**Please make sure original files are not distrubed while doing the following.**

1. Filter out all responses except that of country codes US [United States] and IN [India].
2. There are 36 questions in the survey. Some of the questions that appear later in the survey are very similar or moderately similar to the questions that appear earlier. Using your own ability to comprehend, try to make brackets (any number that you find to be feasible) which contain similar questions that are trying to judge similar or same things from the respondent. Obviously, questions that are related to information like gender, age etc are to be treated differently.While doing this, **please note** thatquestions related to **“partner”** and **“romantic partner”** are to be treated as **different**. In cases where "**romantic"** is not specified exclusively, questions have to be treated as those for general close partners. For example, a person who is very expressive in front of a romantic partner may be very hesitant in front of other close partners. Similarly, a person who worries a lot about romantic relationships may not worry a lot about other close relationships.
3. Once the brackets have been made try to design a “consistency score” for each respondent. The idea behind consistency scores is that a person can be said to be consistent enough only if he answers questions in similar brackets uniformly enough. It then depends on your aptitude how you try to design the consistency score.
4. Once the consistency score has been designed, draw the distribution chart for consistency scores. After drawing the chart, mention how many responses you found to be **extremely inconsistent** after defining a mathematical criteria.
5. Filter out those extremely inconsistent responses and make a new dataset **N1**. Reduce the dimensionality of this new dataset by eliminating similar questions and call it dataset **N2.** In order to reduce the dimensionality, you can remove questions that are repeated and over which most people have been consistent. Mention what criteria you have used to reduce the dimensionality.
6. Using dataset **N1**, test the following hypothesis by **selecting relevant questions** that will help to answer each part.
7. In a romantic partnership, young males (<30) are more anxious and insecure than young females.
8. In a partnership (romantic or non romantic), males are more expressive (less hesitant) than females.
9. An American of age 30+ likes to be closer to romantic partners than an Indian of the same age does.
10. Indians have similar expressiveness towards romantic as well as non romantic partners.

7. Try to perform the same task(6) using reduced dataset **N2** and compare the certainty in results.