



**okcupid**

## FUN FACTS ABOUT SINGLE MEN AND WOMEN

Some exploratory analysis based on OKCupid dataset.  
by Stuart Xu



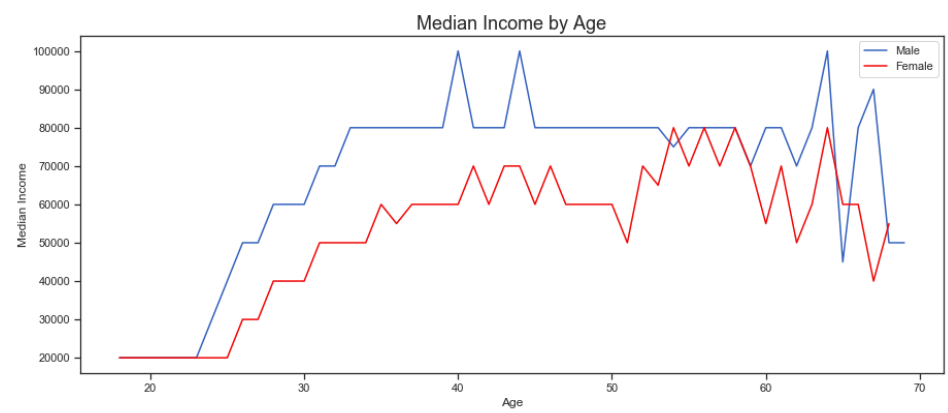
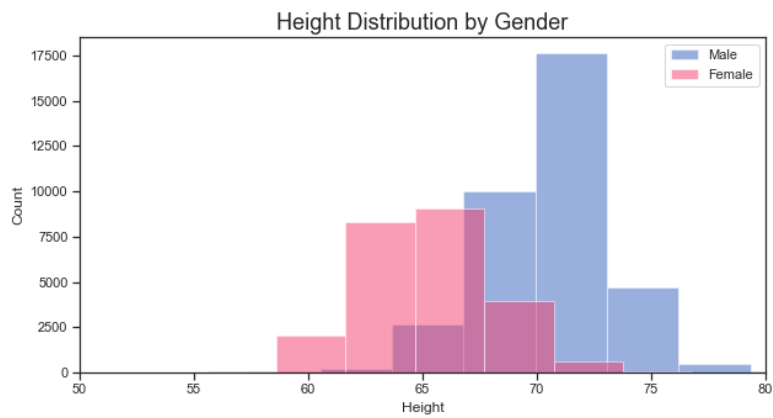
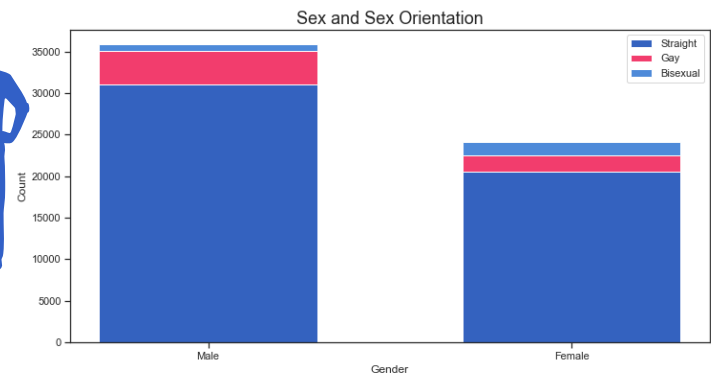
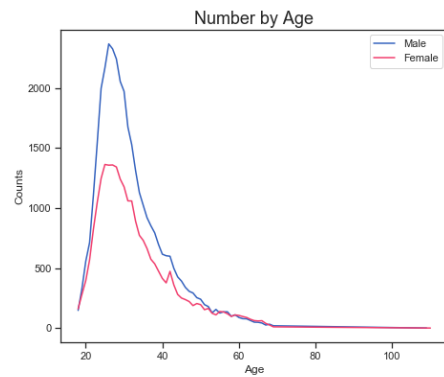
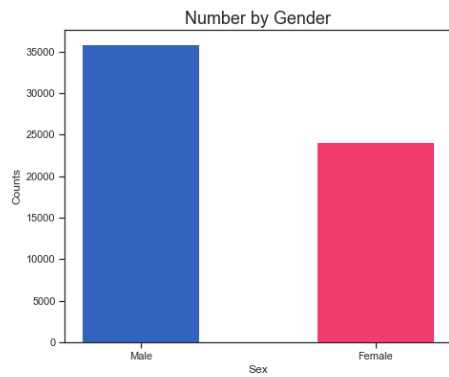
- OKCupid is a popular dating app for people to find mates.
- We are curious about these single male and female's online profiles, self-descriptions, and even their behavior.
- With 10 columns of self-description on their profile, it would be a good training data for future NLP.



