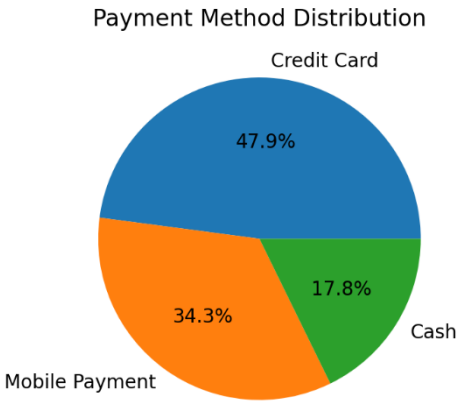
1. Take a look at Customer.py

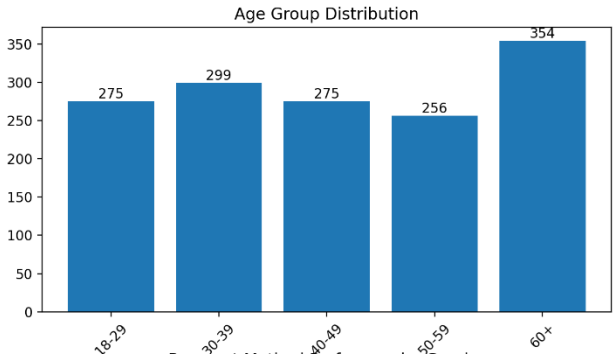
1. *Payment Method Distribution Graph*

- What it is:
  - illustrates the proportion of customers using each payment method
- Key data analysis:
  - Credit Card: 699 customers (47.9%)
  - Mobile Payment: 501 customers (34.3%)
  - Cash: 259 customers (17.8%)
- Trend we observe:
  - Credit Card dominates as the preferred method (~48%), followed by Mobile Payment (~34%), with Cash trailing (~18%)
  - suggests a strong shift toward digital payments, potentially driven by convenience, while cash remains a minority option



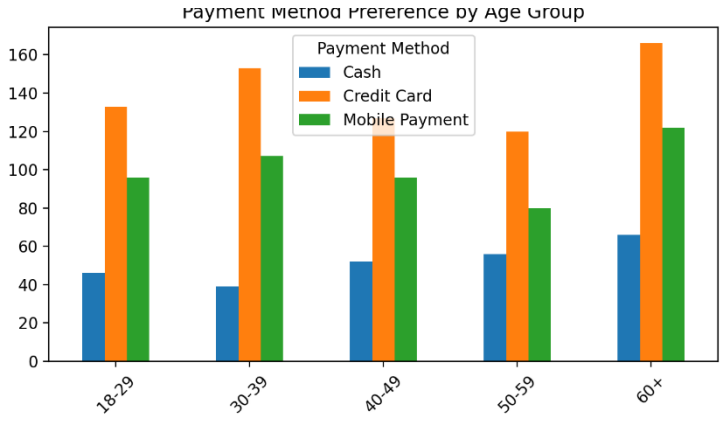
2. *Age Group Distribution Graph*

- What it is:
  - shows the number of customers in each age group
- Key data analysis:
  - 18-29: 275 customers
  - 30-39: 299 customers
  - 40-49: 275 customers
  - 50-59: 256 customers
  - 60+: 354 customers
- Trend we observe:
  - a mature customer base with potential for targeted marketing to seniors while maintaining appeal to working-age adults.



3. Payment Method Preference by Age Group Graph

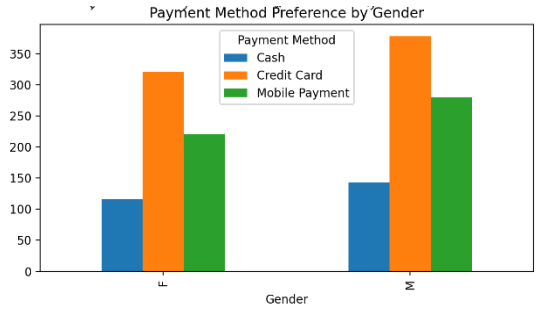
- What it is:
  - breaks down payment method usage within each age group
- Key data analysis:
  - 18-29: Cash 46, Credit Card 133, Mobile Payment 96
  - 30-39: Cash 39, Credit Card 153, Mobile Payment 107
  - 40-49: Cash 52, Credit Card 127, Mobile Payment 96
  - 50-59: Cash 56, Credit Card 120, Mobile Payment 80
  - 60+: Cash 66, Credit Card 166, Mobile Payment 122



- Trend we observe:
  - digital methods prevail, but older groups show marginally higher cash reliance, suggesting tech adoption barriers or preferences.

