Customers segmentation

• • •

September 30, 2025

Overview

This project is an online retail data analysis using SQL and Python. The data is stored in an SQLite database, and queries are executed directly from Python with subsequent visualization of the results using the Plotly. The main goal of the project is to answer key business questions related to customer behavior and sales, and to present the findings in a clear, visual format.

Business questions addressed in this analysis

Q 1

What is the distribution of order values across all customers in the dataset?

Q 2

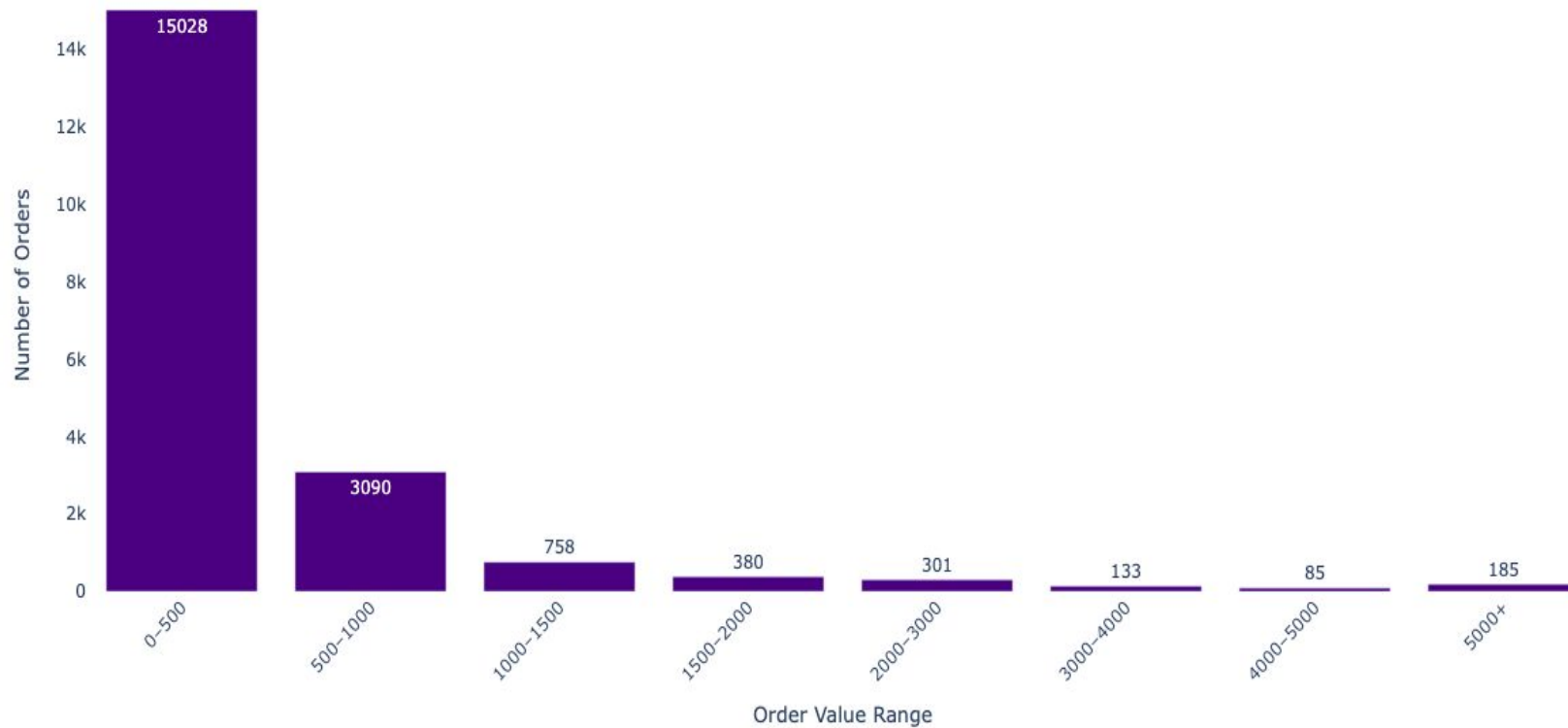How many unique products has each customer purchased?

Q 3

How many customers have only made a single purchase from the company?

Q 4

Which products are most commonly purchased together by customers in the dataset?

