

Author: Meghan Ng

Title: Speed Dating Trends

Summary of research questions:

1. Is there a correlation between people who go out (not necessarily on dates) more often and getting more matches? My hypothesis: yes, participants who go out (not necessarily on dates) more often get more matches than those who don't. Participants ranked themselves on a scale of 1-10 whether they go out (not necessarily on dates) often or not. I will compare their rankings to the number of matches they had. This is to see whether spending time with other people will enhance their chances of finding matches during speed dating events.
Result: There is no correlation.
2. Is there a correlation between people who go out (on dates) more often and getting more matches? My hypothesis: yes, participants who go out (on dates) more often get more matches than those who don't. Participants ranked themselves on a scale of 1-10 whether they go out (on dates) often or not. I will then compare their rankings to the number of matches they had. This is to see whether dating more often will enhance their chances of finding matches during speed dating events.
Result: There is no correlation.
3. Is there a correlation between age of a participant and the number of matches they have at the end of the event? My hypothesis: no, the participant's age does not correlate with the number of matches they have at the end of the event. Participants wrote down their ages. I will compare their ages to the number of matches they had at the end of the event. This is to see whether age plays a role in their chances of finding matches during speed dating events.
Result: There is no correlation.
4. Is there a correlation between a participant's median household income (based on where they are from) and the number of matches they have at the end of the event? My hypothesis: no, the participant's median household income (based on where they are from) does not correlate with the number of matches they have at the end of the event. Participants wrote down the zipcodes they are from, which also determined their median household income in a column in the csv file. I will compare their median household incomes to the number of matches they had. This is to see whether their median household income and possibly their upbringings will play a role in their chances of finding matches during speed dating events.
Result: There is no correlation.

Motivation and background:

Too often, the same questions and thoughts run through the minds of women and men after a breakup or being ghosted by a first date: “Is there something wrong with me? Men/women are so frustrating!” This dataset can be used to debunk myths about the other gender and allow people to better understand their own habits, interests, what they look for in a partner, and what a potential partner could look for in them. Hopefully by learning about these findings, people who are looking for a relationship can do so with the confidence in knowing that they are many steps closer to finding “the one.”

Dataset:

URL to dataset, located at the bottom of the page in the “Files” section: <https://www.kaggle.com/annavictoria/speed-dating-experiment>

Username: meghanng8

Email: ngmeghan8@hotmail.com

Password: password

This dataset was compiled by Sheen Iyengar and Ray Fisman, professors of the Columbia Business School. The experimental speed dating events that produced the data were conducted from October 16, 2002 to April 7, 2004. Participants were asked to fill multiple questionnaires before, during, and after the events to describe themselves (demographics, personal values, habits), what they hoped to see in other participants, and their overall experience with the event. The answers to these questionnaires as well as information about their matches are included in the dataset.

Methodology (algorithm or analysis):

1. Is there a correlation between how frequently a participant goes out (not necessarily on dates) and the number of matches they have at the end of the event?

Participants ranked themselves (on a scale of 1 to 7, with more information related to this scale in the metadata) on how often they go out (not necessarily on dates).

Create a scatterplot of rank vs total number of matches. For each participant, their rank and the total number of matches they had at the end of the event will be plotted.

Calculate the pearson correlation coefficient and interpret it to determine if there is a correlation between the two variables

2. Is there a correlation between how frequently a participant goes out (on dates) and the number of matches they have at the end of the event?

Participants ranked themselves (on a scale of 1 to 7, with more information related to this scale in the metadata) on how often they go out (on dates).

Create a scatterplot of rank vs total number of matches. For each participant, their rank and the total number of matches they had at the end of the event will be plotted. Calculate the pearson correlation coefficient and interpret it to determine if there is a correlation between the two variables.

3. Is there a correlation between age of a participant and the number of matches they have at the end of the event?

Participants wrote down their ages.

Create a scatterplot of age vs total number of matches. For each participant, their age and the total number of matches they had at the end of the event will be plotted. Calculate the pearson correlation coefficient and interpret it to determine if there is a correlation between the two variables.

