

Customer segmentation

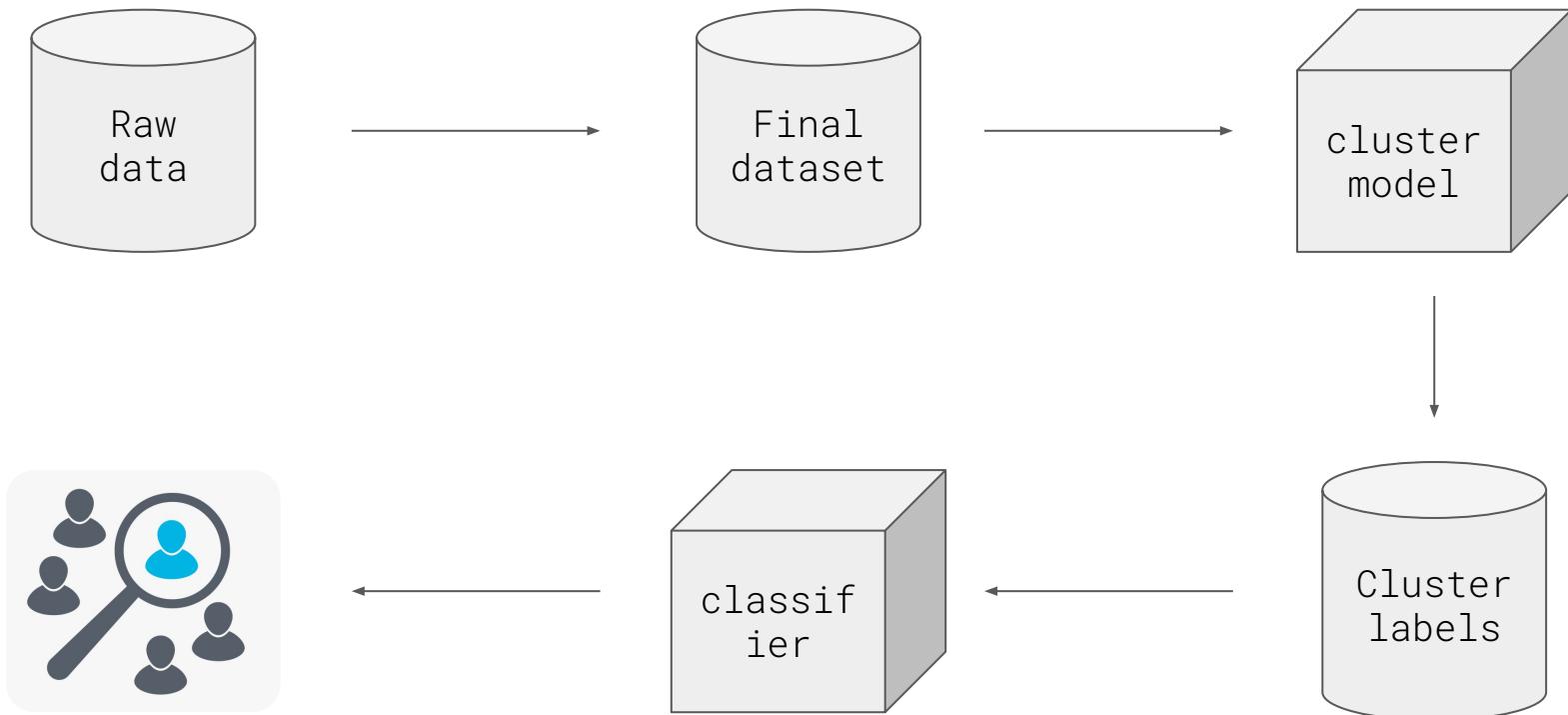
e-commerce transactions dataset



Goals

- Find group of similar customers
- Understand the characteristics of each group
- Make informed decisions based on extracted data (e.g. custom emails to specific customers)

Strategy



The dataset

	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
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047	United Kingdom

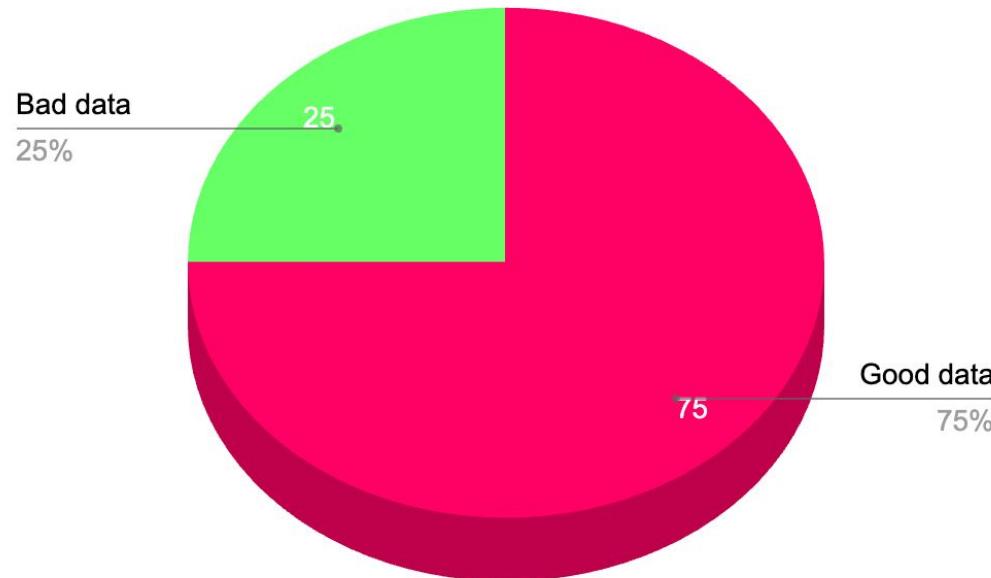
Modelling customers

- **Customer**
 - ID
 - Country
 - <Transaction>
 - Invoice Number
 - Invoice Date
 - <Articles>
 - Description
 - Unit price
 - Quantity bought



