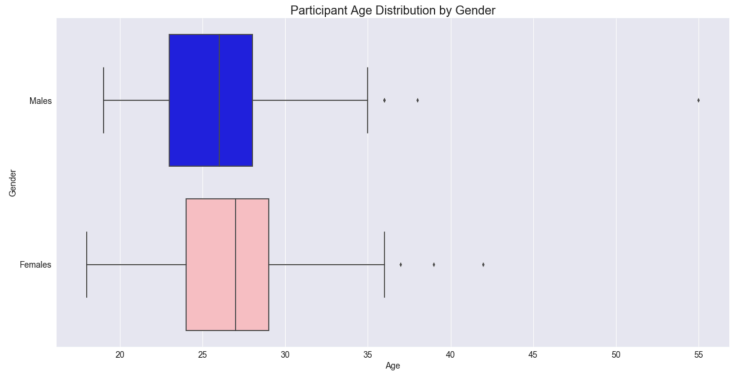
Speed Dating Experiment - Capstone Inferential Statistics

Jose Reyes – Dec. 2017
Springboard Data Science Career Track

This study aims to identify the variables that will predict a yes response from a speed-date partner, indicating that he/she will would like to see the participant again. The raw data contains 195 variables and 8,348 rows which covers data collected from each speed-date encounter, from both the male and female perspective, as well as data collected during various times intervals, namely prior to the speed-date event and in the days after. Because this study is focused around the data collected at the time of each speed-date, the revised data frame is comprised of basic demographics of the participants, namely age and race, their social and dating habits as well as the ratings received from each speed-date partner.

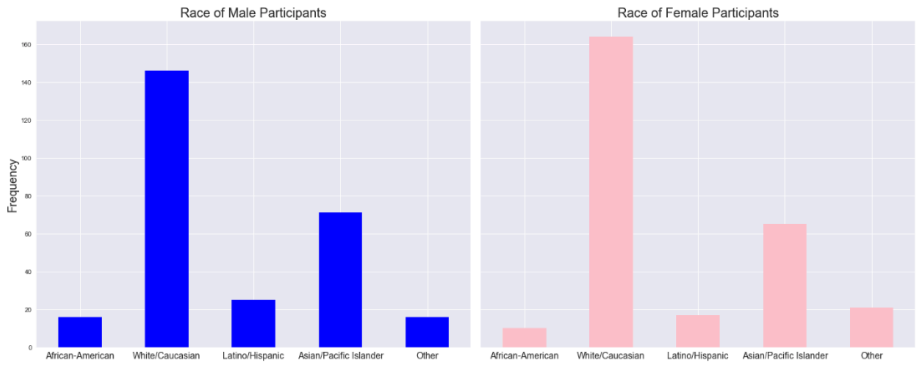
Demographics of the participants

In a total of 21 speed-dating events analyzed, the median age of all participants was 26 and the average age of all participants being 26.37 having a standard deviation of 3.74. The oldest participants identified was a 55-year-old male, assumed to be a graduate student, and the youngest participant was an 18-year-old female. Specifically, males had a median age of 26 and an average age of 26.14 with a standard deviation of 3.96 while females had a median age of 27 and an average age of 26.59 with a standard deviation of 3.50.



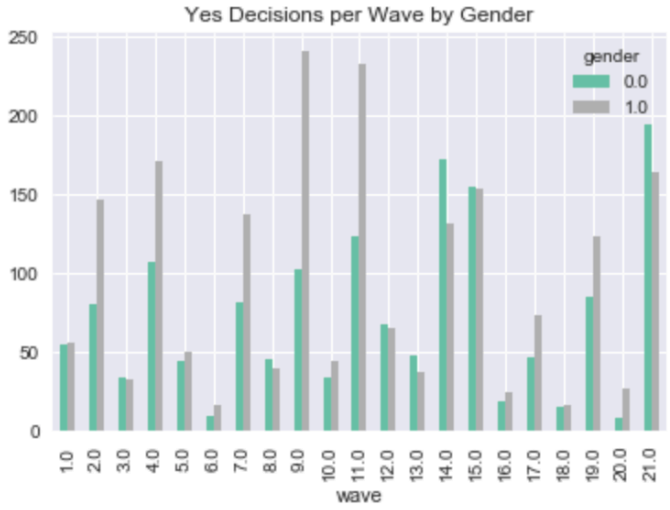
Collectively, the ages of the participants not normally distributed and are positively-skewed. The same was identified as true for males and females, respectively.

