

DATA MINING MIDTERM **PROJECT**

Full Name: Rutuja Bendre

NJIT UCID: rb264

Email: rb264@njit.edu

Programming language: Python

Amazon.csv

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition

Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition,Beginning Programming with Java,,

Android Programming: The Big Nerd Ranch,Beginning Programming with Java, Java 8 Pocket Guide,,

A Beginner's Guide,Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition,,

A Beginner's Guide,Head First Java 2nd Edition,Beginning Programming with Java,,

Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,,

Java For Dummies,Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition,Beginning Programming with Java,

Beginning Programming with Java,Java 8 Pocket Guide,C++ Programming in Easy Steps,,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,HTML and CSS: Design and Build Websites,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Java 8 Pocket Guide,HTML and CSS: Design and Build Websites

Java For Dummies,Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition,,

Java For Dummies,Android Programming: The Big Nerd Ranch,,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,

A Beginner's Guide,Java: The Complete Reference,Java For Dummies,Android Programming: The Big Nerd Ranch,

Head First Java 2nd Edition,Beginning Programming with Java,Java 8 Pocket Guide,,

Android Programming: The Big Nerd Ranch,Head First Java 2nd Edition,,

A Beginner's Guide,"Java: The Complete Reference, Java For Dummies",,,U

BESTBUY.CSV

Desk Top, Printer, Flash Drive, Microsoft Office, Speakers, Anti-Virus

Lab Top,Flash Drive,Microsoft Office,Lab Top Case,Anti-Virus,,,,,

Lab Top,Printer,Flash Drive,Microsoft Office,Anti-Virus,Lab Top Case,External Hard-Drive,,,

Lab Top,Printer,Flash Drive,Anti-Virus,External Hard-Drive,Lab Top Case,,,,,

Lab Top,Flash Drive,Lab Top Case,Anti-Virus,,,,,

Lab Top,Printer,Flash Drive,Microsoft Office,,,,,

Desk Top,Printer,Flash Drive,Microsoft Office,,,,,

Lab Top,External Hard-Drive,Anti-Virus,,,,,

Desk Top,Printer,Flash Drive,Microsoft Office,Lab Top Case,Anti-Virus,Speakers,External Hard-Drive,,

Digital Camera,Lab Top,Desk Top,Printer,Flash Drive,Microsoft Office,Lab Top Case,Anti-Virus,External Hard-Drive,Speakers

Lab Top,Desk Top,Lab Top Case,External Hard-Drive,Speakers,Anti-Virus,,,,,

Digital Camera,Lab Top,Lab Top Case,External Hard-Drive,Anti-Virus,Speakers,,,,,

Digital Camera,Speakers,,,,,

Digital Camera,Desk Top,Printer,Flash Drive,Microsoft Office,,,,,

Printer,Flash Drive,Microsoft Office,Anti-Virus,Lab Top Case,Speakers,External Hard-Drive,,,

Digital Camera,Flash Drive,Microsoft Office,Anti-Virus,Lab Top Case,External Hard-Drive,Speakers,,,

Digital Camera,Lab Top,Lab Top Case,,,,,

Digital Camera,Lab Top Case,Speakers,,,,,

Digital Camera,Lab Top,Printer,Flash Drive,Microsoft Office,Speakers,Lab Top Case,Anti-Virus,,

Digital Camera,Lab Top,Speakers,Anti-Virus,Lab Top Case,,,

KMART.CSV

Decorative Pillow

Bedspread, Embroidered

Bedspread

Embroidered Bedspread,"Shams","Kids Bedding

",Bedding Collections,"Bed Skirts",Bedspreads,"Sheets"

Decorative Pillows,"Quilts

",Embroidered Bedspread,"Shams","Kids Bedding

",Bedding Collections,

Kids Bedding,Bedding Collections,"Sheets",Bedspreads,"Bed Skirts",,

",Bedding Collections,"Sheets","Bed Skirts",Bedspreads,
 Bedding Collections,Bedspreads,"Bed Skirts","Sheets","Shams","Kids Bedding",
 Decorative Pillows,Quilts,,,,,
 "Decorative Pillows","Quilts,Embroidered Bedsread,,,,
 Bedspreads,"Bed Skirts","Shams","Kids Bedding
 ","Sheets",,
 Quilts,Embroidered Bedsread,Bedding Collections,,,,
 Bedding Collections,Bedspreads,"Bed Skirts","Kids Bedding
 ","Shams",Sheets,
 Decorative Pillows,Quilts,,,,,
 Embroidered Bedsread,"Shams",,,,,
 Sheets,"Shams","Bed Skirts","Kids Bedding ",,,
 Decorative Pillows,Quilts,,,,,
 Decorative Pillows,"Kids Bedding
 ","Bed Skirts^[SEP]","Shams^[SEP]",,,
 Decorative Pillows,"Shams^[SEP]","Bed Skirts^[SEP]",,,
 Quilts,Sheets,"Kids Bedding
 ",,,,,
 Shams,"Bed Skirts^[SEP]","Kids Bedding
 ",Sheets,,
 Decorative Pillows,Bedspreads,"Shams^[SEP]",Sheets,"Bed Skirts^[SEP]","Kids Bedding
 ",

NIKE.CSV

Running Shoe,Socks,Sweatshirts,Modern Pants,,,,,
 Running Shoe,Socks,Sweatshirts,,,,,
 Running Shoe,Socks,Sweatshirts,Modern Pants,,,,,
 Running Shoe,Sweatshirts,Modern Pants,,,,,
 Running Shoe,Socks,Sweatshirts,Modern Pants,Soccer Shoe,,,,
 Running Shoe,Socks,Sweatshirts,,,,,

Running Shoe,Socks,Sweatshirts,Modern Pants,Tech Pants,Rash Guard,Hoodies,,

Swimming Shirt,Socks,Sweatshirts,,,,,,

Swimming Shirt,Rash Guard,Dry Fit V-Nick,Hoodies,Tech Pants,,,,,

Swimming Shirt,Rash Guard,Dry Fit V-Nick,,,,,,

Swimming Shirt,Rash Guard,Dry Fit V-Nick,,,,,,

Running Shoe,Swimming Shirt,Socks,Sweatshirts,Modern Pants,Soccer Shoe,Rash Guard,Hoodies,Tech Pants,Dry Fit V-Nick

Running Shoe,Swimming Shirt,Socks,Sweatshirts,Modern Pants,Soccer Shoe,Rash Guard,Tech Pants,Dry Fit V-Nick,Hoodies

Running Shoe,Swimming Shirt,Rash Guard,Tech Pants,Hoodies,Dry Fit V-Nick,,,

Running Shoe,Swimming Shirt,Socks,Sweatshirts,Modern Pants,Dry Fit V-Nick,Rash Guard,Tech Pants,,

Swimming Shirt,Soccer Shoe,Hoodies,Dry Fit V-Nick,Tech Pants,Rash Guard,,,

Running Shoe,Socks,,,,,,

"Socks,",Sweatshirts,Modern Pants,Soccer Shoe,Hoodies,Rash Guard,Tech Pants,Dry Fit V-Nick,,

Running Shoe,Swimming Shirt,Rash Guard,,,,,,

Running Shoe,Swimming Shirt,Socks,Sweatshirts,Modern Pants,Soccer Shoe,Hoodies,Tech Pants,Rash Guard,Dry Fit V-Nick

How to run the file: python project.py datasetname.csv minimum_support(in decimal) minimum_confidence(in decimal)

[illegible]

```

41 - def create_1itemset(data_set):
    """
    Create frequent candidate 1-itemset C1 by scanning data set.
    """
    d1 = {}
    data_set = list(data_set)
    Retrrs = {}
    C1 = {}
    for t in data_set:
        for item in t:
            if item not in C1:
                C1[item] = 1
    return C1

63 - def is_apriori(Ck_item, Lksub1):
    """
    Judge whether a frequent candidate k-itemset satisfy Apriori property.
    """
    Ck_item = list(Ck_item)
    Lksub1 = list(Lksub1)
    Retrrs = {}
    Inc = {}
    for item in Ck_item:
        sub_Ck = Ck_item - [item]
        if sub_Ck not in Lksub1:
            return False

81 - def create_Ck(Lksub1, k):
    """
    Create Ck, a set of frequent k-itemset, which contains all frequent candidate
    k-itemset.
    """
    Lksub1 = list(Lksub1)
    k = int(k)
    Retrrs = {}

```

```

C):set()
%c t) b1: lec(t?tbl)
L.s(ltschl:list(?cb1)
fof i&zacge(iec ttsLb1):

    l1 = list(list_Lksub1[i])
    l2 = list(list_Lksub1[j])
    l1.sort()
    l2.sort()
    if l1[0:k-2] == l2[0:k-2]:
        Ck_item = list_Lksub1[i] | list_Lksub1[j]
        # continue

```

```

def generate Lk by Ck(data set, Ck, min support, support data):

```

```

Lk = set()
item_count = {}
for t in data_set:
    for item in Ck:
        if item.issubse
            if item not
                item_co
            else:
                item_co
t_num = float(len(data_
for item in item_count:
    if (item_count[item
        Lk.add(item)
    support_data[it
return Lk

```



```
def generate_L(data_set, k, min_support):
```

```
    f.append(f 2sub1)
```

```
def generate_big_rules(L, support_data, min_conf):
```

```
    b.rule_list = []
    sub_set_list = []
    for i in range(1, len(L)):
        for free_set in L[i]:
            for sub_set in sub_set_list:
                if sub_set.is_subset(free_set):
                    conf = support_data[free_set] / support_data[free_set - sub_set]
                    big_rule = (free_set - sub_set, conf)
```

OUTPUT

```
Enter the minimum support: 0.2
Enter the minimum confidence: 0.4
=====
frequent 1-itemsets      support
=====
frozenset({'Swimming Shirt'}) 0.55
frozenset({'Dry Fit V-Nick'}) 0.5
frozenset({'Tech Pants'}) 0.45
frozenset({'Rash Guard'}) 0.6
frozenset({'Soccer Shoe'}) 0.3
frozenset({'Sweatshirts'}) 0.65
frozenset({'Modern Pants'}) 0.5
frozenset({'Socks'}) 0.6
frozenset({'Hoodies'}) 0.4
frozenset({'Running Shoe'}) 0.7
=====
frequent 2-itemsets      support
=====
frozenset({'Rash Guard', 'Hoodies'}) 0.4
frozenset({'Rash Guard', 'Socks'}) 0.25
frozenset({'Modern Pants', 'Dry Fit V-Nick'}) 0.25
frozenset({'Running Shoe', 'Hoodies'}) 0.25
frozenset({'Sweatshirts', 'Socks'}) 0.55
frozenset({'Modern Pants', 'Socks'}) 0.4
frozenset({'Sweatshirts', 'Hoodies'}) 0.25
frozenset({'Running Shoe', 'Soccer Shoe'}) 0.2
frozenset({'Modern Pants', 'Hoodies'}) 0.25
frozenset({'Sweatshirts', 'Running Shoe'}) 0.55
frozenset({'Dry Fit V-Nick', 'Socks'}) 0.2
frozenset({'Rash Guard', 'Modern Pants'}) 0.3
frozenset({'Rash Guard', 'Sweatshirts'}) 0.3
frozenset({'Sweatshirts', 'Dry Fit V-Nick'}) 0.25
frozenset({'Sweatshirts', 'Tech Pants'}) 0.3
frozenset({'Dry Fit V-Nick', 'Running Shoe'}) 0.25
frozenset({'Dry Fit V-Nick', 'Swimming Shirt'}) 0.45
frozenset({'Modern Pants', 'Running Shoe'}) 0.45
frozenset({'Modern Pants', 'Swimming Shirt'}) 0.2
frozenset({'Sweatshirts', 'Swimming Shirt'}) 0.25
frozenset({'Soccer Shoe', 'Hoodies'}) 0.25
frozenset({'Rash Guard', 'Running Shoe'}) 0.35
frozenset({'Rash Guard', 'Swimming Shirt'}) 0.5
frozenset({'Dry Fit V-Nick', 'Tech Pants'}) 0.4
frozenset({'Modern Pants', 'Tech Pants'}) 0.3
frozenset({'Running Shoe', 'Tech Pants'}) 0.3
frozenset({'Rash Guard', 'Tech Pants'}) 0.45
frozenset({'Rash Guard', 'Dry Fit V-Nick'}) 0.5
frozenset({'Socks', 'Soccer Shoe'}) 0.2
frozenset({'Running Shoe', 'Swimming Shirt'}) 0.3
frozenset({'Sweatshirts', 'Modern Pants'}) 0.5
frozenset({'Socks', 'Hoodies'}) 0.2
```

```

=====
frequent 3-itemsets      support
=====
frozenset({'Socks', 'Swimming Shirt', 'Tech Pants'}) 0.2
frozenset({'Rash Guard', 'Running Shoe', 'Modern Pants'}) 0.25
frozenset({'Rash Guard', 'Dry Fit V-Nick', 'Hoodies'}) 0.35
frozenset({'Sweatshirts', 'Running Shoe', 'Soccer Shoe'}) 0.2
frozenset({'Modern Pants', 'Soccer Shoe', 'Hoodies'}) 0.2
frozenset({'Rash Guard', 'Soccer Shoe', 'Hoodies'}) 0.25
frozenset({'Sweatshirts', 'Soccer Shoe', 'Tech Pants'}) 0.2
frozenset({'Modern Pants', 'Sweatshirts', 'Tech Pants'}) 0.3
frozenset({'Dry Fit V-Nick', 'Soccer Shoe', 'Tech Pants'}) 0.25
frozenset({'Dry Fit V-Nick', 'Soccer Shoe', 'Hoodies'}) 0.25
frozenset({'Soccer Shoe', 'Rash Guard', 'Modern Pants'}) 0.2
frozenset({'Sweatshirts', 'Soccer Shoe', 'Hoodies'}) 0.2
frozenset({'Modern Pants', 'Sweatshirts', 'Swimming Shirt'}) 0.2
frozenset({'Socks', 'Running Shoe', 'Hoodies'}) 0.2
frozenset({'Swimming Shirt', 'Soccer Shoe', 'Hoodies'}) 0.2
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Soccer Shoe'}) 0.2
frozenset({'Tech Pants', 'Soccer Shoe', 'Hoodies'}) 0.25
frozenset({'Modern Pants', 'Dry Fit V-Nick', 'Rash Guard'}) 0.25
frozenset({'Rash Guard', 'Sweatshirts', 'Soccer Shoe'}) 0.2
frozenset({'Rash Guard', 'Sweatshirts', 'Modern Pants'}) 0.3
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Socks'}) 0.2
frozenset({'Dry Fit V-Nick', 'Running Shoe', 'Tech Pants'}) 0.25
frozenset({'Sweatshirts', 'Socks', 'Hoodies'}) 0.2
frozenset({'Socks', 'Running Shoe', 'Soccer Shoe'}) 0.2
frozenset({'Modern Pants', 'Running Shoe', 'Socks'}) 0.4
frozenset({'Rash Guard', 'Swimming Shirt', 'Tech Pants'}) 0.35
frozenset({'Rash Guard', 'Modern Pants', 'Hoodies'}) 0.25
frozenset({'Modern Pants', 'Socks', 'Rash Guard'}) 0.25
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Rash Guard'}) 0.25
frozenset({'Modern Pants', 'Dry Fit V-Nick', 'Swimming Shirt'}) 0.2
frozenset({'Modern Pants', 'Running Shoe', 'Tech Pants'}) 0.25
frozenset({'Modern Pants', 'Dry Fit V-Nick', 'Socks'}) 0.2
frozenset({'Sweatshirts', 'Tech Pants', 'Hoodies'}) 0.25
frozenset({'Running Shoe', 'Dry Fit V-Nick', 'Socks'}) 0.2
frozenset({'Dry Fit V-Nick', 'Running Shoe', 'Hoodies'}) 0.2
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Tech Pants'}) 0.25
frozenset({'Rash Guard', 'Sweatshirts', 'Swimming Shirt'}) 0.2
frozenset({'Sweatshirts', 'Socks', 'Running Shoe'}) 0.5
frozenset({'Modern Pants', 'Socks', 'Swimming Shirt'}) 0.2
frozenset({'Rash Guard', 'Running Shoe', 'Hoodies'}) 0.25
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Swimming Shirt'}) 0.2
frozenset({'Modern Pants', 'Dry Fit V-Nick', 'Running Shoe'}) 0.2
frozenset({'Modern Pants', 'Tech Pants', 'Hoodies'}) 0.25
frozenset({'Rash Guard', 'Dry Fit V-Nick', 'Soccer Shoe'}) 0.25
frozenset({'Soccer Shoe', 'Swimming Shirt', 'Tech Pants'}) 0.2
frozenset({'Tech Pants', 'Running Shoe', 'Hoodies'}) 0.25
frozenset({'Socks', 'Running Shoe', 'Tech Pants'}) 0.25

