Speed Dating Experiment

Chen Liang

Problem

With the vast development of the whole world, the speed of dating experience is also getting faster than ever, especially for people in their late 20s or 30s, whose spare time is pretty limited considering the very time-consuming fact of dating. So, whether you are more into casual or serious long-term relationships, leaving a good first impression in the first few minutes could be very important.

Data: The dataset is from Kaggle and used to be gathered from participants in an experimental speed dating events from 2002-2004

Potential clients: Single men or women, especially those who would like to find the partner through speed dating or online dating events

Purpose: Find insights about the different preference or thoughts between men and women, and also identify important attributes in the short dating phase

Three approaches for analysis

1. Exploratory data analysis to find the different behavior between men and women base on aggregated results
2. Clustering analysis for each gender group and find the decision/match rate between each cluster pair
3. Gaining insights about attribute importance by building random forest models

Exploratory data analysis

personal preference for 6 attributes by gender
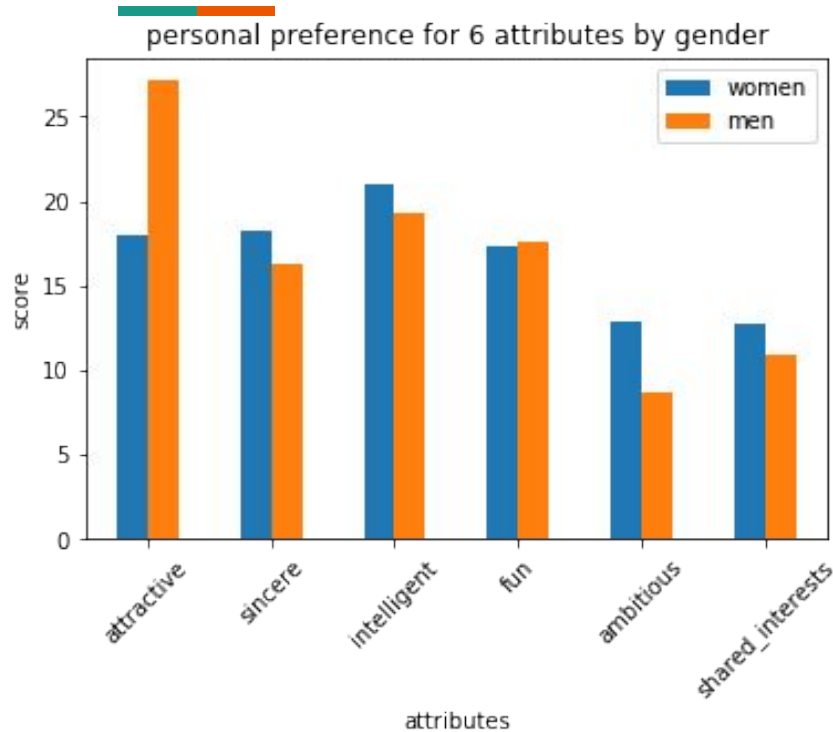
On average, men prefer attractiveness more than women, while women will compensate the outlook by ambitious, shared interests, sincerity and intelligence.

The top 3 valued attributes by male are: attractive, intelligent and fun

The top 3 valued attributes by female are: intelligent, attractive and sincere

Thoughts about fellows' preference by gender
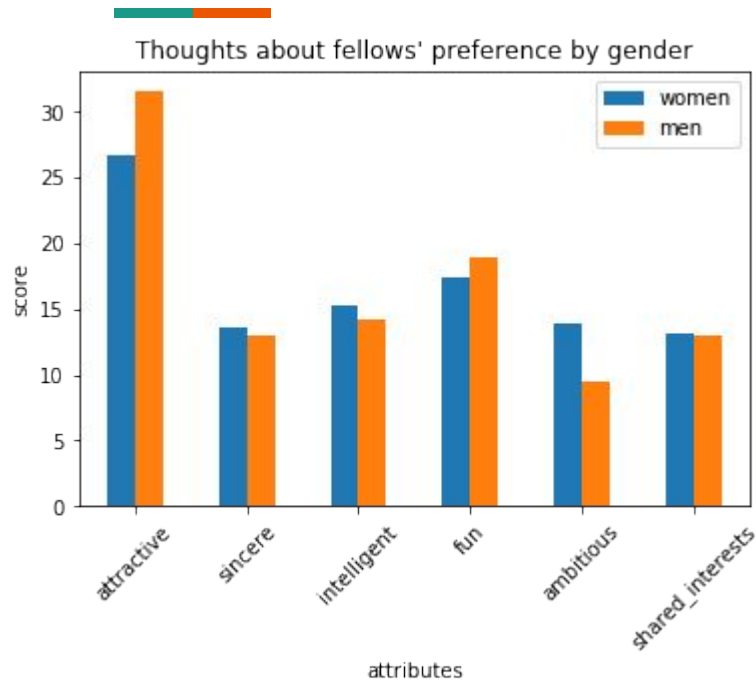
People seem to natively think of their fellow to value attractiveness much more the other attributes, even for women who are said to prefer attractiveness less than men are considered to value the outlook way more important than other attributes like sincere, intelligence by their fellow.

