LOVE AT SECOND SIGHT

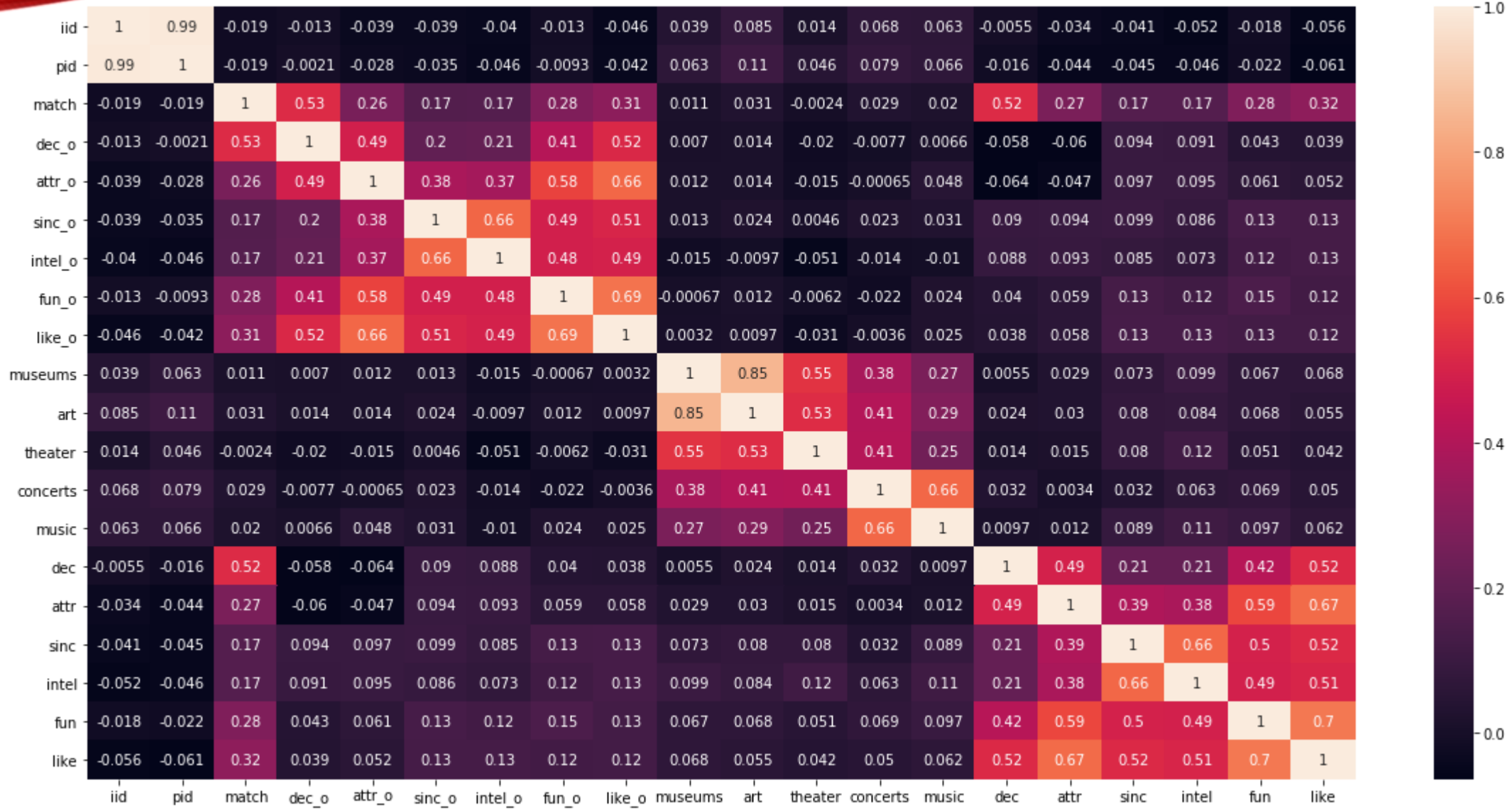
What are the key factors during a date that maximize your chances?

