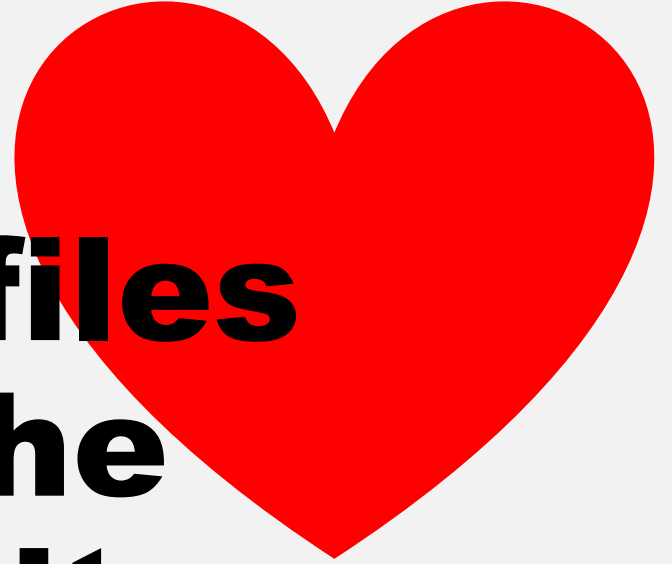


# **Analysing the profiles of members of the OkCupid dating site**

**Data analysis and interpretation**

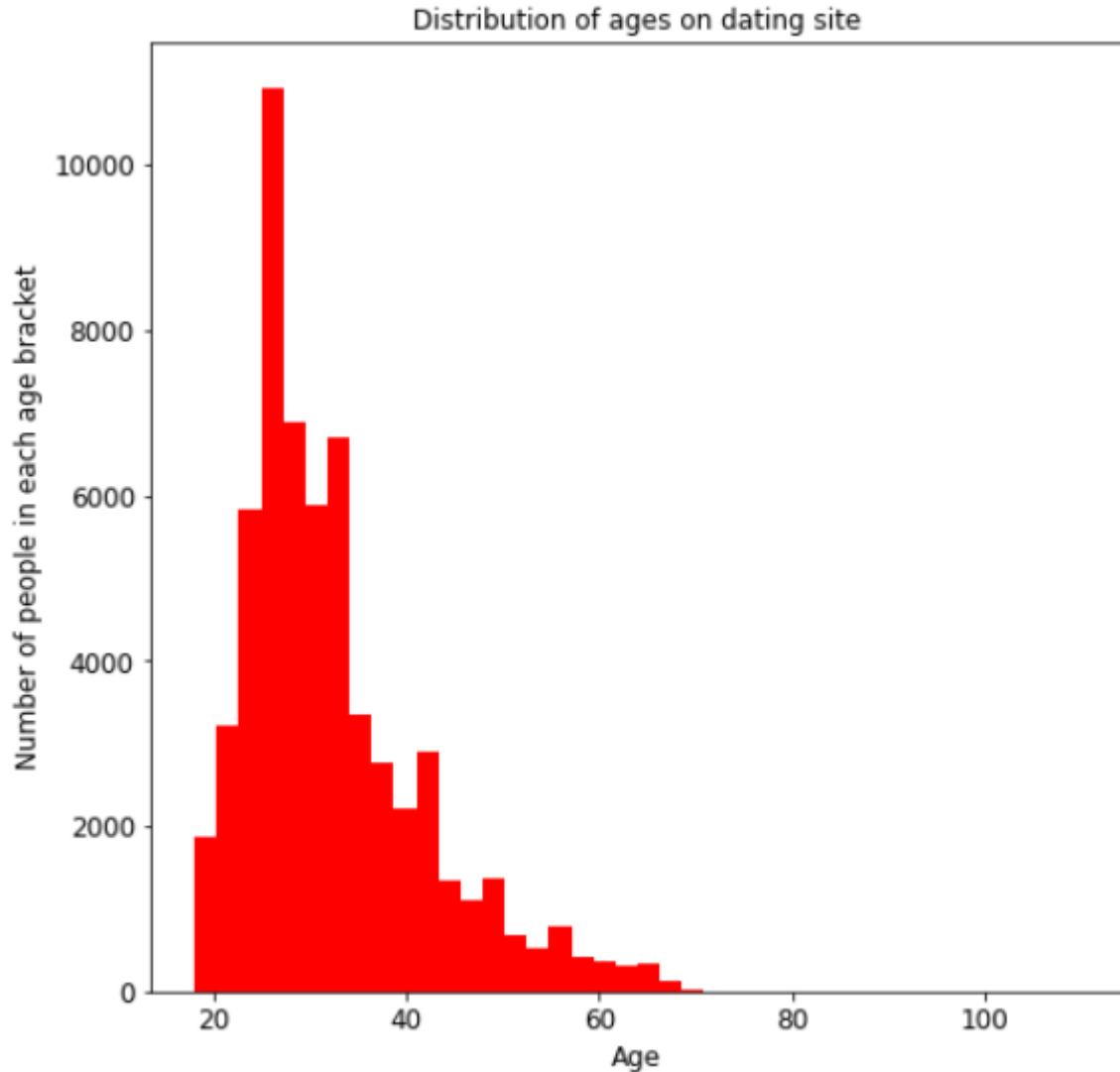
**By Amy Birdee**



# Introduction

- The data consist of circa 60,000 profiles on OkCupid's member database during the 2011-2012 period
- This project aims to segment the data and discover patterns in the member profiles, e.g. what is the age profile of members? What are their eating and drinking habits? Is there a link between their level of education and drug use? Are they even single??
- The main findings in the data have been presented in graphical format and the data analysis has been carried out in Python
- The Python libraries used include Pandas, Numpy, Matplotlib and Seaborn

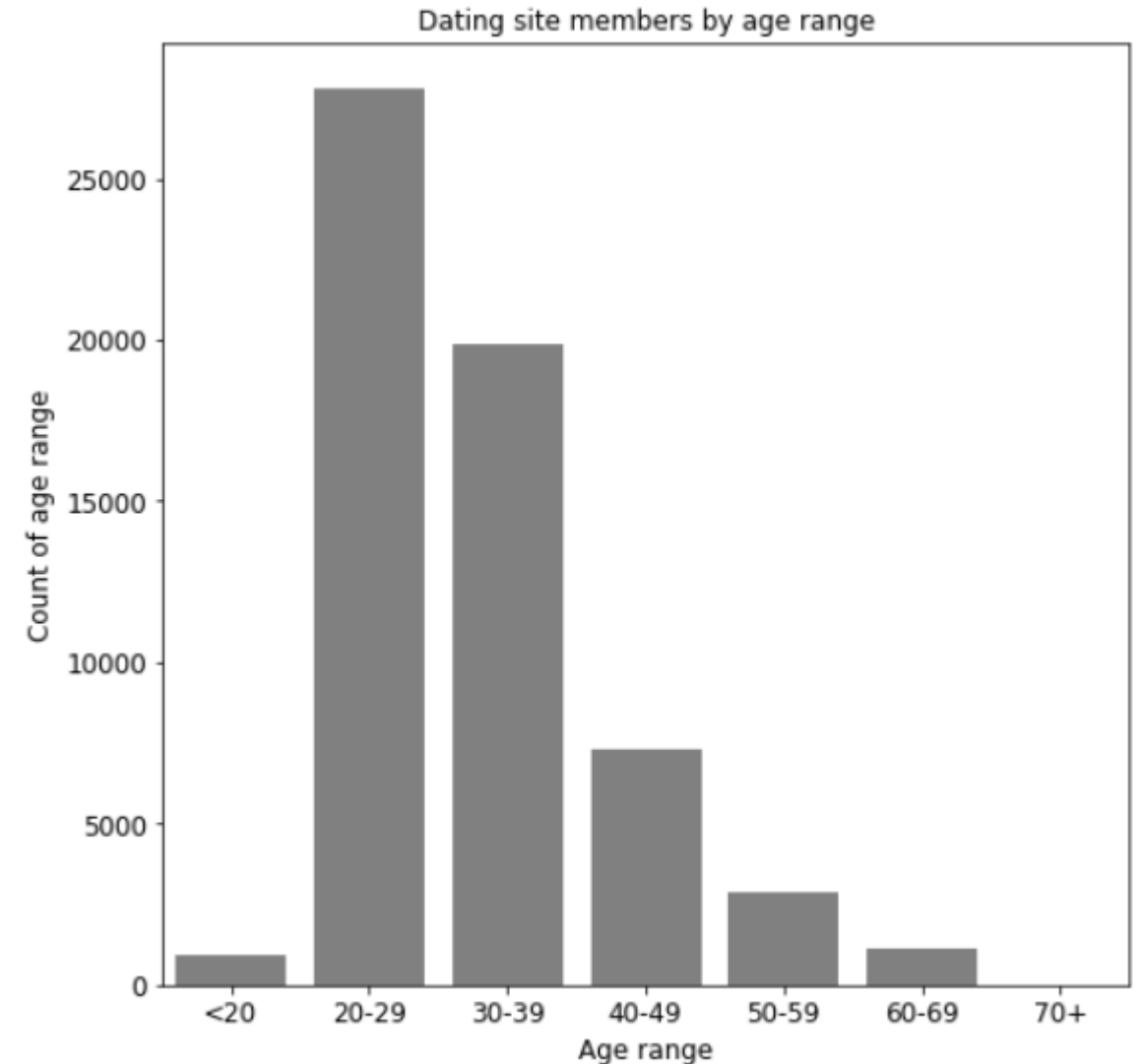
# How old are the members of OkCupid?