Thoughts about the opposite sex's preference in a date
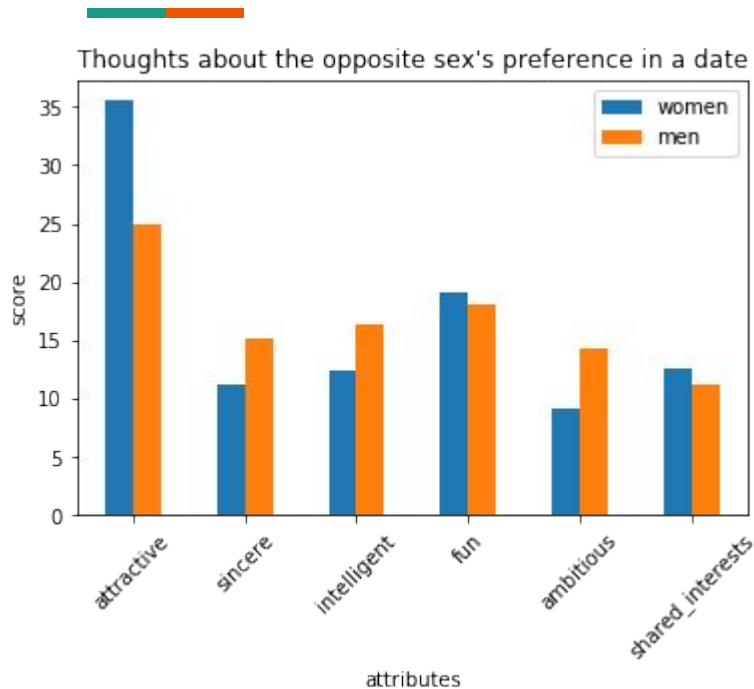
The comparison results between men and women are quite aligned to the first finding for personal preference, however the absolute scores for each attributes differ from the first one in that attractiveness