CAPSTONE BUSINESS REPORT SUBMISSION-2

LIFE INSURANCE SALES

PGP – DSBA (FEB BATCH: 2022-2023)

BY SASHWAT RAJ BAKSHI

INDEX

FIGURE INDEX

1. Model Building and Interpretation

a) Building Various models

Approach:

1. It is important for us to note that since we are looking to build a predictionmodel for a continuous variable, before we feed the data into any model we will require to convert all categorical variables into numerical variables. This has been done in Capstone Project Notes 1.
2. We also need to make sure that the data is scaled/normalized in a way thatthere is no bias. We have already scaled the data using z score, where the mean is now 0, and standard deviation is 1, in the previous milestone. This has also been done in Capstone Project Notes 1.
We can see the statistical description of the data where categorical variablesare converted into numerical, and data is scaled in Table 2
3. We are using the scaled data because our Target Variable - AgentBonus is a continuous variable, and we will be using models like Linear Regression andArtificial Neural Network and since they are distance-based algorithms, we will need data to be scaled uniformly so that our model is not biased.
4. We will also use Random Forest and Decision Trees and compare theirresults with the models
5. The scaled and numerical data will be split in a 70:30 fashion, 70 fortraining the model and 30 for testing the model.
6. We will then go ahead and build the models Linear Regression, DecisionTree Regressor, Random Forest Regressor and ANN Regressor.

	count	mean	std	min	25%	50%	75%	max
AgentBonus	4520.0	0.0	1.0001	-1.7623	-0.7484	-0.1185	0.5626	3.9412
Age	4520.0	0.0	1.0001	-1.4145	-0.7304	-0.1603	0.5238	4.9705
CustTenure	4520.0	-0.0	1.0001	-1.4180	-0.7316	-0.1596	0.5267	4.8738
Channel	4520.0	-0.0	1.0001	-0.6091	-0.6091	-0.6091	0.6515	1.9120
Occupation	4520.0	0.0	1.0001	-4.5280	-0.4103	-0.4103	0.9623	0.9623
EducationField	4520.0	-0.0	1.0001	-1.5892	-0.4379	-0.4379	1.2892	1.2892
Gender	4520.0	0.0	1.0001	-1.2113	-1.2113	0.8256	0.8256	0.8256
ExistingProdType	4520.0	0.0	1.0001	-2.6475	-0.6783	0.3063	0.3063	2.2754
NumberOfPolicy	4520.0	0.0	1.0001	-1.7710	-1.0807	0.3001	0.9904	1.6808
MaritalStatus	4520.0	0.0	1.0001	-1.5398	-0.2408	-0.2408	1.0582	2.3572
MonthlyIncome	4520.0	0.0	1.0001	-1.4303	-0.6224	-0.2555	0.3586	3.2812
Complaint	4520.0	0.0	1.0001	-0.6347	-0.6347	-0.6347	1.5755	1.5755
ExistingPolicyTenure	4520.0	0.0	1.0001	-0.9551	-0.6500	-0.3448	0.2654	6.3682
SumAssured	4520.0	-0.0	1.0001	-1.8591	-0.7193	-0.1637	0.5428	5.0390
Zone	4520.0	-0.0	1.0001	-2.0954	-1.1082	0.8662	0.8662	0.8662
PaymentMethod	4520.0	-0.0	1.0001	-0.7780	-0.7780	-0.7780	1.4162	1.4162
LastMonthCalls	4520.0	-0.0	1.0001	-1.2783	-0.7257	-0.4495	0.9318	3.6945
CustCareScore	4520.0	-0.0	1.0001	-1.5039	-0.7765	-0.0492	0.6782	1.4056
DesignationOrdinal	4520.0	-0.0	1.0001	-0.9633	-0.9633	-0.0718	0.8197	2.6028

Fig 1: Data Description (Scaled and Transformed)

1. Linear Regression:

Linear regression analysis is used to predict the value of a variable based onthe value of another variable. The variable you want to predict is called the dependent variable - in this case the

AgentBonus.

Furthermore, all the predictor variables should be normally distributed withconstant variance and should demonstrate little to no multicollinearity nor autocorrelation with one another, hence indicating that data must be scaled. In regression, it is often recommended to scale the features so that the predictors have a mean of 0. This makes it easier to interpret the intercept term as the expected value of Y when the predictor values are setto their means.

When one variable has a very large scale: e.g., if you are using the Sum Assured and Monthly Income as a predictor. In that case, the regression coefficients will be higher for Sum Assured since it is at least 10 times more than the monthly income, and it will imply that Sum Assured is a more important predictor than Monthly Income, which may not necessarily be true.

Assumptions:

1. Normality - The data should follow a normal distribution.
2. Linearity - The line of the best fit is a straight line and not a curve.
3. No Multicollinearity (VIF - Checked) - No hidden relationships amongvariables.
4. Homogeneity of variance (Homoscedasticity). Homoscedasticity means “having the same scatter”. If the variance of the error term is homoscedastic, the model was well-defined. If there is too much variance, the model may not be defined well. For it to exist in a set of data, the points must be about the same distance from the regression line. The opposite is heteroscedasticity (“different scatter”).

R-Squared

It measures the proportion of the variation in your dependent variableexplained by all of your independent variables in the model.
It assumes that every independent variable in the model helps to explain variation in the dependent variable. In reality, some independent variables(predictors) don't help to explain dependent (target) variable.

Adjusted R-Square

Adjusted R-square should be used while selecting important predictors(independent variables) for the regression model.
Adjusted R-square should be used to compare models with different numbersof independent variables.

Characteristics of R-Squared & Adjusted R-Squared

Every time you add an independent variable to a model, the R-squared increases, even if the independent variable is insignificant.
It never declines. Whereas Adjusted R-squared increases only when independent variable is significant and affectsdependent variable.
Adjusted r-squared can be negative when r- squared is close to zero. Adjusted r-squared value always be less than or equal to r-squared value.

P-value

The statistical test for this is called Hypothesis testing. A low P-value (< 0.05)means that the coefficient is likely not to equal zero. A high P-value (> 0.05) means that we cannot conclude that the explanatory variable affects the dependent variable.

The null hypothesis states that there is no relationship between the two variables being studied. The alternative hypothesis states that the independent variable did affect the dependent

variable, and the results aresignificant in terms of supporting the theory being investigated.

Train-Test Split:

We have done test train split on both the data, the normal data as well as the scaled data by taking test size as 0.30 and random state as 1.

Linear Regression using statsmodels model 1:

Intercept	-0.001788
Age	0.141166
CustTenure	0.148068
Channel	0.003769
Occupation	-0.007737
EducationField	0.005270
Gender	0.010345
ExistingProdType	-0.001357
NumberOfPolicy	0.015643
MaritalStatus	-0.001270
MonthlyIncome	0.115424
Complaint	0.015627
ExistingPolicyTenure	0.086879
SumAssured	0.591159
Zone	0.003040
PaymentMethod	-0.008410
LastMonthCalls	-0.008335
CustCareScore	0.010502
DesignationOrdinal	0.083344
dtype: float64	

OLS Regression Results

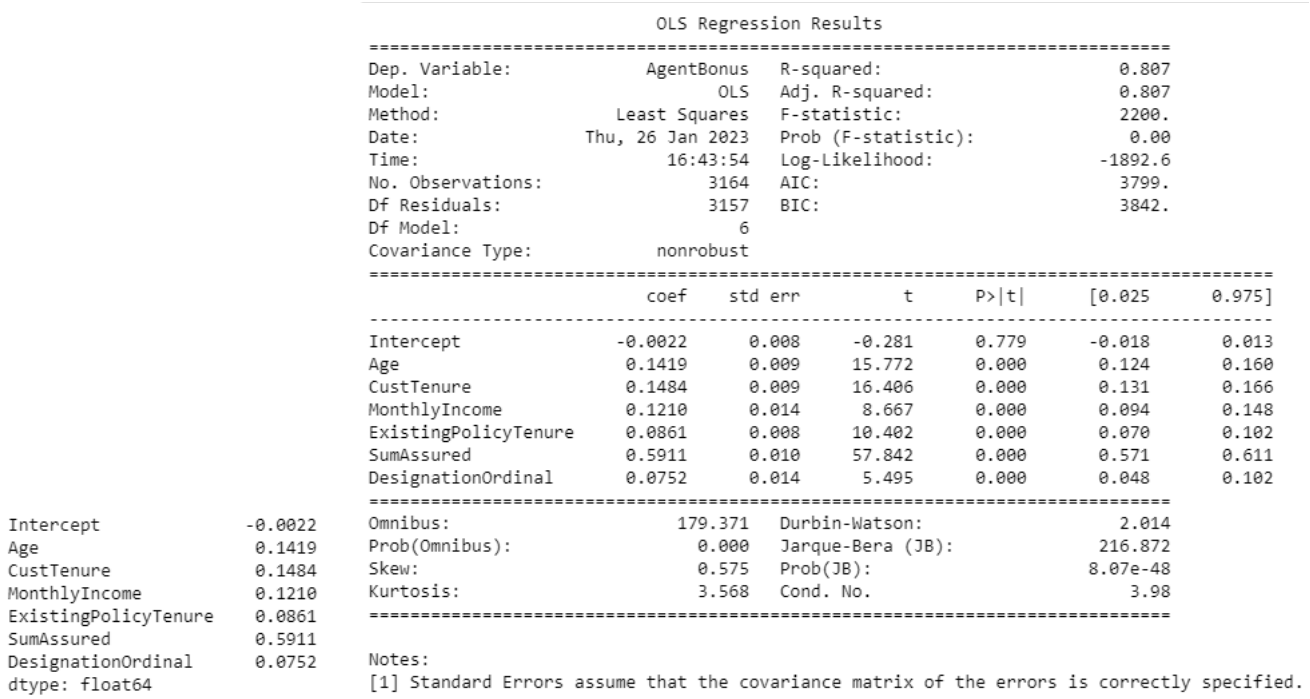
Dep. Variable:	AgentBonus	R-squared:	0.808
Model:	OLS	Adj. R-squared:	0.807
Method:	Least Squares	F-statistic:	734.7
Date:	Thu, 26 Jan 2023	Prob (F-statistic):	0.00
Time:	16:43:53	Log-Likelihood:	-1885.5
No. Observations:	3164	AIC:	3809.
Df Residuals:	3145	BIC:	3924.
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.0018	0.008	-0.228	0.820	-0.017	0.014
Age	0.1412	0.009	15.651	0.000	0.123	0.159
CustTenure	0.1481	0.009	16.353	0.000	0.130	0.166
Channel	0.0038	0.008	0.479	0.632	-0.012	0.019
Occupation	-0.0077	0.009	-0.862	0.389	-0.025	0.010
EducationField	0.0053	0.009	0.596	0.551	-0.012	0.023
Gender	0.0103	0.008	1.317	0.188	-0.005	0.026
ExistingProdType	-0.0014	0.010	-0.141	0.888	-0.020	0.018
NumberOfPolicy	0.0156	0.008	1.939	0.053	-0.000	0.031
MaritalStatus	-0.0013	0.008	-0.160	0.873	-0.017	0.014
MonthlyIncome	0.1154	0.015	7.654	0.000	0.086	0.145
Complaint	0.0156	0.008	1.982	0.048	0.000	0.031
ExistingPolicyTenure	0.0869	0.008	10.442	0.000	0.071	0.103
SumAssured	0.5912	0.010	57.706	0.000	0.571	0.611
Zone	0.0030	0.008	0.386	0.699	-0.012	0.018
PaymentMethod	-0.0084	0.009	-0.919	0.358	-0.026	0.010
LastMonthCalls	-0.0083	0.008	-0.994	0.320	-0.025	0.008
CustCareScore	0.0105	0.008	1.334	0.182	-0.005	0.026
DesignationOrdinal	0.0833	0.015	5.665	0.000	0.054	0.112

Omnibus:	174.261	Durbin-Watson:	2.015
Prob(Omnibus):	0.000	Jarque-Bera (JB):	209.574
Skew:	0.566	Prob(JB):	3.10e-46
Kurtosis:	3.555	Cond. No.	4.55

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Fig 2: 1st Iteration of Multiple Linear Regression Using OLS



```
                            OLS Regression Results
==============================================================================
Dep. Variable:             AgentBonus   R-squared:                       0.807
Model:                            OLS   Adj. R-squared:                  0.807
Method:                 Least Squares   F-statistic:                     2200.
Date:                Thu, 26 Jan 2023   Prob (F-statistic):               0.00
Time:                        16:43:54   Log-Likelihood:                -1892.6
No. Observations:                3164   AIC:                             3799.
Df Residuals:                    3157   BIC:                             3842.
Df Model:                           6
Covariance Type:            nonrobust
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
Intercept                 -0.0022      0.008     -0.281      0.779      -0.018       0.013
Age                        0.1419      0.009     15.772      0.000       0.124       0.160
CustTenure                 0.1484      0.009     16.406      0.000       0.131       0.166
MonthlyIncome              0.1210      0.014      8.667      0.000       0.094       0.148
ExistingPolicyTenure       0.0861      0.008     10.402      0.000       0.070       0.102
SumAssured                 0.5911      0.010     57.842      0.000       0.571       0.611
DesignationOrdinal         0.0752      0.014      5.495      0.000       0.048       0.102
==============================================================================
Omnibus:                      179.371   Durbin-Watson:                   2.014
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              216.872
Skew:                           0.575   Prob(JB):                     8.07e-48
Kurtosis:                       3.568   Cond. No.                         3.98
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
Intercept              -0.0022
Age                     0.1419
CustTenure              0.1484
MonthlyIncome           0.1210
ExistingPolicyTenure    0.0861
SumAssured              0.5911
DesignationOrdinal      0.0752
dtype: float64
```

Fig 3: 13 Iteration of Multiple Linear Regression Using OLS

Channel, Occupation, Education Field, Gender, ExistingProdType, NumberOfPolicy, MaritalStatus, Zone, PaymentMethod and LastMonthCalls have a high p value and hence implying that they do not have a relation with the target variable.

Significant Predictors:

ExistingProdType, MaritalStatus, Zone, Channel, Education Field, Occupation,PaymentMethod, Gender, LastMonthCalls and NumberOfPolicy have a high p-value as per first OLS regression model.

We will go ahead and drop the variables one by one. We will start by ExistingProdType which has a p-value of 0.888, and will then follow with MaritalStatus, Zone, Channel, Education Field, Occupation, PaymentMethod,Gender, LastMonthCalls and NumberOfPolicy.

If p > 0.05 we the coefficient is not significant, thereby signifying the independent variable ExistingProdType is not an important predictor as perthe Regression Model

Final Equation:

(-0.0022) * Intercept + (0.1419) * Age + (0.1484) * CustTenure + (0.121) * MonthlyIncome + (0.0861) * ExistingPolicyTenure + (0.5911) * SumAssured +(0.0752) * DesignationOrdinal

The final equation is the one we have seen performance metrics on as below:

RMSE Train: 0.4400
RMSE Test: 0.4449
MAPE Train: 158.73
MAPE Test: 217.23

Test for Multi-collinearity using VIF:

```
Age ---> 1.3318242945163823
CustTenure ---> 1.330046921603369
Channel ---> 1.0066249825033349
Occupation ---> 1.3034389370049737
EducationField ---> 1.3018772818552433
Gender ---> 1.0139686675306303
ExistingProdType ---> 1.5245358756963587
NumberOfPolicy ---> 1.0659136458632514
MaritalStatus ---> 1.0291129494313538
MonthlyIncome ---> 3.715839760513621
Complaint ---> 1.0036456832086942
ExistingPolicyTenure ---> 1.1306992026239535
SumAssured ---> 1.742812587009907
Zone ---> 1.0086815566075218
PaymentMethod ---> 1.3658213037929479
LastMonthCalls ---> 1.1702907464189178
CustCareScore ---> 1.0100969102619328
DesignationOrdinal ---> 3.583738855018704
```

Fig 4: Multicollinearity using VIF

Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. Mathematically, the VIF for a regression model variable is equal to the ratio of the overall model variance to the variance of a model that includes only that single independent variable.

As we can see, the variance inflation factor is not greater than 5 for any of the variables. We will go ahead and keep all variables.

b. Test your predictive model against the test set using various appropriate performance metrics

Decision Tree:

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithmthat only contains conditional control statements.

We will be initially designing a basic model for Decision Trees

Assumptions:

1. Linearity of Data
2. The input data is continuous
3. The input data contains multiple variables, and each variable has only onelevel.
4. There are no missing values in the input data. The data is distributednormally.

Advantages of Decision Trees:

1. Compared to other algorithms decision trees requires less effort for datapreparation during pre-processing.
2. A decision tree does not require normalisation of data.

3. A decision tree does not require scaling of data as well.
4. Missing values in the data also do NOT affect the process of building adecision tree to any considerable extent.
5. A Decision tree model is very intuitive and easy to explain to technicalteams as well as stakeholders.

Disadvantage of Decision Trees:

1. A small change in the data can cause a large change in the structure ofthe decision tree causing instability.
2. For a Decision tree sometimes, calculation can go far more complexcompared to other algorithms.
3. Decision tree often involves higher time to train the model.
4. Decision tree training is relatively expensive as the complexity and timehas taken are more.
5. The Decision Tree algorithm is inadequate for applying regression andpredicting continuous values.

Random Forest Model:

Assumptions:

It follows the same assumptions of that of Decision Tree models.
Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees ondifferent samples and takes their majority vote for classification and averagein case of regression.

Advantages of Random Forest:

1. It reduces overfitting in decision trees
2. It is flexible to both classification and regression problems
3. It works well with both categorical and continuous values
4. It automates missing values present in the data
5. Normalizing of data is not required as it uses a rule-based approach.

Disadvantages of Random Forest:

1. It requires much computational power as well as resources as it buildsnumerous trees to combine their outputs.
2. It also requires much time for training as it combines a lot of decisiontrees to determine the class.
3. Due to the ensemble of decision trees, it also suffers interpretability andfails to determine the significance of each variable.

ANN Regressor:

Regression ANNs predict an output variable as a function of the inputs. The input features (independent variables) can be categorical or numeric types, however, for regression ANNs, we require a numeric dependent variable.
All variables should be numeric.

The purpose of using Artificial Neural Networks for Regression over Linear Regression is that the linear regression can only learn the linear relationshipbetween the features and target and therefore cannot learn the complex non-linear relationship.

Points to remember:

1. Encoding needs to be done for all ordinal categorical variables

2. Identify Significant Predictors in the model using p value,
3. R-Squared, Adjusted R-Squared, etc
4. Form the Linear Regression Equation Predict Values & Compare with TestData
5. Selection of Best Model
6. Predict Agent Bonus
7. Decide on Engagement Strategy for High Performers and Upskill Programfor Low Performers.

Conclusion:

1. Encoding has been done to ensure all data is numerical
2. There is no Multicollinearity in the data
3. Missing Values and outliers have already been taken care of

We have used OLS Regression for Linear Regression model building exerciseand sklearn as well for all 4 models. Results are shown below.

Linear Regression models can only handle numeric variables. Therefore, if a column has categorical variable, models will only work when it is encoded tonumeric variable.

There are two types of categorical data -
Ordinal: Order based Payment Method
Nominal: Without any order or ranks like city Channel, Marital Status, Zone

Designation - Order can be on the basis from customers with higher designation to lower designation conventionally (i.e., VP > AVP > Senior Manger > Manager > Executive) All other features (Channel, Occupation,Categorical Variable Transformations:
Education Field, Gender, Marital Status, Zone, and Payment Method) will beconsidered as Nominal for further steps. All data types have been transformed to Numeric and are ready for modelling

Scaling: The range of the data features is not across the same scale, and since we are predicting using a Linear Regression Model (Which is distancebased). It is important that we ensure the features are brought to the same scale.

Model Performance Measures:

1. Root Mean Square Error (RMSE)

The root mean squared error (MSE) tells you how close a regression line is toa set of points. It does this by taking the difference between Observed (Actual) & Predicted Values and squaring them followed by a square root of those values.

The squaring is necessary to remove any negative signs. And after that is taken care of to understand the magnitude of closeness the root is taken. Italso gives more weight to larger differences.

It's called the root mean squared error as you're finding the average of a setof errors. The **lower** the **RMSE**, the **better** the **forecast**.

2. Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) is the mean or average of theabsolute percentage errors of forecasts.

Error is defined as actual or observed value minus the forecasted value. Percentage errors are summed without regard to sign to compute MAPE. Thismeasure is easy to understand because it

provides the error in terms of percentages.

Also, because absolute percentage errors are used, the problem of positiveand negative errors canceling each other out is avoided. Consequently, MAPE has managerial appeal and is a measure commonly used in forecasting. The **smaller** the **MAPE** the **better** the **forecast**.

	Train RMSE	Test RMSE	Training Score	Test Score	Training MAPE	Test MAPE
Linear Regression	0.439102	0.446577	0.807881	0.798861	158.280045	214.571432
Decision Tree Regressor	0.000000	0.537807	1.000000	0.708286	0.000000	299.504204
Random Forest Regressor	0.140886	0.380135	0.980222	0.854260	51.128570	194.581091
ANN Regressor	0.062869	0.510206	0.996062	0.737461	25.204602	303.563046

Fig 5: Performance metrics - Basic Modelling (No Model Tuning done)

C) Interpretation of the models

From the above results table we can draw the following conclusions:

- The Decision Tree, Random Forest, and ANN models are over fit as observed in the Training and Test Score
- MAPE is also not consistent amongst train and test sets hence difficult to draw conclusion
- Linear Regression model seems to be a good fit for now
- We will have to go ahead and fine tune these models to generate a better fit and better forecasts

2) Model Tuning & Business Implication

a) Ensemble modelling, wherever applicable

Ensemble Method (XG Boost for Regression)

XGBoost is a powerful approach for building supervised regression models. The validity of this statement can be inferred by knowing about its (XGBoost)objective function and base learners. The objective function contains loss function and a regularisation term. It tells about the difference between actual values and predicted values, i.e how far the model results are from the real values. The most common loss functions in XGBoost for regression problems is reg:linear, and that for binary classification is reg:logistics.
Ensemble learning involves training and combining individual models (knownas base learners) to get a single prediction, and XGBoost is one of the ensemble learning methods. XGBoost expects to have the base learners which are uniformly bad at the remainder so that when all the predictions are combined, bad predictions cancels out and better one sums up to form final good predictions.

Parameters:

(objective ='reg:linear', n_estimators = 10, seed = 123)

Objective determines the loss function to be used like reg: linear forregression problems
n_estimators is number of trees you want to build and **seed** is used
for reproducibility of results.

Performance metrics:

RMSE Train: 0.31
RMSE Test: 0.39
MAPE Train: 116.11
MAPE Test: 207.46

b) Any other model tuning measures (if applicable)

Using Grid Search on Decision Tree:

Parameters:

Result: {'max_depth': 15, 'min_samples_leaf': 3, 'min_samples_split': 55}

max_depth defines the maximum depth of each decision tree created in theforest. Values taken as [10,11,13] in grid search as it ensures model will not underfit or overfit.
min_samples_split parameter will define the number of observations neededin a node to be split further, if the number is less than the minimum, the split will not happen, and the node will become a leaf.
min_samples_leaf parameter checks before the node is generated, that is, ifthe possible split results in a child with fewer samples, the split will be avoided

Performance metrics:

RMSE Train: 0.35
RMSE Test: 0.42
MAPE Train: 129.99
MAPE Test: 245.83

Using Grid Search for Random Forest:

Parameters:

Steps:

1. Used “RandomForestClassifier” function of “sklearn.ensemble” package todefine the estimator with random_state as 123
2. Here, Grid search approach is taken to find the best parameters. Grid isdefined as below:
   {'max_depth': [13,15,17,19],
   'max_features': [8,10,12],
   'min_samples_leaf': [3,4,5],
   'min_samples_split': [8, 10, 12],
   'n_estimators': [100, 200, 300]}

max_depth defines the maximum depth of each decision tree created in theforest. Values taken as [10,11,13] in grid search as it ensures model will not under-fit or overfit.
min_samples_split parameter will define the number of observations neededin a node to be split further, if the number is less than the minimum, the split will not happen, and the node will become a leaf.
min_samples_leaf parameter checks before the node is generated, that is, ifthe possible split results in a child with fewer samples, the split will be avoided

In the Grid taking both the values of min_samples_split, min_samples_leafparameters as 1 or 2 to avoid overfit and underfit
n_estimators indicates the number of trees to build before taking themaximum voting of predictions.

Results:

Iteration 1 - {'max_depth': 13, 'max_features': 8, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 300}
Iteration 2 - {'max_depth': 15, 'max_features': 10, 'min_samples_leaf': 3, 'min_samples_split': 8, 'n_estimators': 200}
Iteration 3 - {'max_depth': 17, 'max_features': 10, 'min_samples_leaf': 3, 'min_samples_split': 10, 'n_estimators': 200}

We will be selecting the 3rd Iteration with:
{'max_depth': 17,
'max_features': 10,
'min_samples_leaf': 3,
'min_samples_split': 10,
'n_estimators': 200}

Performance metrics:

RMSE Train: 0.24
RMSE Test: 0.38
MAPE Train: 87
MAPE Test: 193.40

Using Grid Search for ANN:

Parameters:

Results {'activation': 'relu', 'hidden_layer_sizes': 2000, 'solver': ‘sgd’}

hidden_layer_sizes : The it’s element represents the number of neurons inthe it’s hidden layer.
activation is Activation function for the hidden layer.
‘relu’, the rectified linear unit function, returns f(x) = max(0, x)
Solver is for weight optimization, ’sgd’ refers to stochastic gradient descent.

Performance metrics:

RMSE Train: 0.40
RMSE Test: 0.43

MAPE Train: 151.56
MAPE Test: 199.14

c) Interpretation of the most optimum model and its implication on the business

	Train RMSE	Test RMSE	Training Score	Test Score	Training MAPE	Test MAPE
Linear Regression	0.44	0.45	0.81	0.80	158.28	214.57
Decision Tree Regressor	0.35	0.42	0.88	0.82	129.99	245.83
Random Forest Regressor	0.24	0.38	0.94	0.85	86.51	193.40
ANN Regressor	0.40	0.43	0.84	0.82	151.56	199.14
XGBoost	0.31	0.39	0.90	0.85	116.11	207.46

Fig 6: Performance Metrics Final Table (Models Tuned) with XG BoostEnsemble Model

1. As we can see all the tuned models now seem to be better fit as comparedto the models without any tuning
2. Random forest however still seems a little over fit
3. XGBoost model gives a high score with low rise however MAPE Values arenot consistent given that the model is slightly overfit
4. The Linear Regression and ANN model seem to have the best fit whencompared
5. Although the Training and Testing scores are in the 80's we will go aheadwith either of these two models as consistency of prediction is importantto us
6. Between the Linear Regression and ANN Regressor models, the ANN model has a lower RMSE and a lower MAPE and hence we can go aheadwith either of these two models
7. We will look at the True vs Predicted values for both to get acloser insight on prediction:

X - axis: True values
Y- axis: Predicted Values

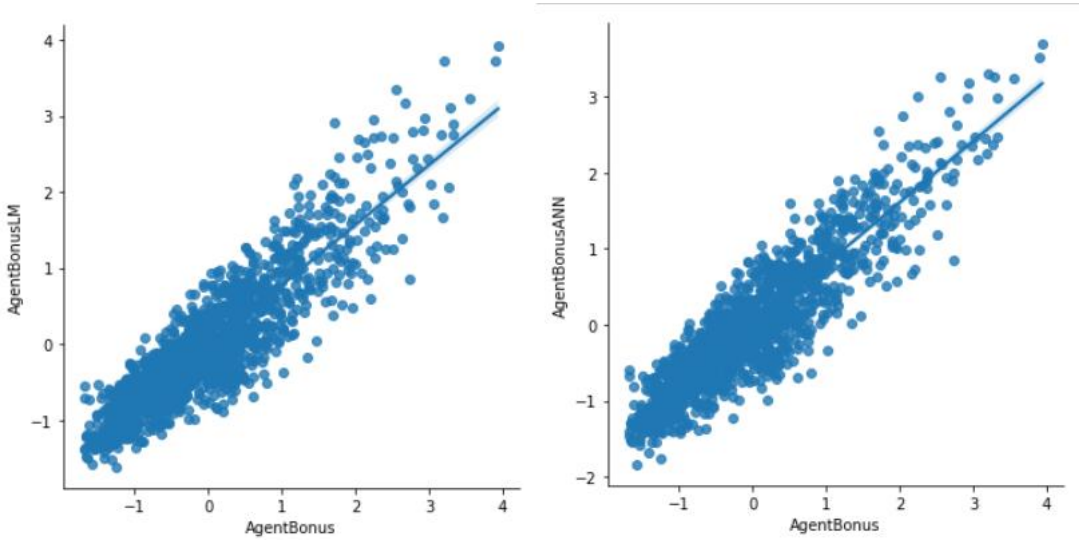


Fig 7: Linear RegressionModel & Artificial Neural Network (True vs Predicted AgentBonus) on Test Data

Final Model Selected:

We go ahead with the **Linear Regression** model and the equation below

Final Equation:

(-0.0022) * Intercept + (0.1419) * Age + (0.1484) * CustTenure + (0.121) * MonthlyIncome + (0.0861) * ExistingPolicyTenure + (0.5911) * SumAssured +(0.0752) * DesignationOrdinal

We can clearly see the trend of the significant features (Sum Assured, Monthly Income, Designation, Existing Policy Tenure, Age and Customer Tenure) in Figure 14, in relation to our target variable - Agent Bonus. This isdone to understand the relationship between significant predictors and theAgent Bonus.

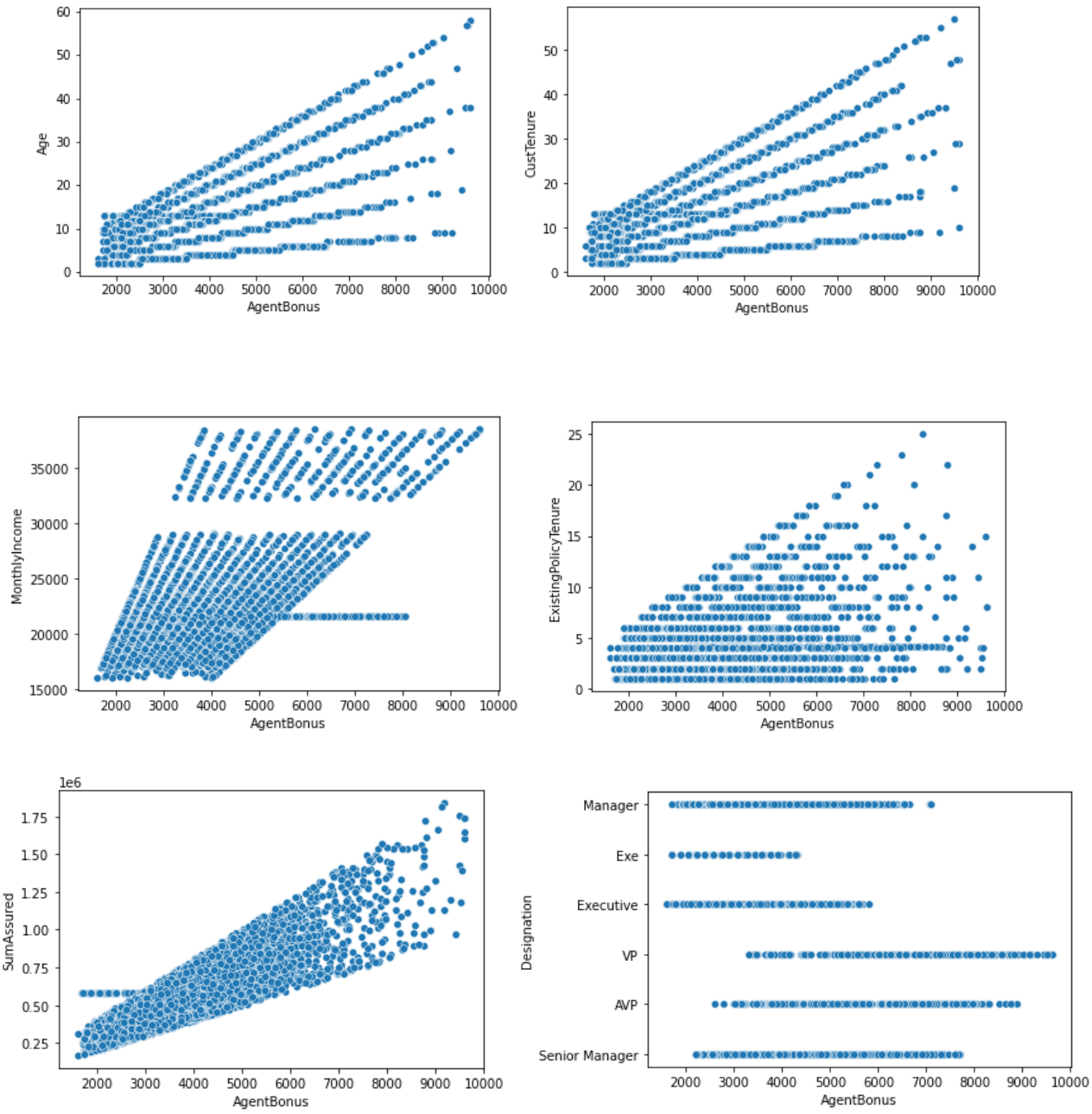


Fig 8: Trend of Significant features with Target Variable (Agent Bonus)

Business Implications:

1) We can create an agent bonus decider on the basis of the weights in theadjacent figure with highest to lowest weights in the following order:
    - SumAssured to the customer is the best predictor for Agent Bonus
    - CustomerTenure and Age of customer is the second highest predictor
    - Monthly Income is the 3rd most important predictor of AgentBonus
    - Designation and Existing Policy tenure play a vital role in deciding agent bonus however it is not as important as the other 4

2) Agent Bonus is highest where average Tenure of customer is highest indicating that reselling policy is strong for these customers. This also is confirmed by the average number of calls made by agents as it is highestacross all 3 customers.
3) Agent Bonus is highest for customer policy's where average age ofcustomer is higher.
4) Long term customers are bringing in a good policy/ reselling.
5) Sum Assured is also higher to customers who have a higher average ageindicating trend. Higher age customers can be pursued more rigorously to upsell policy's and plan upgradation insurance policies.
6) Monthly income is highest for customers where the sum assured is higherindicating right kind of followup done on cream customers which is good time management. Lower performing sales agents can learn this to maximize company realization.
7) There is a higher company presence in Northern and Western India, whereas sales numbers in the Eastern and Southern India is very poor. Market surveys can help us understand what the concern is
8) Sales made by Agents dominates the other two channels (i.e. Third PartyPartner and Online Channels). Better incentives to third party and online channels can increase company presence
9) Most of the customers are Manger, Executive and Senior Manger Level. We can focus on customer leads coming from this area and give extra attention to customers who have not onboarded with insurance policiesfrom this segment as it is cream customer segment.
10) Upskill plan 1- "Product Knowledge":
    - In this plan, the focus will be to develop knowledge about the product categories, frequently asked questions and most frequent customer profiles and their requirements. This will also have a study of building astronger sales pitch practice formally.
11) Upskill Plan 2- "Mimic the top performer"
    - In this plan, it will be a more hands on approach where a low performancesales agent will shadow/mirror the work behavior of a top performer.

--- END ---