School of Engineering and Applied Science
CSE523- Machine Learning

Project Report - 3

Vehicle Insurance Predictor

Submitted to: **Prof. Mehul Raval**
Date of submission**: 2nd Mar 2021**

Members:

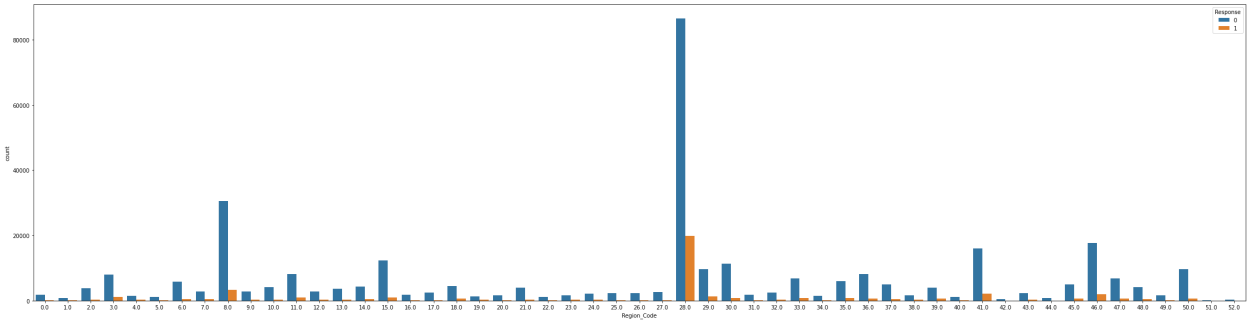
Name	Roll Numbers	Branch
Samarth Shah	AU1841145	B. Tech, ICT
Priyank Sangani	AU1841136	B. Tech, ICT
Yash Patel	AU1841141	B. Tech, ICT
Shaili Gandhi	AU1841012	B. Tech, ICT

Task performed this week:

- Data Preprocessing
- Data Visualization
- Data Insights

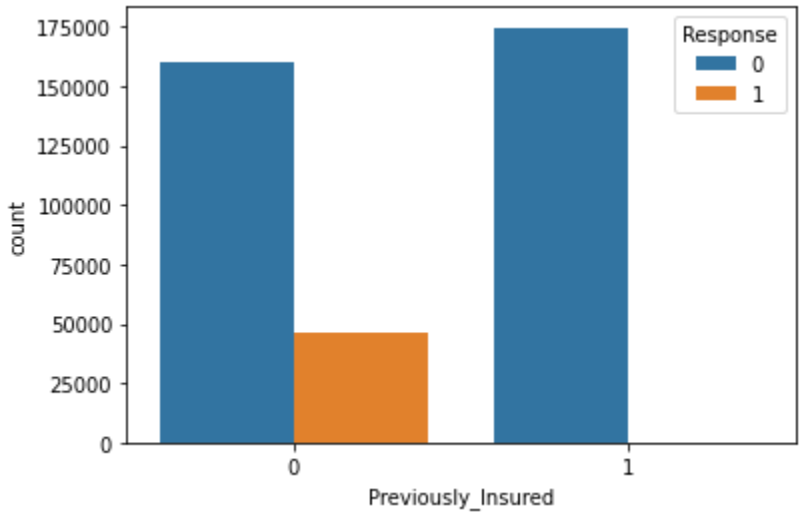
Outcomes of the tasks performed:

1) Plot between Region Code vs Response of the customer



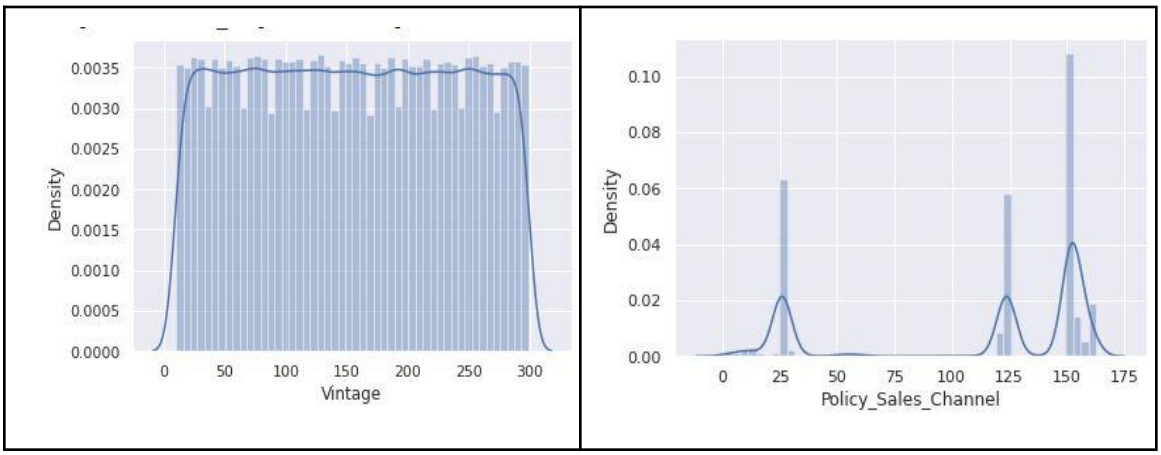
- As observed in the plot, most of the positive response of the customer has come from region code = 28.0.
- Rest of the region codes have few positive responses from the customer.

2) Plot between Previously Insured vs Response of the customer



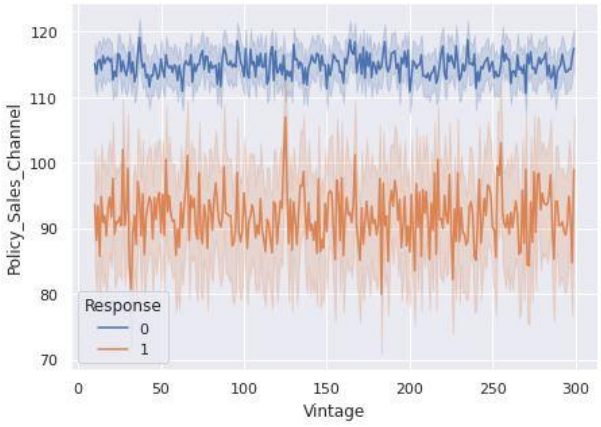
- As shown in the above plot, only customers which had positive responses towards the insurance were the customers which did not have insurance previously.
- Customers which already have vehicle insurance are not interested in buying the new insurance from our company.

3) Plots of Vintage count and Policy sales channel count



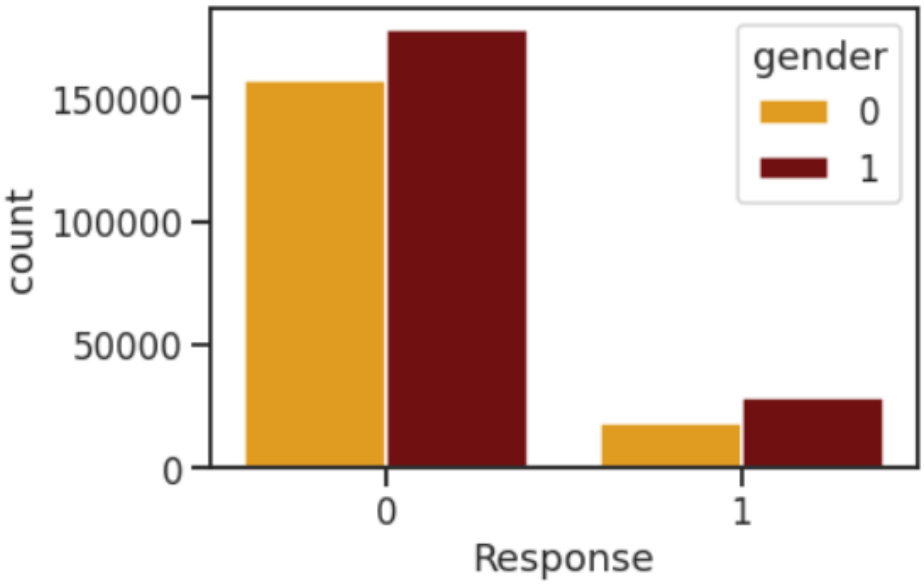
- For the first plot we can say that Vintage distribution is quite similar to uniform distribution. Also there are no outliers as the range is defined.
- For the second plot we can say that most of the distribution is governed by some of the channels (eg. 25, 125, 150, etc). There are no outliers in this plot.

