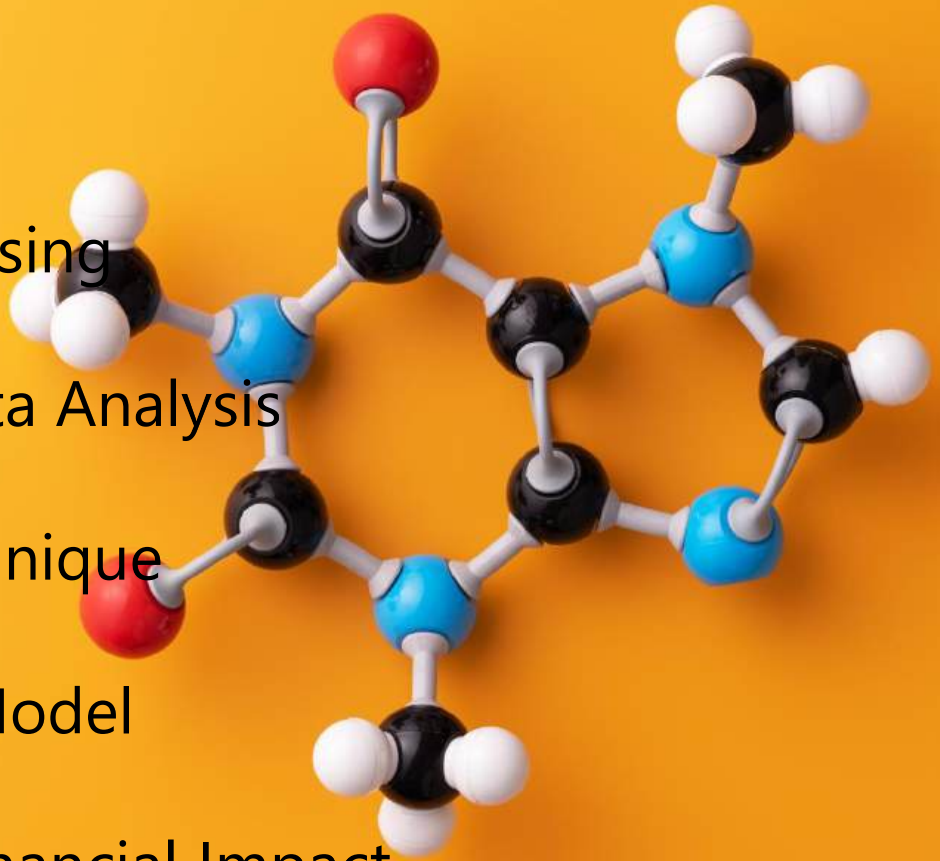
Data-Driven Insights in Auto Insurance: Predicting Claims and Enhancing Customer Retention.

Kaushik Das Lnu.
Srikanth Goud Durgam.
Mayen Udoffia.
Professor Wei Xiong.

Mercer University, 2024

Overview

- Introduction
- Objective
- Data Preprocessing
- Exploratory Data Analysis
- Clustering Technique
- Classification Model
- Strategies & Financial Impact
- Business Implication



Introduction

- In today's competitive market, understanding customer behavior and predicting needs are vital for building loyalty, reducing churn, and maximizing customer lifetime value. However, traditional segmentation approaches rely on static groupings that fail to adapt to the evolving preferences and behaviors of new customers. Without a dynamic, data-driven method to classify these new customers into meaningful segments, businesses risk missing opportunities for personalized engagement and targeted marketing.

- This project addresses this challenge by developing a predictive segmentation model that combines clustering and classification techniques to dynamically assign new customers to predefined clusters. By proactively tailoring engagement strategies, businesses can enhance customer satisfaction, improve retention rates, and maximize lifetime value, gaining a critical edge in a competitive marketplace.

Business Question

"How can predictive models and customer segmentation be used to forecast insurance claims and analyze policyholder behavior, to enhance risk management, reduce claim-related costs, and improve customer retention in the insurance industry?"

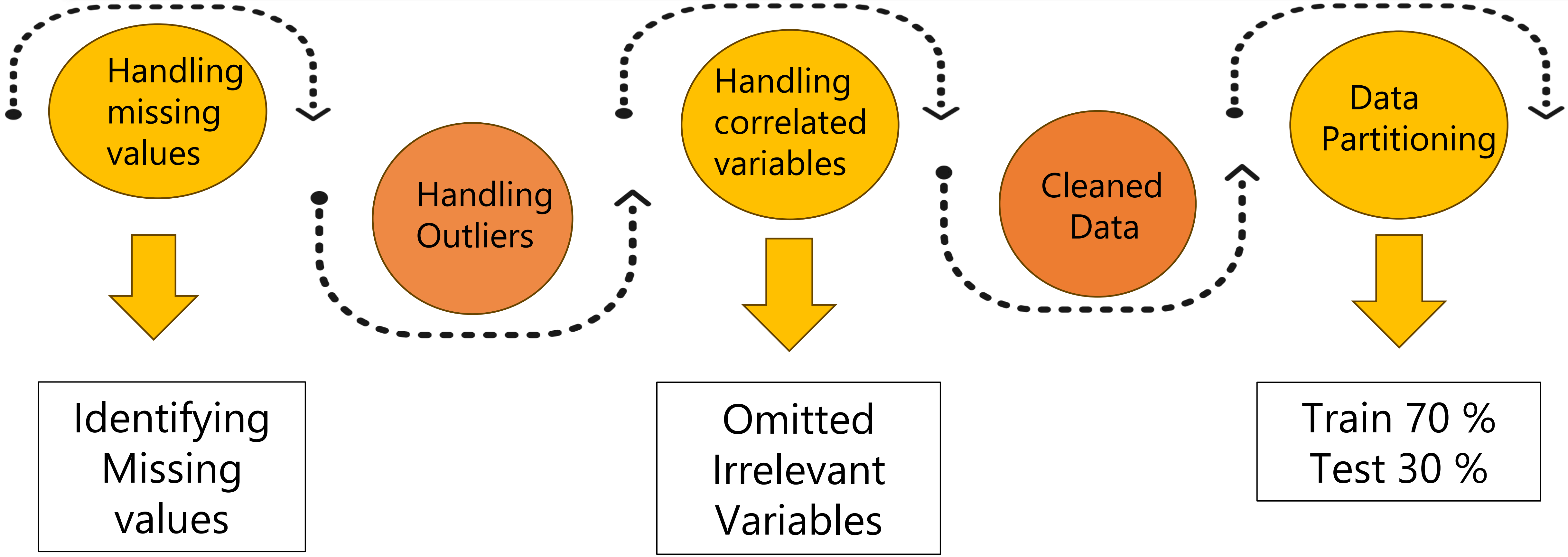
Objectives

- Identify distinct customer segments through behavioral clustering.
- Develop predictive models to dynamically assign new customers to predefined segments.
- Enable businesses to implement tailored engagement strategies to improve customer satisfaction, retention, and lifetime value.
- Provide actionable insights for reducing claim-related costs and enhancing risk management.

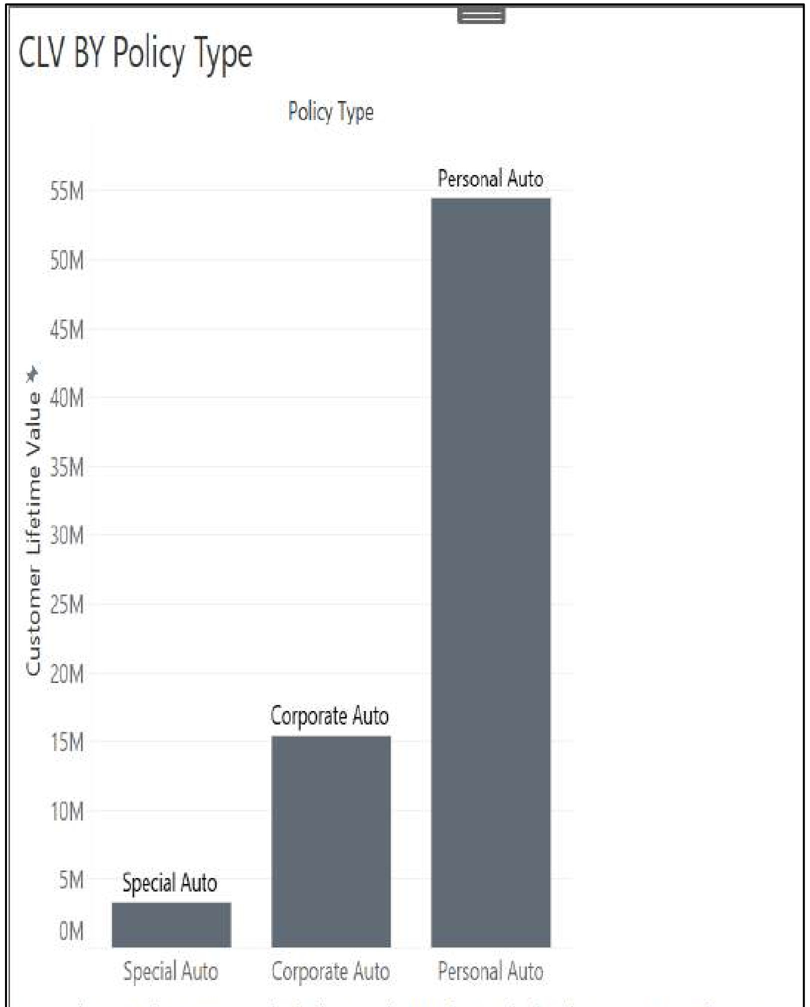
Data Dictionary

Category	Attributes
Customer Information	Customer, Gender, Marital Status, Education, Employment Status, Income
Policy Information	Policy, policy Type, Renew Offer Type, Sales Channel, Number of policies, Number of Complaints
Vehicle Information	Vehicle Class, Vehicle Size
Coverage Information	Coverage, Monthly Premium Auto, Effective To Date
Claim Information	Total Claim Amount, Months Since Last Claim, Months Since policy Inception
Derived Metrics	Customer Lifetime Value, Response
Geographical Info	State, Locational Code

