IBM Naan Mudhalvan Project – Artificial Intelligence – Market Basket Analysis

- ## Problem Statement

"The problem is to perform market basket analysis on a provided dataset to unveil hidden patterns and associations between products. The goal is to understand customer purchasing behavior and identify potential cross-selling opportunities for a retail business. This project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization."

- ## Problem Analysis and Inference [Phase 1]

From analyzing the given problem, the task is to perform market basket analysis with the given dataset to find hidden patterns and relations between products if they exist. This is to understand the purchasing behavior of customers and check for potential cross-selling opportunities for retailers. This problem requires the usage of association analysis techniques, to generate insights for business improvements.

- ## Dataset Details and Description

- **Link to the dataset:** *https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis*
- The dataset contains transactional data from a retail store over a specified time period.
- Each row represents a unique transaction, listing the items purchased by a customer.
- The dataset includes information such as transaction ID, customer ID, and a list of purchased products.
- Product details include product names or IDs, categories, and prices.
- It also contains additional metadata such as country of purchase, date, time and other information.

- ## Design Thinking

- Discover frequent itemsets: Apply the Apriori algorithm to identify which products are frequently purchased together in customer transactions.
- Calculate association rules: Establish association rules, including support, confidence, and lift, to quantify the relationships between products.
- Uncover cross-selling opportunities: Identify product pairs or sets that exhibit strong associations, enabling the retail business to strategically promote and bundle related products.
- Understand customer purchasing behavior: Gain insights into customer preferences and behaviors based on the discovered patterns.
- Optimize business strategies: Utilize the analysis findings to enhance product placement, marketing campaigns, and overall business operations.

- ## Phases of Development
    1. **Phase 1 (Problem Definition and Design Thinking)** – The given problem is read and understood to find possible solutions. An outline of the solution is thought out using Design Thinking. The dataset is downloaded and prepared for analysis.
    2. **Phase 2 (Innovation)** – The dataset is preprocessed and imported in the program. Then the sales data is taken as variables and the products sold the most and highest quantity products are taken and a graph is plotted to visualize the data and to be able to interpret it easily

