MateFinder - The Next Generation of Dating Recommendation System

Chuqiao Ren(cr2826);
Jinyang Yu(jy2803);
Lyujia Zhang(lz2467)

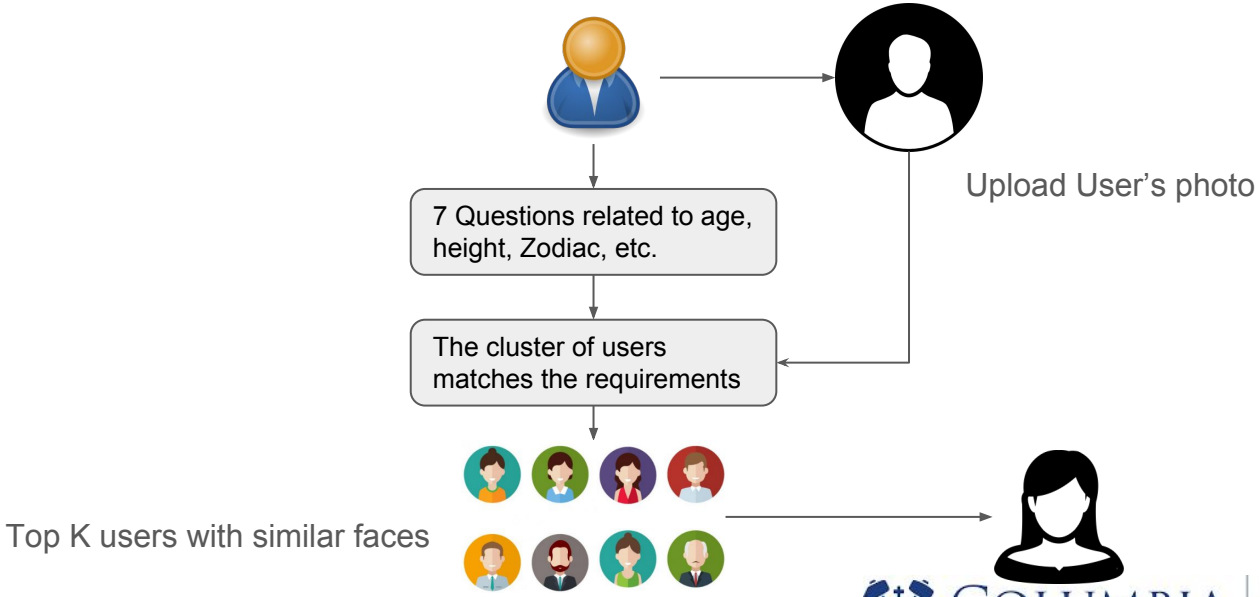
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Motivation

- Large market for dating services
  - 50M single adults in the US
  - 49M people have tried online dating services
- Many dating apps
  - Tinder
  - OKCupid
- Few personalized recommendation services
  - Find the best fit based on your **requirements**
  - Narrow down your search based on your **face**
    - Convergence in the physical appearance of spouses [R. Zajonc et al.]
    - Goleman, Daniel. "Long-married couples do look alike, study finds." *New York Times* (1987).

Overview

- Implement a web application, **MateFinder**, to search the best fit among registered users based on your requirement in real-time.