keep getting overvalued here, while other attributes like intelligence and sincerity are overlooked.
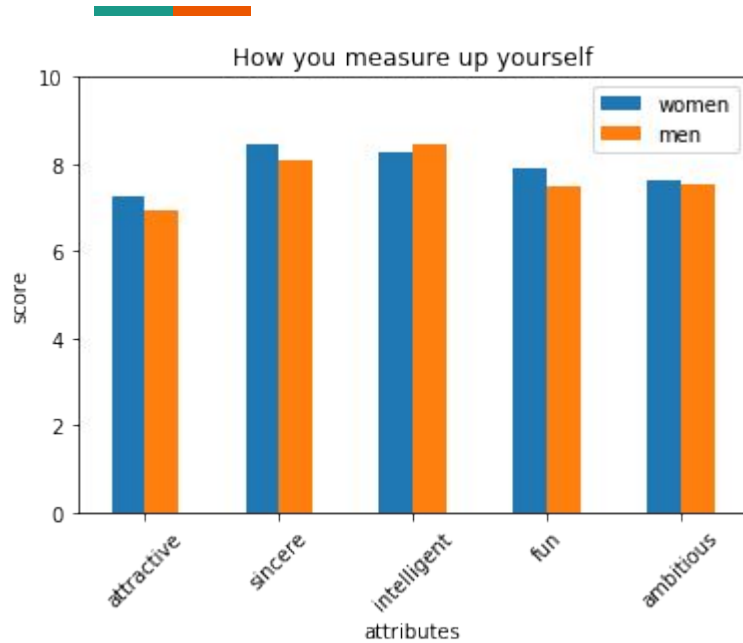
From the measurement result, we do not see any significant difference among the attributes between two genders, however on average women seem to be more