CO1	**Examine the basic concepts of data mining and machine learning concepts**
Task 1:	Implement Apriori and FP-growth algorithm to find all frequent item sets for the chosen dataset and also generate Association Rules. **Platform: Rapidminer, Language: Python**

Tool: Rapidminer

<u>Apriori Algorithm:</u>

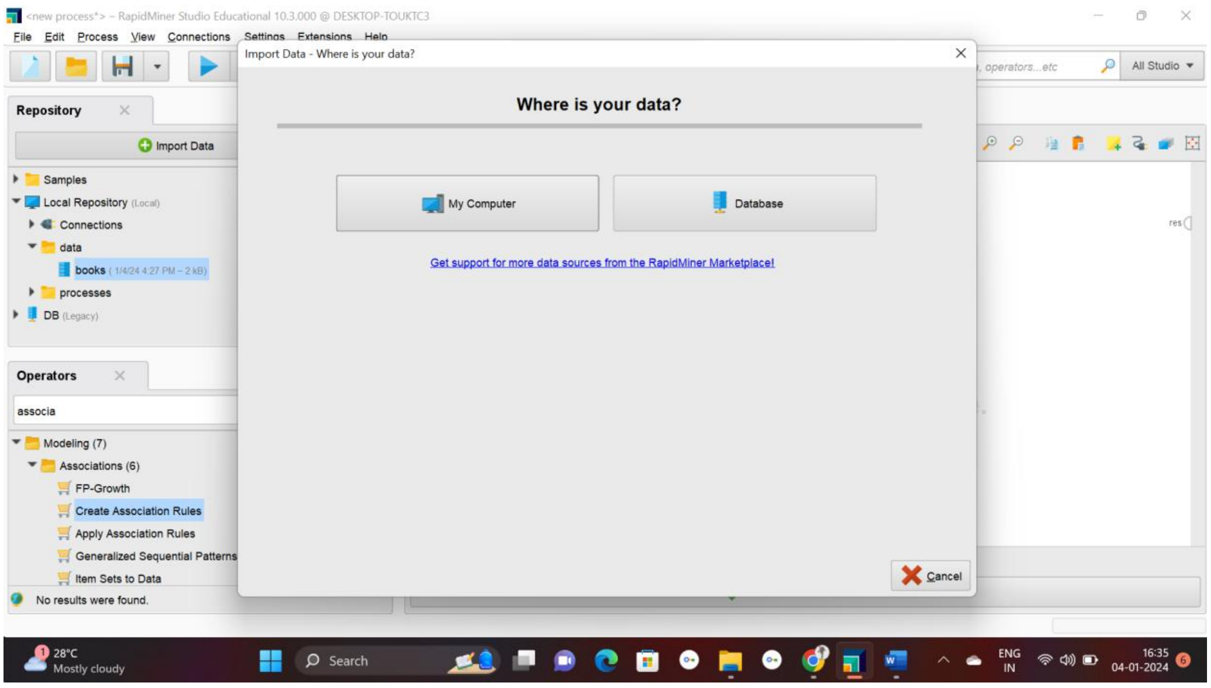
Use Case: Books Data

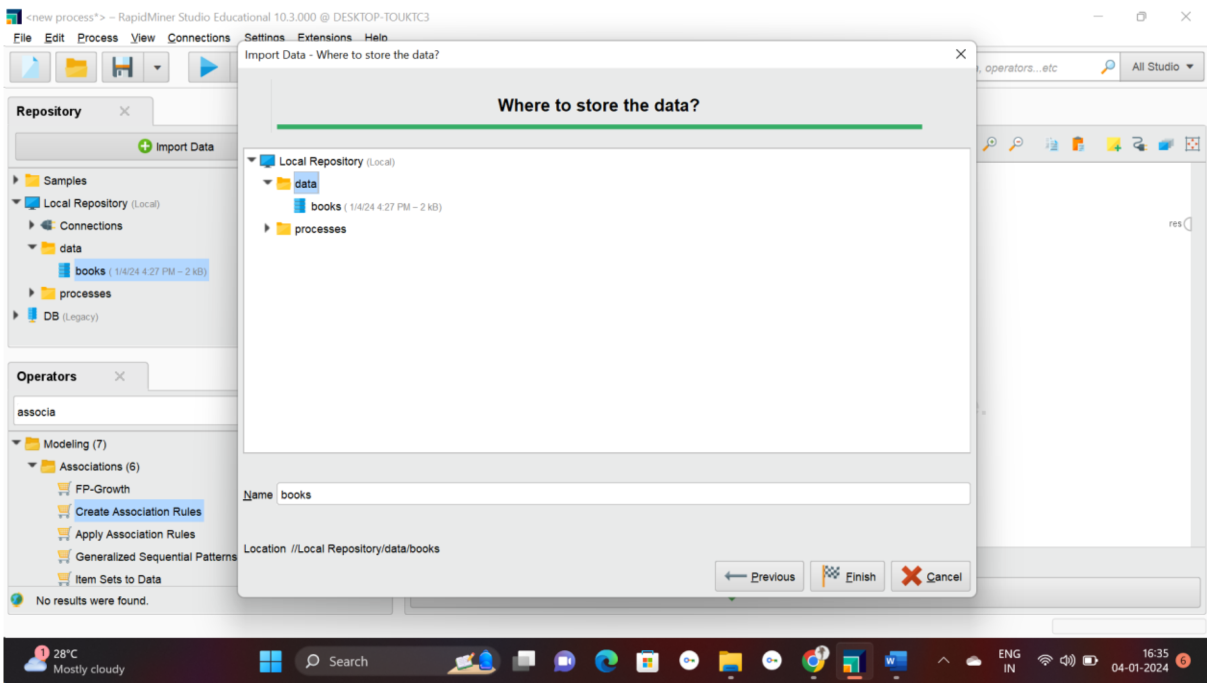
Objective:
A shop wants to analyse customer purchase patterns of books to optimize product placement and run targeted promotions. The goal is to identify frequent item sets and association rules among purchased products.

Implementation:
The Shop has a transaction history that includes the books purchased by customers. Using the Apriori algorithm and association rule mining, the supermarket aims to discover patterns of books that are frequently bought together.