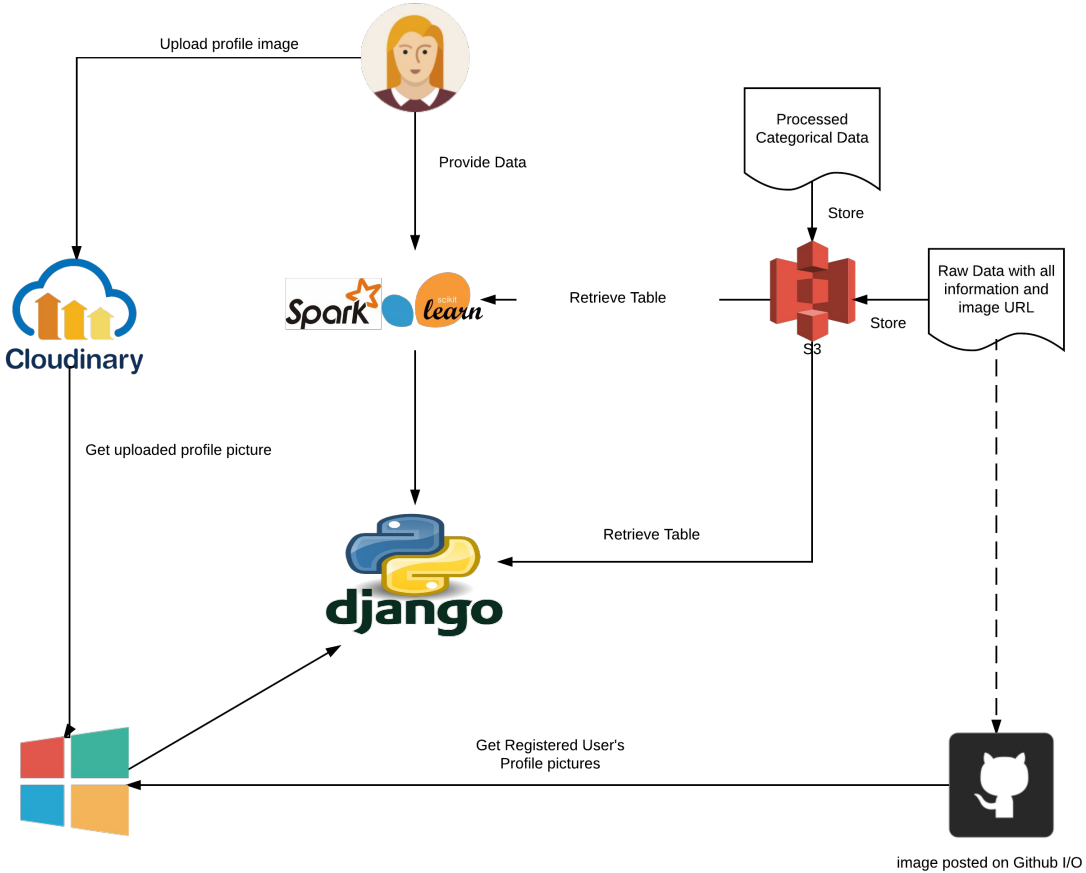
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Technology

1. Preprocess to numerical values
2. K-mean clusters
    a. Spark
    b. Python

Microsoft Face API

1. Recognize face in each photo
2. Find top k similar faces to user's photo in the existing database based on Convolutional Neural Network