Cleaning the dataset

- ~ 400,000 final examples
- ~ 4,000 unique customers



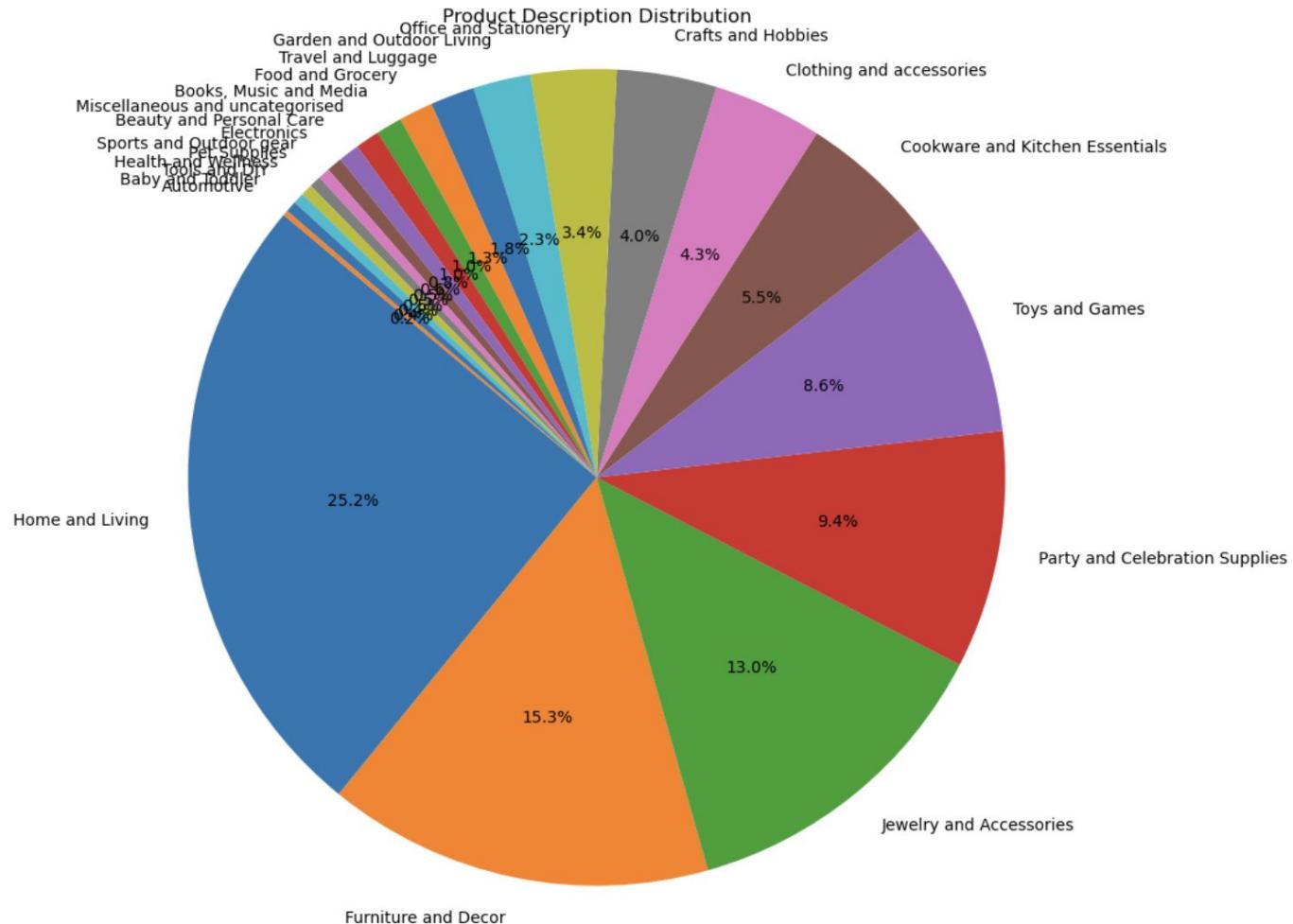
NLP

```
shopping_categories = [  
    "Clothing and accessories",  
    "Electronics",  
    "Home and Living",  
    "Cookware and Kitchen Essentials",  
    "Beauty and Personal Care",  
    "Books, Music and Media",  
    "Sports and Outdoor gear",  
    "Toys and Games",  
    "Automotive",  
    "Health and Wellness",  
    "Food and Grocery",  
    "Office and Stationery",  
    "Pet Supplies",  
    "Jewelry and Accessories",  
    "Furniture and Decor",  
    "Tools and DIY",  
    "Travel and Luggage",  
    "Baby and Toddler",  
    "Crafts and Hobbies",  
    "Party and Celebration Supplies",  
    "Garden and Outdoor Living",  
    "Miscellaneous and uncategorised",  
]
```

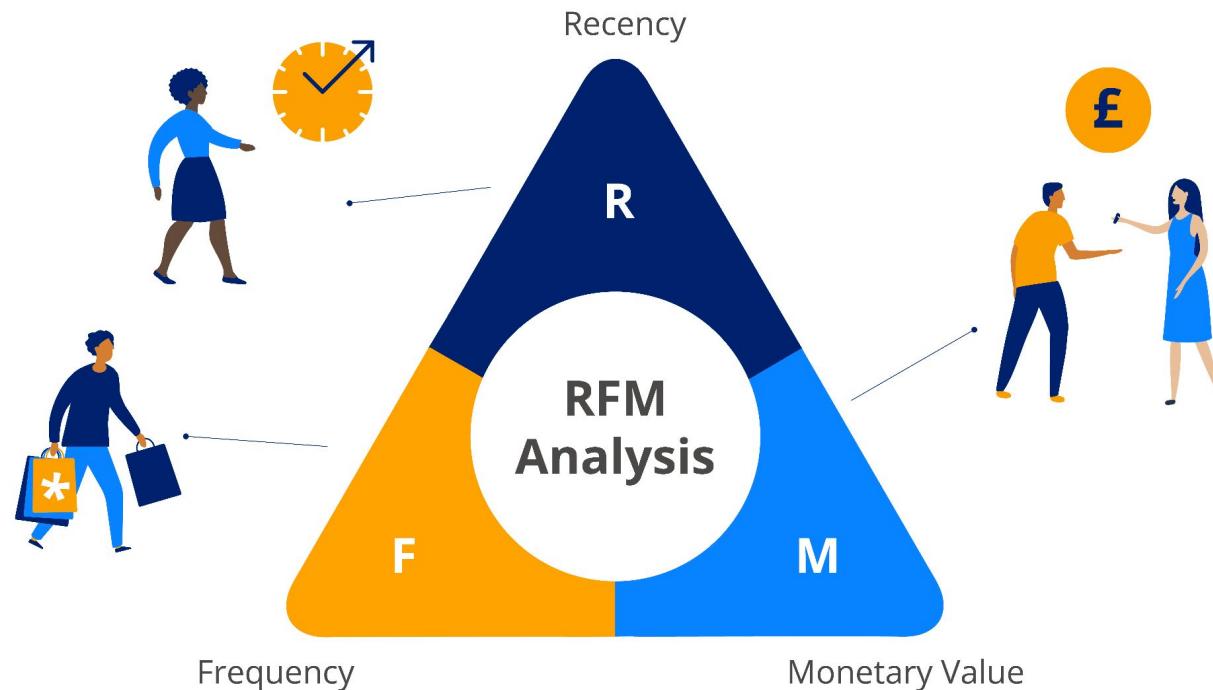
**"BOX OF 6 ASSORTED COLOUR
TEASPOONS"**



Kitchen and Dining

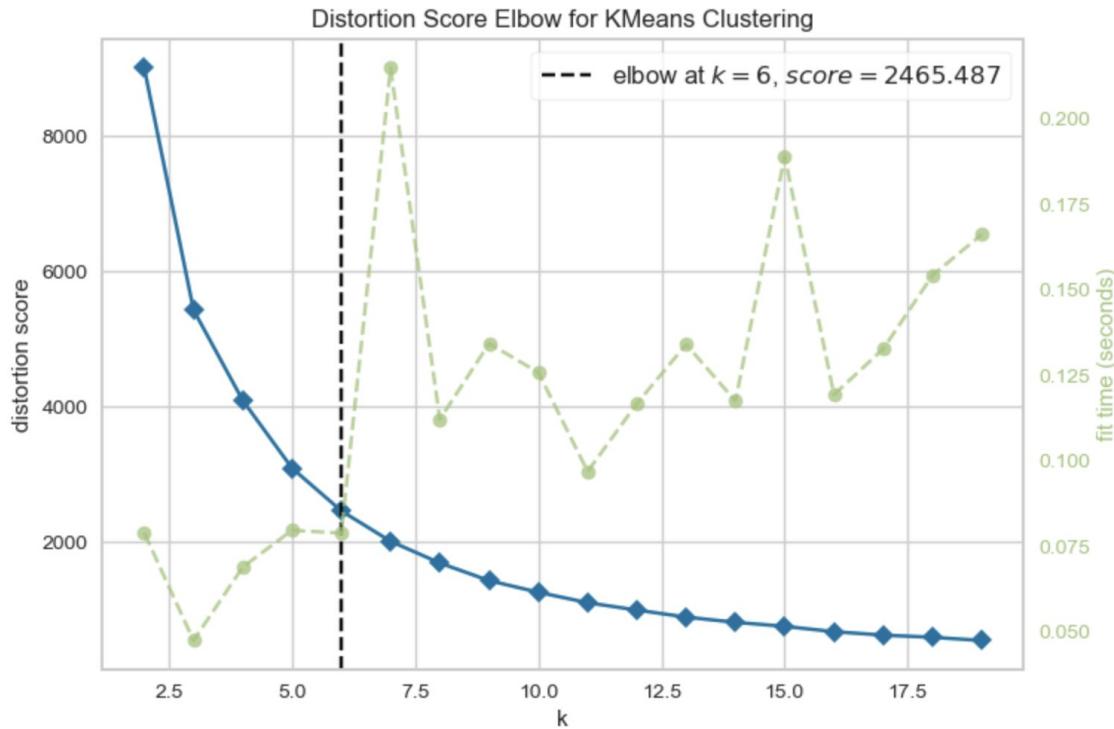


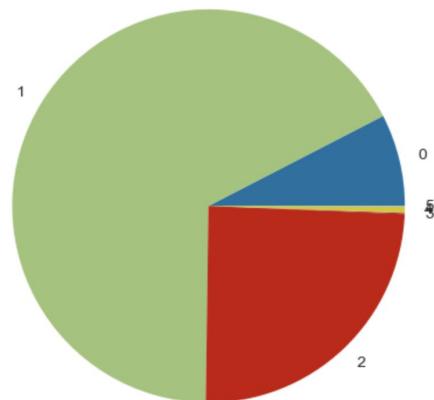
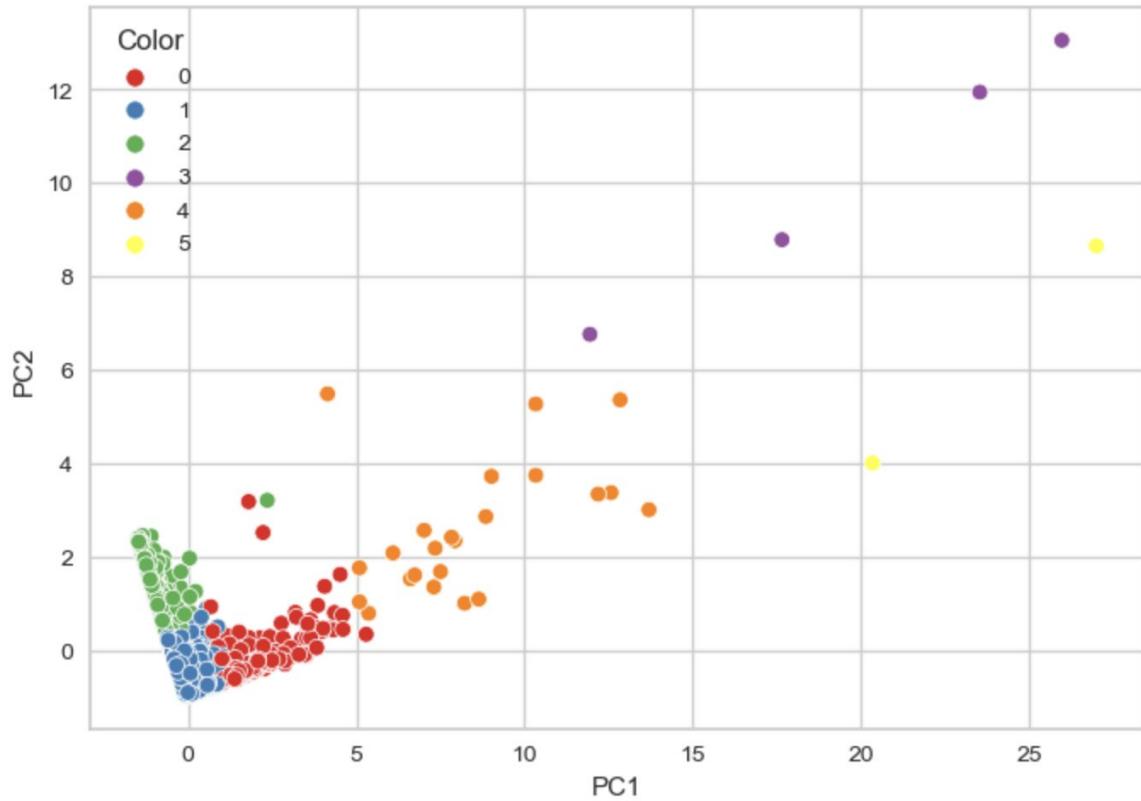
RFM analysis



Pipeline

- Standard scaler
- K-means ($K = 6$)
- PCA (visualization)





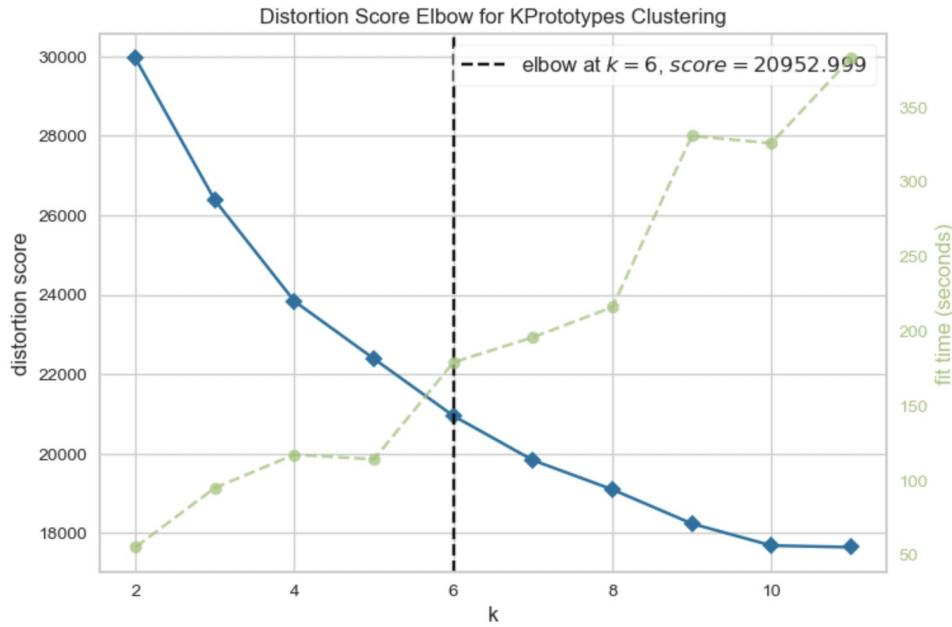


Aggregating user informations

- total expenditure
- product categories (one hot encoded)
- number of transactions
- number of purchases
- mean/std of quantity/unit price/expenditure
- mean/std of the time distance between two transactions

Pipeline

- Standardisation of numerical features
- K-prototypes (K=6)
- t-SNE visualization



K-prototypes = K-means + K-modes

Numerical features

k centroids

find position of centroids that
minimizes the total sum of
distances example-closest centroid

12 norm

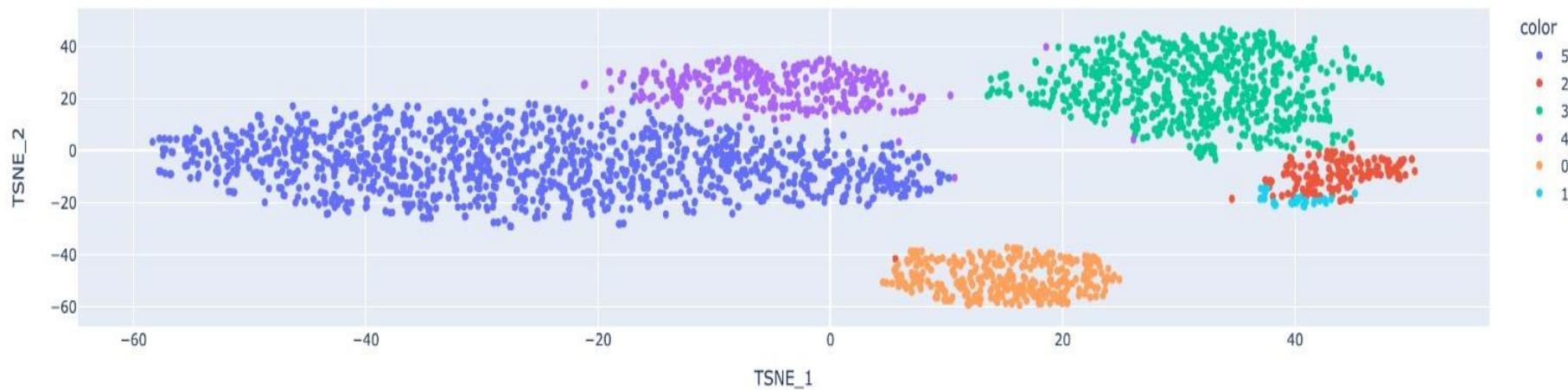
Categorical features

k modes

find modes that minimizes the
total sum of distances
example-closest modes

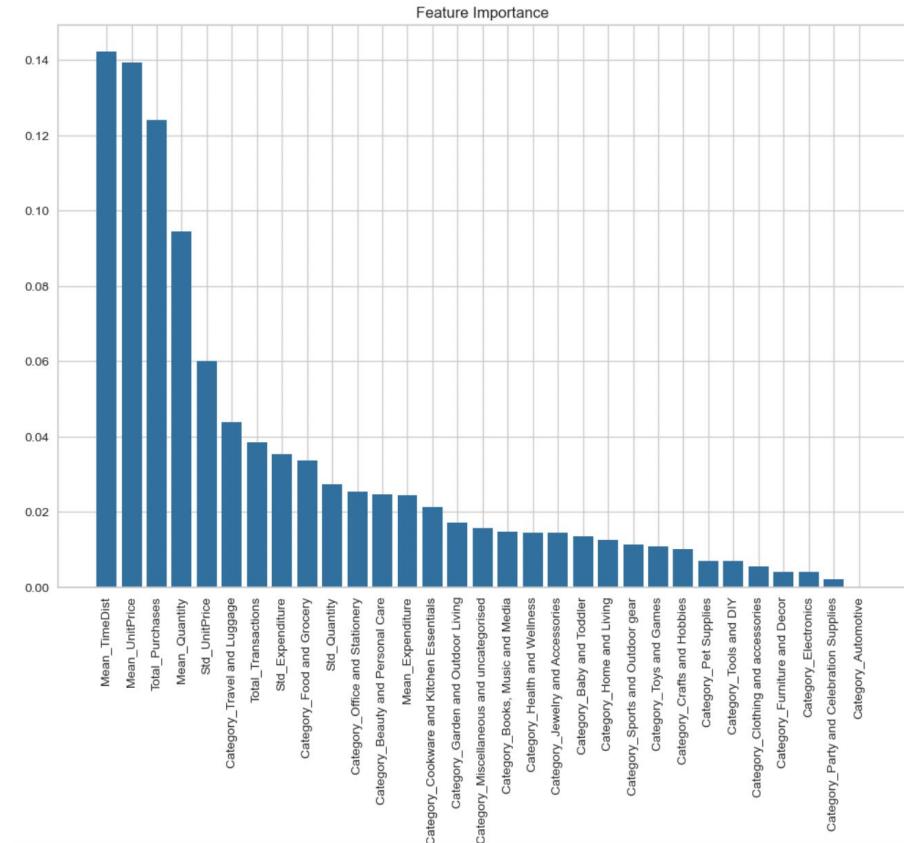
dissimilarity measure

t-SNE + cluster labels



Gradient boosting

- 93% accuracy
- TOP 5:
 - mean time distance
 - mean unit price
 - total purchases
 - mean quantity
 - std unit price
- Shopping categories are less important

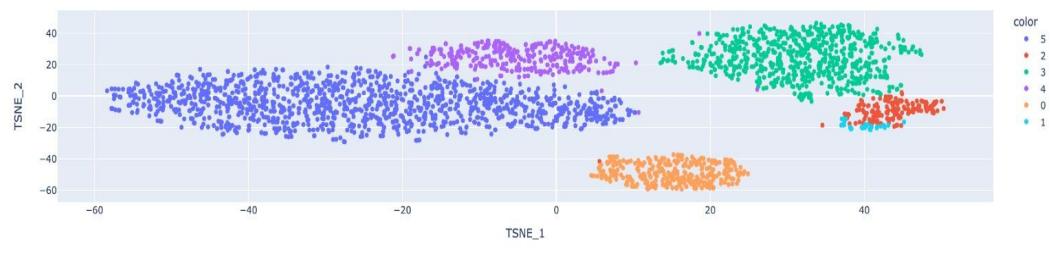


Findings



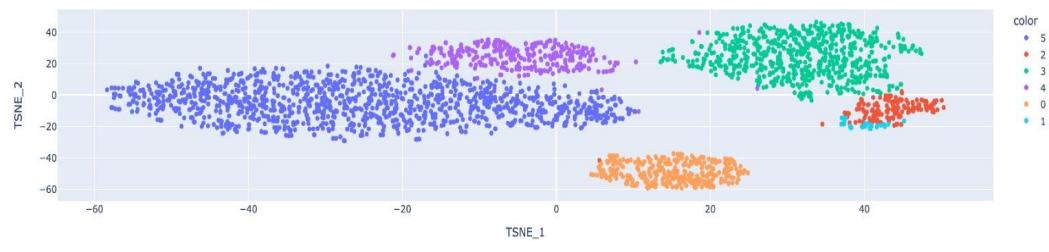
Cluster 0: Diverse Home Products Shoppers

- Customers who frequently shop across various home categories, exhibiting consistent but moderate spending.



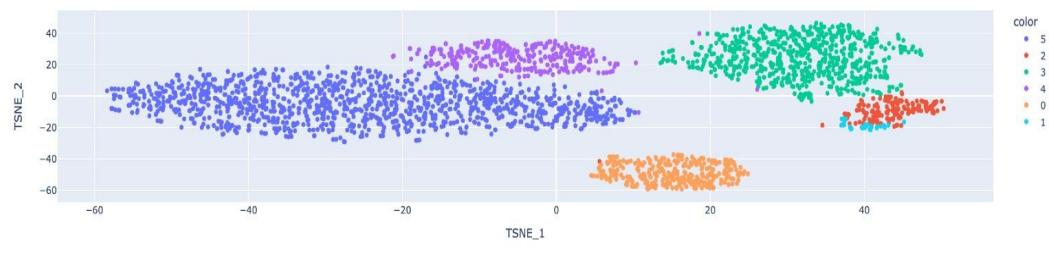
Cluster 1: Premium Party Patrons

- High-value customers with significant spending on party supplies, home decor, and premium products.



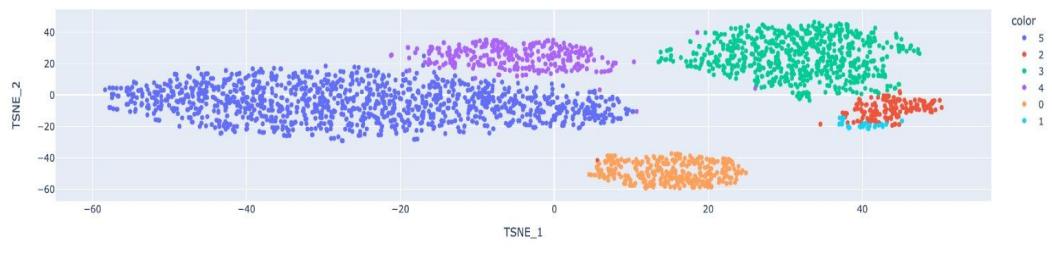
Cluster 2: Frequent Home Decor Enthusiasts

- Regular buyers who frequently purchase home and celebration supplies, with moderate to high spending on decor and essentials.



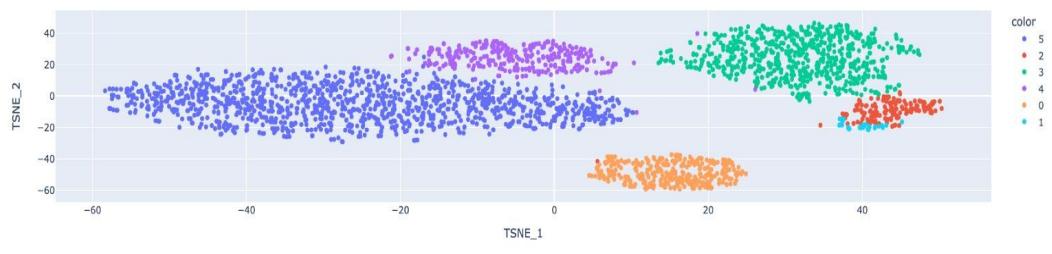
Cluster 3: Budget-Conscious Home Essentials Buyers

- Customers with moderate spending habits focused on essential home items at a reasonable price.



Cluster 4: Infrequent Furniture Seekers

- Customers who occasionally buy furniture and related items with moderate spending but infrequent transactions.



Cluster 5: Occasional Home Shoppers

- Customers who make moderate purchases across various home-related categories occasionally.

