ML to Improve Marketing ROI

MIDS 207 Spring 2025 Section 6 Final Project

Saaketh Gunukula, Shreshta Keta, Vishal Saxena

Table of Contents

Business Context

Customer Experience Excellence (CX)

- Critical component of strategic priorities for companies across sectors
- Must adapt these for all customer demographics — Gen-Z, Millennials, Senior Citizens
- Need to innovate on CX across channels, product lines, and market segments.

Comparison with other work

- Previous work has used Lifetime Value (LFTV) and churn predictive models in industry to reduce Marketing Selling, General, and Administrative (SG&A) tasks
- Ongoing focus for incorporating varying data sets to improve commercial performance of Machine Learning engines for this task continues
- From our research, there is little to no research on the combination of customer segmentation with LFTV and churn models to improve campaign performance

Business Context and Opportunity

Key Definitions

- Customer Experience (CX): Overall perception a customer has across all touchpoints with a brand.
- Churn Prediction: Identifying customers likely to stop engaging or purchasing.
- Lifetime Value (LFTV): Total net profit attributed to the entire future relationship with a customer.

Business Problem

- Companies struggle to personalize CX across diverse demographics.
- Generic marketing campaigns lead to low engagement and poor ROI.
- Churn and LTV insights are often underutilized or siloed in strategy.

Opportunity

- Use machine learning to integrate churn, LTV, and segmentation into one unified pipeline.
- Deliver personalized, data-driven marketing to improve ROI and customer retention.
- Enable smarter allocation of campaign budgets toward high-value customers.

History, Data Challenge, Past research

Customer Experience Excellence (CX)

- Critical component of strategic priorities for companies across sectors
- Must adapt these for all customer demographics — Gen-Z, Millennials, Senior Citizens
- Need to innovate on CX across channels, product lines, and market segments.

Comparison with other work

- Previous work has used Lifetime Value (LFTV) and churn predictive models in industry to reduce Marketing Selling, General, and Administrative (SG&A) tasks
- Ongoing focus for incorporating varying data sets to improve commercial performance of Machine Learning engines for this task continues
- From our research, there is little to no research on the combination of customer segmentation with LFTV and churn models to improve campaign performance

History, Data Challenge, Past research

What Industry Has Tried

- Applied LFTV models to optimize marketing spend and customer prioritization.
- Used churn models to flag at-risk users and reduce SG&A costs.
- Efforts often focus on individual models -segmentation, LFTV, or churn - not all three together.

Ongoing Challenges

- Limited research combining customer segmentation + LFTV + churn into one strategic ML framework.
- No clear pipeline for using these combined insights to drive campaign performance or ROI.

Our Question

Our Goal: Improvement of Marketing Campaign ROI

From this scope, in this work we address two questions:

- How can a Retail Chief Marketing Officer **improve their marketing campaign ROI** by launching effective intervention campaigns informed from Machine Learning models for customer lifetime value (LFTV) calculation and churn prediction?
- Does the use of Machine Learning defined Customer Segmentation improve model accuracy?

Impact: A repeatable ML engine that can continuously incorporate new data to improve campaign performance and learn from the market can be a tremendous asset for any company.

Our Question

Our Goal: Improve marketing campaign ROI through a unified ML-driven approach

From this scope, in this work we address two questions:

- How can a Retail Chief Marketing Officer **improve their marketing campaign ROI** by launching targeted intervention campaigns based on machine Learning models for customer **lifetime value (LFTV) calculation** and **churn prediction**?
- Does integrating machine-learned Customer Segmentation improve the accuracy and effectiveness of these predictive models?

Impact: A repeatable ML engine that can continuously incorporate new data to improve campaign performance and learn from the market can be a tremendous asset for any company.

Our Algorithms

Churn Prediction

Goal: Predicting whether a customer will stop using the company's products

Highest Correlation with Target: Tenure, Complaint, Marital Status

Final model: Random Forest Classifier with Grid Search

Validation Accuracy: 95%
AUC: 98%
Improvement over baseline: 32.51%

Customer Lifetime Value

Goal: Determine customer lifetime value based on purchasing habits (continuous variable)

Highest Correlation with Target: Purchase frequency, membership years, purchase value

Final model: Feed Forward Neural Network with Feature Selection

Validation RMSE: 23.07%
Improvement over baseline: 74.9%

Customer Segmentation

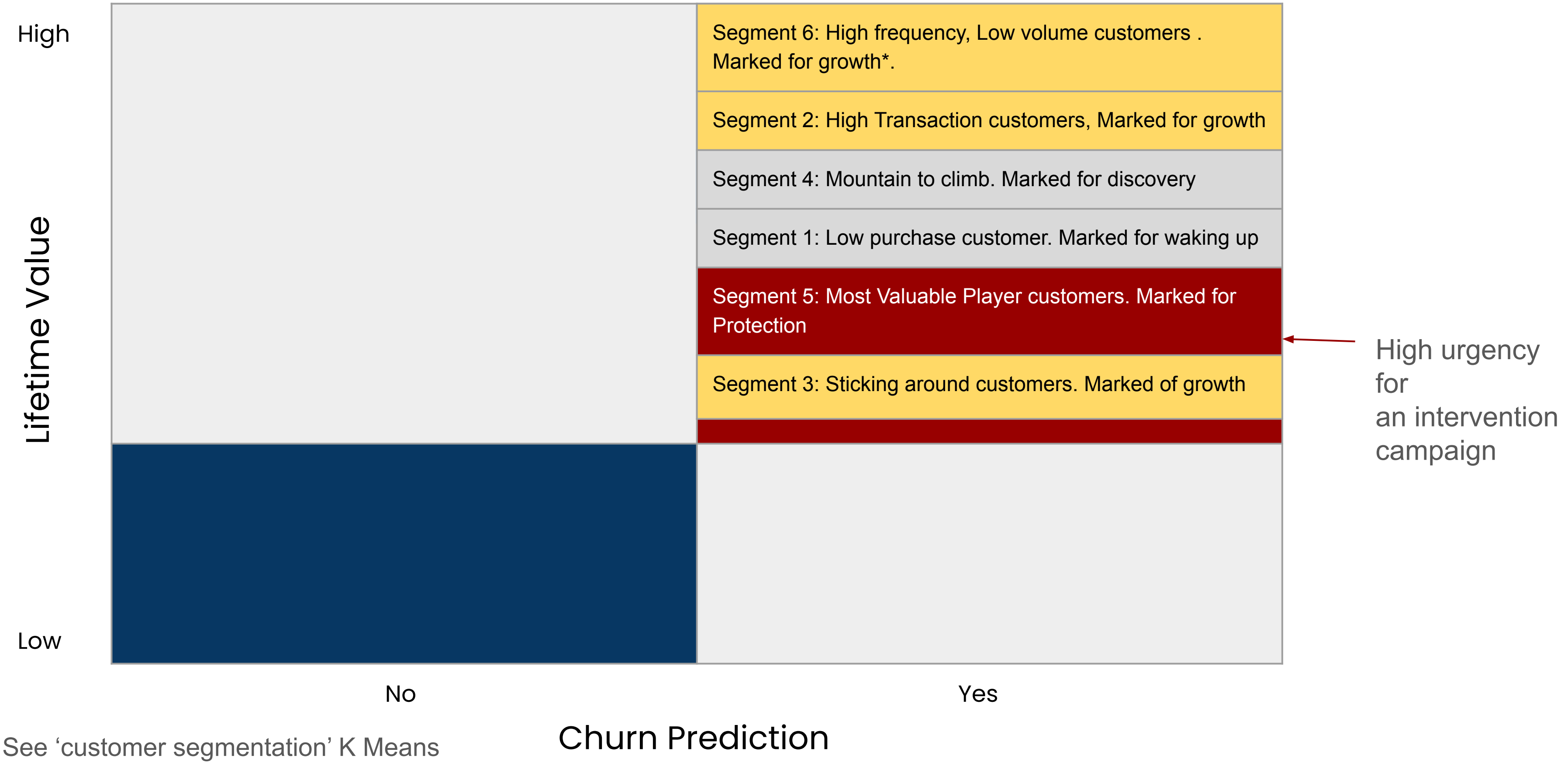
Goal: Understand customer demographics and consumption patterns via clustering algorithms.

Final model: K-Means with 6 clusters

Output: 6 Customer segments with meaningful action steps

Inertia**:** 2.94 on elbow-selected 6 clusters

Vision combines all three – LFTV, Churn, and Segmentation – to identify urgency of marketing action



* See 'customer segmentation' K Means section for complete segment definitions

Churn Prediction

Data Statistics and Preprocessing

Source:
https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction/data

Key Statistics:

Initial Shape:
After cleaning and EDA: (5630, 9)

Churn Model Data Cleaning

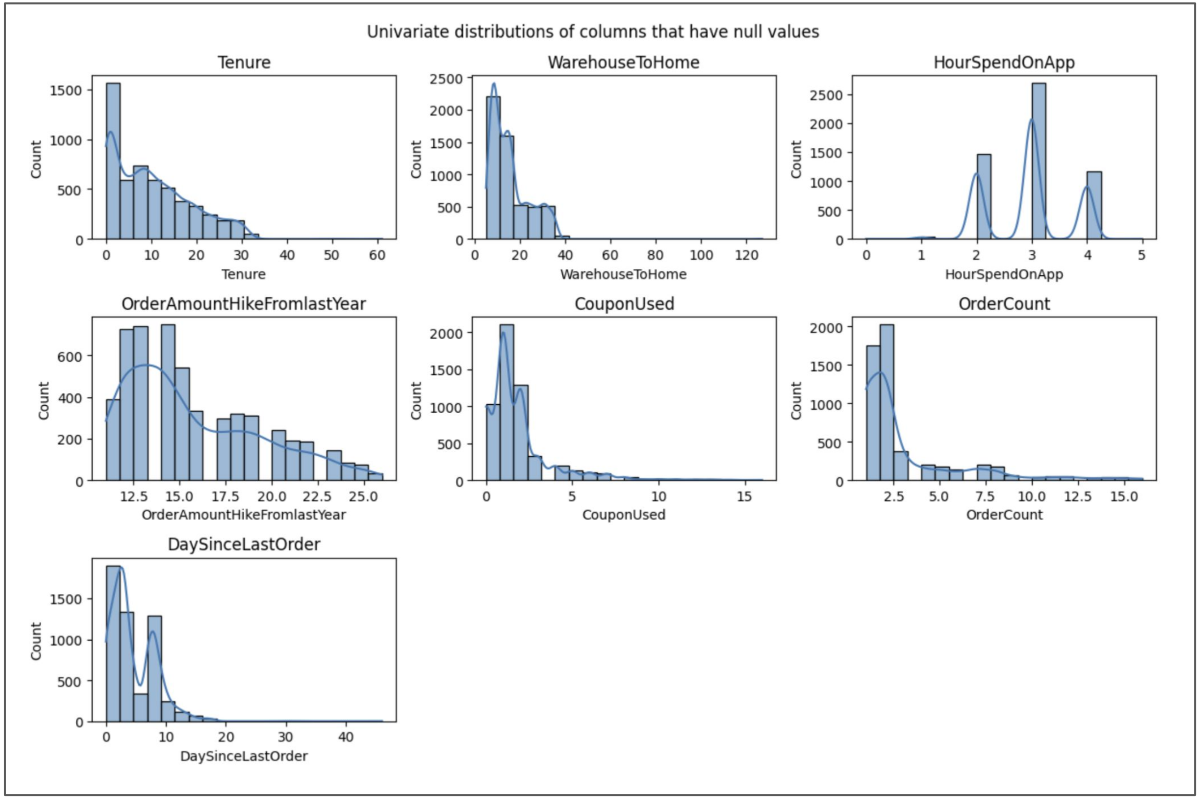
33% of the data had at least one null value in the row

The data to the right are the distributions of variables with null values

Since none had a normal distribution, we **imputed** null values using the **median**, which is better for skewed distributions

Additionally, we **One-Hot Encoded** the following categorical variables:

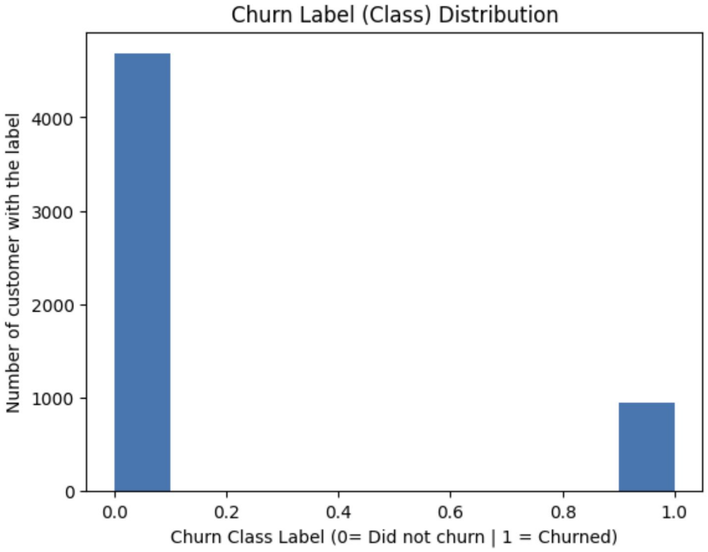
- Preferred Login Device
- Preferred Payment Mode
- Gender
- PreferedOrderCat
- Marital Status'



Churn Model EDA

Distribution of Churn Model Target Variable

~83% not churned, ~16% churned
After SMOTE: ~71% not churned, ~29% churned



Variables with the Highest Correlation to Churn

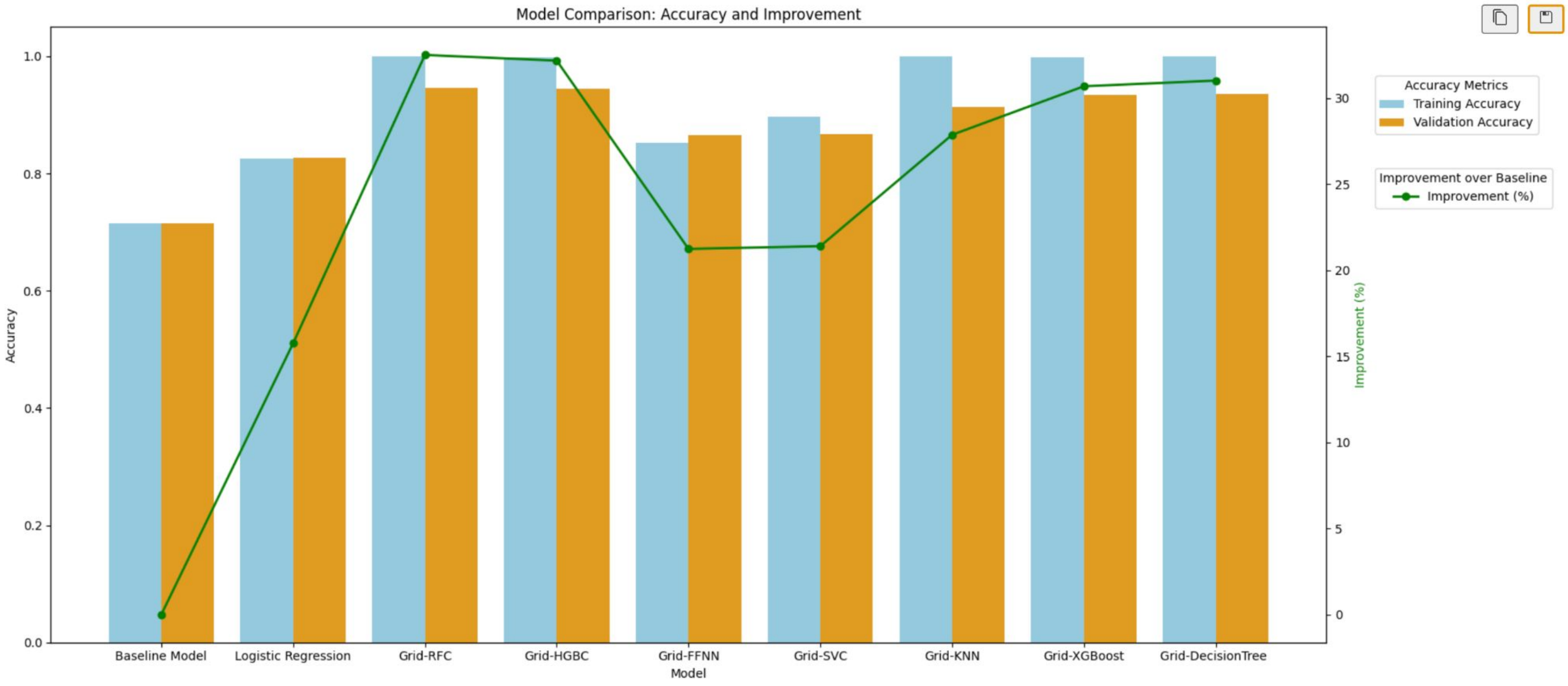
Dropped columns with corr < 0.1 to reduce noise

SatisfactionScore	0.105481
NumberOfDeviceRegistered	0.107939
PreferredLoginDevice_Mobile_Phone	0.111639
PreferedOrderCat_Mobile	0.113364
PreferedOrderCat_Laptop_and_Accessory	0.133353
CashbackAmount	0.154118
PreferedOrderCat_Mobile_Phone	0.154387
DaySinceLastOrder	0.155871
MaritalStatus_Single	0.180847
Complain	0.250188
Tenure	0.337831
Churn	1.000000

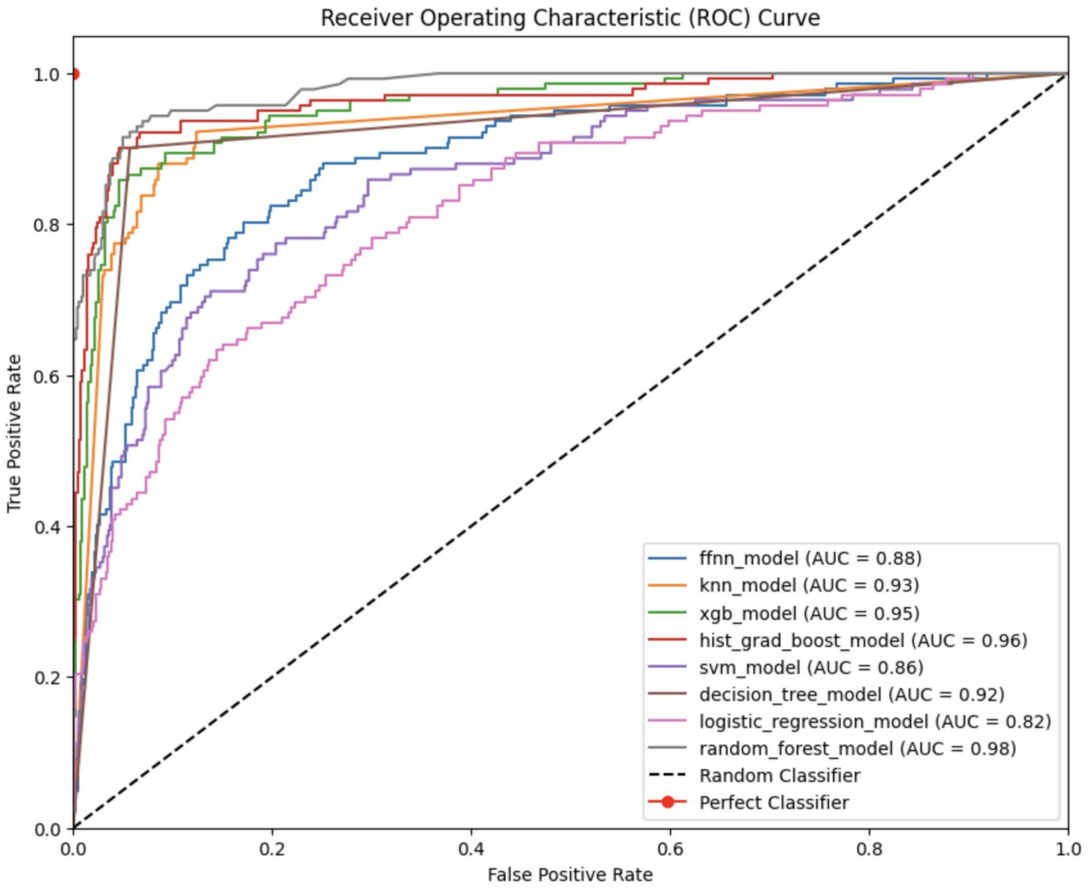
Baseline and Experimentation

Model	**Hyperparameters**	**Training Accuracy**	**Validation Accuracy**	**Improvement over Baseline %**
Baseline: Majority Class Classifier	N/A	0.71	0.71	0.00
Logistic Regression	Sigmoid activation, Cross Entropy Loss, 50 Epochs, Early Stopping	0.83	0.83	15.76
Grid Search Random Forest Classifier	**200 estimators, max depth = 0, max features = sqrt, min split = 2**	**1.00**	**0.95**	**32.51**
HistGradientBoostingClassifier comprehensive grid search	Learning rate = 0.2, max iterations = 300, min samples leaf = 20, l2 reg.	1.00	0.94	32.18
FFNN with Grid Search	2 hidden layers of 32 units, lr = 0.01	0.85	0.87	21.24
Support Vector Machine, grid searched, 5 fold cross validation	C = 10, degree = 3, gamma = auto, kernel = rbf	0.90	0.87	21.40
K-Nearest Neighbors, grid searched, 5 fold cross validation	Manhattan distance, 3 neighbors, p = 1, weights = distance	1.00	0.91	27.87
XGBoost Classifier, grid searched, 5 fold cross validation	Learning rate = 0.2, Max depth = 9, log loss, n_estimators = 150, col sample by tree = 0.7	1.00	0.93	30.69
Decision Tree, grid searched, 5-fold cross-validation	Entropy criteria, no max depth, no max features, min sample split = 2	1.00	0.94	31.02

Churn Prediction - Model Performance



Churn Prediction - Model Performance



Random Forest Model (best model) AUC is 0.98

Customer LFTV

Data Statistics and Preprocessing

Source: Retail Sales and Customer Behavior Analysis

https://www.kaggle.com/datasets/utkalk/large-retail-data-set-for-eda

Key Statistics:

Initial Shape: (1,000,000, 78)
After cleaning and EDA: (500,000, 24)

Feature Engineering to Develop our Ground Truth Customer LFTV:

LFTV = Average Purchase Value x Purchase Frequency x Customer Lifespan (CL)

Customer Lifespan was calculated using conditional logic based on loyalty membership years

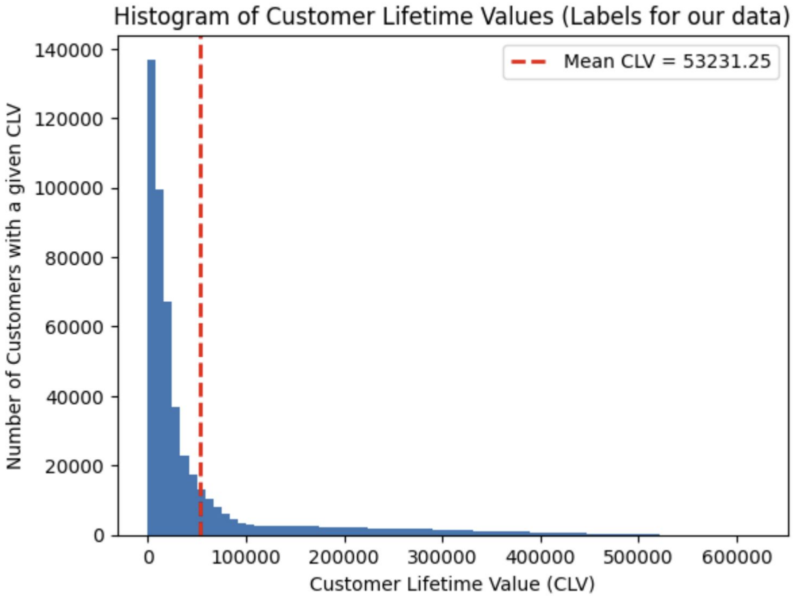
Final LFTV = mean of calculations from APV x PF x CL, and Total sales x CL

LFTV Data Cleaning

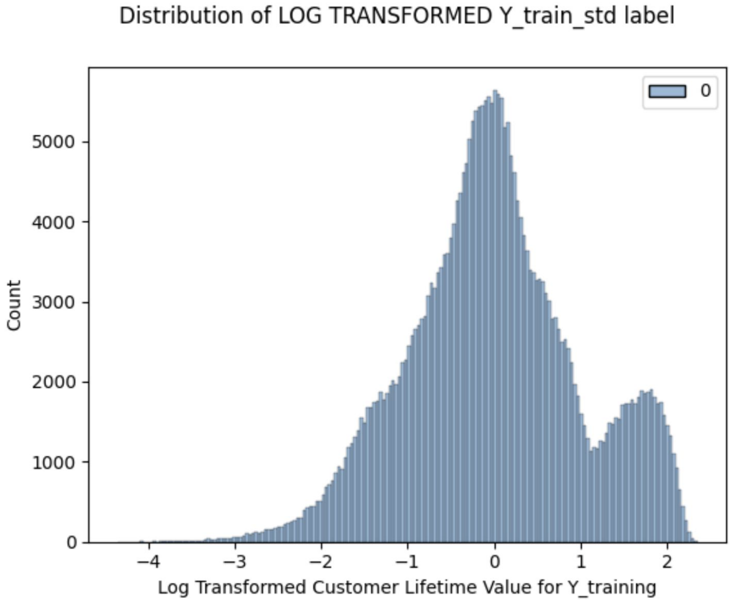
- Add noise-adjusted columns. Drop their 'pure' counterparts
  - Used normal distribution to add noise
  - Drop avg_purchase_value, purchase_frequency, membership_years, total_sales
- One Hot Encode categorical features
  - income_bracket, app_usage, social_media_engagement
- Address skewness, outliers in Y through transformation
  - Log transformation of Y (customer lifetime value)
- Standard scale numeric features

Customer LFTV EDA

Distribution of LFTV Model Target Variable before transformation



Distribution of LFTV Model Target Variable after log transformation



Customer LFTV Experimentation

	Baseline Model: Simple Linear Regression	Feed Forward Neural Network
Hyperparameters	N/A	2 hidden layers with unit sizes 32 and 16, relu activation, adam optimizer, MSE loss function, Early Stopping, 3329 parameters, 100 epochs, batch size of 32
Training RMSE	93.02%	23.00%
Validation RMSE	92.52%	23.07%
Improvement over baseline	N/A	74.9%
Test RMSE	N/A	23.23%

Customer LFTV Experimentation

Model	**Hyperparameters**	**Training RMSE**	**Validation RMSE**
Baseline: Simple Linear Regression with No Feature Selection	N/A	179.01%	178.29%
Simple Linear Regression with Feature Selection	N/A	93.02%	92.52%
Feed Forward Neural Network with No Feature Selection	2 hidden layers, unit sizes 32 and 16, relu activation, adam optimizer, MSE loss function, Early Stopping, 3585 parameters, 50 epochs, batch size of 32	178.73%	178.00%
Feed Forward Neural Network with Feature Selection	**2 hidden layers, unit sizes 32 and 16, relu activation, adam optimizer, MSE loss function, Early Stopping, 3329 parameters, 100 epochs, batch size of 32**	**23.00%**	**23.07%**

Customer Segmentation

Segmentation Data Preprocessing

Same dataset as Customer Lifetime Value, but different feature engineering and selection process, as described below:

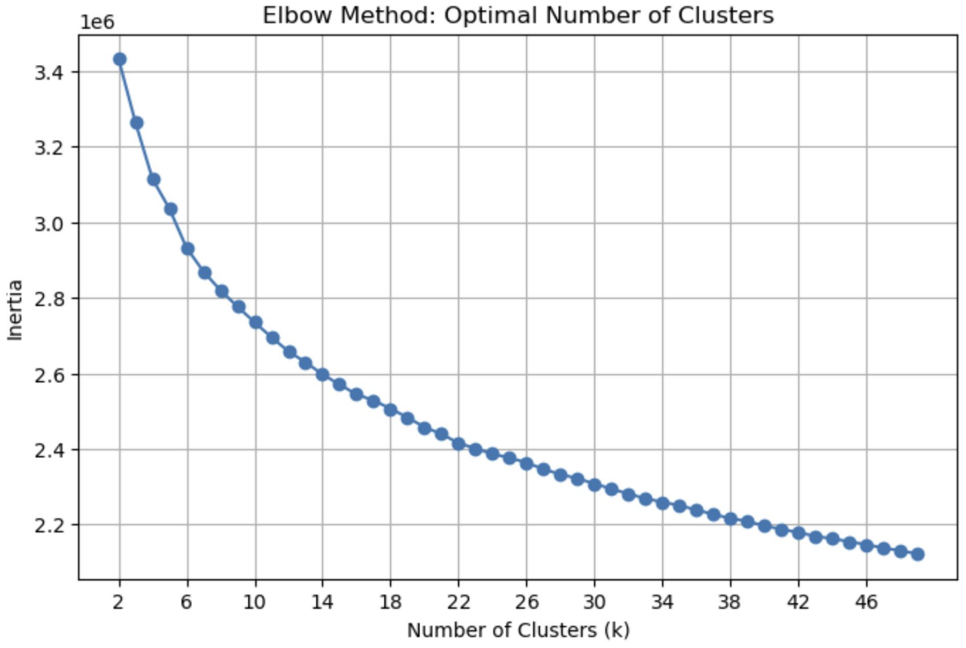
- Created new feature, recency, for days since last transaction
- One Hot Encoded loyalty_program feature (Yes/No)
- Ordinal Encoded the following features:
  - app_usage, social_media_engagement, income_bracket, promotion_effectiveness
- Standard Scale for all numerical features

Column Name	Data Type
recency	int64
frequency	int64
monetary	float64
customer_lifetime_value	float64
customer_lifespan	float64
total_transactions	int64
loyalty_program	object
avg_discount_used	float64
app_usage	object
social_media_engagement	object
income_bracket	object
age	int64
promotion_effectiveness	object
online_purchases	int64
in_store_purchases	int64