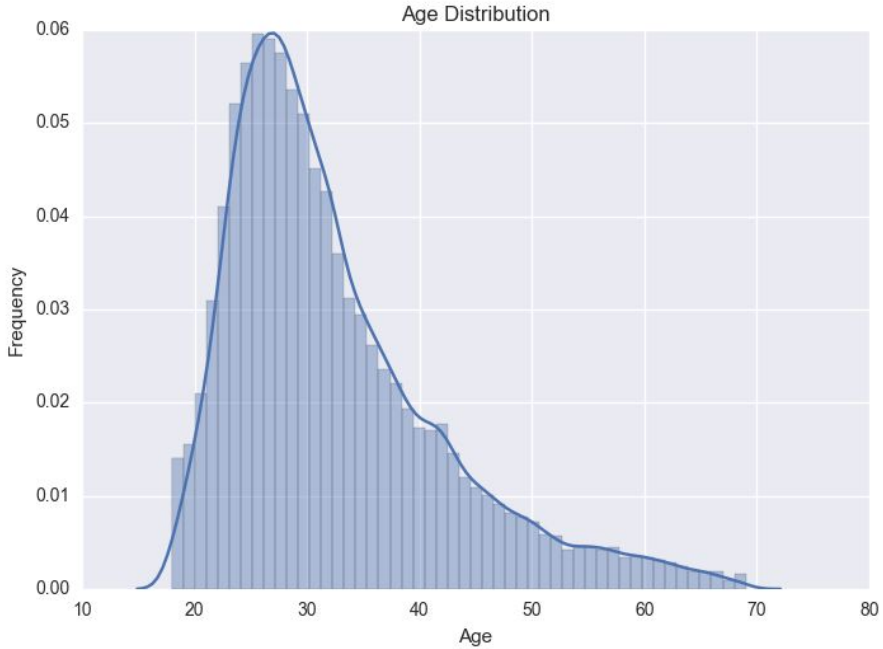


Data Set

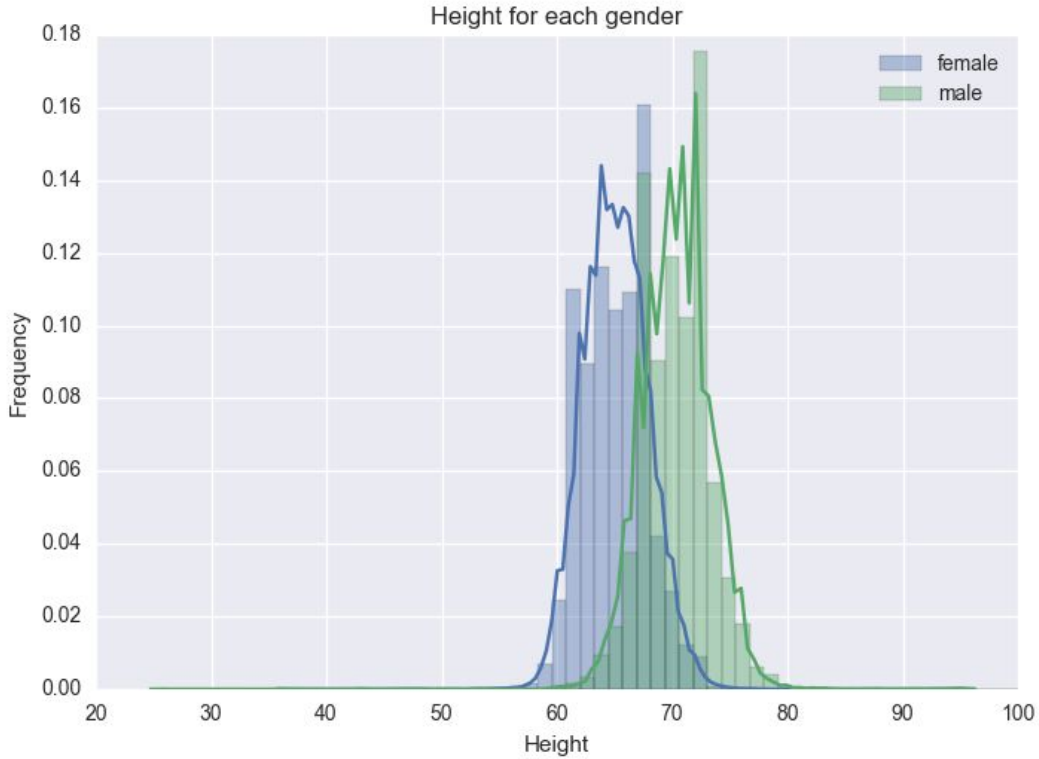
- OKCupid Profile Data
  - ~50k user profile data
  - Data contains user basic information such as height, gender, Zodiac
- Chicago Face Dataset
  - ~1000 photos
- Big Data Analytics Class Face Dataset
- Pre-process
  - Convert text content to numerical values -> upload to AWS RDS MySQL
  - Image content upload to Github io. Query URL stored in AWS with its respective ID.

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Data Visualization



Data Visualization (cont.)