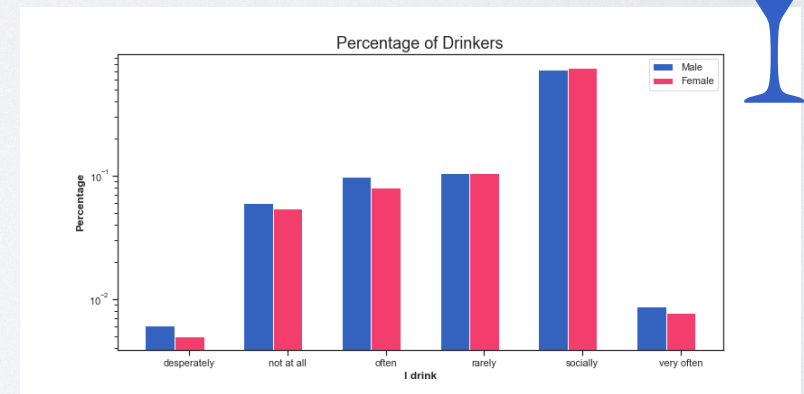
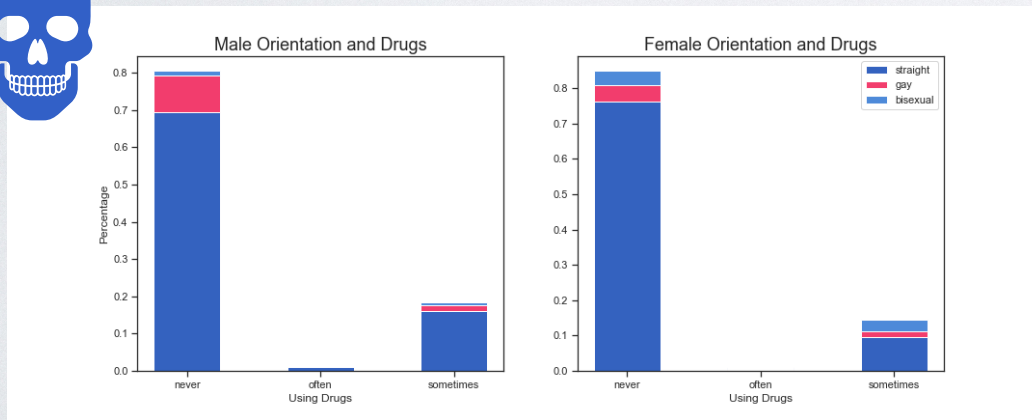
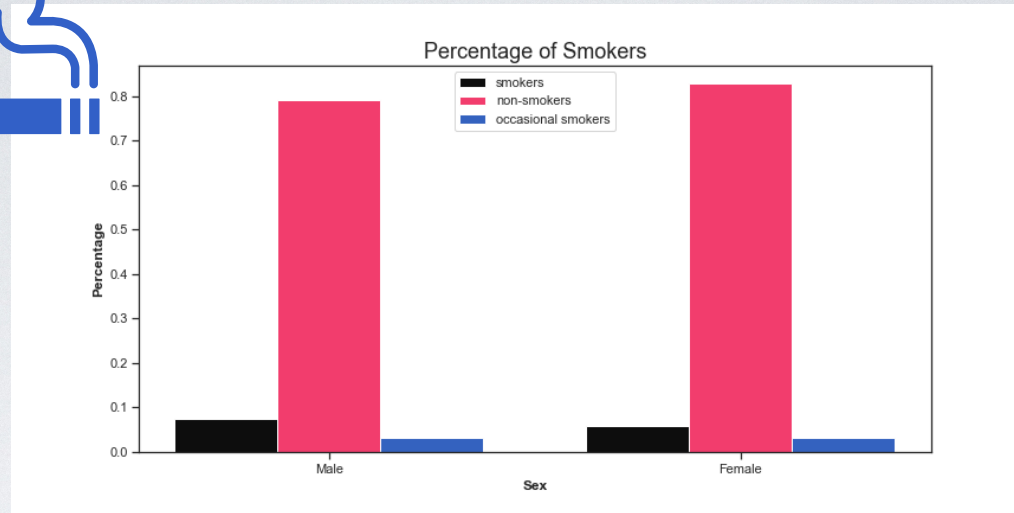
## MOTIVATION