confident on all the attributes (attractive, sincere, fun and ambitious) except (intelligence).
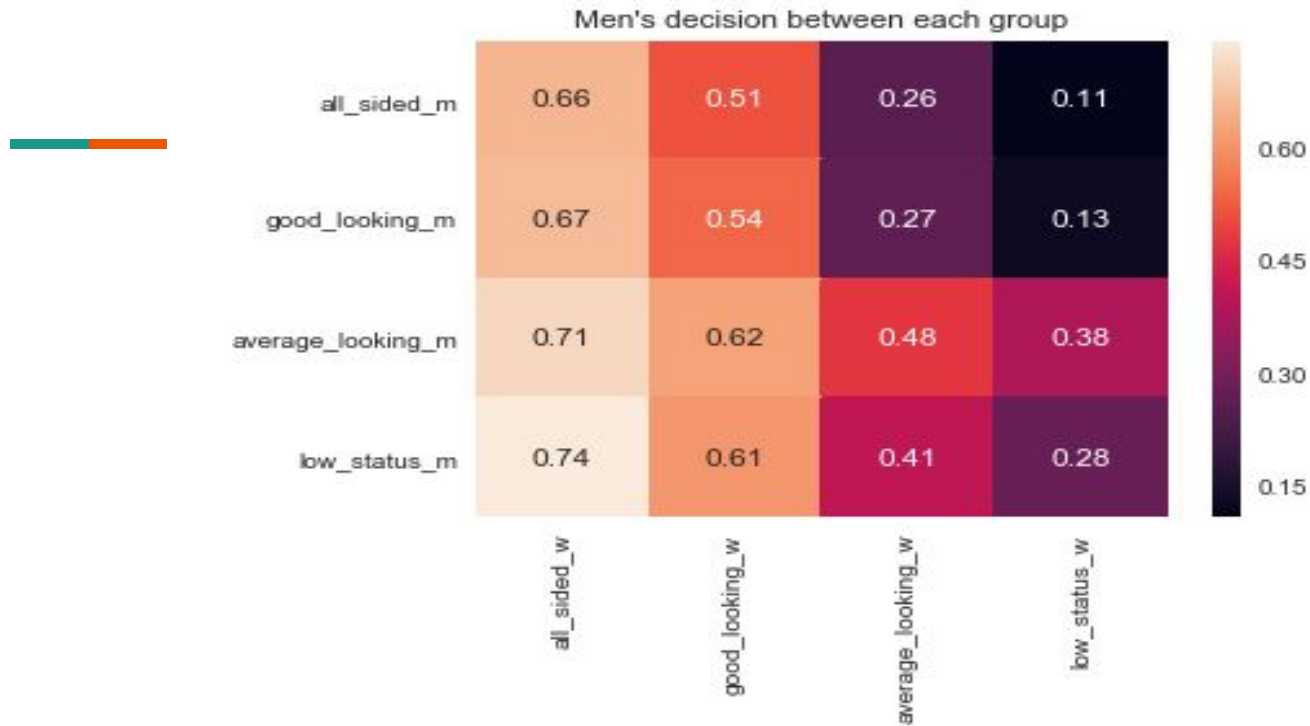
How do you think others perceive you
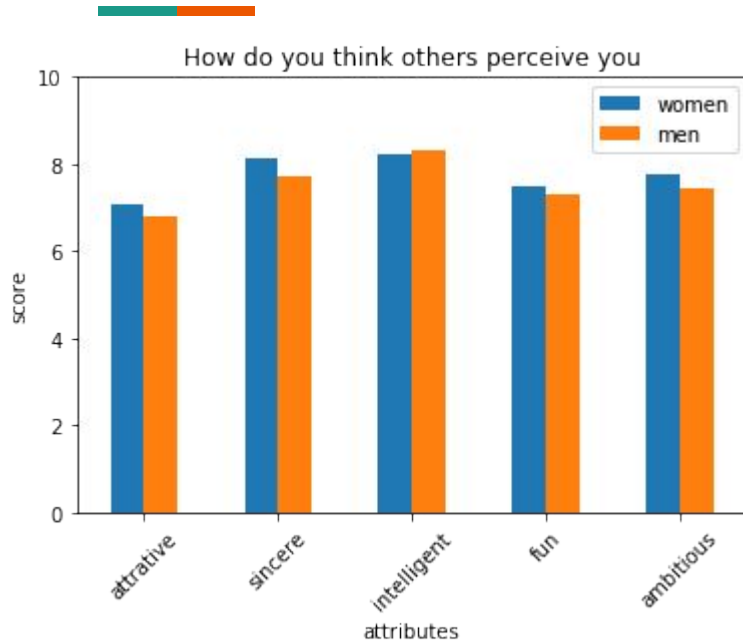
The average result is pretty close to the above result about their measurement about themselves.
