



Exploratory Data Analysis & Visualization with Python

Justine VILLENEUVE

Jedha
Fullstack Data Analysis
dafs-14

Agenda

- 🔥 Introduction: Context, Objective, Dataset
- 🔥 Data overview
- 🔥 EDA: Sample profiling
- 🔥 EDA: Decision distribution
- 🔥 What are the factors that drive mutual interest?
- 🔥 Conclusion & Recommendations

INTRODUCTION

Context

Tinder's marketing team is tackling a project aimed at **reversing a decrease in the number of matches** on the app.

To gather insights, they conducted **a speed dating experiment**. Participants provided extensive personal information, which could reflect potential app profile data.

INTRODUCTION

Context

Tinder's marketing team is tackling a project aimed at **reversing a decrease in the number of matches** on the app.

To gather insights, they conducted **a speed dating experiment**. Participants provided extensive personal information, which could reflect potential app profile data.

Objective

Understand the factors that drive mutual interest between people

INTRODUCTION

Context

Tinder's marketing team is tackling a project aimed at **reversing a decrease in the number of matches** on the app.

To gather insights, they conducted **a speed dating experiment**. Participants provided extensive personal information, which could reflect potential app profile data.

Objective

Understand the factors that drive mutual interest between people

Dataset

- Data gathered during the experiment ran between 2002 and 2004
- 6 attributes scored: Attractiveness, Sincerity, Intelligence, Fun, Ambition, Shared Interests
- Binary decision to meet again the partner: Yes/No

DATA OVERVIEW

8,378
observations
collected throughout
the experiment

195
dimensions
per date

551
subjects

Pre-processing:

- Create columns to label the categorical values (gender, go-out habits, goal, race, etc.)
- Acting on missing values

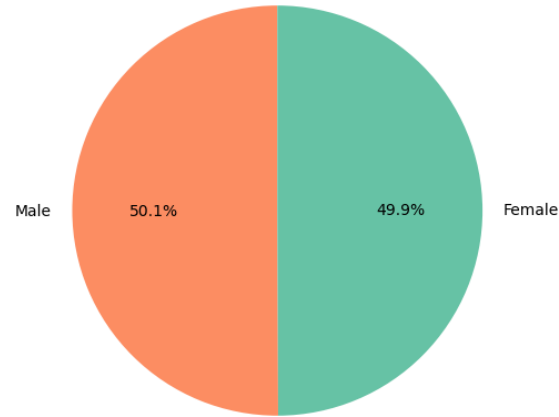
```
# for missing values in age and age_o, filling with median since 75% of subjects are between 24-28 yo
# for missing values in race and race_o, filling with mode
# for missing values in goal, filling with '6' = other
# for missing values in date, go out, imprace, imprelig, filling with median

df.fillna({'age': df['age'].mean(),
          'age_o': df['age_o'].mean(),
          'race': df['race'].mean(),
          'race_o': df['race_o'].mean()},
         inplace=True)
df.fillna({'goal': 6}, inplace = True)

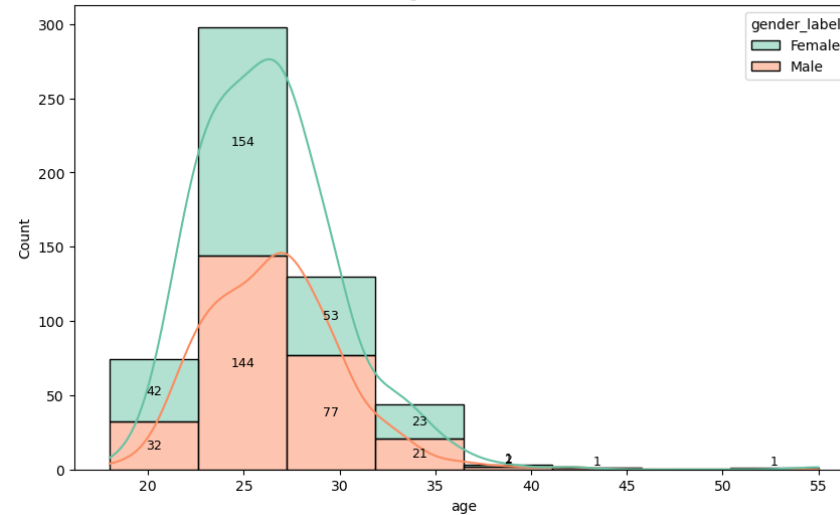
df.fillna({'date': df['date'].median(),
          'go_out': df['go_out'].median(),
          'imprace': df['imprace'].median(),
          'imprelig': df['imprelig'].median()},
         inplace=True)
```

EDA – SAMPLE PROFILING

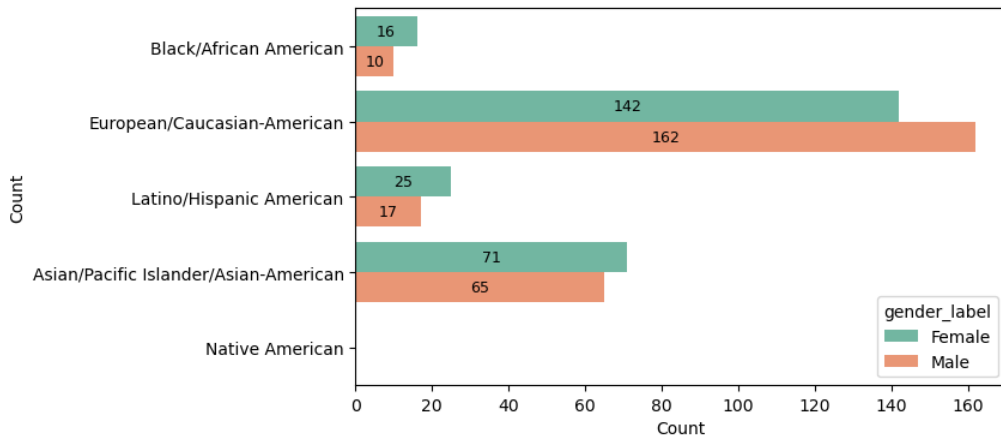
Gender distribution



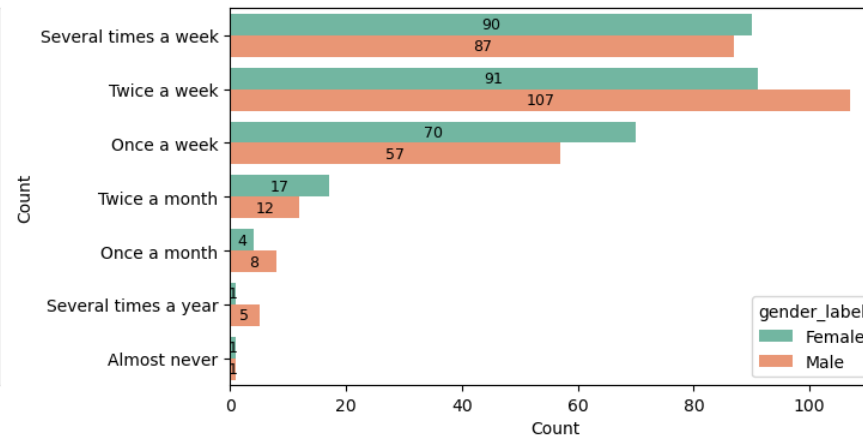
Age distribution



Race distribution



Go out habit distribution

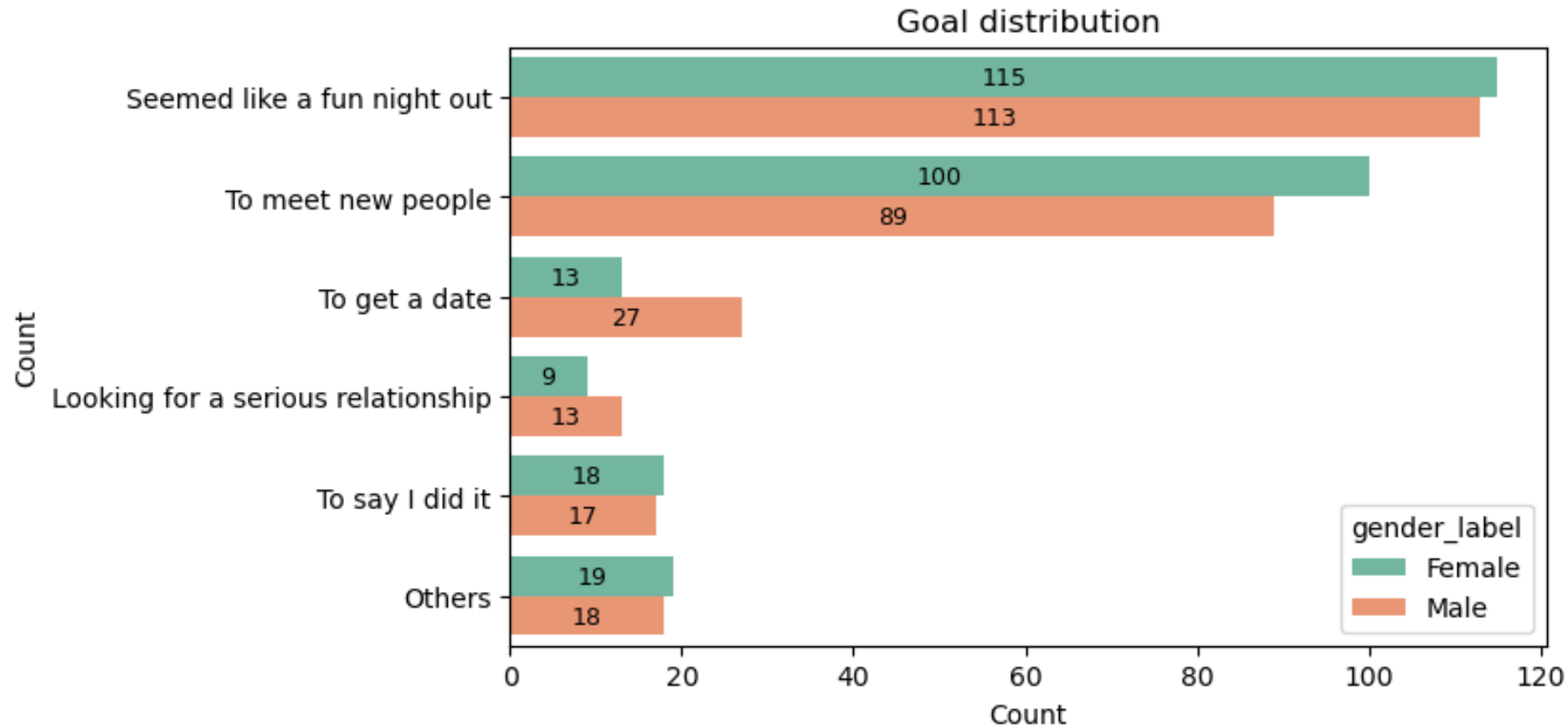


The experiment counts **551 unique subjects**.

The population looks homogeneous in terms of gender across ages, races and go out habits, assuming the sample has been made based on the **statistical distribution of Tinder's audience**.

