MSBA 286 — Capstone Project II (Section 1)

University of the Pacific

Fall 2024

Final Project Instructions

Due: Monday December 9, 2024

Predicting Customer Lifetime Value (CLV) to Enhance Customer Retention and Revenue

Team Number: 1

Group Leader	**Bala Gaurav Reddy Pasam**	**ID: 989467995**
Group Member 1	**Hetvi Rajesh Bhalani**	**ID: 989469635**
Group Member 2	**Vishal Sagre**	**ID: 989469775**
Group Member 3	**Khushbu Singh**	**ID: 989465248**

Abstract:

This study focuses on leveraging predictive analytics to forecast Customer Lifetime Value (CLV) for an insurance company. By utilizing machine learning and visualization tools, the project aims to identify key drivers influencing CLV and provide actionable insights to optimize customer retention and revenue generation. A Random Forest model was developed to predict CLV with an R^2 of 0.69, highlighting the model's reliability. Insights derived from Tableau dashboards uncovered significant trends in customer demographics, sales channels, and policy types, offering targeted strategies for business growth. The findings contribute to the application of data-driven approaches in improving customer-focused decision-making processes.

Keywords:

- Customer Lifetime Value (CLV)

- Predictive Modelling

- Machine Learning

- Business Analytics

- Tableau Dashboards

1.Introduction

1.1 Objective of the Study

Customer Lifetime Value (CLV) is a critical metric for businesses, representing the total revenue a customer generates over their relationship with the company. For an insurance firm, understanding and forecasting CLV is crucial for identifying high-value customers, optimizing marketing strategies, and improving overall profitability. This study aims to:

- Develop a predictive model to forecast CLV using historical data.

- Identify key drivers of CLV through data analysis and visualization.

- Provide actionable recommendations to enhance customer retention and maximize revenue.

1.2. Research Motivation

The insurance industry operates in a highly competitive environment where customer retention is both challenging and essential. This study seeks to empower the business with data-driven strategies by:

- Predicting CLV to prioritize resource allocation.

- Understanding demographic, geographic, and behavioural patterns influencing customer value.

- Uncovering actionable insights that drive strategic decisions.

1.3. Dataset Background

The dataset, sourced from Kaggle's IBM Watson Marketing repository, comprises over 10,000 customer records. It includes attributes like demographic details (e.g., gender, marital status), vehicle information, policy types, and sales channel performance. The target variable, Customer Lifetime Value, is pre-calculated and serves as the foundation for training a predictive model and deriving insights.

1.4. Structure of the Report:

Section	Pages
Title Page	1
Abstract	2
Keywords	2
Introduction	2 - 4
Literature Review	4 - 7
Data and Methods	7 - 13
Discussion	13 - 20
Theoretical and Managerial Implications	20 - 26
Conclusion	27
References	28
Appendix	29

2. Literature Review

2.1 Importance of Customer Lifetime Value

Customer Lifetime Value (CLV) serves as a cornerstone of customer relationship management and strategic decision-making. It quantifies the long-term financial contribution of a customer, enabling businesses to allocate resources effectively, prioritize high-value segments, and enhance profitability.

Gupta et al. (2006) emphasizes that CLV prediction helps align marketing strategies with revenue potential. In the insurance sector, CLV insights are particularly valuable for optimizing policy pricing, designing personalized retention strategies, and identifying cross-selling opportunities.

2.2 Predictive Analytics in CLV Estimation: The evolution of predictive modelling has transformed CLV estimation, shifting from traditional statistical approaches to advanced machine learning techniques.

Each method contributes uniquely to the accuracy and interpretability of CLV predictions:

A. **Regression Models**: Commonly used for their simplicity, regression models have historically been the baseline for CLV prediction.

B. **Random Forest Models**: Introduced by Breiman (2001), Random Forest is widely adopted for its robustness in capturing nonlinear relationships and feature interactions, making it ideal for complex datasets like CLV.

C. **Recent Advances**: Machine learning models, such as gradient boosting and XGBoost, outperform traditional methods in capturing intricate patterns, though they often sacrifice interpretability for performance.

Random Forest strikes a balance between these priorities, offering high accuracy while maintaining explainability, a critical requirement for actionable business insights.

2.3 Visual Analytics for Business Insights

The integration of predictive analytics with visualization tools has revolutionized decision-making. Tableau, a leading dashboard platform, enables users to explore data interactively and translate technical findings into actionable strategies. Few (2006) highlights that effective dashboards:

- Facilitate trend exploration and hypothesis testing.

- Empower decision-makers with intuitive visualizations.

- Bridge the gap between raw data and strategic actions.

This project leverages Tableau to complement predictive modelling, ensuring that insights are both accessible and impactful for stakeholders.

