Online Retail Data

Customer and products analysis & quick wins business recommendation

The UCI Machine Learning Repository

Nugrah Nurrohman / 15-10-2025

Executive Summary

Why?
Identify most valuable customer and product stocking and bundling strategy to increase profitability.
What?

1. Customer could be classified into 4 tiers based on the revenue generated. Elite, Prime, Core, Regular.
2. The Elite segments generate 40x more on total revenue, 10x more frequent and 10x more basket value compare to regular customer.
3. Peaks Month are in September, November and October.

How?

1. Launch loyalty program to The Elite segments.
2. Stock products most sold in peak season.
3. Launch and design campaign to the products that frequently purchased together.

Context Definition

This is a transactional dataset which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail.

The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers.

This analysis will provide insights & strategies that can be used to increase profitability by identifying most valuable customer and product stocking and bundling strategy.

Methodology & Tools Used

Methods:

a. EDA
b. Descriptive statistic
c. Apriori

Tools:

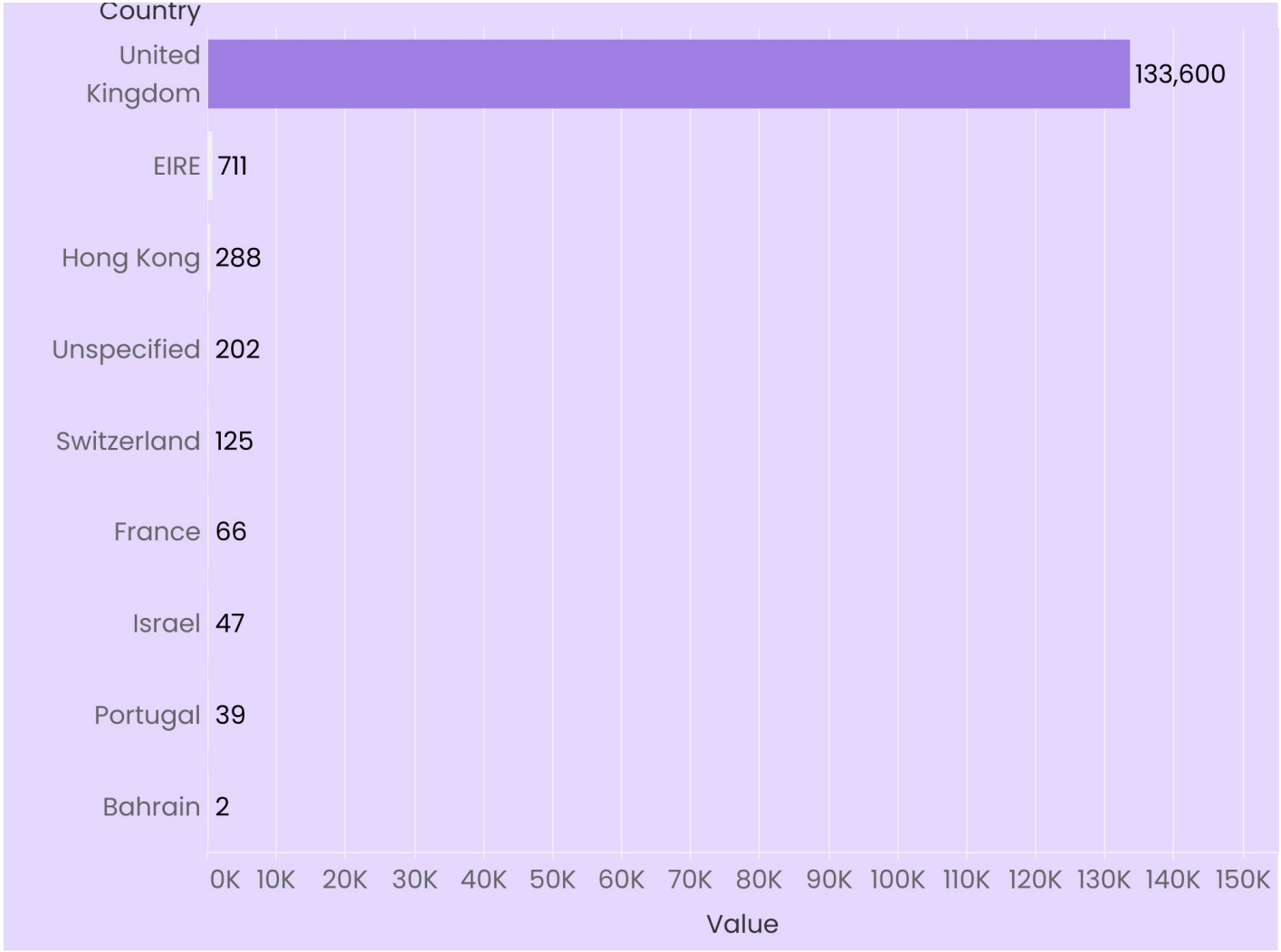
a. postgreSQL
b. R
c. Tableau
d. ChatGPT

Key Findings

Country Missing Values

UK is the Highest

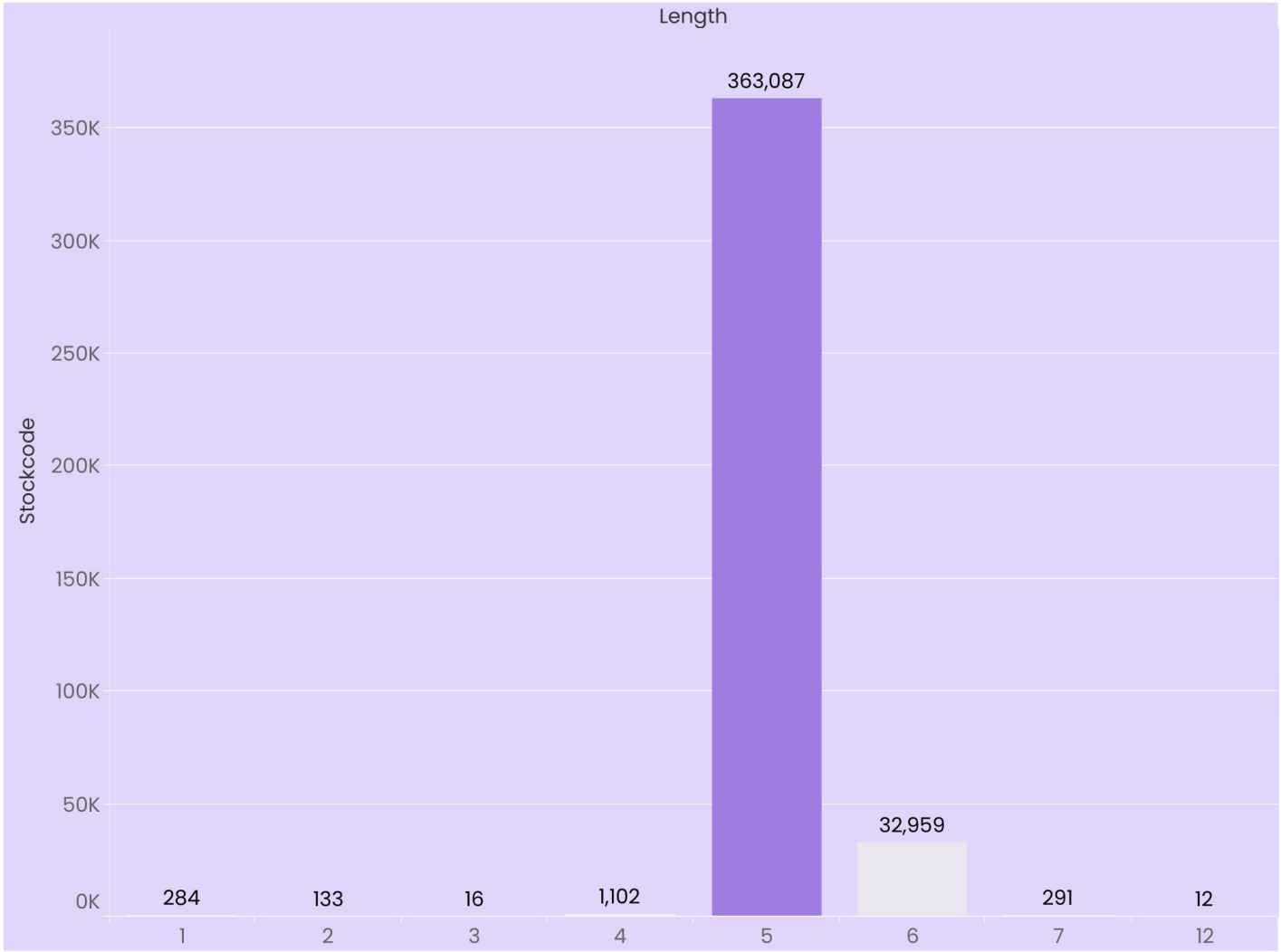
Focus more consistent formatting on UK



Inconsistent product numbering

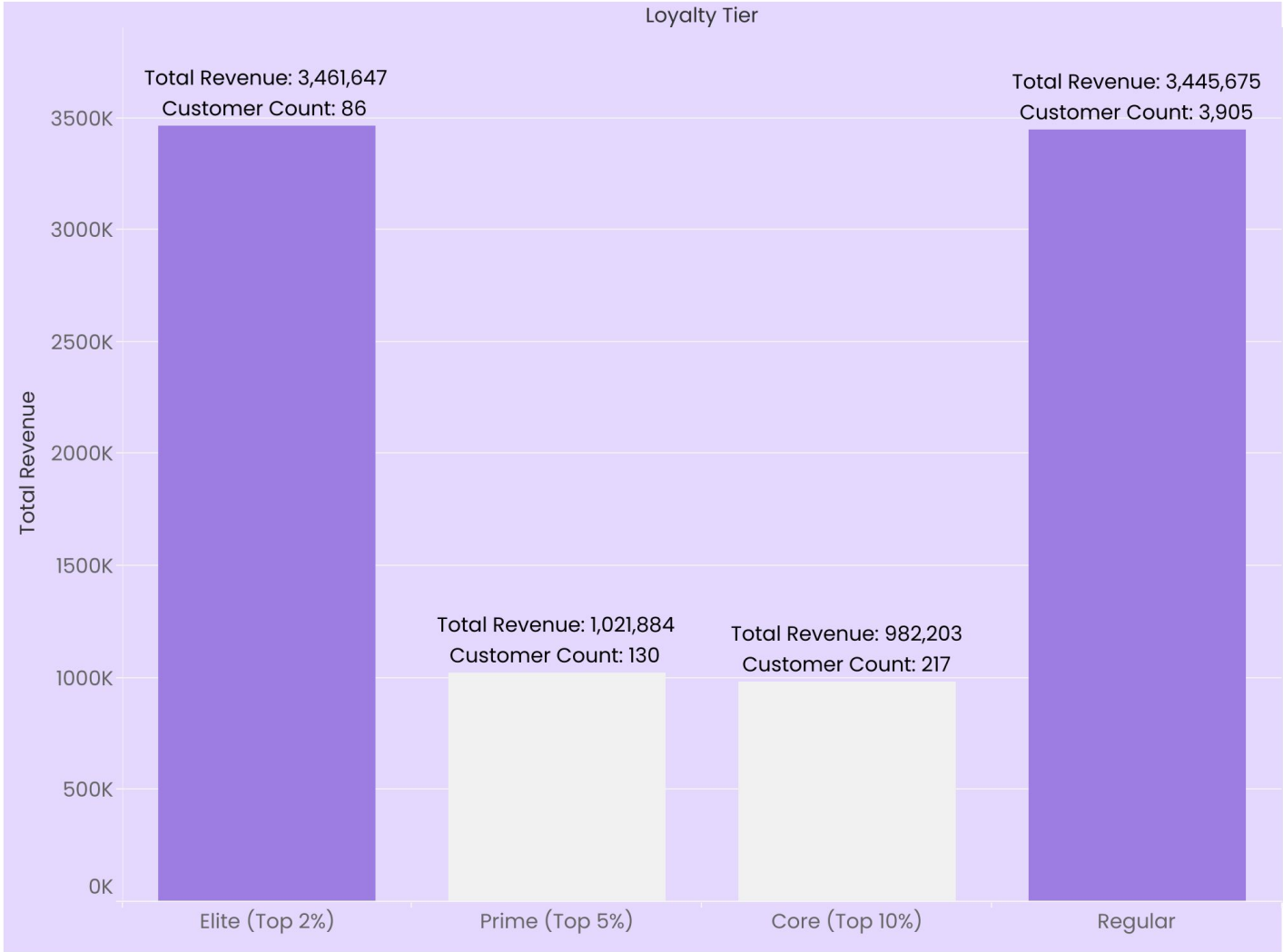
Most are formatted using 5 digit code

Consistent and good formatting will be easier for analysis



Top Customer

Customer could be ranked based on revenue generated

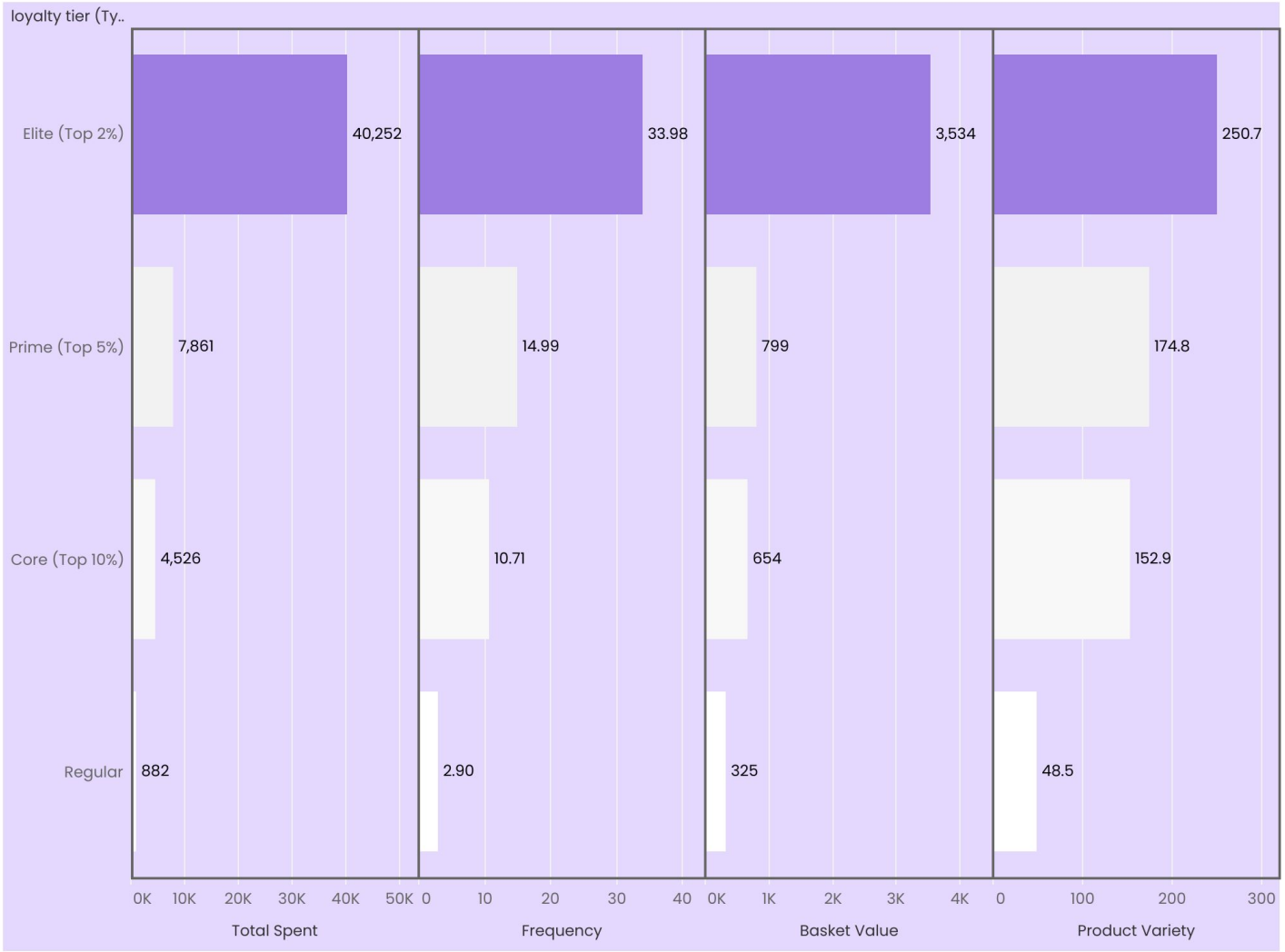


Customer Tier Traits

Average:
-Total Spent,
-Frequency,
-Basket Value/Size,
-Product variety

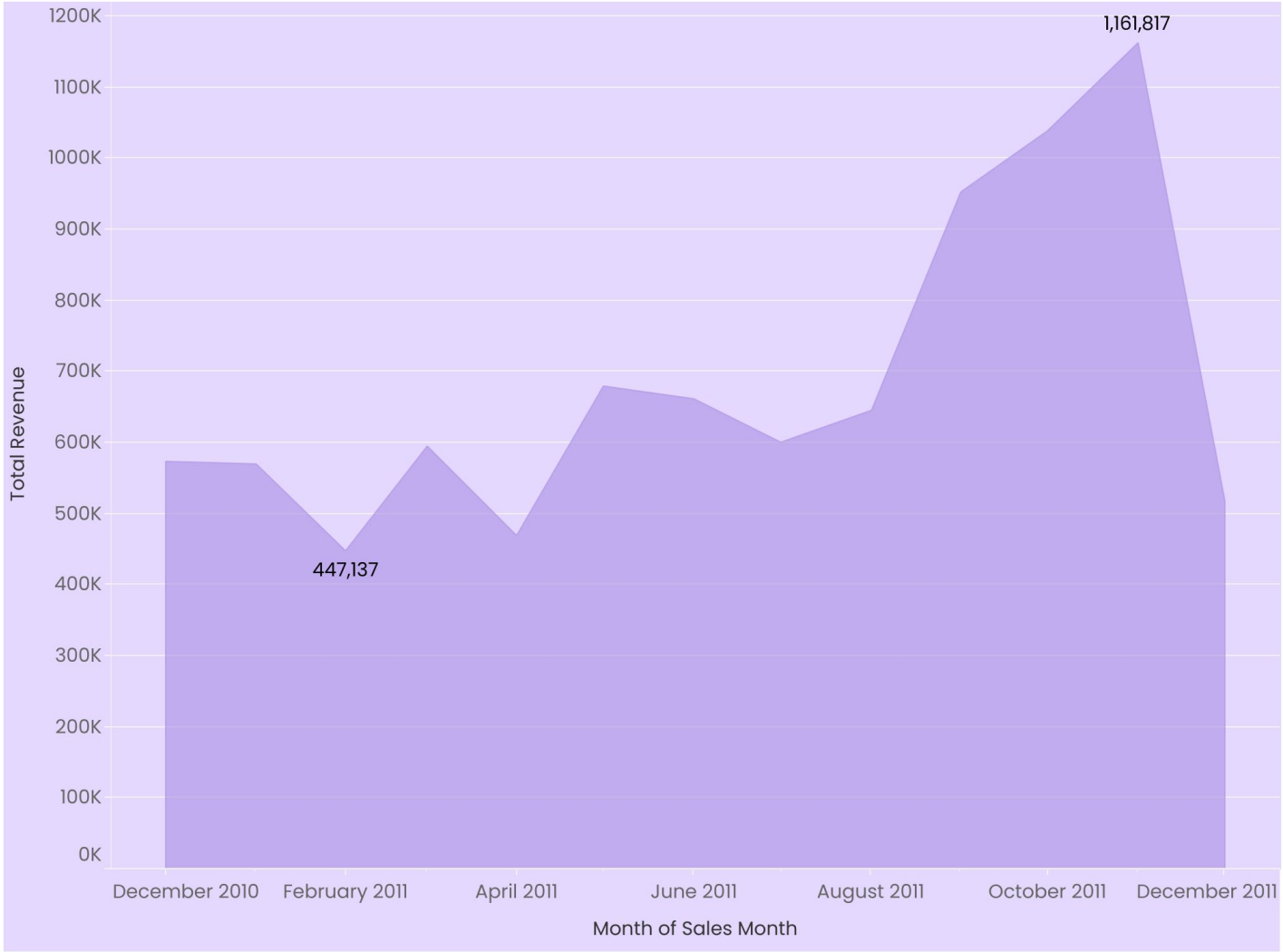
Those elit customer are highly valuable for business.

Prioritise Elite segments by launching relevant loyalty program.



Sales graphic

Sales tend to spike near end of the year



Monthly Sales Change

How's the monthly change?

September, october, november are the peak months

sales_month	total_revenue	change
2010-12-31	572714	
2011-01-31	569445	-0.57%
2011-02-28	447137	-21.48%
2011-03-31	595501	33.18%
2011-04-30	469200	-21.21%
2011-05-31	678595	44.63%
2011-06-30	661214	-2.56%
2011-07-31	600091	-9.24%
2011-08-31	645344	7.54%
2011-09-30	952838	47.65%
2011-10-31	1039319	9.08%
2011-11-30	1161817	11.79%
2011-12-31	518193	-55.40%

Top 10 product in peak months

This is the top product of the peak months

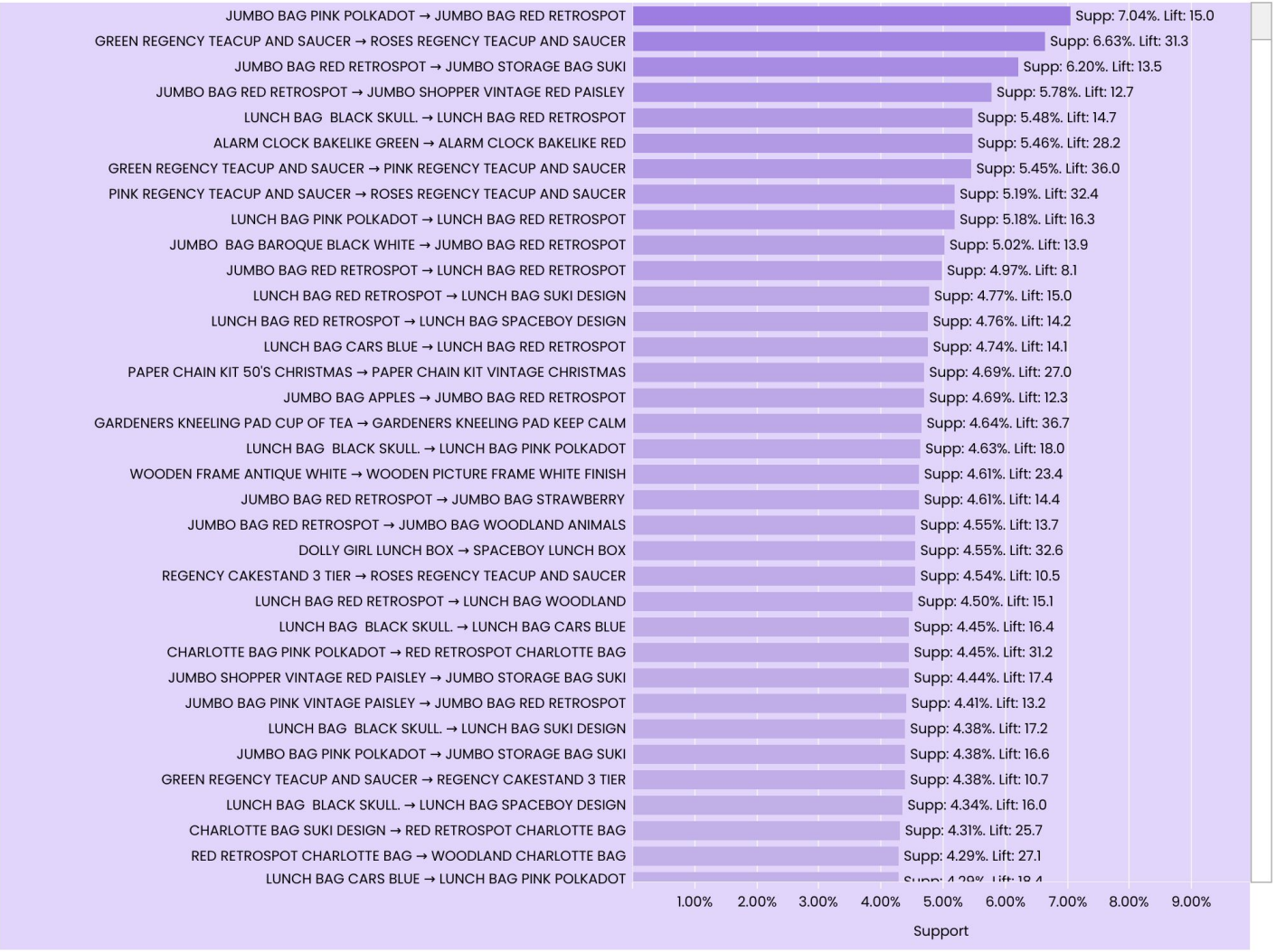
Stocking these products, would be valuable for the business.

	Month		
Description	September 2011	October 2011	November 2011
ASSORTED COLOUR BIRD ORNAMENT			5,190
CHILLI LIGHTS			1,957
DOORMAT KEEP CALM AND COME IN		1,503	
FELTCRAFT DOLL MOLLY		2,447	
HOT WATER BOTTLE KEEP CALM	1,326	1,317	2,102
JUMBO BAG PINK POLKADOT		3,098	
JUMBO BAG RED RETROSPOT	4,177	5,237	5,678
JUMBO BAG VINTAGE DOILY	2,679		
PAPER CHAIN KIT 50'S CHRISTMAS	2,064	3,243	5,919
PAPER CHAIN KIT VINTAGE CHRISTMAS			3,181
PINK BLUE FELT CRAFT TRINKET BOX		4,873	
POSTAGE	279	338	460
RABBIT NIGHT LIGHT		6,249	12,393
REGENCY CAKESTAND 3 TIER	730	1,028	944
SCOTTIE DOG HOT WATER BOTTLE	1,079		
SET OF 3 CAKE TINS PANTRY DESIGN	1,207		
SET OF TEA COFFEE SUGAR TINS PANTRY	2,056		
WHITE HANGING HEART T-LIGHT HOLDER	2,444		4,861

Products Frequently purchased together

High support high lift

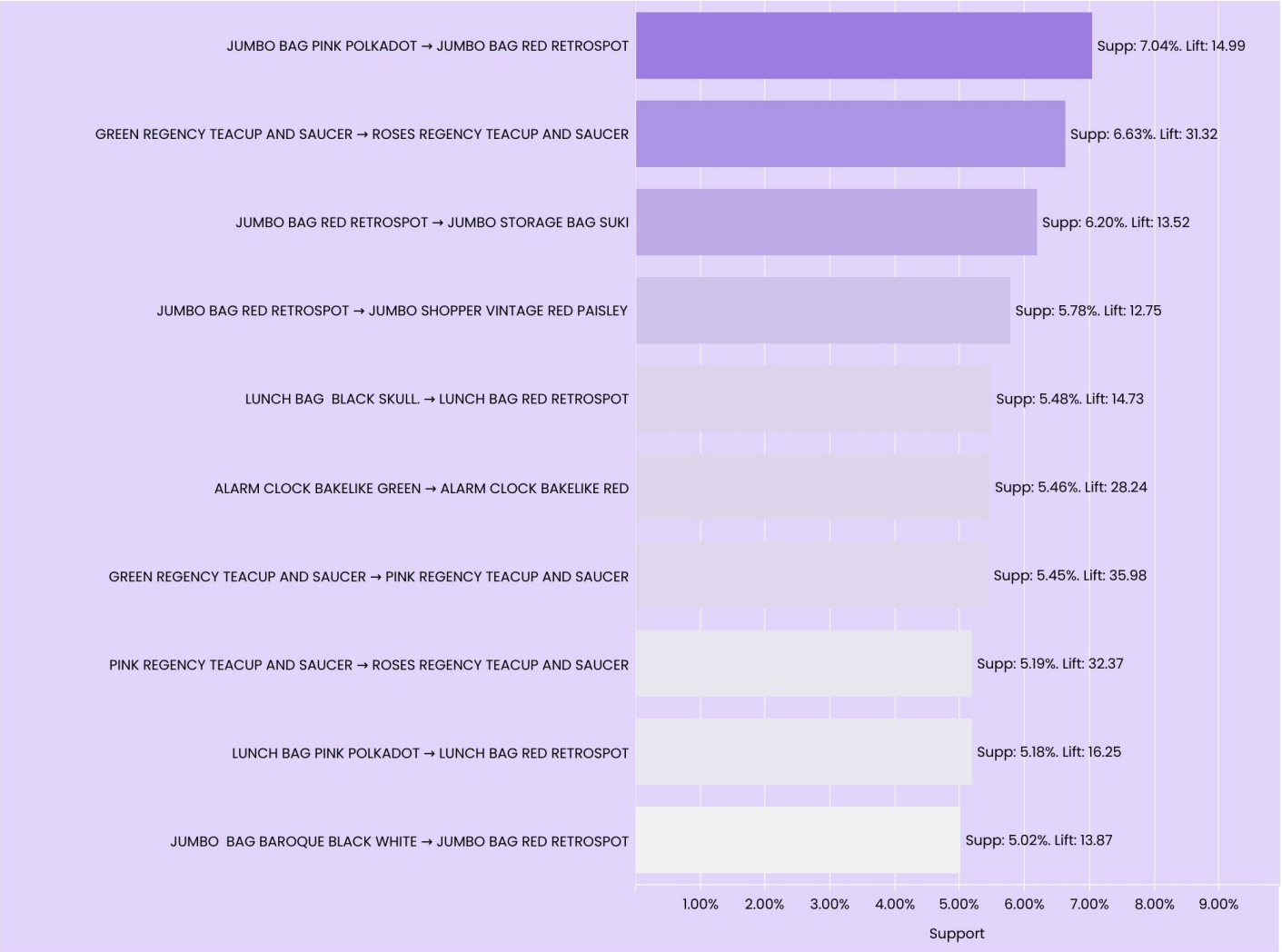
These could increase the basket size of customers



Top 10 Associated Product

What can we do?

1. Thematic promotion bundle. “Romantic color bag: Pink + Red bag Bundle”
2. Close product placement
3. Cross sell in cart when Add A product, add B product bellow the cart
4. Design product page based on co-interest



Next Steps

1. Launch loyalty program to The Elite segments
2. Stock products most sold in peak season
3. Launch and design campaign to the products that frequently purchased together

Appendix

Ranking Customer with sql

```
WITH customer_revenue AS (
  SELECT
    CustomerID,
    SUM(UnitPrice * Quantity) AS total_revenue
  FROM online_retail
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
),
ranked AS (
  SELECT
    CustomerID,
    total_revenue,
    CUME_DIST() OVER (ORDER BY total_revenue DESC) AS revenue_dist
  FROM customer_revenue
)
SELECT
  CustomerID,
  total_revenue,
  CASE
    WHEN revenue_dist <= 0.02 THEN 'Elite (Top 2%)'
    WHEN revenue_dist <= 0.05 THEN 'Prime (Top 5%)'
    WHEN revenue_dist <= 0.10 THEN 'Core (Top 10%)'
    ELSE 'Regular'
  END AS loyalty_tier
FROM ranked
ORDER BY total_revenue DESC;
```

Appendix

Apriori algorithm to product purchased together

```
#install packages
install.packages("arules")
library("arules")

transaction <- read.csv("Cleaned - baskets.csv")

#separating the string as individual string
basket_list <- strsplit(transaction$items, ",\\s*")
transactions <- as(basket_list, "transactions")

#running the algorithm
rules <- apriori(transactions, parameter = list(supp = 0.01, conf = 0.1, maxlen = 3))

#Checking the algorithm
inspect(head(sort(rules, by="lift"), 20))

#exporting into .csv format
rules_df <- as(rules, "data.frame")
rules_df$lhs <- labels(lhs(rules_df))
rules_df$rhs <- labels(rhs(rules_df))
colnames(rules_df) <- c("Rule", "Support", "Confidence", "Lift", "LHS", "RHS")
write.csv(rules_df, "transactions_rules.csv", row.names = FALSE)
```