Number of Quantity Sold

Number of Sales
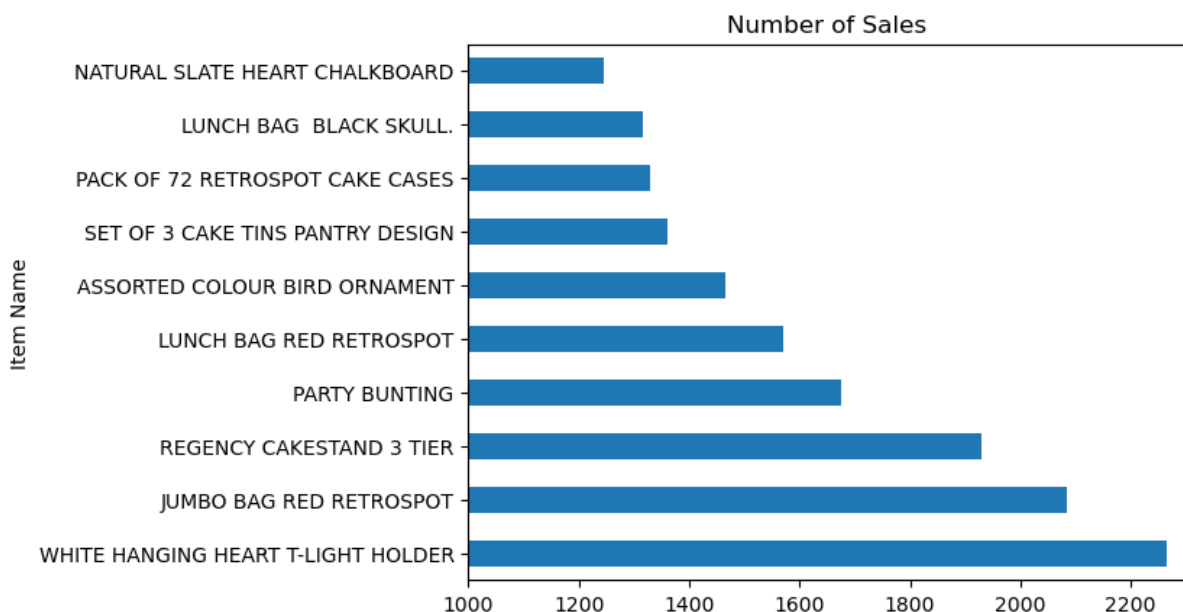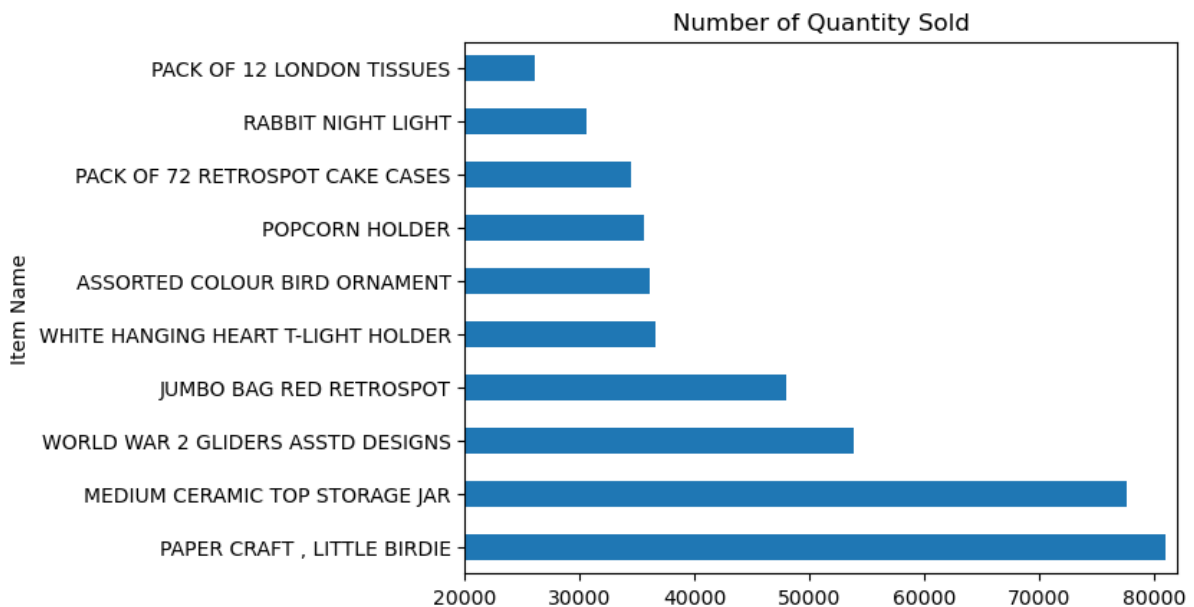
    3. **Phase 3 (AI Development Part 1)** – The apriori algorithm is imported and the association rules are created with a minimum threshold of 0.5% using the frequent itemsets sold.

Frequent Itemsets:

```
      support                                          itemsets
0    0.017370                          (10 COLOUR SPACEBOY PEN)
1    0.013751                (12 MESSAGE CARDS WITH ENVELOPES)
2    0.019653                 (12 PENCIL SMALL TUBE WOODLAND)
3    0.019820            (12 PENCILS SMALL TUBE RED RETROSPOT)
4    0.019597                 (12 PENCILS SMALL TUBE SKULL)
...        ...                                              ...
2467 0.010355   (LUNCH BAG RED RETROSPOT, LUNCH BAG SUKI DESIG...
2468 0.010188   (LUNCH BAG RED RETROSPOT, LUNCH BAG SUKI DESIG...
2469 0.010300   (LUNCH BAG RED RETROSPOT, LUNCH BAG SPACEBOY D...
2470 0.010467   (LUNCH BAG RED RETROSPOT, LUNCH BAG PINK POLKA...
2471 0.011302   (CHARLOTTE BAG PINK POLKADOT, STRAWBERRY CHARL...

[2472 rows x 2 columns]
```

