

INSTACART BASKET ANALYSIS

BACKGROUND

Instacart, an online grocery delivery service, aims to better understand customer purchasing behaviors to enhance its targeted marketing.

OBJECTIVE

This project goal is to conduct exploratory analysis to uncover insights and recommend strategies for enhancing market segmentation.

DATASET

This project utilized multiple open-source datasets from Instacart, including customer profiles, order histories, product information, and department classifications. The data was sourced from Kaggle and is part of the “Instacart Online Grocery Shopping Dataset 2017,” which is publicly available at

<https://www.instacart.com/datasets/grocery-shopping-2017>

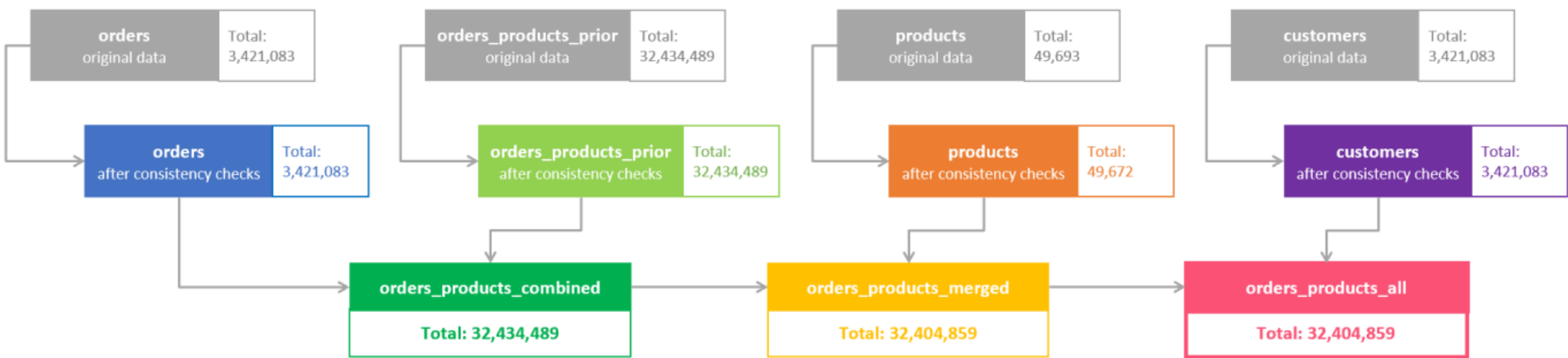
KEY BUSINESS QUESTIONS

- What are the busiest days of the week and hours of the day in terms of order volume?
- Are there specific times of day when customers tend to spend the most money?
- Are there certain types of products that are more popular than others?
- Are there differences in ordering habits based on a customer's loyalty status?
- Are there differences in ordering habits based on a customer's region?
- Is there a connection between age and family status in terms of ordering habits?
- What differences can you find in ordering habits of different customer profiles such as order price, ordering frequency, and the types of products purchased?

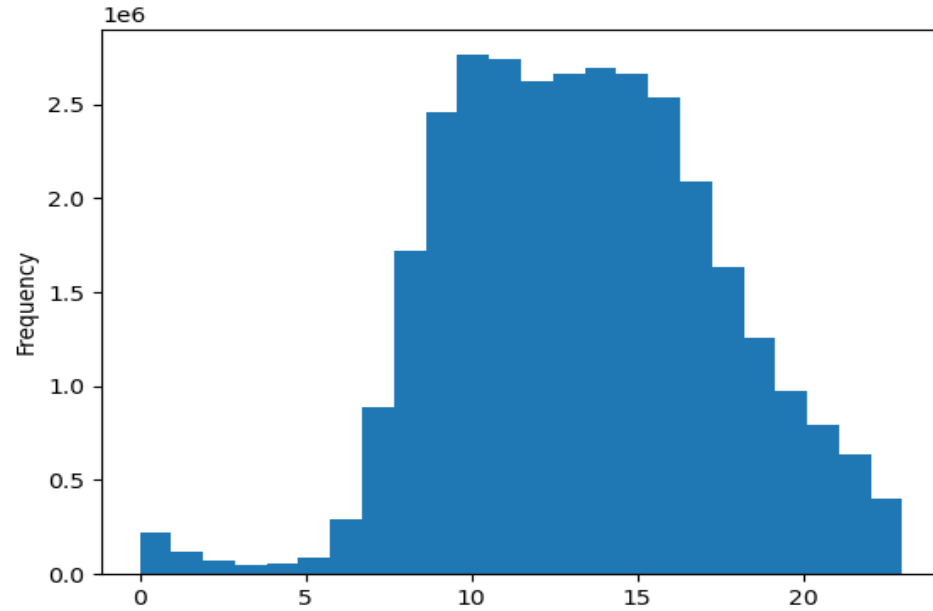
DATA CLEANING

To conduct the exploratory analysis, several datasets were cleaned, merged, and transformed to create a unified dataset that reflects different dimensions of customer behavior.