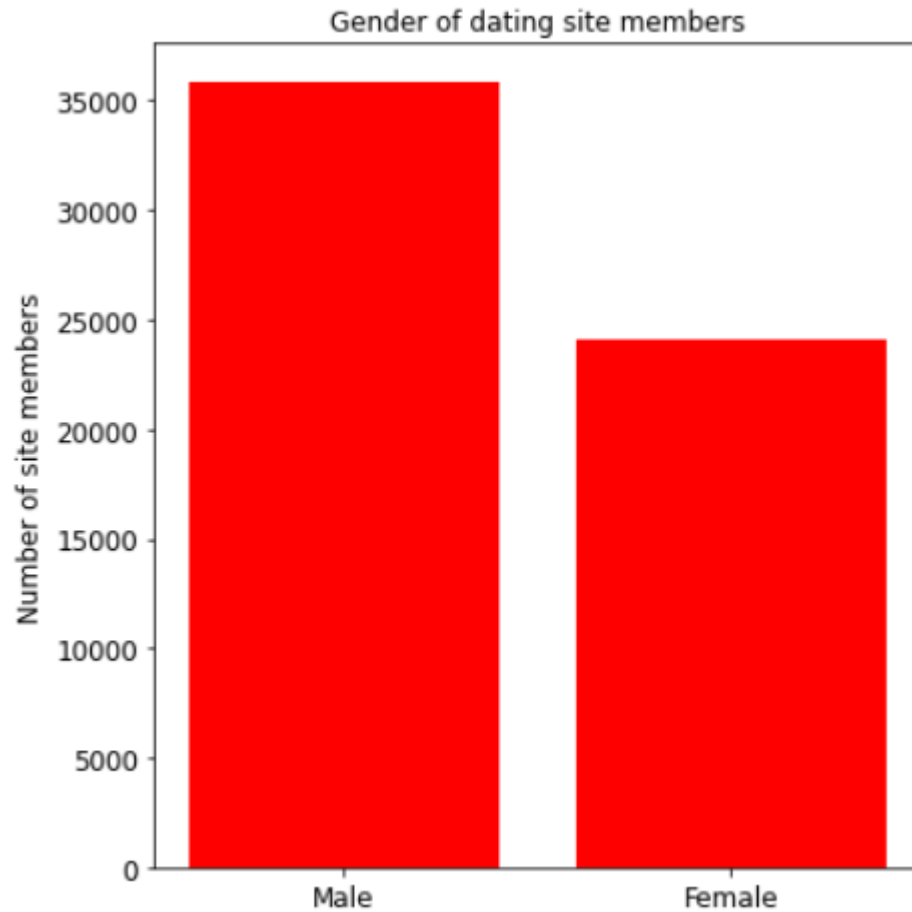
- The histogram shows a right skew and it's evident that the majority of members are in the younger age category although a few members are in the older age bracket
- The statistics show that the average age of members is 32 with ages ranging from 18 to 110!

# How old are the members of OkCupid?

- Despite the average age being 32, the majority of members are in the 20-29 age bracket
- The 20-29 age bracket consists of **27,820 members** which makes up **46%** of all members
- The average age has been pulled up because there are some members in older age groups – **over 1,000 members are over 60** and two of those are **over 100** showing that it's never too late to look for love!

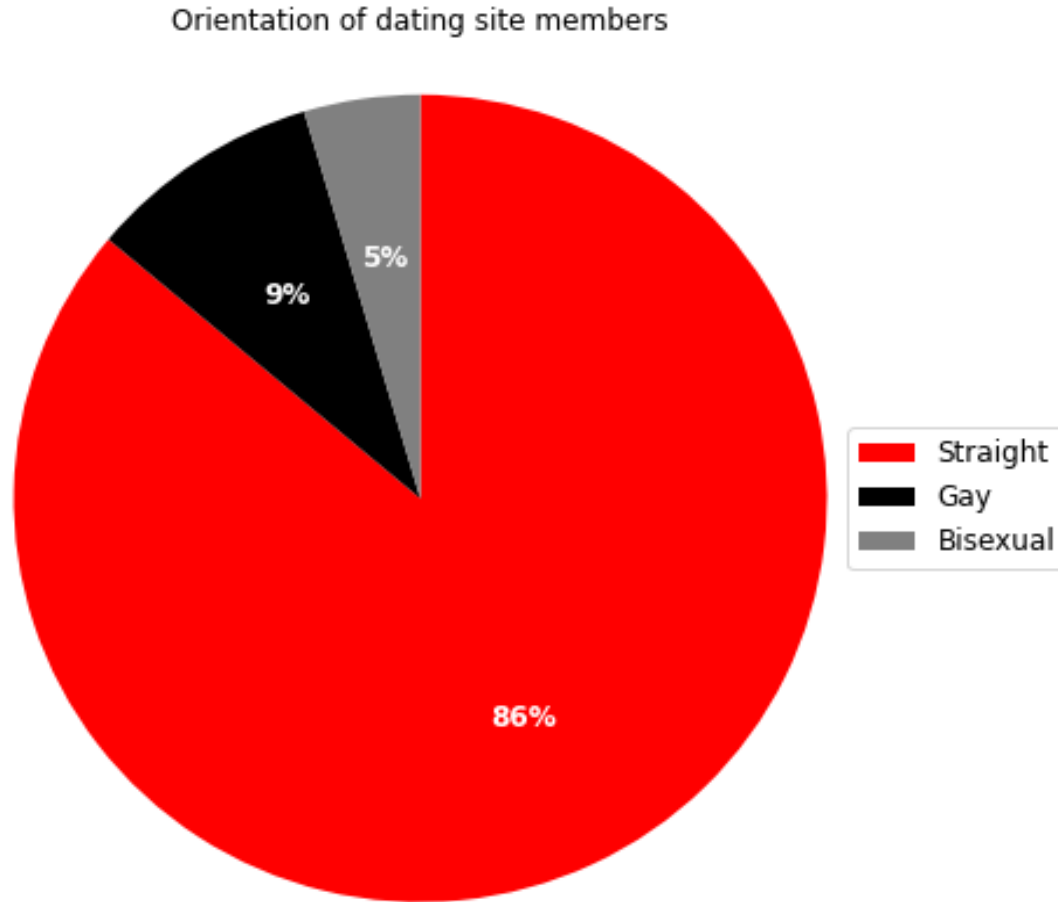


# What is the gender distribution of members of OkCupid?



- The majority of members (60%) are male – there are 11,709 more male members than female
- This could cause a dating imbalance if the majority of members are heterosexual/straight