Association Rules:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
0	(60 CAKE CASES DOLLY GIRL DESIGN)	(PACK OF 72 RETROSPOT CAKE CASES)	0.023160	0.071206	0.013028	0.562500	7.899629	0.011378	2.122958	0.894120
1	(60 TEATIME FAIRY CAKE CASES)	(PACK OF 72 RETROSPOT CAKE CASES)	0.044427	0.071206	0.024218	0.545113	7.655446	0.021054	2.041812	0.909794
2	(ALARM CLOCK BAKELIKE CHOCOLATE)	(ALARM CLOCK BAKELIKE GREEN)	0.023216	0.053558	0.015254	0.657074	12.268575	0.014011	2.759906	0.940321
3	(ALARM CLOCK BAKELIKE CHOCOLATE)	(ALARM CLOCK BAKELIKE PINK)	0.023216	0.042256	0.011691	0.503597	11.917802	0.010710	1.929369	0.937865
4	(ALARM CLOCK BAKELIKE CHOCOLATE)	(ALARM CLOCK BAKELIKE RED)	0.023216	0.057121	0.015811	0.681055	11.923112	0.014485	2.956246	0.937903
...	...	...	...	...	...	...	...	...	...	...
1392	(CHARLOTTE BAG SUKI DESIGN, STRAWBERRY CHARLOT...	(CHARLOTTE BAG PINK POLKADOT, WOODLAND CHARLOT...	0.018483	0.021824	0.011302	0.611446	28.017319	0.010898	2.517477	0.982467
1393	(CHARLOTTE BAG SUKI DESIGN, WOODLAND CHARLOTTE...	(CHARLOTTE BAG PINK POLKADOT, STRAWBERRY CHARL...	0.018595	0.020989	0.011302	0.607784	28.957623	0.010911	2.496105	0.983760
1394	(CHARLOTTE BAG PINK POLKADOT, STRAWBERRY CHARL...	(CHARLOTTE BAG SUKI DESIGN, WOODLAND CHARLOTTE...	0.020989	0.018595	0.011302	0.538462	28.957623	0.010911	2.126378	0.986165
1395	(CHARLOTTE BAG PINK POLKADOT, WOODLAND CHARLOT...	(CHARLOTTE BAG SUKI DESIGN, STRAWBERRY CHARLOT...	0.021824	0.018483	0.011302	0.517857	28.017319	0.010898	2.035738	0.985822
1396	(WOODLAND CHARLOTTE BAG, STRAWBERRY CHARLOTTE ...	(CHARLOTTE BAG PINK POLKADOT, CHARLOTTE BAG SU...	0.022492	0.018261	0.011302	0.502475	27.516648	0.010891	1.973247	0.985832

Frequent Itemsets:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
17	(BEADED CRYSTAL HEART PINK ON STICK)	(DOTCOM POSTAGE)	0.011469	0.039305	0.011190	0.975728	24.824404	0.010740	39.580626	0.970851
614	(HERB MARKER CHIVES, HERB MARKER THYME)	(HERB MARKER PARSLEY)	0.010411	0.012916	0.010077	0.967914	74.938272	0.009942	30.764113	0.997036
607	(HERB MARKER CHIVES, HERB MARKER ROSEMARY)	(HERB MARKER PARSLEY)	0.010355	0.012916	0.010021	0.967742	74.924917	0.009887	30.599599	0.996977
619	(HERB MARKER CHIVES, HERB MARKER ROSEMARY)	(HERB MARKER THYME)	0.010355	0.012916	0.010021	0.967742	74.924917	0.009887	30.599599	0.996977
1217	(HERB MARKER BASIL, HERB MARKER ROSEMARY, HERB...	(HERB MARKER THYME)	0.010578	0.012916	0.010188	0.963158	74.570009	0.010052	26.792276	0.997137
...	...	...	...	...	...	...	...	...	...	...
25	(RED RETROSPOT CUP)	(BLUE POLKADOT CUP)	0.021378	0.018038	0.010689	0.500000	27.719136	0.010304	1.963924	0.984981
1159	(STRAWBERRY CHARLOTTE BAG, RED RETROSPOT CHARL...	(CHARLOTTE BAG PINK POLKADOT, WOODLAND CHARLOT...	0.026834	0.021824	0.013417	0.500000	22.910714	0.012832	1.956352	0.982723
113	(HAND WARMER RED LOVE HEART)	(HAND WARMER SCOTTY DOG DESIGN)	0.021935	0.030286	0.010968	0.500000	16.509191	0.010303	1.939428	0.960496
147	(LOVE HOT WATER BOTTLE)	(HOT WATER BOTTLE KEEP CALM)	0.025832	0.042701	0.012916	0.500000	11.709257	0.011813	1.914597	0.938850
370	(CHARLOTTE BAG PINK POLKADOT, WOODLAND CHARLOT...	(PACK OF 72 RETROSPOT CAKE CASES)	0.021824	0.071206	0.010912	0.500000	7.021892	0.009358	1.857588	0.876722

4. **Phase 4 (AI Development Part 2)** – The association rules, i.e. confidence and support, are calculated and the results are plotted in a bar graph and scatter plot respectively.

