



# PRESENTATION

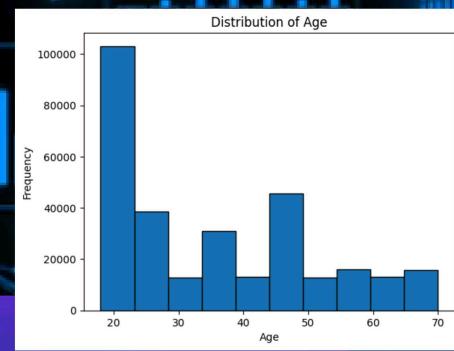
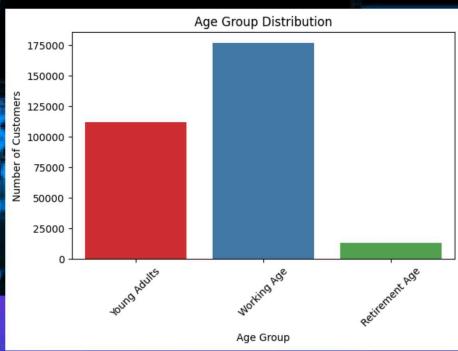
BY TEAM  
GROOT

CHALLENGE -1

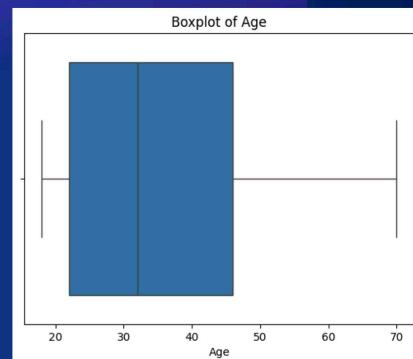


## Age-wise Customer Analysis

```
count    301837.000000
mean      35.481326
std       15.021933
min      18.000000
25%     22.000000
50%     32.000000
75%     46.000000
max      70.000000
Name: Age, dtype: float64
```

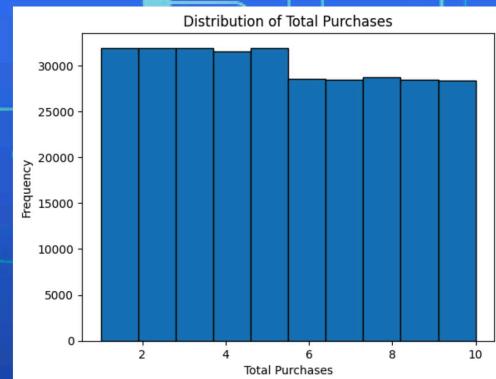


- 👉 The majority of the population/customers fall in the working-age group, making them the dominant segment.
- 👉 Young adults form the second-largest group, showing a strong presence of younger individuals.
- 👉 The retirement-age group is the smallest, indicating fewer older individuals in the dataset.
- 👉 The histogram shows a sharp peak around age 18–22, suggesting many young people enter the dataset at this stage.
- 👉 There is a smaller rise around ages 45–50, but overall, frequencies decline with age



```
count    301649.000000
mean      5.359729
std       2.868575
min      1.000000
25%      3.000000
50%      5.000000
75%      8.000000
max     10.000000
Name: Total_Purchases, dtype: float64
```

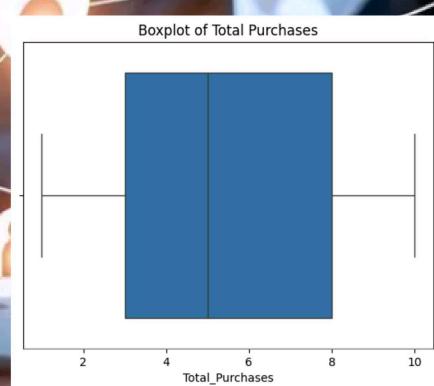
## Distribution of Total Purchases and Purchase Frequency



👉 Purchases are spread evenly between 2 and 10.

👉 Most customers purchase 4 to 8 times.

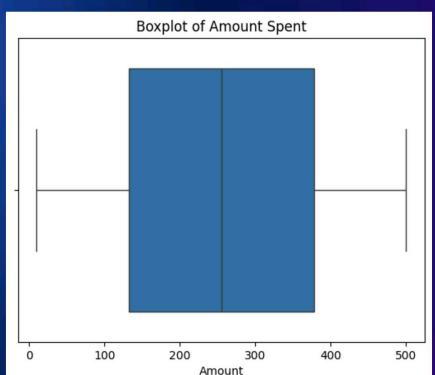
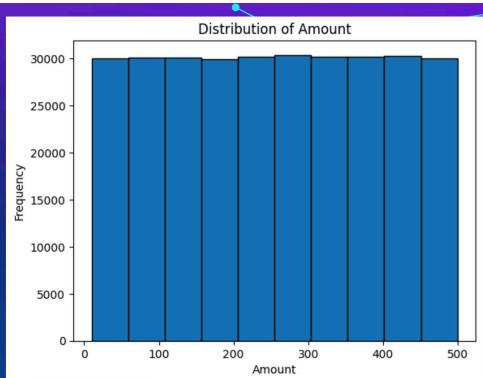
👉 The median is 6, which is right in the middle → distribution is balanced and not skewed.



# CUSTOMER SPENDING PATTERNS

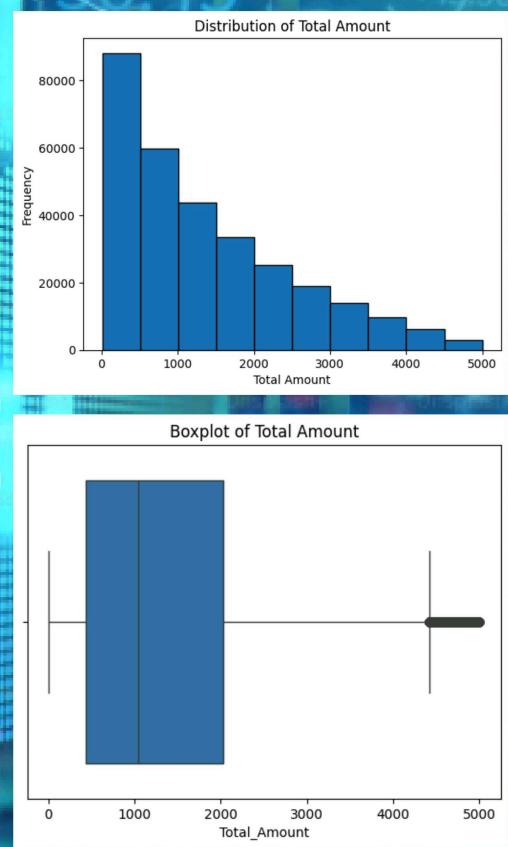
```
count    301653.000000
mean     255.163659
std      141.389640
min      10.000219
25%     132.890764
50%     255.470969
75%     377.672606
max     499.997911
Name: Amount, dtype: float64
```

- Amount spent ranges evenly between 0 and 500.
- Most people spend between 125 and 375.
- The median spend is around 250.



## Total Expenditure Distribution Across Customers

```
count      301660.000000
mean       1367.651156
std        1128.998515
min        10.003750
25%        438.724278
50%        1041.117547
75%        2029.999853
max        4999.625796
Name: Total_Amount, dtype: float64
```



- Most customers spend small amounts (under 1000).
- Median spending is around 2000–2500.
- A few high spenders exist (outliers), but they are rare.
- Data is skewed with a long tail toward higher amounts.

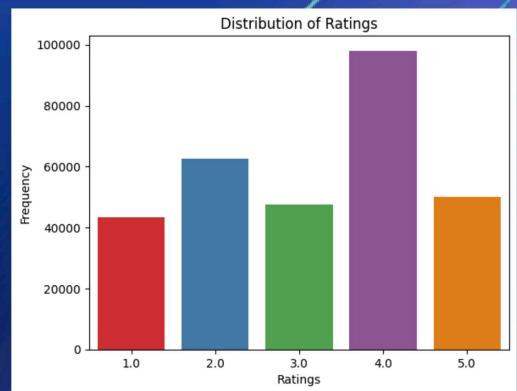
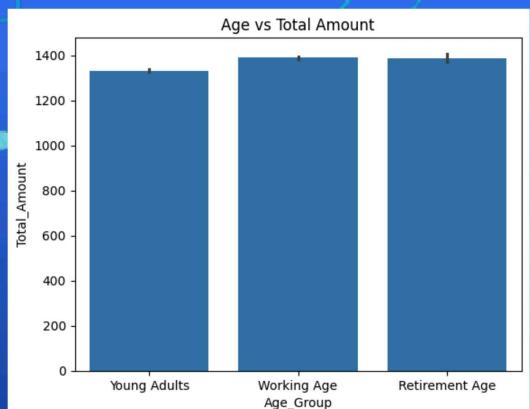
```
count    301826.000000
mean      3.162670
std       1.320827
min      1.000000
25%     2.000000
50%     3.000000
75%     4.000000
max      5.000000
Name: Ratings, dtype: float64
```

## AGE & EXPENDITURE AND RATINGS & PURCHASE FREQUENCY ANALYSIS

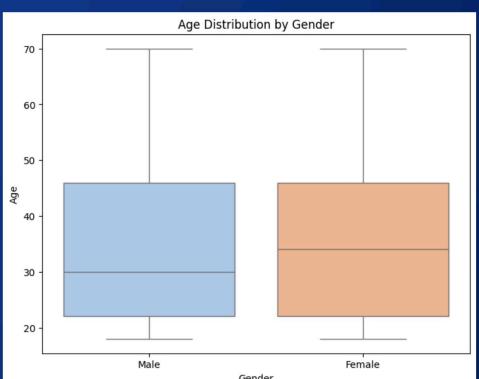
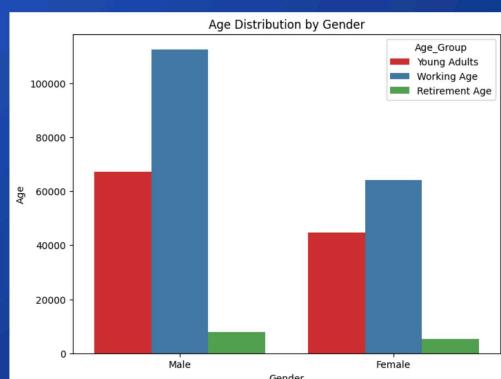
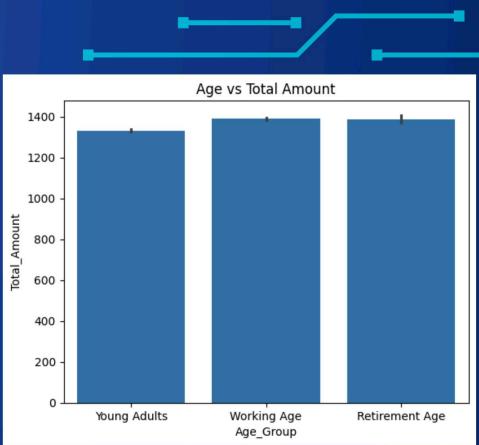
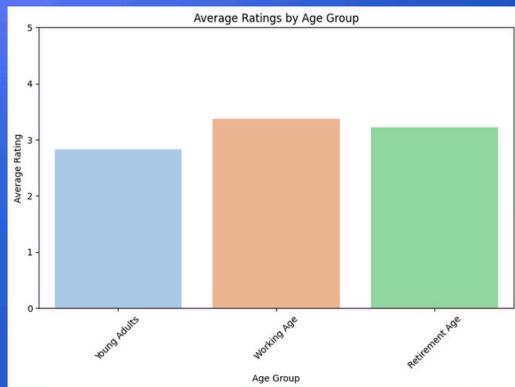
Ratings of 4.0 are the most frequent, with a count of nearly 100,000. This is the most common rating.

Ratings of 5.0 are also very frequent, but less so than 4.0.

The lowest ratings (1.0 and 2.0) are the least frequent.



## Customer Insights by Age: Ratings, Spending, and Gender



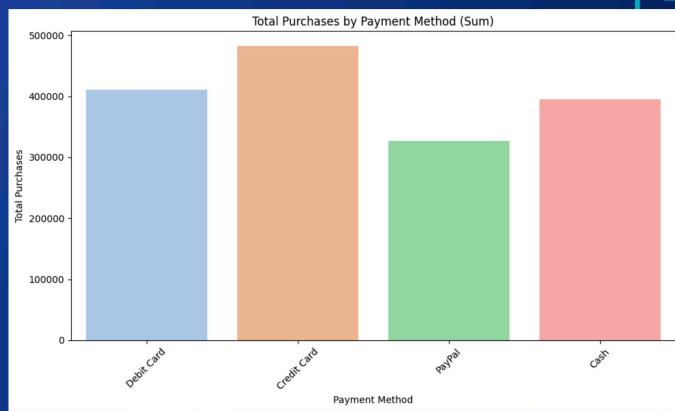
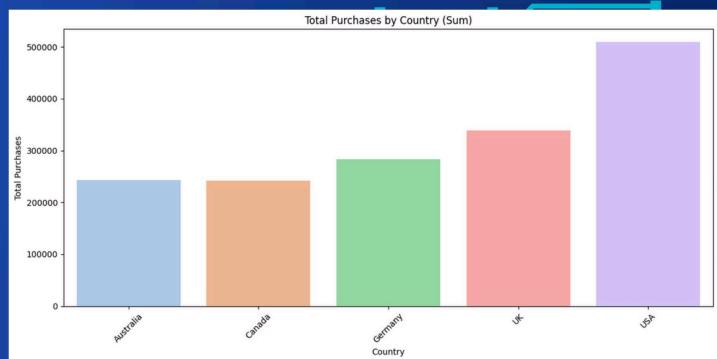
👉 The data suggests a female-dominated user base.

👉 Working professionals form the core demographic.

👉 Financial engagement is consistent across active age groups.

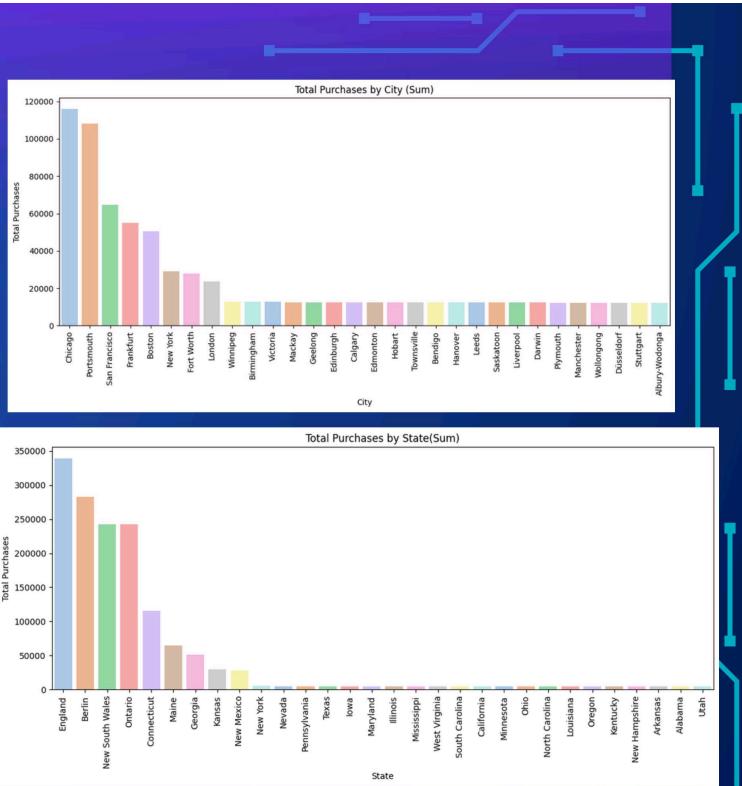
# COUNTRY-WISE AND PAYMENT METHOD-WISE PURCHASE ANALYSIS

- USA leads significantly in total purchases (~\$500k), showing it's the primary market.
- Australia and Canada have similar, lower purchase volumes (~\$240k each)
- Credit Card is the dominant payment method, generating the highest total purchases (~\$480k)
- Debit Card follows as second preferred (~\$410k)
- Cash and PayPal show similar usage levels (~\$400k and ~\$330k respectively)



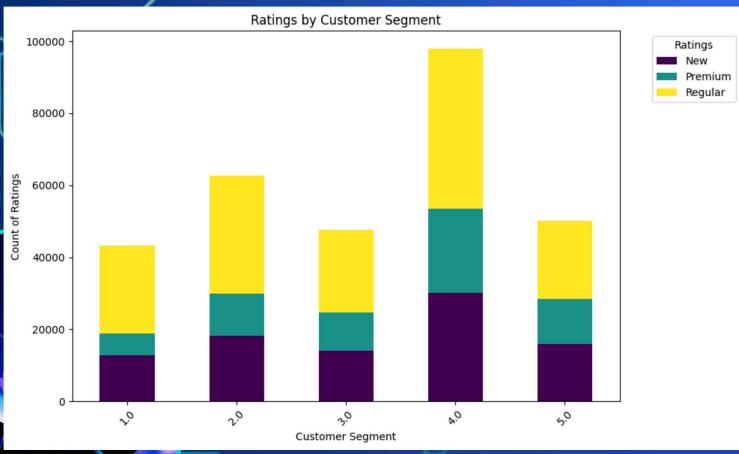
## REGIONAL PURCHASE ANALYSIS: STATE AND CITY INSIGHTS

- Chicago dominates city-level sales, indicating a strong urban market presence
- Portsmouth is the second-highest performing city
- England leads county-level purchases, consistent with UK being a strong country market
- Berlin and Ontario show strong regional performance.
- There's significant geographic diversity with many counties contributing smaller amounts



