РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ

Факультет физико-математических и естественных наук

Кафедра информационных технологий

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

Дисциплина: Интеллектуальный анализ данных

Студент: Гусейнов Вахид Азерович

Группа: НБИбд-01-17

Москва 2020

Вариант № 14

Алгоритм: Eclat

День недели (поле order_dow таблицы orders): “4”

Код департамента (поле department_id таблицы products): “5”

1. При помощи модуля sqlite3 откройте базу данных Instacart в файле instacart.db.

In [1]:
```
import sqlite3

conn = sqlite3.connect('instacart.db')
```

1. При помощи запроса SELECT извлеките из таблицы order_products**train записи, соответствующие указанным в индивидуальном задании дню недели (поле order_dow таблицы orders) и коду департамента (поле department_id таблицы products). Определите количество записей в полученном наборе и определите количество товаров (поле order_id таблицы order_products**train) в транзакциях набора.

In [2]:
```
import pandas as pd

query =\
'''
SELECT DISTINCT train.*, products.product_name
FROM order_products__train as train
JOIN orders
    ON train.order_id = orders.order_id
JOIN products
    ON train.product_id = products.product_id
WHERE orders.order_dow = 4
AND   products.department_id = 5
'''
data = pd.read_sql(query, conn)
```

In [3]:
```
data.head()
```

Out[3]:

	order_id	product_id	add_to_cart_order	reordered	product_name
0	877974	1808	1	0	Champagne
1	1859940	15511	1	1	Draft Sake
2	3409264	2120	3	1	Sauvignon Blanc
3	1881604	10607	6	1	Prosecco
4	1881604	29509	2	1	80 Vodka Holiday Edition

1. Определите количество покупок (транзакций) для пяти наиболее популярных товаров в наборе.

In [4]:
```
popular = data.groupby('product_name')['order_id'].count().nlargest(5)
popular
```

Out[4]:
```
product_name
Beer                  45
Cabernet Sauvignon    43
Sauvignon Blanc       38
Chardonnay            36
India Pale Ale        31
Name: order_id, dtype: int64
```

1. Постройте транзакционную базу данных для поиск ассоциативных правил из полученного набора записей таблицы order_products__train, используя в качестве идентификатора транзакции поле order_id, а в качестве названий товаров - поле product_name из таблицы products, соответствующее полю product_id.

In [5]:
```
transactions = []
for order in data['order_id'].unique():
    products = set(data[data['order_id'] == order]['product_name'])
    pair = list((order, products))
    transactions.append(pair)
```

In [6]:
```
for i in range(3):
    print(transactions[i])
```

```
['877974', {'Champagne'}]
['1859940', {'Draft Sake'}]
['3409264', {'Sauvignon Blanc'}]
```

1. Реализуйте указанный в индивидуальном задании метод построения популярных наборов предметов (Apriori/Eclat/Declat) (3 балла) или используйте метод BruteForce (0 баллов). Протестируйте корректность

In [7]:
```
def prep(database):
    all_items = set()
    for pair in database:
        all_items = all_items.union(pair[1])
    all_items

    res = dict()
    for item in all_items:
        res[item] = set()
        for pair in database:
            if item in pair[1]:
                res[item].add(pair[0])
```

```
    res = list(res.items())
    return res

def eclat(prefix, items):
    while items:
        i, itids = items.pop()
        isupp = len(itids)
        if isupp >= minsup:
            items_freq[frozenset(prefix + [i])] = isupp
            suffix = []
            for j, ojtids in items:
                jtids = itids&ojtids
                if len(jtids)>=minsup:
                    suffix.append((j,jtids))
            eclat(prefix+[i], sorted(suffix, key=lambda item: len(item[1]), reverse=
```

Проверка

In [8]:
```
D_train = [
    [ 1, {"A","B","D","E"} ],
    [ 2, {"B","C","E"} ],
    [ 3, {"A","B","D","E"} ],
    [ 4, {"A","B","C","E"} ],
    [ 5, {"A","B","C","D","E"} ],
    [ 6, {"B","C","D"} ],
]

minsup = 3
items_freq = dict()
eclat([], prep(D_train))

res_lst = []
for key, value in items_freq.items():
    tmp = []
    tmp.append(tuple(key))
    tmp.append(value)
    res_lst.append(tmp)
res_lst = sorted(res_lst, key=lambda x: len(x[0]))

for result in res_lst:
    print(result)
```

```
[('E',), 5]
[('D',), 4]
[('B',), 6]
[('A',), 4]
[('C',), 4]
[('D', 'E'), 3]
[('C', 'E'), 3]
[('A', 'E'), 4]
[('B', 'E'), 5]
[('A', 'D'), 3]
[('D', 'B'), 4]
[('A', 'B'), 4]
[('C', 'B'), 4]
[('A', 'D', 'E'), 3]
[('B', 'D', 'E'), 3]
[('B', 'C', 'E'), 3]
[('B', 'A', 'E'), 4]
[('A', 'D', 'B'), 3]
[('B', 'A', 'D', 'E'), 3]
```

1. При помощи указанного в индивидуальном задании метода или метода BruteForce постройте популярные наборы товаров с минимальной поддержкой, равной половине среднего количества покупок пяти наиболее популярных товаров. В случае

нехватки вычислительных ресурсов (слишком долгой работы программы) при построении популярных наборов товаров оставьте в наборе данных транзакции с 10 наиболее популярными товарами и сокращайте число записей (например, методом деления пополам).

In [9]:
```
minsup = round(popular.mean() / 2)
minsup
```

Out[9]: 19

Возьмем minsup = 2 иначе не получим никаких ассоциативных правил

In [10]:
```
minsup = 2
items_freq = dict()
eclat([], prep(transactions))

res_lst = []
for key, value in items_freq.items():
    tmp = []
    tmp.append(tuple(key))
    tmp.append(value)
    res_lst.append(tmp)
res_lst = sorted(res_lst, key=lambda x: len(x[0]))

print('Полученные популярные наборы:\n')
for result in res_lst:
    print(result)
```

```
Полученные популярные наборы:

[('Sauvignon Blanc Wine',), 3]
[('India Pale Ale Racer 5',), 4]
[('Sonoma Pinot Noir Wine',), 2]
[('Sauvignon Blanc',), 38]
[('Old Rasputin Stout Beer',), 3]
[('12 Oz Lager',), 3]
[('Reserve Shiraz Wine',), 2]
[('Draft Sake',), 2]
[('Verry Special Cognac',), 2]
[("Vintner's Reserve Chardonnay",), 2]
[('Champagne',), 2]
[('Original Tequila Reposado',), 2]
[('Lager',), 5]
[('Bitters Liqueur',), 3]
[('Original Spiced Rum',), 2]
[('Brut Champagne',), 2]
[('Vodka',), 27]
[('Chardonnay',), 36]
[('India Pale Ale',), 31]
[('Pinot Noir, California',), 2]
[('Prosecco Sparkling Wine',), 18]
[('Sparking Brut Classic Wine',), 2]
[('Brut Rosé',), 2]
[('Diamond Collection Silver Label Pinot Noir Wine',), 2]
[('Distilled London Dry Gin',), 3]
[('Beer',), 45]
[('Potato Vodka',), 3]
[('Brut Sparkling Wine',), 6]
[('Draft Beer',), 3]
[('Frontier Straight 95% Rye Mash Whiskey',), 3]
[('Liqueur',), 2]
[('Claret, Black Label, 2006',), 2]
[('Prosecco',), 8]
[('Red Wine, Dark, California, 2013',), 2]
[('Fume Blanc Sonoma County',), 2]
[('Extra Dry California Champagne',), 2]
```

[('Fresh Squeezed IPA',), 3]
[('Chenin Blanc',), 3]
[('London Dry Gin England',), 2]
[('Brown Ale',), 5]
[('Sauvignon Blanc, Napa County',), 2]
[('Belgium Beer',), 9]
[('Amber Ale',), 9]
[('Old Vine Zinfandel',), 4]
[('Clara',), 11]
[('Fat Tire Amber Ale - 12 CT',), 2]
[('Pinot Grigio Wine',), 3]
[('Pale Ale',), 2]
[('Draught',), 3]
[('Scrimshaw Pilsner Style Beer',), 4]
[('Whiskey',), 2]
[('Alto Adige Pinot Grigio',), 3]
[('Lager Beer',), 4]
[('Cabernet Sauvignon, Central Coast, 2011',), 4]
[('Brut',), 6]
[('Variety Pack Hard Cider',), 8]
[('California Chardonnay',), 2]
[('Icelandic White Ale 6 Pack',), 3]
[('California Red Wine',), 7]
[('Handmade Vodka From Austin, Texas',), 3]
[('Coastal Pinot Noir',), 2]
[('805 Blond Ale',), 2]
[('Brut California Champagne',), 3]
[('Moscato',), 2]
[('Especial',), 3]
[("Mixed 12 Pack Lion's Share Ale",), 5]
[('Private Selection Pinot Noir',), 2]
[('312 Urban Wheat Ale',), 3]
[('IPA',), 6]
[("Little Sumpin' Sumpin' Ale",), 10]
[('Cabernet Sauvignon Wine',), 6]
[('Mighty Dry Hard Cider',), 3]
[('Extra Stout Beer',), 3]
[('Tequila, Silver',), 4]
[('Monterey County Pinot Noir',), 2]
[('Silver Tequila',), 4]
[('Merlot',), 7]
[('Handmade Vodka',), 10]
[('Triple Distilled Vodka',), 8]
[('Nuestro Tequila',), 2]
[('Robert Mondavi',), 3]
[('Bottled Imported Premium Light Lager Beer',), 2]
[('Pale Ale Beer',), 2]
[('Malbec',), 7]
[('Pinot Noir Wine',), 10]
[('Brut Cuvee Sparkling Champagne',), 3]
[('Light',), 3]
[('The Original Irish Creme Liquer',), 7]
[('Red Blend',), 5]
[('Ale, Amber',), 3]
[('90 Minute Imperial Ipa',), 3]
[('Belgian Style Wheat Ale',), 5]
[('Premium Belgian Lager',), 7]
[('Cabernet Sauvignon, Central Coast, 2007',), 2]
[('Down To Earth IPA',), 2]
[("Brewmaster's Seasonal Sampler",), 3]
[('London Dry Gin',), 3]
[('Blonde Ale',), 2]
[('Ale, India Pale, Brew Free! Or Die IPA',), 5]
[('Tennessee Whiskey',), 2]
[('California Merlot',), 2]
[('Orange Liqueur',), 2]
[('Tequila Reposado with Glass',), 2]
[('Riverstone Chardonnay',), 2]
[('Brandy',), 2]

```
[('Frontier Whiskey',), 6]
[('Bourbon Kentucky Frontier Whiskey',), 3]
[('Ranger India Pale Ale Bottles',), 2]
[('Gruner Veltliner',), 2]
[('Blended Scotch Whiskey Black Label',), 2]
[('Extra Beer Bottles',), 9]
[('Belgian White Beer',), 3]
[('Valdobbiadene Prosecco',), 2]
[('Premium Lager Beer',), 16]
[('Variety Pack Beer',), 2]
[('Light Beer Cans',), 2]
[('California Pinot Noir',), 3]
[('Merlot, Columbia Valley, 2007',), 2]
[('Zinfandel',), 3]
[('Coffee Liqueur',), 2]
[('Cabernet Sauvignon',), 43]
[('Summer Pack Variety Bottles',), 4]
[('Kentucky Straight Bourbon Whiskey',), 4]
[('Original Belgian Wheat Beer',), 4]
[('Crisp Hard Cider Crisp Apple',), 8]
[('Vintners Reserve Sauvignon Blanc',), 2]
[('Sauvignon Blanc, Mendocino, 2007',), 2]
[('Tequila Silver',), 5]
[('Villager Ipa',), 3]
[('Rose',), 6]
[('Pinot Noir',), 22]
[('Ksa Ko?Lsch Style Ale',), 3]
[('Coronita Beer Extra',), 3]
[('Chardonnay Wine',), 12]
[('Cabernet Sauvignon Sonoma County',), 3]
[('Pils',), 3]
[('Pinot Grigio',), 12]
[('312 Urban Wheat',), 2]
[('Irish Whiskey',), 3]
[('Smokestack Mixed Four',), 2]
[('Longboard Island Lager',), 6]
[('Belgian White Wheat Ale',), 11]
[('80  Vodka Holiday Edition',), 7]
[('Sauvignon Blanc', 'India Pale Ale'), 2]
[('Sauvignon Blanc', 'Beer'), 2]
[('Sauvignon Blanc', 'Fresh Squeezed IPA'), 2]
[('Sauvignon Blanc', 'Chenin Blanc'), 2]
[('Sauvignon Blanc', 'Pinot Noir Wine'), 2]
[('Sauvignon Blanc', 'Pinot Noir'), 2]
[('Sauvignon Blanc', 'Chardonnay Wine'), 2]
[('Sauvignon Blanc', 'Chardonnay'), 3]
[('Sauvignon Blanc', 'Prosecco Sparkling Wine'), 3]
[('Sauvignon Blanc', 'California Red Wine'), 3]
[('Sauvignon Blanc', 'Cabernet Sauvignon'), 7]
[('Beer', 'Lager'), 2]
[('Lager', 'Belgian White Beer'), 2]
[('Bitters Liqueur', 'Ale, India Pale, Brew Free! Or Die IPA'), 2]
[('Vodka', 'Chardonnay'), 2]
[('Vodka', 'Silver Tequila'), 2]
[('Cabernet Sauvignon', 'Vodka'), 2]
[('Vodka', 'Chardonnay Wine'), 2]
[('Pinot Grigio', 'Vodka'), 2]
[('Chardonnay', 'India Pale Ale'), 2]
[('Chardonnay', 'Prosecco Sparkling Wine'), 2]
[('Brown Ale', 'Chardonnay'), 2]
[('Malbec', 'Chardonnay'), 2]
[('Premium Lager Beer', 'Chardonnay'), 2]
[('Beer', 'Chardonnay'), 3]
[('Chardonnay', 'Tequila Silver'), 3]
[('Cabernet Sauvignon', 'Chardonnay'), 8]
[('India Pale Ale', 'Lager Beer'), 2]
[("Mixed 12 Pack Lion's Share Ale", 'India Pale Ale'), 2]
[('India Pale Ale', "Brewmaster's Seasonal Sampler"), 2]
[('Belgian White Beer', 'India Pale Ale'), 2]
```

```
[('India Pale Ale', 'Longboard Island Lager'), 2]
[('Belgian White Wheat Ale', 'India Pale Ale'), 2]
[('India Pale Ale', 'Prosecco Sparkling Wine'), 3]
[('Clara', 'India Pale Ale'), 3]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale'), 3]
[('Pinot Noir Wine', 'India Pale Ale'), 3]
[('Premium Belgian Lager', 'India Pale Ale'), 3]
[('Frontier Whiskey', 'India Pale Ale'), 3]
[('Cabernet Sauvignon', 'India Pale Ale'), 3]
[('Amber Ale', 'India Pale Ale'), 4]
[('India Pale Ale', 'Extra Beer Bottles'), 4]
[('India Pale Ale', 'Variety Pack Hard Cider'), 5]
[('Premium Lager Beer', 'India Pale Ale'), 6]
[('Beer', 'India Pale Ale'), 10]
[('Pinot Noir Wine', 'Prosecco Sparkling Wine'), 2]
[('Pinot Noir', 'Prosecco Sparkling Wine'), 2]
[('Cabernet Sauvignon', 'Prosecco Sparkling Wine'), 3]
[('Beer', 'Variety Pack Hard Cider'), 2]
[('Beer', "Little Sumpin' Sumpin' Ale"), 2]
[('Beer', 'Ale, Amber'), 2]
[('Beer', 'Premium Belgian Lager'), 2]
[('Beer', 'Belgian White Beer'), 2]
[('Beer', 'Belgium Beer'), 3]
[('Beer', 'Clara'), 3]
[('Beer', "Mixed 12 Pack Lion's Share Ale"), 3]
[('Beer', 'Premium Lager Beer'), 3]
[('Beer', 'Cabernet Sauvignon'), 3]
[('Beer', 'Crisp Hard Cider Crisp Apple'), 3]
[('Beer', 'Amber Ale'), 4]
[('Beer', 'Belgian White Wheat Ale'), 5]
[('Cabernet Sauvignon', 'Red Wine, Dark, California, 2013'), 2]
[('Premium Lager Beer', 'Amber Ale'), 2]
[('Clara', 'Premium Belgian Lager'), 2]
[('Cabernet Sauvignon', 'Pinot Grigio Wine'), 2]
[('Scrimshaw Pilsner Style Beer', "Little Sumpin' Sumpin' Ale"), 2]
[('Scrimshaw Pilsner Style Beer', 'Crisp Hard Cider Crisp Apple'), 2]
[('Scrimshaw Pilsner Style Beer', 'Belgian White Wheat Ale'), 2]
[("Little Sumpin' Sumpin' Ale", 'Variety Pack Hard Cider'), 2]
[('Pinot Noir Wine', 'Variety Pack Hard Cider'), 2]
[('Premium Belgian Lager', 'Variety Pack Hard Cider'), 2]
[("Brewmaster's Seasonal Sampler", 'Variety Pack Hard Cider'), 2]
[('Cabernet Sauvignon', 'Variety Pack Hard Cider'), 3]
[('California Red Wine', 'Cabernet Sauvignon Wine'), 2]
[('California Red Wine', 'Cabernet Sauvignon'), 2]
[('Crisp Hard Cider Crisp Apple', 'IPA'), 3]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale"), 2]
[('Premium Belgian Lager', "Little Sumpin' Sumpin' Ale"), 2]
[('Premium Lager Beer', "Little Sumpin' Sumpin' Ale"), 2]
[('Crisp Hard Cider Crisp Apple', "Little Sumpin' Sumpin' Ale"), 2]
[('Belgian White Wheat Ale', "Little Sumpin' Sumpin' Ale"), 2]
[("Little Sumpin' Sumpin' Ale", 'Longboard Island Lager'), 3]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale"), 4]
[('Merlot', 'Cabernet Sauvignon'), 4]
[('Cabernet Sauvignon', 'Handmade Vodka'), 2]
[('Pinot Noir Wine', 'Premium Belgian Lager'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon'), 3]
[('Belgian Style Wheat Ale', 'Extra Beer Bottles'), 2]
[('312 Urban Wheat', 'Belgian Style Wheat Ale'), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager'), 2]
[('Premium Belgian Lager', 'Belgian White Wheat Ale'), 2]
[('Cabernet Sauvignon', "Brewmaster's Seasonal Sampler"), 2]
[('Longboard Island Lager', "Brewmaster's Seasonal Sampler"), 2]
[('Belgian White Beer', 'Extra Beer Bottles'), 2]
[('Premium Lager Beer', 'Crisp Hard Cider Crisp Apple'), 2]
[('Cabernet Sauvignon', 'Chardonnay Wine'), 2]
[('Cabernet Sauvignon', 'Longboard Island Lager'), 2]
[('Cabernet Sauvignon', 'Cabernet Sauvignon Sonoma County'), 3]
[('Cabernet Sauvignon', 'Belgian White Wheat Ale'), 3]
[('Cabernet Sauvignon', 'Pinot Noir'), 6]
```

```
[('Pinot Noir', 'Chardonnay Wine'), 3]
[('Belgian White Wheat Ale', 'Pils'), 2]
[('Sauvignon Blanc', 'Cabernet Sauvignon', 'California Red Wine'), 2]
[('Beer', 'Cabernet Sauvignon', 'Chardonnay'), 2]
[('Beer', "Mixed 12 Pack Lion's Share Ale", 'India Pale Ale'), 2]
[('Longboard Island Lager', 'India Pale Ale', "Brewmaster's Seasonal Sampler"), 2]
[('Beer', 'Belgian White Beer', 'India Pale Ale'), 2]
[('Beer', 'Belgian White Wheat Ale', 'India Pale Ale'), 2]
[('Clara', 'Premium Belgian Lager', 'India Pale Ale'), 2]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale", 'India Pale Ale'), 2]
[("Little Sumpin' Sumpin' Ale", 'Premium Belgian Lager', 'India Pale Ale'), 2]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'India Pale Ale'), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Variety Pack Hard Cider'), 2]
[('Beer', "Little Sumpin' Sumpin' Ale", 'India Pale Ale'), 2]
[('Pinot Noir Wine', 'Premium Belgian Lager', 'India Pale Ale'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'India Pale Ale'), 2]
[('Pinot Noir Wine', 'India Pale Ale', 'Variety Pack Hard Cider'), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager', 'India Pale Ale'), 2]
[('Premium Belgian Lager', 'India Pale Ale', 'Variety Pack Hard Cider'), 2]
[('Cabernet Sauvignon', 'India Pale Ale', 'Variety Pack Hard Cider'), 2]
[('Beer', 'Amber Ale', 'India Pale Ale'), 2]
[('Beer', 'India Pale Ale', 'Variety Pack Hard Cider'), 2]
[('Beer', 'Premium Lager Beer', 'India Pale Ale'), 2]
[('Beer', 'Premium Belgian Lager', 'Belgian White Wheat Ale'), 2]
[('Scrimshaw Pilsner Style Beer', 'Crisp Hard Cider Crisp Apple', "Little Sumpin' Su
mpin' Ale"), 2]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale", 'Variety Pack Hard Cider'), 2]
[('Premium Belgian Lager', "Little Sumpin' Sumpin' Ale", 'Variety Pack Hard Cider'),
