HUNT N' GATHER

an eclectic online store
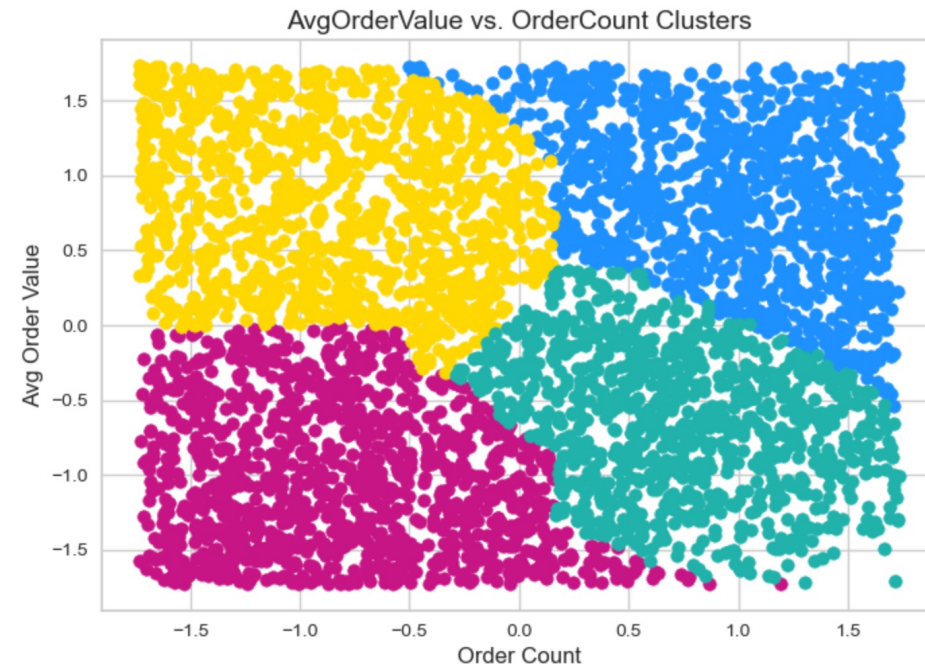
**CUSTOMER SEGMENTATION
K-MEANS ALGORITHM**

BUSINESS SITUATION

The marketing department has established 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

AvgOrderValue vs. OrderCount Clusters

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
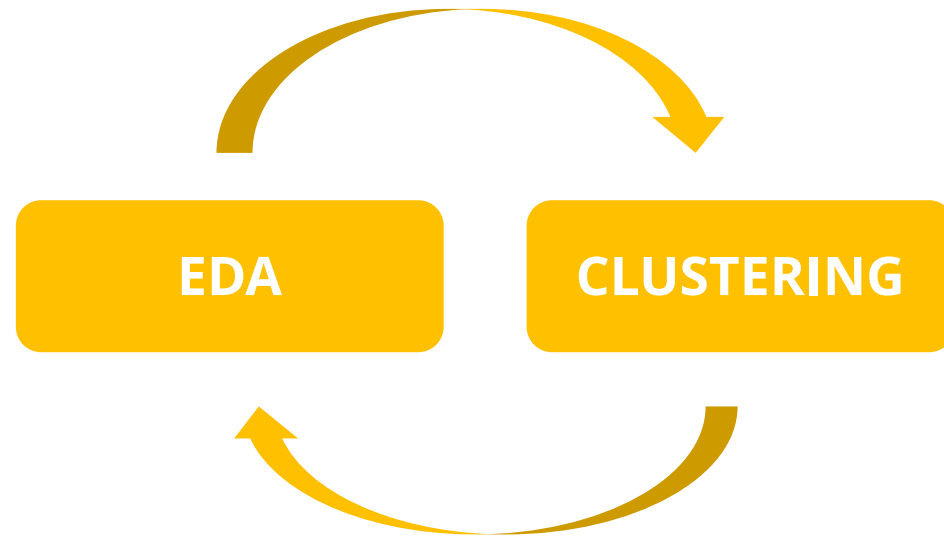Explore other algorithms (Hierarchal, etc) as next phase of work

OBJECTIVES

- Determine buying behaviors to be uncovered through exploratory data analysis

AND

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

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

FIRST LOOK EDA OBSERVATIONS

Finding	Approach
2010: Only December Data Exists	Don't use 2010 Data
2011: Data for all 12 months, Incomplete Data for December (Missing day 10-31)	Use all 2011 data
91% of transactions from UK	Focus on UK
Appears there's also a business customer in addition to retail customer Evidence: - 31% of line item purchases > 10 items - Customer bought 32 assorted color bird ornament, - Customer bought 24 alarm clock bakelike red - Spend range $1 - $337K (2011) - qty purchase range of a single item from 1 – 80,995	Identify business customer to determine if we can learn more about them through clustering
Outlier High Spenders $25K-$337K	Identify these specific people

	Buying Behaviors	**Feature(s) to Engineer** Business Rules + Working Titles
WHO Business vs Retail Customer	**Are there Business and Retail Buying Segments?** Understand what these segments look like **Can we characterize buyers based on quantity of individual line items that 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	**Logic:** It's peculiar for a customer to buy more than 10 items of a given novelty item. Identify them as business buyers to learn about them and see how they compare to others A person could purchase something for a party or social occasion at 10 items. Without any additional insight from these people use an intuitive 10 qty break point. **High =** Purchased > 10 qty of a stock code in a sitting – not cume **Low =** purchased < 10 qty of a stock code in a sitting – not cume **Logic:** Understand different business customers **Business Buyer =** Customers who bought had lots of line items at 10+ qty **Rule:** 80% or more line items are > 10 items per customer **Combo Biz/Retail Buyer** = Some line items are more than 10 qty, but more at less than 10 qty **Rule:** <= 20% of line items are > 10 items per customer **Parking Lot Idea**: Do they tend to buy the same items or a variety of items? **Logic:** Identify Retail Purchasers determine how they cluster and potential characteristics that define them. **Retail Buyer =** Most of the line items a customer purchases are < 10 qty **Rule: >=** 80% of the total line items are < 10 qty
HOW MUCH MONEY 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? - What characteristics do segments have based on annual spend level?	**Logic: Use float value of 2011 annual spend by customer** **Annual Spend = Sales Sum by Customer**
WHEN Are they spending	**Uncover purchase motivators based on when they buy** What specific months do customers tend to purchase? How does this vary between retail and business customers?	- Groupby business customer and month - Get # of unique months

BUYING BEHAVIORS TO EXPLORE

	Buying Behavior (2010 - 2011)	Feature(s) to Engineer
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**