Remove Watermark
Wondershare PDFelement

FINTECH CUSTOMER LIFETIME VALUE DRIVERS

Jonathan Beck

Remove Watermark Wondershare PDFelement

01 BUSINESS PROBLEM & OBJECTIVE

FINTECH

The use of digital tech to improve financial services.

Faster, ubiquitous, and convenient tools for managing money, making payments, lending, and investing (e.g., payment apps, mobile banking services)

CLTV IN FINTECH

- **Core question:**
  - What truly differentiates high-value from low-value customers?
- **Why matters?**
  - Profitability, retention, targeting
- **Dual objective:**
  - Predict High vs Low CLTV (if practical)
  - Identify actionable operational drivers

Remove Watermark Wondershare PDFelement

02 PROJECT GOALS

To understand drivers of high-value customers and apply this knowledge to provide a competitive advantage of the Fintech firm.

01 Identify natural customer segments through unsupervised learning.

02 Characterize behavioral and demographic differences between high-value and low-value customers.

03 Make attempts at developing supervised models that could be deployed to predict high LTV customers and/or their operational drivers.

04 Implement actionable recommendations to provide Fintech firm with a strong financial return based on our discoveries and insights.

Remove Watermark Wondershare PDFelement

03 DATA OVERVIEW & PREPROCESSING

```
$ customer_id               : chr [1:7000] "cust_0000" "cust_0001" "cust_0002" "cust_0003" ...
$ age                       : num [1:7000] 54 67 44 30 58 23 36 54 34 38 ...
$ location                  : chr [1:7000] "Urban" "Suburban" "Urban" "Rural" ...
$ income_level              : chr [1:7000] "Low" "High" "High" "High" ...
$ total_transactions        : num [1:7000] 192 979 329 71 878 741 946 63 297 245 ...
$ avg_transaction_value     : num [1:7000] 184.1 159.9 77.7 180.7 118.8 ...
$ max_transaction_value     : num [1:7000] 662 532 358 196 192 ...
$ min_transaction_value     : num [1:7000] 71.8 24.1 30.2 48 49.9 ...
$ total_spent               : num [1:7000] 35347 156546 25555 12829 104307 ...
$ active_days               : num [1:7000] 140 229 73 299 236 195 88 97 166 296 ...
$ last_transaction_days_ago : num [1:7000] 209 240 21 285 329 93 255 162 303 252 ...
$ loyalty_points_earned     : num [1:7000] 2114 2960 3170 4756 1992 ...
$ referral_count            : num [1:7000] 25 20 0 35 18 5 46 17 36 7 ...
$ cashback_received         : num [1:7000] 2224 4027 1441 4366 4162 ...
$ app_usage_frequency       : chr [1:7000] "Monthly" "Monthly" "Monthly" "Weekly" ...
$ preferred_payment_method  : chr [1:7000] "Debit Card" "UPI" "Debit Card" "Wallet Balance" ...
$ support_tickets_raised    : num [1:7000] 3 17 11 6 18 6 1 13 3 10 ...
$ issue_resolution_time     : num [1:7000] 61.6 60.4 45.3 22 20.6 ...
$ customer_satisfaction_score: num [1:7000] 1 8 4 1 5 10 8 8 10 3 ...
$ ltv                       : num [1:7000] 327955 1437053 241939 128460 956951 ...
```

- **7,000 customers, 20 attributes**
  - No null attributes
  - Feature types:
    - Demographic
    - Behavioral
    - Transactional
    - Service interactions

- **Converted INR → USD**

- **General Preprocessing**
  - One-hot encoding
  - Scaling for clustering
  - Removal of identifiers

