**EECS 4412 Data Mining**

**Assignment 1 Report**

Name: Chun Ho Li

Student ID: 214908800

Date: 22 Feb 2021

**Question 1)**

**Part A**

levelUp.py works by applying Apriori algorithm to mine frequent itemset for a given threshold.Before applying the pre-candidate optimization, the algorithm will join on the prefixes which is created randomly to create pre-candidates, then applying the Apriori property to finalize the candidates. This can lead to a huge set of pre-candidates’ generation, result in a slower running time. Here is the result before optimization by testing with level 5 of the mushroom dataset.



To achieve a more efficient algorithm, we sort the item in the itemset from least to most frequent, such that the pre-candidates’ generation can be more optimized and less pre-candidates will be created. Hence, increasing the performance of the algorithm. Here is the result with optimization, the number of pre-candidates are greatly reduced, and the running time is slightly improved as a result.



**Question 2)**

**Part A)**

**Diagram

Description automatically generated**

**Part B)**

**Text, letter

Description automatically generated**

**Text, table, letter

Description automatically generated**