## DESCRIPTION

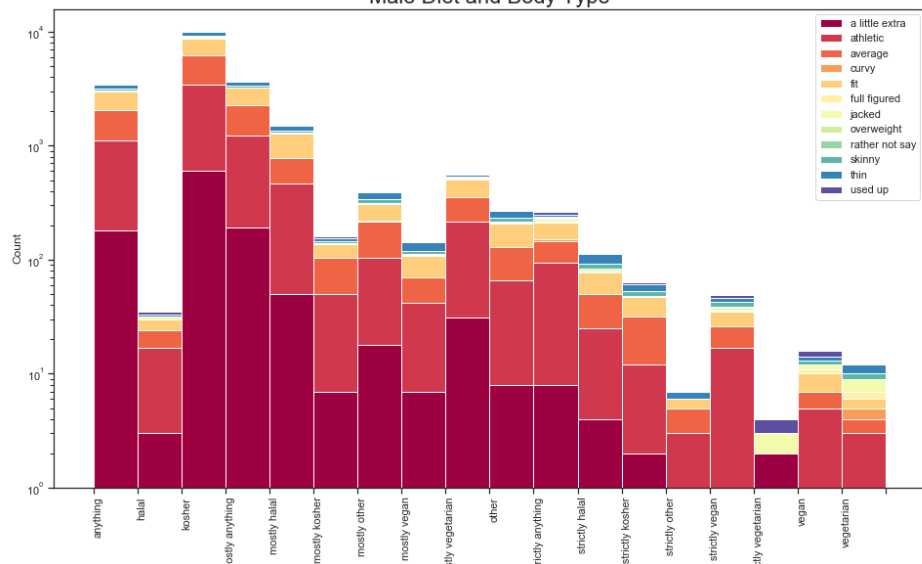
- The OKCupid user profile dataset contains 31 columns with over 50,000 entries collected in 2012 in San Francisco area.
- The analysis focused on the demographic info of the single male and the single female, feature distribution of smokes, drugs usage, drinks, diet and body type, income disparities, etc.
- The 10 essay columns with self-introductory sentences are processed with NLP tools to remove signs and stop words, so word use frequencies are calculated.