Distribution of order values across all customers

Interactive Order Value Distribution

Chart: Bar chart showing Number of Orders by Order Value Range

Order Value Range	Number of Orders
0–500	15028
500–1000	3090
1000–1500	758
1500–2000	380
2000–3000	301
3000–4000	133
4000–5000	85
5000+	185

The range is highly skewed: most orders are small, but there are a few very large ones.
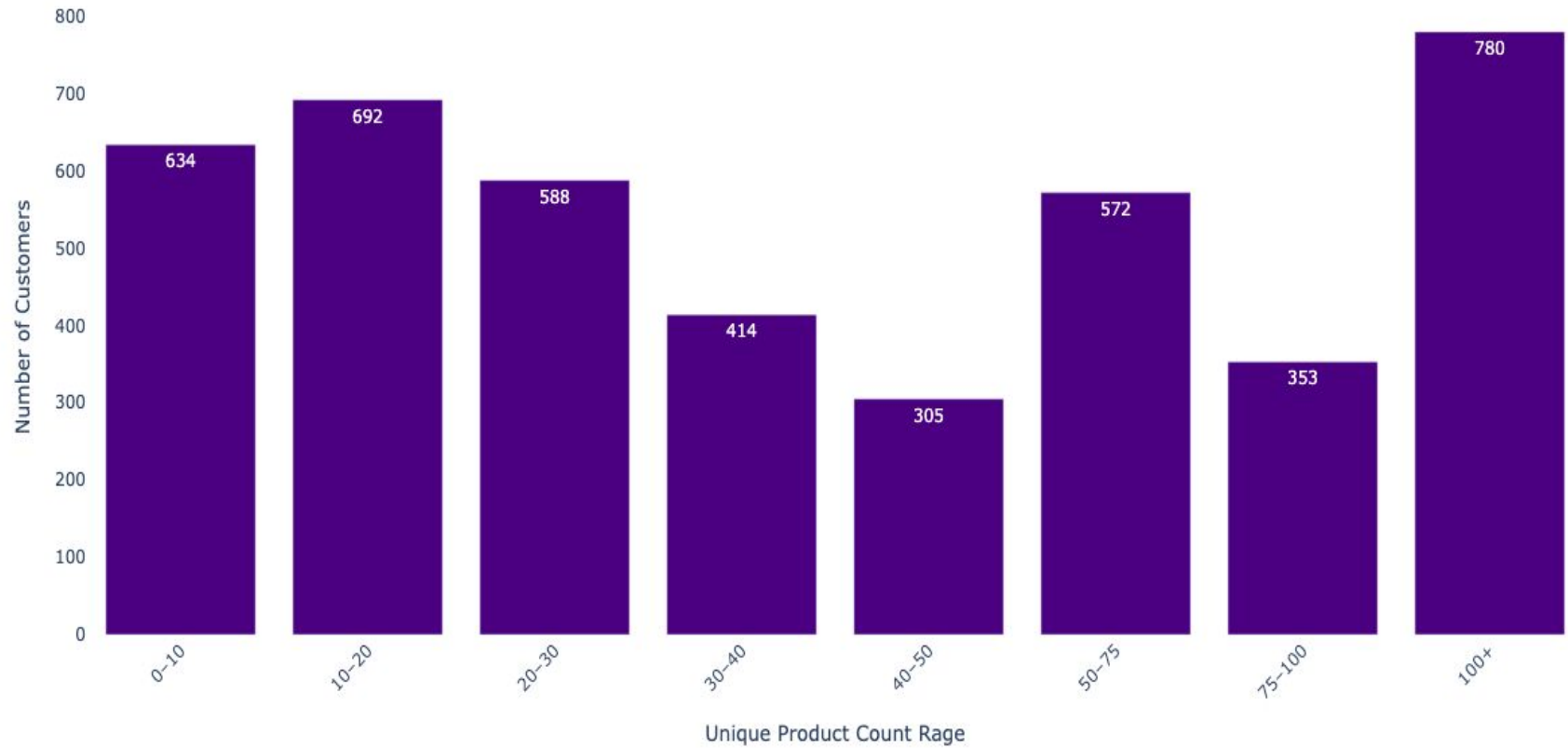
Outliers (extremely large orders) likely represent wholesale or corporate clients.

