HUNT N' GATHER

an eclectic online store

ADVANCED CUSTOMER SEGMENTATION
K-MEANS ALGORITHM

BUSINESS SITUATION

The marketing department has established which customers fall into these classic marketing segments:

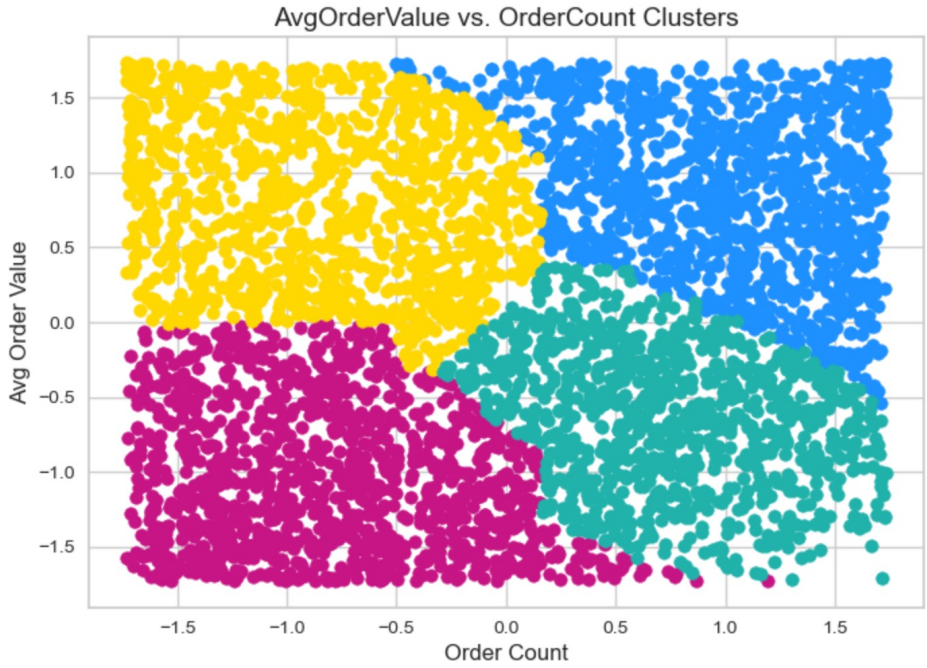
Highest Value Customer:
Highest Avg Order Value / Highest Order Count

High Value Customer:
High Avg Order Value / Lower Order Count

Lower Value Customer:
Lower Avg Order Value / Lower Order Count

Lowest Value Customer:

Lowest Avg Order / Lowest Order Count



WHAT'S MISSING?

We always want to move customers to the magic quadrant.

What's missing are specific buying behaviors that can give us more insight into determining what interventions can be used to change behavior.

OBJECTIVES

- Determine what new buying behaviors can be uncovered AND
- What new segments emerge?

WHY USE A CLUSTERING ALGORITHM?

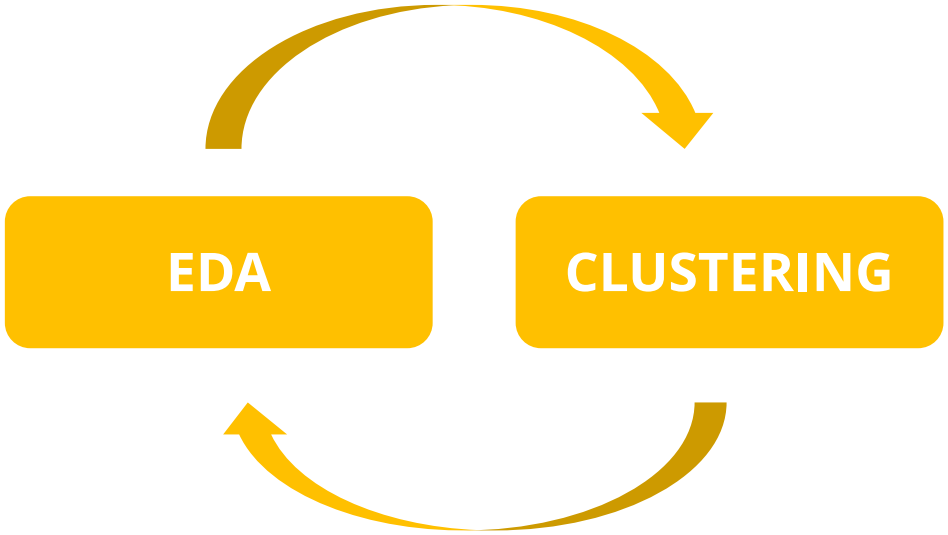
Leverage advanced math and computational power

Process more potential purchasing behaviors

Includes all customers, not just a small sample
Strategies are implemented on the exact customer

Utilize k-Means algorithm as starting point
Explore other algorithms (Hierarchal, etc) as next phase of work

PROCESS



AVAILABLE DATA

Raw Data

Rows
541,909 Invoice Line Items

8 Potential Features:
- Invoice No
- Stock Code
- Product Description
- Customer ID
- Country
- Quantity Ordered
- Invoice Date / Time
- Unit Price

Unique Values
- 25,900
- 4,070
- 4,223
- 4,372
- 38 Countries / 91% transactions come from UK
- Range: 1 – 80,995
- 2010 & 2011, Months 1-12
- NA

	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	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

Workflow to focus on business buyer at first

BUYING BEHAVIORS TO EXPLORE

	Buying Behavior (2010 - 2011)	Feature(s) to Engineer
WHO Business vs Retail Customer	**Business and Retail Buying Segments?** Understand what these segments look like with emphasis of analysis on the business buying segment. Initial data indicated qty purchase range of a single item from 1 – 80,995	**Separate Business Buyers from Retail Shoppers** - Business Buyers > x items per given stock code - Retail Shoppers < x items per given stock code
HOW MANY ITEMS Quantity by item / per invoice	**Who are low, mid, high quantity purchasers, by item - per order?** Understanding how they cluster given other features can give us more dimensions of learning about them.	**Business Buyer Features based on Quantity They Order** Look at distributions of order quantity to determine break points Low, Mid, High features
HOW MUCH Are they spending	**How does spend vary between the newly identified clusters?** This creates a baseline metric from which we know we need to move a given customer segment. Amongst other features we can create additional hypothesis about this customer.	Standard deviation criteria - remove outliers K-means doesn't handle outliers well so this is especially helpful for this algorithm
WHAT ITEMS Type of Product Purchased	**Do certain types of products tend to be bought within low, mid, high, quantity purchasers?** If so, what can we hypothesize about these people and what other items they may want?	**How to do this** - Use stock code to get to this - Use an NLP tactic to understand type of product.
WHEN Cadence of purchasing	**Uncover purchase motivators based on when they buy** What specific months do business customers tend to purchase? What specific months do retail customers tend to purchase?	- Groupby business customer and month - Groupby retail customer and month - Get # of unique months
WHERE Country Purchased	**How do UK customers behave and segment?** Focus on understanding UK customers first, they make up 91% rows of data	[Country] == United Kingdom