Customer Segmentation: Online Retail

A Springboard Capstone Project

Eduardo Mosada Jr.

The Problem:

How can an online retailer target specific groups of customers for its marketing campaigns and strategies based on their purchasing patterns?

Objectives:

- ✓ Valuable insights into customer purchasing behavior
- ✓ Customer-centric data anchored on RFM features
- ✓ Segmented customers based on RFM analysis.

Benefits:

- Better understanding of customers' behaviors and ever-changing needs
  - ❑ Targeted Marketing Campaigns/Strategies
  - ❑ Enhanced Customer Experience
- Drive innovative and creative solutions
  - ❑ Boost product sales and profitability
  - ❑ Promising outlook for the future

The Data:

- UK-based all-occasion gift-ware online retailer
- **500,000+** transactions in 2010
- **4,000+** customers recorded

Data Source: [UCI Machine Learning Repository | Online Retail II]

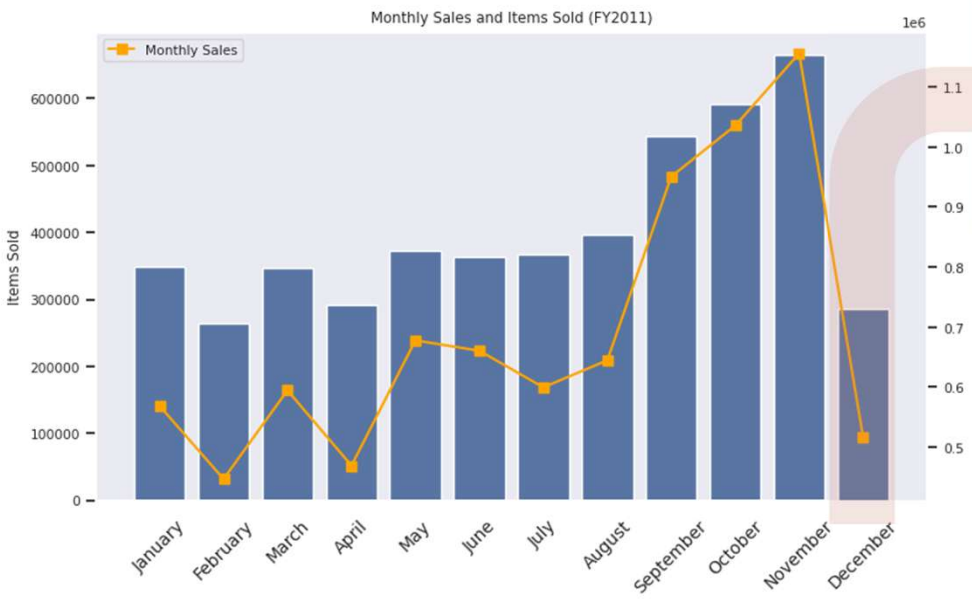
Data Wrangling

- **541,910** Sales transactions in 2010
- **25 %** Blank Customer IDs - 135,080
- **1.6 %** Negative values in quantity – 8,905
- **1 %** Duplicate records – 5,192
- **72.5%:** Records retained – 392,733

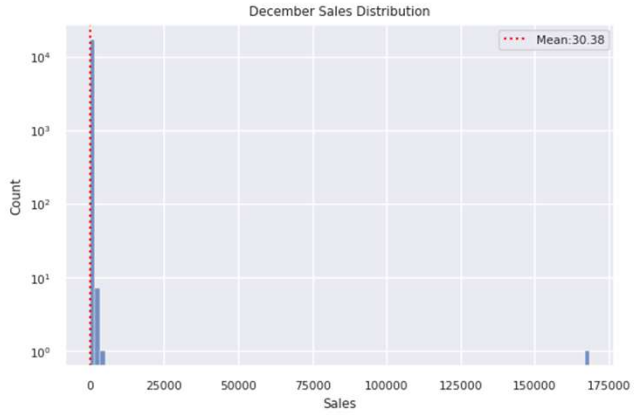
Exploratory Data Analysis

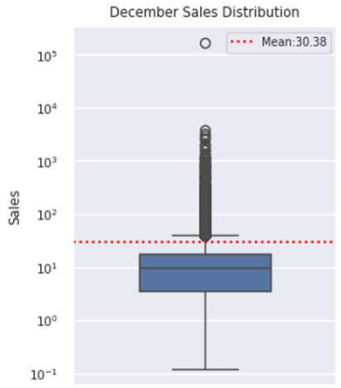
- **392,733** Records for EDA
- **40** Zero-Priced items
- **2,560** Out of range Invoices
- **367,023** Records processed