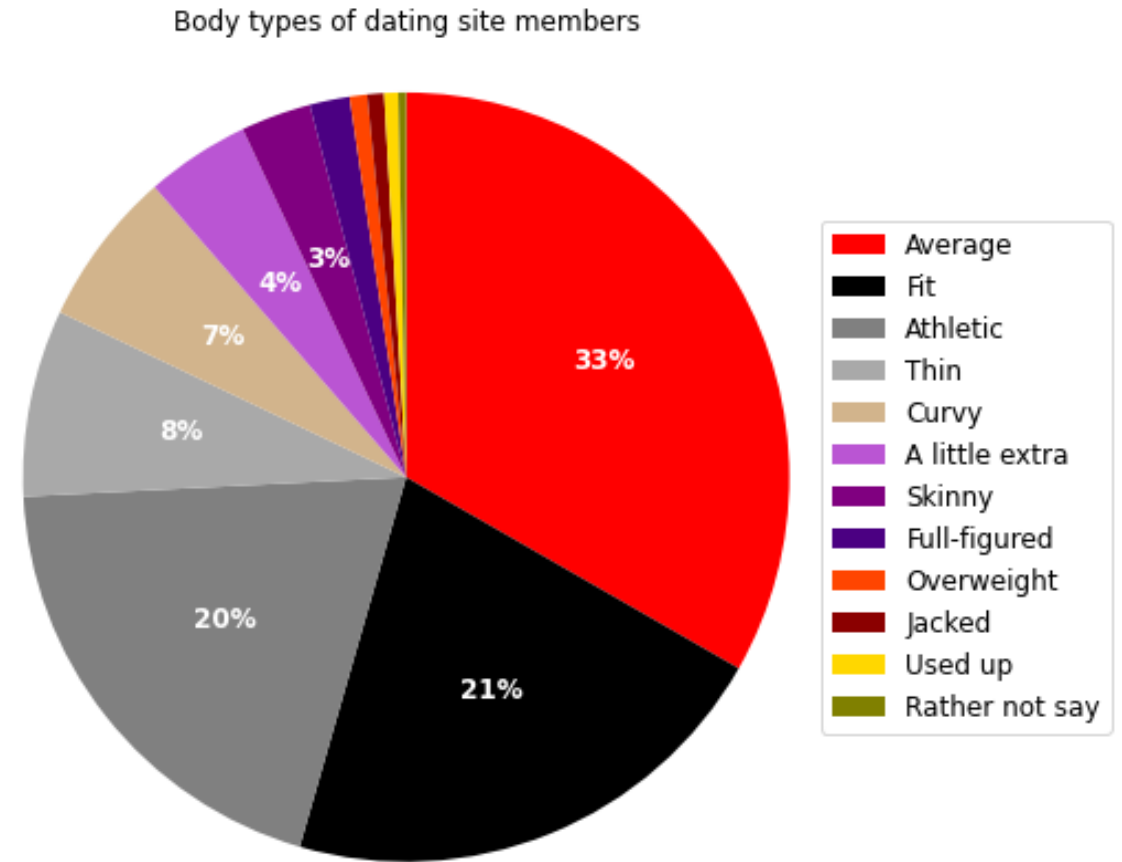
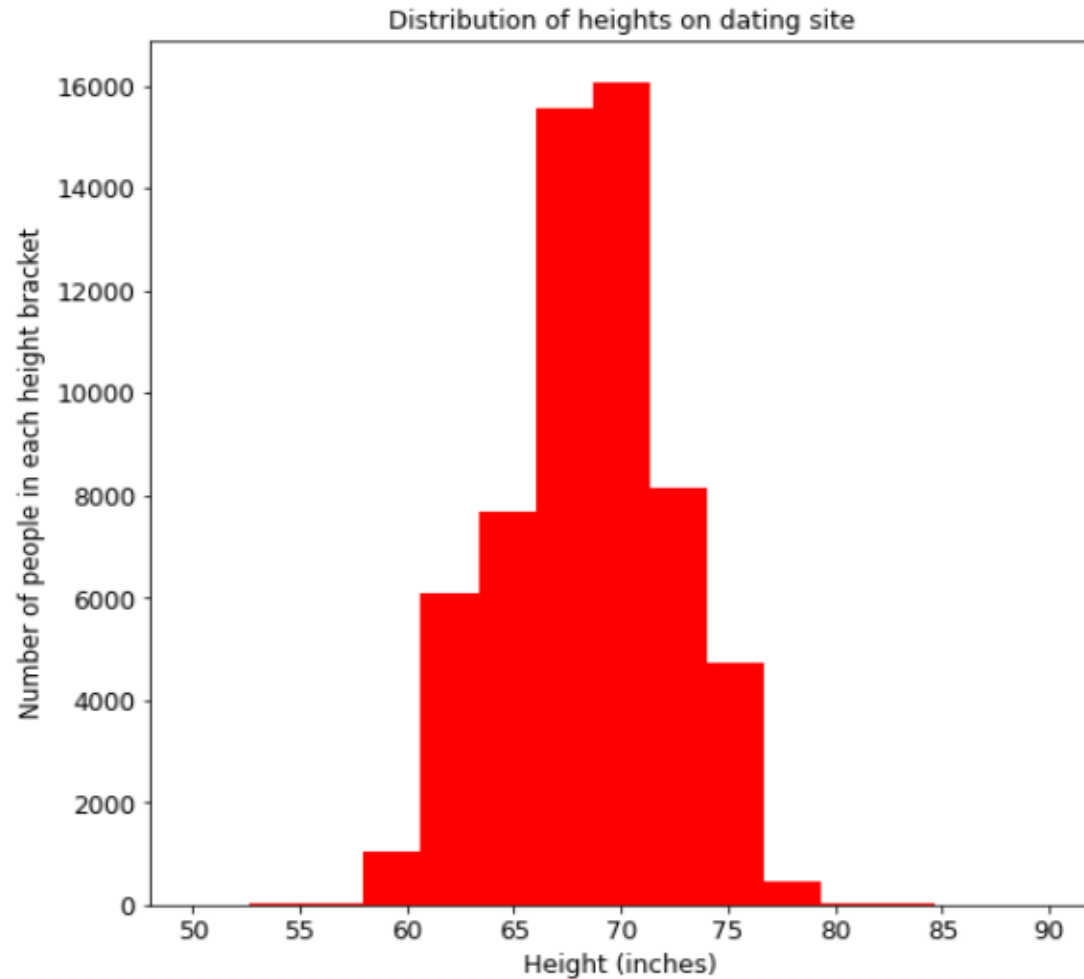
# What is the sexual orientation of members of OkCupid?



- Whilst 86% of members are straight, there are 14% who are gay or bisexual
- This percentage is made up of **8,340 members**
- If the majority of these members are male, this could mean a dating imbalance can be avoided
- If not, OkCupid should work on strategies to attract more females to its site, e.g. a discounted membership for females on International Women's Day