2]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'Variety Pack Hard Cider'), 2]
[('Pinot Noir Wine', 'Premium Belgian Lager', 'Variety Pack Hard Cider'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'Variety Pack Hard Cider'), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager', 'Variety Pack Hard Cider'), 2]
[('Cabernet Sauvignon', "Brewmaster's Seasonal Sampler", 'Variety Pack Hard Cider'),
2]
[('Pinot Noir Wine', 'Premium Belgian Lager', "Little Sumpin' Sumpin' Ale"), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale"), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager', "Little Sumpin' Sumpin' Ale"), 2]
[('Cabernet Sauvignon', 'Belgian White Wheat Ale', "Little Sumpin' Sumpin' Ale"), 2]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'Longboard Island Lager'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'Premium Belgian Lager'), 2]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale", 'Premium Belgian Lager', 'India P
ale Ale'), 2]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale", 'Cabernet Sauvignon', 'India Pale
Ale'), 2]
[('Pinot Noir Wine', "Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Variety Pack H
ard Cider'), 2]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'Premium Belgian Lager', 'Indi
a Pale Ale'), 2]
[("Little Sumpin' Sumpin' Ale", 'Premium Belgian Lager', 'India Pale Ale', 'Variety
Pack Hard Cider'), 2]
[('Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Variety Pac
k Hard Cider'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'Premium Belgian Lager', 'India Pale Al
e'), 2]
[('Pinot Noir Wine', 'Premium Belgian Lager', 'India Pale Ale', 'Variety Pack Hard C
ider'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'India Pale Ale', 'Variety Pack Hard Cide
r'), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager', 'India Pale Ale', 'Variety Pack Har
d Cider'), 2]
[('Pinot Noir Wine', 'Premium Belgian Lager', "Little Sumpin' Sumpin' Ale", 'Variety
Pack Hard Cider'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', "Little Sumpin' Sumpin' Ale", 'Variety Pa
ck Hard Cider'), 2]
[('Cabernet Sauvignon', 'Premium Belgian Lager', "Little Sumpin' Sumpin' Ale", 'Vari
ety Pack Hard Cider'), 2]
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'Premium Belgian Lager', 'Variety Pack Ha
rd Cider'), 2]
```

```
[('Pinot Noir Wine', 'Cabernet Sauvignon', 'Premium Belgian Lager', "Little Sumpin'
Sumpin' Ale"), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Premium Belgian Lager', 'Pinot No
ir Wine', 'Cabernet Sauvignon'), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Premium Belgian Lager', 'Variety
Pack Hard Cider', 'Pinot Noir Wine'), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Variety Pack Hard Cider', 'Pinot
Noir Wine', 'Cabernet Sauvignon'), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Premium Belgian Lager', 'Variety
Pack Hard Cider', 'Cabernet Sauvignon'), 2]
[('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir
Wine', 'Cabernet Sauvignon'), 2]
[("Little Sumpin' Sumpin' Ale", 'Premium Belgian Lager', 'Variety Pack Hard Cider',
'Pinot Noir Wine', 'Cabernet Sauvignon'), 2]
[("Little Sumpin' Sumpin' Ale", 'India Pale Ale', 'Premium Belgian Lager', 'Variety
Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sauvignon'), 2]
```

