

# Todate

Marina

# Project Overview

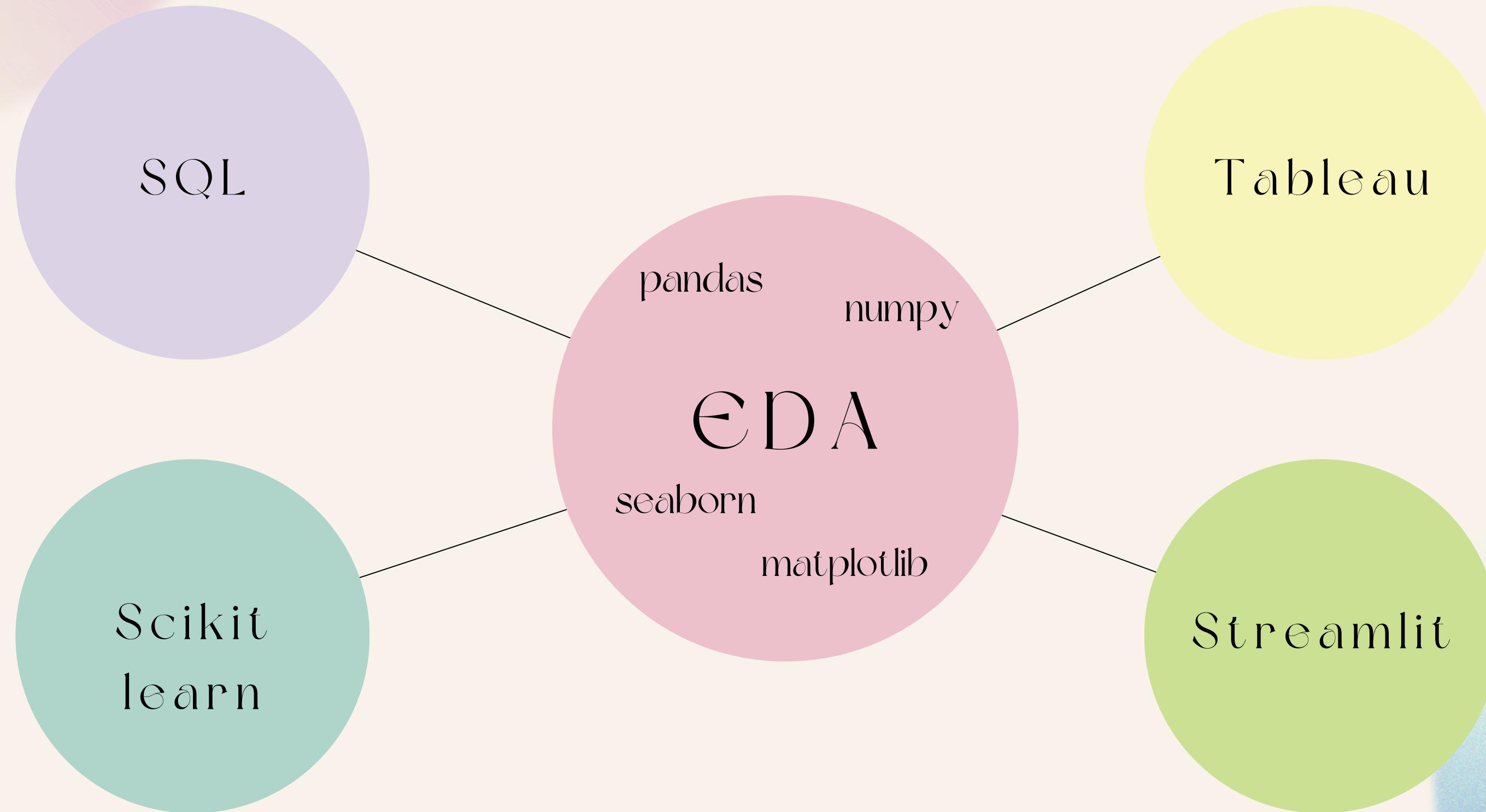
The Todate Dating App, launched in 2023, has transformed online dating market by integrating virtual interactions with real-life meetups through organized events.

By leveraging advanced data analysis techniques, the project will explore various aspects of user behavior, preferences, and interactions to enhance understanding of the online dating landscape.

# ♦ Objectives:

- 1 Data analysis
- 2 Refine predictive models
- 3 Event organization and optimization
- 4 User engagement
- 5 Company growth: new users

# Main tools



# ♦ Data selection and preparation



Original Data frame  
31 columns x 60 k rows  
Only dropped 1 column  
Remaining rows 47k

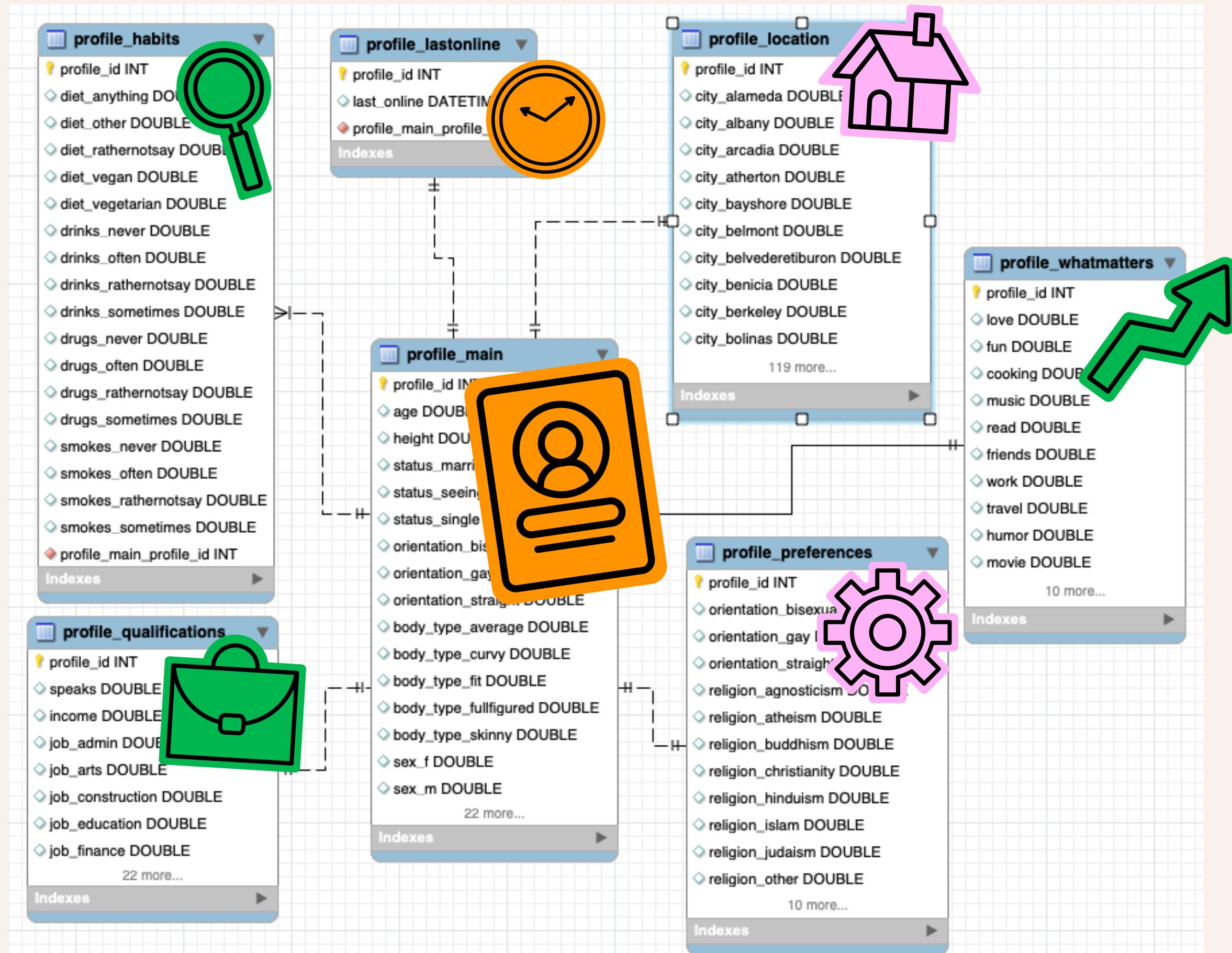


Reduce most categorical  
columns into broader  
classifications.  
(Religion, Offspring, Diet  
or jobs)



