**Health Insurance Cross Sell Prediction**

**Mentored by:**

Mr. Vikash Chandra

**Submitted by:**

1. Mihir.V.Joshi
2. Alan David
3. Kishore P.G
4. Dhayanithi
5. Thomas Loyola

**Abstract:**

Cross-selling involves selling complementary products to existing customers. It is one of the highly effective techniques in the marketing industry.

**1.Introduction:**

Cross-selling is the practice of marketing additional products to existing customers, often practiced in the financial services industry. Cross-selling to existing clients is one of the primary methods of generating new revenue for many businesses, including financial advisors. This is perhaps one of the easiest ways to grow their business, as they have already established a relationship with the client and are familiar with their needs and objectives.

**1.1 Need for cross selling**

The main objective behind this method is to increase sales revenue and profit from the already acquired customer base of a company.

Cross-selling is perhaps one of the easiest ways to grow the business as they have already established a relationship with the client. Further, it is more profitable as the cost of acquiring a new customer is comparatively higher. Financial advisors can often earn additional revenue by cross-selling additional products and services to their existing client base. Care needs to be taken to do this correctly in order to stay clear of regulators and protect the client’s best interests. Advisors who simply make referrals in order to receive additional incentives may find themselves on the receiving end of customer complaints and disciplinary action.

**Literature Survey:**

**Health Insurance Cross Sell Prediction:**

**Author: Wagner A Kamakura**

SELLING

The analytical tools that make cross-selling possible in a CRM context can be classified into two main groups: Acquisition Pattern Analysis and Collaborative Filtering. The main purpose in acquisition pattern analysis is to identify the next logical step for the customer, in terms of product/service acquisition, based on the pattern of previous acquisitions and on the pattern of other customers. For example, a business person who acquires a PDA may next acquire a carrying case, followed by additional memory, software, etc. A cable subscriber may subscribe to on-demand programming, followed by broadband Internet access, followed by VoIP phone service, etc. While the first category of cross-selling tools focuses on the sequence of acquisitions, collaborative filtering looks at the patterns of associations among purchases across customers, to identify suggestions of other items that would go along with the purchased one. For example, as soon as a book is added to the shopping cart, Amazon.com suggests other titles purchased by customers who bought that same book. Similarly, Netflix, would look at a customer’s rentals and ratings to suggest movies that were rented and liked by customers with similar rentals and preferences.

**Data Dictionary:**

The dataset which we are going to deal with is related to Cross Selling. Using this dataset,

we are able to predict whether a customer would be interested in Vehicle Insurance is

extremely helpful for the company because it can then accordingly plan its communication

strategy to reach out to those customers and optimize its business model and revenue.

**Variable Categorization:**

There are about 381109 rows and 12 columns in the dataset in which there are about 9 numerical columns and 3 categorical columns.

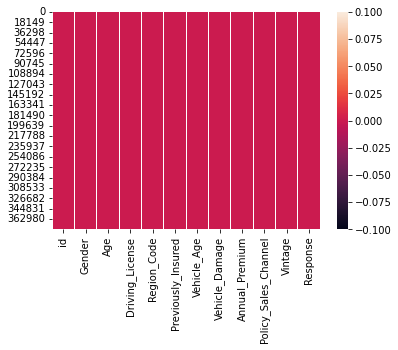
|  |  |
| --- | --- |
| Feature Name | Feature Description |
| id | Unique ID for the customer |
| Gender | Gender of the customer |
| Age | Age of the customer |
| Driving\_License | 0: Customer does not have DL, 1: Customer already has DL |
| Region\_Code | Unique code for the region of the customer |
| Previously\_Insured | 1: Customer already has Vehicle Insurance, 0: Customer doesn't have Vehicle Insurance |
| Vehicle\_Age | Age of the Vehicle |
| Vehicle\_Damage | 1: Customer got his/her vehicle damaged in the past. 0: Customer didn't get his/her vehicle damaged in the past. |
| Annual\_Premium | The amount customer needs to pay as premium in the year |
| PolicySalesChannel | Anonymized Code for the channel of outreaching to the customer i.e. Different Agents, Over Mail, Over Phone, In Person, etc. |
| Vintage | Number of Days, Customer has been associated with the company |
| Response | 1: Customer is interested, 0: Customer is not interested |

**Preprocessing Data Analysis:**

There are no redundant columns in the data.

There are no alternate columns that can supplement the core dataset.

**Null Value Visualization:**



There are no null values in the data.

**Project Justification:**

To optimize customer, reach out process, many insurance worker spend a lot of their time having meeting with prospective client without knowing the probability of that customer to buy the insurance product.

**Project Statement:**

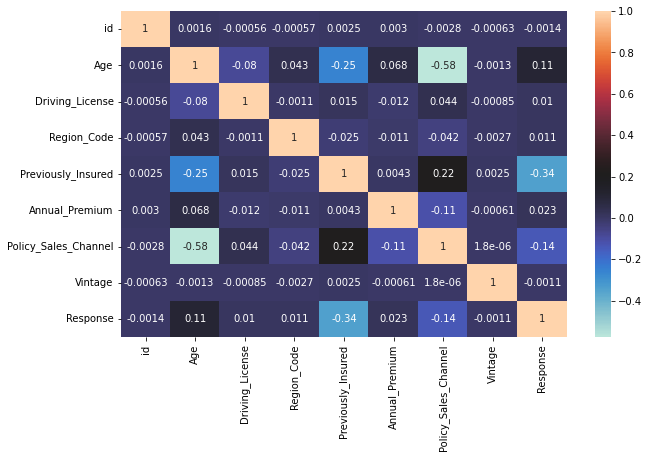
An insurance company that provides Health Insurance to its customers, usually they offer other insurance product to the customers through different kind of marketing channel. In this case we will build a model to predict whether the policyholders (customers) from past year will also be interested in Vehicle Insurance provided by the company.

**Complexity involved:**

There are few columns to predict the target if we have more columns the performance of the model could be improved

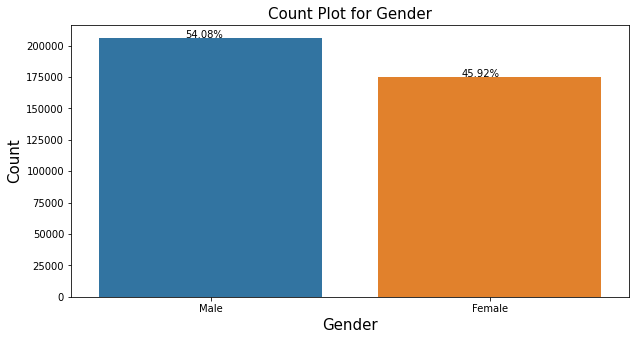
**Project Outcome –Commercial:**

The Business can reduce the time they spend by convincing the non-ideal customers. With the help of the model, they can approach the ideal customers

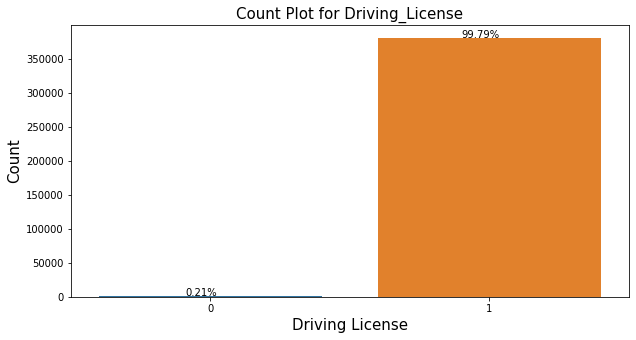
**Exploratory Data Analysis (EDA):**

Among all the columns policy sales channel and age have a high correlation. Previously Insured has some correlation with the response variable.

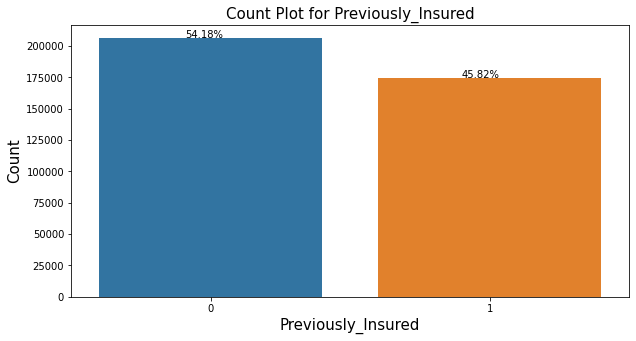
**Univariate Analysis (Gender)**

****

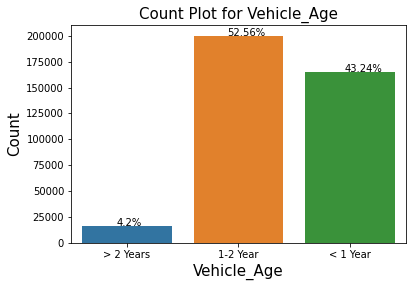
From the above graph we can see that there are more male customers compared to female customers.

**Driving License: **

Majority of Customers have Driving License.

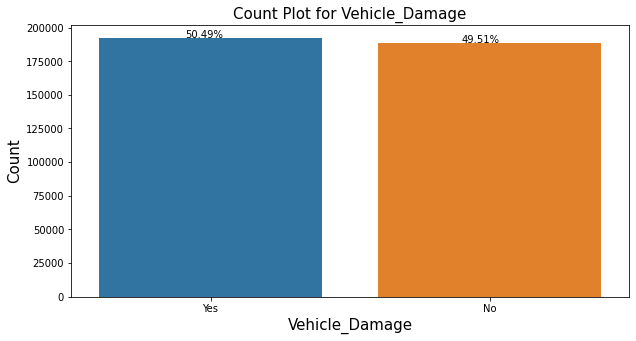
**Previously Insured:**

Around 54.18% of the customers are previously insured

**Vehicle Age:**

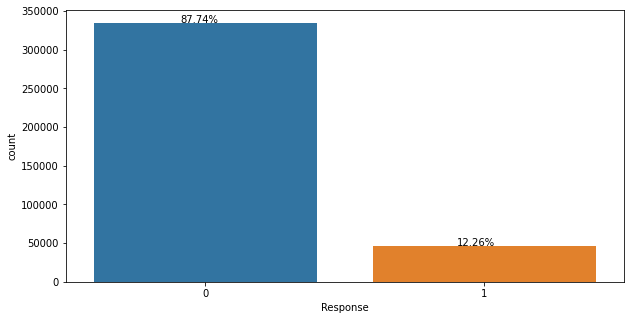
52.56% of the customers vehicle age is between 1-2 year.

**Vehicle Damage:**

****

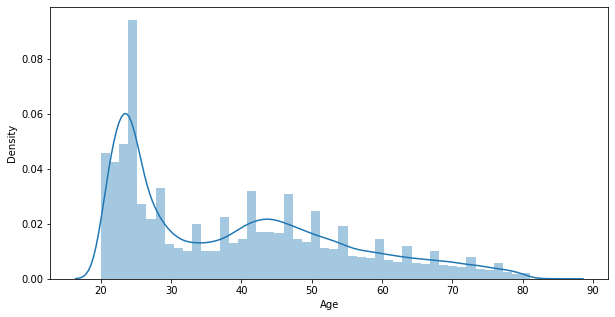
The ratio of customers is equal with and without vehicle damage.

**Response:**



Since Response is our target variable from the above plot we can see that the target variable is severely imbalanced.