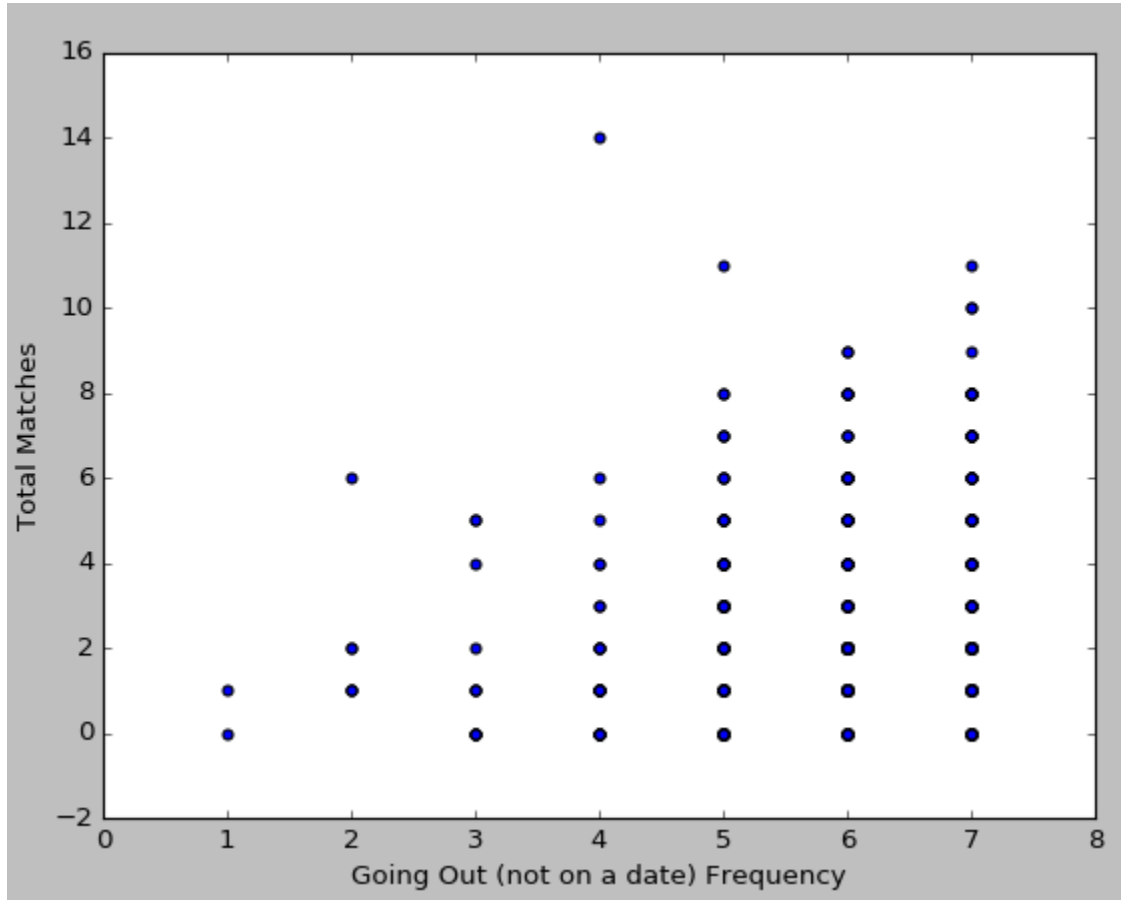
4. Is there a correlation between a participant's median household income (based on where they are from) and the number of matches they have at the end of the event?

Participants wrote down the zipcodes of the places they come from. These zipcodes were then converted to the median household incomes (\$) of those zipcodes.

Create a scatterplot of median household income vs total number of matches. For each participant, their median household income and the total number of matches they had at the end of the event will be plotted. Calculate the pearson correlation coefficient and interpret it to determine if there is a correlation between the two variables.

