Febin Francis
2327216

DOMAIN-SPECIFIC MODEL BUILDING

CIA-I:MLA

1. Business Understanding

Domain Introduction

The vehicle insurance sector plays a crucial role in safeguarding both individual vehicle owners and businesses against financial risks associated with owning and operating vehicles. This sector encompasses various types of insurance policies designed to protect vehicles, drivers, passengers, and third parties in the event of accidents, theft, or damage.

Problem Statement

How to predict customer life time and what all factors influences it for a vehicle insurance company

2. Data Understanding

Data Dictionary

Column	Description	Data Type
clv	Customer Lifetime Value (CLV) - a prediction of the net profit attributed to the entire future relationship with a customer.	Float
Response	Customer response to the offer (Yes/No).	String
Coverage	Type of insurance coverage (Basic/Extended/Premium).	String
Education	Education level of the customer (High School or Below/College/Bachelor/Master).	String
EmploymentStatus	Employment status of the customer (Employed/Unemployed/Medical Leave/Disabled).	String
Gender	Gender of the customer (M/F).	String
Income	Annual income of the customer in USD.	Integer (Nullable)
Location.Code	Type of location where the customer resides (Urban/Suburban/Rural).	String

Marital.Status	Marital status of the customer (Single/Married/Divorced).	String
Monthly.Premium.Auto	Monthly premium for the auto insurance in USD.	Integer
Months.Since.Last.Claim	Number of months since the last claim was made.	Integer
Months.Since.Policy.Inception	Number of months since the policy inception.	Integer
Number.of.Open.Complaints	Number of open complaints the customer has.	Integer
Number.of.Policies	Number of insurance policies the customer holds.	Integer
Policy.Type	Type of insurance policy (Personal Auto/Corporate Auto/Special Auto).	String
Policy	Specific insurance policy (e.g., Personal L1, Corporate L2).	String
Renew.Offer.Type	Type of renewal offer provided to the customer (Offer1/Offer2/Offer3/Offer4).	String
Sales.Channel	Channel through which the insurance was sold (Agent/Branch/Call Center/Web).	String
Total.Claim.Amount	Total amount of claims made by the customer in USD.	Float
Vehicle.Class	Class of the vehicle insured (Two-Door Car/Four-Door Car/SUV).	String
Vehicle.Size	Size of the vehicle insured (Small/Medsize/Large).	String

3. Data Preparation

```
> summary(insurance_data)
     ...1           clv           Response      Coverage                    Education          EmploymentStatus Gender
 Min.   :   1   Min.   : 1898   No :7400   Basic    :5314   Bachelor            :2592   Disabled     : 381   F:4390
 1st Qu.:2280   1st Qu.: 3858   Yes:1230   Extended:2563   College             :2549   Employed     :5366   M:4240
 Median :4548   Median : 5569              Premium : 753   Doctor              : 324   Medical Leave: 407
 Mean   :4561   Mean   : 6725                              High School or Below:2465   Retired      : 264
 3rd Qu.:6853   3rd Qu.: 8456                              Master              : 700   Unemployed   :2212
 Max.   :9134   Max.   :21235

     Income       Location.Code   Marital.Status Monthly.Premium.Auto Months.Since.Last.Claim Months.Since.Policy.Inception
 Min.   :    0   Rural   :1672   Divorced:1285   Min.   : 61.00       Min.   : 0.00           Min.   : 0
 1st Qu.:    0   Suburban:5467   Married :4999   1st Qu.: 68.00       1st Qu.: 6.00           1st Qu.:24
 Median :33817   Urban   :1491   Single  :2346   Median : 81.00       Median :14.00           Median :48
 Mean   :37586                                   Mean   : 91.68       Mean   :15.08           Mean   :48
 3rd Qu.:62251                                   3rd Qu.:108.00       3rd Qu.:23.00           3rd Qu.:71
 Max.   :99981                                   Max.   :298.00       Max.   :35.00           Max.   :99

 Number.of.Open.Complaints Number.of.Policies           Policy.Type                Policy      Renew.Offer.Type       Sales.Channel
 Min.   :0.0000            Min.   :1.000       Corporate Auto:1866   Personal L3 :3234   Offer1:3518        Agent      :3305
 1st Qu.:0.0000            1st Qu.:1.000       Personal Auto :6412   Personal L2 :2001   Offer2:2783        Branch     :2400
 Median :0.0000            Median :2.000       Special Auto  : 352   Personal L1 :1177   Offer3:1355        Call Center:1665
 Mean   :0.3891            Mean   :3.021                             Corporate L3: 960   Offer4: 974        Web        :1260
 3rd Qu.:0.0000            3rd Qu.:4.000                             Corporate L2: 569
 Max.   :5.0000            Max.   :9.000                             Corporate L1: 337
                                                                     (Other)     : 352
 Total.Claim.Amount        Vehicle.Class    Vehicle.Size
 Min.   :   0.099   Four-Door Car:4461   Large  : 903
 1st Qu.: 268.881   Luxury Car   : 127   Medsize:6061
 Median : 376.880   Luxury SUV   : 148   Small  :1666
 Mean   : 427.132   Sports Car   : 432
 3rd Qu.: 542.400   SUV          :1636
 Max.   :2893.240   Two-Door Car :1826
```

4. <u>Modelling</u>

❖ Multiple Linear Regression

```
Coefficients: (2 not defined because of singularities)
                               Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   2.413e+03  5.260e+02   4.588 4.54e-06 ***
ResponseYes                  -1.676e+02  1.195e+02  -1.402  0.16083    
CoverageExtended              2.507e+02  1.462e+02   1.715  0.08644 .  
CoveragePremium               6.374e+02  3.092e+02   2.061  0.03929 *  
EducationCollege              1.642e+02  9.799e+01   1.676  0.09376 .  
EducationDoctor              -6.715e+01  2.078e+02  -0.323  0.74663    
EducationHigh School or Below  2.649e+02  9.947e+01   2.663  0.00776 ** 
EducationMaster               2.831e+02  1.508e+02   1.877  0.06051 .  
EmploymentStatusEmployed      3.304e+02  2.036e+02   1.622  0.10474    
EmploymentStatusMedical Leave  9.018e+00  2.505e+02   0.036  0.97128    
EmploymentStatusRetired      -3.049e+01  2.895e+02  -0.105  0.91612    
EmploymentStatusUnemployed    1.771e+02  2.055e+02   0.862  0.38873    
GenderM                      -5.738e+01  7.619e+01  -0.753  0.45141    
Income                        3.375e-03  2.210e-03   1.527  0.12685    
Location.CodeSuburban         8.650e+01  1.514e+02   0.572  0.56768    
Location.CodeUrban            6.951e+01  1.388e+02   0.501  0.61642    
Marital.StatusMarried        -7.258e+01  1.120e+02  -0.648  0.51713    
Marital.StatusSingle         -2.931e+02  1.298e+02  -2.258  0.02399 *  
Monthly.Premium.Auto          4.020e+01  5.995e+00   6.706 2.13e-11 ***
Months.Since.Last.Claim       1.796e+00  3.767e+00   0.477  0.63366    
Months.Since.Policy.Inception -2.952e+00  1.365e+00  -2.162  0.03062 *  
Number.of.Open.Complaints    -1.625e+02  4.108e+01  -3.957 7.66e-05 ***
Number.of.Policies            2.636e+02  1.555e+01  16.954  < 2e-16 ***
Policy.TypePersonal Auto     -4.121e+02  2.006e+02  -2.054  0.03998 *  
Policy.TypeSpecial Auto       2.746e+02  3.550e+02   0.774  0.43922    
PolicyCorporate L2           -5.375e+02  2.410e+02  -2.231  0.02573 *  
PolicyCorporate L3           -5.020e+02  2.219e+02  -2.262  0.02374 *  
PolicyPersonal L1             3.180e+01  1.193e+02   0.267  0.78978    
PolicyPersonal L2             1.208e+02  9.969e+01   1.212  0.22551    
PolicyPersonal L3                    NA         NA      NA       NA    
PolicySpecial L1             -1.391e+01  5.398e+02  -0.026  0.97944    
PolicySpecial L2             -5.093e+02  4.118e+02  -1.237  0.21623    
PolicySpecial L3                     NA         NA      NA       NA    
Renew.Offer.TypeOffer2       -6.871e+02  9.489e+01  -7.241 4.82e-13 ***
```

```
Renew.Offer.TypeOffer2       -6.871e+02  9.489e+01  -7.241 4.82e-13 ***
Renew.Offer.TypeOffer3       -3.005e+02  1.149e+02  -2.616  0.00891 ** 
Renew.Offer.TypeOffer4       -8.300e+02  1.335e+02  -6.216 5.33e-10 ***
Sales.ChannelBranch          -6.021e+01  9.474e+01  -0.636  0.52509    
Sales.ChannelCall Center      5.918e+01  1.068e+02   0.554  0.57947    
Sales.ChannelWeb             -3.866e+01  1.179e+02  -0.328  0.74292    
Total.Claim.Amount           -3.048e-01  2.779e-01  -1.097  0.27278    
Vehicle.ClassLuxury Car       1.050e+03  8.260e+02   1.271  0.20381    
Vehicle.ClassLuxury SUV       9.205e+02  8.215e+02   1.120  0.26255    
Vehicle.ClassSports Car       4.890e+02  3.114e+02   1.570  0.11637    
Vehicle.ClassSUV              7.233e+02  2.723e+02   2.656  0.00792 ** 
Vehicle.ClassTwo-Door Car     1.569e+02  9.748e+01   1.609  0.10757    
Vehicle.SizeMedsize           1.020e+02  1.253e+02   0.814  0.41566    
Vehicle.SizeSmall             1.659e+02  1.461e+02   1.135  0.25634    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3498 on 8585 degrees of freedom
Multiple R-squared:  0.2138,	Adjusted R-squared:  0.2097 
F-statistic: 53.05 on 44 and 8585 DF,  p-value: < 2.2e-16
```

Splitting of data 70% training 30 % testing output

```
Coefficients: (2 not defined because of singularities)
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   1.544e+03  6.270e+02   2.463 0.013799 *
ResponseYes                  -1.816e+02  1.449e+02  -1.254 0.210055
CoverageExtended              1.720e+02  1.735e+02   0.992 0.321427
CoveragePremium               3.655e+02  3.671e+02   0.996 0.319527
EducationCollege              6.578e+01  1.171e+02   0.562 0.574446
EducationDoctor               4.264e+01  2.426e+02   0.176 0.860474
EducationHigh School or Below 2.359e+02  1.194e+02   1.976 0.048178 *
EducationMaster               2.593e+02  1.776e+02   1.460 0.144370
EmploymentStatusEmployed      5.758e+02  2.473e+02   2.328 0.019948 *
EmploymentStatusMedical Leave 1.915e+02  3.010e+02   0.636 0.524518
EmploymentStatusRetired       4.771e+02  3.510e+02   1.359 0.174145
EmploymentStatusUnemployed    3.580e+02  2.496e+02   1.434 0.151515
GenderM                       3.859e+01  9.086e+01   0.425 0.671078
Income                        2.556e-03  2.618e-03   0.976 0.329022
Location.CodeSuburban         2.046e+02  1.803e+02   1.135 0.256441
Location.CodeUrban            1.617e+01  1.665e+02   0.097 0.922669
Marital.StatusMarried         9.730e-01  1.333e+02   0.007 0.994176
Marital.StatusSingle         -2.518e+02  1.549e+02  -1.625 0.104227
Monthly.Premium.Auto          4.667e+01  7.122e+00   6.553 6.10e-11 ***
Months.Since.Last.Claim       2.214e+00  4.499e+00   0.492 0.622632
Months.Since.Policy.Inception -1.764e+00 1.627e+00  -1.084 0.278337
Number.of.Open.Complaints    -1.686e+02  4.919e+01  -3.427 0.000615 ***
Number.of.Policies            2.646e+02  1.847e+01  14.326  < 2e-16 ***
Policy.TypePersonal Auto     -3.289e+02  2.340e+02  -1.405 0.159988
Policy.TypeSpecial Auto       7.356e+02  4.234e+02   1.738 0.082349 .
PolicyCorporate L2           -4.933e+02  2.837e+02  -1.739 0.082119 .
PolicyCorporate L3           -4.425e+02  2.604e+02  -1.700 0.089251 .
PolicyPersonal L1             4.665e+00  1.414e+02   0.033 0.973688
PolicyPersonal L2             8.083e+01  1.190e+02   0.679 0.496937
PolicyPersonal L3                    NA         NA      NA       NA
PolicySpecial L1             -5.704e+02  6.597e+02  -0.865 0.387287
PolicySpecial L2             -6.999e+02  4.821e+02  -1.452 0.146620
PolicySpecial L3                     NA         NA      NA       NA
```


```
PolicySpecial L1            -5.704e+02  6.597e+02  -0.865 0.387287
PolicySpecial L2            -6.999e+02  4.821e+02  -1.452 0.146620
PolicySpecial L3                    NA         NA      NA       NA
```


```
Renew.Offer.TypeOffer2      -7.474e+02  1.132e+02  -6.603 4.37e-11 ***
Renew.Offer.TypeOffer3      -3.622e+02  1.367e+02  -2.650 0.008076 **
Renew.Offer.TypeOffer4      -8.448e+02  1.593e+02  -5.303 1.18e-07 ***
Sales.ChannelBranch         -5.331e+01  1.129e+02  -0.472 0.636847
Sales.ChannelCall Center     6.596e+01  1.276e+02   0.517 0.605183
Sales.ChannelWeb            -4.627e+01  1.412e+02  -0.328 0.743129
Total.Claim.Amount          -5.141e-01  3.368e-01  -1.527 0.126917
Vehicle.ClassLuxury Car      3.291e+02  9.732e+02   0.338 0.735218
Vehicle.ClassLuxury SUV      4.557e+02  9.722e+02   0.469 0.639267
Vehicle.ClassSports Car      2.699e+02  3.738e+02   0.722 0.470409
Vehicle.ClassSUV             5.438e+02  3.217e+02   1.690 0.091003 .
Vehicle.ClassTwo-Door Car    1.811e+02  1.161e+02   1.561 0.118675
Vehicle.SizeMedsize          1.549e+02  1.503e+02   1.031 0.302759
Vehicle.SizeSmall            1.677e+02  1.748e+02   0.959 0.337445
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3487 on 5997 degrees of freedom
Multiple R-squared:  0.2234,	Adjusted R-squared:  0.2177
F-statistic:  39.2 on 44 and 5997 DF,  p-value: < 2.2e-16

> # Predict using the test data
> predictions_mlr <- predict(model, newdata = test_data)
> # Calculate RMSE (Root Mean Squared Error)
> rmse_mlr <- sqrt(mean((test_data$clv - predictions_mlr)^2))
> print(paste("RMSE MLR: ", rmse_mlr))
[1] "RMSE MLR:  3534.39271015111"
```

❖ Lasso Regression

```
[1] "Lasso Coefficients:"
> print(lasso.coef)
48 x 1 sparse Matrix of class "dgCMatrix"
                               lambda.min
(Intercept)                  1.672920e+03
(Intercept)                  .
ResponseYes                  .
CoverageExtended             1.204017e+01
CoveragePremium              7.556161e+01
EducationCollege             .
EducationDoctor              .
EducationHigh School or Below 8.418062e+01
EducationMaster              4.201135e+01
EmploymentStatusEmployed     2.868884e+02
EmploymentStatusMedical Leave .
EmploymentStatusRetired      .
EmploymentStatusUnemployed   .
GenderM                      .
Income                       6.470442e-04
Location.CodeSuburban        .
Location.CodeUrban           .
Marital.StatusMarried        .
Marital.StatusSingle        -1.650339e+02
Monthly.Premium.Auto         4.708706e+01
Months.Since.Last.Claim      .
Months.Since.Policy.Inception .
Number.of.Open.Complaints   -1.170992e+02
Number.of.Policies           2.513009e+02
Policy.TypePersonal Auto     .
Policy.TypeSpecial Auto      2.473488e+02
PolicyCorporate L2          -1.427301e+01
PolicyCorporate L3          -1.236343e+01
```

```
PolicyPersonal L1             .
PolicyPersonal L2             .
PolicyPersonal L3             .
PolicySpecial L1              .
PolicySpecial L2              .
PolicySpecial L3              4.530357e+02
Renew.Offer.TypeOffer2       -5.407897e+02
Renew.Offer.TypeOffer3       -1.077951e+02
Renew.Offer.TypeOffer4       -5.400941e+02
Sales.ChannelBranch           .
Sales.ChannelCall Center      .
Sales.ChannelWeb              .
Total.Claim.Amount            .
Vehicle.ClassLuxury Car       .
Vehicle.ClassLuxury SUV       .
Vehicle.ClassSports Car       .
Vehicle.ClassSUV              2.997723e+02
Vehicle.ClassTwo-Door Car     2.304763e+01
Vehicle.SizeMedsize           .
Vehicle.SizeSmall             .
```

```
> lasso_model$lambda.min
[1] 40.20132
```

> # Calculate RMSE (Root Mean Squared Error)


```
> lasso_model$lambda.min
[1] 40.20132
```

❖ Ridge Regression

```
[1] "Ridge Coefficients:"
> print(ridge.coef)
48 x 1 sparse Matrix of class "dgCMatrix"
                                lambda.min
(Intercept)                   2.649444e+03
(Intercept)                   .
ResponseYes                  -1.662605e+02
CoverageExtended              4.178857e+02
CoveragePremium               9.319237e+02
EducationCollege              5.157198e+01
EducationDoctor               2.966980e+01
EducationHigh School or Below 2.087491e+02
EducationMaster               2.475342e+02
EmploymentStatusEmployed      4.197925e+02
EmploymentStatusMedical Leave 3.540290e+01
EmploymentStatusRetired       3.053177e+02
EmploymentStatusUnemployed    1.705673e+02
GenderM                       2.903210e+01
Income                        2.613653e-03
Location.CodeSuburban         4.564879e+01
Location.CodeUrban           -7.259905e+01
Marital.StatusMarried         1.079543e+01
Marital.StatusSingle         -2.525121e+02
Monthly.Premium.Auto          3.180833e+01
Months.Since.Last.Claim       2.171214e+00
Months.Since.Policy.Inception -1.511887e+00
Number.of.Open.Complaints    -1.582609e+02
Number.of.Policies            2.550403e+02
Policy.TypePersonal Auto     -1.652850e+02
Policy.TypeSpecial Auto       1.876035e+02
PolicyCorporate L2           -3.957368e+02
PolicyCorporate L3           -3.558635e+02
PolicyPersonal L1            -8.349565e+01
PolicyPersonal L2            -1.501060e+01
PolicyPersonal L3            -8.572937e+01

PolicySpecial L1              6.280041e+00
PolicySpecial L2             -8.314241e+01
PolicySpecial L3              5.948532e+02
Renew.Offer.TypeOffer2       -7.009323e+02
Renew.Offer.TypeOffer3       -3.256507e+02
Renew.Offer.TypeOffer4       -7.837279e+02
Sales.ChannelBranch          -5.378796e+01
Sales.ChannelCall Center      5.759056e+01
Sales.ChannelWeb             -5.693115e+01
Total.Claim.Amount           -8.537728e-02
Vehicle.ClassLuxury Car       1.864894e+03
Vehicle.ClassLuxury SUV       2.021935e+03
Vehicle.ClassSports Car       7.851995e+02
Vehicle.ClassSUV              1.043355e+03
Vehicle.ClassTwo-Door Car     1.493667e+02
Vehicle.SizeMedsize           1.432011e+02
Vehicle.SizeSmall             1.460423e+02
```

```
> ridge_model$lambda.min
[1] 166.1124
```

```
[1] "RMSE Ridge:  3529.33868959327"
```

5. Evaluation

1. **Multiple Linear Regression (MLR)**:
    - **RMSE MLR**: 3534.39
    - This indicates the RMSE (Root Mean Squared Error) obtained from the multiple linear regression model. RMSE measures the average deviation of predicted values from actual values. In this case, an RMSE of 3534.39 suggests that, on average, the model's predictions are approximately 3534.39 units away from the actual values.
2. **Lasso Regression**:
    - **Lasso_model$lambda.min**: 40.20132
    - **RMSE Lasso**: 3531.79
    - Lasso regression is a type of regression that applies L1 regularization, which can lead to sparse models (some coefficients are exactly zero). The **lambda.min** parameter indicates the regularization strength that resulted in the lowest RMSE during cross-validation or model selection. The RMSE of 3531.79 suggests slightly improved predictive accuracy compared to MLR, but the improvement is marginal.
3. **Ridge Regression**:
    - **ridge_model$Lambda.min**: 166.1124
    - **RMSE Ridge**: 3529.34
    - Ridge regression applies L2 regularization, which penalizes the sum of squared coefficients. The **Lambda.min** parameter similarly indicates the optimal regularization strength that minimized RMSE. The RMSE of 3529.34 shows further marginal improvement over both MLR and Lasso, indicating that Ridge regression has slightly better predictive performance in this context.

Lower RMSE indicates better model performance in terms of prediction accuracy. In this case, Ridge regression has the lowest RMSE, suggesting it provides the best predictions among the models evaluated.

Best Model: Ridge

6. <u>Deployment</u>

Model Output Interpretation of Ridge Regression

Intercept:

- The intercept term (Intercept) is approximately 2.649444e+03.

Predictor Variables:

- **ResponseYes**: For respondents answering "Yes", there is a decrease in the response variable, estimated at approximately -166.26.
- **CoverageExtended**: Having extended coverage increases the response variable by approximately 417.89.
- **CoveragePremium**: Premium coverage increases the response variable by about 931.92.
- **Education**: Different levels of education impact the response variable differently:
    - **College**: Increases the response variable by approximately 51.57.
    - **Doctor**: Increases the response variable by approximately 29.67.
    - **High School or Below**: Increases the response variable by about 208.75.
    - **Master**: Increases the response variable by approximately 247.53.
- **EmploymentStatus**: Various employment statuses affect the response variable:
    - **Employed**: Increases the response variable by approximately 419.79.
    - **Medical Leave**: Increases the response variable by about 35.40.
    - **Retired**: Increases the response variable by approximately 305.32.
    - **Unemployed**: Increases the response variable by about 170.57.
- **GenderM**: Being male increases the response variable by approximately 29.03.
- **Income**: Every unit increase in income increases the response variable by approximately 0.00261.
- **Location.Code**: Location codes affect the response variable differently:
    - **Suburban**: Increases the response variable by about 45.65.
    - **Urban**: Decreases the response variable by approximately -72.60.
- **Marital.Status**: Marital status impacts the response variable:
    - **Married**: Increases the response variable by approximately 10.80.
    - **Single**: Decreases the response variable by about -252.51.
- **Monthly.Premium.Auto**: Increases the response variable by approximately 31.81.
- **Months.Since.Last.Claim**: Increases the response variable by about 2.17.
- **Months.Since.Policy.Inception**: Decreases the response variable by approximately -1.51.
- **Number.of.Open.Complaints**: Increases the response variable by approximately -158.26.
- **Number.of.Policies**: Increases the response variable by approximately 255.04.
- **Policy.Type**: Different policy types impact the response variable:

    - **Personal Auto**: Decreases the response variable by approximately -165.28.
    - **Special Auto**: Increases the response variable by about 187.60.
    - **Corporate L2, L3**: Decrease the response variable significantly.
- **Renew.Offer.Type**: Different offer types affect the response variable:
    - **Offer2, Offer3**: Decrease the response variable significantly.
    - **Offer4**: Decreases the response variable by approximately -783.73.
- **Sales.Channel**: Sales channels impact the response variable:
    - **Branch**: Decreases the response variable by approximately -53.79.
    - **Call Center**: Increases the response variable by about 57.59.
    - **Web**: Decreases the response variable by approximately -56.93.
- **Total.Claim.Amount**: Decreases the response variable by approximately -0.085.
- **Vehicle.Class**: Vehicle classes impact the response variable differently:
    - **Luxury Car, Luxury SUV**: Increase the response variable significantly.
    - **Sports Car, SUV, Two-Door Car**: Increase the response variable to varying degrees.
- **Vehicle.Size**: Vehicle size impacts the response variable:
    - **Medsize, Small**: Increase the response variable moderately.

These coefficients indicate how each predictor variable influences the response variable (likely an outcome or prediction), considering the regularization effect of Ridge regression. The sign and magnitude of each coefficient provide insights into the relationships between predictors and the response, adjusted for multicollinearity and model complexity.

Model Interpretation from the Business Point of View

Factors Influencing Customer Lifetime Value (CLV):

1. **Response to Offer (ResponseYes):**
    - Customers responding "Yes" to offers are associated with a decrease in CLV by approximately $166.26. This suggests that targeting customers less responsive to offers might be more cost-effective.
2. **Insurance Coverage (CoverageExtended, CoveragePremium):**
    - Offering extended or premium coverage increases CLV significantly. For instance, premium coverage increases CLV by about $931.92. This indicates potential revenue gains from upselling higher coverage options.
3. **Education and Employment Status:**
    - **Education (College, Doctor, High School or Below, Master):**
        - Higher education levels generally correlate with higher CLV. For example, customers with a Master's degree have a CLV increase of approximately $247.53.

- **Employment Status (Employed, Medical Leave, Retired, Unemployed):**
  - Employed customers tend to have the highest CLV increase ($419.79), followed by retired and unemployed customers. This insight can guide targeted marketing efforts towards employed individuals.

4. **Demographic Factors (Gender, Marital Status):**
   - **Gender (GenderM):**
     - Male customers show a slight increase in CLV ($29.03). This demographic insight can inform gender-specific marketing strategies.
   - **Marital Status (Married, Single):**
     - Married customers have a modest increase in CLV ($10.80), whereas single customers show a decrease in CLV (-$252.51). Targeting married individuals might yield higher CLV opportunities.
5. **Financial Metrics (Income, Total Claim Amount):**
   - **Income:**
     - Higher income correlates with a slight increase in CLV ($0.00261 per unit increase). This suggests that wealthier customers might represent higher CLV.
   - **Total Claim Amount:**
     - Higher total claim amounts decrease CLV (-$0.085). Monitoring and managing claim amounts could help maintain or increase CLV.
6. **Policy and Sales Channel:**
   - **Policy Type (Personal Auto, Special Auto, Corporate L2, L3):**
     - Policies like Special Auto increase CLV, whereas Personal Auto and Corporate L2, L3 decrease CLV. Adjusting product offerings towards policies with higher CLV potential could optimize revenue.
   - **Renew Offer Type (Offer2, Offer3, Offer4):**
     - Renewal offers such as Offer4 significantly decrease CLV (-$783.73). This insight can guide strategic adjustments in renewal offer strategies.
   - **Sales Channel (Branch, Call Center, Web):**
     - The Call Center sales channel increases CLV ($57.59), while Branch and Web channels decrease CLV. Focusing on more effective sales channels can enhance CLV.
7. **Vehicle Details (Vehicle Class, Vehicle Size):**
   - **Vehicle Class (Luxury Car, Luxury SUV, Sports Car, SUV, Two-Door Car):**
     - Vehicles like Luxury Cars and SUVs increase CLV significantly, suggesting potential revenue gains from insuring higher-value vehicles.
   - **Vehicle Size (Medsize, Small):**
     - Medium and small-sized vehicles moderately increase CLV, indicating opportunities across different vehicle segments.

Business Insights and Recommendations:

- **Targeted Marketing Strategies:**
  - Tailor marketing efforts towards customers with higher education levels, employed status, and higher income, as they exhibit higher CLV potential.
  - Focus on promoting premium coverage options to enhance CLV through upselling.
- **Policy and Offer Optimization:**
  - Optimize product offerings towards policies (e.g., Special Auto) and renewal offers (e.g., Offer2) that increase CLV.
  - Review and adjust renewal offer strategies to mitigate the significant CLV decrease associated with Offer4.
- **Sales Channel Effectiveness:**
  - Invest in the Call Center channel, which shows positive CLV impact, while considering improvements or adjustments for Branch and Web channels.
- **Customer Segmentation:**
  - Segment customers based on demographic factors like gender and marital status to better tailor marketing and service offerings, aiming to maximize CLV.
- **Risk Management:**
  - Monitor and manage total claim amounts effectively to minimize their negative impact on CLV, ensuring sustainable profitability.

By leveraging these insights, the vehicle insurance company can optimize its strategies to enhance customer lifetime value, drive revenue growth, and improve overall business performance.

Conclusion

In conclusion, the predictive modeling efforts focused on estimating customer lifetime value (CLV) in the vehicle insurance sector have provided valuable insights into the factors influencing CLV. Through multiple linear regression (MLR), Lasso regression, and Ridge regression, it was determined that Ridge regression yielded the best predictive performance with the lowest RMSE. The model interpretation highlighted significant influences on CLV, such as customer responses to offers, coverage types, demographic variables (education, employment status, gender, marital status), financial metrics (income, total claim amount), policy details, renewal offers, sales channels, and vehicle characteristics. These insights enable strategic recommendations for targeted marketing, optimization of product offerings and renewal strategies, channel effectiveness enhancement, customer segmentation, and risk management to maximize CLV and ensure sustainable business growth in the competitive vehicle insurance market.
GitHub Link: febin-francis/Vehicle-Insurance