From Swipes to Success: An Analysis of Dating App User Behavior

Christina Lee

Background and Motivation

- Artificial dataset generated based on dating app behavior and usage ([Kaggle](#))
- Contains 50,000 records and 19 different variables
- Aims to explore algorithmic and behavioral factors influencing matching success on dating apps
- Motivated by the growing impact of online dating to start relationships
- Provides practical applications for improving the user experience and platform performance for existing/future apps

Variables

- Gender
- Sexual Orientation
- Location Type
- Income Bracket
- Education Level
- Interest Tags
- App Usage (minutes)
- Swipe Right Ratio
- Likes Received
- Mutual Matches
- Profile Pic Count
- Bio Length
- Message Sent Count
- Emoji Usage Rate
- Swipe Time of Day
- Match Outcome
- Age
- Height (cm)
- Weight (kg)
- Zodiac Sign
- Body Type
- Relationship Intent

Summary Statistics

	count	mean	median	mode	std	min	max
app_usage_time_min	50000.0	149.91	150.00	293.00	86.99	0.0	300.00
swipe_right_ratio	50000.0	0.50	0.50	0.53	0.20	0.0	1.00
likes_received	50000.0	99.53	100.00	40.00	58.00	0.0	200.00
mutual_matches	50000.0	13.87	13.00	0.00	9.11	0.0	30.00
profile_pics_count	50000.0	2.99	3.00	2.00	2.00	0.0	6.00
bio_length	50000.0	250.17	250.00	189.00	144.80	0.0	500.00
message_sent_count	50000.0	50.07	50.00	46.00	29.17	0.0	100.00
emoji_usage_rate	50000.0	0.29	0.27	0.19	0.16	0.0	0.94
last_active_hour	50000.0	11.52	12.00	3.00	6.92	0.0	23.00
age	50000.0	38.48	38.00	34.00	12.13	18.0	59.00
height_cm	50000.0	172.58	173.00	177.00	16.13	145.0	200.00
weight_kg	50000.0	72.10	70.30	66.40	17.07	37.8	120.00

	count	mean	median	std	min	25%	50%
age	50000.0	38.48	38.0	12.13	18.0	28.00	38.0
height_cm	50000.0	172.58	173.0	16.13	145.0	159.00	173.0
weight_kg	50000.0	72.10	70.3	17.07	37.8	59.20	70.3
mutual_matches	50000.0	13.87	13.0	9.11	0.0	6.00	13.0
likes_received	50000.0	99.53	100.0	58.00	0.0	49.00	100.0
swipe_right_ratio	50000.0	0.50	0.5	0.20	0.0	0.37	0.5
app_usage_time_min	50000.0	149.91	150.0	86.99	0.0	74.00	150.0
bio_length	50000.0	250.17	250.0	144.80	0.0	125.00	250.0
profile_pics_count	50000.0	2.99	3.0	2.00	0.0	1.00	3.0

	75%	max	missing_pct
age	49.00	59.0	0.0
height_cm	187.00	200.0	0.0
weight_kg	83.80	120.0	0.0
mutual_matches	22.00	30.0	0.0
likes_received	150.00	200.0	0.0
swipe_right_ratio	0.64	1.0	0.0
app_usage_time_min	225.00	300.0	0.0
bio_length	376.00	500.0	0.0
profile_pics_count	5.00	6.0	0.0
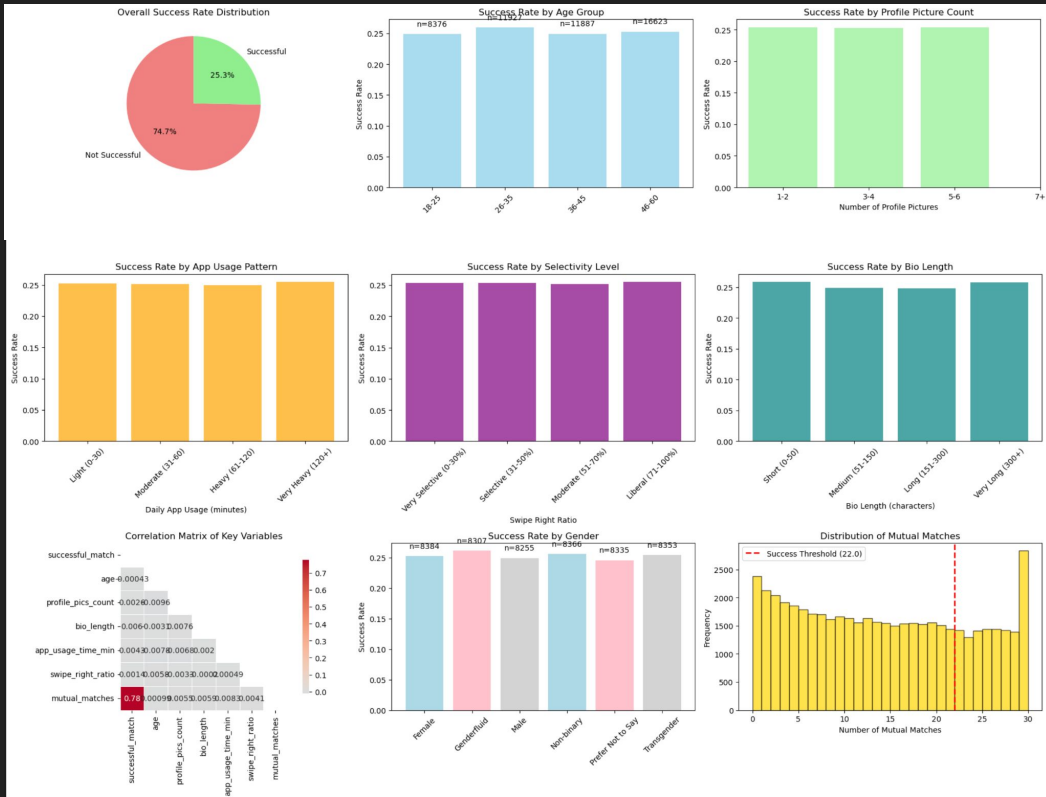
- The artificial dataset seemed to have a lack of variety
- Flat success rates
- Similar correlations across categories