4) Plot of relation between Vintage and Policy sales channel



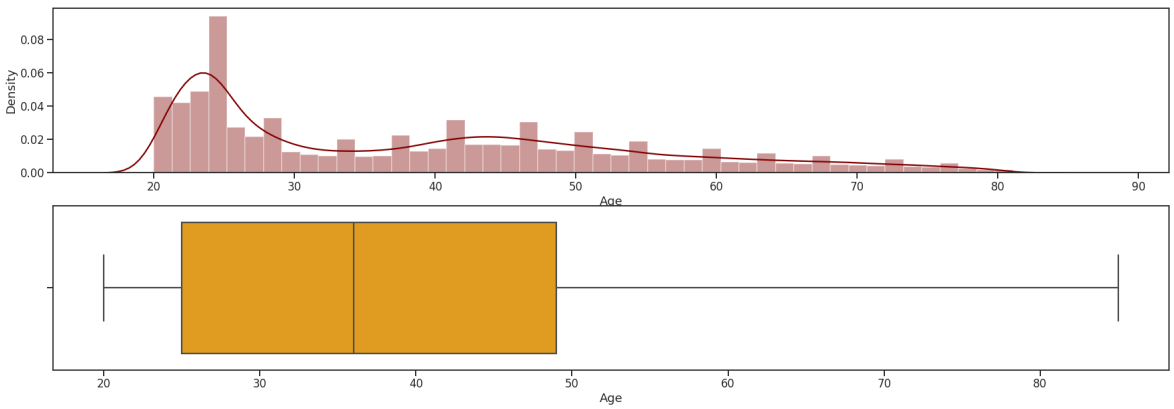
- According to this plot we can say that insurance is guaranteed if the policy sales channel is below 100.

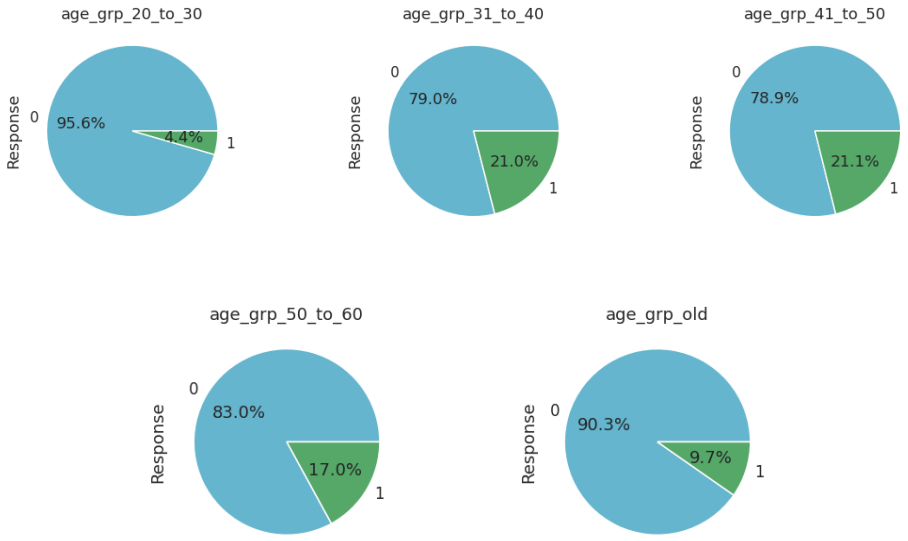
5) Plots of relation between gender and number of response:

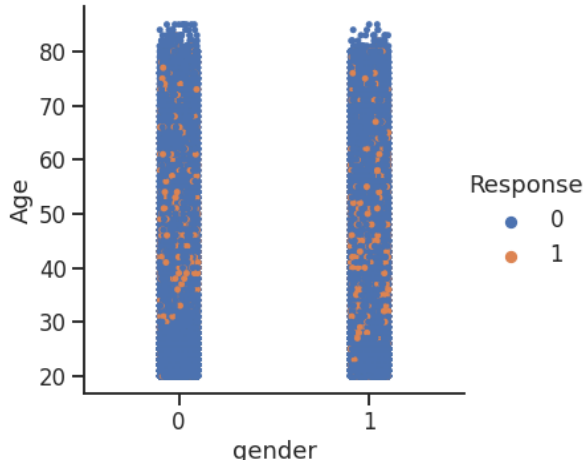


- From the above plot, we can derive that the male and female customers who don't have previously insured are more likely to buy vehicle insurance.
- The count of male customers is greater than the female customers who are likely to buy vehicle insurance.

6) Plots of relation between customer's age and number of response:

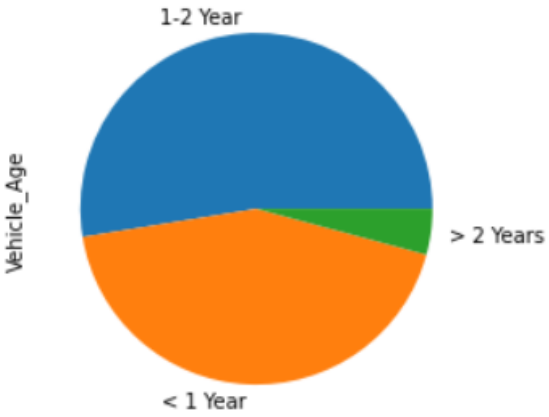




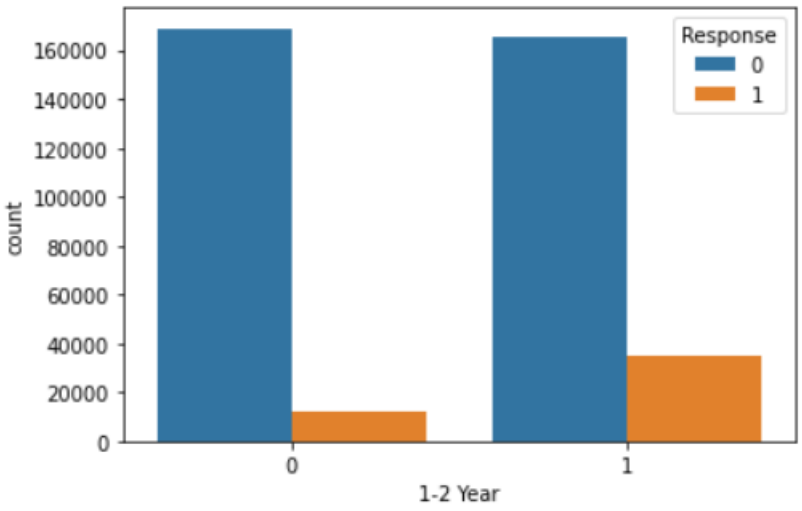


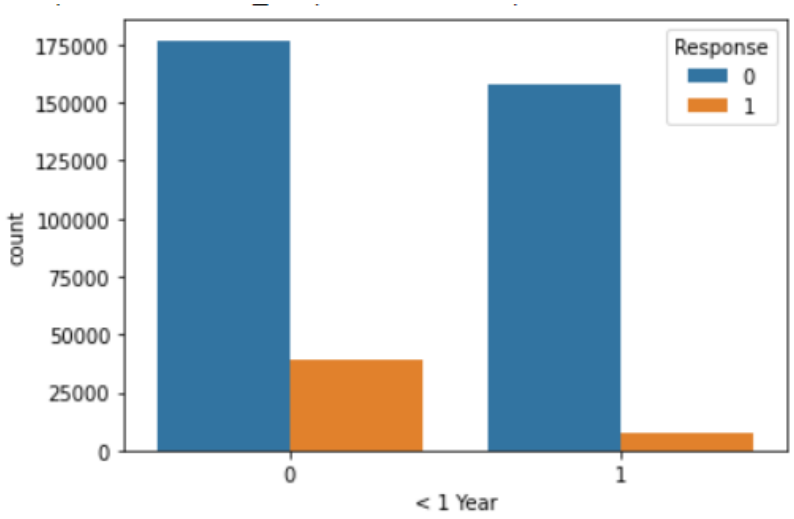
- In almost every age group, 'Male's are more likely to buy insurance.
- Females under age 30 are very less likely to buy insurance
- Costumes of age between 20 to 30 are more likely to buy insurance.
- Customers of age between 30 to 60 are less likely to buy insurance.