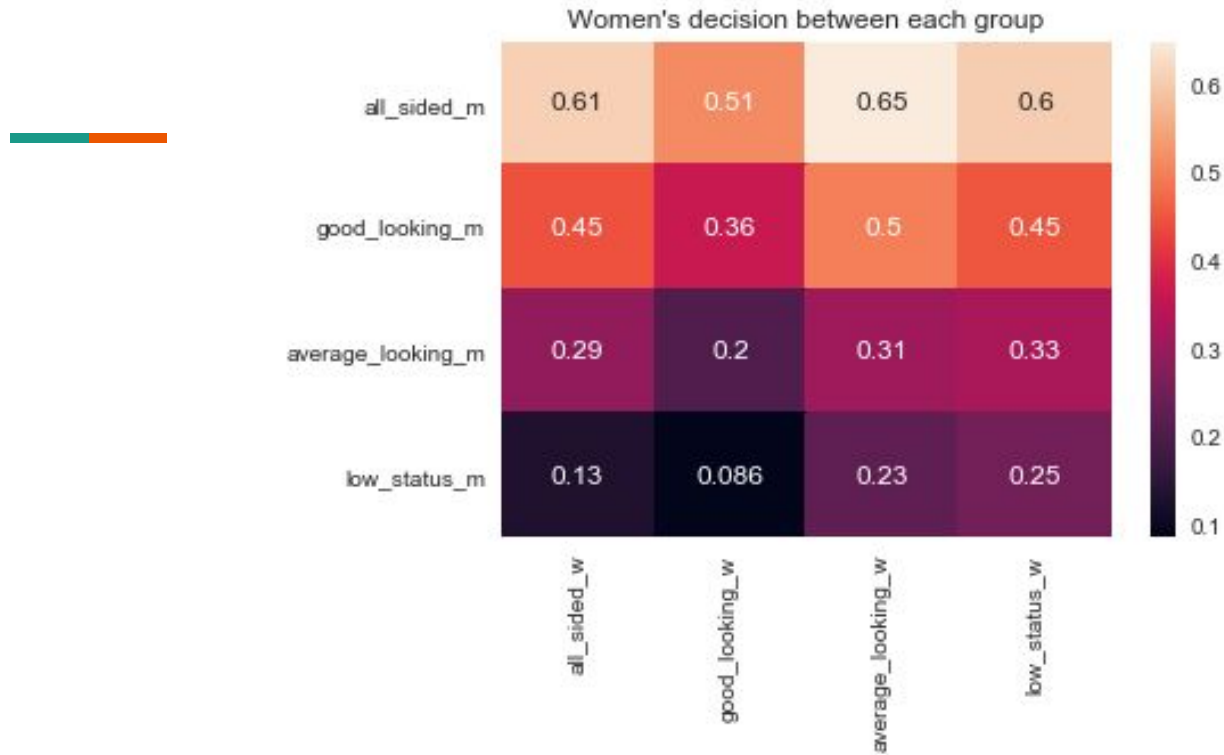
Clustering Analysis
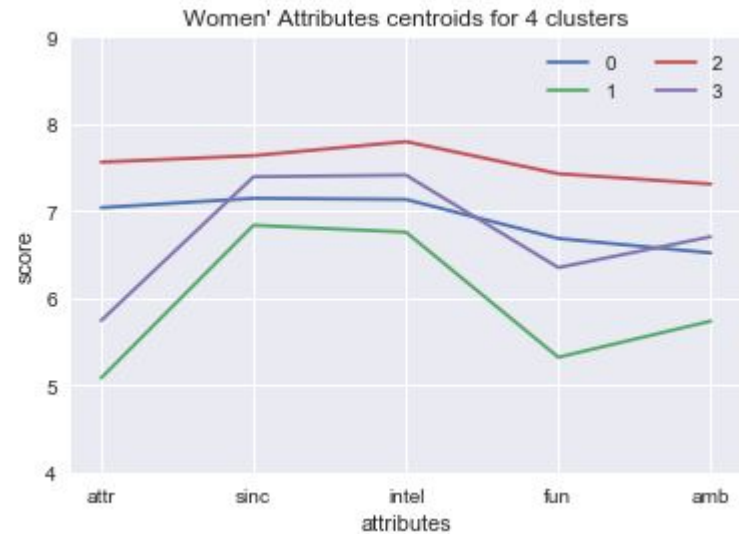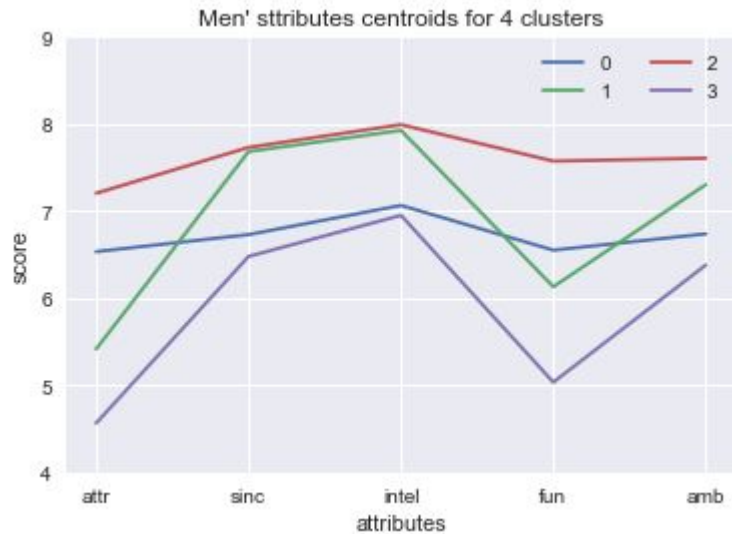
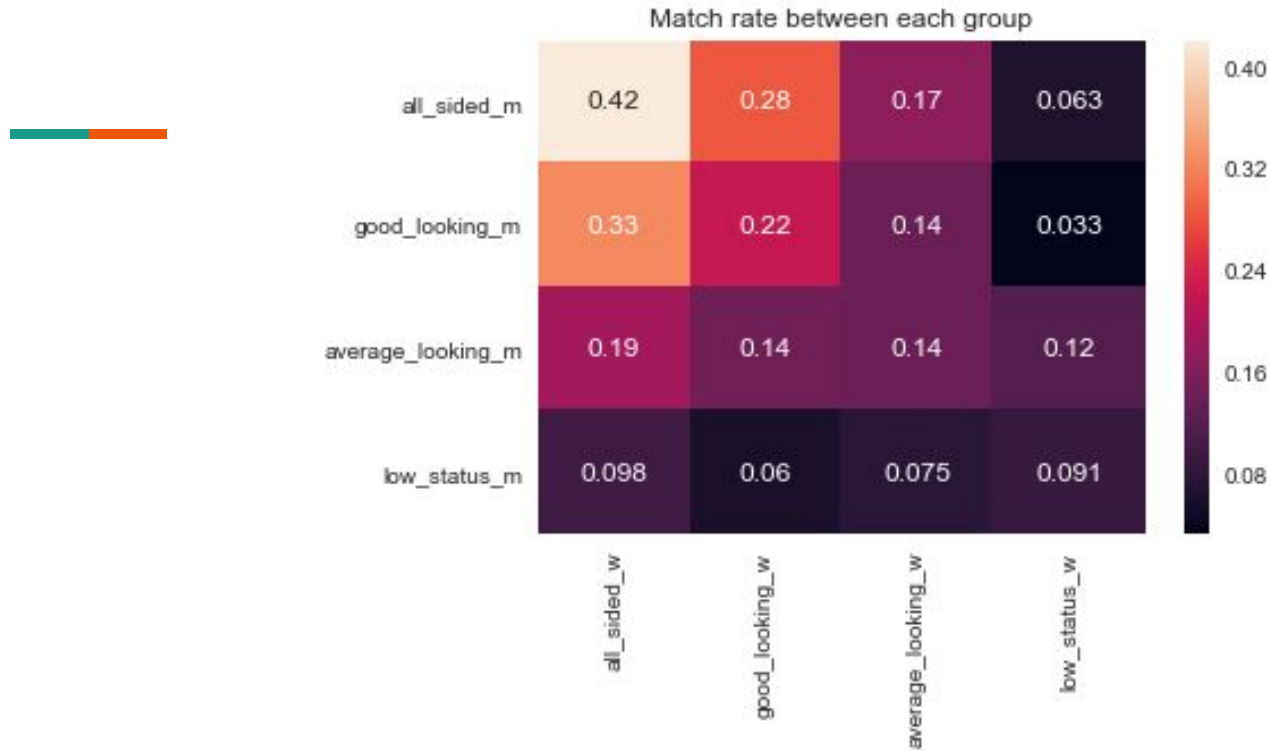For both men and women, we cluster them into 4 groups: all_sided, good_looking, average_looking, low_status

