PURPOSE:

To write a python application that will take a user CSV file and perform analysis on the data

INPUT:

The program will take a user input for a CSV file name, and the file will be an input data for the code

OUTPUT:

1. Number of edges used to identify triads
2. Number of positive (trust) edges
3. Number of negative (distrust) edges
4. Probability p that an edge will be positive: (number of positive edges) / (total edges)
5. The probability that an edge will be negative: $1 - p$
6. Number of triangles of each type and the total number of triangles
1. Expected distribution of triad types (based on p and $1 - p$ applied to the number of triangles in the graph). Show number and percentage:
    a. Trust-Trust-Trust
    b. Trust-Trust-Distrust
    c. Trust- Distrust -Distrust
    d. Distrust- Distrust- Distrust
    e. Total
7. The actual distribution of triad types. Show number and percentage:
    a. Trust-Trust-Trust
    b. Trust-Trust-Distrust
    c. Trust- Distrust -Distrust
    d. Distrust- Distrust- Distrust
    e. Total

WHAT THE PROGRAM DOES:

Upon taking the user input file, the program analyses the data representing relationships in the network. First, the CSV data is converted into graph data type, and then the analysis is performed. The final analysis data is printed on the terminal in a tabular format.

RESULT:

Our result shows the total number of triads formed in the data provided, along with the type of relationship they hold and specifics about types of edges. We also show the expected and actual distribution of triad types based on the probability of positive and negative edges we calculated. We can notice a difference in the expected vs actual distribution. I feel this happens because when we are calculating the expected distribution, we purely look at the probability of an individual edge being positive or negative and then calculate accordingly. While in the actual distribution, we have to examine how the triads are in relation to each other.

ADDITIONAL INFORMATION:

Please install the required package mentioned in the code before executing the program