# What are the body shape profiles of members of OkCupid?

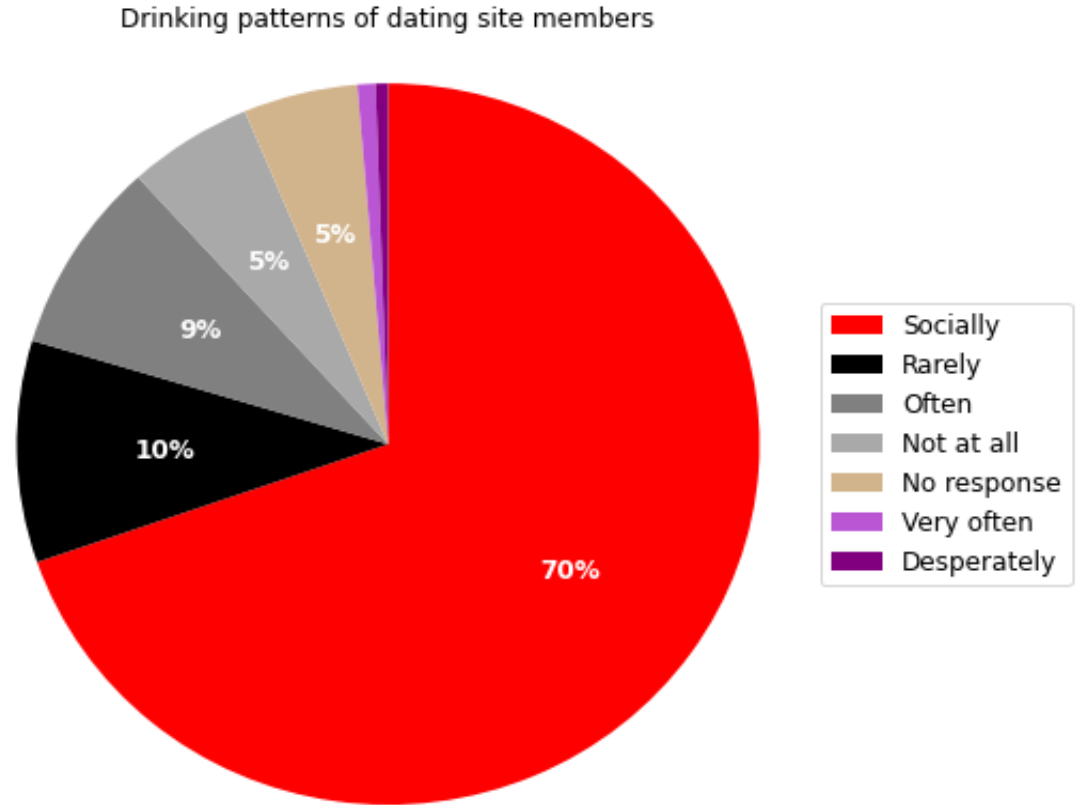
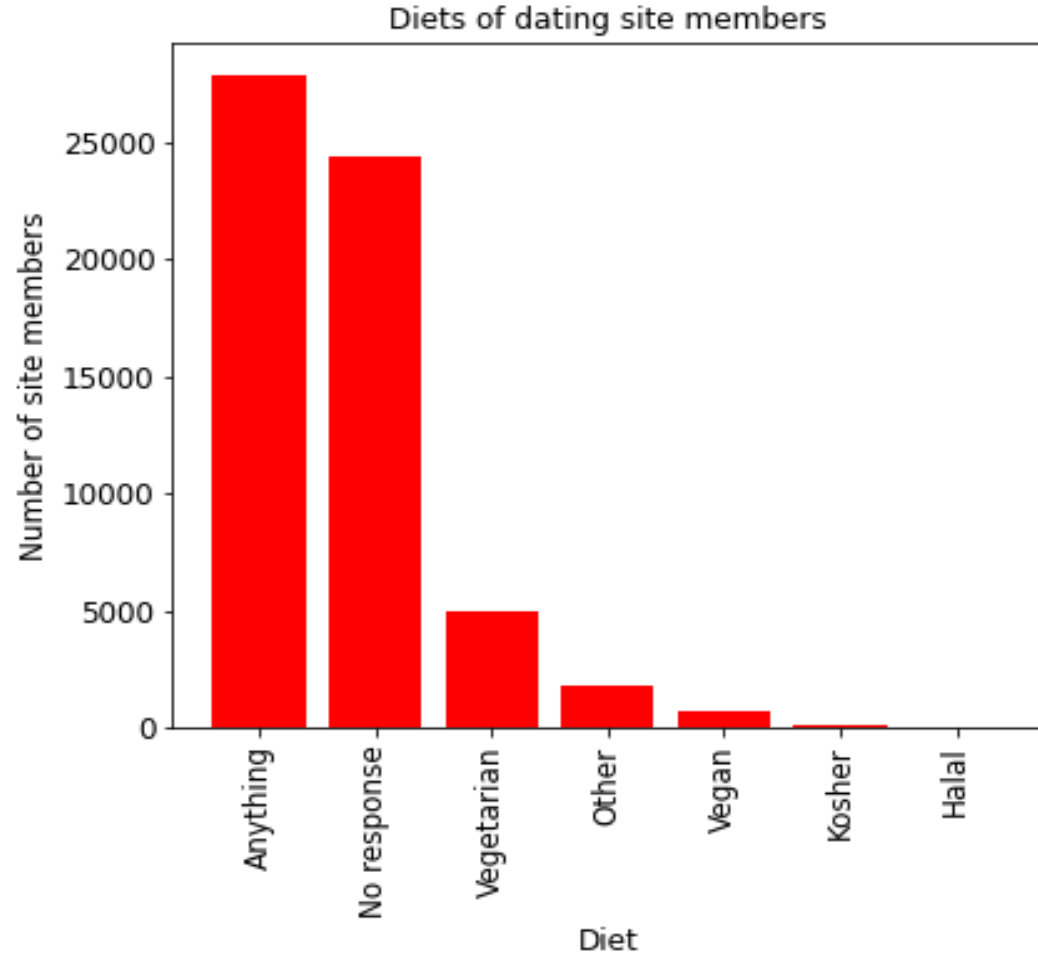
The heights data follow a normal distribution pattern as would be expected in the wider population. Most members are in the 65-75 inches height bracket



The majority of members are of average body type. A minority have said they are 'used up'...maybe best not to ask!

# What are the food and drink habits of members of OkCupid?

The majority of members **eat anything and drink socially**. These data are from 2011-2012 so it's likely that the numbers for vegetarian and vegan diets would be a lot higher today

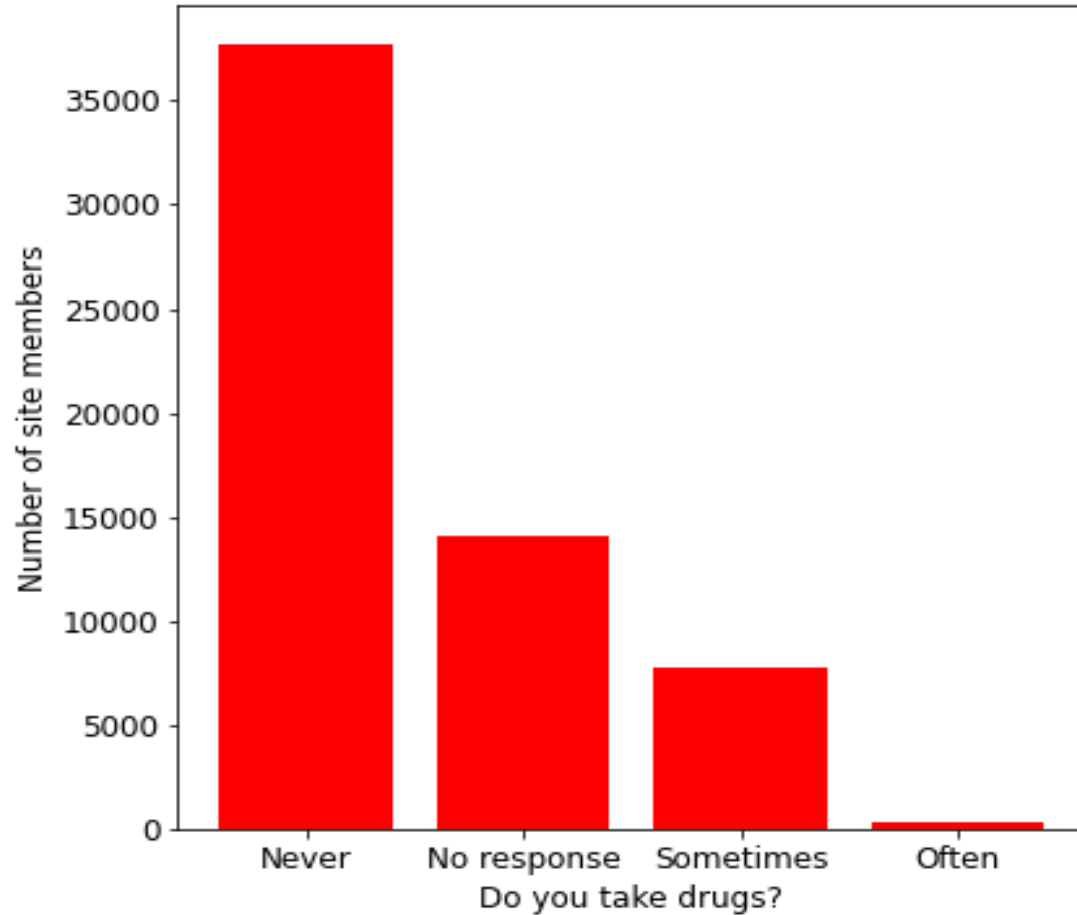


A small percentage of people drink 'desperately' – would it be ethical for OkCupid to ignore this response or should they feel obliged as a company to reach out to these members?