The majority of the participants identify themselves as white/Caucasian though other ethnicities are represented to a slight degree.



Speed-date outcomes

At the end of each 4-minute speed date participants must complete a 10-question scorecard rating their partner's various attributes such as level of attractiveness, sincerity and intelligence among others. The scorecard culminates with a yes/no decision of their date partner, indicating whether or not the participant would like to see them again. Interestingly males appear to be far more selective in their yes responses given that in 15 speed-dating events females made more yes decisions than their male counterparts. This goes against the traditional notion of evolutionary psychology which suggests that females are far more selective than their male counterparts as females pay a "greater reproductive cost by making a wrong choice."[1] However because participants are attending these events primarily for enjoyment and to meet new people, mate selection pressures bare little influence in making a yes response.

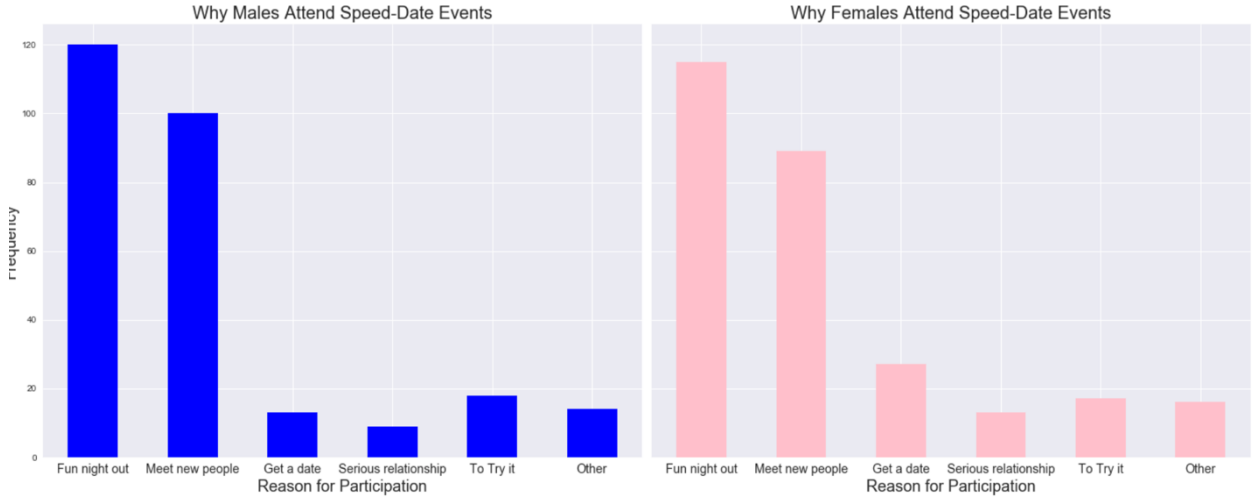


[1] https://www.psychologytoday.com/blog/the-scientific-fundamentalist/200908/are-women-always-more-selective-in-mate-choice-men-i

Yes responses by gender

wave	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Male	55	80	34	107	45	9	82	46	103	34
Female	56	147	33	171	50	16	138	40	241	45

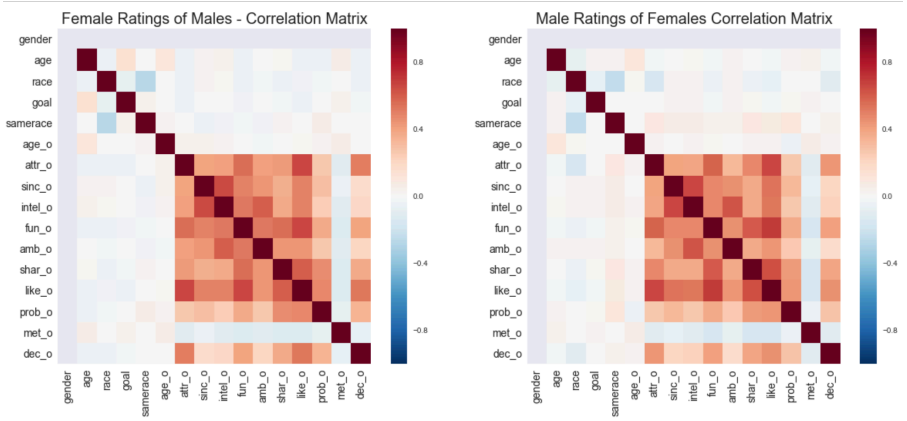
wave	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0	21.0
Male	123	68	48	172	155	19	47	15	85	8	194
Female	233	65	38	132	154	25	74	17	123	27	164