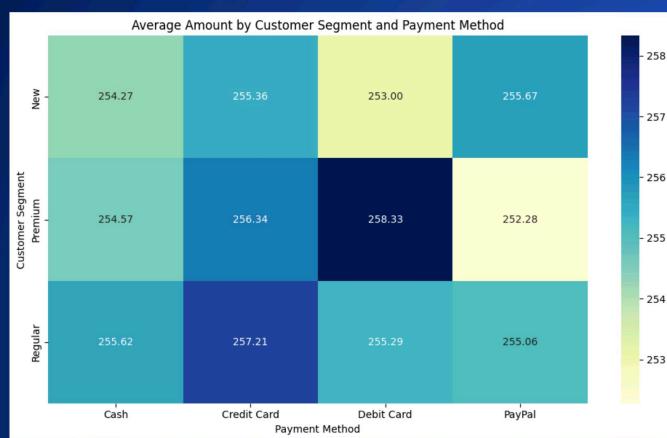
## CUSTOMER SATISFACTION BY SEGMENT

Customer_Segment Ratings	New	Premium	Regular
1.0	12808	6015	24453
2.0	18260	11636	32741
3.0	14065	10694	22840
4.0	30180	23405	44405
5.0	15831	12613	21665



- Overall trend: The highest number of ratings were given to the Customer Segment 4.0.
- Segment breakdown:
  - New customers have a relatively even distribution of ratings across all segments.
  - Premium customers are most heavily concentrated in Customer Segment 4.0.
  - Regular customers give the most ratings to Customer Segments 2.0 and 4.0

### Average Amount by Customer Segment and Payment Method:-

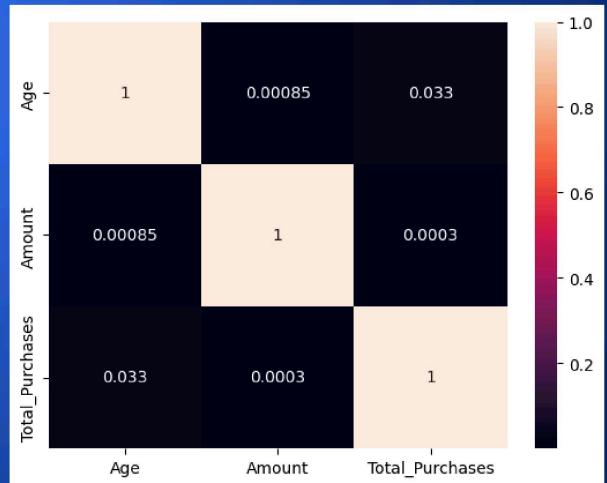


👉 **Highest Ratings:** Premium and Regular customers give the most ratings to Customer Segment 4.0.

👉 **Highest Spending:** Premium customers have the highest average spending, particularly when using a debit card.

👉 **Key Takeaway:** The Premium customer segment appears to be the most valuable, contributing both high ratings and high average spending.

### Heatmap of Age, Amount, and Total Purchases:-

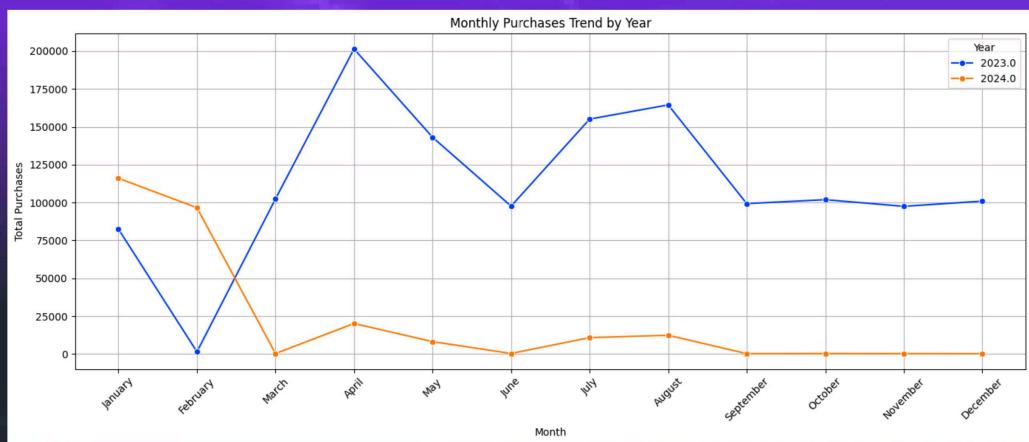


👉 **Age and Amount → no meaningful relationship**

👉 **Age and Total\_Purchases → very weak positive relationship, but practically negligible**

👉 **Amount and Total\_Purchases → almost no correlation, meaning the number of purchases doesn't strongly influence how much is spent on average**

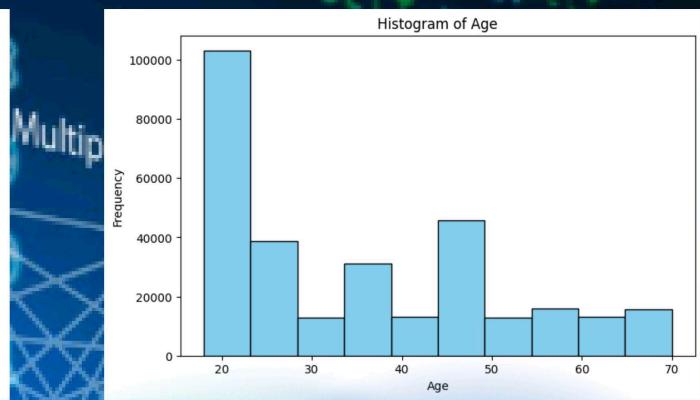
## "Annual Purchase Comparison: 2023 vs 2024 Monthly Trends"



- 2023 showed high volatility with significant peaks and recoveries
- 2024 performance was dramatically lower across all months
- April appears to be a seasonal peak month for both years
- February consistently shows the lowest purchases in both years
- Overall trend suggests a major decline in business performance from 2023 to 2024

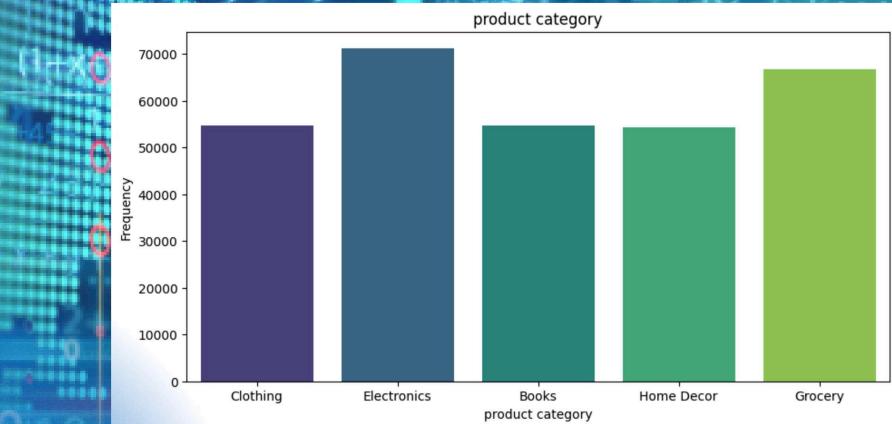
# Population Demographics by Age Group

Age	
count	301837.000000
mean	35.481326
std	15.021933
min	18.000000
25%	22.000000
50%	32.000000
75%	46.000000
max	70.000000



- 👉 Heavy skew toward younger ages: Massive peak in the 20s age group (~103k frequency)
- 👉 Sharp decline: Dramatic drop from 20s to 30s, then gradual decrease
- 👉 Secondary peak: Notable uptick in the 40s (~45k) before declining again
- 👉 Aging population gap: Very low representation in 50s-70s age ranges

# MARKET SHARE BY PRODUCT CATEGORY



- Electronics dominates: Highest sales/frequency at ~72k, leading all categories
- Grocery second: Strong performance at ~68k, close behind electronics
- Mid-tier categories: Books and Home Decor are nearly tied at ~55k each
- Clothing lowest: Underperforms at ~55k, the weakest category

	Product_Category	count
0	Electronics	71196
1	Grocery	66786
2	Clothing	54741
3	Books	54622
4	Home Decor	54382

