

CSE 91, Fall 2017

Project 1, Part 2

Due: Thursday, October 19, 11:59pm

Part 2. Data Cleaning

Data cleaning involves ensuring that all the data you will use in your analysis is of the proper format, and that all the data is meaningful. The main purpose of part two is to get experience with data cleaning.

For this part of the project, you will clean up your data table in the following ways:

- remove all people whose income is not a numerical value
- remove all people whose “steps” data is missing
- remove all people whose income is unusually high (outliers)
- create a consistent way of recording smokers and nonsmokers
- reorder columns and sort rows by ID number

Submission Details

For this project, you may work alone or with one other person in either section of CSE 91, using the pair programming approach.

For help with your projects, instructors are available during the scheduled discussion section times on Thursdays.

To start working, log in to Jupyterhub and then click the assignment link available on the course website and Piazza.

Work through the notebook, and if you need to stop, submit the file, as only the most recent submission will be graded.

To submit your file, run the cell

```
_ = ok.submit()
```

This project is due at 11:59pm on Thursday, October 19.