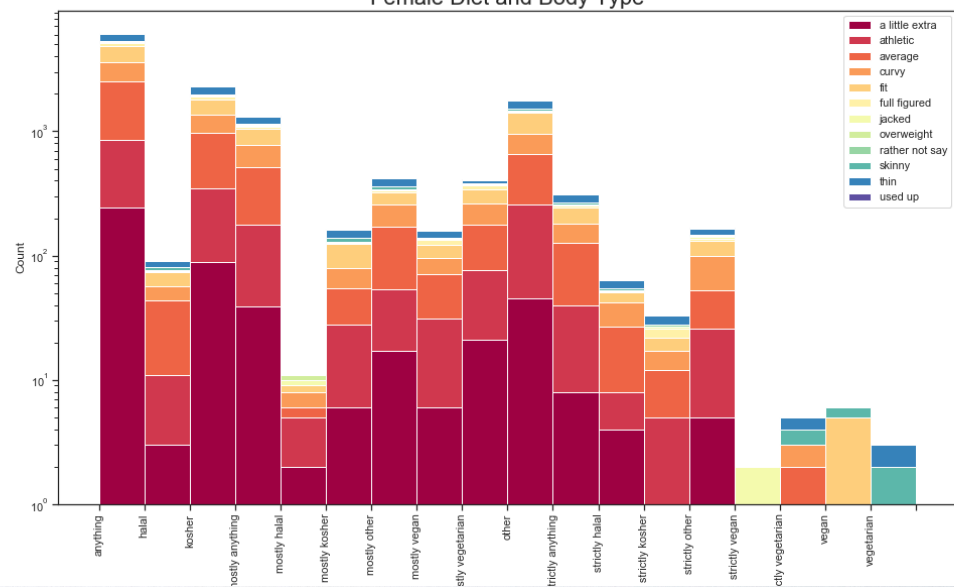




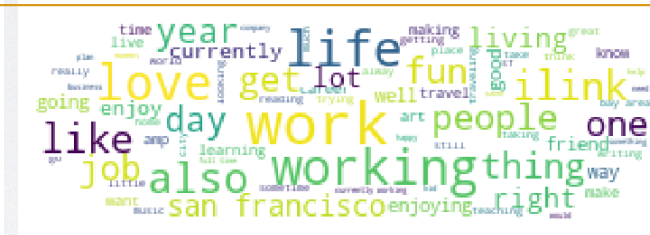
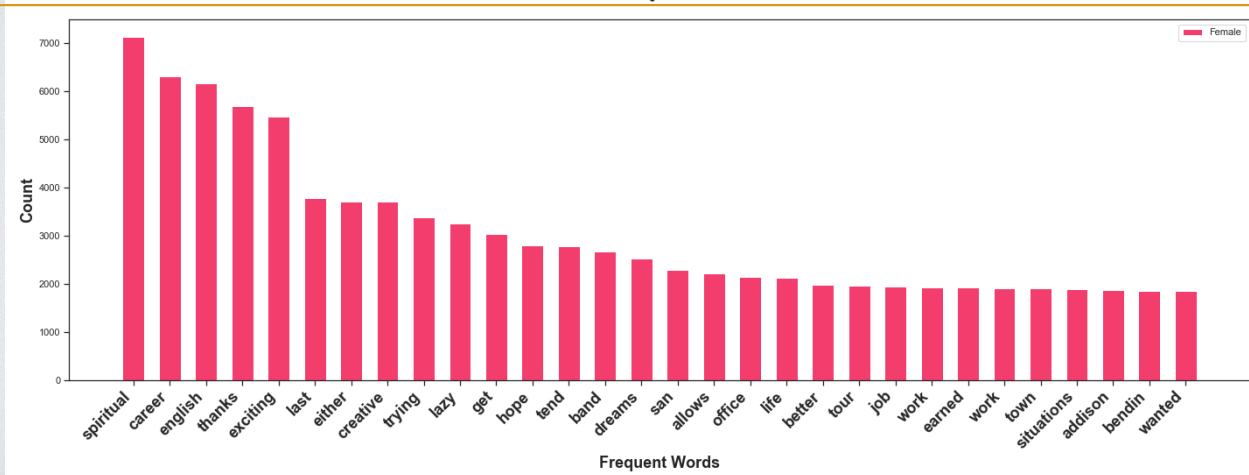
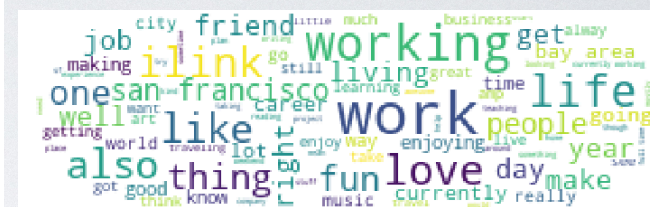
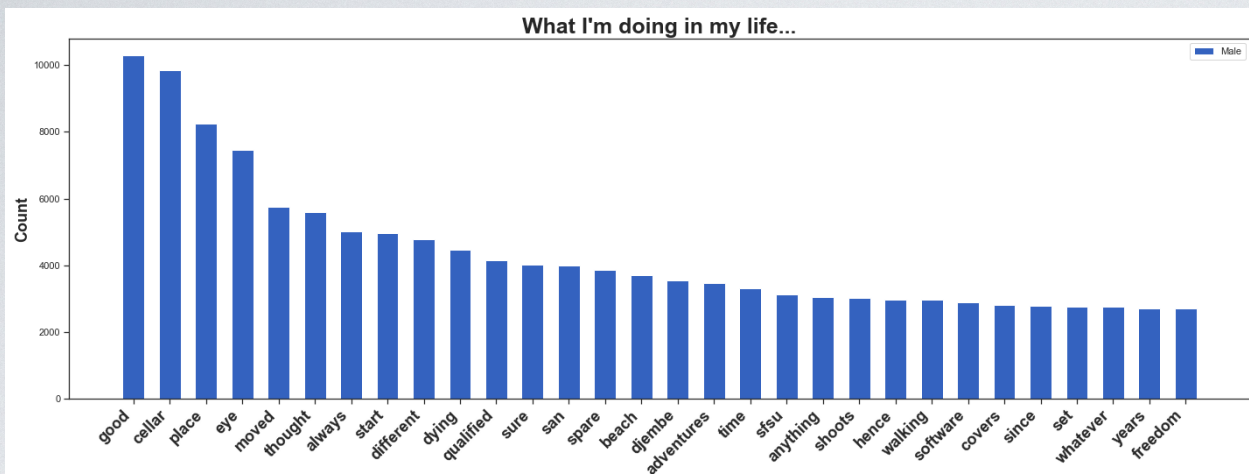
### Male Diet and Body Type

