

# Customer Segmentation Analysis

---

March 8 2024  
Jiazhou Pang

# Project Overview

- Define objectives
- Understanding the dataset
- Data processing
- Segmenting the customers
- Evaluate the customer segments and derive strategies

**Project objective:**  
Create customer  
segments to inform  
customer retention  
strategies

# Understanding the Dataset

- Transactional data spanning 2 years
- **541909** entries
- **8** variables
- Limited customer individualization, feature engineering was needed

	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

# Basket Value

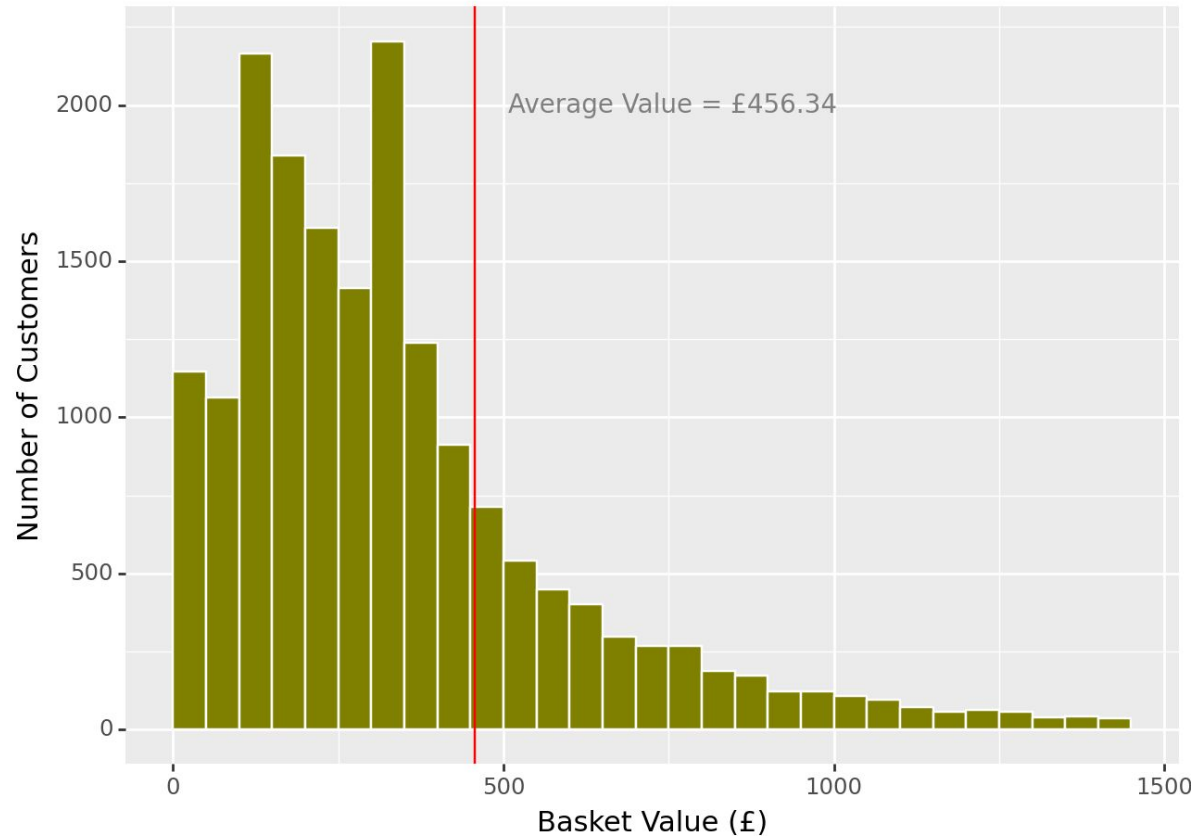
Basket values measures how much a customer spends on one transaction.

On average, customers spend **456.34** Sterling for each transaction.

The graph on next slide shows the distribution of average basket values.

---

The distribution of basket values is skewed to the right



Note: values larger than £1500 were excluded for the readability of the graph

# Text Mining

- To glean more information of customer behavior, clustering the production description would help
- Descriptions were tokenized and stemmed as shown below
- Product clusters were obtained from k means
- 10 clusters in total
- Amt. of money each customer spends on each product category was calculated

	Description	TokenizedDesc	cluster
1	50'S CHRISTMAS GIFT BAG LARGE	's christma gift bag larg	0
15	12 EGG HOUSE PAINTED WOOD	egg hous paint wood	0
27	12 PINK ROSE PEG PLACE SETTINGS	pink rose peg place set	0
50	3 RAFFIA RIBBONS 50'S CHRISTMAS	raffia ribbon 's christma	0
91	5 STRAND GLASS NECKLACE AMBER	strand glass necklac amber	0

## Examples of Product Clusters

## Cluster 0:



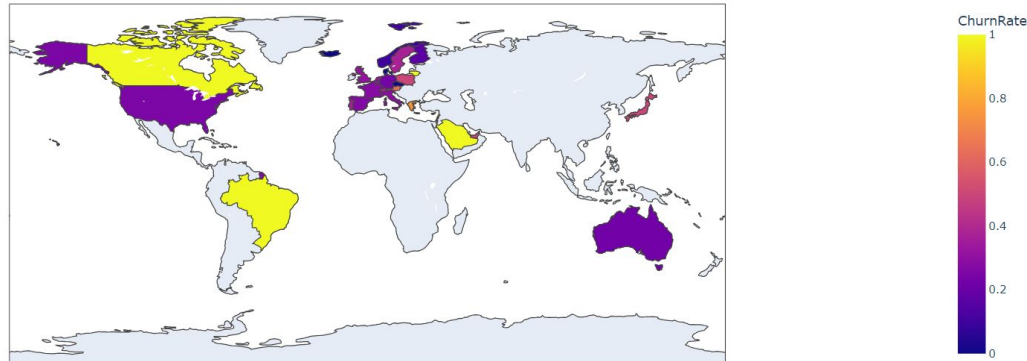
### Cluster 3:





# Defining Churn Rate

- Customers take **72** days to place their next order on average
- Hence, customers who have not placed any orders after Sep 01 2011 (100 days before the last date) as churned
- Overall churn is **31.43%**
- Churn rate relatively uniform across countries
- Countries with very high/low churn rates have very few customers



# Customer Segmentation

## Features Used

- Days since first & last purchase
- No. of transactions
- IsUK
- Cancel rate
- Avg, Max, Min of amount spent
- Sum spent in each product category