The histogram/boxplot shows a strong long-tail distribution.

Conclusion: Sales follow a long-tail pattern — a small number of large orders significantly impact total revenue.

Number of unique products purchased by customers

Number of Unique Products Purchased by Customers

Most customers purchase only a small number of distinct products (1–10).

A smaller group of customers buy a wide variety of products (tens or even hundreds).

Conclusion: Customers can be split into "niche buyers" (low diversity, repeat the same products) and "variety seekers" (broad range of products).

Customers by number of purchases

Proportion of Customers by Number of Purchases

Legend:
- Multiple Purchases
- Single Purchase

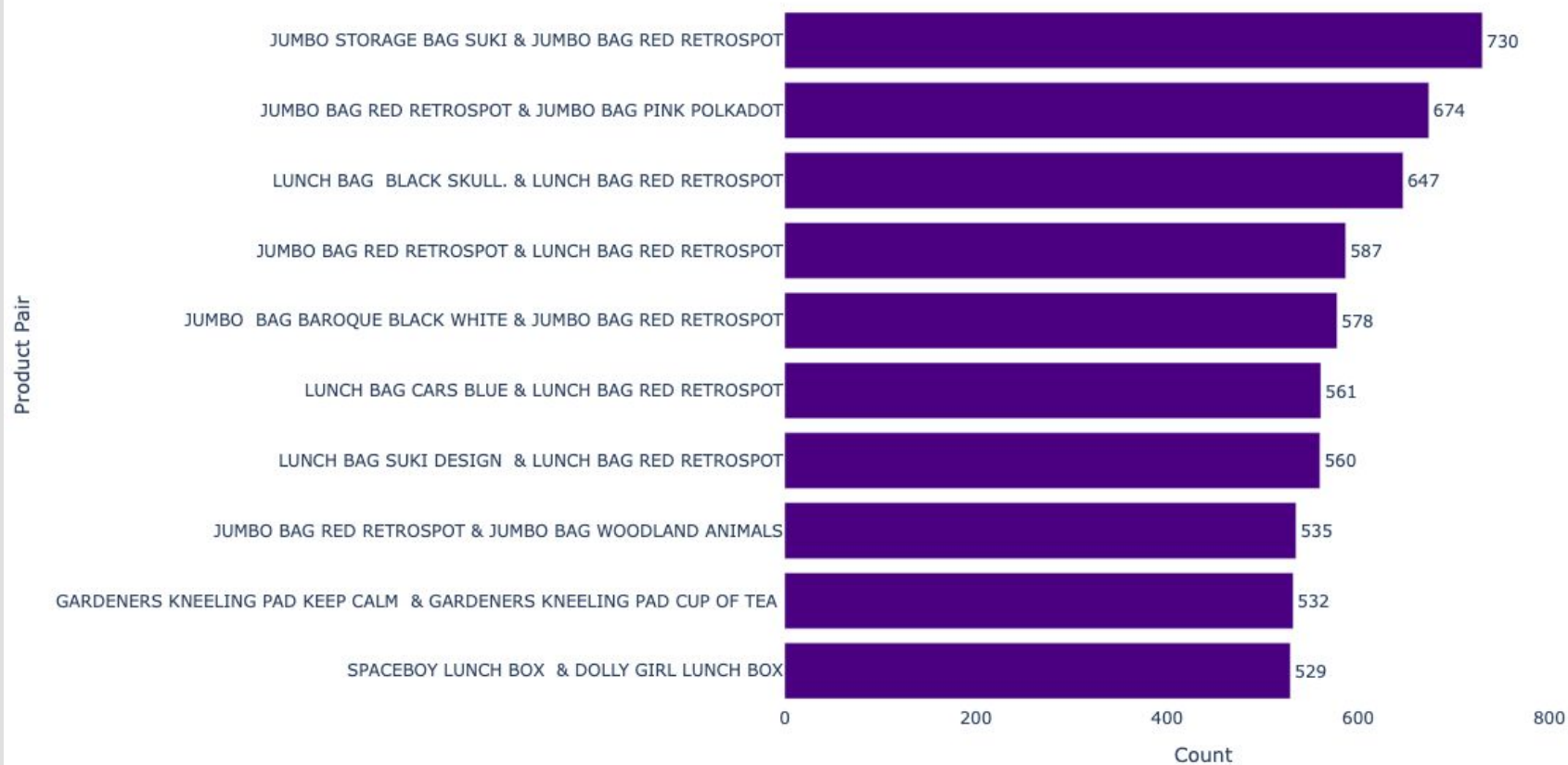Single Purchase
34.4%

Multiple Purchases
65.6%
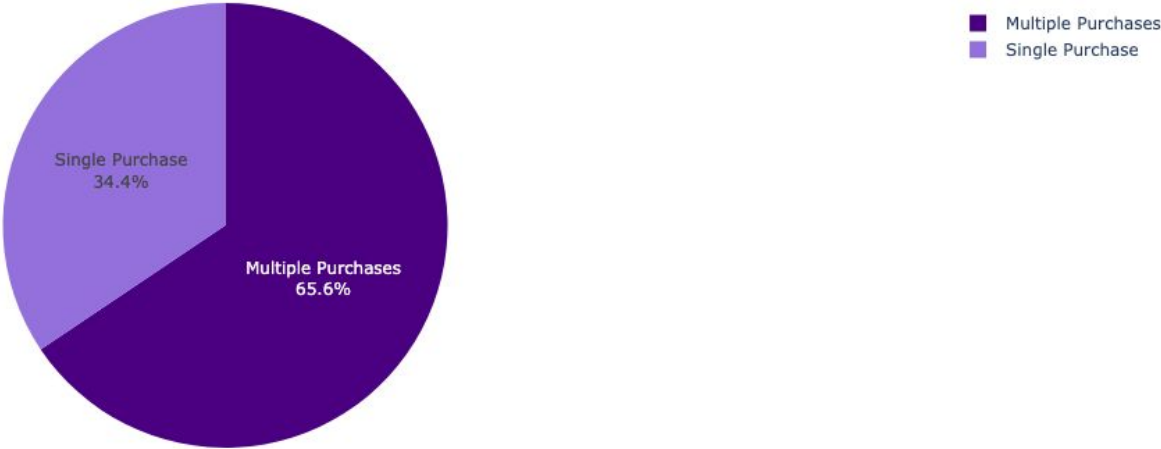
A significant share of customers falls into this category (commonly 40–60% in retail).

This is a critical group for customer retention strategies.

Conclusion: A high percentage of one-time buyers suggests challenges with repeat sales and indicates room for retention improvement.