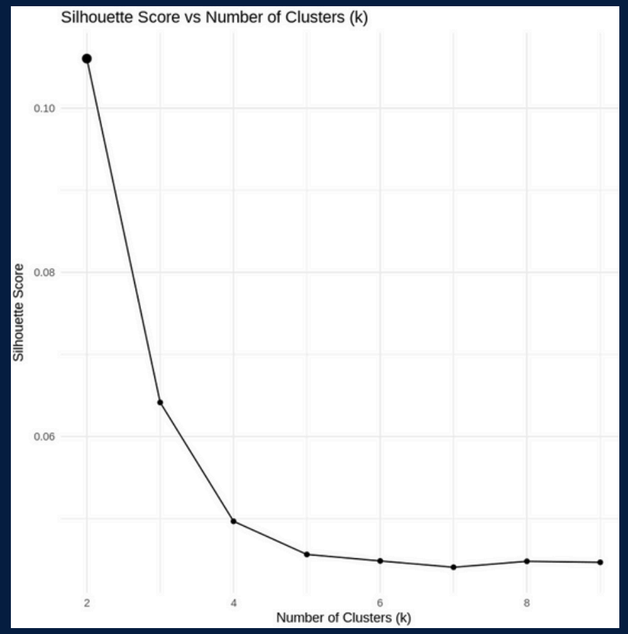
Remove Watermark Wondershare PDFelement

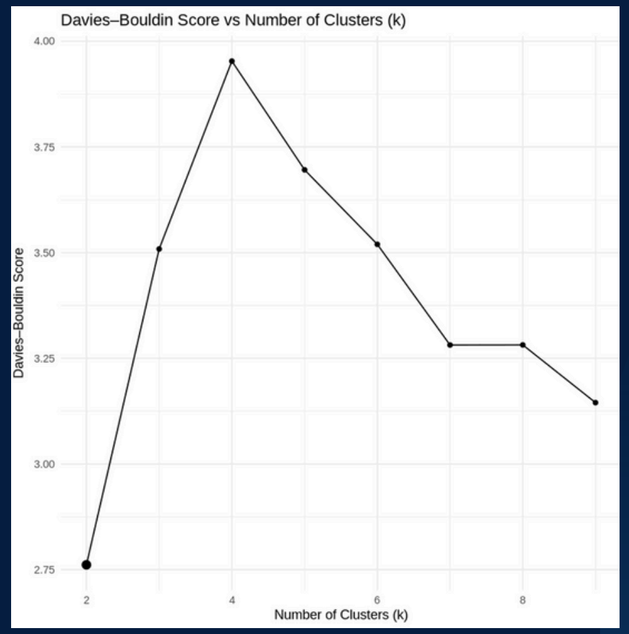
04 UNSUPERSIVED LEARNING: CUSTOMER SEGMENTATION

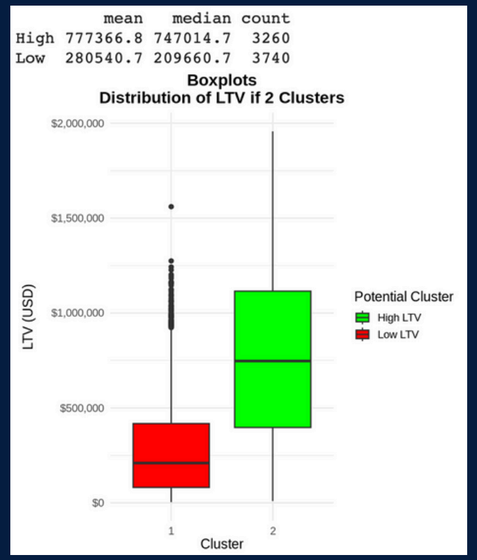
- K-Means Clustering (k = 2-9 evaluated)
  - Evalutation Metrics used:
    - Silhouette Score - cluster compactness
    - Davies-Bouldin Index - distance between clusters
  - Best solution: 2 clusters

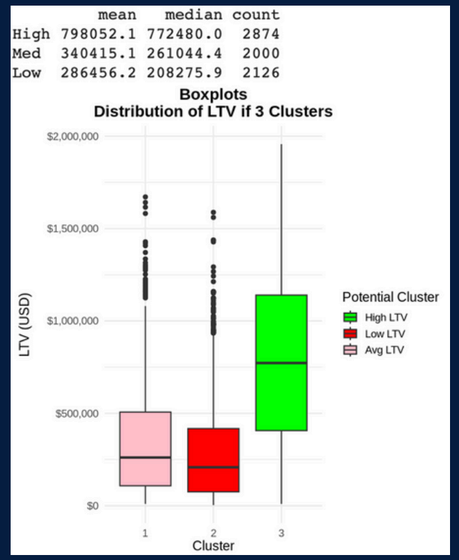
Key Insight:

- Natural segmentation aligns strongly with CLTV differences

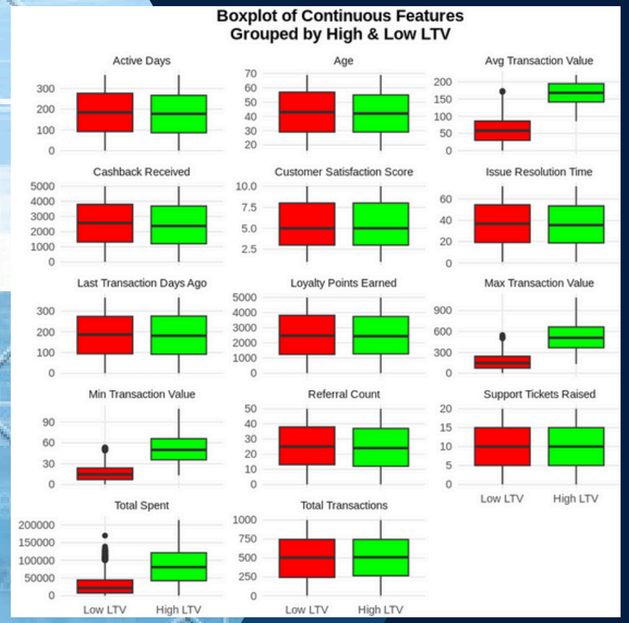


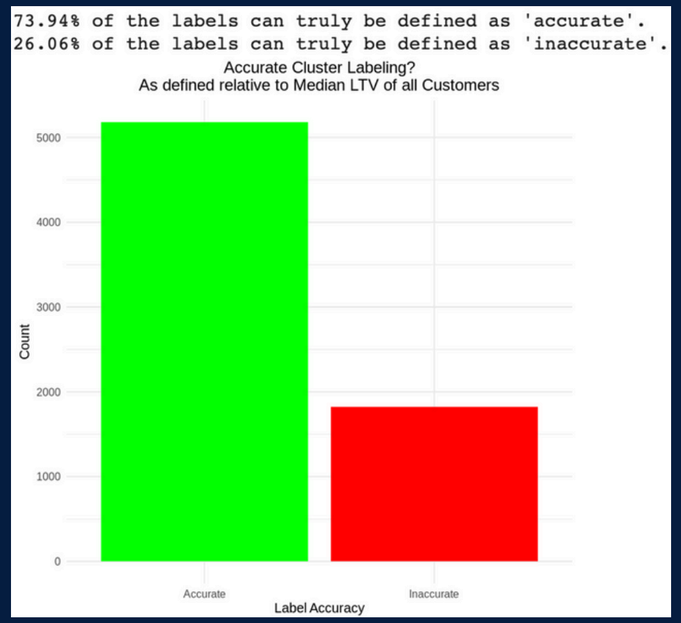






05 CLUSTER INTERPRETATION & VALIDATION





Validation against median CLTV:

- ~ 74% of customers in alignment with true labels

Behavioral differences:

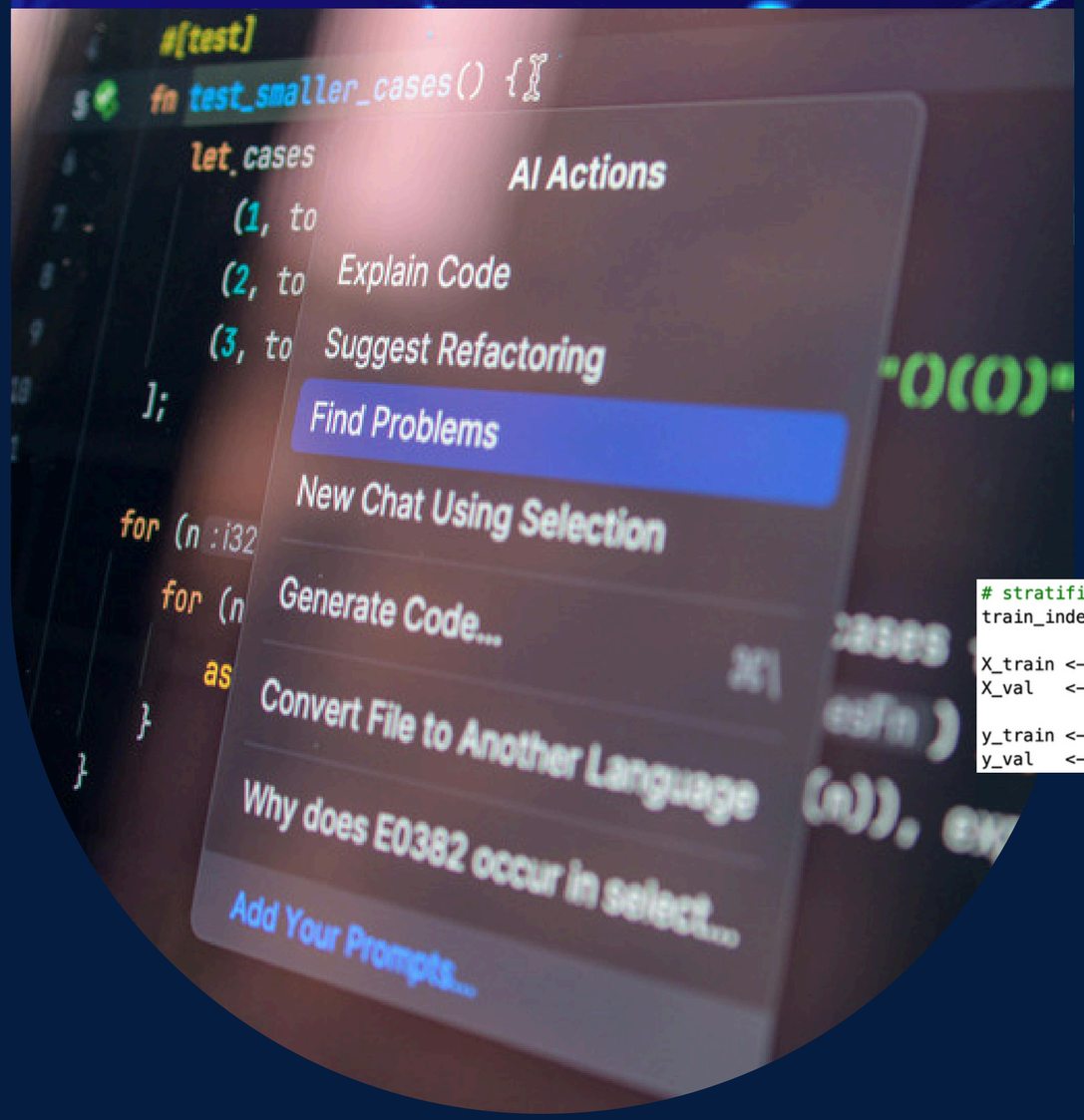
- Avg, Max, & Min Transaction values
- Total Spend
- Transaction frequency

← Key Takeaway:

- CLTV is overwhelmingly driven by monetary behavior

Remove Watermark Wondershare PDFelement

Remove Watermark Wondershare PDFelement



06 TRANSITION TO SUPERVISED LEARNING

```
# Creating binary target ltv_01
# (1 = at/above median LTV, 0 = below)
data_for_supervised <- data_for_supervised %>%
  mutate(
    ltv_01 = if_else(
      ltv >= median(ltv, na.rm = TRUE),
      1L, 0L
    )
  )
```

- Binary target: High vs Low CLTV (median split)

```
50.0% are High LTV
50.0% are Low LTV
```

- Balanced classes (50/50)

```
# stratified split: 90% training, 10% validation
train_index <- createDataPartition(y_class, p = 0.9, list = FALSE)

X_train <- X[train_index, ]
X_val   <- X[-train_index, ]

y_train <- y_class[train_index]
y_val   <- y_class[-train_index]
```

