# **Test 1 – Mental Health**

1. This dataset is drawn from a sample post-intervention. The RCT attempted to improve the participants’ mental well-being. This was measured by multiple scales and combined into a single mental health score (m health). A higher score indicates better mental health. We are interested in finding out the impact of the intervention. The variable treatment indicates whether a person was in the treatment group. (Kindly provide the code for any analysis you perform using the program of your choosing. If using excel, please provide the excel file).
   1. How many participants are there in the treatment group? And how many in the control group?
   2. Answer the below questions only for the ”Red hostel” in “Karnataka” state.
      1. How many missing values are there in the m health variable? How can you deal with the missing values?
      2. How many rooms are there in the hostel?
      3. What is the minimum number of individuals in a room? Which room?
      4. Which floor has the lowest number of rooms?
      5. What is the average number of people per floor?
   3. One of the variables in this dataset records the participant’s education level.
      1. Generate a relationship between income and education.
      2. In the distribution, which level of education has the lowest average income
   4. Run any analysis you seem fit between mental health score and treatment indicator to study impact of the intervention. Clearly state the direction of the effect; whether the effect was significant; if yes, the level of significance. [Note: Kindly provide detailed explanation on how you would approach the question and why using your method.]
   5. Now, your research manager asks you to control for income in the regression. The income can be classified into 3 categories - low (below 30,000), middle (30,000-35,000) and high (above 35,000). Will you use the continuous income variable or convert it into a categorical variable and use it? Why so? [Note: We are not only asking you to code an additional analysis but also provide the explanation for using the appropriate income variable.]

# **Test 2 – Simpon Transfer Log**

1. This dataset is drawn from a production transfer log. Kindly provide a detailed report (max 2 pages) deriving linkages and relationships you can infer from the data. Further, provide the code for any analysis you perform using the program of your choosing. If using excel, please provide the excel file.