Men's decision between each group

	all_sided_w	good_looking_w	average_looking_w	low_status_w
all_sided_m	0.66	0.51	0.26	0.11
good_looking_m	0.67	0.54	0.27	0.13
average_looking_m	0.71	0.62	0.48	0.38
low_status_m	0.74	0.61	0.41	0.28

The above heatmap shows a rather clear pattern that generally the less popular group will give a higher decision rate for the corresponding more popular group.

Women's decision between each group

	all_sided_w	good_looking_w	average_looking_w	low_status_w
all_sided_m	0.61	0.51	0.65	0.6
good_looking_m	0.45	0.36	0.5	0.45
average_looking_m	0.29	0.2	0.31	0.33
low_status_m	0.13	0.086	0.23	0.25

Still a general preference for most popular/attractive male group, but the different response given among the female group is not as distinct as male.

Match rate between each group

	all_sided_w	good_looking_w	average_looking_w	low_status_w
all_sided_m	0.42	0.28	0.17	0.063
good_looking_m	0.33	0.22	0.14	0.033
average_looking_m	0.19	0.14	0.14	0.12
low_status_m	0.098	0.06	0.075	0.091

The heatmap for match rate also shows clear pattern no matter column wise or row wise. The most popular/attractive group will have a higher match rate with any group in the opposite gender.
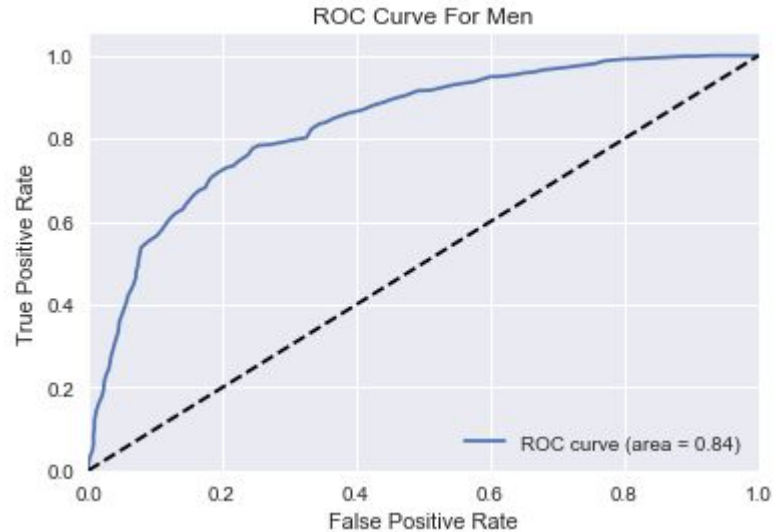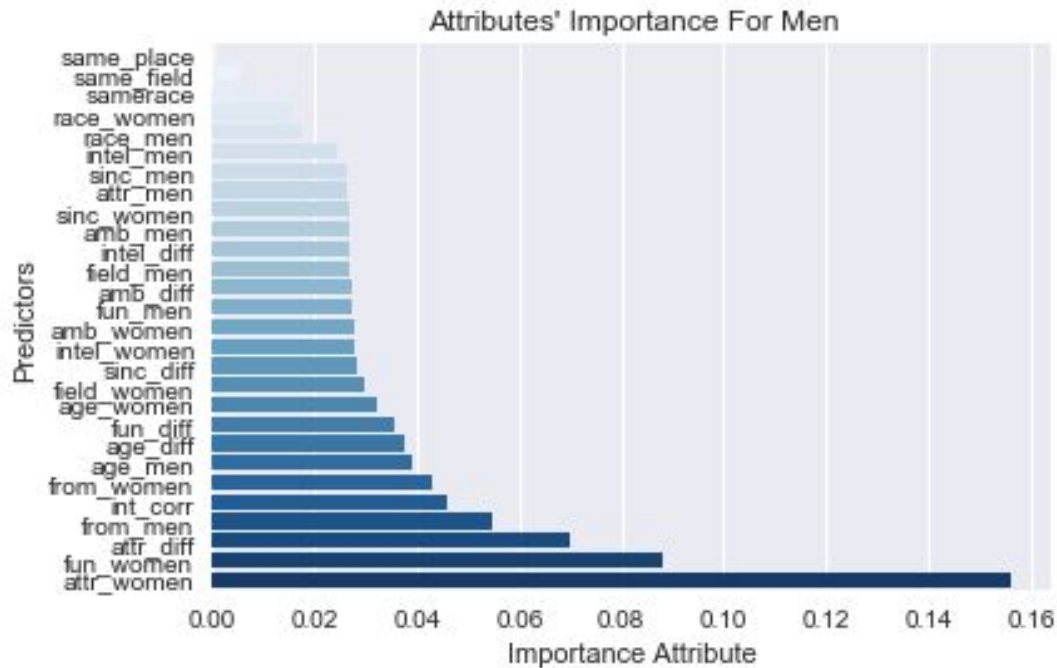
Random Forest Models