- 90/10 train-test split

Initial modeling task: Decision Tree to gauge Feature Importance →

Remove Watermark Wondershare PDFelement

07 FEATURE IMPORTANCE: THE PROXY PROBLEM

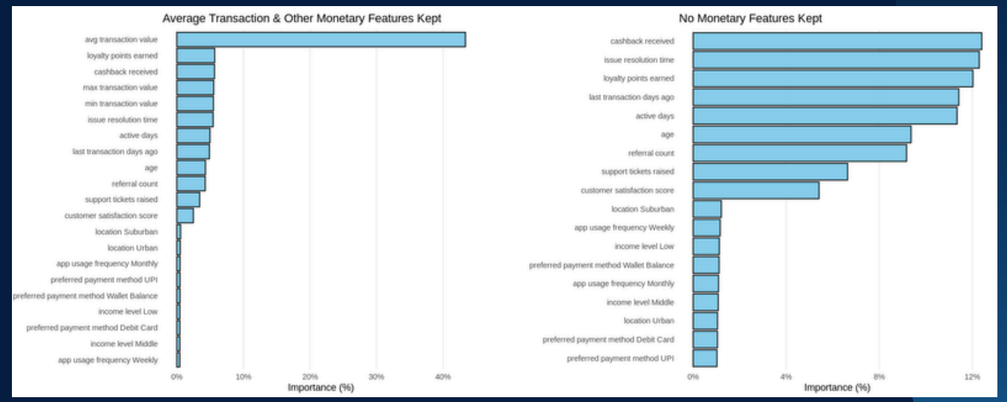
	feature	importance
	<chr>	<dbl>
total_spent	total_spent	3112.1526
avg_transaction_value	avg_transaction_value	1698.3489
total_transactions	total_transactions	1583.3411
max_transaction_value	max_transaction_value	1464.3674
min_transaction_value	min_transaction_value	1457.4273
support_tickets_raised	support_tickets_raised	100.1362
age	age	0.0000

1) Total Spent & Avg Tranaction Value dominate

- Dropped Total Spent & Total Transactions as candidate regressors
  - Allowed us to see influence of more descriptive, demographic features

3) **Decision Tree fit to Training and Validation Splits using No Monetary Features:** Stratified K (5) - Fold CV over Grid

maxdepth	minsplit	minbucket	mean_accuracy
<dbl>	<dbl>	<dbl>	<dbl>
8	2	5	0.5069841



2) Key Hypothesis: Monetary variables = near-perfect proxies for LTV of customer

4) Our best model was evaluated on the holdout validation data split using a confusion matrix

- The results show an overall accuracy of ~47.57%, precision of ~48.44%, and recall of ~75.43%.,
- These performance metrics provide early indication that the removal of monetary features induces an undeployable predictive model.

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0  69  86
         1 281 264

               Accuracy : 0.4757
                 95% CI : (0.4382, 0.5135)
    No Information Rate : 0.5
    P-Value [Acc > NIR] : 0.9071

                  Kappa : -0.0486
```

Remove Watermark Wondershare PDFelement

08 MODELING W/ MONETARY FEATURES: WHY HIGH ACCURACY != BUSINESS VALUE

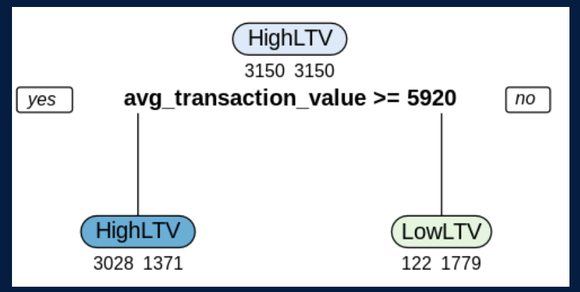
Feature	Coefficient
<chr>	<dbl>
avg_transaction_value	-2.6099187
intercept	1.4768764
min_transaction_value	-0.3064622
max_transaction_value	-0.2854515
age	0.0000000

```
          Reference
Prediction HighLTV LowLTV
   HighLTV     260     90
   LowLTV       90    260

Accuracy: 0.7429
Sensitivity (Recall): 0.7429
Specificity: 0.7429
Precision: 0.7429
F1: 0.7429
```

Elastic Net Regularized Linear Regression

- Penalty essentially removes influence of any nonmonetary features in classifying custmers as High/Low LTV



Confusion Matrix at a glance

- High Accuracy

But:

- Predicts High/Low CLTV at perfect 50/50 split...
  - Predictive performance alone is misleading
  - No strategic insight
  - Not practical

Remove Watermark Wondershare PDFelement

09

REMOVING MONETARY FEATURES: WHAT HAPPENS

