**CS 634 Data Mining**

**Midterm Project: Apriori Algorithm**

**Alfred Zane Rajan Velladurai**

**UCID: ar2257**

**Source Code: (In Python)**

import pandas as pd

support = 50 #float(sys.argv[1])

confidence = 70 #float(sys.argv[2])

file = "db1.txt" #(sys.argv[3])

df = pd.read\_csv( file, skipinitialspace = True, names = ['0', '1','2','3','4','5','6','7','8','9','10'])

df = df.drop('0', axis=1)

df = df.T

for col in df:

df[col] = df[col].str.replace(" ", "")

db = list(list(df[i]) for i in df)

db = [[j for j in i if not pd.isna(j)] for i in db]

#for trans in db:

#item = ["diapers", "sweaters", "tissues", "belts", "water", "noodles", "cereals", "books", "pen", "batteries"]

items = []

for trans in db:

for item in trans:

if not item in items:

items.append(item)

support = support\*len(db)/100

confidence = confidence/100

count = {}

for item in items:

count[item] = 0

for item in items:

for trans in db:

if item in trans:

count[item]+=1

for item in items:

if count[item] < support:

items.remove(item)

itemsets = [{i} for i in items]

while itemsets:

tempsets = []

for iset in itemsets:

for item in items:

if not {item}.issubset(iset):

match = 0

total = 0

for trans in db:

if iset.issubset(trans):

total+=1

if item in trans:

match+=1

if match>=support:

temp = {i for i in iset}

temp.add(item)

if temp not in tempsets:

tempsets.append(temp)

if (match/total)>=confidence:

print(iset," -> ",item)

itemsets = tempsets

for item in items:

flag = 0

for iset in itemsets:

if {item}.issubset(iset):

flag = 1

if flag == 0:

items.remove(item)

P.S.: I had run the program on Spyder 3.2.8 on Windows on my PC. So I have hard coded the support, confidence and input path into the code. But I have also given the code to read from the command line (Commented).

For all Databases:

Support = 50%

Confidence = 70%

**Data Base 1:**

**Input:**

1001, diapers, sweaters, tissues, belts, water, noodles, cereals, books, pen, batteries

1002, noodles, pen, books, sweaters, water

1003, sweaters, tissues, noodles, cereals, belts, books

1004, pen, tissues, batteries, diapers, water

1005, water, diapers, belts, books

1006, sweaters, batteries, cereals, tissues, diapers, books

1007, batteries, water, sweaters, belts, noodles, books, diapers, cereals, tissues, pen