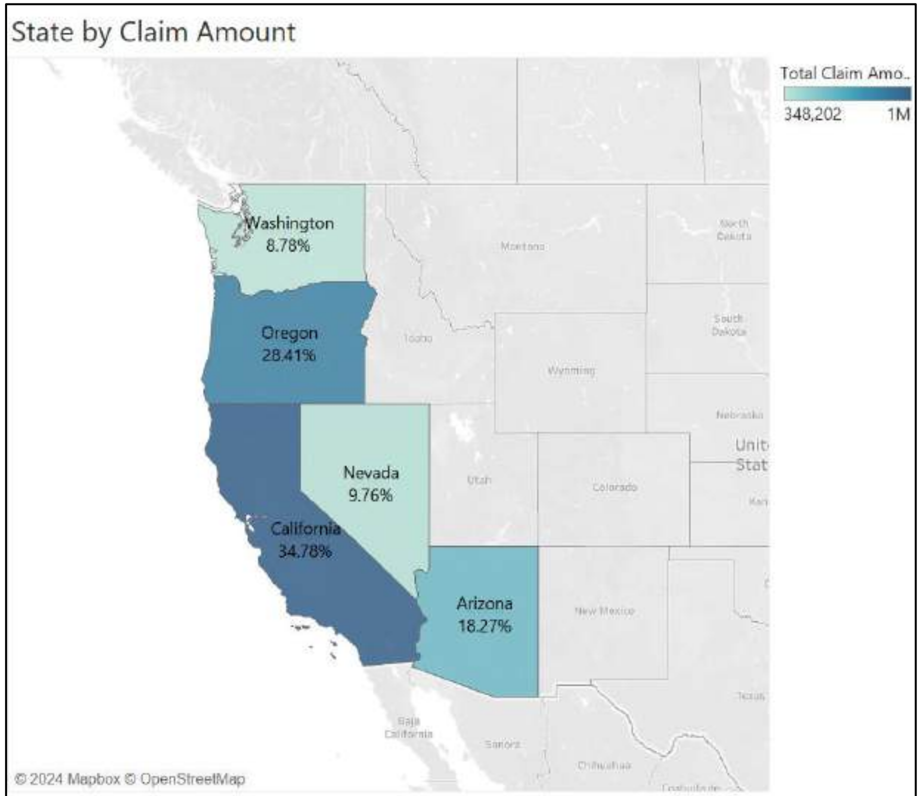
Data Preprocessing



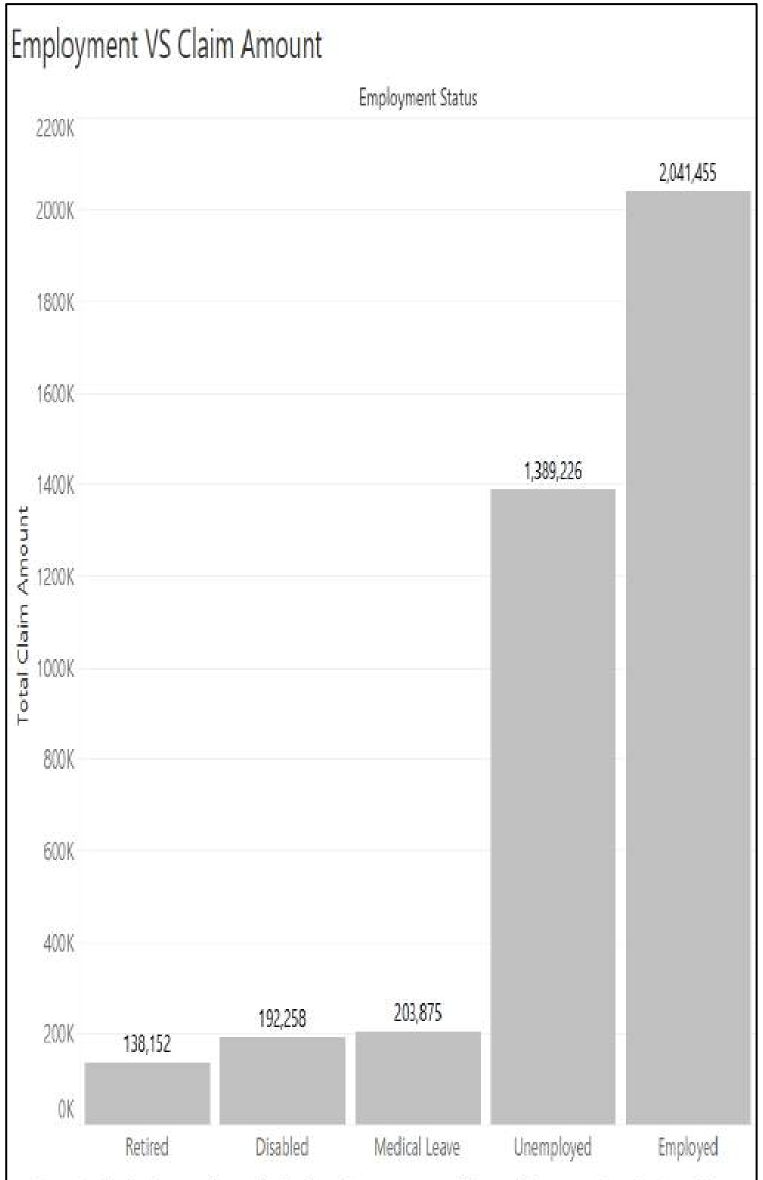
Insights



'Personal Auto' policies generate the highest total CLV, indicating high-value customers in this category.