A brief insight into the Speed Dating Dataset from Kaggle.

THE DATASET

- Downloaded from Kaggle
- Containing 195 variables, describing the date, the people involved and their preconceptions prior to the evening.
- After filtering the dataset so that the variables left have a >50% correlation to another variable we arrived at the correlation matrix on the next page.

Variables whom have at least a 50% correlation with at least 1 other variable



OUR EXPLANATORY VARIABLES

After filtering variables again to having >10% correlation to a successful date (our 'match' variable) we are left with the following:

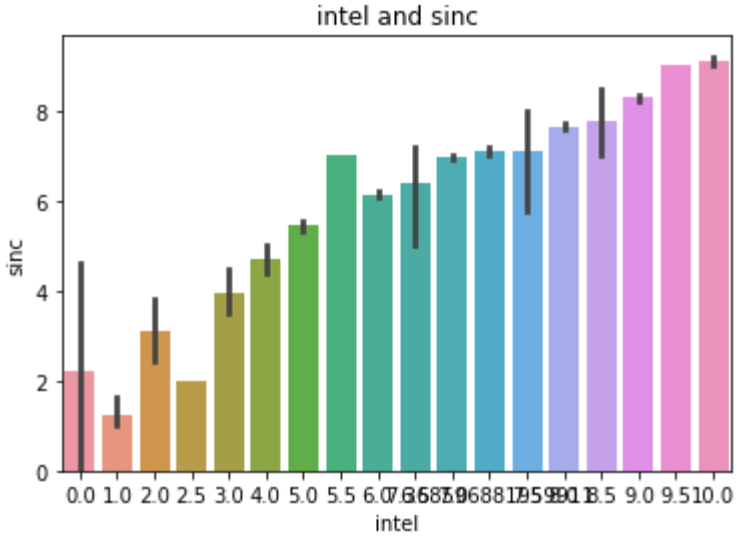
match: 1.000000 – whether both daters decide to match
dec_o: 0.525080 - Decision of partner at end of night,
attr_o: 0.263869 - Rating by partner at night of event on attractiveness,
sinc_o: 0.167217 - Rating by partner at night of event on sincerity,
intel_o: 0.169106 - Rating by partner at night of event on intelligence,
fun_o: 0.278102 - Rating by partner at night of event on being funny,
like_o: 0.312442 - Rating by partner at night of event on being liked,
prob_o: 0.260750 - Rating by partner at night of event on how much they think you liked them,
met_o: -0.114970 - Has the partner met you before,
dec: 0.524295 - Decision at end of night,
attr: 0.265760 - Rate yourself - attractiveness,
sinc: 0.172255 - Rate yourself - sincerity,
intel: 0.171725 - Rate yourself - intelligence,
fun: 0.280478 - Rate yourself - being funny,
like: 0.315384 - Did you like your partner?,
prob: 0.265932 - Rating at end of night of event on how much you think they liked you.