Support of Association Rules

Confidence vs. Support of Association Rules
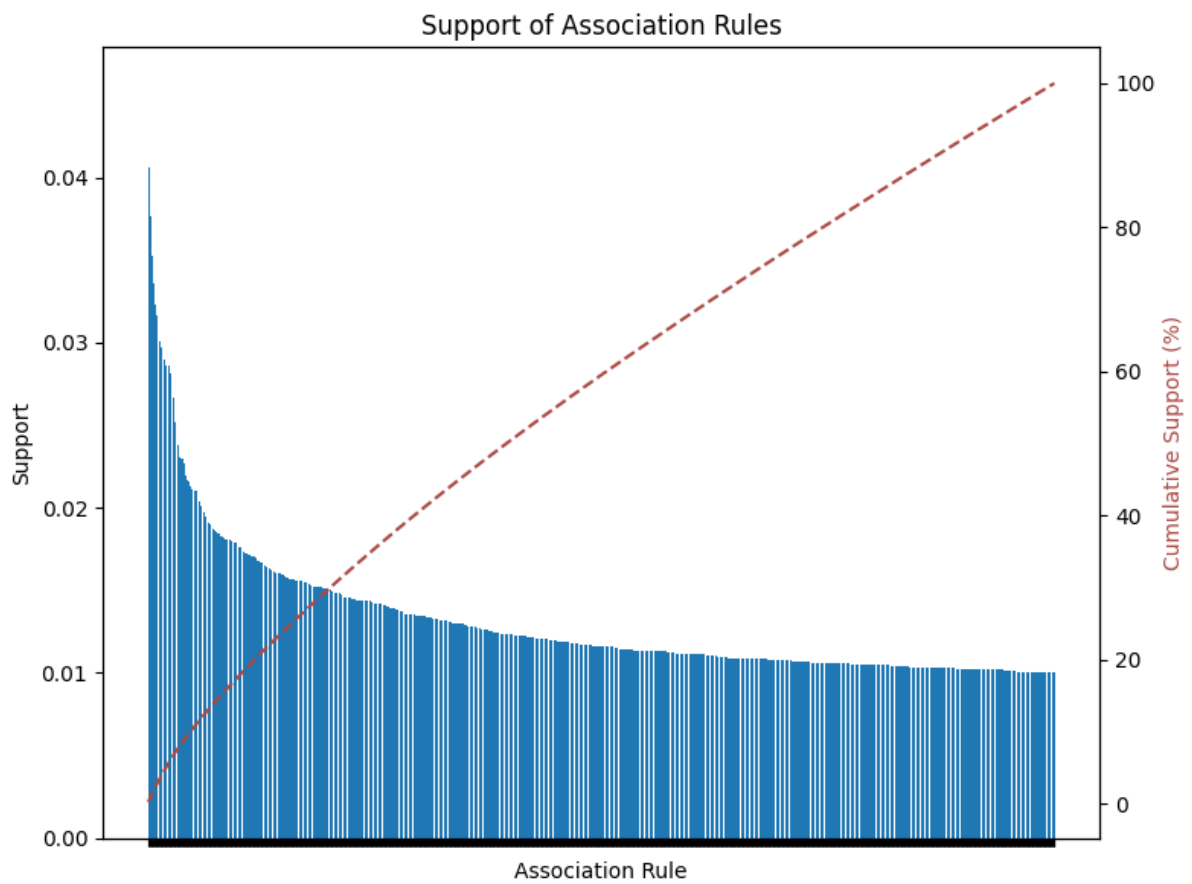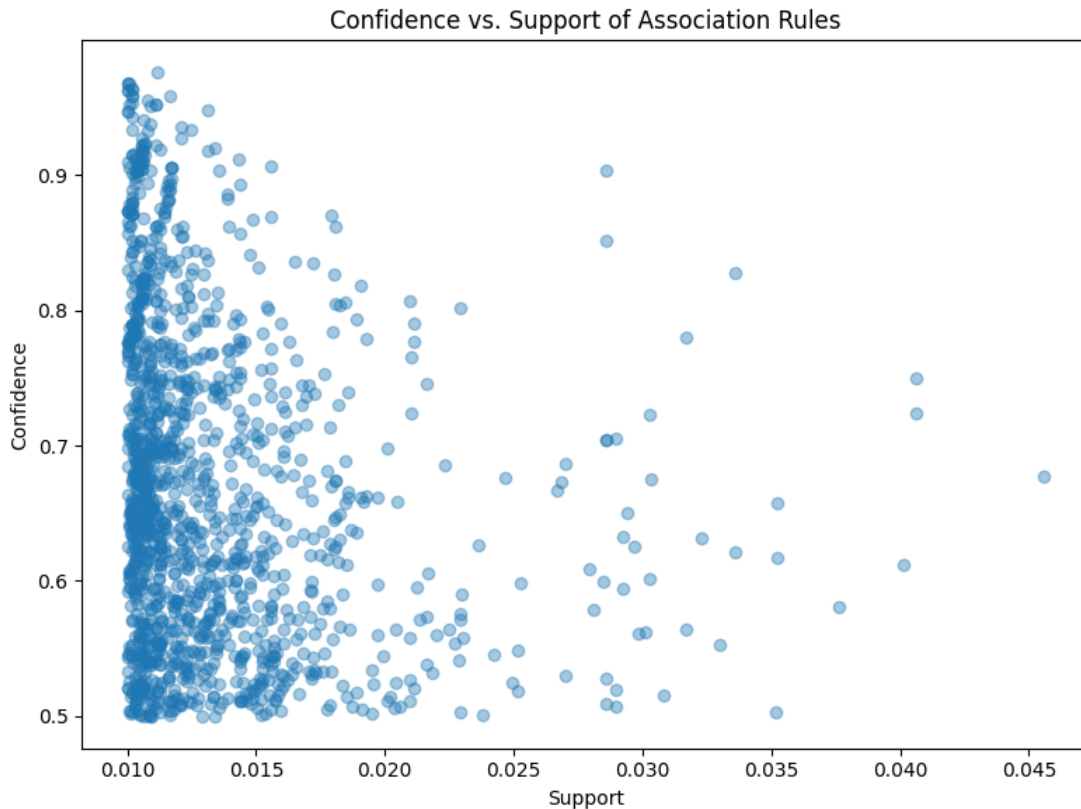
5. **Phase 5 (Project Documentation and Submission)** – Further cross-selling and upselling opportunities are explored and a conclusion is drawn. Now this program is ready to be used.

```
Cross-Selling Recommendations:
Customers who bought 'BEADED CRYSTAL HEART PINK ON STICK' also bought
'DOTCOM POSTAGE'.
Customers who bought 'HERB MARKER THYME' also bought 'HERB MARKER
ROSEMARY'.
Customers who bought 'HERB MARKER ROSEMARY' also bought 'HERB MARKER
THYME'.
Customers who bought 'HERB MARKER CHIVES' also bought 'HERB MARKER
PARSLEY'.
Customers who bought 'REGENCY TEA PLATE PINK' also bought 'REGENCY TEA
PLATE GREEN'.

Upselling Recommendations:
For customers who bought 'HERB MARKER CHIVES', recommend the following
upgrades: HERB MARKER PARSLEY, HERB MARKER THYME.
For customers who bought 'HERB MARKER CHIVES', recommend the following
