Marketing Strategies for XYZ E-Commerce Company

Data Science for Algorithmic Marketing

Final Project

Group 5: Roselyn Rozario, Adela Cho, and Mike Meissner

Agenda

Introduction
Business Context
Marketing Problem

Data
Source and Overview
Data Hygiene (Pre-Processing and Feature Engineering)

Exploratory Data Analysis

Models
Association Rules Algorithm
K-Means Clustering

Marketing Insights

Introduction

Business Context

XYZ E-Commerce Company is seeking to better understand their European customers and their behavior through the products that they typically demand and how to segment customers to develop marketing strategies to increase profits.

ARM Consulting Company will assist in identifying customer clusters to create the best marketing strategies and recommendations necessary to maintain and attract customers for this unique all-occasion gift establishment

Marketing Problem

XYZ E-Commerce Company lacks consistency in the amount of customer transaction throughout the year with more transactions during the winter (holiday) months and a drastic decrease and unsteadiness in the beginning of the year.

The E-Commerce Company hopes to increase sales during these inactive months through an analysis of consumer clusters and targeted marketing strategies.

Data

Data Source and Description

Data acquired from the UC Irvine Machine Learning Repository.

Transnational dataset containing transactions occurring at a UK-based online only retail store between January 12th, 2010 and September 12th , 2011.

The data mainly consists of primary data.

The store mainly specializes in selling unique all-occasion gifts.

Many of the customers of this online store are wholesalers.

Data: https://www.kaggle.com/datasets/carrie1/ecommerce-data

Data Overview

541,909 transactions recorded.

Purchases made from European countries like the United Kingdom, Germany, France and Spain.

4,300 distinct customers utilized the shop during specified time-period.

Dataset Features

Feature	Description
InvoiceNo	Unique Identifier for Each Transaction
StockCode	Product/Item Code
Description	Product Description
Quantity	Number of Items Purchased in Each Transaction
InvoiceDate	Timestamp of Each Transaction
UnitPrice	Product Price Per Unit
CustomerID	Unique Identifier for Each Customer
Country	Customer's Country of Residence

Data Hygiene

This part of the process included preparing the data and performing feature engineering :

Examining data for any limitations (i.e., nulls, etc.)

Addressing/reviewing null values and duplicated rows.

Added numerical variables/columns based on existing data, such as recency based on reference date.

Grouping data by continent to filter European countries.

Exploratory Data Analysis

Descriptive Statistics

Numerical Columns

There are a limited amount of numerical columns.

Customer ID

Given that numbers make up a Customer ID, this column is being interpreted as a numerical column, but should not be included.

Quantity

The numbers provide perspective on various quantiles or sizes of orders.

Unit Price

The numbers provide perspective on various quantiles or prices.

	Quantity	UnitPrice	CustomerID
count	541909.000	541909.000	406829.000
mean	9.552	4.611	15287.691
std	218.081	96.760	1713.600
min	-80995.000	-11062.060	12346.000
25%	1.000	1.250	13953.000
50%	3.000	2.080	15152.000
75%	10.000	4.130	16791.000
max	80995.000	38970.000	18287.000

Quantity, Sales Price and Transactions
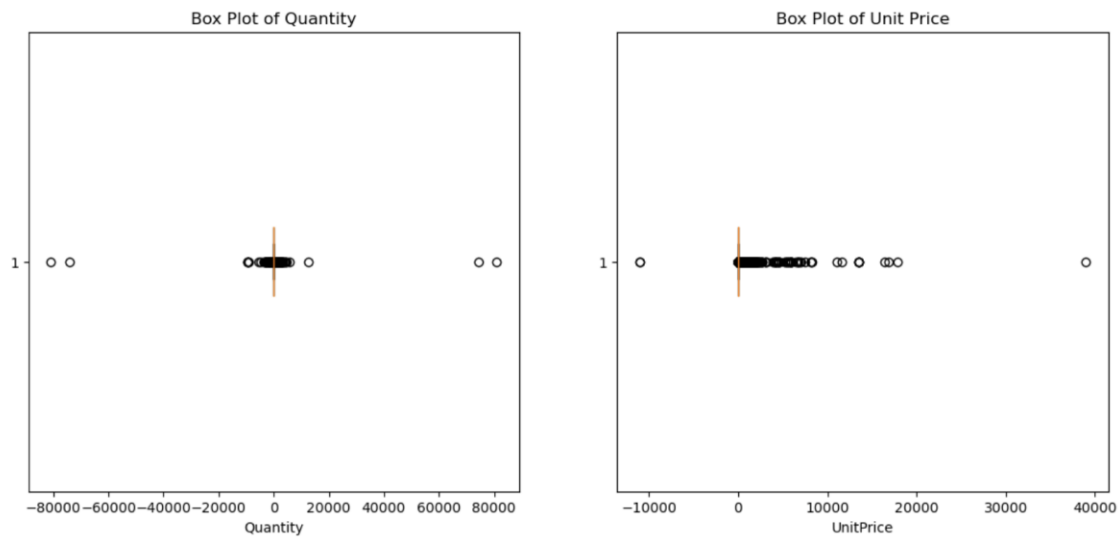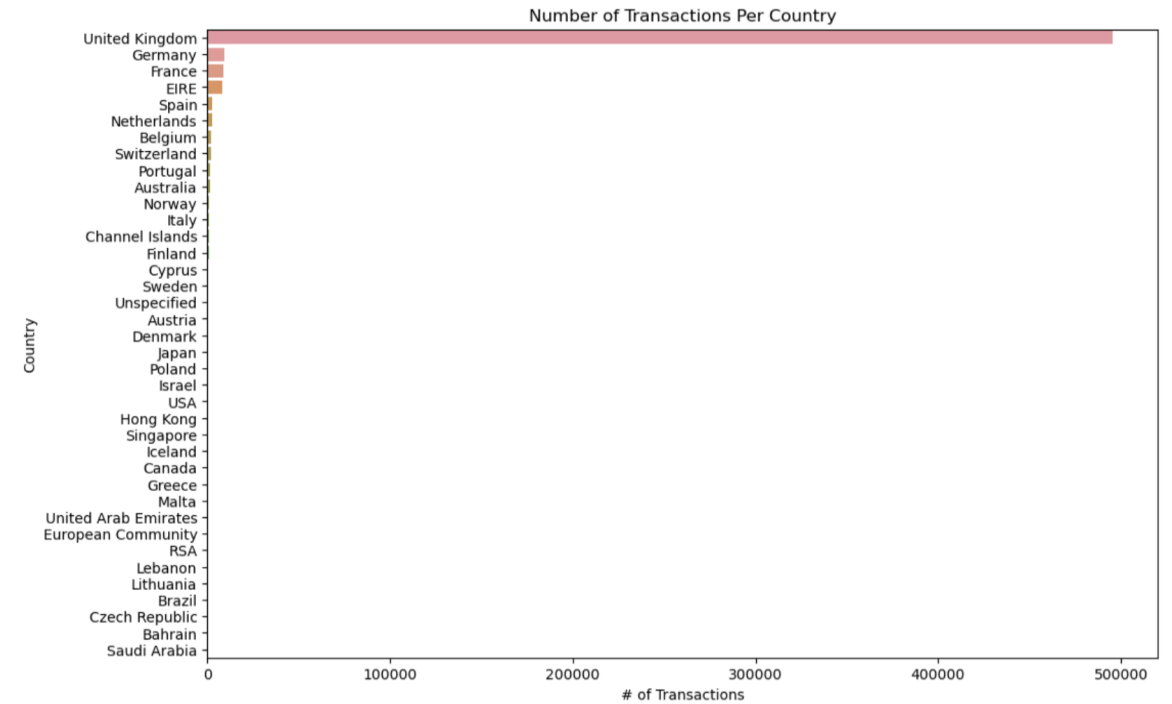

Figure 1

Box Plot of Quantity and Unit Price

Figure 2

Bar Chart of the Number of Transactions Per Country
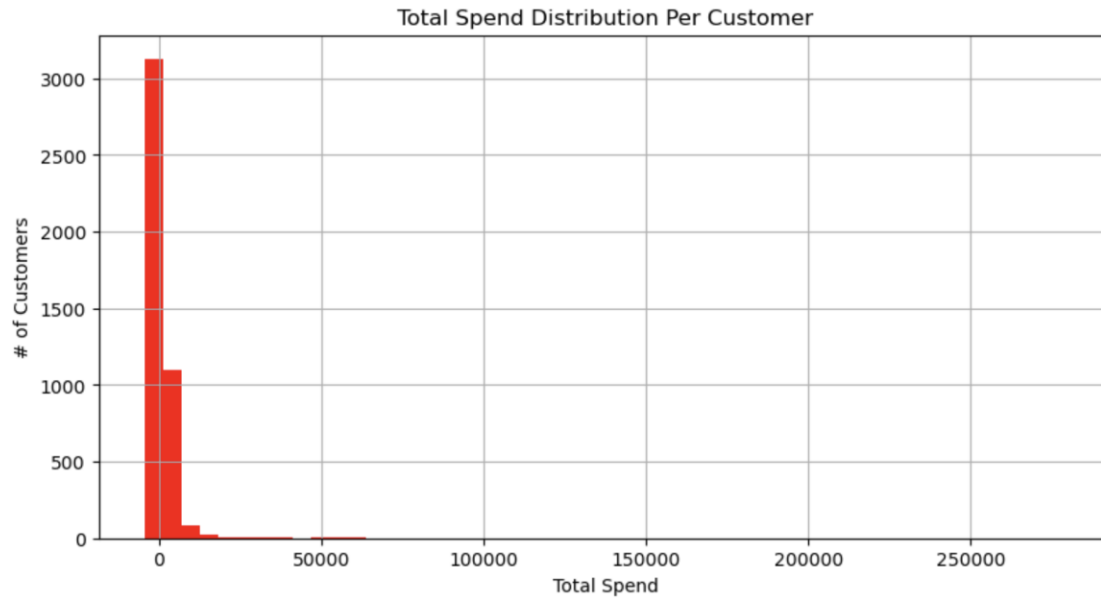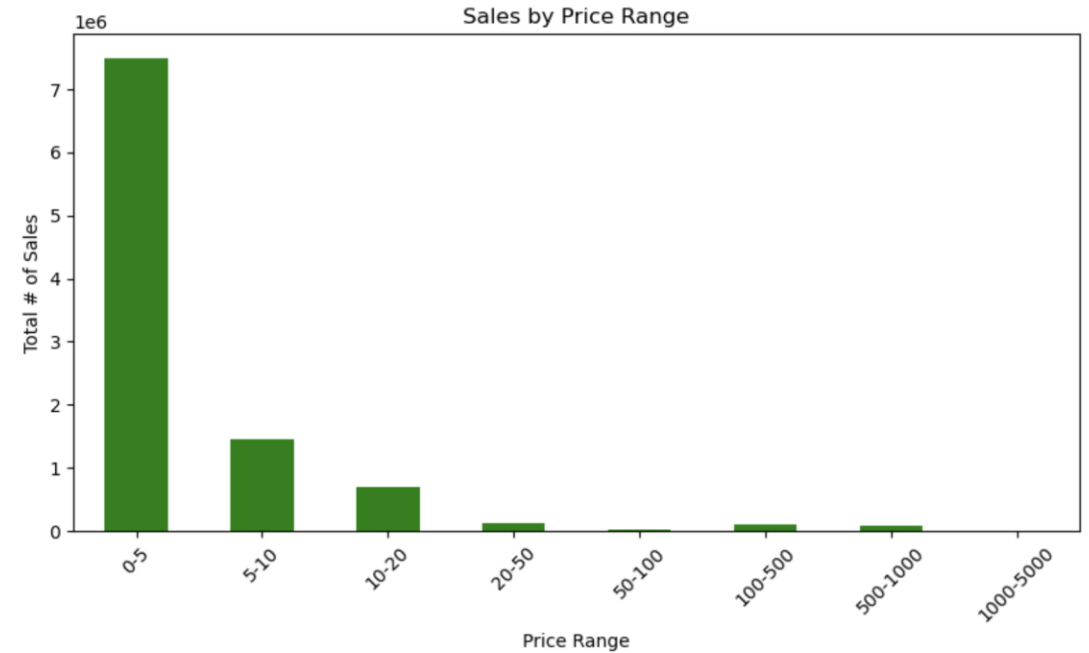
Spend and Sales

Total Spend Distribution Per Customer

Figure 3

Total Spend Distribution Per Customer

Sales by Price Range

Figure 4

Sales by Price Range
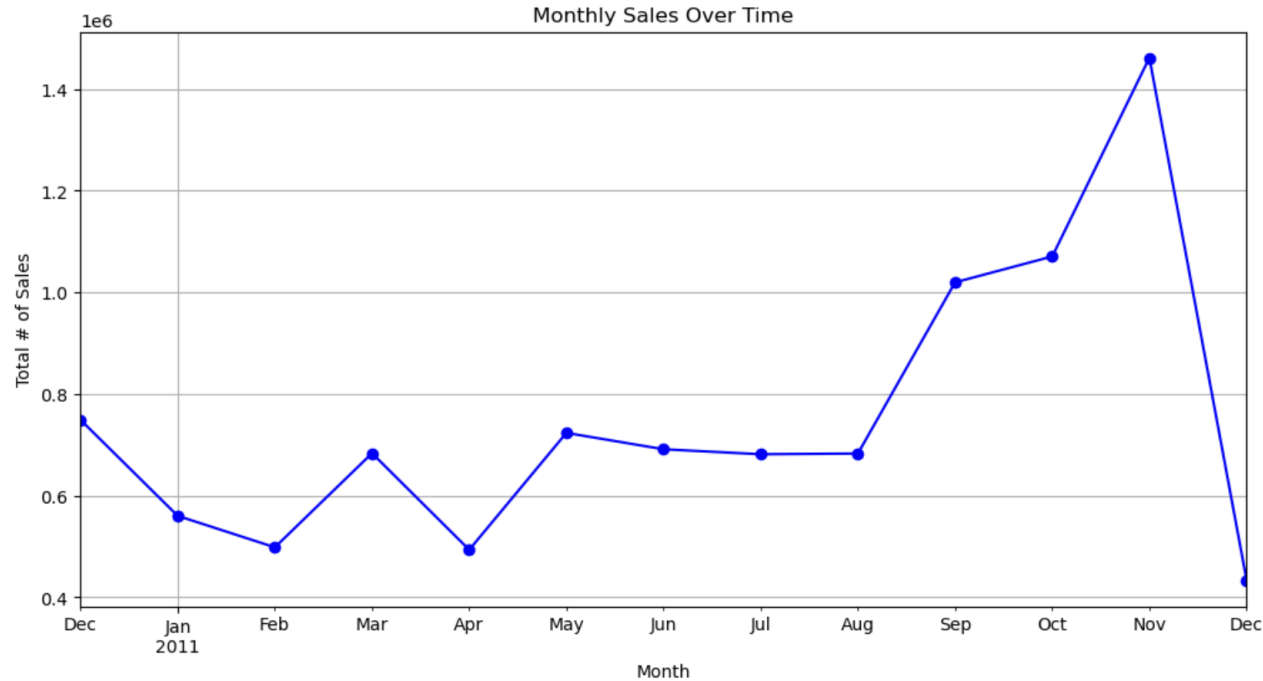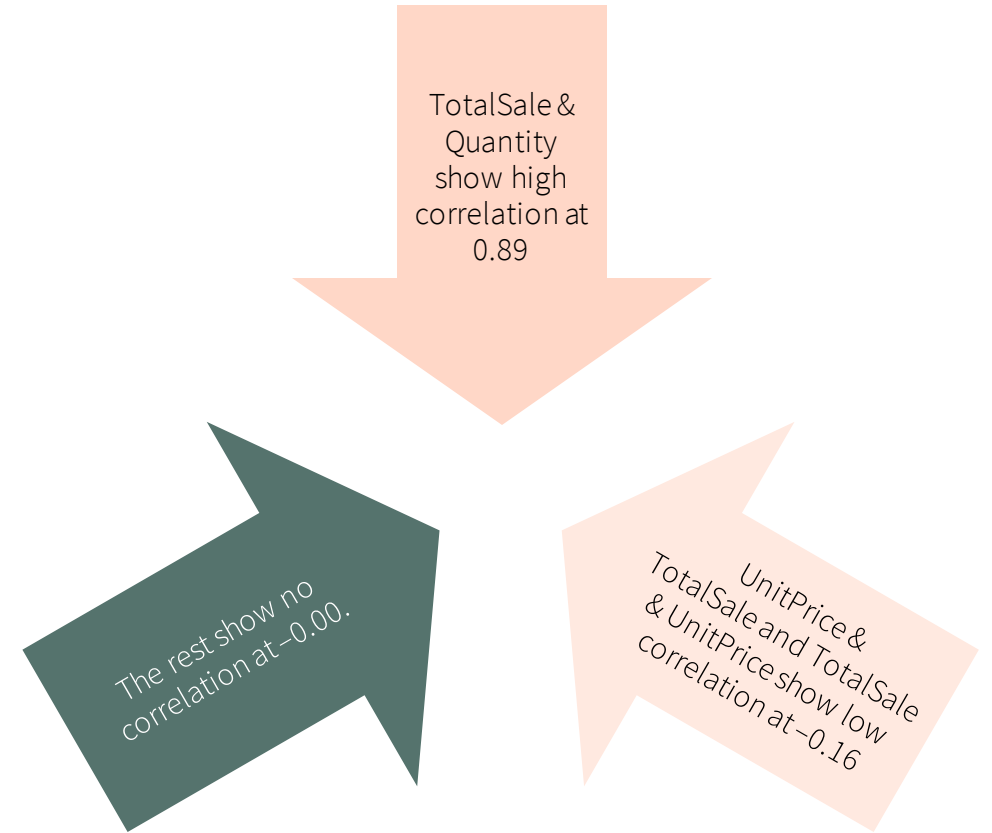
Sales Timeline

Figure 5

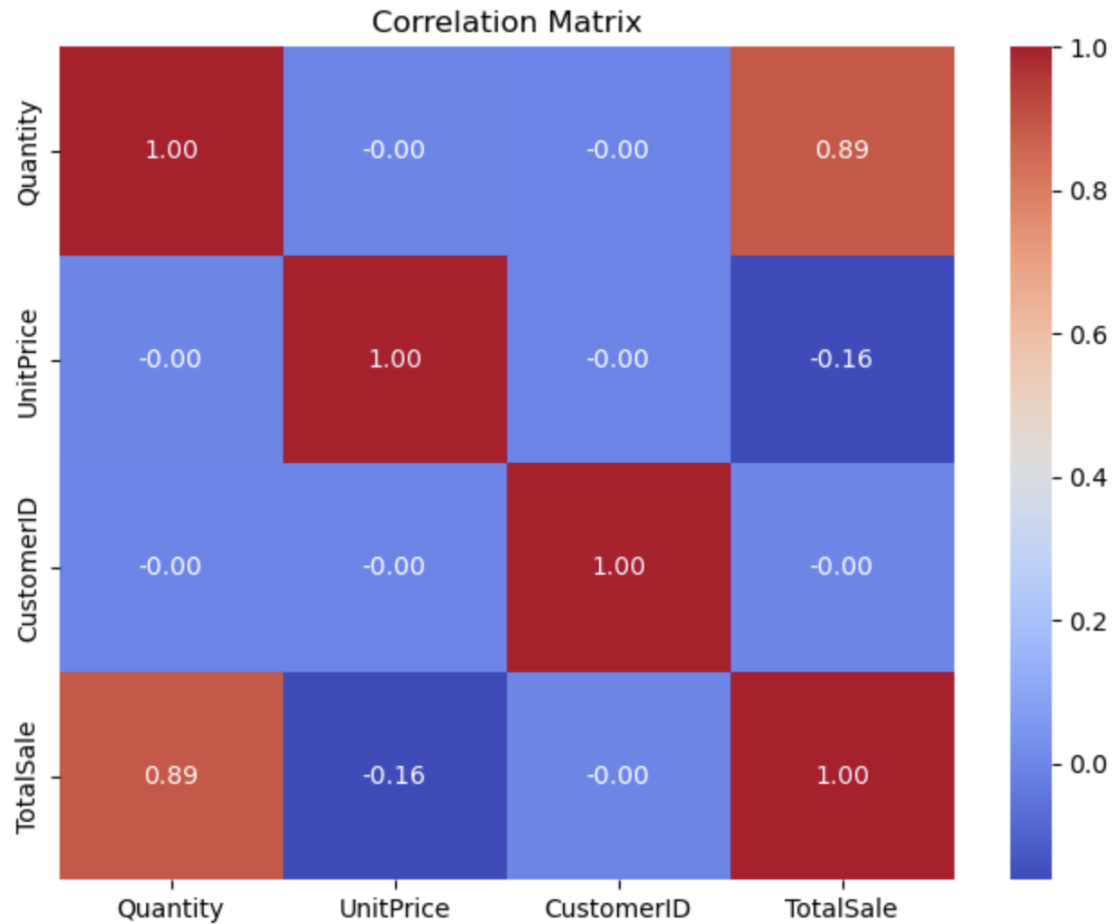Monthly Sales Over Time

High sale
between October to
December

- Potential increase due to holiday gift buying season

Drop in
gift purchasing habits
for consumer
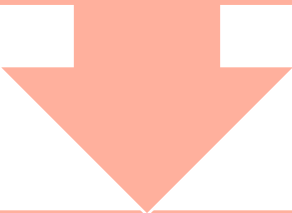after December
month

Correlation Matrix

Correlation Matrix

TotalSale & Quantity show high correlation at 0.89

The rest show no correlation at –0.00.

UnitPrice & TotalSale and TotalSale & UnitPrice show low correlation at –0.16

Models

Association Rules Algorithm

Algorithm examines links/relationships between items.

Approach involved looking at two aspects:

Top 10 Association Rules	Top 10 Frequent Item Sets

Metrics Overview

Antecedents & Consequents

Item Name

Support

Proportion of Transaction Consisting of Item(s)/Item Set

Confidence

The Conditional Probability of the Consequent Item's Inclusion if the Antecedent Item is Included.

Lift

The Likelihood of the Consequent Being Purchased When the Antecedent Is Included Rather Than Not Included.

Top 10 Association Rules

	Antecedents	Consequents	Antecedent Support	Consequent Support	Support	Confidence	Lift
1	Regency Tea Plate Roses	Regency Tea Plate Green	0.10	0.10	0.10	0.70	57.90
2	Regency Tea Plate Green	Regency Tea Plate Roses	0.10	0.10	0.10	0.85	57.90
3	Set/6 Red Spotty Paper Cups	Set/6 Red Spotty Paper Plates	0.10	0.10	0.10	0.83	56.94
4	Set/6 Red Spotty Paper Plates	Set/6 Red Spotty Paper Cups	0.10	0.10	0.10	0.73	56.94
5	Poppy's Playhouse Kitchen	Poppy's Playhouse Bedroom	0.20	0.10	0.10	0.73	51.21
6	Poppy's Playhouse Bedroom	Poppy's Playhouse Kitchen	0.10	0.20	0.10	0.80	51.21
7	Small Marshmallows Pink Bowl	Small Dolly Mix Design Orange Bowl	0.10	0.20	0.10	0.78	48.00
8	Small Dolly Mix Design Orange Bowl	Small Marshmallows Pink Bowl	0.20	0.10	0.10	0.63	48.00
9	Blue Happy Birthday Bunting	Pink Happy Birthday Bunting	0.20	0.20	0.10	0.71	45.01
10	Pink Happy Birthday Bunting	Blue Happy Birthday Bunting	0.20	0.20	0.10	0.71	45.01

Highlights

- Most Antecedent/Consequent pairings appear with the same occurrence.

- In some occasions, the Antecedent or Consequent has a higher occurrence frequency than the other.

- Each pairing of Antecedent/Consequent represents an equal proportion of transactions in the data.

- High Confidence numbers show that the Consequent item is likely to be included if the Antecedent is.

- All Lift scores are extremely high, which means that the likelihood is very high for the Consequent item being purchased when the Antecedent is.

Top 10 Frequent Item Sets*

	Proportion of Transactions of Item Set	Item Set	General Category
1	2.47%	Jumbo Bag Red Retrospot Jumbo Bag Pink Polkadot	Jumbo Bag
2	2.41%	Green Regency Teacup and Saucer Roses Regency Teacup and Saucer	Teacup and Saucer
3	2.38%	Alarm Clock Bakelike Red Alarm Clock Bakelike Green	Alarm Clock
4	2.35%	Lunch Bag Pink Polkadot Lunch Bag Red Retrospot	Lunch Bag
5	2.34%	Lunch Bag Black Skull Lunch Bag Red Retrospot	Lunch Bag
6	2.12%	Wooden Picture Frame White Finish Wooden Frame Antique White	Wooden Frame
7	2.11%	Lunch Bag Red Retrospot Lunch Bag Spaceboy Design	Lunch Bag
8	2.10%	Lunch Bag Pink Holdout Lunch Bag Black Skull	Lunch Bag
9	2.09%	Gardeners Kneeling Pad Keep Calm Gardeners Kneeling Pad Cup of Tea	Gardeners Kneeling Pad
10	2.07%	Lunch Bag Cars Blue Lunch Bag Red Retrospot	Lunch Bag

Highlights

- Items within an Item Set are **similar**.

- Each Item Set can be generally categorized.

- The Proportion of Transactions of Item Set (Support) **equals 22.44%** for the top 10 frequent item sets.

Item Sets are in at least bundles of 2, formed based on relative support.

K-Means Clustering

Unsupervised model that groups data into a predefined number of clusters based on similarity.

Goal is to partition data so that variance within each cluster is minimized.

Centroids: Represents the mean position of all the data points in a cluster. Focal point around which clusters are formed and then optimized.

Allows for great interpretability with many different types of datasets.

Enables us to make marketing insights by allowing us to break customer purchase behaviors into groups.

Results

	1	2	3
Total Amount	907.065	648.643	3592.906
Purchase Frequency	2.474	1.654	7.030
Average Basket Size	236.478	263.585	262.479
Recency of Purchase	36.758	262.699	56.929
Winter	0.076	0.46	0.208
Spring	0.064	0.523	0.18
Summer	0.025	0.016	0.409
Fall	0.835	0.002	0.203

- Modelling led to forming 3 clusters.

- Results show average for each feature per cluster, which can help define customer segments.

Marketing Insights

Household Items

Everyday Use Items

Common Themes of Frequent Items and Item Sets

High Frequency Products

Jumbo Bag
Regency Teacup and Saucer
Alarm Clock
Wooden Frame
Gardeners Kneeling Pad
Lunch Bag

High Frequency Products Make Up Around **25% of** Transactions

Association Pairing Products

Regency Tea Plate
Paper Servingware
Playhouse Rooms
Bowl
Happy Birthday Bunting

Association Pairing Products Make Up **10% of** Transactions

Item Purchasers Profiles

Products
(i.e., Jumbo Bag,
Paper Servingware,
Lunch Bag, etc.)

Household
Items

Everyday
Use Items

Types of Customers
- Families
- Parents with Children
- Students
- Employees
- People with Homes

Example of Customer Behavior/Needs/Wants
- Organizing Birthday Parties
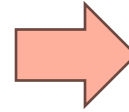- Carry Home Lunch/Food for School/Work
- Purchase Playthings for Children

Common Customer Characteristics
- Fans of Regency Era
- Tea Drinkers
- Preference for Red and Pink

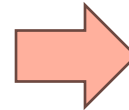# Recommendations for Products

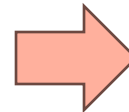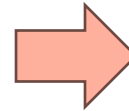| Observations	Recommendations
Antecedents/Consequents Pairings and Frequent Item Sets Consist of Similar Items	Offer Choices (i.e., Different Colors) for Each Item
Household and Everyday Items Commonly Purchased	Offer Items That Have Practical Use
Items Purchased Together Are From Similar Color Families	Strengthen Inventory to Consist of Items From the Same or Similar Color Family
Themed (i.e., Regency) Products Purchased	Offer Items That Are In Line with Trends Ex. Popular Regency Shows May Influence Purchasing Regency Themed Products

Lavish, but Infrequent Spenders

Characteristics & Interpretation:
This cluster has a high spending amount with low purchase frequency and high recency of purchase. With a significantly high average basket size and high spending in winter, this cluster most likely buys gifts in bulk in winter for holidays like Christmas.

Strategic Insights & Recommendations:

o   High recency indicates that recent purchases haven't been made, warranting a customer re-engagement strategies that target high-value customers to return before and during their most active season (winter)

o   Discount bulk purchases: incentivize customers to maintain their habit of bulk purchases and potentially increase the number of items per order

o   Seasonal Promotions: Target the winter season to capitalize on their seasonal buying purchases, offer discounts, promotional codes, or coupons

Moderate Purchasers

Strategic Insights & Recommendations:

o Customer Loyalty Program: As this cluster has moderate spending, frequency and basket size with inconsistencies in purchase recency, a loyalty program that offers additional discounts or deals would incentivize and reward repeated purchases.

o Seasonal Promotions: To target the increase in spending in Spring offer promotional discounts and online showcases of Spring-collection gifts

Characteristics & Interpretation:
The total amount spent, frequency of purchase, and average basket size is moderate for this cluster. This cluster has a more balanced seasonal spending with a slight increase in Spring.

Frequent, but Economical Speakers

Strategic Insights & Recommendations:

- o Subscription Services: Target this cluster's consistent purchasing habits by offering a subscription plan that offers percent discounts for every purchase to incentivize and maintain steady transactions.

- o Promotional Material: To maintain repeated purchase, create promotional material online to promote new items and notify customers of additional items they could expect to see upon their next visit.

Characteristics & Interpretation:
Has lower total amount of spending than cluster 1 but higher than Cluster 2. High purchase frequency indicates a regular purchasing pattern, and the lower average basket size indicates smaller transaction sizes per purchase. They have a relatively distributed seasonal pattern with slightly less purchases in Winter and Summer.

Thank You!