# Data Mining on OkCupid

Investigation by Laura Howard

# How does it work?

---

Start answering questions with…

- Your answer
- How you'd like someone else to answer
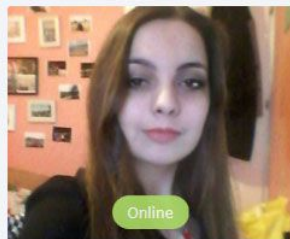- How important the question is to you

Get matches!

89% Match

94% Match

72% Match

90% Match

okcupid

**smeralda06**
23 · Thessaloníki, Greece

| 89% | 9% |
|---|---|
| Match | Enemy |

**kabuki31**
21 · Sofia, Bulgaria

| 42% | 68% |
|---|---|
| Match | Enemy |

**lilika_**
31 · Athens, Greece

| 88% | 11% |
|---|---|
| Match | Enemy |
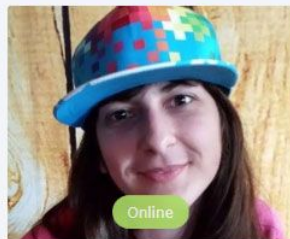
**Dobito12**
26 · Sofia, Bulgaria

| 55% | 61% |
|---|---|
| Match | Enemy |
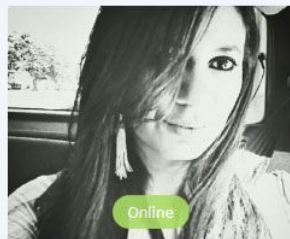
**Fabis_johns**
28 · Athens, Greece

| 43% | 77% |
|---|---|
| Match | Enemy |
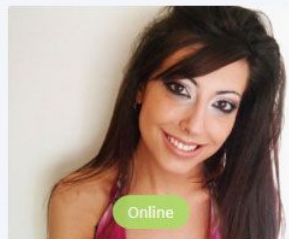
**wildfoxthecat**
26 · Kyustendil, Bulgaria

| 37% | 68% |
|---|---|
| Match | Enemy |

okcupid

35,348 online now

Messages   Matches   Connections   Treasures

You might like...

See more matches

Favorites

*No favorites are online*

Show offline favorites

Our Mobile Apps

iPhone   Android

Your Awards

Share your profile

f  **Facebook**

t  **Twitter**

[blurred name]

28 / M / Straight / Single
Barcelona, Spain

Online Now

About   Photos   Questions   Personality   Tests   Awards

## My Self-Summary

[blurred text]

## My Details

| | |
| --- | --- |
| Last Online | Online now! |
| Ethnicity | White |
| Height | 5' 10" (1.78m). |
| Body Type | A little extra |
| Smokes | No |
| Drinks | Rarely |
| Drugs | Never |

# How is the match percentage calculated?

Person A          Person B
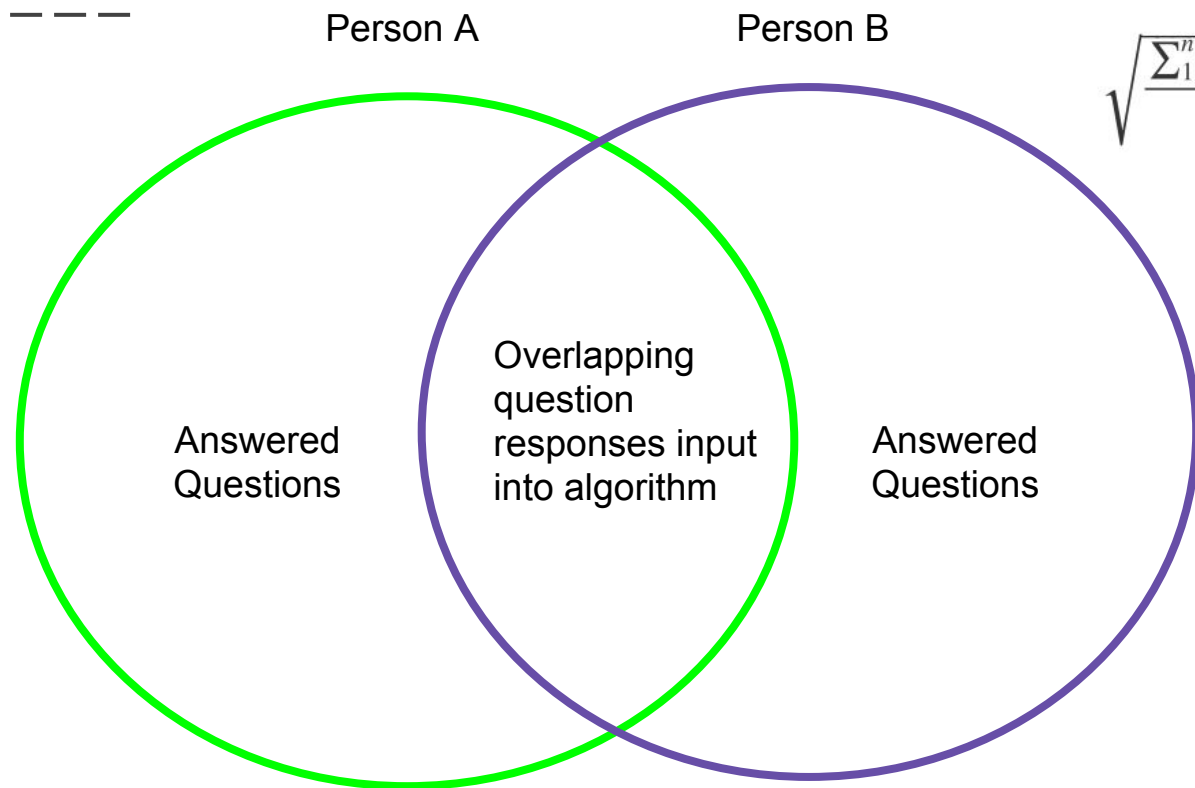
$$\sqrt{\frac{\sum_1^n Ans_A * Imp_B}{\sum_1^n Imp_B} * \frac{\sum_1^n Ans_B * Imp_A}{\sum_1^n Imp_A} - Error}$$