2.4 Literature Summary Table:

Author(s)	Key Focus	Methodology	Key Findings
Gupta et al. (2006)	Importance of CLV for profitability	Regression Models	CLV is critical for aligning resource allocation with customer value.
Breiman (2001)	Introduction of Random Forest models	Ensemble Learning	Random Forest provides high accuracy in predictions with interpretable results.
Kumar et al. (2008)	CLV estimation in the insurance sector	Statistical Analysis	Segmentation by CLV improves marketing efficiency and customer satisfaction.
Few (2006)	Role of dashboards in decision-making	Visual Analytics	Dashboards help stakeholders interact with complex data, enabling actionable insights.
Chen et al. (2020)	Machine learning models for CLV prediction	Random Forest, XGBoost	Machine learning models outperform traditional methods in capturing complex relationships in CLV data.

2.5 Connection to Theories This study builds on established theories while integrating modern methodologies:

A. **Predictive Modeling**: Applies Breiman's Random Forest framework to enhance accuracy and interpretability in CLV estimation.

B. **Customer-Centric Strategies**: Aligns with Gupta's emphasis on resource prioritization for high-value customers.

C. **Visual Analytics**: Leverages Few's principles of effective dashboard design to present findings in an accessible and actionable format.

2.6 Gaps Addressed by This Study

While prior studies have extensively researched CLV prediction, this project addresses key gaps by:

- Integrating predictive modeling with visual dashboards to deliver actionable insights.

- Highlighting underutilized customer segments (e.g., niche demographics, vehicle classes) for strategic interventions.

- Bridging the theory-practice gap by translating technical findings into specific managerial recommendations tailored for the insurance sector.

3. Data and Methods

3.1 Data

The dataset for this project was sourced from Kaggle's IBM Watson Marketing repository, containing over 5,000 customer records. It includes information on customer demographics, vehicle details, policy types, and sales channel performance. The primary objective of using this dataset was to predict the Customer Lifetime Value (CLV) and identify key factors driving its variability.

120   **A. Data Features**

121   a.   **Demographic Attributes**:

122   o   Gender, marital status, and education level.

123   o   Insight into customer profiles and behavioural patterns.

124   b.   **Vehicle Details**:

125   o   Vehicle class (e.g., luxury, economy), size, and type.

126   o   Important for understanding insurance needs and risk factors.

127   c.   **Policy Information**:

128   o   Policy type (e.g., corporate, personal), coverage type (basic, premium), and

129   renewal offers.

130   d.   **Sales and Geography**:

131   o   Sales channel (agent, branch, web, call center) and customer state.

132   o   Provides insight into how and where policies are purchased.

133   e.   **Target Variable**:

134   o   CLV: Represents the total revenue generated by a customer over their lifetime

135   relationship with the company.

```python
categorical_columns = ['State', 'Response', 'Coverage', 'Education', 'EmploymentStatus',
                       'Gender', 'Location Code', 'Marital Status', 'Policy Type',
                       'Policy', 'Renew Offer Type', 'Sales Channel', 'Vehicle Class', 'Vehicle Size']
data_encoded = pd.get_dummies(data_cleaned, columns=categorical_columns, drop_first=True)

# Normalize numerical features for consistency
# Why? Scaling ensures all numerical values are on a comparable scale, improving model performance.
numerical_columns = ['Income', 'Monthly Premium Auto', 'Months Since Last Claim',
                     'Months Since Policy Inception', 'Number of Open Complaints',
                     'Number of Policies', 'Total Claim Amount']
```
136

B. Data Cleaning and Transformation

    a. **Handling Missing Values**: Default imputation strategies were applied for consistency.

    b. **Encoding Categorical Variables**: Categorical data such as Policy Type and Education Level were transformed using one-hot encoding for machine learning compatibility.

    c. **Feature Scaling**: Continuous features like Income and CLV were normalized to ensure consistency and enhance model performance.

    d. **Outlier Detection**: Extreme values in CLV were examined and retained to preserve variability essential for the insurance industry.

3.2 Methods

A. Predictive Modelling with Python

The core of the analysis involved building a predictive model for CLV using Python. A **Random Forest Regressor** was chosen for its ability to handle non-linear relationships and complex feature interactions effectively. The modeling process included:

    a. **Data Preparation**:

- Cleaned and processed the dataset to ensure high-quality inputs.

- Encoded categorical variables and scaled continuous features for uniformity.

    b. **Model Training**:

- Split the data into an 80% training set and a 20% testing set.

157 • Trained the Random Forest model using scikit-learn's implementation,

158 optimizing key hyperparameters such as the number of trees and depth using

159 grid search.