Data Preprocess in Details

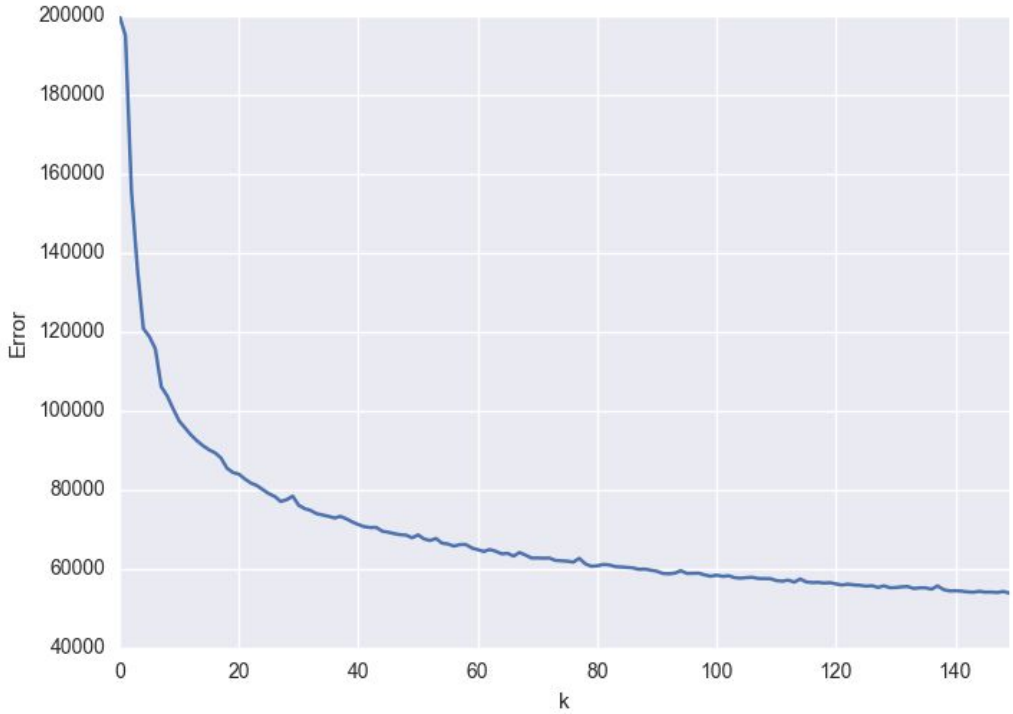
- Age: 13 groups
- Drink: 6 groups
- Education: 5 groups
- Height: partition on Sex, each 7 groups
- Target Sex: 3 groups
- Zodiac: 12 groups
- Smoke: 5 groups