4. Payment Method Preference By Gender

- What it does:
  - compares payment method usage between genders
- Key data analysis:
  - Female (F): Cash 116, Credit Card 321, Mobile Payment 22
  - Male (M): Cash 143, Credit Card 378, Mobile Payment 280

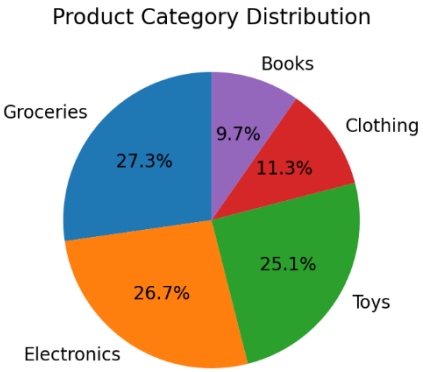


- Trend we observe:
  - minimal differences suggest gender-neutral preferences, though males slightly edge in absolute numbers due to possible higher representation, indicating uniform marketing potential

2. Take a look at products.py

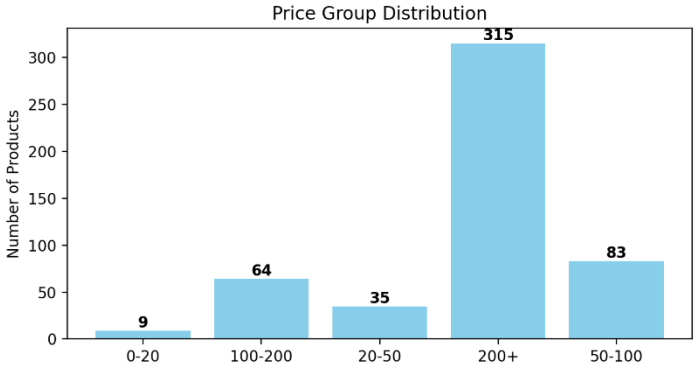
1. *Product Category Distribution*

- What it is:
  - showing proportion of products in each of the five categories across the entire inventory
- Key data analysis:
  - Groceries: 138 products → 27.3%
  - Electronics: 135 products → 26.7%
  - Toys: 127 products → 25.1%
  - Clothing: 57 products → 11.3%
  - Books: 49 products → 9.7%
- Trend we observe:
  - business focuses heavily on groceries, toys and electronics with limited in books, clothes



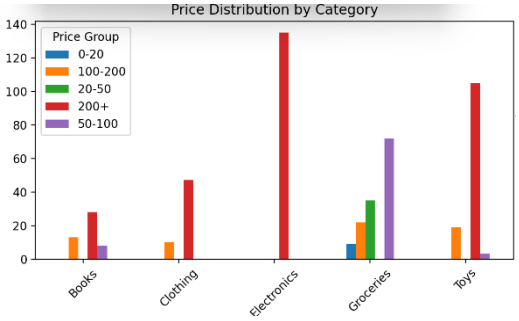
2. *Price Group Distribution*

- What it is:
  - showing how many products fall into five broad price bands
- Key data analysis:
  - 200+: 315 products
  - 50-100: 83 products
  - 100-200: 64 products
  - 20-50: 35 products
  - 0-20: 9 products
- Trend we observe:
  - Extremely skewed toward high-priced items



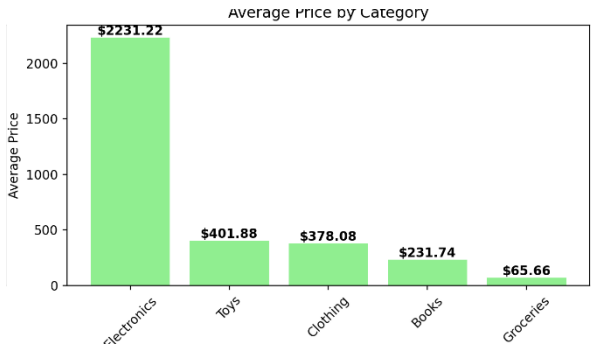
3. *Price Distribution by Category*

- What it is:
  - breaks down the same five price groups for each individual category
- Key data analysis:
  - Books: Mostly 200+, some 100-200, very few below 100
  - Clothing: Almost entirely 200+, Electronics: 100% in 200+
  - Groceries: Spread across lower bands
  - Toys: Overwhelmingly 200+, some 100-200
- Trend we observe:
  - Clear category-specific pricing tiers



4. Average Price by Category

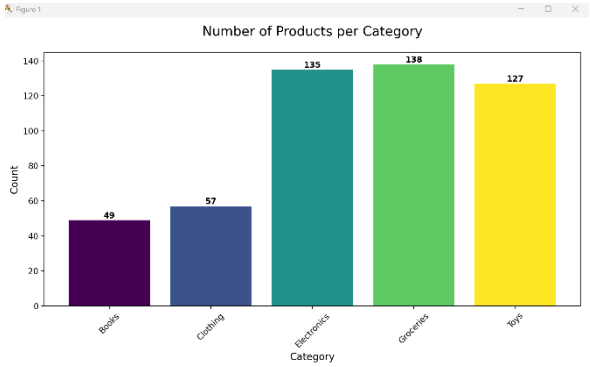
- What it is:
  - displaying the mean price for each category, ordered from highest to lowest
- key data analysis
  - Electronics: $2,231.22
  - Toys: $401.88
  - Clothing: $378.08
  - Books: $231.74
  - Groceries: $65.66



- Trend we observe;
  - Massive price gap between Electronics (extremely expensive) and everything else
  - Toys and Clothing form a mid-premium tier
  - Books are moderate, and Groceries are true low-price items

5. View Products Counts By Category

- What it is:
  - displays the number of products in each category
- Key data analysis:
  - Groceries: 138, Electronics: 135, Toys: 127, Clothing: 57, Books: 49.
  - Total products: 506
- Trend we observe:
  - focus on high-demand categories with potential expansion in lower ones