- **17,132** Invoices
- **4,219** Customer IDs
- **3,596** Stock codes
- **36** Countries



Analysis of 55% Dip in December Sales





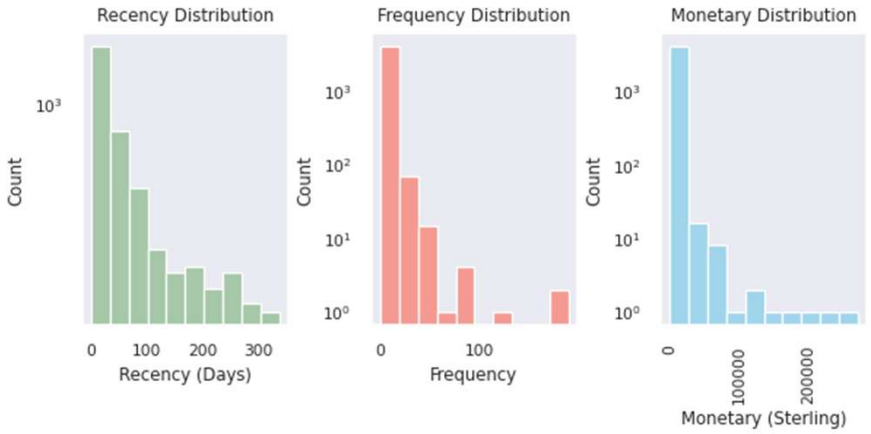
- Huge number of **low-value** transactions
- December only accounted for **5% of the total transactions** for the current year
- Last transaction is **December 9**

Incomplete transactions

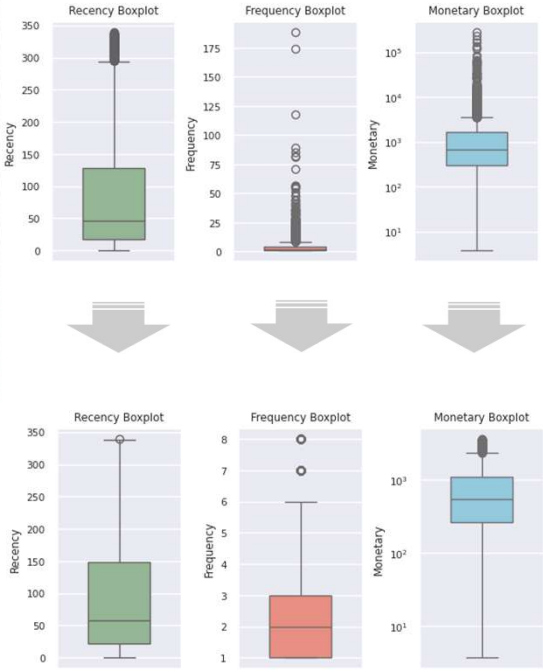
Pre-Processing

Feature Engineering

- **Recency** - How long since the customer's last purchase date?
- **Frequency** - How many transactions did the customer have?
- **Monetary** - How much was spent by the customer?



Dealing with outliers



Outliers for further analysis

- **402** Monetary outliers
- **412** Frequency outliers
- **269** M x F outliers

Non-outliers for modeling

- **87 %** Non-outliers – 3647
- **3,647** – Scaled data

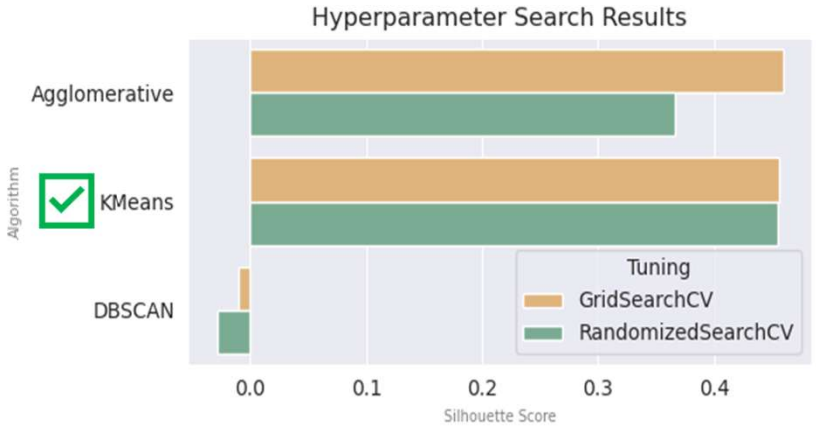
Modeling

Clustering Algorithm Options

- **KMeans Clustering**
- Agglomerative Hierarchical Clustering
- DBSCAN

Hyperparameter Seach Implementation

- **3** clustering algorithms
- **2** Cross-Validation (Nested CV)
- Scoring function: **Silhouette Score**



Modeling cont'd.

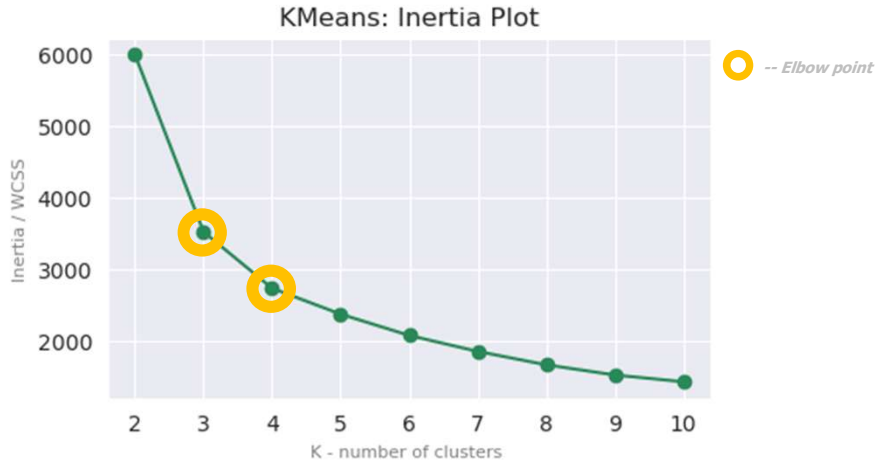
Cross-Validation vs. Sequential Model-Based Optimization SMBO Search

Tuning	Inertia	Silhouette Score
☑ **SMBO**	**5989.735**	**0.432**
RandomizedSearchCV	3521.054	0.453
GridSeachCV	3517.192	0.453

SMBO hyperparameters:

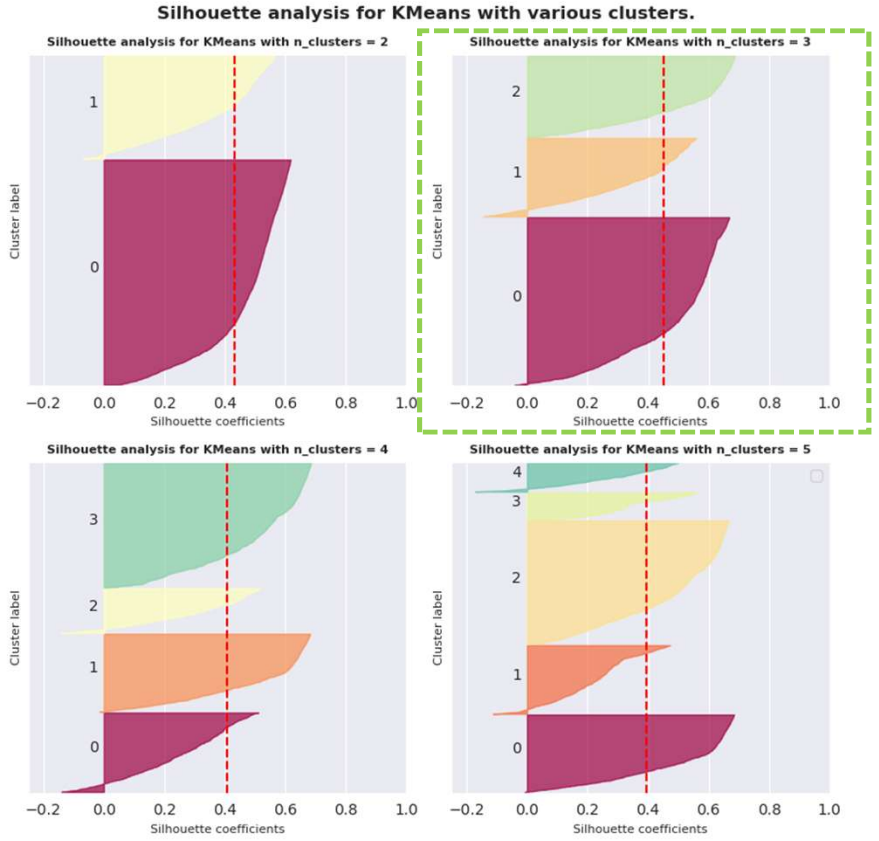
- algorithm : 'lloyd'
- init: 'random'
- max_iter: 645
- n_clusters: 2
- n_init: 5
- tol: 0.1

Optimal K- no. of clusters : Inertia Plot



Deciding between n_clusters= 3 and 4 -> Silhouette Analysis

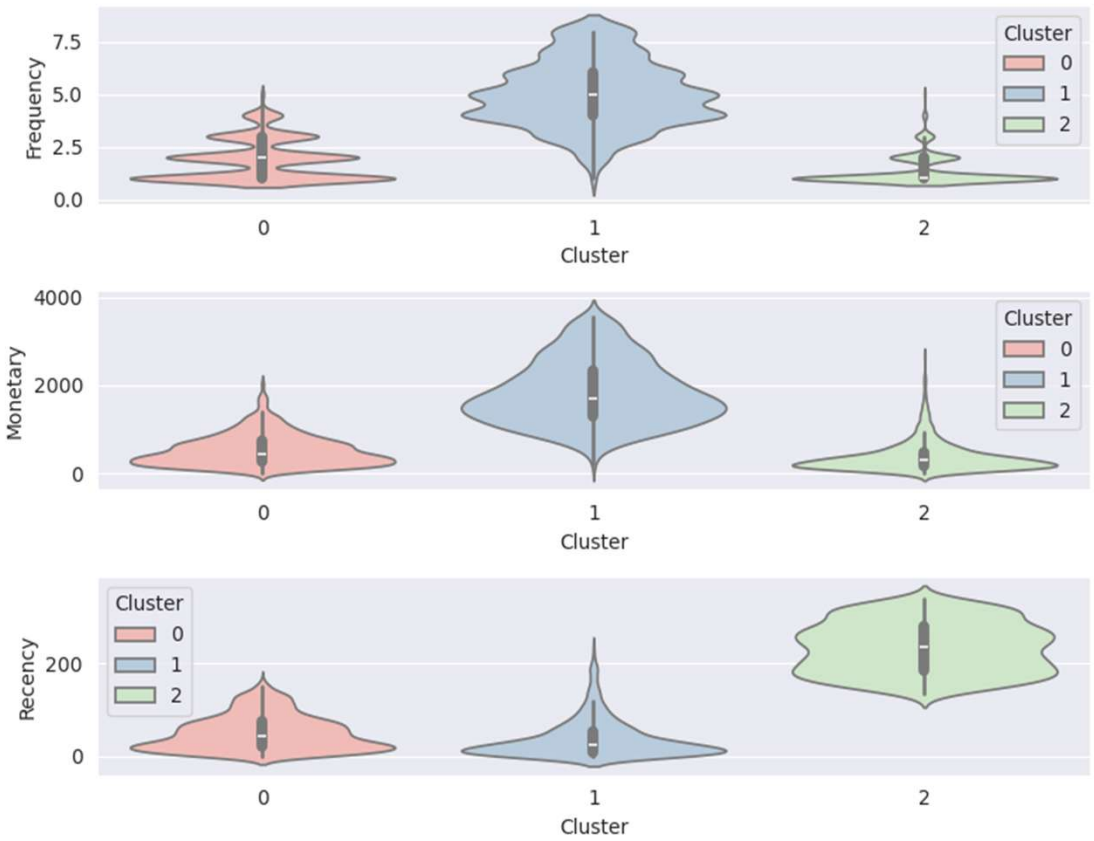
Optimal K- no. of clusters: Silhouette Analysis



Optimal K- no. of clusters: 3

Modeling cont'd.

Cluster Analysis



Assigning Labels

Cluster 0: ***Moderate***

- Moderately frequent buyers
- Not necessarily high spenders
- Majority made recent purchases

Cluster 1: ***Loyal***

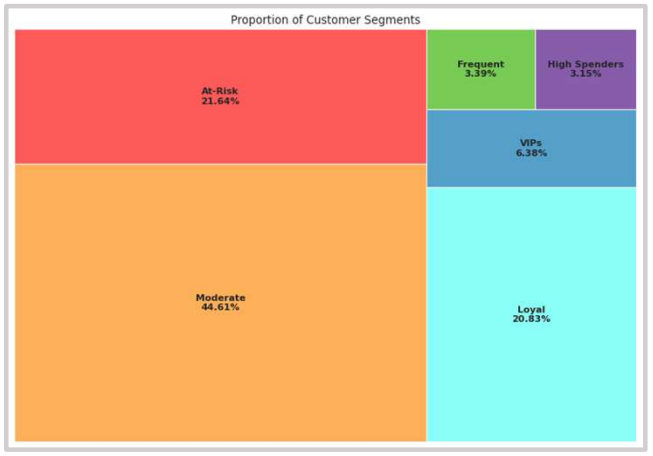
- Frequent buyers
- High spenders
- Majority made recent purchases

Cluster 2: ***At-Risk***

- Least active recently
- Less frequent buyers
- Low-spenders

Summary

Customer Segmentation



Customer Segmentation Analysis and Recommendations

Segment	Description	Recommendations
MODERATE	Moderately frequent buyers that are not necessarily high spenders. Majority in this segment made recent purchases.	▪ Offer subscription on frequently bought items (whenever applicable) ▪ Recommend "Frequently bought together" items ▪ Implement customer retention and loyalty programs
AT-RISK	Least active segment. Less frequent buyers who are low-spenders and have minimal recent purchases.	▪ **Requires attention and re-engaging** ▪ Perform sentiment analysis. Send surveys and ask for reviews or feedback. Identify areas for improvement to enhance the overall buying experience ▪ Whenever possible, drop small rewards, such as reactivation bonuses or discounts on their next purchase. ▪ Tease or entice with the benefits of becoming active or loyal customers
LOYAL	Frequent buyers who are high spenders, and with its majority having purchased recently.	▪ Implement rewards and loyalty programs, and exclusive perks ▪ Run targeted ads for trending items and top items sold ▪ Offer subscription on frequently bought items (whenever applicable)
VIPS	High value, frequent buyers.	▪ Offer exclusive perks and vouchers ▪ Pamper and enhance the shopping experience by offering expedited or free shipping (whenever applicable)
FREQUENT	Very frequent buyers.	▪ Offer subscription on frequently bought items (whenever applicable) ▪ Implement customer retention and loyalty programs ▪ Offer expedited or free shipping vouchers (whenever applicable)
HIGH SPENDERS	High-spending buyers	▪ Implement customer retention and loyalty programs ▪ Offer discounts on bulk purchases or reaching certain amount (whenever applicable) ▪ Enhance customer experience and services

Recommendations and Next Steps:

- Evaluate the feasibility of launching targeted marketing campaigns
- Develop and periodically monitor metrics
  - Effectiveness of marketing campaigns
  - Implementation of new processes
  - Customer responses and sentiments

  > Do sales, profitability, and customer engagement/retention improve over time?
- Explore additional customer segmentation methods beyond RFM analysis
- Sentiment Analysis on existing and future products and offerings
  - Revisit data gathering process
  - IT landscape transformation
  - Omnichannel approach for seamless experience
- Recommendation systems to introduce product recommendations

Sources:

Github Project Repository	Customer Segmentation for an Online Retail
Machine Learning Libraries	Scikit-Learn Clustering Hyperopt Hyperparameter Optimization
Data Source	UCI Machine Learning Repository | Online Retail II

Special Thanks:

Jyant Mahara (Springboard Mentor) https://www.linkedin.com/in/jyant-mahara-95877b93/

Thank you

Eduardo Mosada Jr.


edu.mosada.jr@gmail.com


https://github.com/ds-edu/


https://www.linkedin.com/in/edumosada/