General Instructions

The purpose of this group project is to analyze a data set, using Python to process the data and to generate graphs, figures, and visualizations as necessary, with the ultimate goal of producing a single written report which presents your results and explains their significance.

You may work in groups of two or three for this project. Each week, you should submit the code that you wrote (there may be multiple .py files generated), a written report of your results including the graphs that your code generated, and a small statement of each group member’s contribution to that week’s results. The deadline for turning in all files is 11:59PM on the due date.

The report should include the following:

1. Overview: A description of what this section of the analysis is about.
2. Methodology: How you performed the analysis.
3. Results: The results that your analysis produced.
4. Discussion: Interpreting the results.

The above is a general outline of what you should present, but you are not required to “number” or outline your report using these terms. Just be sure that it is clear that you are providing the required parts.

It should also be noted that there may be multiple ways to interpret the given prompts. The purpose of this assignment is to explore the possibilities of what you can do. Do not be concerned with trying to figure out a single “right” answer. The prompt is a suggested starting point for investigation. You are not being graded on whether or not you come up with a specific answer, but rather how correct your methodology is and on the quality of execution.

For all written work (i.e., the reports, not the .py files), you may use any word processor that you like, but please make sure to export the report as a PDF, so that there will not be any software compatibility issues in grading.

This project will make use of the Speed Dating data set, which is available on the course website.
Due: February 13, 2019

Who wants what?
Participants were asked to identify what qualities they were most interested in for their dating partners. Participants were also asked what they thought the other sex was looking for. Examine this data as it relates to the sexes, visualize it in a meaningful way, and interpret its meaning.

Modest or Conceited?
Participants were asked to rate themselves. They were also rated by the people that they talked to during the speed dating event. In general, how differently do the sexes behave in this exercise? That is, for the different attributes, do men (or women) over- or under-estimate themselves?

Due: February 25, 2019

For clarity in the subsequent sections, we will refer to Participants and Potentials. A Participant is anyone participating in the experiment. A Potential is one of the people that a Participant talks to during the course of the speed dating night. Of course, all Participants are also Potentials for the opposite sex!

Down-to-Earth or Head-in-the-Clouds?
How do male/female Participants rate their Potentials in contrast to how that Potential rates themself? Does it matter if the Participant decides “yes” or “no” on that Potential?

Dating Up (and Down)
How likely is a Participant to want to date a Potential if the Participant rates that Potential higher (or lower) than the Participant rates themself?

Due: March 6, 2019

Who Benefits From Speed Dating?
Is it worth it to attend a speed dating event? Or perhaps it’s only worth it for one gender... or neither... or both! How can we know for sure? We know that Participants made a decision as to whether or not they were interested in a specific Potential. We also know that the organizers would exchange information between two people only if both of those people indicated that they were interested in the other person. But how often did this happen?
Examine this phenomenon from multiple (non-trivial) perspectives. For example, consider that, for any given interaction, either both the man and the woman are interested in each other, or only the man is interested but not the woman, or only the woman is interested but not the man, or neither the man nor the woman is interested. How often does this happen? Is this better to express as a percent or as a number? A different example would be this: Does only “one girl get all the guys”, or is it spread out evenly (and how does it work out for the guys)? Some of these facts may be presented as a table as well as a chart/visualization. It is expected for you to analyze this data from two perspectives of your choosing.

Note: In asking for “two perspectives”, this does not simply mean that you are showing some analysis as it applies to women, and then that same analysis as it applies to men. Rather, this is really only one analysis that uses gender as a factor. If it helps you to understand what is intended, I will clarify that the above suggestions are really just two possible “perspectives” that you can use to examine the data. You may, of course, come up with your own type of analysis instead. The requirement for a “non-trivial” perspective simply clarifies that the analysis that you choose should not be a simple statistic (e.g., the percent of participants who are men).

Put It All Together

Combine your work from Part 1 and 2 with your above analysis and submit it as a single document. It should be submitted as a PDF. You should also submit your code for this week’s work, but you do not need to submit the code from the previous weeks.

Add an Introductory section that introduces the data set itself and gives enough information so that the reader can understand the analysis provided. You can see information about the data set here:
https://www.kaggle.com/annavictoria/speed-dating-experiment/home

Add a Future Work section, in which you discuss additional types of (non-trivial) analysis that could be performed with this data set. This section probably does not need to be more than a paragraph in length.

Add a Conclusion section that provides a perspective on the analysis (a sort of meta-analysis) that you performed across all three Parts of the project. How has this research confirmed or challenged your opinions? What, if any, significant results may be interpreted from the analysis? We are not looking for a specific answer, but rather for you to demonstrate your own critical thinking in regards to the analysis performed.