# What are the drug use habits of members of OkCupid?

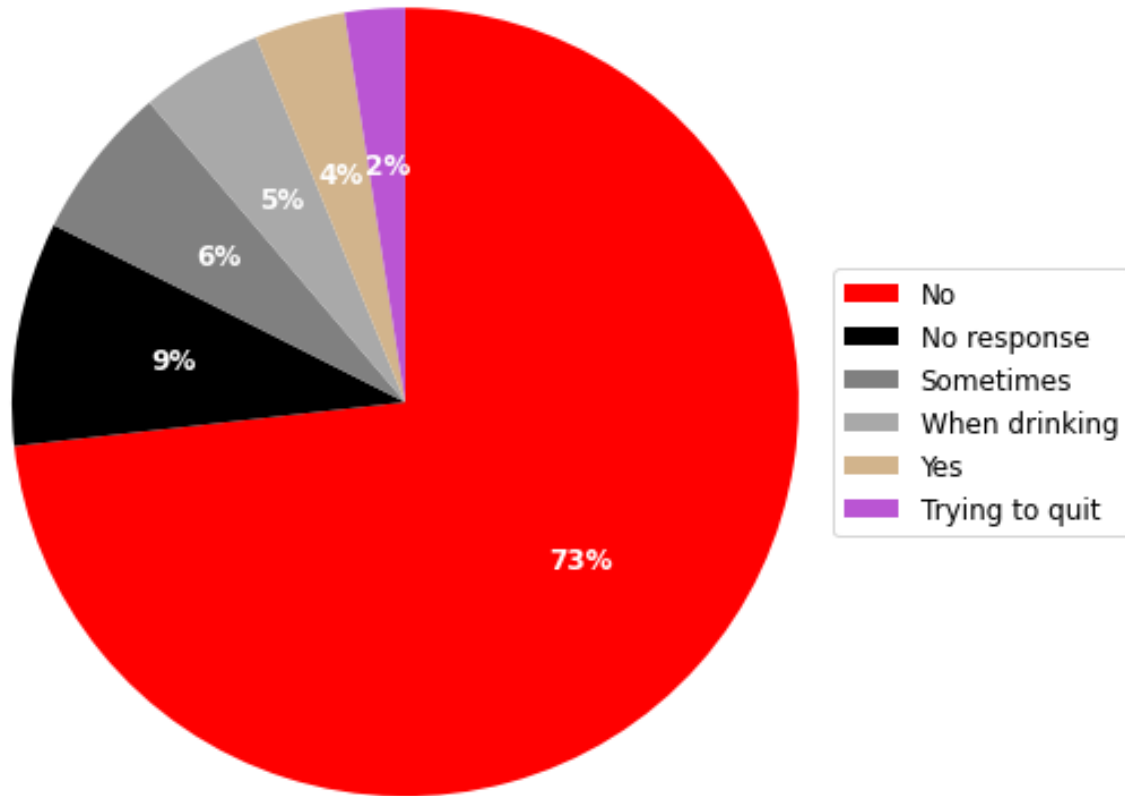
Drug taking habits of dating site members



- Over half of members never take drugs and just over 14,000 members have not answered the question
- This could be because they only filled in the mandatory fields across their whole profile or they may not want to admit to drug taking for fear of being judged by a potential date
- Of all members, 14% say they take drugs sometimes or often

