# Project #1 – Principles of Data Mining – Report

**Team ID:**

**Team members**:

*Answer the following questions. Your answers should be in dark blue. The plots requested should be in this file. Do not submit them as separate files.*

**Question 0**

How did each member of the team contribute to the project?

<answer>

Cite any sources that you used to complete the project.

<answer>

**Question 1**

Describe in detail the data structures that you used (how you stored the data) for:

a. the transactions:

the data structure containing transactions is a list of sets with each set representing a transaction.

b. the frequent itemsets:

the data structure containing the frequent itemsets is a dictionary. the dictionary keys are the length of the itemsets, the dictionary values is a frequent itemset dictionary. The frequent itemset dictionary keys are the frequent itemsets and values are the support counts of the respective itemset

c. rules generated:

the data structure containing the rules is a list of tuples each tuple containing 3 elements. The first two elements of the tuple represents each side of the rule implication respectively, and the last element represents the rules confidence

**Question 2**

Run your code for *minconf=0.8* and the following values for the *minsup = {50, 75, 100, 125, 150, 200}* (note that this is the support count). Plot the amount of **time** required to generate the frequent itemsets for the different values of minimum support. Make sure that each axis is properly annotated with the quantity that it corresponds to.

<plot>

What can you understand from the plot?

<answer>

**Question 3**

Run your code for *minconf=0.8* and the following values for the *minsup = {50, 75, 100, 125, 150, 200}* (same as in Question 2). Plot the **number of frequent itemsets** generated for the different values of minimum support. Make sure that each axis is properly annotated with the quantity that it corresponds to.

<plot>

What can you understand from the plot?

<answer>

**Question 4**

For *minsup=80*, generate the rules at different levels of confidence, i.e., *minconf={0.7, 0.75, 0.8, 0.85, 0.9}.* Plot the **number of rules** that were found for the different values of minconf threshold. Make sure that each axis is properly annotated with the quantity that it corresponds to.

<plot>

What can you understand from the plot?

<answer>

**Question 5**

Submit the files produced when minsup=140 and minconf=0.8.