6. Price Distribution Tables

- What it is:
  - showing the distribution of product prices across 10 equal-interval classes, including frequency (count per range), cumulative frequency, and cumulative relative frequency.
  - filtered by specific category
- Trend we observe:
  - Overall and across categories, prices show a right-skewed pattern with heavy concentration in lower ranges, sparse mid-ranges, and a tail in highs
  - category filters reveal tighter clustering
  - highlighting segmented pricing strategies with premiums dominating non-essentials

Price Distribution Table for Books (10 Classes)

Price Range	Frequency	Cumulative Frequency	Cumulative Relative Frequency
51.68 – 93.11	8	8	0.1633
93.11 – 134.54	7	15	0.3061
134.54 – 175.97	4	19	0.3878
175.97 – 217.40	3	22	0.4490
217.40 – 258.83	6	28	0.5714
258.83 – 300.26	4	32	0.6531
300.26 – 341.69	7	39	0.7959
341.69 – 383.12	4	43	0.8776
383.12 – 424.55	3	46	0.9388
424.55 – 465.98	3	49	1.0000

Price Distribution Table for Clothing (10 Classes)

Price Range	Frequency	Cumulative Frequency	Cumulative Relative Frequency
167.42 – 210.44	11	11	0.1930
210.44 – 253.47	4	15	0.2632
253.47 – 296.49	6	21	0.3684
296.49 – 339.52	3	24	0.4211
339.52 – 382.54	2	26	0.4561
382.54 – 425.56	5	31	0.5439
425.56 – 468.59	4	35	0.6140
468.59 – 511.61	11	46	0.8070
511.61 – 554.64	7	53	0.9298
554.64 – 597.66	4	57	1.0000

Price Distribution Table for Electronics (10 Classes)

Price Range	Frequency	Cumulative Frequency	Cumulative Relative Frequency
1403.43 – 1567.58	13	13	0.0963
1567.58 – 1731.73	15	28	0.2074
1731.73 – 1895.88	15	43	0.3185
1895.88 – 2060.03	11	54	0.4000
2060.03 – 2224.18	11	65	0.4815
2224.18 – 2388.33	9	74	0.5481
2388.33 – 2552.48	18	92	0.6815
2552.48 – 2716.63	16	108	0.8000
2716.63 – 2880.78	11	119	0.8815
2880.78 – 3044.93	16	135	1.0000

Price Distribution Table for Groceries (10 Classes)

Price Range	Frequency	Cumulative Frequency	Cumulative Relative Frequency
14.12 – 24.26	17	17	0.1232
24.26 – 34.39	12	29	0.2101
34.39 – 44.53	12	41	0.2971
44.53 – 54.67	10	51	0.3696
54.67 – 64.80	12	63	0.4565
64.80 – 74.94	15	78	0.5652
74.94 – 85.08	19	97	0.7029
85.08 – 95.22	13	110	0.7971
95.22 – 105.35	11	121	0.8768
105.35 – 115.49	17	138	1.0000

Price Distribution Table for Toys (10 Classes)

Price Range	Frequency	Cumulative Frequency	Cumulative Relative Frequency
85.92 – 149.16	13	13	0.1024
149.16 – 212.40	14	27	0.2126
212.40 – 275.64	9	36	0.2835
275.64 – 338.88	14	50	0.3937
338.88 – 402.12	12	62	0.4882
402.12 – 465.36	15	77	0.6063
465.36 – 528.60	15	92	0.7244
528.60 – 591.84	8	100	0.7874
591.84 – 655.08	11	111	0.8740
655.08 – 718.32	16	127	1.0000

7. Descriptive Statistics

- What it is:
  - computes and displays summary statistics for prices, such as count, mean, median, std. deviation, variance, min, max, and range
  - per-category views
- Trend we observe:
  - per category, Electronics/Toys show high means and variance (wide premium spread), Groceries low and tight (budget focus), with mid-categories like Books/Clothing balanced but leaning moderate, reflecting tiered value propositions

Descriptive Statistics for Price - Books

Statistic	Value
Count	49
Mean	$231.74
Median	$233.13
Std. Deviation	$120.23
Variance	$14456.37
Min	$51.68
Max	$465.98
Range	$414.30

Descriptive Statistics for Price - Clothing

Statistic	Value
Count	57
Mean	$378.08
Median	$399.77
Std. Deviation	$141.03
Variance	$19888.16
Min	$167.42
Max	$597.66
Range	$430.24

Descriptive Statistics for Price - Electronics

Statistic	Value
Count	135
Mean	$2231.22
Median	$2233.55
Std. Deviation	$485.48
Variance	$235690.80
Min	$1403.43
Max	$3044.93
Range	$1641.50

Descriptive Statistics for Price - Groceries

Statistic	Value
Count	138
Mean	$65.66
Median	$69.08
Std. Deviation	$29.52
Variance	$871.65
Min	$14.12
Max	$115.49
Range	$101.37

