جامعة أم القرى
UMM AL-QURA UNIVERSITY

Data Analysis 2

Market Basket Analysis

Dataset:

OnlineRetail

Student Name	ID
Alaa Khalil Bondagji	443000351
Jana Maher Jamal	443010542

Market Basket

Overview

This project analyzed an online retail dataset to find patterns in customer purchases. We identified frequently bought items and relationships between products using the FP-Growth algorithm and association rules. These findings can help businesses with product bundling, targeted marketing, and improving inventory management.

1. Top 10 Best-Selling Items

In our analysis of the OnlineRetail.csv dataset, we used the total quantity sold to determine the Top 10 Best-Selling Items. By calculating the total quantity for each product, we identified the most popular items among consumers. These insights help improve inventory management by ensuring that high-demand products are always stocked. Additionally, businesses can use this information for targeted marketing campaigns that focus on promoting these popular items to maximize sales.

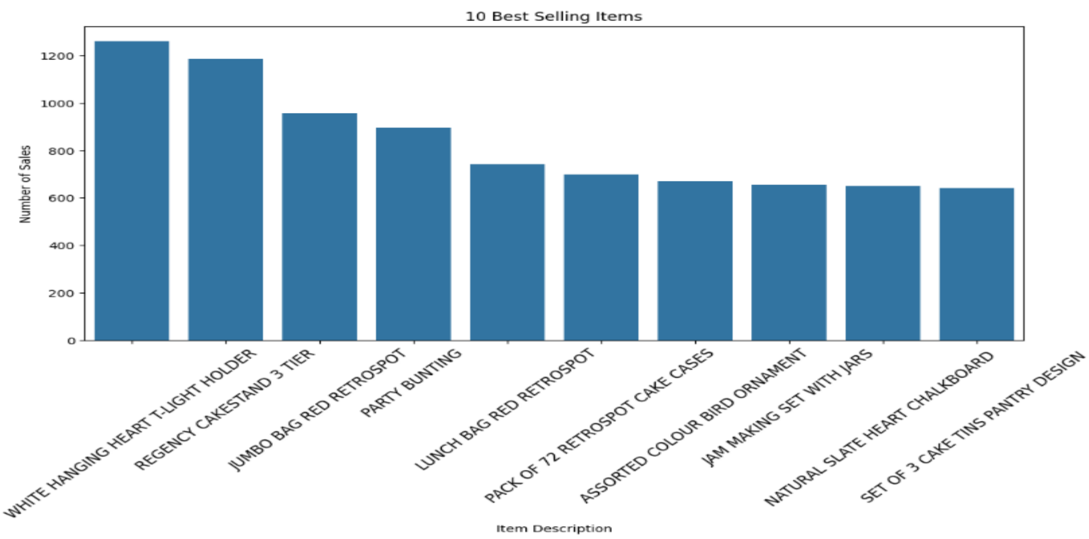
```
import pandas as pd
from mlxtend.frequent_patterns import fpgrowth
from google.colab import drive

(Basket Format)
basket =
data_cleaned.groupby('InvoiceNo')['Description'].apply(list).reset_in
()dex

basket_encoded =
('|')basket['Description'].str.join('|').str.get_dummies

frequent_itemsets = fpgrowth(basket_encoded, min_support=0.01,
use_colnames=True)

frequent_itemsets.sort_values('support', ascending=False).head(10)
```



2. Monthly Trends: Sales

We analyzed the OnlineRetail.csv data to track monthly sales trends by counting transactions per month. This helps identify peak sales periods, allowing businesses to better plan their marketing and inventory for high-demand months.

```
data['InvoiceDate'] = pd.to_datetime(data['InvoiceDate'],
format='%d/%m/%Y %H:%M', errors='coerce')

Add 'year', 'month', 'day', and 'day_name' columns #
data['year'] = data['InvoiceDate'].dt.year
data['month'] = data['InvoiceDate'].dt.month
data['day'] = data['InvoiceDate'].dt.day
()data['day_name'] = data['InvoiceDate'].dt.day_name


import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12,6))

,sns.barplot(x=data['month'].value_counts().index
(y=data['month'].value_counts().values

Customize the plot #
plt.xticks(size=12)
plt.title('Number of Transactions Each Month')
plt.xlabel('Month')
plt.ylabel('Number of Transactions')

Display the plot #
plt.show()
```

Summary of Discovered Rules:

Using the FP-Growth algorithm and association rules from the dataset, several key patterns were discovered in the purchasing behavior of customers:

Frequent Itemsets:

- Certain products are frequently purchased together. For example, the product WHITE HANGING HEART T-LIGHT HOLDER appears in more than 11% of the transactions, often paired with other household decor items.
- Similarly, items like REGENCY CAKESTAND 3 TIER and JUMBO BAG RED RETROSPOT show high support, indicating they are popular items frequently bought together.

Association Rules:

- The rule `REGENCY TEA PLATE ROSES → REGENCY TEA PLATE GREEN` has a **confidence** of 94.85% and a **lift** of 59.97. This indicates that customers who purchase the REGENCY TEA PLATE ROSES are very likely to also purchase the REGENCY TEA PLATE GREEN.

- Another strong rule shows that customers who buy CHARLOTTE BAG PINK POLKADOT are highly likely to purchase RED RETROSPOT CHARLOTTE BAG, with a confidence of 90.47% and a lift of 21.03.

5. Association Rules

Using the FP-Growth algorithm, we identified products frequently bought together in the OnlineRetail.csv dataset. These insights help businesses bundle products, enhance cross-selling, and improve marketing strategies to boost sales.

```
import pandas as pd
from mlxtend.frequent_patterns import fpgrowth, association_rules

frequent_itemsets = fpgrowth(basket_encoded, min_support=0.01,
use_colnames=True)

rules = association_rules(frequent_itemsets, metric='lift',
min_threshold=0.5)


top_rules = rules[['antecedents', 'consequents', 'support',
'confidence', 'lift']].sort_values('confidence',
(١٠)ascending=False).head

top_rules['antecedents'] = top_rules['antecedents'].apply(lambda x:
', '.join(list(x)))
top_rules['consequents'] = top_rules['consequents'].apply(lambda x:
', '.join(list(x)))

import pandas as pd
from IPython.display import display

# عرض النتائج
display(top_rules)
```

Real-World Applications

Product Bundling:

Retailers can use the discovered frequent itemsets to create product bundles that align with customer purchasing behavior. For instance, bundling REGENCY TEA PLATES in different colors could increase sales by offering attractive combinations that customers already tend to buy together.

Targeted Marketing and Recommendations:

The association rules can be used to implement recommendation systems. For example, if a customer buys the CHARLOTTE BAG PINK POLKADOT, the system could automatically suggest the RED RETROSPOT CHARLOTTE BAG due to their high confidence of being purchased together.

Inventory Management:

By understanding which items are frequently bought together, businesses can optimize their inventory to ensure these items are stocked together, reducing the likelihood of running out of popular combinations during peak shopping periods..

Promotional Strategies:

The analysis can be used to inform marketing campaigns. For instance, offering discounts on frequently paired items such as the REGENCY TEA PLATE ROSES and REGENCY TEA PLATE GREEN can encourage customers to buy both items, boosting overall sales.

Conclusion

The rules discovered through frequent itemset mining and association rule analysis provide valuable insights into customer purchasing behaviors. These insights can be leveraged for effective product bundling, inventory management, targeted marketing, and promotional strategies, ultimately enhancing customer satisfaction and business profitability.