جامعة أم القرى
UMM AL-QURA UNIVERSITY

Data Analysis 2

Task2

Dataset:

E-commerce retail

Student Name	ID
Mawaddah Alharthi	443004953
Bayan Alqarni	443002589

Market Basket Analysis (E-commerce retail)

Dataset Overview

The dataset used in this project involves customer transactions. It contains several columns essential for understanding sales patterns, including:

- CustomerID: Unique identifier for customers.
- InvoiceDate: Date and time of each transaction.
- Description: Product description for items sold.
- Quantity: Number of units purchased per product.

This dataset provides insights into the items purchased, transaction frequency, and trends over time.

Data Preprocessing and Cleaning

Handling Date and Time:

The InvoiceDate column was converted into datetime objects to extract useful features such as month and day of the week. New columns were created for month and day name to explore trends over time.

Handling Duplicates and Grouping:

We grouped transactions based on CustomerID, InvoiceDate, and product Description to avoid duplicate entries. A unique Transaction ID was generated for each transaction to simplify further analysis.

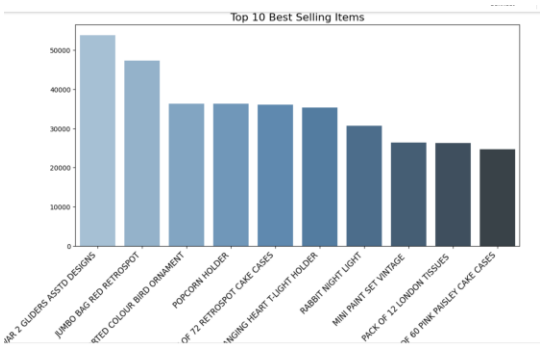
Encoding Data:

A pivot table was created to structure the data, with Transaction IDs as rows and product descriptions as columns. Each entry in the table contains the count of a product in a transaction.

Exploratory Data Analysis (Visualizations)

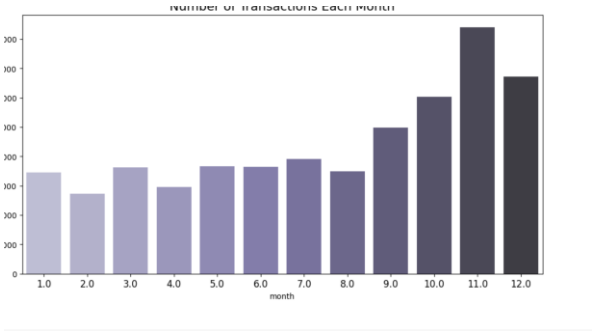
Top 10 Best Selling Items:

A bar chart visualized the most popular products by total quantity sold. This helps in identifying key products driving sales.



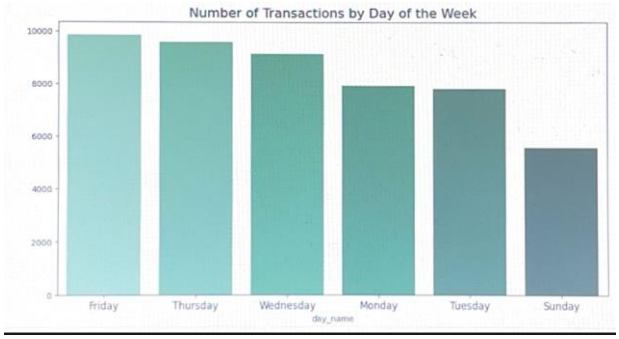
Monthly Transaction Trends:

A bar plot showed the number of transactions for each month, highlighting peak shopping seasons or trends over time. This can help plan promotions or manage inventory during busy months.



Transaction Patterns by Day of the Week:

An additional bar plot explored transaction frequency by weekday. This insight can help businesses adjust staffing and operational hours.



Model Used: Apriori Algorithm for Association Rules

We used the Apriori algorithm to extract frequent itemsets and generate association rules. This technique is part of Market Basket Analysis and identifies relationships between items commonly purchased together.

Evaluation Metrics

Support: Fraction of transactions containing the itemset.
Confidence: Likelihood that the consequent is purchased given the antecedent.
Lift: How much more likely the consequent is bought when the antecedent is present, compared to random chance.

Association Rules

Using the association_rules function on the frequent itemsets, we extracted rules identifying the strongest relationships between products based on Lift values greater than 0.5. The results were filtered to show only rules with the highest Confidence.

	antecedents	consequents	support	confidence	lift
678	(POPPY'S PLAYHOUSE BATHROOM, POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE BEDROOM)	0.011242	1.0	35.581818
803	(POPPY'S PLAYHOUSE BATHROOM, POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE BEDROOM)	0.011242	1.0	35.581818
845	(WHITE HANGING HEART T-LIGHT HOLDER, WOODEN FR...	(WOODEN PICTURE FRAME WHITE FINISH)	0.010731	1.0	25.415584
670	(POPPY'S PLAYHOUSE BEDROOM , POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE KITCHEN)	0.012775	1.0	34.946429
672	(POPPY'S PLAYHOUSE BATHROOM, POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE BEDROOM)	0.012775	1.0	35.581818
775	(ALARM CLOCK BAKELIKE RED , ALARM CLOCK BAKELI...	(ALARM CLOCK BAKELIKE GREEN)	0.010220	1.0	20.819149
683	(POPPY'S PLAYHOUSE BATHROOM, POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE KITCHEN)	0.011242	1.0	34.946429
808	(POPPY'S PLAYHOUSE BATHROOM, POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE BEDROOM , POPPY'S PLAYHOUSE ...	0.011242	1.0	44.477273
527	(ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI...	(ALARM CLOCK BAKELIKE GREEN)	0.014819	1.0	20.819149
801	(POPPY'S PLAYHOUSE BEDROOM , POPPY'S PLAYHOUSE...	(POPPY'S PLAYHOUSE KITCHEN)	0.011242	1.0	34.946429

The findings highlighted that some items, such as POPPY'S PLAYHOUSE BATHROOM and POPPY'S PLAYHOUSE BEDROOM, have a strong association, often purchased together with a confidence level of 1.0 and a lift of 35.58. These insights can support marketing recommendations and cross-selling strategies by pinpointing frequently purchased product combinations.

Real-World Applications

Product Recommendation Systems:

Use identified rules to recommend products to customers during checkout (e.g., 'Customers who bought X also bought Y').

Cross-Selling and Upselling:

Suggest complementary products to boost sales and customer satisfaction.

Inventory Management:

Stock items that are often purchased together to prevent stockouts and ensure smoother operations.

Targeted Marketing Campaigns:

Create promotions based on commonly purchased bundles to attract more customers.

Conclusion

The Apriori algorithm successfully revealed patterns in the transaction data, providing valuable insights for product bundling, inventory planning, and personalized marketing. The visualizations reinforced these insights by highlighting sales trends across time and product categories. Future work could involve testing these rules in a real-world setting to evaluate their effectiveness in driving sales and improving customer experience.