Most commonly purchased products pair

Top 10 Most Commonly Purchased Product Pairs

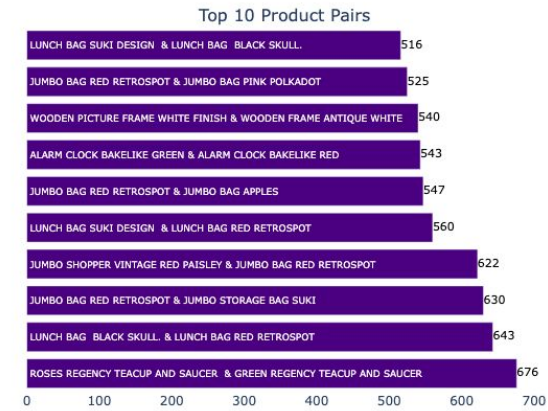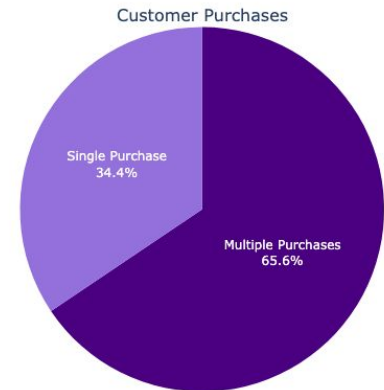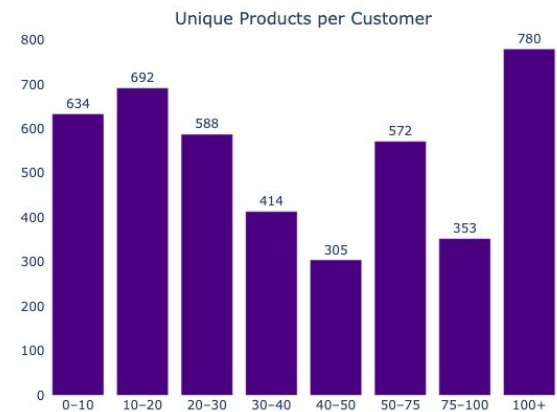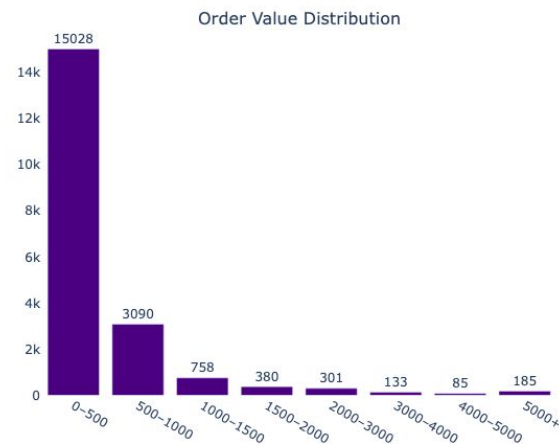| Product Pair	Count
JUMBO STORAGE BAG SUKI & JUMBO BAG RED RETROSPOT	730
JUMBO BAG RED RETROSPOT & JUMBO BAG PINK POLKADOT	674
LUNCH BAG BLACK SKULL. & LUNCH BAG RED RETROSPOT	647
JUMBO BAG RED RETROSPOT & LUNCH BAG RED RETROSPOT	587
JUMBO BAG BAROQUE BLACK WHITE & JUMBO BAG RED RETROSPOT	578
LUNCH BAG CARS BLUE & LUNCH BAG RED RETROSPOT	561
LUNCH BAG SUKI DESIGN & LUNCH BAG RED RETROSPOT	560
JUMBO BAG RED RETROSPOT & JUMBO BAG WOODLAND ANIMALS	535
GARDENERS KNEELING PAD KEEP CALM & GARDENERS KNEELING PAD CUP OF TEA	532
SPACEBOY LUNCH BOX & DOLLY GIRL LUNCH BOX	529

The top-20 pairs reveal strong associations between certain items.

This knowledge can guide cross-selling and product placement strategies.

Conclusion: Frequent co-purchases highlight natural product bundles and opportunities for recommendation systems.

Final dashboard

Customer Segmentation Dashboard

Order Value Distribution

Range	Count
0~500	15028
500~1000	3090
1000~1500	758
1500~2000	380
2000~3000	301
3000~4000	133
4000~5000	85
5000+	185

Unique Products per Customer

Range	Count
0~10	634
10~20	692
20~30	588
30~40	414
40~50	305
50~75	572
75~100	353
100+	780

Customer Purchases

Single Purchase 34.4%

Multiple Purchases 65.6%

Top 10 Product Pairs

Product Pair	Count
LUNCH BAG SUKI DESIGN & LUNCH BAG BLACK SKULL.	516
JUMBO BAG RED RETROSPOT & JUMBO BAG PINK POLKADOT	525
WOODEN PICTURE FRAME WHITE FINISH & WOODEN FRAME ANTIQUE WHITE	540
ALARM CLOCK BAKELIKE GREEN & ALARM CLOCK BAKELIKE RED	543
JUMBO BAG RED RETROSPOT & JUMBO BAG APPLES	547
LUNCH BAG SUKI DESIGN & LUNCH BAG RED RETROSPOT	560
JUMBO SHOPPER VINTAGE RED PAISLEY & JUMBO BAG RED RETROSPOT	622
JUMBO BAG RED RETROSPOT & JUMBO STORAGE BAG SUKI	630
LUNCH BAG BLACK SKULL. & LUNCH BAG RED RETROSPOT	643
ROSES REGENCY TEACUP AND SAUCER & GREEN REGENCY TEACUP AND SAUCER	676

Overall conclusions

Sales are unevenly distributed: dominated by many small orders and a few very large ones.

A large proportion of customers are one-time buyers → retention is a major growth opportunity.

Strong product pairings exist → useful for cross-selling and recommendations.

Customer segmentation by product diversity shows different shopping behaviors worth targeting separately.

———