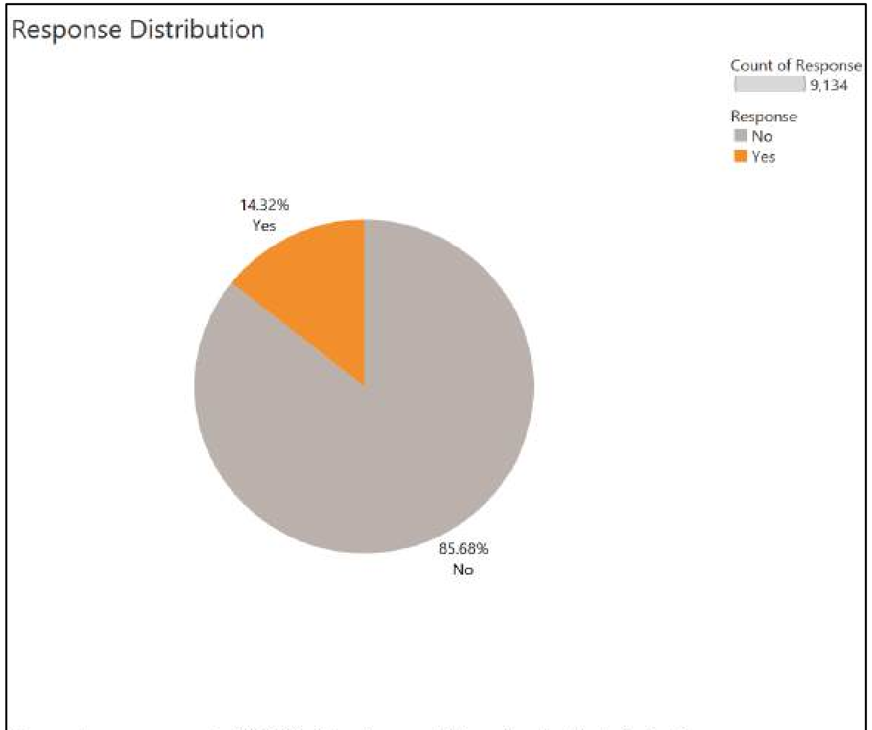


Highlights regional trends in insurance claims distribution, with California and Oregon showing the highest claim amounts

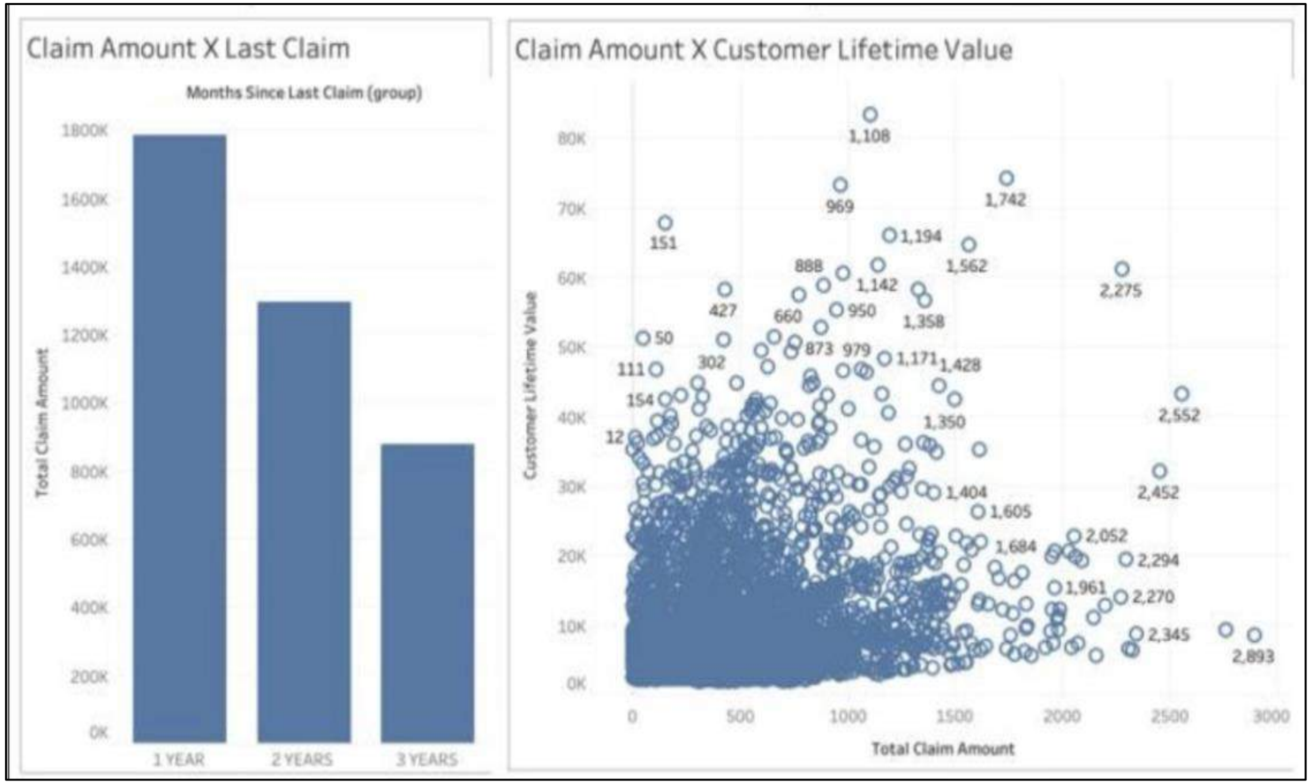


Employed individuals contribute the highest claim amounts, followed by unemployed individual

