Introduction

We started with selecting a data set. Initial inspection revealed the SpeedDating data set was the most complex and had very detailed data on matching so it became our choice.

We did some EDA in tableau: [Link](https://public.tableau.com/app/profile/narendhra.chitturi/viz/CodePuddingEDASpeedDating/HobbyRatings) . We were able to confirm basic data: total participants, gender distribution, profession distribution, Goals of participants. In addition we found that some hobbies were more popular than others : Dining, Movies, Music & Reading were the highest rated. Next was determining some possible routes of exploration:

1. Comparison of how participants rated themselves vs how they were rated by others
2. Analysis of participants responses of the event i.e. did they have enough dates? Was 4 min too long or short?
3. Match rates by profession
4. Match rates by age
5. Match rates if they have met before
6. Further analysis of hobbies
7. What is the ideal number of dates?

We proceed to clean the data and make a log of all the changes made: [Link](https://docs.google.com/spreadsheets/d/1_kirP_4oF0MiHrdLSFat8LJXK-_4LPzuUM6sRqHYGPY/edit?usp=share_link) . Further analysis was done in Juypter Notebooks, data was then saved in .JASON files and sent to Software Engineers.

Description of Visuals

(all column references are from the Raw Data sets)

Participants page:

1. Do other people enjoy speed Dating Events? Aggregated Data on participants' responses to how much they enjoyed the event.
2. How likely are you to get a match? We did a simple bar graph to show how many matches the average participant received, how many calls after the event and how many dates they went on
3. What is the General Age of people who go?: We used column AH ‘age’, aggregated all the ages into a bar graph, 95% of them fell between ages of 21 and 28
4. What professions get the most matches? Used column AJ ‘field\_cd’, all the professions were assigned a code. Number of matches made per total dates was calculated for each profession.
5. What kind of hobbies do people have? Used columns AY to BO. Participants were asked to rank from 1-10 how interested they were in various activities. took an average of how each person ranked them.
6. How likely am I to match with someone I haven’t met before? Used column AG ‘met\_o’.

1- have met before, 2 have not met before. (This column was cleaned as per documentation in [google sheet](https://docs.google.com/spreadsheets/u/0/d/1_kirP_4oF0MiHrdLSFat8LJXK-_4LPzuUM6sRqHYGPY/edit)) Some of the participants were familiar with each other before the event, they had a higher chance of matching.

1. What attributes do people look for in others? Used columns BR to BW. participants were asked to rate the importance of attributes they were looking for in the opposite sex with a budget of 100. Waves 6-9 rated each attribute on a scale of 1-10 ( This was normalized to the above 100 point budget). Average was taken for all 6 attributes.
2. How self-aware are people? Used columns CJ to CN. participants were asked to rate how they measure up in each category on a scale of 1-10.
3. Are people just wanting to have some fun? Used column AT ‘goal’. Participants were surveyed about what their primary goal was in coming?

Organizers page:

1. Do people enjoy these kinds of events? Aggregated Data on participants' responses to how much they enjoyed the event.
2. What ages get the most matches? We used column AH ‘age’. Broke down the match rate of each age.
3. Were people happy with how many dates they went on? This event was conducted in waves ranging from 10-40 participants in size. Participants were asked their opinion on the number of dates they saw.
4. Do people like 4 minute dates? Used column DQ. Participants were asked if 4 min. was too long or short, responses were aggregated into a graph.
5. Does past familiarity increase match rate?Used column AG ‘met\_o’. Some of the participants were familiar with each other before the event, they had a higher chance of matching.
6. What careers are most likely to get a match? Used column AJ ‘field\_cd’, all the professions were assigned a code. Number of matches made per total dates was calculated for each profession.

Conclusions

Our intent was to provide both organizers and participants with useful data they can use. However further analysis can be more interesting and compelling:

A . What factors lead to high match rates among men and women?

B . Do participants select based on attributes they picked in the questionnaire

C . Does who is rotating play a role?