

---

# Customer Lifetime Value Analysis - E-Commerce Sector

Shruti Bhanot E24092

---

## Overview

The Customer Lifetime Value (CLV) analysis for FLO identifies key customer segments and predicts CLV using Ridge Regression. The analysis segments customers into groups like Best Customers and At Risk, assesses model performance, and offers actionable insights for targeted marketing and customer retention strategies.

## Objective

The objective of the CLV analysis is to evaluate and predict customer lifetime value using statistical models, segment customers into actionable groups, and provide insights to enhance marketing strategies and improve customer retention.

## Assigned Task(s)

Customer Lifetime Value Analysis - E-Commerce Sector

## Task Details

- Task 2: The task involves analyzing customer lifetime value (CLV) to gain insights into customer behavior and predict future spending. Key steps include cleaning and preparing the dataset, engineering relevant features, and building predictive models using linear regression and Ridge regression. The analysis focuses on testing model performance with various regularization strengths and validating results through cross-validation. Additionally, customer segments are identified based on CLV metrics. The final deliverable includes a comprehensive report summarizing findings, visualizing data, and offering actionable recommendations for enhancing marketing strategies and customer retention efforts.
- Status: Complete
- Details:

The task involves analyzing customer lifetime value (CLV) to understand customer behavior and predict future value. This includes:

1. Data Preparation: Cleaning and transforming the dataset to ensure accuracy.

2. Feature Engineering: Creating relevant features like average purchase value and tenure.
3. Modeling: Implementing and evaluating predictive models, including linear regression and Ridge regression, to estimate CLV.
4. Regularization Analysis: Testing model performance with varying regularization strengths to prevent overfitting.
5. Segmentation: Identifying and analyzing customer segments based on CLV and other metrics.
6. Validation: Conducting cross-validation to assess model reliability and robustness.
7. Reporting: Summarizing findings, visualizing results, and providing actionable insights for marketing strategies.

## Progress

- Accomplishments :

1. High-Accuracy Models: Achieved an exceptional R-squared value of 1.0 for both standard and Ridge regression models, indicating perfect predictive accuracy on training data.
2. Effective Regularization: Demonstrated robust model performance across varying regularization strengths in Ridge regression, with minimal Mean Squared Error (MSE) and consistent R-squared values.
3. Feature Engineering: Successfully engineered and utilized key features, such as Recency, Frequency, and Monetary values, to enhance model accuracy and interpretability.
4. Cross-Validation Success: Ensured model reliability through rigorous cross-validation, confirming the robustness of predictive performance.
5. Insightful Segmentation: Identified distinct customer segments based on CLV metrics, providing actionable insights for targeted marketing and customer retention strategies.

- Metrics :

1. Model Accuracy:

- Standard Regression:
  - Mean Squared Error (MSE):  $(1.2756 \times 10^{-25})$
  - R-squared: 1.0
- Ridge Regression:
  - Mean Squared Error (MSE):  $(1.6160 \times 10^{-13})$

- R-squared: 1.0
- Ridge with Stronger Regularization:
  - Mean Squared Error (MSE):  $(1.6159 \times 10^{-11})$
  - R-squared: 1.0

## 2. Cross-Validation Metrics:

- Standard Regression:
  - Cross-Validation MSE:  $(2.2535 \times 10^{-25})$
- Ridge Regression:
  - Cross-Validation MSE:  $(1.7086 \times 10^{-13})$
- Ridge with Stronger Regularization:
  - Cross-Validation MSE:  $(1.7085 \times 10^{-11})$

## 3. VIF (Variance Inflation Factor) Values:

- Recency: 1.623
- Frequency: 3.456
- Monetary: 4.058

## 4. Randomized Target Metrics for Ridge Regression:

- Mean Squared Error (MSE): 629,639.481
- R-squared: 0.00031

# Challenges and Solutions

## ● Challenge Faced :

1. Data Quality Issues: Initially encountered missing or incorrect column names, which required adjustments to ensure accurate calculations and analysis.
2. Model Overfitting: Faced high R-squared values and extremely low MSE, suggesting potential overfitting. Addressed by applying regularization techniques and cross-validation.
3. Feature Multicollinearity: Identified through high VIF values, necessitating careful consideration of feature selection and regularization to maintain model robustness.
4. Randomized Target Variability: The randomized target data revealed poor model performance, highlighting the importance of using genuine data for reliable analysis.

## ● Solutions Faced :

1. Data Quality Issues: Adjusted column names and updated the DataFrame to correct missing or incorrect entries, ensuring accurate feature creation and calculations.
2. Model Overfitting: Applied Ridge regression with varying regularization strengths to control overfitting. Conducted cross-validation to verify model performance and avoid excessive fitting to training data.
3. Feature Multicollinearity: Assessed Variance Inflation Factor (VIF) to identify multicollinear features. Used regularization methods to mitigate the impact of multicollinearity and enhance model stability.
4. Randomized Target Variability: Conducted analysis with randomized target data highlights model robustness. Verified results against genuine data to ensure that the model's performance metrics reflect real-world applicability.

## Next Step

- Upcoming Task:
  - Perform further Analysis on this dataset
- Goals :
  - Complete the upcoming task and prepare ppt.

## Conclusion

Summary:

In the CLV analysis, the primary focus was on refining the customer segmentation and predicting Customer Lifetime Value (CLV). The analysis involved:

- Feature Engineering: Created and validated new features, including average order value and tenure-based metrics.
- Model Performance: Achieved excellent results with the initial model, showing very low MSE and high R-squared. Applied Ridge regression to address overfitting and assessed the model's stability with varying regularization strengths.
- Feature Multicollinearity: Evaluated and addressed multicollinearity through VIF analysis, ensuring robust feature inclusion in the model.
- Validation: Verified model robustness by testing with randomized target data, confirming the model's strong performance on genuine data.

Overall, the analysis demonstrated a strong predictive capability with well-calibrated models, providing valuable insights into customer behavior and future value.

Acknowledgments : Thank you for your time and attention