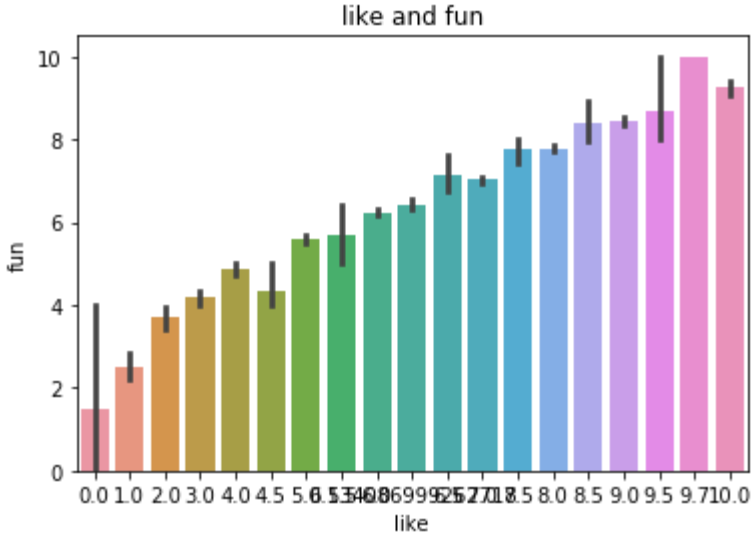
EXPLANATORY VARIABLES

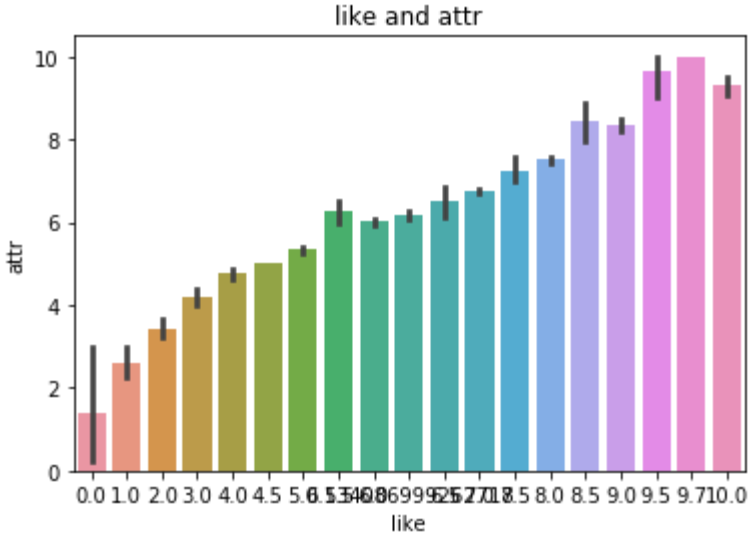
- After looking to see any link between these variables, I found that they all had a positive correlation with each other, i.e. rating someone more attractive means your more likely to rate them highly as sincere or fun. Positive traits complement each other. Some more than others.
- Liking the person more had a very strong link with all other variables
- On the other hand, the rating of probability that they like you has a much weaker link alongside the variables.
- The top 5 strongest links were the following:

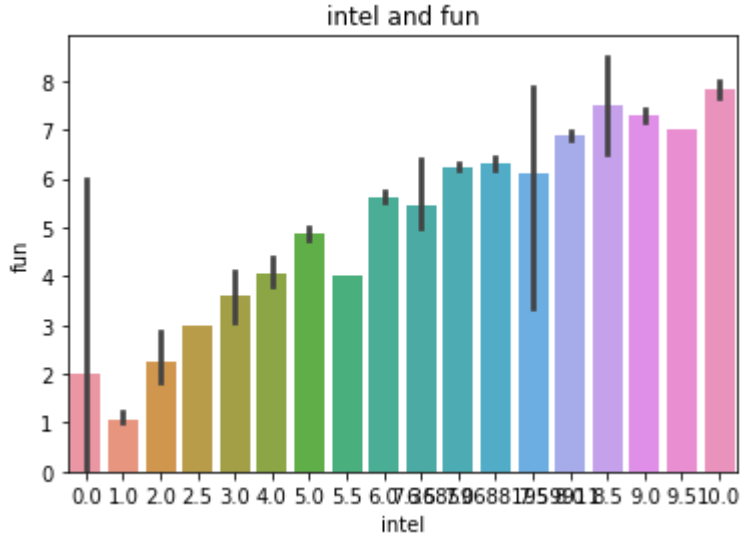
Note: the 1st variable noted has been plotted on the x axis, the 2nd on the y and the number following is the coefficient of a linear regression between the two.