upgrades: HERB MARKER PARSLEY, HERB MARKER MINT.
For customers who bought 'HERB MARKER CHIVES', recommend the following
upgrades: HERB MARKER ROSEMARY, HERB MARKER PARSLEY.
For customers who bought 'HERB MARKER CHIVES', recommend the following
upgrades: HERB MARKER ROSEMARY, HERB MARKER THYME.
For customers who bought 'HERB MARKER THYME', recommend the following
upgrades: HERB MARKER ROSEMARY, HERB MARKER PARSLEY.
```

For customers who bought 'HERB MARKER CHIVES', recommend the following upgrades: HERB MARKER PARSLEY, HERB MARKER THYME.
For customers who bought 'HERB MARKER THYME', recommend the following upgrades: HERB MARKER ROSEMARY, HERB MARKER PARSLEY.
For customers who bought 'HERB MARKER PARSLEY', recommend the following upgrades: HERB MARKER ROSEMARY, HERB MARKER THYME.
For customers who bought 'HERB MARKER ROSEMARY', recommend the following upgrades: HERB MARKER PARSLEY, HERB MARKER THYME.
For customers who bought 'REGENCY TEA PLATE PINK', recommend the following upgrades: REGENCY TEA PLATE GREEN, REGENCY TEA PLATE ROSES.