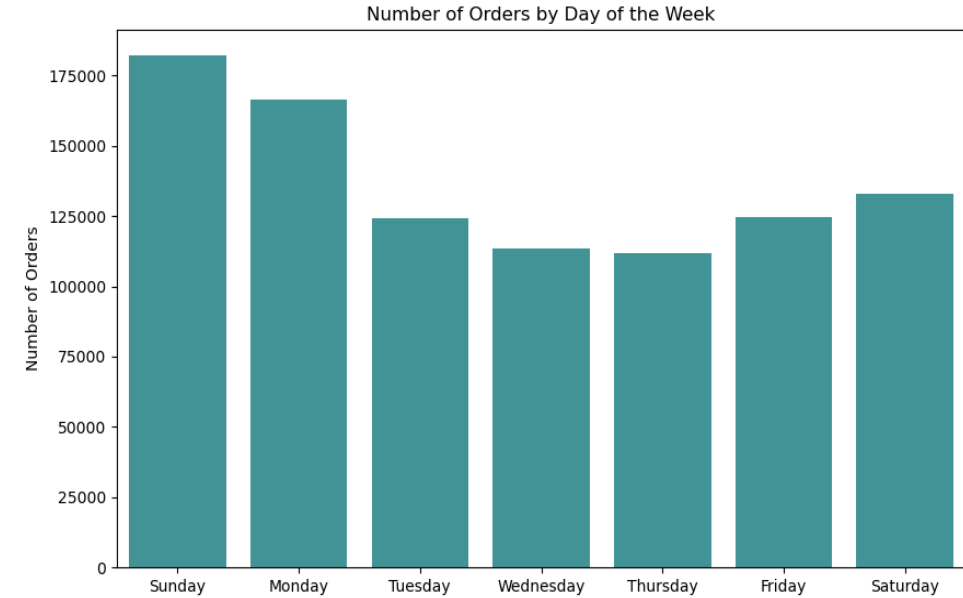
The figure below illustrates the ‘Population Flow’, outlining the sequential process of converting raw data into integrated tables while maintaining consistency and alignment across sources.



DAILY ORDER FREQUENCY & BUSIEST DAYS OF THE WEEK

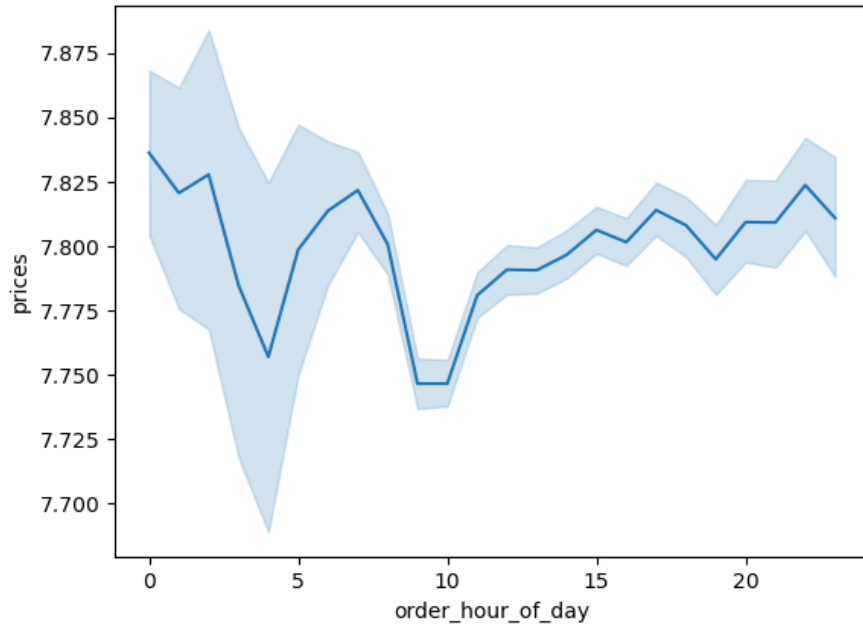


The plot illustrates the distribution of orders across different hours of the day. Order frequency is very low around midnight, indicating minimal activity during that time. Starting around 7 to 8 a.m., the number of orders begins to rise, reaching its peak at 10 a.m. This high order volume continues steadily through the afternoon, up until around 5 p.m. (17:00). After that, the number of orders gradually declines toward the end of the day.