Step 1: Load Dataset

Click ->Import-> select from My computer and load, Save under Local Repository

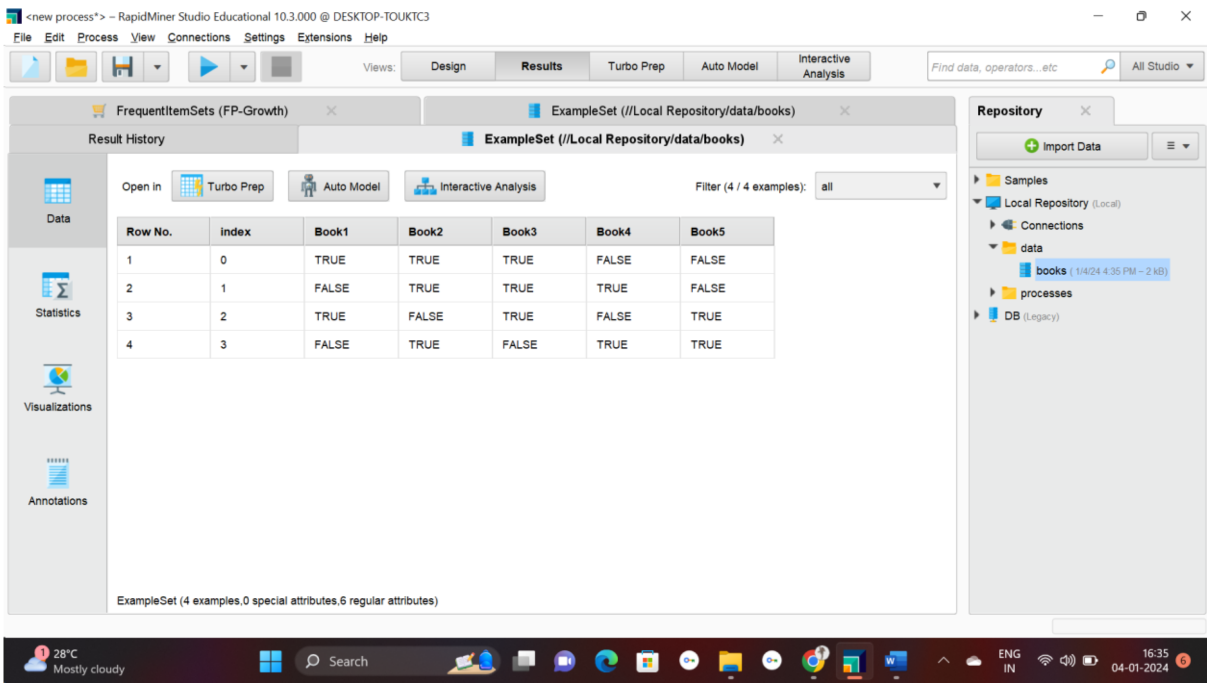



Import Data - Where to store the data?

Where to store the data?

Name: books

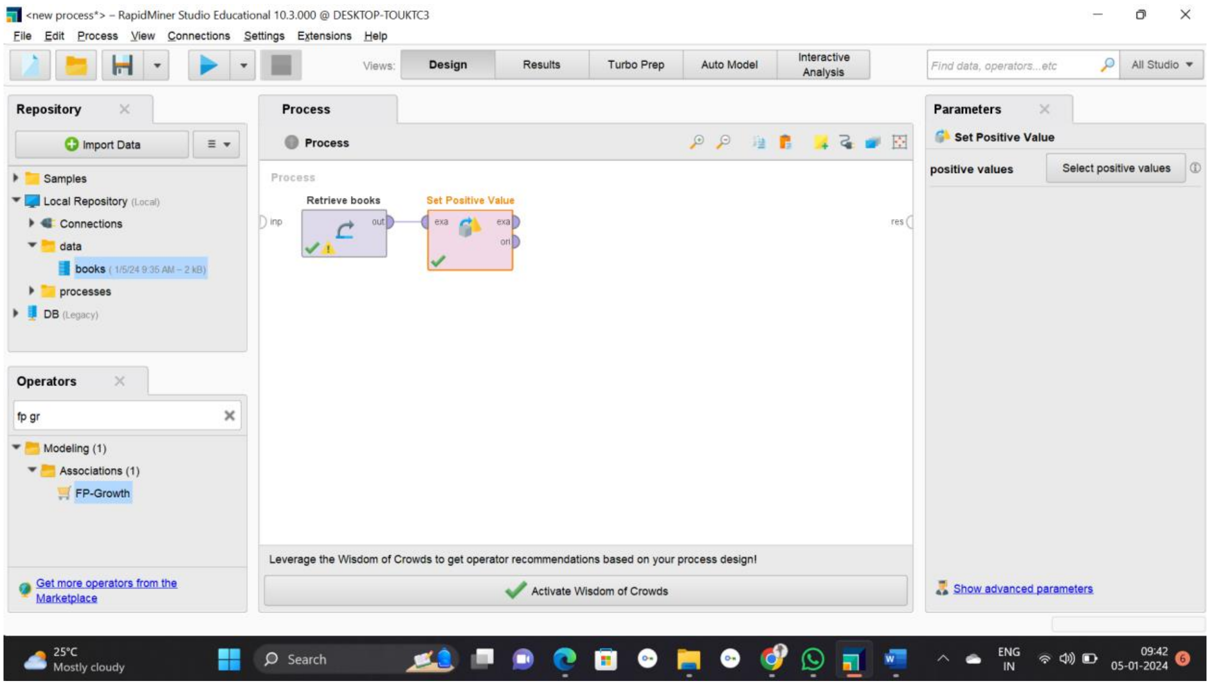
Location: //Local Repository/data/books

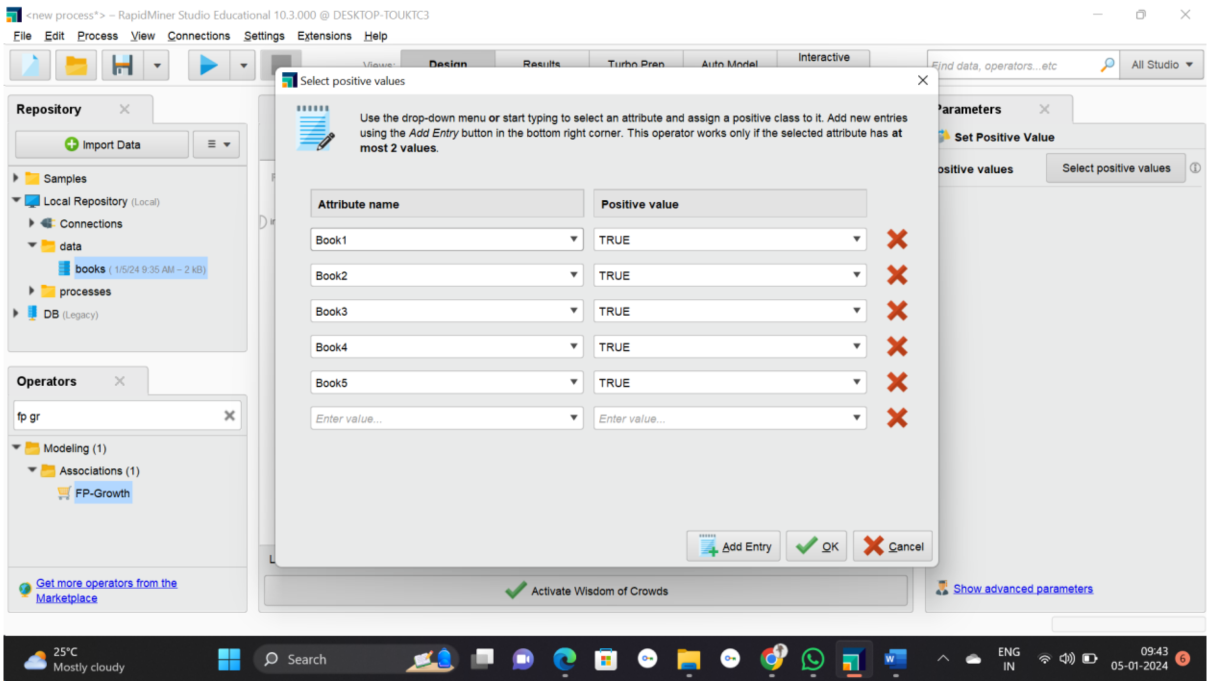

ExampleSet (//Local Repository/data/books)

Row No.	index	Book1	Book2	Book3	Book4	Book5
1	0	TRUE	TRUE	TRUE	FALSE	FALSE
2	1	FALSE	TRUE	TRUE	TRUE	FALSE
3	2	TRUE	FALSE	TRUE	FALSE	TRUE
4	3	FALSE	TRUE	FALSE	TRUE	TRUE

ExampleSet (4 examples,0 special attributes,6 regular attributes)

Step 2: Set Positive Values True to Positive

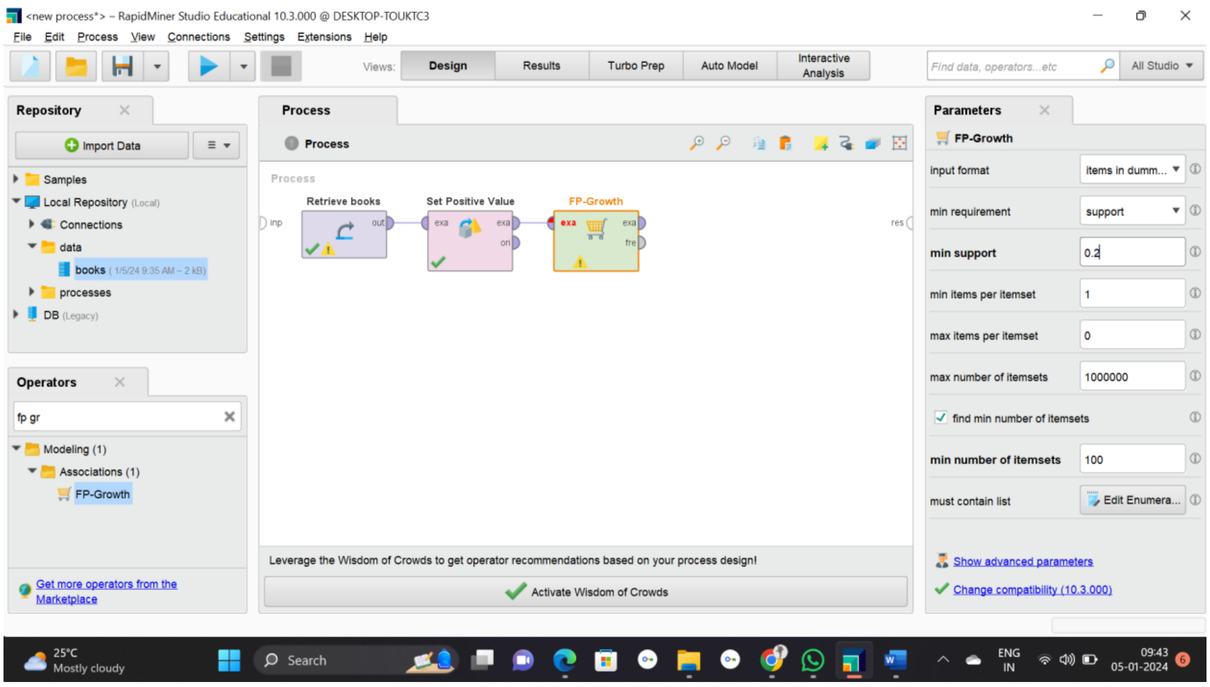




Use the drop-down menu or start typing to select an attribute and assign a positive class to it. Add new entries using the *Add Entry* button in the bottom right corner. This operator works only if the selected attribute has at **most 2 values.**

Attribute name	Positive value
Book1	TRUE
Book2	TRUE
Book3	TRUE
Book4	TRUE
Book5	TRUE
Enter value...	Enter value...

