractive their partner was, but females placed the most weight on fun and shared interests rather than intelligence. Furthermore, sincerity, which was ranked second by females in terms of importance, had the second lowest impact on a partner's likeability. This may be due to the difficulties in determining how sincere someone is after a short meeting. A similar issue may be the cause of the disparities in overall intelligence rankings for the combined dataset. Despite being ranked second highest in priority in the pre-event survey, it was the least important factor for likeability for the full dataset. Finally, both males and females view ambition as a negative factor in how likeable their partner is, despite it being given an average importance of 10 out of 100 in the pre-event survey.

Recommendations

While this data set was compiled for "speed dating", only 3.63% of individuals who participated were interested in a serious relationship and only 7.6% were interested in a date at all. This means that only 11.23% of participants—approximately one-tenth—were at the event with the goal of getting at least one date. The other 88.77% were at the event simply to meet new people, experiment, or have fun. Therefore, we can draw conclusions from this dataset beyond the context of dating, such as for general first impressions or short-term interactions with the opposite sex.

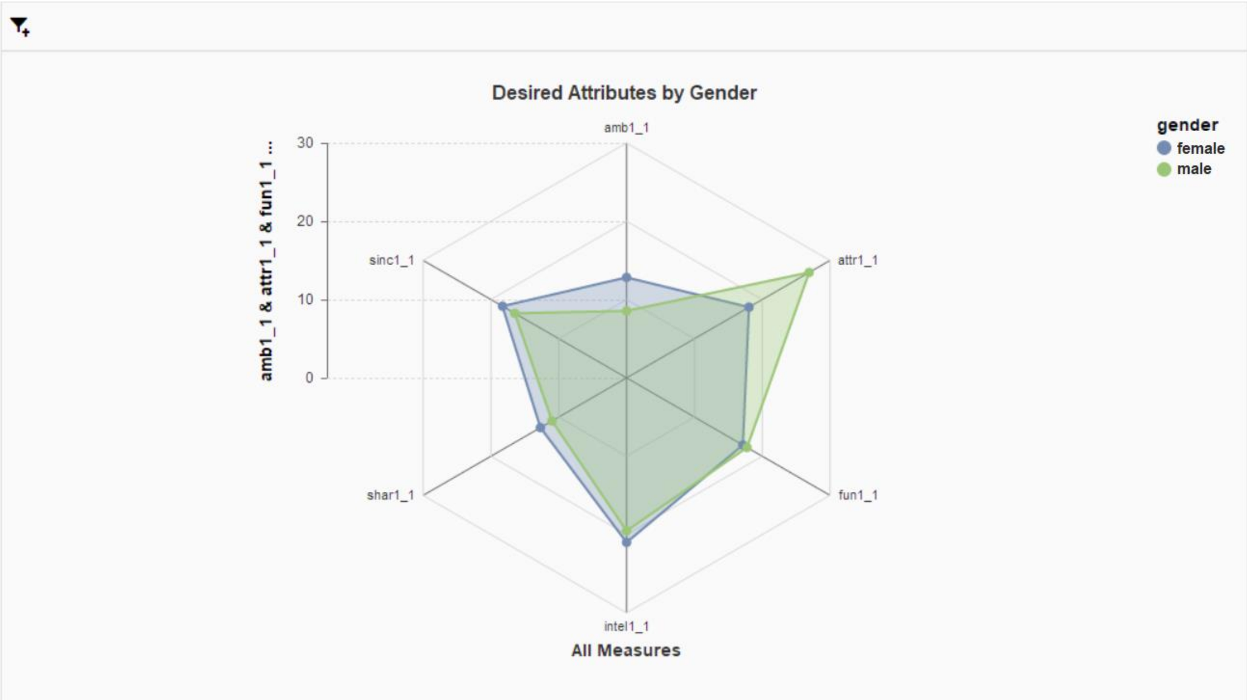
Based on the conclusions above, any time you would like to appear likeable to an individual of the opposite sex in a first meeting lasting less than ten minutes, it is important to focus on shared interests within the conversation. Both males and females view ambition as a detractor for likeability, so it is important to save conversations about long-term ambitions and goals for a later time when there is more time to talk. In a dating context, this would imply waiting for a second or third date to discuss topics that make you appear overly ambitious. Additionally, despite women rating appearance as less important than males when determining likeability, it is still one of the top factors across both genders. Therefore, no matter the meeting type or place, it is important to put thought and care into your appearance.