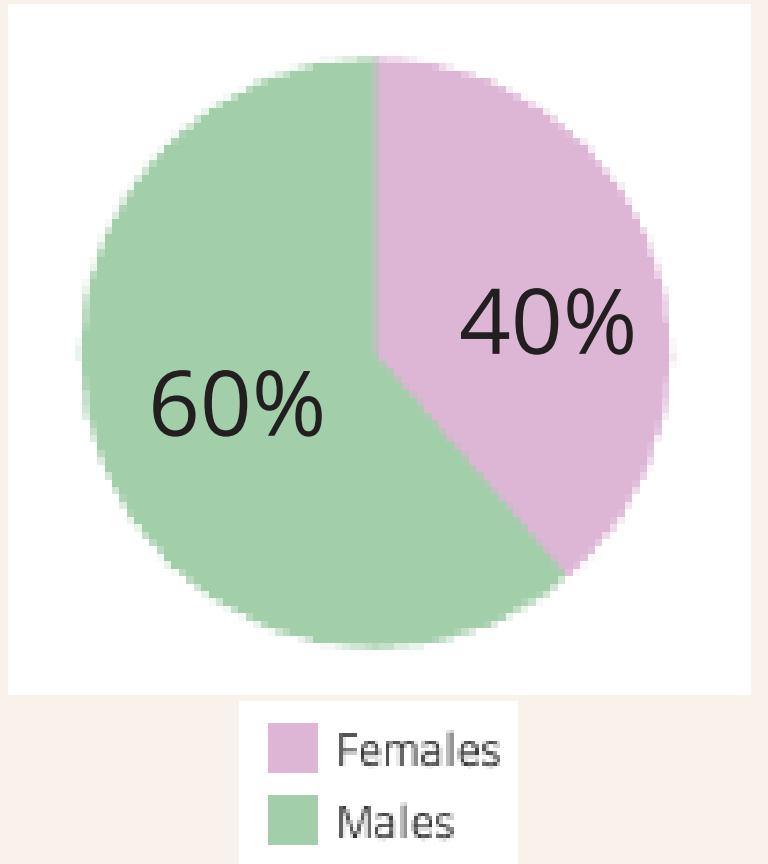
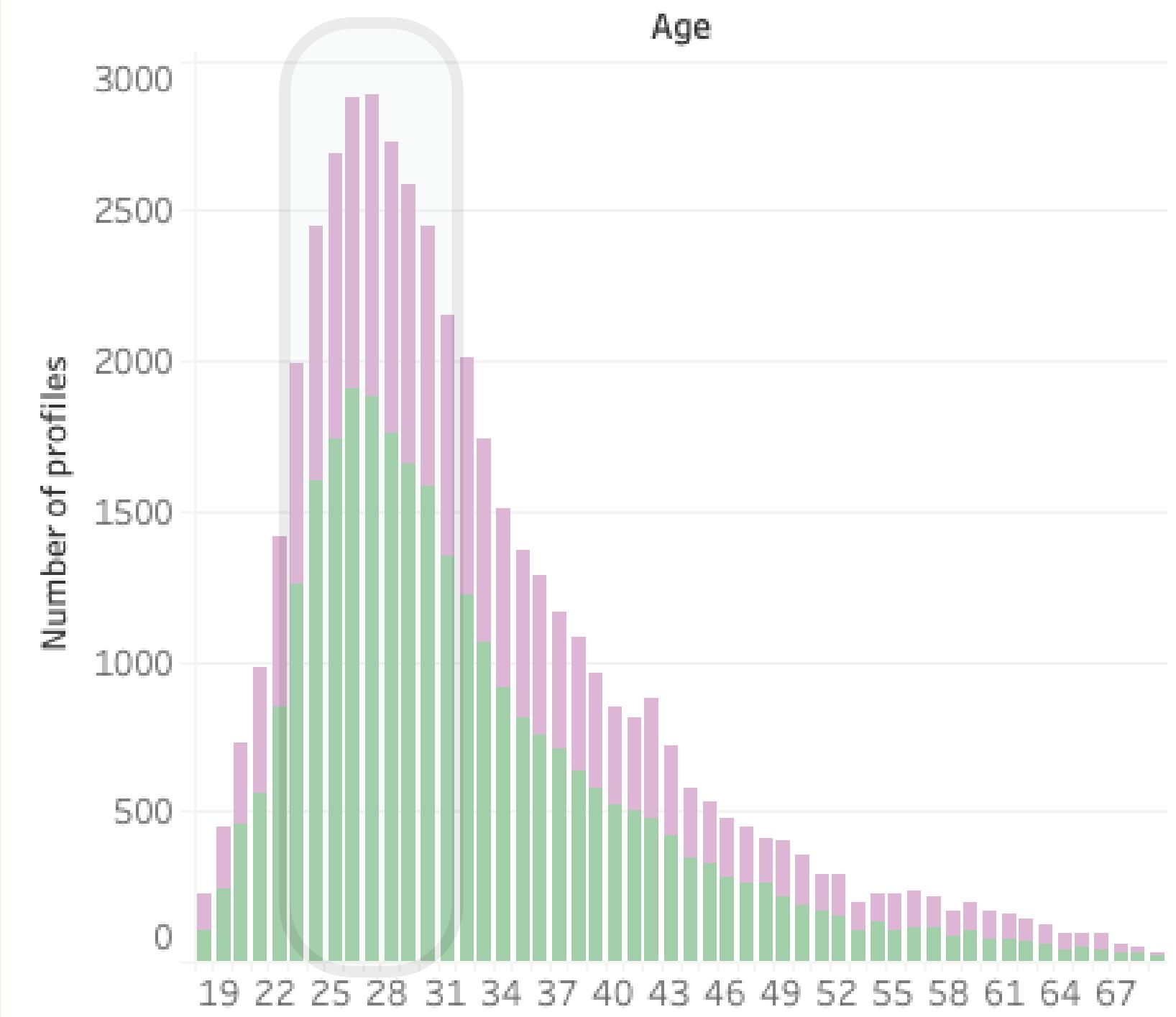
Matplotlib, Seaborn and Tableau  
Understand demographic and  
establish Customer Profiling.

