Assignment 3 - Pattern Mining

COMP3401

Submission Instructions

- Handwritten submission is not allowed. It was a lot of trouble reading the submitted work in the previous assignment.
- You can use Word/Excel/Visio or any other tool that works for you but make the submission easy to read. Otherwise, it may not be graded.
- Submit the following:
  - pdf containing solutions both questions. Convert the code notebooks for q2 to pdf and combine with q1 to submit the final pdf.
  - submit the results (rules.csv) and the code notebook as well for q2.

Question 1 [50]

A dataset consists of 5 separate transactions. Define minimum support (min_sup) as 60% and minimum confidence (min_conf) as 80%.

The following table presents the transaction ID (TID) alongside the respective items purchased:

TID	items_purchased
T1	{A, B, C, D, E, F}
T2	{G, B, C, D, E, F}
T3	{A, H, D, E}
T4	{A, I, J, D, F}
T5	{J, B, B, D, K, E}

Proceed with the following tasks:

Question 1 Part 1 [40]

Identify all frequent itemsets by applying both the Apriori and the FP-growth algorithms.

- Solve the problem using a method analogous to the one demonstrated in Example 4.3 for the Apriori algorithm. Include a diagram akin to Figure 4.2 on page 152 and provide all the necessary calculations and steps. [20]
- For the FPGrowth algorithm, refer to Example 4.5 as your guide. Include a representation of the tree similar to the one in Figure 4.7, and construct a table comparable to Table 4.2. Ensure that all the required work is clearly shown. [20]

Question 1 Part 2 [10]

List all the strong association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and \textit{item}$_i$ denotes variables representing items (e.g., `A`, `B`):

$$\forall X \in \text{transaction}, \ \text{buys}(X, \text{item}_1) \wedge \text{buys}(X, \text{item}_2) \Rightarrow \text{buys}(X, \text{item}_3) \ [s, c]$$

Question 2: Use Apriori to mine for frequent itemsets. [50]

Open the starter.ipynb file provided and write code in the cells that have the `"# Your Code Here"` comment