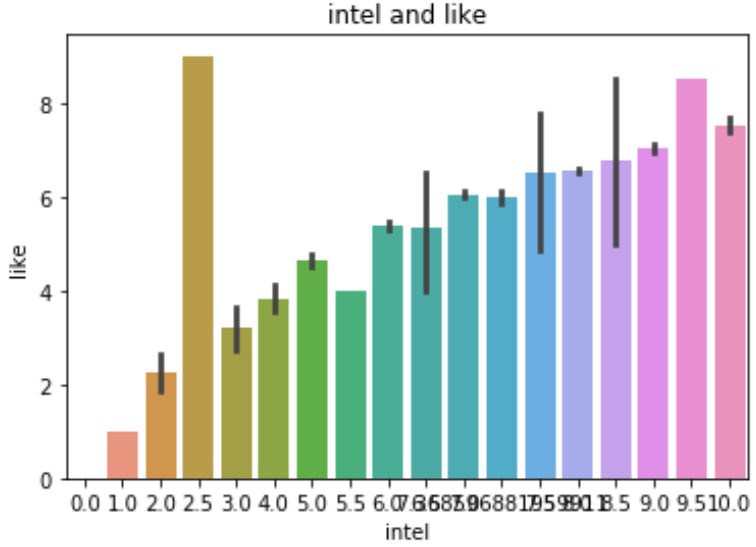
Intelligence and sincerity – 0.73, Liking and finding the person funny – 0.72, Liking and attraction – 0.71, intelligence and finding the person funny – 0.61, intelligence and liking – 0.60