```r
# 75th percentile (top 25%) of LTV
top_25 <- quantile(data$ltv, probs = 0.75, na.rm = TRUE)

# Binary label: 1 = in top 25% of LTV, 0 = otherwise
ltv_high <- ifelse(data$ltv >= top_25, 1, 0)
```

Redefined definition of High CLTV:

- Top 25%
- Class imbalance introduced → 24% customers in each split High LTV

Ridge Regression Logistic Regression

Feature	Coefficient
<chr>	<chr>
intercept	1.09438395
support_tickets_raised	0.00000000
referral_count	0.00000000
last_transaction_days_ago	0.00000000
customer_satisfaction_score	-0.00000000

All coefficients shrunk to almost zero entirely
Intercept-only model

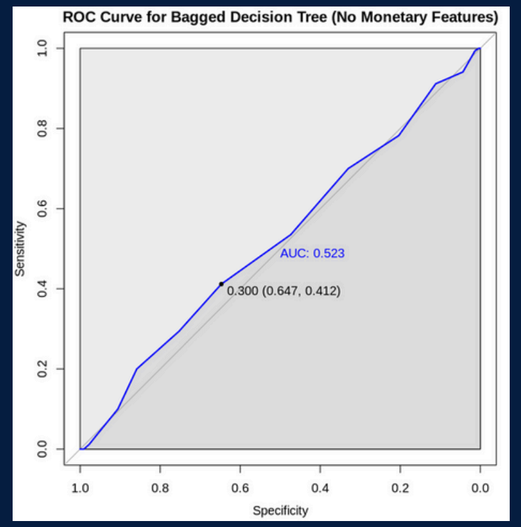
```
          Reference
Prediction HighLTV LowLTV
   HighLTV       0      0
   LowLTV      170    530

Accuracy: 0.7571
Sensitivity (Recall): 0
Specificity: 1
Precision: NA
F1: NA
```

Bagging Ensemble of 100 Base Estimators

```
          Reference
Prediction HighLTV LowLTV
   HighLTV       2     12
   LowLTV      168    518