1008, water, belts, pen, books, tissues

1009, books, tissues, batteries, diapers, noodles

1010, pen, noodles, books, sweaters, batteries, belts, tissues

1011, water, noodles, cereals, sweaters, tissues, diapers, belts

1012, belts, tissues, batteries

1013, belts, water, tissues, batteries

1014, batteries, sweaters, belts, diapers, pen, books, water, cereals, tissues

1015, noodles, batteries, cereals, tissues, pen

1016, noodles, cereals, sweaters, diapers, water, batteries

1017, belts, pen

1018, diapers, sweaters, water, noodles, books, batteries

1019, diapers, tissues, water, noodles, pen, batteries

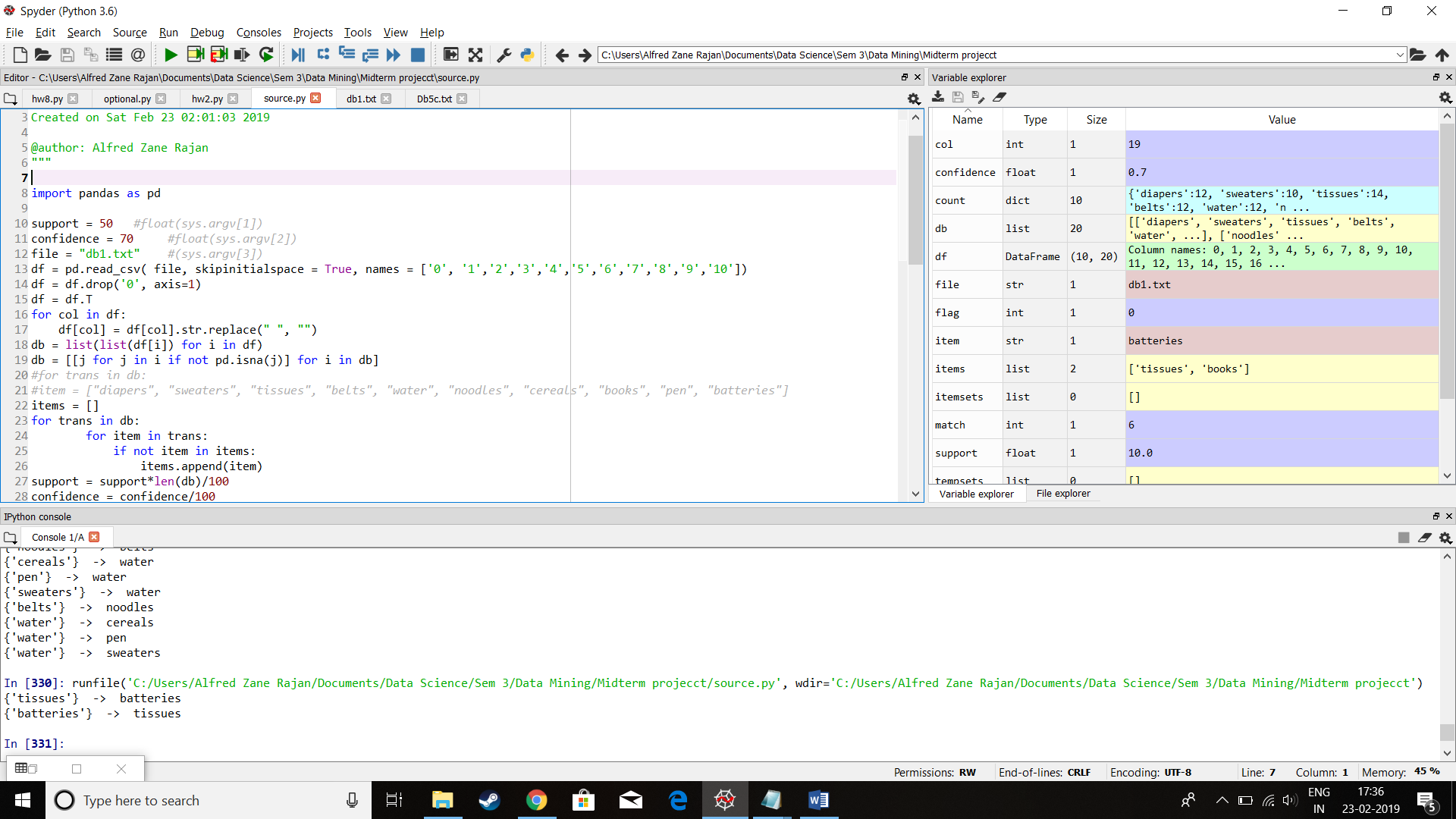
1020, diapers, belts, noodles,

**Output:**

runfile('C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct/source.py', wdir='C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct')

{'tissues'} -> batteries

{'batteries'} -> tissues



**Data Base 2:**

**Input:**

1001, diapers, sweaters, tissues, belts, water, noodles, cereals, books, pen, batteries

