

Descriptive Statistics:

	gender	age	age_o	samerace	attractive_o \
count	978.000000	978.000000	978.000000	978.000000	978.000000
mean	0.490798	25.125767	24.914110	0.415133	6.078221
std	0.500171	3.286013	3.213136	0.492997	1.915371
min	0.000000	18.000000	18.000000	0.000000	0.000000
25%	0.000000	23.000000	22.000000	0.000000	5.000000
50%	0.000000	25.000000	25.000000	0.000000	6.000000
75%	1.000000	27.000000	27.000000	1.000000	7.000000
max	1.000000	35.000000	35.000000	1.000000	10.000000

	sincere_o	intelligence_o	funny_o	ambition_o	shared_interests_o \
count	978.000000	978.000000	978.000000	978.000000	978.000000
mean	7.093047	7.292434	6.164622	6.692229	5.256646
std	1.720647	1.528261	2.021632	1.823412	2.185965
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	6.000000	6.000000	5.000000	6.000000	4.000000
50%	7.000000	7.000000	6.000000	7.000000	5.000000
75%	8.000000	8.000000	8.000000	8.000000	7.000000
max	10.000000	10.000000	10.000000	10.000000	10.000000

	intelligence_partner	funny_partner	ambition_partner \
count	978.000000	978.000000	978.000000
mean	7.379346	6.260736	6.743354
std	1.491107	1.983172	1.773968
min	0.000000	0.000000	0.000000
25%	6.000000	5.000000	6.000000
50%	7.000000	6.000000	7.000000
75%	8.000000	8.000000	8.000000
max	10.000000	10.000000	10.000000

	shared_interests_partner	interests_correlate \
count	978.000000	978.000000
mean	5.302658	0.158067
std	2.035110	0.335550
min	0.000000	-0.630000
25%	4.000000	-0.097500
50%	5.000000	0.160000
75%	7.000000	0.427500
max	10.000000	0.900000

	expected_num_interested_in_me	expected_num_matches	like \
count	978.000000	978.000000	978.000000
mean	5.663599	2.814928	6.109918
std	4.848611	2.389739	1.794719
min	0.000000	0.000000	0.000000
25%	2.000000	1.000000	5.000000
50%	4.000000	2.000000	6.000000
75%	8.000000	4.000000	7.000000
max	20.000000	10.000000	10.000000

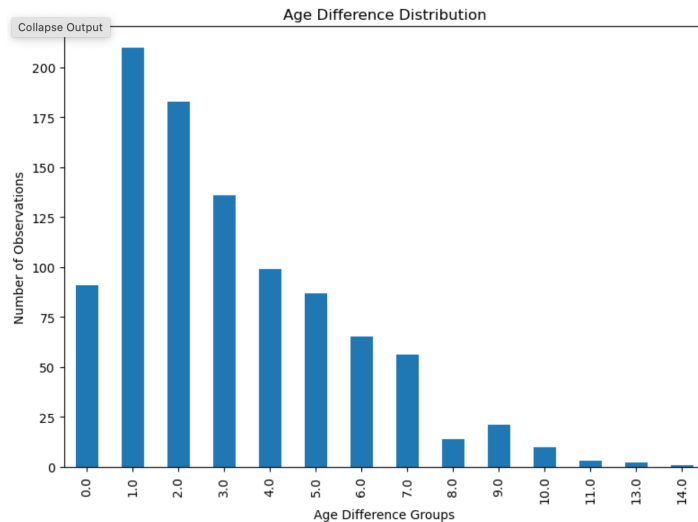
	guess_prob_liked	match
count	978.000000	978.000000
mean	4.925358	0.158487
std	2.187606	0.365384
min	0.000000	0.000000
25%	3.000000	0.000000
50%	5.000000	0.000000
75%	7.000000	0.000000
max	10.000000	1.000000

[8 rows x 33 columns]