Additionally, the highest number of ordering occurs over the weekend followed by Monday, suggesting that customers tend to place larger or more considered orders when they have more free time and when they start their first day of the week.

DAILY EXPENDITURE

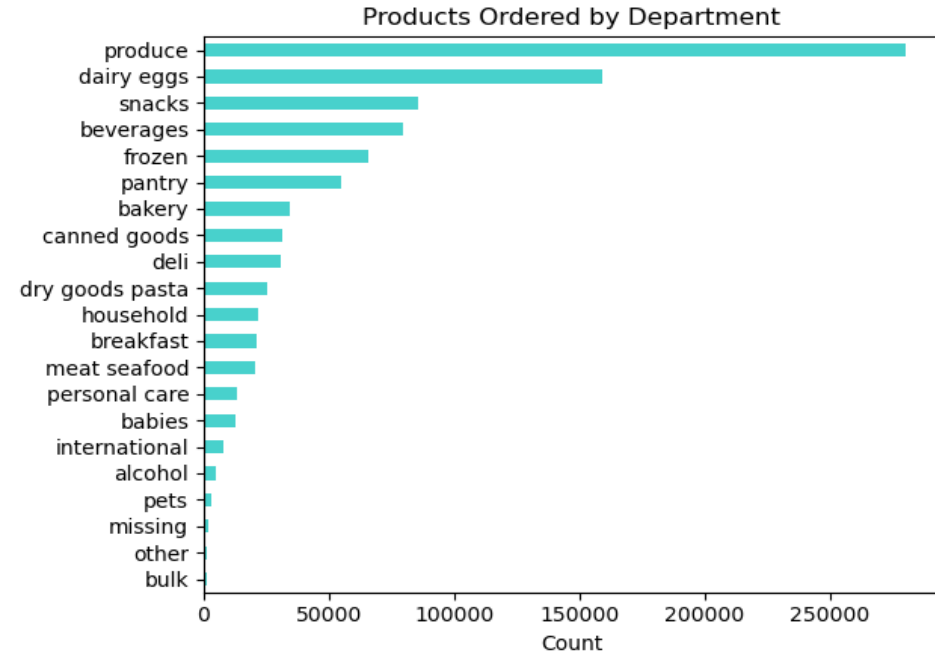


There is a difference in the expenditure depending on the hour of the day.

Customer payments drop significantly between 12:00 AM and 5:00 AM, then gradually increase from 5:00 to 8:00 AM before declining again until 10:00 AM.