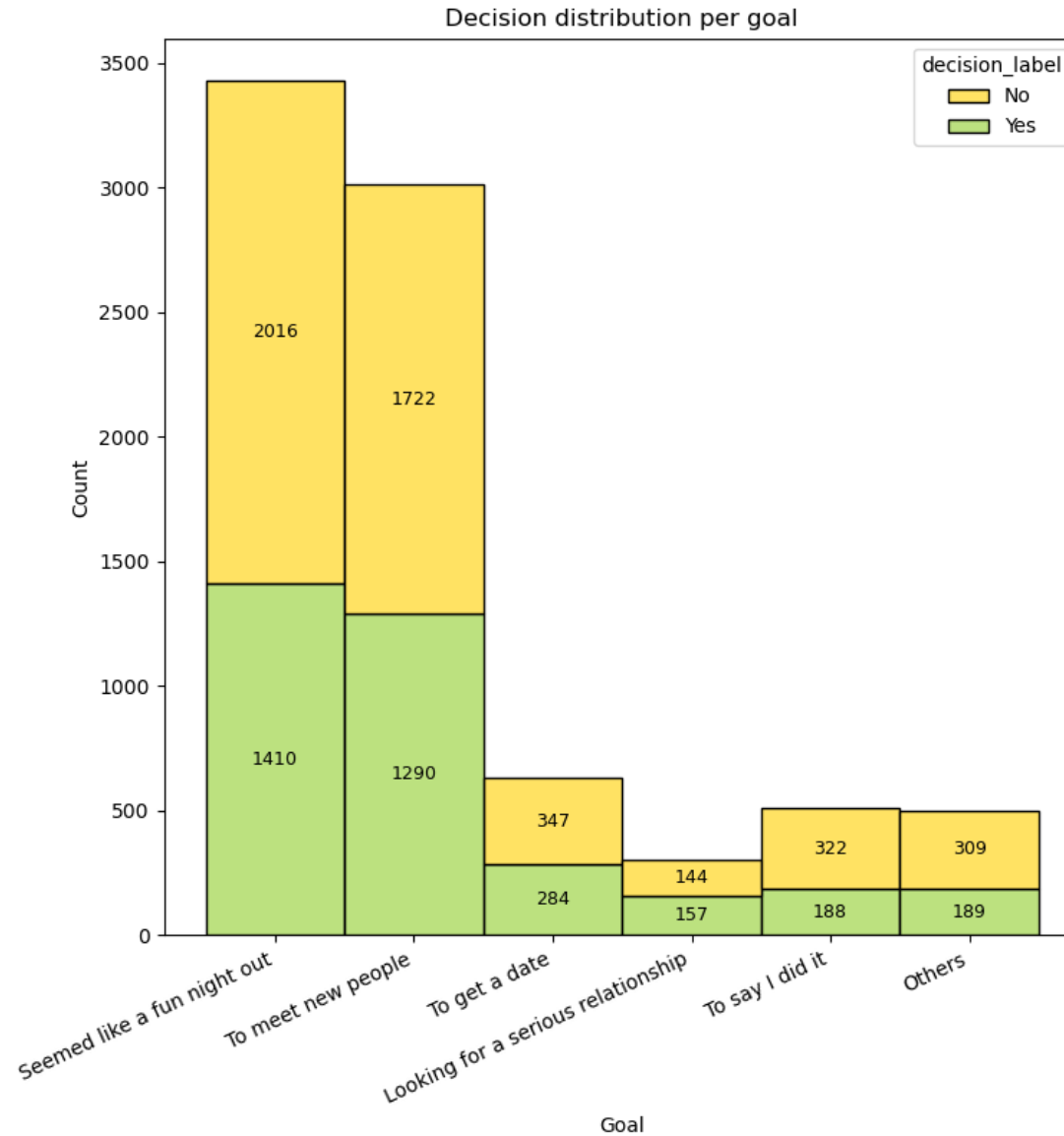
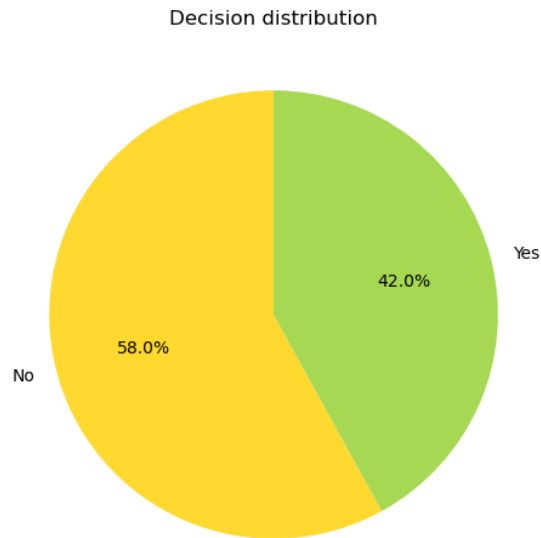
EDA – SAMPLE PROFILING

Let's focus on the **subjects' goal** for our analysis.



They mainly want to have fun and meet new people.