Kmeans - Result

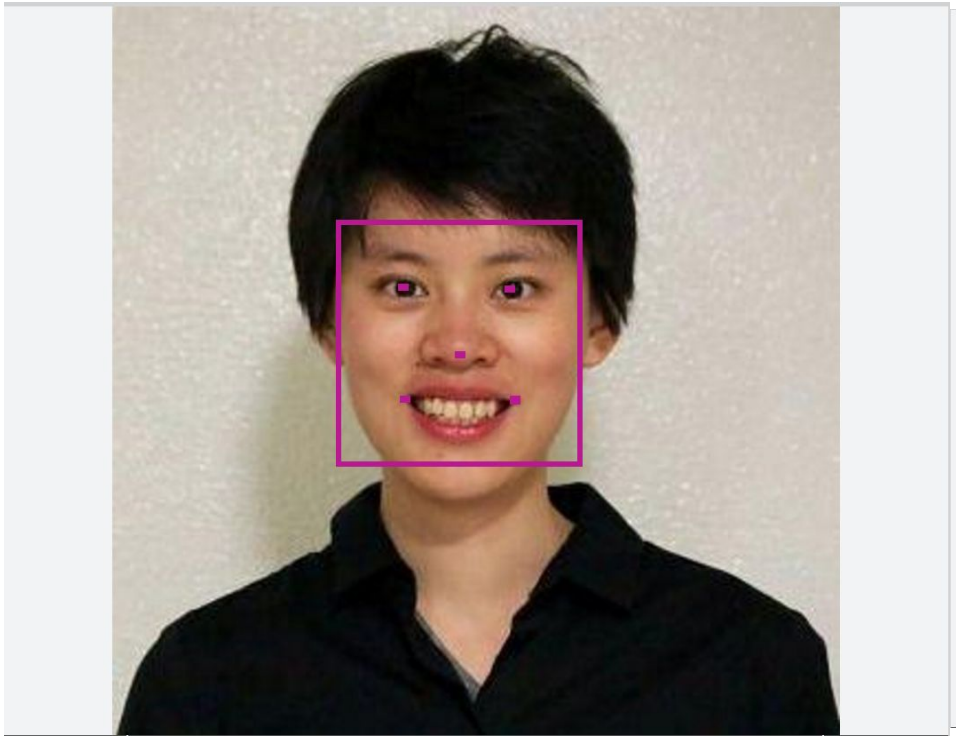
K = 15

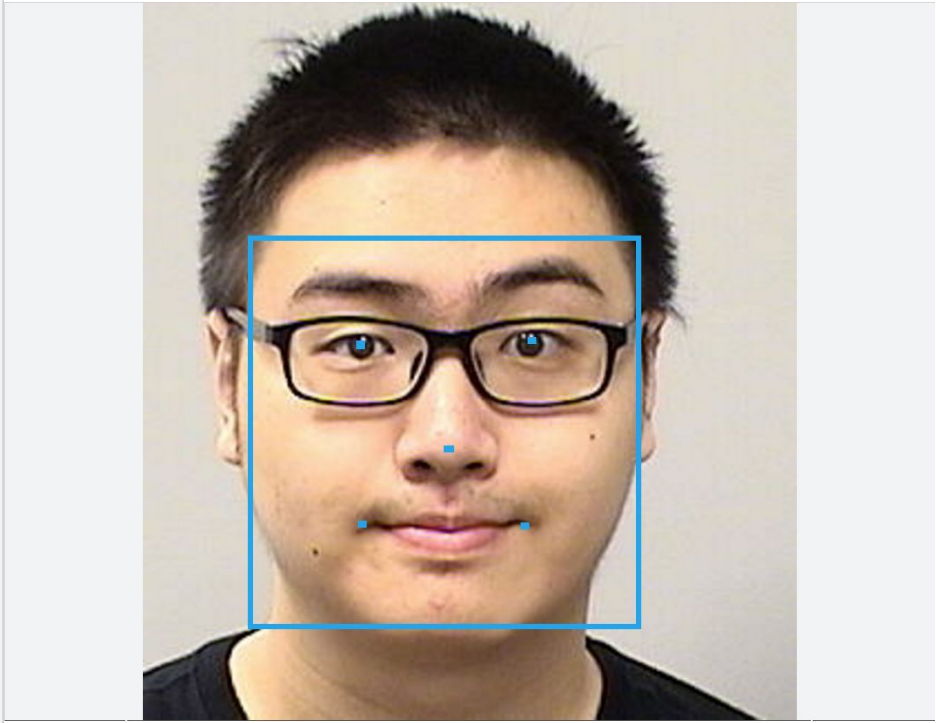
Each cluster has:

- 78.5 male
- 60.5 female

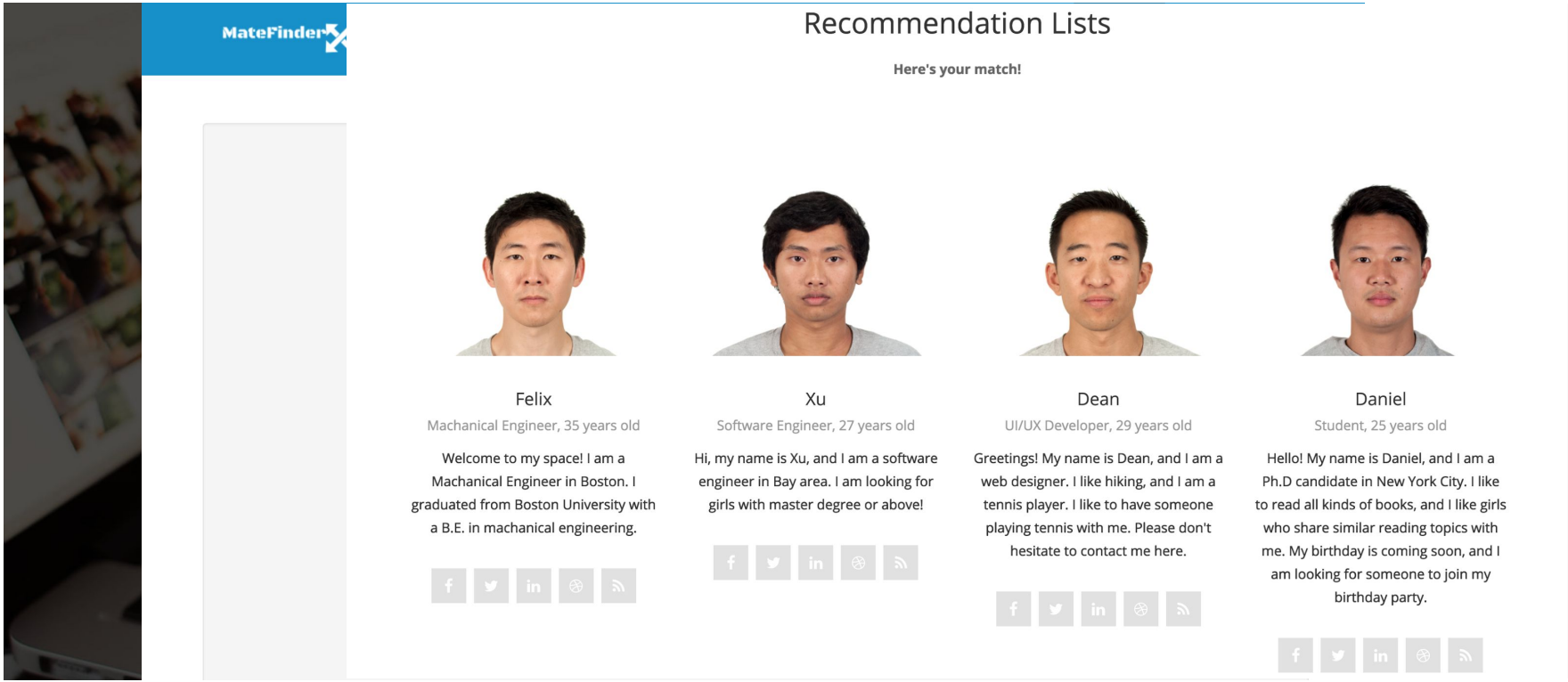


Microsoft Face API - Result





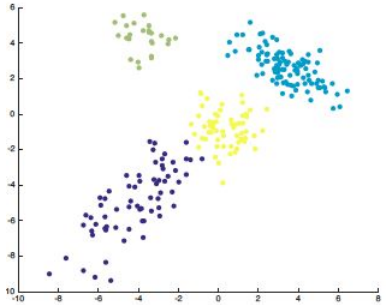
Demo



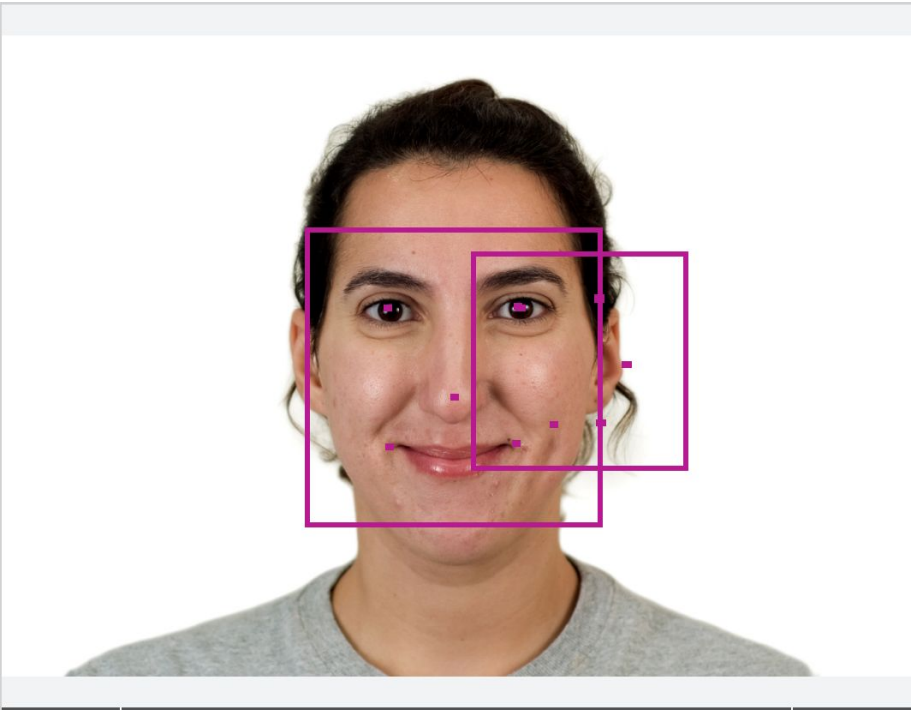
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