1002, sweaters, tissues, noodles, cereals, books, batteries

1003, sweaters, tissues, belts, noodles, cereals, pen, batteries

1004, tissues, belts, water, noodles, cereals, books, pen, batteries

1005, tissues, belts, water, cereals, pen, batteries

1006, diapers, tissues, belts, water, noodles, cereals, books, batteries

1007, sweaters, cereals, books

1008, diapers, belts, water, pen

1009, diapers, sweaters, tissues, water, pen

1010, diapers, belts, water, noodles, books, pen, batteries

1011, diapers, belts, water, noodles, cereals, pen, batteries

1012, diapers, tissues, belts, water, cereals, batteries

1013, sweaters, water, noodles, cereals, books, batteries

1014, belts, pen

1015, sweaters, tissues, belts, water, noodles, cereals, batteries

1016, diapers, tissues, water, noodles, pen

1017, diapers, water, cereals, batteries

1018, diapers, belts, cereals, pen

1019, sweaters, belts, water, noodles, cereals, pen, batteries

1020, diapers, tissues, water, cereals, books, pen

**Output:**

runfile('C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct/source.py', wdir='C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct')

{'diapers'} -> water

{'belts'} -> water

{'belts'} -> cereals

{'belts'} -> pen

{'belts'} -> batteries

{'water'} -> cereals

{'water'} -> batteries

{'noodles'} -> batteries

{'cereals'} -> water

{'cereals'} -> batteries

{'pen'} -> belts

{'pen'} -> water

{'batteries'} -> belts

{'batteries'} -> water

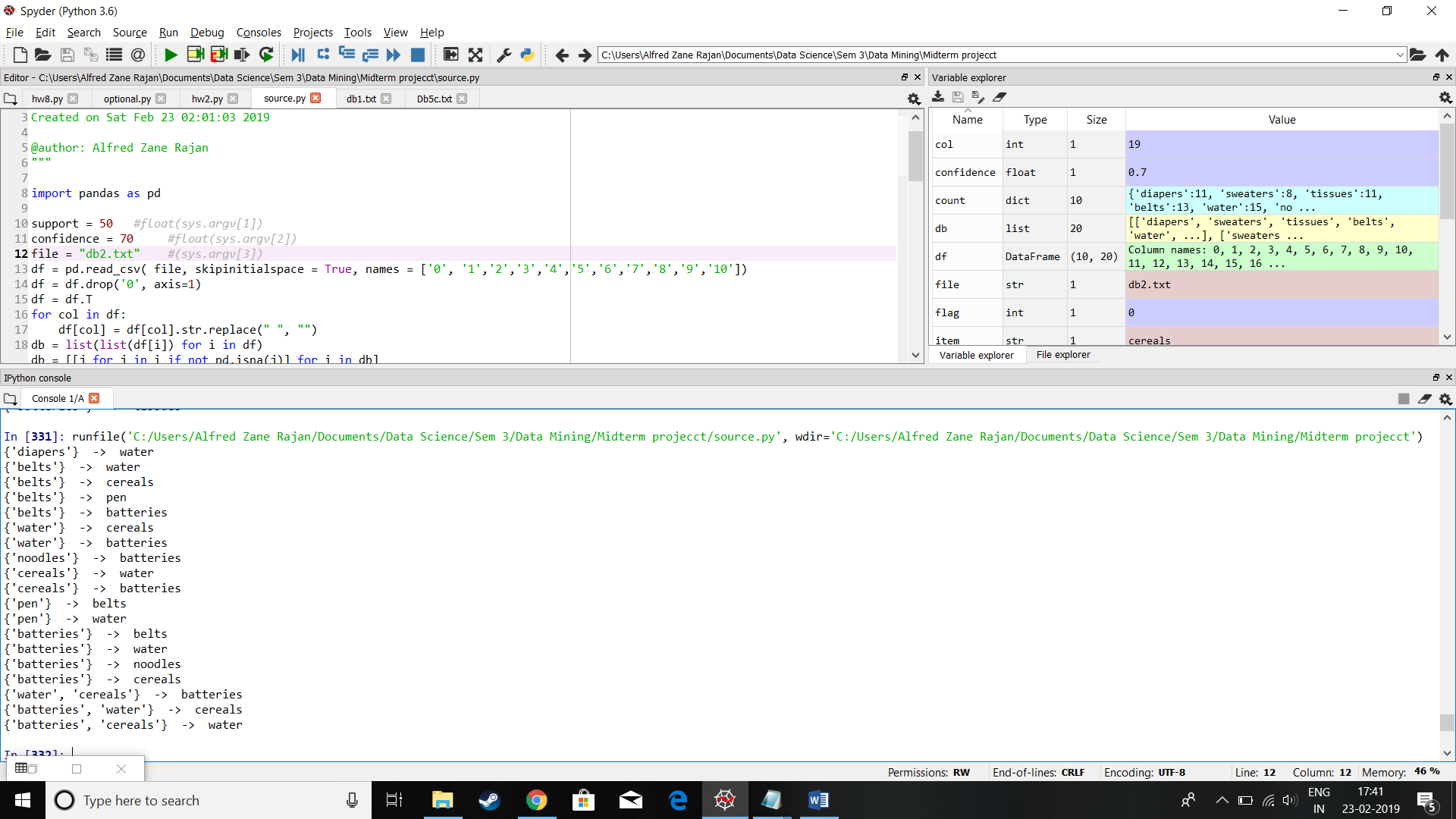
{'batteries'} -> noodles

{'batteries'} -> cereals

{'water', 'cereals'} -> batteries

