COL-870
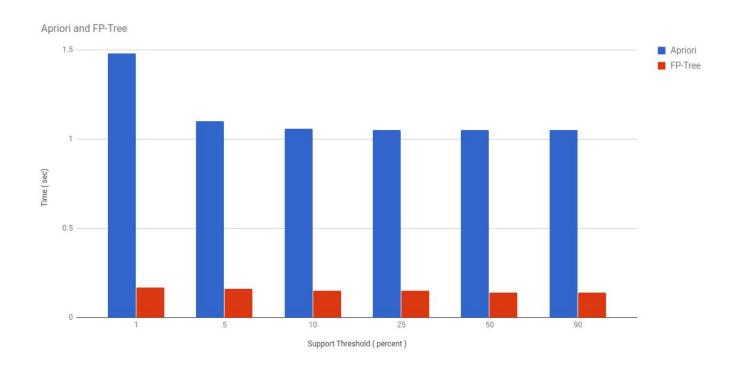
Homework-2

Shashank Yadav

2013CS50799

FP-Growth library : As given on piazza (http://www.borgelt.net/fpgrowth.html)

i) Following is the graph showing the times of apriori and FP-tree algorithms for specified support thresholds. There is an exponential degradation in the performance of apriori as compared to FP tree.
Apriori generates all potential candidates and then prunes them on the basis on min support. This is quite wasteful as the number of generated subsets would be exponential while only some of them would be useful. So, for lower minimum threshold the number of generated candidates would increase exponentially, leading to exponential increase in time. On the other hand, FP tree works with suffixes and hence the growth in complexity is not exponential.

Apriori and FP-Tree

ii) Following is the graph showing the times of apriori and FP-tree algorithms for specified data sizes with min threshold kept at 10%. Again exponential increase in time as the data size is increased. Since apriori generates all subsets, the number of subsets would increase exponentially with linear increase in data and hence time would also increase similarly. Moreover, apriori reads the transaction list several times while FP tree only does it twice, hence larger transaction list size also leads to a decrease in performance for apriori compared to FP tree.