Step 3: Add FP-Growth and set Support Value

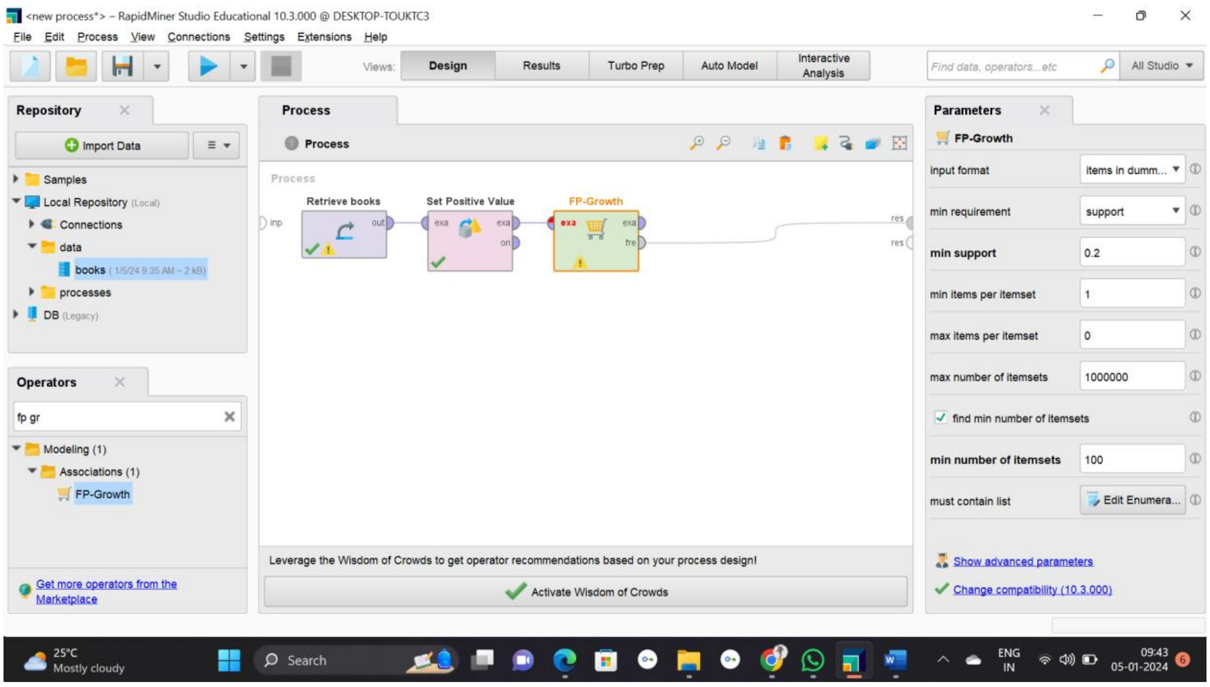

File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
Operators
fp gr
Modeling (1)
Associations (1)
FP-Growth
Get more operators from the Marketplace
Process
Retrieve books
Set Positive Value
FP-Growth
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Show advanced parameters
Change compatibility (10.3.000)

<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3

Parameters — FP-Growth

input format	items in dumm...
min requirement	support
min support	0.2
min items per itemset	1
max items per itemset	0
max number of itemsets	1000000
✓ find min number of itemsets	
min number of itemsets	100
must contain list	Edit Enumera...

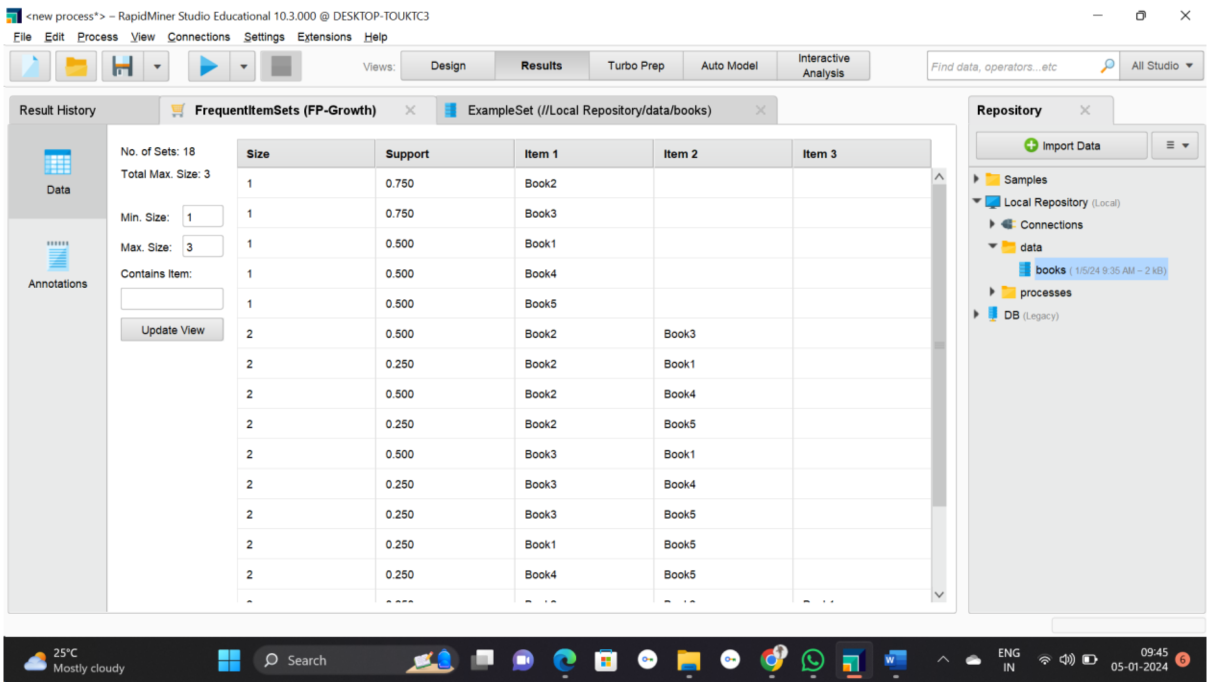
Step 4: View the Frequent Itemset


File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
Operators
fp gr
Modeling (1)
Associations (1)
FP-Growth
Get more operators from the Marketplace
Process
Retrieve books
Set Positive Value
FP-Growth
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Show advanced parameters
Change compatibility (10.3.000)