Correlation Heatmap

160
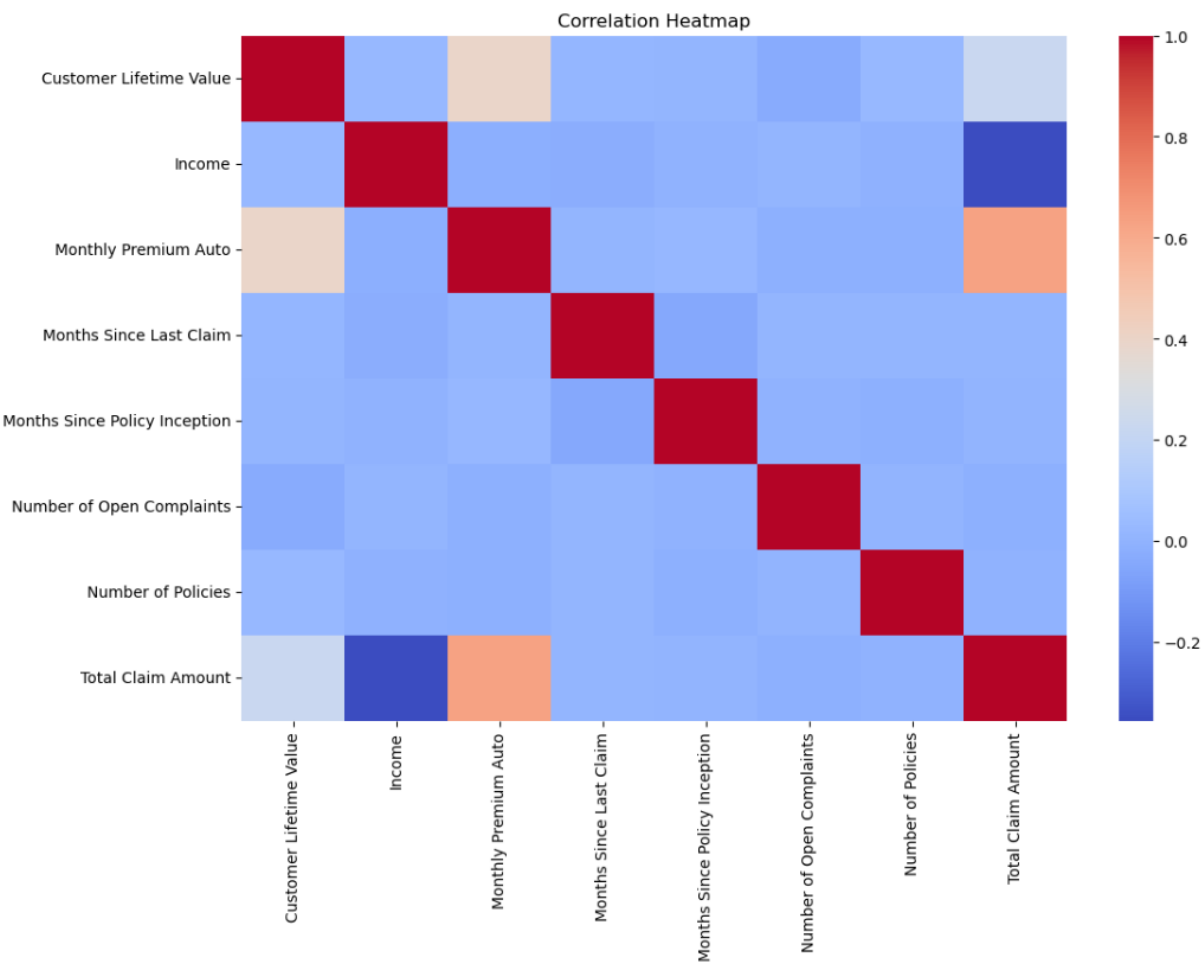
161 c. **Evaluation**:

162 • Achieved an R² score of 0.69, indicating that the model explains 69% of the

163 variance in CLV.

164 • The Mean Absolute Error (MAE) of $1,473 highlights the model's reliability

165 in predicting CLV values.

```
Mean Absolute Error (MAE): 1473.8469485605506
R² Score: 0.6897674423512135
```
166

167 d. **Results Integration**:

168   • Predicted CLV values were appended to the dataset, allowing for direct

169   comparisons with existing values and facilitating deeper insights in Tableau.

```
# Confirmation message
print(f"Predicted results saved to: {full_path}")

Predicted results saved to: D:\UNIVERSITY OF PACIFIC\MSBA 286\Final Project\Predicted_CLV.csv
```
170

171 **B. Visualization and Business Analytics**

172 To complement the machine learning results, Tableau dashboards were developed, offering a

173 user-friendly interface for exploring CLV trends and patterns. These dashboards provided:

174 • Interactive visuals for demographics, policy types, sales channels, and geographic

175   distribution.

176 • Clear identification of actionable business opportunities.

177 **3.3 Analytics and Visualization**

178 **A. Process Overview**

179 a. **Exploratory Data Analysis (EDA)**:

180   • Used Python libraries like pandas and matplotlib to identify trends,

181   correlations, and anomalies in the data.

182   • Key insights, such as the dominance of suburban customers and the popularity

183   of basic coverage plans, informed subsequent analyses.

184 b. **Model Development**:

185   • Built and evaluated the Random Forest model to predict CLV accurately.

186   • Exported predictions for visualization and further analysis.

c. **Tableau Dashboards**:

- Integrated model outputs with interactive visuals to highlight key trends, such as the impact of education level and vehicle class on CLV.