Results:

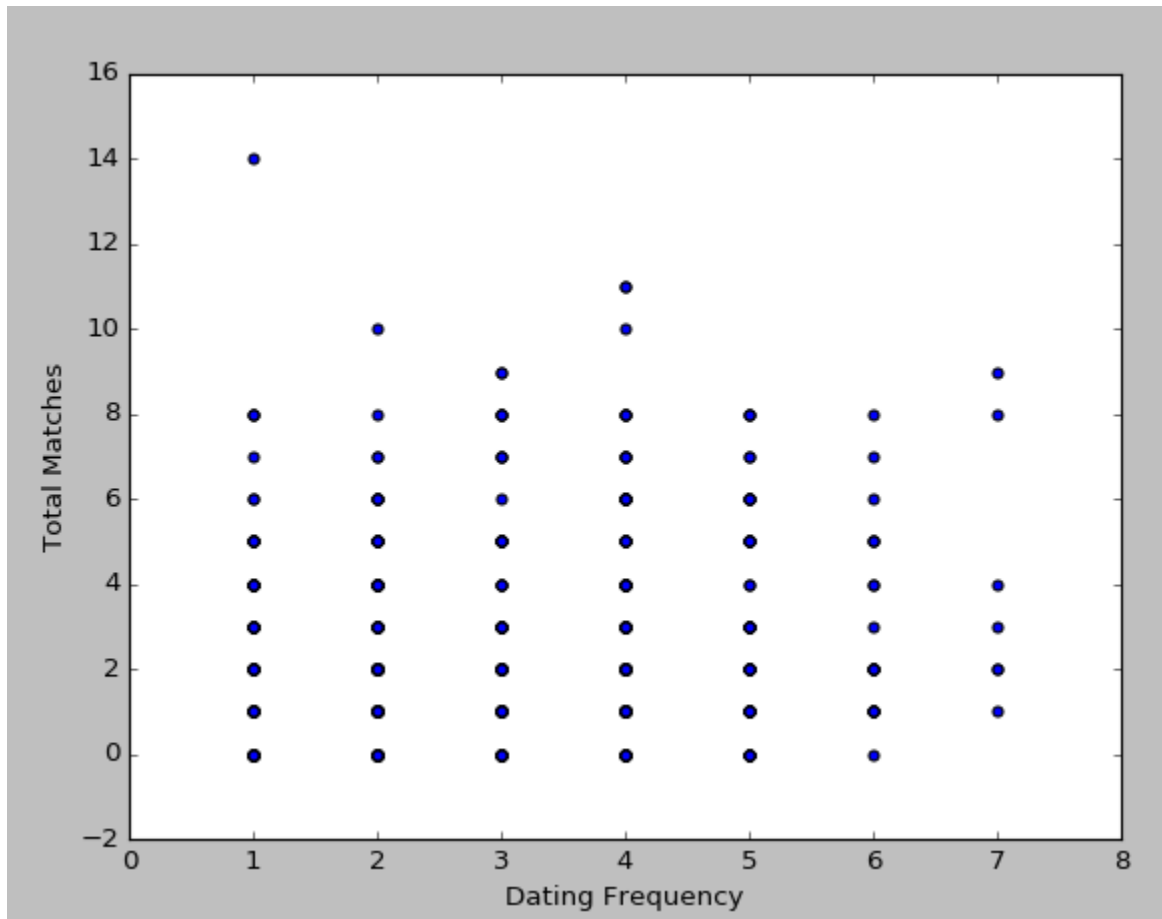
1. Is there a correlation between how frequently a participant goes out (not necessarily on dates) and the number of matches they have at the end of the event?



Resulting pearson correlation coefficient: 0.135125083467

Based on the sample tested, there is no correlation between how frequently a participant goes out (not necessarily on dates) and the number of matches they have at the end of the event. This means that how frequently a participant goes out (not necessarily on dates) does not determine how many matches they will get at a speed dating event.

2. Is there a correlation between how frequently a participant goes out (on dates) and the number of matches they have at the end of the event?