<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3

Parameters — FP-Growth

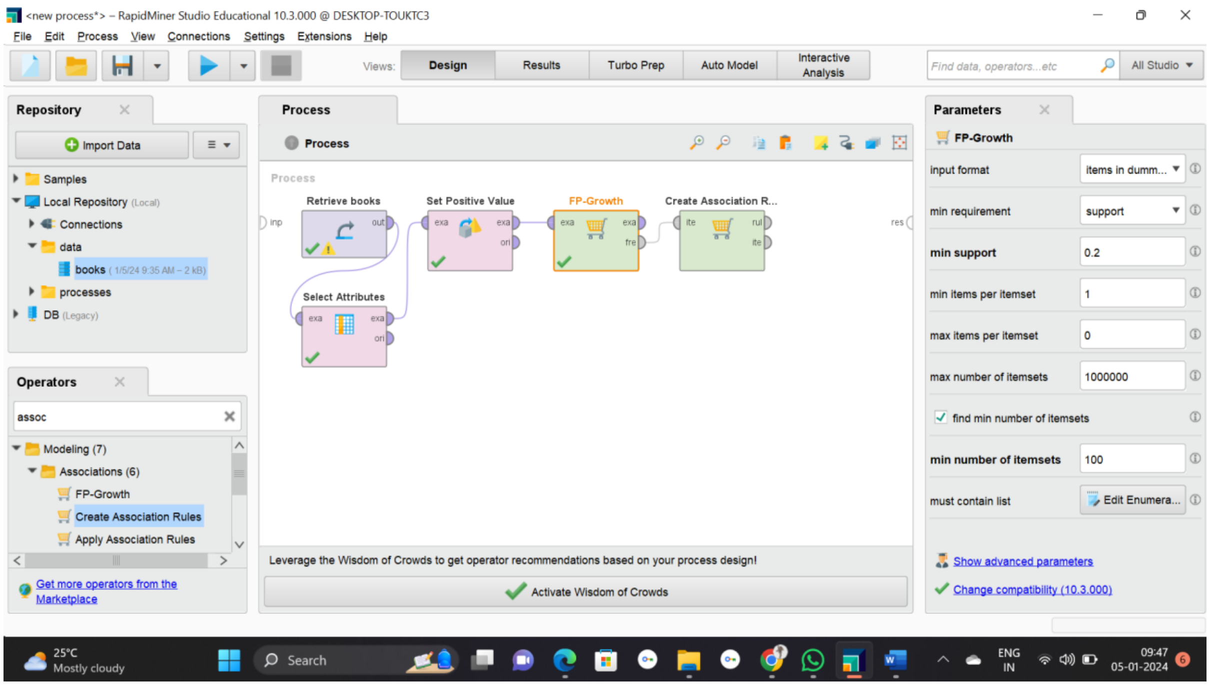
input format	items in dumm...
min requirement	support
min support	0.2
min items per itemset	1
max items per itemset	0
max number of itemsets	1000000
✓ find min number of itemsets	
min number of itemsets	100
must contain list	Edit Enumera...


<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Result History
FrequentItemSets (FP-Growth)
ExampleSet (//Local Repository/data/books)
Data
Annotations
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)

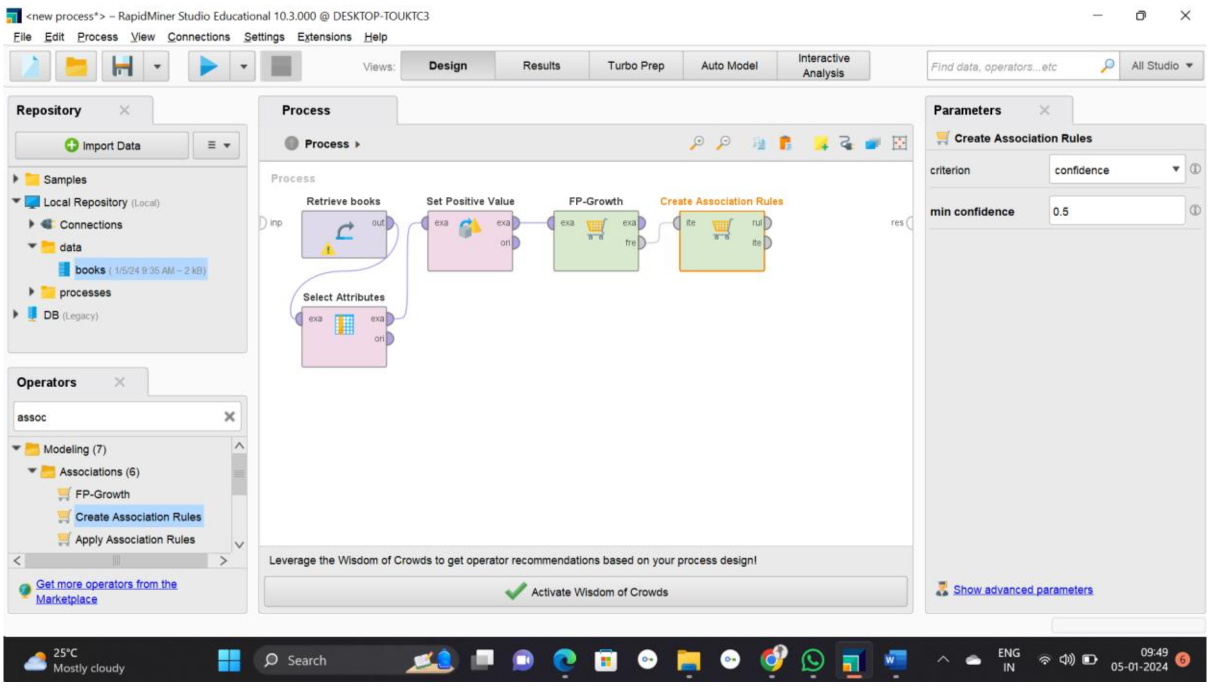
No. of Sets: 18
Total Max. Size: 3
Min. Size: 1
Max. Size: 3
Contains Item:
Update View

Size	Support	Item 1	Item 2	Item 3
1	0.750	Book2		
1	0.750	Book3		
1	0.500	Book1		
1	0.500	Book4		
1	0.500	Book5		
2	0.500	Book2	Book3	
2	0.250	Book2	Book1	
2	0.500	Book2	Book4	
2	0.250	Book2	Book5	
2	0.500	Book3	Book1	
2	0.250	Book3	Book4	
2	0.250	Book3	Book5	
2	0.250	Book1	Book5	
2	0.250	Book4	Book5	

Step 5: Add Create Association and set the confidence value


<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
Operators
assoc
Modeling (7)
Associations (6)
FP-Growth
Create Association Rules
Apply Association Rules
Get more operators from the Marketplace
Process
Retrieve books
Set Positive Value
FP-Growth
Create Association R...
Select Attributes
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
FP-Growth
Show advanced parameters
Change compatibility (10.3.000)

