Homework Problem Set

Tasks

1. Column Creation

a) Calculate the 75th percentile for the following specified columns:

- `PovertyLT100_FPL_score`
- `Single_Parent_Fam_score`
- `Education_LT12years_score`
- `HHNo_Vehicle_score`
- `HHRenter_Occupied_score`
- `HHCrowding_score`
- `Nonemployed_score`

b) Create new columns in the DataFrame that flag whether each value in the specified columns exceeds the calculated 75th percentile.

2. Frequent Itemsets Analysis

a) Select the columns with flag data created in the previous step for analysis.

b) Convert the selected flag data into a binary format suitable for mining frequent itemsets.

c) Run the Apriori algorithm on the formatted data to identify frequent itemsets, using a minimum support threshold of 0.05.

3. Association Rule Mining

a) Generate association rules from the frequent itemsets identified in the previous step.

b) Rank generated rules by metrics such as lift.

c) Present the top 25 association rules based on the chosen metric (highest lift values).

4. Results Analysis

a) Interpret the generated association rules by explaining the relationships between the antecedents and consequents.

b) Discuss the practical significance of the top rules, and how they might be applied or understood in the context of the dataset.

5. Optional Visualization Task

a) Visualize the relationships between the antecedents and consequents of the top association rules using a network graph.

b) Label the graph with relevant metrics, such as lift values, and provide an interpretation of the visualization to help explain the connections shown in the graph.