جامعة أم القرى
UMM AL-QURA UNIVERSITY

Data Analysis Project

"Market Basket REPORT"

NAME:	ID:
Wed Alshehri	444015020
Ftoon Alshmaimri	444003233

Dr. Omima Fallatah

1. Data Preparation and Cleaning

1. **Data Loading and Initial Exploration**:
    - Multiple datasets, including `items`, `pay`, `pro`, `order`, and `catname`, were loaded. These files contained detailed information about order items, payments, products, orders, and product categories respectively.
2. **Identifying Missing Values**:
    - Missing values were analyzed in each dataset. Missing values in essential fields, such as `order_id` in `order`, were handled by dropping affected rows, while missing values in `pay` were filled with zeros to preserve data integrity.
3. **Duplicate Removal**:
    - Duplicate rows in each dataset were identified and removed to prevent data redundancy.
4. **Dataset Merging**:
    - The datasets were merged using `order_id` as a key to obtain a comprehensive view, connecting product, order, and customer information.

2. Exploratory Data Analysis (EDA)

1. **Top-Selling Products**:
    - A merged dataset with English-translated product category names was created, and the top 10 best-selling products were identified.
    - Visualization: A bar chart showcased the sales counts of these top products, indicating product popularity and potential customer preferences.
2. **Monthly Sales Analysis for 2017 and 2018**:
    - Order data was filtered by year to analyze monthly sales trends in 2017 and 2018.
    - Visualization: Monthly sales were visualized using bar charts. The month with the highest sales in each year was highlighted, providing insights into seasonal demand trends.
3. **Annual Sales Distribution**:
    - The sales distribution between 2017 and 2018 was illustrated using a pie chart, revealing the proportion of orders by year and indicating overall sales growth or decline.

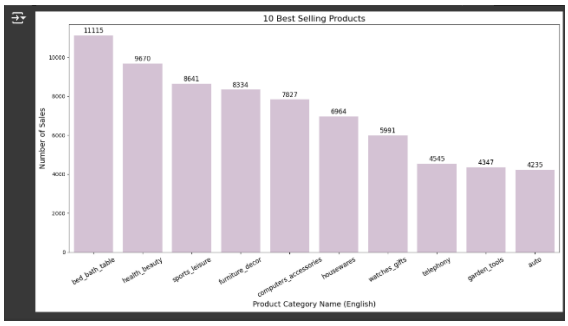
3. Market Basket Analysis and Association Rules

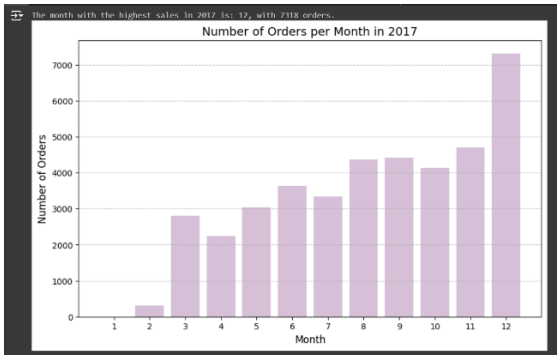
1. **Data Transformation for Association Analysis**:
    - A pivot table for `customer_id` and `product_id` was created, enabling a one-hot encoded format suitable for association rule mining.
    - A binary matrix (with 1 indicating purchase and 0 otherwise) was developed to facilitate analysis using the FP-growth and Apriori algorithms.
2. **Frequent Itemsets Extraction**:
    - Using Apriori and FP-growth, frequent itemsets were extracted with a minimum support threshold of 1% initially, and then refined to 0.5%.
    - The top 10 itemsets based on support were displayed to understand common product bundles and purchasing patterns.
3. **Association Rule Mining**:
    - Association rules were derived from the frequent itemsets using `lift` as the metric. Key association rules with high confidence and lift values were identified, showcasing strong relationships between products.
    - The top 10 association rules provided insights into products that are often purchased together, valuable for cross-selling strategies.

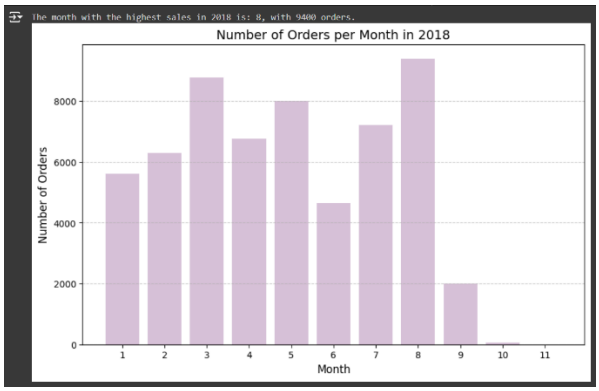
4. Summary and Key Insights

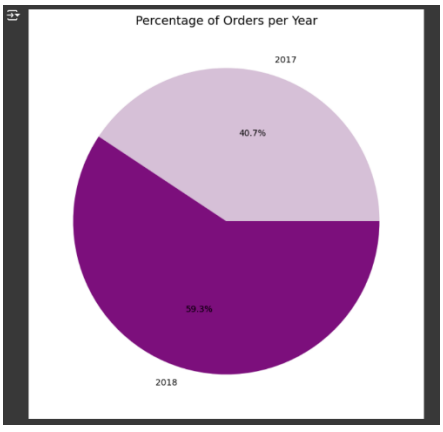
1. **Top Products and Customer Preferences**:
    - Analysis of top products provided insights into customer preferences, indicating that specific categories consistently had high sales volumes.
2. **Seasonal Sales Patterns**:
    - Monthly and yearly analysis of sales trends suggested that demand might be influenced by specific times of the year, useful for inventory management and promotional timing.
3. **Product Associations**:
    - Market basket analysis highlighted frequent product associations, revealing which products are likely to be bought together. These associations can inform targeted marketing and upsell opportunities.

5. Visualization









6. Conclusion

The analysis effectively combined EDA and association rule mining to offer actionable insights into customer behavior, product popularity, and seasonal trends. This data-driven approach can support strategic decision-making, including inventory planning, marketing, and personalized recommendations.

جامعة أم القرى

UMM AL-QURA UNIVERSITY