{'batteries', 'water'} -> cereals

{'batteries', 'cereals'} -> water



**Data Base 3:**

**Input:**

1001, diapers, sweaters, tissues, belts, water, noodles, cereals, books, pen, batteries

1002, diapers, tissues, belts, water, noodles, cereals, books, pen, batteries

1003, diapers, belts, water, noodles, cereals, pen, batteries

1004, diapers, tissues, belts, water, noodles, books, pen, batteries

1005, sweaters, belts, noodles, cereals, books, pen, batteries

1006, diapers, tissues, water, cereals, pen

1007, diapers, sweaters, belts, water, cereals, pen, batteries

1008, water, cereals, books, pen, batteries

1009, diapers, sweaters, tissues, belts, cereals, books, pen, batteries

1010, sweaters, tissues, belts, water, cereals, books, pen

1011, diapers, sweaters, tissues, belts, water, noodles, cereals, pen, batteries

1012, sweaters, tissues, belts, water, noodles, cereals, books, pen, batteries

1013, diapers, sweaters, tissues, noodles, cereals, pen, batteries

1014, tissues, water, noodles, cereals, pen, batteries

1015, diapers, sweaters, tissues, belts, water, noodles, pen

1016, sweaters, belts, water, noodles, pen, batteries

1017, diapers, sweaters, belts, water, cereals, books, pen, batteries

1018, diapers, tissues, water, noodles, pen

1019, diapers, tissues, water, noodles, cereals, books, pen, batteries

1010, diapers, tissues, water, noodles, cereals, books

**Output:**

runfile('C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct/source.py', wdir='C:/Users/Alfred Zane Rajan/Documents/Data Science/Sem 3/Data Mining/Midterm projecct')

