

Capstone Project
Health Insurance Cross
Sell

Team - Dream

Team Members
Indugopal Maity

-
- Importance of health insurance in India**

Importance of health insurance in India

What is Health Insurance??

Health insurance is a type of insurance that covers medical expenses that arise due to an

illness.

The health insurance concept was first suggested in the year 1694 by Hugh the Elder Chamberlen from Peter

Chamberlen family.

Currently, there are around 10 different types of schemes are available around the globe



What is Vehicle Insurance Insurance??

Vehicle insurance is insurance for cars, trucks, motorcycles, and other road vehicles.

Gilbert J. Loomis holds the distinction of being the first person to buy an automotive liability insurance policy

in 1897

Currently, there are around 15 different types of schemes available around the globe



How does Health Insurance work??

Health insurance is a legal agreement between you and a health insurance company. That agreement includes a Health plan that helps you pay for certain medical care and services, so you don't have to pay all the costs on your own.

Benefits: Insured can have

1. Direct Payment(Cashless).
2. Reimburses the expenses
3. Protection for you and your family
4. Reduce stress during difficult time

Family Health Insurance Plan



What is Cross Sell??

Cross-selling is to sell related or complementary products to an existing customer

CROSS-SELLING



before



after

Where can we find Health Insurance??

Health Insurance are available

1. Online(Different Websites)
2. Offline(Different Agents)



Data Overview:

- Data set obtain from “**TRAIN-HEALTH INSURANCE CROSS SELL PREDICTION.csv**”
- **Provided Feature Set**
 - **Id, Gender, Age, Driving_License, Region_Code, Previously_Insured, Vehicle_Age, Vehicle_Damage, Annual_Premium, Policy_Sales_Channel, Vintage, Response**

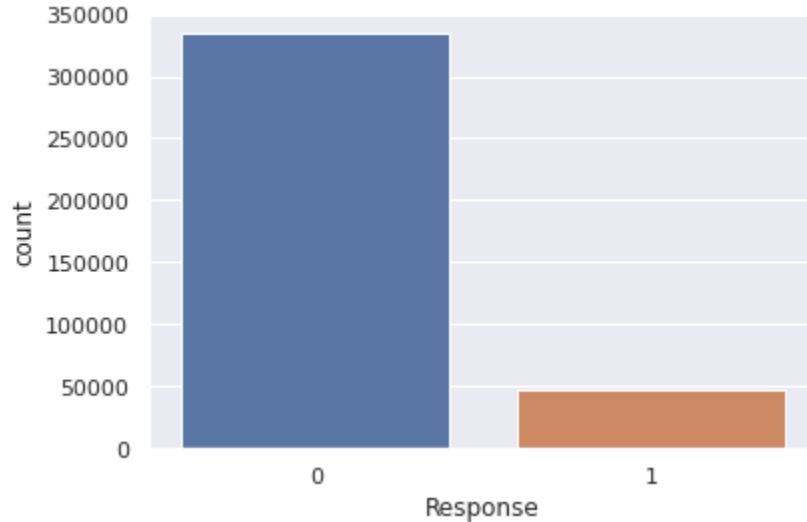
	id	Gender	Age	Driving_License	Region_Code	Previously_Insured	Vehicle_Age	Vehicle_Damage	Annual_Premium	Policy_Sales_Channel	Vintage	Response
0	1	Male	44	1	28.0	0	> 2 Years	Yes	40454.0	26.0	217	1
1	2	Male	76	1	3.0	0	1-2 Year	No	33536.0	26.0	183	0
2	3	Male	47	1	28.0	0	> 2 Years	Yes	38294.0	26.0	27	1
3	4	Male	21	1	11.0	1	< 1 Year	No	28619.0	152.0	203	0
4	5	Female	29	1	41.0	1	< 1 Year	No	27496.0	152.0	39	0



Exploratory Data Analysis (EDA)

Relationship: Insurance to Response Count

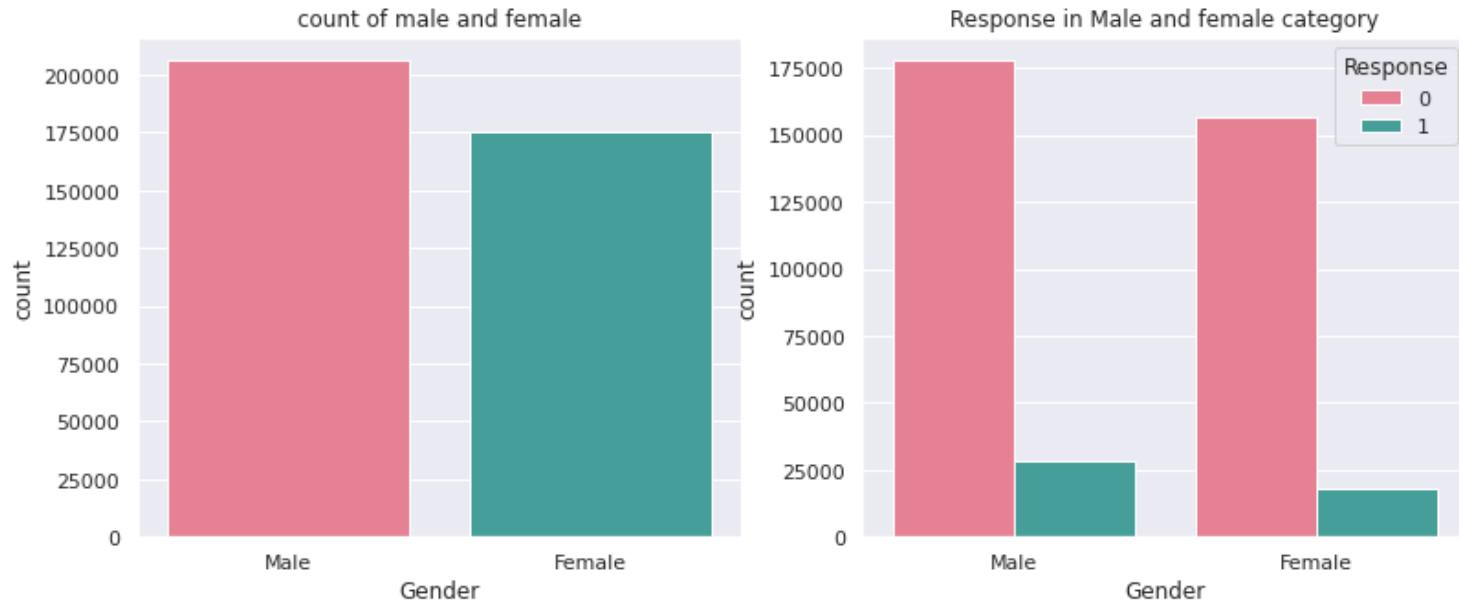
AI



We can see the data is highly imbalanced. People like to go for the Insurance is less Than **“50,000”** and people don't like to go for Insurance is near to **“3,50,000”**.

Relationship: Response Count to Gender

AI

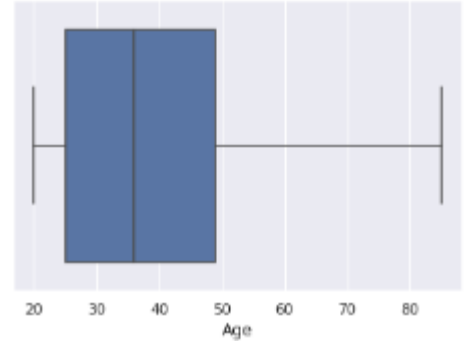
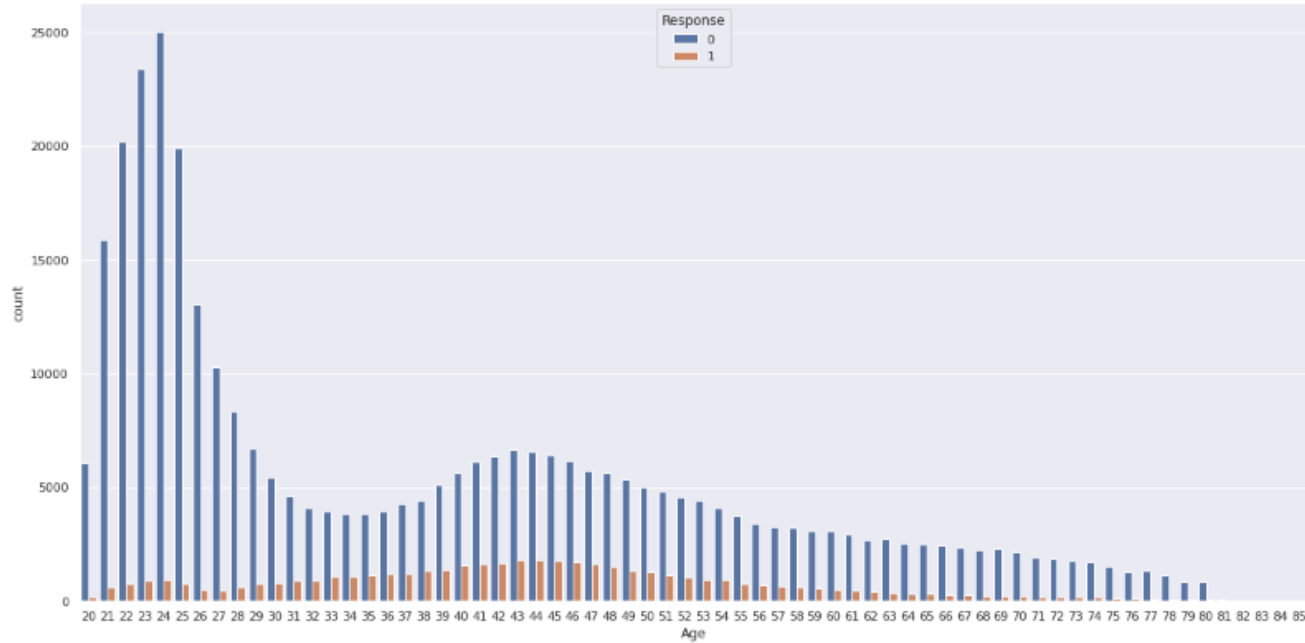


Gender variable are almost equally distributed, Male are little higher than Female.

Chances of Buying Insurance for Men is more than "25,000" and for Female is near to "20,000"

Relationship: Response Count to Age

AI



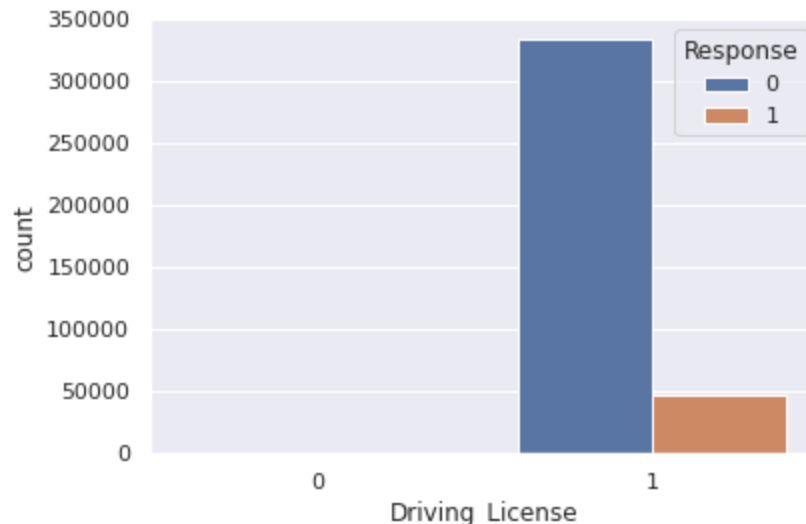
Young people below 28 are less interest in Insurance, may be due to “Less Experience”

People aged between “30-60” are more likely to interest in Insurance.

From boxplot, there is no outlier in the data.

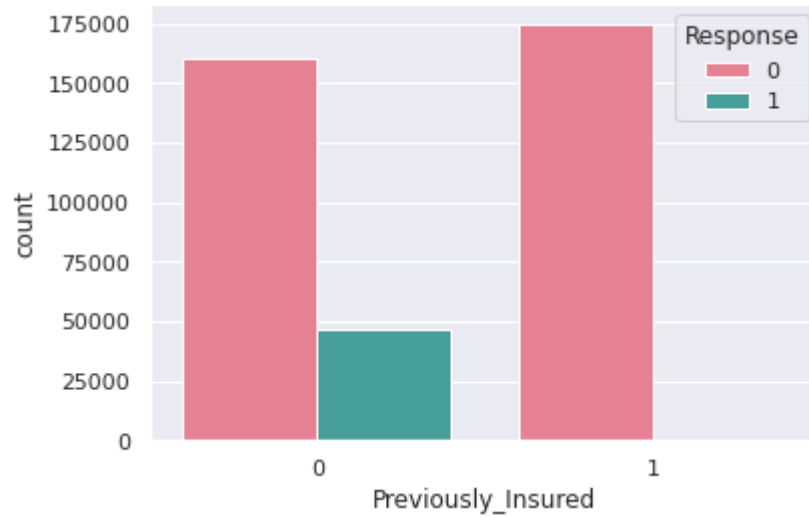
Relationship: Health Insured to Vehicle Insurance

AI



Customers who are interested in **Vehicle Insurance** almost all have **Driving License**.

Relationship: Previously Insured to Response Count AI

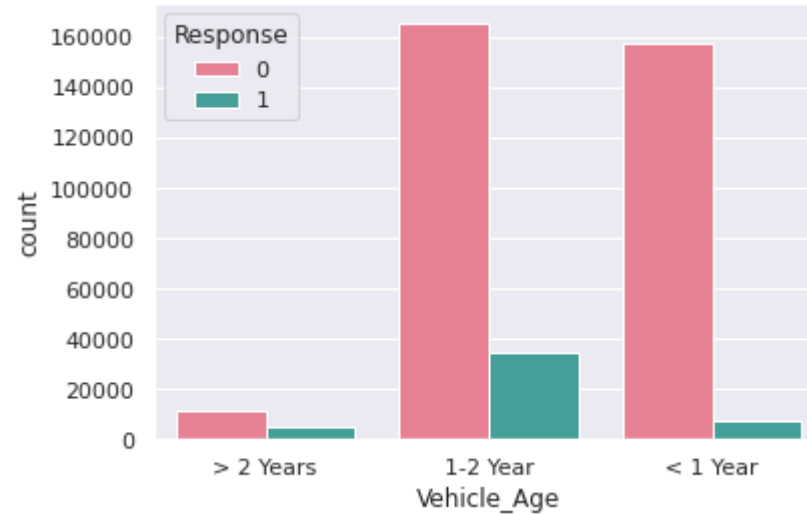


Customers who are **not previously insured**, are likely to be **interested**.

Customers who are **previously insured**, **not interested** to buy **Vehicle Insurance**.