Descriptive Statistics for Price - Toys

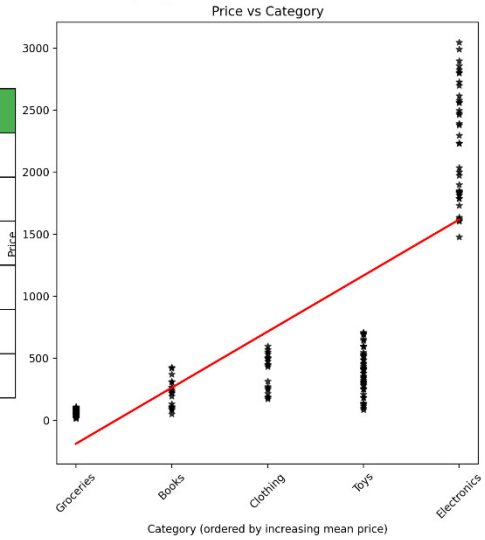
Statistic	Value
Count	127
Mean	$401.88
Median	$408.68
Std. Deviation	$182.82
Variance	$33423.50
Min	$85.92
Max	$718.32
Range	$632.40

8. Linear Regression Analysis

- What it is:
  - runs a linear regression modeling price as a function of category
- Trend we observed:
  - Strong positive R^2 (~0.91) shows category strongly predicts price; upward line from low (Groceries) to high (Electronics),
  - with points clustering tightly per category but increasing variance in premiums, indicating a clear escalating price trend tied to category type from essentials to luxuries

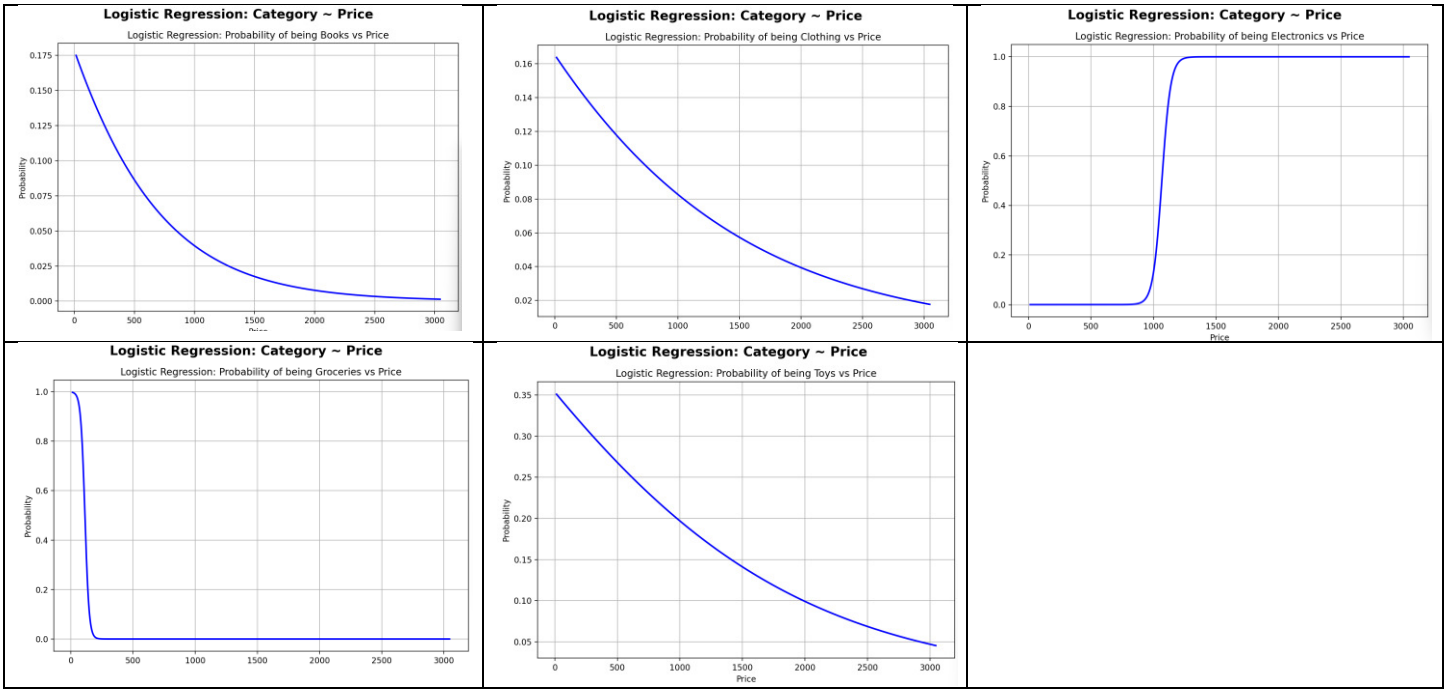
Linear Regression: Price ~ Category

Statistic/Category	Value
R^2	0.9132
Groceries	65.66
Books	231.74
Clothing	378.08
Toys	401.88
Electronics	2231.22



9. Logistic Regression Analysis:

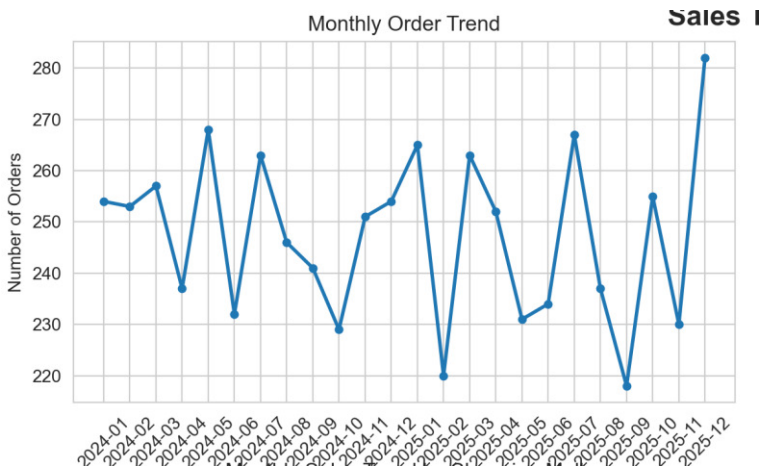
- What it is:
  - fits a binary logistic regression for each selected category using price as the predictor
  - evaluate price's discriminatory power
- Trend we observe:
  - Accuracy varies by category
  - Give a price, can estimate which category the products belong to