# SQL database (ERD)

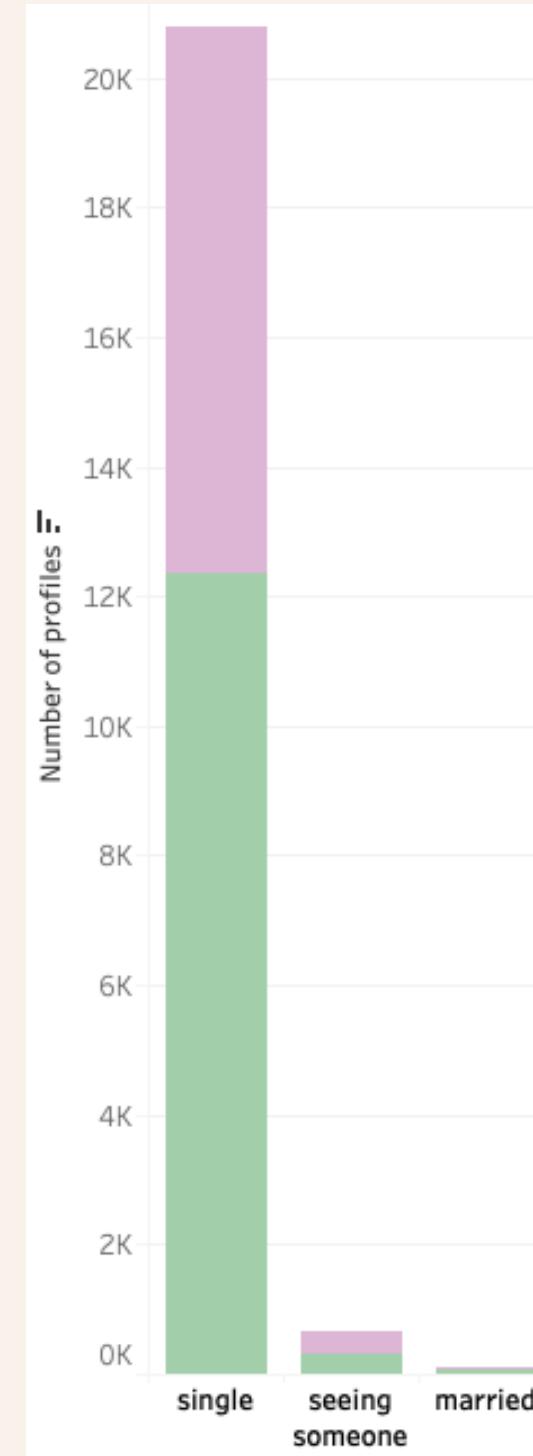
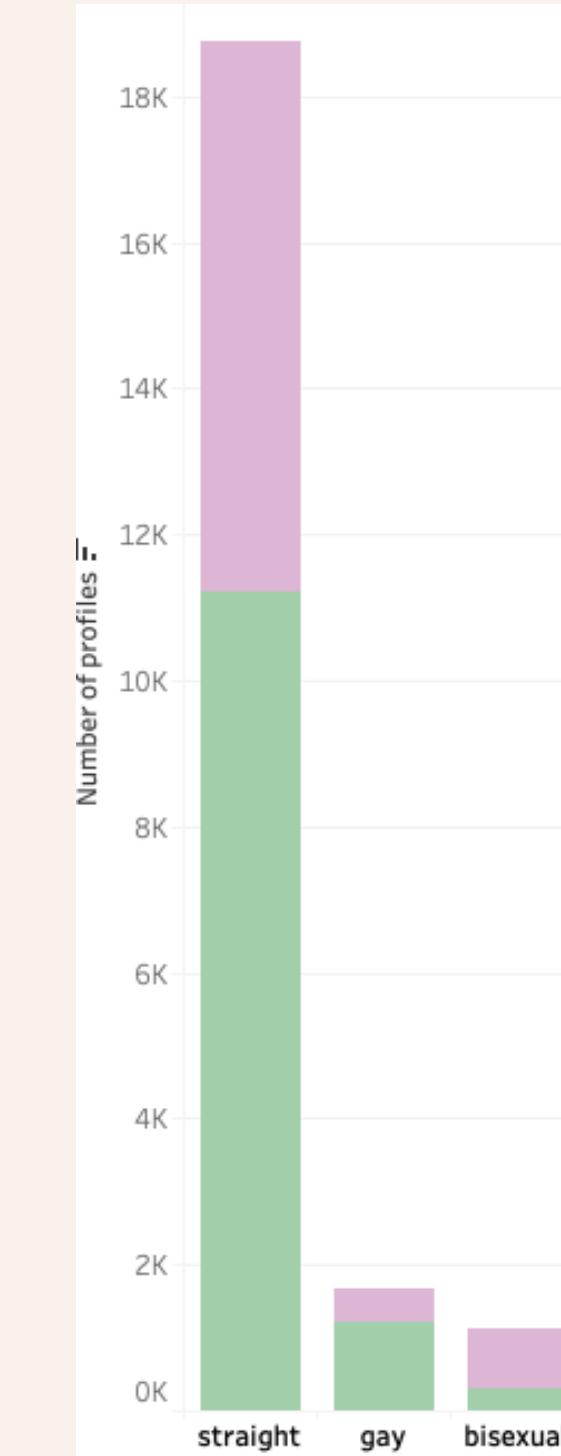