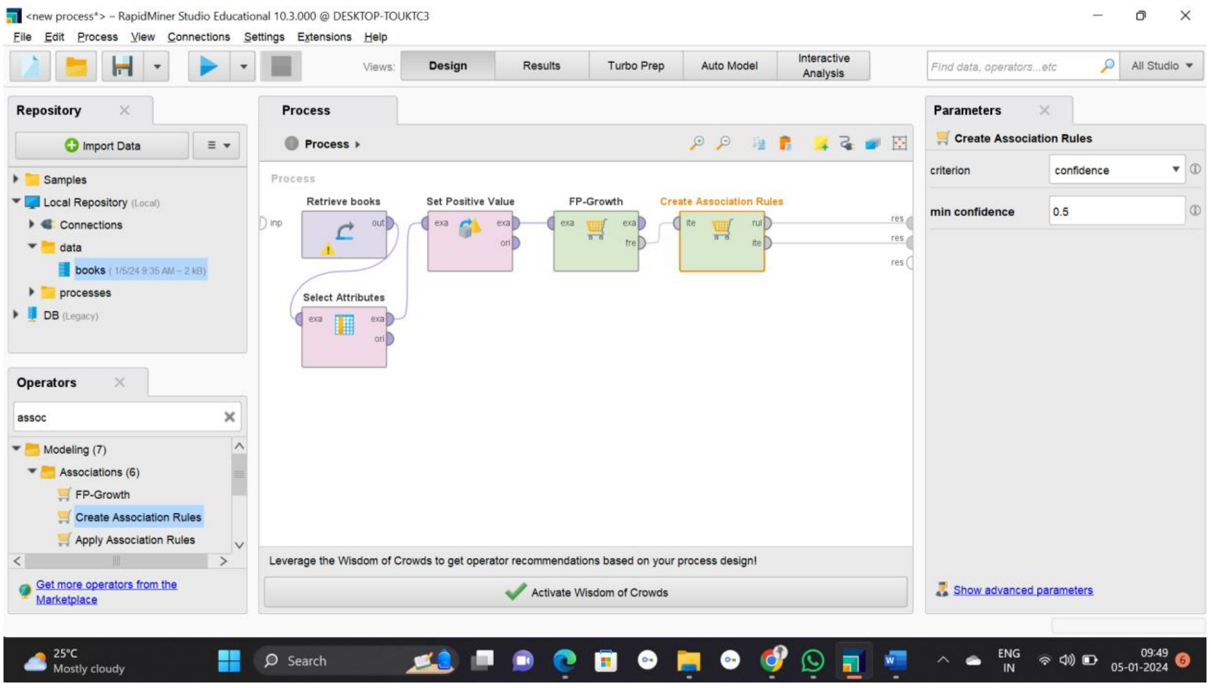
input format	items in dumm...
min requirement	support
min support	0.2
min items per itemset	1
max items per itemset	0
max number of itemsets	1000000
✓ find min number of itemsets	
min number of itemsets	100
must contain list	Edit Enumera...



<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
Operators
assoc
Modeling (7)
Associations (6)
FP-Growth
Create Association Rules
Apply Association Rules
Get more operators from the Marketplace
Process
Retrieve books
Set Positive Value
FP-Growth
Create Association Rules
Select Attributes
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Create Association Rules
Show advanced parameters

criterion	confidence
min confidence	0.5

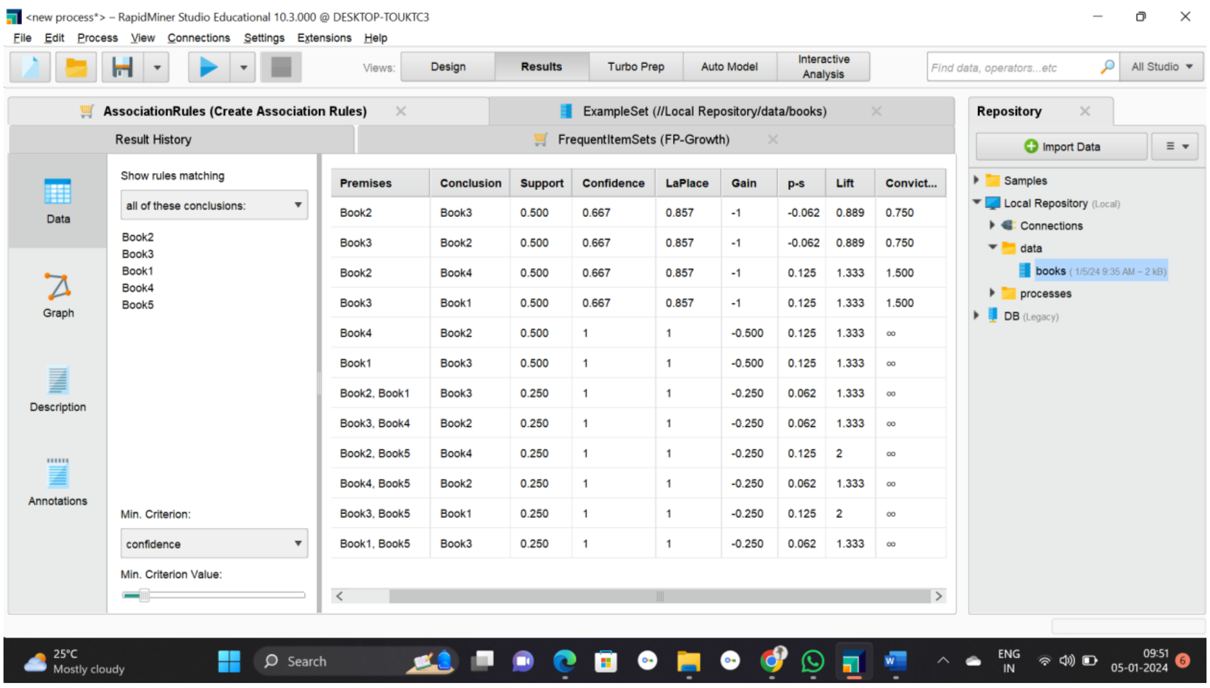
Step 6: View both rules and Item set



<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
Operators
assoc
Modeling (7)
Associations (6)
FP-Growth
Create Association Rules
Apply Association Rules
Get more operators from the Marketplace
Process
Retrieve books
Set Positive Value
FP-Growth
Create Association Rules
Select Attributes
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Create Association Rules
Show advanced parameters

criterion	confidence
min confidence	0.5

Step 7: Execute the Model



<new process*> – RapidMiner Studio Educational 10.3.000 @ DESKTOP-TOUKTC3
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Interactive Analysis
Find data, operators...etc
All Studio
AssociationRules (Create Association Rules)
ExampleSet (//Local Repository/data/books)
Result History
FrequentItemSets (FP-Growth)
Data
Graph
Description
Annotations
Repository
Import Data
Samples
Local Repository (Local)
Connections
data
books (1/5/24 9:35 AM – 2 kB)
processes
DB (Legacy)
25°C Mostly cloudy
Search
ENG IN
09:51 05-01-2024