Segmentation Experimentation

Initially ran K Means with 5 clusters as a baseline

Used **inertia score** to select the number of clusters via **Elbow Plot**

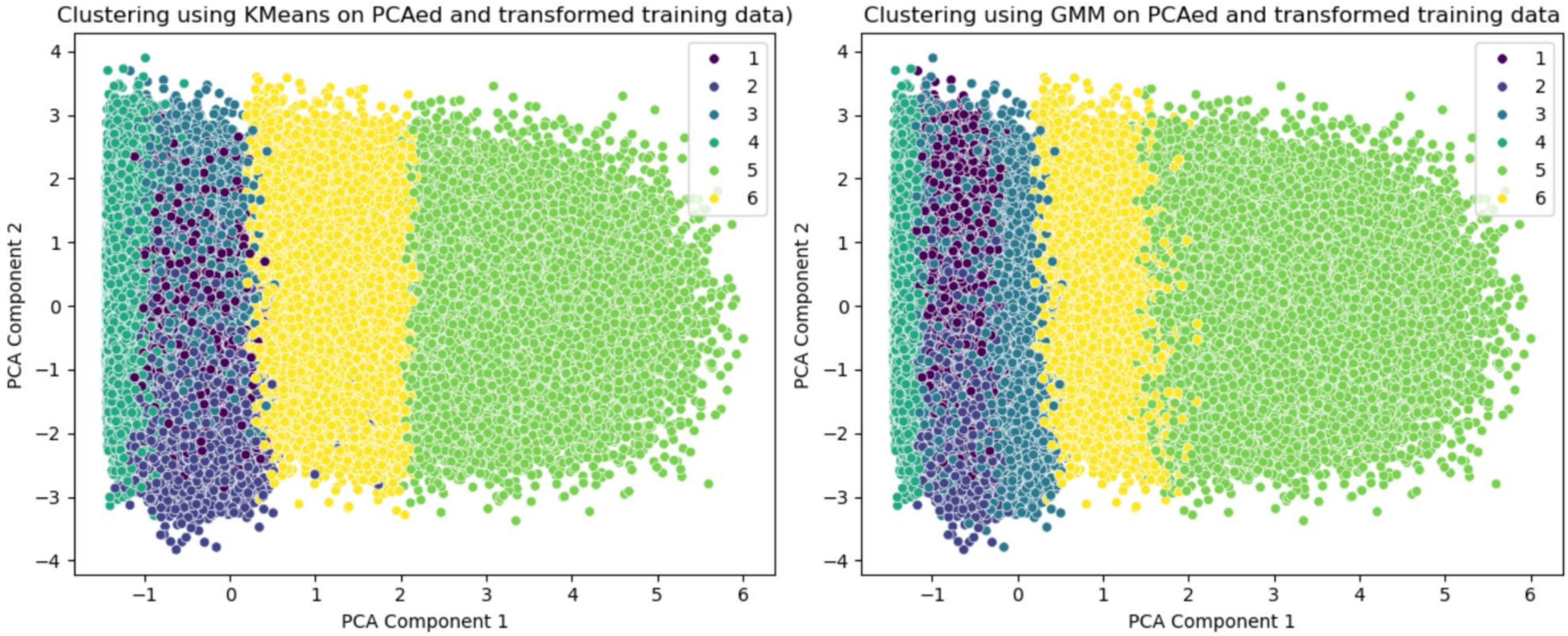


Based on these results, we ran K Means on 6 clusters, which has inertia 2.94. This is 38% higher than lowest inertia (highest accuracy).

To keep number of customer segments meaningful, we accept this accuracy, and chose 6 as an optimal number of clusters.

GMM actually outperformed KMeans

Clusters 1, 2, and 3 that have poor separation in K-Means have better quality cluster separation using GMM



Better machine would run GMM on full data set, w/o PCA. Absent that, we stick with K Means

Silhouette Plot (K Means)

Interpreting the plot

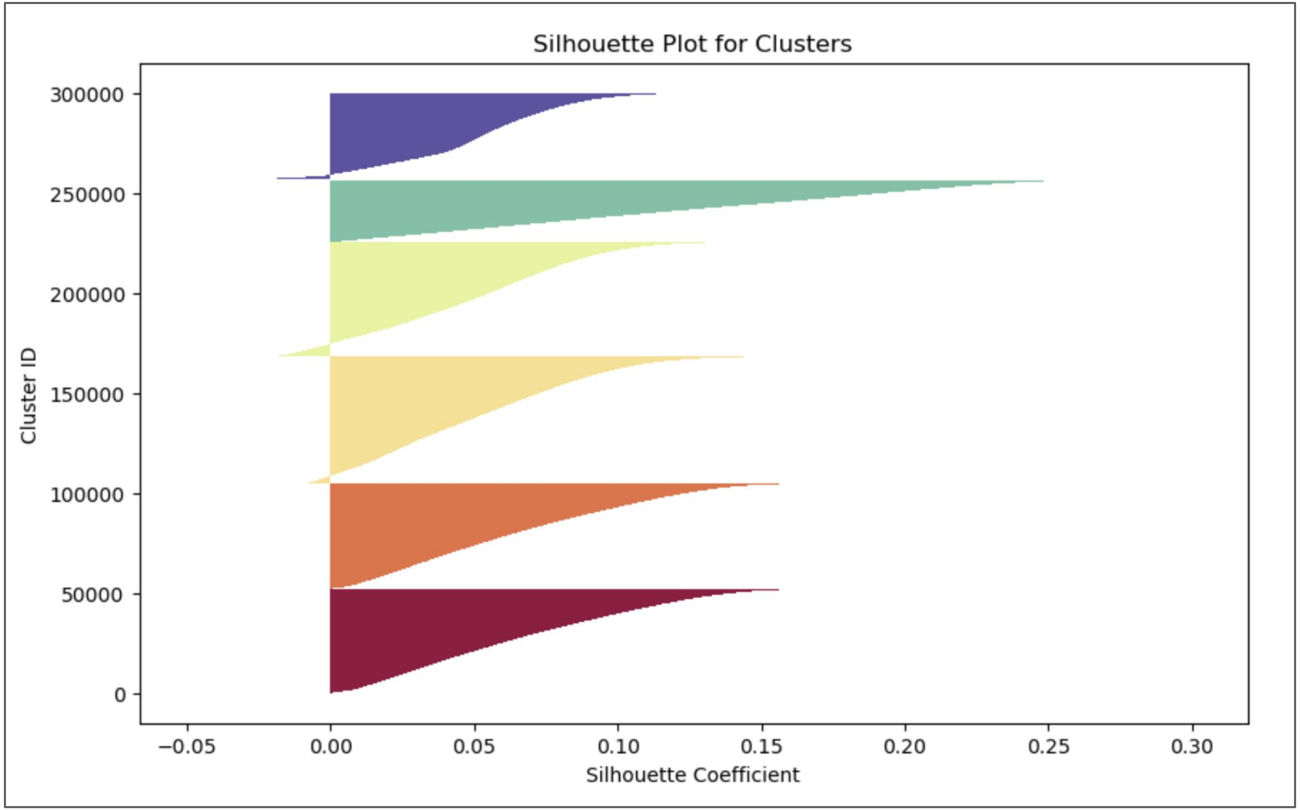
Scores range from -1 to +1

+1: The point is very well clustered, far from neighboring clusters.