Model for men's decision

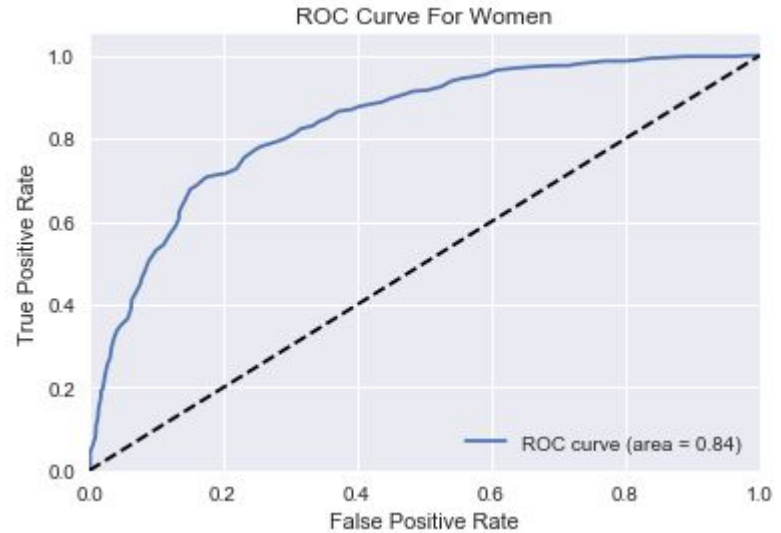OOB error: 0.76468579235

Test error: 0.759554140127

Attributes' Importance For Men

Top 3: women's attractiveness, fun and attractive difference between two people
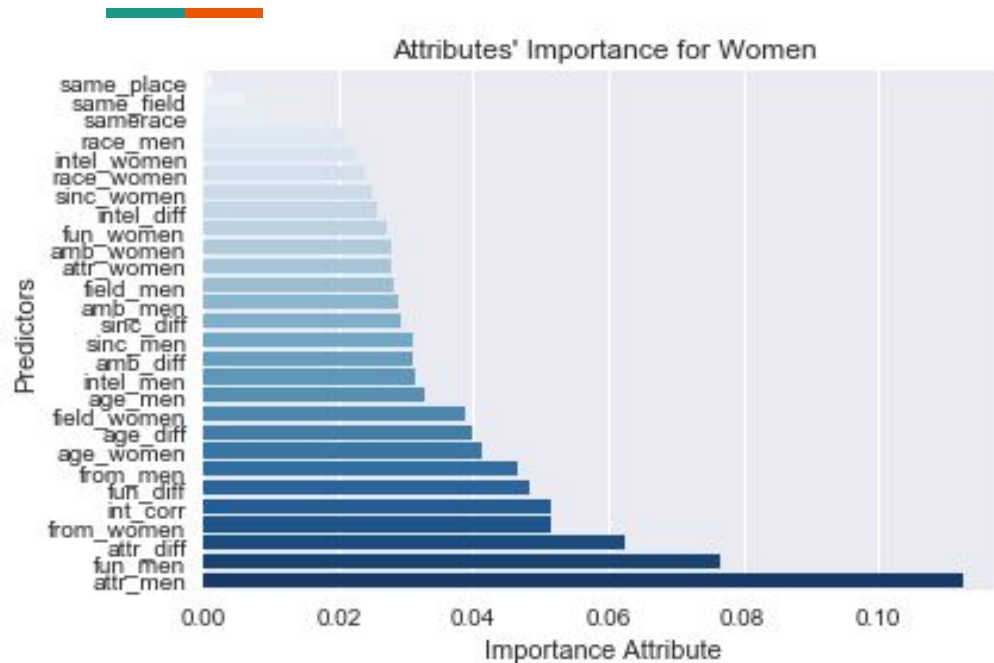
Bottom 3: same_place, same_field and same_race