1. Для какого-либо из полученных популярных наборов товаров постройте набор ассоциативных правил.

In [11]:
```
sample, _ = res_lst[-3]
print(sample, len(sample))
```

```
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir W
ine', 'Cabernet Sauvignon') 5
```

In [12]:
```
from itertools import chain, combinations

def ComputeSupport(X, D):
    supX = 0
    for _, itemset in D:
        if X.issubset(itemset):
            supX += 1
    return supX

def powersetk(iterable, k):
    xs = list(iterable)
    return list(chain.from_iterable(combinations(xs, n) for n in range(k, len(xs) +

def AssociationRules(D, Z_set, minconf):
    A_rules = []
    supZ = ComputeSupport(set(Z_set), D)
    A_set = powersetk(Z_set, 1)
    while len(A_set)>0:
        X_set = A_set[-1]
        A_set.pop()
        conf = supZ / ComputeSupport(set(X_set), D)
        if conf >= minconf:
            Y_set = sorted(list(set(Z_set) - set(X_set)))
            A_rules.append([X_set, Y_set])
        else:
            for W_set in powersetk(X_set, 1):
                if W_set in A_set:
                    A_set.remove(W_set)
    return A_rules

rules = AssociationRules(transactions, sample, 0.9)
print('Полученные ассоциативные правила:\n')
for rule in rules:
    print('rule:')
    for x in rule:
        print(x)
    print()
```

```
Полученные ассоциативные правила:
```

```
rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sa
uvignon')
['India Pale Ale']

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sauvigno
n')
['Premium Belgian Lager']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['Variety Pack Hard Cider']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Cabernet Sau
vignon')
['Pinot Noir Wine']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir W
ine')
['Cabernet Sauvignon']

rule:
('Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['India Pale Ale', 'Premium Belgian Lager']

rule:
('Premium Belgian Lager', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['India Pale Ale', 'Variety Pack Hard Cider']

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Cabernet Sauvignon')
['India Pale Ale', 'Pinot Noir Wine']

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale']

rule:
('India Pale Ale', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['Premium Belgian Lager', 'Variety Pack Hard Cider']

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Cabernet Sauvignon')
['Pinot Noir Wine', 'Premium Belgian Lager']

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'Premium Belgian Lager']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Cabernet Sauvignon')
['Pinot Noir Wine', 'Variety Pack Hard Cider']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'Variety Pack Hard Cider']

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider')
['Cabernet Sauvignon', 'Pinot Noir Wine']

rule:
('Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale', 'Premium Belgian Lager']

rule:
```

```
('Premium Belgian Lager', 'Cabernet Sauvignon')
['India Pale Ale', 'Pinot Noir Wine', 'Variety Pack Hard Cider']

rule:
('Premium Belgian Lager', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale', 'Variety Pack Hard Cider']

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider')
['Cabernet Sauvignon', 'India Pale Ale', 'Pinot Noir Wine']
```

