📊 Ecommerce - Market Basket Analysis

> 🚀
> ## TL;DR Takeaways:
> - 🚀 **Sales and customer spending peak around Christmas**, but **dip in April and May**. Consider promotions during these months!
> - 🏆 **Top-selling items** include 'PAPER CRAFT, LITTLE BIRDIE' and 'MEDIUM CERAMIC TOP STORAGE JAR'. Ensure they remain in stock!
> - 🛠️ **Leverage the new tool to quickly identify top items to pair in campaigns, optimising cross-selling strategies with ease.**

Initial Exploration 🕵️

In this analysis, a comprehensive exploration of the dataset was conducted to understand its structure and quality:

- **Column Values**: Each column was examined to ascertain its representation and significance.
- **Missing Values**: An assessment was made to identify any data gaps.
- **Data Quality**: The overall quality was scrutinised to detect any anomalies.

Data Cleaning 🧹

To enhance the accuracy of the analysis, the dataset underwent a thorough cleaning process:

- **Removing Duplicates**: Duplicate entries were eliminated to ensure data integrity.

- **Handling Null Values**: Missing data was either filled or removed where necessary.

- **Negative Values**: Negative quantities and prices were excluded, likely indicative of cancelled orders.

- **Non-Product Items**: Items such as postage fees were removed to focus on actual product sales.

Calculating Sales Totals 💰

The total sales amount for each item was calculated using the formula: `Quantity * Unit Price`. This provided a clear view of revenue streams.

Visualisation of Sales Over Time 📈

The sales and average customer spending over time were visualised, revealing the following insights:

- **Sales Trends**: Total sales gradually increased, peaking around Christmas. 🎄

- **Customer Spending**: Similarly, customer spending followed this trend, with a noticeable dip in April and May. 🌸
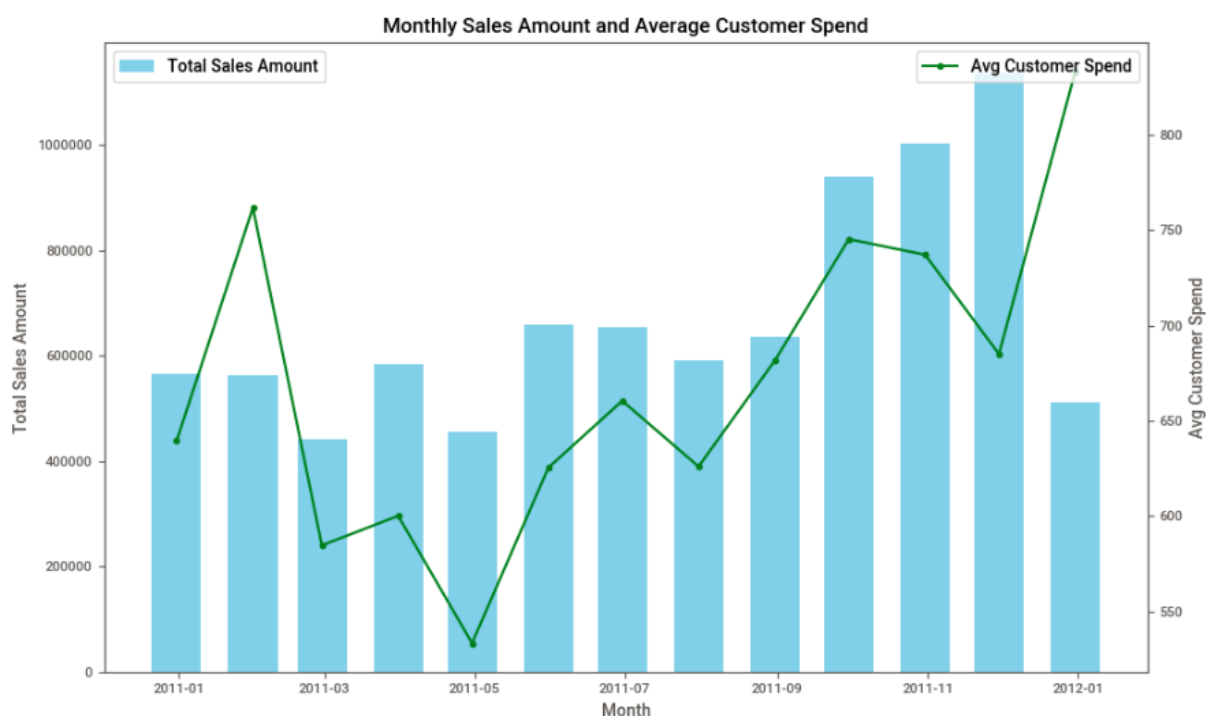
Recommendations:

🚀
- **Addressing Dips**: It is recommended to launch promotions or campaigns in April and May to boost sales during these slower months.