Bayesian Logistic Regression

Feature Importance for Dating Success (Logistic Regression)
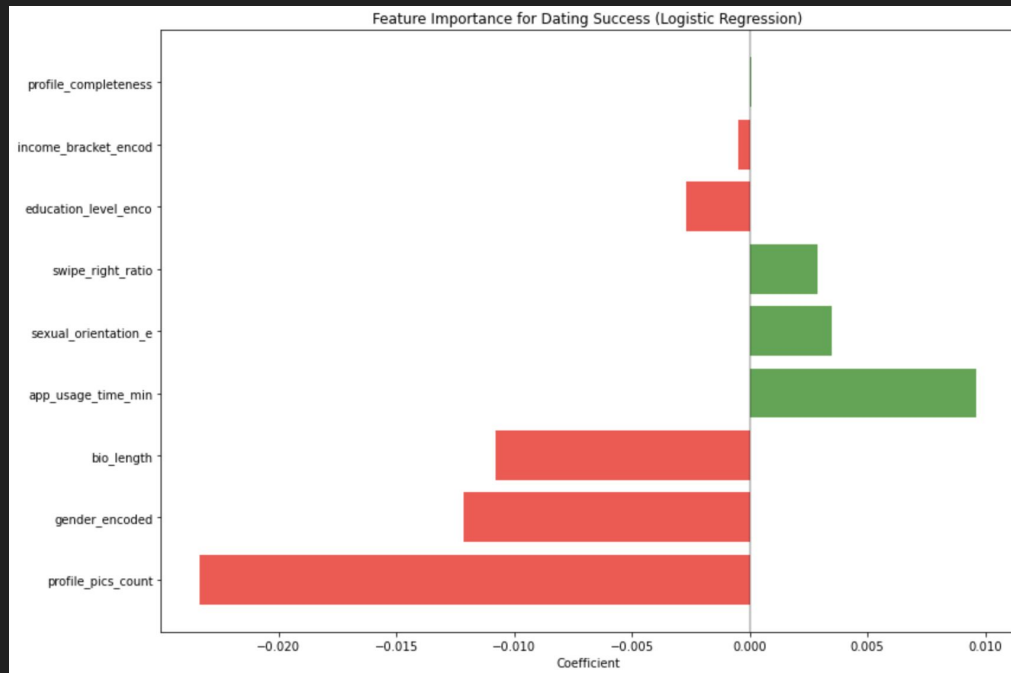
Factors That Increase Success
- App usage time (min): 1.01× increased odds