B. Key Visualizations

a. **CLV Distribution**: Reveals the skewed nature of customer lifetime value, with a small proportion of high-value customers.

b. **Geographic Insights**: Heatmaps showing state-wise CLV contributions, highlighting California's dominance and the potential in Oregon.

c. **Demographic Patterns**: Charts showing married women and divorced customers as high-value segments.

d. **Policy and Coverage Trends**: Analysis of popular and profitable plans, with actionable insights for improving premium plan adoption.

C. Managerial Implications of the Methodology

The insights derived from this analysis are critical for decision-making:

a. **Resource Allocation**: Predicted CLV values enable managers to prioritize high-value customers for retention campaigns.

b. **Targeted Marketing**: Visual trends guide personalized marketing strategies for specific customer demographics, such as luxury vehicle owners or high-educated retirees.

c. **Policy Optimization**: Analysis of popular vs. profitable plans highlights opportunities for upselling and developing new policy tiers.

208   d.   **Geographic Strategy**: Heatmaps identify underperforming regions like Nevada,

209   informing targeted campaigns for growth.

210   This integrated methodology not only addresses the technical objective of predicting CLV but

211   also provides a robust framework for actionable business strategies.

212   **D. Our Tableau Dashboard:**

213
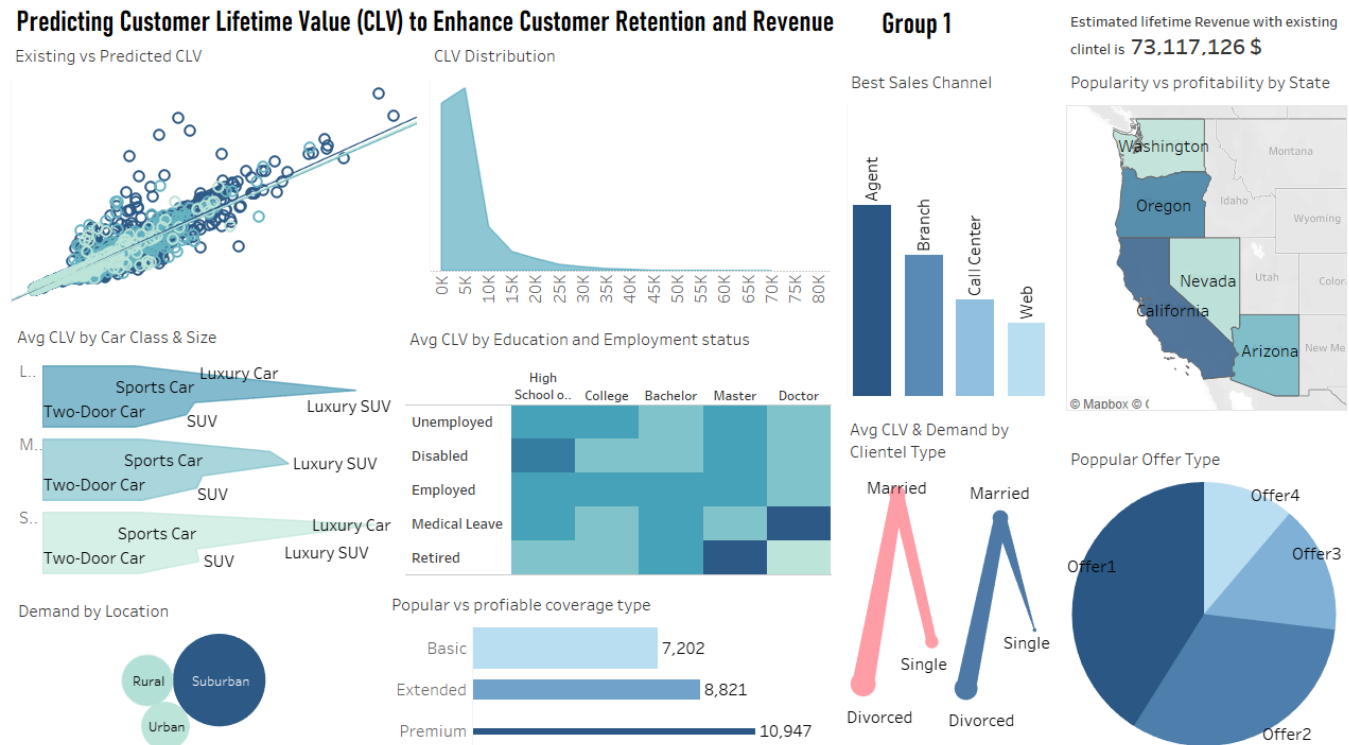
214   **4. Discussion**

215   **4.1 Model Performance**

216   The predictive model for Customer Lifetime Value (CLV) was developed using a Random

217   Forest Regressor, chosen for its robustness in handling large datasets and capturing non-

218   linear relationships. The model achieved an R^2 score of 0.69, meaning it could explain 69%

219   of the variance in CLV, demonstrating substantial predictive power. The Mean Absolute

220   Error (MAE) of $1,473 indicates that, on average, the predictions deviated by this amount

221   from actual values.