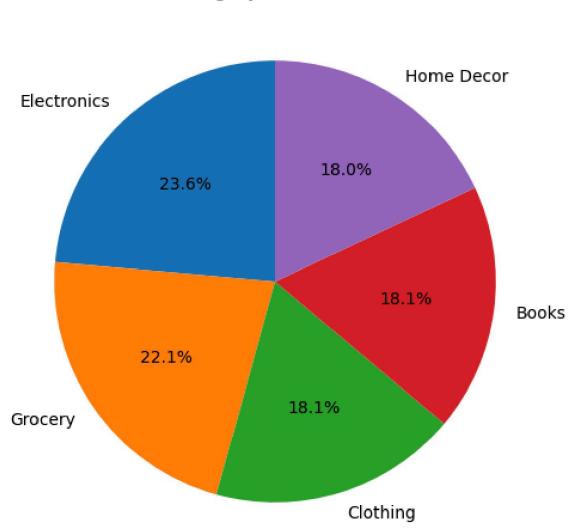
## Product Category Distribution :

### Key Insights :

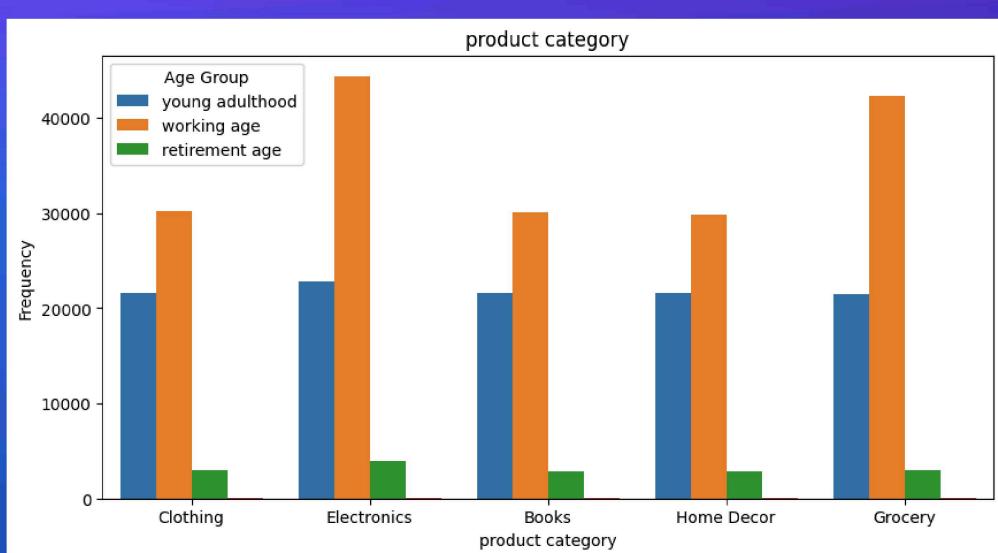
- Electronics is the most purchased category.
- Grocery is a close second, indicating essential/recurring demand.
- Clothing, Books, and Home Decor have similar and lower purchase counts, showing moderate interest.
- Clear preference for tech and daily essentials over lifestyle items.

	Product_Category	count
0	Electronics	71196
1	Grocery	66786
2	Clothing	54741
3	Books	54622
4	Home Decor	54382

Category Distribution



## Product Category Distribution based on age group



### Key Insights:

- Electronics dominates across all age groups.
- Working age group is the largest consumer base—key target for most categories.
- Young adults show diverse interest, especially in Books and Home Decor.
- Retirees are active consumers, especially in Electronics and Clothing—possibly for communication, entertainment, and comfort.

5	retirement age	Electronics	3985
6	retirement age	Clothing	2935
7	retirement age	Grocery	2934
8	retirement age	Home Decor	2895
9	retirement age	Books	2873
10	working age	Electronics	44366
11	working age	Grocery	42279
12	working age	Clothing	30202
13	working age	Books	30076
14	working age	Home Decor	29818
15	young adulthood	Electronics	22802
16	young adulthood	Books	21641
17	young adulthood	Home Decor	21640
18	young adulthood	Clothing	21566
19	young adulthood	Grocery	21542

# PRODUCT CATEGORY VS GENDER

Gender	Product_Category	count
Female	Grocery	25893
Female	Electronics	24679
Female	Books	21189
Female	Clothing	21111
Female	Home Decor	21107
Male	Electronics	46447
Male	Grocery	40835
Male	Clothing	33565
Male	Books	33379
Male	Home Decor	33205

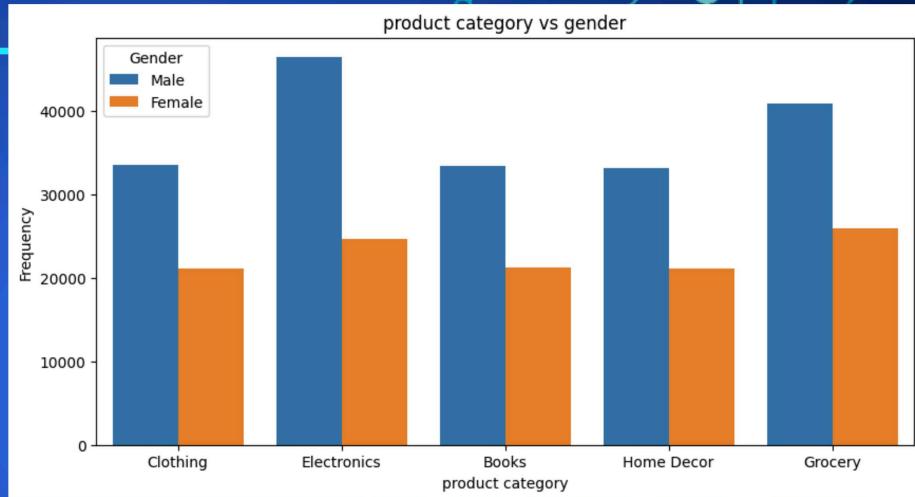
## Key Insights by Gender:

### Female Shoppers:

- Top Category: **Grocery (25,893)** – Suggests focus on daily needs and essentials.
- **Electronics** is the second most purchased (24,679), showing strong tech interest.
- Other categories like Books, Clothing, and Home Decor have nearly equal purchase counts (~21,000 each), indicating balanced shopping behavior across lifestyle and leisure items.

### Male Shoppers:

- Top Category: **Electronics (46,447)** – Dominant preference, more than any other category or demographic.
- **Grocery (40,835)** is next, showing practical shopping behavior.
- High numbers in Books, Clothing, and Home Decor, but less balanced than females – more skewed toward tech and essentials.



# CUSTOMER SEGMENT DISTRIBUTION

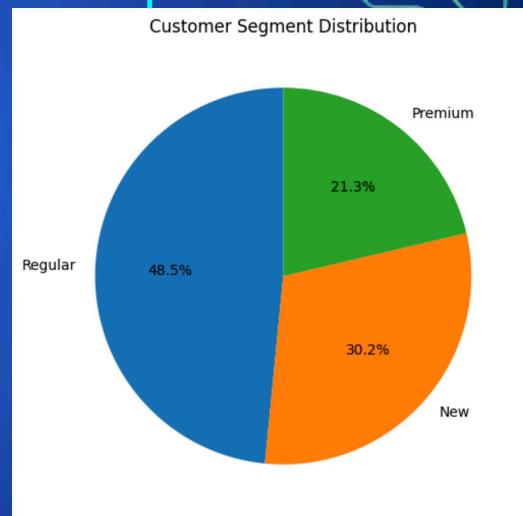
## Key Insights:

- Regular customers form the largest segment (~47.5%):
  - Indicates strong customer retention and repeat usage.
  - This is the core user base.
- New customers make up ~30%:
  - Healthy customer acquisition rate.
  - Important to focus on converting them to Regular or Premium through onboarding and engagement.
- Premium customers are ~21%:
  - Smaller group but likely high value.
  - Important to retain and upsell this group—loyal, high-paying customers.

## Business Implications:

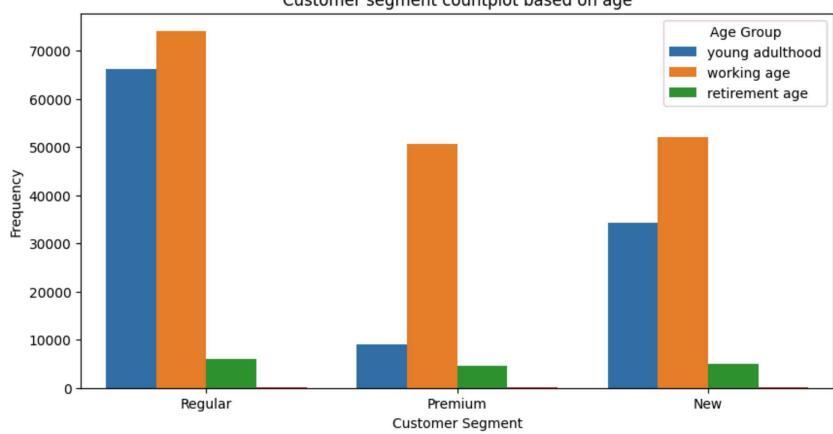
- Focus on converting New → Regular, as nearly 1 in 3 users is still new.
- Encourage Regular → Premium upgrades via exclusive perks, better service, or loyalty rewards.
- Maintain satisfaction among Premium users to prevent churn.
- Regular users are a valuable middle segment that drives consistent business—ideal target for loyalty campaigns.

Customer_Segment	count	
0	Regular	146221
1	New	91187
2	Premium	64387



# CUSTOMER SEGMENT DISTRIBUTION

Customer segment countplot based on age



3	retirement age	Regular	5993
4	retirement age	New	4956
5	retirement age	Premium	4665
6	working age	Regular	74034
7	working age	New	52008
8	working age	Premium	50691
9	young adulthood	Regular	66130
10	young adulthood	New	34173
11	young adulthood	Premium	8972

- Key Insights:

- 🚀 Working Age (Most Valuable Segment):
- Highest count in all customer segments:

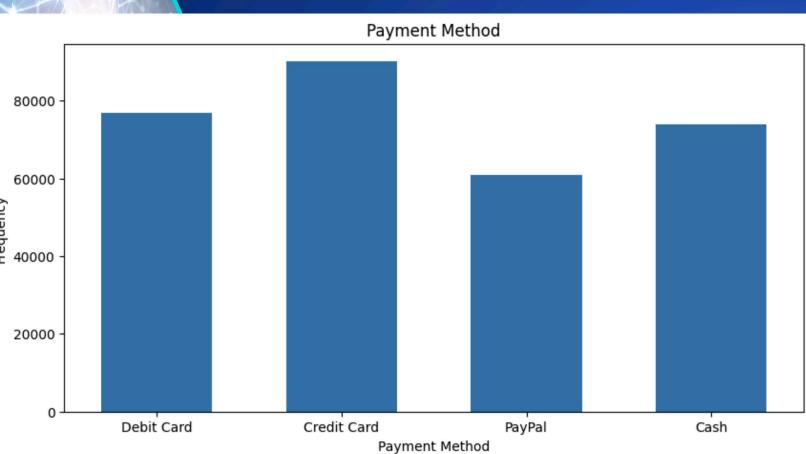
- This group represents the largest and most engaged user base.
- They're likely the most financially stable and tech-savvy, making them the best target for upselling to Premium.
- 🧑 Young Adulthood:
- Very strong Regular base (66,130) but low Premium conversion (8,972).
- Indicates they engage well but are less likely to pay for premium services—likely due to budget sensitivity.
- Opportunity: Offer student discounts, entry-level Premium tiers, or loyalty-based upgrades.
- 🏖 Retirement Age:
- Moderate numbers across all segments:
- Shows good Premium conversion rate relative to their total.
- Likely value service quality, simplicity, and trust, so focusing on reliability and customer support could help further grow this segment.

## Key Takeaways :

- Working-age users are the most engaged and show high Premium conversion – your core audience.
- Young adults are loyal (high Regular) but low on Premium – target them with affordable Premium plans.
- Retirement-age users show balanced engagement – focus on trust and simplicity to grow.

.

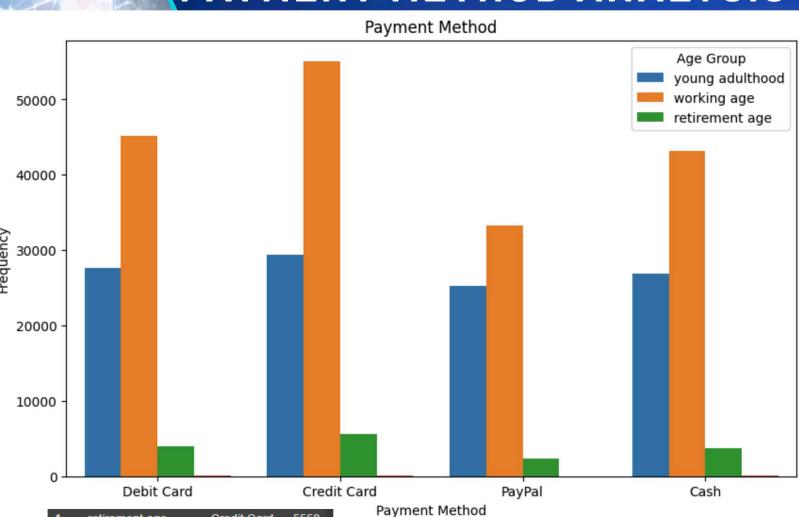
## PAYMENT METHOD DISTRIBUTION



	Payment_Method	count
0	Credit Card	90116
1	Debit Card	76790
2	Cash	73810
3	PayPal	60997

- Key Insights:
- Payment Method Usage
- Credit Card is the most used payment method (90,116 transactions):
  - Indicates customer trust in the platform and preference for convenience or rewards.
  - Possibly used more for larger or recurring purchases.
- Debit Card (76,790) and Cash (73,810) are close in usage:
  - Suggests a balanced mix of digital and traditional payment behavior.
  - Cash usage remains strong, indicating offline or cash-on-delivery transactions are still relevant.
- PayPal (60,997) is the least used, but still significant:
  - Shows some users prefer secure third-party gateways.
  - Opportunity to promote PayPal for privacy-conscious or international users.

## PAYMENT METHOD ANALYSIS ACCORDING TO AGE GROUP



4	retirement age	Credit Card	5558
5	retirement age	Debit Card	4034
6	retirement age	Cash	3697
7	retirement age	PayPal	2337
8	working age	Credit Card	55092
9	working age	Debit Card	45112
10	working age	Cash	43187
11	working age	PayPal	33329
12	young adulthood	Credit Card	29402
13	young adulthood	Debit Card	27601
14	young adulthood	Cash	26885
15	young adulthood	PayPal	25306

- Key Insights:
- Payment Method Usage
- Credit Card is the most used payment method (90,116 transactions):
- Indicates customer trust in the platform and preference for convenience or rewards.
- Possibly used more for larger or recurring purchases.
- Debit Card (76,790) and Cash (73,810) are close in usage:
- Suggests a balanced mix of digital and traditional payment behavior.
- Cash usage remains strong, indicating offline or cash-on-delivery transactions are still relevant.
- PayPal (60,997) is the least used, but still significant:
- Shows some users prefer secure third-party gateways.
- Opportunity to promote PayPal for privacy-conscious or international users.

## Feedback Analysis :

**Overwhelmingly Positive:** The majority of the feedback is positive, with "Excellent" and "Good" being the most frequent categories.

### *Clear Descending Pattern:*

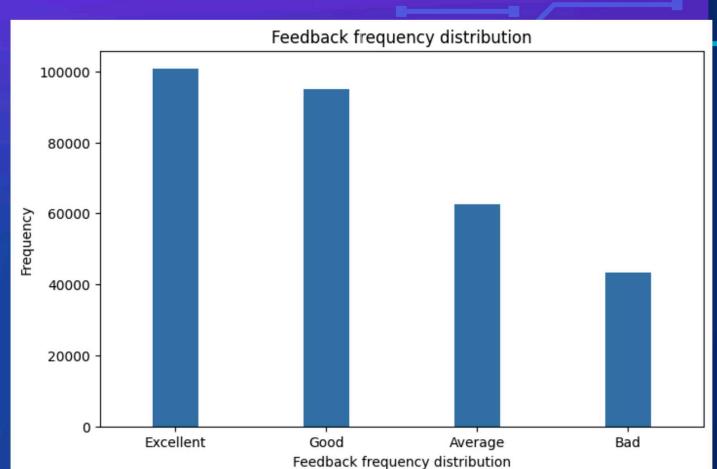
The frequency of feedback decreases as the sentiment becomes more negative, following a clear trend from "Excellent" down to "Bad".

### *Significant Negative Feedback:*

Despite the positive trend, there is a substantial amount of "Average" and "Bad" feedback (totaling over 100,000 responses), indicating a large group of users with suboptimal experiences.

### *Actionable Insights:*

While the overall picture is good, the volume of "Average" and "Bad" feedback highlights specific areas where improvements can be made.



