Supermarket

MARKET BASKET ANALYSIS

Data analysis

By:

Rawan Alharthi	444001029
Mashael Abdali	444001062
Shaimaa Alghamdi	444000746

Objective

The objective is to conduct market basket analysis using the (Apriori) algorithm to uncover relationships between products in a retail sales dataset.

Data Processing

Loading the Dataset:

The dataset was loaded from the (Supermarket) dataset.

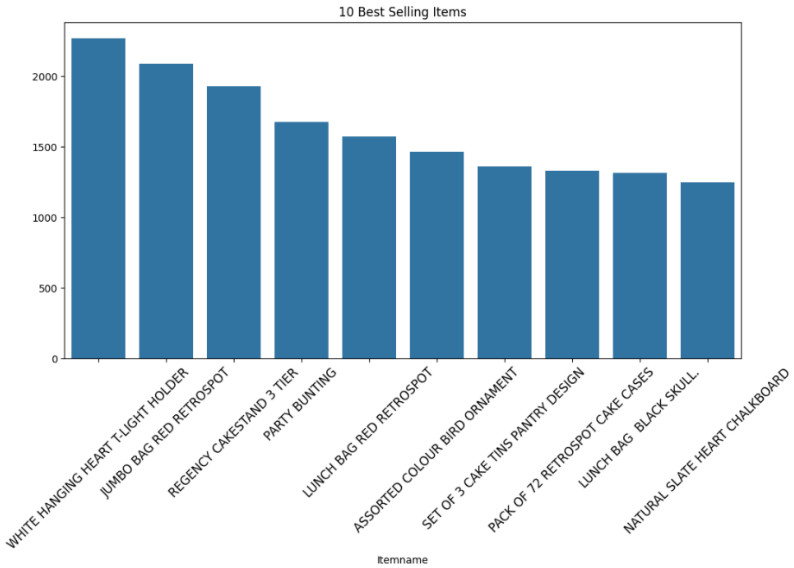
Data Cleaning:

- **Removing Missing Rows.**
- **Handling Numeric Values: Commas in the "Price" column were replaced with periods (.) to convert the values into the correct data type (float), facilitating numerical operations.**
- **Displaying Cleaned Data: The first five rows of the cleaned dataset were displayed to verify the success of the cleaning process.**

	BillNo	Itemname	Quantity	Date	Price	CustomerID	Country
0	536365	WHITE HANGING HEART T-LIGHT HOLDER	6	01.12.2010 08:26	2.55	17850.0	United Kingdom
1	536365	WHITE METAL LANTERN	6	01.12.2010 08:26	3.39	17850.0	United Kingdom
2	536365	CREAM CUPID HEARTS COAT HANGER	8	01.12.2010 08:26	2.75	17850.0	United Kingdom
3	536365	KNITTED UNION FLAG HOT WATER BOTTLE	6	01.12.2010 08:26	3.39	17850.0	United Kingdom
4	536365	RED WOOLLY HOTTIE WHITE HEART.	6	01.12.2010 08:26	3.39	17850.0	United Kingdom

Exploratory Data

Bar plot to display the top 10 best-selling items.



Transaction Aggregation

We aggregate the items in each transaction (based on BillNo) to create a basket of items purchased in each transaction.

```
# Transaction Aggregation

basket = (data_cleaned.groupby(['BillNo', 'Itemname'])['Quantity']
          .sum().unstack().reset_index().fillna(0)
          .set_index('BillNo'))

basket.head(10)
```

Itemname	*Boombox Ipod Classic	*USB Office Mirror Ball	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 DAISY PEGS IN WOOD BOX	12 EGG HOUSE PAINTED WOOD	12 HANGING EGGS HAND PAINTED	12 IVORY ROSE PEG PLACE SETTINGS	12 MESSAGE CARDS WITH ENVELOPES	12 PENCIL SMALL TUBE WOODLAND	...	wrongly coded 20713	wrongly coded 23343	wrongly coded- 23343	wrongly marked	wrongly marked 23343	wrongly marked carton 22804	wrongly marked. 23343 in box	wrongly sold (22719) barcode
BillNo																			
536365	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536366	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536367	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536368	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536369	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536370	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536371	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536372	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536373	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536374	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

10 rows × 4185 columns

Data Converting:

For the Apriori algorithm, we need to represent whether an item was purchased (1) or not (0) in each transaction.

