

Analysis on e-Commerce Data

Presented by:

Lisa Patel

Binny Chauhan

Vishakha Srivastava

Hithesh Sekhar Bathala

E-Commerce Data

The dataset contains the information about customer's purchases of an UK online retailer 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.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850	United Kingdom

Attributes

1. Invoice number: 6-digit unique number assigned to each transaction
2. Stockcode: Product (item) code
3. Description: Product (item) name
4. Quantity: The quantities of each product (item) per transaction
5. Unit price: Product price per unit in sterling
6. Invoice date: The day and time when each transaction was generated
7. Country: The name of the country where each customer resides
8. Customer_id: 5-digit unique number assigned to each Customers

EDA and Data Cleaning

Insights

Exploratory Data Analysis

1. 89% of the orders are from United Kingdom.
2. 16.47% of the orders are cancelled orders.
3. The top 5 countries to receive the highest number of orders are: United Kingdom, Germany, France, Ireland, Spain
4. 79.01% of revenue is generated by 20% of the products.
5. Most products are sold on Thursday in contrast Saturday and Sunday which have very low transactions.
6. Pre-christmas season starts in September and peaks in November.
7. 16.76% of the product descriptions are invalid.
8. 1.09% of the total products have unit price as zero.
9. For the month of December 2011, we have only 2 weeks of data (Weeks 50,51,52 are missing).
10. Maximum number of rewards are given in the month of November which is also the month of maximum number of sales.

Data Cleaning

- The data had 3.1% values as NA
- Invoice starting with 'C' are the orders that are cancelled. 16.47% of all orders are cancelled orders.
- There were some quantity outliers is due to cancelled orders, which led to negative quantities.
- Products with unit price zero
- Stockcodes with length less than 5 & greater than 6 were invalid, Ex: Post, Cruik
- Most of the descriptions were in uppercase. Description with lowercase characters had some strange descriptions, Ex: Bank charges, High resolution image

Trend Analysis for UK based customers:

Weekly Trends:

- Thursday is the day with maximum sales
- In contrast Saturday and Sunday have very low sales



Monthly Trends:

- The pre-Christmas season starts in September and shows a peak in November.
- February and April are months with very low sales



Trend Analysis for Non UK based customers:

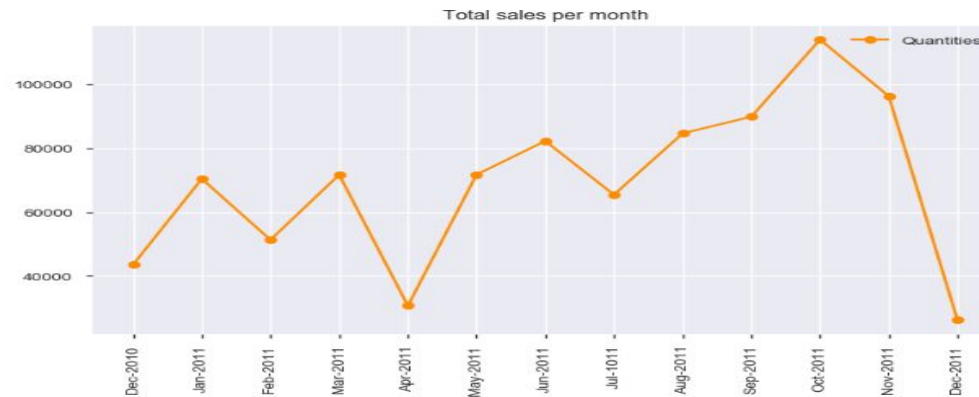
Weekly Trends:

- Similar to that of UK based customers with Thursday being the day with maximum sales

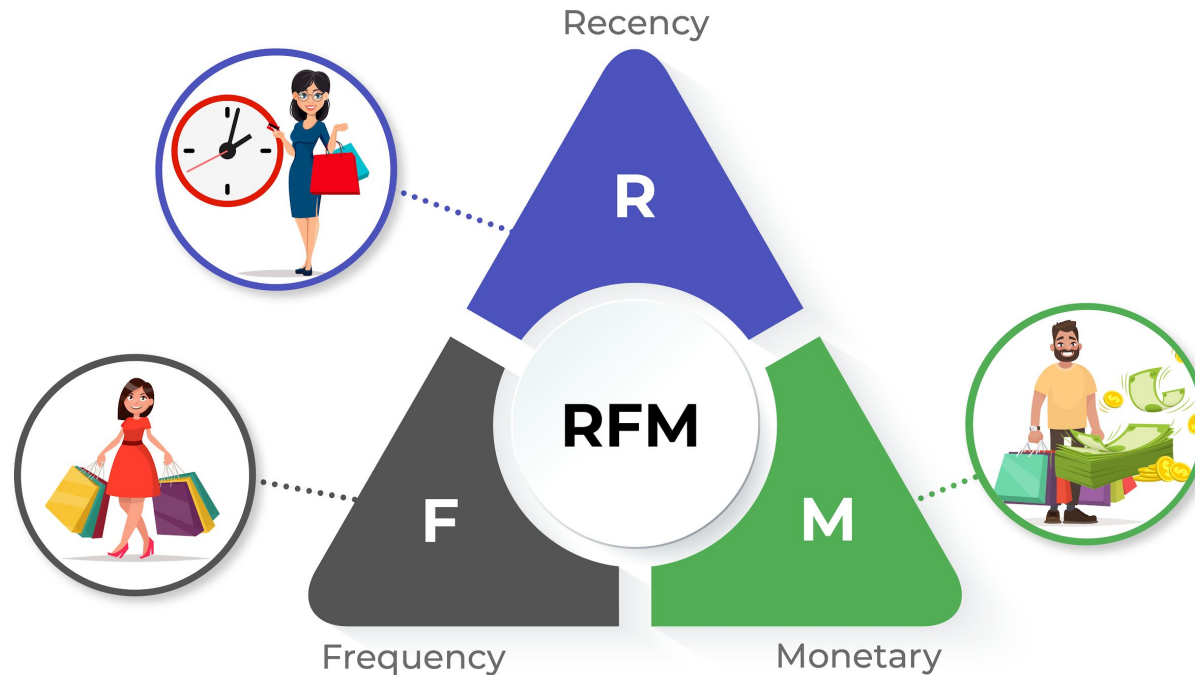


Monthly Trends:

- The pre-Christmas season starts in September and shows a peak in October.
- February and April are month with very low sales



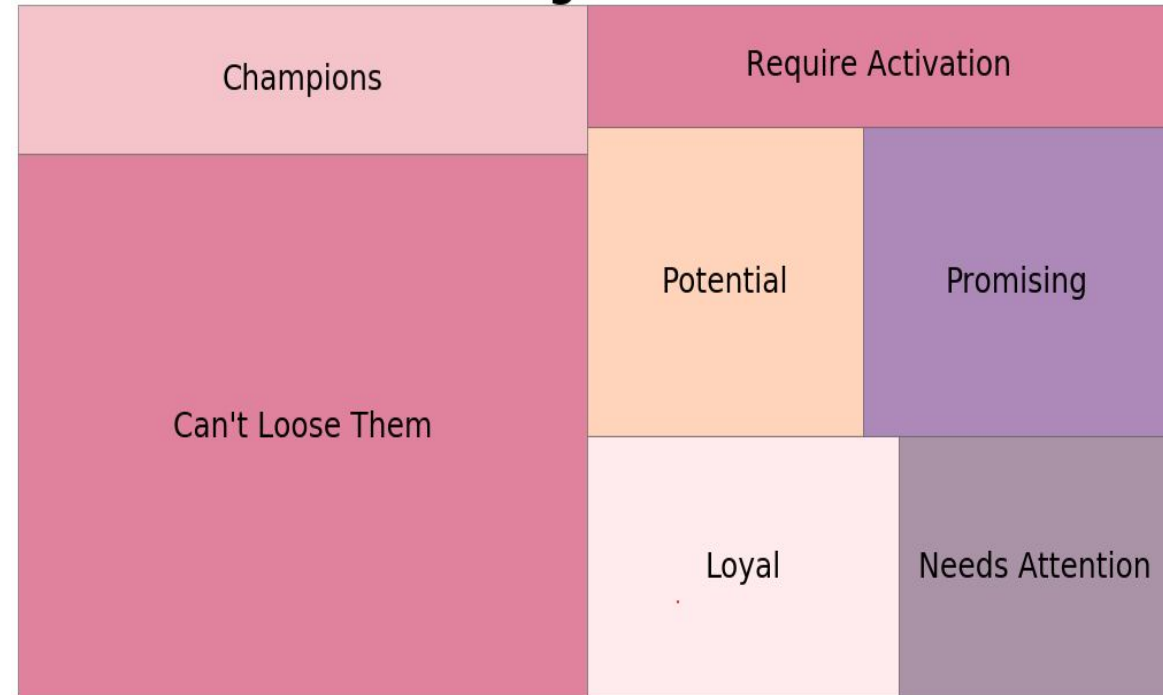
RFM Analysis



- **Recency:** How recently the purchase was made?
- **Frequency:** How frequent is the customer?
- **Monetary:** How much money is spent?

RFM Segmentation

- 60% of the customers are in the top tier RFM levels (**Can't lose them, Champions, Loyal**)
- **Potential** - the store should distribute freebies to get more attention from them!
- **Promising** — it has been a while since they last bought something from you.
- **Needs Attention** — made some initial purchase but have not seen them since. They could have had a bad customer experience or product-market fit
- **Require Activation** — Poorest performers of our RFM model. They might have went with the competitors for now



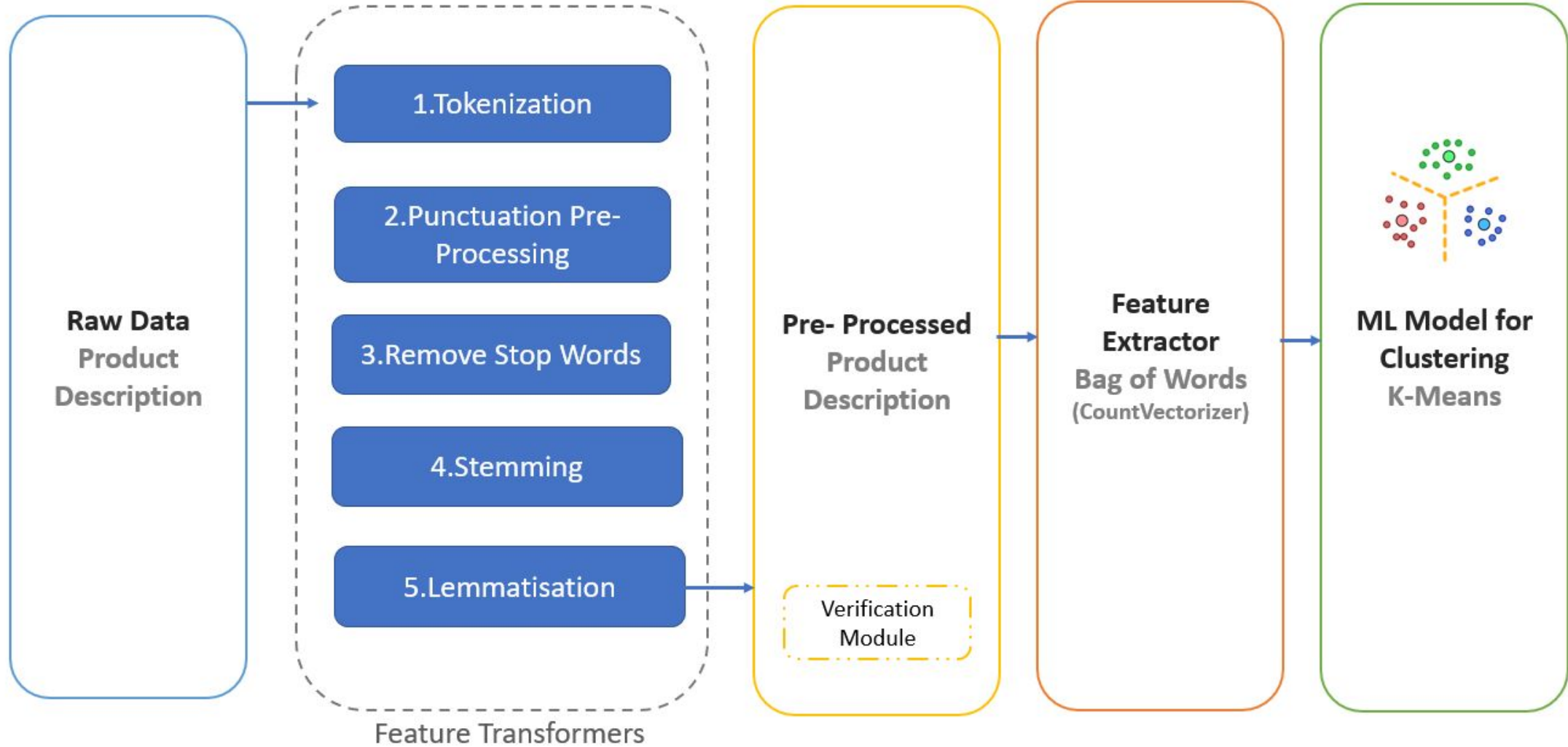


Product Marketing suggestion
based on purchase behaviour



Tool for Sellers/ Managers to club
products to make basket sale

NLP Pipeline



EDA and Data Cleaning

Insight 3

Sellers could use the below module to find the right products for grouping to boost sales

Product

Submit

Project Sale Lift by combining with :Small chocolates pink bowl