222   These metrics highlight the reliability of the model for practical application. However, the

223   presence of certain outliers (notably in the "Personal Auto" category) suggests areas for

224   refinement. These findings support the use of the model in segmenting customers and driving

225   targeted strategies, while also pointing to opportunities for further improvements in predictive

226   accuracy.

227   **4.2 Existing vs. Predicted CLV**

Existing vs Predicted CLV

228
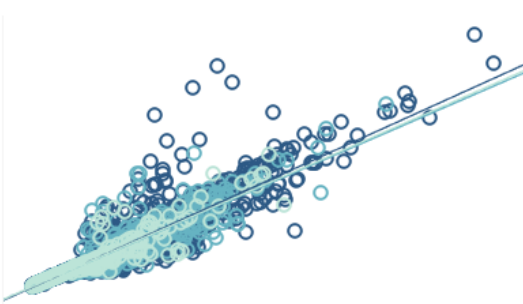
229   The scatterplot comparing existing and predicted CLV illustrates a strong alignment between

230   the two, reflecting the model's accuracy. However, noticeable outliers—particularly among

231   "Personal Auto" customers—highlight discrepancies in certain predictions. These outliers

232   warrant further investigation to understand unique customer behaviors or data irregularities

233   contributing to the deviations.

234   The alignment of most data points along the line of perfect prediction reinforces the model's

235   reliability in forecasting CLV for strategic decision-making. By addressing the outliers, the

236   model's accuracy could be enhanced, further strengthening its applicability for customer

237   segmentation and personalized interventions.

238   **4.3 CLV Distribution**

CLV Distribution

239
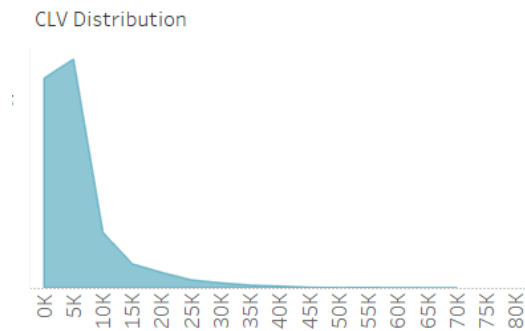
240   The distribution of CLV is heavily skewed, with the majority of customers having a CLV

241   below $15,000. Beyond this threshold, the count of customers declines steeply. This pattern

242   underscores a concentration of value within a smaller segment of high-value customers.

243   This insight emphasizes the need to design strategies that not only retain these high-value

244   customers but also uplift low-to-mid CLV segments through upselling, cross-selling, and

245   loyalty programs. Enhancing the value of lower-CLV customers could result in a more

246   balanced distribution and increased revenue potential across the customer base.

247   **4.4 Best Sales Channel**

Best Sales Channel

248
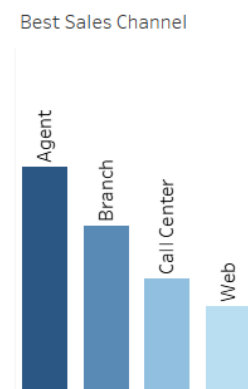
249   The analysis of sales channels reveals stark differences in performance:

250   • **Agents** dominate in both client acquisition and total CLV contribution, making them

251   a critical component of the sales strategy.

- **Call Centres** achieve the highest average CLV, reflecting their ability to attract high-value customers despite a smaller client base.

- **Web Channels**, in contrast, underperform in both customer count and average CLV, indicating the need for enhanced digital marketing and user experience.

These findings highlight the importance of optimizing underperforming channels while leveraging the strengths of agents and call centers to maximize value generation.

4.5 Geographic Distribution:


Popularity vs profitability by State
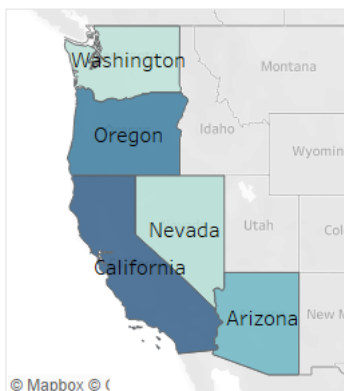
The geographic analysis provides key insights:

- **California** leads in both total CLV and client count, reaffirming its role as the most valuable market.

- **Oregon** stands out with strong performance in total CLV, maintaining its position as the second-highest contributor.

- **Nevada** leads in average CLV per client, highlighting its profitability despite contributing less in terms of total CLV and client count.

- **Arizona** has the lowest contribution in both metrics, presenting an opportunity for targeted campaigns to improve penetration.

269 These findings guide resource allocation, with a focus on retaining high-value clients in

270 leading states while driving growth in underperforming regions.

271 **4.6 Average CLV by Car Class and Size**


Avg CLV by Car Class & Size

272
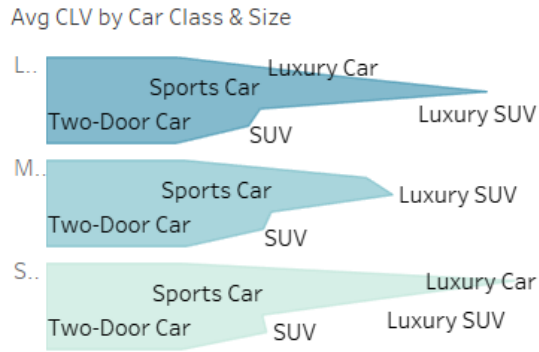
273 The analysis of car class and size highlights significant differences in CLV generation:

274 • **Luxury Vehicles**, especially Small Luxury Cars and Large Luxury SUVs, yield the

275 highest average CLV ($22,637 and $21,294, respectively). This finding reinforces the

276 profitability of targeting high-end vehicle owners.

277 • Economy vehicles dominate in customer volume but generate significantly lower

278 average CLV, reflecting their mass-market appeal but limited profitability.

279 These insights underscore the need to prioritize high-value segments like luxury vehicle

280 owners with premium offerings while exploring ways to enhance the value proposition for

281 economy vehicle customers.

282 **4.7 Demand by Location**

Demand by Location

283
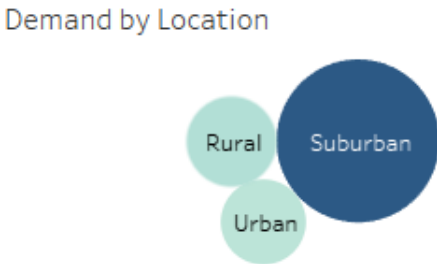
284   The geographic distribution of demand reveals that suburban areas dominate in client density,

285   followed by urban and rural regions. The relatively lower client density in rural areas

286   represents untapped potential, which can be addressed through tailored marketing strategies

287   and localized offerings. Strengthening the suburban customer base while expanding into rural

288   markets could drive significant growth.

289   **4.8 Popular vs. Profitable Coverage Type**

Popular vs profiable coverage type

|	
Basic	7,202
Extended	8,821
Premium	10,947

290
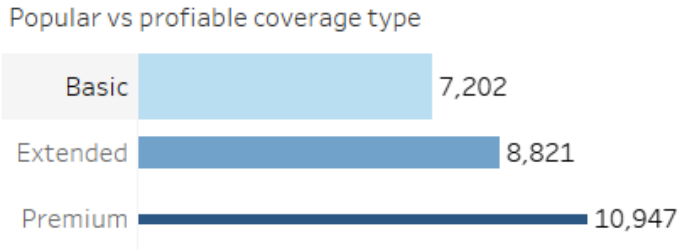
291   The analysis of coverage types reveals a clear distinction between popularity and

292   profitability:

293   • **Premium Coverage**: Generates the highest average CLV ($10,947), making it the

294   most profitable plan, but has the lowest client count, indicating limited adoption.

295   • **Basic Coverage**: Attracts the highest client count but contributes the lowest average

296   CLV, reflecting its mass-market appeal but limited profitability.

297   This imbalance underscores the importance of targeted upsell strategies to transition basic

298   policyholders to premium plans. Additionally, introducing mid-tier options could bridge the

299   gap and attract more customers toward profitable plans.

4.9 Average CLV and Demand by Clientele Type


Avg CLV & Demand by Clientel Type
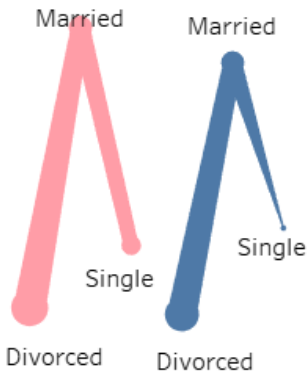
The segmentation by gender and marital status reveals distinct trends:

- **Married Women** dominate in client count, suggesting an opportunity to design family-oriented policies.

- **Divorced Women** generate the highest average CLV, presenting a niche high-value market for tailored offerings. These findings highlight the importance of understanding client demographics to design effective marketing strategies.

4.10 Popular Offer Type

Poppular Offer Type
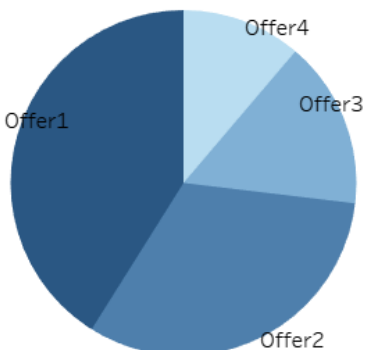
Offer1 outperforms other offers in client count and total CLV contribution. Despite similar average CLV across all offer types, the higher adoption of Offer1 suggests features or

312   incentives that resonate well with customers. By analyzing the success factors of Offer1,

313   businesses can replicate its appeal across other offers to improve overall performance.