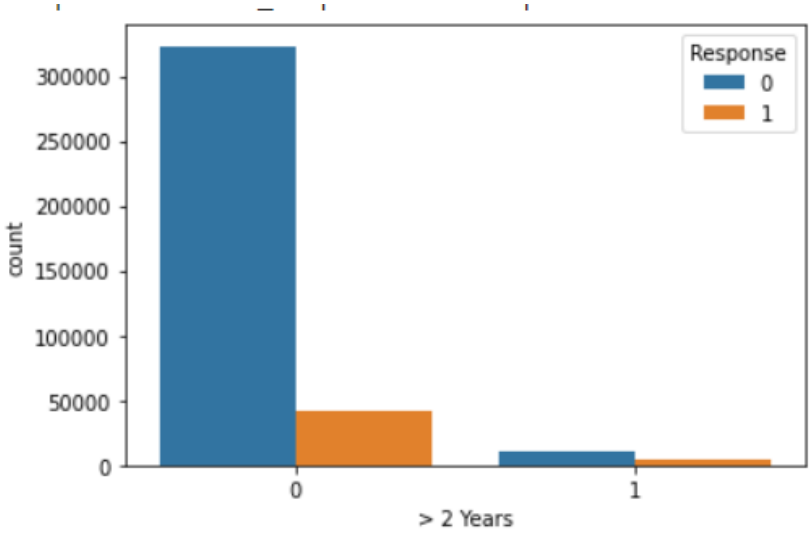
7) Relation between the age of vehicle and number of responses:



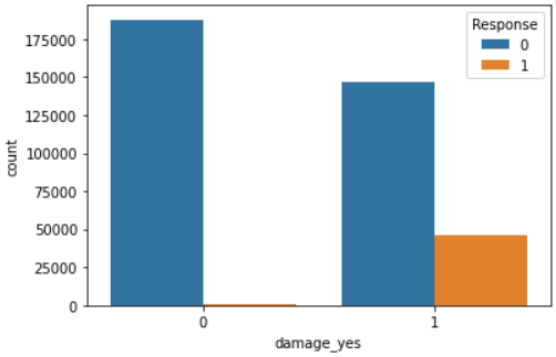
- Half of the customers have the vehicle with age 1-2 years.
- The proportion of customers with a vehicle age of more than 2 years are very less.
- Out of the customers with vehicle age of more than 2 years, those buying the insurance are considerable proportionally.
- Out of the customers with vehicle age of less than 1 year, those buying the insurance are less proportionally.