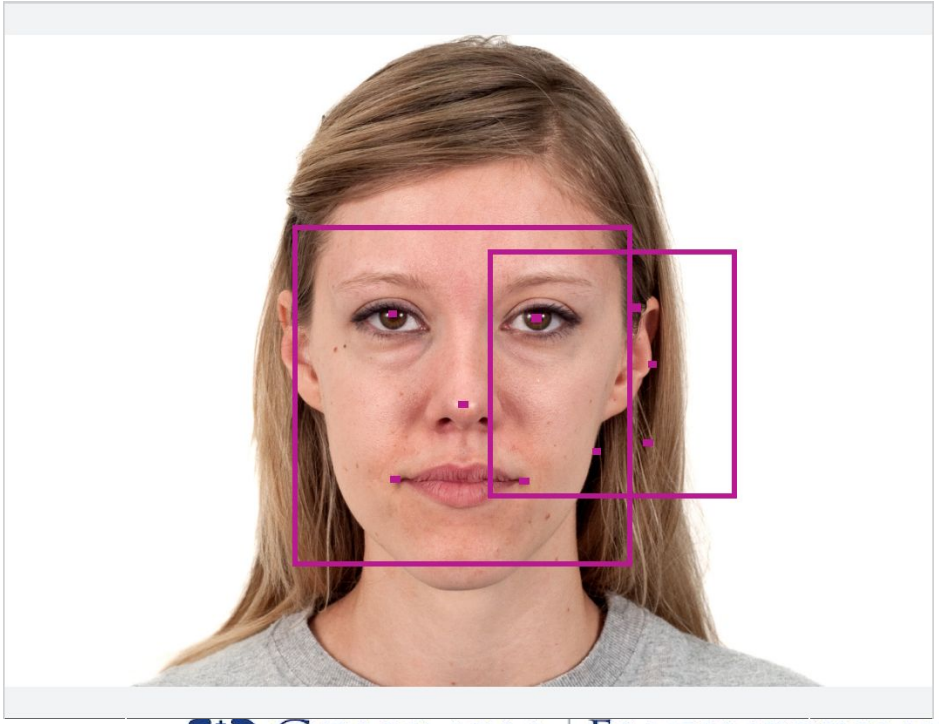
Challenges

- Public available dataset does not contain photo
  - Use Chicago Dataset and class profile photo instead
- Real-time calculation is time consuming
  - Decision tree takes longer time
  - Move to unsupervised k-mean clustering
- Fake data
  - Crawling all information from OKCupid.com without filtering
  - Normal distribution to filter out the noise (e.g. a user with 1 feet in height or 1 lb in weight)



Issues with Microsoft Face API





Conclusion

- Implemented a web application to recommend the best fit user based on your **requirements** and your **outlook**
- Used non-trivial **Spark** techniques learned in class to implement the core algorithm
- Learned how to use Django, Bootstrap 3, Node.js and more

Next Step

- More users post information and their requirements to validate our result
- Extend our product from a web application to mobile end
- Testing on photo with non-frontal faces or smiling faces

Thank you!

Chuqiao Ren(cr2826); Jinyang Yu(jy2803); Lyujia Zhang(lz2467)

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science