Accuracy: 0.7429
Sensitivity (Recall): 0.0118
Specificity: 0.9774
Precision: 0.1429
F1: 0.0217
```



Critical Insights

- High accuracy driven buy majority-class bias
- Non-monetary features alone cannot predict CLTV directly

Remove Watermark Wondershare PDFelement

STRATEGIC PIVOT IN MODELING OBJECTIVE

Shift from "predict CLTV" → "predict drivers of CLTV"

- Rather than attempting to predict Hight LTV directly, we've chosen to examine how non-monetary features can be leveraged to predict three foundational drivers of CLTV :

01 Support Tickets Raised

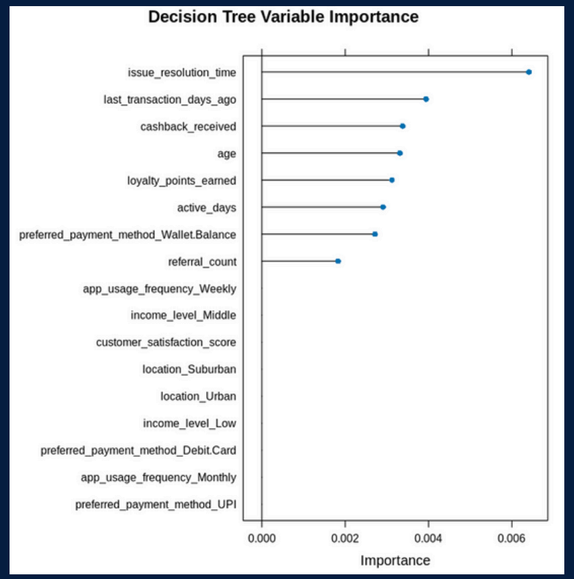
02 Customer Satisfaction

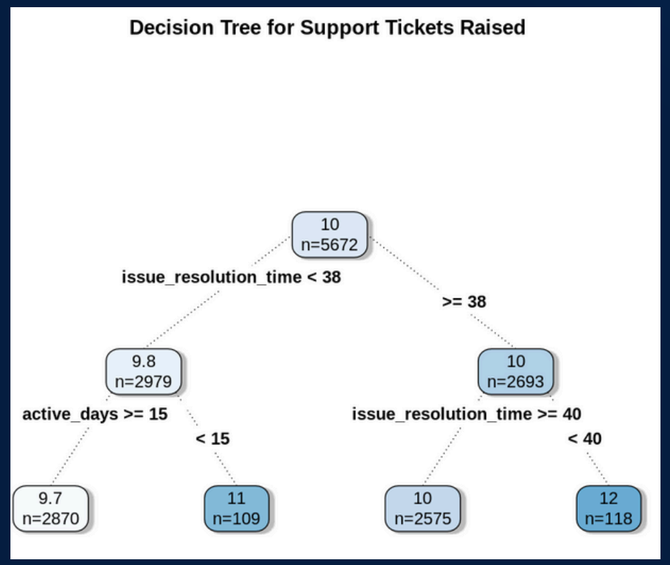
03 Issue Resolution Time

Key Reasoning → These are operationally controllable

Remove Watermark Wondershare PDFelement

11 SUPERVISED OPERATIONAL MODEL #1 - SUPPORT TICKETS





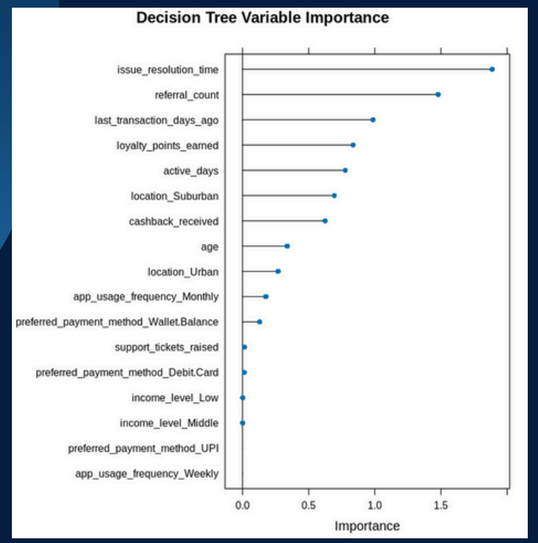
Hypothesis: Slow resolution times escalates customer frustration

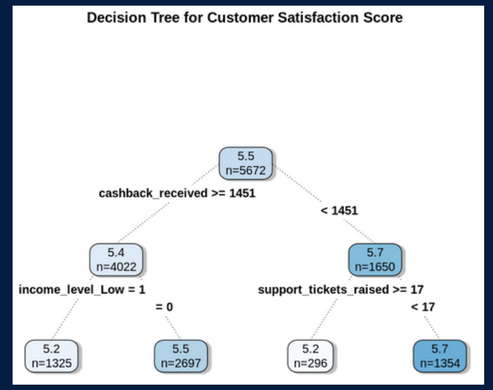
Frustruated customers → incentive to escalate discrepancies → higher volume of support tickets

Decision Tree Interpretation:

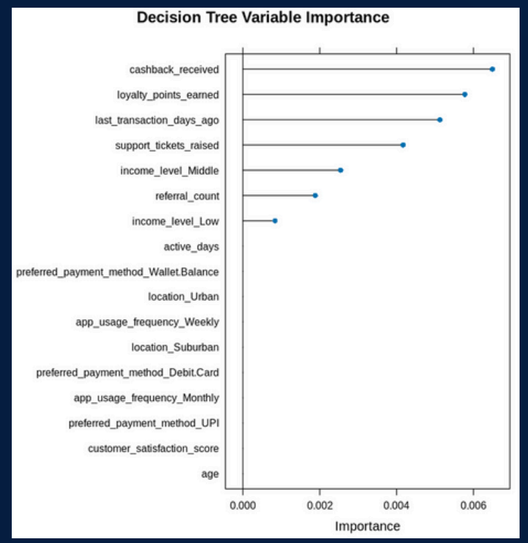
- Customers who have issues resolves in less than 38 days and spend more time on app (>=15 days) raise less support tickets on average

SUPERVISED OPERATIONAL MODEL #2 AND #3 - CUSTOMER SATISFACTION AND ISSUE RESOLUTION





	Coefficient
support_tickets_raised	0.084960066
preferred_payment_method_UPI	-0.068711295
cashback_received	-0.061643871
app_usage_frequency_Monthly	0.056947008
referral_count	-0.051915493
app_usage_frequency_Weekly	0.047281467
loyalty_points_earned	-0.036870483
income_level_Low	-0.035621101
last_transaction_days_ago	0.024767267
location_Suburban	-0.021939838
customer_satisfaction_score	0.019000320
age	-0.018699055
location_Urban	0.015107390
preferred_payment_method_Wallet.Balance	-0.014379929
income_level_Middle	-0.008908482
active_days	0.007846215
preferred_payment_method_Debit.Card	0.005104719



Customer Satisfaction Score:

- Again, driven by Issue Resolution Time
- When forced depth to 2, frequency of support tickets clearly matter

Issue Resolution Time:

- Strongly correlated with volume of Support Tickets Raised