EDA – DECISIONS' DISTRIBUTION



Key insights from decision distribution:

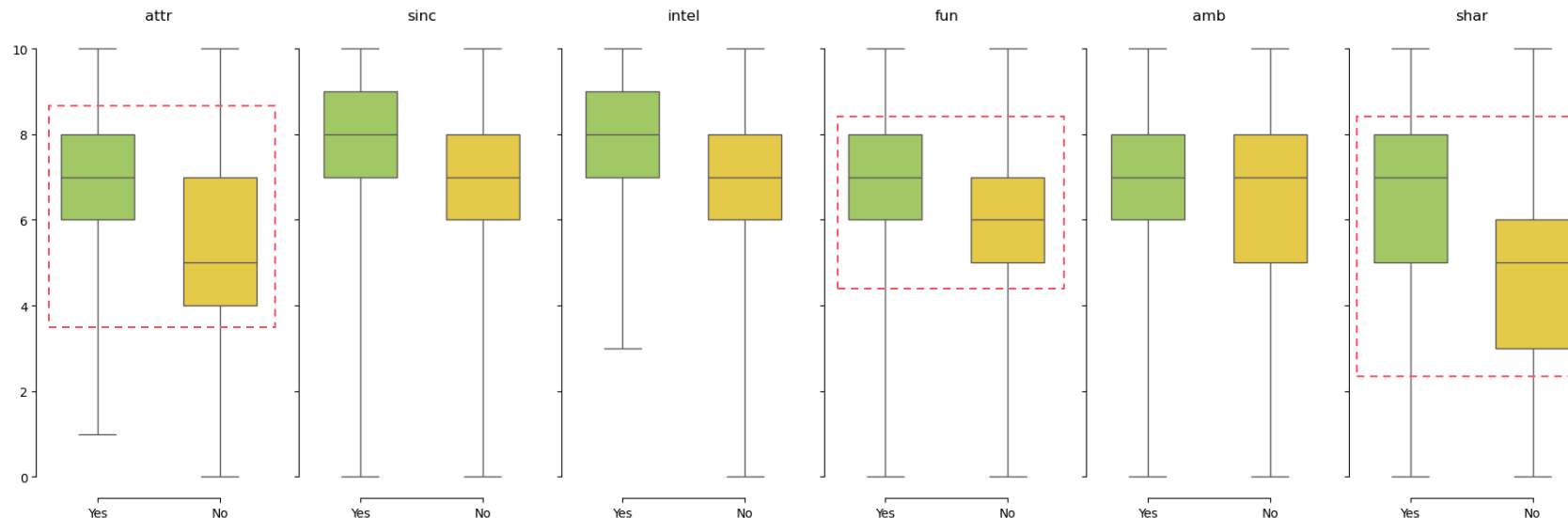
- 58% of interactions end in a "No", 42% in a "Yes".
- Users looking for fun or meeting new people generate the most interactions, but also the most rejections.
- More serious goals show slightly higher proportions of "Yes", though with fewer participants.

WHAT ARE THE FACTORS THAT DRIVE MUTUAL INTEREST BETWEEN PEOPLE?

Methodology: compare scores' mean per attribute depending on the decision (Yes/No)