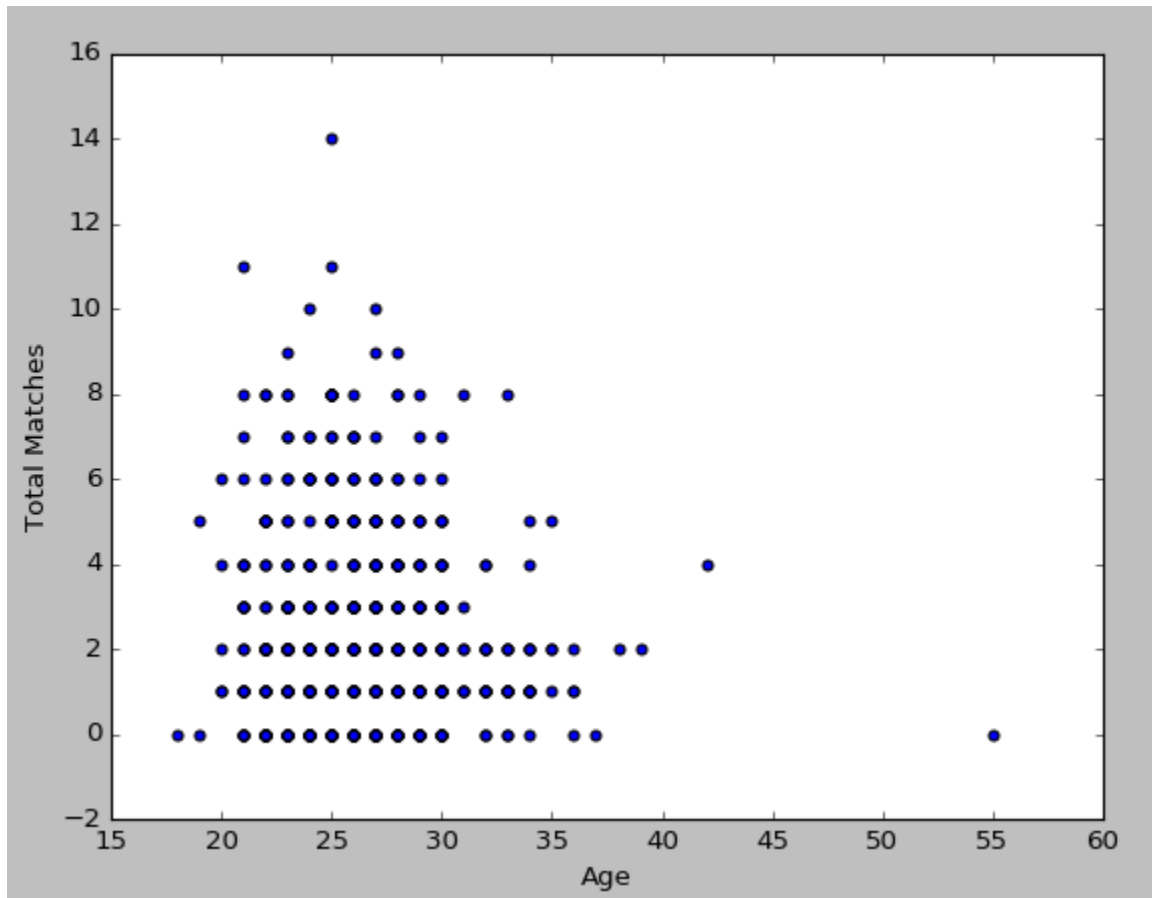


Resulting pearson correlation coefficient: 0.137244948127

Based on the sample tested, there is no correlation between how frequently a participant goes out (on dates) and the number of matches they have at the end of the event. This means that how frequently a participant goes out (on dates) does not determine how many matches they will get at a speed dating event.

The most surprising results were those of questions 1 and 2. I would have thought that people who go out and socialize more would be the ones to easily get matches during speed dating events. There may be other factors, such as pickiness, that could result in the number of matches they get.