Model for women's decision

OOB error: 0.749658469945

Test error: 0.778662420382
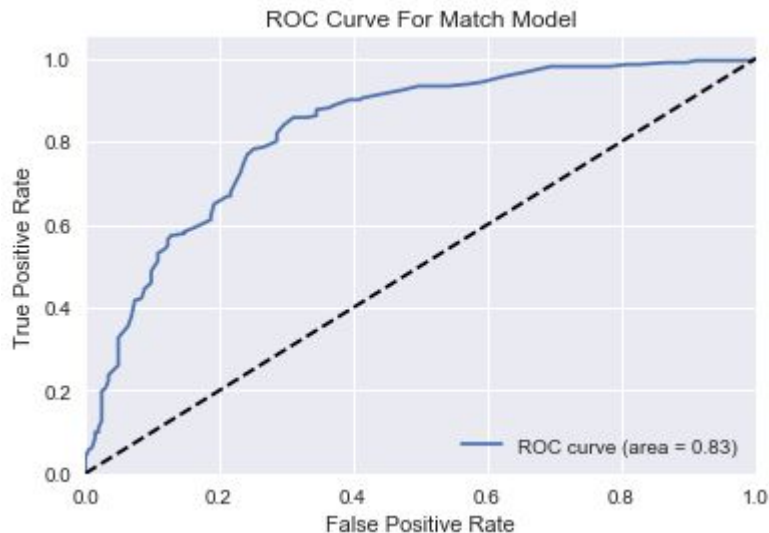

ROC Curve For Women

Attributes' Importance for Women

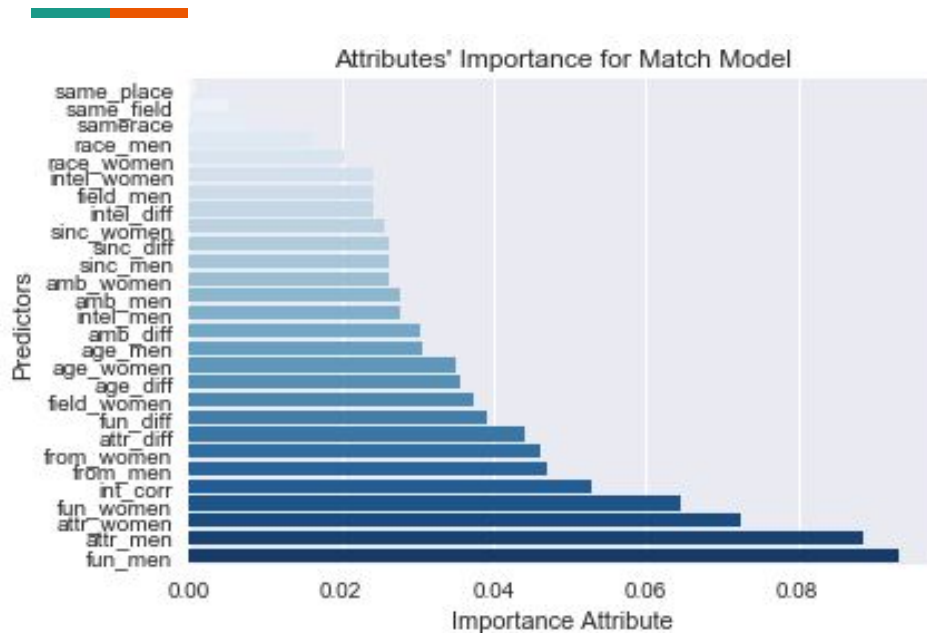Top 3: men's attractiveness, fun and attractive difference between two people

Bottom 3: same_place, same_field and same_race

Model for match rate (after downsampling)

OOB error: 0.749482401656

Test error: 0.729468599034

ROC Curve For Match Model

Attributes' Importance for Match Model

Top few: fun_men, attr_men, attr_women, fun_women, int_corr, from_men and from_women

Bottom 3: same_place, same_field and same_race

Summary

1. Attractiveness and fun are still the two most important attributes people look for during the dating period.
2. If people come from the same place, share same study background or race are not important as expected
3. The exact place where men or women come from individually also plays an important role for the final decision.
4. For the final match purpose, it seems men's attributes (attractiveness, fun) are more important that for women.