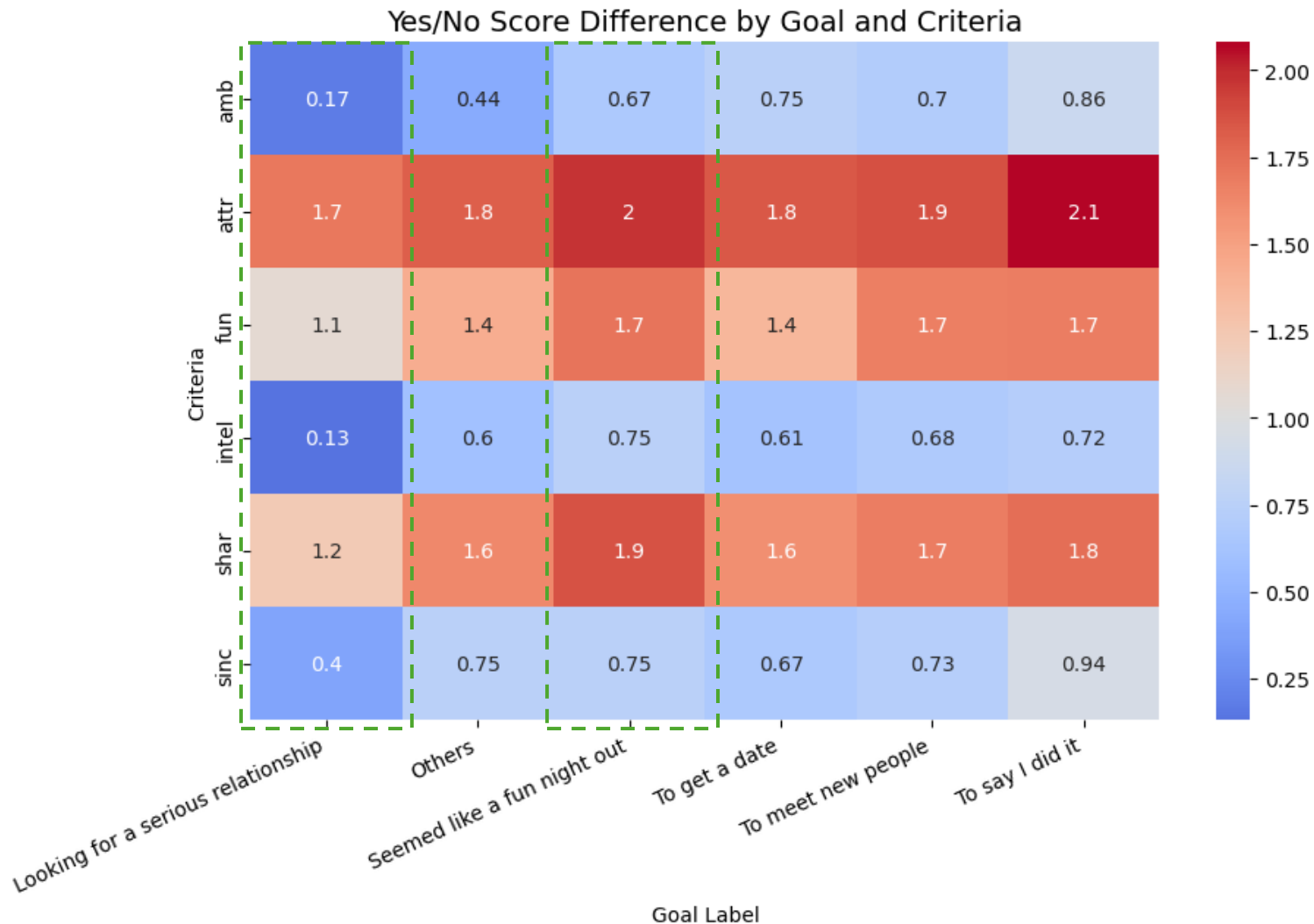
	# No	# Yes	# Difference
attr	5.37	7.29	1.92
shar	4.72	6.46	1.74
fun	5.7	7.33	1.63
sinc	6.86	7.6	0.74
intel	7.08	7.76	0.68
amb	6.49	7.16	0.67

What makes the difference in the willingness to meet someone again?



Attractiveness, shared interests, and fun consistently show the largest differences between positive and negative responses.

HOW PEOPLE'S GOAL IMPACT THE WAY THEY SCORE THEIR PARTNERS



Segmentation by goal_label shows that the relative importance of these attributes depends on the participants' objective.

Those seeking a serious relationship show more moderate differences, while those with a more exploratory or playful objective reinforce the weight of attractiveness, fun, and shared interests.

CONCLUSION & RECOMMENDATION

Top drivers of “see again” decisions:

- Attractiveness
- Shared Interests
- Fun

Role of user goals:

- Serious relationship: effects are more moderate
- Casual/exploratory goals: stronger influence of attractiveness, shared interests, fun

Marketing implications:

- Emphasize high-impact attributes in app features (profiles, matching prompts, interest tags).
- Tailor marketing campaigns to user goals by highlighting fun, chemistry, or shared interests depending on the segment.