# Demographics

Age distribution



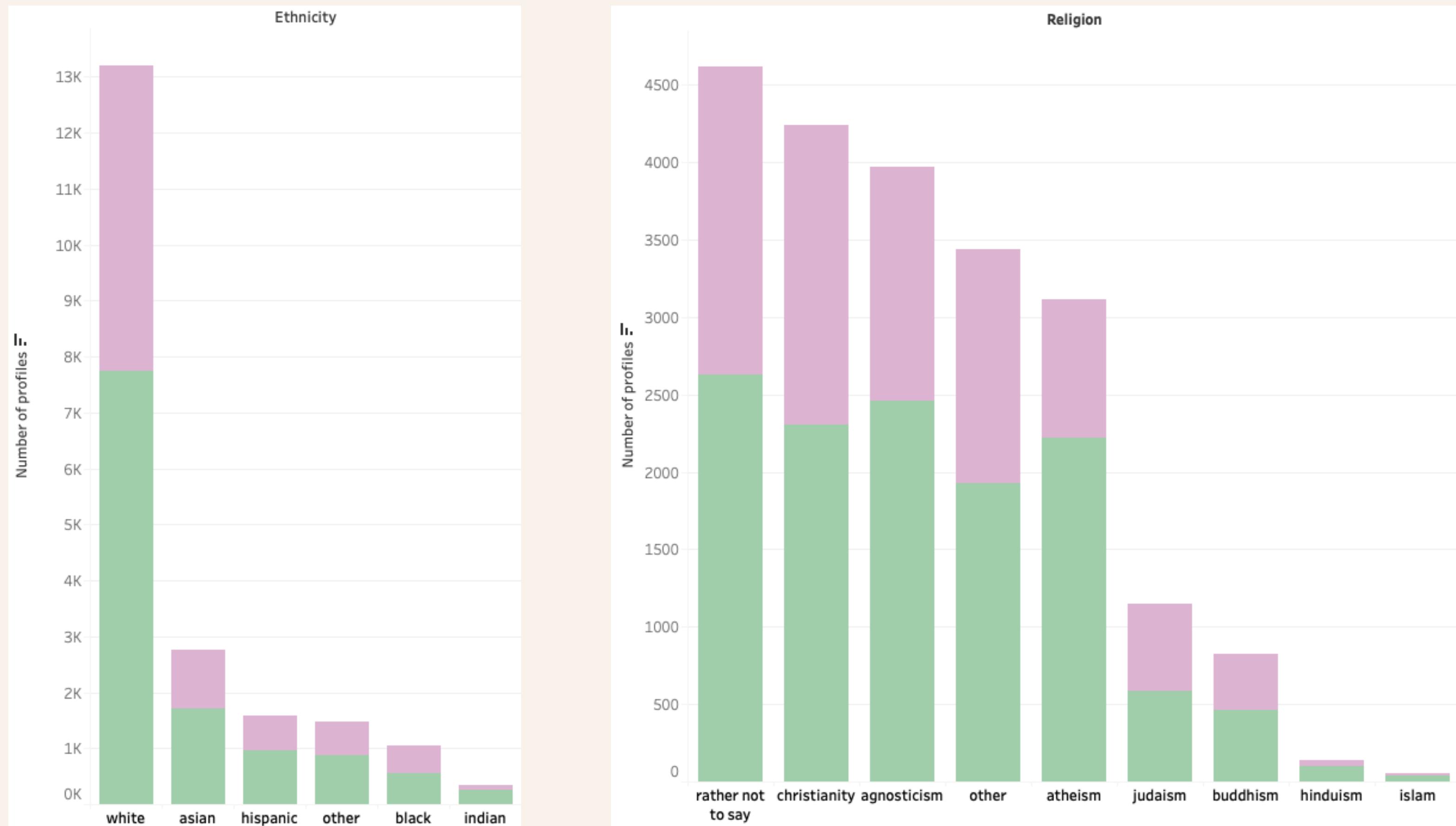
Orientation      Status



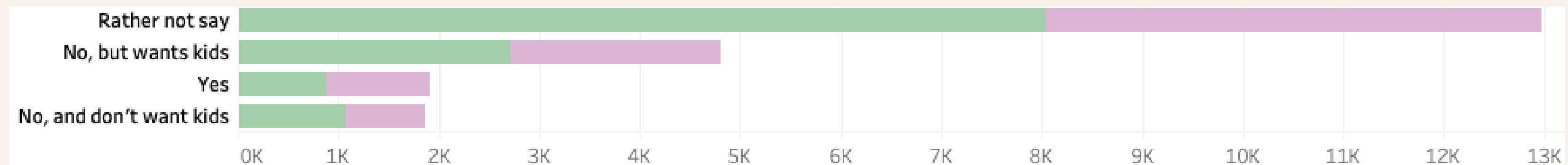
Generations:

62% Millennials  
30% Gen Z  
8% Boomers

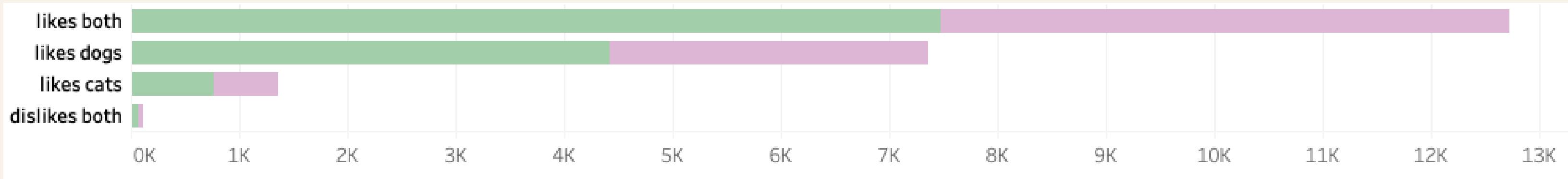
# Ethnicity and Religion by sex



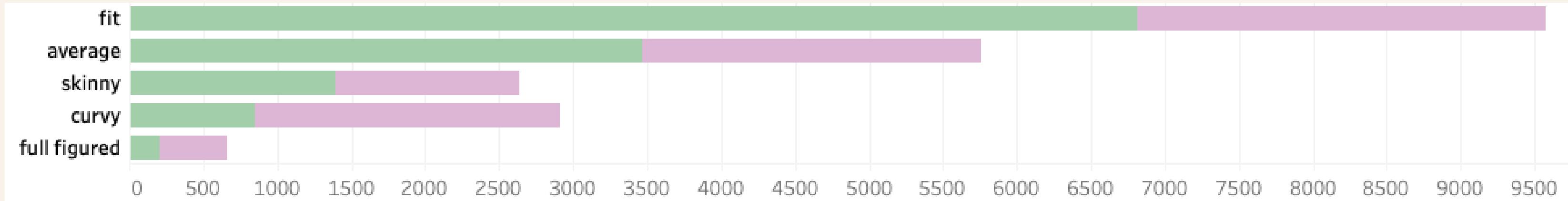
# Offspring



# Pets preferences

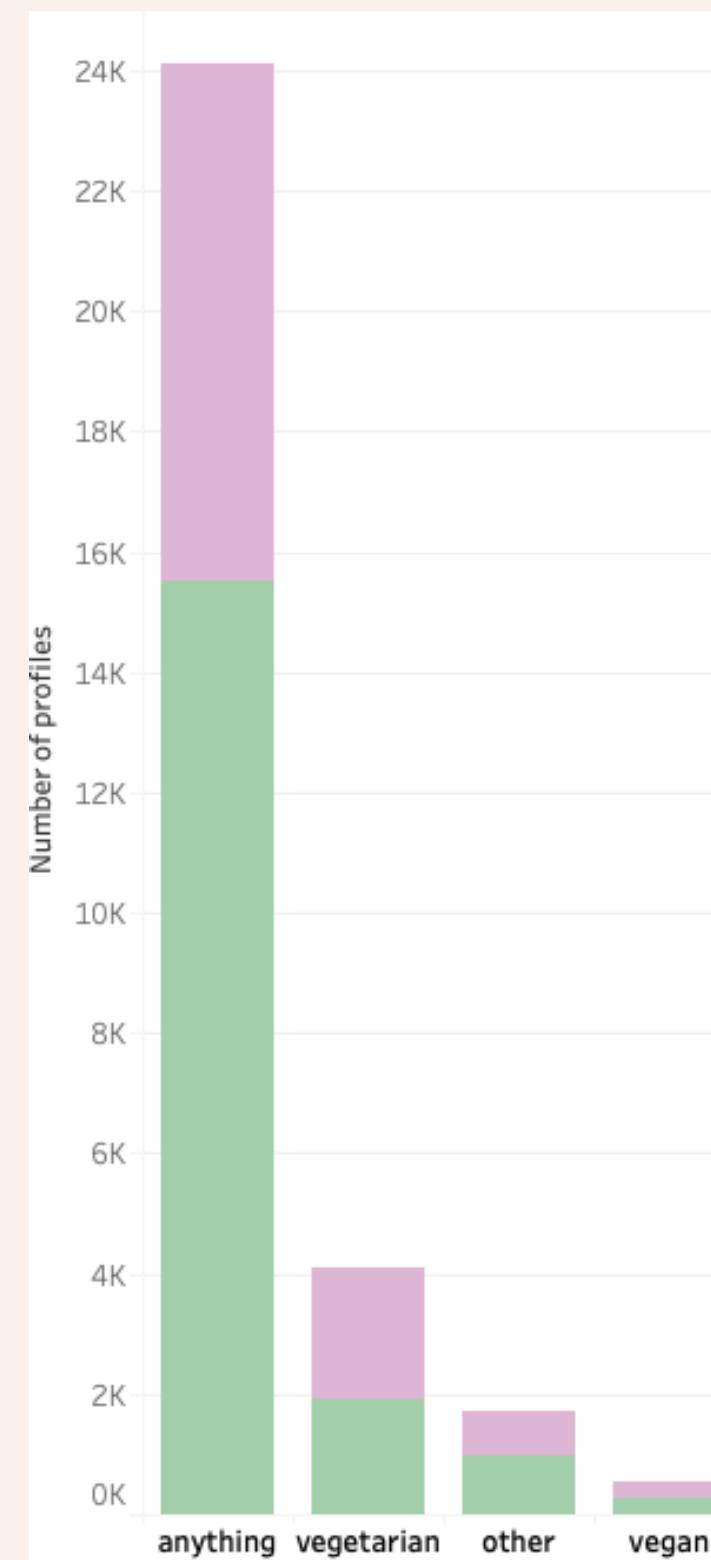