```

```
frozenset({'Sweatshirts', 'Dry Fit V-Nick', 'Running Shoe'}) 0.2
frozenset({'Sweatshirts', 'Socks', 'Tech Pants'}) 0.25
Big Rules
frozenset({'Hoodies'}) => frozenset({'Rash Guard'}) conf: 1.0
frozenset({'Rash Guard'}) => frozenset({'Hoodies'}) conf: 0.6666666666666667
frozenset({'Socks'}) => frozenset({'Rash Guard'}) conf: 0.4166666666666667
frozenset({'Rash Guard'}) => frozenset({'Socks'}) conf: 0.4166666666666667
frozenset({'Modern Pants'}) => frozenset({'Dry Fit V-Nick'}) conf: 0.5
frozenset({'Dry Fit V-Nick'}) => frozenset({'Modern Pants'}) conf: 0.5
frozenset({'Hoodies'}) => frozenset({'Running Shoe'}) conf: 0.625
frozenset({'Socks'}) => frozenset({'Sweatshirts'}) conf: 0.9166666666666667
frozenset({'Sweatshirts'}) => frozenset({'Socks'}) conf: 0.8461538461538461
frozenset({'Socks'}) => frozenset({'Modern Pants'}) conf: 0.6666666666666667
frozenset({'Modern Pants'}) => frozenset({'Socks'}) conf: 0.8
frozenset({'Hoodies'}) => frozenset({'Sweatshirts'}) conf: 0.625
frozenset({'Soccer Shoe'}) => frozenset({'Running Shoe'}) conf: 0.6666666666666667
frozenset({'Hoodies'}) => frozenset({'Modern Pants'}) conf: 0.625
frozenset({'Modern Pants'}) => frozenset({'Hoodies'}) conf: 0.5
frozenset({'Running Shoe'}) => frozenset({'Sweatshirts'}) conf: 0.7857142857142858
frozenset({'Sweatshirts'}) => frozenset({'Running Shoe'}) conf: 0.8461538461538461
frozenset({'Dry Fit V-Nick'}) => frozenset({'Socks'}) conf: 0.4
frozenset({'Modern Pants'}) => frozenset({'Rash Guard'}) conf: 0.6
frozenset({'Rash Guard'}) => frozenset({'Modern Pants'}) conf: 0.5
frozenset({'Sweatshirts'}) => frozenset({'Rash Guard'}) conf: 0.4615384615384615
frozenset({'Rash Guard'}) => frozenset({'Sweatshirts'}) conf: 0.5
frozenset({'Dry Fit V-Nick'}) => frozenset({'Sweatshirts'}) conf: 0.5
frozenset({'Sweatshirts'}) => frozenset({'Tech Pants'}) conf: 0.4615384615384615
frozenset({'Tech Pants'}) => frozenset({'Sweatshirts'}) conf: 0.6666666666666666
frozenset({'Dry Fit V-Nick'}) => frozenset({'Running Shoe'}) conf: 0.5
frozenset({'Dry Fit V-Nick'}) => frozenset({'Swimming Shirt'}) conf: 0.9
frozenset({'Swimming Shirt'}) => frozenset({'Dry Fit V-Nick'}) conf: 0.8181818181818181
frozenset({'Running Shoe'}) => frozenset({'Modern Pants'}) conf: 0.6428571428571429
frozenset({'Modern Pants'}) => frozenset({'Running Shoe'}) conf: 0.9
frozenset({'Modern Pants'}) => frozenset({'Swimming Shirt'}) conf: 0.4
frozenset({'Swimming Shirt'}) => frozenset({'Sweatshirts'}) conf: 0.4545454545454545
frozenset({'Hoodies'}) => frozenset({'Soccer Shoe'}) conf: 0.625
frozenset({'Soccer Shoe'}) => frozenset({'Hoodies'}) conf: 0.8333333333333334
frozenset({'Running Shoe'}) => frozenset({'Rash Guard'}) conf: 0.5
frozenset({'Rash Guard'}) => frozenset({'Running Shoe'}) conf: 0.5833333333333334
frozenset({'Rash Guard'}) => frozenset({'Swimming Shirt'}) conf: 0.8333333333333334
frozenset({'Swimming Shirt'}) => frozenset({'Rash Guard'}) conf: 0.9090909090909091
frozenset({'Tech Pants'}) => frozenset({'Dry Fit V-Nick'}) conf: 0.8888888888888889
frozenset({'Dry Fit V-Nick'}) => frozenset({'Tech Pants'}) conf: 0.8
frozenset({'Modern Pants'}) => frozenset({'Tech Pants'}) conf: 0.6
frozenset({'Tech Pants'}) => frozenset({'Modern Pants'}) conf: 0.6666666666666666
frozenset({'Running Shoe'}) => frozenset({'Tech Pants'}) conf: 0.4285714285714286
frozenset({'Tech Pants'}) => frozenset({'Running Shoe'}) conf: 0.6666666666666666
frozenset({'Rash Guard'}) => frozenset({'Tech Pants'}) conf: 0.75
frozenset({'Tech Pants'}) => frozenset({'Rash Guard'}) conf: 1.0
frozenset({'Rash Guard'}) => frozenset({'Dry Fit V-Nick'}) conf: 0.8333333333333334
```