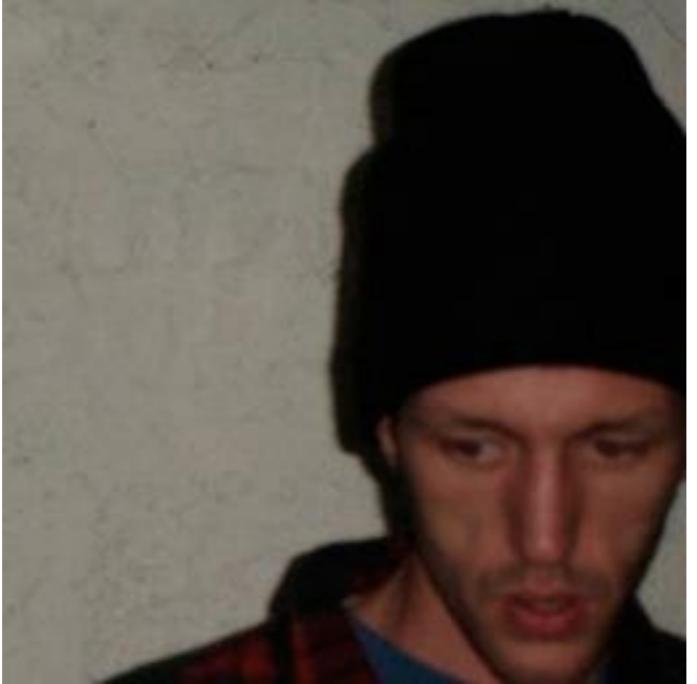


# Supervised and Unsupervised Learning Techniques to Increase the Frequency of Meet-Cutes





This is a story of boy meets girl. But you should know upfront that this is not a love story.

It's a math story

# Chris's Problem

## Goal

Have more compatible women reach out to him on OK Cupid

## Target of Optimization

NumberOfMessages = f(MatchPercentage(Survey Questions Answered, Answers to Survey Questions, Importance of Survey Questions))

# The Strategy

To figure out how to design his profile to attract messages from compatible women, Chris needed to:

1. Gather a dataset of profiles of Los Angeles women
2. Identify the most compatible women (MCW) for him in the dataset
3. Learn which survey questions the MCW answered—and how importantly they rated those questions
4. Optimize his OK Cupid profile to attract the MCW based on his predictive model—without lying about his answers

# Step 1: Gather Data

**Strategy:** Create OK Cupid profiles for the sole purpose of scraping the profiles of women in LA

**Hiccup:** OK Cupid realized that his profiles were just automated bots, so Chris created new bots that mimicked the brow