# Body Type

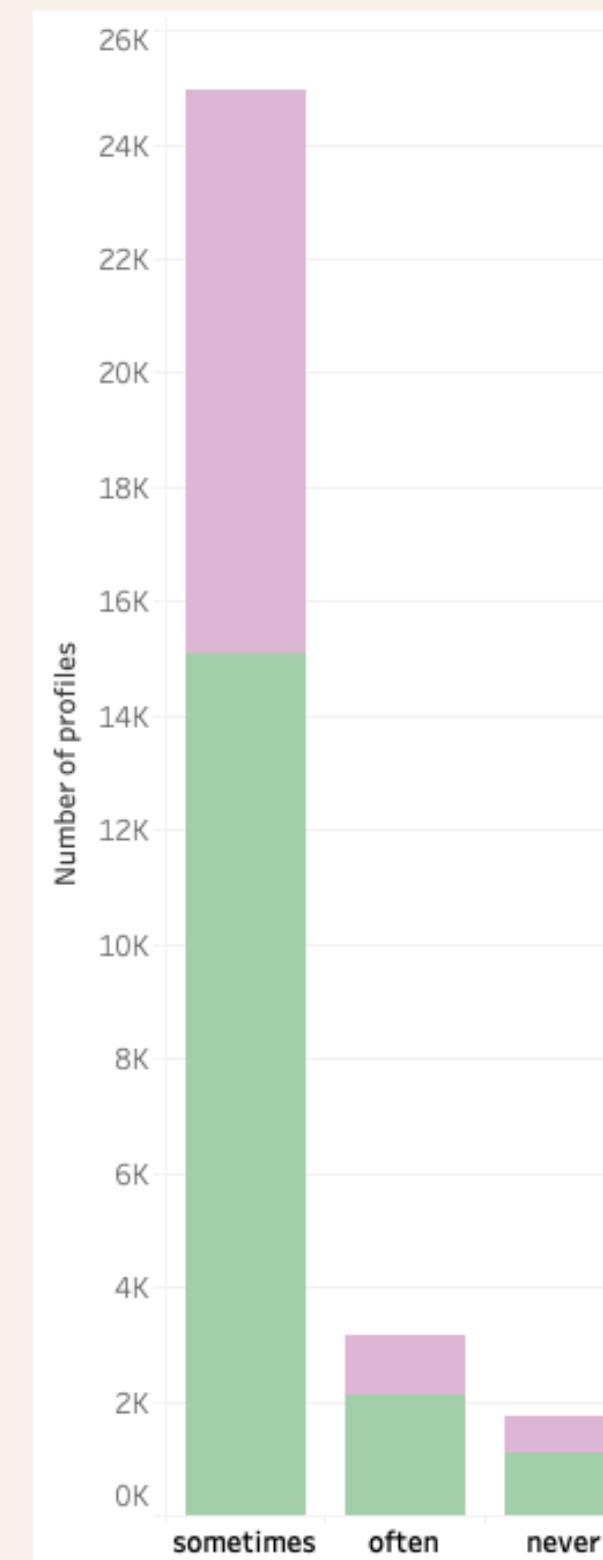


# Habits and lifestyle by sex

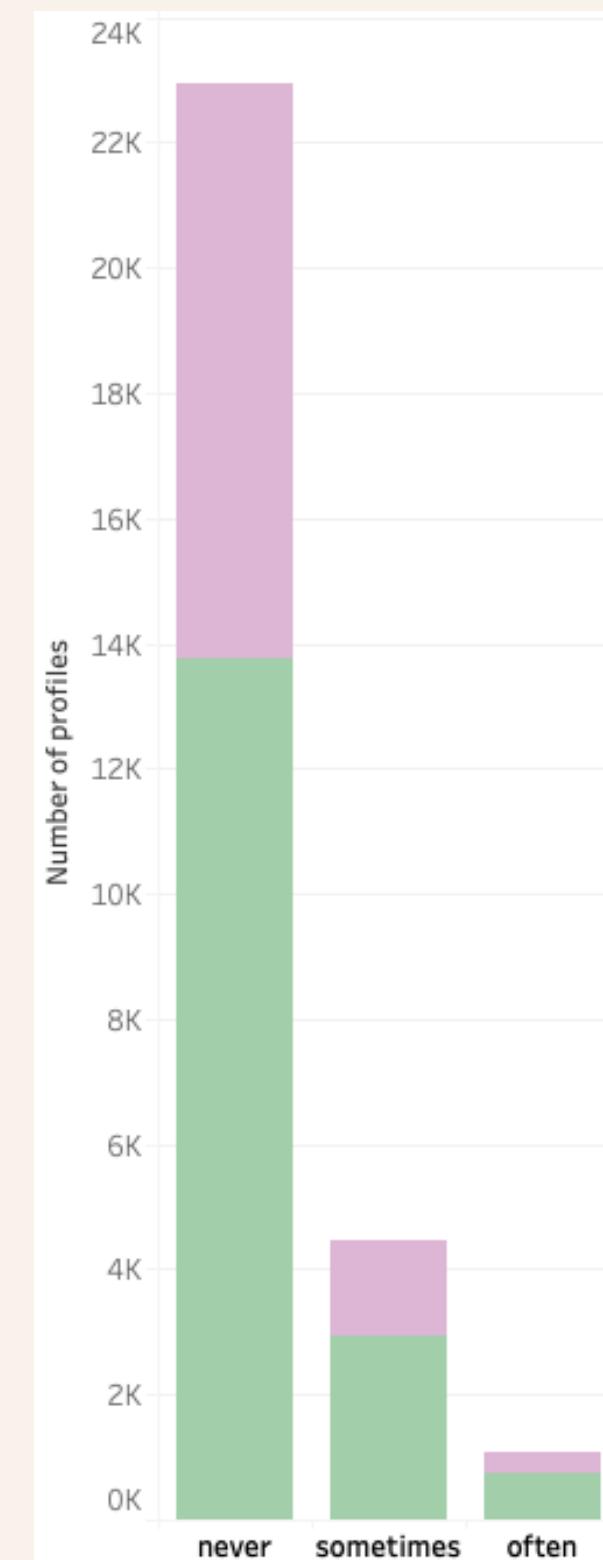
## Diet



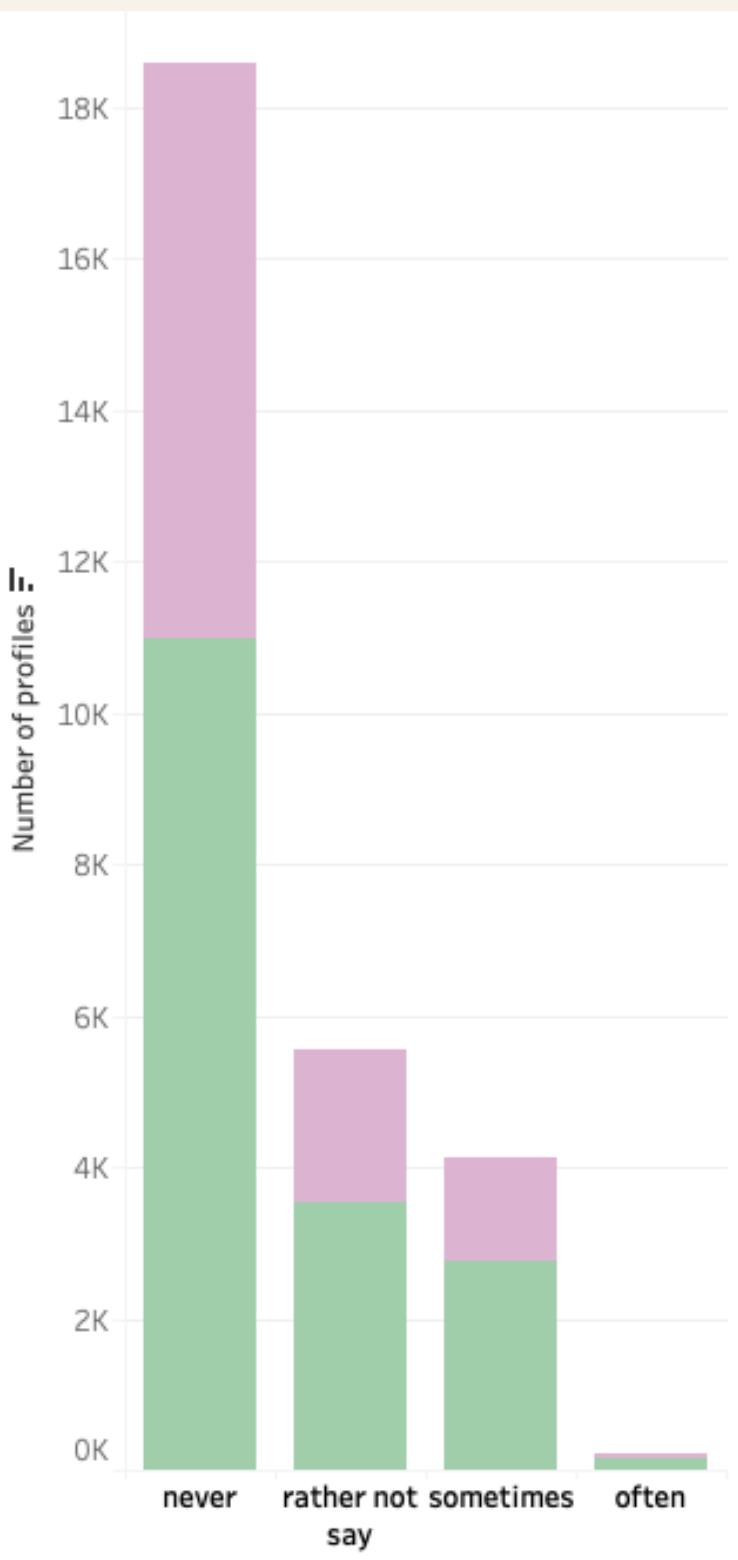