3. Take a look at Sales.py

1. Monthly Order Trend Graph

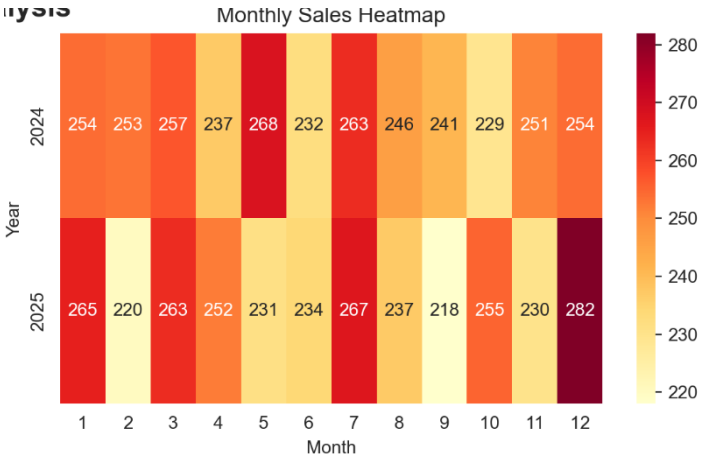
- What it does:
  - tracks the number of orders over time
  - visualizes overall sales volume trends to identify seasonal patterns or growth/declines
- Key data analysis:
  - Total orders: 5,939.
  - Average monthly orders: 247.46.
  - Peak: 282 (2025-12), Low: 218 (2025-09).
- Trend we observe:
  - peaks in May-2024 and December-2025
  - dips in April-2024 and September-2025
  - Slight upward from 2024-2025



2. Monthly Sales Heatmap Graph

- What it does:

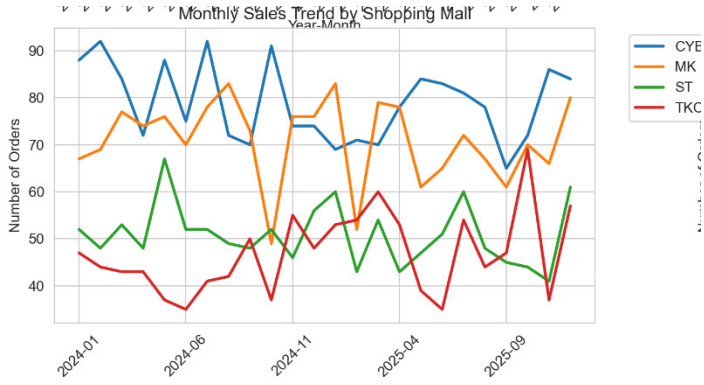
- Displays order counts by year and month with colour intensity representing volume
- highlights seasonal variations and year-over-year comparisons for sales activity



- Key data analysis:
  - Highest: 282 (Dec 2025)
  - Lowest: 218 (Sep 2025)
- Trend we observe:
  - year-over-year slight decline but consistent seasonal peaks in summer and winter,
  - indicating potential holiday or promotional influences.

3. Monthly Sales Trend by Shopping Mall Graph

- What it does:
  - Show order counts per mall over Year-month
  - compares performance across locations to spot mall-specific trends
- Key data analysis:
  - CYB: Peaks at Feb/Jul 2024, low at Sep 2025
  - MK: Peaks at Aug 2024/Jan 2025, low at Oct 2024
  - ST: Peaks at May 2024, low at Nov 2025
  - TKO: Peaks at Oct 2025, low at Jun 2024/2025

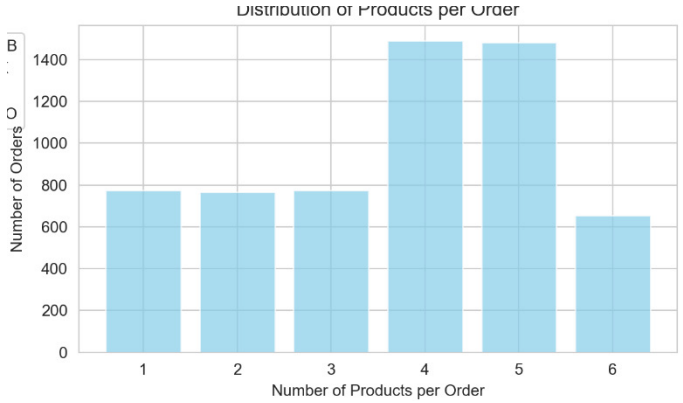


- Trend we observe:
  - All dips in mid-2025, with upward in late 2025
  - suggesting location-specific factors like foot traffic affecting synchronized but scaled trends.