Further Analysis

Due to the nature of this project, our group had to limit the scope of our analysis to a single implicit relationship between attribute ratings and likeability, but the robust nature of this dataset would yield to a wide variety of additional studies. As detailed in the implicit relationships section, there is the potential for analysis on the relationship between how an important race was to an individual participant and whether they said yes or no on individuals of the same race. An exploration of this relationship could uncover subconscious bias on the part of the participant or give credence to the belief that individuals will misrepresent themselves on surveys if they believe their beliefs to be unpopular. Another potential analysis could be done on how career impacts the overall likeability or match rating for participants in the study. Do careers that are traditionally seen as more successful (e.g. law, medicine) play a role in how often the individual is matched? This analysis could prove more interesting due to the fact that our analysis concluded that ambition is viewed as a negative factor for overall likeability, and these career fields draw individuals with a higher level of ambition. Given the size of the dataset, the opportunities for analysis are only limited by the willingness to explore and learn from the data.

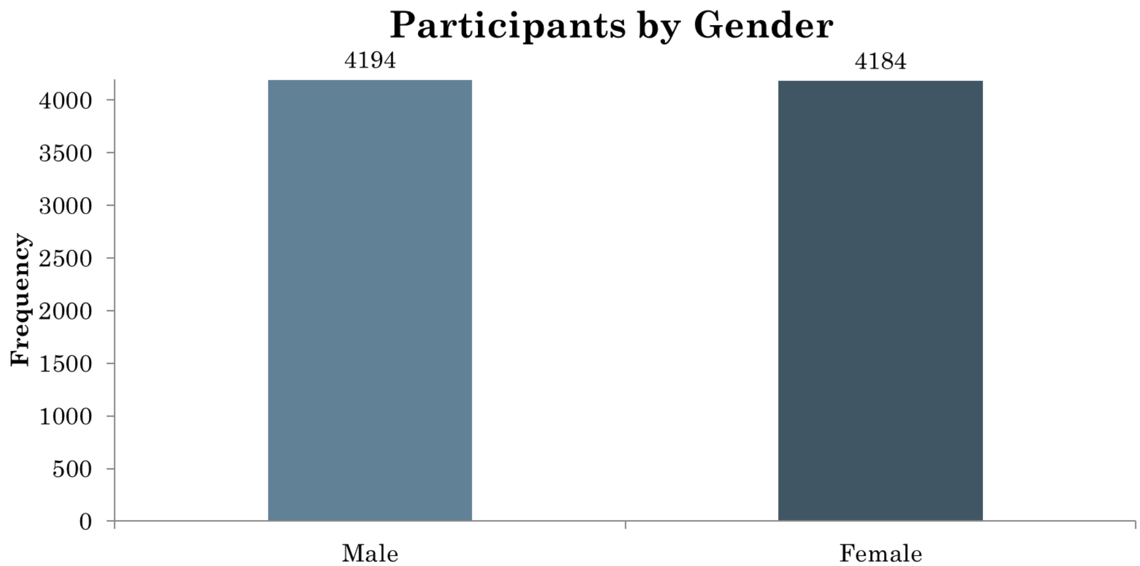
Appendix A

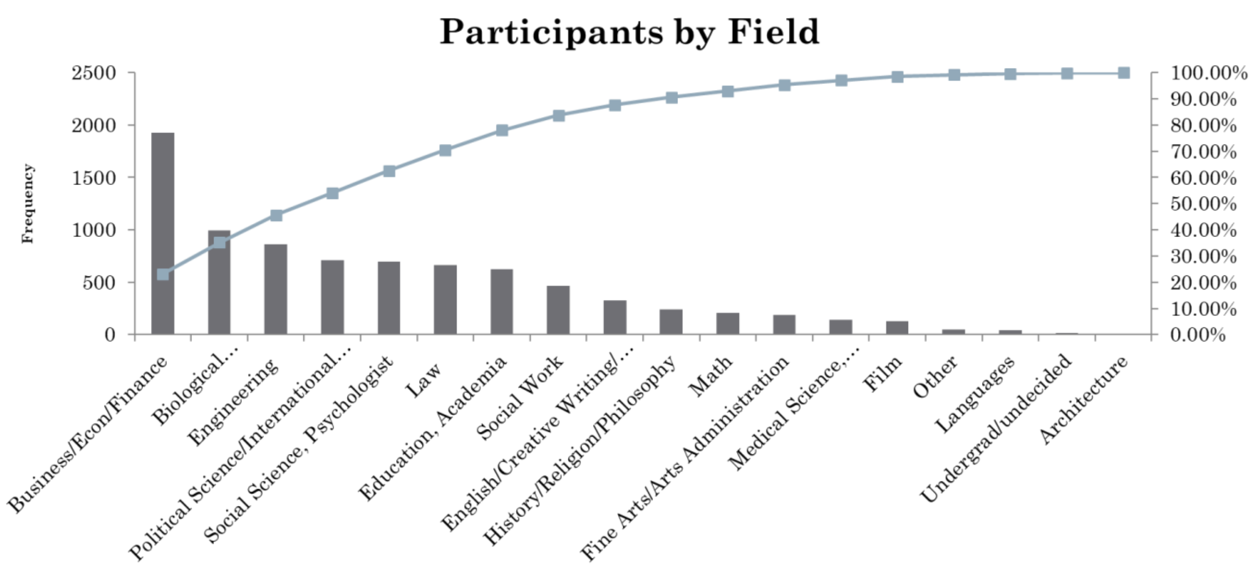
A display of the desired attributes (attribute1_1) by gender.

