DATA-DRIVEN INSIGHTS & PERFORMANCE ANALYSIS

COMPELLING STATISTIC RELATED

BY: Ajay Suyal
Intern

BUSINESS CONTEXT

- The client, a leading retail chain in India, seeks data-driven insights.
- CRM, marketing, campaign and sales strategies.
- Measure, manage, and analyze business performance.
- Seasonality impact, cross-selling, customer satisfaction, cohort analysis, and store-level analysis.

Business Objective :

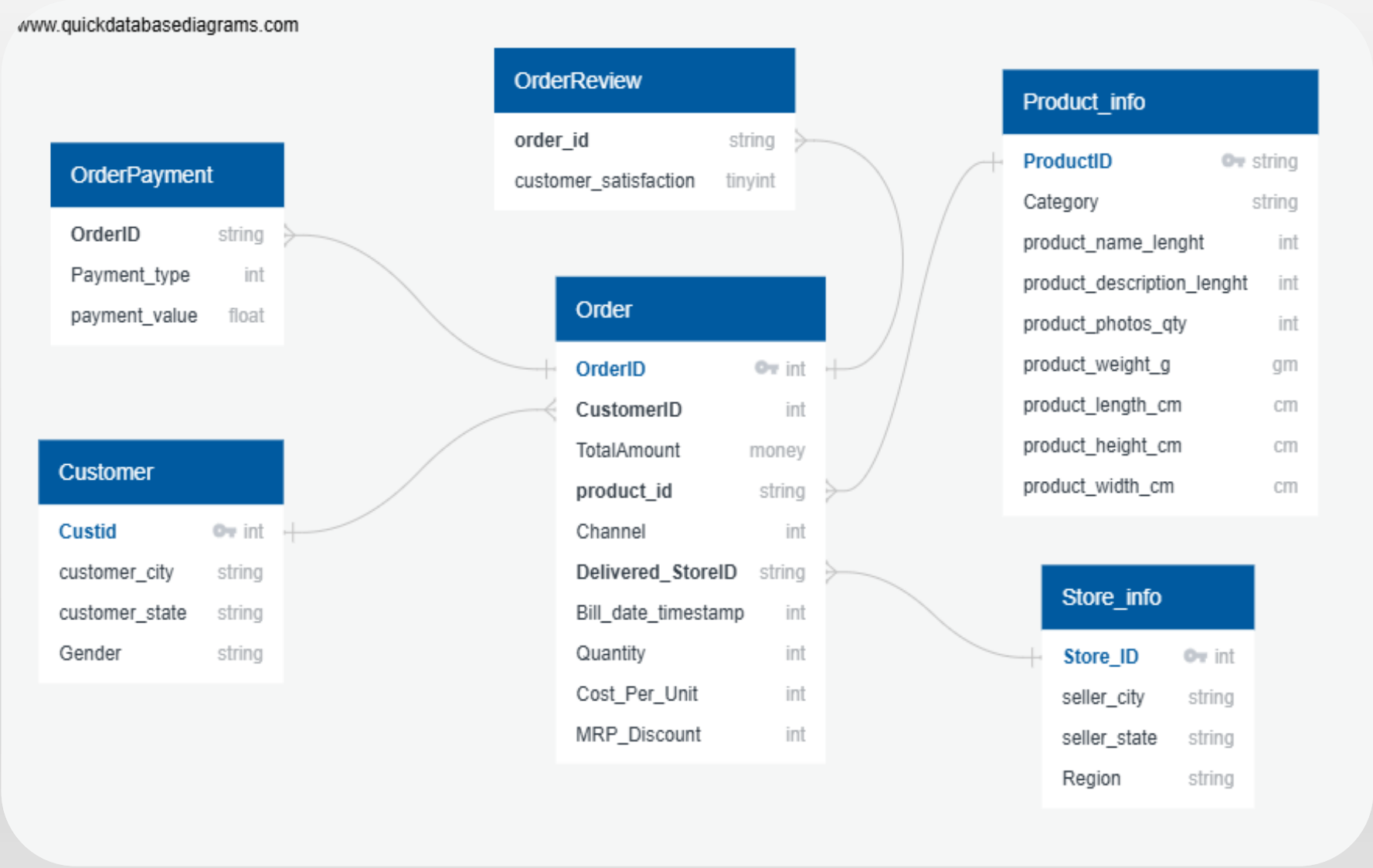
Retail client is one of leading retail chain in India facing challenge in their business :

- Identifying the Sales of company from the store chain
- Understanding the acquire of new customer
- Understanding customer behaviour, product analysis
- Understand the trends/seasonality of sales, quantity, store performance
- Understand the behaviour of discount seekers & non discount seekers customer

Overview Of Tables

TABLE NAME	NO. OF COLUMN	NO. OF ROW	PRIMARY KEY	FOREIGN KEY	DESCRIPTION
CUSTOMER	4	99,441	CUSTID	NA	Customer table have id, address, gender
ORDER	11	1,12,650	ORDER ID	CUST ID,STORE ID,PRODUCT ID	Order Table have order related
ORDER PAYMENT	3	1,03,886	NA	ORDER ID	In this have payment for the order id
ORDER RATING	2	1,00,000	NA	ORDER ID	Rating of customer satisfaction
PRODUCT INFO	9	32,951	PRODUCT ID	NA	This table contain product details
STORES INFO	4	535	STORE ID	NA	Store table have store id and their location

Entity Relationship Diagram



- Customers place Orders (One-to-Many)
- Orders have Payments (One-to-One)
- Orders have Reviews (One-to-One)
- Orders include Products (Many-to-Many)
- Stores fulfill Orders (Many-to-Many)

Data Overview Slide

- Data has been provided from Sep 2021 to Oct 2023
- We have provide 6 files in which have customer and order details
- The given Data from 39 stores out of 535
- Their have multiple Payment method
- Mostly Customers are female
- Avg. rating of order is 4
- Total product category is 13
- Wrong datatype of Datetime column in order

Data Audit 1/3

Customer Data:

Columns: Cust ID, Customer City, Customer State, Gender

- No missing values found, No zero value
- No special character
- 866 customer not purchased any thing

Data Audit 2/3

Order Data:

Columns: Order Id, customer Id, Product ID, Channel, Delivered Store ID, Bill date timestamp, Quantity, Cost per Unit, MRP, Discount, Total amount

- ❑ Wrong Date time format
- ❑ No negative value, No encoding value, missing value
- ❑ 7088 records found data inconsistency
- ❑ No mismatch of data
- ❑ Total Deliver storied are 37
- ❑ Discount is more than 100 it may be Discount value

Order Payment Data:

Columns: Order Id, Payment Value, Payment Type

- ❑ 615 duplicated records found
- ❑ 2961 records found data inconsistency
- ❑ 9 payment value are zero
- ❑ No encoding, missing value

Data Audit 3/3

Order Review Rating:

Columns: Order Id, Customer Satisfaction

- 555 record found data inconsistency
- 315 record found duplicate
- 776 reviews different orders from order table

Product info:

Columns: Product ID, Category, Product name Length, Product Description Length, Product photos qty, Product Weight, Product length, Product height, Product width

- 623 product category are missing
- No duplicated value

Store Info:

Columns: Store id, Seller city, Seller State, Region

Data Discrepancies 1/2

- **payment table found 2961 order id different payment type, Payment value**

Order id	payment_type	payment_value
0016dfedd97fc2950e388d2971d718c7	voucher	17.92000008
0016dfedd97fc2950e388d2971d718c7	credit_card	52.63000107
002f19a65a2ddd70a090297872e6d64e	voucher	33.18000031
002f19a65a2ddd70a090297872e6d64e	voucher	44.11000061
0071ee2429bc1efdc43aa3e073a5290e	voucher	92.44000244
0071ee2429bc1efdc43aa3e073a5290e	voucher	100

- **In the order table 1071 order id have different delivery store id**

order_id	Delivered_StoreID
fe64170e936bc5f6a6a41def260984b9	ST129
fe64170e936bc5f6a6a41def260984b9	ST106
fe678293ea3bb6607a15b2e320e91722	ST167
fe678293ea3bb6607a15b2e320e91722	ST133

Data Discrepancies 2/2

- Found zero(0) payment value in Payment table

Order id	Payment type	Payment value
8bcbe01d44d147f901cd3192671144db	voucher	0
fa65dad1b0e818e3ccc5cb0e39231352	voucher	0
6ccb433e00daae1283ccc956189c82ae	voucher	0

- For the similar order and product id the no of quantity increasing

Customer id	Order id	Product id	Channel	Quantity	Cost Per Unit	MRP	Discount	Total Amount
1115885858	713eda1fb337fff2cccfae60fd0b411e	781afe929e3016a667f5f439afd55fce	Instore	1	109.9	122.17	0	122.17
1115885858	713eda1fb337fff2cccfae60fd0b411e	781afe929e3016a667f5f439afd55fce	Instore	2	109.9	122.17	0	244.34
1115885858	713eda1fb337fff2cccfae60fd0b411e	781afe929e3016a667f5f439afd55fce	Instore	3	109.9	122.17	0	366.51
1115885858	713eda1fb337fff2cccfae60fd0b411e	781afe929e3016a667f5f439afd55fce	Instore	4	109.9	122.17	0	488.68

Data Cleaning Steps

- In Orders table we have multiple order id have increasing no of quantity we consider only the max quantity
- For similar order id have multiple store id we can choose Only one of them
- Delete these rows which are not between the specific range
- Replace the #N/A in product category with other and Null with 0
- Aggerate the order review table on order id
- Aggerate the

Key Performance Indicators

SALES INFORMATION

Total Revenue	15.4 M
Total Cost	13.2 M
Total Profit	2.17 M
Total Discount	48.99 K
Discount Percent	3.18%

STORE INFORMATION

No Of State	7
No Of City	37
No Of Region	4
Product Category	14

Key Performance Indicators

CUSTOMER INFORMATION

Total Customers	96,803
No Of Customer City	4,092
Avg Spend Per Customer	159.15
Avg Discount Per Customer	5.06
Repeated Buyers	36
Avg Inactive Days	253
Avg Rating	4.09

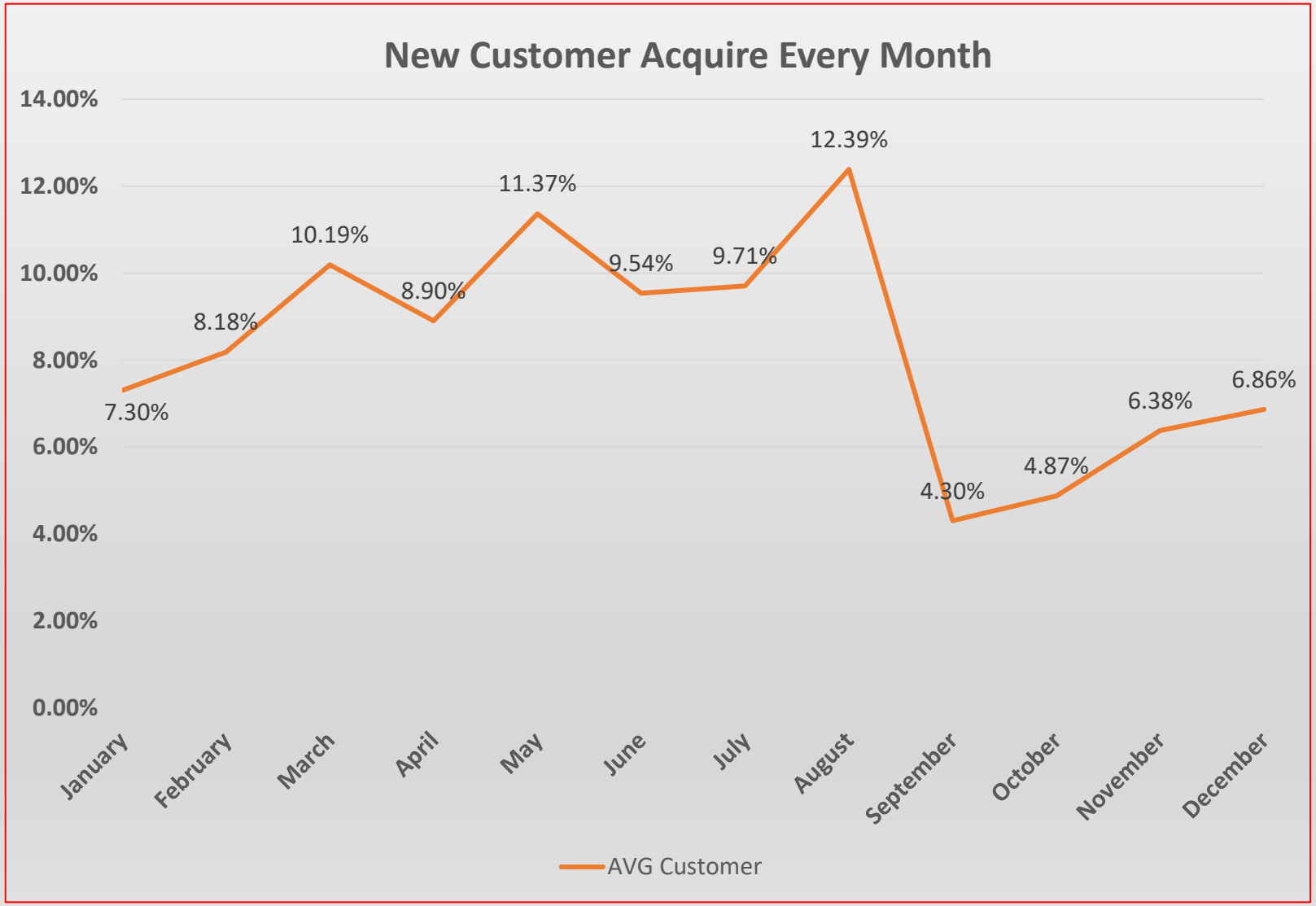
ORDER INFORMATION

Total Orders	96,894
Total Quantity	1,08,388
Avg Quantity Per Order	1.12
Avg Discount Per Order	5.05
Avg Order Value	159.00
Avg Product Per Order	1.01
Avg Rating	4.09

Exploratory Data Analysis

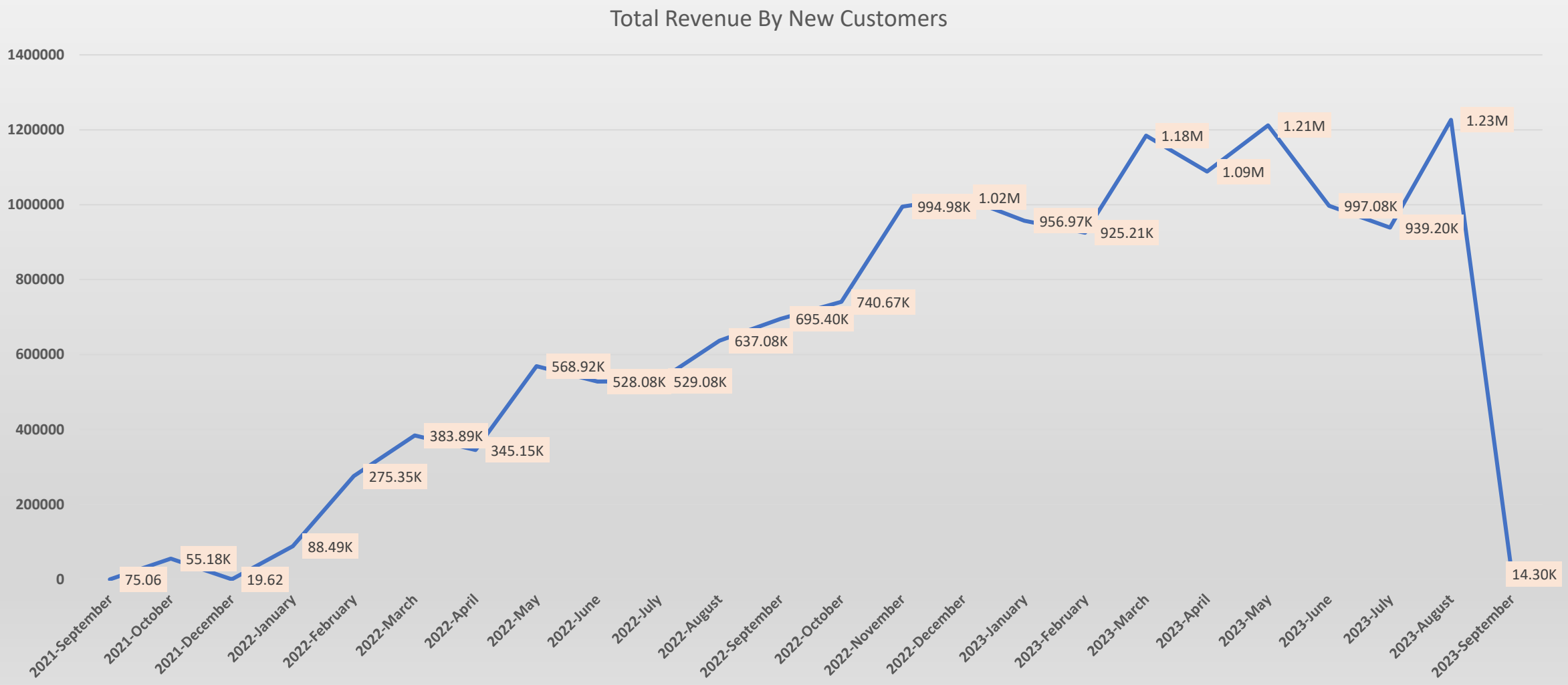
- Top Performer: ST103 is the highest-revenue store, generating 3.29M, Profit is 489.96K
- The gap between revenue and profit across most stores suggests potential areas for cost reduction, pricing optimization
- Store 103 have highest revenue but the avg order value is 131.92

CUSTOMER ACQUISITION

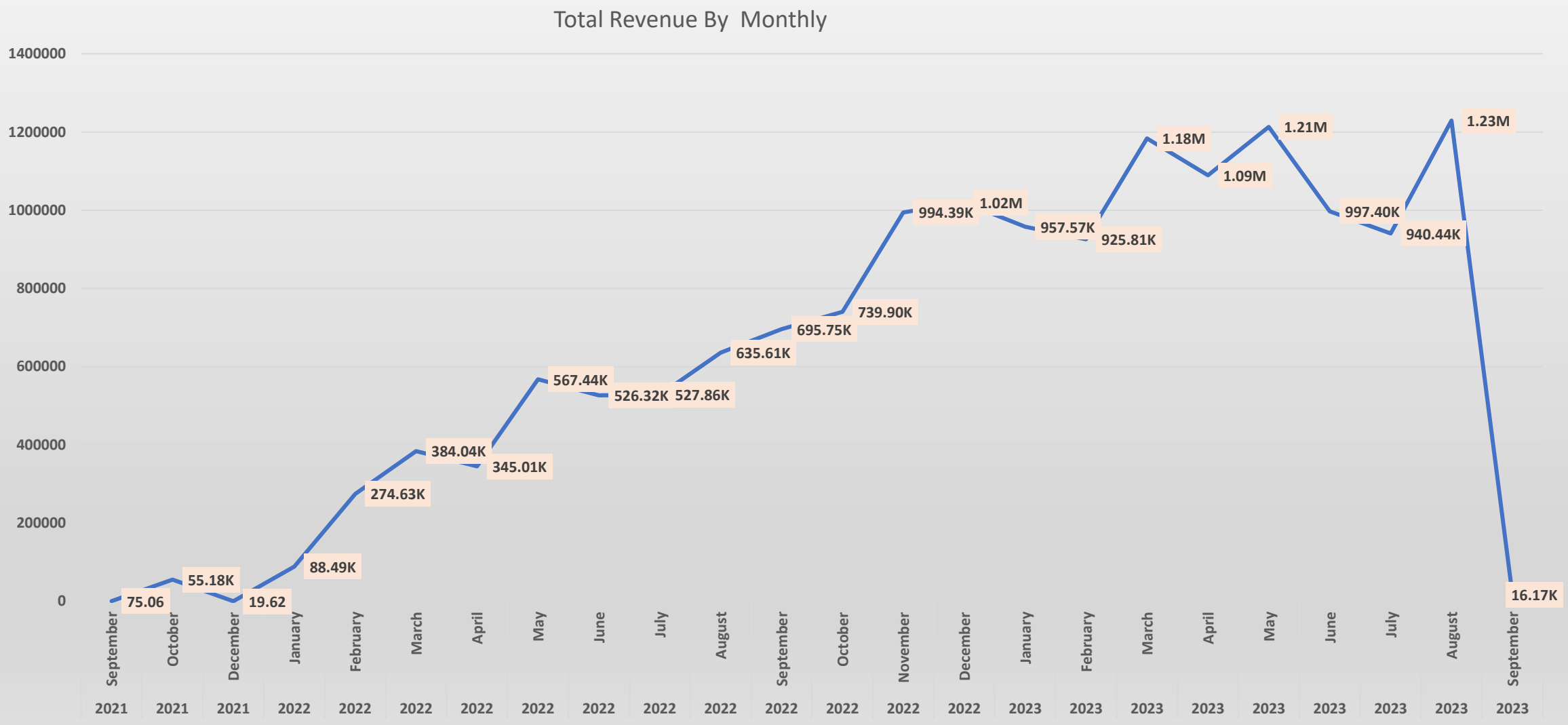


- May (11.37%) and August (12.39%) saw the highest new customer percent
- September (4.30%) and October (4.87%) had the lowest acquisitions.

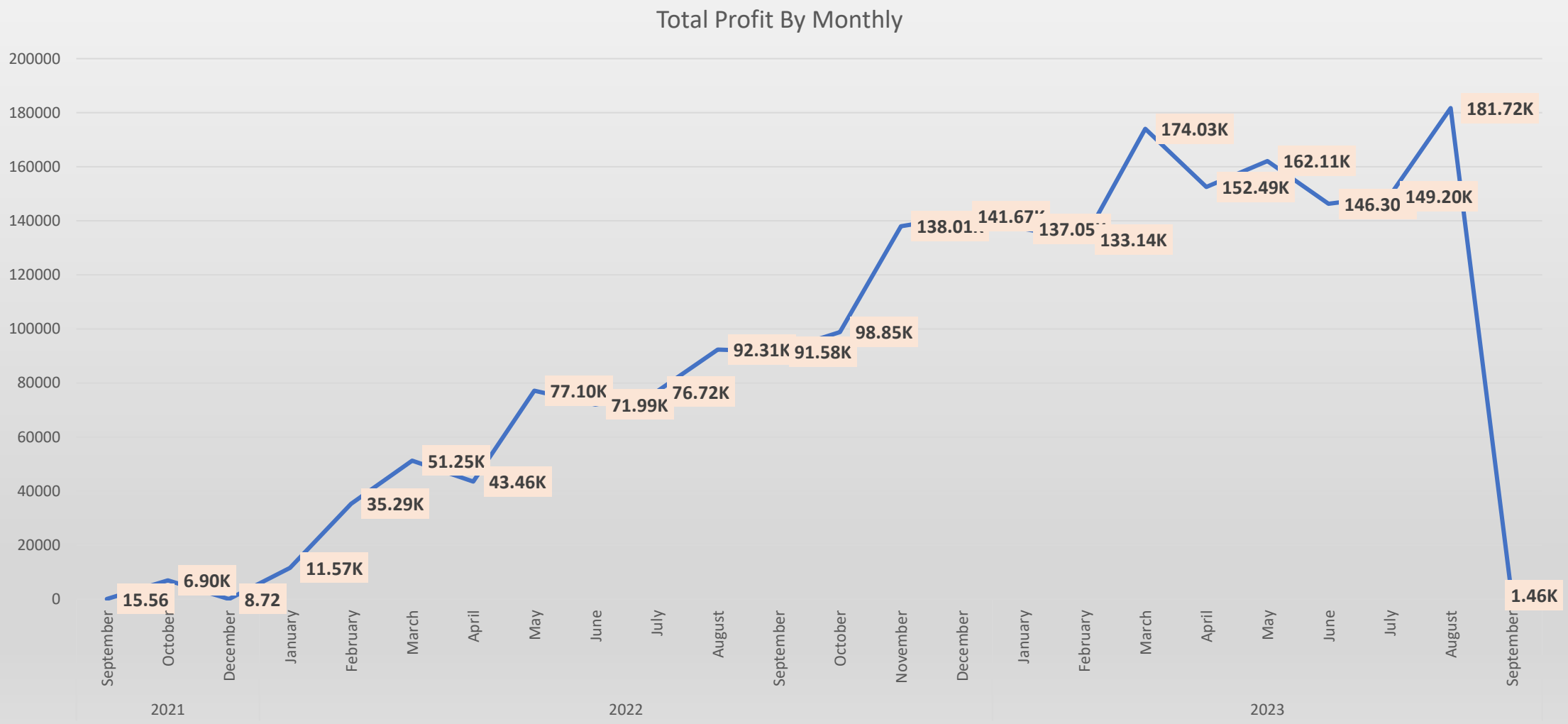
Monthly Revenue By New Customers



Monthly Revenue BY Customers



Monthly Profit By Customers



Exploratory Data Analysis

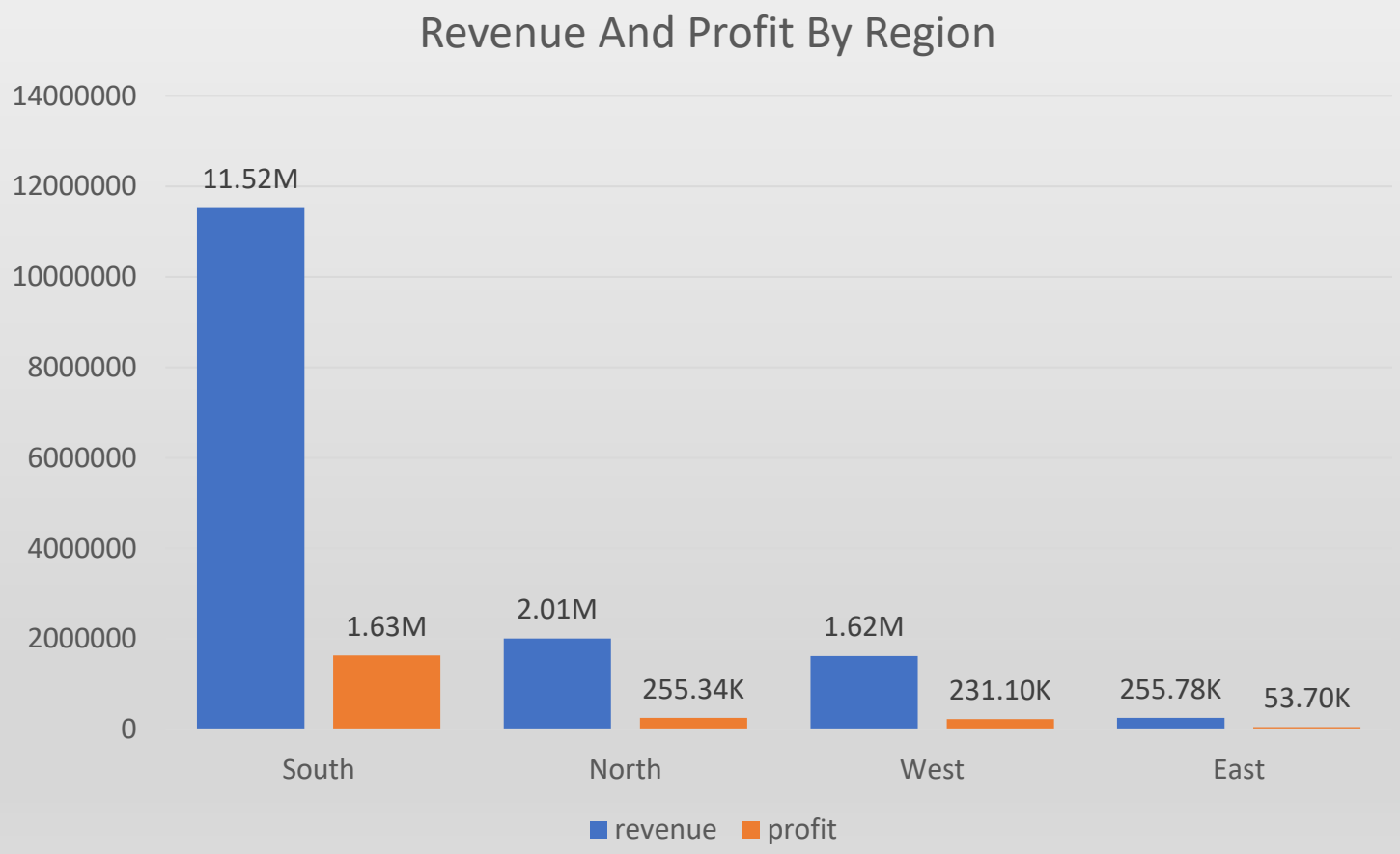
- Insights :

  - ❖In the chart we can see that the Revenue from new customers are increasing.
  - ❖Total revenue from customers are increasing but the no share is major from the new customers
  - ❖ From march to august the Revenue shown some decline this may be the seasonal.
  - ❖The profit collection follow the revenue

- Recommendations:

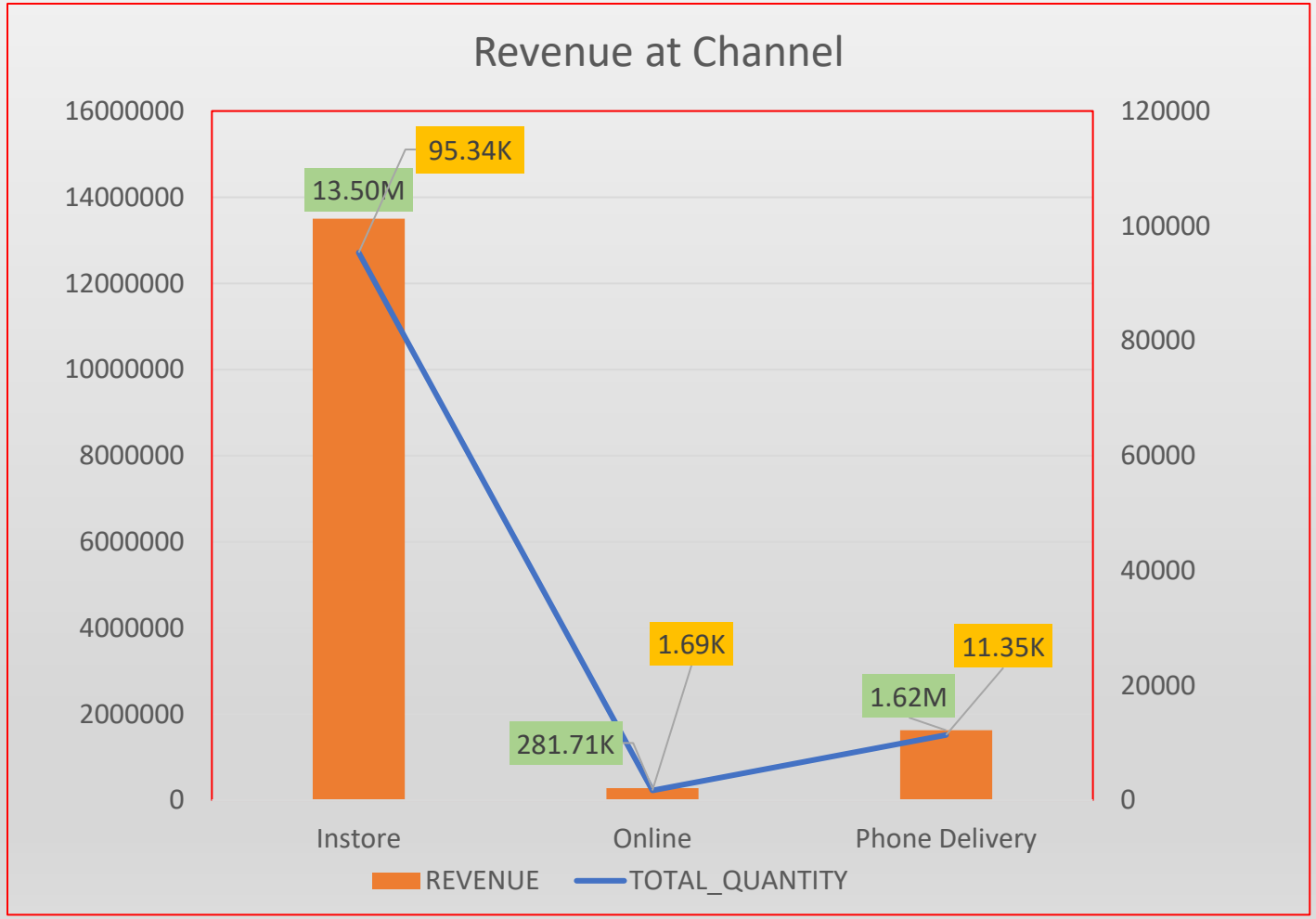
  - ❖ Need to focus on the customer retention to increase the revenue from old customers.
  - ❖Profit follow the revenue But margin are low we should focus on to increase the margin.
  - ❖Customer acquisition are increasing but some fall shown in may to august we can focus at that time

Revenue And Profit By Region



- South region dominates revenue (11.5M), whereas East has the least (255K)
- North and West are nearby same at revenue generated
- Increase the presence in north and west for increase the revenue.

Channel Used By Customers



- In-store contributes the highe4st revenue(13.50M)
- Lowest contributes by Online then Phone-Delivery
- Increase marketing spend online and promotion

Top 10 Selling product

Exploratory Data Analysis

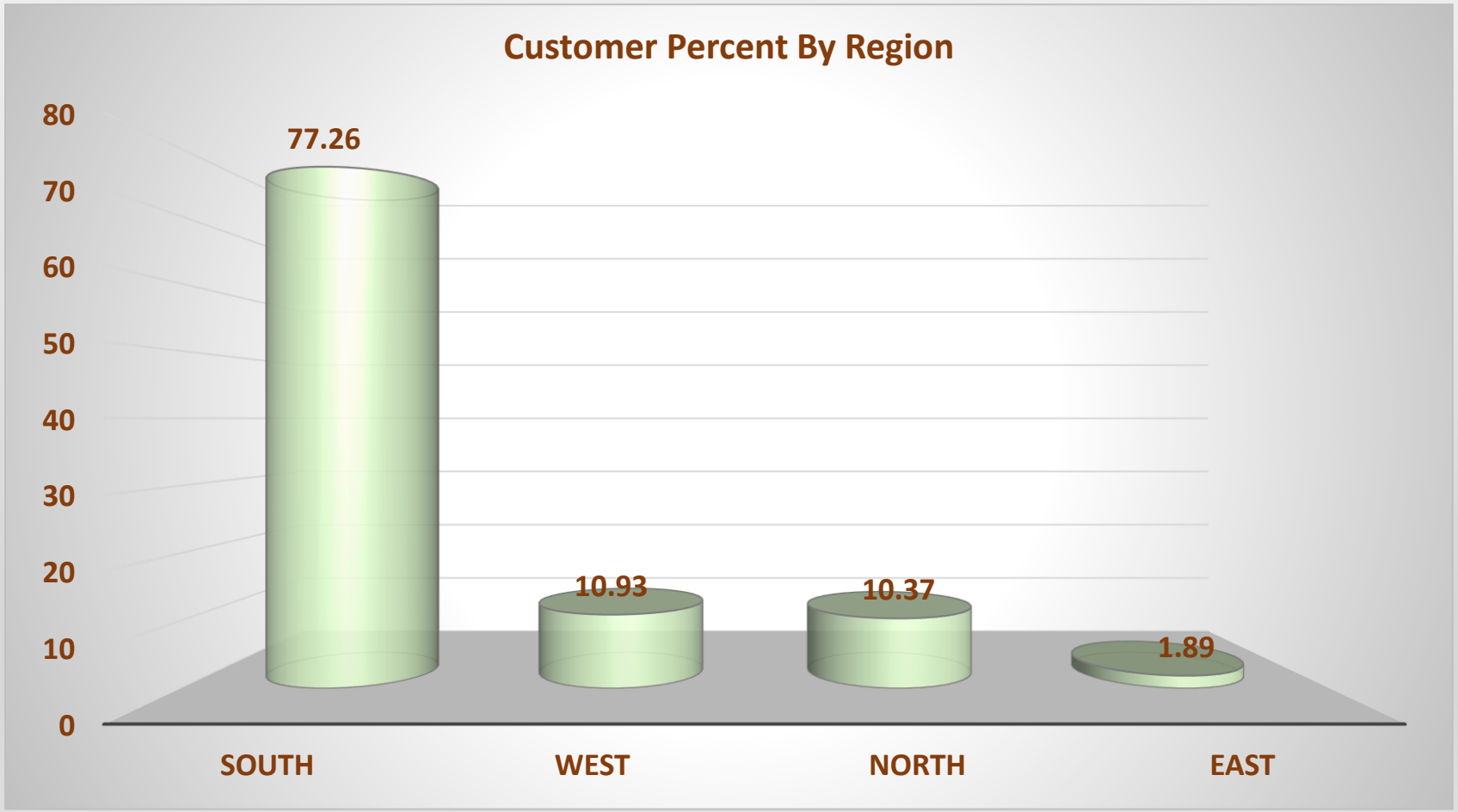
Insights :

- March & April, May & August had the highest new customers .
- September–October acquisition drops sharply during these months.
- Monthly revenue increase which is good for Company.
- Measure revenue from new customers

Recommendations:

- Company focus to increase the Top 10 selling product quantity.
- Seasonal promotions to acquisition of new customers.
- Focus to target old customer to increase revenue.
- Optimize operational costs at top-performing stores to boost profits.
- Expand product offerings in low-performing stores/regions.

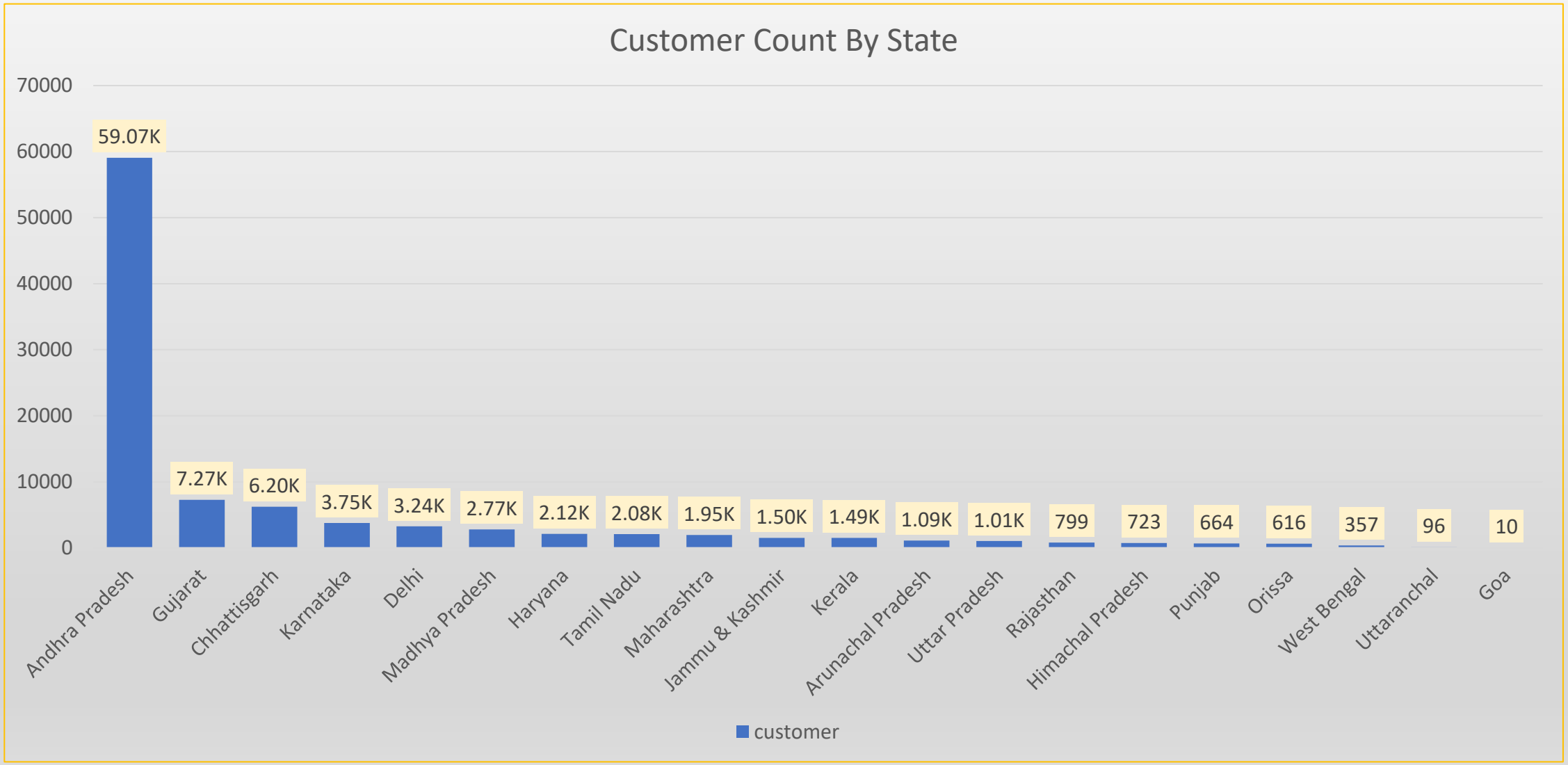
Customer Behaviour



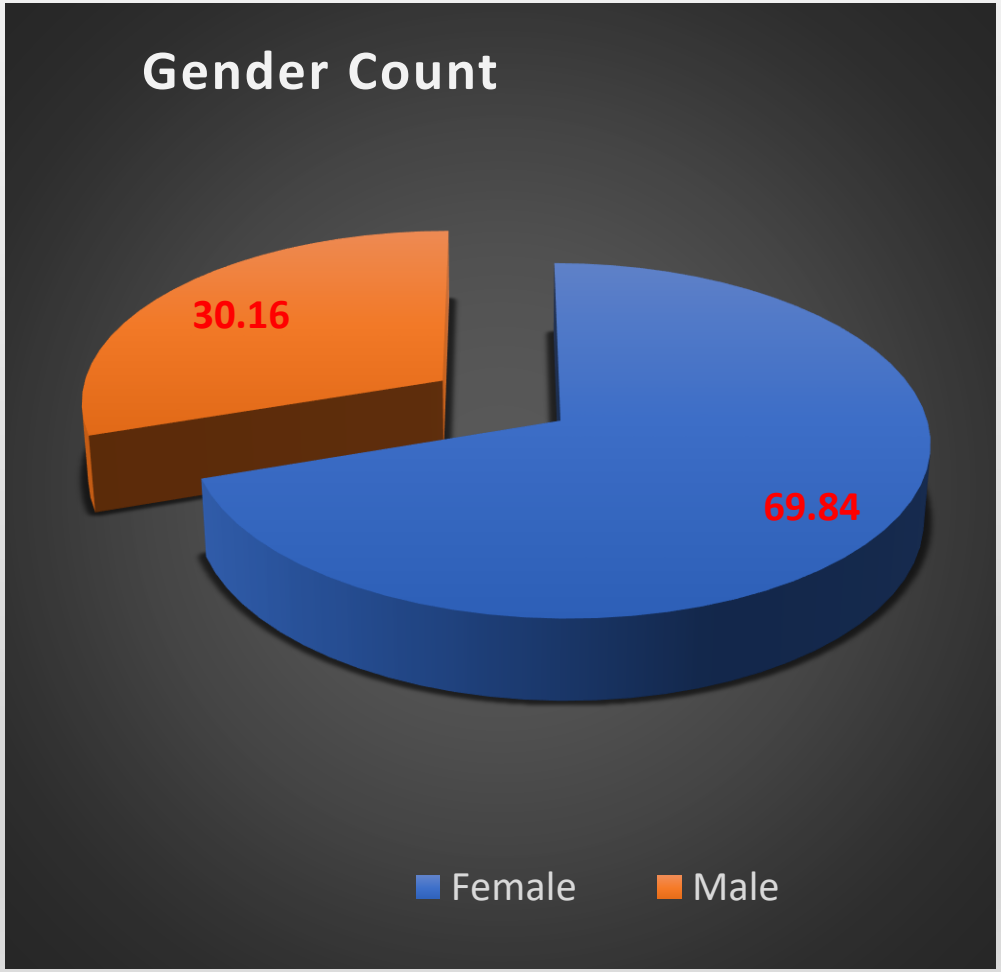
The South region dominates with the highest customer percentage (77.26) and the East region with least percent (1.69)

Company have to focus on increase the customer present at every region

Customer Behaviour



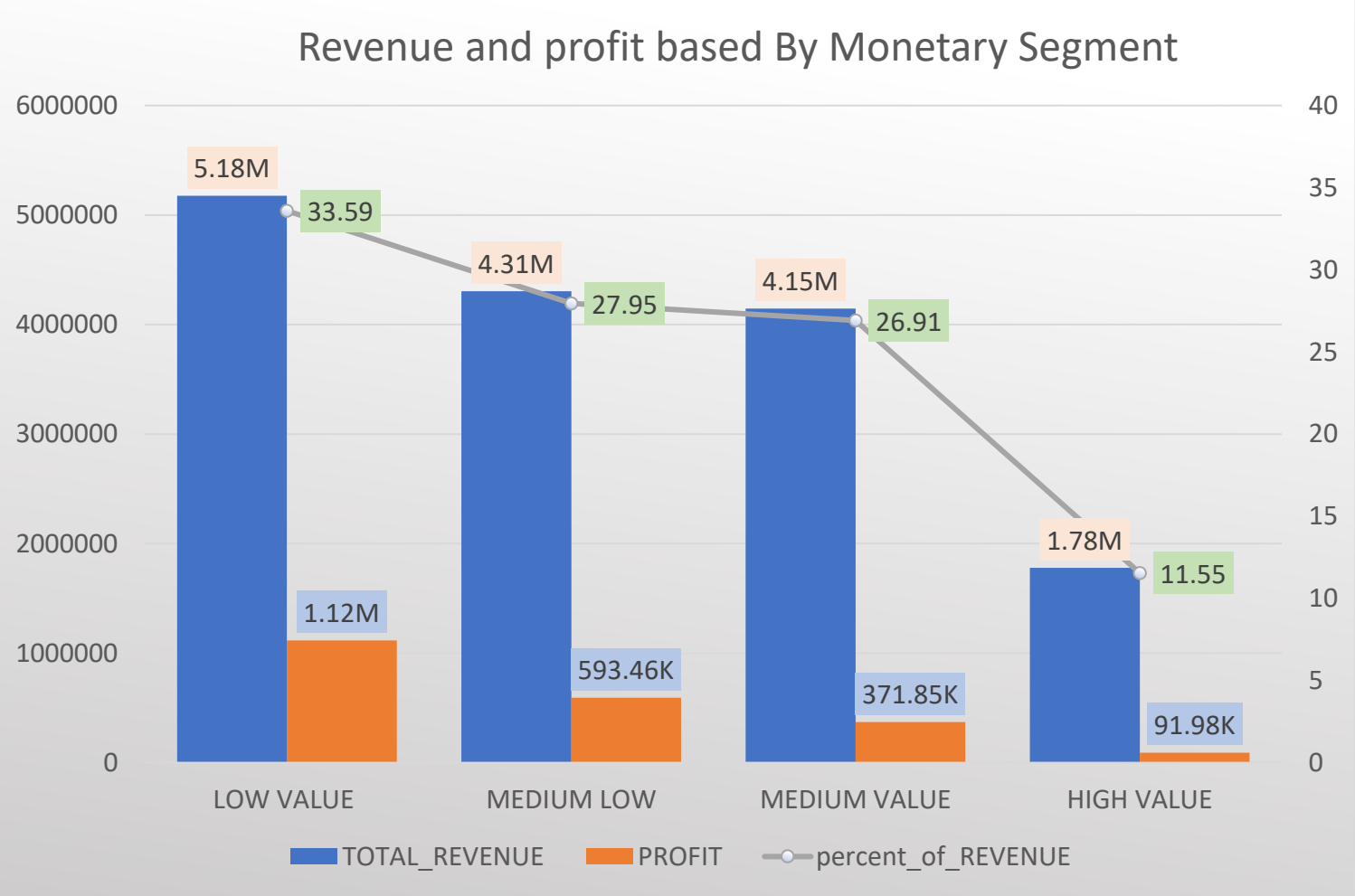
Customer Count By Gender



Gender	FEMALE	MALE
TOTAL CUSTOMER	67,608	29,195
REVENUE	10.78 M	4.64 M
DISCOUNT	343.73K	146.25 K
TOTAL ORDERS	67,679	29,215

- Mostly Customers are Female
- Female Customers are contributing 10.78 M in revenue
- We have to increase the customer base mostly focus on Female they are mostly in the no.

Customer Behaviour



- High Revenue Customers Having Revenue (>=1000)
- Medium Revenue Customers Having Revenue Between (300 To 1000)
- Med-low Revenue Customers Having Revenue Between (150 To 300)
- Low Revenue Customers Having Revenue Less Than 150

Customer Behaviour

RFM Segmentation

- ❑ RECENCY : How Recently Customer Purchased
- ❑ FREQUENCY : No. Of Orders Customer Makes A Certain Period
- ❑ MONETARY : How Much Customer Spend On Purchase In Certain Period

Recency Segment

- Active(Score=4) If Days < 90
- Semi-Active(Score=3) If Days Between 90 to 180
- At Risk(Score=2) If Days Between 180 to 270
- Churn(Score=1) If Days < 270

Frequency Segment

- High(Score=4) if Orders > 4
- Mid (Score=3) if Orders = 3 & 4
- Mid-Low(Score=2) if Orders = 2
- Low(Score=1) if Orders = 1

Monetary segment

- High(Score=4) if Revenue >1000
- Mid (Score=3) if Revenue Between 300 to 1000
- Mid-Low(Score=2) if Revenue Between 150 to 300
- Low(Score=1) if Revenue < 150

RFM Segmentation Steps

Each customer is assigned a score for Recency, Frequency, and Monetary by dividing into four groups.

➢ R Score: A score for recency, where the most recent buyers get higher scores(4).

➢ F Score: A score for frequency, where customers who have made more purchases get higher scores(4).

➢M Score: A score for monetary value, where customers who have spent more money get higher scores(4).

➢ Total scoring: This step combines the R, F, and M scores into a Final score by adding R Score, F Score, and MScore.

➢ A customer with the lowest scores for recency, frequency, and monetary will have a total score of 3 (1 + 1 + 1), while a customer with the highest scores in all three categories will have a total score of 12 (4 + 4 + 4)

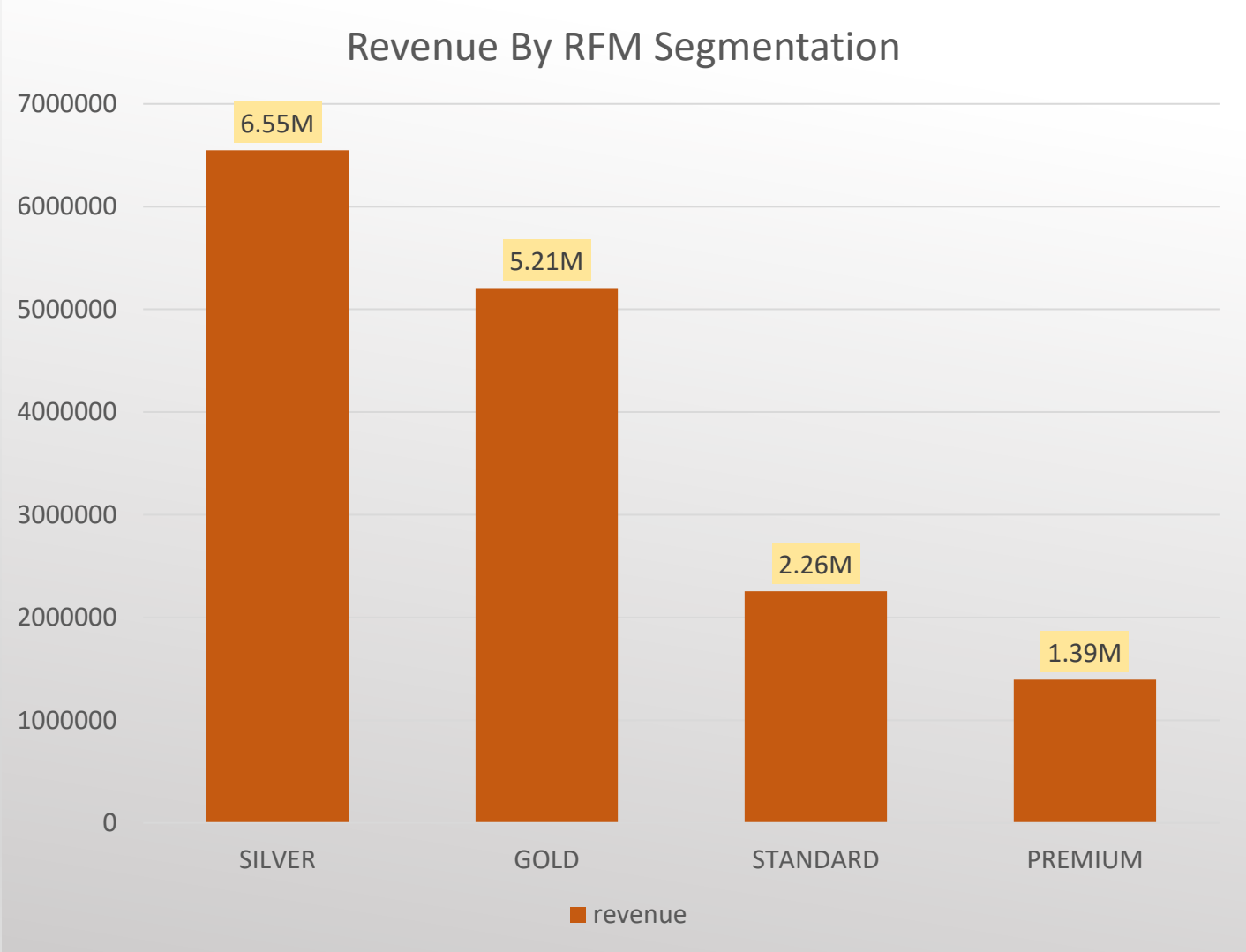
➢ Based on the Final score, customers are classified into different segments:

- Premium: Total score between 12 and 08
- Gold: Total score between 06 and 08.
- Silver: Total score between 04 and 06
- Standard: Total score less than 04

Customer Behaviour

RFM SEGMENT	NO OF CUSTOMER	CUSTOMER PERCENT	TOTAL ORDERS	TOTAL REVENUE
PREMIMUM	1,804	1.86	1,862	1,394,131.29
GOLD	22,874	23.63	22,899	5,208,502.98
STANDARD	28,755	29.7	28,755	2,256,297.58
SILVER	43,370	44.8	43,378	6,547,542.58

Customer Behaviour



❖Most revenue came from Silver segment (6.55M) not Premium or Gold.

❖Silver Segment (Score 4–6): Largest group with 43.37K customers.

❖Churn risk is high for Standard customers (score <4), who form a substantial portion.

CUSTOMER BEHAVIOUR

RFM Segmentation :

Insights :

- The majority of customers fall into the Silver segment 43.37K(44.8%)
- The distribution indicates that a large portion of customers are in the lower-value (Silver & Standard) segment, with fewer customers in the higher-value (Gold & Premium) segments.
- The maximum revenue generated from Silver segment Customers.

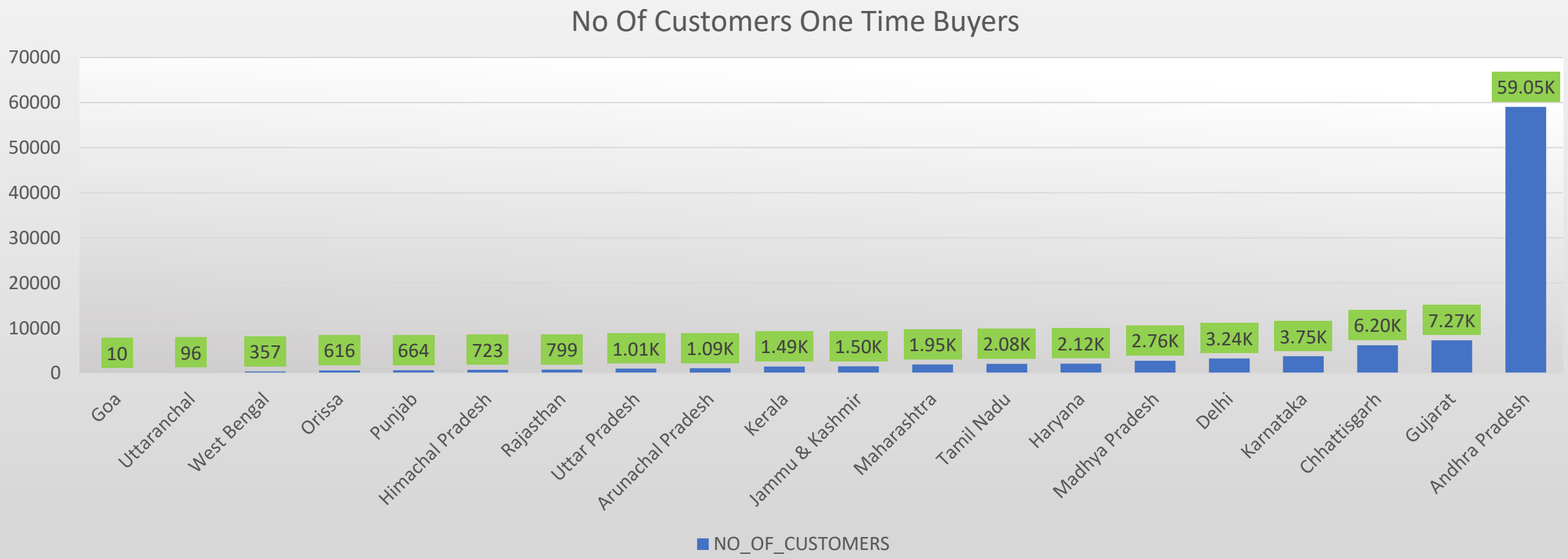
Recommendations:

- Develop strategies to move customers from the Standard & Silver segment to higher-value segments like Gold & Premium.
- Implement loyalty programs & Incentives for Gold and Platinum customers to maintain their loyalty and encourage higher spending.
- Improve after sales service to moving Standard to higher-value segments. Highlight the advantages of frequent purchases and higher spending.

Customer Behaviour

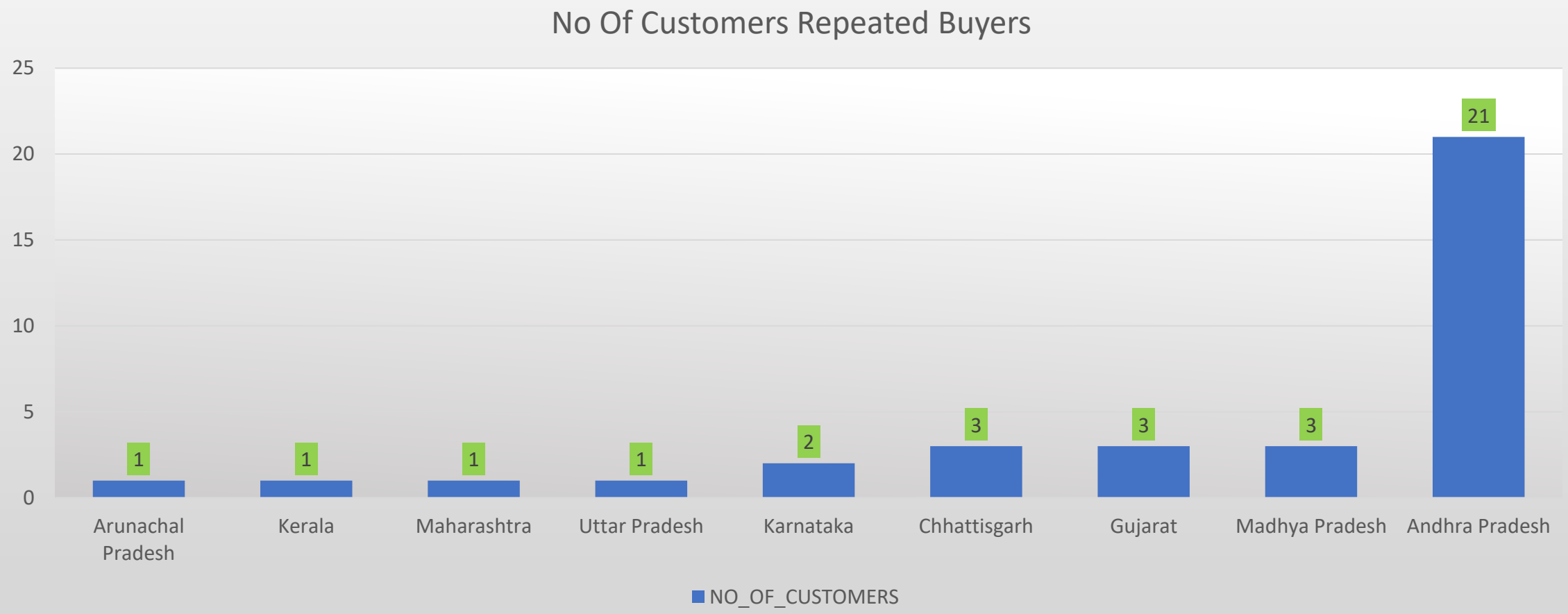
Types Of Customers	No Of Customers	Percentage
Discount Seeker Customers	**57,781**	59.7%
Non Discount Seeker Customers	**39,022**	40.3%
One-time Buyer	**96,767**	99.96%
Repeated Buyer	**36**	0.0372%
Transaction With Voucher	**3,708**	3.83%
Customer With Multiple Channel	**9**	0.0093%

Behaviour Of One Time Buyers



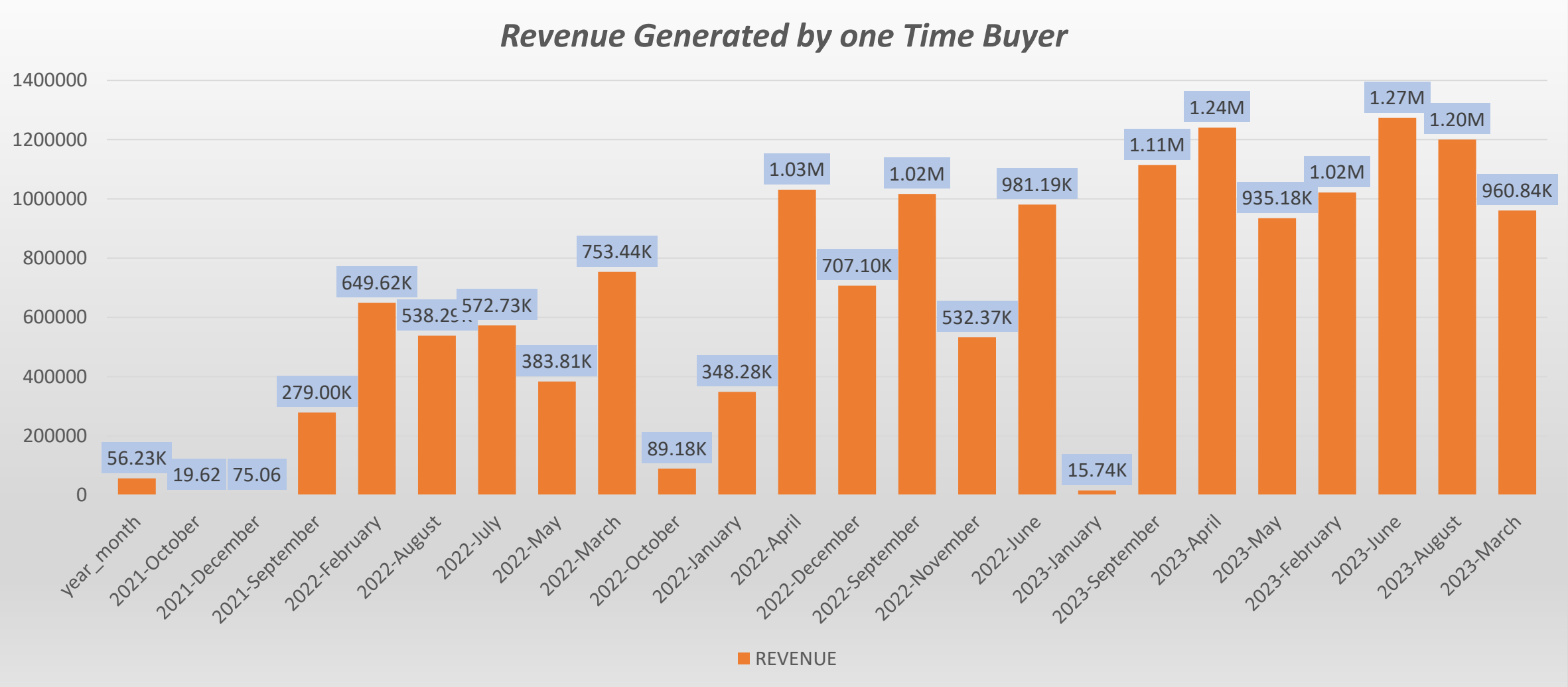
- The one-time Buyer mostly from Andhra Pradesh follow as the total no of customers.
- The least no of customer in one time Buyer Goa(10) which are the total customer of Goa.
- The last 5-6 total customers are one time buyer it may be their no presence of store.

Behaviour Of Repeated Buyers

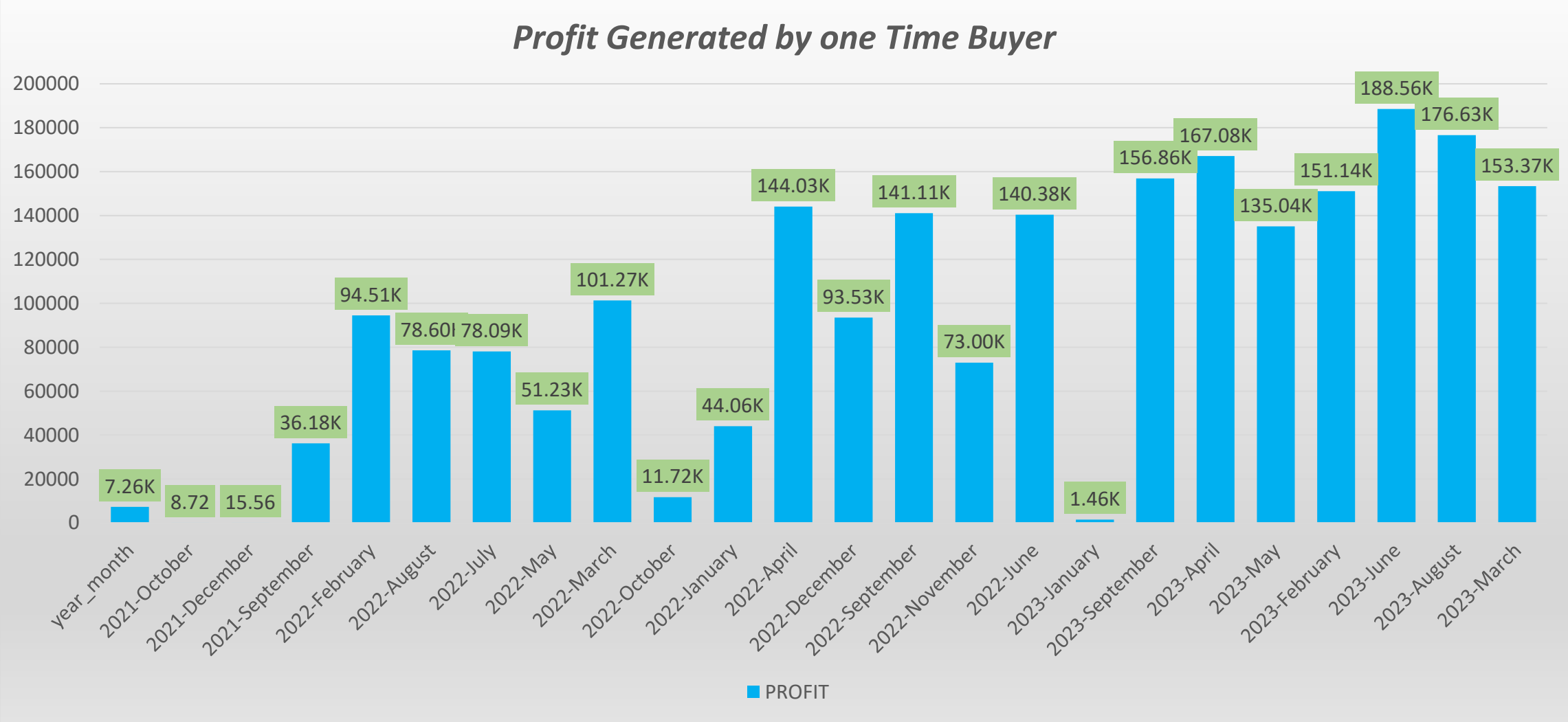


- The repeated Buyer mostly from Andhra Pradesh similar to on-time Buyer.
- The repeated buyer customer are very less max 21 and min 1.
- The last 4 state we see that there are only one customer.

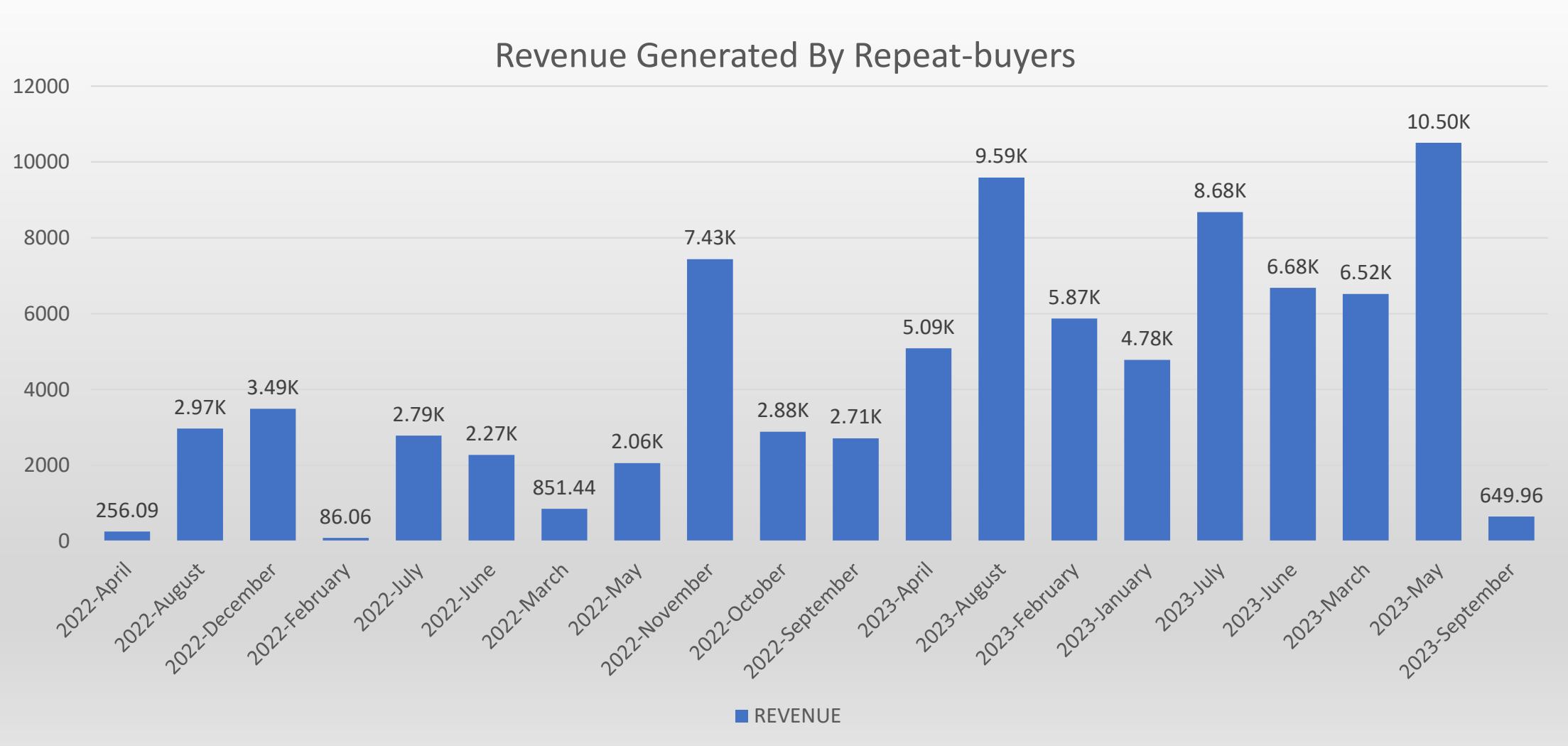
Behaviour Of One Time Buyers



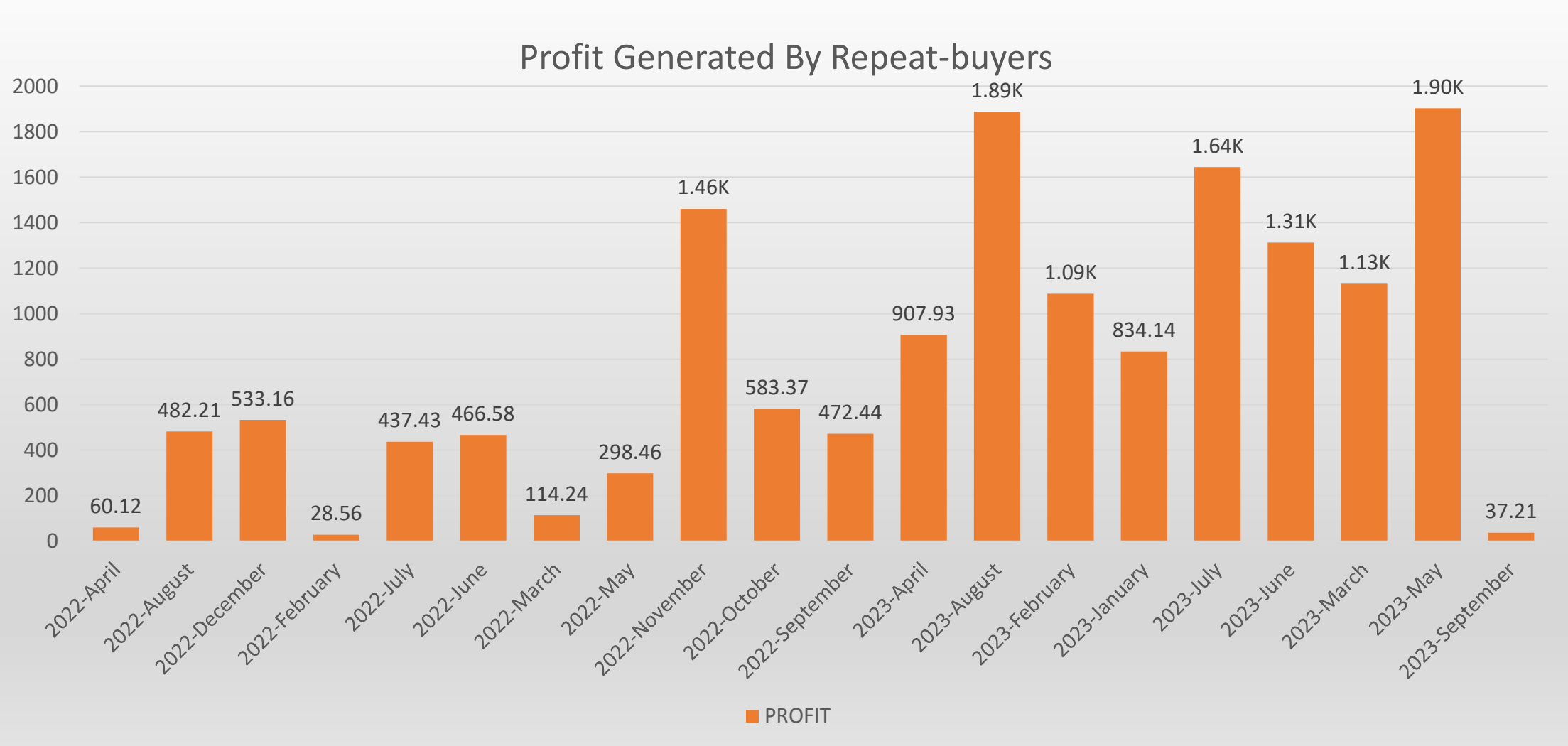
Behaviour Of One Time Buyers



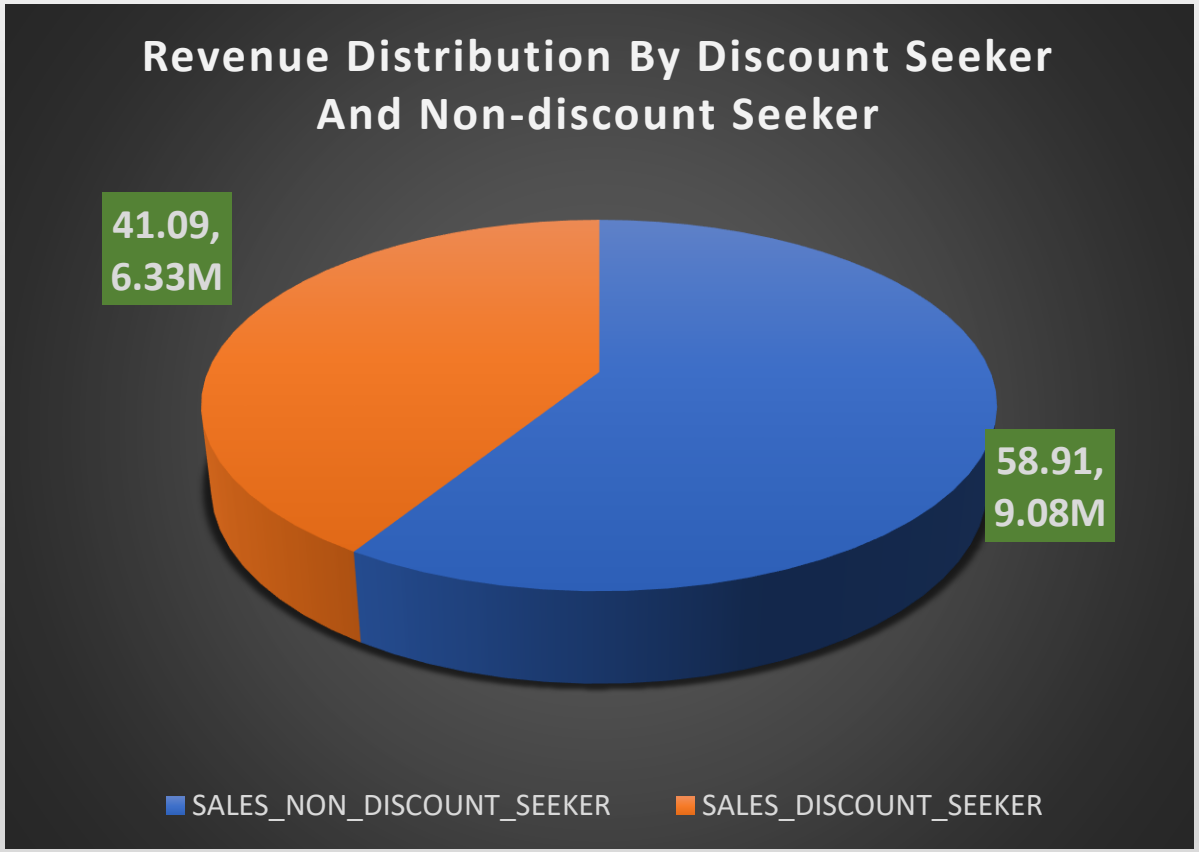
Behaviour Of Repeat Buyers



Behaviour Of Repeat Buyers

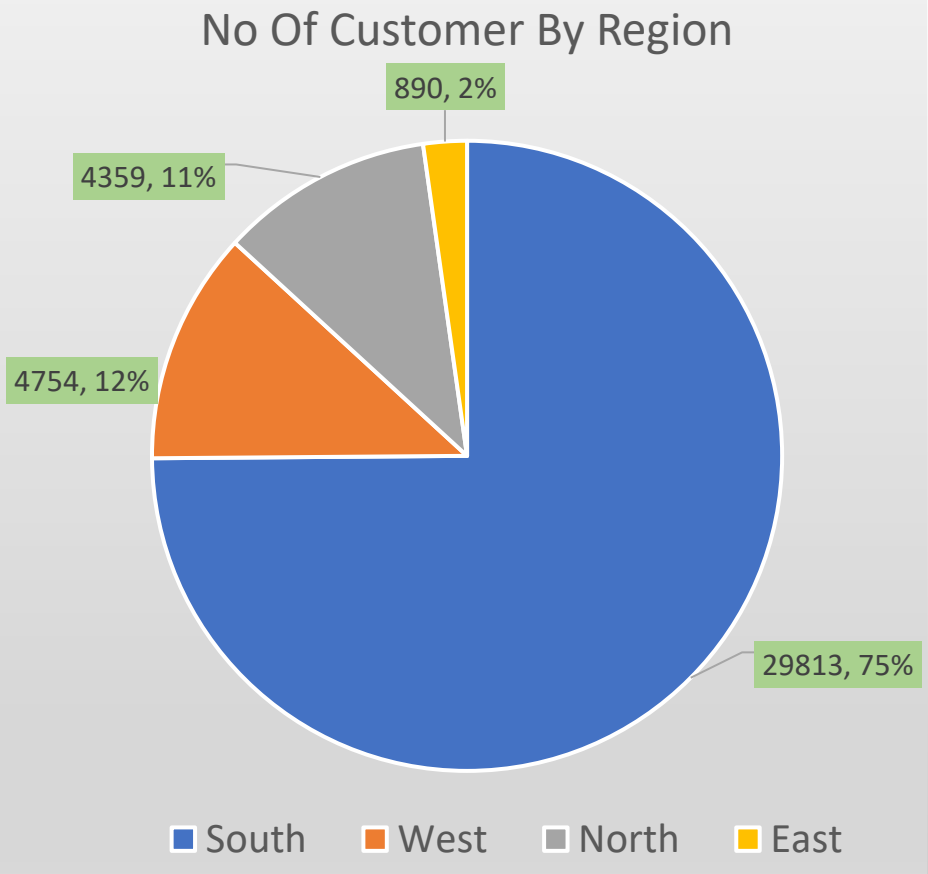


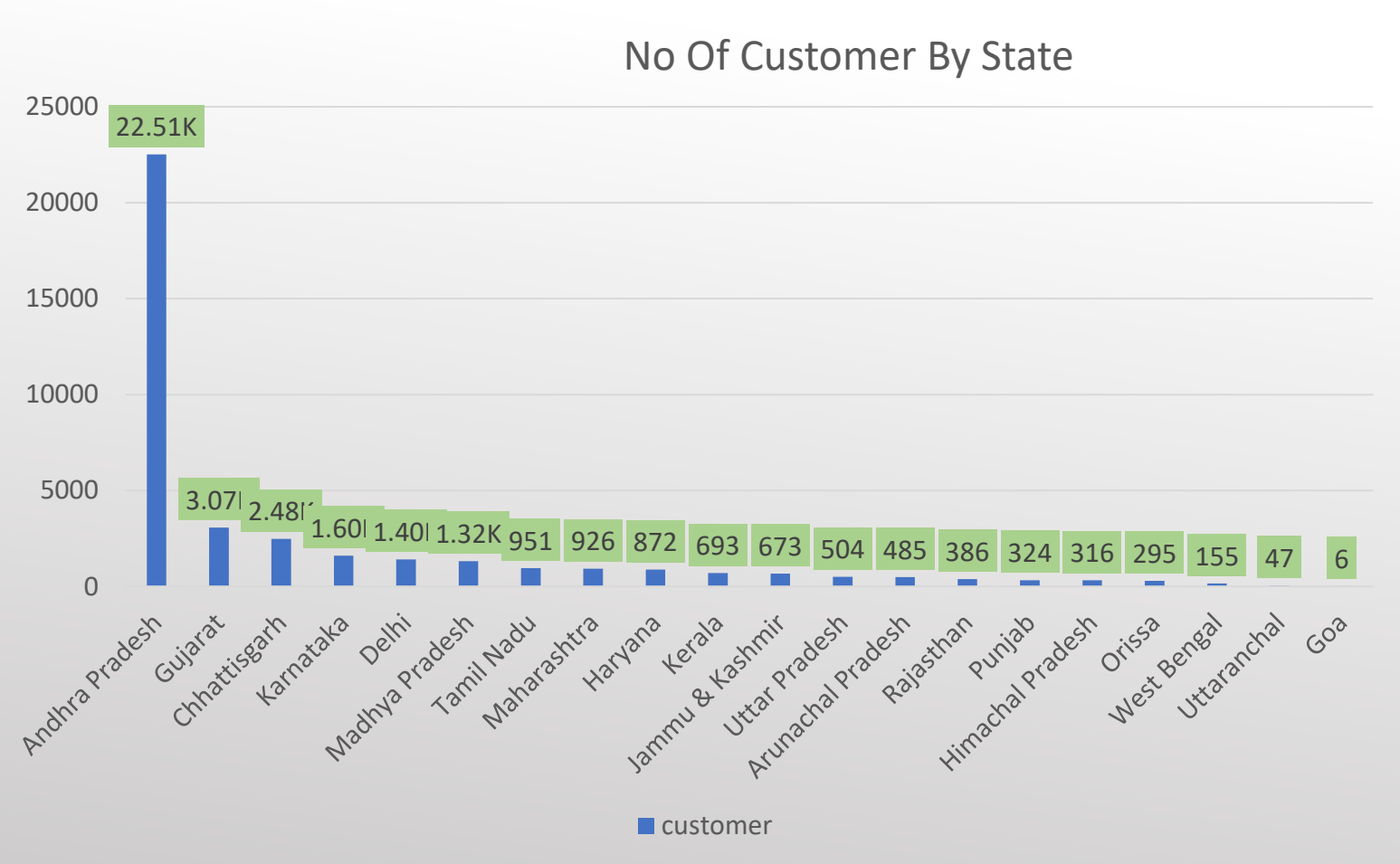
Behaviour Of Discount Seeker Vs Non-discount Seeker



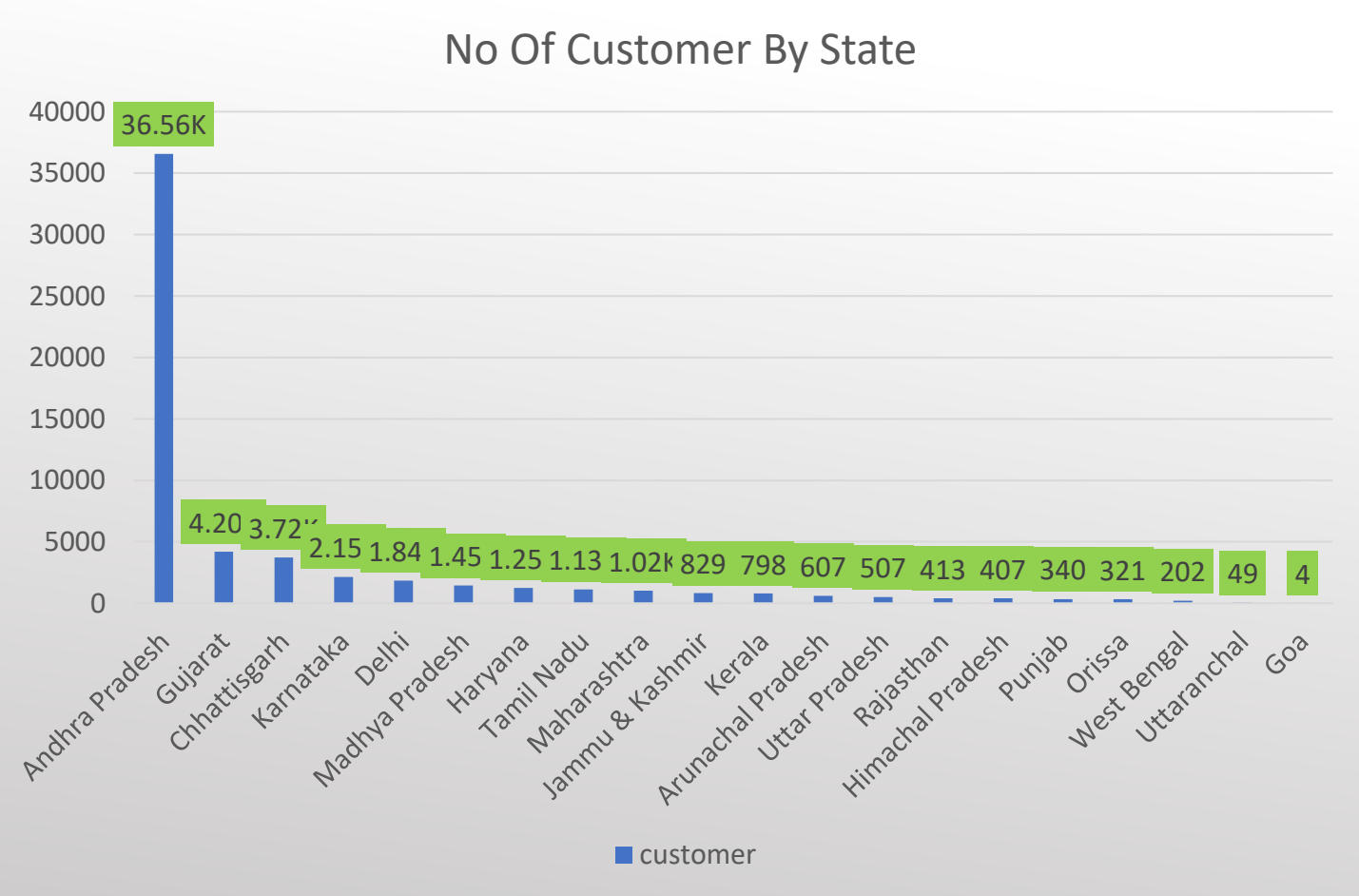
CUSTOMER TYPE	DISCOUNT SEEKER	NON DISCOUNT SEEKER
NO OF CUSTOMER	57,781	39,022
TOTAL QUANTITY	65,096	39,022
TOTAL REVENUE	9.07 M	6.33 M
AVG ORDER VALUE	109.57	232.57
TOTAL PROFIT	1.11 M	1.06 M

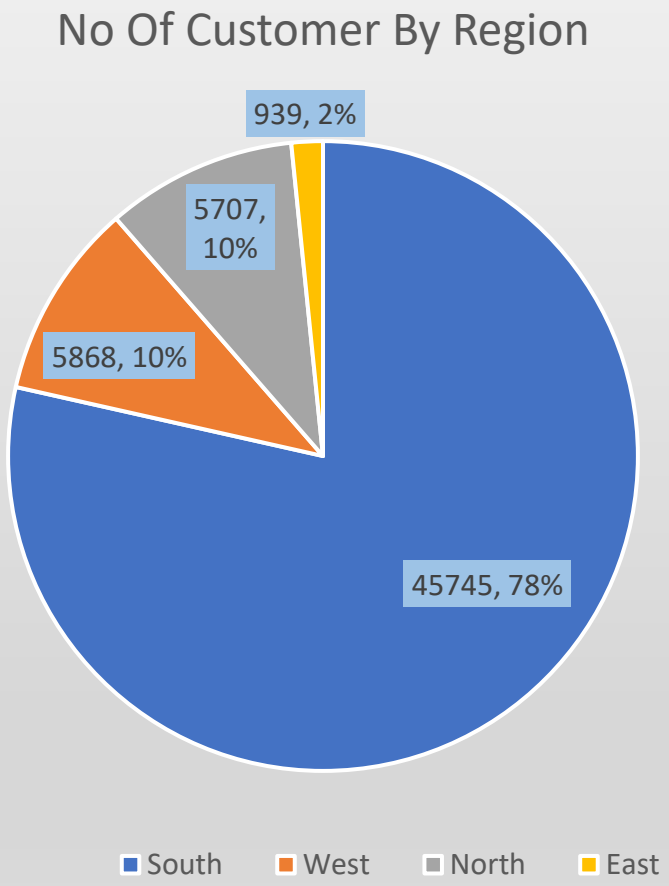
Behaviour Of Discount Seeker





Behaviour Of Non-Discount Seeker





CUSTOMER BEHAVIOUR

Insights :

- ❖ Female customers dominate the user base at 69.84%, with males at only 30.16%.
- ❖ 96.7K are one-time buyers.
- ❖ Only 36 are repeat buyers.
- ❖ Majority of customers fall in the low-revenue segment (revenue < ₹150)
- ❖ Medium (₹300–₹1000) and high (₹1000+) revenue customers are significantly fewer but likely more profitable.
- ❖ Many customers are discount-driven, with noticeable behavioral differences between discount seekers and non-seekers.

Recommendations :

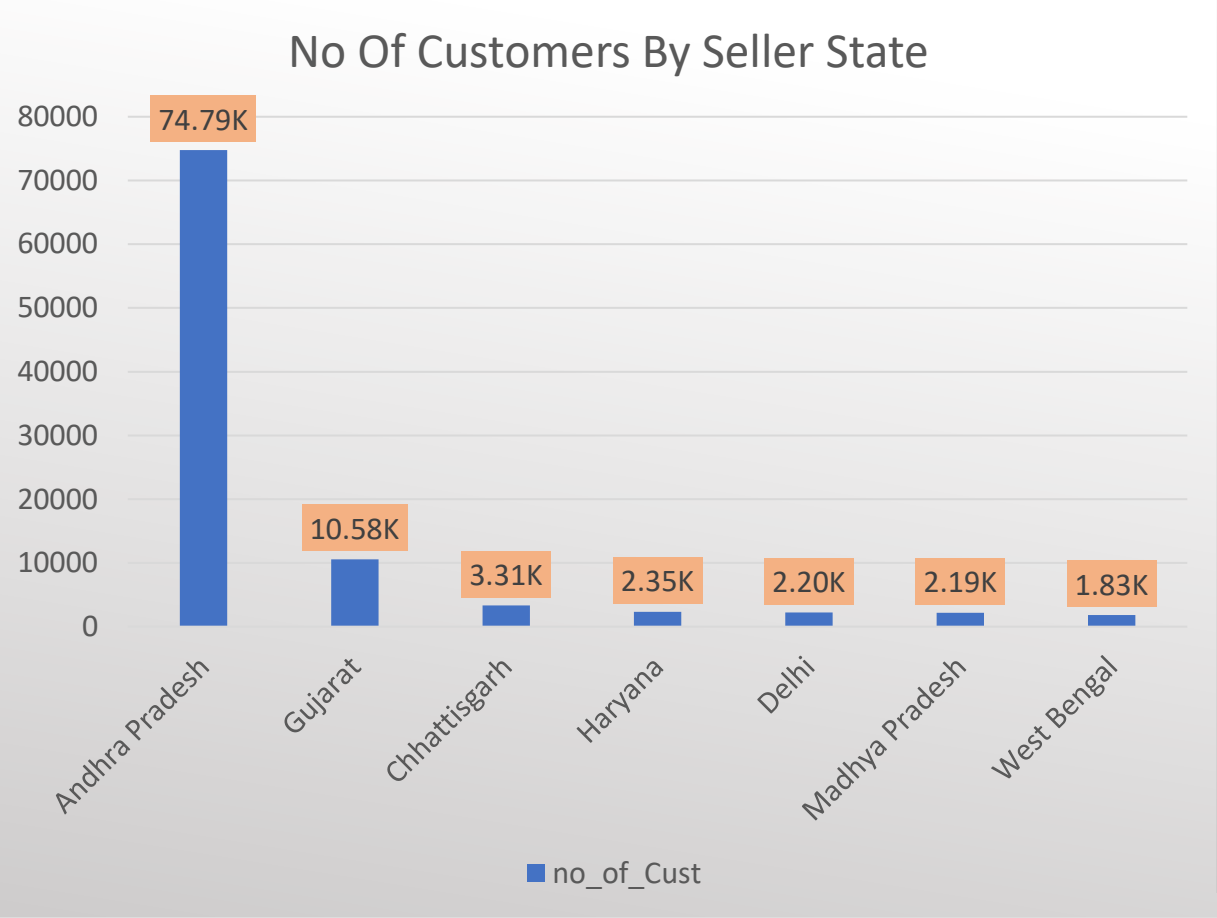
❖Targeted Marketing : Focus marketing efforts on converting low-revenue customers to medium-revenue customers by offering personalized promotions and loyalty programs.

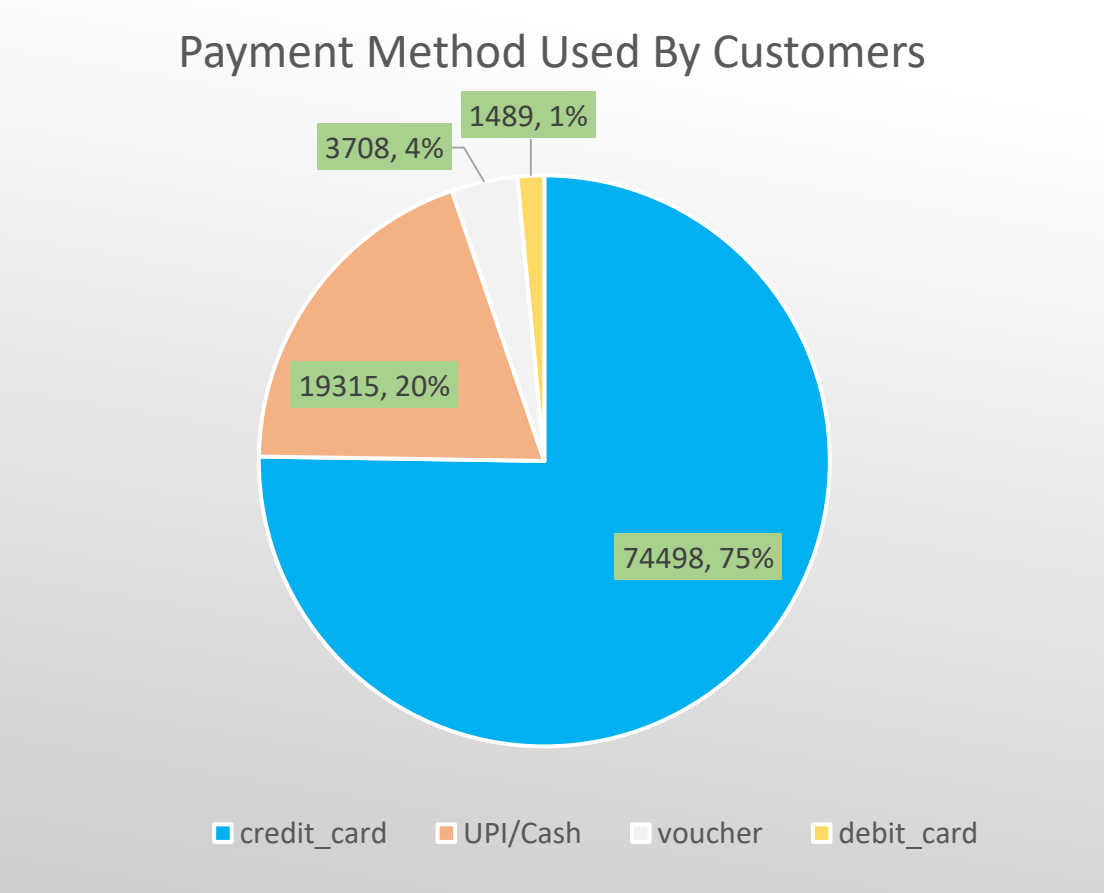
❖Implement strategies to retain medium-revenue customers and encourage them to move to the high-revenue segment.

❖Focus to increase no of customers at every region and state.

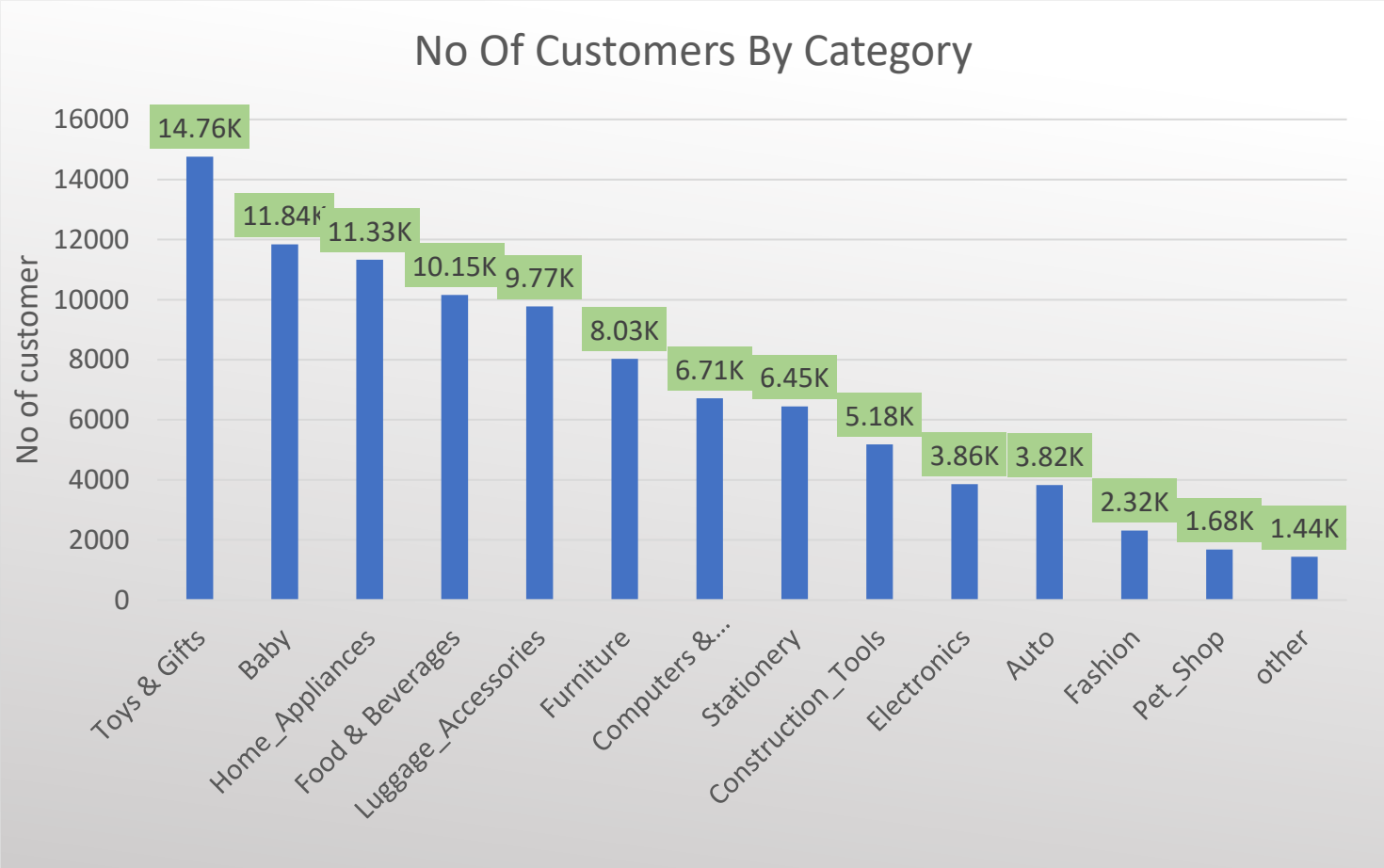
❖Customer Retention: Implement strategies to retain Customers to increase no of orders.

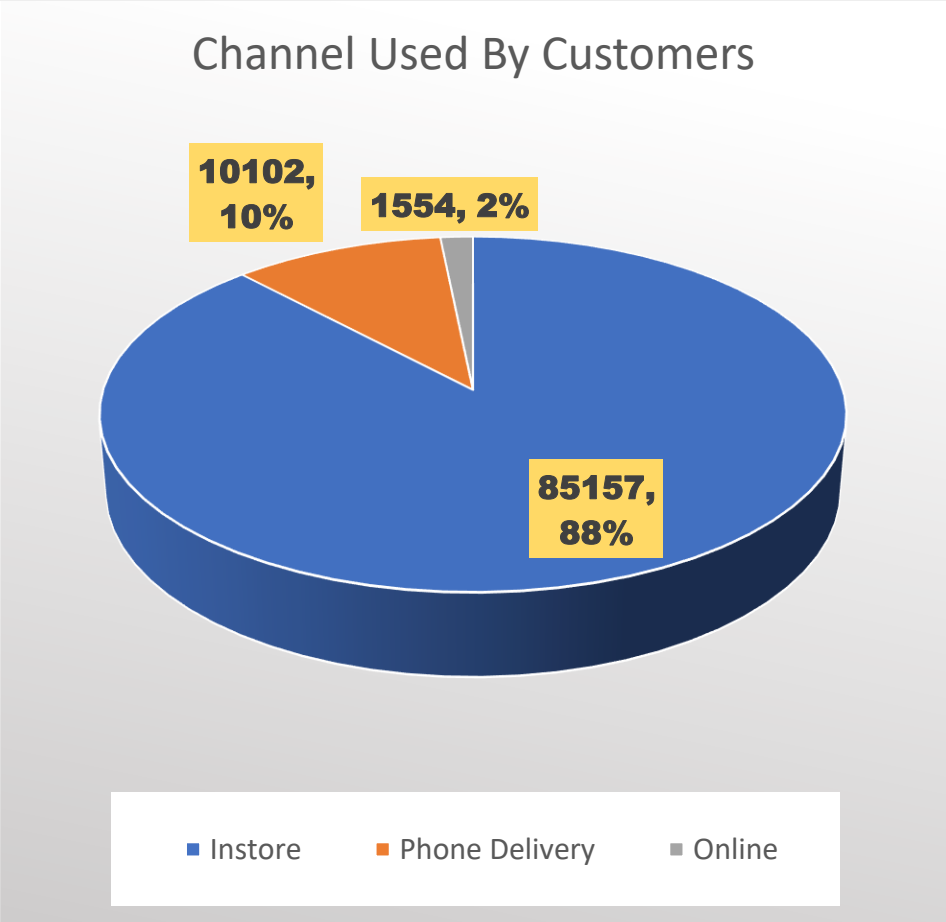
Customer Preferences Analysis



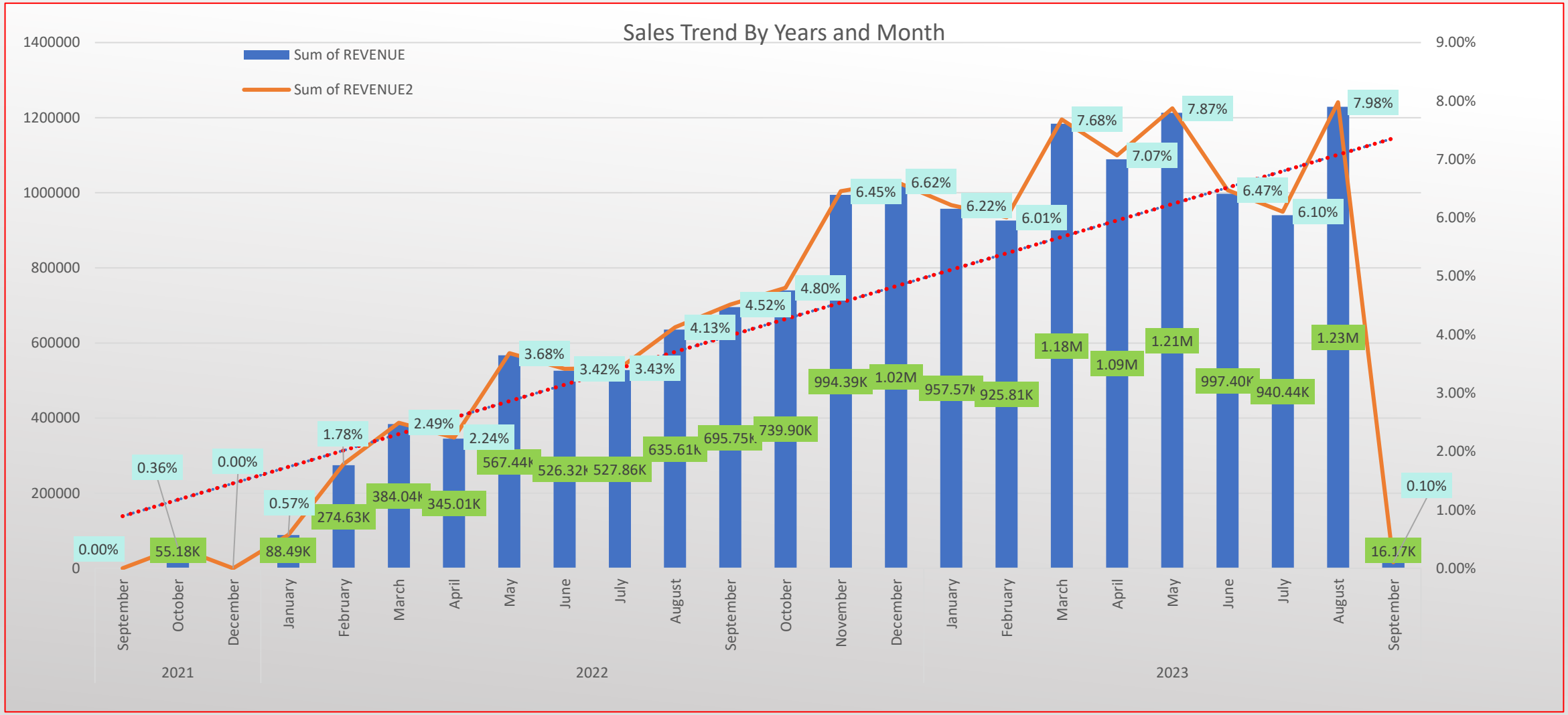


Customer Preferences Analysis

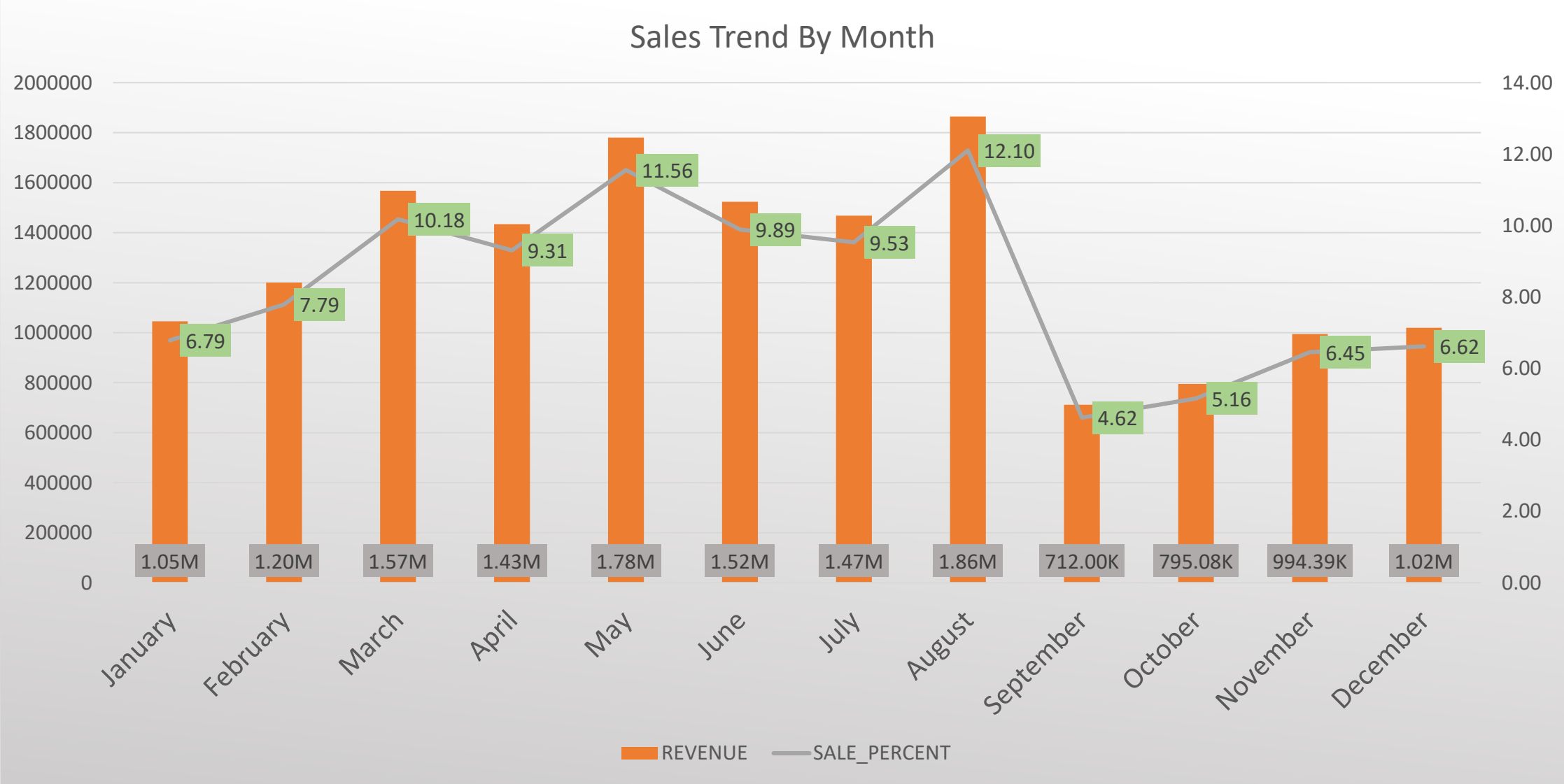




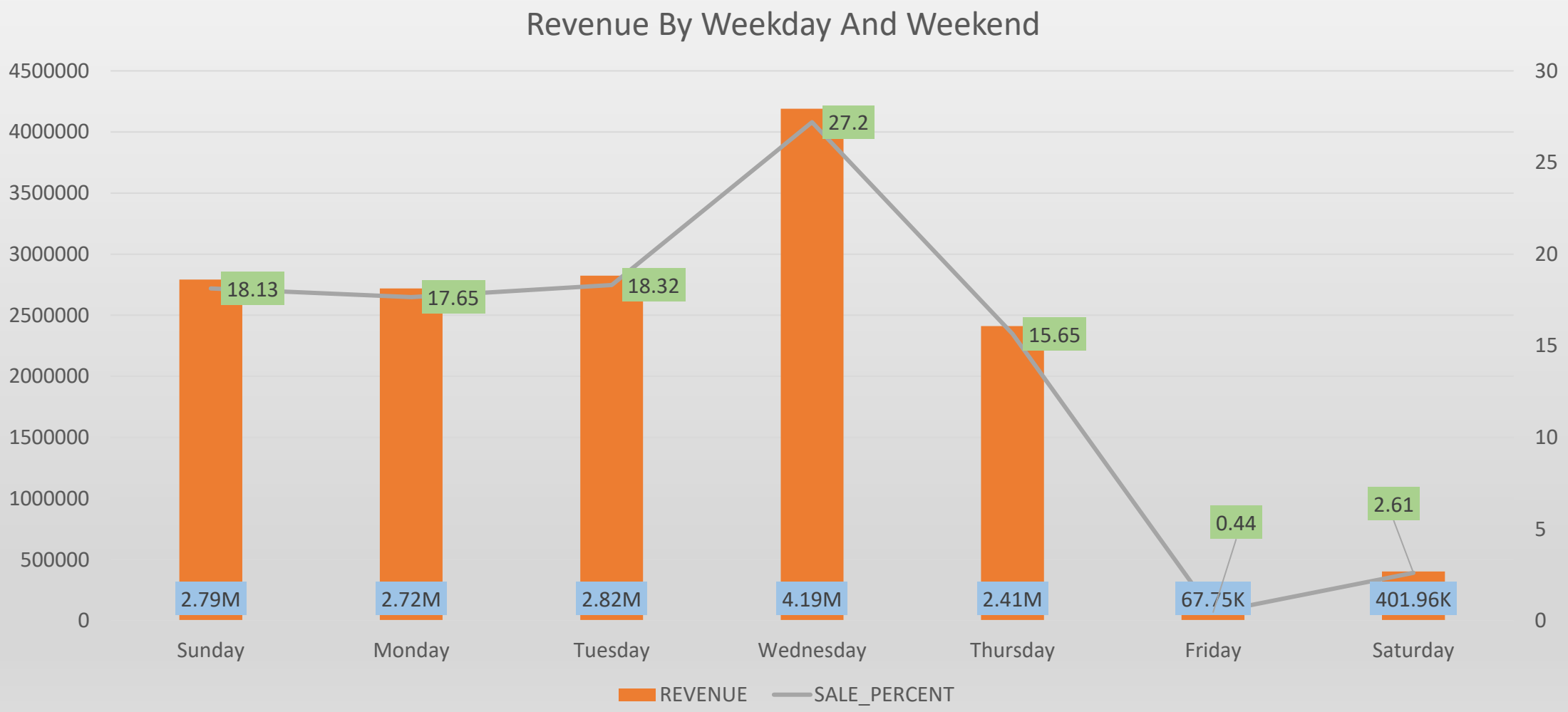
Sales Trends, Patterns & Seasonality



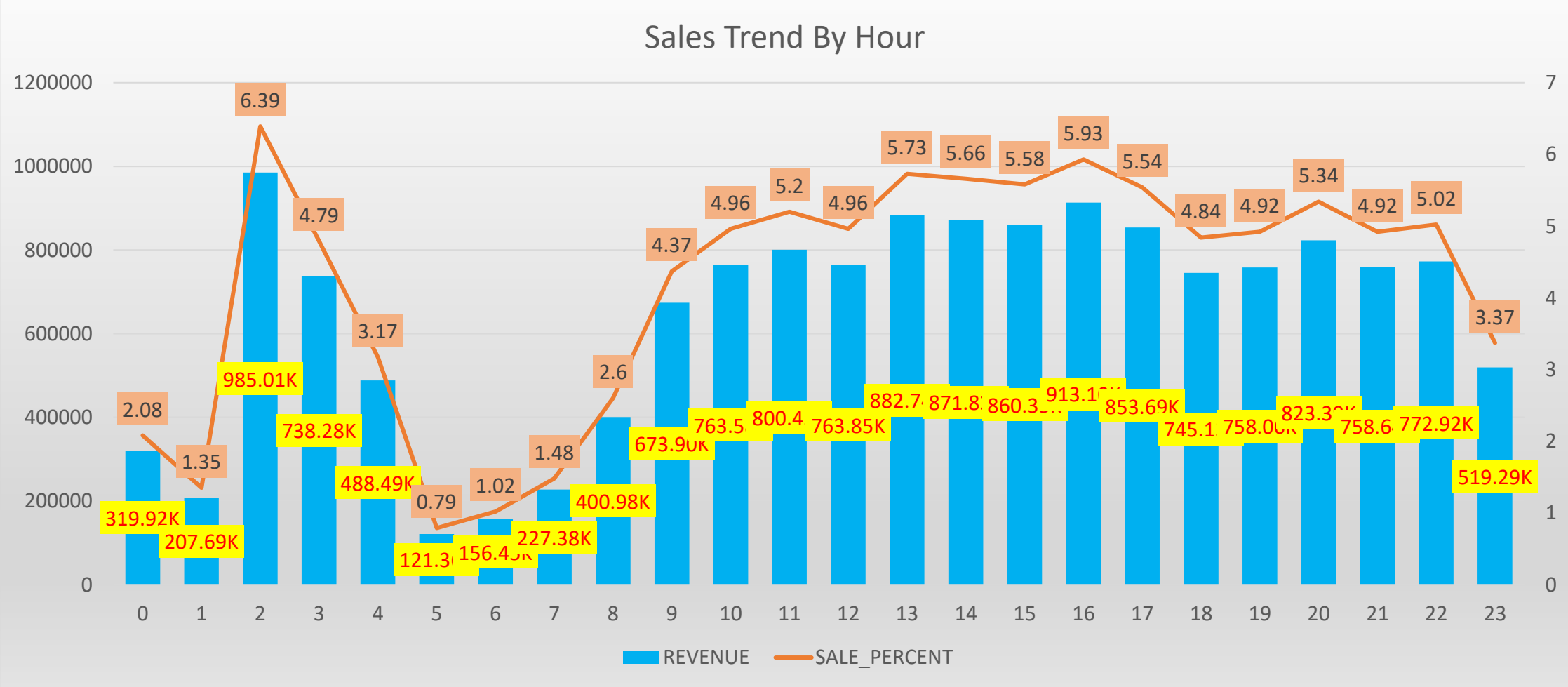
Sales Trends, Patterns & Seasonality



Sales Trends, Patterns & Seasonality



Sales Trends, Patterns & Seasonality



Sales Trends, Patterns & Seasonality

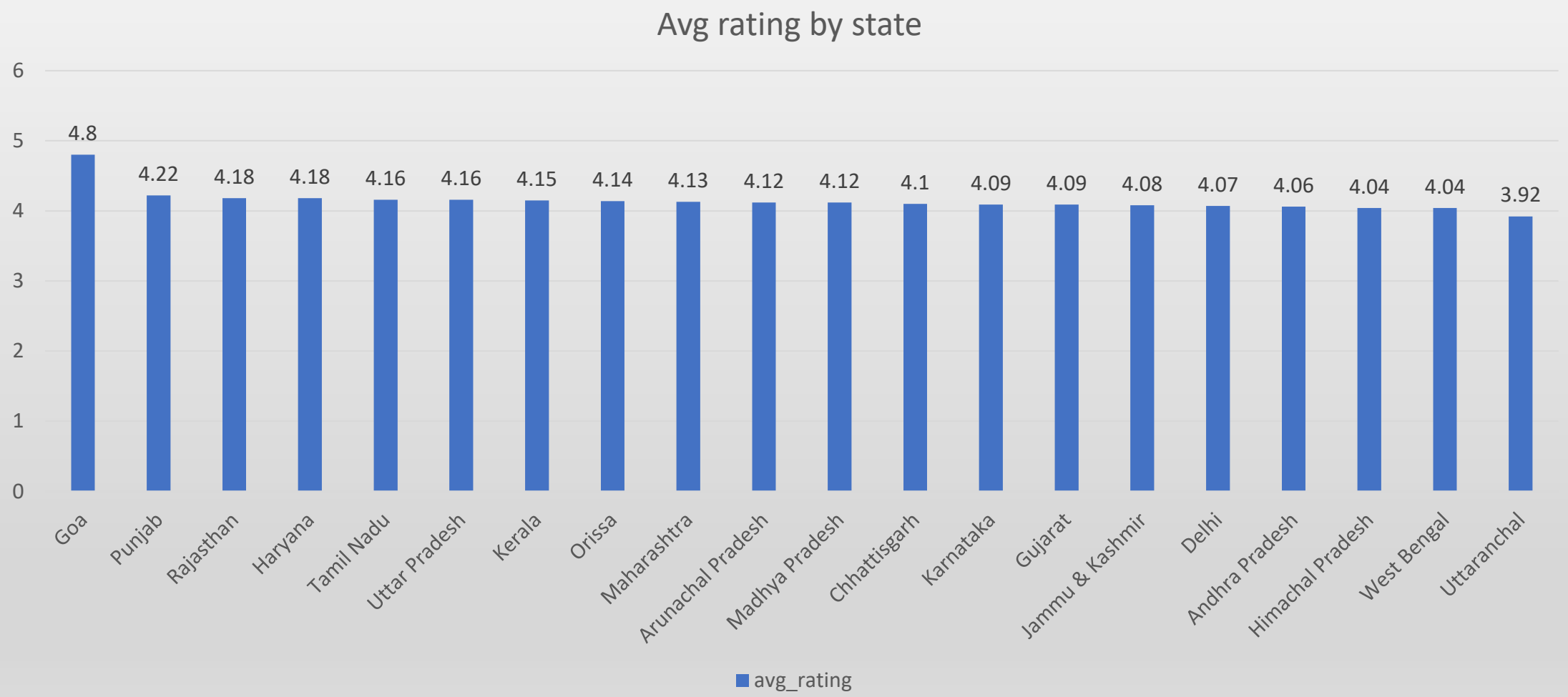
Insights :

- The upward trend suggests that the business is growing or expanding.
- In monthly bases the sales of last four months have decrease it may be the data we have oct2023.
- In weekly bases the sales trend at weekend going downward.
- In hourly bases the sales increase during morning (7-11 AM), afternoon(1-4PM), evening(7-9PM)
- The recurring peaks and troughs suggest that there are factors influencing sales that are cyclical or seasonal. These could be related to factors like holidays, weather, or economic conditions.

Recommendations:

- Explore the impact of factors such as industry-specific events, regulatory changes, or competitive dynamics on sales trends.
- Review operational costs and identify areas for improvement to enhance profitability.
- Adjust weekend sales strategies to counter declining weekend purchases.
- Develop seasonal promotions to boost engagement in low-acquisition months.
- Campaign Timing Capitalize on high acquisition months (May & August).

Average Rating By State



Average Rating By Category

Average Rating

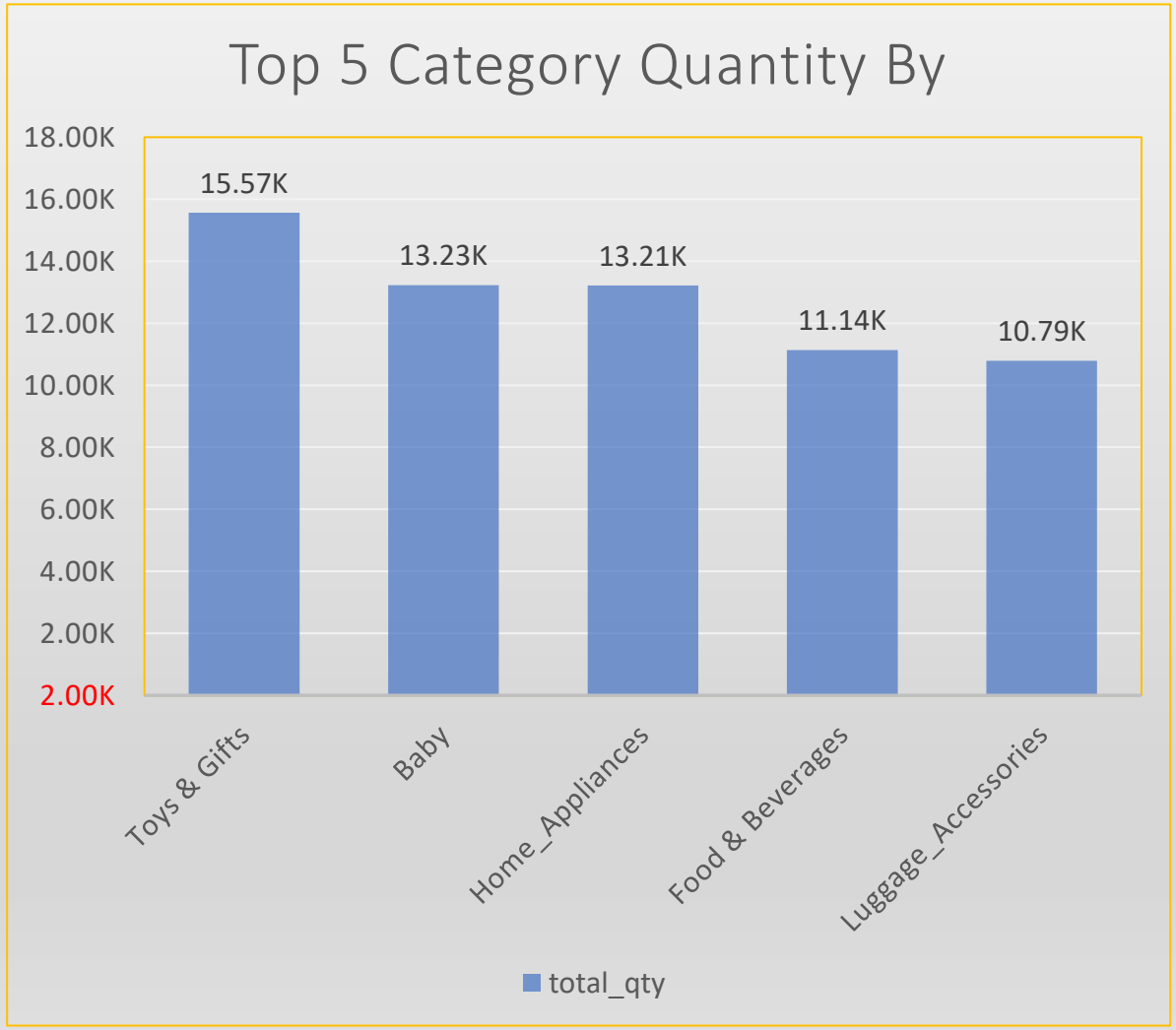
Insights :

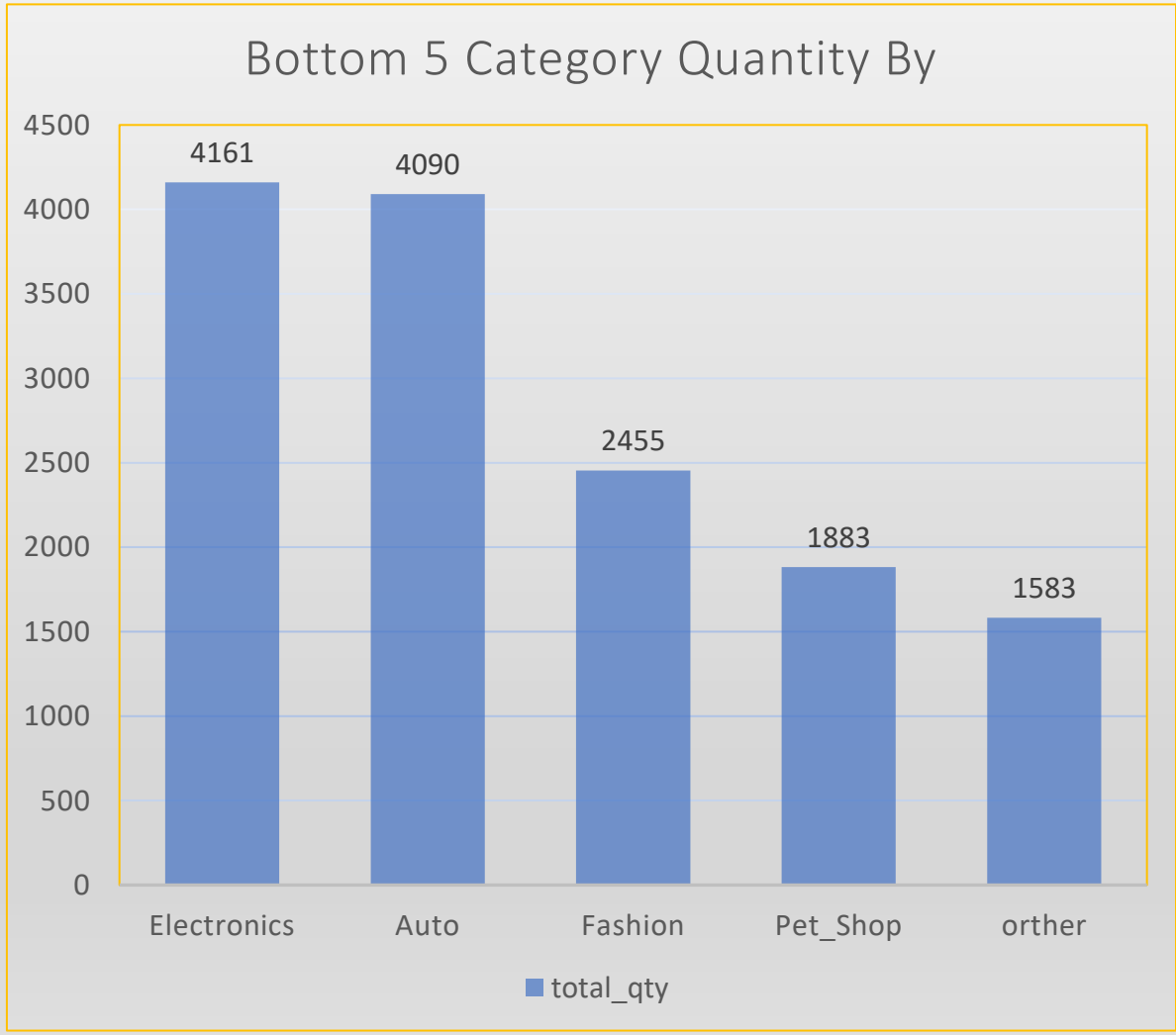
- Pet Shop category has the highest satisfaction score at 4.22, indicating strong customer satisfaction in this category.
- “Other” category and some states like Uttaranchal have the lowest satisfaction (~3.9)
- Most categories have scores above 4.0, indicating that overall customer satisfaction is good.
- States like Goa, Punjab, and Rajasthan have high average ratings (>4.15).

Recommendations :

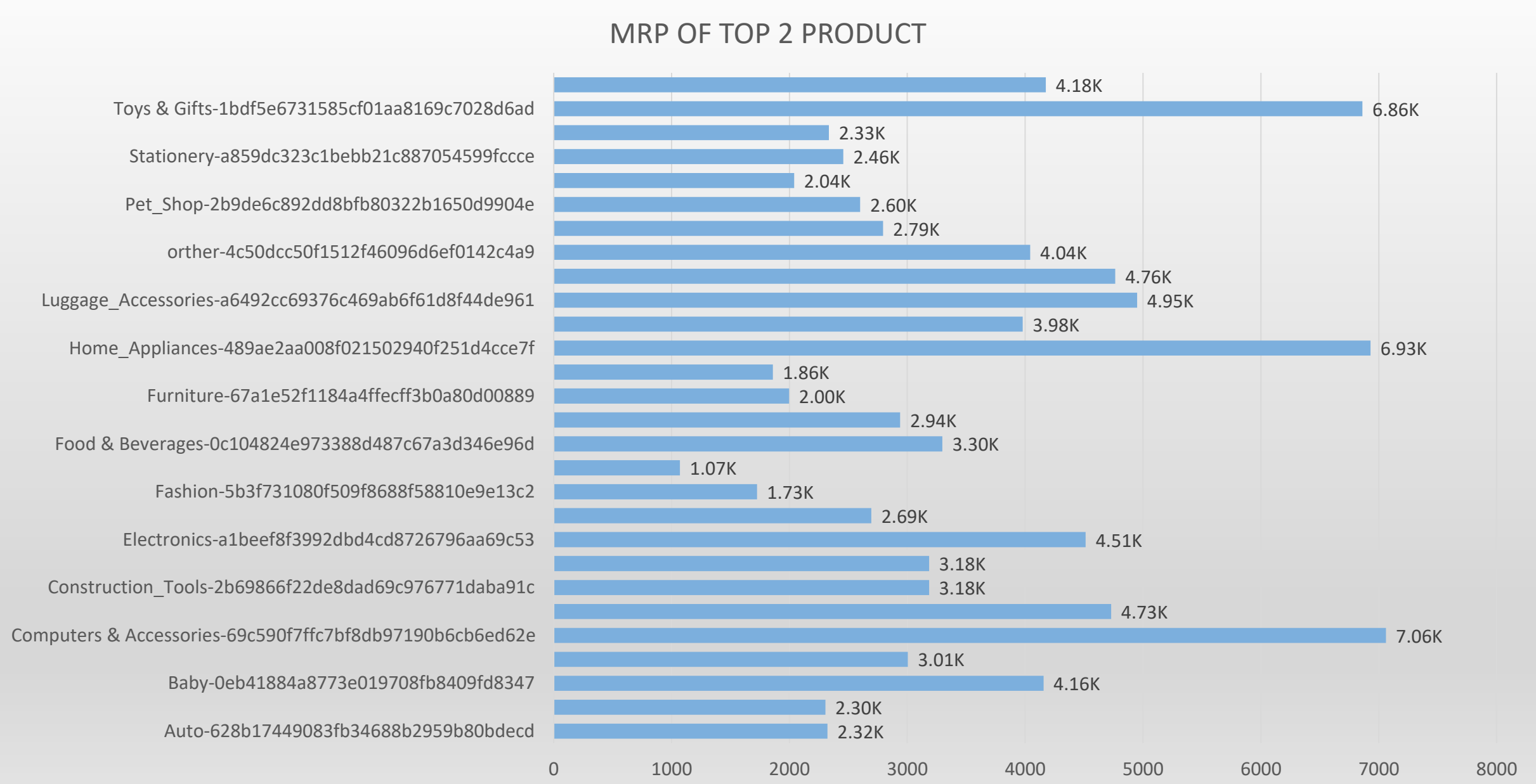
- Consider improving product quality, delivery times, or customer service in the Furniture, Baby, and # NA categories to boost satisfaction.
- Launch customer satisfaction improvement programs across all categories with a focus on personalization, better customer service.
- For low-rated regions like Uttaranchal, run a Voice of Customer (VOC) initiative to capture dissatisfaction reasons and implement improvements.

Top 5 And Bottom 5 selling Category

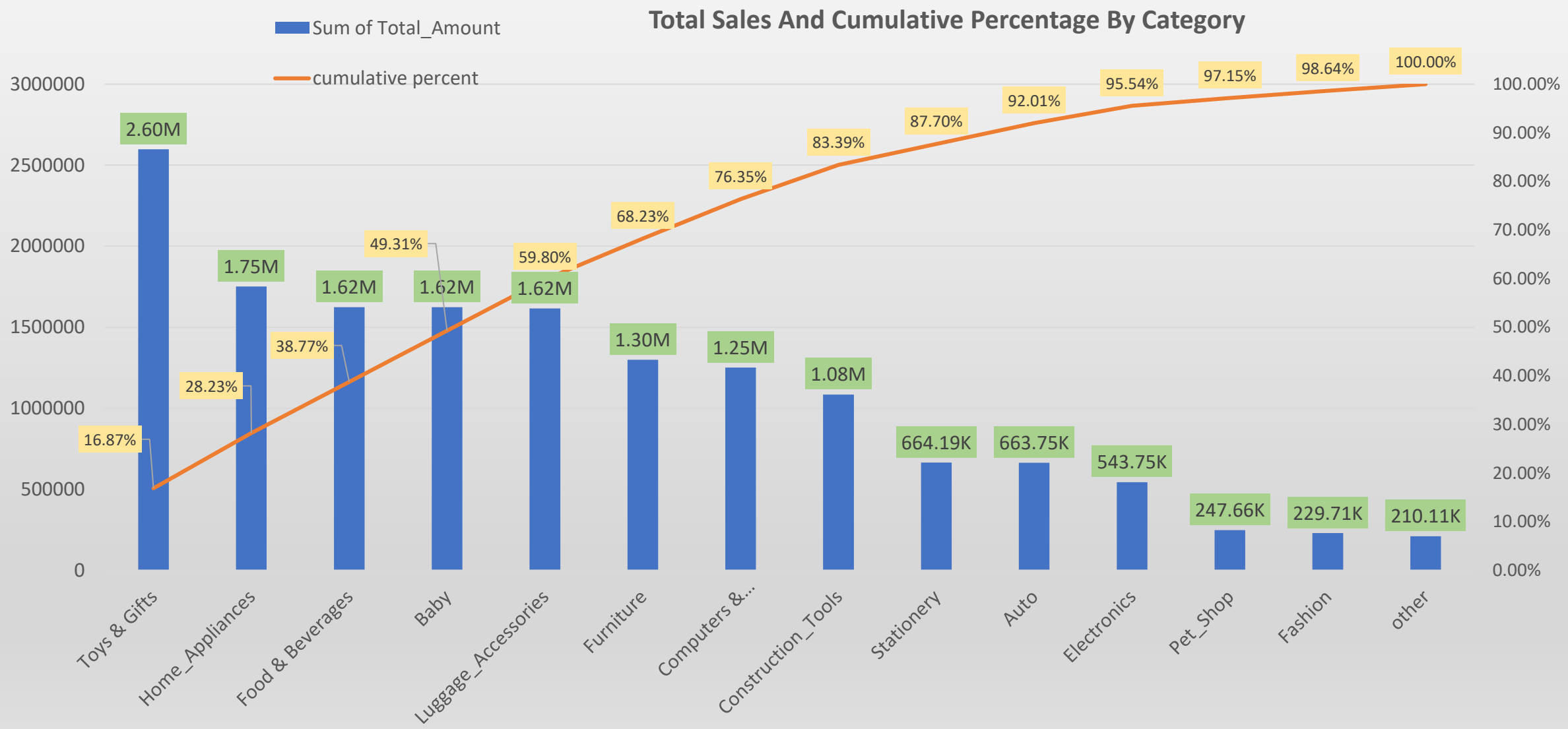




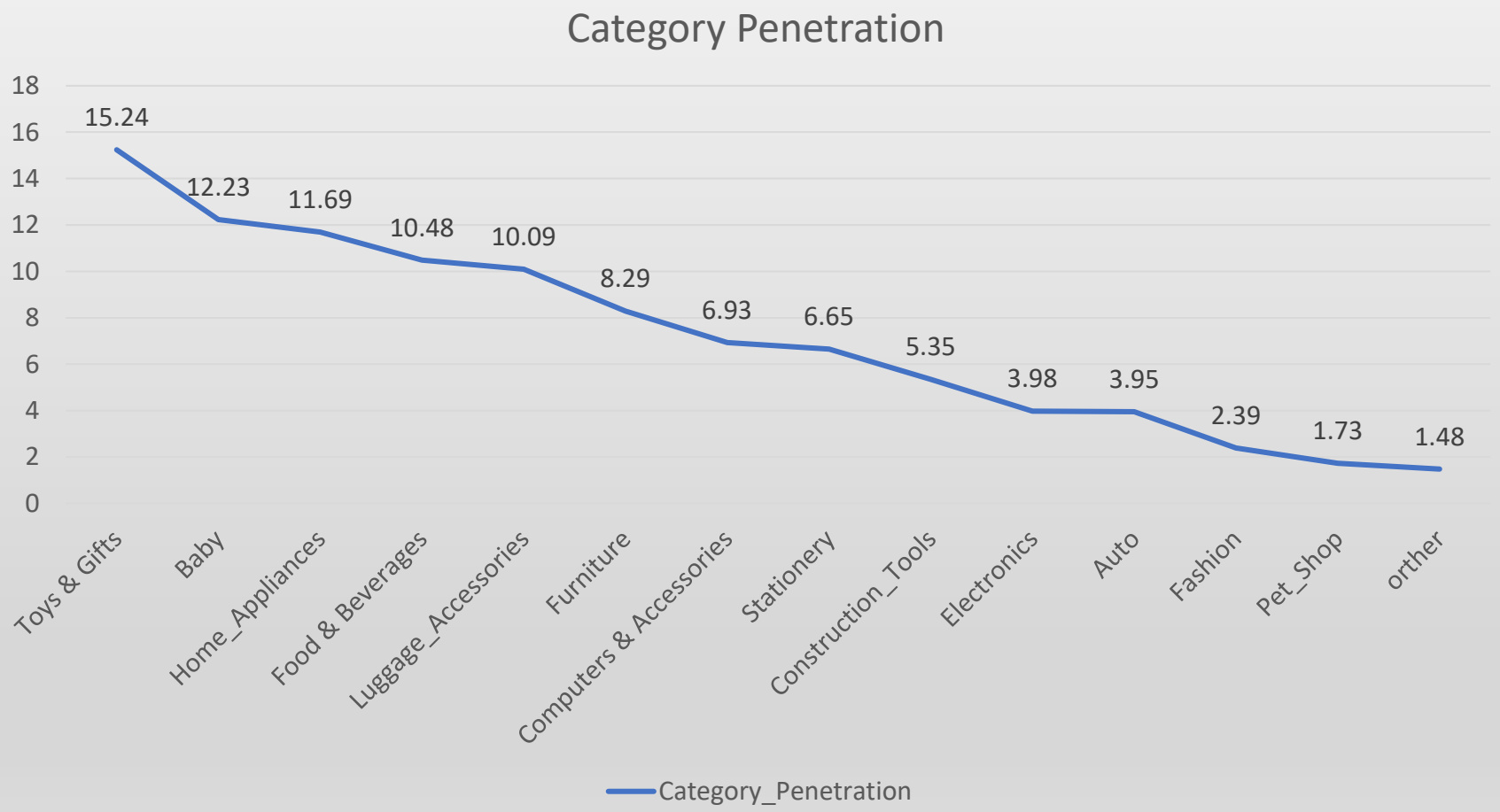
Top 2 Most Expensive Products By Category



Category wise Sales (Pareto Analysis)



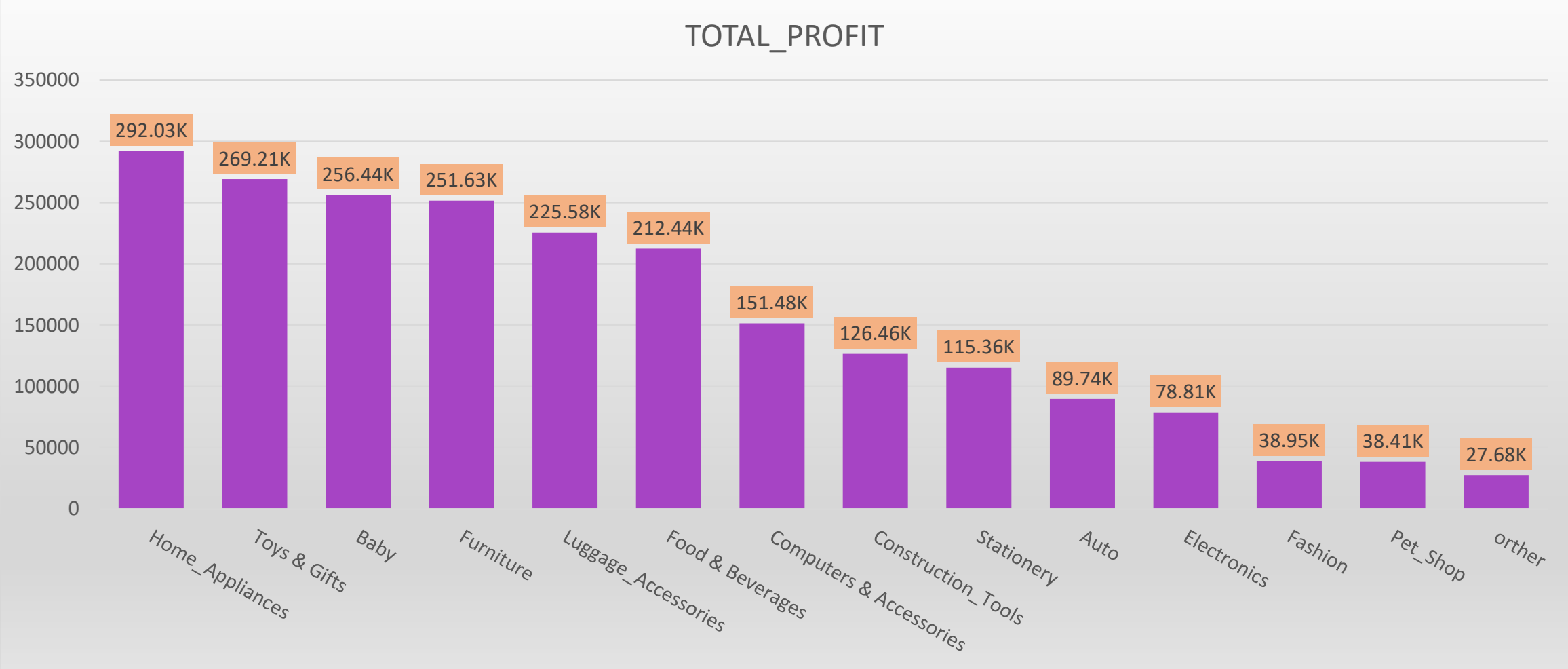
Category Penetration Analysis



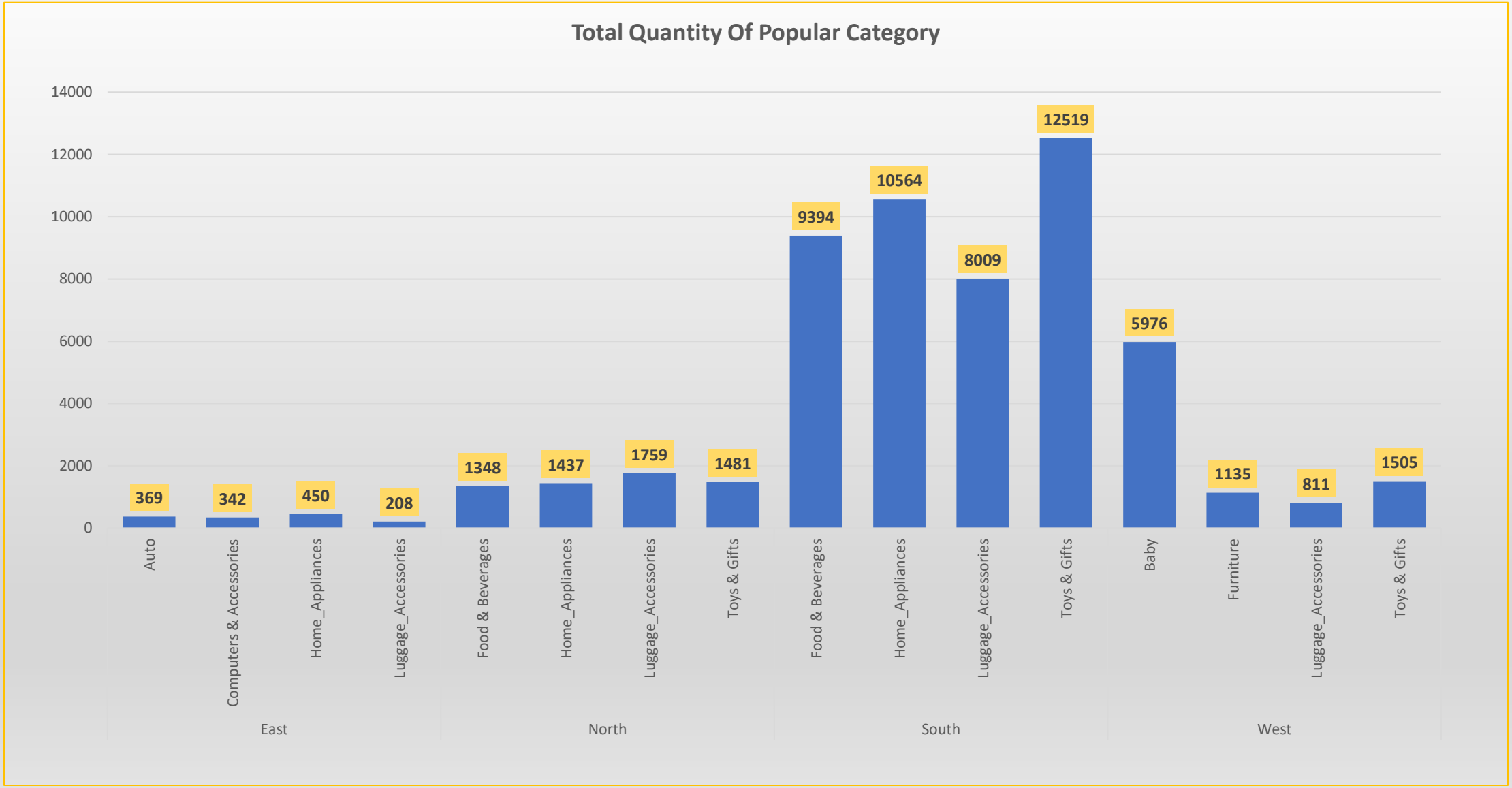
- Toys & gifts are leading with the 15.24 %
- Categories that typically involve frequent purchases or are gift-related and Home while other have lower Penetration
- Focus on low-performing categories like Fashion and Pet Shop

Total Discount By Product Category

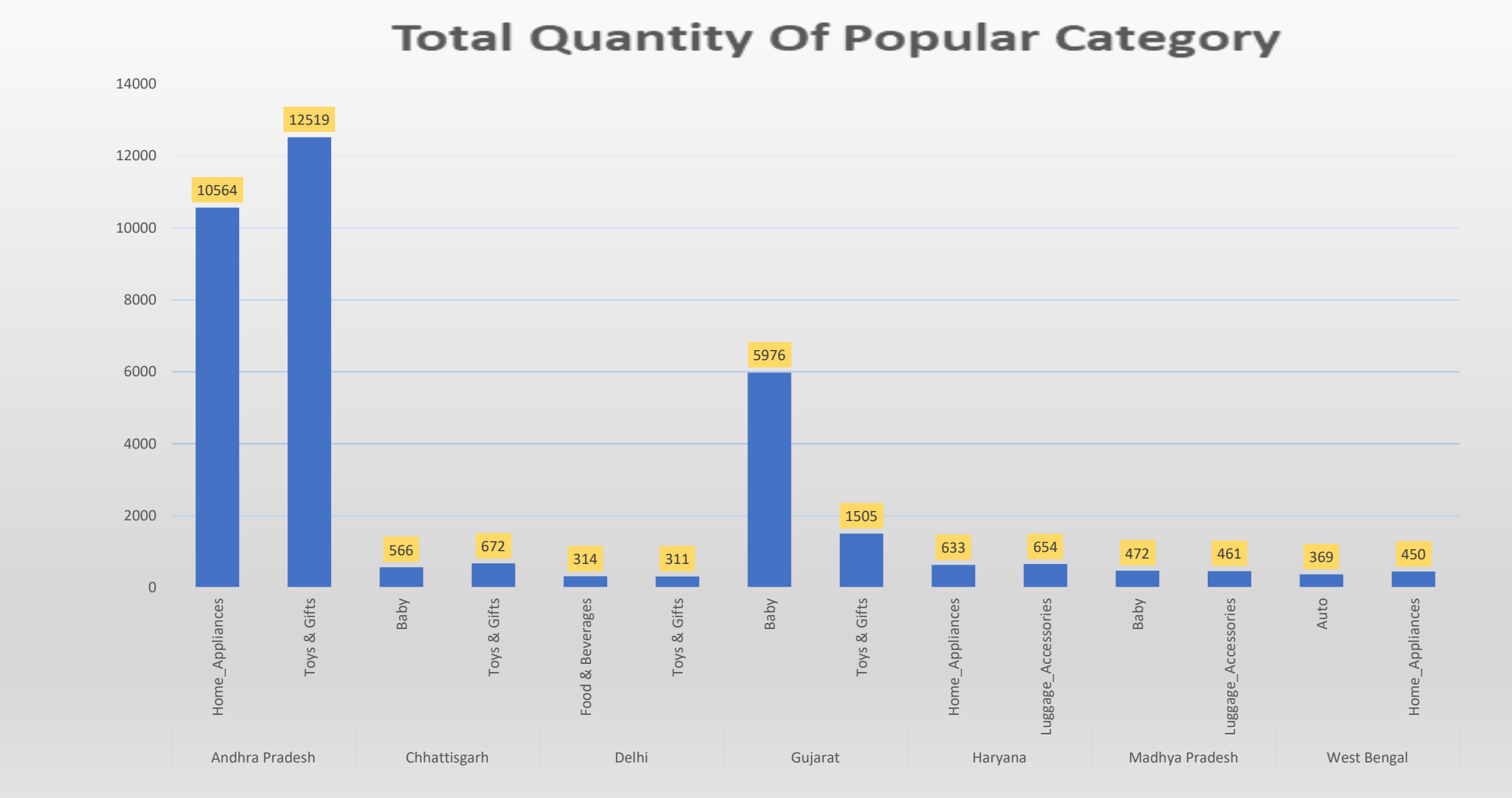
Total Profit By Category



Popular Category By Region



Popular Category By State



SALES AND TRENDS ANALYSIS

Insights :

- A small number of categories contribute the majority of revenue (as shown in the Pareto analysis).
- These top-performing categories are likely driving the business—possibly Fashion, Grocery.
- Some categories(Home applications, Baby) consistently show very high discount levels, impacting margins.
- The bottom 5 categories generate minimal sales.
- Toys & Luggage Accessories have demand at every Region.

Recommendations:

- Focus on High-Performing Categories
- Identify categories with unsustainable discounting and apply pricing optimization.
- Some Low product category have seasonal demand evaluate these.
- Focus on Product which demand at every Region.