Target Selection Campaign & Customer Profiling for e-com business

UNSUPERVISED LEARNING

- TEAM 5
- PML JANUARY 2022

Anisha, Anwesha, Arkadip, Darakshan, Naveen, Parth, Yatish

12 April 2022

Background story

An E-Commerce company over the years have collected its customer data. To accelerate the business growth, company wants to study customers behaviour and plan its business accordingly

Our team will profile its customer base, based on their buying pattern, spendings, volume. This would enable the company to have a better understanding about the users and top SKU(s). Which in turn will help in better strategy planning like campaign management for targeted customers, increase in sales or strategies to retain loyal customers and attract new customers.

To identify the persons, who does frequent / infrequent shopping. Understand their buying pattern, choices. Retain using promotional offers. Gain new leads for the business.

DATA DICTIONARY

Source

https://www.kaggle.com/code/fabiendaniel/customer-segmentation/data

InvoiceNo - Unique transaction id of a customer's purchase

StockCode - Unique identifier of a product / SKU

Description - Description of SKU or Product

Quantity - Number of products purchased in an order

InvoiceDate - Date of transaction

UnitPrice - Per unit price of the product

CustomerID - Unique identifier for a customer/ entity

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
581483	23843	PAPER CRAFT , LITTLE BIRDIE	80995	12/09/11 9:15	2.08	16446	United Kingdom
541431	23166	MEDIUM CERAMIC TOP STORAGE JAR	74215	1/18/2011 10:01	1.04	12346	United Kingdom
578841	84826	ASSTD DESIGN 3D PAPER STICKERS	12540	11/25/2011 15:57	0	13256	United Kingdom
573008	84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	4800	10/27/2011 12:26	0.21	12901	United Kingdom
554868	22197	SMALL POPCORN HOLDER	4300	5/27/2011 10:52	0.72	13135	United Kingdom
556231	85123A	RED HARMONICA IN BOX	4000	06/09/11 15:04	0		United Kingdom
544612	22053	EMPIRE DESIGN ROSETTE	3906	2/22/2011 10:43	0.82	18087	United Kingdom
560599	18007	ESSENTIAL BALM 3.5g TIN IN ENVELOPE	3186	7/19/2011 17:04	0.06	14609	United Kingdom
540815	21108	FAIRY CAKE FLANNEL ASSORTED COLOUR	3114	01/11/11 12:55	2.1	15749	United Kingdom
536830	84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	2880	12/02/10 17:38	0.18	16754	United Kingdom
542505	79063D		2560	1/28/2011 12:04	3	432	United Kingdom
543669	22693	GROW A FLYTRAP OR SUNFLOWER IN TIN	2400	02/11/11 11:22	0.94	16029	United Kingdom
544152	18007	ESSENTIAL BALM 3.5g TIN IN ENVELOPE	2400	2/16/2011 12:10	0.06	14609	United Kingdom

DATA CLEANING

- Removed null Customer ids
- Removed transactions that had 0 Unit price of a product
- Removed transactions where quantity of products purchased was less than 0, they were the returned items/transactions. Also, the transactions that had quantity of purchase more than 5000 units were dropped, they were the outliers.
- Removed Stock Codes that contained only alphabets. These had invalid product description like Manual, DOTCOM Postage, Discount etc.

FEATURE ENGINEERING

- Total Quantity – Total number of products purchased by a customer=Sum of Quantity group by customers
- Total Price – Total amount of purchase(£) by a customer = Quantity*UnitPrice
- Recency – How recently the customer has made a transaction, measured in days=(Max transaction date of data-Last transaction date of each customer)
- Frequency – How frequently a customer has purchased in the time frame=Count of unique transactions per customer
- Average Transaction Value – Average value per transaction of a customer (£) = Total Price/Frequency
- Average Transaction (Unit) - Average number of products purchased by customer
- Time in between orders (days) - Time between first and last purchase of customer in the time frame, measured in days=(Min. Transaction date of customer-Max. Transaction date of customer)
- Average Time between orders (days) - Average time of order placed by customer, measured in days=Time in between orders (days)/Frequency