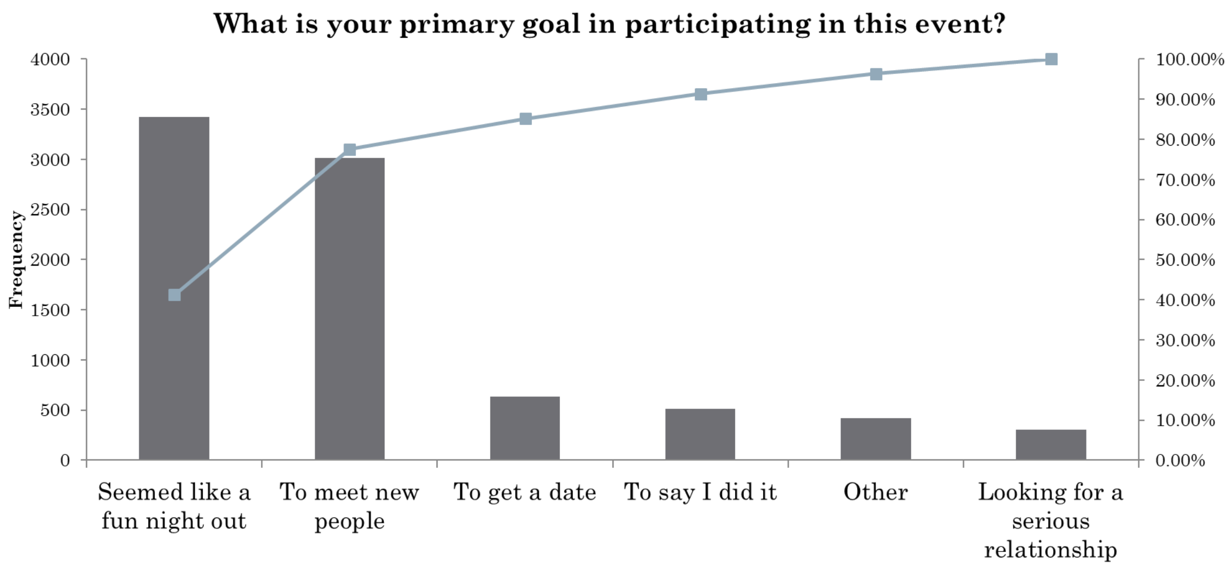


Appendix B

Additional descriptive statistics histograms for gender, field, and goal for attending event.