4. Distribution of Products per Order Graph

- What it does:
  - shows the frequency of orders by number of products
  - analyses basket sizes to understand purchasing behaviour

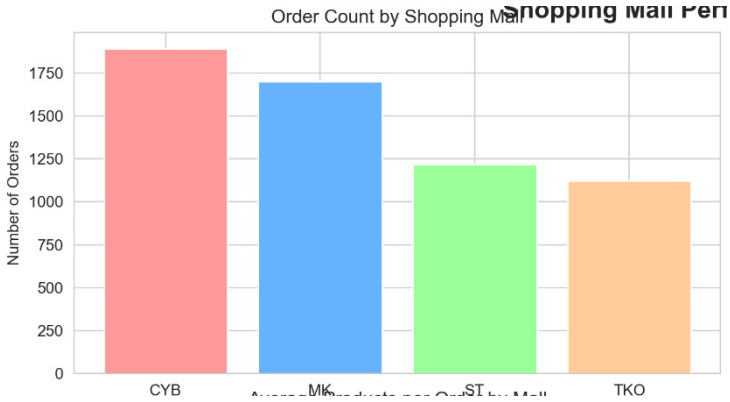
- Key data analysis:
    - Counts:
        - 1 product: 773,
        - 2 products: 767
        - 3 products: 774
        - 4 products: 1,490
        - 5 products: 1,481
        - 6 products: 654



    - Most common: 4 products (1,490 orders, 25.1%).
    - Average products per order: 3.69.
    - Total orders: 5,939.
- Trend we observe:
    - right-skewed toward larger baskets
    - indicating customers often buy multiple items
    - with potential for upselling to increase from low-end orders

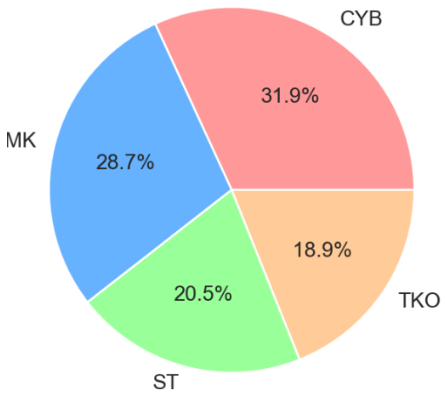
5. Order Count by Shopping Mall Graph

- What it does:
    - compares total orders per mall with count
    - ranks mall performance by volume
- Key data analysis:
    - CYB: 1,893 (31.9%).
    - MK: 1,702 (28.7%).
    - ST: 1,220 (20.5%).
    - TKO: 1,124 (18.9%).
    - Total: 5,939.
- Trend we observe:
    - Descending order CYB > MK > ST > TKO, with CYB/MK dominating



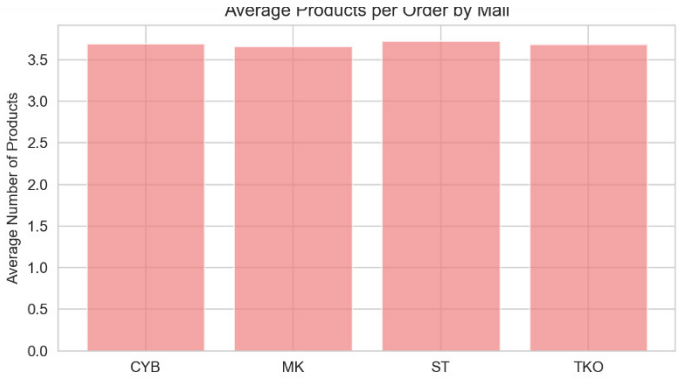
- suggesting resource allocation toward high-performers while investigating underperformers

6. *Market Share by Shopping Mall Graph*



- What it does:
  - shows percentage of total orders per mall
  - visualizes relative contributions to overall sales
- Key data analysis:
  - Shares: CYB 31.9%, MK 28.7%, ST 20.5%, TKO 18.9%.
- Trend we observe:
  - indicates market concentration in top two
  - with opportunities to boost lower shares through targeted promotions

7. *Average Products per Order by Mall Graph*

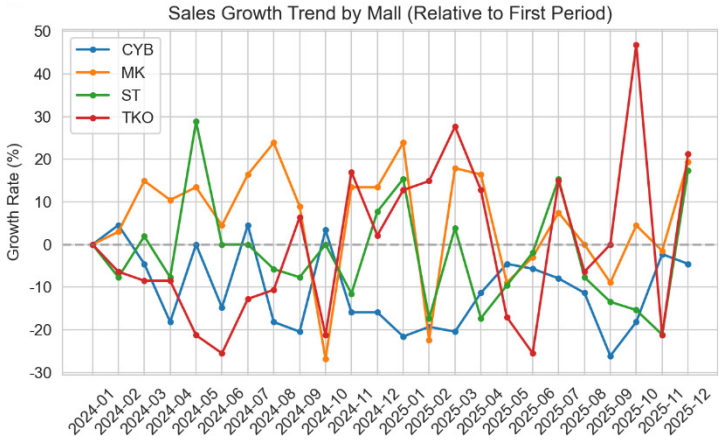


- What it does:
  - displays average products per order per mall with average
  - assesses basket size variations by location
- Key data analysis:
  - Averages:
    CYB 3.69, MK 3.66, ST 3.73, TKO 3.69.
  - Overall average: 3.69.
- Trend we observe:
  - Nearly uniform across malls
  - consistent buying behavior regardless of location

8. *Sales Growth Trend by Mall Graph*

- What it does:
  - monthly growth rates (%) relative to first period per mall
  - tracks performance changes over time
- Key data analysis:

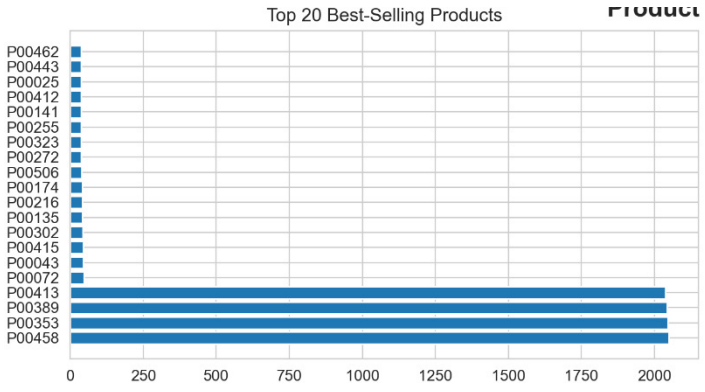
- CYB: Peaks +4.55% (Feb 2024), lows -26.14% (Sep 2025).
- MK: Peaks +23.88% (Aug 2024/Jan 2025), lows -26.87% (Oct 2024).
- ST: Peaks +28.85% (May 2024), lows -21.15% (Nov 2025).
- TKO: Peaks +46.81% (Oct 2025), lows -25.53% (Jun 2024/2025).
- Averages: CYB -10.48%, MK +5.07%, ST -2.56%, TKO -1.60%.



- Trend we observe:
  - highlights concentration on few SKUs, suggesting reliance on staples with potential to promote mid-tier products

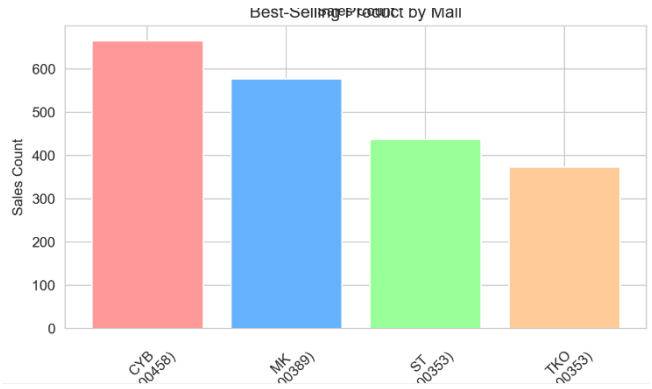
9. Top 20 Best-selling Products Graph

- What it does:
  - ranks top 20 products by sales count
- Key data analysis:
  - Top: P00458: 2,049, P00353: 2,045, P00389: 2,044, P00413: 2,038
- Trend we observe:
  - Extreme skew with top 4 >>2,000 each (dominant)



10. Best-selling Products by Mall Graph

- What it does:
  - shows sales count for the top product per mall
- key data analysis:
  - CYB: P00458 (666), MK: P00389 (577), ST: P00353 (437), TKO: P00353 (374)
- Trend we observe:
  - Shows the highest sales of products in each mall

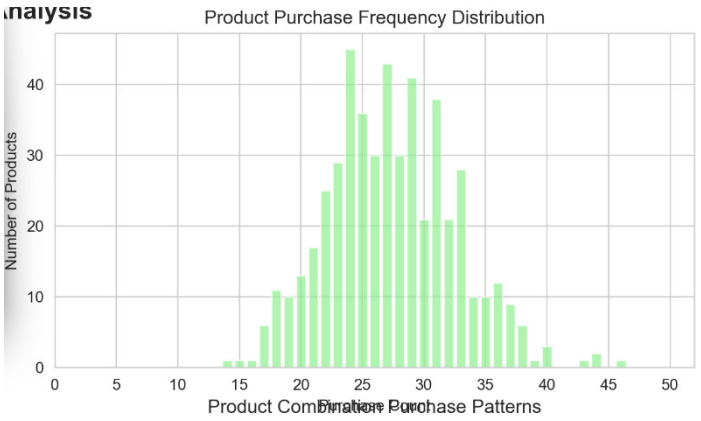


11. Product Purchase Frequency Distribution Graph

- What it does:
  - distribution of product purchase frequencies

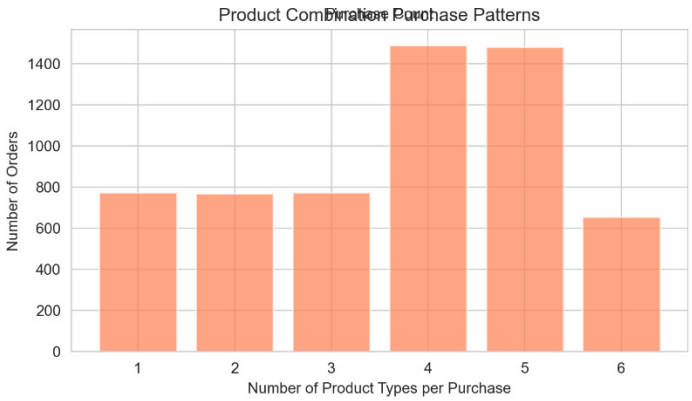
- binned to reveal how often items are bought

- Key data analysis:
  - Frequencies: Most products 14-115 (502 products), then sparse
  - top bin 1,947-2,049: 4 products
- Trend we observe:
  - most items low-volume, few bestsellers drive sales, suggesting inventory optimization for highs