314   **4.11 Average CLV by Education and Employment Status**

315
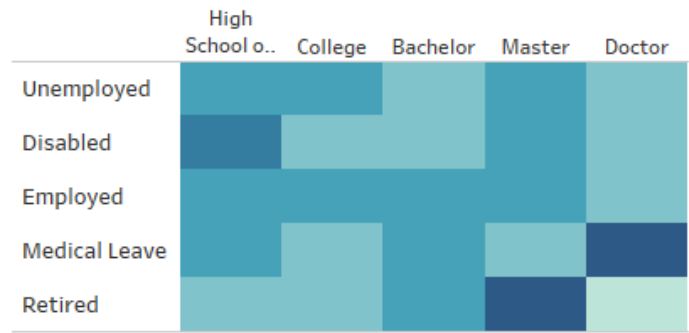
316   The analysis of education and employment status reveals that:

317   • **Doctorate holders on medical leave** and **Master's degree retirees** generate the

318   highest average CLV, representing niche but highly profitable customer segments.

319   • Higher education levels generally correlate with higher CLV, emphasizing the

320   importance of targeting educated customers with tailored packages.

321   These findings suggest opportunities to design specialized offerings for highly educated and

322   temporarily unemployed professionals, as well as retirees with substantial purchasing power.

323   **5. Theoretical and Managerial Implications**

324   **5.1 Theoretical Implications**

325   Our findings make significant contributions to the understanding of **Customer Lifetime**

326   **Value (CLV)**, providing a deeper understanding of its drivers, patterns, and actionable

327   segmentation strategies:

328   **A. Advancing CLV Models**:

329   The predictive model showcases how demographic, geographic, and product-specific

330 factors interact to influence CLV. This aligns with and extends existing frameworks

331 on customer profitability by integrating high-dimensional predictive analytics.

332 - For instance, the observed high average CLV among luxury vehicle owners
333   and educated retirees confirms theories that higher disposable income drives
334   customer value. However, the niche finding of medical leave customers
335   having high CLV suggests new areas for further study in customer behavior
336   during life transitions.

337 **B. Understanding Profitability by Segment**:

338 Insights such as suburban dominance, premium coverage profitability, and gender-

339 based preferences highlight the interplay between product type, location, and

340 demographics. These findings can refine segmentation theory and help academics

341 develop more nuanced customer models.

342 **C. Exploring Underrepresented Segments**:

343 Outliers like rural regions or small luxury vehicle owners suggest opportunities for

344 further research into niche or underserved markets. These insights help address gaps

345 in traditional CLV studies, which often focus on the mainstream customer base.

346 **5.2 Managerial Implications**

347 The findings provide a roadmap for strategic decision-making to optimize customer retention,

348 maximize profitability, and expand market share. Here are actionable recommendations for

349 each insight:

350 **A. Leveraging Model Insights**

351 a. **Implication**: The CLV model equips managers with a predictive tool for

352   segmentation and personalization.

b. **Actions**:

- Develop **high-CLV customer retention strategies**: Offer loyalty rewards and premium benefits to top-tier customers identified by the model.

- Use **CLV-based segmentation** to prioritize marketing spend: For example, allocate higher budgets to customers with high predicted CLV and offer cross-sell opportunities to mid-CLV segments.

B. Sales Channel Optimization

a. **Implication**: Call centers show the highest average CLV but lower client count, while web channels underperform across both metrics. Agents and branches dominate sales count but require better targeting strategies.

b. **Actions**:

- **Call Center**: Invest in training for call center staff to promote premium products effectively. Leverage their high-CLV potential by routing high-value customer inquiries directly to them.

- **Web Channel**: Implement AI-driven recommendations and dynamic pricing on the web platform. Enhance the user experience to convert more visitors into high-CLV customers.

- **Agent Sales**: Focus agent campaigns on upselling policies to existing customers or targeting luxury segments.

C. Geographic Strategy

a. **Implication**: California and Oregon dominate in client count and CLV, while Nevada shows the highest CLV per customer. Arizona underperforms in both metrics.

b. **Actions**:

- **Retention in High-Performing States**: Reinforce loyalty programs in California and Nevada to retain existing high-value customers.

- **Expand in Underrepresented Markets**: Launch localized campaigns in Arizona and rural regions offering affordable bundles to attract new customers.

- **Targeting High-CLV States**: For Nevada, focus on high-value segments like luxury vehicle owners and premium policyholders.

D. Product-Based Strategies (Car Class and Size)

a. **Implication**: Luxury vehicles generate significantly higher average CLV, with small luxury cars leading the segment. Economy cars dominate in count but lag in profitability.

b. **Actions**:

- Develop **exclusive premium services** (e.g., concierge assistance, luxury repair networks) tailored to luxury vehicle owners.

- Create **value-based add-ons** for economy car owners, such as accident forgiveness or extended warranties, to increase their CLV.

E. Policy Type Optimization

a. **Implication**: Premium policies generate the highest average CLV but have the lowest adoption. Basic policies dominate in count but offer the lowest profitability.

b. **Actions**:

- **Upsell Strategies**: Educate basic policyholders on the benefits of premium coverage through targeted marketing campaigns and testimonials.

- **Mid-Tier Options**: Introduce an intermediate policy bridging basic and premium coverage to appeal to customers reluctant to jump to premium.

- **Reward Systems**: Provide discounts or cashback incentives for upgrading policies.

