

Untitled2

April 3, 2022

```
[ ]: import PAMI.extras.dbStats.transactionalDatabaseStats as tds
import PAMI.extras.graph.plotLineGraphFromDictionary as plt

obj = tds.transactionalDatabaseStats('transactional_BMS_POS.csv', sep=',')
# obj = tds.transactionalDatabaseStats(inputFile, sep=',') #override default
  ↳tab seperator
obj.run()

print(f'Database size : {obj.getDatabaseSize()}')
print(f'Total number of items : {obj.getTotalNumberOfItems()}')
print(f'Database sparsity : {obj.getSparsity()}')
print(f'Minimum Transaction Size : {obj.getMinimumTransactionLength()}')
print(f'Average Transaction Size : {obj.getAverageTransactionLength()}')
print(f'Maximum Transaction Size : {obj.getMaximumTransactionLength()}')
print(f'Standard Deviation Transaction Size : {obj.
  ↳getStandardDeviationTransactionLength()}')
print(f'Variance in Transaction Sizes : {obj.getVarianceTransactionLength()}')

itemFrequencies = obj.getSortedListOfItemFrequencies()
transactionLength = obj.getTransactionLengthDistribution()
obj.storeInFile(itemFrequencies, 'itemFrequency.csv')
obj.storeInFile(transactionLength, 'transactionSize.csv')

plt.plotLineGraphFromDictionary(itemFrequencies,50,'item frequencies', 'item
  ↳rank', 'frequency')
#above command plots the graph with top 50% of the elements in the dictionary
plt.plotLineGraphFromDictionary(transactionLength,100,'distribution of
  ↳transactions', 'transaction length', 'frequency')
#above command plots the graph with top 100% of the elements in the dictionary
```

```
Database size : 3367020
Total number of items : 515598
Database sparsity : 0.9999961210090031
Minimum Transaction Size : 2
Average Transaction Size : 2.0
Maximum Transaction Size : 2
Standard Deviation Transaction Size : 0.0
Variance in Transaction Sizes : 0
```

[]: