Shopamore

Shopamore Buyer purchase behaviour analysis

Executive Summary

Project Description:

Use customer-segment analytics to increase repeat-purchase rate and average order value through personalized promotions, targeting a 10 % lift in repeat purchases within six months.

Objectives:

10% increase in repeat purchases over the next 6 months

Tools:

- *Data Cleaning and analysis : SQL Bigquery, Python, spreadsheet*
- *Visualization : Tableau, Spreadsheet*

Metrics:

- *Buyer Rate*
- *Basket Size & Value*
- *Retention Rate*

Analysis Result:

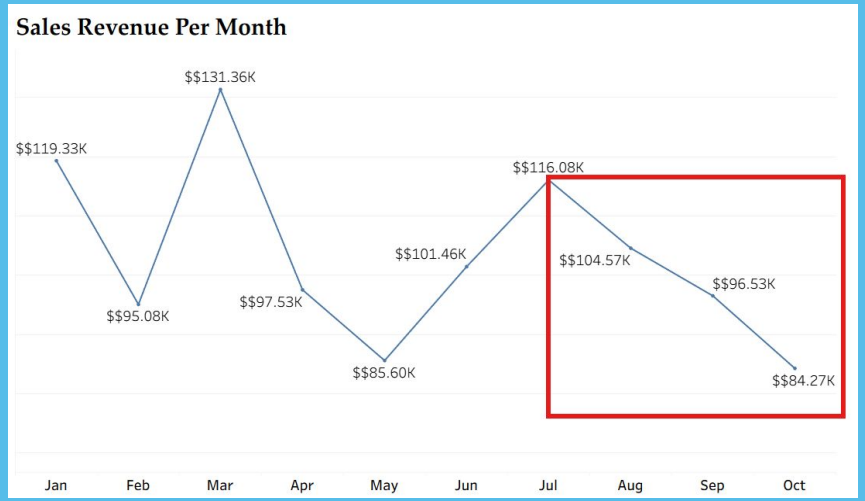
Focus on how to make first time buyer and single-time buyer to become repeat buyer and growing members than growing basket size per order, since average orders already multi-items.

Recommendation:

- *Focus on Loyalty Program*
- *follow-up first buyer*
- *Win Back inactive buyer*
- *Long Term : Bundling project*

Background of Project

- Company Overview : Shopamore is an Australian-based online clothing store offering a diverse range of apparel for various tastes and preferences.
- Data Information : This shopping cart database includes transaction-level data capturing which products were purchased together, primarily focusing on buyer purchase behavior over time.Dataset covers only Jan-Oct 2021 and excludes marketing-spend data.
- Data Source : This Analysis is based on public kaggle dataset : shopping-cart-database



"Over the last quarter, monthly revenue has shown a sustained decline, with the most recent month recording the lowest sales $84.27K, figure in the observed time frame."

Methodology

01
Problem Understanding
Understand the buyer churn from the data sales provided

02
Data Preparation & Data Cleaning
Changing the right data type, checking double data and null values by Python.

03
Exploratory Data Analysis
Conduct Initial Investigation on dataset to define metrics and test for hypothesis

04
Data Analysis
Conduct the hypothesis testing with the metrics that decided on EDA with python and excel.

05
Data Visualization
Visualize the analysis into chart, table, and dashboard by using Tableau

06
Insight & Recommendation
Use the result of data analysis to determined the insight and business recommendation to solve the problem

Business Problem

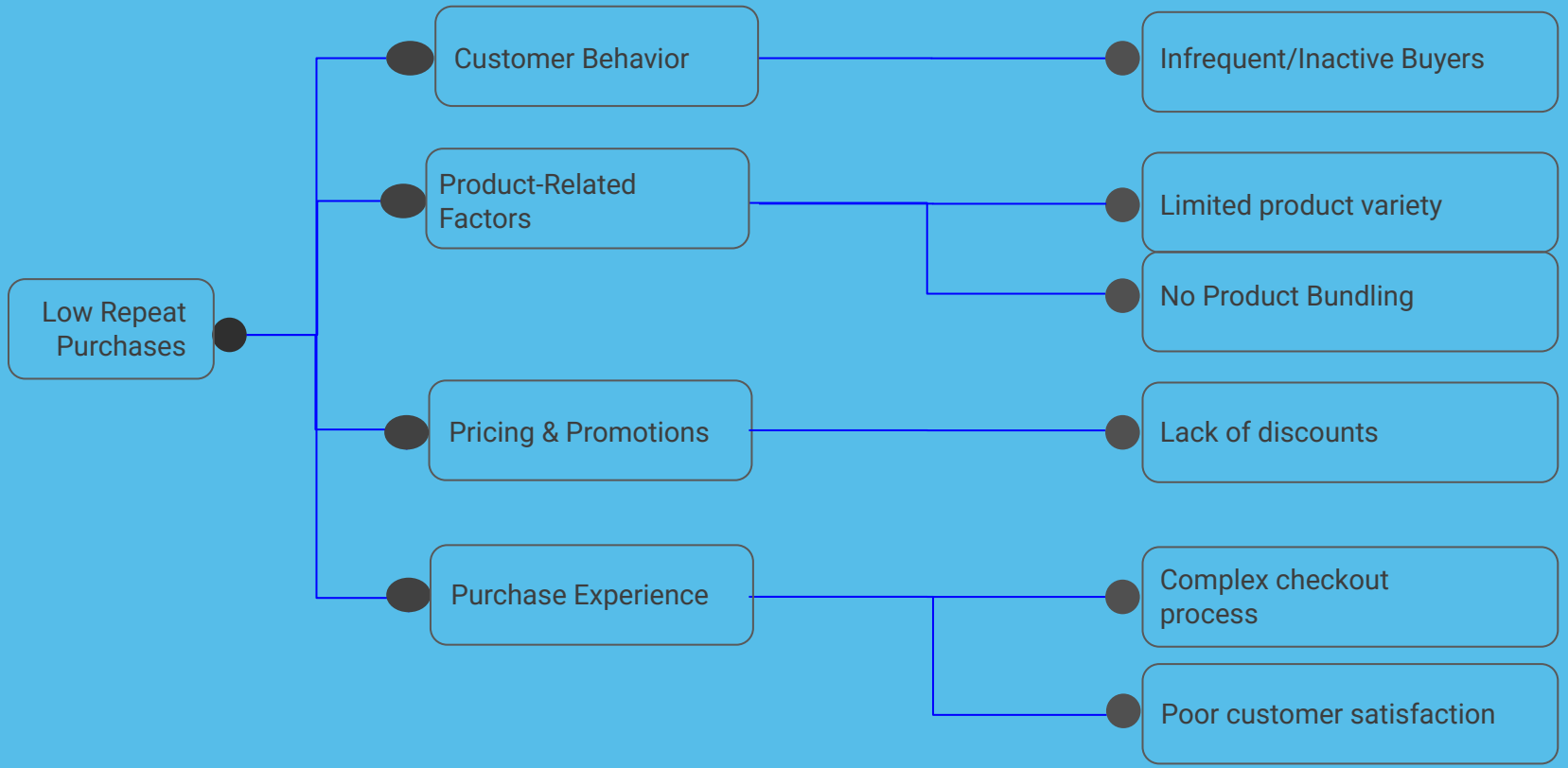
Problem Statement

The sales in the last quarter drop 27% from $116.08K at July to $84.27K at October. For better revenue, we aim to enhance customer purchase frequency and average order value by identifying key product bundles and customer segments through data-driven insights, targeting a 10% increase in repeat purchases over the next 6 months

Objective of analysis

Conduct a data-driven analysis of buyer behavior to identify the factors driving the recent revenue decline and to recommend the most effective strategies for revenue recovery.

Business Problem
Root Cause Analysis & Issue Tree



Business Problem

Hypothesis and Metrics

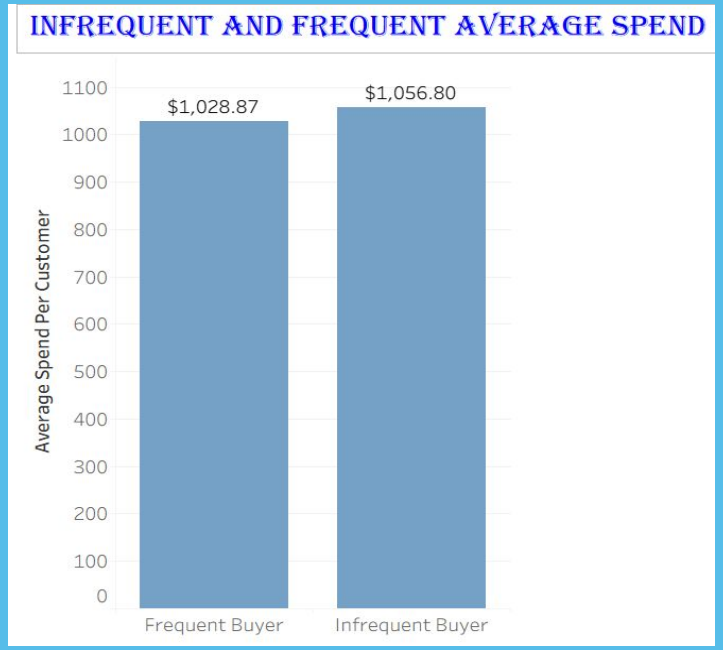
Issue Branch	Hypothesis	Metrics
Infrequent shoppers	Demographic differences, Risk aversion and sensitivity to online shopping risks, Different priorities in purchasing decisions.	- Buying Rate: Measure the average amount spent by infrequent shoppers per purchase.
Poor customer satisfaction	Customer Experience Issues (slow website, poor UI), Pricing & Promotions (no discounts)	Purchase Frequency: Measure the number of times a customer makes a purchase within a specific timeframe. Since the data is only 10 months, so monthly.
Limited product variety	Customers are less engaged and transact less because the shopamore offers limited product variety, reducing its ability to meet diverse customer needs.	Total Spend → total spend per transaction. Category Diversity per Basket → Number of different categories included in one cart. Base on Infrequent and frequent buyer.

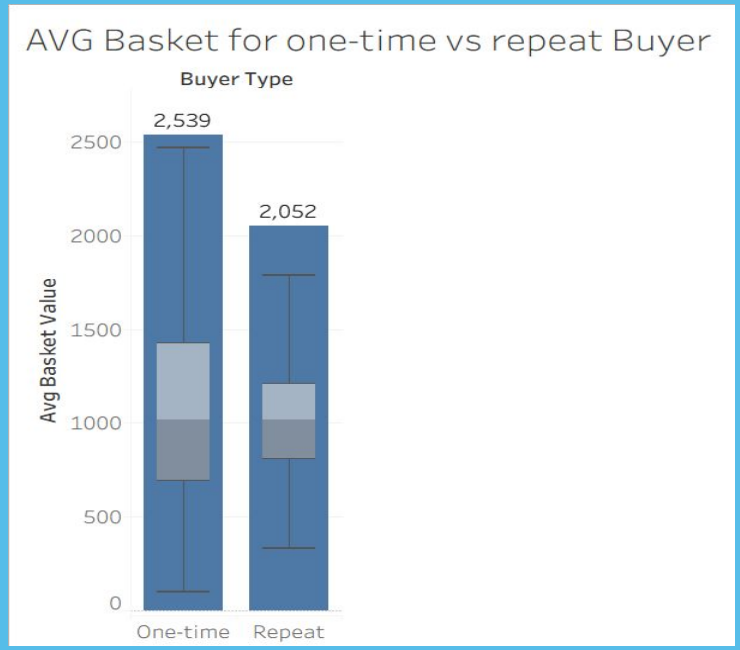
Business Problem

Hypothesis and Metrics

Issue Branch	Hypothesis	Metrics
No Product Bundling	Customers spend less per transaction because Shopamore does not offer product bundling, which reduces opportunities for cross-selling and increasing basket size.	Multi-Category Basket % → % of carts that include products from more than one category.
Lack of discounts	Customers buy less frequently and spend less per transaction because Shopamore does not offer discounts or promotions, reducing purchase motivation.	Repeat Purchase Rate → % of customers who return for multiple purchases, compared to : Average Basket Value → Average amount spent per transaction.
Complex checkout process	Customers abandon their carts and complete fewer purchases because Shopamore's checkout process is too complex, reducing conversion.	Orders with Only 1 Item →% of transactions that contain only one item vs multi item because of the complexity

Buying Rate (avg order per purchase)





Because frequent buyer's individual order value is not dramatically larger, lifetime value comes from additional orders over time, not bigger baskets.

Buying Rate : T-Test Method

H0 : Frequent buyer order = Infrequent buyer order
H1 : Frequent buyer order ≠ Infrequent buyer order
∝ : 0.05

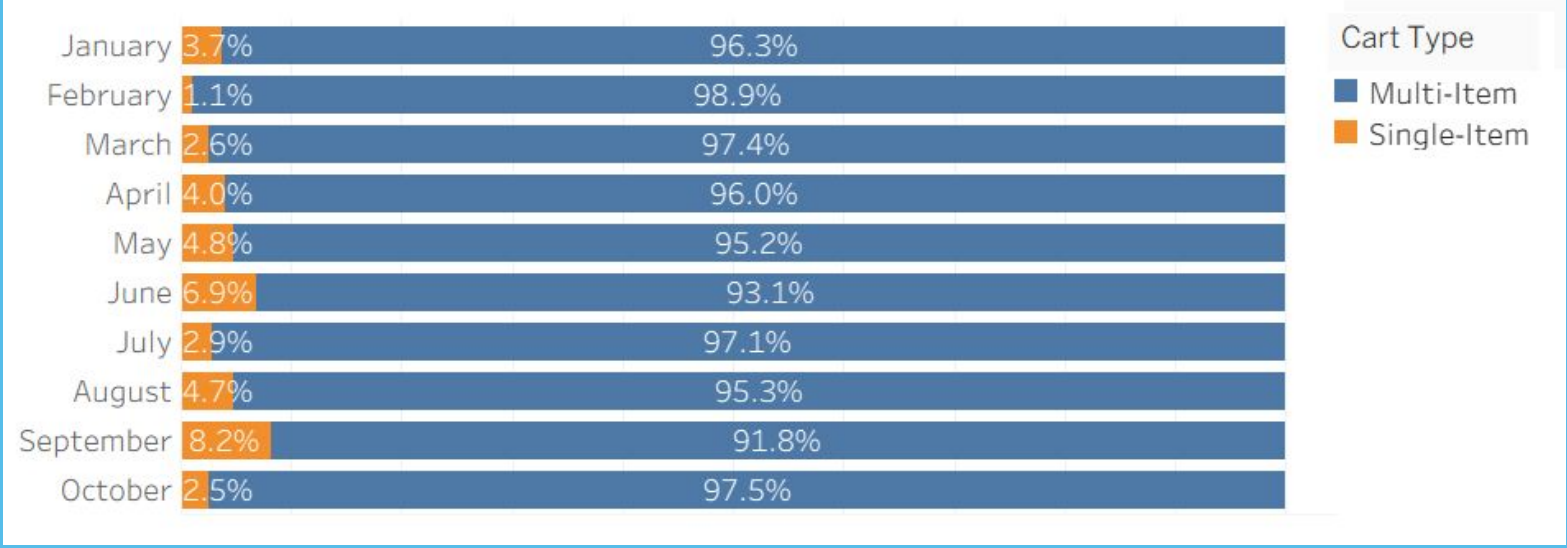
	Variable 1	*Variable 2*
Mean	1028.866	1056.804469
Variance	100024.7	247390.69
Observations	258	358
Pooled Variance	185708.2	
Hypothesized Mean Difference	0	
df	614	
t Stat	-0.79387	
P(T<=t) one-tail	0.213789	
t Critical one-tail	1.647339	
P(T<=t) two-tail	0.427577	
t Critical two-tail	1.963835	

43%

P-value

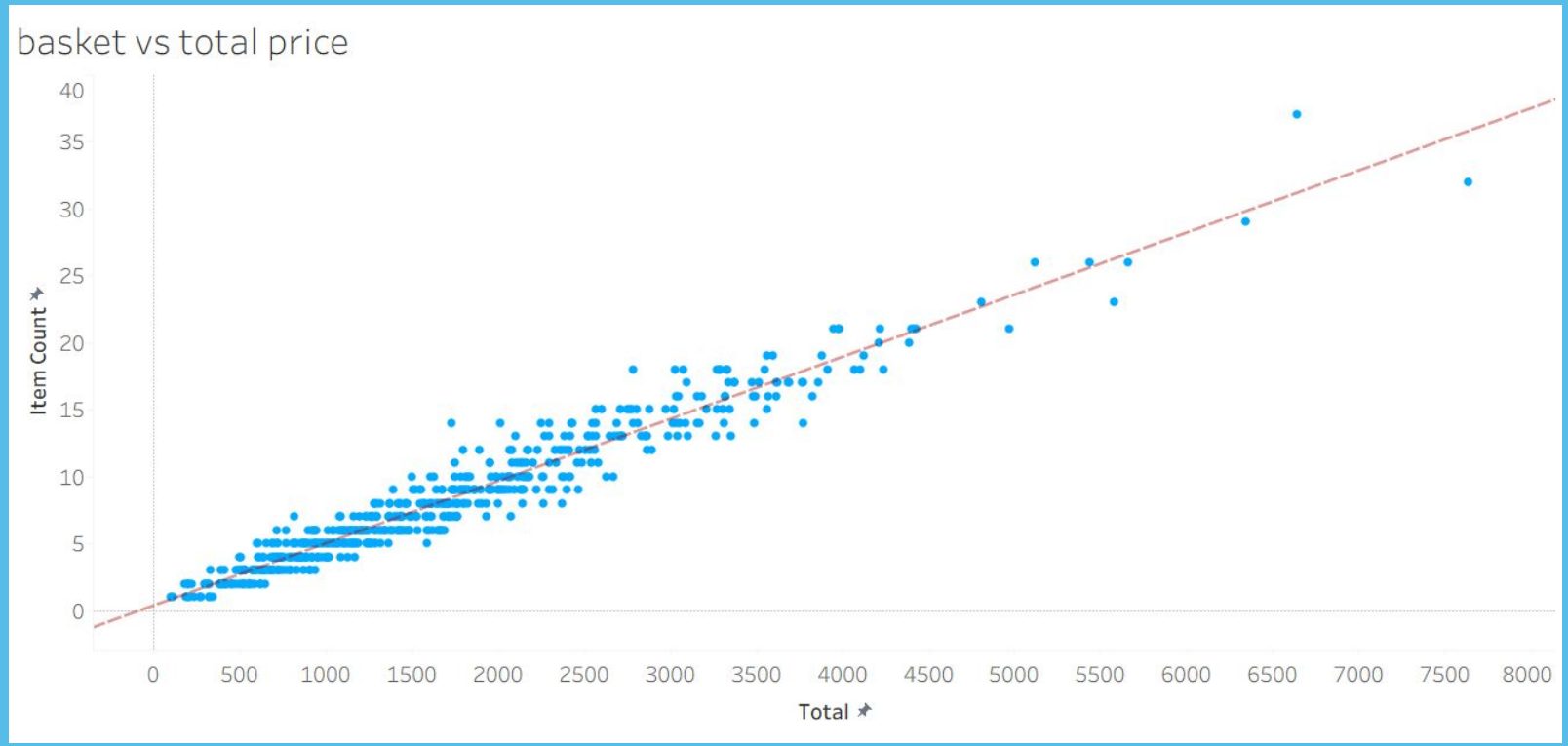
P-value = 0.423 > 0.05, then H0 accepted. There's no difference between frequent and infrequent buyer order.
Recommendation : Both groups behave similarly for that variable, so targeting them differently for this metric may not be effective.

Multi Item vs single Item in one Order



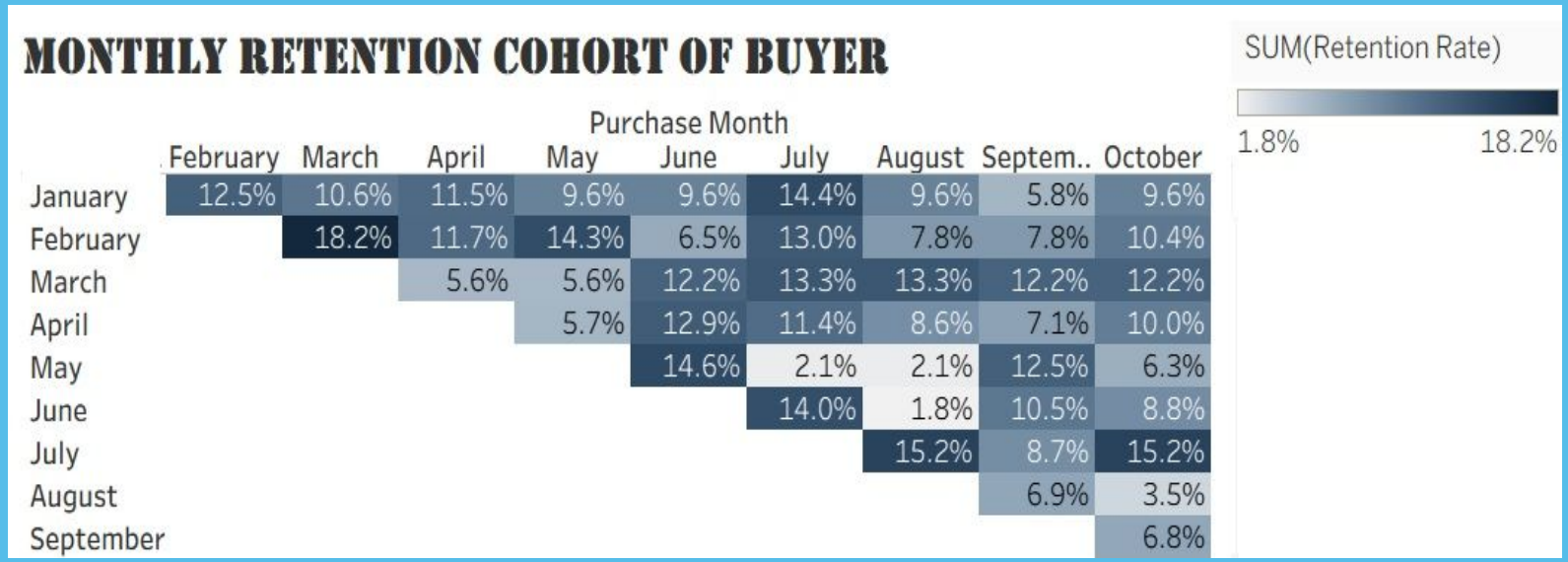
Monthly percentage for multi item in one cart stays very high- between 91% to 99%, while single-item cart stay in low percentage < 10%.
This is shows that buyer usually bought more than 1 item per order due to some reason

Correlation analysis between basket and total spend per customer



Strong Positive Correlation: The scatter plot shows that as item count increases, total price also increases almost linearly.
Recommendation : Encourage Larger Baskets like Offer free shipping or small discounts above certain item thresholds and Bundle High-Margin Products.

Purchase Frequency



February 2021 Strong Start – Highest Month 2 retention at 18.2%, better than January's 12.5% → likely due to a successful campaign.
General Decline – Most cohorts drop steadily after the first month.
Notable Spikes –

- July 2021 cohort high at 15.2% in Month 2.
- January 2021 saw a jump in July (14.4%).
- May 2021 dropped to 2.1% in July but rebounded to 12.5% in September.

Weak Cohorts

- August 2021 low at 6.9% Month 2.
- September 2021 low start at 6.8%.

Seasonality Signs – Retention rises around July–August and Sept–Oct, likely seasonal or promo-driven.

Retention problem : Conclusion

Retention falls by half or more after the first month.

After the initial drop, retention stabilizes at low single digits—indicating limited long-term engagement.

CONCLUSION:
The first month after a purchase are the critical window to turn one-time buyers into repeat customers.

Insight Summary & Recommendation

1. **Retention is key** ⇨ Persuade Infrequent to become frequent buyer with promotion loyalty - focused.

2. **Retention is front-loaded** ⇨ The first month after a purchase are the critical window to turn one-time buyers into repeat buyers.

3. **Most shoppers naturally bundle products.** ⇨ Additional promotion for bundling items.

Business Recommendation

The Priority – For buyer retention case :

- Re-engagement Programs – (High Impact / Low Effort)Launch tailored email or push campaigns targeting one-time buyers within 30 days of purchase.
- Loyalty Program – (High Impact / High Effort) Offer tiered rewards for repeat purchases (e.g., free shipping after the second order).
- First-purchase follow-up – (High Impact / Low Effort) Deploy automated emails, loyalty credits, or time-limited coupons within the first two weeks to capture that crucial Month-1 return.
- Personalized win-back flows – (High Impact / Low Effort) For customers inactive 30+ days, launch tailored offers or reminders to re-engage before they lapse completely.

Business Recommendation

For better revenue per order (*Non Priority, but nice to have*) :

- Bundle discounts – (High Impact / Low Effort) Offer “buy 3 different categories, get 10% off” promotions to encourage diversity and higher spend in one trip.

- Upsell single-category shoppers – (Low Impact / Low Effort) Target the small single-category segment with “complete your basket” prompts to raise their basket diversity.

- Free Shipping – (High Impact / Low Effort) Offer “free shipping on orders more than $150”. Encourages Buyers to add one or two more items to reach the minimum.

Dashboard Overview

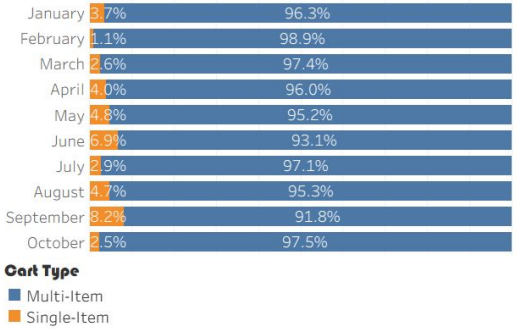
Total Revenue	Total Item Sold	Total Orders
$1,031,800.00	9,962 Pcs	993

Month of Order Date: (All)

State: (All)

Sales Revenue Per Month

Multi Item vs Single Item



basket diversity VS Order Spend

MONTHLY RETENTION COHORT OF BUYER

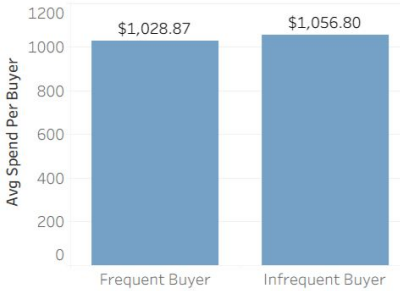
Purchase Month

	February	March	April	May	June	July	August	September	October
January	12.5%	10.6%	11.5%	9.6%	9.6%	14.4%	9.6%	5.8%	9.6%
February		18.2%	11.7%	14.3%	6.5%	13.0%	7.8%	7.8%	10.4%
March			5.6%	5.6%	12.2%	13.3%	13.3%	12.2%	12.2%
April				5.7%	12.9%	11.4%	8.6%	7.1%	10.0%
May					14.6%	2.1%	2.1%	12.5%	6.3%
June						14.0%	1.8%	10.5%	8.8%
July							15.2%	8.7%	15.2%
August								6.9%	3.5%
September									6.8%

Retention Rate

AVG Basket for one-time vs repeat Buyer

One-TIme (Infrequent) vs Repeat(Frequent) Average Spend



Appendix

Dataset Structure and Relationships

Master Table

products
product_id(PK)
product_type
product_name
size
colour
price
quantity
description

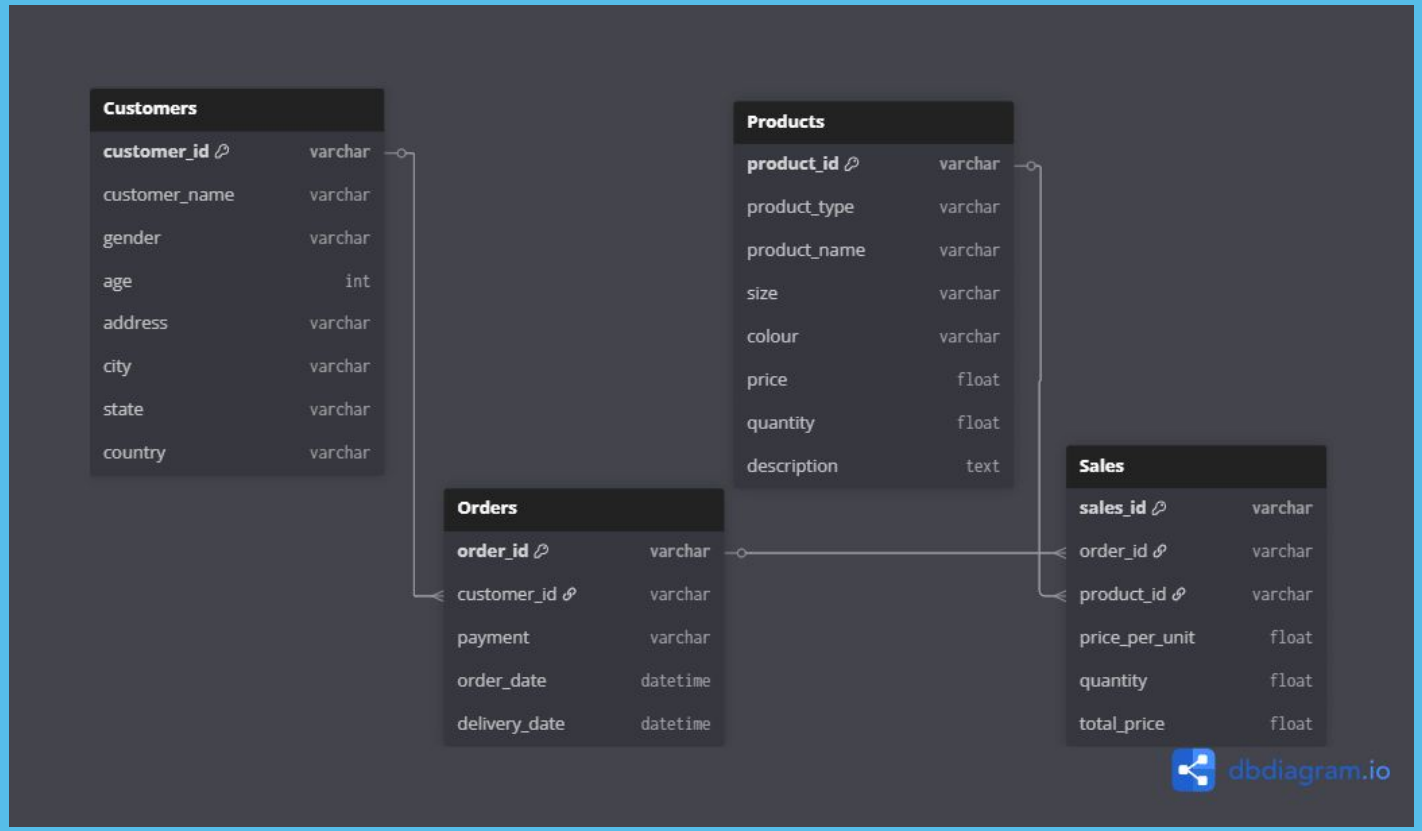
customers
customer_id(PK)
customer_name
gender
age
home_address
zip_code
city
state
country

Transactional Table

orders
order_id(PK)
customer_id
payment
order_date
delivery_date

sales
sales_id(PK)
order_id
product_id
price_per_unit
quantity
total_price

Appendix
Dataset Structure and Relationships



Appendix
Reminder List : Analysis Methods

Issue Branch	Metric	Analysis Method
Infrequent shoppers	Buying Rate (avg spend per purchase)	t-test (compare frequent vs infrequent shoppers' avg spend)
Poor customer satisfaction	Purchase Frequency	Time-series analysis (trend of purchases per month)
Limited product variety	Average Basket Size, Category Diversity	Correlation analysis (check relationship between basket diversity and total spend)
No Product Bundling	Multi-Category Basket %	Chi-square test (compare proportion of single vs multi-category carts)
Lack of discounts	Repeat Purchase Rate, Average Basket Value	t-test (compare avg basket value of repeat vs one-time buyers)
Complex checkout process	Sessions with Only 1 Item	Chi-square test (proportion of single-item vs multi-item carts)

Appendix : Calculation and Test Method

Metrics : Buying Rate : T-Test

H0 : Frequent buyer order = Infrequent buyer order
H1 : Frequent buyer order ≠ Infrequent buyer order
∝ : 0.05

	Variable 1	*Variable 2*
Mean	1028.866	1056.804469
Variance	100024.7	247390.69
Observations	258	358
Pooled Variance	185708.2	
Hypothesized Mean Difference	0	
df	614	
t Stat	-0.79387	
P(T<=t) one-tail	0.213789	
t Critical one-tail	1.647339	
P(T<=t) two-tail	0.427577	
t Critical two-tail	1.963835	

43%

P-value

P-value = 0.423 > 0.05, then H0 accepted. There's no difference between frequent and infrequent buyer order.
Recommendation : Both groups behave similarly for that variable, so targeting them differently for this metric may not be effective.

Appendix : Calculation and Test Method

Metrics : Purchase Frequency

SUM of retention_rate / purchase_month	2021-02-01	2021-03-01	2021-04-01	2021-05-01	2021-06-01	2021-07-01	2021-08-01	2021-09-01	2021-10-01
2021-01-01	12.5	10.58	11.54	9.62	9.62	14.42	9.62	5.77	9.62
2021-02-01		18.18	11.69	14.29	6.49	12.99	7.79	7.79	10.39
2021-03-01			5.56	5.56	12.22	13.33	13.33	12.22	12.22
2021-04-01				5.71	12.86	11.43	8.57	7.14	10
2021-05-01					14.58	2.08	2.08	12.5	6.25
2021-06-01						14.04	1.75	10.53	8.77
2021-07-01							15.22	8.7	15.22
2021-08-01								6.9	3.45
2021-09-01									6.82

February 2021 Strong Start – Highest Month 2 retention at **18.18%**, better than January's **12.5%** → likely due to a successful campaign.
General Decline – Most cohorts drop steadily after the first month.
Notable Spikes –

- July 2021 cohort high at **15.22%** in Month 2.
- January 2021 saw a jump in July (**14.42%**).
- May 2021 dropped to **2.08%** in July but rebounded to **12.5%** in September.

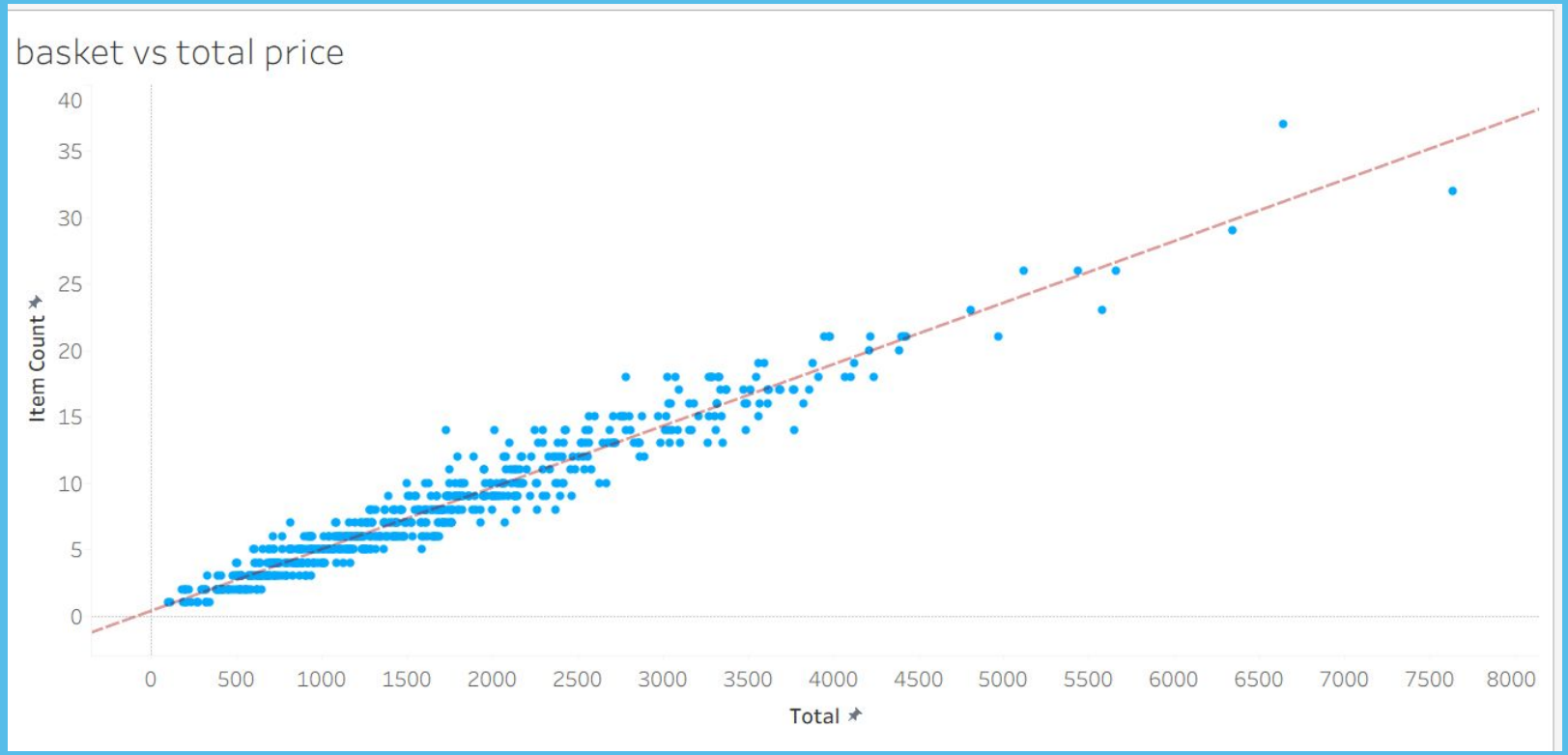
Weak Cohorts

- August 2021 low at **6.9%** Month 2.
- September 2021 low start at **6.82%**.

Seasonality Signs – Retention rises around **July–August** and **Sept–Oct**, likely seasonal or promo-driven.

Appendix : Calculation and Test Method

Method : Correlation analysis between basket and total spend per customer



Strong Positive Correlation: The scatter plot shows that as item count increases, total price also increases almost linearly.
Recommendation : Encourage Larger Baskets like Offer free shipping or small discounts above certain item thresholds and Bundle High-Margin Products.

Appendix : Calculation and Test Method

Metrics : Multi-Category Basket - chi square method

NB* The data used as category is product type from table products.

```
SELECT
    s.order_id,
    CASE WHEN COUNT(DISTINCT p.product_type) = 1 THEN 'Single'
         ELSE 'Multi' END AS category_type
FROM `turnkey-banner-458413-i3.shopamore.orders` s
JOIN `turnkey-banner-458413-i3.shopamore.products` p
    ON s.product_id = p.product_id
GROUP BY s.order_id;
```

```
df = pd.read_csv("category_type_per_order.csv")

# Create frequency table
contingency_table = pd.crosstab(index=df['category_type'], columns="count")

# Chi-square test
chi2, p, dof, expected = chi2_contingency(contingency_table)

print("Chi-square statistic:", chi2)
print("p-value:", p)
if p < 0.05:
    print("Significant difference in proportion of single vs multi-category carts.")
else:
    print("No significant difference in proportion.")
```

```
Chi-square statistic: 0.0
p-value: 1.0
No significant difference in proportion.
```

Reccomendation : Use purchase history to push cross-selling within customer's preferred categories, since category count per basket doesn't differ between groups.

Appendix : Calculation and Test Method

Repeat Purchase Rate by Average Basket Value : T-Test Method

H0 : Frequent buyer basket = Infrequent buyer basket
H1 : Frequent buyer basket ≠ Infrequent buyer basket
∝ : 0.05

	Variable 1	*Variable 2*
Mean	12.37597	5.047486
Variance	24.50012	4.846478
Observations	258	358
Pooled Variance	13.07284	
Hypothesized Mean Difference	0	
df	614	
t Stat	24.81933	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.647339	
P(T<=t) two-tail	0.00	
t Critical two-tail	1.963835	

0%

P-value

P-value = 0 < 0.05, then H1 accepted. There's significant difference between frequent and infrequent buyer order.
Recommendation : Strengthen loyalty for frequent buyers (VIP perks, rewards) and reactivate infrequent buyers (targeted promos, reminders).

```
Notes : second test is Mann-Whitney U method.
Result :
Mann-Whitney U statistic: 86661.5
p-value: 0.0000000
Significant difference between repeat and one-time buyers.
Conclusion : same result with t-test method
```

Appendix : Calculation and Test Method

Sessions with Only 1 Item vs Multi-Item : chi square method

```
SELECT
    s.order_id,
    CASE WHEN COUNT(s.product_id) = 1
THEN 'Single-item'
        ELSE 'Multi-item' END AS cart_type
FROM `turnkey-banner-458413-i3. sales`
s
GROUP BY s.order_id;
```

Recommendation : Shift segmentation to Average Order Value (AOV), product mix, or margin per order instead of cart size

```
# Load SQL output
df = pd.read_csv("cart_type_per_order.csv")

# Create frequency table
contingency_table = pd.crosstab(index=df['cart_type'], columns="count")

# Chi-square test
chi2, p, dof, expected = chi2_contingency(contingency_table)

print("Chi-square statistic:", chi2)
print("p-value:", p)

if p < 0.05:
    print("Significant difference in proportion of single-item vs multi-item carts.")
else:
    print("No significant difference in proportion.")
```

```
Chi-square statistic: 0.0
p-value: 1.0
No significant difference in proportion.
```

Appendix : List of Link

You can check the tableau public here

dataset : shopping-cart-database

Colab File

Connect to My LinkedIn