Market Basket Insights

MARKET BASKET INSIGHTS PYTHON PROGRAM:

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules

def find(data):
    data=list(data.apply(lambda x:x.split(" ")))

    te = TransactionEncoder()
    te_data = te.fit(data).transform(data)
    df = pd.DataFrame(te_data,columns=te.columns_)

    df1 = apriori(df,min_support = 0.01,use_colnames = True)
    print(df1.sort_values(by = "support",ascending = False))

    df_ar = association_rules(df1, metric = "confidence",min_threshold=0.5)
    print(df_ar)

#pd.set_option("display.max_columns",None)
#pd.set_option("display.max_rows",None)

data=pd.read_csv("Book1.csv")
find(data["Itemname"])
find(data["Country"])
```

MARKET BASKET INSIGHTS PYTHON OUTPUT:

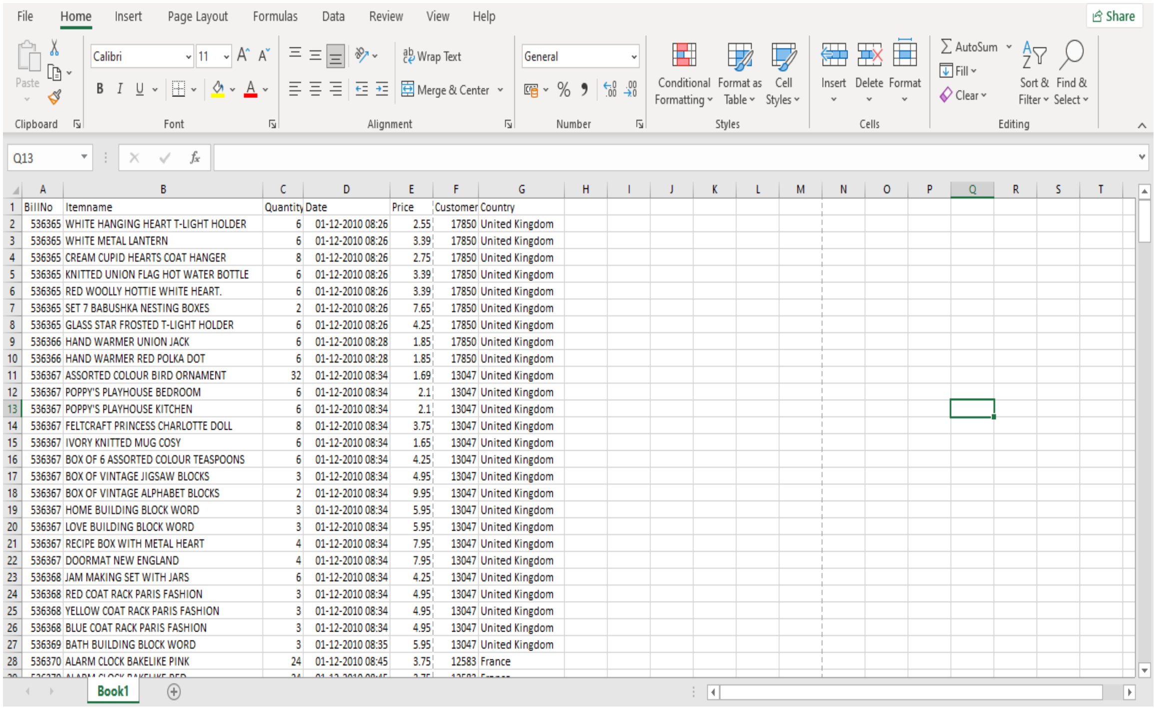
```
     support                 itemsets
119  0.133779                   (RED)
143  0.113712                 (WHITE)
126  0.090301                   (SET)
121  0.083612             (RETROSPOT)
69   0.083612                 (HEART)
..        ...                     ...
467  0.010033  (ANTIQUE, FRAME, WHITE)
465  0.010033        (ANT, S/3, WOOD)
464  0.010033       (ANT, WHITE, S/3)
461  0.010033      (ANT, FINISH, S/3)
411  0.010033             (S/3, WOOD)

[822 rows x 2 columns]
           antecedents  ... zhangs_metric
0             (LIGHTS)  ...      1.000000
1                 (10)  ...      0.996622
2            (CABINET)  ...      0.993174
3                  (2)  ...      0.986441
4                  (2)  ...      0.981921
...                ...  ...           ...
3733  (BOTTLE, KNITTED)  ...      1.000000
3734   (KNITTED, WATER)  ...      1.000000
3735            (UNION)  ...      1.000000
3736             (FLAG)  ...      1.000000
3737          (KNITTED)  ...      1.000000

[3738 rows x 10 columns]
```

```
support          itemsets
2  0.886288         (Kingdom)
3  0.886288          (United)
4  0.886288  (Kingdom, United)
1  0.066890          (France)
0  0.046823       (Australia)
  antecedents consequents  ...  conviction  zhangs_metric
0   (Kingdom)    (United)  ...         inf            1.0
1    (United)   (Kingdom)  ...         inf            1.0