# What are the smoking habits of members of OkCupid?

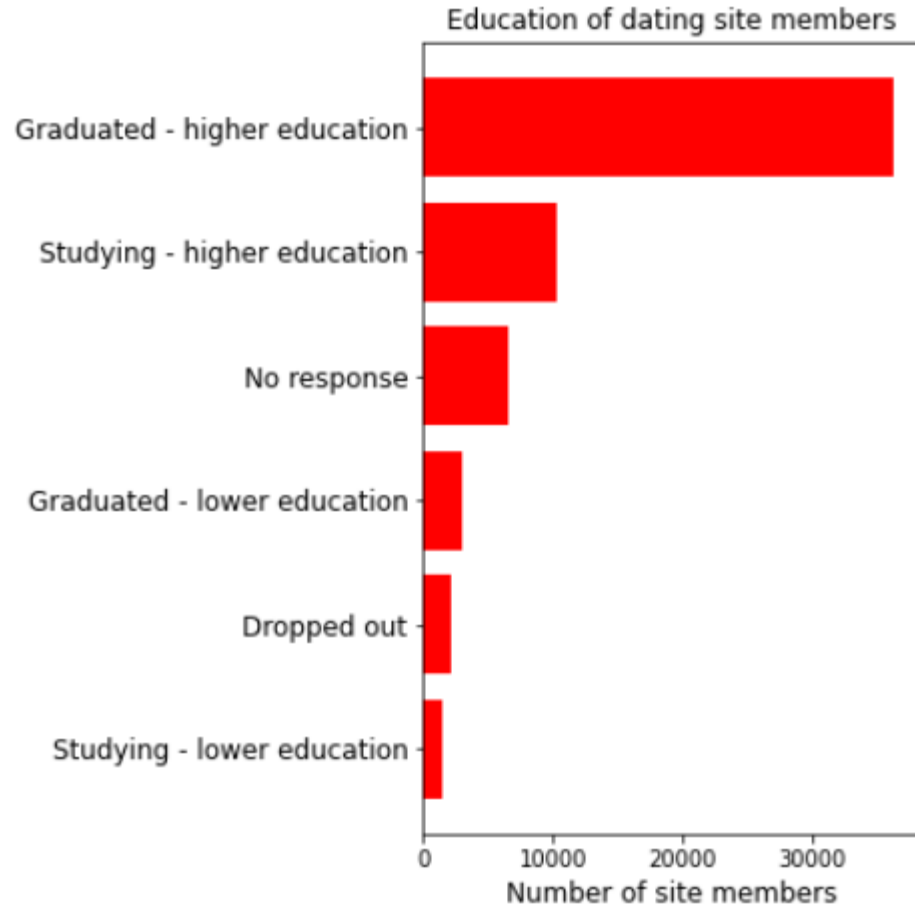
Smoking status of dating site members



The majority of members do not smoke and only **4%** say that they do

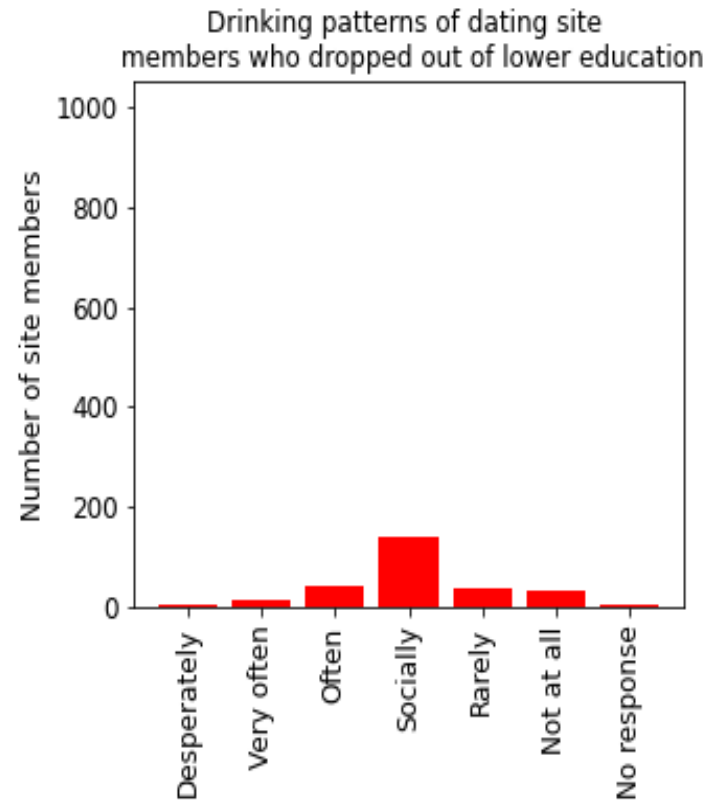
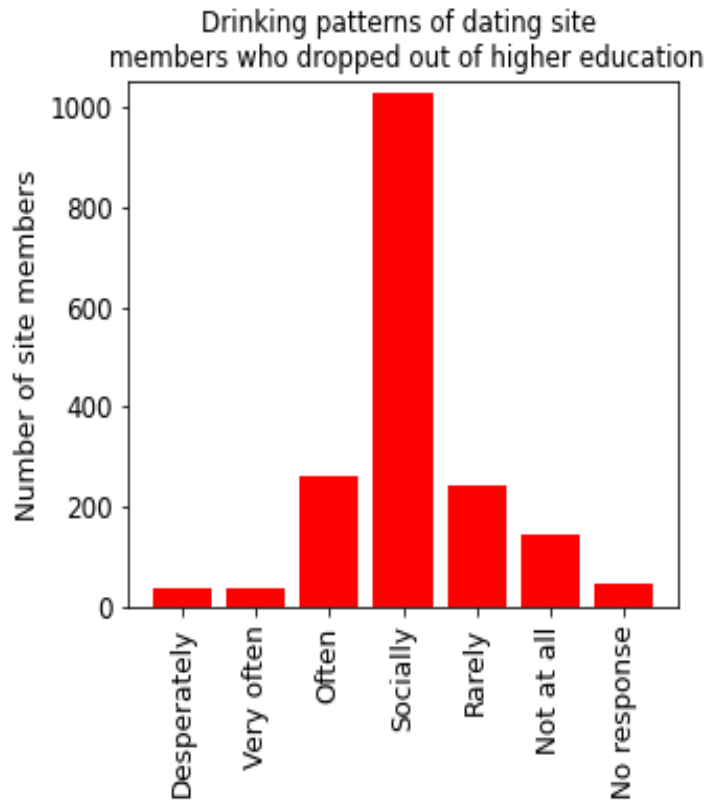
But if this figure is combined with 'sometimes', 'when drinking' and 'trying to quit', the smoking figure increases to 17% which is **nearly a fifth** of the dating population

# What is the education level of members of OkCupid?



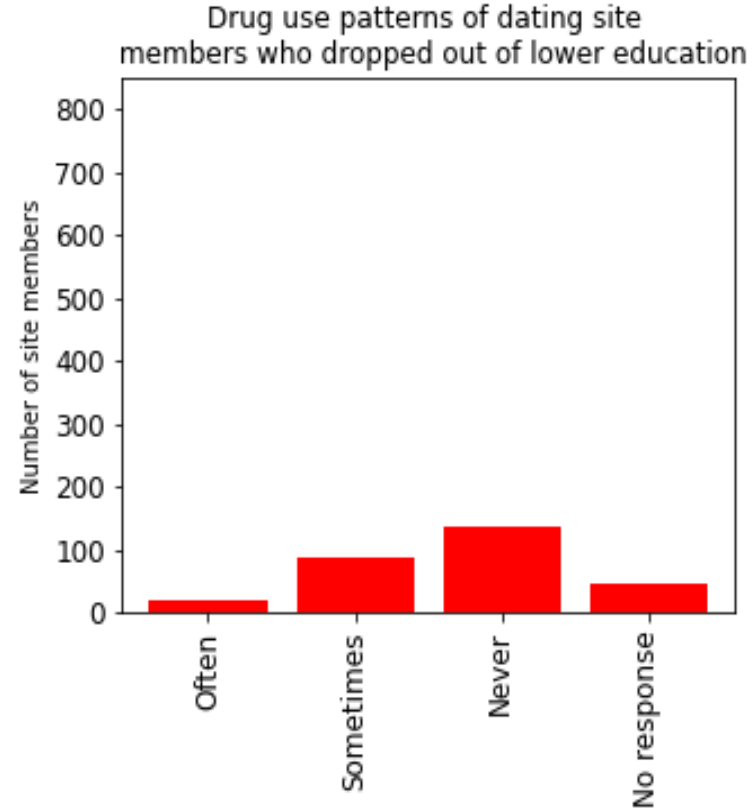
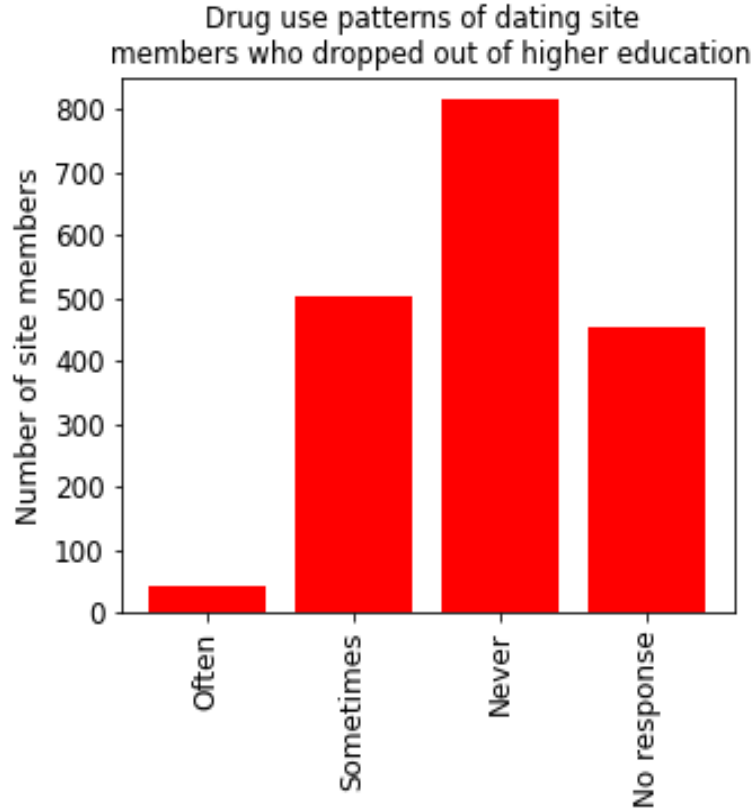
- The majority of members are highly educated or currently studying in higher education
- Higher education includes college / university, masters degree, phd, law school, medical school, space camp
- Lower education includes two-year college and high school
- Some members dropped out of education – is there a link between education dropouts and their drink and drug habits we saw on the previous slides?