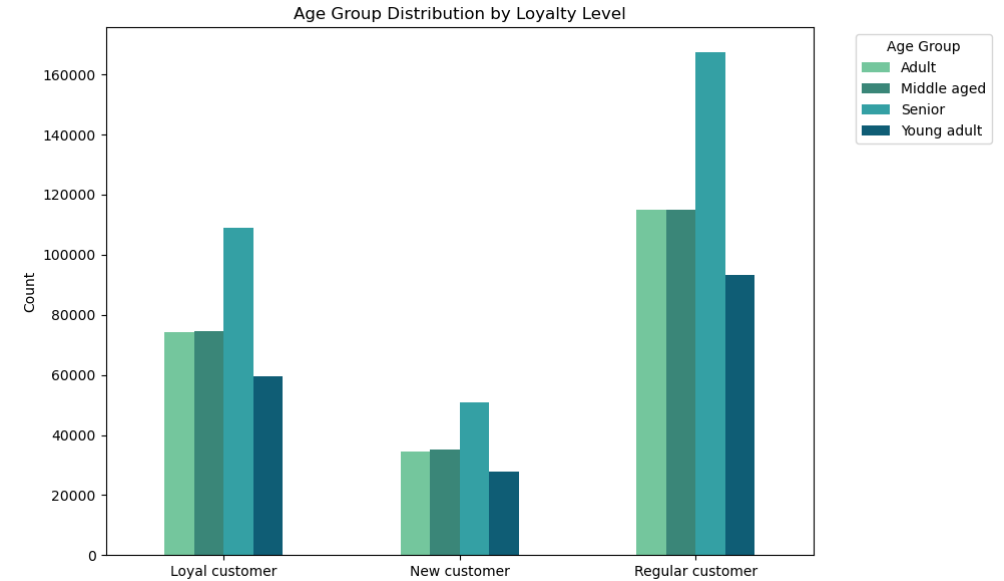
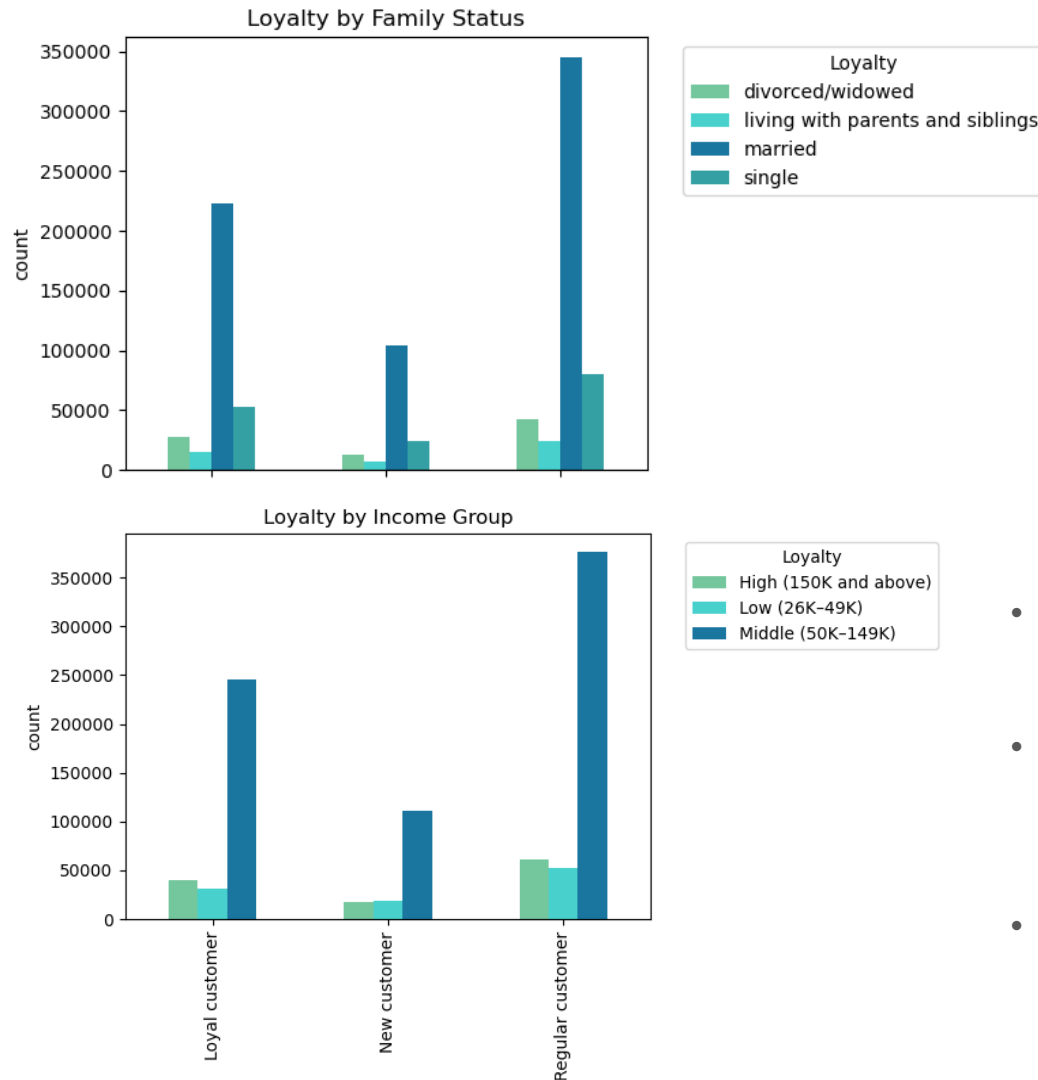
The lowest payment activity is observed around 5:00 AM and 10:00 AM.