Answered Questions

Overlapping question responses input into algorithm

Answered Questions

| Level of Importance | Point Value |
|---|---|
| Irrelevant | 0 |
| A Little important | 1 |
| Somewhat important | 10 |
| Very important | 250 |

# How many overlapping questions are needed to create strong matches?

— — —

| Number of Overlapping Questions | Margin of Error | Highest Possible Match |
| --- | --- | --- |
| 2 | 0.50 | 50% |
| 5 | 0.20 | 80% |
| 10 | 0.10 | 90% |
| 50 | 0.02 | 98% |
| 100 | 0.01 | 99% |
| 1000 | 0.001 | 99.9% |

# Let's take an example

— — —

**Question 1:**

How messy are you?

- Very messy
- Average
- Very organized



$$\text{sqrt}(^{10}/_{11} * {}^{250}/_{251}) - \text{Error}(2) = \text{Match \%}$$

$$\text{sqrt}(91\% \times 99.5\%) - 50\% = \sim 45\%$$

**Question 2:**

Do you like beer?

- Yes
- No

# "Hacking" the System



Meet Chris McKinlay,
PhD in Mathematics from UCLA

Scraped 20,000
female profiles

Trained K-Modes
clustering model

Identified 7
distinct clusters

Indie musician/artist types



Young professionals in
creative jobs

# Finding High Matches (aka Swiping Right)

———
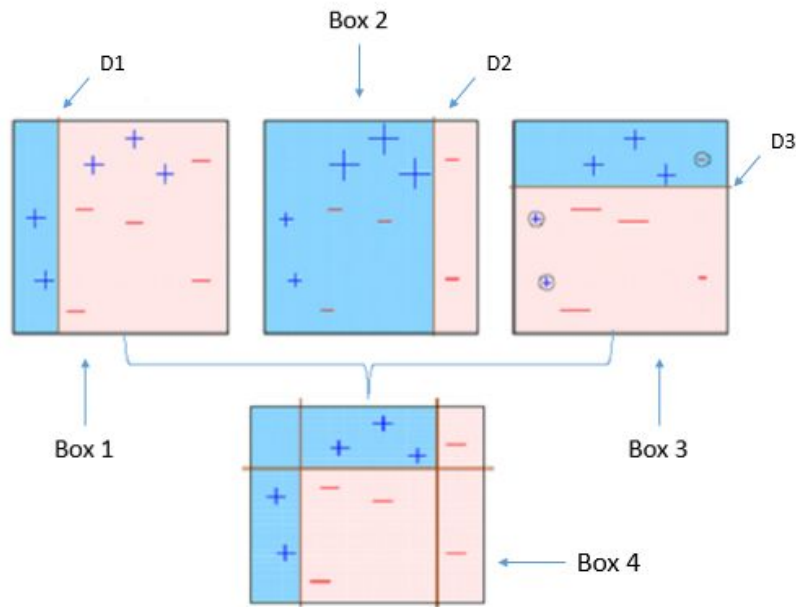
Honestly answered the 500 most popular questions

Created two profiles, one for each cluster

Used adaptive boosting to derive the best importance weightings on every question for each of his two profiles

10,000 women matched at > 90% !

# What next?

———

Created a program to visit
the pages of his top matches

# Thousands of messages and 55 dates later...



## Mrs. 91% right

# Best Predictors of Relationships According to OkC

___

- Do you like horror movies?

- Have you ever traveled around another country alone?

- Wouldn't it be fun to chuck it all and go live on a sailboat?

"After analyzing the profiles of nearly 35,000 successful couples who met on OkCupid, researchers found that a third of lucky lovers agreed on all three questions. That's 3.7 times the rate of simple coincidence."

# My Favorite Questions...

— — —

If a clone was made of you, would you sleep with it?
- Yes
- No

While in the middle of the best lovemaking of your life, if your lover asked you to squeal like a dolphin, would you?
- I don't think I could make that noise if I wanted to
- No way
- Absolutely
- Maybe

Which is bigger?
- The earth
- The sun

Your a b****!
What bothers you more about the above sentence?
- The profanity
- The grammar
- They both bother me equally
- Neither bothers me

Do you believe in dinosaurs?
- Yes
- No

How do you eat your cupcake?
- Always start with frosting
- Always starts with the cake
- I don't have a consistent way of eating a cupcake.

# Questions?

# Sources

—  —  —

https://www.wired.com/2014/01/how-to-hack-okcupid/all/

https://okcupid.desk.com/customer/portal/articles/2458380-match-percentages---a-detailed-explanation

http://abcnews.go.com/Technology/dating-site-okcupid-shares-revealing-date-questions/story?id=12874977

https://mic.com/articles/85297/these-3-simple-questions-can-predict-if-an-okcupid-date-will-succeed#.A9KkjyRol