{'water'} -> pen

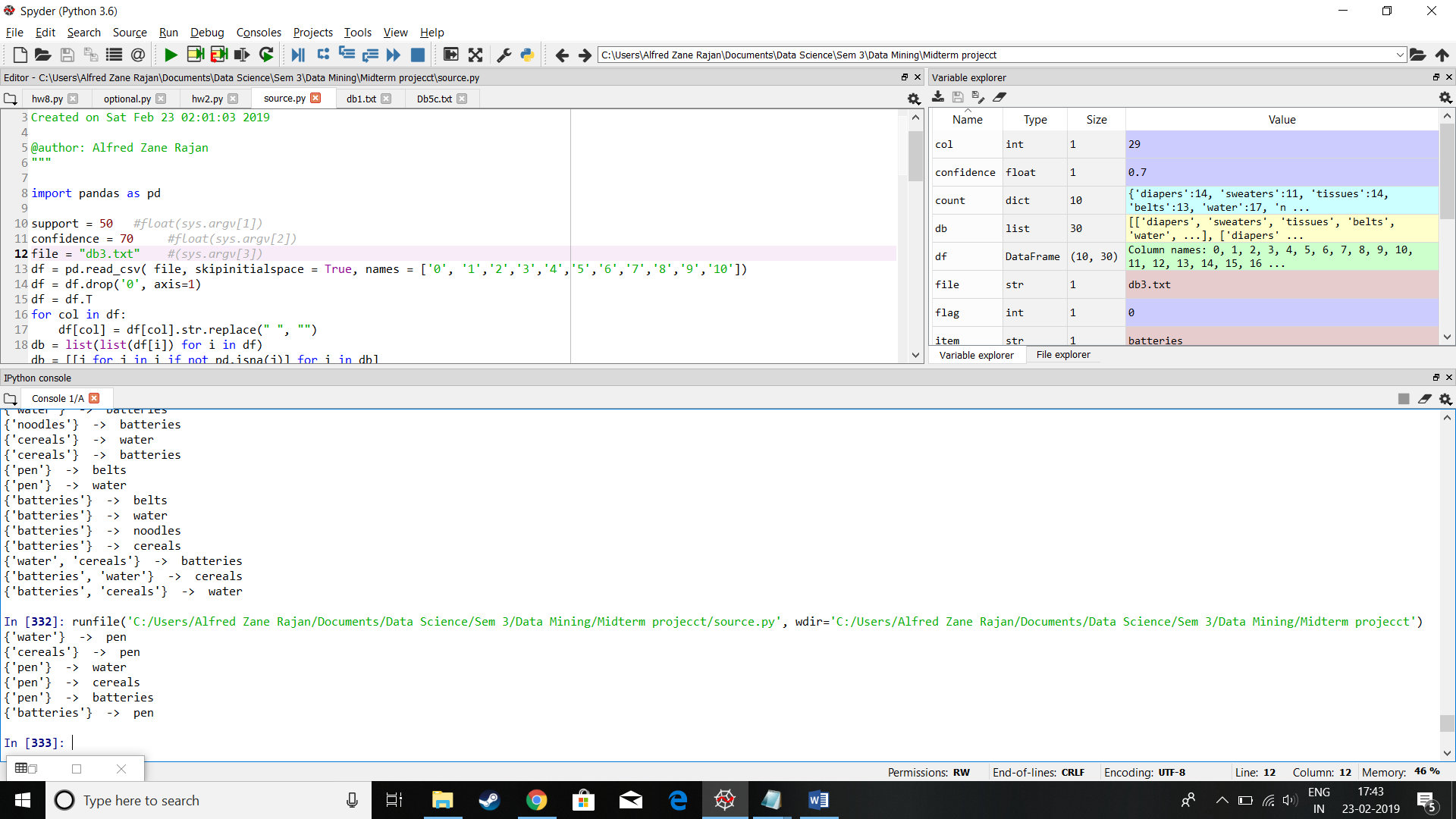
{'cereals'} -> pen

{'pen'} -> water

{'pen'} -> cereals

{'pen'} -> batteries

{'batteries'} -> pen



**Data Base 4:**

**Input:**

1001, diapers, sweaters, tissues, belts, water, noodles, cereals, books, pen, batteries