Show rules matching

all of these conclusions:

Book2
Book3
Book1
Book4
Book5

Min. Criterion:

confidence

Min. Criterion Value:

Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Convict...
Book2	Book3	0.500	0.667	0.857	-1	-0.062	0.889	0.750
Book3	Book2	0.500	0.667	0.857	-1	-0.062	0.889	0.750
Book2	Book4	0.500	0.667	0.857	-1	0.125	1.333	1.500
Book3	Book1	0.500	0.667	0.857	-1	0.125	1.333	1.500
Book4	Book2	0.500	1	1	-0.500	0.125	1.333	∞
Book1	Book3	0.500	1	1	-0.500	0.125	1.333	∞
Book2, Book1	Book3	0.250	1	1	-0.250	0.062	1.333	∞
Book3, Book4	Book2	0.250	1	1	-0.250	0.062	1.333	∞
Book2, Book5	Book4	0.250	1	1	-0.250	0.125	2	∞
Book4, Book5	Book2	0.250	1	1	-0.250	0.062	1.333	∞
Book3, Book5	Book1	0.250	1	1	-0.250	0.125	2	∞
Book1, Book5	Book3	0.250	1	1	-0.250	0.062	1.333	∞

Apriori Algorithm:

Tool : Google Colab

Use Case: Books Data

Objective:
The main objective is to identify sets of books that frequently occur together in the dataset. These sets are known as frequent itemsets, from the frequent itemsets, generate association rules that express relationships between different sets of books and Visualize the generated association rules as a directed graph.

Implementation:
Using the Apriori algorithm and association rule mining, the prediction is brought to discover patterns that are frequently happens together.

Program:

```
# Install necessary libraries
!pip install mlxtend
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules

# Sample dataset (replace this with your actual dataset)
buying_books_data = [
    ['Book1', 'Book2', 'Book3'],
    ['Book2', 'Book3', 'Book4'],
    ['Book1', 'Book3', 'Book5'],
    ['Book2', 'Book4', 'Book5'],
]

# Convert the dataset to a one-hot encoded format
te = TransactionEncoder()
te_ary = te.fit(buying_books_data).transform(buying_books_data)
df_buying_books = pd.DataFrame(te_ary, columns=te.columns_)

# Apply the Apriori algorithm
min_support = 0.2  # Adjust as needed
frequent_itemsets = apriori(df_buying_books, min_support=min_support,
use_colnames=True)

# Generate association rules
min_confidence = 0.5  # Adjust as needed
rules = association_rules(frequent_itemsets, metric='confidence',
min_threshold=min_confidence)

# Display frequent itemsets
```

```
print("Frequent Itemsets:")
print(frequent_itemsets)

# Display association rules
print("\nAssociation Rules:")
print(rules)
```

OUTPUT:

```
Frequent Itemsets:
    support               itemsets
0      0.50                (Book1)
1      0.75                (Book2)
2      0.75                (Book3)
3      0.50                (Book4)
4      0.50                (Book5)
5      0.25         (Book2, Book1)
6      0.50         (Book3, Book1)
7      0.25         (Book5, Book1)
8      0.50         (Book3, Book2)
9      0.50         (Book4, Book2)
10     0.25         (Book5, Book2)
11     0.25         (Book4, Book3)
12     0.25         (Book5, Book3)
13     0.25         (Book5, Book4)
14     0.25  (Book3, Book2, Book1)
15     0.25  (Book5, Book3, Book1)
16     0.25  (Book4, Book3, Book2)
17     0.25  (Book4, Book5, Book2)
```

```
Association Rules:
       antecedents     consequents  antecedent support  consequent
support  \
0          (Book1)         (Book2)                    0.50
0.75
1          (Book3)         (Book1)                    0.75
0.50
2          (Book1)         (Book3)                    0.50
0.75
3          (Book5)         (Book1)                    0.50
0.50
4          (Book1)         (Book5)                    0.50
0.50
5          (Book3)         (Book2)                    0.75
0.75
6          (Book2)         (Book3)                    0.75
0.75
7          (Book4)         (Book2)                    0.50
0.75
8          (Book2)         (Book4)                    0.75
0.50
9          (Book5)         (Book2)                    0.50
0.75
10         (Book4)         (Book3)                    0.50
0.75
```

```
11          (Book5)         (Book3)                   0.50
0.75
12          (Book5)         (Book4)                   0.50
0.50
13          (Book4)         (Book5)                   0.50
0.50
14  (Book2, Book3)          (Book1)                   0.50
0.50
15  (Book3, Book1)          (Book2)                   0.50
0.75
16  (Book2, Book1)          (Book3)                   0.25
0.75
17          (Book1)  (Book2, Book3)                   0.50
0.50
18  (Book5, Book3)          (Book1)                   0.25
0.50
19  (Book5, Book1)          (Book3)                   0.25
0.75
20  (Book3, Book1)          (Book5)                   0.50
0.50
21          (Book5)  (Book3, Book1)                   0.50
0.50
22          (Book1)  (Book5, Book3)                   0.50
0.25
23  (Book3, Book4)          (Book2)                   0.25
0.75
24  (Book2, Book4)          (Book3)                   0.50
0.75
25  (Book2, Book3)          (Book4)                   0.50
0.50
26          (Book4)  (Book2, Book3)                   0.50
0.50
27  (Book5, Book4)          (Book2)                   0.25
0.75
28  (Book2, Book4)          (Book5)                   0.50
0.50
29  (Book5, Book2)          (Book4)                   0.25
0.50
30          (Book4)  (Book5, Book2)                   0.50
0.25
31          (Book5)  (Book2, Book4)                   0.50
0.50
```

```
    support  confidence      lift  leverage  conviction  zhangs_metric
0      0.25    0.500000  0.666667   -0.1250        0.50      -0.500000
1      0.50    0.666667  1.333333    0.1250        1.50       1.000000
2      0.50    1.000000  1.333333    0.1250         inf       0.500000
3      0.25    0.500000  1.000000    0.0000        1.00       0.000000
4      0.25    0.500000  1.000000    0.0000        1.00       0.000000
5      0.50    0.666667  0.888889   -0.0625        0.75      -0.333333
6      0.50    0.666667  0.888889   -0.0625        0.75      -0.333333
7      0.50    1.000000  1.333333    0.1250         inf       0.500000
8      0.50    0.666667  1.333333    0.1250        1.50       1.000000
9      0.25    0.500000  0.666667   -0.1250        0.50      -0.500000
10     0.25    0.500000  0.666667   -0.1250        0.50      -0.500000
11     0.25    0.500000  0.666667   -0.1250        0.50      -0.500000
12     0.25    0.500000  1.000000    0.0000        1.00       0.000000
13     0.25    0.500000  1.000000    0.0000        1.00       0.000000
```

```
14      0.25    0.500000  1.000000    0.0000         1.00        0.000000
15      0.25    0.500000  0.666667   -0.1250         0.50       -0.500000
16      0.25    1.000000  1.333333    0.0625          inf        0.333333
17      0.25    0.500000  1.000000    0.0000         1.00        0.000000
18      0.25    1.000000  2.000000    0.1250          inf        0.666667
19      0.25    1.000000  1.333333    0.0625          inf        0.333333
20      0.25    0.500000  1.000000    0.0000         1.00        0.000000
21      0.25    0.500000  1.000000    0.0000         1.00        0.000000
22      0.25    0.500000  2.000000    0.1250         1.50        1.000000
23      0.25    1.000000  1.333333    0.0625          inf        0.333333
24      0.25    0.500000  0.666667   -0.1250         0.50       -0.500000
25      0.25    0.500000  1.000000    0.0000         1.00        0.000000
26      0.25    0.500000  1.000000    0.0000         1.00        0.000000
27      0.25    1.000000  1.333333    0.0625          inf        0.333333
28      0.25    0.500000  1.000000    0.0000         1.00        0.000000
29      0.25    1.000000  2.000000    0.1250          inf        0.666667
30      0.25    0.500000  2.000000    0.1250         1.50        1.000000
31      0.25    0.500000  1.000000    0.0000         1.00        0.000000
```