Feedback	count
Excellent	100755
Good	95098
Average	62674
Bad	43299

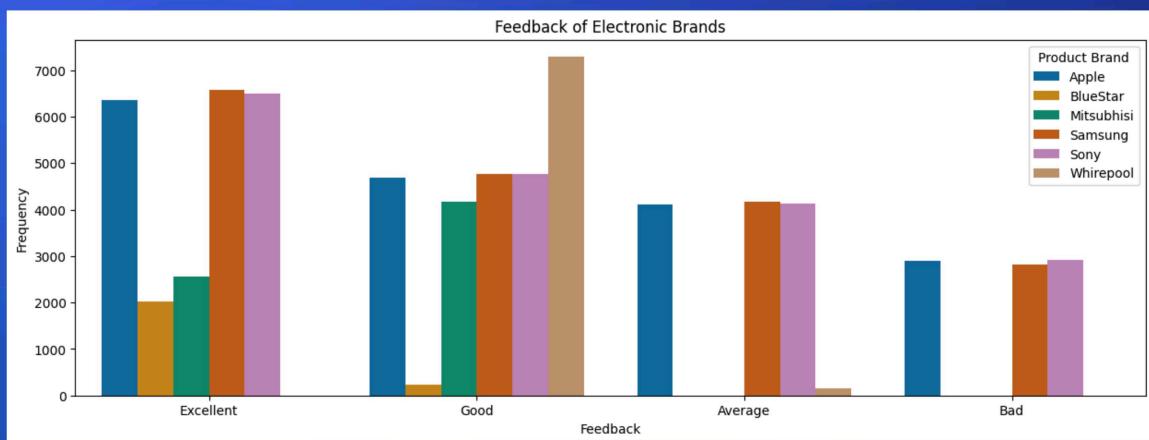
## Feedback Analysis Of Clothing Brands :



Product_Brand	Feedback	count	
0	Adidas	Excellent	6377
1	Adidas	Good	4760
2	Adidas	Average	4089
3	Adidas	Bad	2976
4	Nike	Excellent	6425
5	Nike	Good	4690
6	Nike	Average	4078
7	Nike	Bad	2917
8	Zara	Excellent	6365
9	Zara	Good	4771
10	Zara	Average	4293
11	Zara	Bad	2911

This chart shows that Adidas, Nike, and Zara all receive very similar and mostly positive customer feedback. For all three brands, "Excellent" is the most common rating. While performance is highly competitive, Nike and Adidas have a slight edge in "Excellent" reviews, and Zara receives slightly more "Average" feedback.

## Feedback Analysis Of Electronics Brands :



	Product_Brand	Feedback	count
0	Apple	Excellent	6353
1	Apple	Good	4692
2	Apple	Average	4111
3	Apple	Bad	2889
4	BlueStar	Excellent	2030
5	BlueStar	Good	226
6	Mitsubishi	Good	4171
7	Mitsubishi	Excellent	2553
8	Samsung	Excellent	6569
9	Samsung	Good	4774
10	Samsung	Average	4177
11	Samsung	Bad	2818
12	Sony	Excellent	6498
13	Sony	Good	4762
14	Sony	Average	4128
15	Sony	Bad	2908
16	Whirlpool	Good	7292
17	Whirlpool	Average	149
18	Whirlpool	Excellent	3
19	Whirlpool	Bad	1

## Feedback Analysis Of Electronics Brands :

### Key Insights

#### Overall Performance:

Samsung has the most excellent feedback with 6569 counts, followed closely by Apple (6353) and Sony (6498). Samsung also has the highest overall feedback, with a combined count of 18,338 (Excellent, Good, Average, and Bad feedback). Apple has a high number of excellent and good reviews but also has a significant amount of average and bad feedback, indicating mixed customer experiences . Sony performs well, with a strong showing in excellent and good feedback categories.

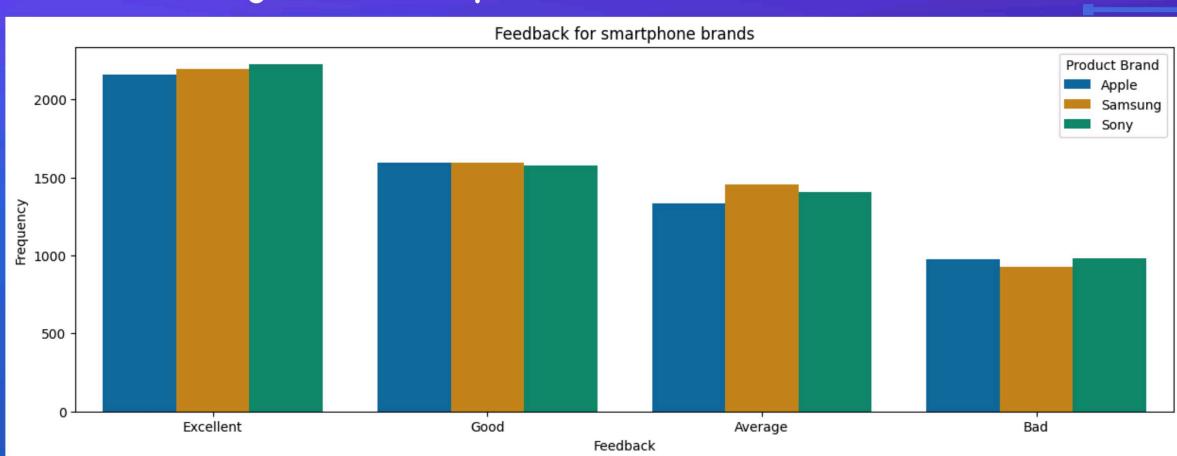
#### Negative Feedback:

Samsung has the highest count of bad feedback at 2818, indicating a potential area for improvement in their products or customer service. Sony and Apple also have a considerable amount of bad feedback, with 2908 and 2889 respectively, suggesting these top brands face similar challenges. Whirlpool has a very low count for bad feedback, but this is likely because their overall feedback count is exceptionally low.

#### Anomalies :

Whirlpool has an unusually low feedback count across all categories, with only 3 excellent and 1 bad feedback. This suggests the data for this brand is either incomplete or represents a much smaller sample size than the others. BlueStar and Mitsubishi have fewer feedback entries compared to Samsung, Sony, and Apple, but they show a good number of excellent and good reviews, which is positive..

## Feedback Analysis Of Smartphone Brands



Product_Brand	Feedback	count
Apple	Excellent	2159
Apple	Good	1594
Apple	Average	1334
Apple	Bad	973
Samsung	Excellent	2197
Samsung	Good	1595
Samsung	Average	1453
Samsung	Bad	928
Sony	Excellent	2226
Sony	Good	1578
Sony	Average	1404
Sony	Bad	982

Here is an analysis of customer feedback for three major smartphone brands:

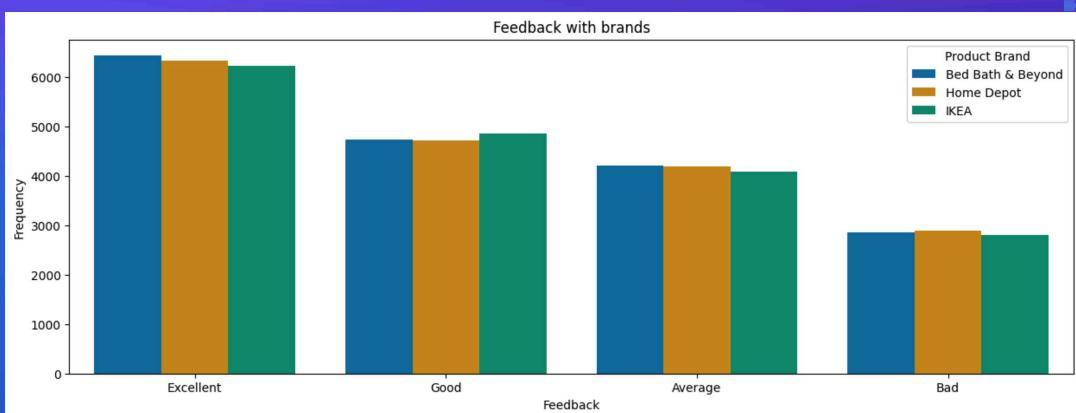
**Apple, Samsung, and Sony.**

Samsung leads in excellent feedback, with 2197 counts, slightly ahead of Sony (2226) and Apple (2159). This suggests a strong positive reception for its products.

Sony has the lowest bad feedback count, at 982, indicating better customer satisfaction in the "bad" category compared to Apple (973) and Samsung (928), a significant amount for all.

All three brands have comparable feedback counts across categories, suggesting they are highly competitive. The data shows no single brand has a definitive advantage in overall customer sentiment. The "Good" and "Average" feedback counts are also very similar, reinforcing the fierce competition among these brands in the smartphone market.

## Feedback Analysis Of Home Decor Brands



Product_Brand	Feedback	count
0 Bed Bath & Beyond	Excellent	6430
1 Bed Bath & Beyond	Good	4724
2 Bed Bath & Beyond	Average	4201
3 Bed Bath & Beyond	Bad	2859
4 Home Depot	Excellent	6322
5 Home Depot	Good	4709
6 Home Depot	Average	4193
7 Home Depot	Bad	2890
8 IKEA	Excellent	6215
9 IKEA	Good	4859
10 IKEA	Average	4087
11 IKEA	Bad	2802

**Overall Performance:** All three brands—Bed Bath & Beyond, Home Depot, and IKEA—have a strong positive reception, with "Excellent" feedback counts ranging from 6215 to 6430. This suggests a high level of customer satisfaction across the board.

**Minimal Negative Feedback:** The number of "Bad" reviews is relatively low for all brands, particularly when compared to the high volume of "Excellent" and "Good" feedback. This indicates that customers are generally happy with their purchases and that major issues are not widespread.

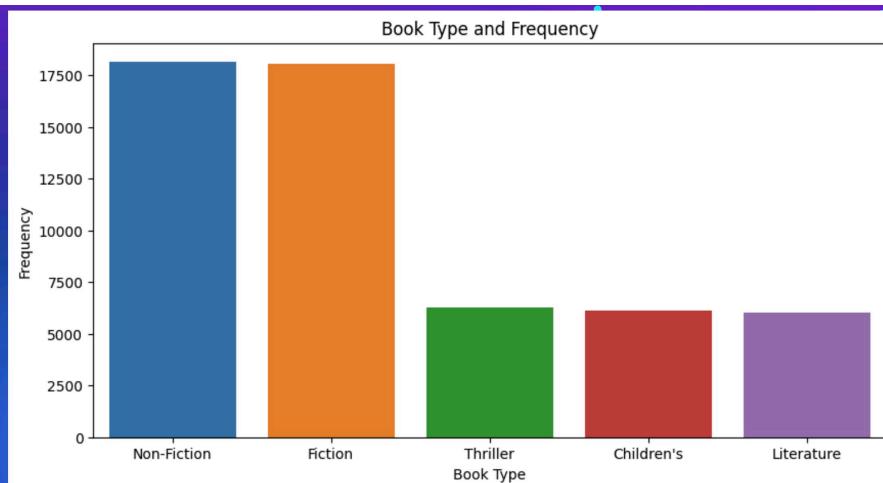
**IKEA's Strong Showing:** IKEA stands out with the lowest "Bad" feedback count at 2802. This, combined with high "Excellent" and "Good" scores, suggests that while they may not have the highest volume of positive reviews, they are effective at minimizing negative experiences.

**Competitive Landscape:** The data is very close for all three brands, with their feedback counts across all categories being remarkably similar. This suggests a highly competitive market where no single brand holds a significant advantage in customer sentiment. The slight differences in the "count" column are not substantial enough to declare a clear winner.

## Analysis on Book Selling :

### Key Insights

- Non-Fiction and Fiction are the top sellers. The data shows that Non-Fiction books have the highest frequency of sales, with a count of 18,148, followed very closely by Fiction at 18,028. This indicates that these two genres are the most popular and in-demand book types.
- Thrillers have a significant following. The Thriller genre is the third most popular with a count of 6,258. This suggests that books in this category, known for suspense and mystery, are a strong performer in the market.
- Children's and Literature genres have a lower selling frequency. The data shows Children's books with a count of 6,143 and Literature with 6,045. These genres have a considerably lower selling frequency compared to Non-Fiction and Fiction, indicating a smaller market share in this specific dataset.



Product_Type	count
Non-Fiction	18148
Fiction	18028
Thriller	6258
Children's	6143
Literature	6045

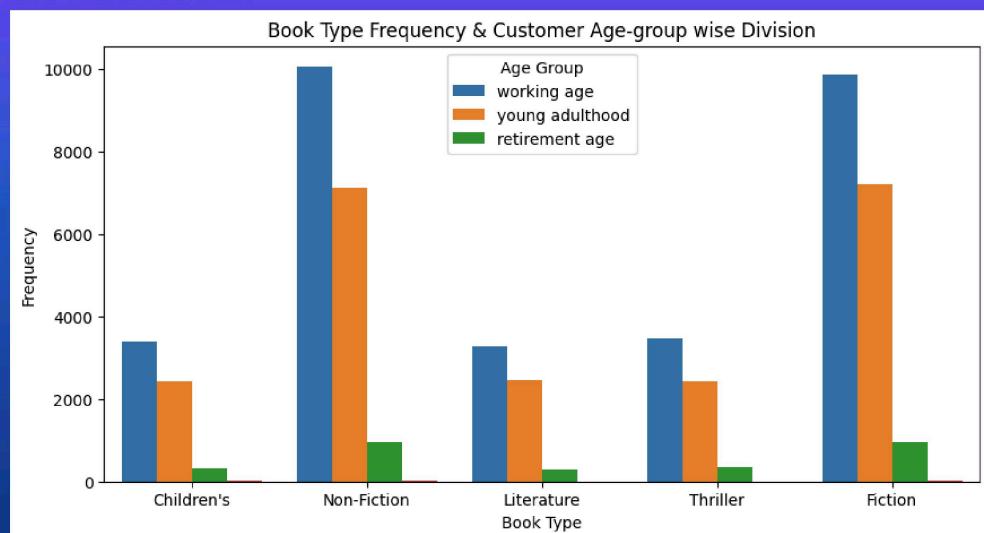
## Sold Books And Their Age Group Wise Distribution

### Age Group Insights:

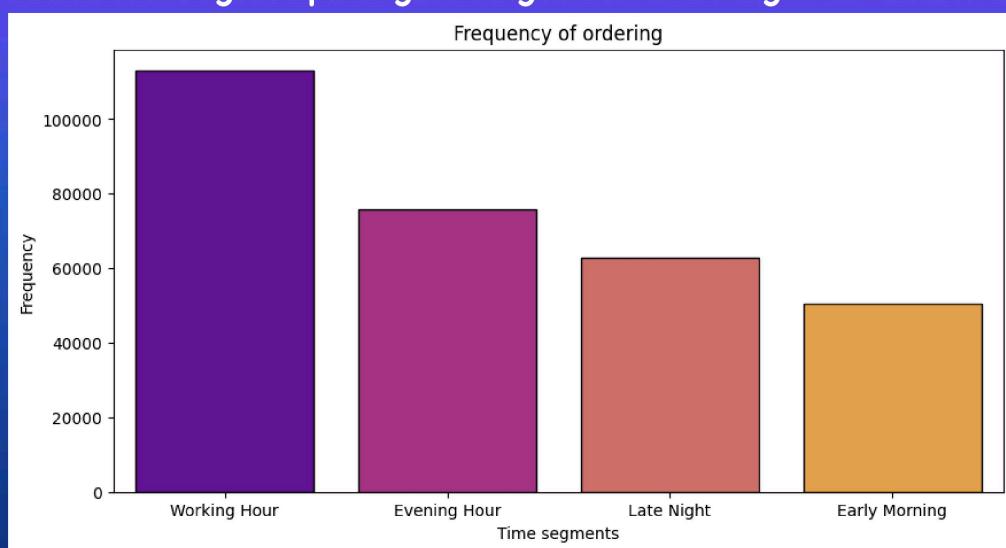
**Working Age :** This age group consistently has the highest frequency of book purchases across all genres. They are the primary consumers of Non-Fiction and Fiction, with their purchase frequency exceeding 10,000 for Non-Fiction and 10,000 for Fiction.

**Young Adulthood :** This group also shows a high frequency of purchases for Non-Fiction and Fiction, but at a lower rate than the working age group. Their frequency is around 7,500 for both genres.

**Retirement Age :** This age group has the lowest frequency of purchases across all book types. Their engagement with books is minimal compared to the other two age groups, with the highest frequency just above 1,000 for Fiction.



## Product Ordering Frequency Through Different Segments Of The Day :

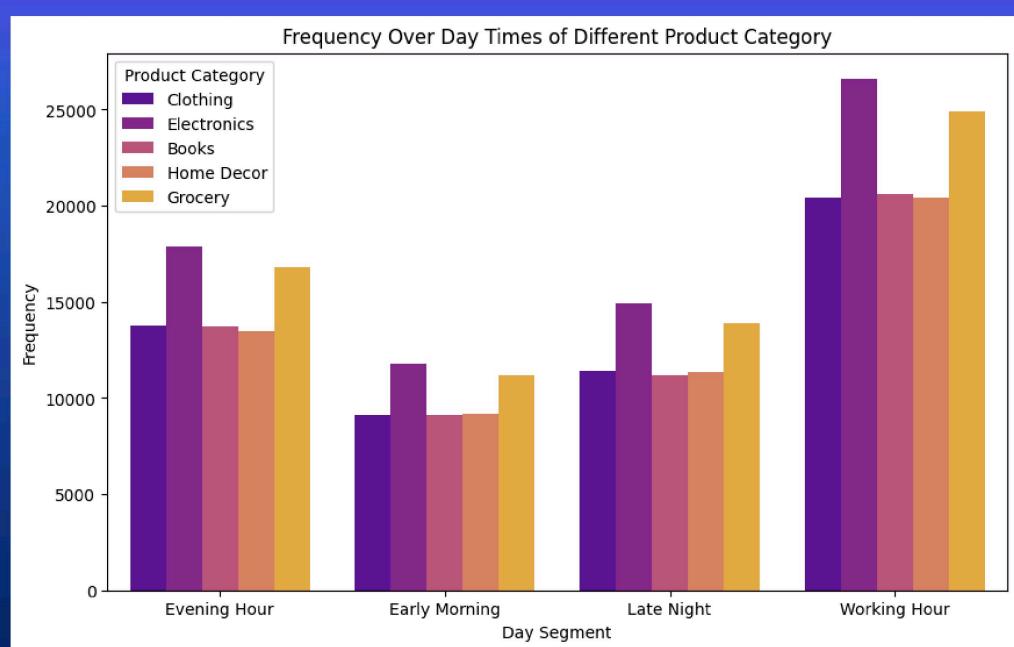


Day_segment	count
0 Working Hour	113081
1 Evening Hour	75681
2 Late Night	62801
3 Early Morning	50447

### Key Insights :

- The Working Hour is the most active for placing orders, which may seem counterintuitive but aligns with habits like shopping during lunch breaks or fulfilling work-related needs.
- Evening and Late Night show strong consumer activity as people unwind.
- Early Morning is naturally the least active due to sleep hours.