Remove Watermark Wondershare PDFelement

Remove Watermark Wondershare PDFelement

13

CROSS-MODEL CHAIN OF EVENTS, NODE ANALYSIS, FINANCIAL IMPACT & BUSINESS RECOMMENDATIONS

1) Issue Resolution Time → Support Tickets Raised

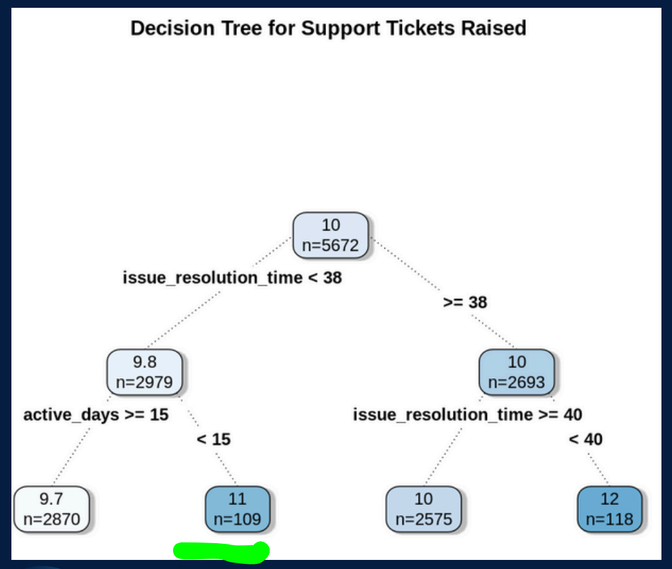
- Faster resolutions == less disgruntled customers raising issues

2) Issue Resolution Time → Customer Satisfaction Score

- Lengthier resolution times == less satisfied customers

3) Support Tickets Raised → Customer Satisfaction Score

- More tickets being raised == less satisfied customers



Node	Avg_Total_Spent	Avg_Avg_Transaction_Value	Avg_LTV	Count
<int>	<dbl>	<dbl>	<dbl>	<int>
3	4990092.	9827.	510630.	2870
4	5442271.	10284.	556812.	109
6	4977441.	9951.	509538.	2575
7	5328203.	10033.	544445.	118

Node-level LTV differences:

- $33K - $47K uplift per customer if can move customers to nodes with higher on average spending behaviors
  - Example: Moving 1,000 customers from Node 3 to Node 4 → ~$46M value

Recommendations to do so:

- Improve first-contact resolution
- Increase staffing
- Invest in technology infrastructure
- Enhanced employee training and onboarding procedures
- Investment in enhaced user experience (downtime issues, app layout, etc.)

Above tree nodes reveal a major insight: customers with lower resolution times and fewer days active, on average, have the highest LTV, total spend, and average transaction value. Notably, customers with lower resolution times and more active days reveal the lowest average monetary metrics. This pattern suggests that even when issues are resolved relatively quickly, extended active engagement with the platform correlates with diminishing economic value, perhaps due to ongoing unresolved problems.

Remove Watermark Wondershare PDFelement

CONCLUSION & STRATEGIC IMPLICATIONS

CLTV PREDICTIONS ≠ CLTV UNDERSTANDING

- Monetary proxies used as regressors inflate accuracy but kill insight
- Operational excellence is the true lever
- Data investment priorities:
  - Better non-monetary signals
  - Earlier lifecycle indicators

Improving support operations is the most reliable path to increasing customer lifetime value.

THANK YOU

Remove Watermark
Wondershare PDFelement