1. Для построенного набора ассоциативных правил вычислите показатели: support, confidence, lift, leverage, conviction и выведите на экран.

In [13]:
```python
import numpy as np

N = len(transactions)
supps = []
confs = []
lifts = []
leverages = []
convictions = []
for rule in rules:
    x1 = set(rule[0])
    x2 = set(rule[1])
    supp_x1_x2 = ComputeSupport(x1.union(x2), transactions)
    supp_x1 = ComputeSupport(x1, transactions)
    supp_x2 = ComputeSupport(x2, transactions)
    conf = supp_x1_x2 / supp_x1 # confidence
    lift = supp_x1_x2 / (supp_x1 * supp_x2) # lift
    leverage = supp_x1_x2 - supp_x1 * supp_x2 # leverage
    if conf == 1:
        conv = np.inf
    else:
        conv = (1 - supp_x2) / (1 - conf) # conviction
    supps.append(supp_x1_x2)
    confs.append(conf)
    lifts.append(lift)
    leverages.append(leverage)
    convictions.append(conv)
```

In [14]:
```python
for i in range(len(rules)):
    print('rule:')
    for x in rules[i]:
        print(x)
    print('Support =', supps[i])
    print('Confidence =', confs[i])
    print('Lift =', lifts[i])
    print('Leverage =', leverages[i])
    print('Conviction =', convictions[i])
    print()
```

```
rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sa
uvignon')
['India Pale Ale']
Support = 2
Confidence = 1.0
Lift = 0.03225806451612903
Leverage = -60
Conviction = inf

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sauvigno
```

```
n')
['Premium Belgian Lager']
Support = 2
Confidence = 1.0
Lift = 0.14285714285714285
Leverage = -12
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.125
Leverage = -14
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Cabernet Sau
vignon')
['Pinot Noir Wine']
Support = 2
Confidence = 1.0
Lift = 0.1
Leverage = -18
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir W
ine')
['Cabernet Sauvignon']
Support = 2
Confidence = 1.0
Lift = 0.023255813953488372
Leverage = -84
Conviction = inf

rule:
('Variety Pack Hard Cider', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['India Pale Ale', 'Premium Belgian Lager']
Support = 2
Confidence = 1.0
Lift = 0.3333333333333333
Leverage = -4
Conviction = inf

rule:
('Premium Belgian Lager', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['India Pale Ale', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.2
Leverage = -8
Conviction = inf

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Cabernet Sauvignon')
['India Pale Ale', 'Pinot Noir Wine']
Support = 2
Confidence = 1.0
Lift = 0.3333333333333333
Leverage = -4
Conviction = inf

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale']
Support = 2
Confidence = 1.0
```

```
Lift = 0.3333333333333333
Leverage = -4
Conviction = inf

rule:
('India Pale Ale', 'Pinot Noir Wine', 'Cabernet Sauvignon')
['Premium Belgian Lager', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Cabernet Sauvignon')
['Pinot Noir Wine', 'Premium Belgian Lager']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('India Pale Ale', 'Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'Premium Belgian Lager']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Cabernet Sauvignon')
['Pinot Noir Wine', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.3333333333333333
Leverage = -4
Conviction = inf

rule:
('India Pale Ale', 'Premium Belgian Lager', 'Variety Pack Hard Cider')
['Cabernet Sauvignon', 'Pinot Noir Wine']
Support = 2
Confidence = 1.0
Lift = 0.3333333333333333
Leverage = -4
Conviction = inf

rule:
('Variety Pack Hard Cider', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale', 'Premium Belgian Lager']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('Premium Belgian Lager', 'Cabernet Sauvignon')
```

```
['India Pale Ale', 'Pinot Noir Wine', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('Premium Belgian Lager', 'Pinot Noir Wine')
['Cabernet Sauvignon', 'India Pale Ale', 'Variety Pack Hard Cider']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf

rule:
('Premium Belgian Lager', 'Variety Pack Hard Cider')
['Cabernet Sauvignon', 'India Pale Ale', 'Pinot Noir Wine']
Support = 2
Confidence = 1.0
Lift = 0.5
Leverage = -2
Conviction = inf
```

In [ ]: