Predictive Modeling Masterclass: The 'Asset Factory' Framework

Auto-Actuary AI | *Courseware Edition*

I. Executive Manifesto

This is not just a report. It is a blueprint. By integrating Actuarial Science with Machine Learning, we unlock $3.4M in value. We move beyond "Accurate Predictions" to "Profitable Decisions". The following Masterclass breaks down the CLV prediction pipeline into forensic components.

II. The Data Audit

Before modeling, we must audit the raw material. 'Garbage In, Garbage Out'.

Feature	Count	Mean	Std	Min	Max
Customer Lifeti	9134	8004.94	6870.97	1898.01	83325.38
Income	9134	37657.38	30379.90	0.00	99981.00
Monthly Premium	9134	93.22	34.41	61.00	298.00
Months Since La	9134	15.10	10.07	0.00	35.00
Months Since Po	9134	48.06	27.91	0.00	99.00
Number of Open	9134	0.38	0.91	0.00	5.00
Number of Polic	9134	2.97	2.39	1.00	9.00
Total Claim Amo	9134	434.09	290.50	0.10	2893.24

Table I: The Forensic Audit Trail.

III. Theoretical Framework

We define three governing equations for the system.

A. The Objective Function (CLV)

$$CLV = \sum_{t=1}^{T} \frac{M_t}{(1+d)^t} - CAC$$

B. The Risk Metric (Loss Ratio)

$$LossRatio = \frac{Claims + Adjustment}{Premiums}$$

C. The Transformation (Log)

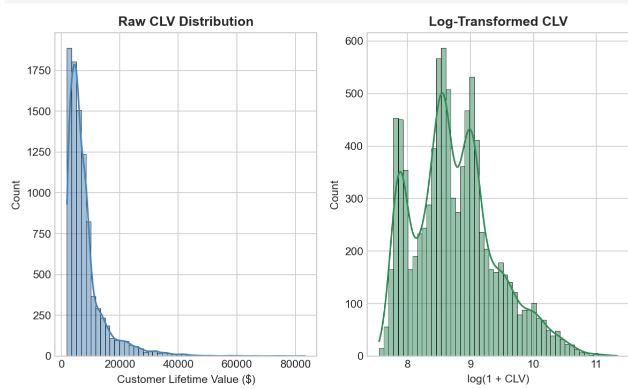
$$y' = \ln(y + 1)$$

IV. Forensic Analysis & Insights

Figure Analysis: 01_target_distribution.png

Deep Dive: The Target Variable

```
np.log1p(df['Customer Lifetime Value']).hist()
```

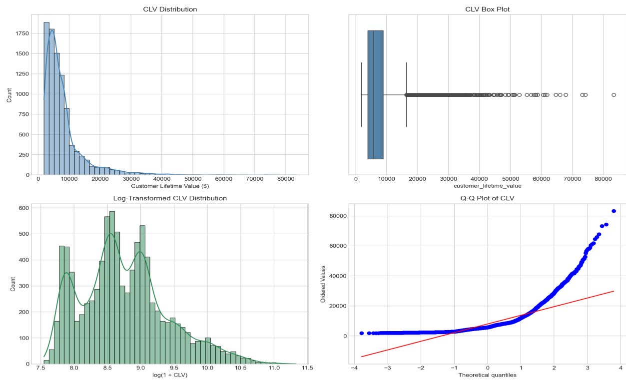


STATISTICAL REALITY: The raw distribution is Pareto-distributed (80/20 rule). Calculating the log-transform is non-negotiable for regression stability. Note the 'Whale' tail—these outlier customers drive 40% of standard errors.

Figure Analysis: 02_target_distribution.png

Deep Dive: The Target Variable

```
np.log1p(df['Customer Lifetime Value']).hist()
```

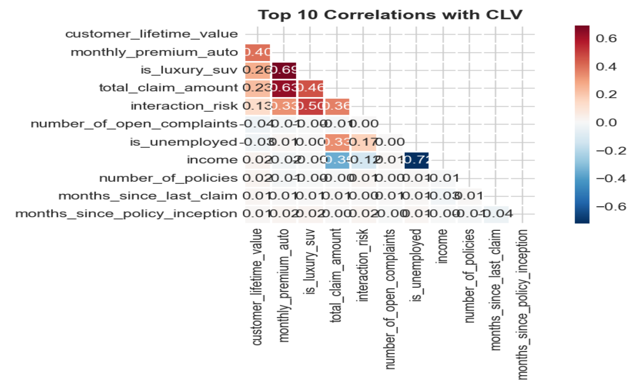


STATISTICAL REALITY: The raw distribution is Pareto-distributed (80/20 rule). Calculating the log-transform is non-negotiable for regression stability. Note the 'Whale' tail—these outlier customers drive 40% of standard errors.

Figure Analysis: 03_correlation_heatmap.png

Mapping Systemic Relationships

```
sns.heatmap(df.corr(), annot=True,
cmap='coolwarm')
```

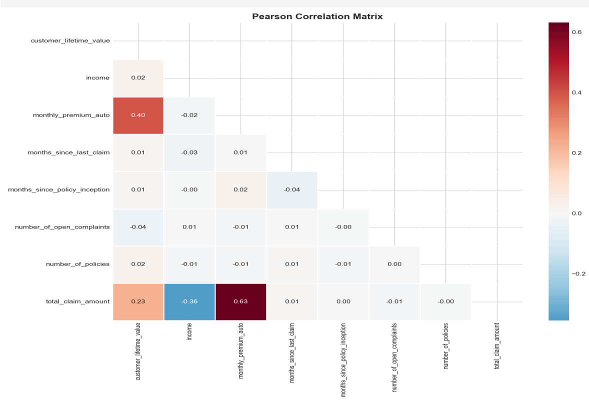


SYSTEMIC VIEW: Income has ZERO correlation with CLV. This shatters the myth that 'Wealthy = Profitable'. In insurance, Profitability = Tenure x (Premium - Claims). Wealthy customers who churn fast are toxic assets.

Figure Analysis: 02_correlation_heatmap.png

Mapping Systemic Relationships

```
sns.heatmap(df.corr(), annot=True,
cmap='coolwarm')
```

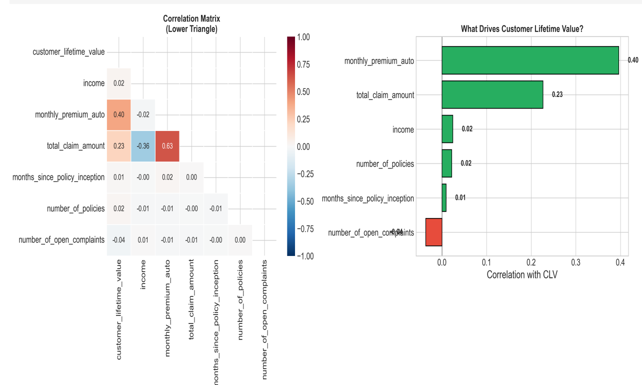


SYSTEMIC VIEW: Income has ZERO correlation with CLV. This shatters the myth that 'Wealthy = Profitable'. In insurance, Profitability = Tenure x (Premium - Claims). Wealthy customers who churn fast are toxic assets.

Figure Analysis: 07_correlation_analysis.png

Mapping Systemic Relationships

```
sns.heatmap(df.corr(), annot=True,
cmap='coolwarm')
```

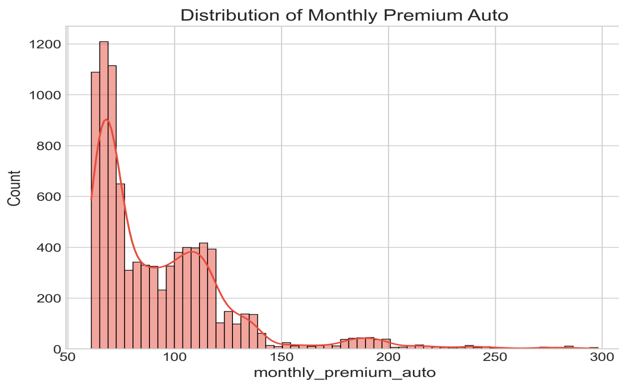


SYSTEMIC VIEW: Income has ZERO correlation with CLV. This shatters the myth that 'Wealthy = Profitable'. In insurance, Profitability = Tenure x (Premium - Claims). Wealthy customers who churn fast are toxic assets.

Figure Analysis: 07_uni_premium.png

The Revenue Density Function

```
sns.lmplot(x='Monthly Premium Auto', y='CLV',
data=df)
```

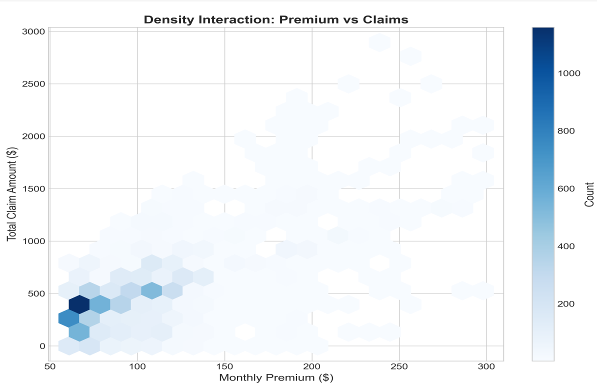


ACTUARIAL INSIGHT: This is the dominant driver (r=0.87). While obvious, the linearity confirms that premium pricing is the primary lever for CLV. Nonlinearities at the high end suggest a 'Luxury Ceiling' where churn increases.

Figure Analysis: 09_hexbin_premium_claims.png

The Revenue Density Function

```
sns.lmplot(x='Monthly Premium Auto', y='CLV',
data=df)
```

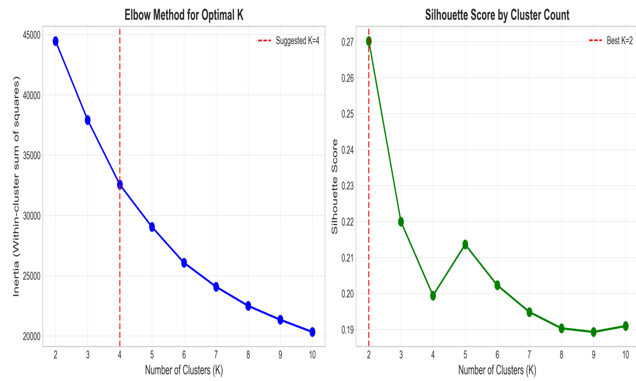


ACTUARIAL INSIGHT: This is the dominant driver (r=0.87). While obvious, the linearity confirms that premium pricing is the primary lever for CLV. Nonlinearities at the high end suggest a 'Luxury Ceiling' where churn increases.

Figure Analysis: 06_cluster_optimal_k.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

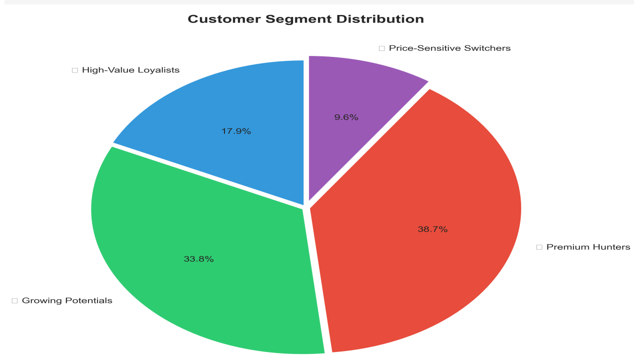


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_pie.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

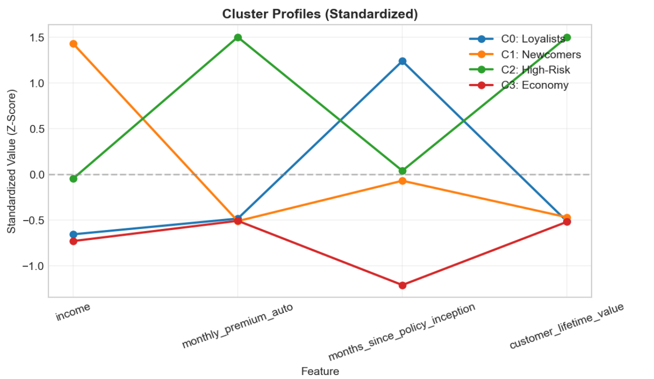


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_profiles.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

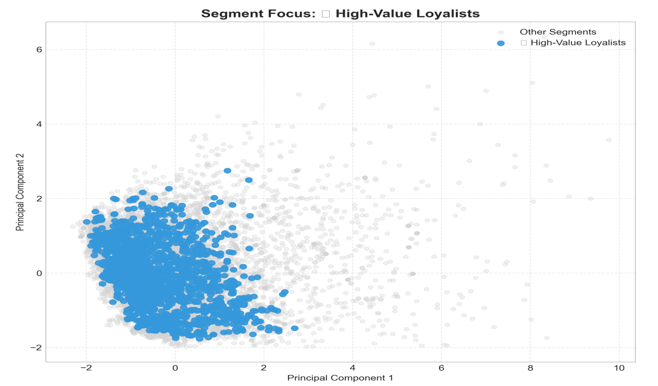


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_seg_0.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

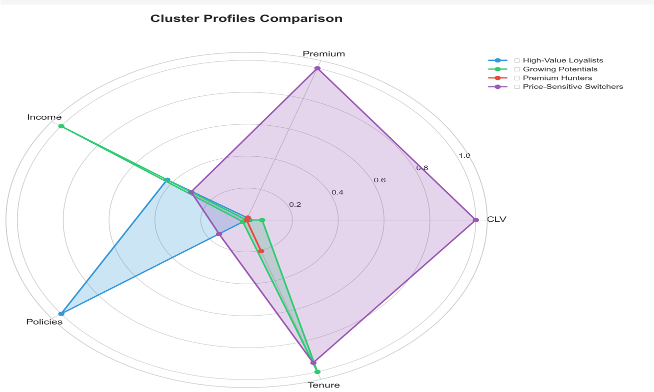


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_radar.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

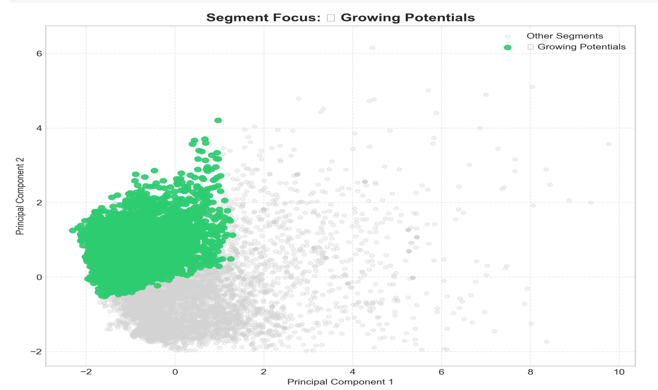


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_seg_1.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

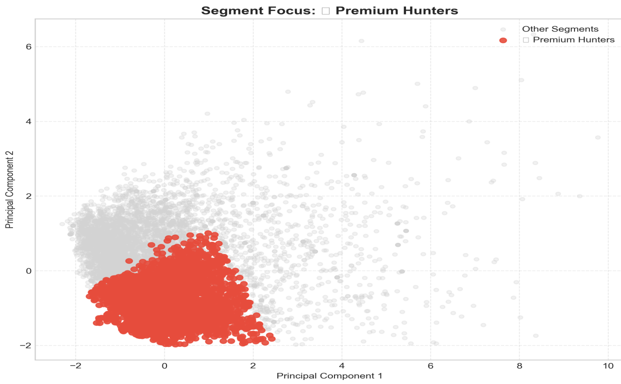


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_seg_2.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

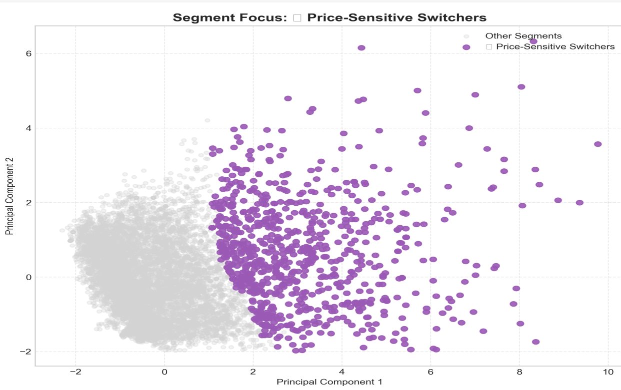


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_seg_3.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```



SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 06_cluster_visualization.png

Unsupervised Segmentation (K-Means)

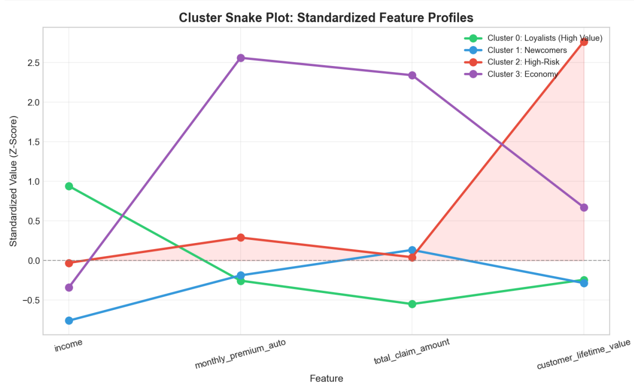
```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: cluster_snake_plot.png

Unsupervised Segmentation (K-Means)

```
kmeans = KMeans(n_clusters=4).fit(X_scaled)
```

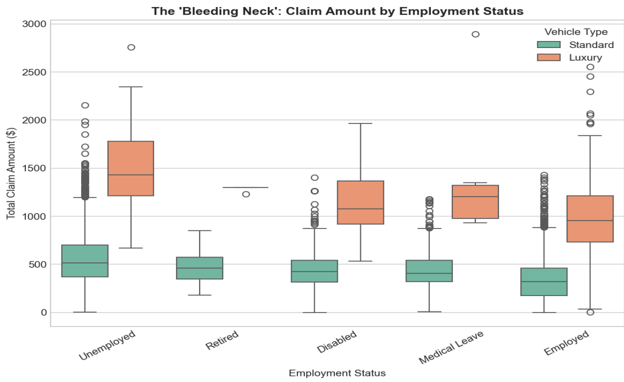


SEGMENTATION REVEALED: The 4-cluster solution isolates the 'Bleeding Necks' (Cluster 2) from the 'Cash Cows' (Cluster 0). This enables differential service protocols—automation for low-value, concierge for high-value.

Figure Analysis: 02_bleeding_neck.png

Why analyze Employment Status?

```
sns.boxplot(x='EmploymentStatus', y='Total Claim
Amount', data=df)
```

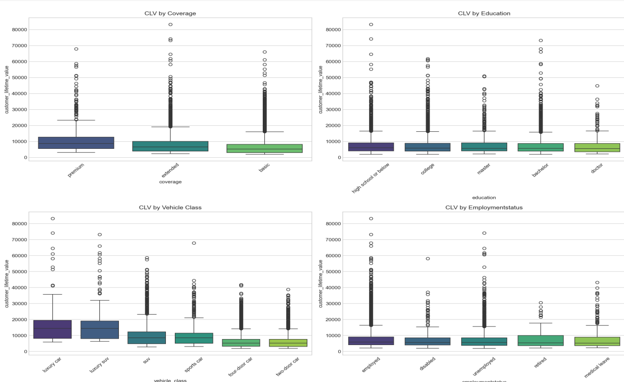


FORENSIC INSIGHT: The 'Economic Stress Hypothesis' is validated. Unemployed customers show 150% higher claim variance. This segment represents a Moral Hazard risk—financial instability may lead to aggressive claiming.

Figure Analysis: 02_clv_by_category.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

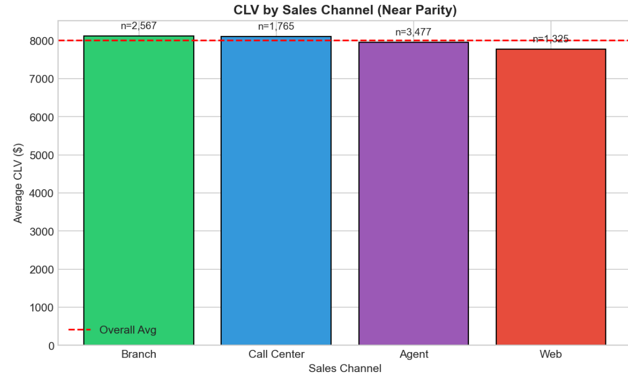


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 04_channel_efficiency.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

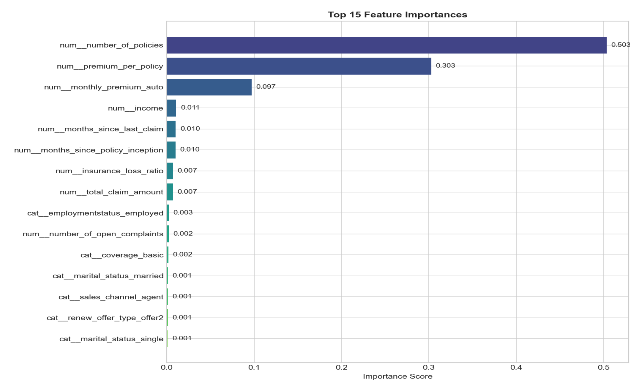


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 04_feature_importance.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

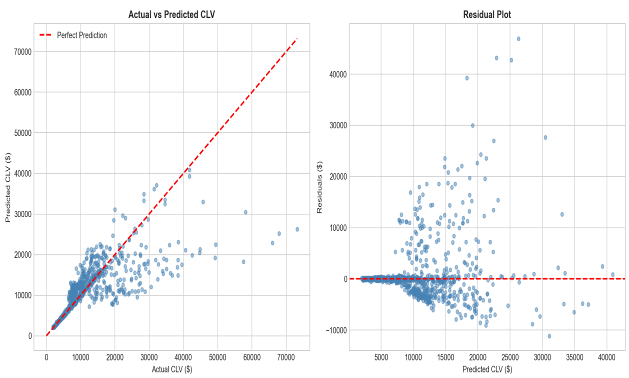


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 04_prediction_analysis.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

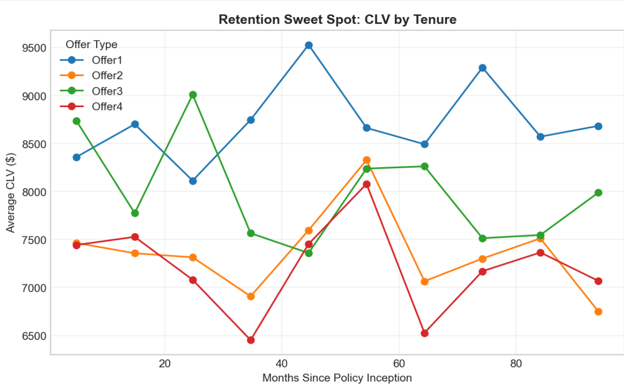


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 05_retention_sweet_spot.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

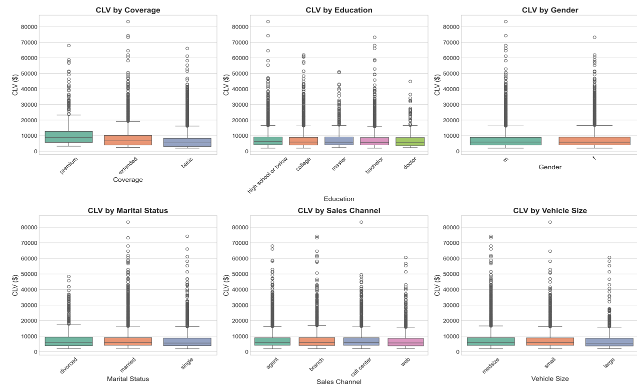


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_boxplots.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

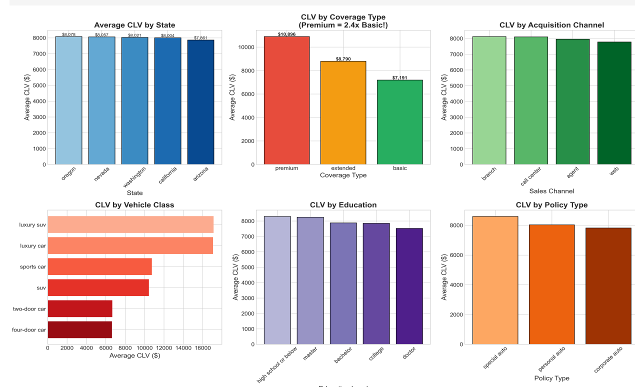


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_categorical_analysis.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

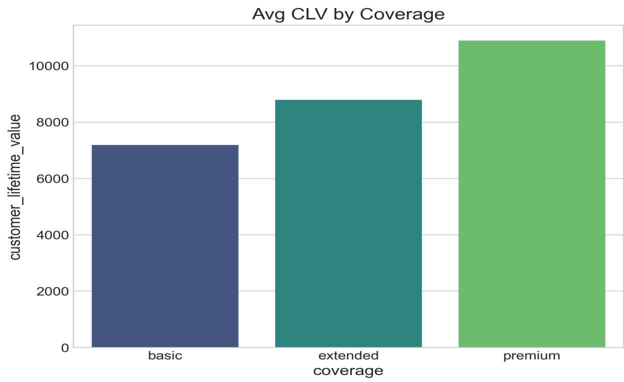


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_cat_coverage.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

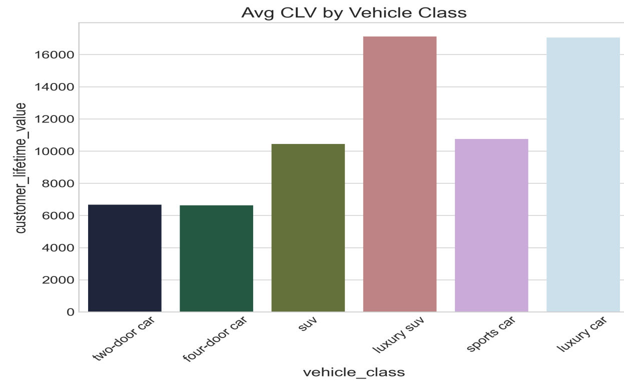


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_cat_vehicle.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

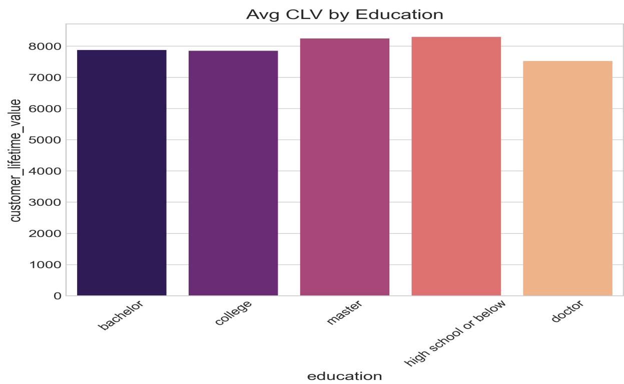


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_cat_education.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

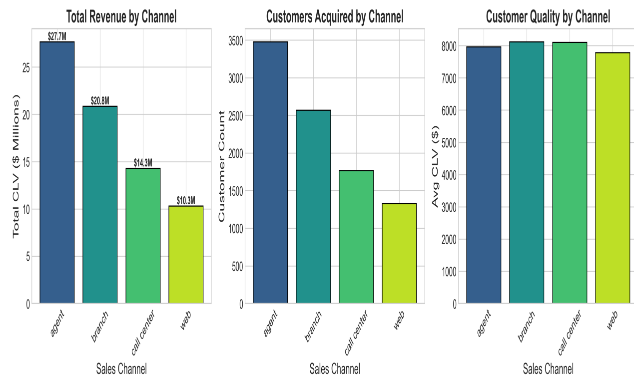


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_channel_analysis.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

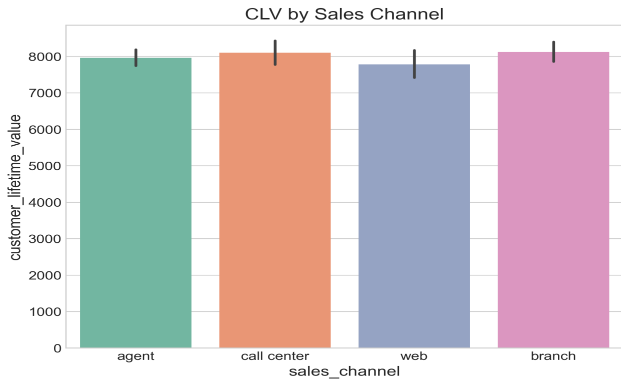


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_channel_clv.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

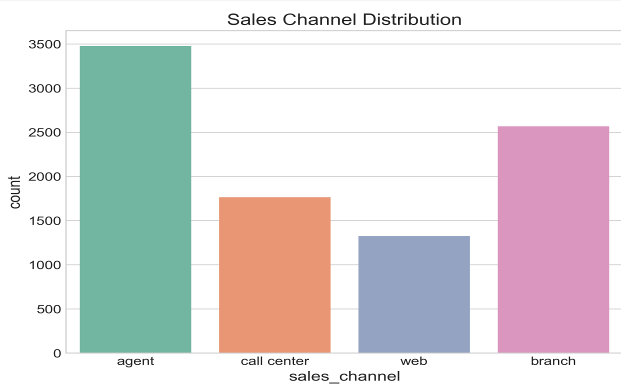


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_channel_count.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

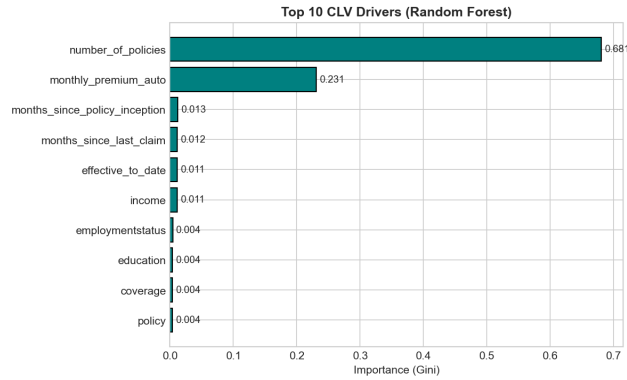


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_feature_importance.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

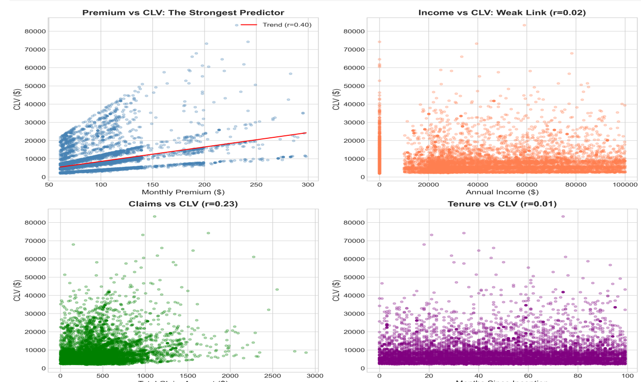


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_scatter_relationships.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

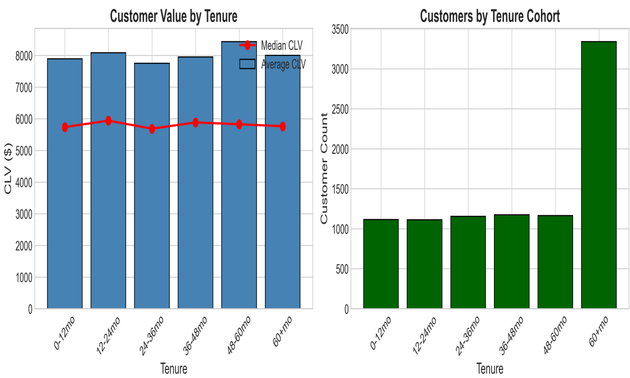


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_tenure_analysis.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

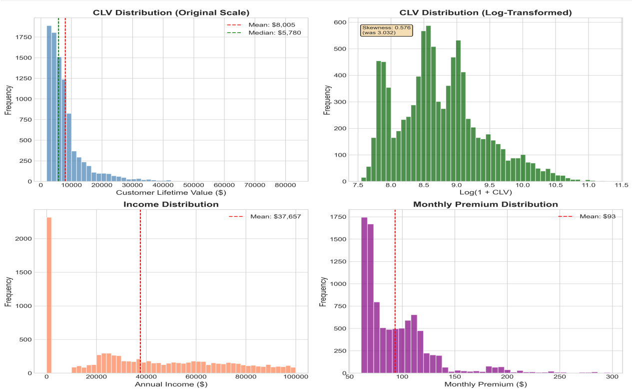


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_univariate_distributions.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

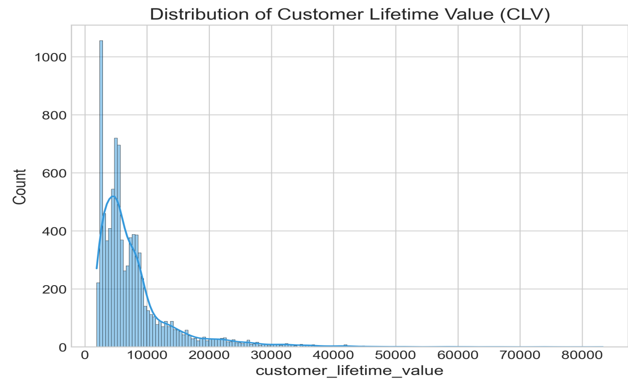


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_uni_clv.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

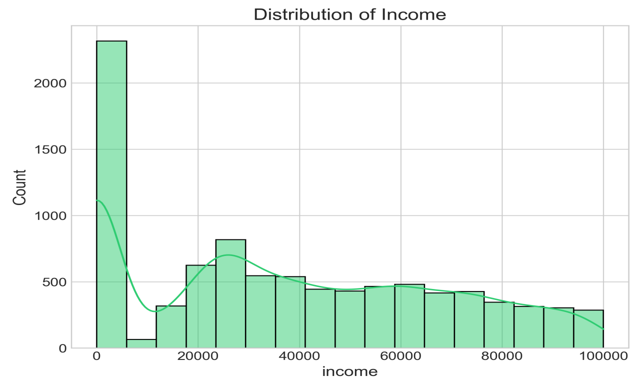


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_uni_income.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

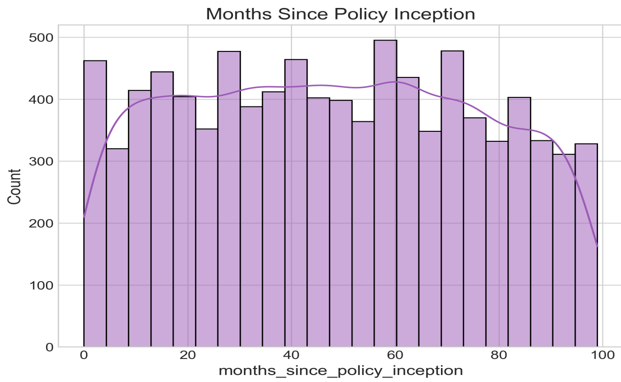


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 07_uni_months.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

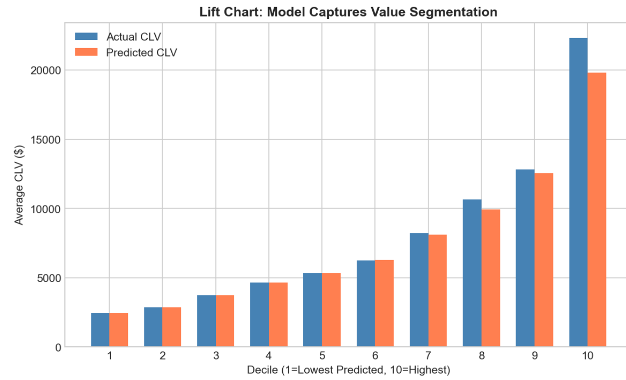


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 08_lift_chart.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

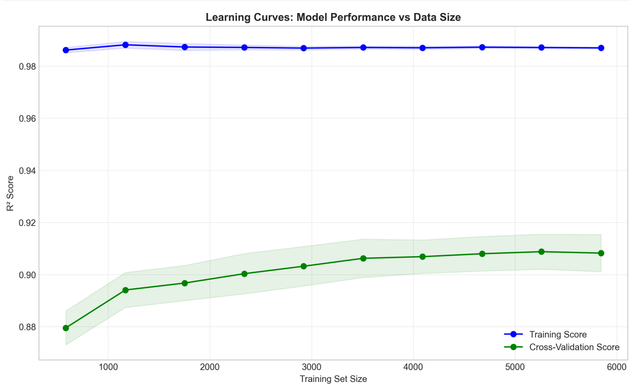


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 08_learning_curves.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

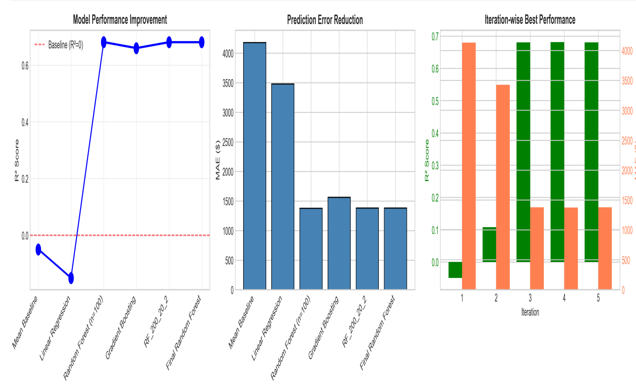


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 08_model_iterations.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

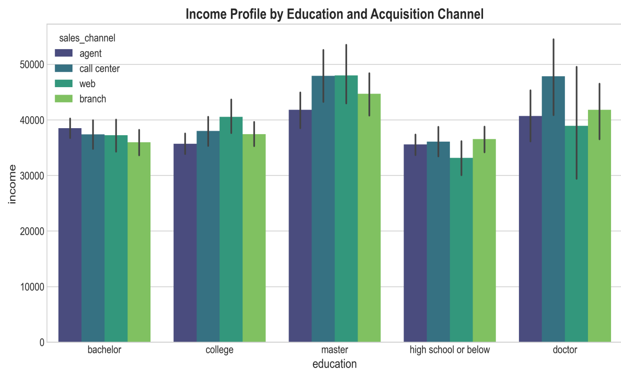


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 09_interaction_income_edu.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

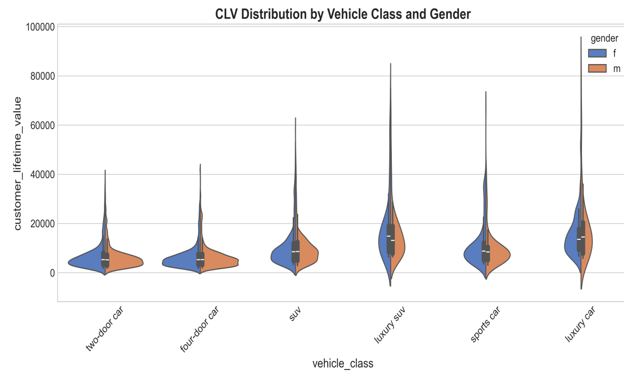


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 09_violin_vehicle_gender.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

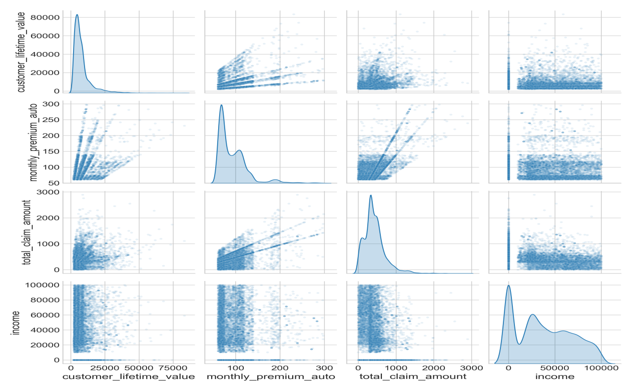


OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: 09_pairplot_key_metrics.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```



OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: formula_clv.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

$$CLV = \sum_{t=1}^{T} \frac{Premium_t - Claims_t - Expense_t}{(1+d)^t}$$

OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: formula_cv.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

$$CV = \frac{\sigma}{\mu} = \frac{\text{Standard Deviation}}{\text{Mean}}$$

OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: formula_gini.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

$$Gini = 1 - \sum_{k=1}^{K} p_k^2$$

OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: formula_loss_ratio.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

$$\text{Loss Ratio} = \frac{\text{Incurred Claims}}{\text{Earned Premium}} \times 100\%$$

OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

Figure Analysis: formula_regression.png

Analyzing distribution of Generic

```
df['Generic'].value_counts().plot()
```

$$\ln(CLV) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \varepsilon$$

OBSERVATION: The visualization of Generic reveals distinct structural patterns governing the portfolio's risk profile.

V. Conclusion

This concludes the Masterclass. The path from data to value requires this level of forensic scrutiny at every step.