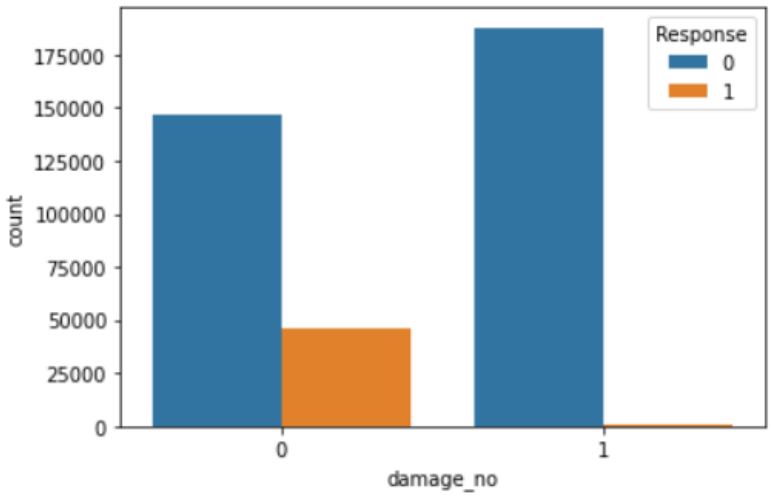






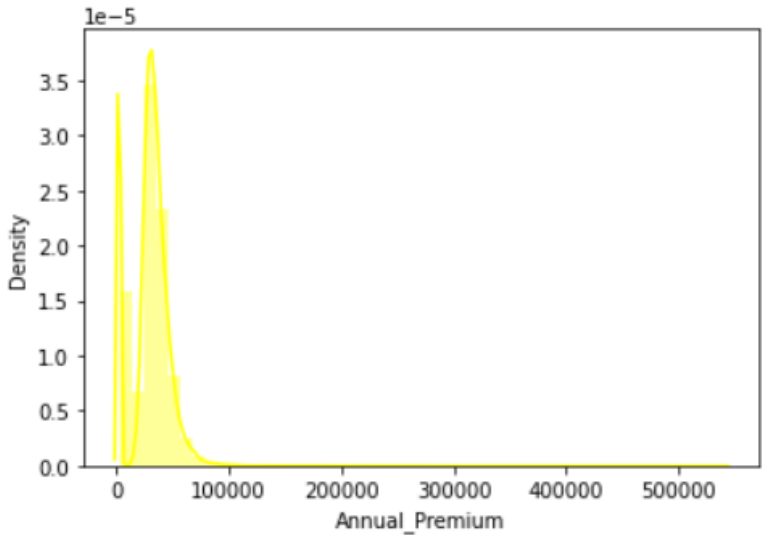
8) Relation between the Vehicle Damage and the number of responses

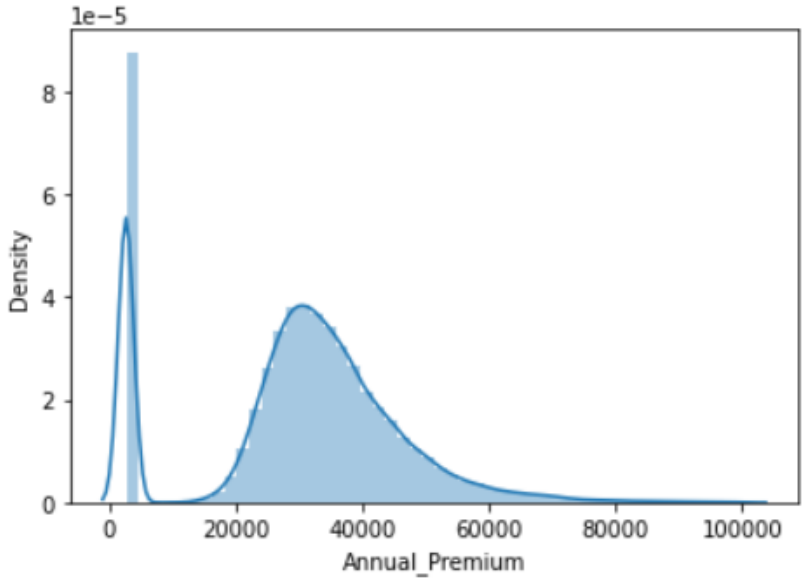




- Customers without vehicle damage are less than 1% likely to buy insurance.
- The number of customers with vehicle damage and without damage are equal.
- Customers with vehicle damage are comparatively more likely to buy the insurance.

9) Annual Premium and the density of customers:





- The data is highly left skewed.
- The annual premium for most of the customers is in the range of (0,10000) and (20000,50000).

Tasks to be performed in the upcoming week:

1) After visualisation plots we would like to go for exploring different types of classification models.
2) Also we would like to search and understand the equations for logistic regression models.