F. Demographic Insights

a. **Implication**: Married women dominate in policy count, while divorced women generate the highest average CLV.

b. **Actions**:

- Design **family-centric marketing campaigns** targeting married women, emphasizing long-term benefits like family protection.
- Develop **customized offers for divorced women**, such as flexible payment plans or enhanced coverage, to tap into their high-CLV potential.

G. Education and Employment Focus

a. **Implication**: Doctorate holders on medical leave and Master's degree retirees show unexpectedly high CLV. Higher education correlates with higher CLV overall.

b. **Actions**:

- Create **special plans for retirees and medical leave customers**, offering flexibility and long-term security.

- Offer **education-based discounts** or tiered plans to attract and retain high-CLV educated customers.

H. Offer Type Analysis

a. **Implication**: Offer1 dominates in client count, but other offers show similar average CLV.

b. **Actions**:

- Investigate the unique features of Offer1 that make it appealing, and replicate those across other offers.

- Reevaluate underperforming offers and enhance their attractiveness through additional perks or simplified terms.

I.   Rural Market Potential

a. **Implication**: Rural regions lag significantly in both client count and CLV.

b. **Actions**:

- Develop **community-specific bundles** to address affordability and access barriers.

- Partner with local organizations to increase trust and drive awareness in rural markets.

5.3 Limitations and Future Research

While the findings provide robust insights, several limitations must be acknowledged:

A. **Data Dependency**:

- The model is trained on historical data, which may not account for sudden market changes or emerging trends.

438   - The lack of behavioral data (e.g., loyalty metrics) limits the granularity of
439   predictions.

440 B. **Niche Segment Challenges**:

441   - Smaller sample sizes for segments like rural customers or medical leave
442   policies may skew results.

443 C. **Operational Scalability**:

444   - Translating insights into actionable strategies may require significant
445   operational changes, such as upgrading call centers or introducing new policy
446   tiers.

447 **Future Research Directions**:

448 a. Explore the impact of customer engagement and loyalty metrics on CLV prediction.

449 b. Conduct market experiments in underrepresented regions or segments to validate
450   recommendations.

451 c. Incorporate external economic factors (e.g., inflation, unemployment) into CLV
452   prediction models for richer insights.

6. Conclusion

This study utilized predictive analytics and advanced visualizations to estimate and analyze Customer Lifetime Value (CLV), focusing on uncovering key drivers and actionable insights for business optimization. The results underscore the pivotal role of predictive modeling in understanding customer segments and identifying opportunities to enhance profitability. Key findings include:

A. **Significant Insights on Coverage Types**: Premium coverage plans generate the highest average CLV, suggesting a need for targeted upsell strategies. Conversely, basic plans dominate in client count but underperform in profitability, highlighting untapped potential for growth.

B. **Geographic Trends**: States such as California lead in both total CLV and client count, while regions like Nevada boast high average CLV per client, offering an opportunity for geographically tailored strategies.

C. **Client Demographics and Preferences**: High-CLV segments include educated retirees and married women, presenting a niche for customized policies and marketing efforts.

D. **Channel and Product Performance**: Call centers exhibit the highest average CLV, although they lag in client volume. Luxury vehicles, particularly small luxury cars and SUVs, significantly drive higher CLV.

These findings provide a foundation for implementing focused, data-driven strategies to improve customer retention, enhance resource allocation, and maximize profitability. The integration of machine learning models like Random Forest with tools such as Tableau enables organizations to translate complex data into actionable business solutions.

7. References

- Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

  https://doi.org/10.1023/A:1010933404324

- Chen, T., & Guestrin, C. (2020). XGBoost: A scalable tree boosting system.
  *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge
  Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

- Few, S. (2006). *Information dashboard design: The effective visual communication of
  data.* O'Reilly Media.

- Gupta, S., Lehmann, D. R., & Stuart, J. A. (2006). Valuing customers. *Journal of
  Marketing Research, 41*(1), 7–18. https://doi.org/10.1509/jmkr.41.1.7

- Kumar, V., & Reinartz, W. (2008). *Customer relationship management: Concept,
  strategy, and tools.* Springer.

Dashboard:

https://public.tableau.com/views/FinalProject286/Dashboard2?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Predictive Model:

https://drive.google.com/file/d/1cKsTP9lzWYADkC48YB04Z_nhX-aNodM7/view?usp=sharing

Final Data:

https://drive.google.com/file/d/1e8CwXfaWbRLCyFWQxjYSGaG0kET4INXI/view?usp=sharing

Full Project Folder:

https://drive.google.com/drive/folders/1PJh0dv_H0nteBg0rfUmNZdfPJdD-NnnH?usp=sharing

500 **8.Appendix**

Time	Contents
Saturday, October 19, 2024	Project Group Formation
Saturday, November 2, 2024	Proposal Submission
Saturday, November 16, 2024	Data Cleaning and Model Training
Saturday, November 30, 2024	Dashboard Creation and Initial Findings
Saturday, December 5, 2024	Final Report Draft and Review
Monday, December 9, 2024	Final Project Paper Submission and Presentation Slides

501