	Fun night out	Meet new people	Get a date	Serious relationship	To Try it	Other
Male	120	100	13	9	18	14
Female	115	89	27	13	17	16

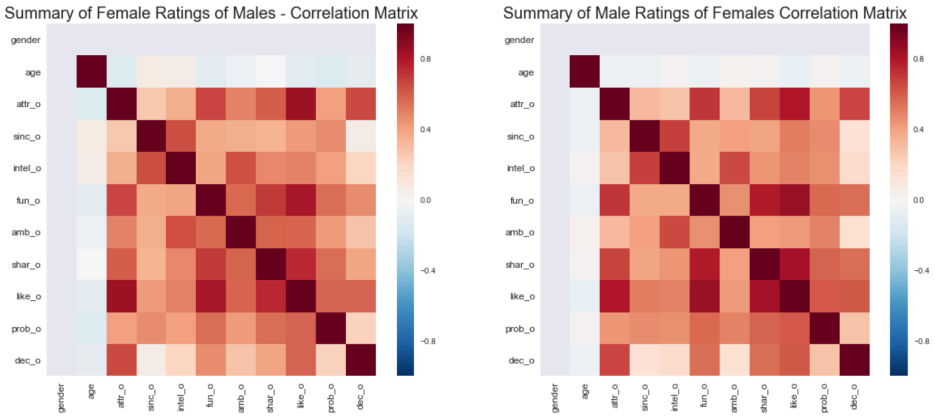
Correlation studies

The dataset was split into male and females for further analysis and for determining correlations between the **dec_o** variable, indicating a partner's yes/no response, and various other attributes. Surprisingly, participant demographics and motives for participation had no correlation with a yes/no response from their partner. Unsurprisingly however, at the speed-date encounter level, if a participant is deemed attractive and given a high "like" rating by their partner it correlates with a yes response. However, the correlations for attractiveness and like ratings of 0.44 and 0.45, respectively, from male participants and 0.51 and 0.52, respectively, from female participants are significant at best and not as strong as initially expected.



	age_o	attr_o	sinc_o	intel_o	fun_o	amb_o	shar_o	like_o	prob_o	met_o	dec_o
Ratings_F_of_M	-0.006218	0.514399	0.177262	0.200603	0.394551	0.205370	0.369519	0.527882	0.336398	-0.067802	1.0
Ratings_M_of_F	0.018818	0.440536	0.208706	0.222494	0.399644	0.167687	0.389185	0.458861	0.282691	-0.096556	1.0

Aggregating the average ratings for each participant and correlating it with the number of yes decisions received, a stronger correlation of the attractiveness and like ratings emerges. As expected, an additional attribute, **fun_o** (a partner's rating of the participant's level of fun) also emerges with a stronger correlation. The correlations for attractiveness, like, and fun ratings of 0.66, 0.61, and 0.55, respectively, from male participants and 0.65, 0.58, and 0.46 respectively, from female participants are stronger than the date-level correlations and will be considered as primary features in the predictive algorithm.



	gender	age	attr_o	sinc_o	intel_o	fun_o	amb_o	shar_o	like_o	prob_o	dec_o
Summary_Ratings_F_of_M	NaN	-0.089238	0.651337	0.058880	0.210556	0.464848	0.283609	0.386909	0.585656	0.233987	1.0
Summary_Ratings_M_of_F	NaN	-0.048572	0.662869	0.126484	0.156710	0.547985	0.147924	0.557225	0.610379	0.286299	1.0