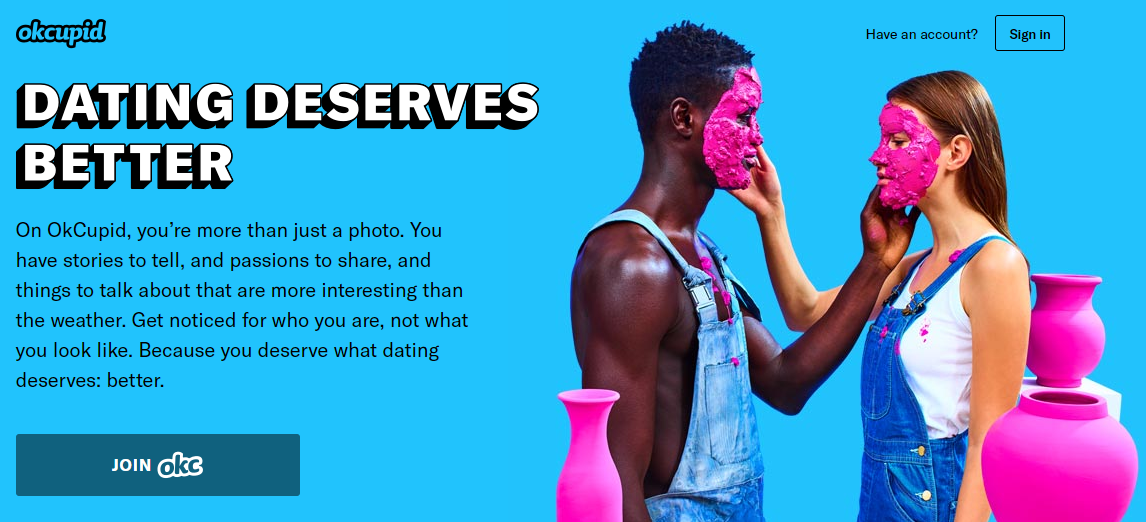
## Scenario



You are part of a business intelligence team at okcupid.com. The team has been asked to make an in-depth exploratory analysis of site users. The goal of the marketing team is to create micro segments and personas for future campaigns. Keep in mind, interesting data correlations may not be beneficial in a marketing context. For example, identifying 5 users with very specific attributes may be interesting but hardly a segment worth attracting.

**You are asked to examine the data, clean it, use supplemental data to enrich the data then identify 4 or more interesting insights from the user data. All relevant cleaning, enriching and EDA steps along with the 4 insightful data nuances should be organized into a presentation. Your team will present to the head of marketing who is looking for an “ah –ha” persona or previously unknown data relationship. As the head of marketing, relevant information is consumed visually instead of in table form. Thus, your presentation should include visualizations when appropriate. Your submission will include code and PowerPoint slides.**

## Data

Source: <https://www.researchgate.net/project/The-OKCupid-dataset-A-very-large-public-dataset-of-dating-site-users>

This data set was scraped from user profiles. At the time, OKCupid did not authorize the data to be collected. After the data was released as part of academic literature, the data was authorized to be used by OKCupid.com .

***As a result, there is some moral ambiguity related to the use of the dataset.***

The data set provided has been authorized, cleaned and anonymized. The profiles are located in profiles.csv. The original data, publication, code, and codebook can be found at <https://github.com/rudeboybert/JSE_OkCupid>”

## Example *Abridged* Data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **age** | **body\_type** | **diet** | **drinks** | **education** | **height** | **income** | **…** | **status** |
| 22 | a little extra | strictly anything | socially | working on college/university | 75 | *NA* | … | single |
| 35 | average | mostly other | often | working on space camp | 70 | 80000 | … | single |
| 38 | thin | anything | socially | graduated from masters program | 68 | *NA* | *…* | available |
| 23 | thin | vegetarian | socially | working on college/university | 71 | 20000 | … | single |
| 29 | athletic | *NA* | socially | graduated from college/university | 66 | *NA* | … | single |

## Course Scripting Supplemental

You will receive an initial script with code examples to get you started since this is the first case of the course.

## The Submission

The submission will include business analyst slides covering the problem, data and insights. Without a presentation, the “organization” section of the rubric will be 0.

The submission will include either a recorded screen narration of the business presentation, a text file with a URL to a recording (like youtube video) or the audio is embedded into the slide deck. Without this the “delivery” section of the rubric will be 0.

An R script covering all data munging and visualization construction used to create the presentation and come to the case recommendations.

## Criteria for Success

The presentation will be evaluated on a 50 pt scale with the following criteria.

* **Organization** – Was the presentation well organized?
* **Delivery** – Was the content delivered clearly and persuasively with the audience in mind?
* **Documentation** – Was the data mined to support the conclusion?
* **Data Mining Proces**s – Is the code submission organized and commented appropriately.

## Another resource may be a public R-Studio examination of the data

*Keep in mind this may not be helpful but code can be examined for additional ideas.* ***Submitting these visuals and code alone will not result in a good learning outcome or rubric socre.***

[*https://rstudio-pubs-static.s3.amazonaws.com/209370\_b62220c849b946088b463fdbec935848.html*](https://rstudio-pubs-static.s3.amazonaws.com/209370_b62220c849b946088b463fdbec935848.html)