Ye Chen

Data mining

Homework 1

Question 1:

A: if the transformation is run on one document , it will pick out the important word within that document and the data will only relate to that documents only, however if it is run using all document, then the data would be relate towards all documents instead of only that one.

B: Tf-idf , is use to calculate how important a word is in a document , for example search engine use this to pick out the website you were looking for

Question 2:

Homework1(Q2)1.py

This pie chart shows the total amount sold percentage compared Furniture 27.1% vs Electrotonic is 72.9%, it shows that people are buying more Electronics compared to Furniture

Homework1(Q2)2.py

This is a bar chart shows the total unit of electronics vs furniture sold and the results is electronics out sold Furnitures

Homework1(Q2)3.py

This line chart shows the overall performance of each category of items throughout the year

Homework1(Q2)4.py

This pie chart shows the performance of individual category items

Question 3:

Homework1Q3.py