intel and sinc
sinc
intel


like and fun
fun
like

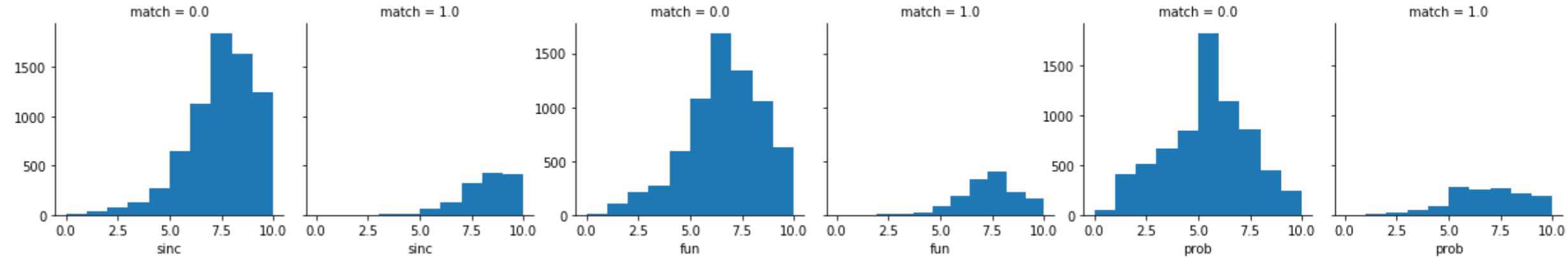

like and attr
attr
like


intel and fun
fun
intel


intel and like
like
intel

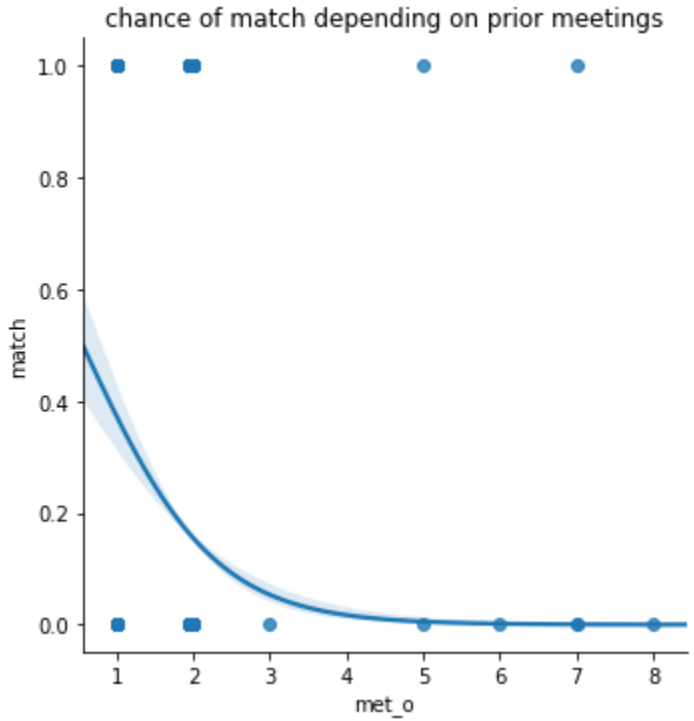
INTERESTING CORRELATIONS

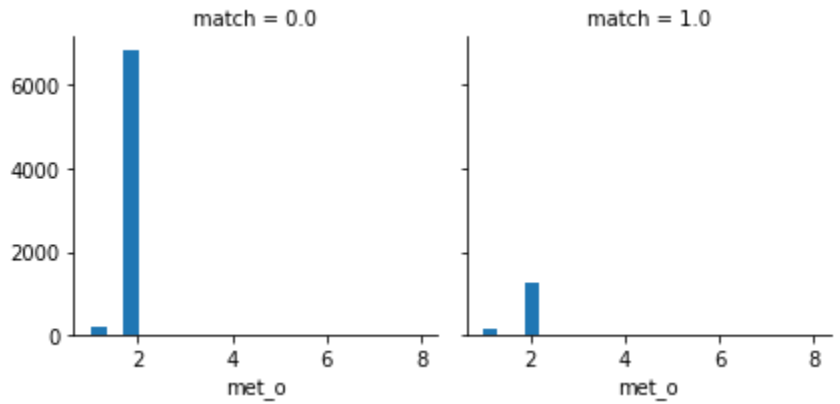
- Even though there seemed to be no strong links in a linear plot of our variables vs success, most followed the intuitive distribution plot, where only higher ratings of personal attributes lead to a match.
- However, most variables (concerning success) had a positive skew, showing people in general vote nicely. The clearest skew being sincerity and intelligence, whilst liking the partner and finding the partner fun had slightly less skew. Attraction even less so, and the probability that you think your partner liked you, had very little skew.



INTERESTING CORRELATIONS

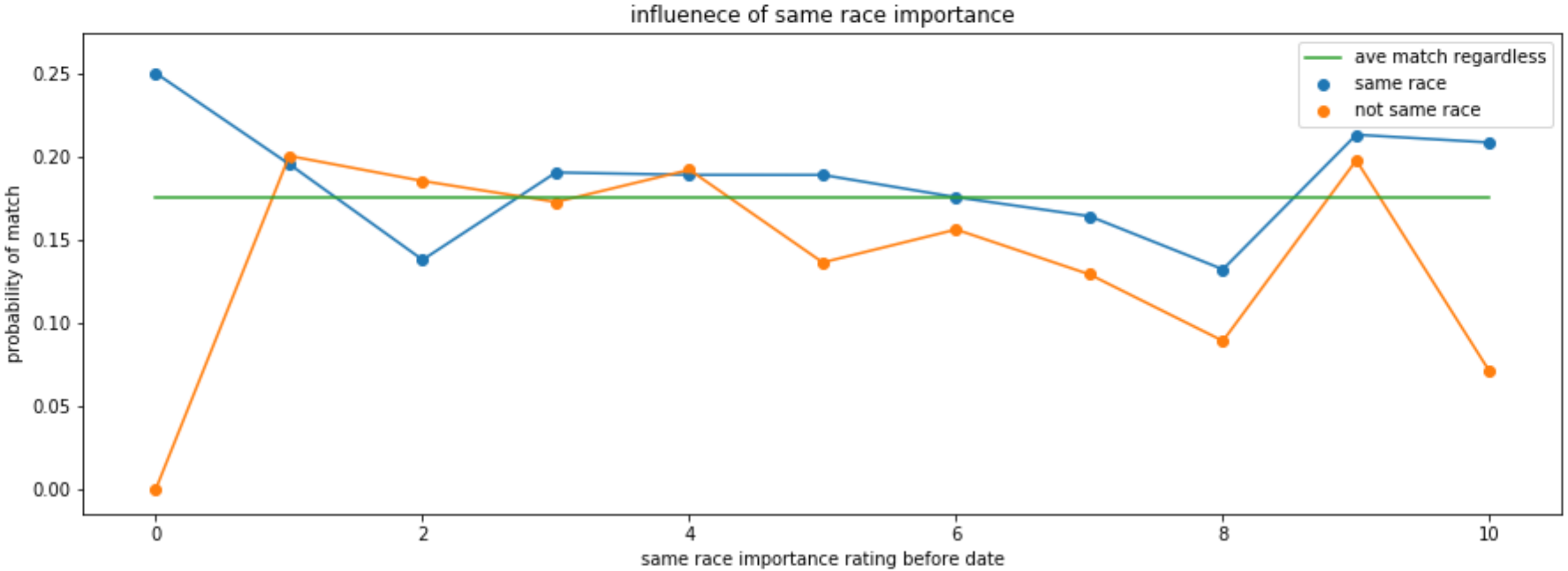
- The met variable – how many times the people have met before the date, is the only variable showing a negative correlation. The more you meet someone platonically or otherwise, the less likely the date is going to be successful.