Insights



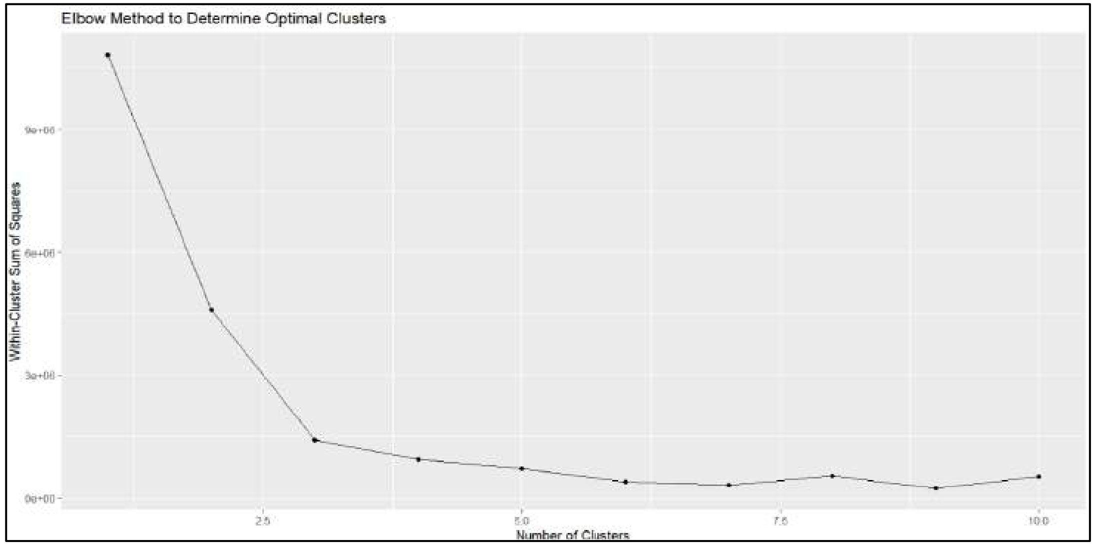
Shows a low response rate (85.68% did not respond) to outreach efforts, indicating room for improvement in marketing strategies.

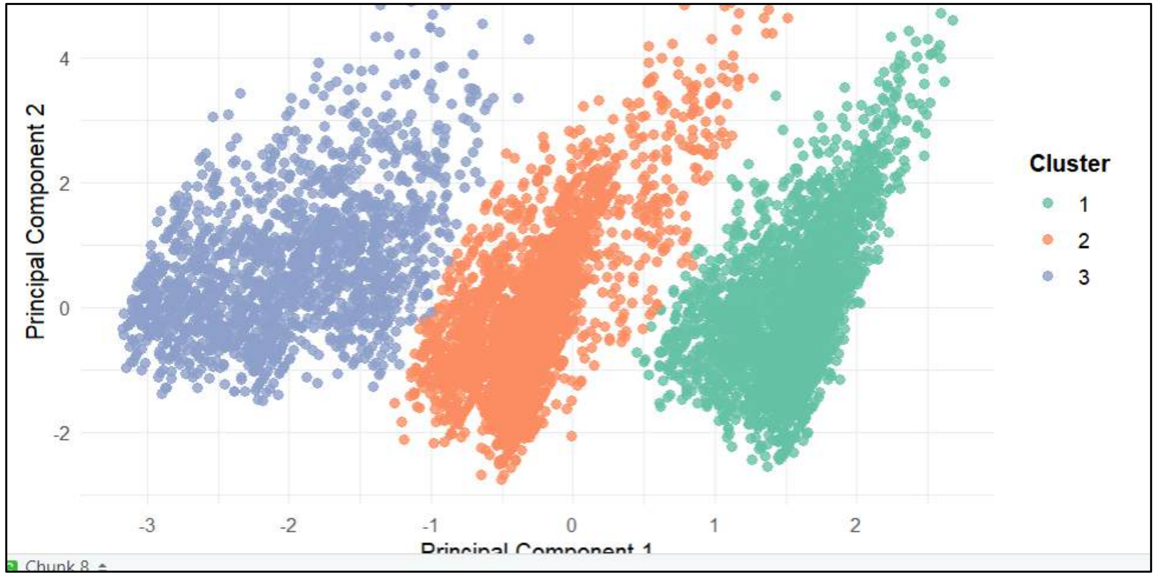


# Customer Count	# Customer Lifetime Value	# Total Claim Amount
9,134	73,117,126	3,960,710

Model #1
Clustering for Customer Segmentation

- To identify distinct customer groups based on **behavioral patterns.**



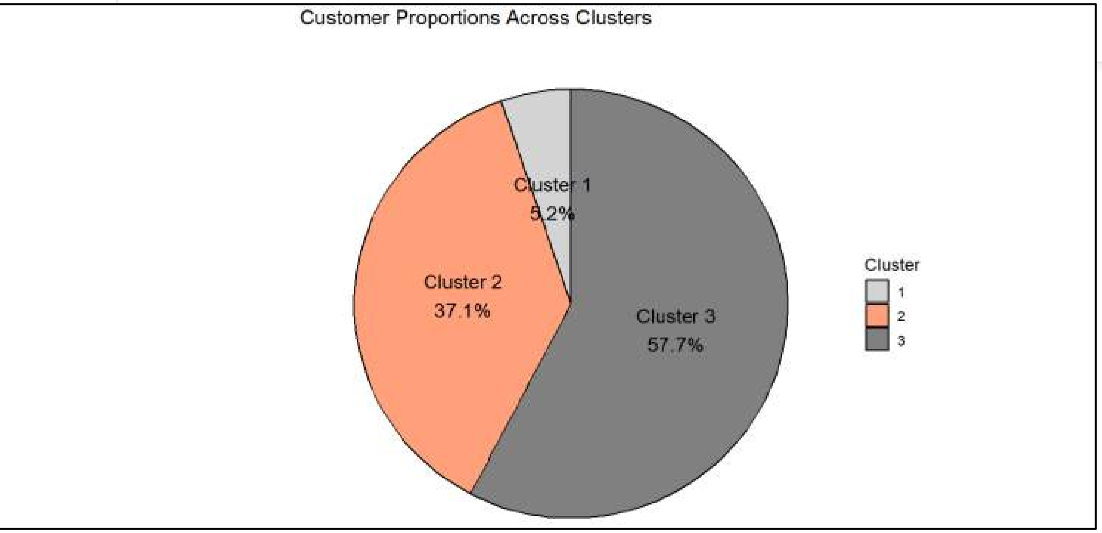


Variables Used:

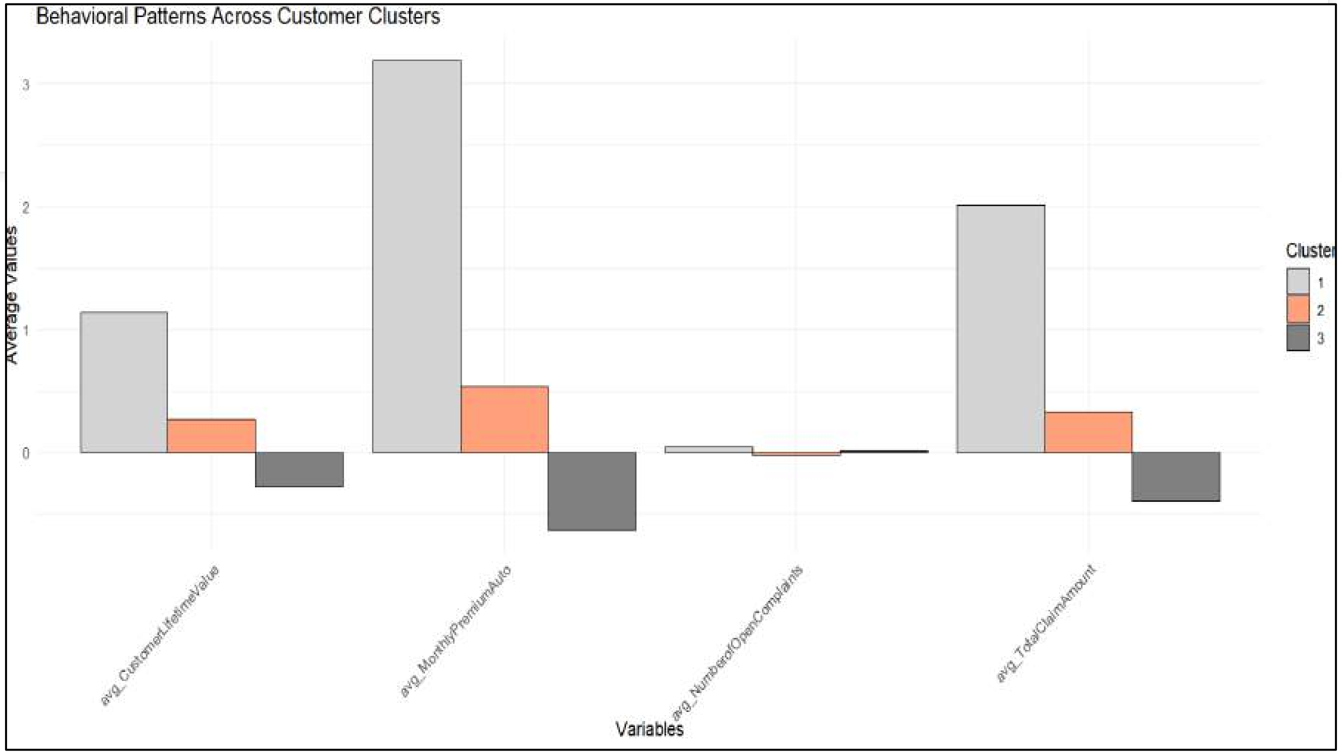
- Customer Lifetime Value
- Total Claim Amount
- Monthly Premium Auto
- Number of Open Complaints

- **Method Used: Kmeans & Silhouette**
- **Number of Clusters = 3**

Cluster Insights



Cluster 3 has the largest proportion of customers (57.7%), followed by Cluster 2 (37.1%) and Cluster 1 (5.2%).