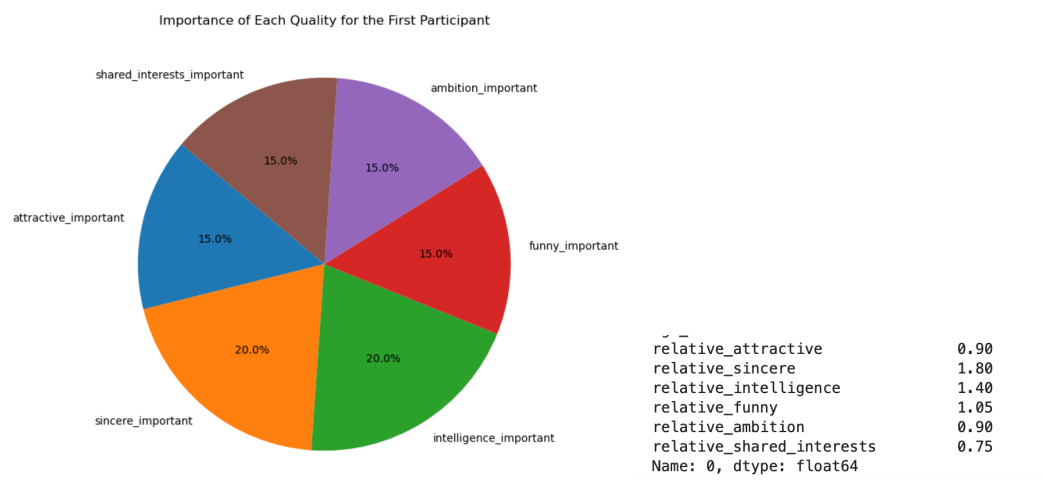
Histograms of Numerical Variables



The dataset contains 978 observations, and the gender distribution is nearly balanced (mean of gender ≈ 0.49). The average age of participants (age) and their partners (age_o) is approximately 25 years. The interquartile range indicates that most ages fall between 23 and 27 years, which is also confirmed by the histograms. Variables such as attractive_o, sincere_o, and intelligence_o have mean values between 6 and 7. The interquartile range is roughly 5–8, suggesting that ratings are generally moderate to high. Histograms show these distributions are concentrated in the mid-range. Self-assessment and partner ratings (e.g., intelligence_partner, funny_partner) also average between 6 and 7, indicating overall positive evaluations. The match variable has a mean of only 0.158, meaning the match rate is about 15.8%, which suggests that successful matches are relatively rare. The mean of interests_correlate is approximately 0.16, indicating low correlation in interests between participants. The expected number of matches averages 2.8, which is much higher than the actual match rate.



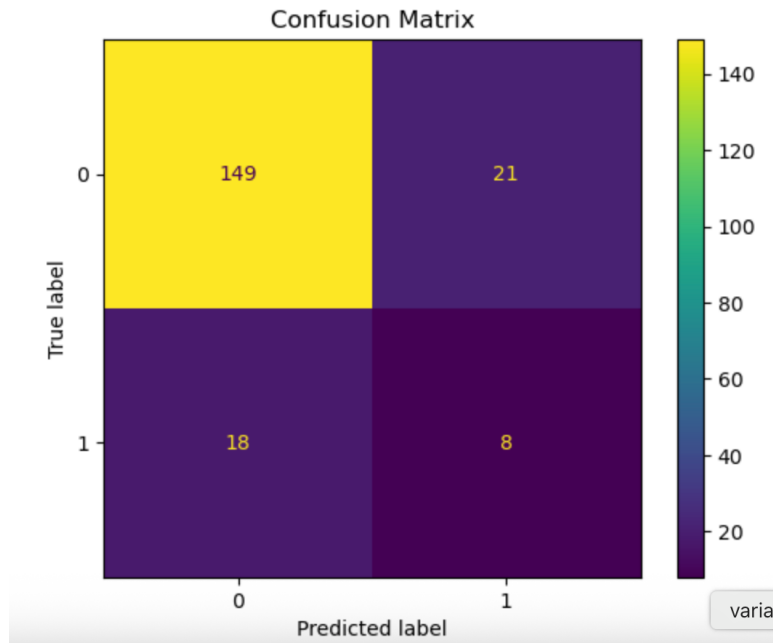
The bar chart shows that most age differences fall within 0–5 years, with 1-year difference being the most frequent, exceeding 200 observations. As the age gap increases, the frequency decreases significantly, and differences greater than 10 years are rare. This indicates that participants generally have small age gaps.



The pie chart shows that the first participant values “sincerity” and “intelligence” the most (20% each), while “attractiveness,” “funny,” “ambition,” and “shared interests” each account for 15%. The relative scores for the first participant indicate that sincerity (1.80) and intelligence (1.40) have the greatest influence, while shared interests (0.75) and ambition (0.90) have less impact. This suggests that sincerity and intelligence are key factors for this participant in forming a match.

Classification Report:

	precision	recall	f1-score	support
0	0.89	0.88	0.88	170
1	0.28	0.31	0.29	26
accuracy			0.80	196
macro avg	0.58	0.59	0.59	196
weighted avg	0.81	0.80	0.81	196



The classification report shows an overall accuracy of 80%, but due to class imbalance, the model performs poorly on predicting the “match” class (1): precision is only 0.28, recall 0.31, and F1-score 0.29

The confusion matrix indicates that the model predicts the “no match” class (0) well (149 correct predictions), but only 8 correct predictions for the “match” class, with a high misclassification rate. This suggests the model is biased toward the majority class and could be improved by using more advanced models or applying techniques to handle class imbalance.