CAPSTONE PROJECT REPORT BY RISHAV KUMAR

Tableau sheets Link:

https://public.tableau.com/app/profile/rishav.kumar6172/viz/Capstone_Rishav/Inc_Edu_Des?publish=yes

INDEX

Graphs and tables

1 Problem Statement: Life Insurance Data

The dataset belongs to a leading life insurance company. The company wants to predict the bonus for its agents so that it may design appropriate engagement activity for their high performing agents and upskill programs for low performing agents.

1a Introduction

The insurance sector is a service based sector where insurance plans are sold to customers by its agents through different modes. The insurance company needs to add maximum number of customers, cross sell their products to its existing customers and engage the customers to maximum tenure.

To increase efficiency of the company we need to increase efficiency of insurance agents by giving bonus to agents efficiently. A model needs to be created based out of given data which will predict bonus to be given to the agents and ensure that fair bonus amount is given to each agent.

This will also ensure that bonus is distributed to agents as per a model based out of figures which can be trusted upon by the agents thereby increasing satisfaction level of the agents. An agent receiving fair amount of bonus and competitive to other companies would increase the agent's engagement with the concerned insurance company in comparison to the others insurance companies. The more engagement of the agents with the concerned company would give an edge over other insurance companies and would help in increasing its share in the insurance sector.

The model should also ensure the bonus distribution as per guidelines of the regulatory body of the insurance company. Further best performing agents can be rewarded accordingly whereas poor performing agents can be engaged in upskill programs.

The model thus designed will enable us to segregate the agents in groups according to which activities can be planned for different group based out of performance. This model can also be used universally to empower existing agents with fair bonus and encourage new agents to connect with the insurance companies. An increase in the agents would help in penetrating the left out population who had still not enrolled in any insurance plan. The insurance helps the insurer by protecting his family, assets/property and self also from financial risk/losses. Thus study/model not only optimizes the bonus of the agents but also helps in inviting more people in the insurance umbrella.

1b What to Achieve

-To make a suitable model through available dataset and establish relationship of Agent Bonus with other independent variables.

-To ensure that justified bonus amount is paid to the agents as per the regulations.

-To engage the agents as per their performance through specially designed programmes.

-To make a universal model which can empower existing agents with fair bonus distribution and encourage new agents to connect with the insurance companies.

-To increase inclusion of more agents that would help in penetrating the left out population who had still not enrolled in any insurance plan.

2 Exploratory Data Analysis

The dataset have 4520 rows and 20 columns. Each row contains amount of bonus received by an agent for a given month with respect to each customer and details of particulars of those customers.

Fig 1: Brief of Dataset provided

Variables	**No of Null values**	**Minimum**	**Mean**	**Maximum**
LastMonthCalls	0	0	4.63	18
NumberOfPolicy	45	1	3.57	6
ExistingPolicyTenure	184	1	4.13	25
Age	269	2	14.49	58
CustTenure	226	2	14.47	57
AgentBonus	0	1605	4077.84	9608
MonthlyIncome	236	16009	22890.31	38456
SumAssured	154	168536	619999.7	1838496
CustID	0	NA	NA	NA
Channel	0	NA	NA	NA
Occupation	0	NA	NA	NA
EducationField	0	NA	NA	NA
Gender	0	NA	NA	NA
ExistingProdType	0	NA	NA	NA
Designation	0	NA	NA	NA
MaritalStatus	0	NA	NA	NA
Complaint	0	NA	NA	NA
Zone	0	NA	NA	NA
PaymentMethod	0	NA	NA	NA
CustCareScore	52	NA	NA	NA

The above table contains data variables, their number of null values, minimum values, mean and maximum values. The Agent bonus is our target variable which we needs to predict and is continuous in nature.

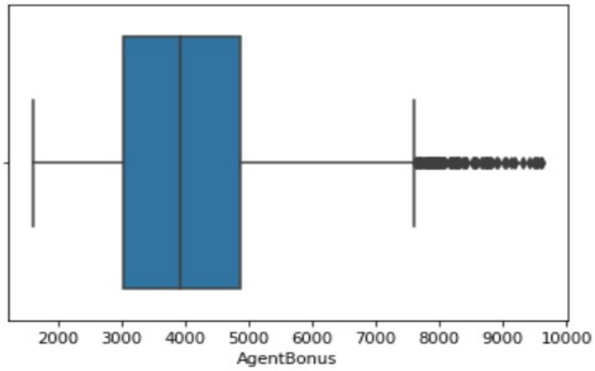
Age, Customer Tenure, Monthly income, Last month calls, Sum Assured and Existing Policy Tenure are also continuous in nature. Other remaining variables are Categorical in nature.

In a given row the data contains the customer id of a customer, bonus amount given to the agent in the last month for that particular customer, current age of the customer, the number of years the customer is engaged with the insurance company and other demographical details of the customers like occupation, age, gender, zone and marital status. The dataset also contains details of the current insurance policy such as number of policy, policy tenure, sum assured and customer tenure with the insurance company.

The data is collected separately for each customer for the last month, also having history of the customer like the number of the customer with the organization and number of insurance policies already taken by the customer. The data is summary of raw data being produced by the company for the last month.

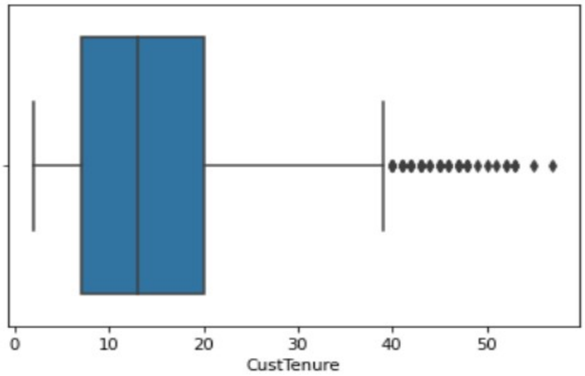
As the Customer Id is unique in nature and will not contribute in predicting our dependent variable Agent Bonus, therefore the variable Customer ID is removed from the dataset.

Fig 2: Boxplot of Agent Bonus



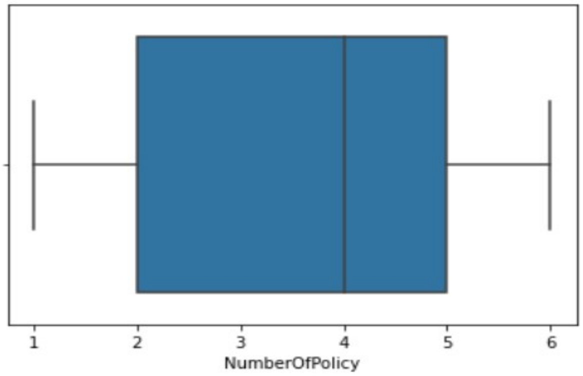
We see that average bonus amount is 4077, the maximum bonus amount is 9608 whereas maximum(75 %) of agents are receiving bonus below 5000.

Fig 3: Boxplot of Cust Tenure



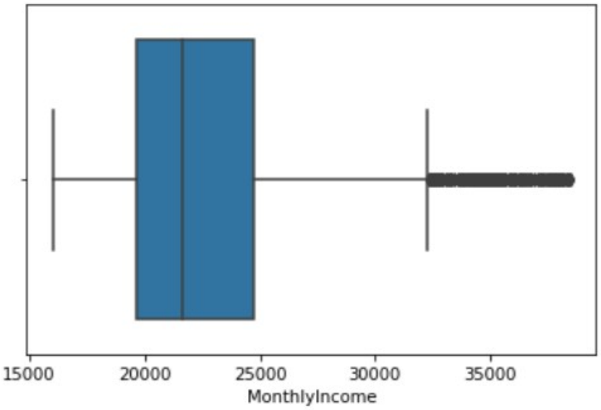
Maximum customers are under age of 20 which shows that by and large young generation is more aware about the insurance whereas customer of age of 58 is also recorded in the dataset.

Fig 4: Boxplot of Number of policy



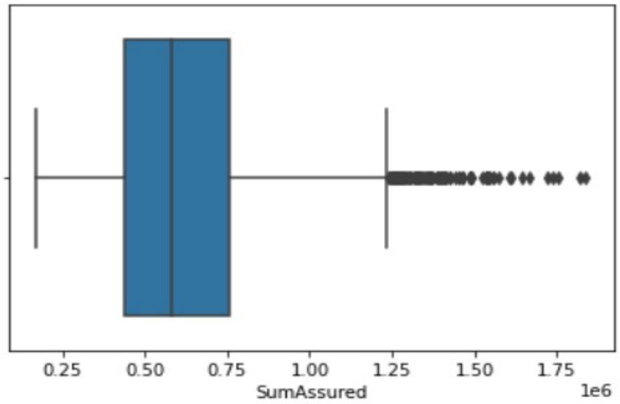
The existing customers have already subscribed an average of 3-4 number of policies by the customers which shows good engagement of the customers with the insurance company.

Fig 5: Boxplot of Monthly Income



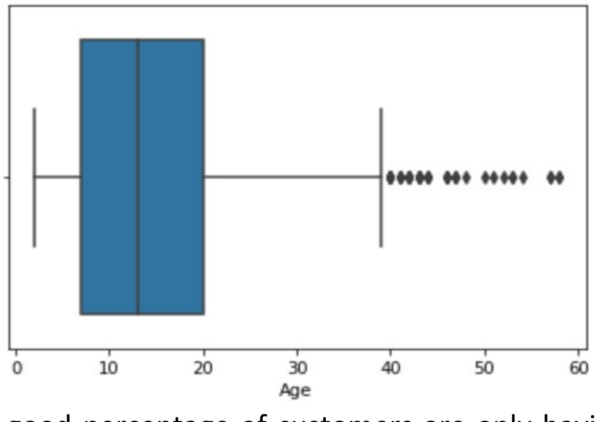
Average monthly Income of the customers is 22890. The monthly income of the customers ranges from 16009 to 38456 whereas 75 % of the customers are having income below 25000. Monthly income is concentrated around 20,000 to 25,000 whereas high income customers are very few. The high income customers are needed to be more focussed on as the premium amount of such customers would also be at a higher side adding to our revenue.

Fig 6: Boxplot of Sum Assured



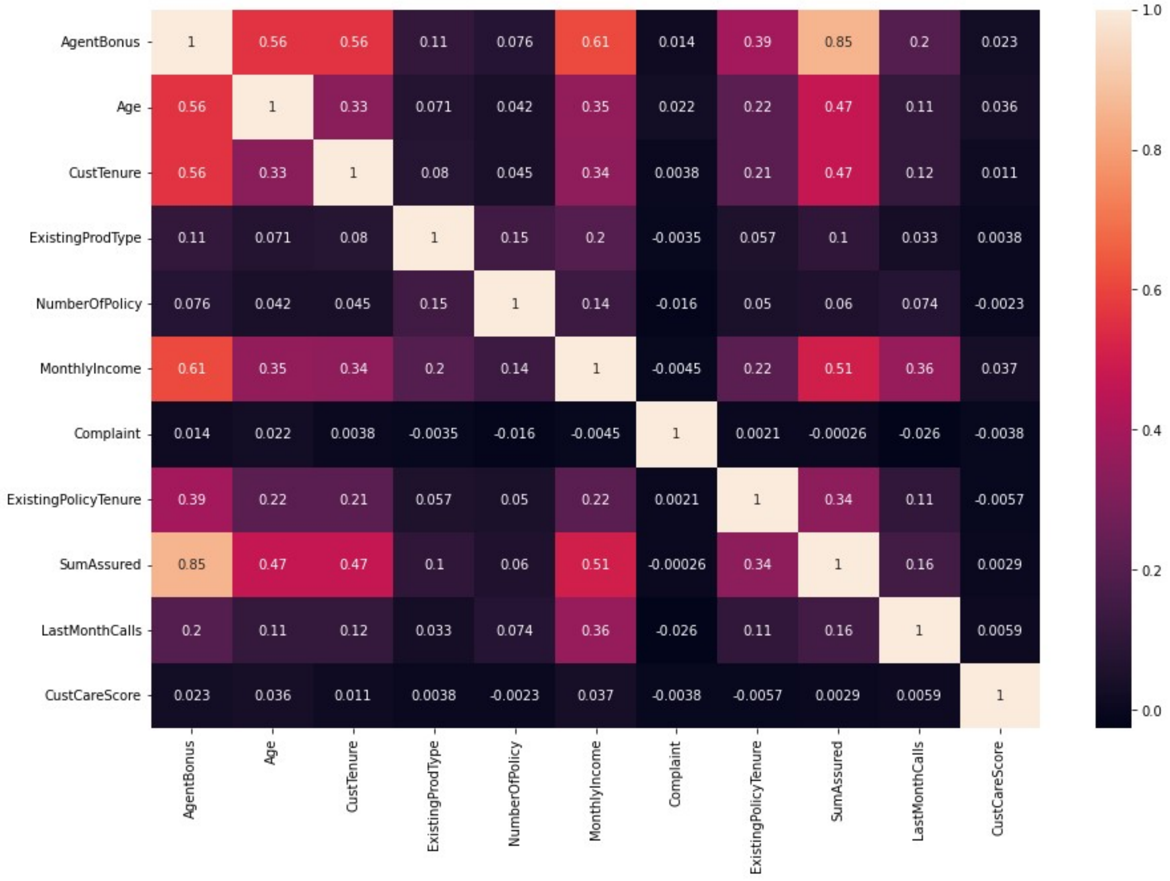
The average sum assured is approx. 6,20,000. The sum assured ranges from 1,68,536 to 18,38,496 whereas 75 % of customers are having sum assured below 7,50,000. We see that the range of the policy is good and maximum citizen can be added to the insurance umbrella which have such vast range(sum assured) of policies.

Fig 7: Boxplot of Age



A good percentage of customers are only having age upto 20 only and above that the age of insurers sharply drops. The current target people is young generation as per the dataset but we can tap on middle aged people(30-35) with various awareness programmes about the need and benefits of insurance as the people of such age are financially stable and can purchase insurance plan easily.

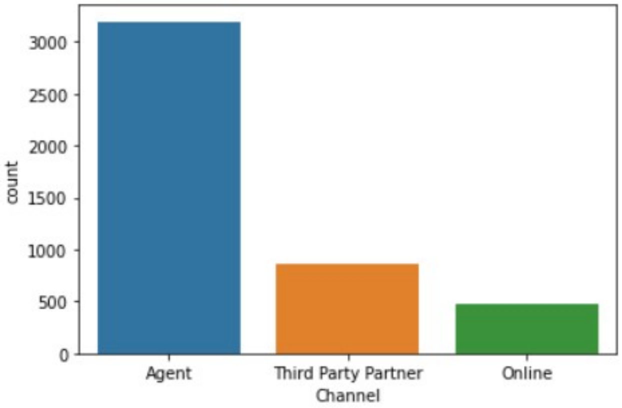
Fig 8: Multicollinearity in the dataset



We see from the heatmap that multicollinearity exists between some variables which will be treated at later stage.

The agent bonus can be seen with highly proportional relation with the Sum assured and monthly income as per above Figure.

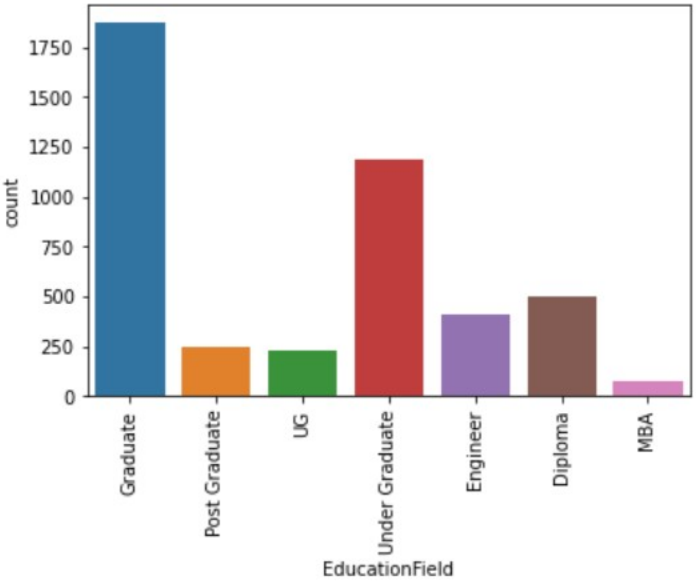
Fig 9: Bar plot of Channel with percentage



Channel	**Percentage**
Agent	0.71
Third Party Partner	0.19
Online	0.1

The Maximum customers are acquired through agents(71%) which shows a good performance by the agents whereas the company should add more efforts on online platform which can be a driving force of the sales in future.

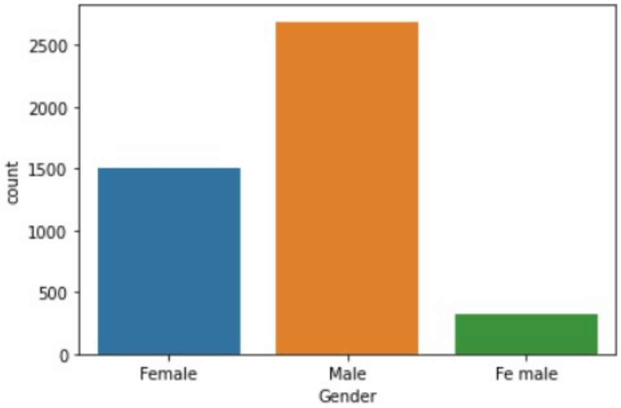
Fig 10: Bar plot of Education with percentage



Education Field	Percentage
Graduate	0.41
Under Graduate	0.26
Diploma	0.11
Engineer	0.09
Post Graduate	0.06
UG	0.05
MBA	0.02

We see that the maximum number of customers are having a graduation(41 %) or are under graduate and customers having MBA are very less as our customers. This may be because a very few people are having MBA degree generally in comparison to graduation.

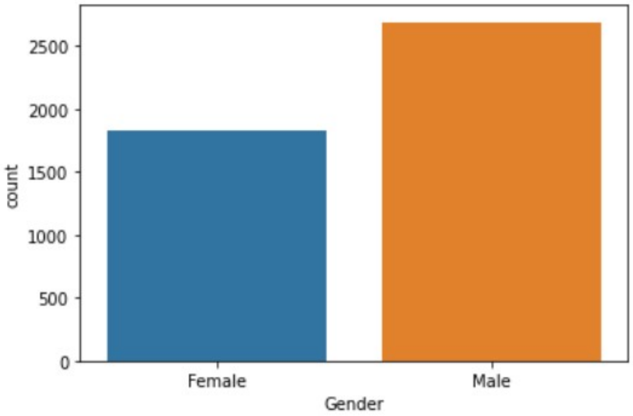
Fig 11: Bar plot of Gender



We see that the data needs correction as the Fe male is showing as a separate field which should be Female only.

After correction we see the below shown figure.

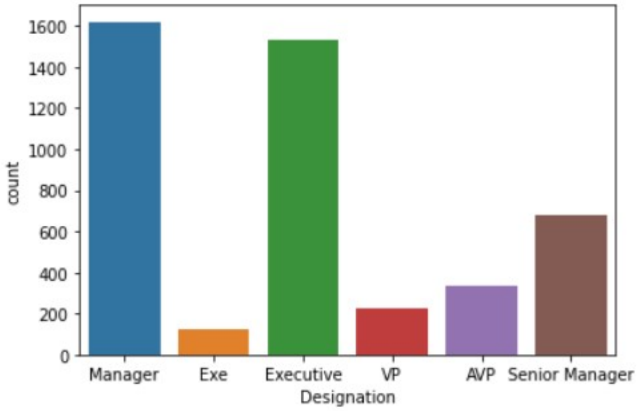
Fig 12: Bar plot of Gender after rectification with percentage



Gender	Percentage
Male	0.59
Female	0.41

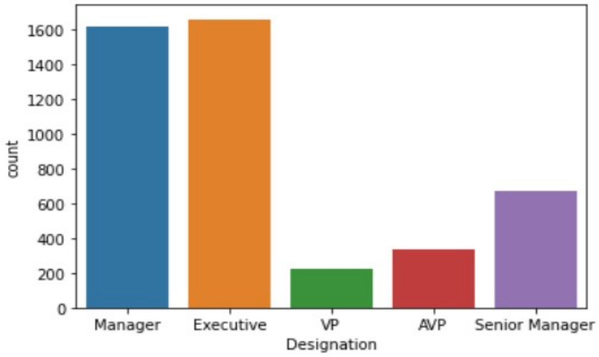
We can see that more number of male customers(0.59%) are insured in comparison to the female customers. The female population should be also given a priority to boost insurance penetration to each level of the population.

Fig 13: Bar plot of Designation



The field Exe is similar to Executive which needs to be rectified and therefore Exe is replaced with Executive.

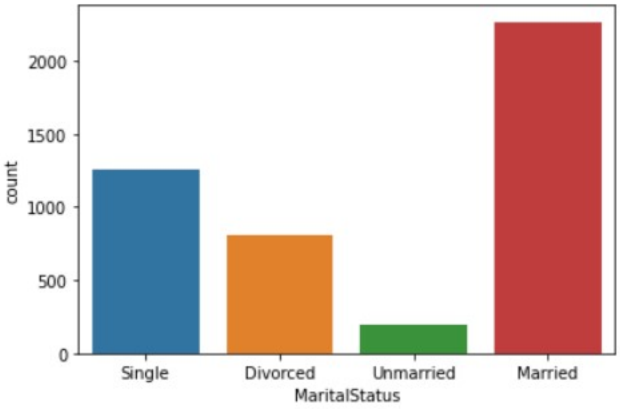
Fig 14: Bar plot of Designation after rectification with percentage



Designation	Percentage
Executive	0.37
Manager	0.36
Senior Manager	0.15
AVP	0.07
VP	0.05

Managers and Executives are having high percentage in the data set which can be due to lesser number of VP and AVP designated customers in a company's hierarchy.

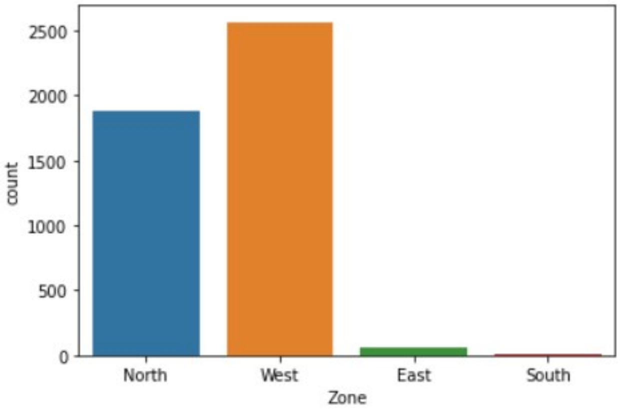
Fig 15: Bar plot of Marital Status with percentage



Marital Status	**Percentage**
Married	0.5
Single	0.28
Divorced	0.18
Unmarried	0.04

We see that married people (0.50%) are more concern about the insurance which may be because of married customer's concern to protect his family from any financial risk.

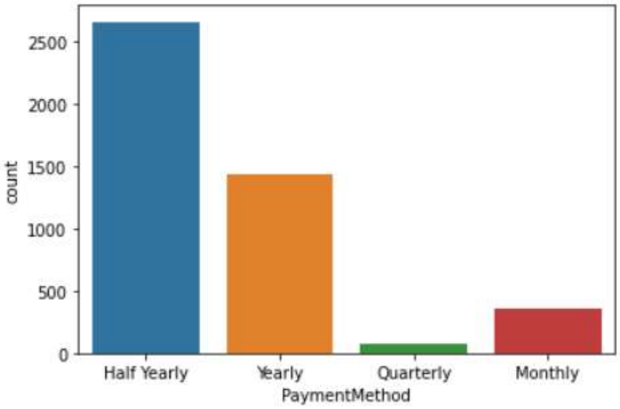
Fig 16 : Bar plot of Zone with percentage



Zone	**Percentage**
West	0.57
North	0.42
East	0.01
South	0

A very low number of customers from East and South zones(0.01%) are included in the data set. To get a better understanding of the patterns followed by the customers we should add more customers from these two zones. The better demographic distribution in the dataset would also help us to get a more correct model.

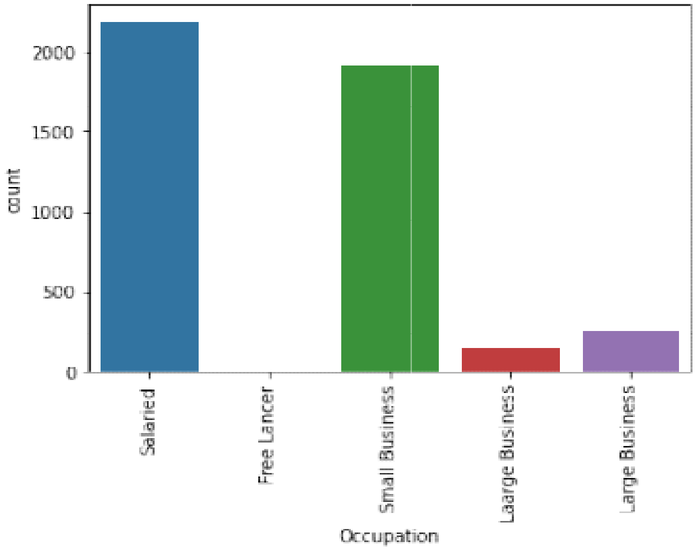
Fig 17: Bar plot of Payment method with percentage



Paymentmethod	Percentage
Half Yearly	0.59
Yearly	0.32
Monthly	0.08
Quarterly	0.02

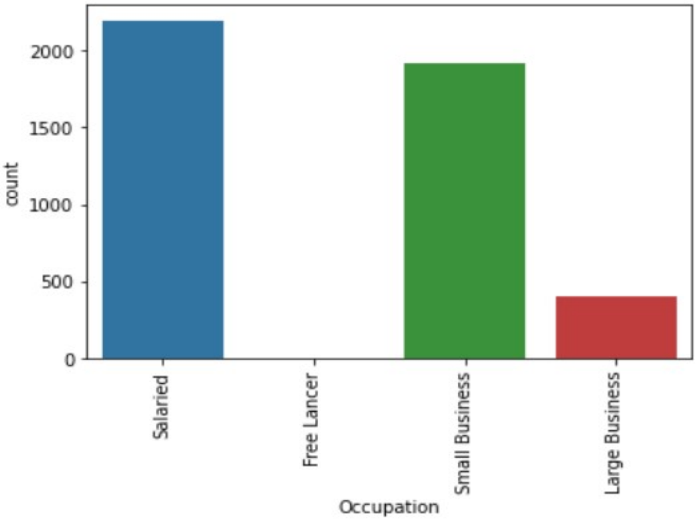
Maximum number of customers have opted for a half yearly(0.59%) and yearly payment(0.32%) of the premium.

Fig 18a: Bar plot of Occupation



The value Laarge business seems to be a spelling mistake therefore is replaced with Large Business.

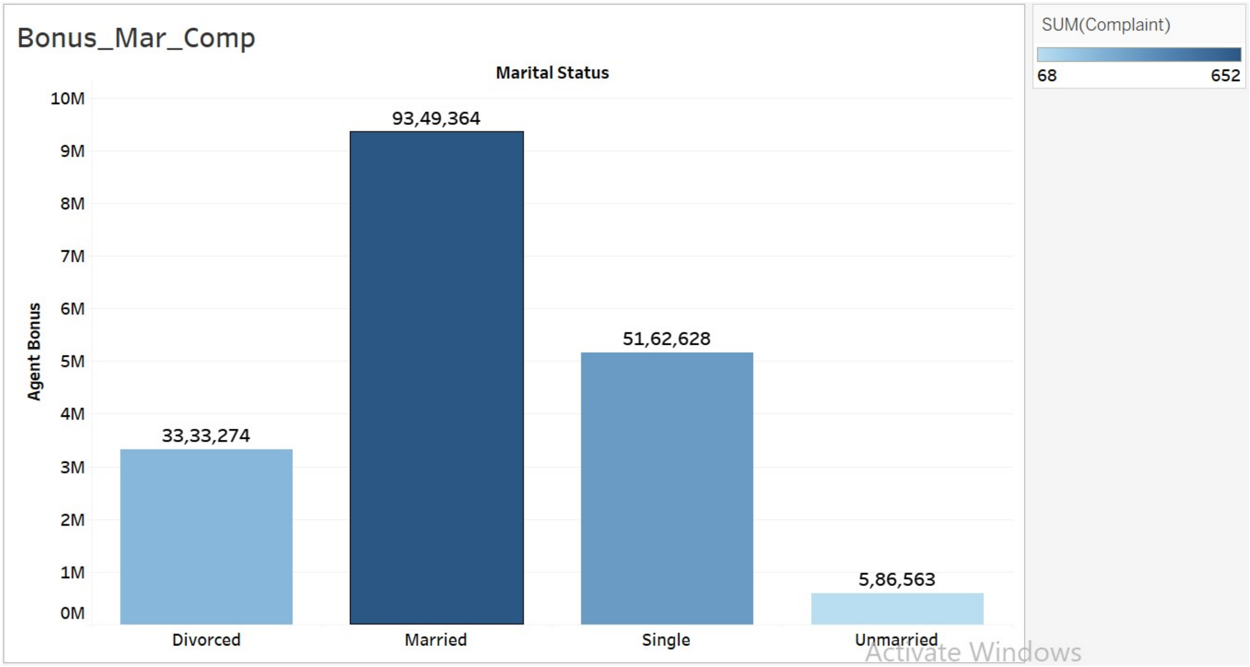
Fig 18b: Rectified Bar plot of Occupation with percentage



Occupation	Percentage
Salaried	0.48
Small Business	0.42
Large Business	0.09
Free Lancer	0

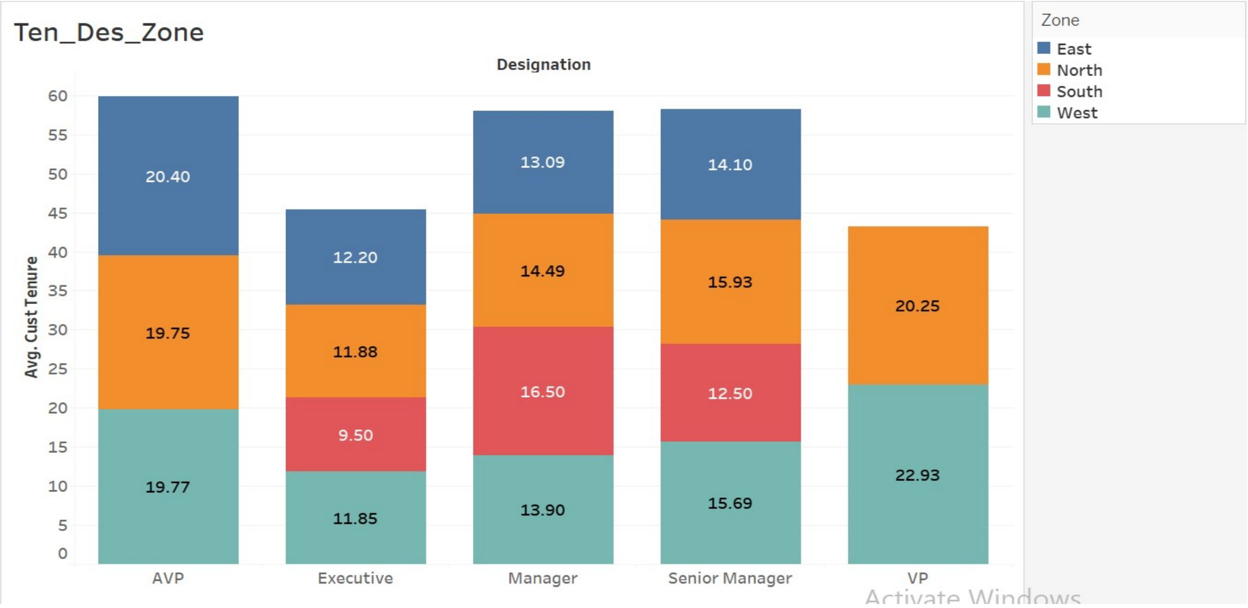
The Salaried(0.48%) and small business(0.42%) customers have high percentage of insurance subscription in comparison to the customers of Large Business and Free Lancer.

Fig 19: Agent Bonus vs Marital Status and Complaints



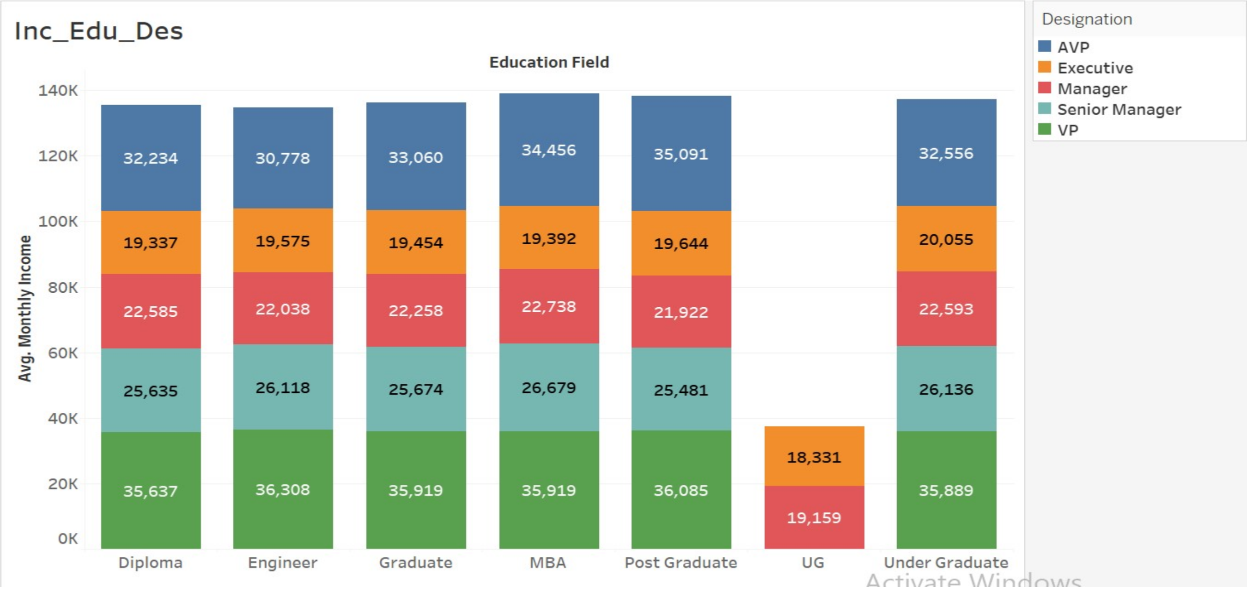
From the Fig 19 we can deduce that Married customers has produced highest of Agent Bonus and has also caused highest number of the complaints.

Fig 20: Avg Cust Tenure vs Designation and Zone



From Figure 20 we can deduce that Average tenure of the insurance policy is highest for AVP and VP designate customers. The agents should focus heavily on such customers. From Figure 21 we can deduce that Average monthly income is also highest for AVP and VP designate customers.

Fig 21: Avg Monthly Income vs Education Field and Designation



2a Business implications from EDA

-AgentBonus and SumAssured exhibit a strong correlation of 0.85 with each other. We can interpret that our model is strongly dependent on Sum Assured with respect to other variables therefore our focus should be to add more policies with higher Sum assured of the policy.

-We also found that the dataset suffers from severe imbalance, as exemplified by the limited representation of certain categories. For instance, the "Zone" category "South" has disproportionately low representation, similar to the "Occupation" category "Freelancer." To address this, we either require additional data or must implement data upscaling techniques. A comparable situation arises with the "EducationField_MBA," where data augmentation is necessary to ensure unbiased decision-making. However, upscaling the data could introduce repetition, potentially compromising accuracy in predictions.

-We need to add new variables such as "Premium," establishing a direct correlation with AgentBonus. This addition would facilitate the identification of high-performing and low-performing agents. Those responsible for generating higher premiums are likely contributing positively to the firm's success, while those with lower premium contributions may require additional attention and support.

3 Data Cleaning and Pre-processing

3a Handling Null values

Fig 22: Variables with null values

Variables	**No of Null values**
NumberOfPolicy	45
ExistingPolicyTenure	184
Age	269
CustTenure	226
MonthlyIncome	236
SumAssured	154
CustCareScore	52

We see that Age, Customer tenure, Number of policy, Monthly income, Existing policy tenure, sum assured and Cust Care score are having null values.

As the cells with null values are very low, we would replace these null values with mean(MonthlyIncome, SumAssured) and mode(Age, CustTenure, NumberOfPolicy, ExistingPolicyTenure, CustCareScore).

Imputing null values allows us to maintain the size and structure of our dataset, ensuring that we don't lose valuable information or observations. Removing rows with missing data can result in a loss of data, which may not be desirable if the missing data is informative or essential. Many statistical analyses and machine learning algorithms require complete data to function properly. Imputing missing values enables us to perform these analyses without running into issues related to incomplete datasets. Machine learning models typically cannot handle missing data directly. Imputation can make it possible to include potentially important features in our model, which can lead to better model performance and predictive accuracy.

3b Handling Outliers

As the outliers can have a disproportionate impact on statistical analyses and machine learning models. All the outliers has been removed from the variables Age, CustTenure, MonthlyIncome, ExistingPolicyTenure, SumAssured and AgentBonus with range capped with lower_range: Q1-(1.5*IQR) and upper_range: Q3+(1.5*IQR).

We have used IQR method because dataset provided to us has a skewed distribution, where the mean and standard deviation may not be robust measures of central tendency and spread. In such cases, the IQR method is a better choice because it relies on quartiles, which are less affected by extreme values.

3c Handling Categorical Values

As we are going to use regression model where we have to predict a continuous value i.e. Agent Bonus, we have used One hot encoding for categorical columns: (Channel, EducationField, Gender, Designation, MaritalStatus, Zone, PaymentMethod, Occupation and ExistingProdType).

After applying one hot encoding to above mentioned variables we have to remove one column thus formed from each categorical value to remain free from multicollinearity. The Columns thus removed are: Channel_Third Party Partner, EducationField_UnderGraduate, Gender_Male, Designation_VP, MaritalStatus_Unmarried, Zone_West, PaymentMethod_Yearly, Occupation_Small Business, ExistingProdType_5.

One-hot encoding is a technique used to convert categorical data into a binary format by creating a binary feature for each category or label in the original categorical variable. One-hot encoding is particularly useful in various situations, and it is the preferred choice in many cases due to its ability to handle nominal categorical variables effectively. One-hot encoding is compatible with a wide range of machine learning algorithms, including linear models, tree-based models (decision trees, random forests) and support vector machines.

3d Scaling the data set

We have used Standard Scaler method to scale the dataset. Standard scaling is often a good choice for machine learning algorithms that are sensitive to the scale of input features. These include linear regression, logistic regression, k-nearest neighbors, support vector machines. Scaling helps these algorithms perform better and converge faster during training.

3e Handling multicollinearity

To treat the variables having high multicollinearity among the continuous variables, we have used variance inflation factor method. As VIF values for all the continuous variables are below 5, therefore we proceed with all the variables.

Fig 23: Variables with their VIF value

Variables	VIF value
SumAssured	1.69
MonthlyIncome	1.52
Age	1.37
CustTenure	1.36
LastMonthCalls	1.15
ExistingPolicyTenure	1.09
NumberOfPolicy	1.02
CustCareScore	1
Complaint	1

The dataset is now divided in train and test set in ratio of 70:30 respectively for independent and dependent variables separately.

Further significance of remaining variables is to be checked so that the extra dimensions can be removed from the model. We have used OLS method of regression for the same. The variables with p value greater than 0.05 are removed from the dataset. The remaining variables with the p value is mentioned in the table shown below.

Fig 24: Variables with their p value

Variables	p value
const	0.928
Age	0
CustTenure	0
MonthlyIncome	0
ExistingPolicyTenure	0
SumAssured	0
Designation_Executive	0

Designation_Manager	0
Designation_Senior Manager	0
MaritalStatus_Divorced	0.008
MaritalStatus_Married	0.047
MaritalStatus_Single	0.003
PaymentMethod_Monthly	0.014

4 Model Building

As our target variable ‘Agent Bonus’ is a continuous variable and we have to describe the relationships between the Agent Bonus and independent variables, we will use various Regression models and select the best of them.

4a Regression models used:

-Linear Regression
Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. It is one of the simplest and most widely used techniques in statistics and machine learning for understanding and predicting the relationships between variables.

-Decision Tree Regression
It is used to predict a continuous numeric output variable based on a set of input features. Decision trees are tree-like structures where each node represents a decision or a test on a feature, and each branch represents an outcome or a decision based on the test. In the case of regression, the outcome at each leaf node is a numeric value. We have used the parameters for the model as: (criterion = 'squared_error', max_depth = 10, min_samples_split=2, random_state =5)

-K-Nearest Neighbour Regression
It is often referred to as K-NN regression or simply K-NN regression. K-NN regression is a non-parametric, instance-based machine learning algorithm that predicts a continuous numeric target variable based on the values of its nearest neighbors in the feature space.
We have used error plot to identify the optimum value of k which comes to be 5 for our dataset and then fit our model with k =5.

-Random Forest Regression
A Random Forest is a powerful ensemble machine learning model that combines the predictions of multiple decision trees to produce a more accurate and robust result. We have used Gridsearch hyper-parameter optimization to find the optimal combination of hyper-parameters (max_depth: 10, max_features: 6, min_samples_leaf: 50, min_samples_split: 100, n_estimators: 501) and then proceeded to fit the model.

4b Performance of the models

Fig 25: Performance Parameters of the Regression models

Models	R square train	R square test	RMSE train	RMSE test	MSE train	MSE test
Simple Linear regression	0.80	0.79	0.45	0.46	0.20	0.21
Decision tree regression	0.90	0.79	0.31	0.46	0.10	0.21
KNN regression	0.84	0.76	0.40	0.50	0.16	0.25
Random forest regression	0.84	0.82	0.41	0.43	0.16	0.18

We see that **Linear Regression model is best model** as it has similar R-squared for both train and test data, RMSE values also indicating good generalization. It also has low MSE values. Additionally, Linear Regression model is also simpler and more interpretable model compared to other models and provides a simplified proportional relation between the dependent variable and independent variables.

We have tried two tuning methods in Linear Regression model to tune the performance of the data set.

Tuning method 1: Elastic Net regression model
Tuning method 2: Changing Train and test ratio to 75:25

The performance of both the tuning methods with respect to the simple linear regression is as follows:

Fig 26: Performance parameters before and after tuning

Models	R square train	R square test	RMSE train	RMSE test	MSE train	MSE test
Simple Linear regression	0.80	0.79	0.45	0.46	0.20	0.21
Tuning method 1	0.78	0.78	0.46	0.47	0.22	0.22
Tuning method 2	0.80	0.79	0.45	0.46	0.22	0.21

We see that the performance remains same by and large for both the tuning methods therefore we will proceed with the initially built simple linear Regression model.

The coefficients for simple linear regression are:
The coefficient for const is 0.0
The coefficient for Age is 0.13
The coefficient for CustTenure is 0.15

The coefficient for MonthlyIncome is 0.11
The coefficient for ExistingPolicyTenure is 0.08
The coefficient for SumAssured is 0.58
The coefficient for Designation_Executive is -0.14
The coefficient for Designation_Manager is -0.16
The coefficient for Designation_Senior Manager is -0.08
The coefficient for MaritalStatus_Divorced is 0.04
The coefficient for MaritalStatus_Married is 0.04
The coefficient for MaritalStatus_Single is 0.05
The coefficient for PaymentMethod_Monthly is 0.01

The equation of the model for Simple Linear Regression model is as follows:
Agent Bonus = (Age*0.13) + (CustTenure*0.15) + (MonthlyIncome*0.11) + (ExistingPolicyTenure *0.08) + (SumAssured*0.58) + (Designation_Executive*-0.14) + (Designation_Manager *-0.16) + (Designation_Senior Manager * -0.08) + (MaritalStatus_Divorced *0.04) + (MaritalStatus_Married *0.04) + (MaritalStatus_Single* 0.05) + (PaymentMethod_Monthly *0.01)

5 Model validation

5a Through Performance parameters (refer Fig 25)

5a1 Performance of Simple Linear Regression

- The Linear Regression model has a decent fit to the data, with R-squared values around 0.80 for both the training and test sets, indicating that it explains a good portion of the variance in the target variable.
- The RMSE values are relatively low, suggesting that the model's predictions are close to the actual values, but it's essential to compare these values with the scale of our target variable to assess their practical significance.

5a2 Performance of Decision Tree Regression

- The Decision Tree Regression model shows excellent performance on the training data, with a very high R-squared value and low RMSE and MSE values, indicating that it fits the training data extremely well.
- However, the drop in R-squared and the increase in RMSE on the test data suggest potential overfitting. The model may not generalize as well to unseen data.

5a3 Performance of KNN Regression

- The KNN Regression model has reasonable performance, with R-squared values above 0.75 for both training and test data.
- However, it seems to suffer from overfitting, as indicated by the drop in R-squared and the increase in RMSE and MSE on the test data compared to the training data.

5a4 Performance of Random Forest Regression

- The Random Forest Regression model shows good performance. It has relatively high R-squared values on both training and test data, suggesting that it captures a significant portion of the variance in the target variable.
- The RMSE and MSE values are also relatively low, indicating that it provides accurate predictions.

Overall, based on these metrics, the simple Linear Regression model appears to perform well. It has a good balance between explaining the variance in the target variable (as indicated by high R-squared values) and making accurate predictions (as indicated by low RMSE and MSE values). Additionally, the model's performance on the test data is consistent with its performance on the training data, which suggests it generalizes effectively and doesn't suffer from overfitting. However, Linear Regression model is also simpler and more interpretable model compared to other models and provides a simplified proportional relation between the dependent variable and independent variables.

5b Through K Fold Cross Validation

By performing K-fold cross-validation, we obtain a more reliable estimate of our Regression model's performance compared to a single train-test split, as it ensures that your model is evaluated on multiple different subsets of data, reducing the impact of randomness in the data splitting process. It helps in gaining a better understanding of how well your model will perform on unseen data and whether it might be overfitting or underfitting the training data.

Fig 27: MSE scores of K fold cross validation

Models/MSE Scores	**Test MSE**	**K fold 1 MSE**	**K fold 2 MSE**	**K fold 3 MSE**	**K fold 4 MSE**	**K fold 5 MSE**	**Avg of K Fold**
Simple Linear Regression	0.21	0.21	0.21	0.2	0.21	0.19	0.20
Decision tree regression	0.21	0.19	0.18	0.2	0.20	0.17	0.19
KNN regression	0.25	0.22	0.24	0.2	0.24	0.23	0.23
Random forest regression	0.18	0.17	0.18	0.2	0.18	0.16	0.17

We have chosen K value as 5 for our validation test for all the models. We see from the above shown table that the Test MSE score of the models and average of the MSE scores of all K fold models are almost same and doesn’t deviate much. This confirms that our models are performing well and we can proceed with the interpretation from the selected model(Simple Linear Regression).

6 Final interpretation / recommendation

6a Important Variables and its effects

-Age (Coefficient: 0.13): As age increases by one unit, the bonus of agents is expected to increase by approximately 0.13 units, assuming all other variables are held constant. This suggests that older customers tend to generate higher bonuses.

-Customer Tenure (CustTenure) (Coefficient: 0.15): An increase in customer tenure by one unit is associated with an increase in agent bonuses by approximately 0.15 units. Longer-lasting customer relationships appear to be positively correlated with agent bonuses.

-Monthly Income (Coefficient: 0.11): For every one-unit increase in monthly income, the agent's bonus is expected to increase by approximately 0.11 units.
Agents with customers who have higher monthly incomes tend to receive higher bonuses.

-Existing Policy Tenure (Coefficient: 0.08): A one-unit increase in the tenure of existing policies is associated with a 0.08 increase in agent bonuses. Agents who canvass policies with higher policy tenure tend to receive higher bonuses.

-Sum Assured (Coefficient: 0.58): The sum assured has a significant positive impact on agent bonuses, with a coefficient of 0.58 and is highly significant. Agents dealing with policies with higher sum assured amounts receive substantially higher bonuses.

-Designation (Executive, Manager, Senior Manager): Negative coefficients (e.g., -0.14 for Executive, -0.16 for Manager, -0.08 for Senior Manager) suggest that the customers with these designations tend to generate lower bonuses compared to customers with other designations.

-Marital Status (Divorced:0.04, Married: 0.04, Single: 0.05): Positive coefficients suggest that customers with certain marital statuses (e.g., Divorced, Married, or Single) tend to generate higher bonuses compared to others.

-PaymentMethod_Monthly (Coefficient: 0.01): Customers paying their premium of insurance monthly are generating slightly higher bonus for the agents.

6b Recommendations

-The dataset exhibits uneven distribution among its variables and necessitates correction through the incorporation of a more comprehensive dataset. This will enable the development of a universally robust model.

-The insurance company ought to include additional variables in the dataset that may demonstrate a strong correlation with agent bonuses, such as:
*Premium amount received by the company.
*Premiums missed in last one year.
*Number of dependents of the customers.
*Classification of residential of customers in urban, rural and semi urban.

-Agents should prioritize reaching out to customers with higher levels of sum insured, greater monthly income, increased age, longer policy tenure, and individuals holding the titles of AVP and VP, as these are the customers of greater significance to them.

-Agents who fall into the lower bonus bracket within a specified time frame should participate in ongoing training and skill development programs. This will help them remain current on industry trends, insurance products, target customer demographics, and sales techniques.

-Introduce incentive and recognition programs that incentivize agents for selling policies with higher sums assured, which is the most pivotal factor. These programs can serve as powerful motivators for agents, encouraging them to concentrate on policies that align with the company's objectives.

-Empower agents to educate customers about the value of insurance and how it can protect their financial well-being. Agents should receive encouragement and support to host camps and seminars. The goal is to reach out to previously untapped segments of the market and enroll as many new customers as possible for insurance coverage.