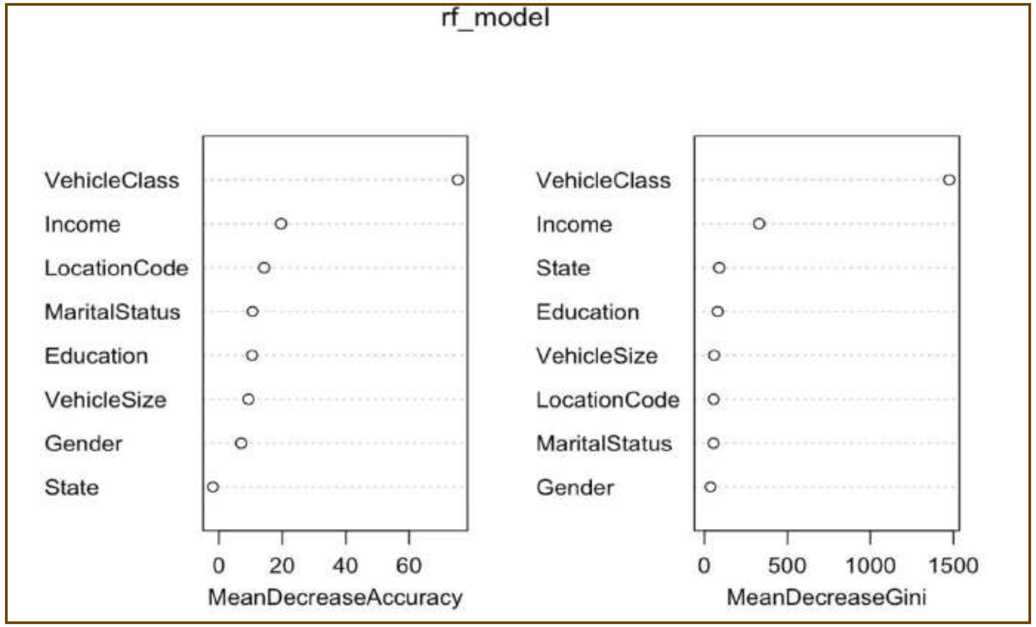
Average behavioral patterns across customer clusters

Cluster 1(**Gold Tier):** High-Value Loyal Customers, high monthly premiums, low complaints.
Cluster 2(**Silver Tier):** Risk-Prone Claimants: customers with moderate premiums but high claim activity, requiring risk management strategies.
Cluster 3(**Bronze Tier)** Cost-Conscious, low-Value Customers ,low premiums, low claims, and minimal engagement.

Model #2
Predictive Customer Segmentation

- **Enables immediate engagement with new customers.**

  - Vehicle type may predict customer risk
  - Income indicates purchasing power and policy preferences.
  - State of residence affects risk exposure and lifestyle.
  - Education level influences decision-making and insurance interest.



Accuracy - 84.4%

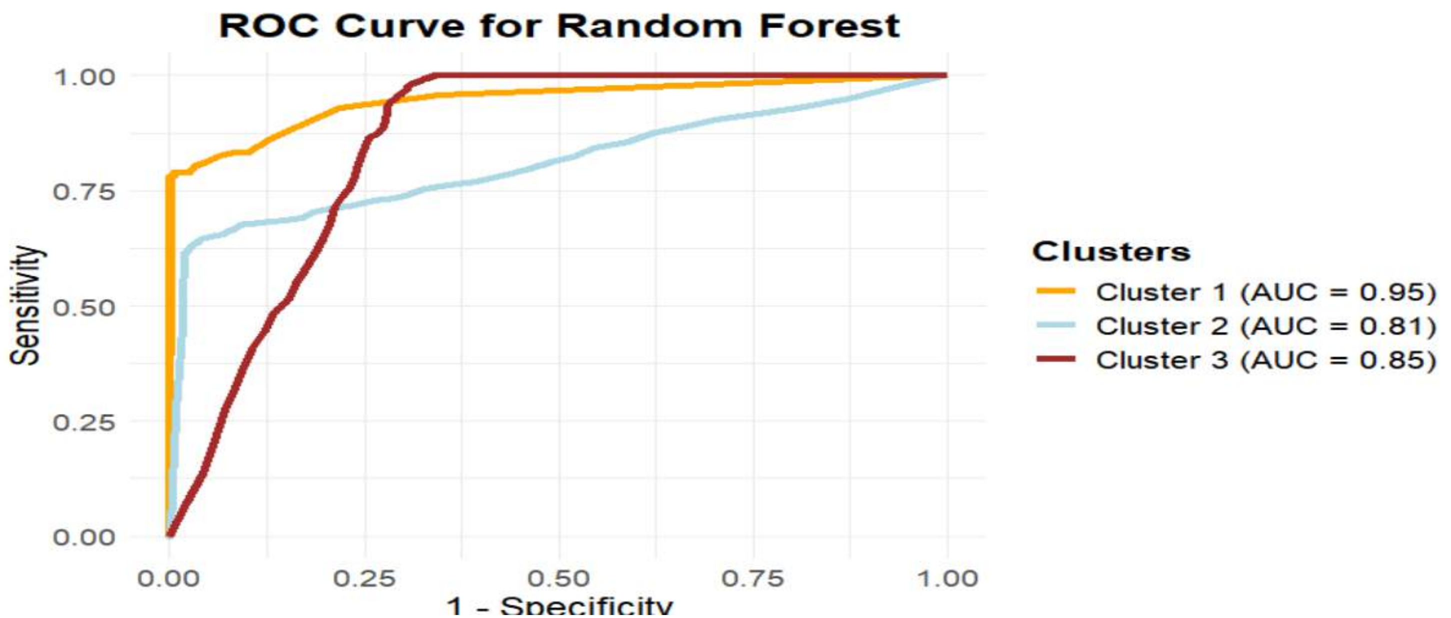
Model Insights: Random Forest

Enables immediate engagement with new customers.