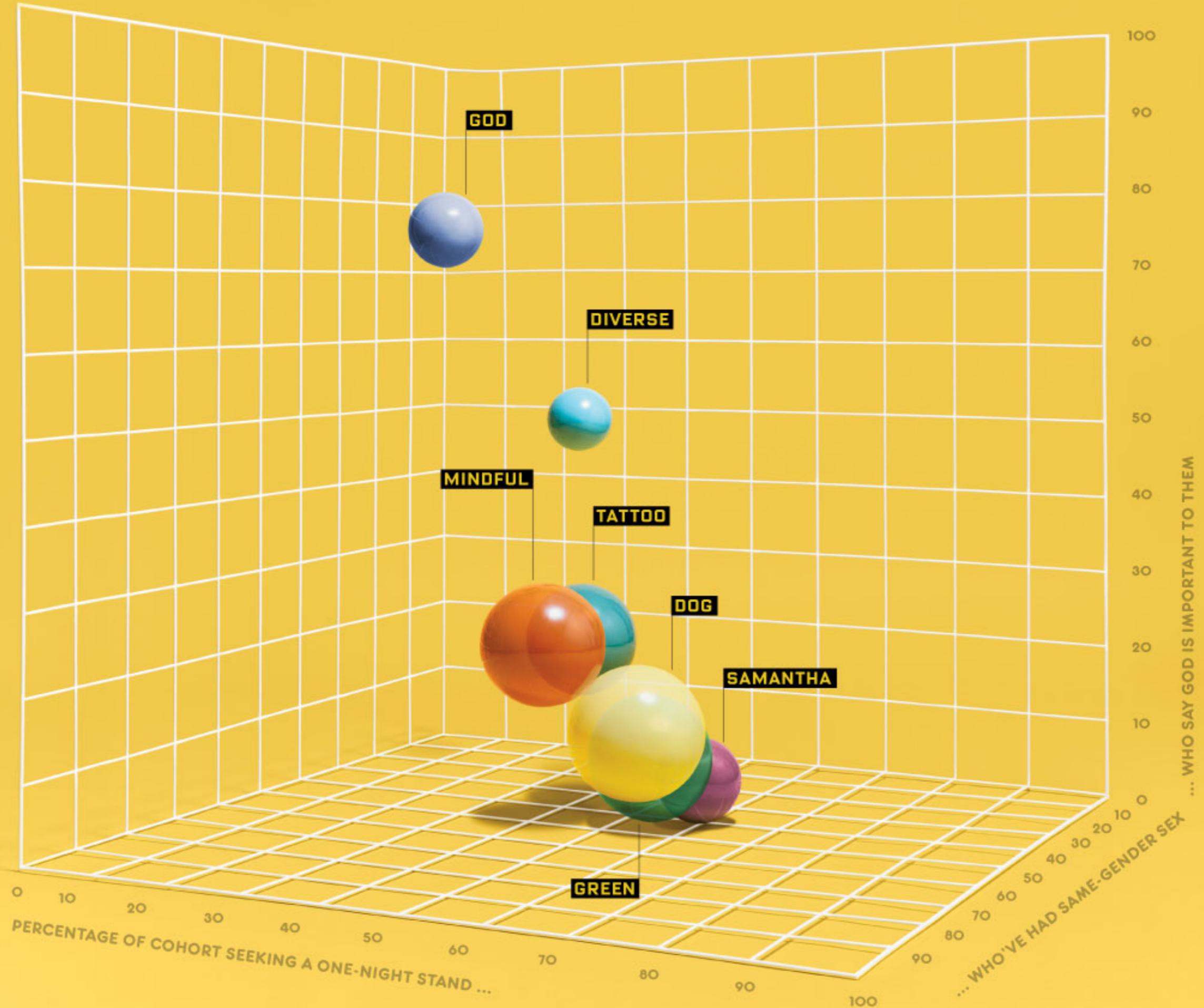
**Results:** Gathered data about more than 6 million survey questions from more than 20,000 women

# Step 2: Find the MCW

**Difficulty:** Without going on dates with the 20,000 women in the dataset—or at least reading through 20,000 profiles—it was hard to predetermine how compatible a particular woman would be

**Solution:** Use an unsupervised learning algorithm\* to cluster the 20,000 women, then focus on the cluster that appeared to have the most MCW