## Frequency Over Day Times of Different Product Category



Day_segment	Product_Category	Total_Purchases	
0	Early Morning	Books	48802.0
1	Early Morning	Clothing	48829.0
2	Early Morning	Electronics	63379.0
3	Early Morning	Grocery	59711.0
4	Early Morning	Home Decor	48487.0
5	Evening Hour	Books	73707.0
6	Evening Hour	Clothing	73700.0
7	Evening Hour	Electronics	95522.0
8	Evening Hour	Grocery	89883.0
9	Evening Hour	Home Decor	72184.0
10	Late Night	Books	59545.0
11	Late Night	Clothing	61714.0
12	Late Night	Electronics	79632.0
13	Late Night	Grocery	73827.0
14	Late Night	Home Decor	60197.0
15	Working Hour	Books	110021.0
16	Working Hour	Clothing	109286.0
17	Working Hour	Electronics	142877.0
18	Working Hour	Grocery	133827.0
19	Working Hour	Home Decor	110146.0

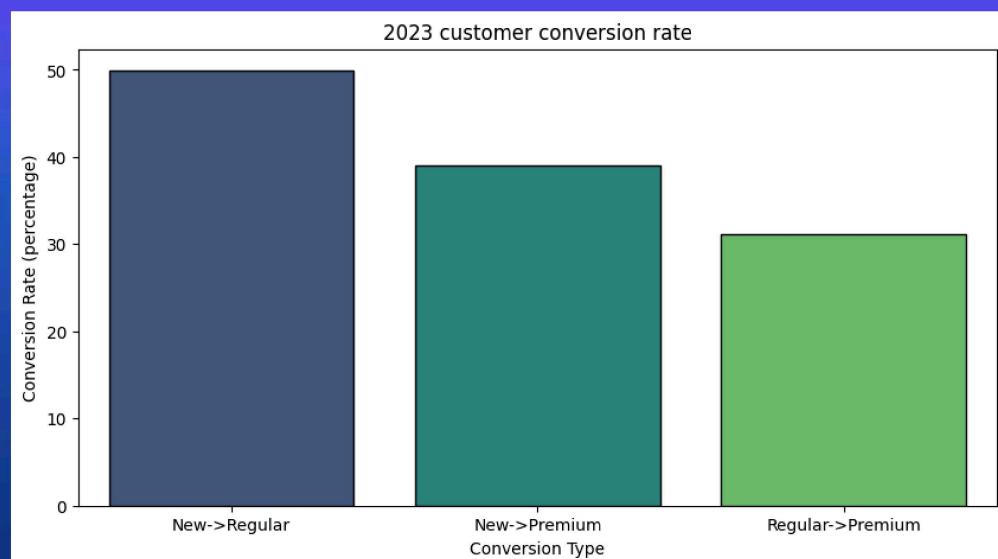
## Frequency Over Day Times of Different Product Category

### **Key Insights:**

#### **Time of Day vs Product Category Purchases**

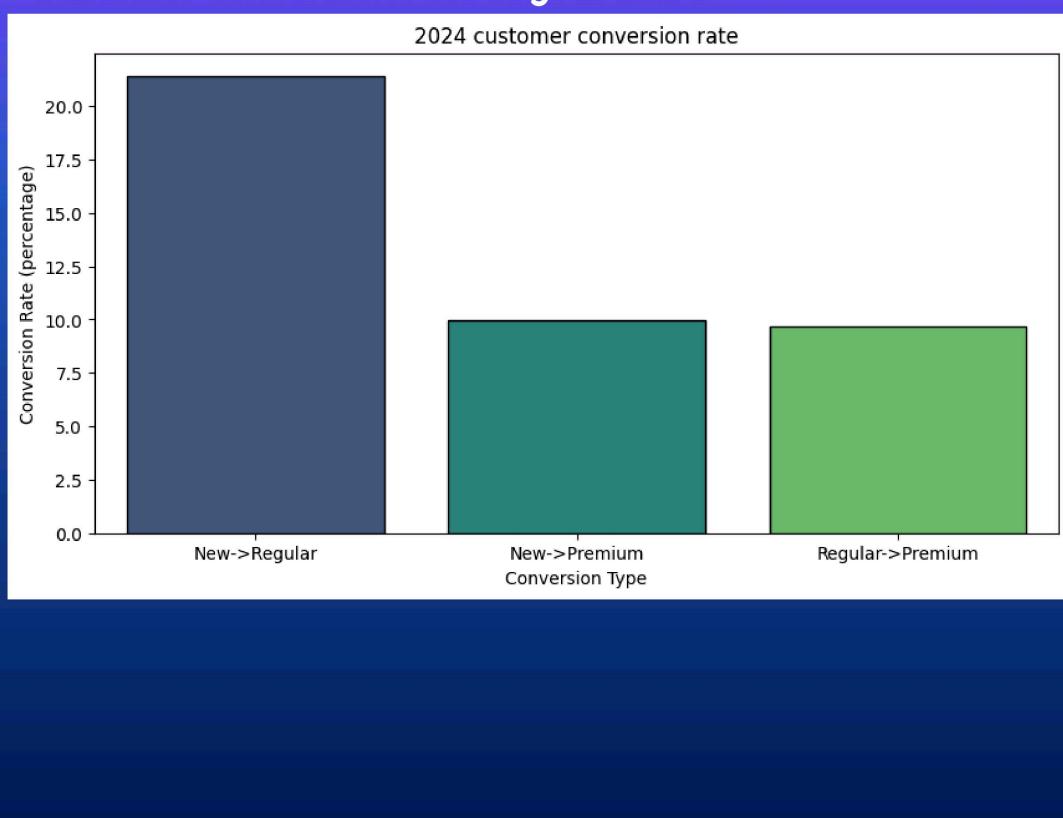
- **Working Hour** is the peak time for purchases across all product categories.
- **Electronics** is the most purchased category overall, in every time segment.
- **Early Morning** shows the lowest purchase volume across all product categories.
- **Evening Hour** is the second-busiest time for shopping, especially for Electronics and Grocery.
- **Books and Home Decor** consistently have lower purchase volumes, indicating they may be less urgent or impulse-based purchases.
- Purchases are fairly evenly spread across most categories during **Evening and Late Night**, suggesting diverse shopping interests.
- **Clothing and Grocery** also show high purchase activity, especially during **Working and Evening Hours**.
- The consistent trend across time segments suggests consumer behavior is time-sensitive, and could be optimized with targeted marketing and inventory management.

## Customer conversion rate in year 2023.



Conversion_Type	Conversion_Rate
New->Regular	49.91
New->Premium	39.06
Regular->Premium	31.11

## Customer conversion rate in year 2024 :



Conversion_Type	Conversion_Rate	
0	New->Regular	21.41
1	New->Premium	9.96
2	Regular->Premium	9.68

## Customer conversion rate Analysis and Insights :

### *Key Insights :*

- **High initial conversion:** Nearly 50% of new users become regular, showing effective early engagement.
- **Strong early premium adoption:** 39% of new users go directly to premium, showing that premium features are attractive even to newcomers.
- **Lower conversion from Regular to Premium (31.1%)** may suggest that:
  - a. Regular users feel comfortable with free features.
  - b. There's less perceived added value in upgrading.
  - c. There might be an opportunity to optimize upselling tactics for this group.

### **Business Implications:**

- Enhance upsell strategies for regular users to push them toward premium (e.g., exclusive content, loyalty discounts).
- Continue promoting the value of premium plans to new users, since direct conversion is already strong.
- Analyze why 50% of new users don't convert at all — identify pain points or drop-offs in the customer journey.

# TEAM



Jibon Mondal



Ratul Mishra



Biswajit Mondal



Laxmi Mahata



# THANK YOU