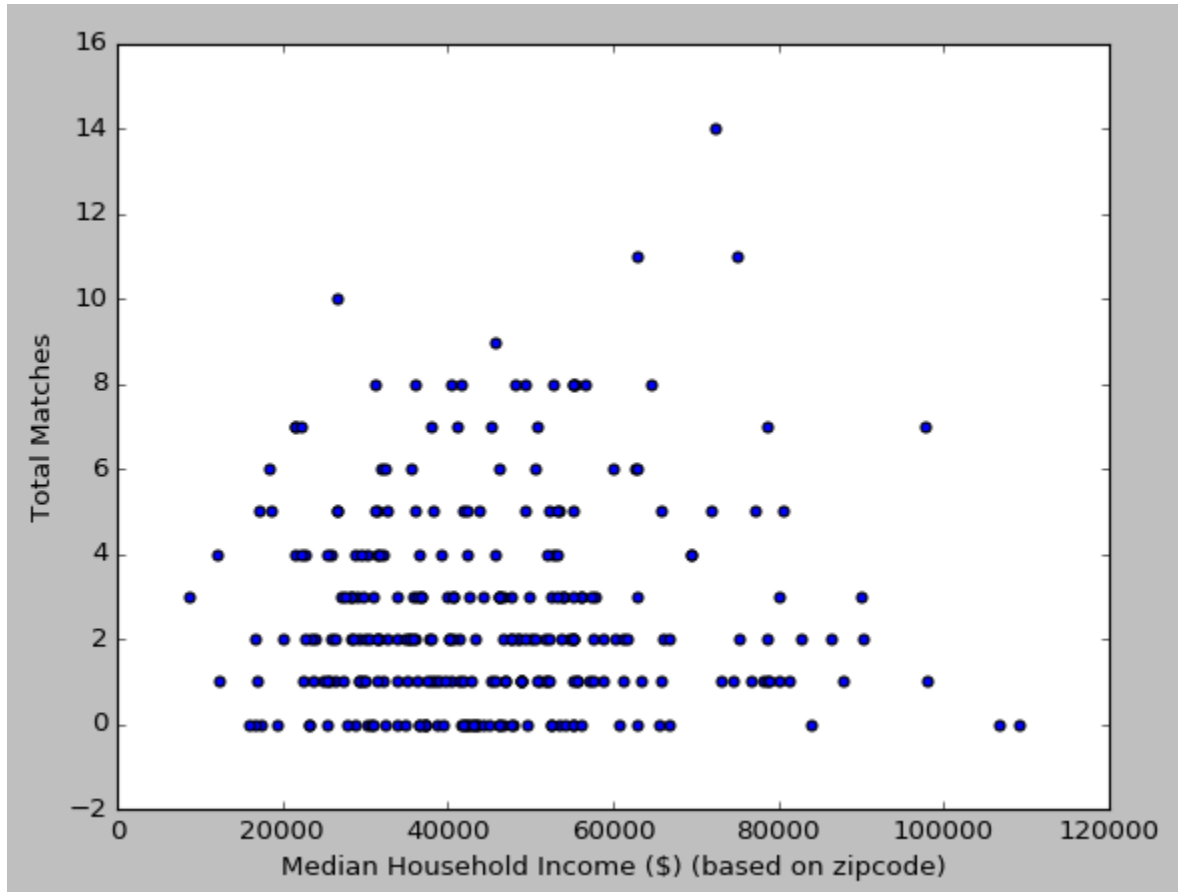
3. Is there a correlation between age of a participant and the number of matches they have at the end of the event?



Resulting pearson correlation coefficient: -0.0820486155657

Based on the sample tested, there is no correlation between age and the number of matches they have at the end of the event. This means that a participant's age does not determine how many matches they will get at a speed dating event.

4. Is there a correlation between a participant's median household income (based on where they are from) and the number of matches they have at the end of the event?



Resulting pearson correlation coefficient: 0.036282651745

Based on the sample tested, there is no correlation between median household income and the number of matches they have at the end of the event. This means that a participant's median household income does not determine how many matches they will get at a speed dating event.

Reproducing the Results:

1. Download the dataset and program into a folder
2. In the command prompt, navigate to this folder
3. Type in “python final.py” and press enter

Interpreting the correlation coefficient:

- If the coefficient is equal to -1 or 1, the xaxis and yaxis variables are said to be correlated.
- If the coefficient is near or equal to 0, the xaxis and yaxis variables are said to not be correlated.

Work Plan Evaluation:

It was very difficult to stick with the work plan I had. My first step was to pseudocode the program, then create the functions, but as I had very little practice with this beforehand, I decided to write the functions first and design the code as I went along. I will hopefully work on my program designing skills in the future.

Testing:

I used asserts to test my code. At times when it was doable, I also manually checked the values of the results of some functions with the values within the csv file. In the future and when I understand tests more, I would love to use different ways to do more testing to further ensure that my results are correct.

I will do a video of my presentation.

Collaboration: I did not collaborate with anyone else.