\*K-Modes, available in the R package klaR

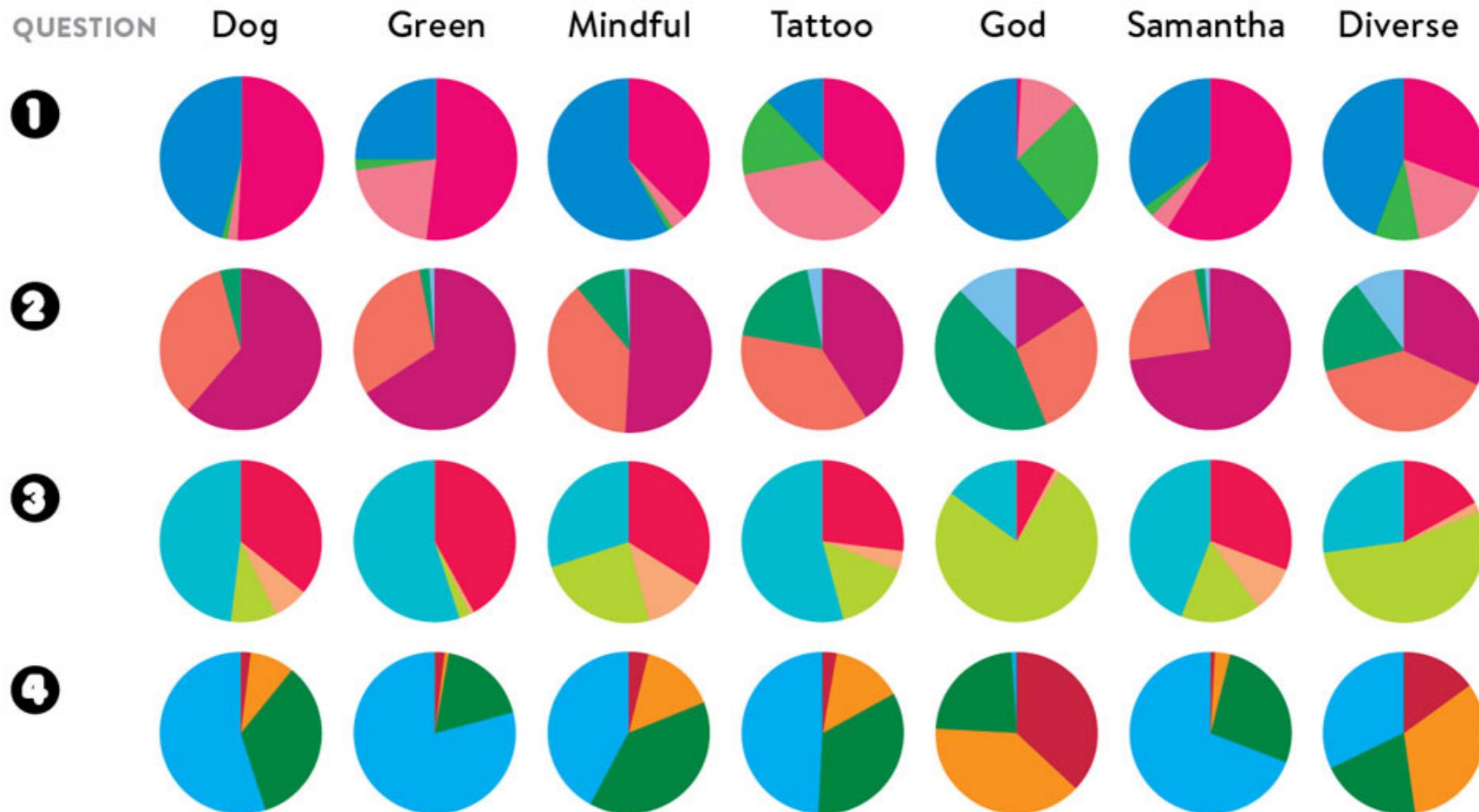


**(1) About how long do you want your next relationship to last?** ■ One night ■ A few months to a year ■ Several years ■ The rest of my life

**(2) Say you've started seeing someone you really like. As far as you're concerned, how long will it take before you have sex?** ■ 1-2 dates ■ 3-5 dates ■ 6 or more dates ■ Only after the wedding

**(3) Have you ever had a sexual encounter with someone of the same sex?** ■ Yes, and I enjoyed myself ■ Yes, and I did not enjoy myself ■ No, and I would never ■ No, but I'd like to

**(4) How important is religion/God in your life?** ■ Extremely important ■ Somewhat important ■ Not very important ■ Not important at all



# 3) Predict Match %

**Strategy:** For the women in preferred clusters, use a supervised learning algorithm to predict the match percentage he'd obtain as a function of the questions he answered and how importantly he rated those questions.

**Algorithm:** AdaBoost with decision trees, an ensemble learning technique

# 4) Optimize the Profile

## Results

From Wired: “When the last question was answered and ranked, he ran a search on OkCupid for women in Los Angeles sorted by match percentage. At the top: a page of women matched at 99 percent. He scrolled down ... and down ... and down. Ten thousand women scrolled by, from all over Los Angeles, and he was still in the 90s.”

# Epilogue



**BuzzFeed FWD** LOL

NEWS ENTERTAINMENT LIFE VIDEOS MORE Get Our App

PARTNER

## Sorry, Guy: Math Can't Get You A Girlfriend

Why OkCupid (and, more generally, love) can't really be "hacked."

posted on Jan. 24, 2014, at 10:15 a.m.

 **Katie Heaney**  
BuzzFeed Staff [Follow](#)

[f](#) [Twitter](#) [Email](#) [Pinterest](#) [g+](#)