# Is there a link between members who dropped out of education and their drinking habits?



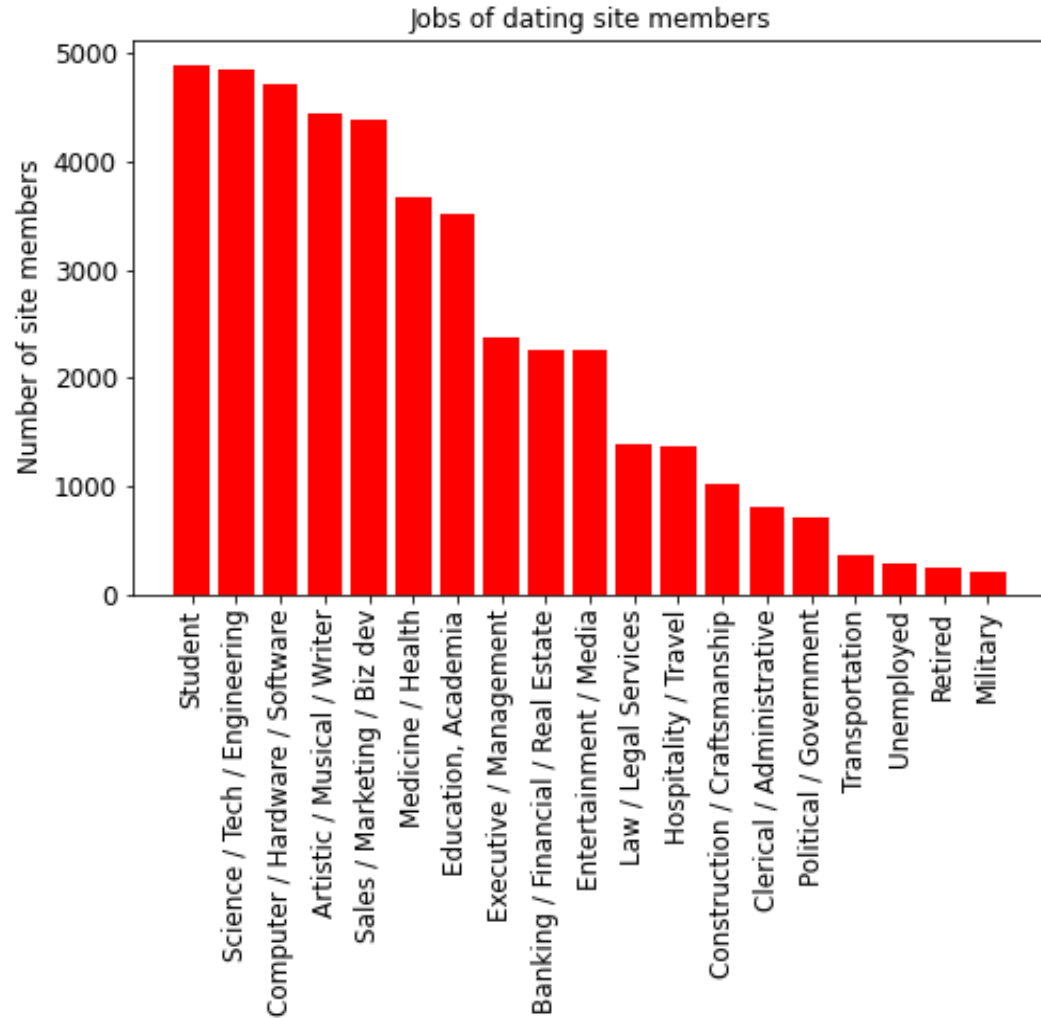
- The pattern between higher and lower education dropouts is similar to all members – the majority are social drinkers in each group
- But the numbers are a lot higher for the higher education group. This is likely to be because more people were in this group in the first place
- Again, a few people say they drink 'desperately' and this figure is higher for the higher education group and makes up **13%** of the wider dating site population

# Is there a link between members who dropped out of education and their drug use habits?



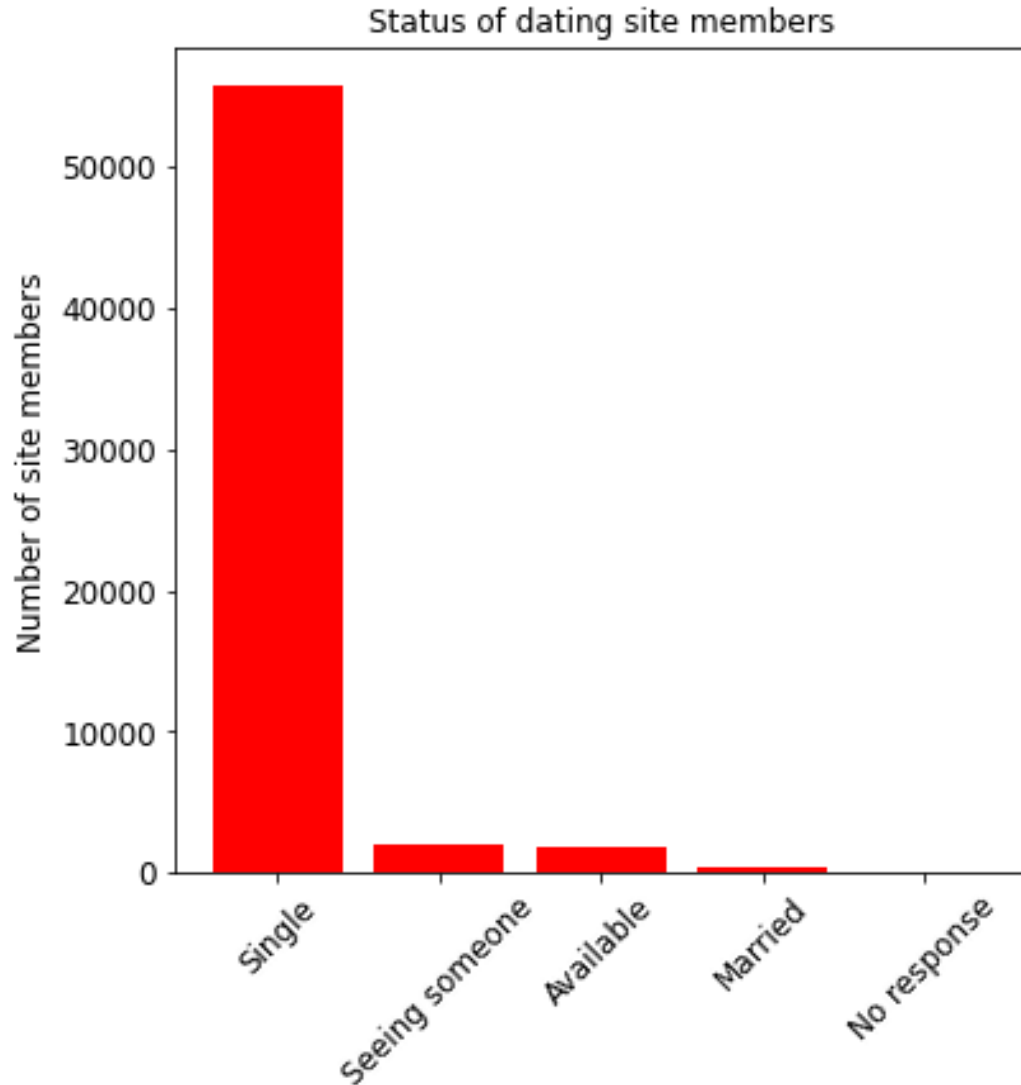
- The pattern between higher and lower education dropouts is similar to all members – the majority never take drugs in each group
- But the numbers are a lot higher for the higher education group
- A few people in each group say they take drugs often and this figure is higher for the higher education group and makes up **10%** of the wider dating site population