Outlier Treatment – Outliers were capped at 10th percentile on lower side and 99th percentile on higher side

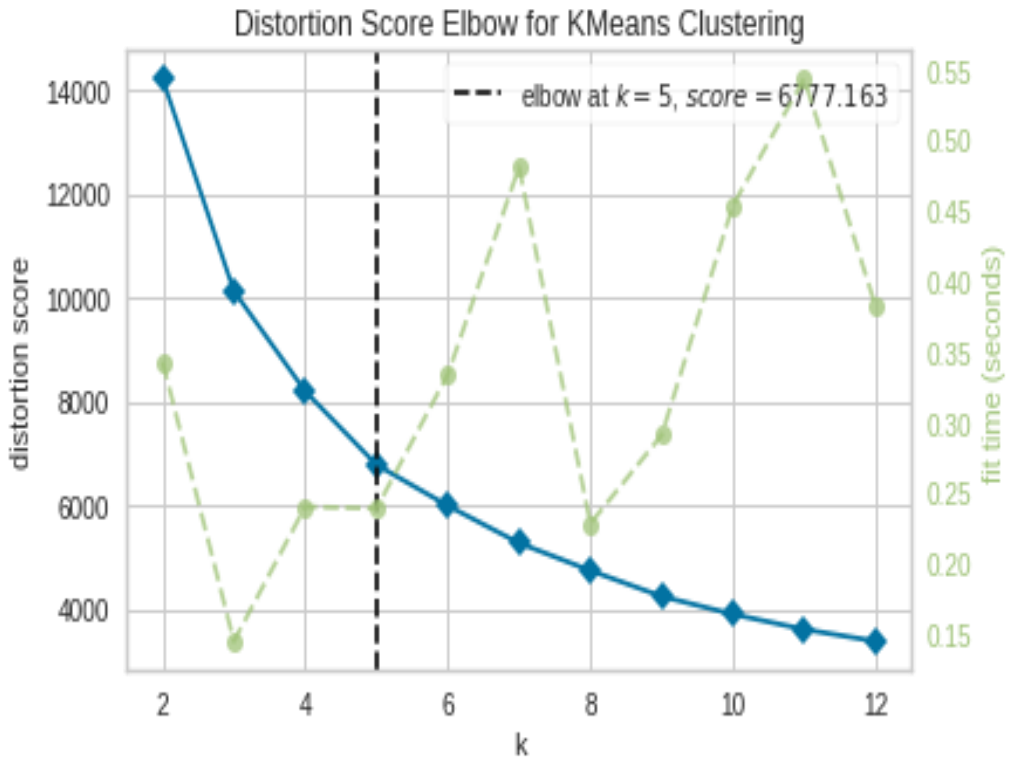
Number of customers for final analysis - 3915

Note: Features marked in green were used for clustering algorithms.

K M E A N S

- Scaled data using Standard Scalar
- Used “elbow” method to select the optimal number of clusters
- Used silhouette scores to calculate the goodness of clustering
- **Number of Clusters: 5, Silhouette Score: 0.2897**
- Also tried Mini-Batch K-means which had a similar silhouette Score of 0.2486