1002, diapers, tissues, belts, water, noodles, cereals, books, pen, batteries

1003, diapers, belts, water, noodles, cereals, pen, batteries

1004, diapers, tissues, belts, noodles, books, pen, batteries

1005, belts, noodles, cereals, books, pen, batteries

1006, diapers, tissues, water, cereals, pen

1007, diapers, sweaters, water, cereals, batteries

1008, water, cereals, books, pen, batteries

1009, diapers, sweaters, tissues, belts, books, pen, batteries

1010, sweaters, belts, water, cereals, books, pen

1011, diapers, belts, water, noodles, cereals, pen, batteries

1012, sweaters, belts, water, noodles, cereals, books, pen, batteries

1013, diapers, tissues, noodles, cereals, pen, batteries

1014, tissues, water, noodles, cereals, pen, batteries

1015, diapers, sweaters, tissues, belts, water, noodles, pen

1016, sweaters, belts, water, noodles, pen, batteries

1017, diapers, sweaters, belts, cereals, books, pen, batteries

1018, diapers, tissues, water, noodles, pen

1019, diapers, water, noodles, cereals, books, pen, batteries

1010, diapers, tissues, water, noodles, cereals, books

**Output:**

{'diapers'} -> water

{'diapers'} -> noodles

{'diapers'} -> cereals

{'diapers'} -> pen

{'diapers'} -> batteries

{'belts'} -> pen

{'belts'} -> batteries

{'water'} -> noodles

{'water'} -> cereals

{'water'} -> pen

{'noodles'} -> diapers

{'noodles'} -> water

{'noodles'} -> cereals

{'noodles'} -> pen

{'noodles'} -> batteries

{'cereals'} -> water

{'cereals'} -> pen

{'cereals'} -> batteries

{'books'} -> pen

{'pen'} -> water

{'pen'} -> noodles

{'pen'} -> cereals

{'pen'} -> batteries

{'batteries'} -> noodles

{'batteries'} -> cereals

{'batteries'} -> pen

{'belts', 'pen'} -> batteries

{'belts', 'batteries'} -> pen

{'noodles', 'water'} -> pen

{'water', 'cereals'} -> pen

{'pen', 'water'} -> noodles