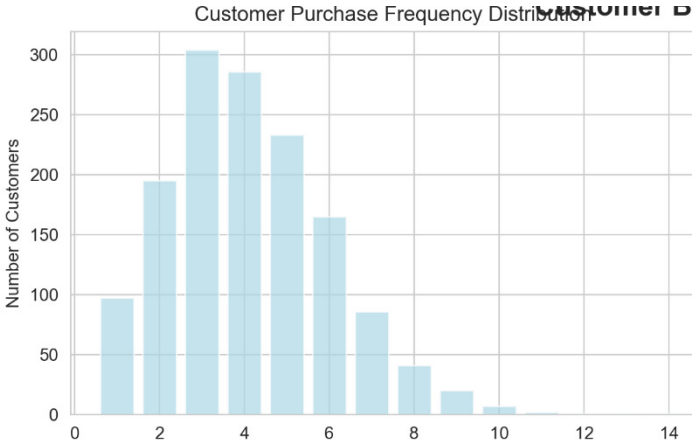
12. Product Combination Purchase Patterns Graph

- What it does:
  - displays orders by number of unique categories per purchase
  - examines cross-category buying
- Key data analysis:
  - 1: 1,007, 2: 1,326, 3: 2,818, 4: 755, 5: 33
- Trend we observe:
  - skewed toward multi-category purchases
  - indicating bundling opportunities to encourage diverse baskets

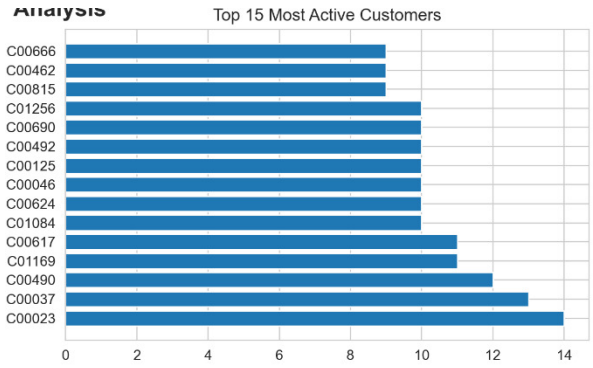


13. Customer Purchase Frequency Distribution Graph

- What it does:
  - shows customers by number of purchases
- Key data analysis
  - Counts: 1: 97, 2: 195, 3: 304, 4: 286, 5: 233, 6: 165, 7: 86, 8: 41, 9: 20, 10: 7, 11: 2, 12: 1, 13: 1, 14: 1
  - Average: 4.13 purchases
- Trend we observe:
  - Right-skewed, Peak at 3-5, drop-off after 6
  - most semi-frequent, few high-repeat, suggesting loyalty programs to shift low-frequency upward

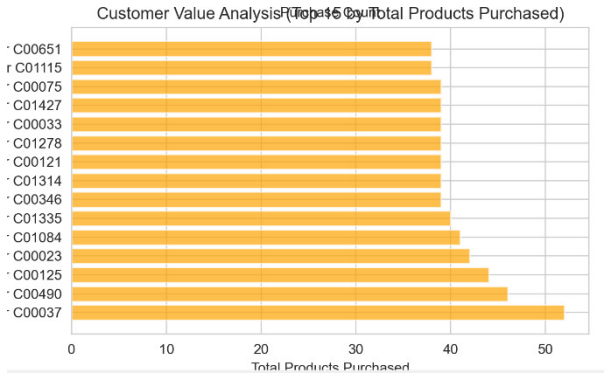


14. Top 15 Most Active Customers Graph



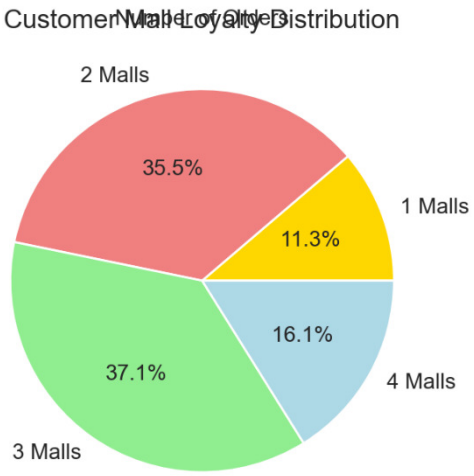
- What it does:
  - ranks top 15 customers by purchase count
- Key data analysis:
  - Top: C00023: 14, C00037: 13, C00490: 12, C01169/C00617: 11
- Trend we observe:
  - See top 15 most active customers

15. Customer Value Analysis Graph



- What it does:
  - shows top customers by total products purchased
- key data analysis:
  - Top: C00037: 52, C00490: 46, C00125: 44, C00023: 42, C01084: 41
- Trend we observe:
  - emphasizing high-value customers buy more items
  - potential for personalized offers to amplify

16. Customer Mall Loyalty Distribution Graph



- What it does:
  - shows percentage of customers by number of malls shopped
- key data analysis:
  - Shares: 3 malls: 37.1%, 2: 35.5%, 4: 16.1%, 1: 11.3%
- Trend we observe:
  - Largest shares in 2-3 malls (~73%), smaller in 1/4; trend toward moderate multi-mall shopping
  - indicating some loyalty but opportunities for cross-promotions to increase exclusivity or expansion