Relationship: Vehicle Age to Response Count

AI

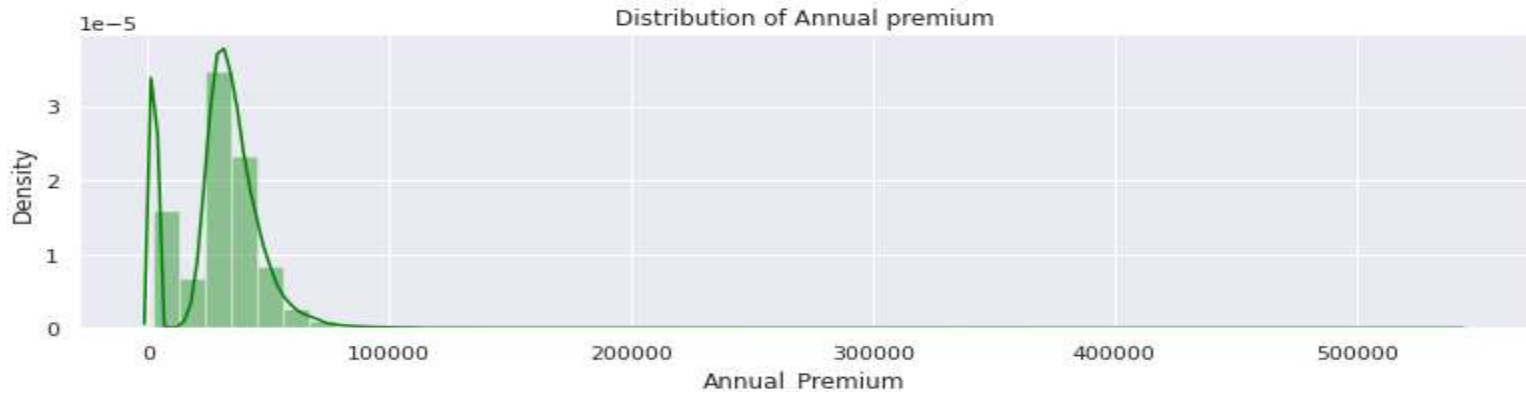


Customers with **Vehicle age 1-2 years** are more likely to **interested** as **compared to other two**.

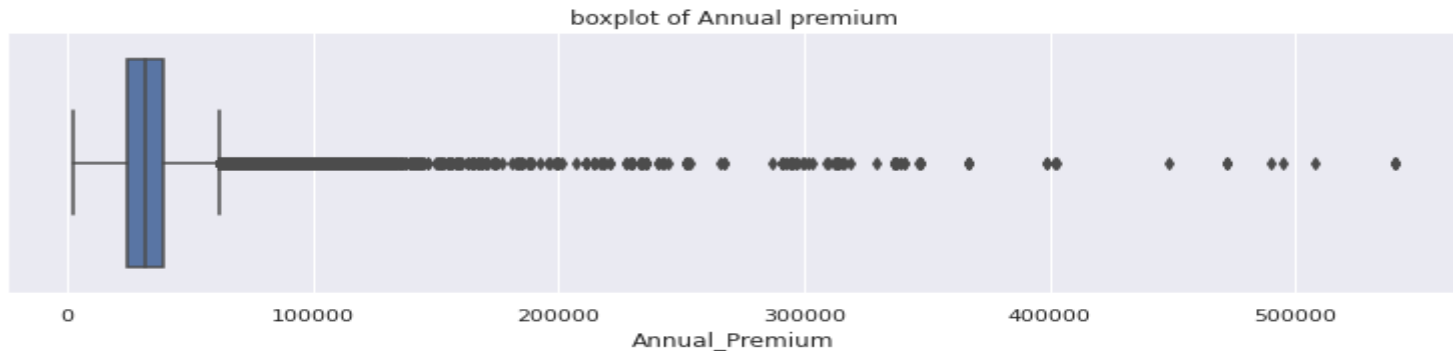
Customers with **Vehicle age <1 years** have **less chances** of buying Insurance.

Relationship: Annual Premium to Response Count

AI



Distribution variable says **annual premium** is **right skewed**, means it **goes down**

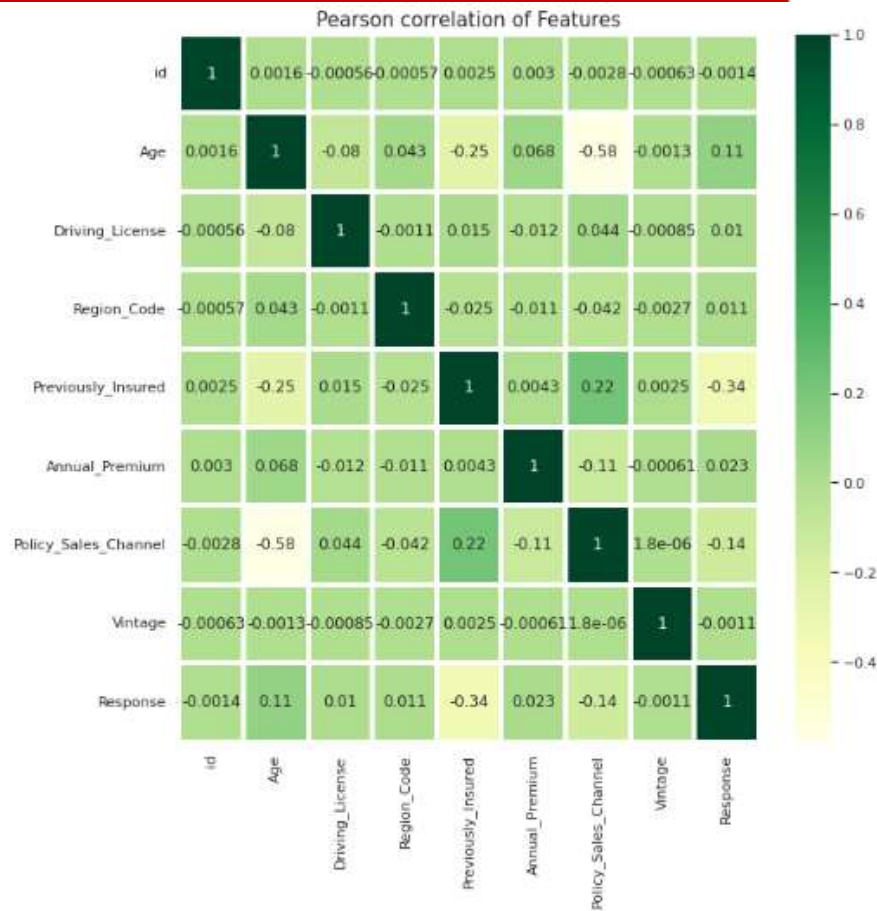


Boxplot says **lot of outliers** in the

Correlation Analysis

Correlation Analysis

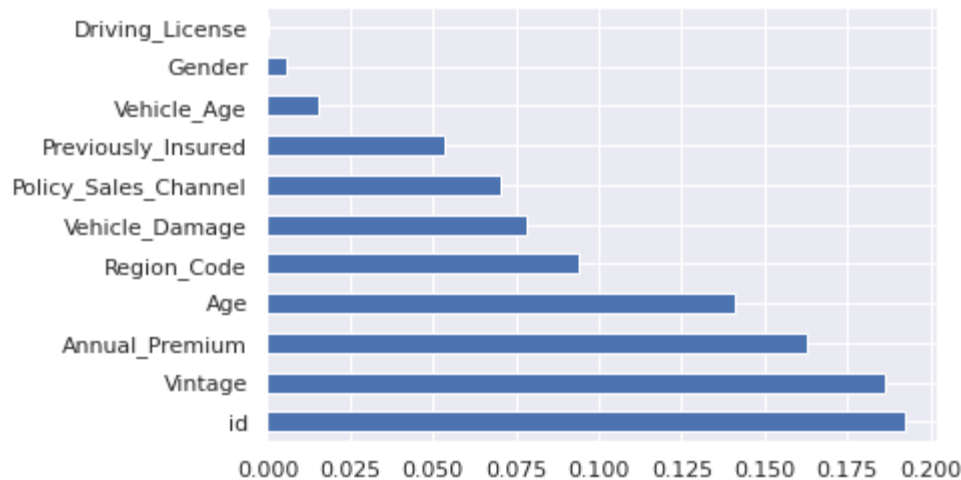
- Target variable is not much affected by vintage variable
- We can drop most co-related value.





Feature Selection

Feature Selection

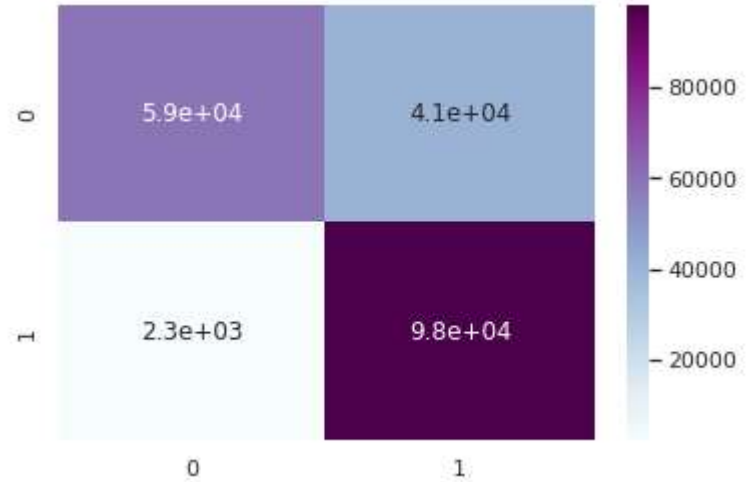
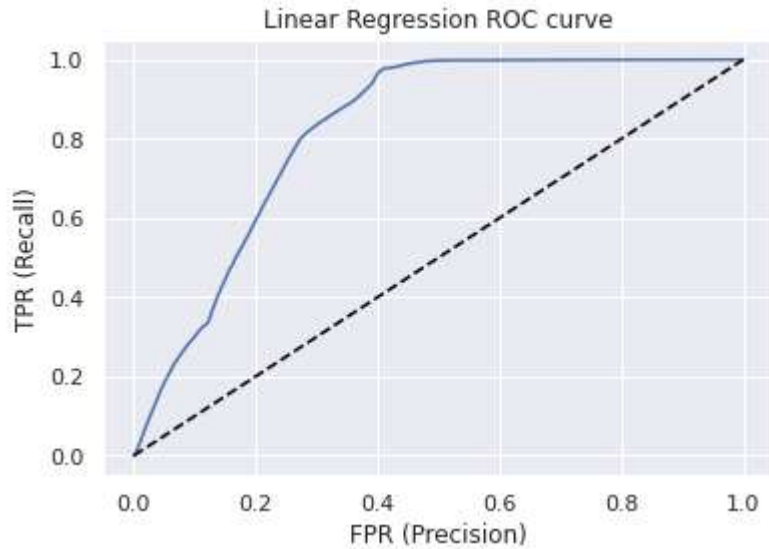


We can remove the less important features like “Driving_License” and “Gender”.

A large, dark, textured brushstroke background, resembling a thick application of black paint with visible bristles and some lighter areas, creating a sense of movement and depth.

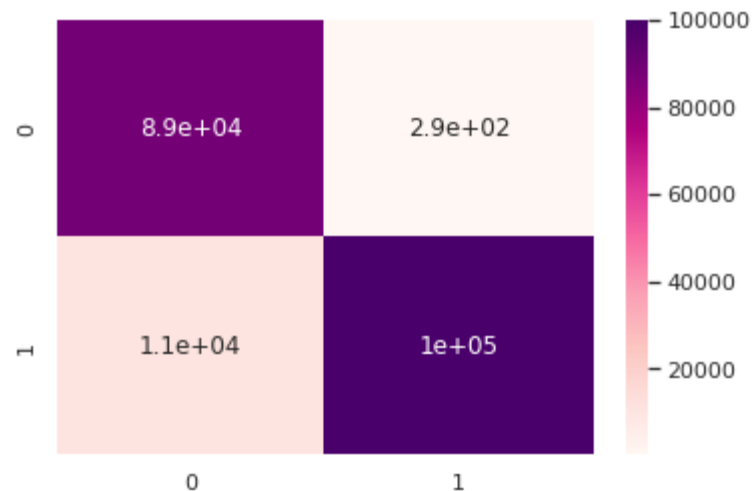
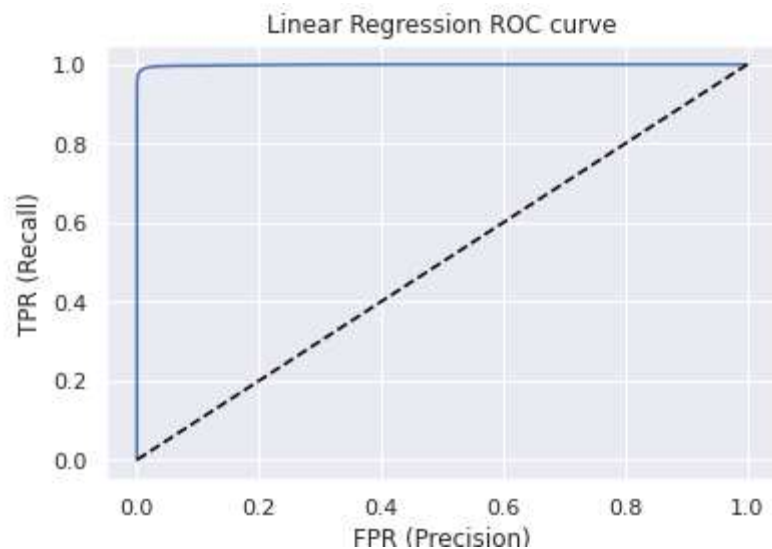
Regression

Logistic Regression

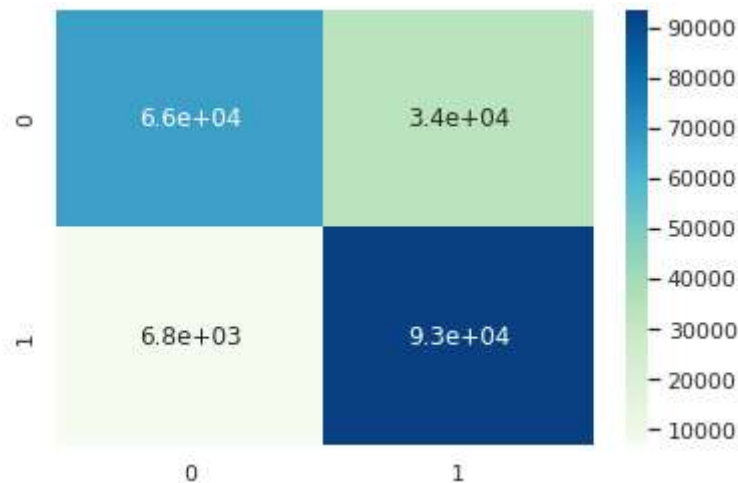
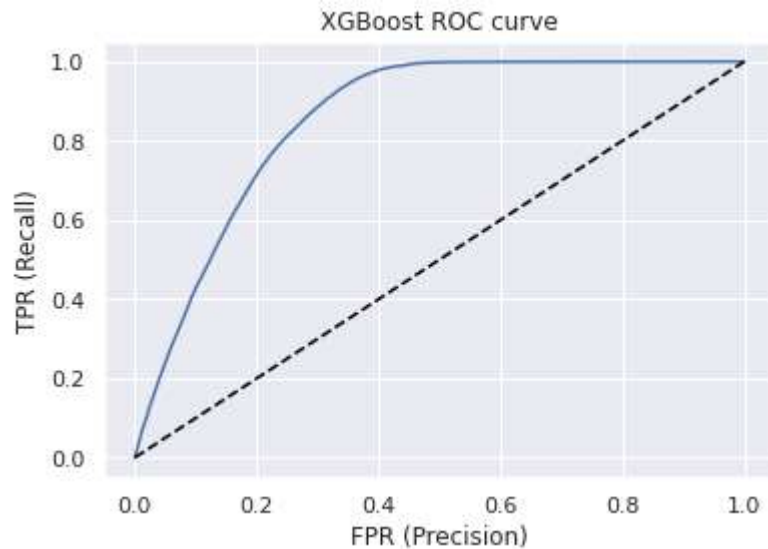


RandomForest Classifier

AI



XGB Classifier



Model Comparison

	Accuracy	Recall	Precision	f1_score	ROC_AUC
Logistic regression	0.783543	0.977111	0.704391	0.818635	0.833526
Randomforest	0.944318	0.997079	0.901901	0.947105	0.949318
XGBClassifier	0.796885	0.931843	0.733765	0.821026	0.820213

The ML model for the problem statement was created using python with the help of the dataset, and the ML model created with Random Forest and XGBClassifier models performed better than Logistics Regression model. Thus, for the given problem, the models created by Random Forest and XGBClassifier.

- Customers of age between 30 to 60 are more likely to buy insurance.
- Customers with Driving License have higher chance of buying Insurance.
- Customers with Vehicle_Damage are likely to buy insurance.
- The variable such as Age, Previously_insured, Annual_premium are more affecting the target variable.
- comparing ROC curve we can see that Random Forest model perform better. Because curves closer to the top-left corner, it indicate a better performance.

Q & A