DIFFERENT ANGLE

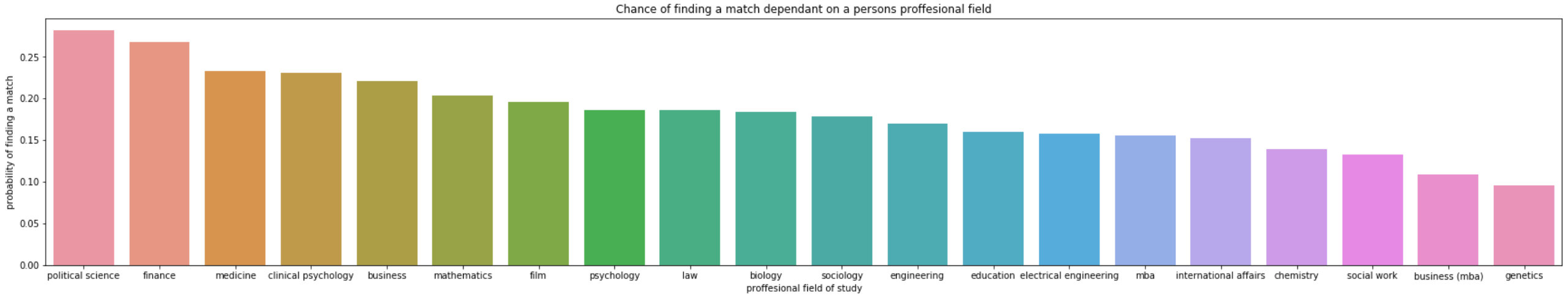
- Rather than just looking at the strongest variables linked to a match, I wanted to know whether people's preconceptions were 'real', as in affected their chance of a match.
- A variable named imprace, was a rating out of ten on how much the person valued dating someone of the same race
- The binary variable samerace was whether the date was between two people of the same race.
- Working out the probabilities of the match per group (i <= imprace < i + 1) for both cases of samerace, we see the line plot on the next page.


influenece of same race importance
ave match regardless
same race
not same race
probability of match
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
same race importance rating before date

- Oddly, no data was found for partners with a race importance rating of 0 who dated a different race.
- We can see, a rating of 10 has a very clear difference in whether the date predicts a match.
- Interestingly, a 9 rating of preferring same race lead to an increase in match probability amongst a different race date. This seems to be an outlier or a magical love number.
- Otherwise the only noteworthy trend is the higher the rating/ preference, the less likely they are to find a match. Picky people being picky. Unless they really care (9 or 10 rating) and are dating someone of the same race.

PROFESSIONAL FIELD

- Looking whether people from specific fields are more likely to match, and sorting from the top down (only accepting fields that at least 50 people are in).



CONCLUSION

- To summarize: be intelligent, be funny, be attractive, be in political science and don't meet the person before. You'll rate people higher than you should (from a normal distribution point of view) and only really be objective when positive and negative connotations are out the window. Either that or you're preparing yourself for reality.
- Happy dating! Use these secrets wisely ;)