## Drinks



## Smokes

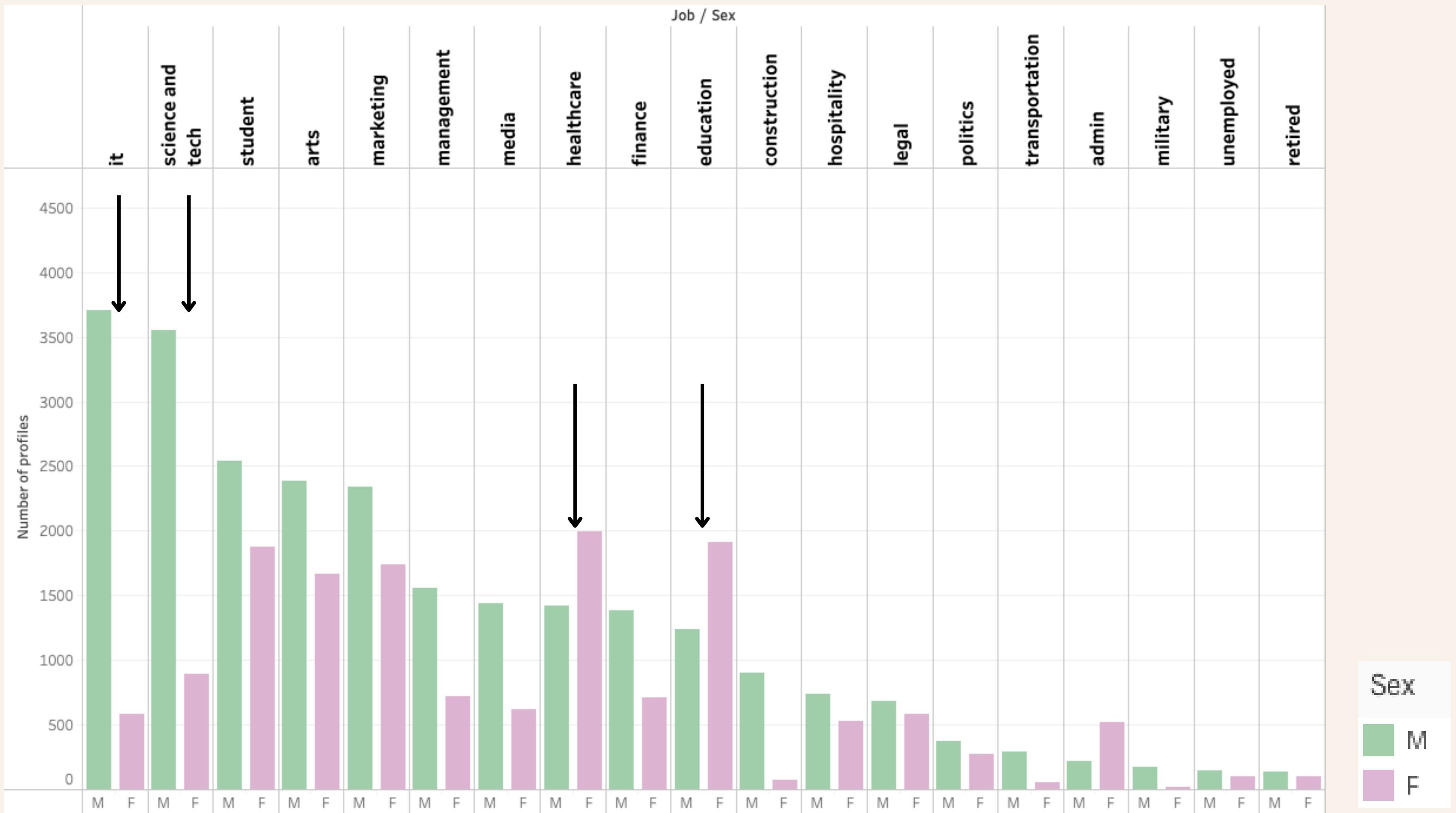


## Drugs

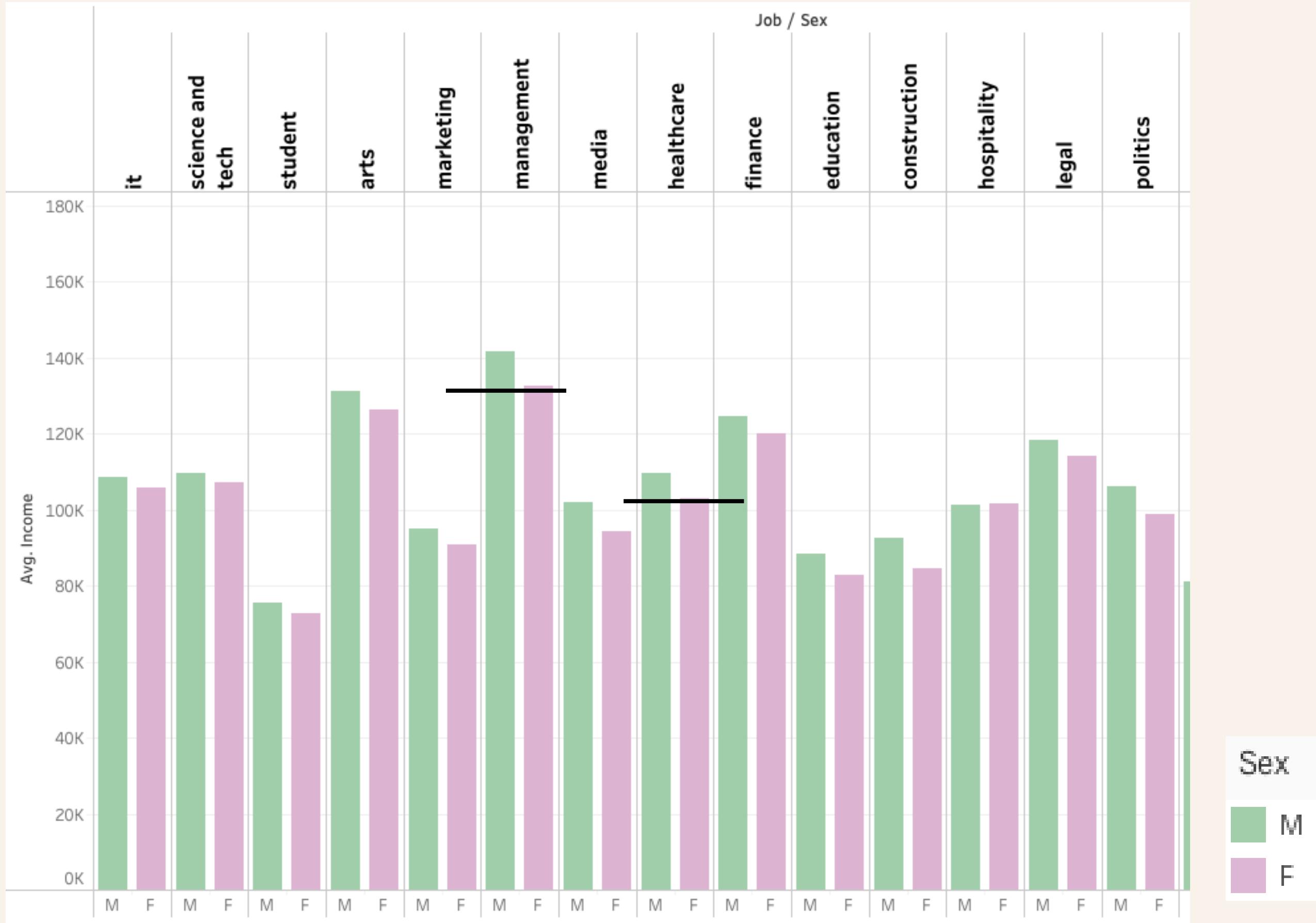


Sex  
M  
F

# Jobs distribution by sex



# Avg. income by industry by sex



T-TEST:

Healthcare

t-stat 2.16

p value 0.03

Management

t-stat 2.28

p value 0.02

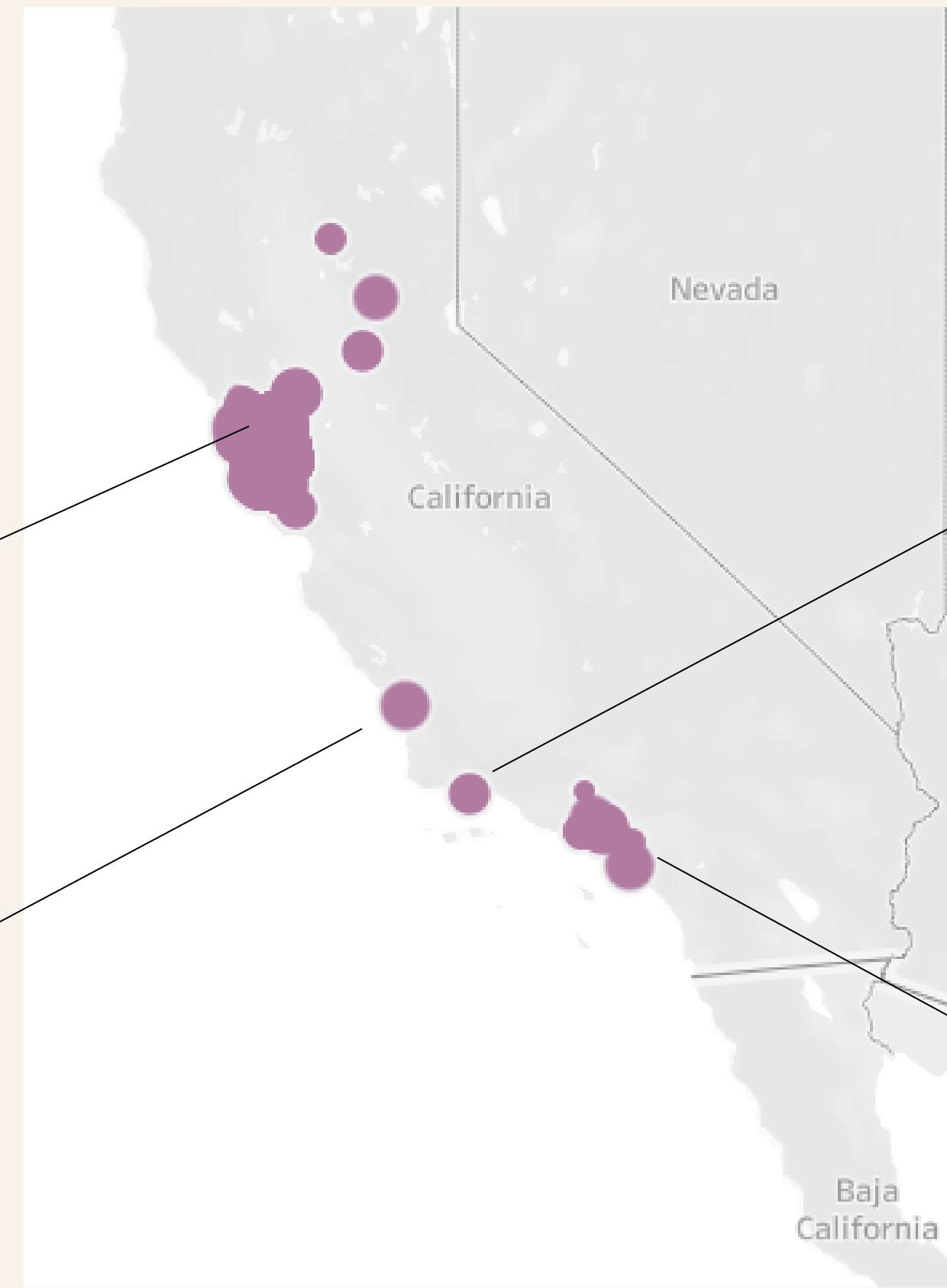
# ♦ Location

60% San Francisco

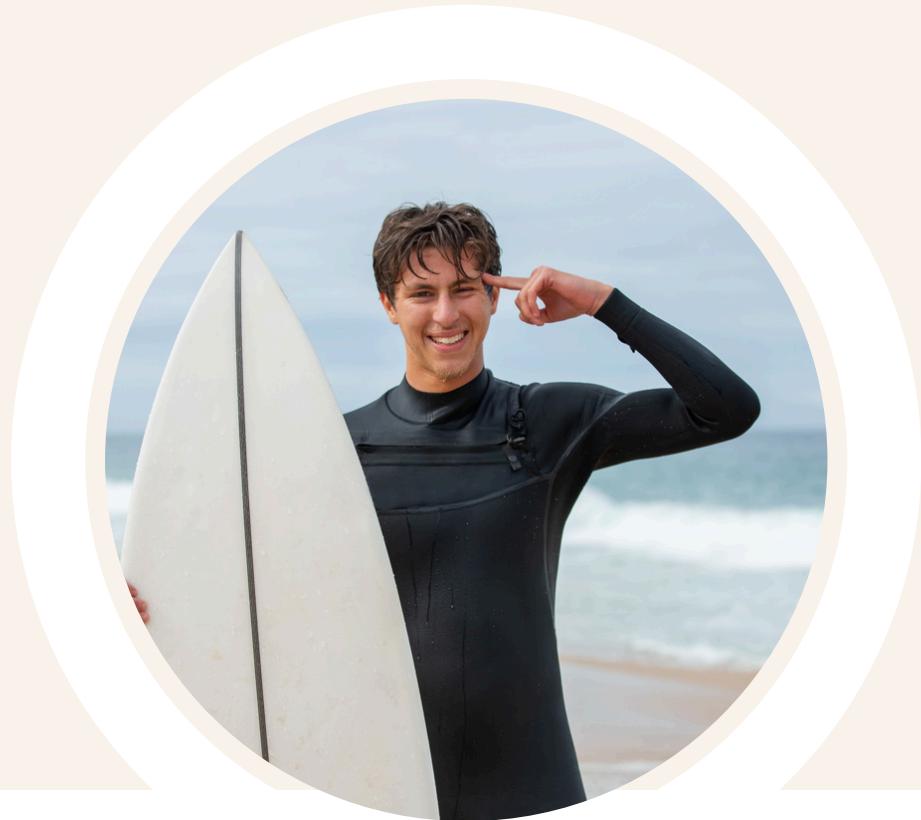
7% Santa  
Barbara

5% Santa  
Barbara

28% Los Angeles



# Customer profiling – females



**Sam, 23**

Biology student  
Love to cook. Looking for someone  
who loves to eat (and pay  
compliments).

Read my soul



**Sheldon, 28**

Software engineer  
Looking for a partner who's my kind of  
nerd.

Expect the unexpected



**Ryan, 33**

Project manager  
Only pet lovers need apply

If you believe, you believe

# Customer profiling – females



**Monica, 24**

Nurse

I'm willing to lie about how we met.

**Eat Pray Love**



**Rasha, 28**

Highschool teacher

I am here because I believe life begins  
at the end of your comfort zone.

**Make it happen**



**Tiffany, 32**

Accountant

Being both strong and soft is a  
combination I have mastered.

**Cherish yesterday, dream  
tomorrow, live today**

# ♦ Next steps

ML models

Create a model that finds best matches based on user preferences

Streamlike set up

Create an app for Todate in order to capture new customers and expand database.

# Machine learning models

**OneHotEncoder:** Utilized this function in order to transform categorical values into numerical, enabling the performance of ML models and saved in joblib library.

Normalize all the values by using the **Standard Scaler**.

**Train test split:** Setting our target column (Profiles ID) and pre-selecting the rest of the features that needed to be considered . Setting test size to 20%

## **K – MEANS CLUSTERS**

Used to group profiles into clusters based on their feature similarity, in order to find similar profiles within the same cluster.

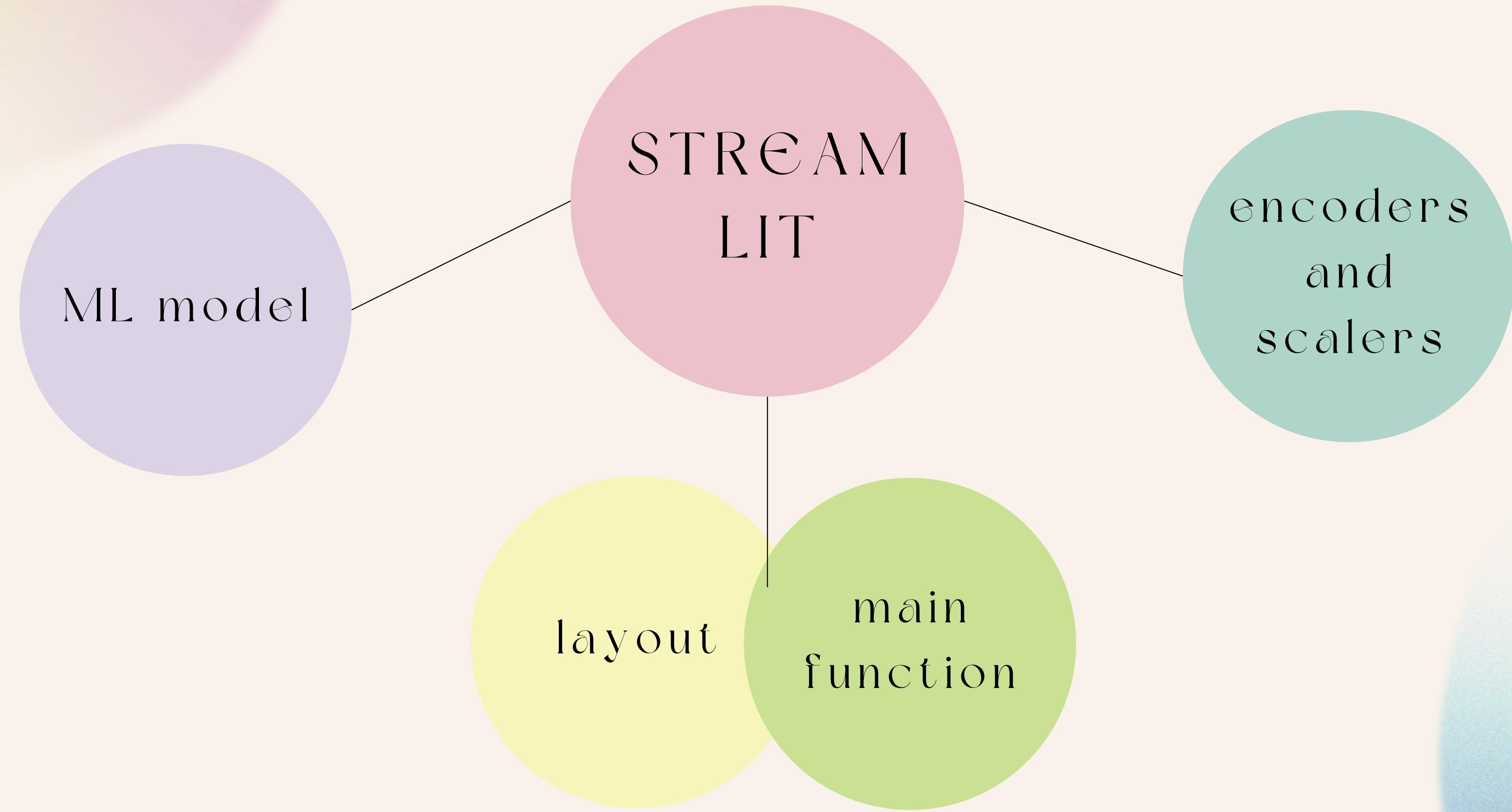
Evaluation metric: Silhouette score

Score obtained: 0.10 (after changing features and adjusting the number of clusters)

Indicates that the current clustering solution is not effective because they are poorly separated, may be due to a high dimensional data.

## **K – NEAREST NEIGHBOURS**

Unsupervised machine learning technique in order to find the 20 most similar profiles to a given one, based on the relevant features.



## ◆ Conclusions:

After a thorough data analysis in order to have a deep insight on our customer database and machine learning techniques, Todate aims to set new standards in the online dating industry, making the search for love a more personalized adventure.

# ♦ Other actions to consider:

- 1 Continuous Improvement of Matchmaking Algorithms
- 2 Data privacy and security
- 3 Community building
- 4 Global expansion
- 5 Partnership and collaboration
- 6 Company growth

Thank you!

Marina