Stock Up on Bestsellers: The top-selling items are:

1. PAPER CRAFT, LITTLE BIRDIE
2. MEDIUM CERAMIC TOP STORAGE JAR
3. WORLD WAR 2 GLIDERS ASSTD DESIGNS
4. JUMBO BAG RED RETROSPOT
5. WHITE HANGING HEART T-LIGHT HOLDER
6. ASSORTED COLOUR BIRD ORNAMENT
7. PACK OF 72 RETROSPOT CAKE CASES
8. POPCORN HOLDER
9. RABBIT NIGHT LIGHT
10. MINI PAINT SET VINTAGE

Revenue Generators: Consider focusing on these top revenue-makers:

1. PAPER CRAFT, LITTLE BIRDIE
2. REGENCY CAKESTAND 3 TIER
3. WHITE HANGING HEART T-LIGHT HOLDER
4. JUMBO BAG RED RETROSPOT
5. MEDIUM CERAMIC TOP STORAGE JAR
6. PARTY BUNTING
7. ASSORTED COLOUR BIRD ORNAMENT
8. RABBIT NIGHT LIGHT
9. CHILLI LIGHTS
10. PAPER CHAIN KIT 50'S CHRISTMAS

Market Basket Analysis 🛒

What is it? 🤔

Market Basket Analysis is a technique used to uncover relationships between products. It's akin to discovering that customers who purchase bread often buy butter as well!

Key Metrics:

- **Support**: Indicates how frequently an itemset appears in the dataset.

- **Confidence**: Measures the likelihood of purchasing a consequent given the antecedent.

- **Lift**: Evaluates the strength of a rule over random chance.

Sample of the Extracted Rules:

- Below is a snapshot of a randomly selected 5 association rules derived from the analysis. These rules highlight key product relationships, helping to identify potential cross-selling opportunities and optimise product placement strategies.

- Each rule is evaluated based on the metrics of support, confidence, and lift, providing insights into their significance and effectiveness.

Antecedents	Consequents	Support	Confidence	Lift
['REGENCY TEA PLATE GREEN ']	['REGENCY TEA PLATE ROSES ']	0.01	0.85	47.62
['LUNCH BAG BLACK SKULL.', 'LUNCH BAG PINK POLKADOT']	['LUNCH BAG WOODLAND']	0.01	0.43	9.58
['HOT WATER BOTTLE TEA AND SYMPATHY']	['HOT WATER BOTTLE I AM SO POORLY']	0.01	0.36	13.27
['CHARLOTTE BAG SUKI DESIGN']	['WOODLAND CHARLOTTE BAG']	0.01	0.41	14.35
['JUMBO STORAGE BAG SKULLS']	['JUMBO BAG PINK POLKADOT']	0.01	0.44	9.29

Rules, Aggregation and Pruning:

- This analysis is crucial for identifying product affinities and optimising cross-selling strategies. However, it can generate an overwhelming number

of rules, necessitating careful selection.

- Generally speaking, rules are aggregated into categorical levels to simplify management and enhance robustness. Pruning is also applied to eliminate insignificant rules.

- To streamline this process, a tool has been added that allows you to input any item—even with a partial name—and receive the top recommended items to pair with in campaigns. This tool leverages metrics like lift to provide quick, actionable insights, eliminating the need for extensive aggregation and pruning. It serves as an efficient resource for marketing and campaign management, offering a rapid go-to solution.

    - Below example showcases when the word *"playhouse"* queried as an item:

```python
recommended_items = recommend_items(results_df, 'playhouse')
recommended_items
```

- **Top 3 Best Campaign Items to Match with 'POPPY'S PLAYHOUSE LIVINGROOM ':**

- POPPY'S PLAYHOUSE KITCHEN, POPPY'S PLAYHOUSE BEDROOM 53.51

- POPPY'S PLAYHOUSE BEDROOM 50.42

- POPPY'S PLAYHOUSE KITCHEN 48.29

- **Top 3 Best Campaign Items to Match with 'POPPY'S PLAYHOUSE KITCHEN':**

- POPPY'S PLAYHOUSE LIVINGROOM 53.51

- POPPY'S PLAYHOUSE BEDROOM 50.42

- POPPY'S PLAYHOUSE LIVINGROOM , POPPY'S PLAYHOUSE BEDROOM 48.29

- **Top 3 Best Campaign Items to Match with 'POPPY'S PLAYHOUSE BEDROOM ':**

- POPPY'S PLAYHOUSE LIVINGROOM 53.51

- POPPY'S PLAYHOUSE LIVINGROOM , POPPY'S PLAYHOUSE KITCHEN 50.42

- POPPY'S PLAYHOUSE KITCHEN 48.29

Recommendation:

🚀 **Utilise this tool to enhance your marketing campaigns, as it provides immediate and strategic insights, helping you to optimise cross-selling opportunities effectively.**