Apriori Algorithm

The Apriori algorithm was applied to discover frequent itemsets in the shopping basket data, where the data was converted to a boolean format (True/False) and a minimum support threshold of 2% was used to identify frequent items, displaying the first five discovered frequent itemsets.

	support	itemsets
0	0.022613	(3 STRIPEY MICE FELTCRAFT)
1	0.022959	(4 TRADITIONAL SPINNING TOPS)
2	0.046611	(6 RIBBONS RUSTIC CHARM)
3	0.020633	(60 CAKE CASES DOLLY GIRL DESIGN)
4	0.029688	(60 CAKE CASES VINTAGE CHRISTMAS)

Generate Association Rules:

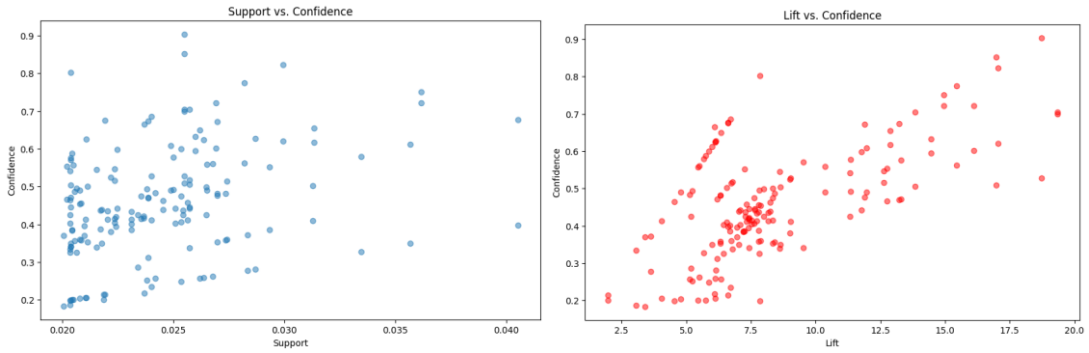
This analysis allows for understanding the relationships between items in the dataset by generating association rules. These rules can be used to make strategic decisions in marketing and promotion, such as determining which items should be displayed together or promoted.

threshold=0.5

- **Indicates strong relationships: Lift values greater than 1 show strong associations.**
- **Filters out unimportant rules: Helps avoid rules that don't reflect clear relationships.**

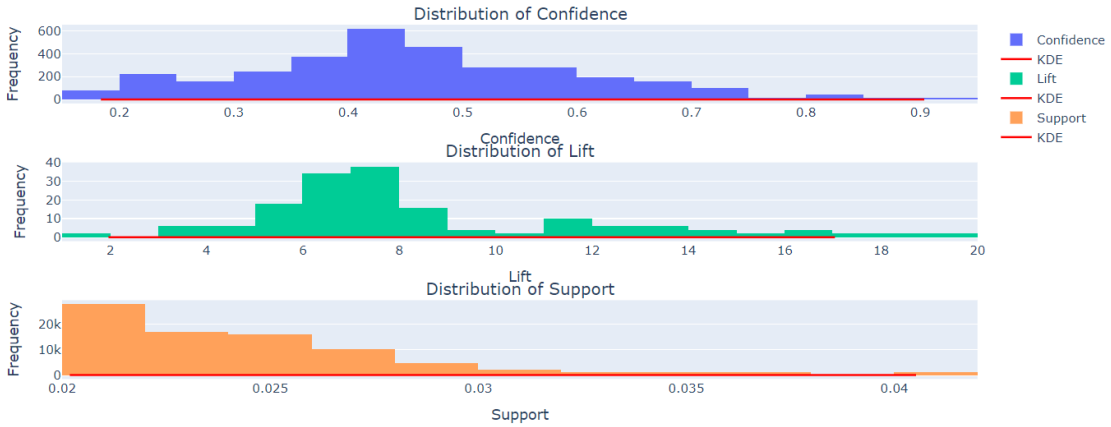
	antecedents	consequents	support	confidence	lift
153	(ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY...	(GREEN REGENCY TEACUP AND SAUCER)	0.025482	0.903509	18.728115
154	(GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY...	(ROSES REGENCY TEACUP AND SAUCER)	0.025482	0.851240	16.982778
25	(PINK REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.029936	0.822011	17.038810
158	(JUMBO BAG PINK POLKADOT, JUMBO STORAGE BAG SUKI)	(JUMBO BAG RED RETROSPOT)	0.020386	0.801556	7.848573
139	(PINK REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.028204	0.774457	15.450905
29	(GREEN REGENCY TEACUP AND SAUCER)	(ROSES REGENCY TEACUP AND SAUCER)	0.036170	0.749744	14.957866
28	(ROSES REGENCY TEACUP AND SAUCER)	(GREEN REGENCY TEACUP AND SAUCER)	0.036170	0.721619	14.957866
22	(GARDENERS KNEELING PAD CUP OF TEA)	(GARDENERS KNEELING PAD KEEP CALM)	0.026917	0.721485	16.111845
10	(CHARLOTTE BAG PINK POLKADOT)	(RED RETROSPOT CHARLOTTE BAG)	0.025730	0.704607	13.852246
152	(ROSES REGENCY TEACUP AND SAUCER, GREEN REGENC...	(PINK REGENCY TEACUP AND SAUCER)	0.025482	0.704514	19.345428

Visualizing support & lift vs. confidence:



Distribution of Confidence, Lift, and Support:

This allows for visualizing the different distributions of confidence, lift, and support in association rules. This understanding aids in analyzing relationships between items and making strategic decisions in areas such as marketing.



Conclusion

The analysis using the Apriori algorithm helped uncover important associations between products frequently bought together. These results can be used to optimize:

- **Product Placement: Group items that are frequently purchased together.**
- **Promotions: Bundle frequently associated items to increase cross-selling opportunities.**