

A LOOK UNDER THE HOOD OF

okcupid

By Bob, Kyle, Ron & Dan

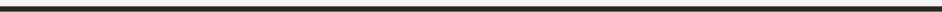
The Data:

- User profile data 68,000+ active OKCupid users
- 2600+ questions and responses
- November 2014 – March 2015



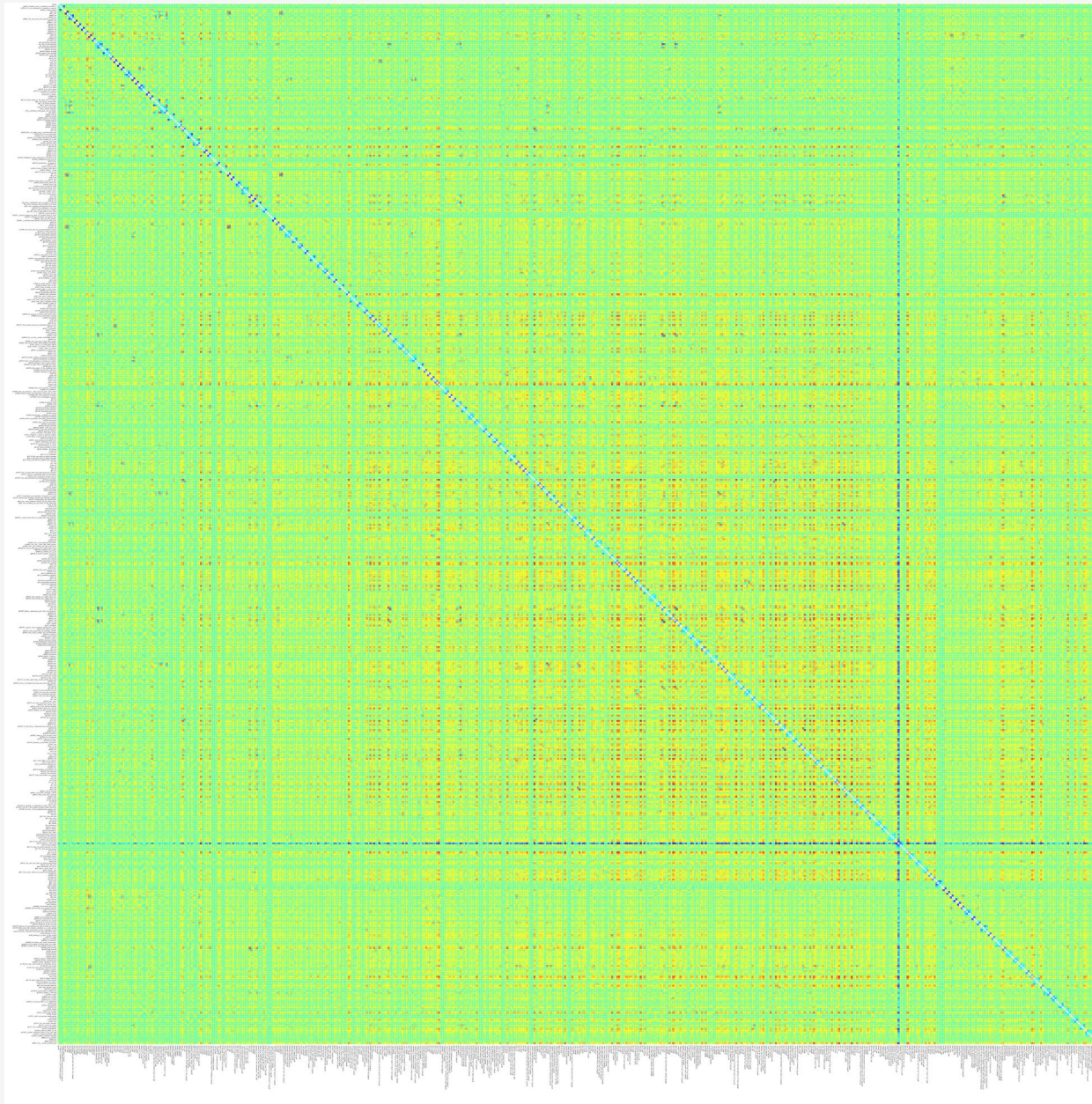
The goal:

- Classify users based on their responses.
- Predict answers to future questions.
- Deploy tools for visualization and prediction



Correlation matrix

(only the top 200
most commonly
answered questions)



This Face When:
We made the
correlation Matrix



This Face When:
We tried to look
at the correlation
Matrix



Rocket Ronnie
Sadness Index:
Off the charts

SOLUTION: BUILD LOTS OF MODELS

Pick the most
predictive questions out
of the bunch

Results:

With a 20 question input,
our model is able to predict
a users:

- Binary gender (only 0.3 % identify as non-binary gender)
- Feelings toward monogamy
- Feelings about rec. drug use
- Proclivity towards research prior to first date.
- Are you “Intense” or “Carefree”
- Feelings on guns
- Is Art important to you?
- Did you get picked on a lot in school?
- Do you spend more on clothes or food?

Results:

Mean AUC score: **0.72**

- We selected 9 models to run in parallel via our Flask web application.
- Created Interactive visualizations for users to explore.

Lets take a look!