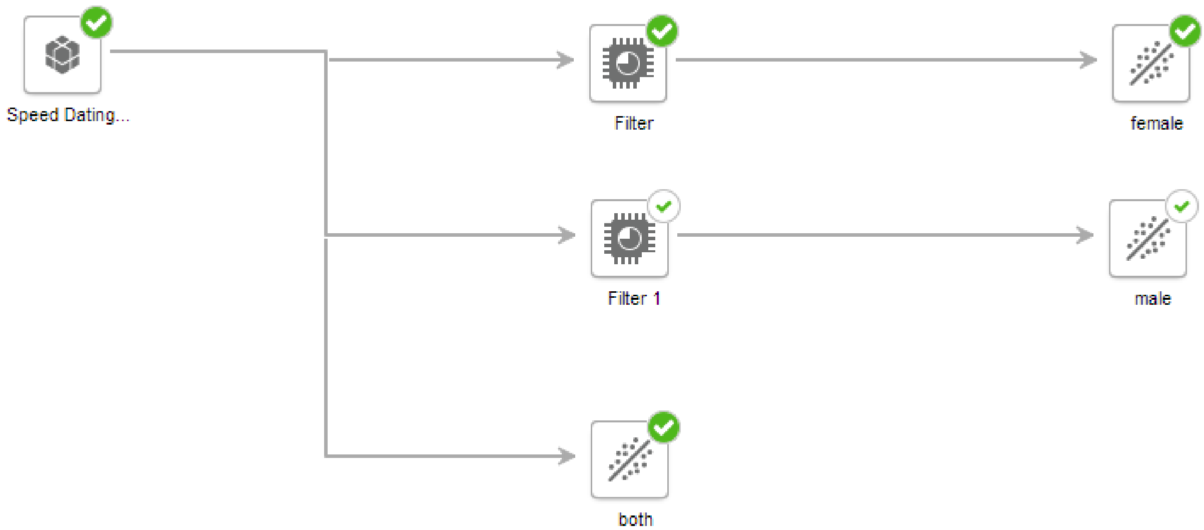





What is your primary goal in participating in this event?
Frequency
4000
3500
3000
2500
2000
1500
1000
500
0
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Seemed like a fun night out
To meet new people
To get a date
To say I did it
Other
Looking for a serious relationship

Appendix C

Model set up in SAP Predictive Analytics.



Appendix D

Regression Model for Combined, Males, and Females respectively

```
Summary of the Model

Call:
lm(formula = like_o ~ attr_o + sinc_o + intel_o + fun_o + amb_o + 
    shar_o, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.8703 -0.5371  0.0620  0.6335  3.2328 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.36695    0.13842   2.651 0.008102 ** 
attr_o       0.26832    0.01771  15.153  < 2e-16 ***
sinc_o       0.10852    0.02076   5.228 1.93e-07 ***
intel_o      0.08960    0.02600   3.446 0.000582 ***
fun_o        0.25858    0.02023  12.784  < 2e-16 ***
amb_o       -0.05278    0.01987  -2.657 0.007970 ** 
shar_o       0.25759    0.01578  16.322  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.106 on 1675 degrees of freedom
  (334 observations deleted due to missingness)
Multiple R-squared:  0.6339,	Adjusted R-squared:  0.6326 
F-statistic: 483.4 on 6 and 1675 DF,  p-value: < 2.2e-16
```

```
Summary of the Model

Call:
lm(formula = like_o ~ attr_o + sinc_o + intel_o + fun_o + amb_o + 
    shar_o, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.8349 -0.5089  0.0638  0.6217  3.1740 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.53379    0.19294   2.767  0.00578 ** 
attr_o       0.31743    0.02302  13.792  < 2e-16 ***
sinc_o       0.09340    0.02940   3.177  0.00154 ** 
intel_o      0.06134    0.03474   1.766  0.07777 .  
fun_o        0.25255    0.02730   9.249  < 2e-16 ***
amb_o       -0.07877    0.02705  -2.912  0.00368 ** 
shar_o       0.26677    0.02108  12.652  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.085 on 932 degrees of freedom
  (170 observations deleted due to missingness)
Multiple R-squared:  0.6274,	Adjusted R-squared:  0.625 
F-statistic: 261.6 on 6 and 932 DF,  p-value: < 2.2e-16
```

```
Summary of the Model

Call:
lm(formula = like_o ~ attr_o + sinc_o + intel_o + fun_o + amb_o +
    shar_o, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max
-4.4771 -0.5225  0.1011  0.6378  3.0260

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.19429    0.20013   0.971 0.331947
attr_o       0.20298    0.02772   7.322 6.42e-13 ***
sinc_o       0.11413    0.02977   3.834 0.000137 ***
intel_o      0.12344    0.03916   3.152 0.001687 **
fun_o        0.27828    0.03019   9.218  < 2e-16 ***
amb_o       -0.02598    0.02964  -0.877 0.380910
shar_o       0.25070    0.02379  10.538  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.125 on 736 degrees of freedom
  (164 observations deleted due to missingness)
Multiple R-squared:  0.6464,    Adjusted R-squared:  0.6435
F-statistic: 224.2 on 6 and 736 DF,  p-value: < 2.2e-16
```

Works Cited

Brown, Nell Porter. "The Road to Romance". Harvard Magazine (March–April 2003): 28–28A. https://web.archive.org/web/20061030035039/http://www.harvardmagazine.com/on-line/030382.html

Montoya, Anna. “Speed Dating Experiment.” Kaggle, 9 May 2016, www.kaggle.com/annavictoria/speed-dating-experiment.