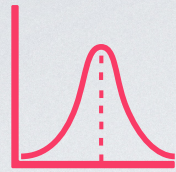


### Female Diet and Body Type







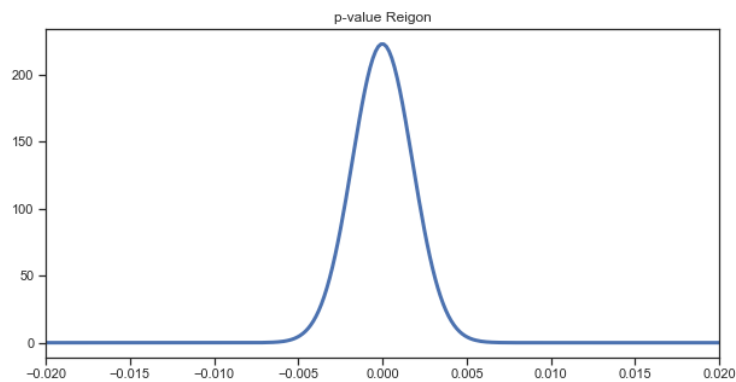


## Hypothesis Test

$$\begin{cases} H_0 : P_m - P_f = 0 \\ H_1 : P_m - P_f > 0 \end{cases}$$

### Two Sample Z-test

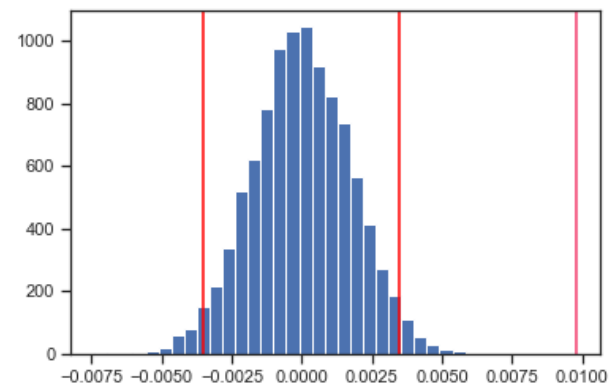
P-value = 0



### Bootstrap Sampling

The confidence 95% interval is  
 (-0.003522081886218326,  
 0.0034644278376134343).

P-value = 0





“Don’t use bootstrap sampling.”

– Hamit



## Conclusions:

1. There is difference in many aspects between the single male and single female. 😜
2. A lot of single men are good at frisbee... (really?) 😏
3. Both of them think they are good at cooking. 🤔
4. Gay men have a much higher percentage of drug users than lesbians. 🤯
5. Since we rejected the  $H_0$ , we can say the single male are more likely to smoke than the single female. 🤔



## Advice to myself:

Don't be lazy...

Always have a goal and keep the goal in mind during the project.