CS 378
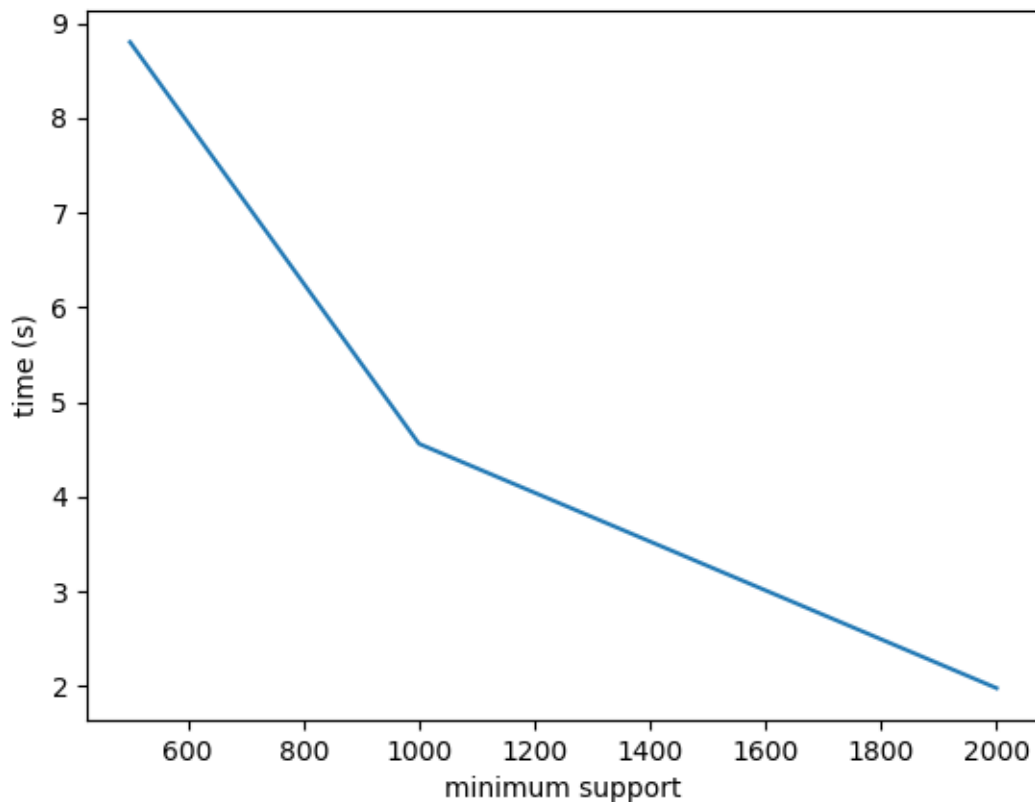
Homework 1 Report

Yicheng (Jason) Wang

1. The performance of my implementation of Apriori Algorithm is as follow:

Performance on T10I4D100K data set (running on my own computer)

2. The optimization method is mainly just pruning the dataset so that item which is not frequent will not appear in the whole transaction list. Since I implement the method in python, most of the operations are just set operations. Therefore, I am able to write most of the methods in one or two lines. I use a lot of python list comprehension for succinctness and also speed, so it might be hard for you to read the source code.

   However, I structure the code in a friendly way to read, I hope it's not too messy for you to grade it. Thanks.