Label	Count
0	351
1	928
2	36
3	581
4	2019



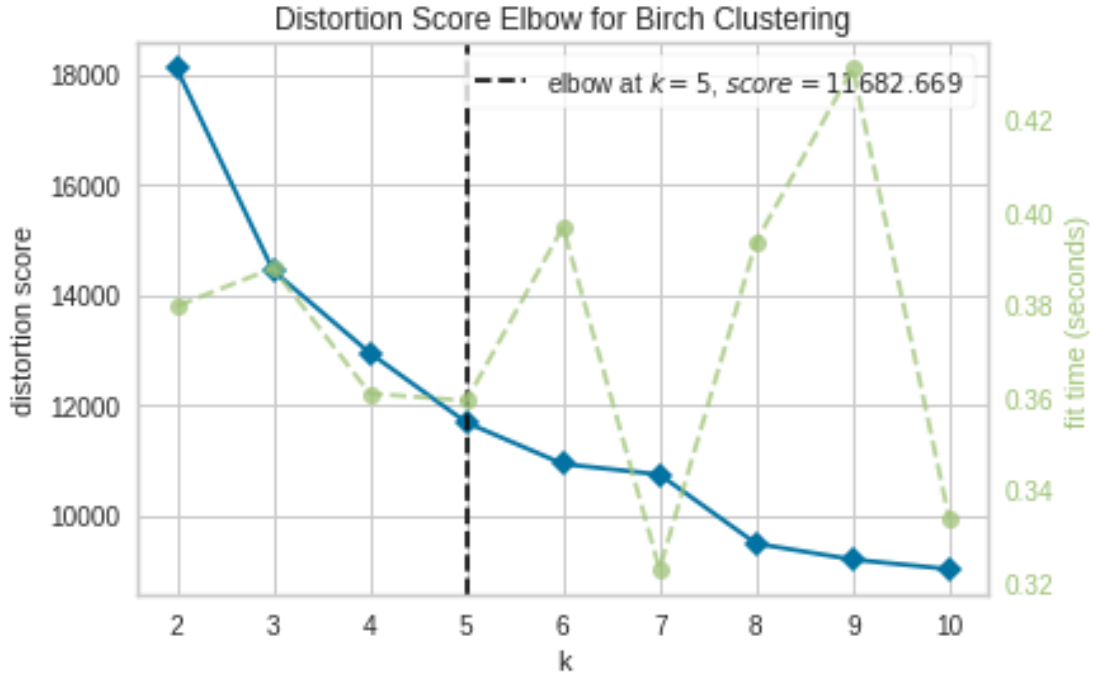
BIRCH

(Balanced Iterative Reducing and Clustering using Hierarchies)

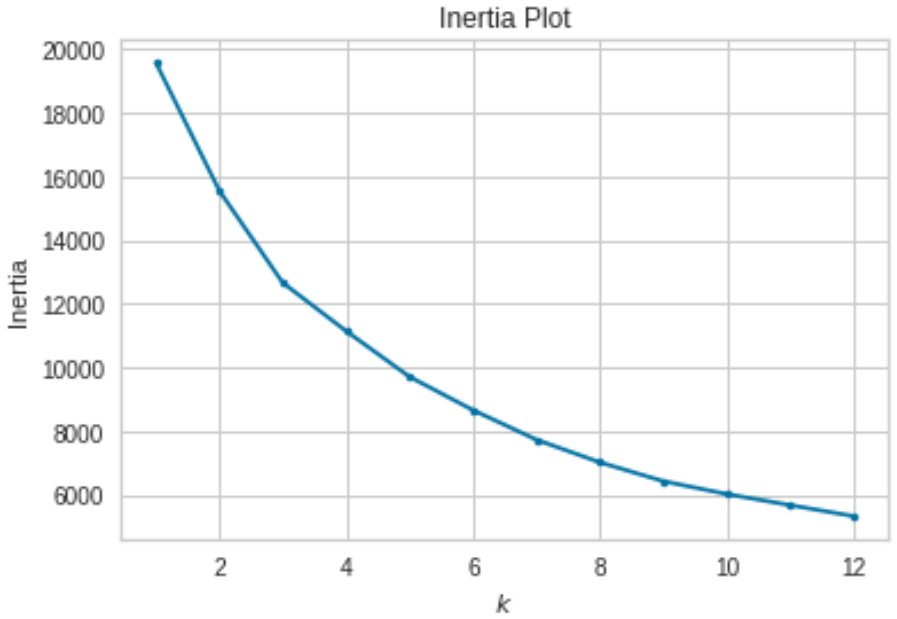
Label	Count
0	354
1	6
2	3359
3	119
4	77

Silhouette Score: 0.4020

Clusters: 5



AGGLOMERATIVE CLUSTERING



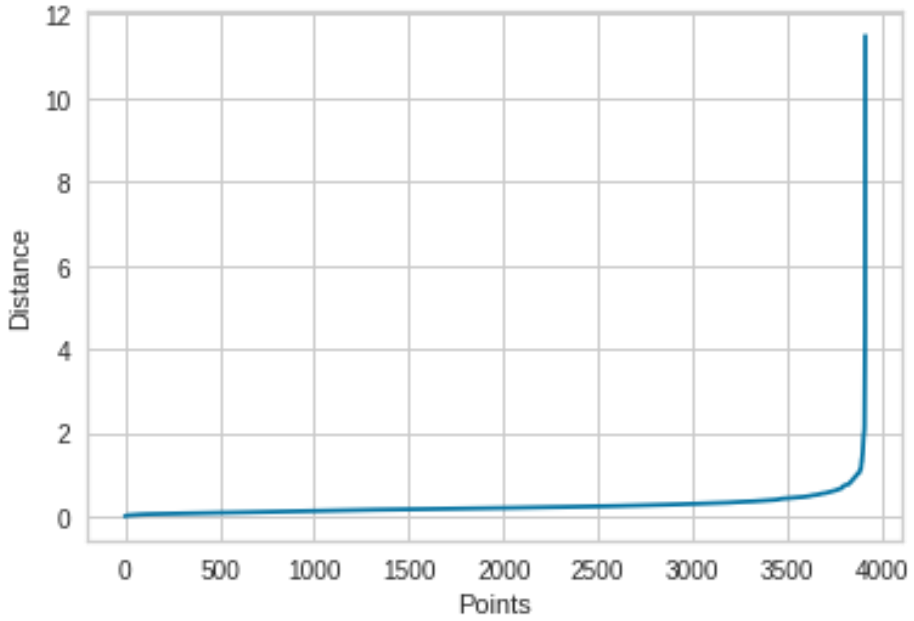
Number of Clusters = 4

Label	Count
0	2516
1	318
2	917
3	164

Silhouette Score=0.26181

DBSCAN

Neighbors=10



Epsilon= 1.1

Min Samples=3

Label	Count
-1	43
0	3851
1	4
2	6
3	8
4	3

Silhouette Score=0.5031

Number of Clusters=5

CONCLUSION

Algorithm	Number of Custers	Silhouette Score
KMEANS	5	0.2897
BIRCH	5	0.4020
AGGLOMERATIVE	4	0.26181
DBSCAN	5	0.5031

PRODUCT DEMAND ANALYSIS

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	InvoiceDate1	Total Price	Cluster label	Purchase Amount
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-01-12 08:26:00	2.55	17850.0	United Kingdom	2010-01-12	15.30	1	15.30
1	536365	71053	WHITE METAL LANTERN	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	2010-01-12	20.34	1	20.34
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-01-12 08:26:00	2.75	17850.0	United Kingdom	2010-01-12	22.00	1	22.00
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	2010-01-12	20.34	1	20.34
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-01-12 08:26:00	3.39	17850.0	United Kingdom	2010-01-12	20.34	1	20.34

	StockCode	Description	Unique Customers	Quantity	Total Revenue	Unique Invoice	Repeated Demand
1163	22502	PICNIC BASKET WICKER SMALL	25	324	41012.35	31	6
2780	85099B	JUMBO BAG RED RETROSPOT	91	18980	33633.72	219	128
1091	22423	REGENCY CAKESTAND 3 TIER	98	1826	20845.65	197	99
2786	85123A	WHITE HANGING HEART T-LIGHT HOLDER	70	7801	19945.61	173	103
253	21137	BLACK RECORD COVER FRAME	14	4860	16517.21	33	19
797	22086	PAPER CHAIN KIT 50'S CHRISTMAS	85	5662	15213.58	138	53
1060	22386	JUMBO BAG PINK POLKADOT	44	8500	15022.35	101	57

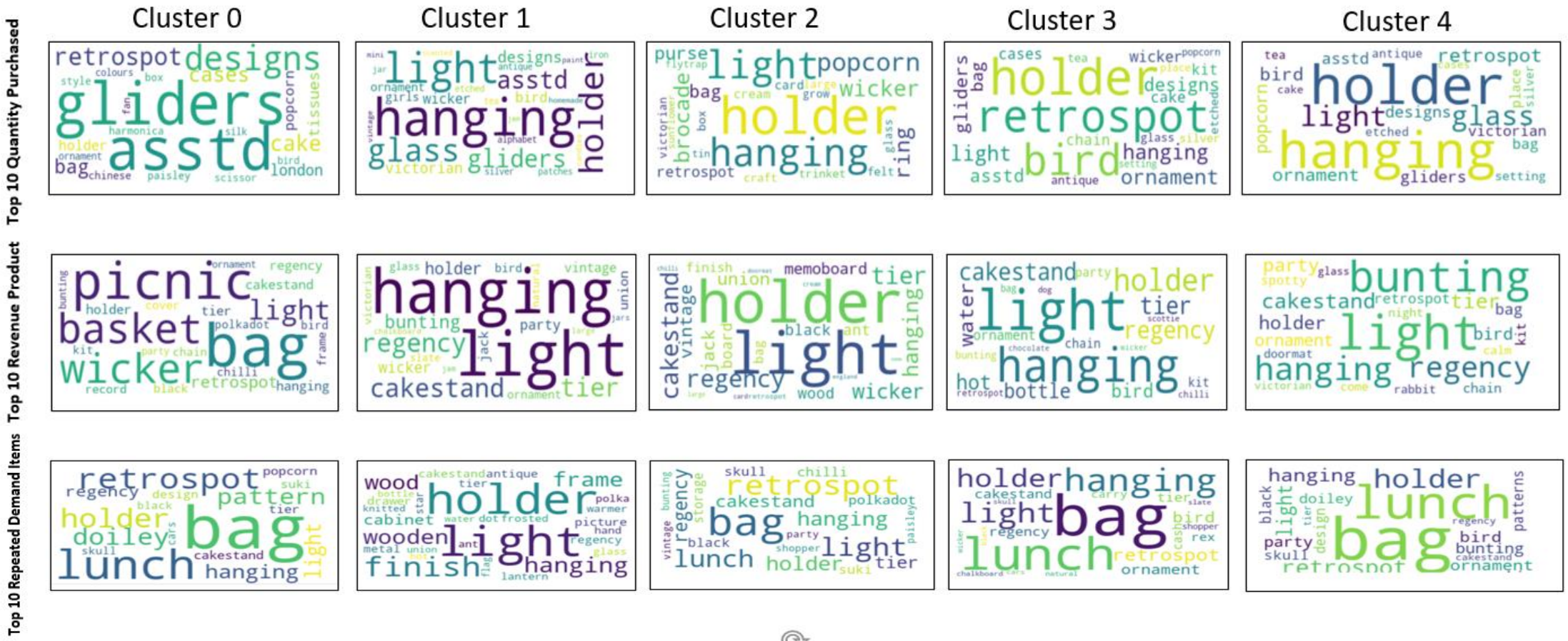
Feature Engineering is done for customer segmented product/demand analysis

- Unique Customers – Customer who have made at-least one purchase against item
- Quantity refers to total quantity demanded against each item
- Total Revenue – Total sales (in Pounds) against each item
- Unique Invoice – Total no. of times item ordered
- Repeated Demand = Unique Customers – Unique Invoice

- Feature Engineering is applied to each of 5 clusters
- TOP 10 Items with high demand "Quantity Wise" in each cluster is analyzed
- Similarly, TOP 10 items with High Revenue in each cluster and Items with High Repeated demand in each cluster is analyzed to understand customers behavior by the product they have purchased

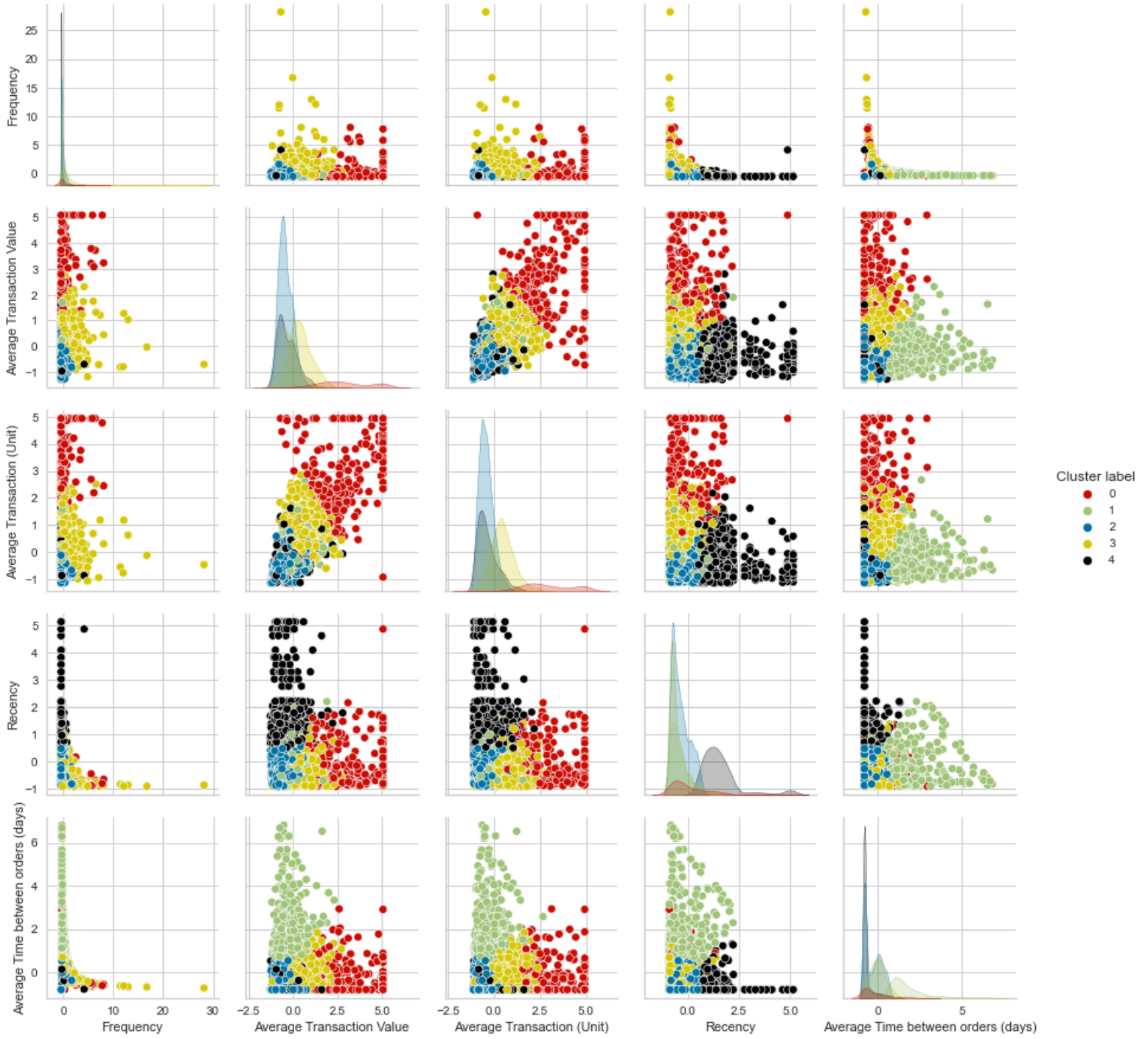
	StockCode	Description	Unique Customers	Quantity	Total Revenue	Unique Invoice	Repeated Demand
2283	84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	60	30631	6827.13	109	49
2780	85099B	JUMBO BAG RED RETROSPOT	91	18980	33633.72	219	128
1269	22616	PACK OF 12 LONDON TISSUES	32	18252	4987.04	72	40
894	22197	SMALL POPCORN HOLDER	53	15101	11272.21	149	96
725	21977	PACK OF 60 PINK PAISLEY CAKE CASES	67	12837	5663.09	121	54
694	21915	RED HARMONICA IN BOX	45	12238	13237.48	86	41
14	16014	SMALL CHINESE STYLE SCISSOR	6	11670	3749.40	11	5
9	15036	ASSORTED COLOURS SILK FAN	28	10041	6943.79	47	19
2341	84879	ASSORTED COLOUR BIRD ORNAMENT	59	9523	14306.35	111	52
298	21212	PACK OF 72 RETROSPOT CAKE CASES	82	8837	3943.65	141	59

