Choice of language:

The preferred language chosen for the assignment was python3 since I am more flexible and accustomed to it. The GUI application is done in tkinter since it comes inbuilt with python3 and does not require installing libraries or dependencies.

Response :
The current program registers the response of only one participant. The data which is displayed at the end of the 4 sessions can be dumped into either a database like SQL or mongoDB or can even stored as a tab-separated/ comma-separated text file (like an Excel file).

I'd prefer to store in the database for faster query operations, better indexing and to handle concurrent transactions. A database is ideally suited to handle larger volumes of data. Nevertheless the 1000 participants form an ideal scenario as well, because 2 people might wish to take the test at the same time, resulting in concurrent access, which a database is better equipped to handle.
We consider the entire data submitted by one participant as one single quanta of information, which along with a unique id forms a row in the database. The unique_id is determined by the timestamp of access for each participant. It is initialized to 1. As each new participant take the study, the id is incremented by 1 and is indexed along with the information submitted by the person. Consequently each row of the database has the following information.

<Unique ID>, <Favourite Color>, <Favourite Drink>, <Favourite Transport> <Value of color in S1>, <Value of color in S2>, <Value of color in S3>, <Value of color in S4>,<Value of drink in S1>, <Value of drink in S2>, <Value of drink in S3>, <Value of drink in S4>, <Value of transport in S1>, <Value of transport in S2>, <Value of transport in S3>, <Value of transport in S4>, <Time for S1>, <Time for S2>, <Time for S3>, <Time for S4>

The variables that could be of interest are the following :

1) Favourite color
2) Favourite drink
3) Favourite transport
4) The absolute time elapsed for S1, S2, S3, S4.

The variables in all of the cases correspond to a separate distribution
We can find the mean, and sd of the 4 sessions distribution and the mean of the favourite variables., which is nothing but a Bernoulli distribution

It would be interesting to see the trend that follows for the time elapsed (for each of the sessions). It is to be noted that the last session has a longer time than others due to the audio files.
Moreover, we can notice if the time elapsed for different sessions have some correlation with the random choices assigned to that particular session and the person's favourite choice.

The results for trends and distributions and correlation can be plotted using some visual libraries like gnuplot, matplotlib etc.