To answer the questions below, access the "books.csv" file from Canvas. The dataset contains 500 transactions of customers across 5 categories of books. An entry of 1 indicates purchase and an entry of 0 indicates no purchase.

1. Compute the following (based on the 5 categories):
    a. Euclidean distance between customers 245 and 431
    b. Manhattan distance between customers 82 and 197
    c. Centroid of the first 50 customers
2. Which two genres of books have:
    a. the highest co-occurrence?
    b. the lowest co-occurrence?
3. Suppose we cluster the customers based on the total number of books purchased. What is the size of each cluster?
4. Compute the support of the following itemsets:
    a. {fiction}
    b. {non_fiction}
    c. {fiction, self_help}
5. Compute the confidence of the following association rules:
    a. {fiction} → {mystery}
    b. {non_fiction} → {self_help}
    c. {fiction, self_help} → {childrens_books}
6. Compute the lift of the following association rules:
    a. {fiction, self_help} → {childrens_books}
    b. {fiction} → {non_fiction}
    c. {non_fiction} → {self_help}
7. Explain the meaning of the following:
    a. Support of {fiction, self_help}
    b. Confidence of {fiction, self_help} → {childrens_books}
    c. Lift of {fiction, self_help} → {childrens_books}