TOP 10 High demand "Quantity Wise" for Cluster0



Cluster	No. of customers	Quantity	Revenue	Repeat Demand
Cluster 0	351	Bulk orders for Glider design , Red Bag Retrospot , Tissues , Popcorn Holder	Very high revenue from picnic basket , Red Bag Retrospot, cakestand & light holder	Less
Cluster 1	928	Orders for Gifting items (very less Buyers)	Low	Least
Cluster 2	36	Bulk orders for popcorn holder,purse & light holder	Very high revenue from cakestand , light holder , memo Board , Black board	Moderate for Bag ,light holder & cakestand
Cluster 3	581	Orders for Gifting items (Not much , less Buyers)	Moderate from Party Items	less
Cluster 4	2019	Bulk orders for Gifting items	High revenue cakestand , light holder, party items , hanging lights	High for light holder, lunch box & Bag

LABELLED CLUSTER PAIR-PLOTS



CUSTOMER PROFILING USING CLUSTER CENTROIDS

Cluster	0	1	2	3	4
Count of CustomerID	351	928	36	581	2019
Average Time between orders (days)	29	6	12	118	28
Average Transaction (Quantity)	728	134	382	161	176
Average Transaction Value	1065.71	238.15	749.31	282.71	293.16
Recency	92	265	10	65	49
Frequency	4	1	54	3	5

- Cluster 0: Few-time/monthly customers with high transaction quantity
- Cluster 1: One-time/Non-repeating customers
- Cluster 2: Loyal Bulk Buyers with moderate repeat demand
- Cluster 3: Seasonal/Festive Buyers (mostly buy party/gift items)
- Cluster 4: Average Buyer (largest cluster)

UNDERSTANDING YOUR BUSINESS WITH CUSTOMER SEGMENTATION

Customer segmentation helps

- **Describing types of customers in a common way across go-to market , product.** Enables the sales team to give segmented customer feedback to product leaders
- **Understanding our most and least engaged customers at a granular level.**
- **Enabling you to make tactical decisions with a holistic view of your customers**
- **Assessing progress on marketing strategy.** For example, Finance team is now able to confirm whether new customer growth is up in our target segments.

FUTURE ROADMAP

- The next step for us was to capture **how customers differ by the problems** they're trying to solve and their willingness to pay. This can take the form of segments or personas, layered on top of our customer segmentation:
  - **Market-based segments:** Based on observable psychographic, firmographic and demographic attributes.
  - **Needs-based segments and personas:** Based on business needs, use cases, and jobs-to-be-done.
  - **Value-based segments and personas:** Based on willingness to pay versus the status quo.

THANK YOU

QUESTION/ FEEDBACK

Notebook -
https://colab.research.google.com/drive/1outrmpAivgGTarOkTGpf9jujOZHmlpvL?usp=sharing#scrollTo=B7c1fmY1TMR5

Team 5

Anwesha, Anisha, Arkadip, Darakshan, Naveen, Parth, Yathish