•Class 1 demonstrates the strongest model performance with an AUC of 0.95.

•Class 3 achieves moderate performance with an AUC of 0.85

•Class 2 has the lowest AUC of 0.81

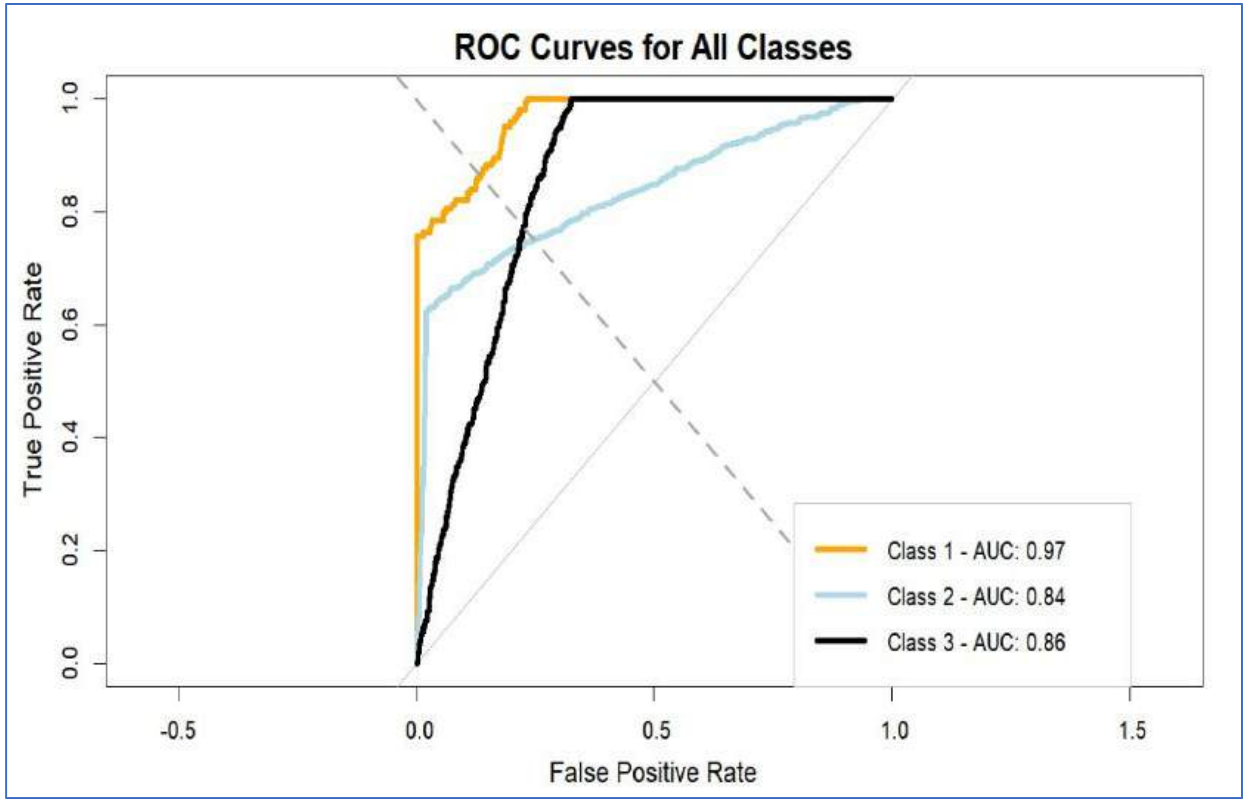


Model #3
Predictive Customer Segmentation: SVM

- High AUC for Cluster 1 indicates strong predictive capability for this segment, enabling more confident targeting

- SVM achieves slightly better AUC for the clusters compared to the Random Forest model

Accuracy: 85%



Strategies & Financial Impact

Goal: Optimize profitability with tailored strategies for each customer cluster

- Gold Tier: Loyalty Programs, Luxury Add-Ons, Fraud Prevention – **Net Impact: +$1.3M annually.**
- Silver Tier: Telematics, Risk-Based Pricing, Education Campaigns – **Net Impact: +$5.8M annually.**
- Bronze Tier: Automated Processes, Upsell Entry-Level Add-ons, Longevity Rewards – **Net Impact: +$2.3M annually.**

Key Assumptions:

Retention programs increase retention by **10%-12%** (industry benchmark)

- Telematics reduces claims by **15%** (based on real-world UBI outcomes).
- Fraud detection saves **8% of payouts**.
- Upsell campaigns achieve **10%-15% adoption rates**.
- Process automation reduces admin costs by **25%**.

References:
Kurylowicz, M. (2016). *Usage-based insurance: The impact of telematics on the insurance industry*. Polish Insurance Association. Retrieved from https://piu.org.pl/wp-content/uploads/2017/05/WU-2016-04-09-Kurylowicz-en.pdf

Business Implication

Gold Tier : +$301,843 annually.

• Focus: Loyalty Programs, Luxury Add-Ons, Fraud Prevention.

Silver Tier: +$2,148,000 annually.

- **Focus: Risk-Based Pricing, Claims Management, Education Campaigns.**

Bronze Tier: +$8,455,000 annually.

• Focus: Automated Processes, Reward Longevity, Upsell Entry-Level Add-Ons.

THANK YOU