BAN 5743-Exercise 4 (10 Points) Solution

Exercise Description:

You will use the Medicare data for this assignment. The description for the data is found at the end of this assignment.

The Stillwater HHA was very excited about the information in your last report. However, they are not sure that the decision tree is the best way to look at this information. They are really wanting to expand their reach into others areas of the country and has asked you to re-examine the data provided by Medicare. Instead of looking a business rules, the Stillwater HHA is interested in predicting whether services are profitable and how profitable services can be, along with the important variables that predict these two items. Your task is to use SAS Enterprise Miner to import the **MedicareHomeHealthProvider.jmp** data table and use logistic regression to make answer the business questions.

Business Questions:

1. Are home health agencies profitable? (Profit_b 1=Yes/ 0 = No)
   a. What variables make a home health agencies profitable?
   b. Are certain geographic locations associated with profitable home health agencies?
2. How profitable can home health agencies be? (Proft = $ amount)
   a. What variables make home health agencies more profitable?
   b. Are their services that make home health agencies less profitable?
   c. Are certain geographic locations associated with high and low profitable home health agencies?

Instructions:

- Create a new project in SAS® Enterprise Miner™.
- Create a new library and import the dataset provided.
- Make sure the model roles and measurement levels are shown as in the table at the end of this document.
- Create a new diagram and add the imported dataset to the diagram.
- Add a data partition node to the diagram and connect it to the data source node. Assign the 65% of the data for training dataset and 35% of the data for validation dataset.
- Apply appropriate transformations to variables when asked using the Transform Node.
- When asked to impute, use count as input method for unknown class variable values, the mean for unknown interval variable values, and create unique imputation indicators (to be used in models) for all imputed inputs.

- Run a series of regressions to answer the business questions.
  a) Regression with imputation only
     i) Selection Model: Forward
     ii) Selection Criteria: Validation Error
  b) Regression with transformation and imputation
     i) Selection Model: Forward
     ii) Selection Criteria: Validation Error
  c) Full second-degree polynomial model plus all two-factor interactions
     i) Selection Model: Forward
     ii) Selection Criteria: Validation Error

Think about these items when you create your report write-up for the Stillwater Home Health Agency.

Your response should be written in business language using paragraphs and not responses to the below bullets only.

- **In preparation for regression, are any missing value imputations needed in this data? Why was the imputation of missing values not needed for decision trees?**
- **How many interval input variables have missing values in the whole data? How many class input variables have missing values in the whole data?**
- **What imputed value was used to replace the missing values? How many such missing values were replaced for each of the affected variables in the training data?**
- **In preparation for regression, are any transformations of the input variables (Interval) warranted in the data? Why or why not?**
- **Which metric would you use to evaluate the models? Which variables are included in the final model?**
- **Did the transformations result in better distributions for these variables?**
- **Are the selected variables different among the regressions? How about the validation ASE and Validation Misclassification rate? What do these metrics tell you about the impact of transformations on model performance?**
- **How many interaction variables are selected in the final model?**

Deliverables (please follow these instructions):

- *As you complete the exercise, create a report in Microsoft Word. In this report, answer the questions in the exercise description.*
- *Make sure you comment or explain and not just provide snapshots of data.*
- *Limit your report to no more than **7 pages** including tables and diagrams.*
- *Copy and paste or screen shot supporting tables/diagrams as needed to justify any of your answer. You may need to shrink your table/ diagrams but please ensure they are readable.*
- *Include any required data sets or codes/projects as requested as separate files.*
- *Make sure you print your name, student ID#, student email on the cover page of the report and turn-in the report as communicated by your instructor.*
- *Please also put a running header/footer with your name, on each page of your exercise solution report.*

Failure to follow these instructions will result in deduction of points

Data Description:

Medicare is the federal government program in the United States that provides health care coverage (health insurance) for citizens who are 65 years of age or over, under 65 and receiving Social Security Disability Insurance (SSDI) for a certain amount of time, or under 65 and with End-Stage Renal Disease (ESRD). Further, Medicare Part A (Hospital Insurance) and/or Medicare Part B (Medical Insurance) covers eligible home health services for beneficiaries (covered persons) like part-time or "intermittent" skilled nursing care, physical therapy, occupational therapy, speech-language pathology services, medical social services, and part-time or intermittent home health aide services (personal hands-on care). Usually, a home health care agency coordinates the services a beneficiary's doctor orders.

While the costs for home health services for Medicare beneficiaries (recipients of Medicare) is typically $0.00, there are significant charges incurred by the home health agencies. For example, a nurse, therapist or social worker may cost $70.00 to $100.00 an hour. An aide to take care of daily living needs, so called activities of daily living, may cost $10.00 to $25.00 an hour. This means that billing Medicare accurately and efficiently is extremely important for home health agencies to continue operating as a business whether they are for-profit or not.