PRODUCT POPULARITY



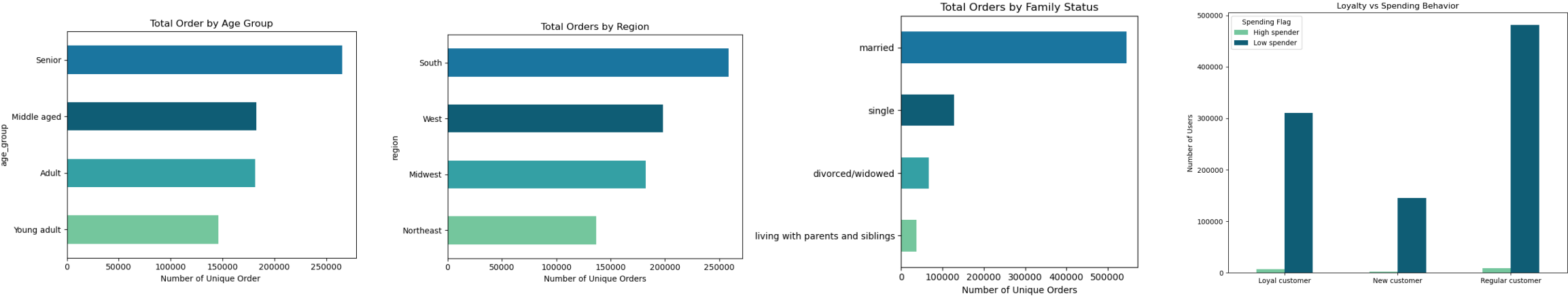
The most popular products are 'Produce' and 'dairy' followed by 'snacks' with the highest order frequency in our dataset by far.

LOYALTY GRUP PROFILE



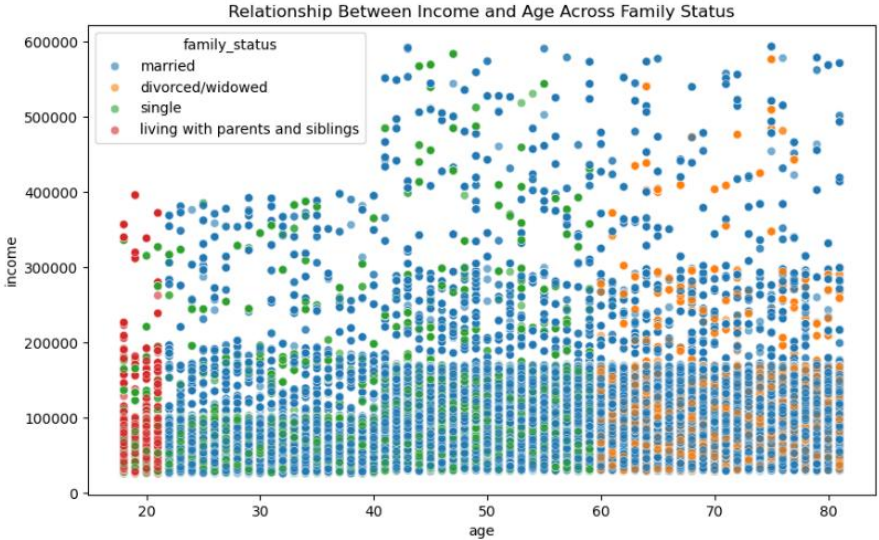
- Seniors represent the largest proportion of members, while young adults have the smallest. Middle-aged and adult groups appear to be similar in size.
- Married individuals make up the largest share of members, while those living with their parents represent the smallest group. Additionally, the divorced or widowed segment is smaller than the single population.
- The middle-income class accounts for the largest share of members. In contrast, the high-income group represents a notably small segment, while the low-income group is the smallest overall. However, the difference in member count between the high- and low-income groups is quite minimal.

ORDERING HABITS BASED ON AGE, REGION, FAMILY STATUS



Married and senior customers have the highest number of orders. The largest number of orders is the south region.

All three customer groups show low spending, with Regulars spending the least, followed by Loyal customers.



Married customers span ages 21–80, with varied incomes:

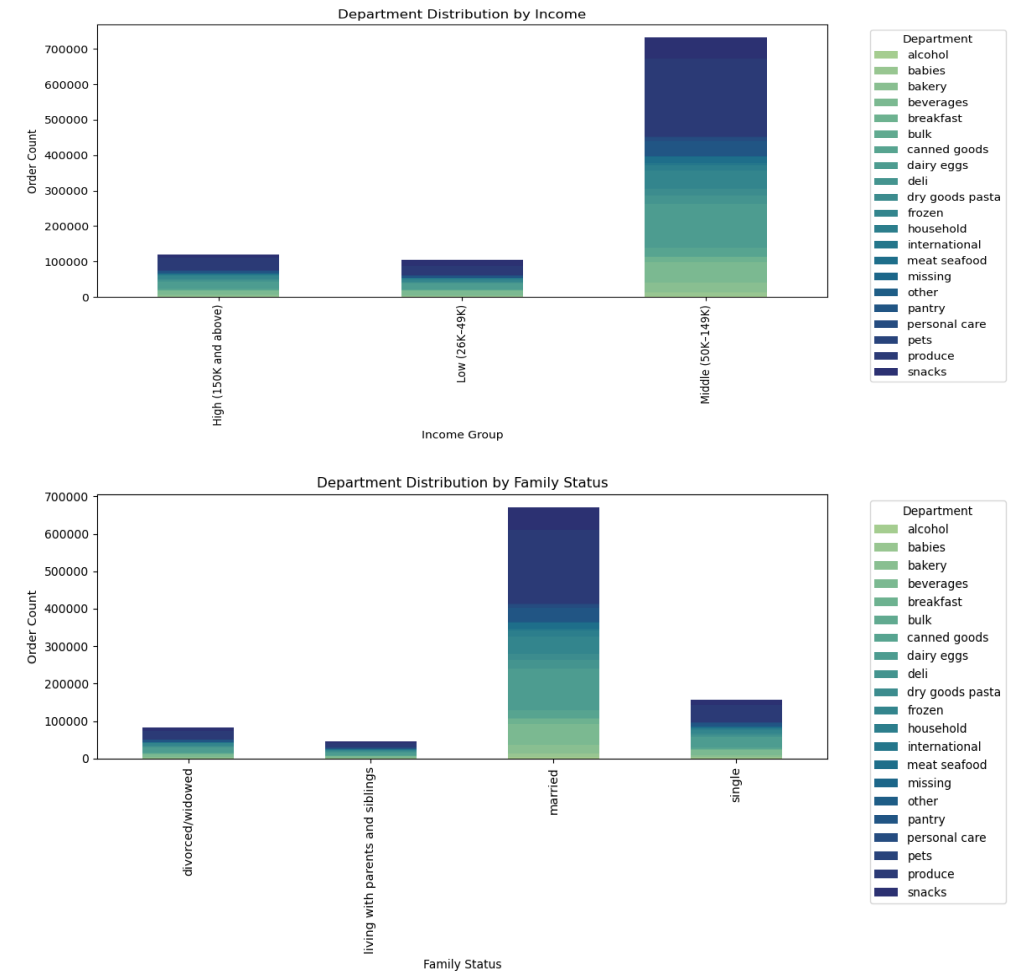
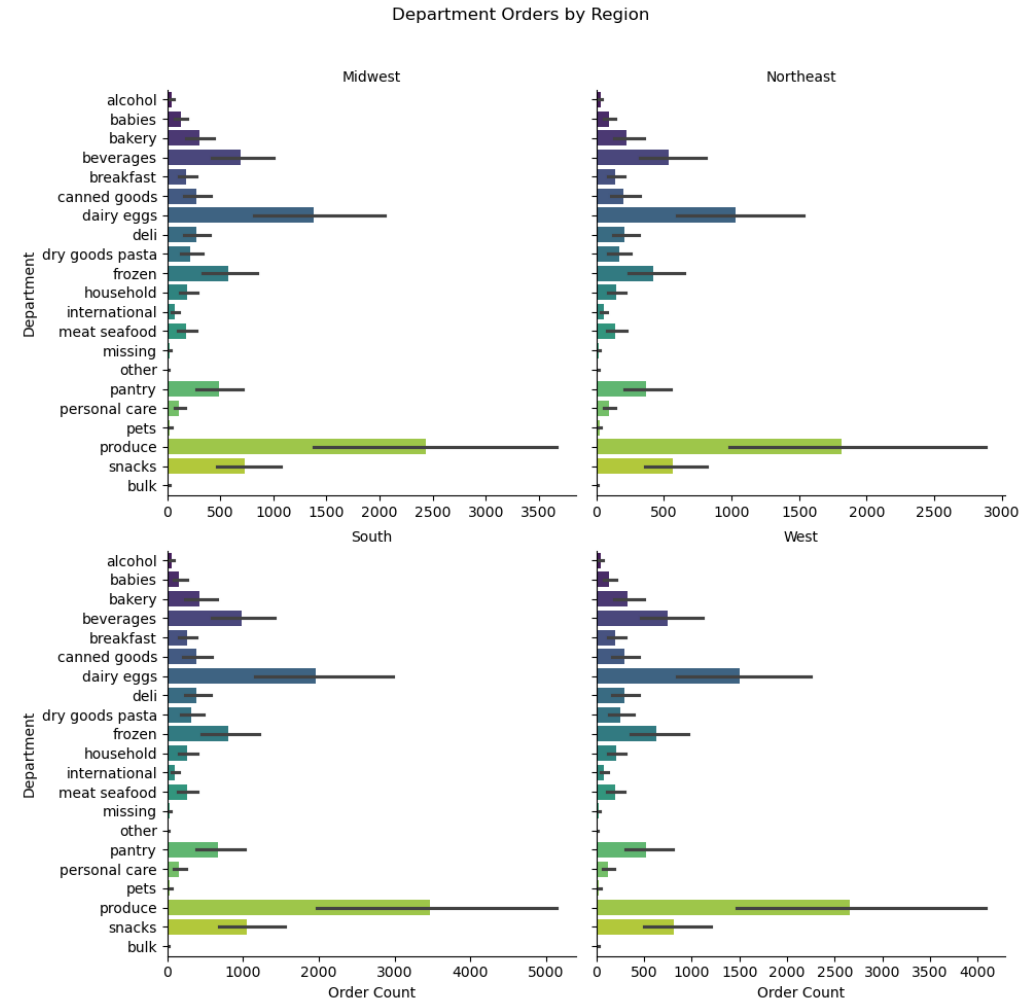
Ages 21–40 mostly earn up to \$100K.