# What do the members of OkCupid do for a living?



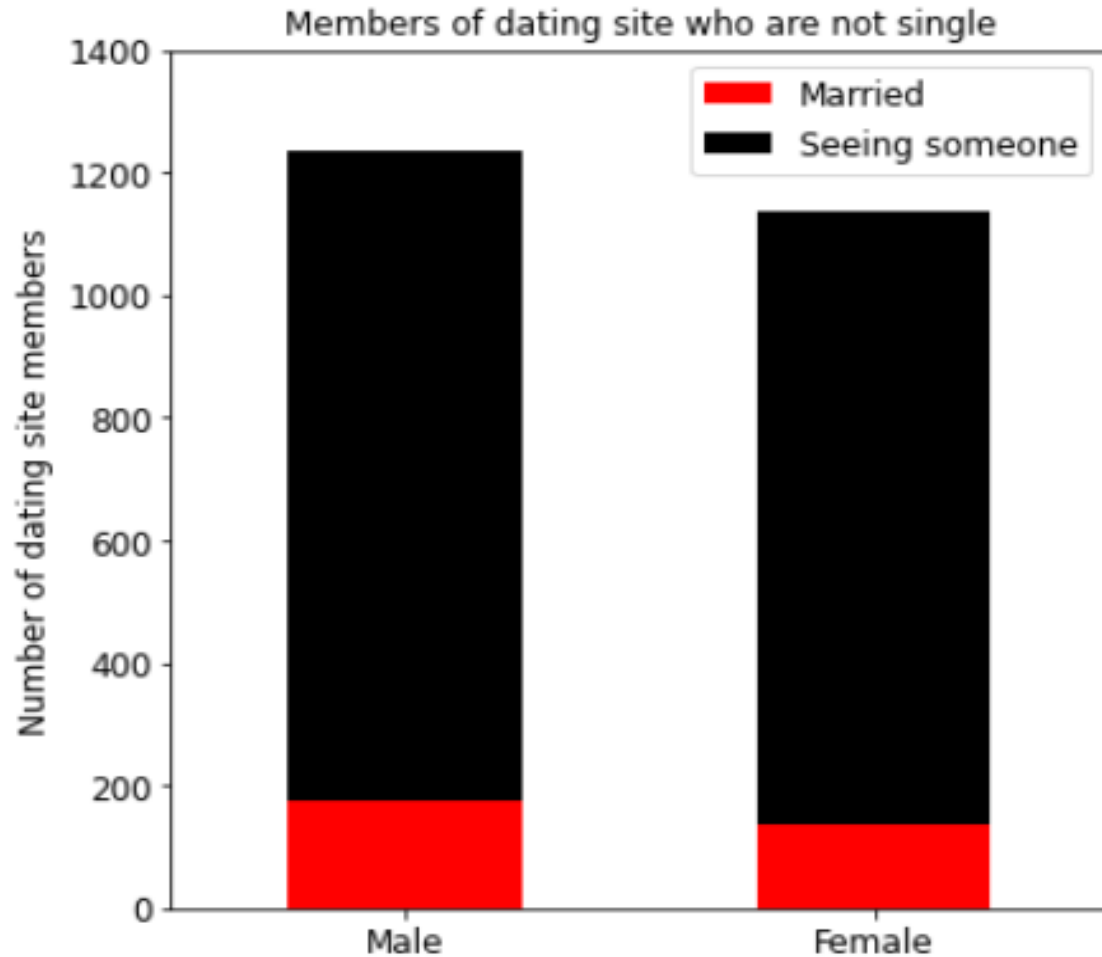
- The majority of members are students which ties in with the younger age range of most members
- There are also a high proportion of members in science, tech and computer fields – all of these jobs would require a high level of education and the education data back up the fact that **most members are highly educated**
- Although not shown here (because most members left the field blank), a high proportion of members earn income in the \$80,000-\$100,000 bracket. It is likely that many of these members work in the science, tech or computer fields

# What is the status of OkCupid members?



- The majority of members are single, as one would expect on a dating site
- However, **over 2,000** people admit they are currently seeing someone else and **just over 300** members admit they are married
- It is likely that these members are looking for something much more casual than those singletons seeking a genuine relationship
- What is the makeup of these attached people – are they mainly men or women?

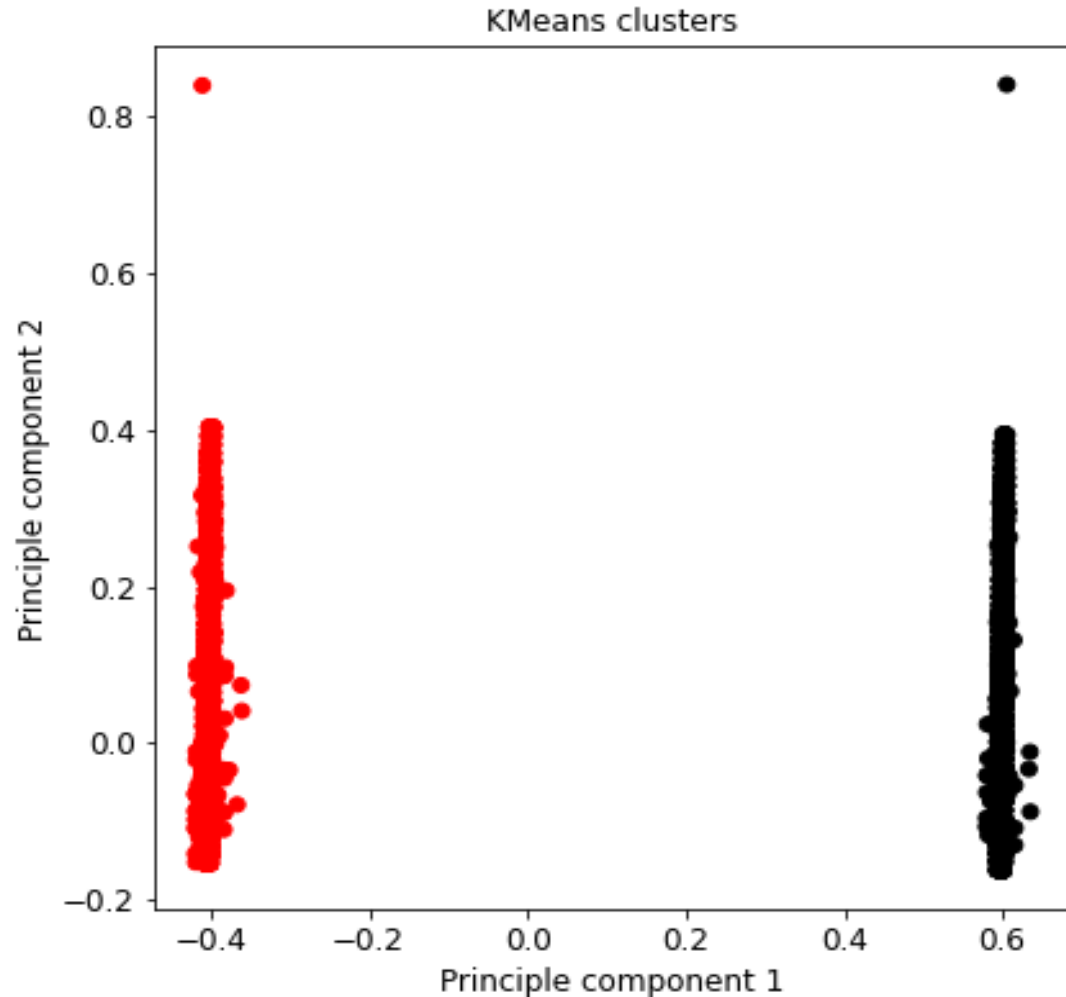
# Who are the non-single members on OkCupid?



- There are more men on the OkCupid dating site who are not single but there isn't a huge difference in the numbers
- There are **175 married men** on the site compared to **135 married women**
- There are **1,061 male** members who are seeing other people compared to **1,003 women**
- This doesn't necessarily mean these members are cheating on their partners (although some are likely to be). It could be that they are in open relationships
- One can also not conclude from this that men are more likely to stray than women because the sample size is too small to verify this



# Clustering the OkCupid data using KMeans



- Three features were used in the KMeans model – age, height and sex
- Adding more features resulted in the curse of dimensionality with the elbow score being very high at nine features
- This may be because KMeans performs better on numerical variables and many of the features of this dataset are categorical – it may be better to try a KModes model on the categorical data
- The KMeans model used here performed best with two clusters (as determined by the elbow curve) and accurately classified each profile as male or female regardless of their age or height

# Conclusion

- It can be seen from the data that OkCupid is mainly targeted at younger members although a few members are also in older age brackets
- There are over 10,000 more male members than female which could result in a dating imbalance. If there is a risk of this being the case, OkCupid could engage in marketing activities to target more female members to its site
- Many of the members are highly educated and this ties in with the fact that a high proportion work in very highly skilled jobs
- Some members have dropped out of education but there does not appear to be a link between these members and their drink and drug habits – their habits follow the same pattern as those of the wider dating population
- Not everyone on the site is single – some are married and others are seeing someone. The figures are similar for men and women although very slightly higher for men
- There were many fields in the dataset which were not completed by all members. It may serve OkCupid well to make many of these fields mandatory – not only will this help its members make a more informed decision about a potential date, it will also help the company understand its members better
- This KMeans model accurately clustered males and females but since much of the dataset consists of categorical values, the modelling can be improved by testing a KModes model on the data

# Thank you

**Contact details:**

**Amy Birdee**

**[amybirdee@gmail.com](mailto:amybirdee@gmail.com)**