Limits:

- Data based on an experiment that differs from the Tinder’s experience
- Declared scores
- American subjects only

What’s next ?

- Multivariate analysis
- ML



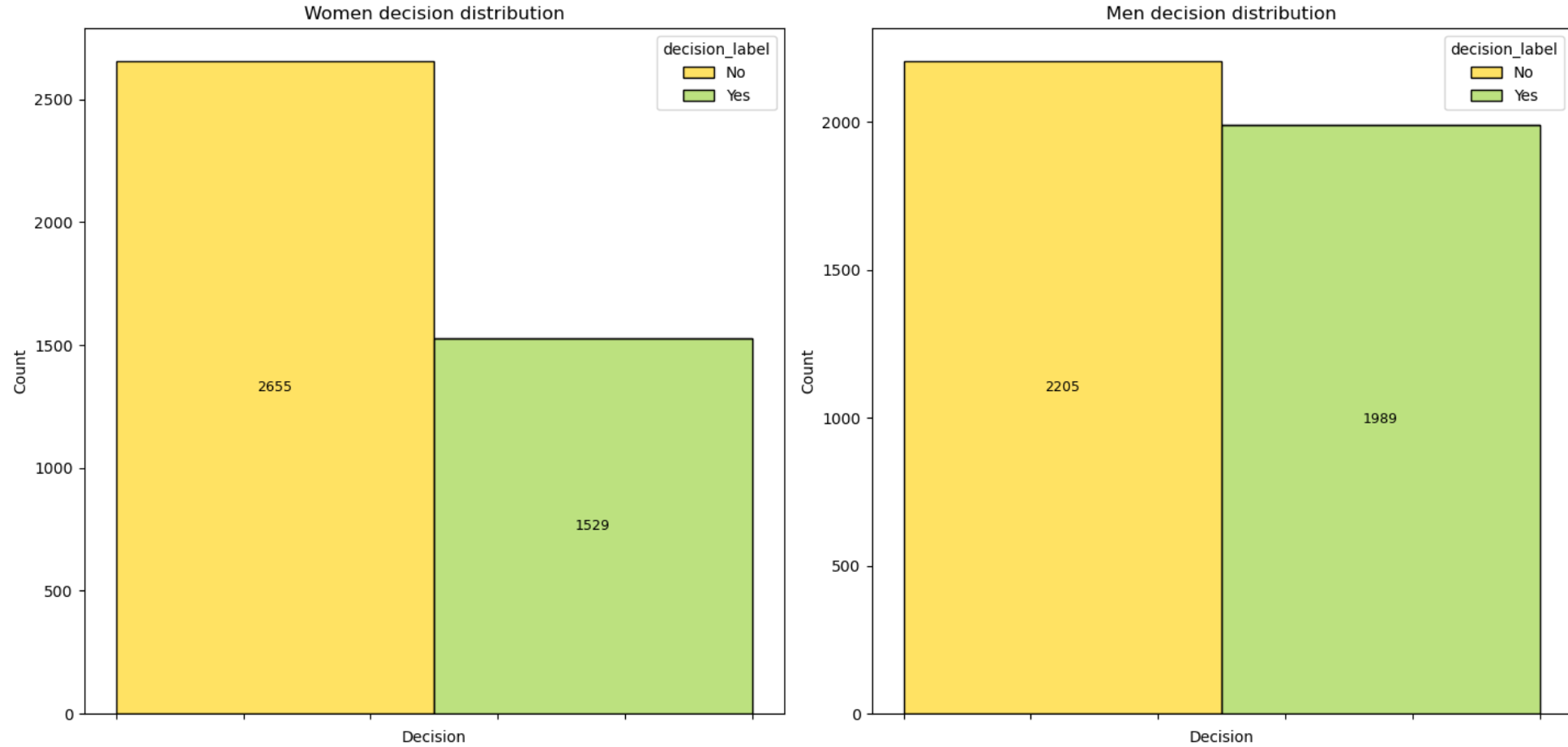
Thank you

Do you have any questions ?