Ages 40–80 generally earn up to \$200K.

Customers living with parents are around age 20 form a small group.

Divorced/widowed customers are aged 60–80, with incomes up to \$200K.

DISTRIBUTION OF DEPARTMENT PRODUCTS BASED ON INCOME, FAMILY STATUS, AND ACROSS REGIONS

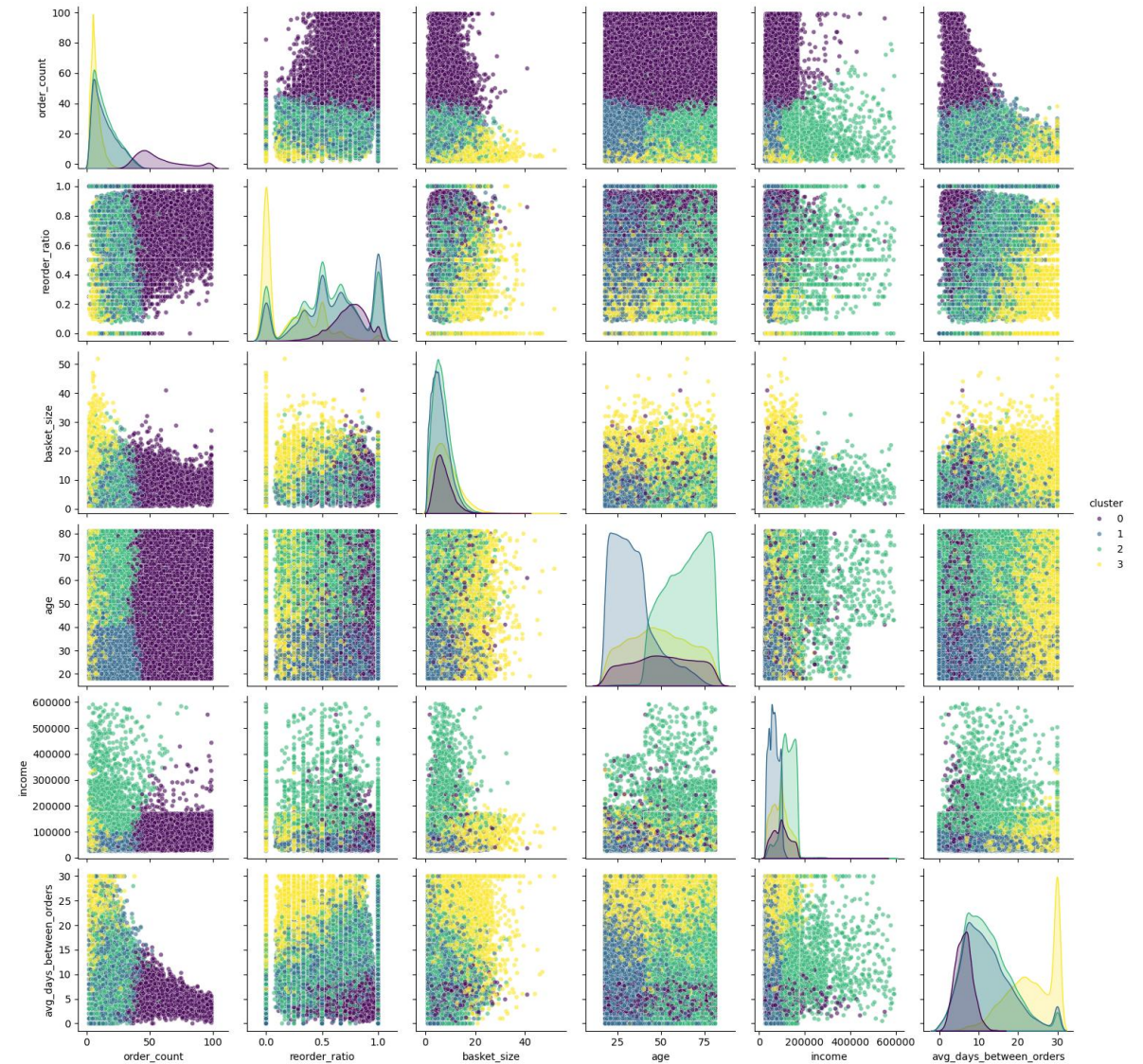


Produce, dairy & eggs, snacks, and beverages are the most popular product categories across all four regions.

There appears to be no significant difference in product choices across different geographic location, family status and income groups.

PROFILING ANALYSIS

- Cluster 0 consists of high-frequency, high-value shoppers. They place frequent orders, have large basket sizes, and short intervals between purchases, making them valuable repeat buyers.
- Cluster 1 represents steady, consistent buyers with moderate order frequency, average basket sizes, and balanced reorder behavior.
- Cluster 2 shows low engagement, with infrequent orders, small baskets, and minimal product loyalty.
- Cluster 3 orders rarely with long gaps but consistently repurchases the same products—showing strong niche loyalty.
- These patterns highlight the need for tailored strategies:
- retain and reward high-value buyers (Cluster 0), nurture the steady base (Cluster 1), activate low-engagement users (Cluster 2), and support loyal but infrequent shoppers (Cluster 3).



RECOMMENDATIONS

Timing & Operations

Staff fully from 10 AM to 4 PM to manage peak orders.

Focus marketing during 5 AM–8 AM and after 10 AM when spending rises.

Run promotions during busy hours to boost sales.

Prepare inventory for high demand on Sundays and Mondays.

Regional Focus

Focus efforts in the South: high order volume, local influencers, faster delivery.

Test localized promotions tied to regional events or holidays.

Use regional data to optimize ad spend and monitor ROI.

Customer Segmentation

Prioritize seniors and married families with loyalty rewards and premium services.

Offer family-sized bundles to middle-aged married customers.

Provide budget-friendly deals for younger married adults.

Use intro offers to build awareness among students.

Target regular customers with re-engagement and loyalty program invites.

Inventory & Product Strategy

Keep produce, dairy, and frozen items well-stocked.

Analyze top-selling weekend products to guide stocking and promotions.

Promote snacks, beverages, and bundles to increase basket size.

Use consistent product offerings across regions and demographics, focusing on pantry, produce, and personal care.

Marketing & Campaigns

Prioritize Sundays for targeted campaigns to increase spending.

Promote product bundles to loyal customers.

Use personalized campaigns to recommend premium products post-purchase.

Test age-based personalization (e.g., family meals vs. wellness products).

Highlight value and convenience in campaigns for seniors and families.

CONTACT

NASIM MAHDINAZHAD | DATA ANALYST

GERMANY

nasimmahdinazhad@gmail.com