The unit of payment under the HH PPS is a 60-day episode of care. A split percentage payment is made for most HH PPS episode periods. There are two payments – initial and final. The first payment is made in response to a Request for Anticipated Payment (RAP), and the last payment is paid in response to a claim. Added together, the first and last payments equal 100 % the permissible payment for the episode – usually a 60/40 split. If there are four or fewer visits provided to a patient in a 60-day episode (continuous period of time), payment to the home health agency is based on a national standardized per visit payment by discipline instead of an episode payment.

Variable	Label	Description	Type	Role
Provider ID	Provider ID	The 6-digit identification number for the home health agency on the claim.	Plain Text	ID
Agency Name	Agency Name	The home health agency name, as reported in the POS file.	Plain Text	Label
Street Address	Street Address	The home health agency address, as reported in the POS file.	Plain Text	Rejected
City	City	The city where the home health agency is located, as reported in the POS file.	Plain Text	Input
State	State	The state where the home health agency is located, as reported in POS file. The fifty U.S. states and the District of Columbia are reported by the state postal abbreviation.	Plain Text	Input
Zip Code	Zip Code	The home health agency's zip code, as reported in the POS file.	Plain Text	Input
TotalEpisNonLUPA	Total Episodes (non-LUPA)	Total count of non-LUPA episodes provided by a specific home health agency or in a unique HHRG category in the calendar year.	Number	Input
DistBenNonLUPA	Distinct Beneficiaries (non-LUPA)	Number of distinct Medicare beneficiaries receiving at least one non-LUPA home health episode in the calendar year. Beneficiaries may receive multiple home health episodes per year but are only counted once in this field.	Number	Input
AvgTotalVPENonLUPA	Average Number of Total Visits Per Episode (non-LUPA)	Average number of total visits provided by the HHA during a non-LUPA episode.	Number	Input
AvgSNVPENonLUPA	Average Number of Skilled Nursing Visits Per Episode (non-LUPA)	Average number of skilled nursing visits provided by the HHA during a non-LUPA episode.	Number	Input
AvgPTVPENonLUPA	Average Number of PT Visits Per Episode (non-LUPA)	Average number of physical therapy visits provided by the HHA during a non-LUPA episode.	Number	Input
AvgOTVPENonLUPA	Average Number of OT Visits Per Episode (non-LUPA)	Average number of occupational therapy visits provided by the HHA during a non-LUPA episode.	Number	Input

AvgSTVPENonLUPA	Average Number of ST Visits Per Episode (non-LUPA)	Average number of speech therapy visits provided by the HHA during a non-LUPA episode.	Number	Input
AvgHHAVPENonLUPA	Average Number of Home Health Aide Visits Per Episode (non-LUPA)	Average number of home health aide visits provided by the HHA during a non-LUPA episode.	Number	Input
AVGMedSocVPENonLUPA	Average Number of Medical-Social Visits Per Episode (non-LUPA)	Average number of medical-social visits provided by the HHA during a non-LUPA episode.	Number	Input
TotalHHAChargeNonLUPA	Total HHA Charge Amount (non-LUPA)	Total charges that the home health agency submitted for non-LUPA episodes.	Number	Rejected
TotalHHAMedPayNonLUPA	Total HHA Medicare Payment Amount (non-LUPA)	Total amount that Medicare paid for non-LUPA episodes. Home health services do not have any cost-sharing requirements and the Medicare payment amount will equal the allowed amount.	Number	Rejected
TotalHHAMedStandPayNonLUPA	Total HHA Medicare Standard Payment Amount (non-LUPA)	Total amount that Medicare paid for non-LUPA episodes adjusted for geographic differences in payment rates.	Number	Rejected
OutPayPercMedPayNonLUPA	Outlier Payments as a % Medicare Payment Amount (non-LUPA)	The % total Medicare payments for non-LUPA episodes paid to an HHA for outlier episodes.	Number	Input
TotLUPAEpis	Total LUPA Episodes	Total count of low utilization payment amount episodes provided by a specific HHA in the calendar year.	Number	Input
TotHHAMedPayLUPA	Total HHA Medicare Payment Amount for LUPAs	Total amount that Medicare paid for LUPA episodes provided by a specific HHA in the calendar year.	Number	Rejected
AvgAge	Average Age	Average age of beneficiaries. Beneficiary age is calculated at the end of the calendar year or at the time of death.	Number	Input
MaleBen	Male Beneficiaries	Number of male beneficiaries.	Number	Input
FemaleBen	Female Beneficiaries	Number of female beneficiaries.	Number	Input
NondualBen	Nondual Beneficiaries	Number of Medicare beneficiaries qualified to receive Medicare only benefits. Beneficiaries are classified as Medicare only entitlement if they received zero months of	Number	Input

		any Medicaid benefits (full or partial) in the given calendar year.		
DualBen	Dual Beneficiaries	Number of Medicare beneficiaries qualified to receive Medicare and Medicaid benefits. Beneficiaries are classified as Medicare and Medicaid entitlement if in any month in the given calendar year they were receiving full or partial Medicaid benefits.	Number	Input
WhiteBen	White Beneficiaries	Number of non-Hispanic white beneficiaries.	Number	Input
BlackBen	Black Beneficiaries	Number of non-Hispanic black or African American beneficiaries.	Number	Input
APIBen	Asian Pacific Islander Beneficiaries	Number of Asian Pacific Islander beneficiaries.	Number	Input
HispBen	Hispanic Beneficiaries	Number of Hispanic beneficiaries.	Number	Input
AIANBen	American Indian or Alaska Native Beneficiaries	Number of American Indian or Alaska Native beneficiaries.	Number	Input
OtherUnkBen	Other/ Unknown Beneficiaries	Number of beneficiaries with race not elsewhere classified.	Number	Input
AvgHCCScore	Average HCC Score	Average Hierarchical Condition Category (HCC) risk score of beneficiaries. Please refer to the "Additional Information" section of the Methodology document for more details on HCC risk scores.	Number	Input
AtrialFib	% Beneficiaries with Atrial Fibrillation	% beneficiaries meeting the CCW chronic condition algorithm for atrial fibrillation.	Number	Input
Alzheimers	% Beneficiaries with Alzheimer's	% beneficiaries meeting the CCW chronic condition algorithm for Alzheimer's, related disorders, or dementia.	Number	Input
Asthma	% Beneficiaries with Asthma	% beneficiaries meeting the CCW chronic condition algorithm for Asthma.	Number	Input
Cancer	% Beneficiaries with Cancer	% beneficiaries meeting the CCW chronic condition algorithms for cancer. Includes breast cancer, colorectal cancer, lung cancer and prostate cancer.	Number	Input
CHF	% Beneficiaries with CHF	% beneficiaries meeting the CCW chronic condition algorithm for heart failure.	Number	Input

KidneyDis	% Beneficiaries with Chronic Kidney Disease	% beneficiaries meeting the CCW chronic condition algorithm for chronic kidney disease.	Number	Input
COPD	% Beneficiaries with COPD	% beneficiaries meeting the CCW chronic condition algorithm for chronic obstructive pulmonary disease.	Number	Input
Depression	% Beneficiaries with Depression	% beneficiaries meeting the CCW chronic condition algorithm for depression.	Number	Input
Diabetes	% Beneficiaries with Diabetes	% beneficiaries meeting the CCW chronic condition algorithm for diabetes.	Number	Input
Hyperlipidemia	% Beneficiaries with Hyperlipidemia	% beneficiaries meeting the CCW chronic condition algorithm for hyperlipidemia.	Number	Input
Hypertension	% Beneficiaries with Hypertension	% beneficiaries meeting the CCW chronic condition algorithm for hypertension.	Number	Input
IHD	% Beneficiaries with IHD	% beneficiaries meeting the CCW chronic condition algorithm for ischemic heart disease.	Number	Input
Osteoporosis	% Beneficiaries with Osteoporosis	% beneficiaries meeting the CCW chronic condition algorithm for osteoporosis.	Number	Input
RAOA	% Beneficiaries with RA/OA	% beneficiaries meeting the CCW chronic condition algorithm for rheumatoid arthritis/osteoarthritis.	Number	Input
Schizophrenia	% Beneficiaries with Schizophrenia	% beneficiaries meeting the CCW chronic condition algorithm for schizophrenia and other psychotic disorders.	Number	Input
Stroke	% Beneficiaries with Stroke	% beneficiaries meeting the CCW chronic condition algorithm for stroke.	Number	Input
Revenue	Total Revenue from Medicare	Total revenue generated from Medicare payments	Number	Rejected
Profit	Revenue – Charges	Total revenue minus total charges	Number	Target2
Profit_b	Profitable or not	Was the HHA profitable	Binary	Target1

Where Are Home Health Agencies Profitable?

Background:

We will look into predicting whether services are profitable and how profitable services can be, along with the important variables that predict these two items .

Also, we will examine additional models to review the information available. This will be in additional to the decision tree models shared in the last weeks model.

Objective:

2) Look into whether the home health agencies are profitable or not. Also investigate variables that make them profitable. Further investigate any geographical locations impact on whether home health agencies are profitable or not.
3) Investigate how much profits are made by home health agencies. Also investigate whether any variables influence amount of profit that an agency making and significance of any geographical factors. And figure out services which are less profitable.

Analysis:

Following models were ran to determine the above objectives –

Variable : Profit (Yes/No)	ASE (Validation data)	RMSE (Validation data)	Model degree of freedom	Misclassification Rate (Validation data)
Regression with imputation only	0.198	0.445	29	0.307
Regression with transformation and imputation	**0.194**	**0.442**	**22**	**0.305**
Full second-degree polynomial model plus all two-factor interactions	0.190	0.436	46	0.294

Variable : Profit	ASE (Validation data)	Model degree of freedom	R-Square
Regression with imputation only	7.16E11	1	-
Regression with transformation and imputation	6.90E11	18	0.0797
Full second-degree polynomial model plus all two-factor interactions	**6.59E11**	**34**	**0.1563**

<u>Based on the above models for both the target variable following are the champion models:</u>

1. Variable : Profit (Yes/No) - Regression with transformation and imputation
    - Reason to select: The RMSE and Misclassification Rates for this model are the 2nd best with considerably lower model degree of freedom i.e. the model has higher prediction power with lower complexity

2. Variable : Profit - Full second-degree polynomial model plus all two-factor interactions
    - Reason to select : The R-Square for this model is the best with high model degree of freedom but with the increased complexity the R-Square is doubled, and the ASE also reduces

<u>Few points related to the chosen model and comparison with previous week's report:</u>

- For regression model we had to impute missing values, this additional step was not required during decision tree as the tree model considers missing values as a category of data and includes in the analysis. If we do not impute the data for the regression model, these records will be ignored for the analysis.
- 23 interval variables have missing values in the data set. Refer to the appendix for additional information.
- For missing values in the interval variables, we have used hot imputation method by replacing the values with the mean of the remaining values. For identification and analysis purposes we have retained the identifier column and used it in the analysis.
- For the regression model outlier or skewness or kurtosis impacts the result of the model. There were also leverage or influential points too in the data. All these issues were resolved by transforming the data – <u>Following transformation techniques were used</u>:
    - Log
    - Square
    - Square root
    - Range Standardizing
- These transformations have resulted in reduced skewness or kurtosis or outlier.
- To evaluate the logistic model to predict the Profit (Yes/No) target we have used the following metrics:
    - Model Degree of freedom or significant variables
    - Root Mean Square Error
    - Misclassification Rate
    - Average Square Error
- To evaluate the linear regression model to predict the Profit target we have used the following metrics:
    - Model Degree of freedom or significant variables
    - R-Square
    - Average Square Error

Result:

Are home health agencies profitable or not ?

58.5% of the home health agencies are profitable.

Factor important in deciding whether a home health agency profitable or not.

<u>Demographical information:</u>

- Number of Asian Pacific Islander beneficiaries
- Number of non-Hispanic black or African American beneficiaries
- Number of Hispanic beneficiaries.
- Average age of beneficiaries

<u>Benefit information</u>

- Number of Medicare beneficiaries qualified to receive Medicare and Medicaid benefits
- Number of Medicare beneficiaries qualified to receive Medicare only benefits

<u>Disease information</u>

- % beneficiaries meeting the CCW chronic condition algorithm for depression
- % beneficiaries meeting the CCW chronic condition algorithm for Alzheimer's, related disorders, or dementia
- % beneficiaries meeting the CCW chronic condition algorithm for rheumatoid arthritis/osteoarthritis

<u>Type of services information</u>

- Average number of medical-social visits provided by the HHA during a non-LUPA episode
- Average number of skilled nursing visits provided by the HHA during a non-LUPA episode
- Average number of speech therapy visits provided by the HHA during a non-LUPA episode
- Average number of occupational therapy visits provided by the HHA during a non-LUPA episode
- Average number of physical therapy visits provided by the HHA during a non-LUPA episode

<u>Other information</u>

- Number of distinct Medicare beneficiaries receiving at least one non-LUPA home health episode in the calendar year
- Total count of non-LUPA episodes provided by a specific home health agency or in a unique HHRG category in the calendar year
- Total amount that Medicare paid for non-LUPA episodes adjusted for geographic differences in payment rates
- Total count of low utilization payment amount episodes provided by a specific HHA in the calendar year
- Average number of total visits provided by the HHA during a non-LUPA episode

No geographical information determined whether a home health agency is profitable or not. However, based on the demography and patient information, regions can be identified where we can extend our business. When region is included in the analysis, the following three regions were profitable: Region I, Region II, and Region VIII.

How profitable can home health agencies be?
On an average a home health agency can be profit $12,600.

<u>Factor important in deciding how much profit a home health agency can make.</u>
<u>Diseases which impacts profit made by home health agencies:</u>

- % beneficiaries meeting the CCW chronic condition algorithm for depression
- % beneficiaries meeting the CCW chronic condition algorithm for rheumatoid arthritis/osteoarthritis
- % beneficiaries meeting the CCW chronic condition algorithm for Alzheimer's, related disorders, or dementia

<u>Demography of home health agencies which contribute to profit:</u>

- Number of Asian Pacific Islander beneficiaries
- Number of Hispanic beneficiaries
- Number of non-Hispanic black or African American beneficiaries
- Number of American Indian or Alaska Native beneficiaries
- Number of non-Hispanic white beneficiaries
- Average age of beneficiaries

<u>Services which contribute to profit:</u>

- Average number of skilled nursing visits provided by the HHA during a non-LUPA episode
- Average number of medical-social visits provided by the HHA during a non-LUPA episode
- Average number of home health aide visits provided by the HHA during a non-LUPA episode
- Average number of physical therapy visits provided by the HHA during a non-LUPA episode
- Average number of speech therapy visits provided by the HHA during a non-LUPA episode
- Average number of home health aide visits provided by the HHA during a non-LUPA episode

<u>Number of beneficiaries based on benefits:</u>

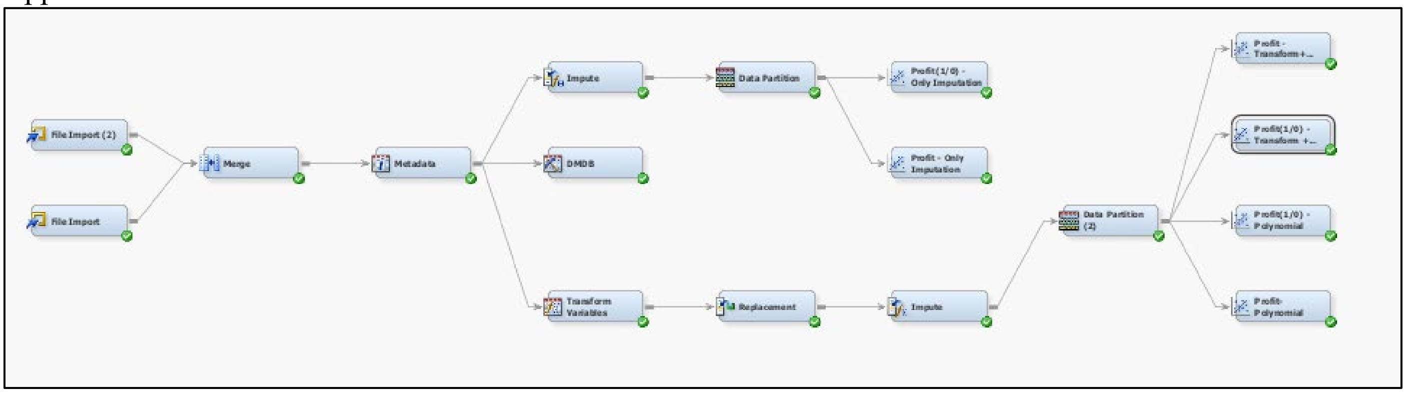
- Number of Medicare beneficiaries qualified to receive Medicare and Medicaid benefits
- Number of Medicare beneficiaries qualified to receive Medicare only benefits

<u>Additional factors impacting based on benefits:</u>

- Total count of non-LUPA episodes provided by a specific home health agency or in a unique HHRG category in the calendar year
- The % total Medicare payments for non-LUPA episodes paid to an HHA for outlier episodes
- Number of distinct Medicare beneficiaries receiving at least one non-LUPA home health episode in the calendar year
- Average number of total visits provided by the HHA during a non-LUPA episode.

No geographical information determined whether a home health agency is profitable or not. However, based on the demography and patient information, regions can be identified where we can extend our business. Further Region I and Region VIII were the most profitable. Region VIII contains Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming, which is the Midwest part of US. And geographically expand to Region I, which is in the far NE United States. While other region such as Region II, Region VII show negative effects so company should focus on improving the facilities in these regions.

Appendix:



Data Statistic for interval variable:

Variable	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
AIANBen	3,186	7,855	0	410	1	10	19	553
APIBen	5,071	5,970	0	2,104	9	46	27	1,051
AVGMedSocVPENonLUPA	0	11,041	0	3	0	0	4	28
Alzheimers	70	10,971	0	1	0	0	1	1
Asthma	0	11,041	0	1	0	0	1	4
AtrialFib	0	11,041	0	1	0	0	0	0
AvgAge	0	11,041	46	91	76	5	-1	2
AvgHCCScore	0	11,041	1	7	2	0	2	8
AvgHHAVPENonLUPA	0	11,041	0	60	3	4	4	25
AvgOTVPENonLUPA	0	11,041	0	12	1	1	2	6
AvgPTVPENonLUPA	0	11,041	0	26	5	3	1	2
AvgSNVPENonLUPA	0	11,041	0	86	10	5	4	35
AvgSTVPENonLUPA	0	11,041	0	4	0	0	4	22
AvgTotalVPENonLUPA	0	11,041	6	117	19	7	3	18
BlackBen	3,550	7,491	0	6,249	61	135	17	621
CHF	44	10,997	0	1	0	0	0	1
COPD	6	11,035	0	1	0	0	0	1
Cancer	0	11,041	0	1	0	0	1	3
Depression	117	10,924	0	1	0	0	0	0
Diabetes	239	10,802	0	1	0	0	0	0
DistBenNonLUPA	0	11,041	11	35,500	318	638	19	868
DualBen	2,118	8,923	0	16,294	123	238	37	2,407
FemaleBen	1,243	9,798	11	22,984	223	421	19	898
HispBen	4,956	6,085	0	5,555	48	104	26	1,298
Hyperlipidemia	1,221	9,820	0	1	1	0	-1	1
Hypertension	10,784	257	0	1	1	0	-2	4
IHD	833	10,208	0	1	1	0	0	0
KidneyDis	19	11,022	0	1	0	0	0	0
MaleBen	1,243	9,798	0	12,516	133	248	16	659
NondualBen	2,118	8,923	0	19,206	259	497	10	272

Osteoporosis	2	11,039	0	1	0	0	1	2
OtherUnkBen	7,503	3,538	0	813	3	16	38	1,872
OutPayPercMedPayNonLUPA	0	11,041	0	0	0	0	1	1
RAOA	2,442	8,599	0	1	1	0	-1	1
Schizophrenia	15	11,026	0	1	0	0	3	13
Stroke	0	11,041	0	0	0	0	1	2
TotLUPAEpis	3,607	7,434	0	7,249	78	156	17	623
TotalEpisNonLUPA	0	11,041	11	53,408	550	998	17	738
WhiteBen	1,724	9,317	0	20,741	280	532	10	271
profit	0	11,041	-29,937,772	7,740,695	10,169	762,082	-10	301

Statistic for transformed variables:

Variable	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
AIANBen	7,855	3,186	0	6	0	1	6	32
APIBen	5,970	5,071	0	8	1	1	2	1
AVGMedSocVPENonLUPA	11,041	0	0	1	0	0	2	8
Asthma	11,041	0	0	1	0	0	1	2
AvgHCCScore	11,041	0	1	2	1	0	1	2
AvgHHAVPENonLUPA	11,041	0	0	4	1	1	1	0
AvgSNVPENonLUPA	11,041	0	0	4	2	0	0	5
AvgSTVPENonLUPA	11,041	0	0	2	0	0	2	6
AvgTotalVPENonLUPA	11,041	0	2	5	3	0	1	3
BlackBen	7,491	3,550	0	9	3	2	-1	0
DistBenNonLUPA	11,041	0	2	10	5	1	0	0
DualBen	8,923	2,118	0	10	4	1	0	1
FemaleBen	9,798	1,243	2	10	5	1	0	0
HispBen	6,085	4,956	0	9	3	2	0	-1
MaleBen	9,798	1,243	0	9	4	1	0	0
NondualBen	8,923	2,118	0	10	5	1	0	0
OtherUnkBen	3,538	7,503	0	7	0	1	3	6
Schizophrenia	11,026	15	0	1	0	0	2	9
TotLUPAEpis	7,434	3,607	0	9	4	1	-1	1
TotalEpisNonLUPA	11,041	0	2	11	6	1	0	0
WhiteBen	9,317	1,724	0	10	5	1	0	0
AtrialFib	11,041	0	0	1	0	0	0	0
CHF	10,997	44	0	1	1	0	0	1
Cancer	11,041	0	0	1	0	0	1	3
Depression	10,924	117	0	1	1	0	0	0
Diabetes	10,802	239	0	1	1	0	0	0
Hyperlipidemia	9,820	1,221	0	1	1	0	-1	1
Hypertension	257	10,784	0	1	1	0	-2	4
IHD	10,208	833	0	1	1	0	0	0
KidneyDis	11,022	19	0	1	1	0	0	0
OutPayPercMedPayNonLUPA	11,041	0	0	1	0	0	1	1
RAOA	8,599	2,442	0	1	1	0	-1	1
Region	56	10,985	0	1	0	0	0	-1
Stroke	11,041	0	0	1	0	0	1	2
Alzheimers	10,971	70	1	1	1	0	0	1
AvgOTVPENonLUPA	11,041	0	1	4	1	0	1	1

AvgPTVPENonLUPA	11,041	0	1	5	2	1	0	0
COPD	11,035	6	1	1	1	0	0	1
Osteoporosis	11,039	2	1	1	1	0	1	2
AvgAge	11,041	0	2,209	8,464	5,913	717	-1	1

Results for whether a model is profitable or not.

Profit (Yes/No) – Only Imputation

Statistics Label	Train	Validation
Akaike's Information Criterion	8363.036	.
Average Squared Error	0.198295	0.198198
Average Error Function	0.578587	0.585117
Degrees of Freedom for Error	7148	.
Model Degrees of Freedom	29	.
Total Degrees of Freedom	7177	.
Divisor for ASE	14354	7728
Error Function	8305.036	4521.787
Final Prediction Error	0.199904	.
Maximum Absolute Error	0.989404	0.999999
Mean Square Error	0.199099	0.198198
Sum of Frequencies	7177	3864
Number of Estimate Weights	29	.
Root Average Sum of Squares	0.445303	0.445194
Root Final Prediction Error	0.447106	.
Root Mean Squared Error	0.446205	0.445194
Schwarz's Bayesian Criterion	8562.517	.
Sum of Squared Errors	2846.322	1531.673
Sum of Case Weights Times Freq	14354	7728
Misclassification Rate	0.310018	0.307195

Effect		Point Estimate
AVGMedSocVPENonLUPA		0.275
Asthma		10.127
AvgAge		1.060
AvgHCCScore		1.301
AvgHHAVPENonLUPA		0.562
AvgOTVPENonLUPA		0.637
AvgPTVPENonLUPA		0.644
AvgSNVPENonLUPA		0.524
AvgSTVPENonLUPA		0.218
AvgTotalVPENonLUPA		1.497
IMP_APIBen		1.004
IMP_CHF		0.463
IMP_COPD		0.377
IMP_Depression		4.565
IMP_Diabetes		1.552
IMP_DualBen		1.001
IMP_Hyperlipidemia		3.348
IMP_KidneyDis		0.318
IMP_Osteoporosis		2.112
IMP_RAOA		0.402
IMP_Schizophrenia		3.665
M_COPD	0 vs 1	999.000
M_HispBen	0 vs 1	1.163
M_RAOA	0 vs 1	0.688
M_TotLUPAEpis	0 vs 1	0.834
M_WhiteBen	0 vs 1	0.806
OutPayPercMedPayNonLUPA		999.000

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	65.5309	52.3749	827	21.4027
1	0	34.4691	19.0372	435	11.2578
0	1	28.9008	47.6251	752	19.4617
1	1	71.0992	80.9628	1850	47.8778

Profit (Yes/No) – Transformation then Imputation

Statistics Label	Train	Validation
Akaike's Information Criterion	8106.104	.
Average Squared Error	0.191666	0.194169
Average Error Function	0.561741	0.567608
Degrees of Freedom for Error	7154	.
Model Degrees of Freedom	22	.
Total Degrees of Freedom	7176	.
Divisor for ASE	14352	7730
Error Function	8062.104	4387.612
Final Prediction Error	0.192845	.
Maximum Absolute Error	0.976021	0.96503
Mean Square Error	0.192256	0.194169
Sum of Frequencies	7176	3865
Number of Estimate Weights	22	.
Root Average Sum of Squares	0.437797	0.440646
Root Final Prediction Error	0.439141	.
Root Mean Squared Error	0.43847	0.440646
Schwarz's Bayesian Criterion	8257.431	.
Sum of Squared Errors	2750.793	1500.927
Sum of Case Weights Times Freq	14352	7730
Misclassification Rate	0.305323	0.305821

Effect		Estimate
IMP_LOG_APIBen		1.229
IMP_LOG_BlackBen		1.166
IMP_LOG_DualBen		1.231
IMP_LOG_HispBen		1.151
IMP_LOG_NondualBen		0.812
IMP_LOG_TotLUPAEpis		0.771
IMP_RANGE_Depression		4.843
IMP_SQRT_Alzheimers		0.167
LOG_AVGMedSocVPENonLUPA		0.281
LOG_AvgSNVPENonLUPA		0.378
LOG_AvgSTVPENonLUPA		0.291
LOG_AvgTotalVPENonLUPA		0.023
LOG_DistBenNonLUPA		1.647
LOG_TotalEpisNonLUPA		0.558
M_LOG_HispBen	0 vs 1	1.184
M_LOG_TotLUPAEpis	0 vs 1	0.623
M_RANGE_RAOA	0 vs 1	0.704
RANGE_OutPayPercMedPayNonLUPA		18.968
SQRT_AvgOTVPENonLUPA		1.560
SQRT_AvgPTVPENonLUPA		1.631
SQR_AvgAge		1.000

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	65.6273	55.1466	884	22.8719
1	0	34.3727	20.4686	463	11.9793
0	1	28.5544	44.8534	719	18.6028
1	1	71.4456	79.5314	1799	46.5459

Profit (Yes/No) – Transformation then Imputation – Regression with interaction term

Statistics Label	Train	Validation
Akaike's Information Criterion	7803.196	.
Average Squared Error	0.182205	0.190617
Average Error Function	0.537291	0.558945
Degrees of Freedom for Error	7130	.
Model Degrees of Freedom	46	.
Total Degrees of Freedom	7176	.
Divisor for ASE	14352	7730
Error Function	7711.196	4320.645
Final Prediction Error	0.184556	.
Maximum Absolute Error	0.98472	0.997165
Mean Square Error	0.18338	0.190617
Sum of Frequencies	7176	3865
Number of Estimate Weights	46	.
Root Average Sum of Squares	0.426855	0.436597
Root Final Prediction Error	0.4296	.
Root Mean Squared Error	0.428229	0.436597
Schwarz's Bayesian Criterion	8119.607	.
Sum of Squared Errors	2615.005	1473.467
Sum of Case Weights Times Freq	14352	7730
Misclassification Rate	0.280936	0.294437

```
Parameter                                              DF    Estimate

Intercept                                               1      1.5709
M_SQRT_COPD                                     0       1      3.9903
M_LOG_AIANBen*M_LOG_HispBen                     0 0     1     -0.0631
M_LOG_AIANBen*M_RANGE_Depression                0 0     1      0.1106
M_LOG_HispBen*M_LOG_WhiteBen                    0 0     1      0.1389
M_LOG_Schizophrenia*M_RANGE_RAOA                0 0     1     -1.1037
M_LOG_TotLUPAEpis*M_RANGE_CHF                   0 0     1     -0.5447
M_LOG_TotLUPAEpis*M_SQRT_Alzheimers             0 0     1      0.2892
M_RANGE_RAOA*M_SQRT_Osteoporosis                0 0     1      0.9507
IMP_LOG_APIBen*IMP_RANGE_Depression                     1      0.3625
IMP_LOG_BlackBen*IMP_LOG_BlackBen                       1     -0.0344
IMP_LOG_BlackBen*IMP_LOG_DualBen                        1      0.0849
IMP_LOG_BlackBen*LOG_AvgHHAVPENonLUPA                   1     -0.0708
IMP_LOG_HispBen*IMP_LOG_Schizophrenia                   1      1.3730
IMP_LOG_HispBen*LOG_Asthma                              1     -0.2601
IMP_LOG_NondualBen*IMP_RANGE_CHF                        1     -0.8571
IMP_LOG_NondualBen*IMP_RANGE_IHD                        1      0.0246
IMP_LOG_NondualBen*SQR_AvgAge                           1    0.000051
IMP_LOG_Schizophrenia*SQRT_AvgOTVPENonLUPA              1     -2.2086
IMP_LOG_TotLUPAEpis*LOG_TotalEpisNonLUPA                1     -0.0571
IMP_LOG_WhiteBen*LOG_AvgHHAVPENonLUPA                   1      0.0638
IMP_RANGE_CHF*LOG_AvgSNVPENonLUPA                       1      1.4916
IMP_RANGE_Depression*LOG_AvgSTVPENonLUPA                1      5.2153
IMP_RANGE_Depression*LOG_DistBenNonLUPA                 1      1.2868
IMP_RANGE_Depression*LOG_TotalEpisNonLUPA               1     -0.8724
IMP_RANGE_Depression*SQRT_AvgPTVPENonLUPA               1     -0.3091
IMP_RANGE_Hyperlipidemia*LOG_AvgHHAVPENonLUPA           1      0.6100
```

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	66.8113	57.6419	924	23.9069
1	0	33.1887	20.2918	459	11.8758
0	1	27.3570	42.3581	679	17.5679
1	1	72.6430	79.7082	1803	46.6494