[2 rows x 10 columns]
```



	BillNo	Itemname	Quantity	Date	Price	Customer	Country
2	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010 08:26	2.55	17850	United Kingdom
3	536365	WHITE METAL LANTERN	6	01-12-2010 08:26	3.39	17850	United Kingdom
4	536365	CREAM CUPID HEARTS COAT HANGER	8	01-12-2010 08:26	2.75	17850	United Kingdom
5	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	01-12-2010 08:26	3.39	17850	United Kingdom
6	536365	RED WOOLLY HOTTIE WHITE HEART.	6	01-12-2010 08:26	3.39	17850	United Kingdom
7	536365	SET 7 BABUSHKA NESTING BOXES	2	01-12-2010 08:26	7.65	17850	United Kingdom
8	536365	GLASS STAR FROSTED T-LIGHT HOLDER	6	01-12-2010 08:26	4.25	17850	United Kingdom
9	536366	HAND WARMER UNION JACK	6	01-12-2010 08:28	1.85	17850	United Kingdom
10	536366	HAND WARMER RED POLKA DOT	6	01-12-2010 08:28	1.85	17850	United Kingdom
11	536367	ASSORTED COLOUR BIRD ORNAMENT	32	01-12-2010 08:34	1.69	13047	United Kingdom
12	536367	POPPY'S PLAYHOUSE BEDROOM	6	01-12-2010 08:34	2.1	13047	United Kingdom
13	536367	POPPY'S PLAYHOUSE KITCHEN	6	01-12-2010 08:34	2.1	13047	United Kingdom
14	536367	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01-12-2010 08:34	3.75	13047	United Kingdom
15	536367	IVORY KNITTED MUG COSY	6	01-12-2010 08:34	1.65	13047	United Kingdom
16	536367	BOX OF 6 ASSORTED COLOUR TEASPOONS	6	01-12-2010 08:34	4.25	13047	United Kingdom
17	536367	BOX OF VINTAGE JIGSAW BLOCKS	3	01-12-2010 08:34	4.95	13047	United Kingdom
18	536367	BOX OF VINTAGE ALPHABET BLOCKS	2	01-12-2010 08:34	9.95	13047	United Kingdom
19	536367	HOME BUILDING BLOCK WORD	3	01-12-2010 08:34	5.95	13047	United Kingdom
20	536367	LOVE BUILDING BLOCK WORD	3	01-12-2010 08:34	5.95	13047	United Kingdom
21	536367	RECIPE BOX WITH METAL HEART	4	01-12-2010 08:34	7.95	13047	United Kingdom
22	536367	DOORMAT NEW ENGLAND	4	01-12-2010 08:34	7.95	13047	United Kingdom
23	536368	JAM MAKING SET WITH JARS	6	01-12-2010 08:34	4.25	13047	United Kingdom
24	536368	RED COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
25	536368	YELLOW COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
26	536368	BLUE COAT RACK PARIS FASHION	3	01-12-2010 08:34	4.95	13047	United Kingdom
27	536369	BATH BUILDING BLOCK WORD	3	01-12-2010 08:35	5.95	13047	United Kingdom
28	536370	ALARM CLOCK BAKELIKE PINK	24	01-12-2010 08:45	3.75	12583	France

EXPLANATION:

1.Importing Libraries:

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

- pandas is a powerful data manipulation library.
- mlxtend is a library that provides utilities for various tasks in machine learning, including association rule mining.

2. Defining the Function:

```
def find(data):
```

- Defines a function named find that takes a pandas Series (data) as an argument.

3.Data Preprocessing:

```
data = list(data.apply(lambda x: x.split(" ")))
```

- Splits each entry in the input data Series by space and converts it into a list.

4.Transaction Encoding:

```
te = TransactionEncoder()
te_data = te.fit(data).transform(data)
```

- TransactionEncoder is used to convert the list of lists into a one-hot encoded format.
- fit is used to fit the encoder to the data, and transform is used to transform the data into a binary matrix.

5.Creating DataFrame:

```
df = pd.DataFrame(te_data, columns=te.columns_)
```

- Creates a pandas DataFrame from the one-hot encoded data with columns representing unique items.

6.Apriori Algorithm:

```
df1 = apriori(df, min_support=0.01, use_colnames=True)
```

- Applies the Apriori algorithm to find frequent itemsets with a minimum support of 0.01.
- use_colnames=True uses the actual item names in the resulting DataFrame.

7.Displaying Frequent Itemsets:

print(df1.sort_values(by="support", ascending=False))

- Prints the frequent itemsets sorted by support in descending order.

8.Association Rules:

df_ar = association_rules(df1, metric="confidence", min_threshold=0.5)

- Generates association rules from the frequent itemsets with a minimum confidence of 0.5.

9.Displaying Association Rules:

print(df_ar)

- Prints the generated association rules.

10.Pandas Display Options:

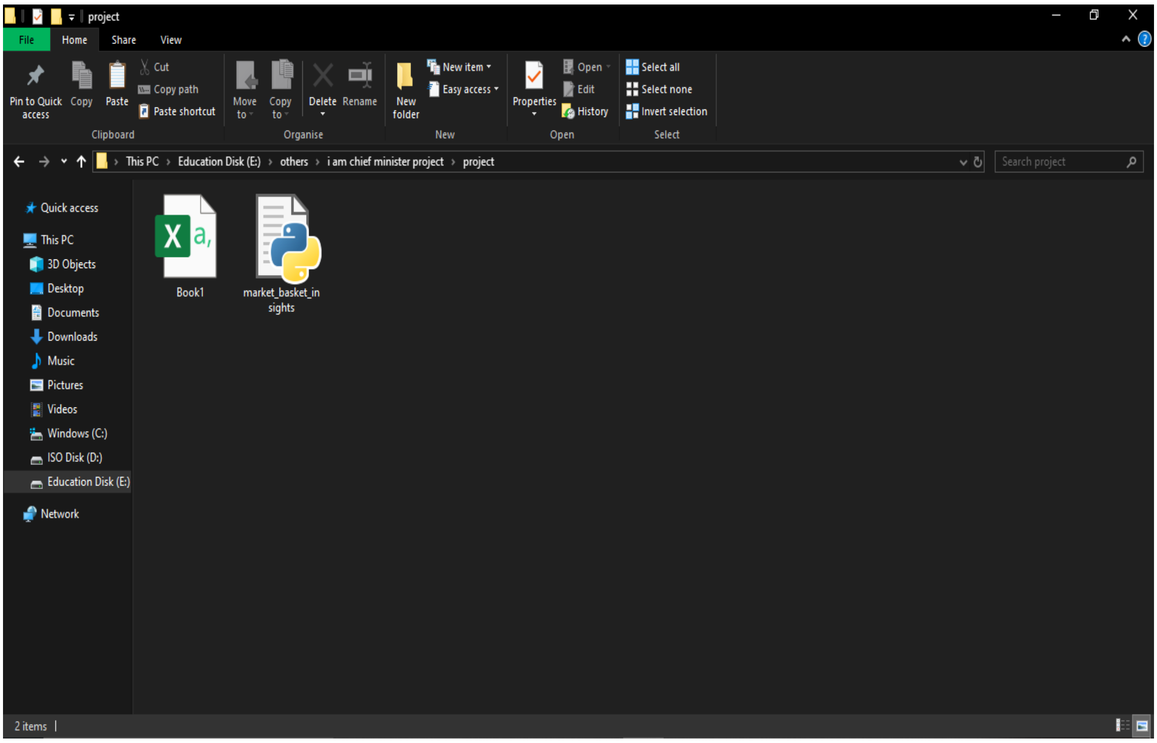
pd.set_option("display.max_columns", None)
pd.set_option("display.max_rows", None)

- Sets pandas display options to show all columns and rows without truncation.

11.Reading Data and Applying the find Function:

```
data = pd.read_csv("Book1.csv")
find(data["Itemname"])
find(data["Country"])
```

- Reads a CSV file ("Book1.csv") into a pandas DataFrame.
- Applies the find function to the "Itemname" and "Country" columns of the DataFrame.



STEPS TO LOAD AND PREPROCESS THE TRANSACTION DATASET FOR MARKET BASKET INSIGHTS:

Data Collection:

Obtain the transaction dataset. This data typically consists of records of items purchased together in various transactions, like shopping carts in a supermarket.

Data Loading:

Use appropriate tools or libraries (e.g., Python with pandas) to load the transaction dataset into my project environment.

Data Exploration:

Familiarize myself with the dataset to understand its structure and contents. Check for missing values or anomalies.

Data Preprocessing:

Perform the following preprocessing steps:

- Data Cleaning: Handle missing values or outliers, if any.
- Transaction Identification: Group transactions by a unique identifier (e.g., receipt or order ID).
- Item Identification: Identify unique items or products in the dataset.

- Data Transformation: Convert the data into a suitable format for association analysis. Typically, this involves creating a binary matrix where rows represent transactions, columns represent items, and the cells indicate whether an item was purchased in a transaction (1 for yes, 0 for no).

- Remove Duplicates: Ensure that duplicate items in a transaction are handled correctly.

Association Analysis:

we can use popular algorithms like Apriori to discover associations between items in the dataset. These algorithms will help you find frequent itemsets and generate association rules.

Support and Confidence Thresholds:

Set appropriate support and confidence thresholds to filter and focus on significant associations.

Interpretation:

Analyze the generated association rules to gain insights into customer behavior, product recommendations, or marketing strategies.

Conclusion:

Market basket analysis is a powerful technique for uncovering hidden patterns and associations between products.

By understanding customer purchasing behaviour and identifying potential cross-selling opportunities, retailers can make more informed decisions about product placement, marketing campaigns, and inventory management.