**Age:**



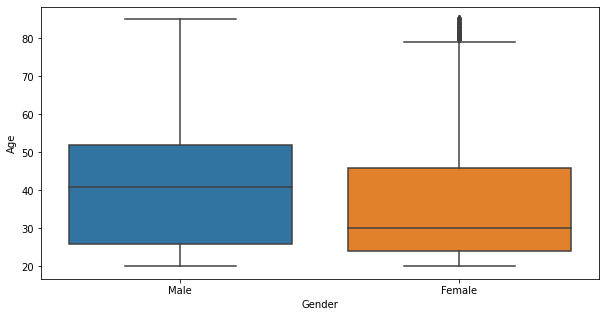
We can see that there are more number of customers in the age range of 20 to 30.

**Annual Premium:**



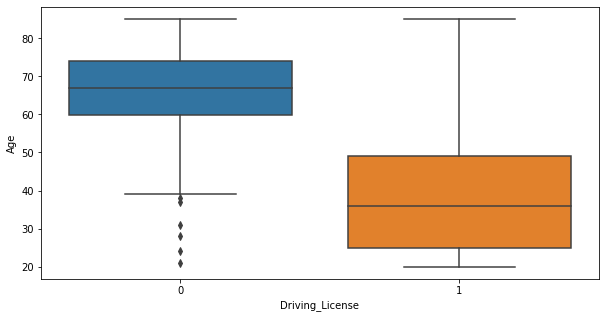
Almost all the values are in the range of 0 to 100000.

**Bivariate Analysis ( Age vs Gender):**

****

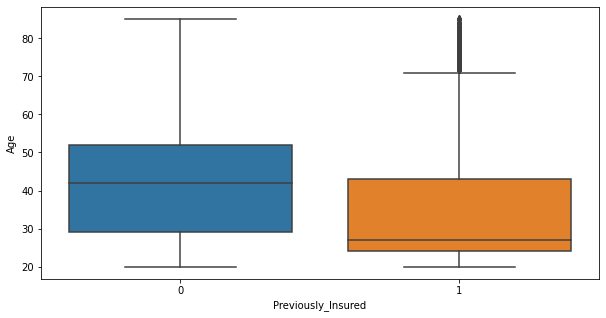
The median age of male customers is around 40 and the median age of female customers is around 30

**Age vs Driving License:**



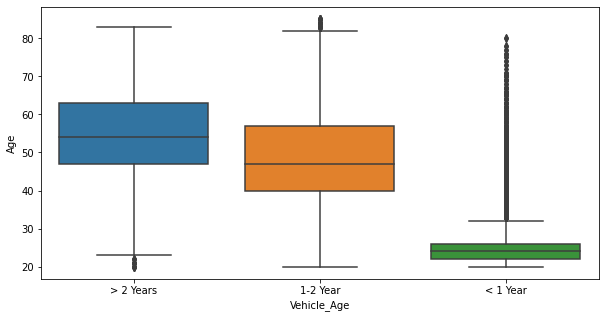
Customers of Age group between 25 to 50 have Driving License.

**Age vs Previously Insured:**

****

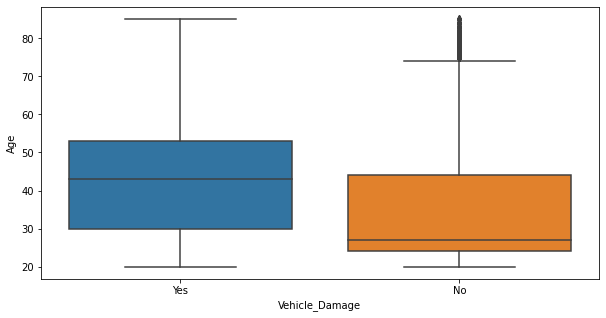
Customers of age group (40-50) have not insured previously.

**Age vs vehicle age:**

****

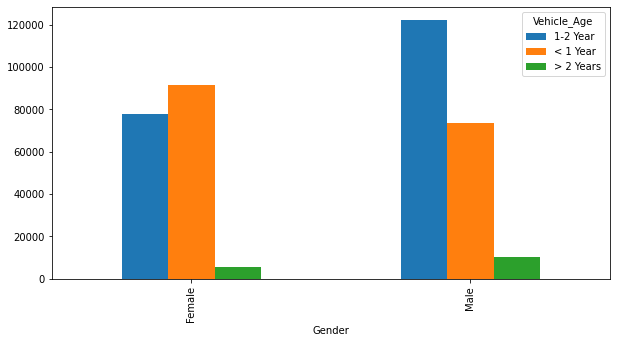
customer vehicle with less than 1 year have least age.

**Age vs Vehicle Damage:**

****

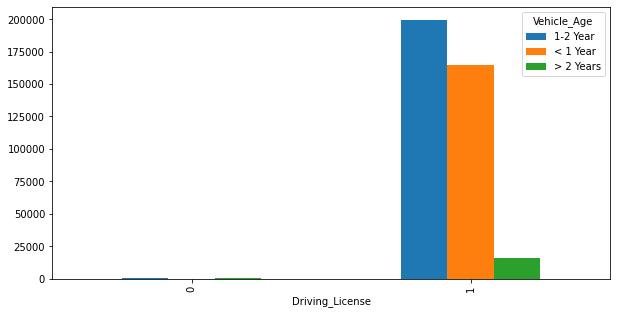
Customers of age group 40-50 have damaged their vehicles previously.

**Gender vs Vehicle Age:**

****

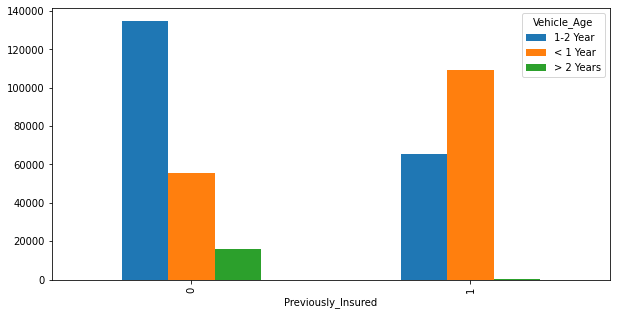
There are more number of female customers whose vehicle age is less than 1 year

**Driving License vs Vehicle Age:**

****

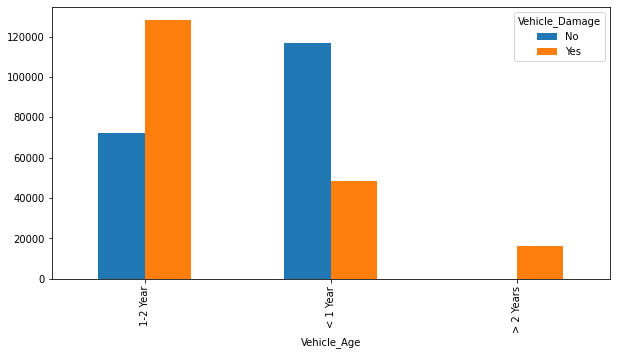
Most of the customers having driving license has a vehicle age between 1-2.

**Vehicle Age vs Previously Insured:**

****

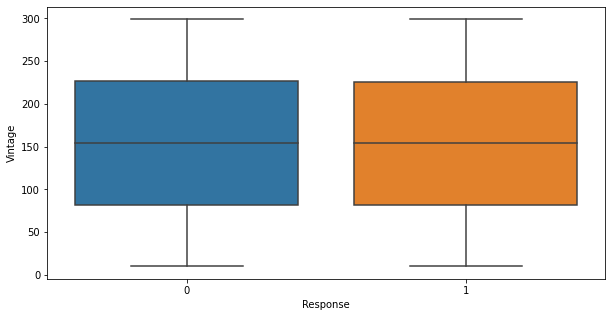
Most of the customers whose vehicle age is between 1-2 are not previously insured

**Vehicle Age vs Vehicle Damage:**



Most of the customers whose vehicle age is less than 1 has no vehicle damage.

**Response Vs Vintage:**



From the above graph we can see that vintage will not be a good feature for our prediction.

**Presence of outliers and its treatment: (Before outlier treatment)**



**After outlier treatment: ( Capping and IQR)**