Justine VILLENEUVE

Appendix

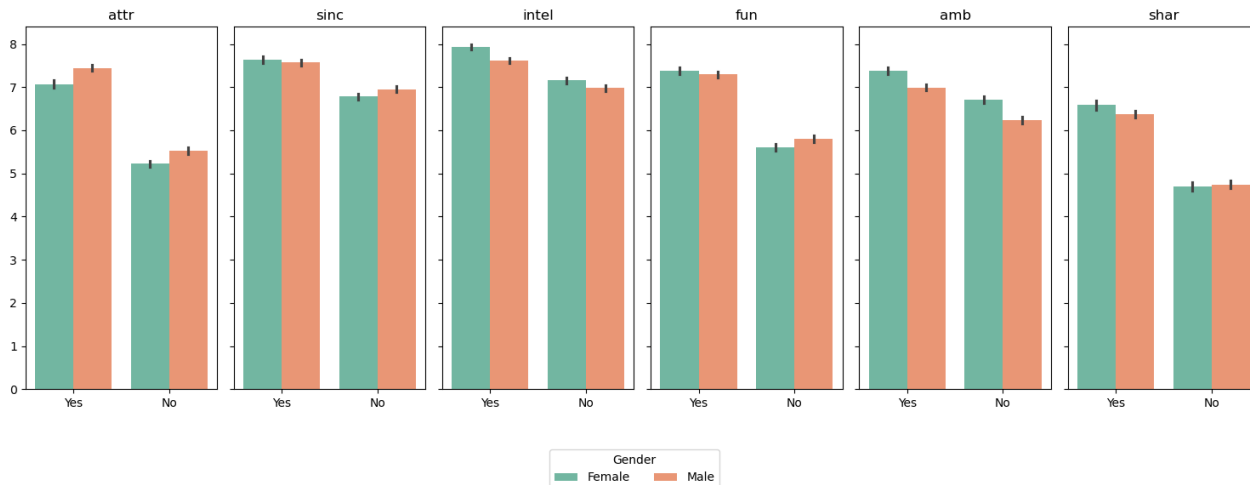
EDA – DECISIONS' DISTRIBUTION



Women filter more. They actually impact more the chance to match.

HOW WOMEN SCORE MEN & HOW MEN SCORE WOMEN

Women scoring:			
decision_label	W-Avg Score (No)	W-Avg Score (Yes)	W-Score Difference
shar	4.70	6.58	1.88
attr	5.23	7.07	1.84
fun	5.61	7.38	1.77
sinc	6.78	7.63	0.85
intel	7.16	7.93	0.77
amb	6.70	7.37	0.67
Men scoring:			
decision_label	M-Avg Score (No)	M-Avg Score (Yes)	M-Score Difference
attr	5.53	7.45	1.92
shar	4.75	6.37	1.62
fun	5.80	7.30	1.50
amb	6.24	7.00	0.76
intel	6.98	7.62	0.64
sinc	6.95	7.57	0.62



Top 3 scoring attributes for women (meaning women rating men) are:

1. [intel] Intelligence (7.93)
2. [sinc] Sincerity (7.63)
3. [amb] Ambition (7.37)

But when we look at the scores with a highest difference between Yes and No, we have another attributes ranking:

1. Shared Interests (+1.88 ppt)
2. Attraction (+1.84 ppt)
3. Fun (+1.77 ppt)
4. Sincerity (+0.85 ppt)
5. Intelligence (+ 0.77 ppt)
6. Ambition (+ 0.67 ppt)

Top 3 scoring attributes for men (meaning men rating women) are:

1. [intel] Intelligence (7.62)
2. [sinc] Sincerity (7.57)
3. [attr] Attraction (7.45)

But when we look at the scores with a highest difference between Yes and No, we have another attributes ranking:

1. Attraction (+ 1.92 ppt)
2. Shared Interests (+ 1.62 ppt)
3. Fun (+1.50 ppt)
4. Ambition (+ 0.76 ppt)
5. Sincerity (+0.64 ppt)
6. Intelligence (+ 0.62 ppt)