HUNT N' GATHER

an eclectic online store

CUSTOMER SEGMENTATION
K-MEANS ALGORITHM

BUSINESS SITUATION

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

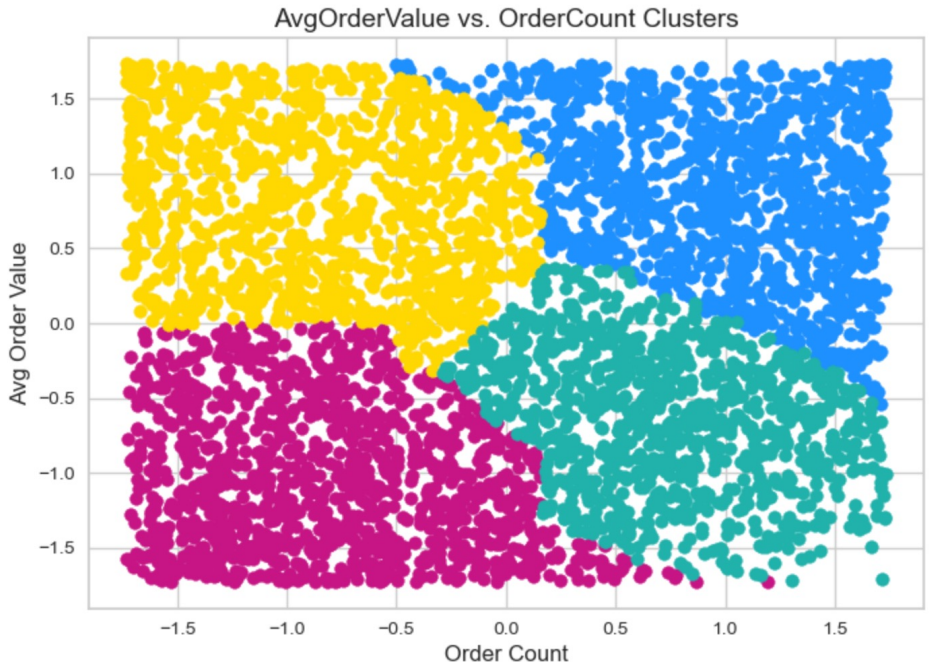
Highest Value Customer:
Highest Avg Order Value / Highest Order Count

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

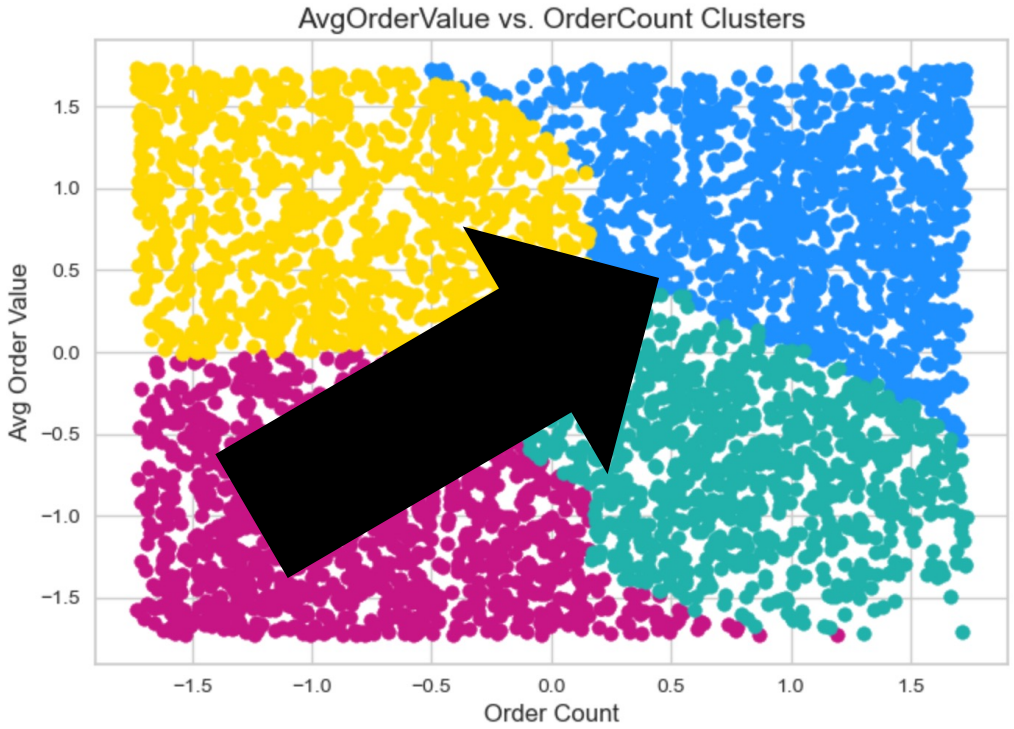


WHAT'S MISSING?

We always want to move customers to the magic quadrant.

What's missing is identifying specific buying behaviors that can be incorporated into a clustering algorithm

This provides a more nuanced picture of the segmentation from which specific marketing interventions can be ideated



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

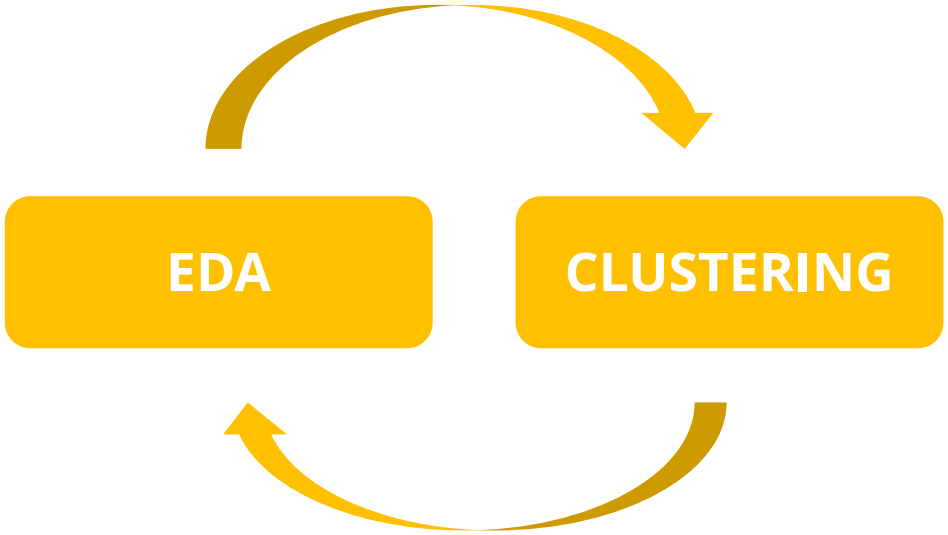
OBJECTIVES

- Determine buying behaviors to be uncovered through exploratory data analysis

AND

- Determine what new segments emerge via k-Means clustering?

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

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 Initial data indicated qty purchase range of a single item from 1 – 80,995	**Business Buyer =** Purchased > 10 items of a stock code, > 2 different stock codes **Retail Buyer =** purchased < 10 items of a stock code
HOW MANY ITEMS Quantity by item / per invoice	**Can we characterize business buyers based on qty they buy?** Are there low, mid, high quantity business buyers? If so, what other features characterize these people, how do they cluster?	**Lower Qty Business Buyer** **Mid Qty Business Buyer** **High Qty Business Buyer** **Outlier Qty Business Buyer**
HOW MUCH Are they spending	**What observations can be made about the annual spend of the segments?** Which segments, and corresponding behaviors, need to most be moved based on spend level?	Spend features based on Standard deviation criteria - remove outliers Create feature for high spend outliers. **Lower Spend – Business Buyer** **Mid Spend – Business Buyer** **High Spend – Business Buyer** **Outlier High Spend - Business Buyer**
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 other items they may want?	**High Frequency Stock Codes** **Mid Frequency Stock Codes** **Low Frequency Stock Codes** - Use an NLP tactic to understand type of product.
WHEN 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