{'pen', 'water'} -> cereals

{'pen', 'noodles'} -> water

{'pen', 'noodles'} -> batteries

{'batteries', 'noodles'} -> pen

{'pen', 'cereals'} -> water

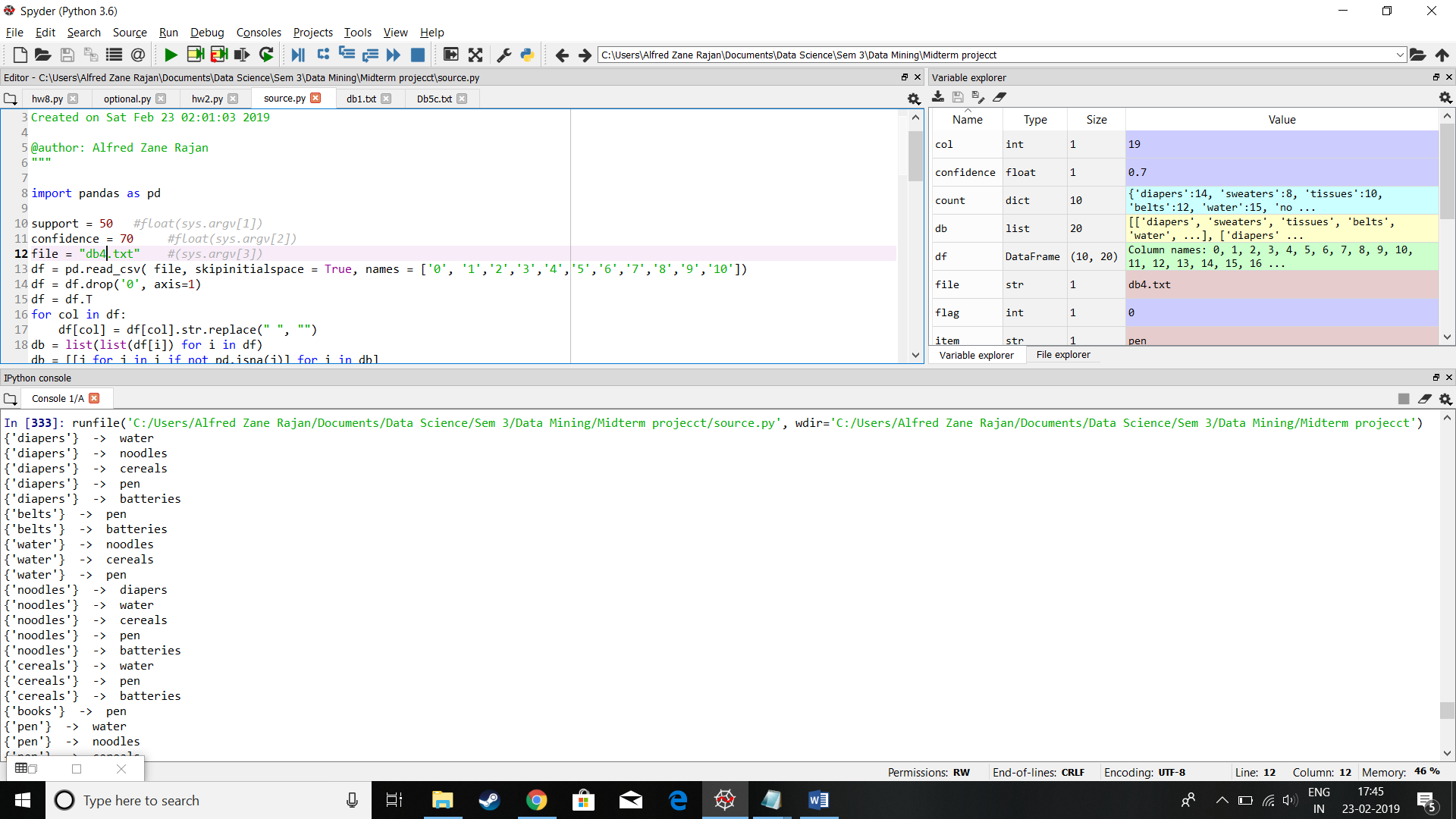
{'pen', 'cereals'} -> batteries

{'batteries', 'cereals'} -> pen

{'pen', 'batteries'} -> belts

{'pen', 'batteries'} -> noodles

{'pen', 'batteries'} -> cereals



**Data Base 5:**

**Input:**

1001, diapers, tissues, noodles, cereals, books, pen, batteries

1002, diapers, sweaters, belts, water, noodles, cereals, books, pen, batteries

1003, sweaters, tissues, water, noodles, cereals, batteries

1004, diapers, sweaters, tissues, cereals

1005, sweaters, tissues, belts, water, noodles, pen

1006, belts, water, noodles, cereals, books, pen, batteries

1007, diapers, sweaters, tissues, water, cereals, pen,

1008, sweaters, tissues, belts, water, noodles, cereals, books, batteries

1009, diapers, tissues, water, cereals, books, pen, batteries

1010, diapers, tissues, belts, water, noodles, cereals, books, pen, batteries

1011, diapers, sweaters, tissues, water, books, pen, batteries

1012, sweaters, tissues, belts, noodles, books, batteries

1013, diapers, sweaters, books, pen, batteries

1014, diapers, belts, noodles, cereals

1015, diapers, sweaters, belts, water, noodles, cereals, books, pen

1016, sweaters, tissues, belts, water, noodles, cereals

1017, diapers, sweaters, tissues, belts, water, pen

1018, sweaters, tissues, water, cereals, pen, batteries

1019, sweaters, belts, water, noodles, cereals, pen, batteries

1010, sweaters, water, cereals

**Output:**

{'tissues'} -> sweaters

{'tissues'} -> water

{'noodles'} -> cereals

{'noodles'} -> belts

{'cereals'} -> water

{'pen'} -> water

{'sweaters'} -> water

{'belts'} -> noodles

{'water'} -> cereals

{'water'} -> pen

{'water'} -> sweaters

