# Mining Frequent Itemsets Apriori

## Running the program

Ensure the database is in the same folder as the source code and then, in the command line, input: python (sourceCode) (databaseName) (supportlvl). Do not forget the extension.

So in my case: python apriori.py retail.dat 10000

Failure to include the database name or give a non-integer value will result in a run-time error.

## Gathering results

To collect data, I made use of python’s time module to record the time elapsed while the program is working. At the end, the time in both seconds and minutes are outputted.

## Notes

Caution: Don’t attempt a threshold of 2. I was forced to restart due way too many frequent items being found. I picked a threshold of 50 and I was fine. It’s safe to say that this implementation does not handle huge amounts of data.

Update 2: A threshold of 50 doesn’t crash but it takes way too long to get a result (8 minutes for only 20K buckets…)

Update 3: Started using Pypy. Drastic decrease in time.

## http://puu.sh/nz4oT/1f40e6eadf.png

Figure - Before Pypy