- Sexual orientation: 1.00× increased odds

- Swipe right ratio: 1.00× increased odds
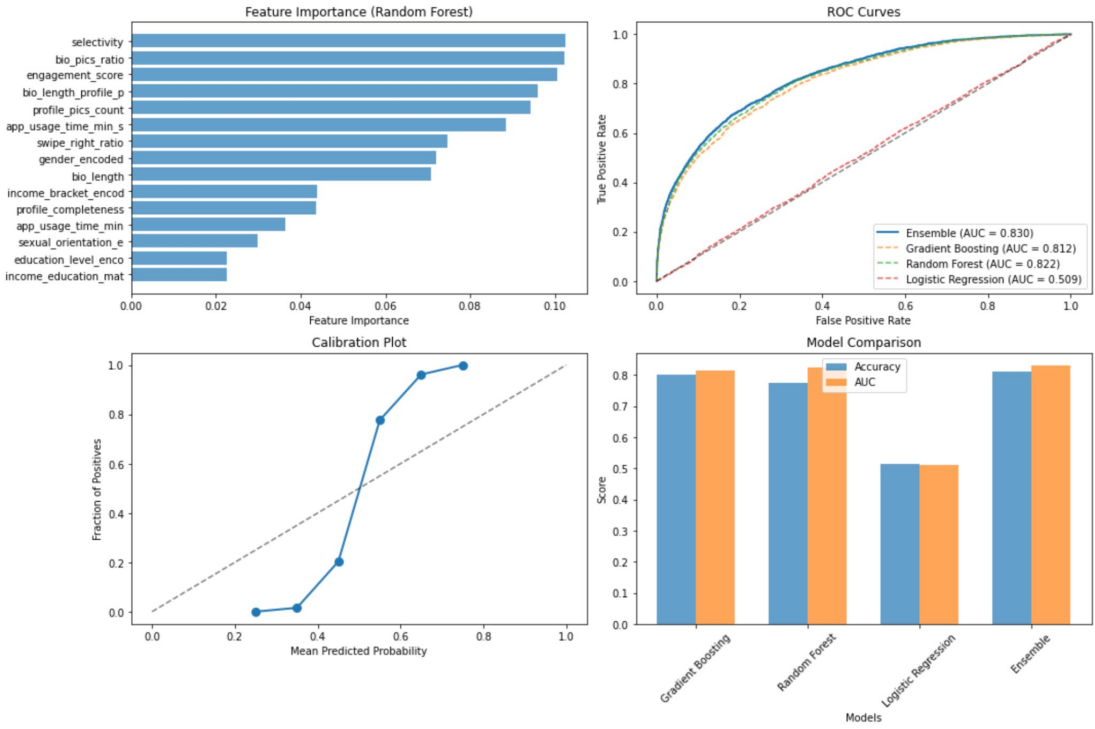
- Profile completeness: 1.00× increased odds

Factors That Decrease Success
- Profile picture count: 0.98× decreased odds

- Gender: 0.99× decreased odds

- Bio length: 0.99× decreased odds

- Education level: 1.00× decreased odds

- Income bracket: 1.00× decreased odds

Overall Model Performance
- Test Accuracy: 74.7%

- Success Base Rate: 25.3%

Improved Model

Ensemble Model Performance:

```
              precision    recall    f1-score    support

         0        0.81       0.97        0.88       7468
         1        0.79       0.35        0.48       2532

  accuracy                               0.81      10000
 macro avg        0.80       0.66        0.68      10000
weighted avg       0.81       0.81        0.78      10000
```

Confusion Matrix:

```
                    Predicted
                   No      Yes
Actual    No     7233      235
          Yes    1649      883
```

Success Rate by Prediction Confidence:
  0.0-0.3: 0.0% (9 samples)
  0.3-0.5: 18.6% (8873 samples)
  0.5-0.7: 78.9% (1116 samples)

Results

- The ensemble model achieved 81.2% test accuracy and 0.830 AUC, which outperformed all individual classifiers
- Precision for predicting successful users: 0.79, which indicates reliable positive predictions
- Recall for successful outcomes: 0.35, which reflects a conservative prediction strategy (predicts success only when it is very confident)
- Weighted average F1-score across classes: 0.78, which demonstrates consistent overall performance

Conclusion

Top Predictors of Success:

- Selectivity/Bio-pics ratio - Being selective about who you swipe on
- Engagement score - Active, meaningful participation
- Optimal bio length - Not too short, not too long
- Multiple profile pictures - But not excessive (3-4 photos seem optimal)
- Moderate app usage time - Consistent but not obsessive usage

Factors That Surprisingly Matter Less:

- Demographics (age, gender, education, income) have minimal impact
- Physical attributes (height, weight) show weak correlations
- Profile completeness alone does not drive success

Success appears more about behavioral strategy (being selective, engaged, and balanced) rather than inherent characteristics. The most successful users are those who use the app thoughtfully rather than frantically swiping or being completely passive. Most demographic factors do not create significant advantages; it is about how you use the platform.