<u>FP-growth algorithm</u>

<u>Tool : Google Colab</u>

Program:

```
!pip install pyfpgrowth

import pyfpgrowth

# Sample dataset (replace this with your actual dataset)
buying_books_data = [
    ['Book1', 'Book2', 'Book3'],
    ['Book2', 'Book3', 'Book4'],
    ['Book1', 'Book3', 'Book5'],
    ['Book2', 'Book4', 'Book5'],
]

# Convert the dataset to a list of transactions
transactions = [tuple(transaction) for transaction in
buying_books_data]

# Apply the FP-growth algorithm
min_support = 2  # Adjust as needed
patterns = pyfpgrowth.find_frequent_patterns(transactions, min_support)

# Generate association rules
min_confidence = 0.5  # Adjust as needed
rules = pyfpgrowth.generate_association_rules(patterns, min_confidence)

# Display frequent itemsets
print("Frequent Itemsets:")
print(patterns)
```

```
# Display association rules
print("\nAssociation Rules:")
print(rules)
```

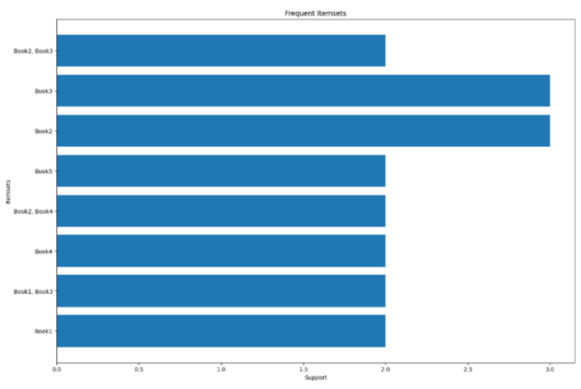
```
itemset_labels = [', '.join(map(str, itemset)) for itemset in
patterns.keys()]
```

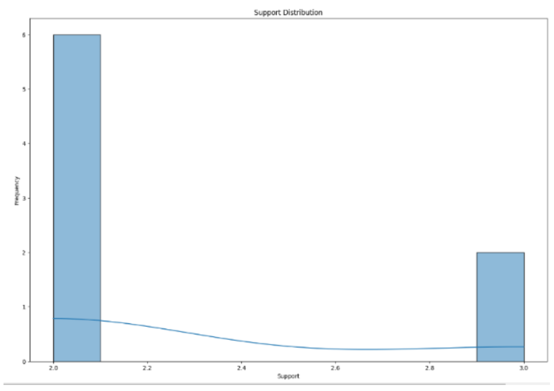
```
plt.figure(figsize=(36, 24))
plt.subplot(2, 2, 1)
plt.barh(itemset_labels, list(patterns.values()))
plt.xlabel('Support')
plt.ylabel('Itemsets')
plt.title('Frequent Itemsets')

import seaborn as sns

plt.subplot(2, 2, 2)
sns.histplot(list(patterns.values()), bins=10, kde=True)
plt.xlabel('Support')
plt.ylabel('Frequency')
plt.title('Support Distribution')
```

OUTPUT:

```
Frequent Itemsets:
{('Book1',): 2, ('Book1', 'Book3'): 2, ('Book4',): 2, ('Book2',
'Book4'): 2, ('Book5',): 2, ('Book2',): 3, ('Book3',): 3, ('Book2',
'Book3'): 2}
```

```
Association Rules:
{('Book1',): (('Book3',), 1.0), ('Book3',): (('Book2',),
0.6666666666666666), ('Book2',): (('Book3',), 0.6666666666666666),
('Book4',): (('Book2',), 1.0)}
```


Frequent Itemsets
Book2, Book3
Book3
Book2
Book5
Book2, Book4
Book4
Book1, Book3
Book1
Itemsets
Support


Support Distribution
Frequency
Support