RECENTY



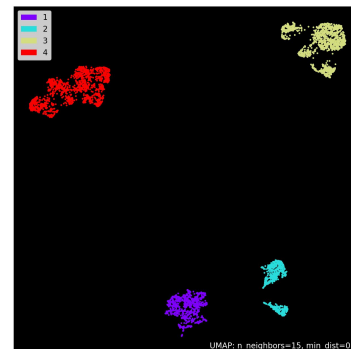
FREQUENCY



MONETARY

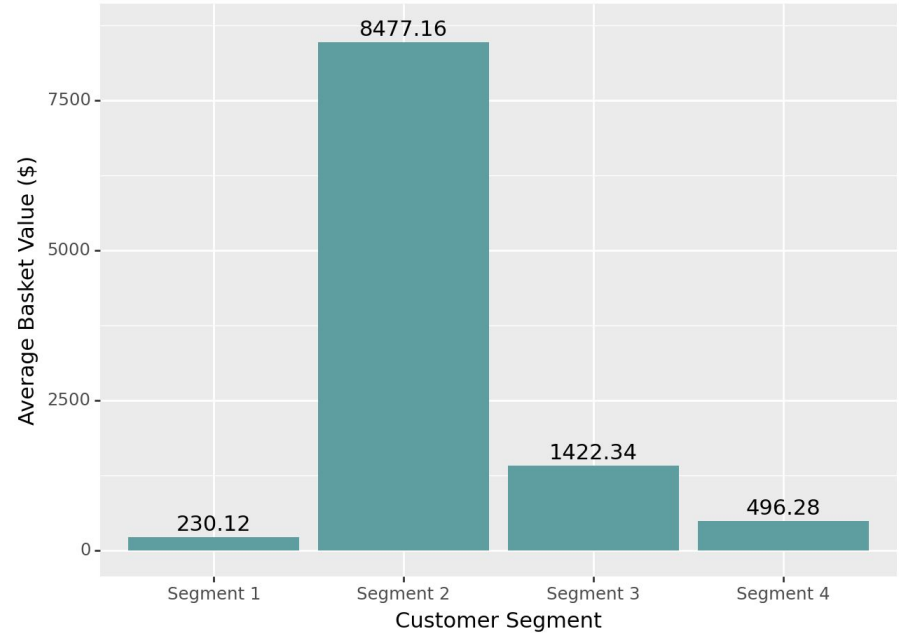
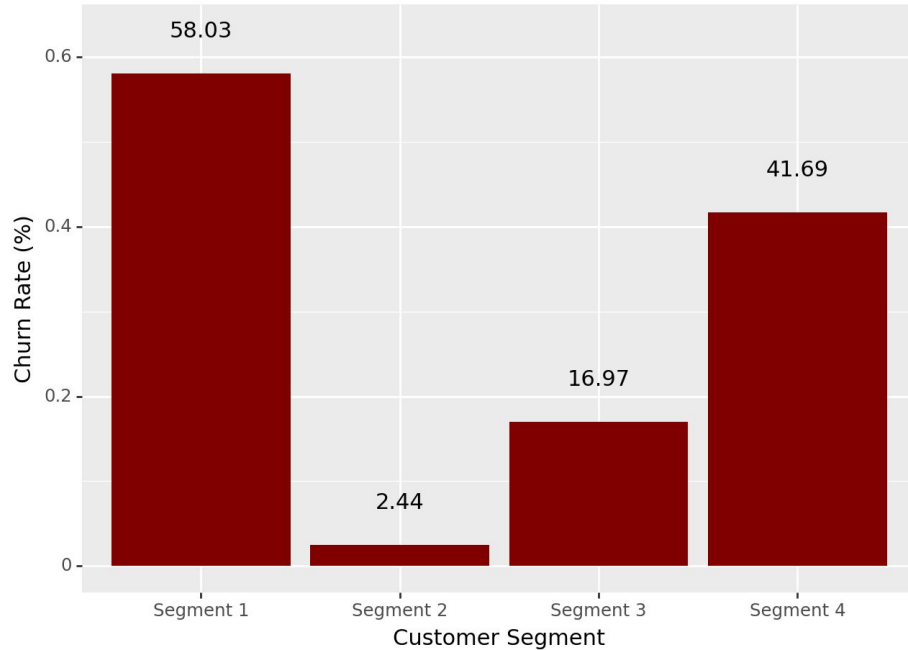
## Clustering

- HDBSCAN clustering optimizing DBCV, only two stable clusters obtained
- K Means produced 4 stable clusters. It is the method chosen for this analysis. Silhouette score = 0.151



# Characteristics of the customer segments

- There is a positive correlation between churn rate and average basket value
  - The differences in churn rates are drastic



# Understanding the Customer Segments

## Customer Segment 1

- **58% Churn, 865** customers
- **Almost lost**
- Relatively more likely to purchase products from cluster **0** and **4**
- Least likely to cancel orders (9.2% Avg. cancel rate)

## Customer Segment 2

- **2.4% churn, 655** customers
- **Loyals**
- Relatively more likely to purchase products from cluster **6** and **9**
- Most likely to cancel orders (22% Avg. cancel rate)

## Customer Segment 3

- **17% churn, 1332** customers
- **Promising**
- Relatively more likely to purchase products from cluster **2** and **3**

## Customer Segment 4

- **42% churn, 1487** customers
- **Needs attention**
- Relatively more likely to purchase products from cluster **1, 3** and **9**
- Days since first purchase similar to segment 1

# Recommendation

- Focus on improving sales generated by customer segments 3 and 4
- Promotions / advertise on products frequently purchased by aforementioned customers
- Examine the order cancellation process as it negatively correlates with customer retention rates
- Examine products that are more popular among customer segments 1 and 4

# Moving Forward

- A/B testing on promotion
  - Account for seasonality in the dataset
  - Deploy a customer classification model to automate marketing
-