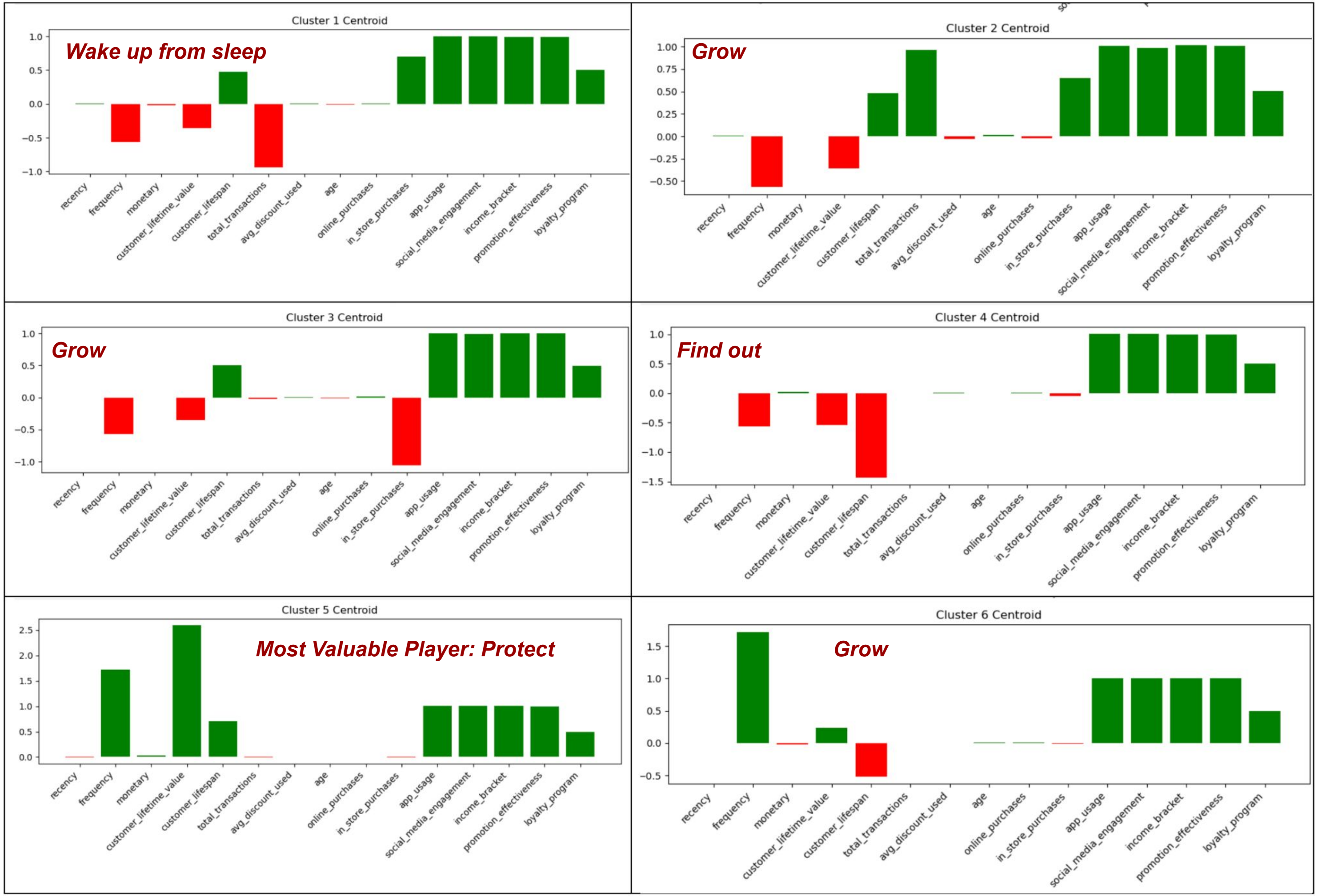
0: The point is on or near the decision boundary between clusters.

-1: The point is likely in the wrong cluster, as it's closer to a neighboring cluster.

As we can see from the plot, most of our points are close to a decision boundary, with only a few likely misclassified.



Customer Segments from K Means



Cluster 1, Segment: **Moderation Molly, Digitally active, Doesn't buy**
Action: Use social media, app to increase online purchases and monetary, frequency values.

Cluster 2, Segment: **High transaction, low frequency buyer.**
Action: Launch promotions to increase frequency (that should drive up monetary scores as customer tends to buy a lot when she makes a purchase).

Cluster 3, Segment: **In-store hating, low lifetime value customer who somehow still sticks around**
Action: Launch promotions (online and digital channels) to increase frequency and monetary scores.

Cluster 4, Segment: **'Mountain to climb' customer segment.**
Most difficult segment: lowest lifespans, low customer lifetime value, low frequency.
Action: Conduct surveys to discover reasons for low traction.

Cluster 5. Segment: **Most Valuable Players Highest Lifetime Value, high frequency customers**
Action: Drive interventions if needed to minimize churn. Launch new campaigns to increase loyalty scores

Cluster 6, Segment: **High frequency, low volume buyers.** Purchases frequently
Action: Use online and digital channels for campaigns to increase retention rates, monetary scores.

Conclusion

Key Results

XXX

Takeaways

XXX

Future Work

XXX

Thank you!